Thesis submitted for the degree of Doctor of Philosophy in Materials Science

“Applications of Focal-series Data in Scanning-Transmission Electron Microscopy”

Lewys Jones

June 2013

Abstract

Since its development, the scanning transmission electron microscope has rapidly found uses right across the material sciences. Its use of a finely focussed electron probe rastered across samples offers the microscopist a variety of imaging and spectroscopy signals in parallel. These signals are individually intuitive to interpret, and collectively immensely powerful as a research tool. Unsurprisingly then, much attention is concentrated on the optical quality of the electron probes used. The introduction of multi-pole hardware to correct optical distortions has yielded a step-change in imaging performance; now with spherical and other remnant aberrations greatly reduced, larger probe forming apertures are suddenly available. Probes formed by such apertures exhibit a much improved and routinely sub-Angstrom diffraction-limited resolution, as well as a greatly increased probe current for spectroscopic work.

The superb fineness of the electron beams and enormous magnifications now achievable make the STEM instrument one of the most sensitive scientific instruments developed by man, and this thesis will deal with two core issues that suddenly become important in this new aberration-corrected era. With this new found sensitivity comes the risk of imaging-distortion from outside influences such as acoustic or mechanical vibrations. These can corrupt the data in an unsatisfactory manner and counter the natural interpretability of the technique. Methods to identify and diagnose this distortion will be discussed, and a new technique developed to restore the corrupted data presented. Secondly, the subtleties of probe-shape in the multi-pole corrected STEM are extensively evaluated via simulation, with the contrast-transfer capabilities across defocus explored in detail. From this investigation a new technique of STEM focal-series reconstruction (FSR) is developed to compensate for the small remnant aberrations that still persist – recovering the sample object function free from any optical distortion.

In both cases the methodologies were developed into automated computer codes and example restorations from the two techniques are shown (separately, although in principal the scan-corrected output is compatible with FSR). The performance of these results has been quantified with respect to several factors including; image resolution, signal-noise ratio, sample-drift, low frequency instability, and quantitative image intensity. The techniques developed are offered as practical tools for the microscopist wishing to push the performance of their instrument just that little bit further.

Acknowledgments

This work would not have been possible without the kindness and support of others.

I wish to emphatically thank my supervisor, Professor Pete Nellist, for the research freedoms afforded me, without which we probably never would have stumbled on so many wacky ideas.

I would like to acknowledge Angus Kirkland, Dirk van Dyck and all the participants of the ATOM workshop series. Our discussions were always educational and my ideas certainly benefited from such 'robust' scrutiny during their formative stages.

I would like to thank Peng Wang for many reasons; for his on-column training, the epiphany of seeing images as arrays of values, and for his introduction to Matlab. I would also like to thank Sarah Haigh, for first introducing me to focus as a free parameter albeit in a coherent manner.

I would like to acknowledge my entire research group for their camaraderie, as well as all my family and friends who, long after everyone else had left for the 'real-world' indulged my continued studying.

Lastly, and most strongly, I wish to thank my darling wife Mayra to whom this thesis is dedicated. Not only for moving continents to be with me during these studies, but for being there throughout the whole spectrum of emotions commensurate with a research degree reliant on temperamental hardware. I hope that to the reader my enthusiasm for microscopy is apparent but apologise that this passion will always remain a close second-place.

Lewys Jones MEng

Associated Publications

Journal Articles & Refereed Proceedings

Lewys Jones and Peter D. Nellist. “Identifying and Correcting Scan Noise and Drift in the Scanning Transmission Electron Microscope” *Microscopy and Microanalysis* vol. 18 (2013) doi:10.1017/S1431927613001402.

Lewys Jones and Peter D Nellist. “STEM Image Post-processing for Instability and Aberration Correction for Transfer Function Extension”. *Journal of Physics: Conference Series* vol. 371 (2012) doi:10.1088/1742-6596/371/1/012001

Lewys Jones, Peng Wang and Peter D Nellist. “Three-Dimensional Crystal Structure Mapping by Diffractive Scanning Confocal Electron Microscopy (SCEM)”. *Journal of Physics: Conference Series* vol. 371 (2012) doi:10.1088/1742-6596/371/1/012003

Conference Articles

Lewys Jones and Peter D. Nellist. “Focal Series Reconstruction in Annular Dark-Field STEM”, *Microscopy and Microanalysis* Supplement S2 (2012)

Lewys Jones and Peter D. Nellist. “Post-processing of STEM Data for Instability and Drift Compensation”. *European Microscopy Congress Proceedings* vol. 2 (2012)

Lewys Jones and Peter D Nellist. “A three-dimensional investigation of the STEM-probe/sample interaction by annular dark-field focal series”. *European Microscopy Congress Proceedings* vol. 2 (2012)

Commercialised Software Products

The scan-noise compensation methods described in section 3.2.2 were programmed in the MatLab environment. These have since been developed into a commercially available software package as a plug-in for Digital Micrograph (additional information can be found in Appendix 8.3). Readers interested in knowing more about this should contact innovation@isis.ox.ac.uk quoting reference MCR/7869. Similarly, the focal-series restoration software described in section 5 should be available from late 2014.

Contents

Introduction

The electron microscope is one of the most powerful investigative tools available to the modern material scientists and such instruments can now be found at research institutions all over the globe. These instruments fall into three categories, the transmission electron microscope (TEM), the scanning electron microscope (SEM) and the scanning transmission electron microscope (STEM). It is the last of these three which this thesis will focus on, specifically dark-field STEM.

In a modern STEM equipped with hardware aberration-correctors, a highly focussed sub-Angstrom sized electron probe is rastered over the sample generating a variety of imaging and spectroscopic signals simultaneously. As a result, each and every type of information that can be given by this instrument is dependent on the performance and control of this electron probe.

This work begins with a review of the literature, with special attention paid to the description, diagnosis and correction of aberrations. The effect of various aberrations on STEM probe shapes is explored as well as the lateral and depth resolution attainable with the current generation of hardware aberration correctors.

Next, the main body of this thesis will focus on four main themes as outlined below.

What are the information-transfer capabilities of aberration corrected STEM probes?

In Chapter two, the research explores the information-transfer behaviour of various simulated electron probes in the presence of differing optical distortions (aberrations). These are evaluated using a hybrid 'reciprocal-lateral / real-defocus' coordinate space that intuitively reveals details of their non-round nature along with any variations in optical-transfer with respect to defocus. This work outlines the framework of what information *may* be possible to be retrieved from STEM focal-series. An example is presented that shows how, in the presence of small remnant aberrations, focal-series experiments may be planned so as to capture the whole range of optical-transfer behaviour.

What effects do instrumental instabilities have on recorded STEM data, and can they be counteracted by post-processing?

In Chapter three, the experimental control of electron probes is discussed. Here the sensitivity of the STEM instrument to outside sources of instability is explored, before presenting methods to identify and correct for so called 'scan-noise' and image-drift. To evaluate the performance of the image-processing methodologies developed, automated tools to quantitatively evaluate the resolution, signal-noise ratio, and the amount of image drift are also described here.

What is the information content of STEM focal-series, to what extent do they agree with the two-dimensional object function model, and can they be used for aberration diagnosis?

In Chapter four the information content of focal-series imaging data is explored, via the variation of Fourier-transform-spot moduli with respect to defocus. The aim of the research presented in this section, is to experimentally test the validity of the two-dimensional object function assumption for incoherent STEM imaging, and subsequently, to determine whether from this focal-series data it may be possible to diagnose and correct for remnant illumination aberrations from experiment.

Given an 'over-determined' STEM focal-series data-set can the illumination's effects be separated to leave only an image of the sample free from any imaging distortions?

In chapter five a new method of dark-field STEM focal-series reconstruction is presented which can determine the sample object-function from experimental focal-series data, free from the effects of lens aberrations and certain other experimental distortions. The performance of this new method is extensively and quantitatively evaluated with example reconstructions presented.

1. Review of the Literature

This review will begin by first briefly introducing the STEM instrument, before moving to discuss more detailed topics progressing in the same sequence that the electron path follows. The source of the electron illumination, the electron gun, will be described with attention paid to its finite physical size and the energy-spread of its emission. Next, as the STEM uses a rastered focussed illumination, the key metrics of the electron probe's width and depth sensitivity will be discussed. These performance measures are of key importance to STEM operators as it is these which determine the three-dimensional resolutions achievable. Now with a sharply waisted electron probe at the microscopists disposal, the behaviour of samples consisting of oriented crystalline solids must be discussed; specifically with regard to electron channelling behaviour that allows them to be described as two-dimensional objects. With the illumination and the sample then set-up the various imaging and analytical signals that result will be discussed. These have differing formation mechanisms and the mathematical framework for their information content will be presented.

In the latter half of this review, performance limiting optical distortions known as aberrations will be introduced with attention paid to how these may be observed and diagnosed. A special case of balancing spherical aberration with defocus will be discussed before more general techniques for aberration mitigation are explored. A variety of hardware approaches intended for the correction of these optical distortions will be reviewed with special attention paid to the two hardware systems that have been successfully realised commercially.

To complete this literature review, some image-processing methods intended to compensate for any remaining optical-distortions in post-processing will be discussed.

1.1. Overview of the STEM Instrument

The main components of the STEM instrument are well described by Crewe [1], Figure 1.

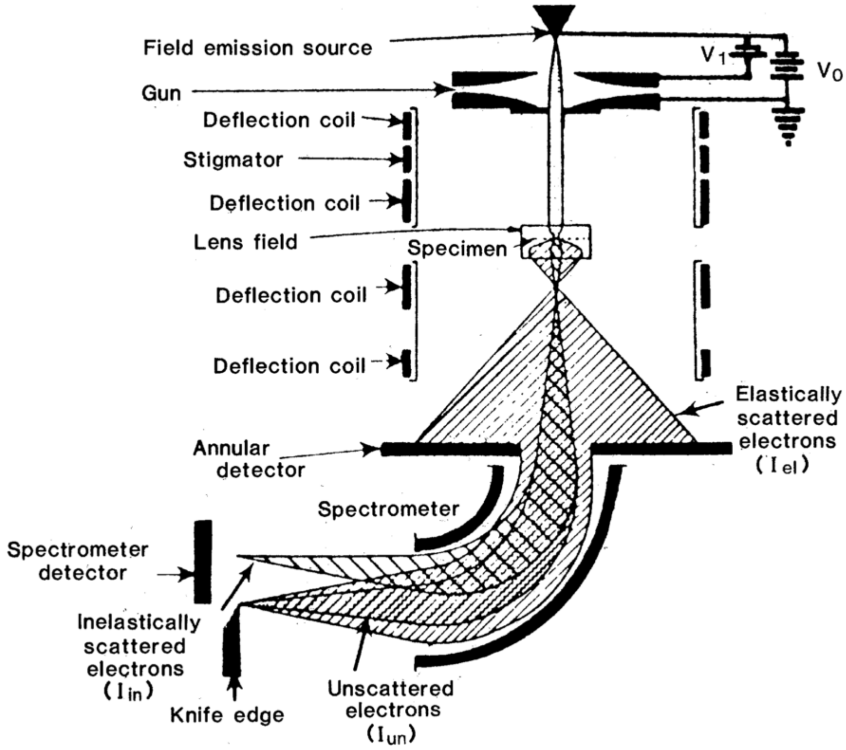


Figure 1 – Main components of a STEM instrument. Image credit: reference [1].

In a STEM an electron source produces illumination which is focussed by pre-specimen lenses to a focussed probe. The waist of this hourglass shaped probe is scanned, or rastered, across the sample surface using pairs of deflector coils. Various beam-specimen interactions occur and can be recorded with a variety of detectors. As the probe is scanned across the sample the signals detected are built up pixel-by-pixel into an image.

Figure 1 shows a 'top down' type arrangement. This type of diagram will be most familiar to TEM users as it has the electron source at the top. However as the electron source is the heaviest assembly in the microscope for dedicated STEM instruments a 'bottom up' design is favoured to aid the overall mechanical stability.

1.2. Finite Sizes & Energy-Spreads of Electron Sources

The electron source is the starting point for the illumination as well as, unavoidably, the point where we must begin to consider the introduction of possible sources of imaging distortions. Figure 2 shows schematically the thermionic-emission gun (Schottky) and field-emission gun (FEG) designs of electron sources [2].

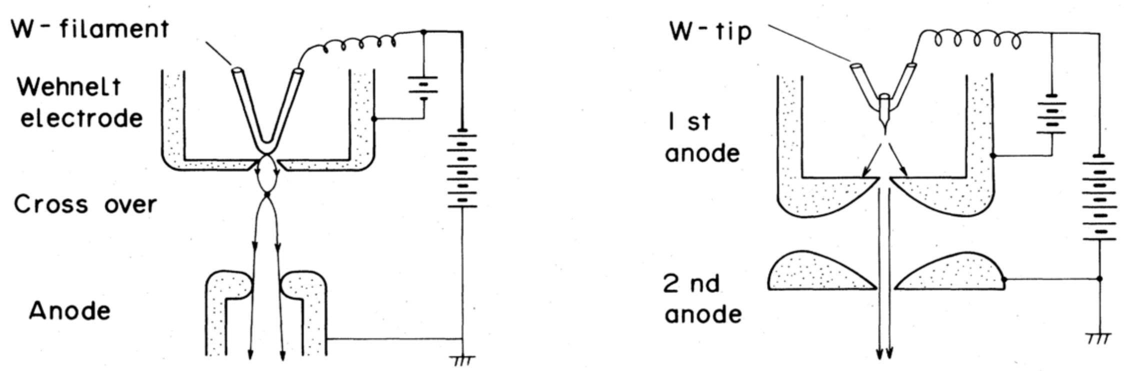


Figure 2 – Schematics of thermionic (left) and field-emission (right) electron guns. Image credit: reference [2].

In both cases two anodes are used to extract and accelerate the electrons. However beyond that the behaviour of the two is vastly different.

In the case of the thermionic gun the electron source, either a tungsten metal 'hairpin' or a LaB_6 crystal, is heated to the point that electrons 'boil off' at their tip (≈2600-3000K and ≈1400-2000K respectively [3]). These are then accelerated and focussed by the successive anodes. For the field emission gun an extremely sharp metal tip is unheated. Using the first anode electrons are instead extracted from this sharp tip by quantum tunnelling (requiring an improved vacuum system design versus thermionic setups [4]). This yields electrons of a far more consistent energy originating from a highly localised point on the tip. These electrons are then accelerated by the second anode. A variant of this is a tungsten tip with a multi-walled carbon nano-tube attached; with the even finer nano-tube acting as the emitting point [5]. These tips have demonstrated some promising brightness results but are not commonly found installed, perhaps because of their limited lifetime. An excellent description of many additional aspects of a range of electron sources is included in references [6] and [7].

An ideal electron source would take the form of an infinitely sharp point. Unfortunately no such source exists and all have a finite size. Many methods exist to measure the width and distribution of source size and several mathematical fittings have been recommended, including; Gaussian [8], [9], Lorentzian [10] and bivariate Gauchy distributions [11].

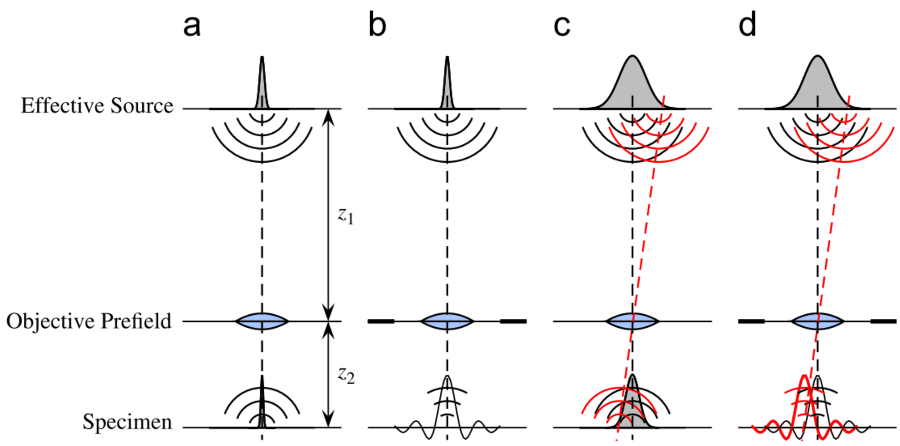


Figure 3 – An ideal probe-forming system (a) where the effective source is perfectly imaged at the specimen plane, is then modified by the finite size of any apertures used (b) as well as the finite size of the electron source (c) resulting in a probe shape that incorporates both effects (d) [10].

Additionally the finite energy spread of all types of electron sources must be considered [12]. This is because the strength of focussing performed by electron lenses is dependent on the energy of the electrons they act upon. Alamir discusses in detail the subtle variation in focussing, magnification and image rotation that can be associated when electron lenses operate on illumination containing such an energy spread [13]. This behaviour is the source of so called 'chromatic aberrations', Figure 4.

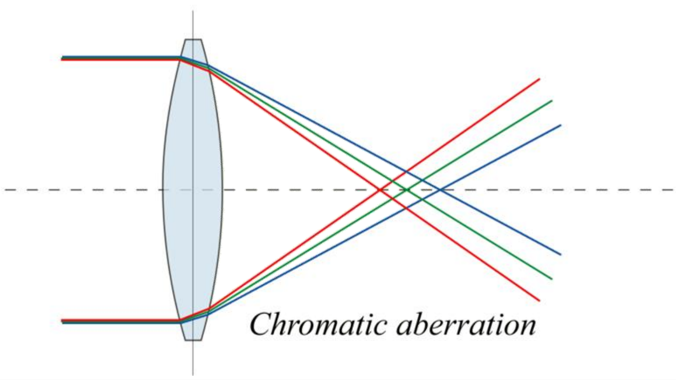


Figure 4 – Ray path diagram through a chromatically aberrated lens.

Here lower energy rays, $E_0 - {}^{\Delta E}/_2$, (shown in red) are focussed to a crossover closer than the average, whereas higher energy rays , $E_0 + {}^{\Delta E}/_2$, (blue) are focussed more weakly to a cross over further from the lens. Wang et al shows a direct measurement of this spread in an electron-optical device [14]. Mathematically this imposes a limit on the lateral resolution given by [15]:

$$d_c = C_c \, . \alpha \left(\frac{\Delta E}{E_0}\right) \tag{1}$$

where: C_c is the coefficient of chromatic aberration,

ΔE the energy spread of the emitted electrons, and

E_o the mean energy of the emitted electrons (given by the operating voltage of the microscope).

Similarly the loss of depth resolution (that is the ability to use focus to discriminate between features at different heights), Δz_{chrom}, is given by [16]:

$$\Delta z_{chrom} \approx \frac{C_c \ \Delta E}{1.76 \ E_o} \tag{2}$$

Equations 1 and 2 also highlight another important side-effect of these chromatic aberrations relevant to the field of low-voltage microscopy. This field is growing very rapidly at the moment with enormous interest in beam sensitive materials. However this loss of depth resolution will be 2.5 times greater for images recorded at 80 kV compared with those at 200 kV [12].

The result of these two effects is that finite source-size acts so as to impose an additional *lateral* blurring of the hourglass shaped STEM probe, whereas chromatic effects introduce a defocus, or *vertical*, blurring only [7]. Finally, an additional defocus blurring can be realised if any fluctuation in the gun's high-voltage supply or in the objective lens current is present [17].

Each of these can be studied by incorporation in 3D probe simulations by consecutive convolution of the otherwise un-blurred probe with an appropriately shaped function. Borisevich calculates and plots the variation of probe intensity with z (defocus) for three typical energy spread conditions of 0.6, 0.3 and 0.1 eV [16], Figure 5.

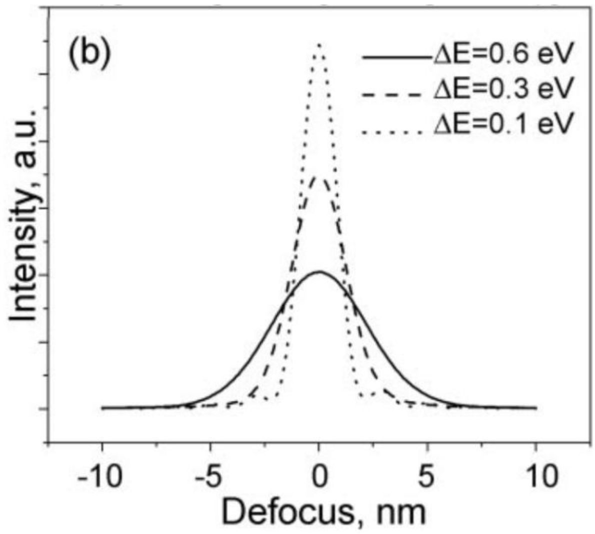


Figure 5 – Defocus intensity variations with varying energy spread electron sources. Image credit: reference [16].

For a 200kV aberration corrected instrument these spreads correspond to the expected emissions from thermionic-emission, cold-field-emission and monochromated guns respectively [18], [19].

The majority of the results presented in this thesis were acquired at 200kV and 300kV and fortunately at this voltage C_c is rarely the most performance limiting effect, instead geometric aberrations (discussed later) dominate. However, if the techniques presented in the subsequent chapters are explored at lower primary-beam voltages revisiting chromatic aberration would be merited.

1.3. Lateral Resolution and Depth Sensitivity in the STEM

The highly focussed nature of STEM probes allows them to be used for studying materials in three-dimensions. For that reason, in this section, the lateral resolution and depth (vertical) sensitivity of the illumination will be discussed.

Lateral Resolution

Rayleigh noted in his seminal work on optics that *"the actual finiteness of λ* (wavelength) *imposes a limit upon the ... resolving power of an optical instrument"* [20]. For a focussed illumination microscope all its illumination would ideally be brought to a single point which is infinitely small both along and perpendicular to its axis. However, Rayleigh shows that for a circular aperture the diameter, d, of the smallest probe that can be formed (narrowest waist in the hourglass) is given by:

$$d = \frac{1.22 \ . \ \lambda \ . \ f}{D} \tag{3}$$

where:	λ	is the wavelength of the illumination,
	f	the focal length of the probe forming lens, and
	D	the diameter of the aperture defining the illumination.

This ratio of focal length to aperture diameter is something familiar to light microscopists and photographers as the numerical aperture, NA, where for small angles [21]:

$$NA = \frac{n \ . \ D}{2 \, f} = n \, . \sin \alpha \tag{4}$$

where:	n	is the refractive index of the material between the lenses, and
	α	the aperture angle of the objective lens.

The relationship in equation 4 shows the expression by Rayleigh to be similar to those below by Abbe. In a transmission electron microscope, for near parallel (more coherent) illumination, the minimum distance between two objects that may be resolved, d, is given by [21], [22]:

$$d = \frac{0.82 \ \lambda}{\sin \alpha} \tag{5}$$

Alternatively for larger illuminating apertures, and hence more incoherent illumination, diffraction is again the limiting factor and d is given by [21], [22]:

$$d = \frac{0.61 \ \lambda}{\sin \alpha} \tag{6}$$

Immediately from these equations we can see two important points for microscopists. Firstly, the resolution limit imposed by the wavelength of the illumination used (the key motivation behind the development of electron optics as a successor to light microscopy) and secondly, the limit to resolution imposed by the aperture angle used. This has driven the development of higher quality electron lenses, which will be discussed later in Section 1.8.

Depth Resolution (Sensitivity)

Depth of field is a familiar concept to photographers and describes the distance range over which a subject will be in focus. When a larger aperture is used objects further from the focus point are less clearly resolved. Adjusting the focus value would bring a different depth into sharp focus and this discrimination is the basis for optical sectioning.

The amount of the field of view which is in focus is called the depth of field. In the above photographic example, this is equivalent to the uncertainty in the distance between the viewer and the object that is in focus. For an incoherent imaging mode this depth of field, Δz, was defined by Born and Wolf as [21]:

$$\Delta z = \frac{\lambda}{\alpha^2} \tag{7}$$

An illustration of the depth sensitivity of a microscope can be obtained by comparing the intensity of a signal recorded as a function of specimen height, or equally defocus, Figure 6 [23], [24].

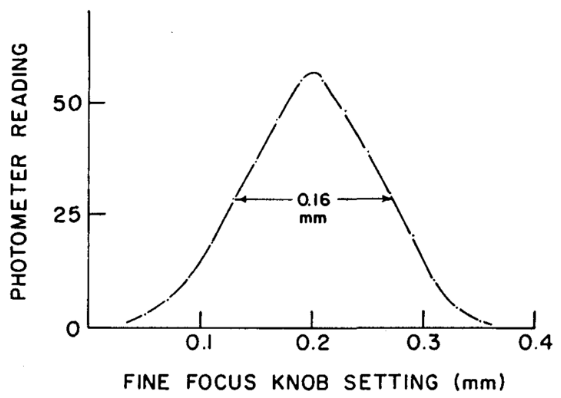


Figure 6 – Koester's *"illumination image"*, reflected light from scanning mirror microscope. Image credit: ref. [23]**.**

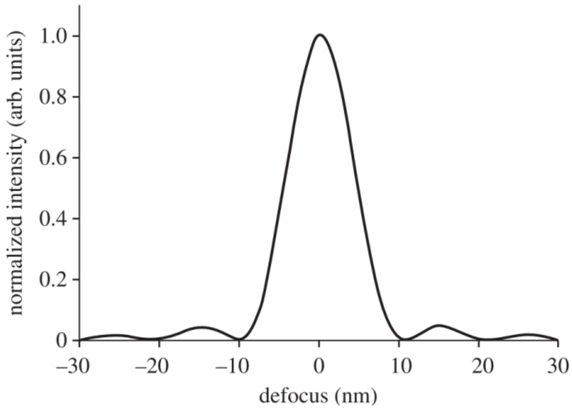


Figure 7 – Example STEM electron probe vertical illumination intensity profile. Image credit: reference [24]**.**

Figure 6 shows an early example of a defocus-intensity plot for an optical microscope. The width shown is 0.16 mm but modern instruments are usually able to reach nearer to hundreds of nanometers [25]. Figure 7 shows a typical defocus-intensity profile of a conventional STEM instrument with a width of the order of 10 nm. Comparing the above figures we see a remarkable similarity of their illumination profiles. However, the absolute value of the widths, suggests a 100 times improvement in vertical resolution using electrons. The intensity of this illumination along the optic axis indicates the achievable depth sensitivity. The full-width at half-maximum (FWHM) of this intensity for probe type illumination, as in the STEM, is given by Nellist et al [26]:

$$\Delta z_{fwhm} = 1.77 \frac{\lambda}{\alpha^2} \tag{8}$$

Again, as for the achievable lateral resolution, both reducing the illumination wavelength and increasing the aperture angle is expected to increase depth sensitivity.

However, one cannot simply enlarge the apertures used. To create an image with interpretable contrast, the rays allowed through an aperture must be within a limited phase discrepancy. Increasing the size of the aperture is expected to improve both the lateral and depth resolution (equations 6 & 8), however this would also allow deviated rays to contribute to the image. The result of this is that for any optical system there exists an aperture size that balances these two phenomena where the best performance is achieved. Reducing aberrations then allows for larger apertures to be used, for probes to be formed with tighter waists, Figure 8.

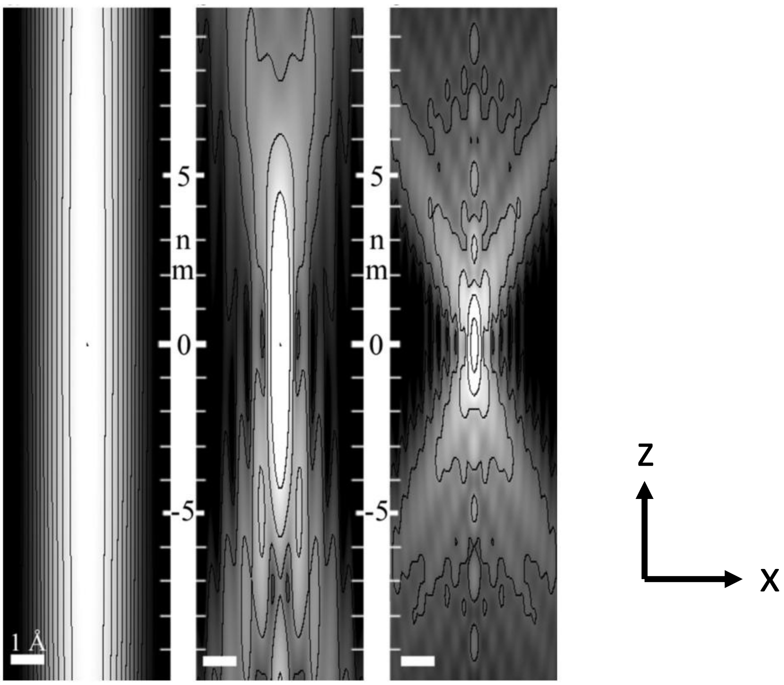


Figure 8 – Simulated probe shapes for left) an uncorrected, centre) a third order corrected, and right) a C3/C5 corrected STEM probe (x-scaling has been stretched by a factor of 10 relative to z-scaling). Image credit: reference [27].

Figure 8 shows some probe shapes simulated by Borisevich et al with increasing degrees of aberration correction, which would yield both higher lateral resolution images and more sensitive depth sectioning [27]. The resolution of electron-optical systems can be severely limited by aberrations to about 50 times the illumination wavelength in the same way that early light-optical microscopists were [28].

1.4. Instrument Stability

In the previous section electron probe formation was discussed, the next technological challenge is controlling such probes to utilise them to their fullest. In the STEM the sub-Angstrom sized probes are carefully rastered by electrostatic deflection (Figure 1). However, it is immediately easy to imagine that any unwanted fluctuating electric fields near the instrument may disrupt the imaging [29]. Furthermore this concept can be extended to include varying magnetic fields, unwanted mechanical or acoustic vibrations, air-flow around the instrument, thermal changes in the column, or even the accuracy to which microscope lens modules are assembled [30], [31], [32]. All of these factors can introduce imaging distortions to the recorded data both separately and cumulatively.

It is perhaps unsurprising then, that every few years, when some other area of the STEM technique advances, such as aberration correction or monochromation, that discussions of stability and image restoration must be revisited [33], [34]. Whilst it is always preferable to wholly eliminate any of these external factors, this is not always practical and in Chapter 3.2 a newly devised method to restore data corrupted by these effects is presented.

1.5. Channelling and the 2D Object Function

For thin TEM/STEM specimens a two-dimensional approximation of the object may be acceptable. However, for thicker samples this is insufficient. Thicker specimens often exhibit an effect known as 'electron channelling' which we will discuss here.

The tilt-dependent interaction of electron beams and specimens is well known to users of scanning electron microscopes (SEM) and gives rise to so-called channelling contrast. For the SEM imaging in backscattered electron (BSE) mode a simple ray-particle explanation is sufficient [35]. However for very thin specimens imaged at high resolution as in STEM this explanation is clearly insufficient. Berry shows mathematically that electron rays entering a specimen surface experience that specimen as a series of potential wells [36]. These rays then repeatedly *"traverse the bottom of the potential well"* as they travel through the crystal, Figure 9.

This electron is then said to have channelled down that atomic column. Van Dyck et al show that for ordered alloys the potential surface that defines this behaviour is equivalent to the projected potential of the whole thickness of the sample and for the HREM focal-series reconstruction (FSR) case leaves the image contrast obtained virtually insensitive to specimen thickness and defocus [38].

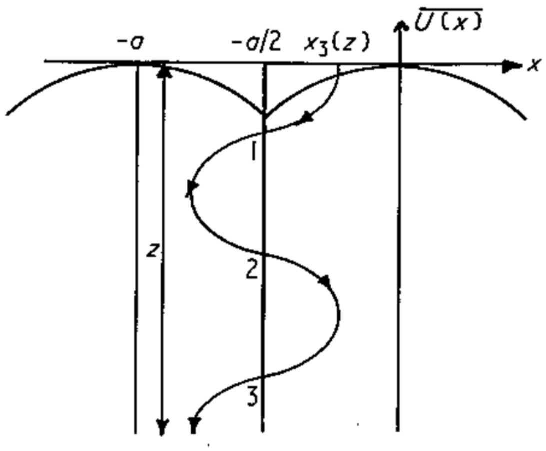


Figure 9 – Electron ray path through a specimen expressed as a series of potential wells. Image credit: reference [36].

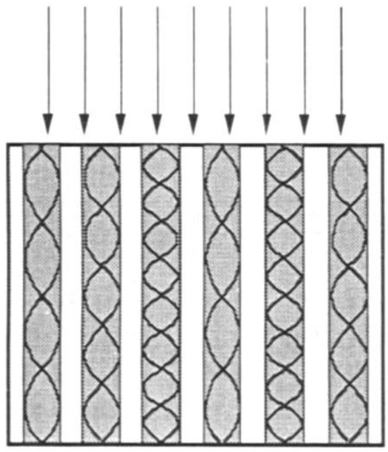


Figure 10 – Schematic representation of the electron channelling along atomic columns. Different mean atomic mass columns yield differing channelling-wavelengths with heavier species being the shorter. Image credit: reference [37]

For STEM this ‘potential projection’ effect is even more important as the focussed probe is of the same order of size as the atom columns themselves [39]. This similarity of length-scale means that there is a very strong coupling effect between the entrance surface of the crystal and the in-focus STEM probe. This coupling then strongly localises the object function (which describes the sample mathematically in the image) to this entrance surface. The effect of this is an object function that appears largely collapsed into two dimensions (collapsed with respect to z / defocus).

Van Dyck & Opdebeeck go on to show that the periodicity of this electron wave-guiding / channelling is dependent on the atomic mass of the atoms of which the columns are comprised [37]. Figure 10 shows a schematic representation of this channelling effect along atomic columns of various relative masses.

This channelling is highly dependent on the angles at which the rays enter the specimen. Multi-slice simulation by Maccagnano-Zacher et al then shows how this channelling effect is responsible for the strong dependence of image contrast on specimen tilt [39]

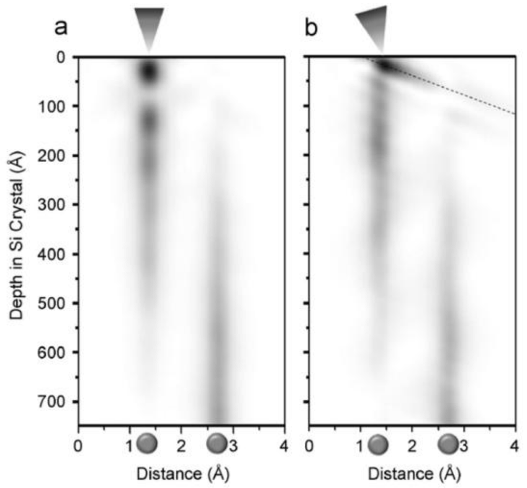


Figure 11 – Dependence of beam propagation on specimen tilt angle. Calculated beam propagation intensities into Si [110] for a) untitled and b) 15 mrad tilted illumination. Image credit: reference [39].

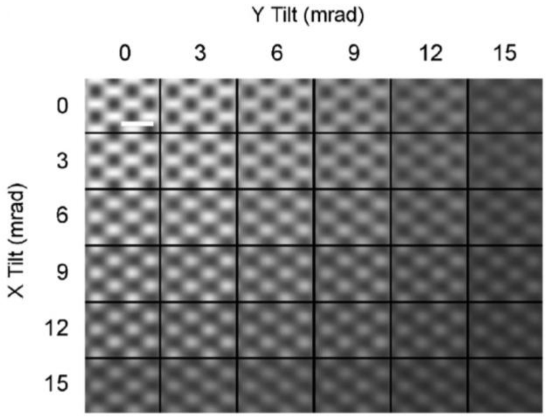


Figure 12 – Simulated ADF STEM images of Si for various small tilt angles about the [110] zone axis. All greyscales normalised to the incident beam, scale bar is 5Å. Image credit: reference [39].

Figure 11 shows the distribution of electron density along adjacent atomic columns for both un-tilted and tilted illumination [39]. This shows the increased de-channelling (intensity transfer off column) resulting from tilted illumination. From this we would then expect that specimens far away from a zone axis orientation will incur a penalty in the resolution of the images obtainable. Maccagnan-Zacher et al present a detailed study of the effects of tilting away from a zone axis on the recorded image intensity and show that being off-axis both reduces the on-column intensity and increases the background intensity. This has a highly deleterious effect on the contrast ratio in the image, which would in the presence of noise in real experiments, limit the observable detail (resolution) in the images. Their work concluded that de-channelling was reduced, and the best contrast achieved, within only one or two degrees of a zone axis.

The result of this electron channelling for well-oriented specimens is very important. This suggests that the localisation of signal (contrast) is highly dependent on the entrant surface of a sample and where probes are focussed on (or just within) the entrant surface of a sample strong contrast is produced.

1.6. Different Interactions, Signals and Detectors in the STEM

One of the advantages of the STEM instrument over a TEM is that many position-sensitive signals can be recorded *simultaneously* with the potential to combine these data, or data from other instruments, to great effect [27], [40]. Each of these arise from a different illumination-specimen interaction and are recorded with a different detector.

1.6.1. Bright field imaging (BF)

Looking again at Figure 1 we can see three modes of electron scattering coming from the specimen. Just as for TEM the unscattered electrons can be used to form a bright field image [6]. By reciprocity this imaging mode is identical to that of bright field imaging in the TEM [41] and has identical transfer functions and contrast behaviours. The image intensity, $I(r)$, can be expressed as [42]:

$$I(\underline{r}) = |P(\underline{r}) \otimes \varphi(\underline{r})|^2 \tag{9}$$

where: $P(r)$ is the complex wave function of the illuminating STEM probe, and

$\varphi(r)$ is the complex transmission function of the sample.

It is then the Fourier transform of the complex illumination $P(r)$ that yields the bright field contrast transfer function, Figure 13 [43].

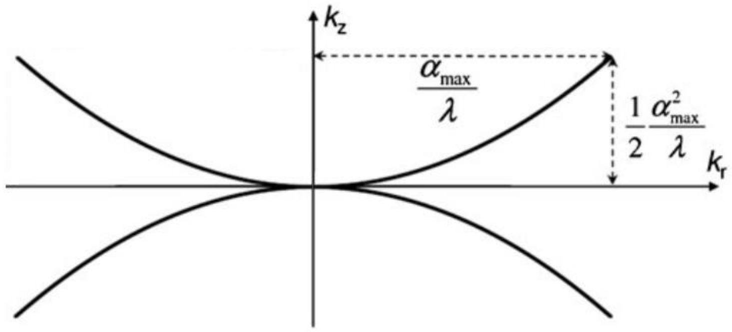


Figure 13 – STEM Bright Field Contrast Transfer Function. Redrawn from reference [43].

The CTF then extends to both approximately the lateral and vertical resolution limits predicted by equations 6 and 8.

1.6.2. Annular dark field (ADF) & Z-number contrast

Again considering Figure 1 we see some electrons are elastically scattered to larger angles. These electrons are then recorded using an annular detector situated around the optic axis (Figure 14) and are used to form a so called annular dark field (ADF) image.

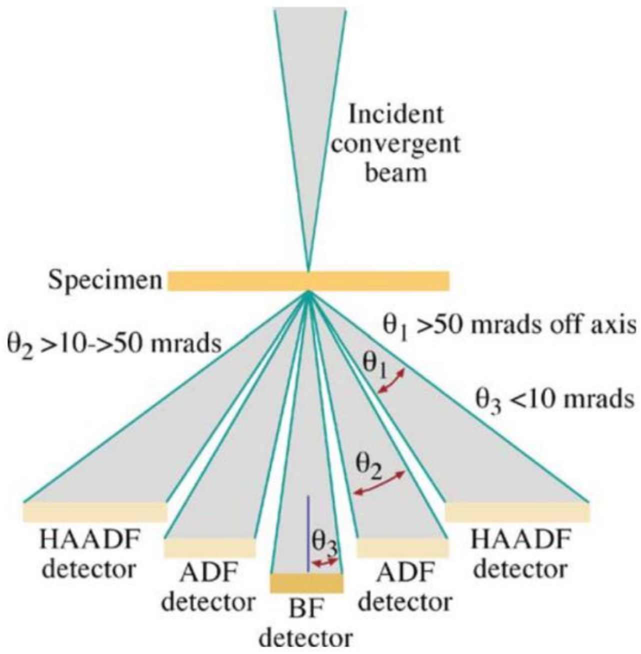


Figure 14 – Example STEM detector collection angles. Image credit: reference [6]**.**

Now in this set-up we must consider both the illumination and the detector properties to understand the expected imaging performance.

The Incoherence of ADF Imaging and the ADF Optical Transfer Function (OTF)

Figure 8 shows some examples of various probe shapes. If each of these probes were used to image a single point in space the resulting image would be a convolution of this point and the probe. This will always be larger, or more spread out, than the point imaged. This is often referred to as the point-spread-function (PSF). The form of this PSF then determines what special frequencies can be transferred by such a probe. This is often visualised with the Fourier transform of the PSF which is called the contrast transfer function (CTF). The CTF however only describes the illumination and in the case of incoherent imaging, detector geometry will affect the recordable detail as well.

Should such an annular detector record signal that has been scattered elastically it will still yield an incoherent signal. The mathematics of this mechanism are well described by Jesson & Pennycook [44].

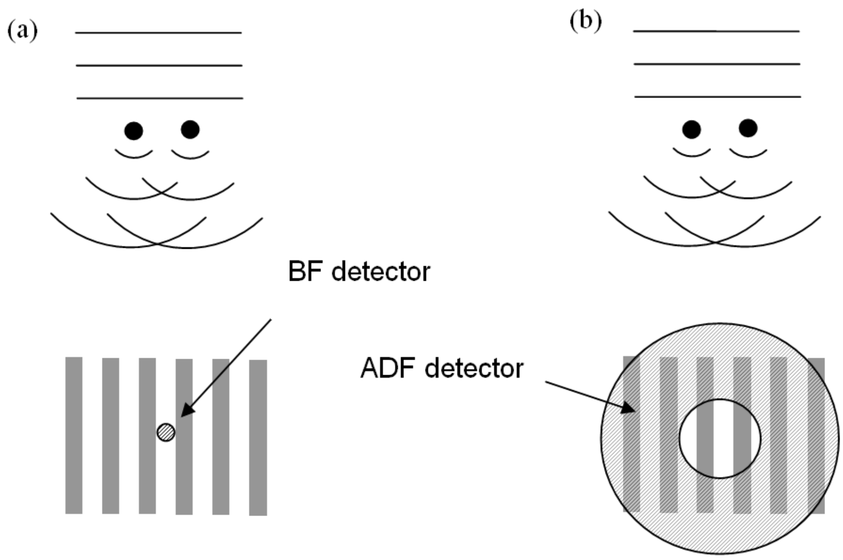


Figure 15 – ADF detector integration of Young's fringes. Image credit: reference [7].

Consider Figure 15 where two adjacent atoms are illuminated by a plane-wave illumination, which is highly coherent in its nature. The scattering from the two atoms will produce interference fringes which appear in the detector plane as parallel lines. The bright field detector (a) retains then its sensitivity to the position of these fringes, and hence phase. The ADF detector (b) however, integrates this signal over a large area. The intensity recorded is entirely insensitive to the position of the fringes and hence phase.

This schematic case can be extended to a more general case described by the overlapping disk model [45]. If we consider the diffraction pattern generated not by plane-wave (TEM like) illumination, but instead by the convergent beam illumination of the STEM probe we observe a convergent-beam electron-diffraction (CBED) pattern.

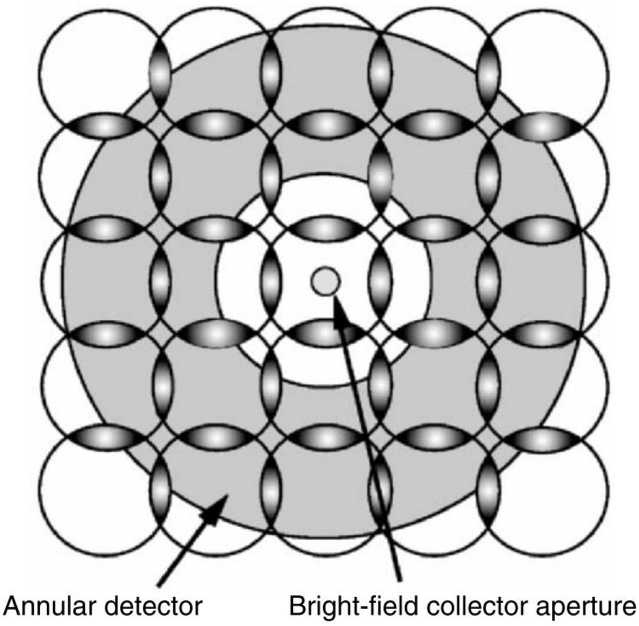


Figure 16 Convergent-beam electron-diffraction pattern and the relative sizes of the STEM bright-field and dark-field detectors. Image credit: reference [45]**.**

At typical accelerating voltages and sample lattice parameters the angle between the central (unscattered) beam and the first diffracted disks is of the order of 10 – 20 mrad [45], however the convergence angle of the illumination is generally slightly larger at 15 – 30 mrad. As a result there is often some small overlapping region (shown in dark-grey in Figure 16). Within these overlapping regions the same interference behaviour as was introduced in Figure 15 will be observed. Again a small bright-field detector will not experience any interference (and no intensity change / amplitude contrast) whereas the large annular detector will collect the summation of the interference fringes – generating amplitude contrast at the expense of coherence.

As an aside, it is also worth noting that the amplitude of the contrast in the overlap regions will be depended on the phase of the coherent wave-functions of the transmitted and diffracted beams [46] – this is in turn dependent on the defocus used and the residual optical distortions present in the probe forming lenses (aberrations). It can be said then that in some way the aberration information is encoded in the recorded images and this theme is revisited in section 4.3.

In either model, the result is an incoherent illumination, where now an optical transfer function (OTF) is used to describe the available frequency information transferred. The image intensity, $I(\boldsymbol{r})$, for ADF is given by:

$$I(\boldsymbol{r}) = |P(\boldsymbol{r})|^2 \otimes O(\boldsymbol{r}) \tag{10}$$

where: $O(\boldsymbol{r})$ is the (real) object function of the sample.

By the theory of reciprocity, to TEM users this would be analogous to high-angle hollow-cone illumination[47]. Unlike the STEM bright-field case it is then the Fourier transform of $|P(\boldsymbol{r})|^2$ that yields the ADF optical transfer function (OTF) Figure 17 [43].

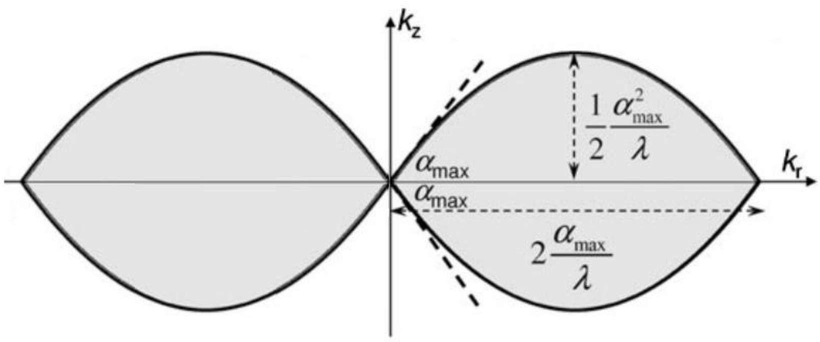


Figure 17 – STEM Annular Dark Field Optical Transfer Function. Redrawn from reference [43].

This OTF is also the convolution of the previous wide-field illumination CTF with itself, indicating the importance of both the illumination *and* the detector geometry in creating this OTF. Nellist shows an excellent graphical illustration of this convolution as well as supporting mathematics [42]. The vertical resolution remains unchanged, whilst the lateral resolution here is doubled as expected from such a convolution.

The use then of probe type illumination and large detector angles together circumvents difficulties encountered in the high-resolution TEM phase contrast imaging mode when lattice periodicity is broken at, for example, crystal defects or nano-particle edges [48]. One illustration of this simplified contrast is to compare the PCTF of coherent imaging (such as BF STEM) and the incoherent OTF, Figure 18 [7], [49].

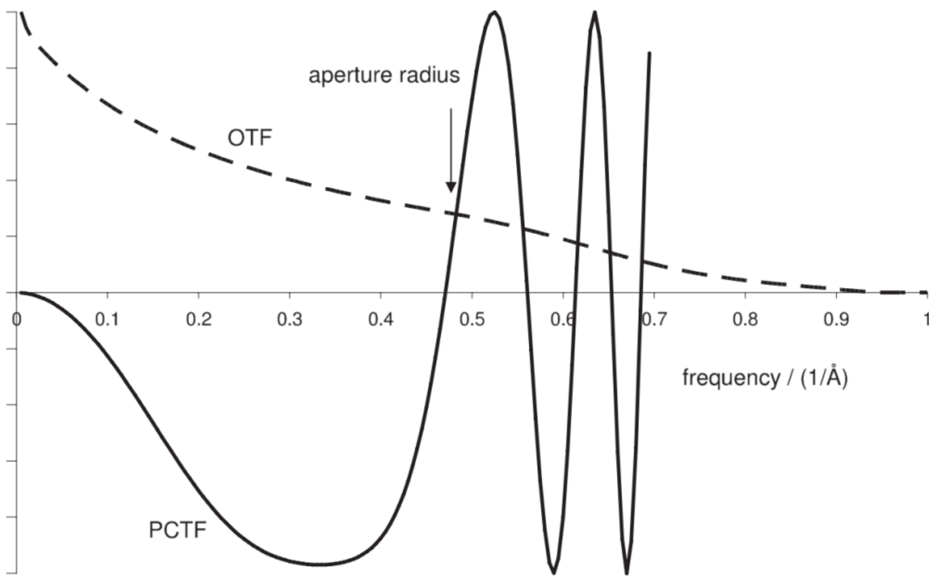


Figure 18 – Comparison of BF STEM PCTF and ADF OTF. Image credit: reference [7].

Here several features can be identified. Firstly the incoherent OTF is above the horizontal axis across its whole range, giving in white-atom contrast and no contrast reversals. Secondly the region over which this contrast is monotonic is twice as wide for the incoherent case. Finally, this simplified OTF, with its lack of contrast reversals, also has the advantage of facilitating more simple probe deconvolution methods.

Choice of ADF detector angles

The choice of inner and outer angles in these annular detectors is important and can be selected by installing different detectors or changing the camera length to optically couple the physically fixed detector to different angular regions. Figure 14 shows an expanded view of some example collection angles for various STEM detectors in cross section including the most commonly used, the high-angle annular dark-field (HAADF) detector [6].

The dark field detector angles are normally selected to exclude any Bragg and high-order-Laue zone scattering [50]. Kotaka shows an example of this where the elastic and inelastic interactions for the atoms in $SrTiO_3$ were calculated [51]. Such simulation then allows for the appropriate detector angles to be chosen and signals to be recorded.

ADF Z-contrast

A significant benefit of using the incoherent imaging afforded by the ADF detector is that it filters the complicated beam-specimen interactions, largely only recording signal from the highly localised 1s electrons about each atom [47], [52]. In addition this scattering interaction is described by the Rutherford scattering equation which gives atomic number, or Z, contrast [48]:

$$\frac{d\sigma(\theta)}{d\Omega} = \frac{e^4 \ Z^2}{16\,(E_0)^2 \ . \ sin^4\left(\frac{\theta}{2}\right)} \tag{11}$$

where:

$\frac{d\sigma(\theta)}{d\Omega}$	is the differential scattering cross-section as a function of the scattering angle θ,
e	the charge on an electron,
E_0	the energy of the incident beam, and importantly
Z	the atomic number of the scattering nucleus.

From this we would expect the intensity recorded in ADF images to follow a Z squared relationship but experimental data yields a relationship nearer $Z^{1.7}$ at the detector inner-angles commonly used [53], [54]. Irrespective of the precise relationship, in many specimens the atomic masses of the two species of interest are often sufficiently separated that they are easily discernable. For the case of low atomic number materials with a constant thickness this Z contrast can even be used to perform atom-by-atom chemical analysis by simply summing the intensity of each atom in the ADF image, Figure 19 [53].

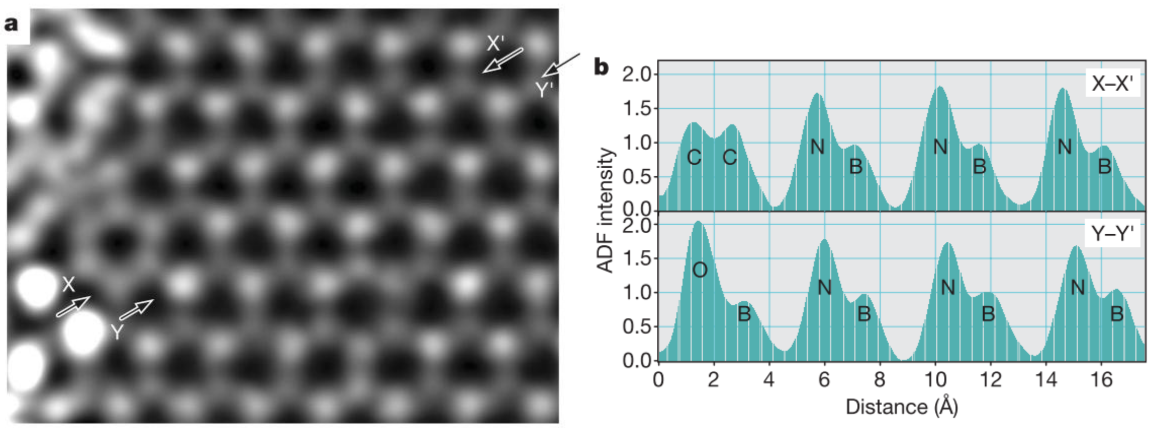


Figure 19 – a) STEM ADF image of a boron-nitride monolayer (smoothed), and b) atomic species identified from line intensity profiles. Image credit: reference [53].

As a result of the localised, Z-contrast, nature of the 1s states contributing to the signal the images that are recorded are often more easy to interpret than their HRTEM equivalents with a monotonic, if not entirely linear, relationship with thickness and no confusing contrast reversals [47], [48], [55].

Electron energy loss spectroscopy (EELS)

A third group of electrons are those scattered inelastically. These have lost only a small amount of energy and remain on the optic axis. To separate these from the unscattered electrons an energy spreading device, beam bending magnet, called a spectrometer is used. These can be either 'in column' such as the Ω-filter design or more commonly 'post-column' as in Figure 1 [56]. Once dispersed from the so called 'zero-loss' electrons into a spectrum characteristic energy losses can be identified corresponding to the atomic species present in the specimen [57]. Early applications of this were reliant on long acquisition times to get enough signal to noise ratio and were generally limited to point spectra [58]. However with the advent of aberration corrected microscopes an increase in probe current of about a factor of ten has been made available. This coupled with new and more efficient 'parallel-EELS' detectors has lead to a 100 fold increase in the available signal [59], [60]. Early 'atomic resolution' measurements by Browning et al should actually have been called 'lattice resolution' with point spectra being collected normal to a crystal interface [61]. Later work by Varela et al show that this technique is sensitive enough to detect individual substitutional atoms [62]. More recent work by Muller show genuinely atomic resolution making use of this increased signal to record spectra at different probe positions and then process these data to map the locations of these characteristic signals [58]. An example of this technique is shown in Figure 20 [60].

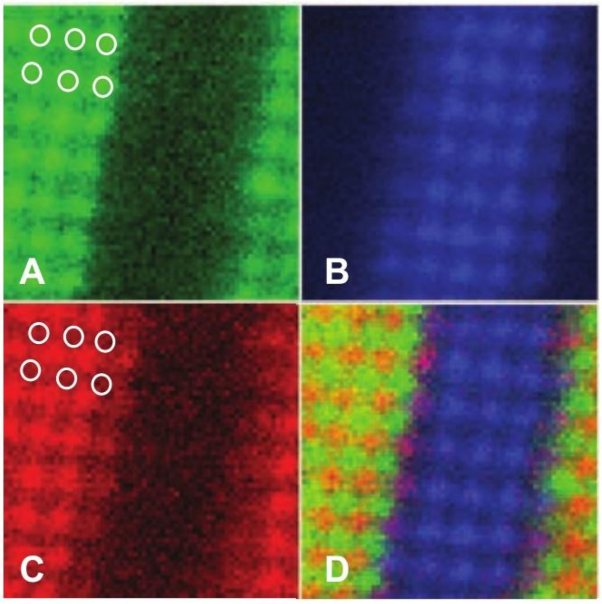


Figure 20 – 64x64 pixel EELS spectrum image of $La_{0.7}Sr_{0.3}MnO_3$. a) Ma M edge, b)Ti L edge, c) Mn L edge and d) the resulting red-green-blue false colour spectrum image. Image credit: reference [60].

Energy dispersive x-ray spectroscopy (EDX)

Energy dispersive X ray spectroscopy (EDX) utilises the detection of characteristic X rays produced following excitation by the electron beam to identify the elements present in a sample.

Various calibration methods exist for EDX [63],[64]. The multi-element reference sample method is most reliable and is preferred but is not always practical for certain materials. Fully standard-less methods are often quicker and more convenient but care should always be taken here to ensure the reliability of the data obtained. Pure element standards methods offer a compromise and are widely used.

Performance of EELS versus EDX

EELS and EDX are both complimentary and competitor micro-analysis techniques. EELS is generally found to perform slightly better for light elements (Z < 17) than EDX and to have a higher resolution for spectrum imaging [65]. However as beam currents are gradually increased through aberration correction and various companies continue to develop new higher performance detectors, for example the Gatan Quantum EELS or Tecnai Osiris systems, we can expect to see continued progress for both these techniques.

1.7. Lens Aberrations: Origins, Effects, Diagnosis and Impacts

Across all fields of microscopy, improved resolution is continually and vigorously pursued. Under nearly all conditions aberrations have a detrimental effect on the images recorded and an understanding of their nature is essential to any microscopist.

1.7.1. The Origins of Aberrations in Electron Lenses

In the collected scientific papers of Rayleigh, the reflection or refraction of rays of light from surfaces with *"any irregularities"* was identified to give rise to *"a retardation* (or advancement) *in the wave-front"* [20]. These are aberrations. Rayleigh also notes the importance of recognizing this *"small departure of the wave-surface from its ideal plane or spherical form"* and the potential for limiting its detrimental effects by using smaller apertures [20]. A similar description is offered in the works of Born & Wolf noting that any *"lack of homocentricity of the image-forming pencil* (rays) *is accompanied by departures of the associated wave-fronts from the* (ideal) *spherical form."* [21]. This is equally true for electron optics.

For glass lenses materials can be selected that reduce or counteract chromatic aberrations and the glass can be ground in an unrestricted manner to a shape which does not introduce any geometrical aberrations. Electron lenses are however much more restricted in their design, necessarily only containing smoothly varying magnetic fields.

Scherzer showed that round magnetic lenses will always contain aberrations (owing to the impossibility of magnetic monopoles) [22], [66], [67] He also supposes methods of correcting for these inevitable aberrations with non-round (multi-pole) lenses or lenses with on-axis charge such as mirrors or foils. All these methods have been attempted and will be discussed here. Scherzer also suggests the possibility of using time-varying fields but this has so far received little technological attention, possibly due to the time needed for lens strength changes to settle.

The development of the modern *"iron-clad...polschuhlinse* (pole-piece lens)*"* is described by Ruska with two quintessential components, the wire coils and the-piece, yellow and blue in Figure 21 respectively [68].

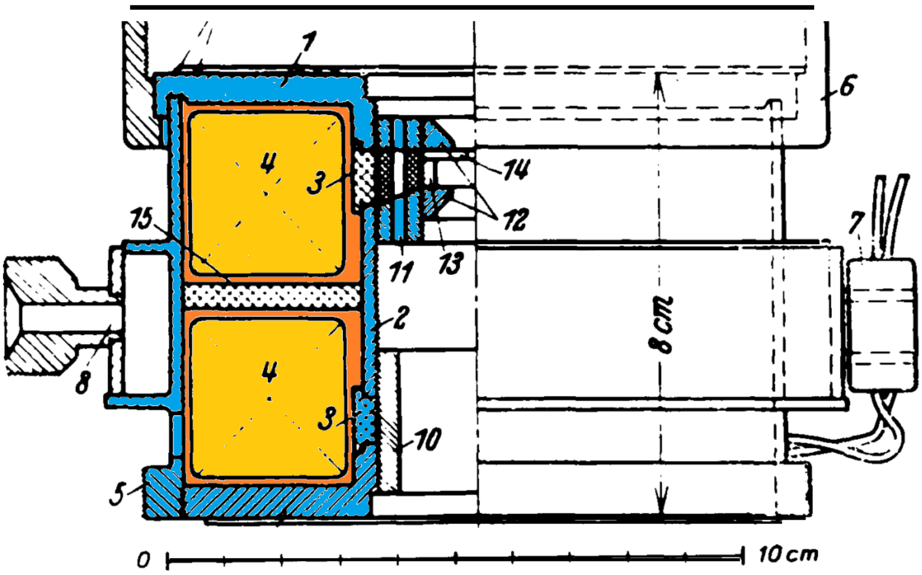


Figure 21 – Pole-piece Lens Schematic. Modified from reference [68].

The wound coils form the magnetic field which is then concentrated by the iron cores; the actual lensing region is then where the magnetic circuit is completed within the so called 'pole-piece gap' [2].

The manufacturing tolerances of these lenses is very tight but advances in computer machining mean that sub-micron milling accuracies are achievable. The overall performance of these lenses is more usually limited by Gauss's law and the shapes of the magnetic field that can be formed within such a pole-piece. Geometrical aberrations can be either round, in the case of spherical aberration or defocus, or non-round in the case of the various astigmatisms or star aberrations.

Aberration surface plots are used to show the difference between the ideal and aberrated wave-fronts. All electron optical elements will introduce some form of aberration and have an associated surface plot. Born & Wolf give examples of the forms of the most commonly occurring aberrations [21].

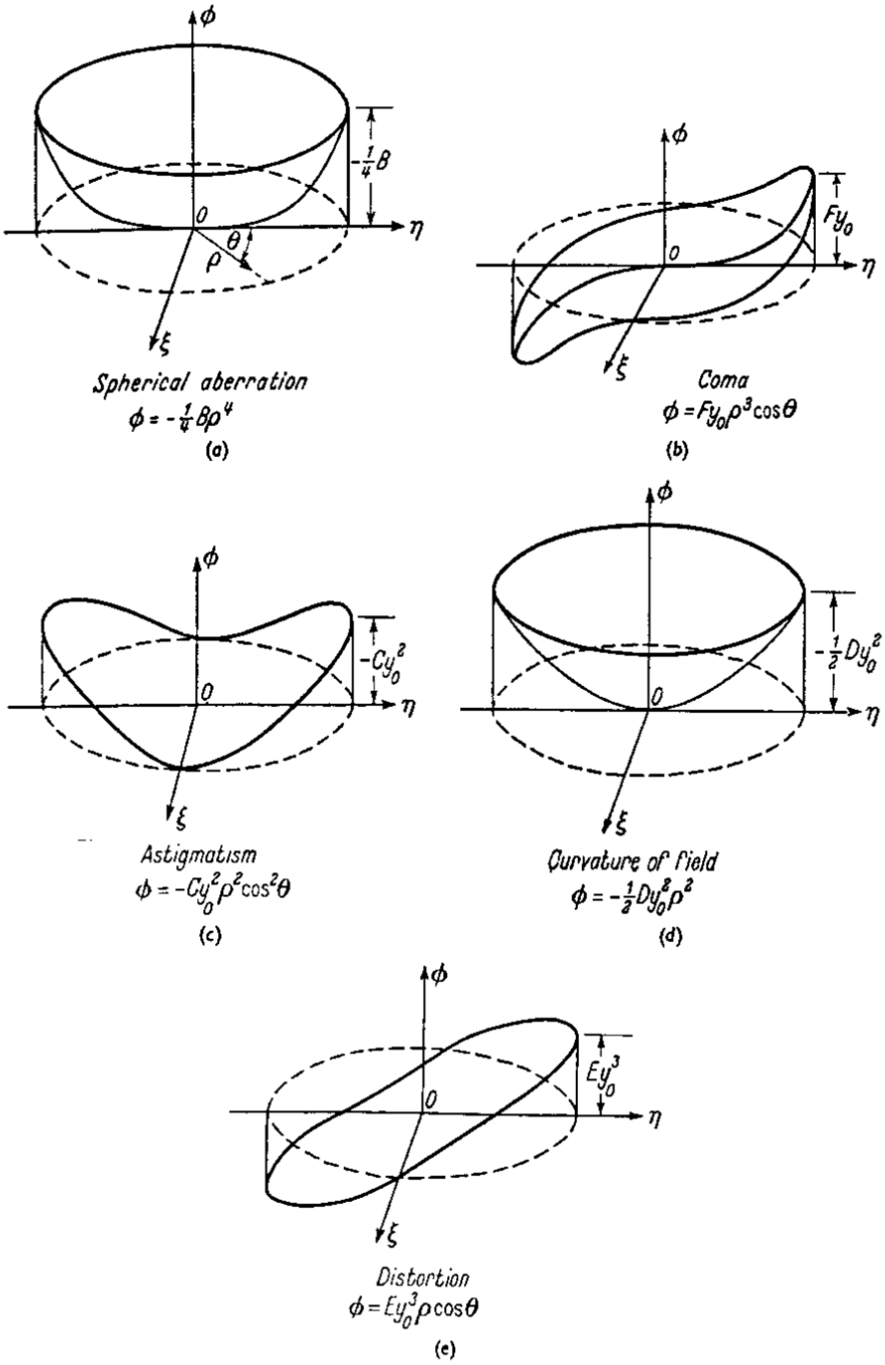


Figure 22 – Aberration surface diagrams and names of the five most commonly observed aberrations. Image credit: reference [21].

Figure 22 a) represents the shape of the aberration surface for spherically aberrated lenses [12] This is usually the most limiting aberration in un-corrected systems and receives much attention in the literature. Figure 23 shows the equivalent ray diagram.

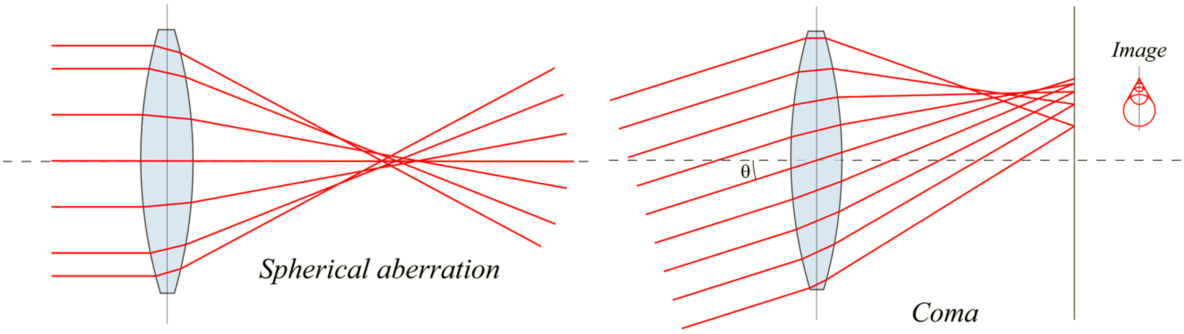


Figure 23 – Ray path diagram through a spherically aberrated lens and one with coma.

Such spherical aberration limits the achievable resolution, d_s, with this effect worsening rapidly with the inclusion of rays from higher angles, α, such that [15]:

$$d_s = \frac{1}{2} C_s \alpha^3 \tag{12}$$

where: C_s is the coefficient of spherical aberration.

The most common compensation technique for this is to add some (negative) defocus, Figure 22 d), and produce a roughly flat region near to the optic axis. Nellist shows a phase plot of such compensation in reference [7]. The resultant surface is then cropped using an aperture to exclude any part of the surface more than a quarter wavelength either side of ideal.

Figure 22 b) shows the effect of coma, usually the result of the electron beam travelling through the optical system at a slight angle just off the optic axis. Compare the teardrop in the image plane in Figure 23 with the in focus illustration for $C_{2,1}$ in Figure 24.

The aberration here is not necessarily the fault of the lens, more in the alignment of the optic axis through it. In electron optical systems this is usually remedied using pairs of beam deflectors to ensure that the optic axis of the illumination passes as close to the so called 'coma-free axis' of the lenses.

Figure 22 c) shows the effect of astigmatism in the lens system. This is a result of the non-roundness of the lenses and is corrected technologically with additional stigmator coils which can be adjusted by the operator. Similarly Figure 22 e) shows image shift.

1.7.2. Describing aberrations: nomenclature and notations

Various notations have evolved in each of the fields where aberrations are of importance. A thorough comparison of the notation schemes used is shown by Sawada et al [69]. Here however, only the main notations used by HRTEM and STEM microscopists will be summarised, Table 1.

Table 1. Commonly used HRTEM (top) and STEM (bottom) aberration notations. Tables modified from reference [70].

		Order n						
Name	Multiplicity	1	2	3	4	5	6	7
Defocus	$\nu = 0$	C_1						
nth order spherical aberration	$\nu = 0$			C_3		C_5		C_7
nth order axial coma	$\nu = 1$		B_2		B_4		B_6	
nth order star aberration	$\nu = 2$			S_3		S_5		S_7
nth order three-lobe aberration	$\nu = 3$				D_4		D_6	
nth order rosette aberration	$\nu = 4$					R_5		R_7
nth order pentacle aberration	$\nu = 5$						F_6	
nth order chaplet aberration	$\nu = 6$							G_7
νfold axial astigmatism	$\nu = n + 1$	A_1	A_2	A_3	A_4	A_5	A_6	A_7

		Order n						
Name	Multiplicity	1	2	3	4	5	6	7
Defocus	$\nu = 0$	$C_{1,0}$						
nth order spherical aberration	$\nu = 0$			$C_{3,0}$		$C_{5,0}$		$C_{7,0}$
nth order axial coma	$\nu = 1$		$C_{2,1}$		$C_{4,1}$		$C_{6,1}$	
nth order star aberration	$\nu = 2$			$C_{3,2}$		$C_{5,2}$		$C_{7,2}$
nth order three-lobe aberration	$\nu = 3$				$C_{4,3}$		$C_{6,3}$	
nth order rosette aberration	$\nu = 4$					$C_{5,4}$		$C_{7,4}$
nth order pentacle aberration	$\nu = 5$						$C_{6,5}$	
nth order chaplet aberration	$\nu = 6$							$C_{7,6}$
νfold axial astigmatism	$\nu = n + 1$	$C_{1,2}$	$C_{2,3}$	$C_{3,4}$	$C_{4,5}$	$C_{5,6}$	$C_{6,7}$	$C_{7,8}$

[a]All four real-valued coefficients with multiplicity $\nu = 0$ and 19 complex-valued coefficients with multiplicity $\nu = 1,\ldots,8$ are listed.

The so called Krivanek notation shown in the lower half of Table 1 shows the aberration coefficients with subscripts denoting, first, the order with respect to the radial direction and second, the rotational symmetry of that aberration. This is the notation that will be adopted in this work but readers may wish to refer back to this chart for the CEOS notation.

Effect on probe shapes

The usefulness of the subscripts used in the Krivanek notation can be immediately seen when the form of the aberrated probe is examined. Uno et al show simulated images of probes, which contain only a single aberration, Figure 24 [71]

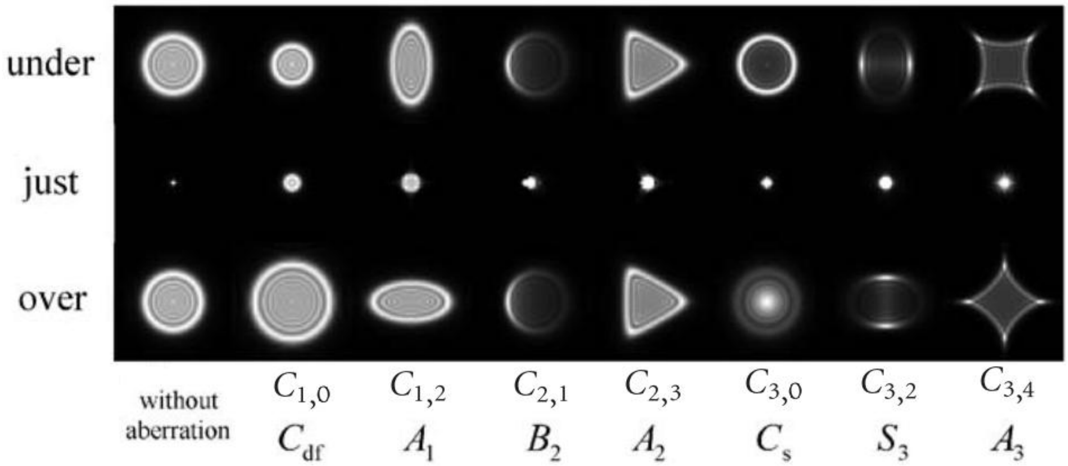


Figure 24 – Effects aberrations on STEM probe shape (of up to 3rd order) on under-focussed, over-focussed and just-focussed probes. Image modified from reference [71].

The rotational symmetry, given by the second subscript, is clearly visible. The radial order, given by the first subscript, is useful as the lower orders must be dealt with primarily before the effects of the higher ordered aberrations become visible. Note also that in Figure 24 whilst the obvious shape changes of the over and under focus conditions are clear, the more subtle changes in the optimum focus row must not be ignored. In all cases the focussed probe was rendered larger by the inclusion of the indicated aberrations. This will affect both imaging but also analytical work such as EELS or EDX (see chapters 5.5 and 5.7 for details).

1.7.3. Measuring Aberrations

The number of different ways that images are affected by aberrations is reflected in the number of ways that have been devised to observe, record and quantify this behaviour. This quantification is an essential first step in the operation of hardware aberration correctors and a selection of the most common methods will be briefly discussed here.

Thon Rings

The Fourier transform of an HRTEM phase-contrast image of an amorphous specimen will produce a ring pattern closely related to the contrast transfer function of the imaging system; these are known as Thon rings [72]. A direct analysis of the shape of the Thon rings enables the two and three fold astigmatism to be determined (see Figure 28) [73], [74]. More recently various computer programmes have been developed to perform this analysis but these are again limited to low order astigmatisms [75], [76].

Ronchigram Analysis

Lin & Cowley (1986) introduce the concept of the electron Ronchigram, an image formed using the mutual interference of many diffracted beams [69], [77], [78]. The image formed at the screen is a so called 'shadow-image' of the specimen. Comparison of such images with simulated ones then allows for the spherical aberration to be determined. For amorphous specimens the beam is brought to focus just before the specimen [7]. Under these conditions the effective magnification varies over the angular range of the illumination. When the aberrations present make the source appear coincident with the specimen a 'ring of infinite magnification' can be seen in the image; equally when aberrations present are minimal a large region of constant magnification is visible, Figure 25 [69].

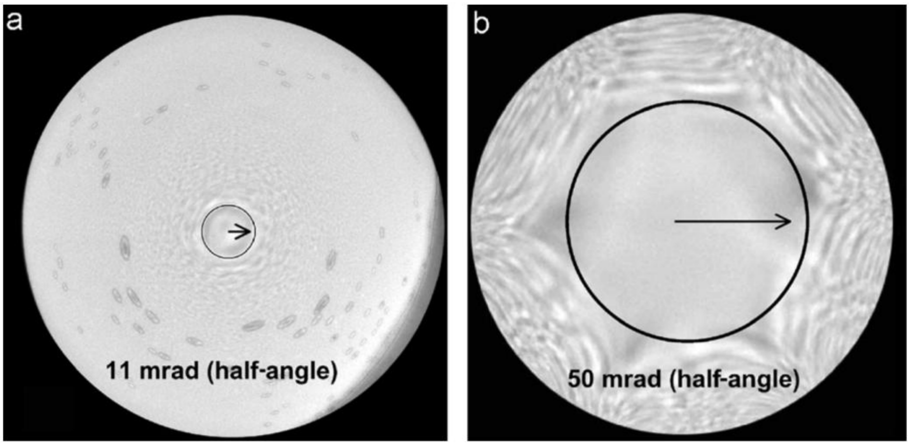


Figure 25 – Ronchigrams from an amorphous foil, uncorrected (left) and aberration corrected (right). Image credit: reference [69].

This method also has the advantage that the region of acceptable phase deviation is immediately apparent in the centre of the Ronchigram. This is the so called 'sweet-spot' and is the optimum size for the illumination aperture [79].

Ronchigram auto-correlation

Alternative methods of calculating aberration constants from Ronchigrams have been suggested. The auto-correlation method uses one image taken at an over-focus and under-focus condition [69]'[78]. These images are then divided into smaller patches, for which each a diffractogram is calculated, Figure 26.

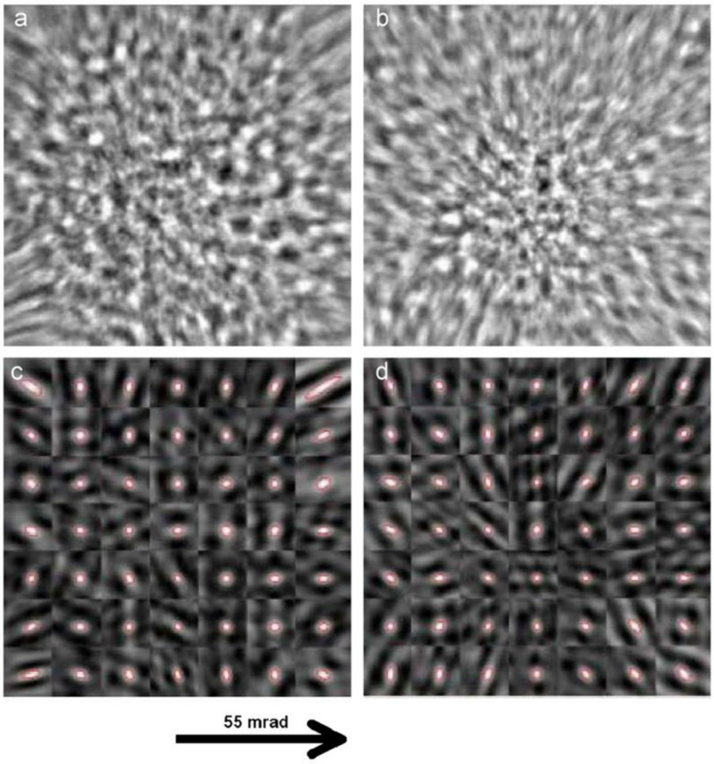


Figure 26 – Magnified centre region of Ronchigrams from a) 128 nm under-focus and b) 128 nm over-focus and their respective auto-correlation patch analyses c) and d). Image credit: reference [69].

This analysis is capable of rapidly determining up to the fifth order aberration constants [69], however, necessarily requires the use of an amorphous sample which may present an unacceptable experimental limitation. This method also performs relatively poorly when only small shifts are observed as is the case for low-order aberrations.

A similar method has been presented by Lupini, using a hybrid crystal-Ronchigram method [80]. This method may alleviate the need for an amorphous specimen but requires a large field of view of perfect crystal – potentially just exchanging one restriction for another.

Image Cross Correlation

Krivanek demonstrates an automated aberration determination software and its results based on the tilt induces shift method, Figure 27 [79].

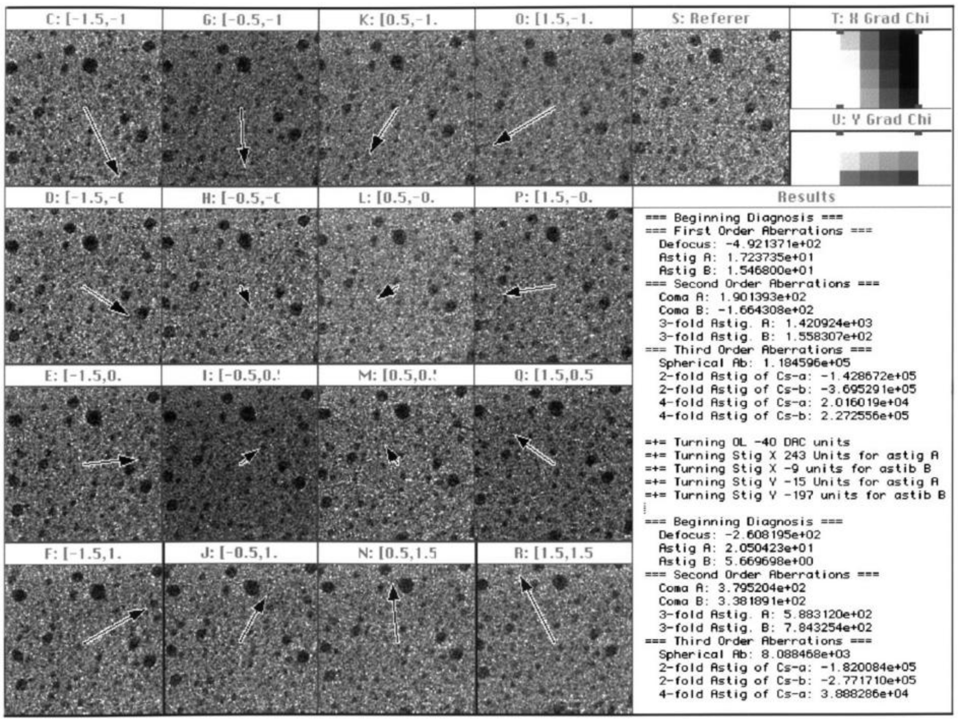


Figure 27 – Tilt induced shift automated aberration determination software. Image credit: reference [79].

Here a reference sample, usually a fine dispersion of similarly sized heavy metal nanoparticles, is imaged at a series of tilt angles. Each appears to have a slightly different defocus and astigmatism and it is the analysis of these that allows the determination of the aberration coefficients. Krivanek demonstrates this software to determine the aberrations, and subsequently drive the multi-pole correctors such that remnant coma, two and three fold astigmatism are all less than 1 μm, and C_s less than 0.12 mm [79].

Diffractogram tableau

In the TEM a method similar to the recording of a Ronchigram is that of the diffractogram. Here the Fourier transform (diffractogram) of a phase-contrast image of an amorphous specimen is calculated. Any lens aberrations will cause the resultant ring pattern to have an asymmetric form. The symmetry elements introduce to the otherwise circular pattern will necessarily match the symmetry of the aberration present. The method of observing this is the diffractogram tilt-tableau. Here the beam is tilted to a known angle off of the optic axis and a corresponding diffractogram recorded, Figure 28 [74], [81], [82], [83].

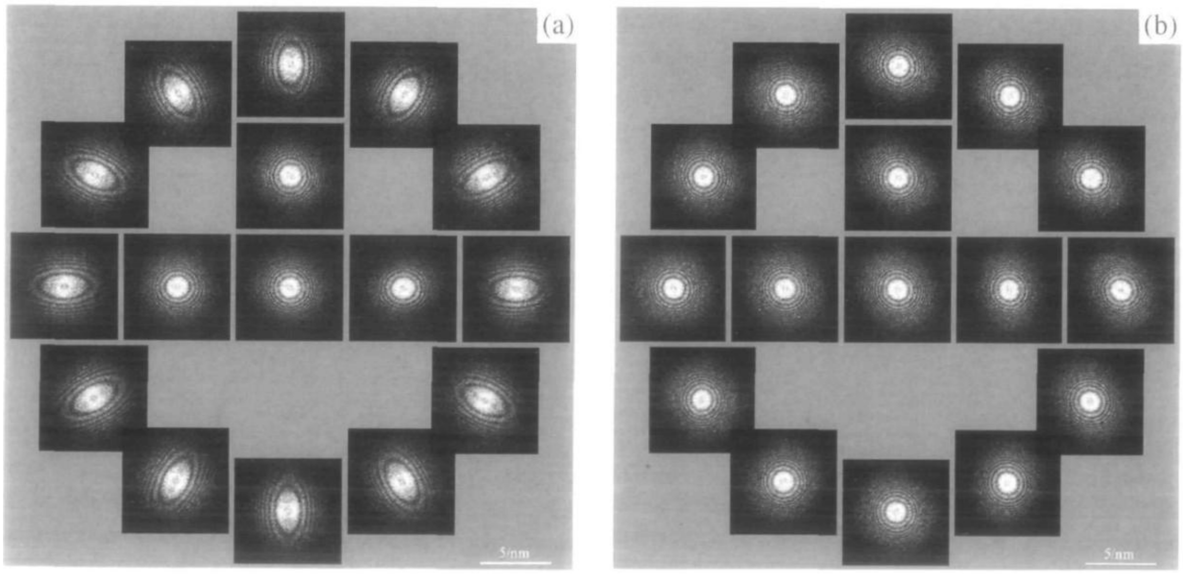


Figure 28 – Diffractogram tableau for a) the uncorrected and b) the corrected microscope. Image credit: reference [74].

For the uncorrected microscope, Figure 28 a), there is a significant elongation of the diffractogram away from the optic axis at all beam tilts. This is indicative of the presence of spherical aberration. Following the correct excitement of the hexapoles (see Figure 31) however, Haider demonstrates the roundness of the diffractograms out to 10.8 mrad* [74], [81]. A similar analysis by Zemlin, after whom such tableaus are often named, demonstrates the detection and correction of axial coma [82]. It is worth noting that by reciprocity [41], using a tableau based method to correct to larger tilt angles is equivalent to correction to a larger radius in the STEM Ronchigram.

* This was performed in 1998; more modern correction extends easily to 20 mrad and beyond.

When first applied this semi-automatic method took approximately fifteen minutes, which itself introduces issues with specimen / focus drift. However in more modern systems, with improvements in both digital detectors and computer power, this time has been greatly reduced taking about one minute.

Recent work by Meyer et al has demonstrated a variation of this technique for HRTEM where a series images at three (or more) focus values have been recorded at each of the tilt positions in the tableau [84], [85]. As would be expected acquisition times for such a method are higher but with the extra, oversampled, information gathered C_1 and A_1 values were determined to within 1 nm.

Saxton presents a similar method to that above, but measures the deviation in the angle of mirror planes in a diffractogram tableau rather than the actual form of each pattern being analysed [73]. This method does not require the exact defocus to be known in advance, nor the operating voltage or magnification; however the spherical aberration constant must be known in advance as well as the accurate calibration of tilt coils, making this method somewhat limited.

Focal-Series Power-Spectrum / Fourier Transform Analysis

In the STEM the contrast transferred in an image is known to vary as a function of defocus [55], [86]. As a result, if the expected behaviour of the imaging data can be predicted either analytically or from simulation then inspection of focal series can be used to measure some aberrations.

In the bright-field case, the power-spectra (Fourier transforms) of frames from a focal-series yield patterns of concentric rings equivalent to the Thon rings in TEM [76], [87]. An example of this is shown in Figure 29 [88].

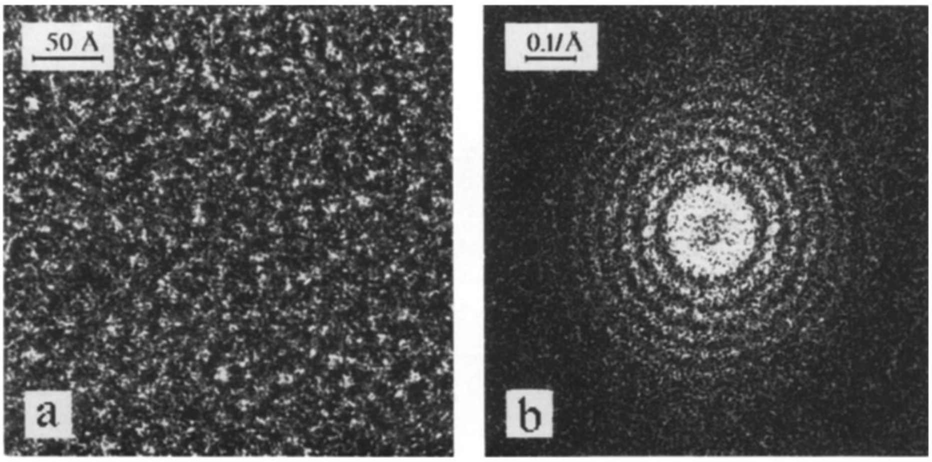


Figure 29 – Left: a bright-field STEM image of an amorphous carbon film ≈6 nm thick, and right: the corresponding power-spectrum (Fourier transform) shown with a logarithmic greyscale. Image credit: reference [88].

This power-spectrum can then be analysed to determine at least the approximate defocus and astigmatism of the illumination; and this is indeed the basis of the focus auto-tuning method presented by Mallic [76]

Similar work by Baba et al. show the equivalent dark-field case, again for amorphous specimens, Figure 30 [89].

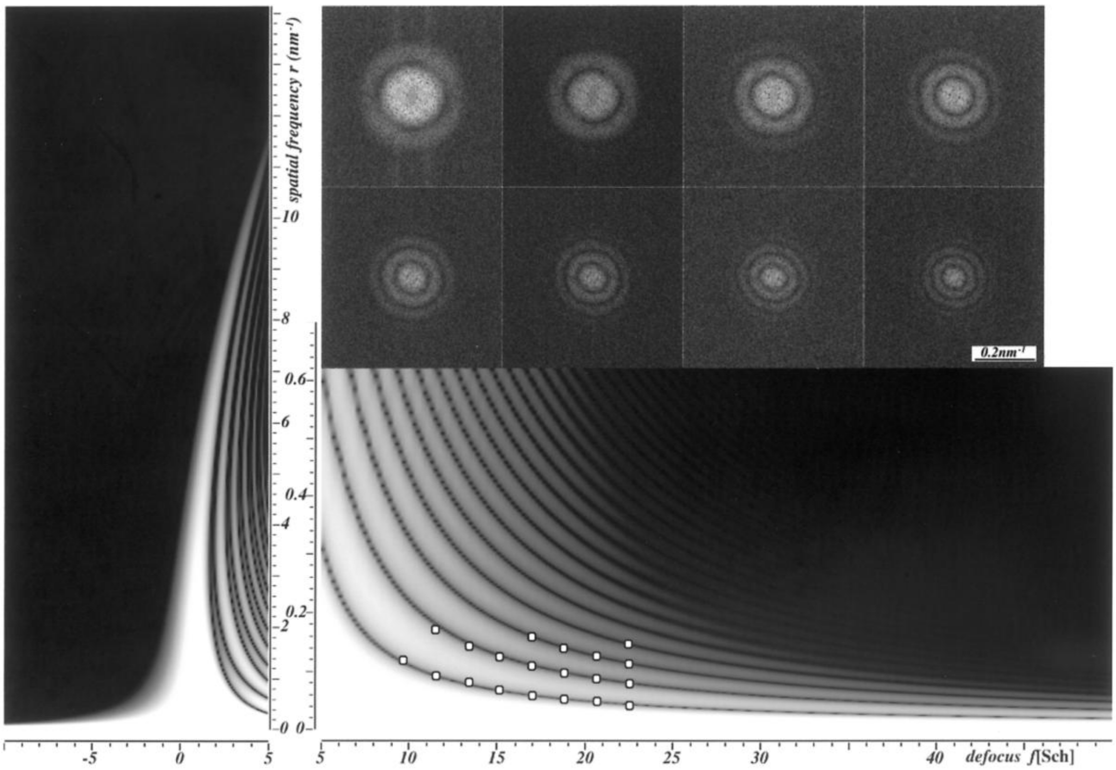


Figure 30 – Simulated transfer functions for dark-field STEM amorphous diffractograms. Note the enlarged scale of the right hand figure. Inset shows experimental diffractograms, the minima of which are marked as white circles on the simulation data. Image credit: reference [89]

The simulation data presented by Baba et al. applies only for amorphous specimens and as a result there are never any inversions of the contrast (no negative diffractogram intensity). The sweeping towards larger defocus with increasing spatial frequency can be seen in the left of Figure 30 and is the result of spherical aberration.

Equally, where crystalline specimens are used in the dark-field STEM case, variation in the transfer function caused by aberrations leads to variations in the appearance of the spots in the Fourier transforms of atomic resolution data. Recording these atomic-resolution images through focus again leads to a possible aberration measurement route [90].

1.8. Reducing Aberrations to Acceptable Levels

Haider presents a thorough analysis of the effects of various aberrations on the resolution attainable in the STEM [32], [91]. These can be considered as upper limits of acceptability, but also as indications of the required accuracies of the measurement (and compensation) of each of the various aberrations. More recent work by Bischoff & Riegers expands upon the data given by Haider pushing the calculations to an updated relevance corresponding to a 50 pm ceiling [90]. Interestingly in their work the state of the art across all current aberration auto-measurement methods is also tabulated, Table 2.

Table 2. Maximum allowable aberration coefficients for various resolution performance levels in a 300kV STEM microscope. Underlined values indicate required aberration tolerance that exceed the current accuracy of automated diagnosis methods. Image credit: reference [90]**.**

allowed maximum aberration coefficient limits to attain the π/4 limit at an acceleration voltage of 300 kV, and best (automated) practice.

Aberration		Resolution [pm] 80	70	60	50	Accuracy of automated methods
Defocus	C_1 [nm]	***<u>0.81</u>***	***<u>0.62</u>***	***<u>0.46</u>***	***<u>0.32</u>***	1
Twofold astigmatism	A_1 [nm]	***<u>0.81</u>***	***<u>0.62</u>***	***<u>0.46</u>***	***<u>0.32</u>***	1
Threefold astigmatism	A_2 [nm]	49.5	33.2	20.9	***<u>12.1</u>***	20
Axial coma	B_2 [nm]	16.5	11.1	***<u>6.96</u>***	***<u>4.03</u>***	10
Spherical aberration	C_3 [μm]	2.68	1.57	0.85	***<u>0.41</u>***	0.5
Fourfold astigmatism	A_3 [μm]	2.68	1.57	0.85	0.41	0.1
Star aberration	S_3 [μm]	0.67	0.39	0.21	0.10	0.1
Fivefold astigmatism	A_4 [μm]	136	69.8	32.2	13.0	2
Axial coma	B_4 [μm]	27.2	14.0	6.46	2.59	2
Three lobe aberration	D_4 [μm]	27.2	14.0	6.46	2.59	2
Spherical aberration	C_5 [mm]	6.63	2.97	1.18	0.40	0.2
Sixfold astigmatism	A_5 [mm]	6.63	2.97	1.18	0.40	0.05

Ultimately then, where auto-tuning is being used to diagnose any remnant aberrations, any ‘diagnosis ceiling’ may present an equally important barrier to performance as any aberration correction methodology. Equally it should be stressed that these aberration limits apply to an otherwise perfect microscope and do not include any experimental instabilities (such as lens-current, accelerating-voltage or mechanical vibrations). In the presence of such small instabilities the aberration tolerances maybe stricter still. These instability effects were collectively described by Haider as an additional contrast damping envelope [32] although their origins are not extensively discussed. A more thorough evaluation of the origins of these is presented by Muller [30], [31].

1.8.1. Multi-pole Hardware Aberration Correctors in Commercial Use

Succesful hardware aberration correction was first demonstrated by Haider et al for TEM [81] and Krivanek at al. for STEM [79] at roughly the same time in the late 1990s. Since then there has been rapid and continued development of this technology.

The hardware routes available for the correction of spherical aberration lay with two main manufacturers, Nion and CEOS, producing different designs [6], [92], [93]. These are the combined quadrupole-octopole [16] and the multiple hexapole [16] designs which are generally intended for dedicated STEMs and TEM/STEMs respectively [92]. This difference is reflected in the orientation of the ray diagrams shown in Figure 31 with the convention for dedicated STEMs having the gun at the base of the column.

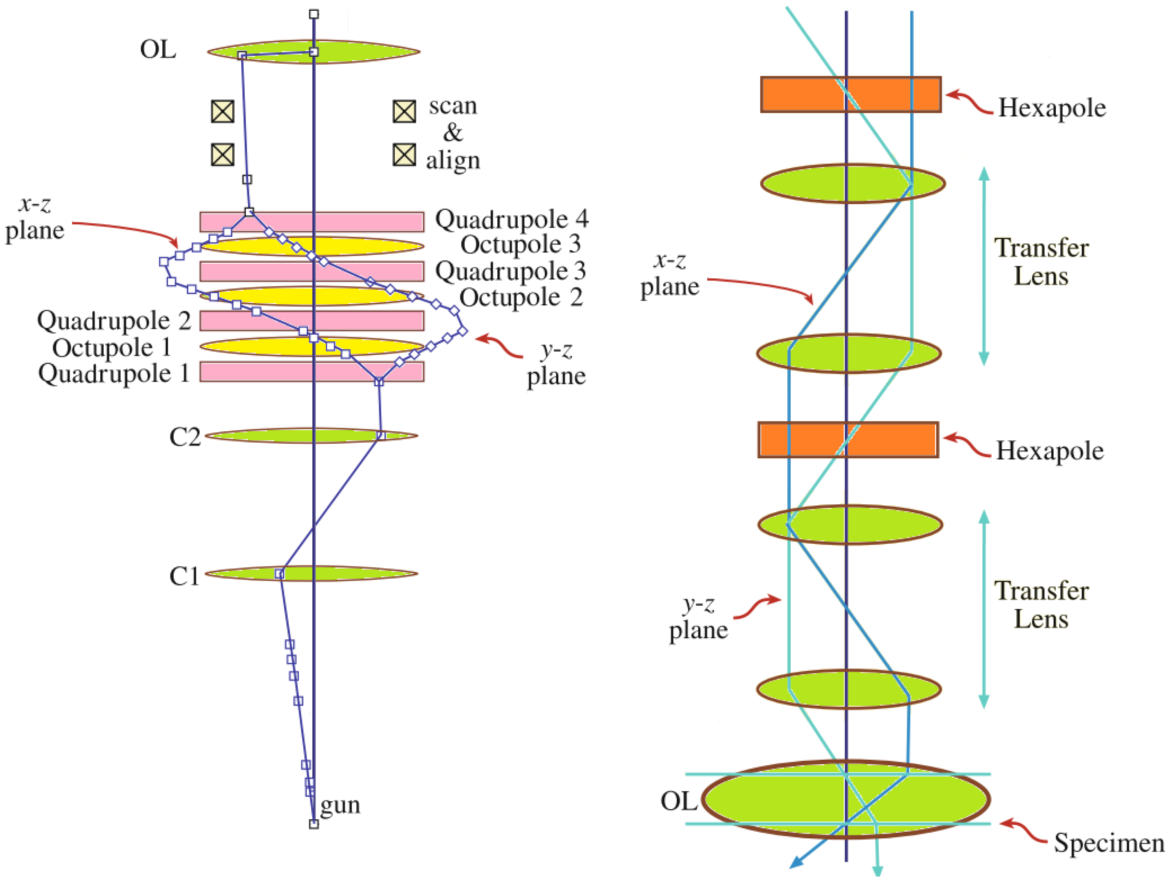


Figure 31 – Hardware Cs corrector technologies, left) Nion multiple quadrupole-octopole and right) CEOS hexapole / transfer-lens designs. Image redrawn from reference [6].

Both the Nion and CEOS designs are in column and require many additional lenses. These extra lenses are generally computer tuned to near-optimum conditions with the user performing fine tuning.

In addition to the added complexity of operating such a machine these extra lenses themselves introduce additional aberrations. Hardware correctors can never be placed at the source of the aberrations they are correcting, but by keeping this distance to a minimum the introduction of additional higher order and chromatic aberrations can be kept to a minimum [94].

The paired hexapole design most commonly installed in TEM/STEM instruments operates on the principle that the second haexapole compensates for the second order aberrations introduced by the first, leaving only a residual third-order aberration which is rotationally symmetrical. This is then of the same form, but opposite sign, as the spherical aberration introduced elsewhere in the microscope [74], [81]. The most advanced Nion hardware aberration correctors are now able to correct both third and fifth order aberrations being limited then by eight-fold astigmatism $C_{7,8}$ [94]. This aberration then defines the available illumination angles, θ_{max}:

$$\theta_{max} = \left(\frac{2\,\lambda}{C_{7,8}}\right)^{\frac{1}{8}} \tag{13}$$

For a typical 200Kv instrument this would give an illumination angle of 67 mrad, however this is only true if chromatic aberration is ignored and a more reasonable expected convergence angle would be ≈ 40 mrad [94].

1.8.2. Other Postulated Aberration Correction Systems

In addition to the two systems described above that have been successfully commercialised, a number of other approaches have been suggested that have not yet been fully realised.

Spherical Aberration Correction with Electrostatic Mirrors

For light optics Rayleigh suggest the use of bent mirrors as means to compensating for the aberrations. Mirrors themselves cannot be used for electron optical systems but electrostatic reflection has been suggested for use in the STEM by Crewe ,Figure 32 [95].

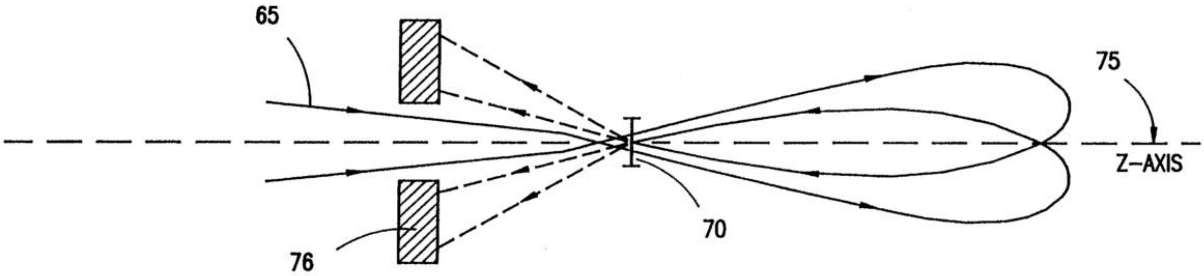


Figure 32 – Electrostatic mirror type aberration corrector. Image credit: reference [95]**.**

Here the electron probe (65) is pre-focussed near the specimen (70) beyond which and electric field reflects the beam in such a way as to oppose the previously introduced aberrations and refocus the illumination to a point on the far side of the specimen. Image recording would then be by the annular detector (76) in a dark field mode.

Spherical Aberration Correction by Conducting Foils

Scherzer suggested in 1949 the use of *"a lens system containing a conductive foil on which electrostatic charges are influenced by neighboring electrodes"*, Figure 33 [22].

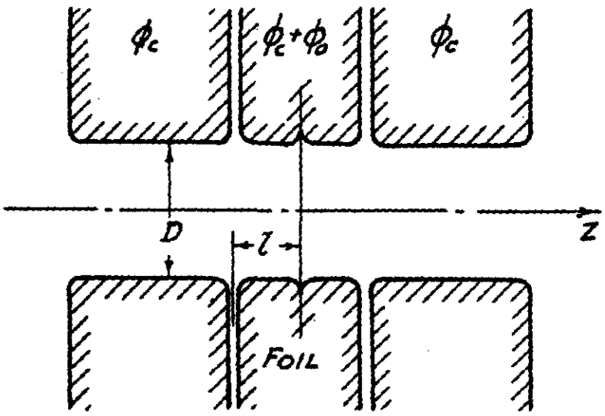


Figure 33 – Schematic design for a charged foil type corrector. Image credit: reference [22]**.**

This idea has been largely ignored in favour of electromagnetic lens type correctors although it may reappear in the future as it has the possibility of being more simple than some of the very complicated correctors currently being produced.

Aberration Correction with Diaphragms, Precession and Image Processing

Uno et al have demonstrated that for the case where aberrations are only corrected in one radial direction at a time, they can be corrected further along that direction than if correction of the entire aberration surface were attempted, Figure 34 [96].

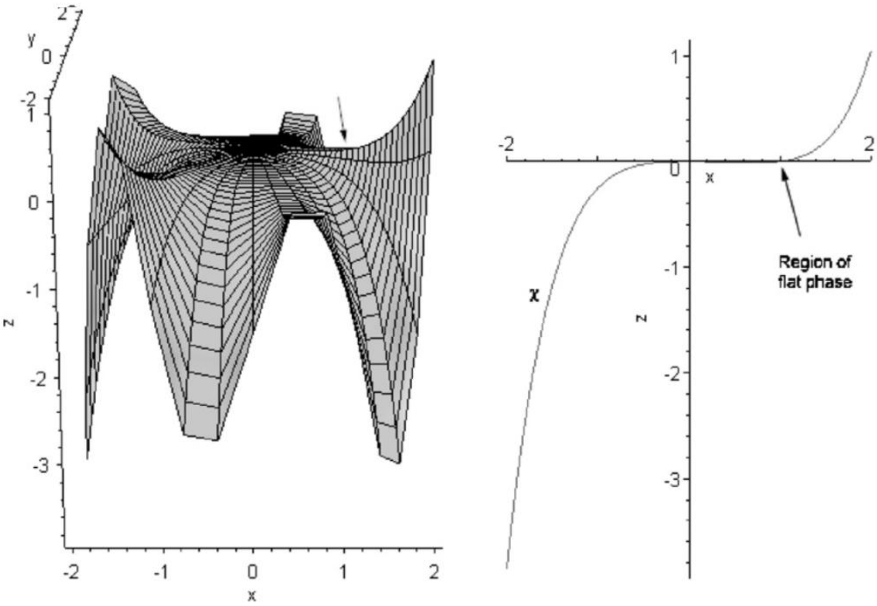


Figure 34 – Flat phase region created (right) along the indicated direction in the aberration surface (left). Image credit: reference [96].

This work however only demonstrates this technique for the upper (objective lens) aberrations and does not consider any lower optics as well as acknowledging that successful inversion of the collected diffraction data has not yet been achieved. Should this be successfully achieved a microscope equipped for precession diffraction type work with an aperture such as that stipulated by Mast and Jong may be one possibility for this work, Figure 35 [97].

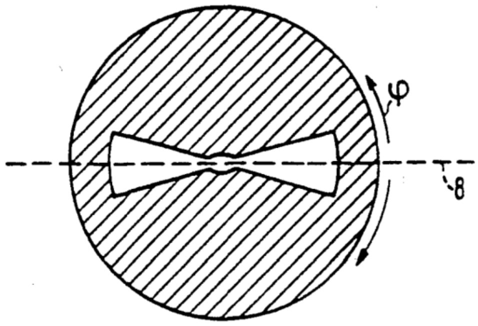


Figure 35 – Electron blanking diaphragm. Image credit: reference [97].

However, even if such an instrument and image-processing route can be achieved there are likely to still be complications resulting from focal or specimen drift, or sample damage, during the long beam-exposure times needed for the collection of the multiple images .

1.8.3. Comparison of Uncorrected and Corrected Instrument Performance

Looking again at Figure 25, we see that aberration correction gives a much wider usable region of the Ronchigram (or equivalently a flatter phase surface) and increases the usable aperture angles from around 10 mrad to 40 – 60 mrad. The use of larger apertures then improves both lateral and depth resolution as given by equations 6 and 8. In Figure 8 some some example x-z views of an uncorrected STEM probe as well as those achievable with third and fifth order correctors [16], [27]. To evaluate these more quantitatively, it is useful to plot the intensity of these probes through their waists to observe both the intensity and the width of the probe, Figure 36 [39].

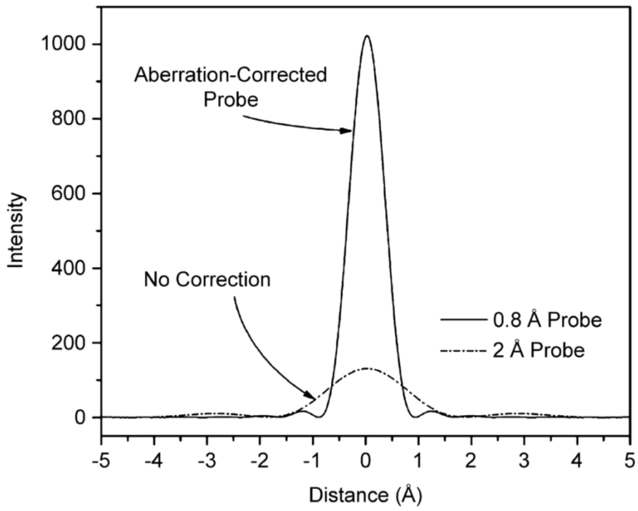


Figure 36 – Uncorrected and corrected probes normalised to show equal total intensity. Simulation details and image credit: reference: [39].

In this view we can see that aberration correction leads to a smaller and more intense probe. However, this one-dimensional view is only part of the story; we must also inspect the form of these probes in the plane normal to the optic axis. Muller et al show some simulated probes both with and without aberration correction, Figure 37.

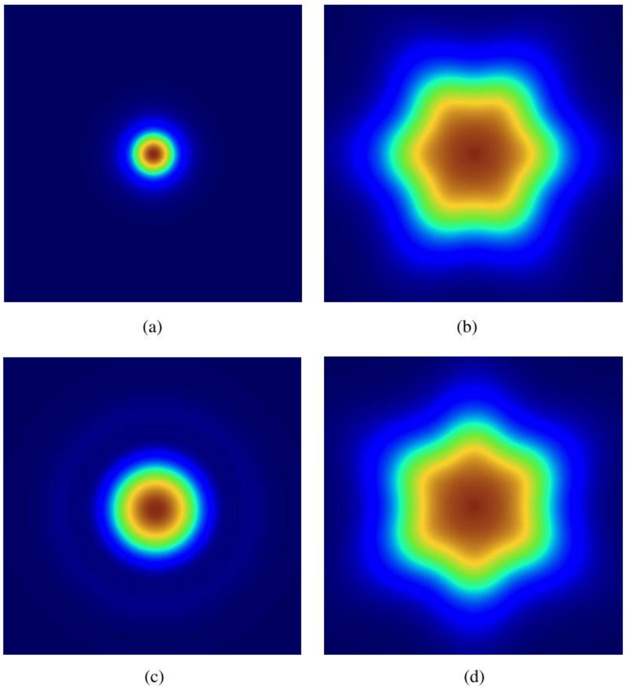


Figure 37 – Simulated 300kV STEM probes a) hexapole corrected (Gaussian focus), b) 8 nm under-focus, d) 8 nm over-focus and c) the uncorrected probe for comparison. Squares all show 0.63 nm field of view. Image credit: reference [70].

Again Figure 37 shows the corrected probe (a) to be smaller than the uncorrected one (c). However, the use of a multiple hexapole corrector leaves the STEM probe with a residual 6-fold shape and would itself require a do-decapole (12 poles) corrector to remedy. This behaviour is difficult to identify when the probe is at focus but readily seen when defocussed slightly (Figure 37 b and d).

Aberration correction then has yielded sharper and more intense probes, but at the expense of losing the simplicity of the previously round probe shapes and hence OTFs. For this reason the non-round optical-transfer properties of aberration corrected probes becomes worthy of study and this is investigated extensively by simulation in section 2.1.

1.9. Compensating for Remnant Aberrations in Post-processing

In Table 2 the current state of the art for auto-aberration tuning was shown, where in the majority of cases this falls short of the require tolerances. It is unsurprising then that attempts have been made in the literature to compensate for these residual aberrations using data post-processing techniques, and three such examples are presented here.

1.9.1. De-convolution of HAADF STEM Data

From equation 10 we saw that the recorded STEM image can be considered as the convolution of the 2D object-function representing the sample with the point spread function (PSF) incident upon it. It is reasonable then that various routes to de-convolve some or all of the blurring effect of the finite sized probe from the recorded data have been attempted [46], [98]. In this situation the probe's PSF can be estimated by blind-deconvoluiton from a region of perfect crystal, maximum entropy approaches [99], or by complimentary probe simulation [98], [100]. However, some caution should be taken when attempting de-convolution of image data. Where off-axial aberrations persist then the PSF of the illumination may not be equal across the entirety of a field of view [101] leading to a possible source of error. Equally where the data is not calibrated in a quantitative sense [102] it is possible to alter the contrast recorded and the so-called z-exponent in dark-field STEM [53].

1.9.2. 'Defocus-Modulation' In TEM

Taniguchi presents a method for the recording of through focal TEM bright field images, from which an analysis of the diffractograms and comparison with simulation (Figure 38) is used to design restorative image modulation functions (Figure 39)– where the data is weighted in both real and reciprocal space before summation to yield a spherical-aberration compensated image [103].

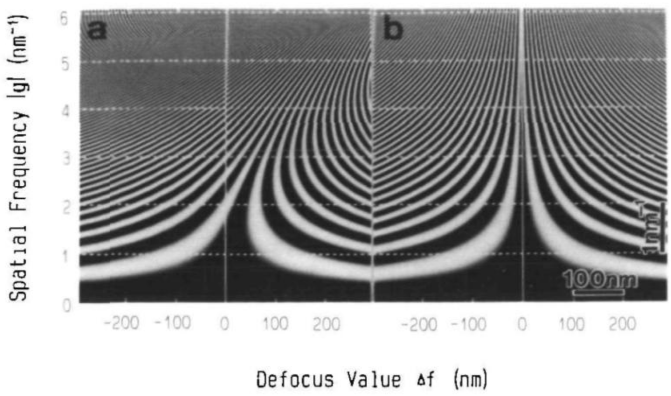


Figure 38. Simulated Thon diagrams for BF TEM with a) C_s = 0.85 mm, and b) Cs = 0 mm. Image credit: reference [103].

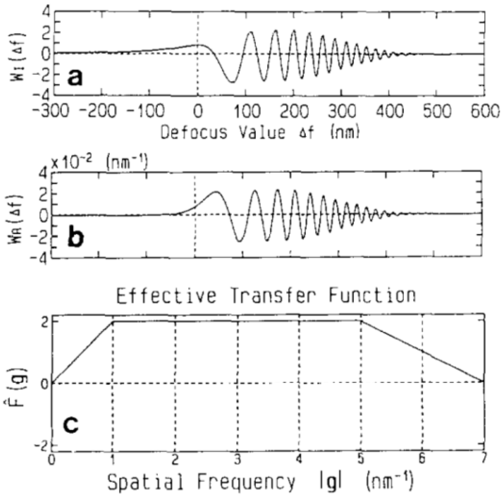


Figure 39. Image credit: reference [104]

This was initially presented for an amorphous carbon film [103] but later for crystalline gold specimens [104]. This technique enabled a phase representation of a weak-phase object specimen to be recovered as well as compensating for much of the spherical aberration observed. In this way it can be said that this defocus-modulation was the predecessor the full exit-wave restoration.

1.9.3. Exit-wave Restoration in HR-TEM

Whilst some utility has been presented for HRTEM image de-convolution [105], it is always preferable to record as much information as possible initially and use image processing sparingly. This is especially true where experimental image resolution and noise may limit the de-convolution performance [105]. More constrained deconvolution practices that are 'semi-blind' may be more tolerant of image noise but require a-priori knowledge about the illumination and careful consideration of the expected image properties both in real and reciprocal space [106].

In TEM a more powerful alternative is to record a series of images of the same specimen with a varying defocus or illumination tilt. This yields image data where each individual frame will have been subject to differing optical distortions (aberrations) but importantly the same object was imaged each time. What is produced then is a so-called 'over-determined' data-set where there are only two unknowns (the complex wave-function at the object's exit surface and the illumination properties) but many experimental images. From this data-set the complex exit-wave-function can be extracted using carefully designed 'restoring filters' [107], or 'parabolic filter functions' [108]. Details of the calculation of these filters is beyond the scope of this review but is presented concisely elsewhere by Kirkland & Meyer [107].

The word complex here is key and highlights one further advantage of this method relevant to coherent imaging. Using this technique, data recorded on electron detectors which record the image intensity only, can be used to retrieve the phase of the specimen exit-surface wave-function.

One example of such a restoration is presented in Figure 40 [109].

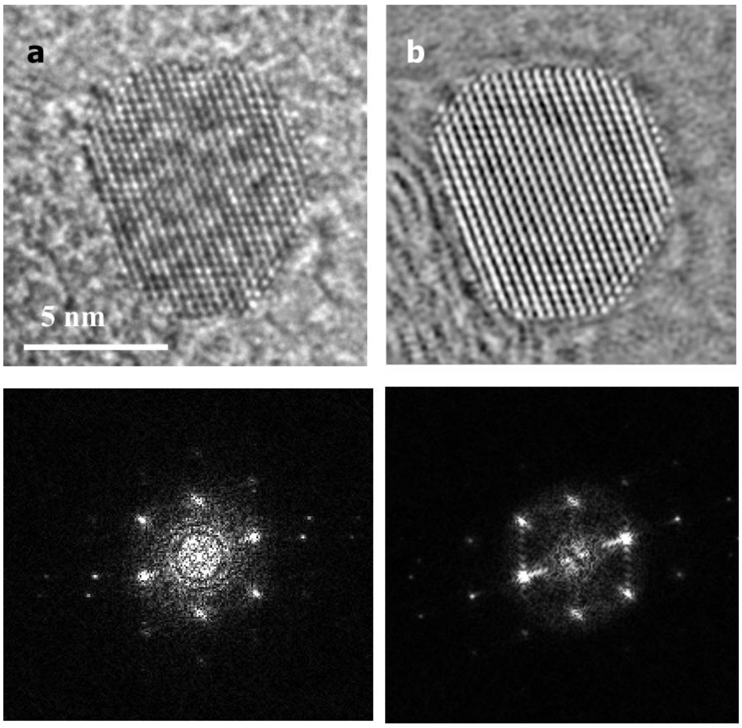


Figure 40. Top-left: an HRTEM image of a Pt nanoparticle on a carbon support recorded with 30 μm of negative C_s. Top-right: the restored phase of the exit wave from a 20 image defocus series of the same particle. Bottom row: the corresponding Fourier transforms indicating the additional information transfer in the restored exit wave. Image credit: reference [109].

For thin and / low atomic number objects, where the weak phase object approximation in TEM holds valid, individual image frames are expected to have very little contrast (especially when close to focus).This can be verified in Figure 40, where the carbon support in the lower-left of the field of view displays virtually no contrast in the experimental image but appreciable structure in the restored phase map.

Finally, whilst this practice does not necessarily require an aberration corrected instrument [110] it is generally preferred to begin from the best possible optical condition and aberration corrected data dominates the exit-wave restoration literature.

1.10. Conclusions from the Literature Review

The STEM instrument with its sharply focussed probe yields many useful imaging and analytical signals simultaneously. The majority of these can be described by the incoherent imaging model, described well by the convolution of the sample's 2D object-function with the point-spread function of the illumination. The achievable resolution of the microscope is then directly dependent on the width of the probes that can be formed; and while larger probe-forming apertures would seem an obvious answer – their use is limited by the aberrations present.

To readily achieve atomic resolution, a key tool for modern materials science, aberration diagnosis and compensation become essential tools for the modern STEM operator. While many methods were initially explored to correct for positive spherical aberration the industry has settled on two designs [74], [79], both utilising multi-pole optics. These additional lenses allow, in theory at least, bespoke amounts of several different aberrations to be induced – each mutually compensating for one another. However, with the introduction of the extra optical elements commensurate with multi-pole aberration correctors (75 elements or greater! [111]) automated tuning becomes essential, especially for higher-order higher-symmetry aberrations.

The attention of the field rapidly turned to exploring a plethora of different aberration diagnosis methods for auto-alignment of these highly complex optical devices. There is no shortage of possibilities here, with most permutations of crystalline / amorphous sample, position-series / tilt-series / focus-series based acquisitions, and real-space / reciprocal-space analysis approaches all being evaluated. Of these methods all seem to attain a similar precision that has been found to be just insufficient for the lower-order most commonly observed aberrations [80], [90].

There exists then opportunities in STEM imaging, for either improvement of the accuracy of the aberration diagnosis / compensation procedure, or alternatively to directly extract image data from the over-determined data-sets collected during these tuning procedures.

2. Three-dimensional Contrast Transfer in the STEM

The wavelengths of electrons typically used for STEM would be expected to yield a diffraction limited resolution somewhere in the pico-meters scale. Instead performance of most instruments is around 100 times worse – demonstrating the importance of both the non-zero (finite) size of the electron source, and the effect of imperfections in the electron lenses. In this section these two factors will be discussed in terms of their role in generating the three-dimensional electron probe used in STEM imaging. The performance of various simulated illuminations will be evaluated via their optical transfer functions, and both the modulus and phase of this complex commodity will be explored. Three-dimensional visualisations as well as lateral and vertical sections will be used to clarify features of the optical performance with especial attention afforded to variations with respect to defocus.

2.1. Contrast Transfer Functions in the Aberration-Corrected STEM

Spherical-aberration (Cs) can often be performance limiting in optical systems. The effect of spherical aberration on the transfer of image contrast in light-optical systems (such as telescopes, microscopes, and photographic equipment) has been described in detail by Black and Linfoot [112], who presented the first numerically calculated optical transfer functions (OTFs) for systems afflicted by spherical aberration; both as a function of defocus across spatial frequency, and for select lateral spatial-frequencies across defocus.

In the scanning transmission electron microscope (STEM), hardware aberration correctors can now correct for the positive spherical aberration of round electron lenses. This in turn allows larger probe-forming apertures to be used which both improves the expected diffraction-limited resolution available for imaging and increases the probe current available for analytical work. However, these hardware correctors necessarily require the use of multi-pole optics, meaning the final probe shape and image resolution can be limited by higher radial- or azimuthal-order aberrations. The resultant STEM probes then often exhibit an azimuthally-varying lateral-resolution and focus-sensitivity.

In electron-optics, Loane, Xu and Silcox predicted by simulation, and subsequently verified experimentally, that it is possible to go beyond the Scherzer-focus resolution limit of an uncorrected STEM instrument by defocusing sufficiently to raise the transfer of a given spatial-frequency [55]. Later, but still using uncorrected hardware, this technique was expanded by Nellist and Pennycook to achieve sub-Angstrom resolution [113].

In recent years these techniques of manipulating defocus to achieve higher spatial-resolution, have been superseded by the advent of hardware aberration correctors and such hardware is becoming increasingly commonplace. It is, however, important to remember that for Cs-corrected instruments, the C_s can be tuned to compensate against the next highest order of round aberration – the fifth-order spherical aberration ($C_{5,0}$). Additionally, as Scherzer demonstrated the impossibility of creating a spherical-aberration-free round electron-lens [66], these hardware correctors necessarily include non-round multi-pole lenses such as the multiple-hexapole [74] 5 or quadrapole-octopole [79] 6 systems currently available commercially. Whilst the development of hardware C_s correctors has enabled the use of larger objective apertures, with a subsequent reduction in overall probe size and an increase in the available probe current for analytical work, we have traded the previously simple C_s-dominated aberration surface for a far more complicated one. The introduction of non-round lenses now means that the objective-plane phase-surface, and ultimately the diameter of the STEM probe, is now often controlled by non-round aberrations [71], [91].

During microscope operation, the final aberration to be adjusted is defocus, and this is usually fine-tuned by hand. Accurate correction of defocus therefore demands an understanding of how the microscope transfer varies with this defocus in the presence of non-round aberrations, as previously calculated for C_s alone by Black and Linfoot [112]. Furthermore, several experiments in the aberration-corrected STEM now utilise the sharply waisted shape of the electron probe, including: optical-sectioning, scanning confocal geometries and focal-series reconstruction methods. As a result, an understanding of the full three-dimensional (3D) form of the expected optical-transfer is essential in knowing what performance to expect from an instrument / experiment.

In light of the current, higher-order limiting aberrations in electron optics, it now seems timely to revisit the work of Black and Linfoot [112], and to explore the form of the OTF when higher radial- or azimuthal-order aberrations are incorporated.

To investigate the effect of various aberrations on the OTF of STEM probes, simulations were performed with a range of aberrations incorporated both separately and concurrently. Here the $C_{n,m}$ aberration notation presented by Krivanek is used as it most intuitively describes the angular dependence and rotational symmetries present [79]. Readers familiar with different notations may find the conversion table presented by Sawada useful [69]. Calculation steps were as follows:

1. Definition of the wavefront deviation, relative to a perfect spherical wave, χ, as a function of the angle from the optic axis, θ, and the azimuthal angle around the optic-axis, φ[†]:

$$\begin{aligned}\chi(\theta,\varphi) = {}& \tfrac{1}{2}\theta^2\left[C_{1,0} + C_{1,2}cos(2\varphi-\varphi_{1,2})\right] + \ldots \\ & \tfrac{1}{3}\theta^3\left[C_{2,1}cos(\varphi-\varphi_{2,1}) + C_{2,3}cos(3\varphi-\varphi_{2,3})\right] + \cdots \\ & \tfrac{1}{4}\theta^4\left[C_{3,0} + C_{3,2}cos(2\varphi-\varphi_{3,2}) + C_{3,4}cos(4\varphi-\varphi_{3,4})\right] + \cdots \\ & \tfrac{1}{5}\theta^5\left[C_{4,1}cos(\varphi-\varphi_{4,1}) + C_{4,3}cos(3\varphi-\varphi_{4,3}) + C_{4,5}cos(5\varphi-\varphi_{4,5})\right] + \cdots \\ & \tfrac{1}{6}\theta^6\left[C_{5,0} + C_{5,2}cos(2\varphi-\varphi_{5,2}) + C_{5,4}cos(4\varphi-\varphi_{5,4}) + C_{5,6}cos(6\varphi-\varphi_{5,6})\right] + \cdots\end{aligned} \tag{14}$$

2. Multiplication to give the complex aberrated wavefront, and masking with a circular top-hat function to represent the objective aperture that defines a convergence angle, α:

[†] The strengths and orientation of aberrations can equivalently be expressed in a $C_{n,ma}$ + $C_{n,mb}$ form as in [101].

$$W(\theta,\varphi) = \begin{cases} e^{\left(-\frac{2\pi i}{\lambda}\chi\right)} & if\ \theta\ \leq\ \alpha \\ 0 & otherwise \end{cases} \tag{15}$$

where λ is the electron wavelength.

3. Conversion of this 2D polar coordinate-system to a Cartesian one to facilitate numerical Fourier transform computation:

$$W(\boldsymbol{r}) = W(\boldsymbol{k_x}, \boldsymbol{k_y}) = W(\theta,\varphi) \tag{16.1}$$

where: $\boldsymbol{k_x} = \frac{\theta}{\lambda} . \cos\varphi$, and (16.2)

$$\boldsymbol{k_y} = \frac{\theta}{\lambda} . \sin\varphi$$

4. Fourier transformation of the complex-wavefront to yield the real-space probe, Ŵ. For incoherent imaging modes in the STEM (such as: dark field imaging, EDX or EELS signals) the square of the modulus of the Fourier transform represents the intensity of the real-space probe as a function of the physical position vector, ***r***, in a plane perpendicular to the optic axis:

$$I(\boldsymbol{r}) = \left|\widehat{W}\right|^2 \tag{17}$$

5. Finally, to visualise the strength that various spatial frequencies are imaged, an addition Fourier transformation yields the complex optical transfer function (OTF) as a function of the reciprocal-space vector, **r***:

$$OTF(\boldsymbol{r}^*) = \hat{I} \tag{18}$$

The simulated STEM probes were calculated using the Matlab® programming environment. A little care is required in setting up the numerical calculations as sufficiently fine-sampling and lateral field of view are required in both real and reciprocal space. To prevent the introduction of artefacts when performing the Fourier transform to yield the OTF, it is important that the real-space probe intensity should be negligible at the edges of the simulation bounds (Figure 41, c)). Upper limits also exist on the aperture-size and defocus that can be simulated practically [101].

Visualisations of the four stages of the probe simulation are illustrated in Figure 41.

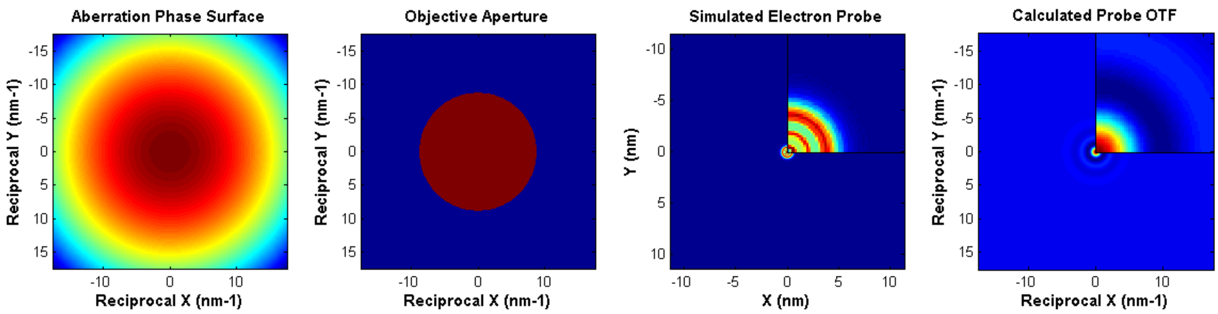


Figure 41. Illustrations of the probe simulation steps. From left to right: the aberration phase surface, the objective aperture mask, the real-space STEM probe and an amplitude plot of the complex OTF. Quadrant insets show the central part magnified by a factor of 8x to visualise the fine detail.

This process is repeated across a sweep of defocii with the 2D probe-intensity image being recorded for each. Subsequently, stacking together these individual views then yields a three dimensional (3D) representation of the real-space probe. Calculation of the aberration phase-surfaces at each defocus step is computationally simplified by the fact that only defocus varies and the contributions from all other aberrations need not be recalculated.

Whilst it is not the primary focus of this work, additional deleterious aspects of the illumination can also be incorporated in the simulations such as; convolution of the real-space probe laterally to incorporate the effects of finite source-size [10], [11] and vertically (along the optic-axis) to incorporate the effects of chromatic aberration (C_C) and temporal coherence [17], [101]. These effects are discussed at the end of this work in relation to 'real' probes. However, not being part of the illumination formation, resolution loss from noise, instabilities or sample (radiation) damage will not be discussed here.

Once the 3D real-space probe is calculated it can be Fourier transformed into reciprocal space to give the three-dimensional complex OTF. Computationally it is more efficient to perform this operation in a single step (without multiple disk read-write operations) so long as sufficient computer memory is available, and this imparts a limitation on the array sizes available for the numerical calculation. Arrays of 1024x1024 pixels at each of 256 defocii can be computed on a desktop PC and give a more than acceptable pixel sampling.

A vertical section through the 3D STEM probe reveals the hour-glass shape of the focussed illumination and this is shown in Figure 42. Here, and throughout this entire work, the simulation reflects a 200kV illumination and a 22 mrad objective aperture.

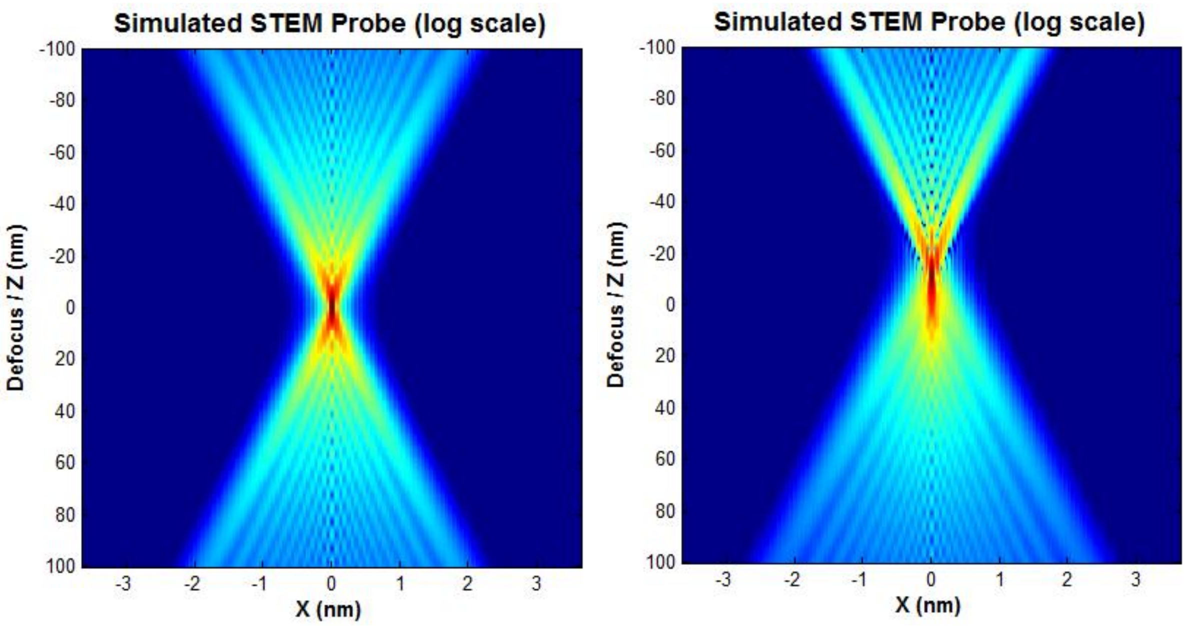


Figure 42. X-Z real-space views of simulated STEM probes (false-colour of intensity, log-scale). Left with no speherical aberration; right, with 50µm C_s. Note, for clarity the scale of the plot has been expanded horizontally by a factor of ≈30.

In the case of a totally aberration free probe (Figure 42, left) we see that the illumination is symmetric about the zero defocus plane, and the lateral resolution is expected to be limited by the Rayleigh diffraction limited only [20] and the depth, or z, resolution for optical sectioning is only limited by the size of the aperture angle [21]. When spherical aberration is introduce (as in Figure 42, right) we see that the cross-over shifts towards negative defocus; the compensation of which will be familiar to many microscope operators.

The 3D Fourier transform of a given real-space probe then gives its complex optical transfer function (OTF). Inspecting the amplitude of the complex OTF we can see that this amplitude has bounds displaying a toroidal shape with a missing cone [114], [115]. The size of this missing cone becomes significant owing to the small convergence angles used in electron microscopy [24]. A vertical section through the reciprocal-x / reciprocal-z plane is shown in Figure 43 corresponding to the two real-space probes shown in Figure 42. Note, to produce a fairly comparable phase plot through the complex OTF, the aberrated probe (Figure 42, right) was shifted with respect to defocus so as to move its cross-over to the centre of the real-space data-cube. Failure to re-centre the real-probe in this way will simply introduce a vertical phase ramp in the OTF section but this may complicate interpretation so it is best avoided.

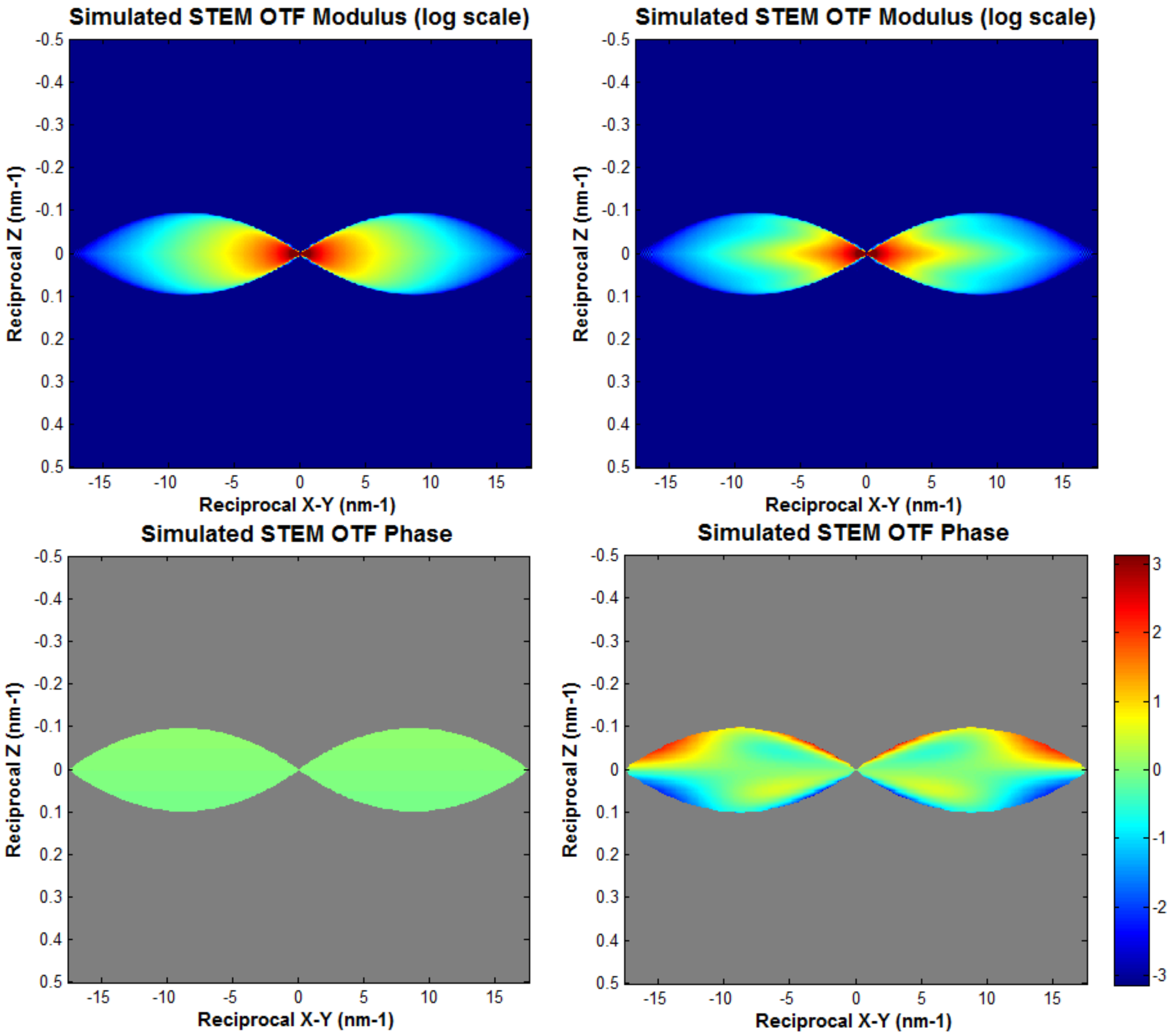


Figure 43. Amplitude plots (top, false-colour of modulus, log-scale) and phase plots (bottom, false-colour of phase, linear scale) of the OTF of the probes shown in Figure 42. Note, for clarity the scale of the plot has been expanded vertically by a factor of ≈35.

In the case of the OTF from the perfect probe we see a monotonic reduction in the lateral information transfer (reciprocal-z, or z* = 0) from the centre out to the information limit (edge of figure). For the OTF corresponding to the probe with C_s this information transfer again decays monotonically but is more rapidly damped than the previous case. In both cases the bounds of the missing-cone are defined by the aperture angle.

Whilst plotting the amplitude of the OTF is a useful plot for representing perfect probes its usefulness is limited when inspecting or comparing practical, aberrated probes. If we take the most simple example of a probe only affected by C_s as in Figure 42 (right) then the amplitude of the OTF remains symmetrical about the z = 0 plane (necessarily because it is the Fourier transform of a real quantity). However, it is clear that the real-space probe is not symmetrical about this plane. This asymmetry can be seen in the phase of the OTF but this is not as readily interpretable and is rarely plotted. What is needed then is a figure which combines the direct interoperability of viewing the lateral optical-transfer in frequency-space but that faithfully represents the genuinely asymmetric nature of aberrated probes with respect to defocus. The stack of the individual 2D optical transfer functions (OTFs) plotted as a function of defocus yields exactly such a figure and two example reciprocal-x versus real-defocus sections through two such stacks are shown in Figure 44.

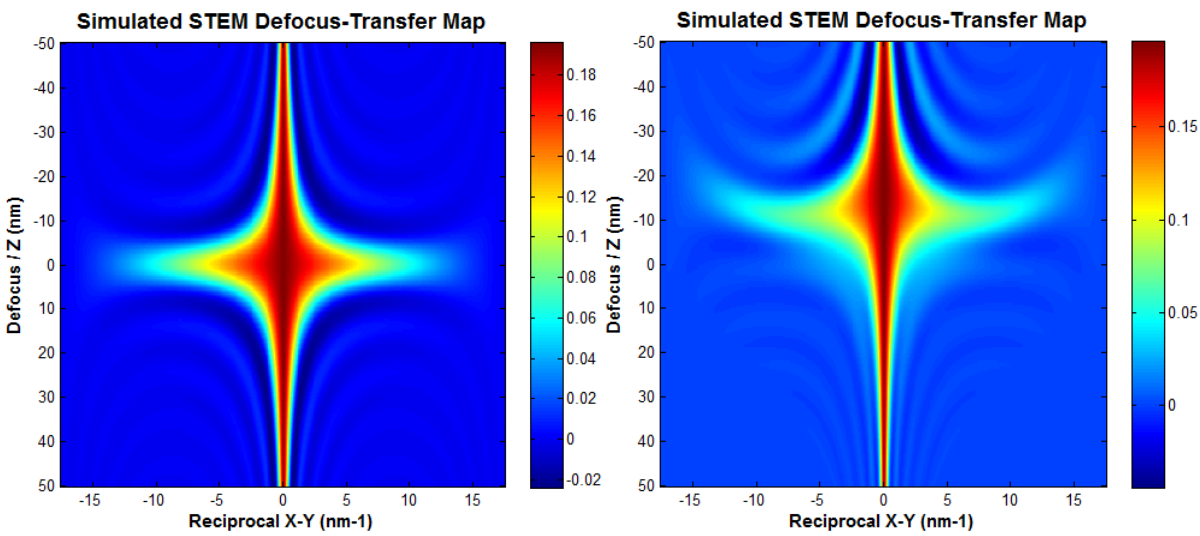


Figure 44. Reciprocal-X versus defocus plots corresponding to the simulated probes shown in Figure 42, intensity shows the strength of optical-transfer (false-colour of intensity, linear-scale).

These figures directly reveal the role of aberrations in the redistribution of contrast with respect to defocus. In Figure 44 spherical aberration causes an upward sweeping that, as microscope operators will be familiar, requires the introduction of defocus to compensate. These figures share some similarities with the work of Baba et al. [89] (Figure 30), however where that work was concerned with diffractogram contrast from amorphous specimens, these figures represent only the properties of the illumination and in this more general consideration negative contrast (destructive image interference) for certain spatial frequencies is observed.

Figures such as Figure 44 can be generated in a few seconds on a laptop computer using the readout of the microscope's aberration corrector, and can provide immediate guidance to the operator. This is again a strong motivation for retaining defocus as the vertical axis in these representations as this is the most frequently adjusted microscope parameter.

If desired, the two-dimensional vertical section shown above can be further sectioned either horizontally (at constant defocus) or vertically (at a given spatial frequency). Extracting constant defocus sections yields 1D optical transfer functions as in Loane, Xu and Silcox's Figure 3 [55]; while sectioning the data for a fixed spatial frequency yields 'defocus response-curves' as in Black and Linfoot's Figure 6 [112]. In the case of the aberration free probe this contrast-defocus figure is again symmetric about zero defocus, however in the presence of C_s the contrast transferred is asymmetric with respect to defocus as expected. In the presence of spherical aberration the side 'wings' of the contrast-defocus plot sweep up indicating the focus at which any given spatial frequency is transferred most strongly (similar to Black & Linfoot – Figure 8). The figure also shows the well-known feature that underfocus can partially compensate for positive spherical aberration.

These sections through the spatial-frequency versus defocus volume are useful for comparing the perfect and spherically-aberrated probes in Figure 44, however in the age of hardware corrected STEM probes we expect to be increasingly limited by azimuthally non-round aberrations. As a result of this increasingly complex behaviour sets of line-plots (such as in [55], [112]) cannot intuitively represent this information and a full three-dimensional visualisation of the spatial-frequency versus defocus volume is needed. Fortunately with modern computing power this is now trivial to calculate and visualise and such a volume is shown in Figure 45.

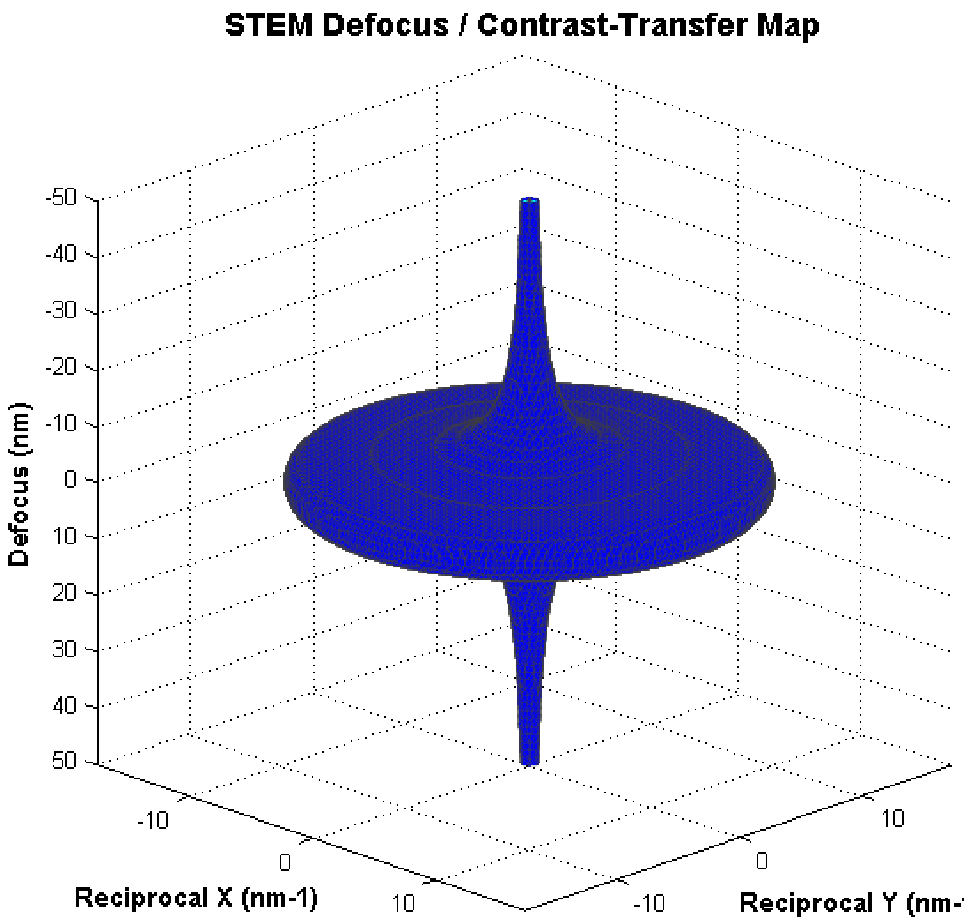


Figure 45. 3D visualisation of the spatial-frequency versus defocus optical-transfer volume corresponding to Figure 44 (left). Surface shows the bounds of the 10% optical-transfer region.

In such figures the 10% optical-transfer limit is an appropriate bounding surface to plot as below this level contrast is likely to be masked by the finite signal to noise ratio of experimental data. Put another way, voxels within this bounding surface represent spatial-frequency and defocus combinations with at least a 10% optical transfer modulus.

The shape of this bounded volume is dominated by two main features; the lateral disk and the central rod. In the case of the defocus-transfer volume for the aberration-free probe (Figure 45), the disk centred about zero defocus is flat and rotationally symmetric. This means that all directions of a given lateral spatial-frequency are transferred equally, and, that all lateral spatial-frequencies in the image are transferred most strongly at the same defocus – namely zero. It is the effect on the rotational symmetry of this central disk (on its z-curvature, z-spread and lateral bounds) that will be examined in detail now for each of the common residual aberrations that can be found in the hardware-corrected STEM. The central core running vertically across focus (at $\mathbf{r}_x = \mathbf{r}_y = 0\ \mathrm{nm}^{-1}$) through Figure 45 signifies the zero spatial-frequency component of the OTF which incorporates any constant background in real-space. This persists over all defocii and is not discussed further.

As we go on to discuss the higher order remnant aberrations a consistent notation is essential. The aberrations that will be examined here are summarised in Table 3.

Table 3. Notations for the first four real (round aberrations – strength only) and the first 19 complex valued (strength and direction) aberrations in the STEM (after [70]). Six aberrations that are commonly observed experimentally are marked with an asterisk indicating those discussed in detail in the text.

Name	Azimuthal Symmetry (multiplicity – m)	Radial Order (n)						
		1	2	3	4	5	6	7
Defocus	0	$C_{1,0}$						
m-fold axial astigmatism	n + 1	$C_{1,2}$*	$C_{2,3}$*	$C_{3,4}$*	$C_{4,5}$	$C_{5,6}$*	$C_{6,7}$	$C_{7,8}$
n^{th} order spherical aberration	0			$C_{3,0}$*		$C_{5,0}$		$C_{7,0}$
n^{th} order axial coma	1		$C_{2,1}$*		$C_{4,1}$		$C_{6,1}$	
n^{th} order star aberration	2			$C_{3,2}$*		$C_{5,2}$		$C_{7,2}$
n^{th} order three-lobe aberration	3				$C_{4,3}$		$C_{6,3}$	
n^{th} order rosette aberration	4					$C_{5,4}$		$C_{7,4}$
n^{th} order pentacle aberration	5						$C_{6,5}$	
n^{th} order chaplet aberration	6							$C_{7,6}$

Defocus is shown in Table 3 as an aberration but, just as at the microscope, this is treated as a free parameter and is used as the vertical axis in the following figures. In the following sections amounts of the aberration under investigation were included that are large enough to exhibit the behaviour of interest but not so large they are unrepresentative of the types of real probes often formed experimentally in the aberration-corrected STEM. The reader may also wish to compare these figures to the real-spaces probe images concisely presented by Uno [71].

3rd Order Spherical Aberration – $C_{3,0}$

Figure 46 shows the simulated three-dimensional optical transfer response for an otherwise perfect probe with +50 microns of third-order spherical-aberration ($C_{3,0}$) present. This is the 3D equivalent to the 2D plot shown in Figure 44 (right).

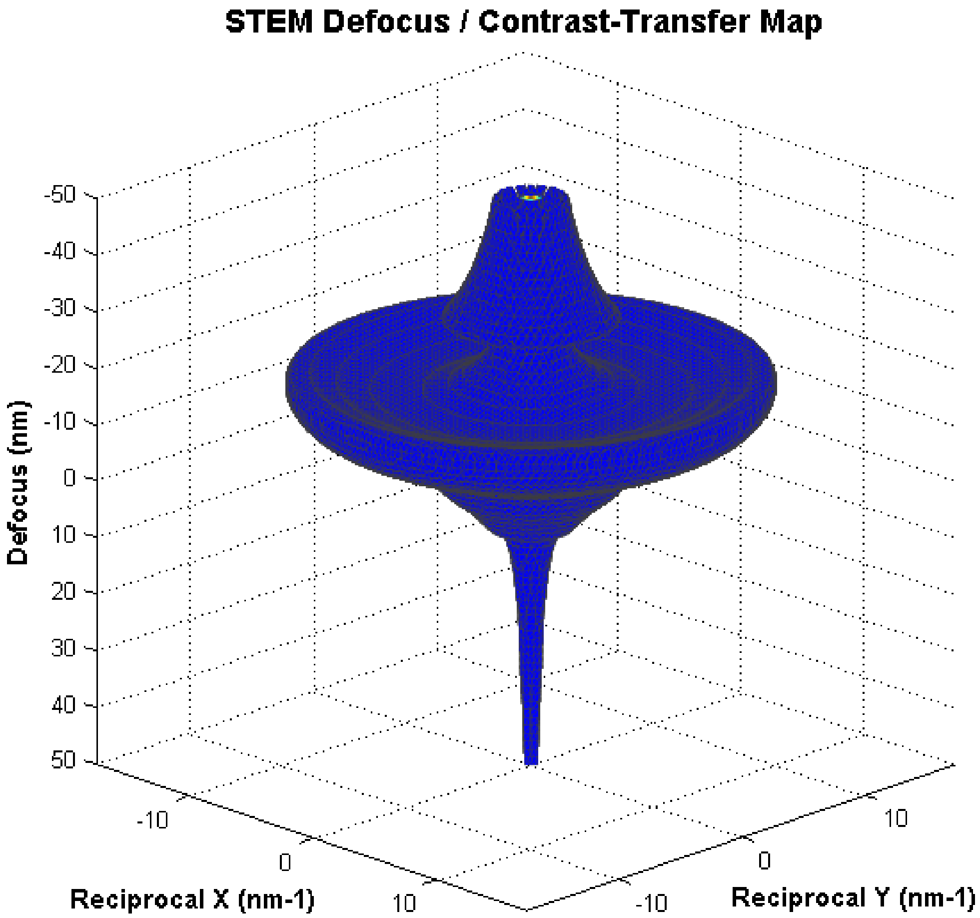


Figure 46. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +50µm $C_{3,0}$. Bounding surface shows 10% contrast level .

In this case the aberration simulated, and hence the optical-transfer response, is fully round about the z axis. As a result, in this instance the 3D representation does not differ in form from the 2D plot so both can be discussed in parallel.

Just as was the case for the real-space probe simulation shown in Figure 42 (right), in this optical-transfer response figure, the introduction of a positive spherical aberration has broken the over-focus/under-focus symmetry. In Figure 42 this is observed as the upward sweeping wings in the right hand figure (compare with Figure 4 in [55]), the 3D equivalent of which is the upward curving 'bowl' shape in Figure 46.

We also see the appearance of a fragmented hollow cone centred around the zero-frequency core on the negative defocus region (top half) of the figure. This corresponds to the secondary wings above the main ones in Figure 44 and arises from the point-spread-function of the negatively defocused probes exhibiting a significant secondary maxima. In this iso-surface view the secondary maxima is just greater than the rendering threshold and appears fragmented even though the secondary maxima are continuous azimuthally. On the opposite side of focus the first secondary maxima to appear are far weaker and far further from Gaussian focus (see Figure 44). Again as expected the zero-frequency core persists across the whole figure.

For odd radial-order aberrations the resultant real-space probes possess an even-fold azimuthal symmetry [71], as a result we would expect the phase of the OTF to also exhibit this even-fold symmetry. This can be inspected by sectioning the complex three-dimensional OTF at some relevant defocus; Figure 47 below shows the spherically aberrated 3D OTF from Figure 46 sectioned in the X-Y plane at Scherzer defocus conditions (-12.1 nm).

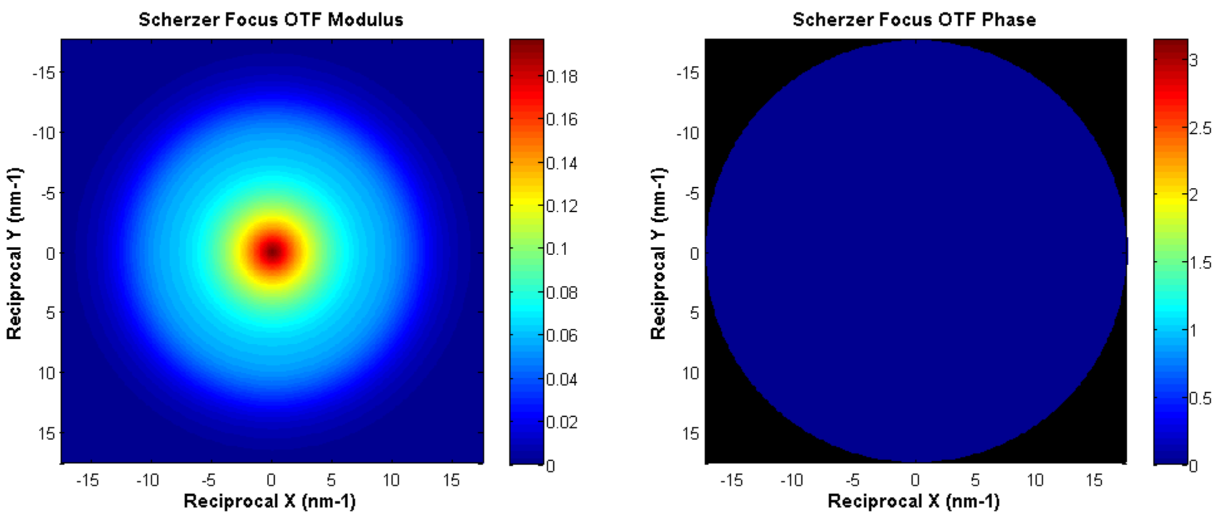


Figure 47. The modulus (left) and phase (right) of the X-Y section through the optical transfer function shown in Figure 46 at Scherzer defocus (-12.1 nm defocus). Black areas in the phase-map represent undefined values corresponding to points of zero modulus from beyond the Nyquist frequency.

1st Order Two-fold Astigmatism – $C_{1,2}$

One of the most commonly observed aberrations in both uncorrected and aberration corrected electron optics is the two-fold astigmatism. It is difficult to avoid this aberration both because of the precision needed in its correction and also due to fact that the tuning of hardware-correctors drifts over time. The drift of the corrector tuning is worse for lower order aberrations and users will be familiar with the need to retune the two-fold astigmatism many times during an operating session.

Figure 48 shows the optical-transfer response simulated with 2.5nm of two-fold astigmatism. This non-round aberration has both magnitude and direction. Here and in all following figures the orientation of the simulated aberration was zero degrees and hence the primary component always runs along the line x = 0 towards positive y.

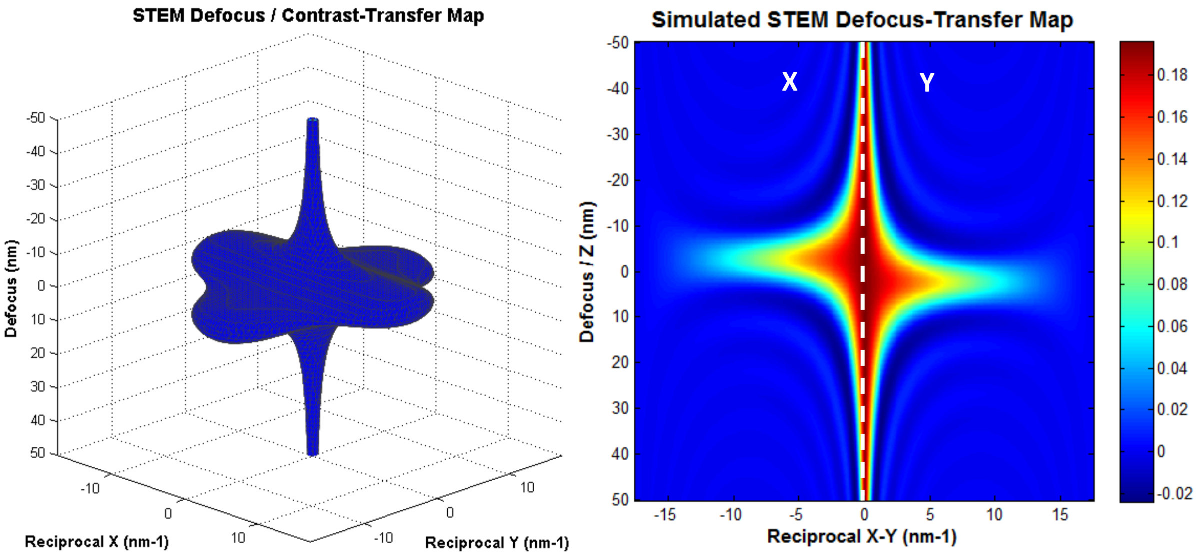


Figure 48. Left, the 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +2.5nm $C_{1,2}$. Bounding surface shows 10% optical level. Right, sections through the 3D OTF in the X-direction (left half) and at 90° from this, in the Y-direction (right half).

As expected this non-round aberration yields a non-round optical-transfer response. What was a round flat central disk in the perfect probe (Figure 45) now appears as two lozenges at ninety degrees from one another and offset with respect to defocus. The shape of this optical transfer response can be explained easily if one recalls that in this spatial-frequency visualisation, points far from the centre laterally represent directions with better resolution – conversely where the transfer bound is closer to the centre (such as at 45°, 135° and so on) the resolution is poorer.

In the presence of a two-fold astigmatism the real-space probe is widened – in one direction over-focus and widened in the perpendicular direction under-focus. This causes two focal-planes to be defined one for each of the sets of perpendicular rays. In each of these two planes, resolution in that direction is maximised and the transfer bounds of Figure 48 are extended. The separation of these two planes is proportional to the strength of the two-fold astigmatism and in Figure 48 each of the extended lozenges of transfer is centre on ±2.5nm corresponding to the magnitude of the simulated aberration [101]. Half way between these two planes (at zero defocus) we often refer to the disk of least confusion. In this plane the transfer bounds is round and circular giving the smallest circular probe and making image interpretation easier. However this is always at a poorer resolution that the aberration free probe Figure 45. Interestingly the resolution performance in each perpendicular direction does persist – but unfortunately the two can never be recorded in a single image simultaneously.

This figure also directly illustrates the spatial-frequency dependent depth resolution. This depth-resolution is the sensitivity to variations in z-height or equivalently variations in defocus. It is an essential property for optical-sectioning in the STEM. In the case of two-fold astigmatism the depth-resolution along either of the principal directions of the astigmatism is equal to that of a perfect probe. However, considering the acquisition as a whole the apparent depth-resolution will be degraded.

As the 3D-OTF for this two-fold astigmatism remains centred at zero with respect to defocus it is convenient to inspect the modulus and phase of the OTF at this X-Y plane (Figure 49).

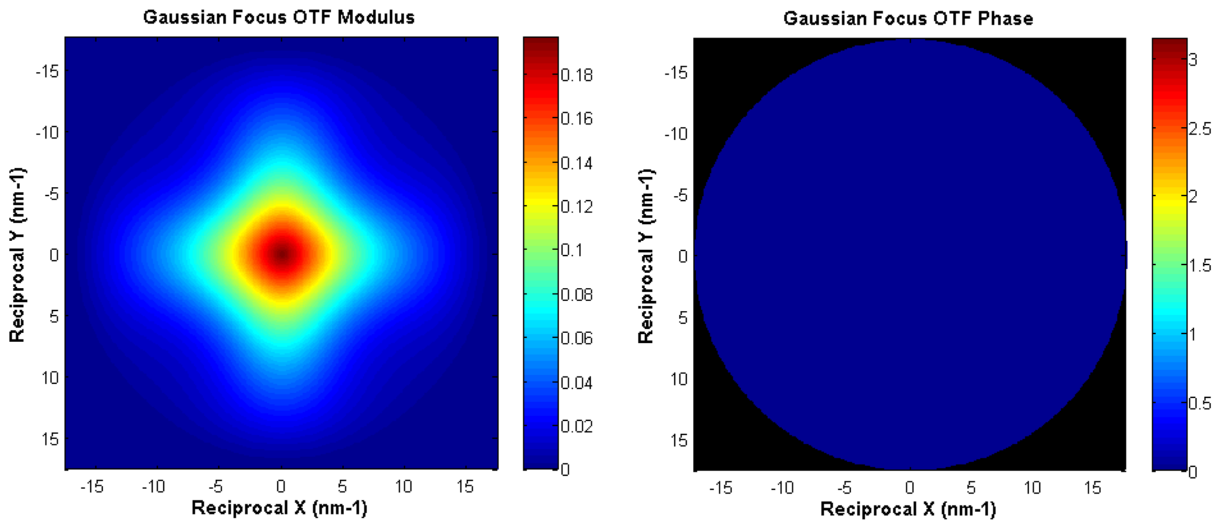


Figure 49. The modulus (left) and phase (right) of the X-Y section through the optical transfer function shown in Figure 48 at Gaussian focus (zero-defocus).

The modulus plot shown above concisely illustrates the four-fold azimuthal symmetry of the contrast transfer at Gaussian defocus. This may exist in any arbitrary orientation and when imaging a sample with a four-fold crystallography may have a greater or lesser deleterious effect.

2nd Order Three-fold Astigmatism – $C_{2,3}$

Simulated probes containing three-fold astigmatism display a somewhat triangular shape. Unlike two-fold or four-fold astigmatism the orientation of the triangular shape is the same both over and under focus [71]. Figure 50 shows the OTF from such simulated probes with 250 nm of $C_{2,3}$.

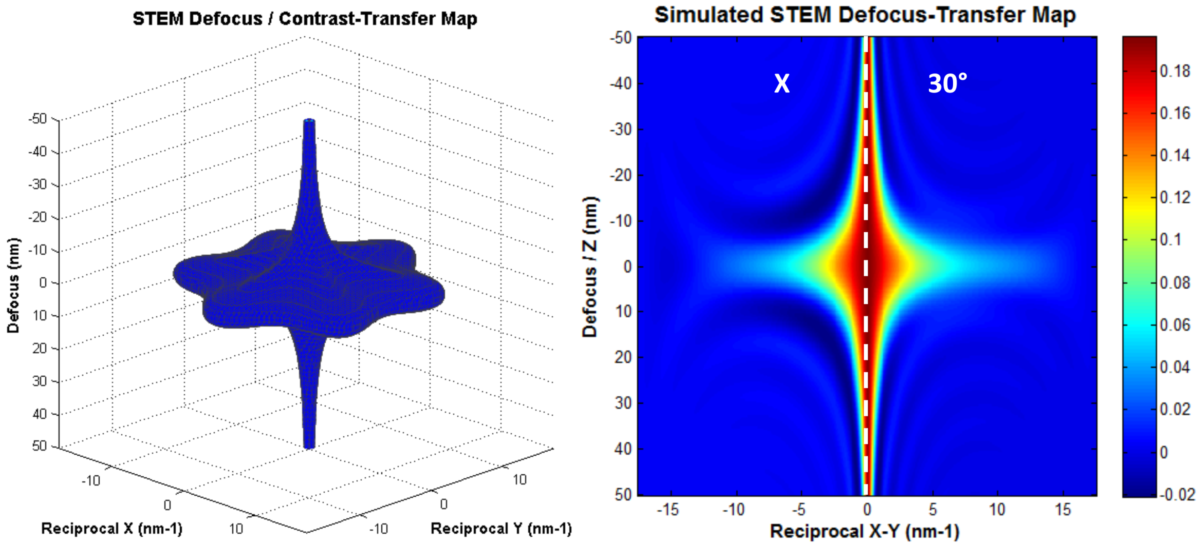


Figure 50. Left, the 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +250nm $C_{2,3}$. Bounding surface shows 10% optical level. Right, sections through the 3D OTF in the X-direction (left half) and at 30° from this direction (right half).

When this aberration is introduced into an otherwise perfect, or diffraction limited only, probe the modulus and the phase of the OTF is still symmetrical with respect to defocus (Figure 51). In the optical-transfer plot this is exhibited as the mid-plane of the central disk remaining symmetrical with respect to z. The lateral extent of the central disk though takes on a six-fold symmetry. This arises from Friedel's law which states that the Fourier transform of a real quantity has the same amplitude at diametrically opposite points, and which necessarily imparts a two-fold symmetry to the three-fold symmetry of the real-space probe. We can also see from Figure 50 that the occurrence of the optical-transfer minima also varies significantly. The left-hand OTF section shows the transfer behaviour along the X-direction and exhibits significant negative transfer at around 10 nm either side of Gaussian focus. In the right-hand of the OTF section the case thirty degree from this is shown. Now the lateral resolution is extended and the negative transfer wholly eliminated. This remarkable non-roundness of behaviour is worth bearing in mind, especially when imaging samples with inherent six-fold symmetry at high resolutions.

As this aberration is non-round and produces a triangular real-space probe at both under- and over- focus [71], it is in this case interesting to examine the phase of the OTF at its mid-plane (Gaussian focus), Figure 51.

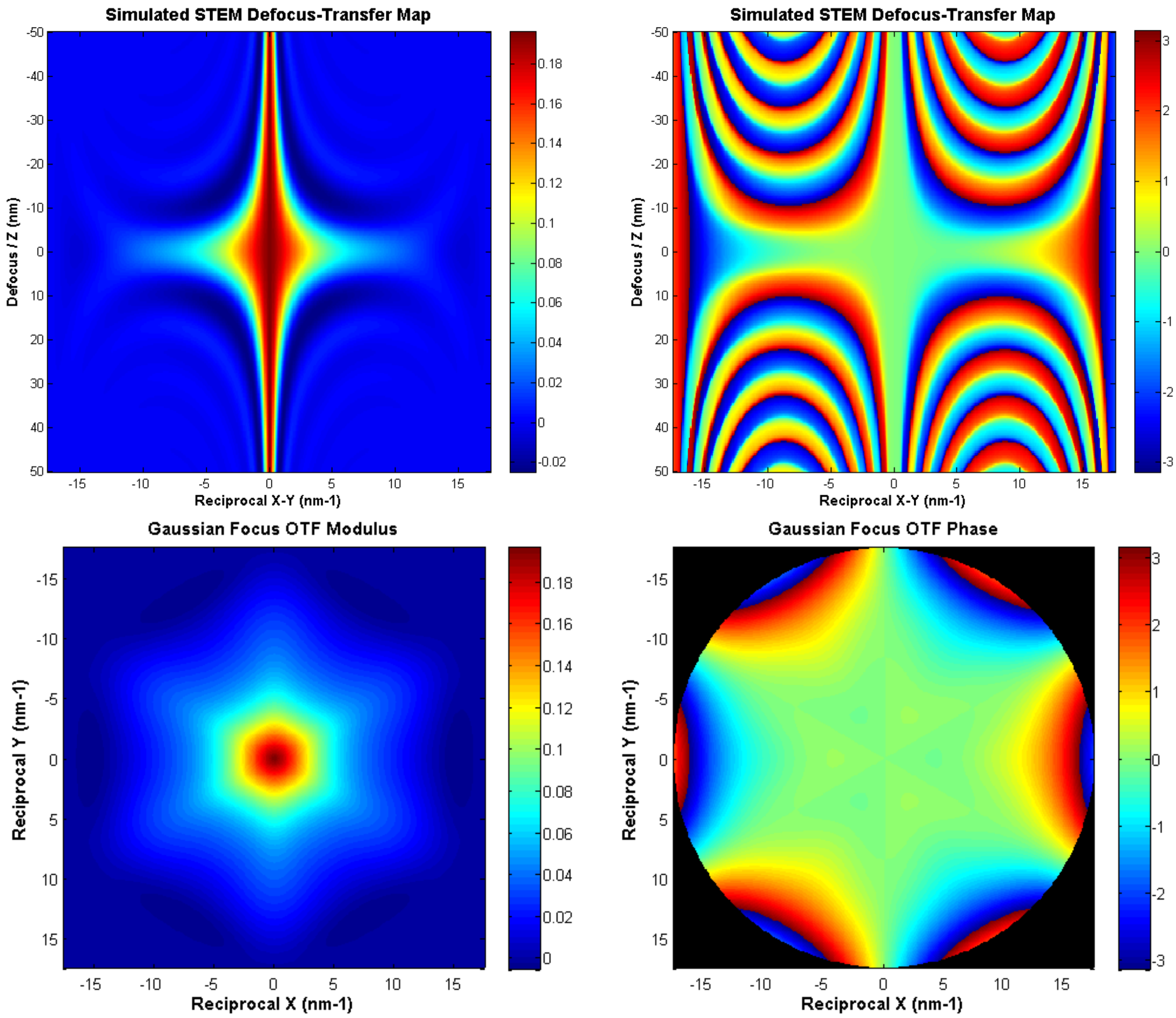


Figure 51. Sections through the complex OTF shown in Figure 50 in the X-Z plane (top) and the X-Y plane(bottom) separated into its modulus (left) and phase (right) components.

In the phase map shown in Figure 51, the three fold-symmetry of the OTF is revealed. This is the source of the triangular appearance of recorded image features and together with the 6-fold symmetry of the modulus plot, allows for the potential diagnosis of three-fold astigmatism from STEM focal-series. Furthermore, from the X-Z section phase-map we can see that the phase varies only slowly within the first few nanometres either side of Gaussian defocus. This region matches the focal-range of strong contrast transfer in the modulus plot and indicates the defocus-region from which phase-consistent image data may be restored from focal-series.

3rd Order Four-fold Astigmatism – $C_{3,4}$

Figure 52 shows the optical-transfer response of a STEM probe suffering from 25 microns of four-fold astigmatism.

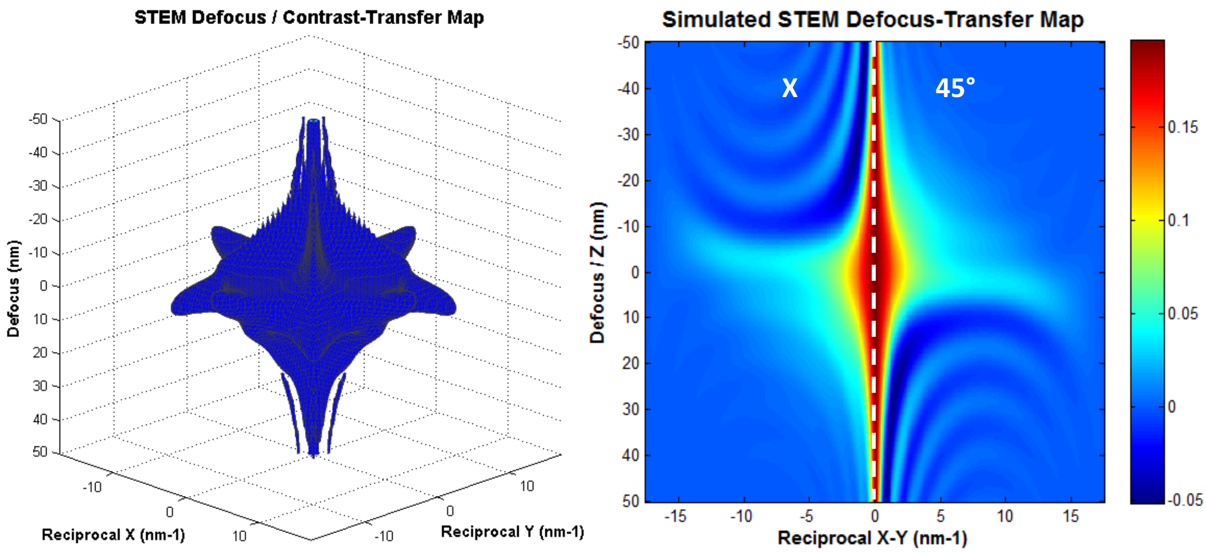


Figure 52. Left, the 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +25µm $C_{3,4}$. Bounding surface shows 10% optical level. Right, sections through the 3D OTF in the X-direction (left half) and at 45° from this, in the XY-direction (right half).

The real-space probe appears square either side of focus [71], but unlike three-fold astigmatism its orientation does rotate through focus (45° in this case). This gives the normally hourglass shaped STEM probe the appearance of two square pyramids touching at their tips. After conversion to the frequency domain in the lateral dimensions to generate the optical transfer function this creates a figure with the appearance of two intersecting square-based pyramids who intersect near their bases. The four tips of these square bases also appear to sweep upwards/downwards with respect to defocus (similar to the $C_{3,0}$ case) and this can be more clearly seen in the OTF sections shown, Figure 52. The left-hand of the OTF section shows the upward sweeping of the optical-transfer strength bounded closely on the negative defocus side by negative transfer strength. The right hand panel shows the opposite; with the transfer sweeping down (towards positive defocus). This opposing behaviour is the combined result of the rotation of the real-space probe either side of defocus, coupled with the third-order radial behaviour of the aberration.

This 45 degree rotation of the orientation of the real-space probe either side of zero-focus also has a very deleterious effect on the depth resolution available for optical sectioning. Figure 52 shows that, for say five reciprocal nanometres lateral feature spacings, there is appreciable contrast transfer from -20 nm to +20 nm.

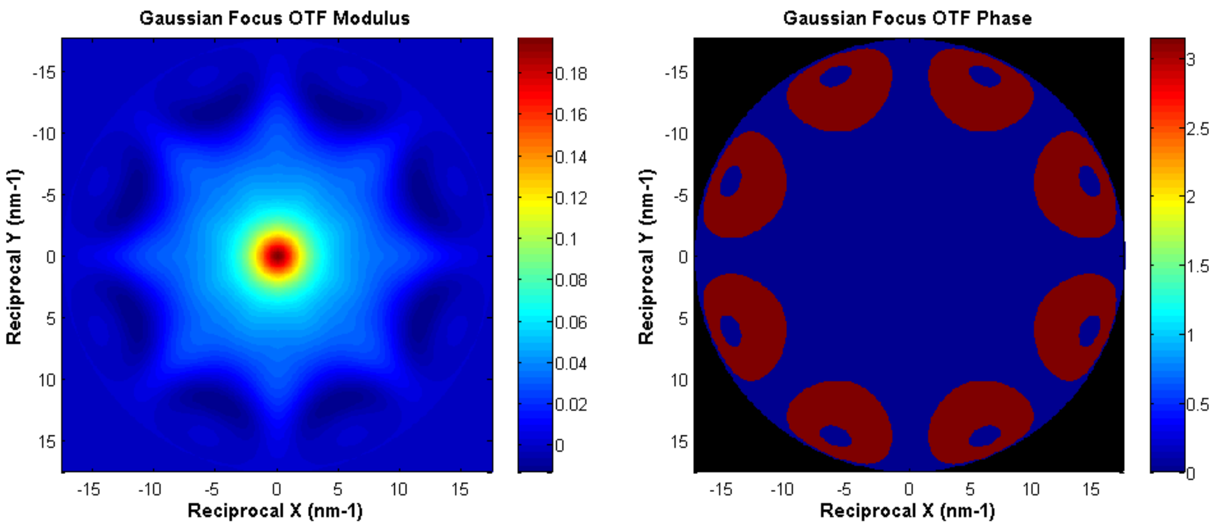


Figure 53 The modulus (left) and phase (right) of the X-Y section through the optical transfer function shown in Figure 52 at Gaussian focus (zero-defocus). Here we can readily correlate regions of negative transfer-modulus in the left figure with areas of π phase change (anti-phase) in the right figure.

Figure 53 also illustrates that the practical limit to lateral resolution with such a probe is reduced to approximately ten reciprocal-nanometres (≈ 1 Å in real-space), significantly less than the diffraction limited case. The cause of this lies in the alternating positive and negative modulus transfer (destructively interfering contrast) that lies beyond that point.

5th Order Six-fold Astigmatism – $C_{5,6}$

The six-fold astigmatism aberration is one that frequently persists as the residual aberration in hardware corrected STEM instruments. Figure 54 shows the optical transfer response from a simulated probe with 50 mm of six-fold astigmatism.

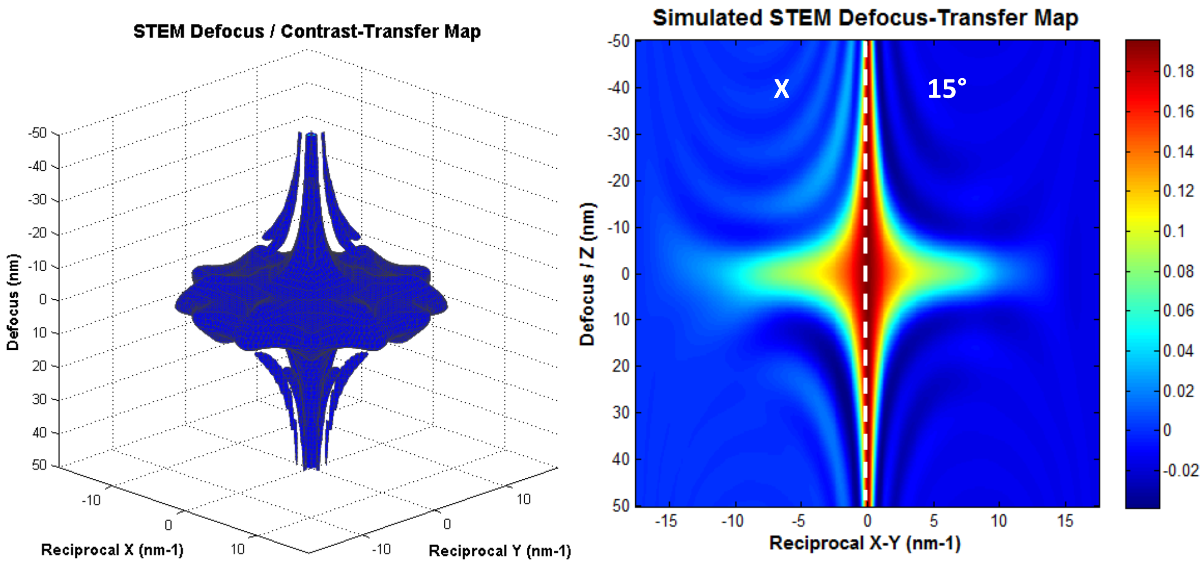


Figure 54. Left, the 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +50mm $C_{5,6}$. Bounding surface shows 10% optical level. Right, sections through the 3D OTF in the X-direction (left half) and at 15° from this direction (right half).

Similar to four-fold astigmatism the shape of the real-space probe in six-fold astigmatism at over/under focus exhibits a rotation (although here by 30°). The effect is the same with the optical-transfer function being shaped like two intersecting hexagonal-pyramids. At zero-defocus the overall lateral bounds of the transfer function adopt a twelve-fold symmetry (six-fold otherwise); this approaches a good approximation to a circularly symmetric response where all directions of spatial frequency are transferred equally. This explains how it can be acceptable for the STEM tuning to be six-fold astigmatism limited at Gaussian focus (Figure 54 OTF section mid-plane).

The OTF sections again reveal a difference in both the lateral extent of usable optical-transfer and also significant differences in the behaviour of secondary-maxima and negative regions of optical transfer.

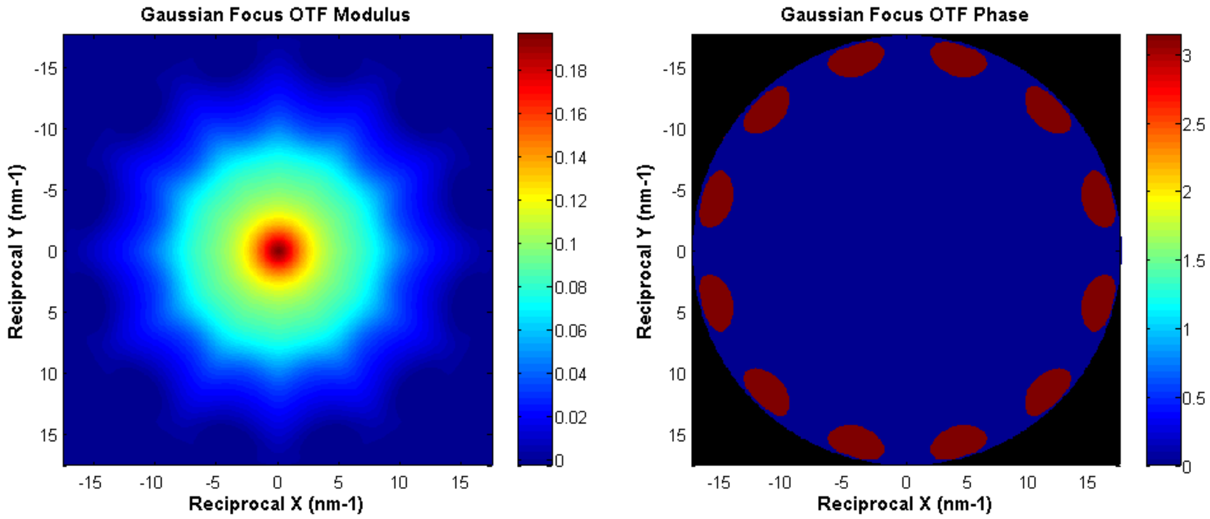


Figure 55. The modulus (left) and phase (right) of the X-Y section through the optical transfer function shown in Figure 54Figure 48 at Gaussian focus (zero-defocus).

In Figure 55 the regions of negative modulus (or equivalently π phase-shift) are localised more to the periphery than in Figure 53. This, coupled with the higher azimuthal symmetry of the modulus plot (12-fold), yield a probe that appears increasingly circular. In fact with a 12-fold symmetric modulus, detecting this limiting aberration from real-space images from samples with either a four-fold or six-fold intrinsic symmetry would not be possible.

2^{nd} Order Axial Coma – $C_{2,1}$

Second-order axial-coma can be easily introduced if any slight mis-alignment exists in the electron beam's trajectory through a hardware aberration corrector and is frequently observed experimentally. This is again corrected manually by the operator by, for example, ronchigram tuning. Figure 56 shows the optical-transfer response for a simulated probe with 150 nm of this axial coma.

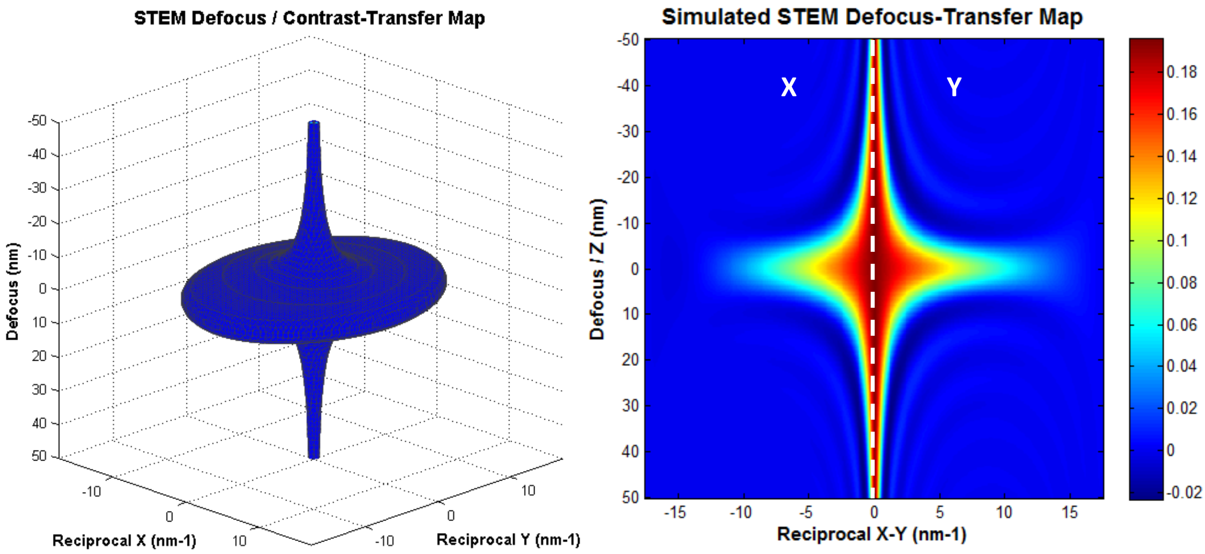


Figure 56. Left, the 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +150nm $C_{2,1}$. Bounding surface shows 10% optical level. Right, sections through the 3D OTF in the X-direction (left half) and at 90° from this, in the Y-direction (right half)

Friedel's law imparts a necessary two-fold rotational-symmetry on the modulus of the Fourier transform. This causes the 'one-fold' axial-coma to yield a two-fold transfer modulus iso-surface as shown above. In the simulated probe used to generate Figure 56 the orientation of the axial-coma was towards the positive x direction (in the plane of y = 0). Comparing this figure with that of a perfect probe (Figure 45) we can see that the shape of the transfer-function has only contracted with respect to x-direction spatial frequencies. Spatial frequencies along x = 0 have not be affected and this directly corresponds to the shape of the real-space probe not being degraded along that direction.

Again as coma yields a probe with no change in its 2D orientation at over/under focus the defocus sensitivity remains centred, and is symmetrical about, the plane z = 0. Defocus resolution is also degraded along the plus/minus x direction but not along the line x = 0. The sections through the OTF also shows the first-minima of negative transfer appears very similarly in the X and Y directions, Figure 56.

Friedel's law does not however enforce a two-fold azimuthal symmetry on the phase of the OTF (only that it be conjugate). This can be confirmed by examining the phase maps from vertical and lateral slices through the complex 3D OTF, Figure 57.

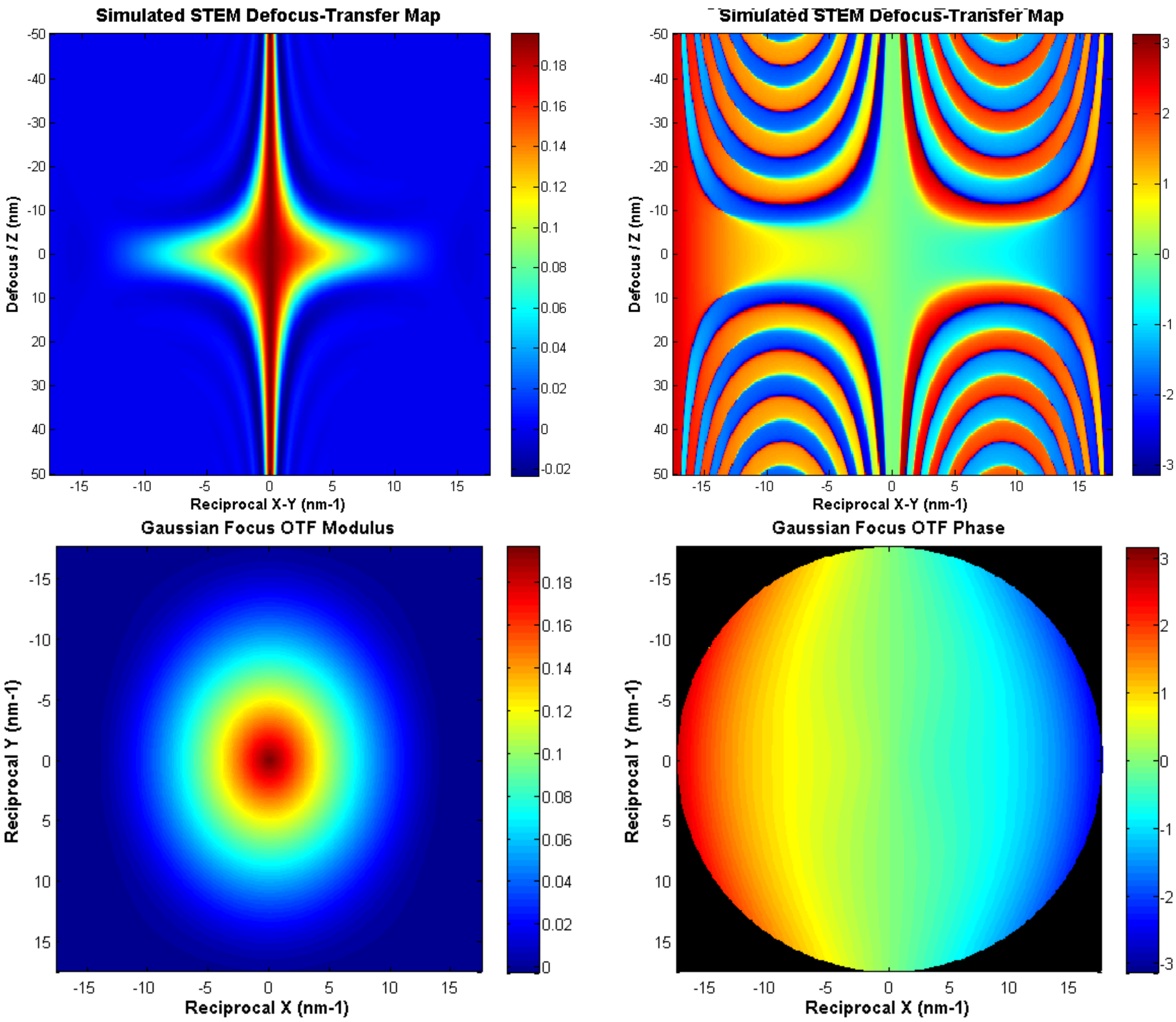


Figure 57. Sections through the complex OTF shown in Figure 56 at in the X-Z plane (top) and the X-Y plane (bottom) separated into its modulus (left) and phase (right) components.

Similar to the case of three-fold astigmatism, the even-fold symmetry of the OTF modulus iso-surface shown in Figure 57, is betrayed by the Gaussian focus phase-map. This directionality to the phase surface in the OTF is expected to persist in experimental STEM data and may be diagnosed from inspection of the phase of image's Fourier transforms. Post-processing of recorded images for the detection and compensation of small residual coma should then in theory be possible as part of a STEM focal-series reconstruction technique.

3rd Order Two-Fold Astigmatism – $C_{3,2}$

Finally the occasionally observed, but frequently present experimentally, third-order two-fold astigmatism. In tuning, from say a Ronchigram or crystal image, this can appear to have similar properties to the first-order two-fold astigmatism. That is to say that there is a loss of resolution in one direction on one side of focus, and an equal loss of resolution at ninety-degrees to that direction on the other side of focus. The difference that a user could observe, although practically this is very difficult, is that this loss of resolution (or the elliptical eccentricity of crystal features) is not linear with focus but goes with defocus cubed. It is this similar appearance and difficult observation that make this aberration one of the more frequently present and incompletely corrected aberrations experimentally.

Figure 58 shows the optical-transfer response for a simulated probe with 20 µm of third-order two-fold astigmatism.

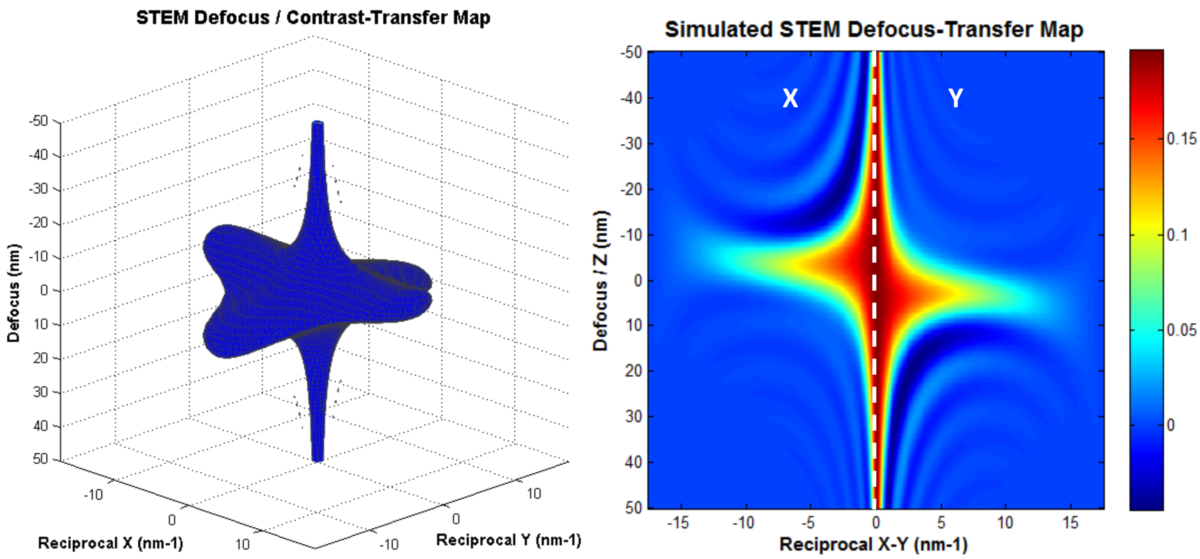


Figure 58. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +20µm $C_{3,2}$. Bounding surface shows 10% optical level. Right, sections through the 3D OTF in the X-direction (left half) and at 90° from this, in the Y-direction (right half).

It is useful to compare this figure with Figure 48. In the case of the first-order two-fold astigmatism the lozenges of strong optical transfer in the perpendicular were simply offset from one another in defocus. Here however, the lozenges are both offset but also curving away from one another (the curvature now more similar to the $C_{3,0}$ case). We can also inspect this similarity of behaviour in the X-Y OTF sections, Figure 59.

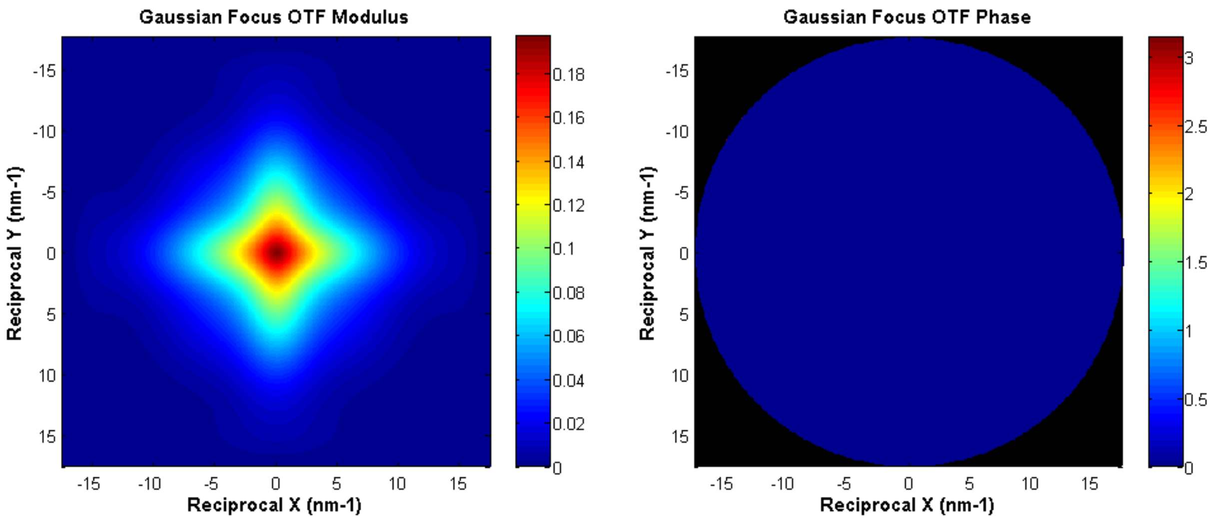


Figure 59. Sections through the complex OTF shown in Figure 58 at in the X-Z plane (top) and the X-Y plane (bottom) separated into its modulus (left) and phase (right) components.

The similarities of the appearances of the first and third order astigmatism OTFs (Figure 49 and Figure 59) demonstrates how easily these two aberration can be confused when tuning or when analysing images, it is also a reason why caution should be advised when attempting to correct arbitrary crystal feature eccentricity in image post-processing by deconvolution.

2.2. Using Simulated STEM OTFs to Plan Focal-series Experiments

When planning focal-series experiments in the STEM, for say optical-sectioning [101], [116], to ensure that the series captures the full behaviour of interest, it is desirable to begin and end the series with the sample just out of focus. Equally focal-steps should be spaced close enough that any subtleties in the OTF behaviour are not missed. However, the OTFs presented thus far have illustrated each of several different aberrations individually, and in practice a mixture of aberrations will be present with a variety of strengths and orientations.

While for focal-series experiments the total focal-range is important, simply recording finely stepped images over a wide range of focus will require considerable imaging time. In the presence of specimen drift this may take the region of interest out of the field-of-view, cause contamination deposition, or for beam-sensitive materials may change the sample entirely via beam (radiation) damage.

For optimal planning of focal-series experiments then, the total focal-range needed should be known in advance; the spacing of the focal-steps can then be determined by the dose-tolerance of the sample. Calculating this from probe-simulation (using the readout from the hardware-corrector tuning) can avoid having to determine this by trial and error and can help keep the dose introduced to the sample to a minimum.

To evaluate this, an experimental hardware-corrector readout was taken following aberration tuning on a JEM 2200MCO operated at 200 kV with a 22 mrad objective aperture (using a standard reference sample of gold balls on amorphous carbon). In addition reasonable estimates were used for source size, chromatic aberration and high-voltage ripple. The values used are shown in Table 4.

Table 4. Aberration values and microscope parameters from experimental read-out.

Aberration Notation	Value	Aberration Name
$C_{1,2}$ (A1)	1.63 nm @ 66.7 °	2-fold astigmatism
$C_{2,1}$ (B2)	40.46 nm @ 80.3°	Axial coma
$C_{2,3}$ (A2)	63.79 nm @ -161.7°	3-fold astigmatism
$C_{3,0}$ (C3)	4.135 µm	Spherical aberration
$C_{3,2}$ (S3)	4.429 µm @ -11.3°	3rd-order 2-fold astigmatism
$C_{3,4}$ (A3)	2.196 µm @ 106.3°	4-fold astigmatism
$C_{5,0}$ (C5)	-3.271 mm	5th order spherical aberration
$C_{5,6}$ (A5)	1.552 mm @ 84.5°	6-fold astigmatism
Source Size	10 pm *	*after demagnification
Source Energy Spread	0.3 eV	Intrinsic source energy spread
C_c	1.3 mm	Chromatic aberration
HT-ripple	1 ppm	Parts-per-million energy stability

Entering this set of relatively arbitrary aberrations and microscope parameters into the probe-simulation code yields the transfer volume shown in Figure 60.

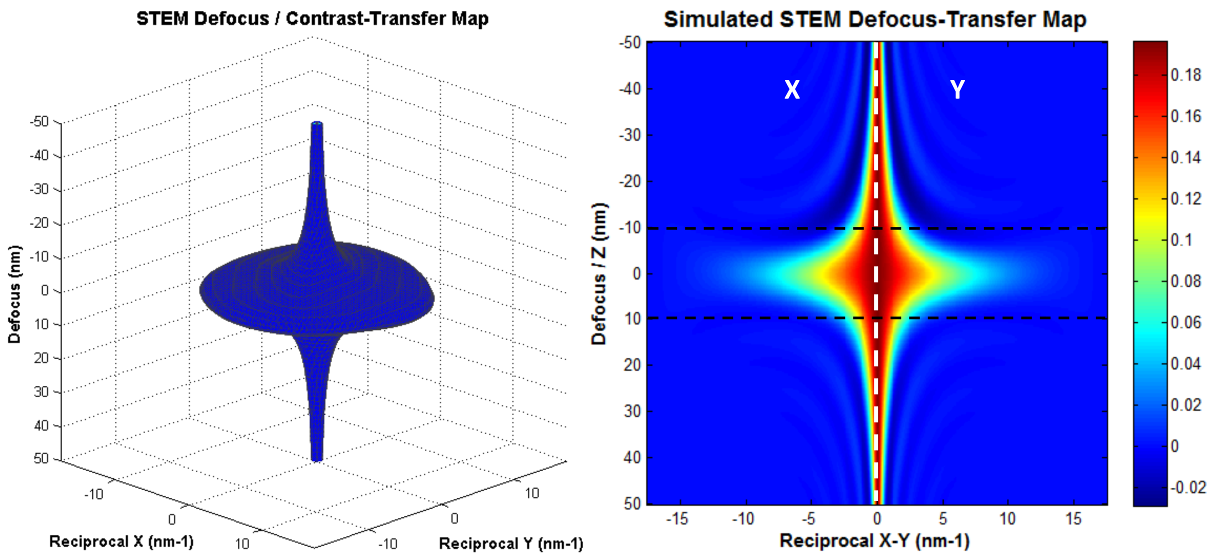


Figure 60. Left, the 3D optical-transfer versus defocus iso-surface for a simulated probe with a 'real' mixture or aberration strengths, orientations and other microscope properties. Bounding surface shows 10% optical level. Right, sections through the 3D OTF in the X-direction (left half) and at 90° from this, in the Y-direction (right half).

Figure 60 gives an intuitive and direct readout of the required focal-range required to entirely capture the defocii that exhibit high-resolution spatial information. The lateral optical transfer volume here extends to roughly 12.5nm^{-1} equivalent to a potential information-limit (in the absence of instabilities / contamination) of approximately 0.8Å. The total spread of defocus over which lattice resolution may be expected (resolution better than four reciprocal nanometres or 2.5 Å) was roughly 20 nm (Gaussian focus ±10 nm) and is indicated by the dashed black lines in Figure 60. As this probe was relatively well tuned the overall appearance of the OTF is largely circularly symmetric (azimuthally) and defocus symmetric (flat).

3. Quantifying the Quality and Recording Performance of STEM Images

In Section 2 some example optical transfer functions were presented. However, these OTFs describe the ultimate ceiling to imaging performance that could be achieved with the microscope, and not necessarily the performance that *will* be delivered.

In any recording system with only finite illumination current, random noise from counting statistics must be considered [49], [117]. Additionally, the STEM instrument can be highly sensitive to environmental or instrumental disturbances such as acoustic, mechanical or electromagnetic interference. This interference can introduce distortions to the images recorded and degrade both signal-noise and resolution performance [29], [31], [118]. Finally, sample or stage drift can cause the images to appear warped and leads to unreliable lattice angles and spacings being exhibited.

In this section tools to quantify the signal-noise ratio (SNR), resolution and drift present in STEM images will be discussed. Once those tools are in place, the image artefacts arising from instability-induced scan-distortion will be discussed both in terms of how they may be diagnosed and subsequently compensated. To compensate for scan-distortion image processing software was written and throughout this chapter the performance of the output of this software will be evaluated quantitatively.

3.1. Quantifying STEM Image Quality

The three most commonly quoted performance metrics when scrutinising STEM images are signal-noise ratio, resolution and sample/stage drift. The SNR of an image affects the clarity of features within images, as well as the ability to reliably observe weak features. The resolution of images affects the minimum ‘close-ness’ of features that can be reliably identified as discrete /

separate from one another and also the degree to which information (imaging / analytical signals) is localised onto a specific atomic column and not diffuse / spread across neighbouring columns. Lastly, the amount of drift present in an image affects the angles between, and spacings of, features that can be measured from an image. As a result having means to measure these three factors affecting image precision, accuracy and fidelity are very important. These metrics can equally be measured in real-space or in reciprocal-space (in Fourier transforms of image data); however, Fourier-space measurements of noise-levels and of the reliable identification of high-spatial-frequency spots are not readily interpretable and are a potential source of errors. For this reason the methods presented in the subsequent sections are all based on real-space approaches measured directly from the raw images. As a result of this, these techniques are only relevant to lattice resolution images.

3.1.1. Robust Peak Finding in STEM Data

The measurement of the three metrics (SNR, resolution and drift) each begins similarly, with an automated peak finding routine that determines the locations of all the atomic columns within an image. It is possible to identify the feature-positions manually (for example, by an operator clicking with a computer mouse), however as images may contain hundreds or thousands of atomic-columns this rapidly becomes impractical for high-throughput materials science studies. Once all the atomic columns are located, lattice planes are chosen to locate line profiles for measuring the SNR, the angles between these lattice planes are used to determine drift and finally each feature is in turn inspected to determine resolution.

Before describing the method developed in these studies it is worth a brief consideration of the various peak-finding approaches that have been described in the literature including:

(where >> denotes sequential processing steps)

- Wiener-filtering >> Bragg-filtering >> band-pass-filtering >> template cross-correlation [119],
- Bragg-filtering >> 8-connected local-maxima search [120], and
- Image-smoothing >> image-deconvolution >> ‘intuition’ [121].

However, each of these methods begins with a global filtering operation, which in the case of Bragg filtering, may even unwittingly enforce a periodicity onto an image and hence affect the peak positions. Additionally, in these methods the filters used have at least one tuneable parameter that must be optimised for the specific micrograph under investigation. If analysing large numbers of micrographs then tuning these variables each time a different sample or magnification is used becomes impractical, and worse, potentially subjective. Even without the use of image filtering, intensity-maxima based approaches can be misled by single-pixel random noise in the images as they do not make use of any a-priori knowledge of the expected appearance of the true image features.

Instead we can expect from the findings in section 2 and from the literature [71] some a-priori knowledge, that the shape of a relatively well tuned probe is approximately Gaussian. It is this knowledge of the expected appearance of image features that then makes manual analysis (using the immense pattern recognition powers of the human brain) such a reliable albeit time consuming technique.

In this investigation we developed an entirely new real-space peak-finding routine that does not require any filtering of the image data, works equally well for images of single-crystal, grain boundaries, dislocations, and nano-particle images (where the field of view may not be filled by the sample), and *does* make use of this 'expected appearance' knowledge. This method was designed with HAADF STEM in mind, but would work equally well in other techniques where image features are generally of a Gaussian, or some similar, form.

Other than the first few border pixels around the image edges, a square region of interest (ROI) is defined centred around each pixel in the image in turn (Figure 61). The width of the ROI should be around one atomic-spacing wide to ensure it contains the majority of any given atomic feature (precise selection of the width will be discussed later). Within this square ROI a two-dimensional (2D) circularly-symmetric Gaussian is fitted[‡].

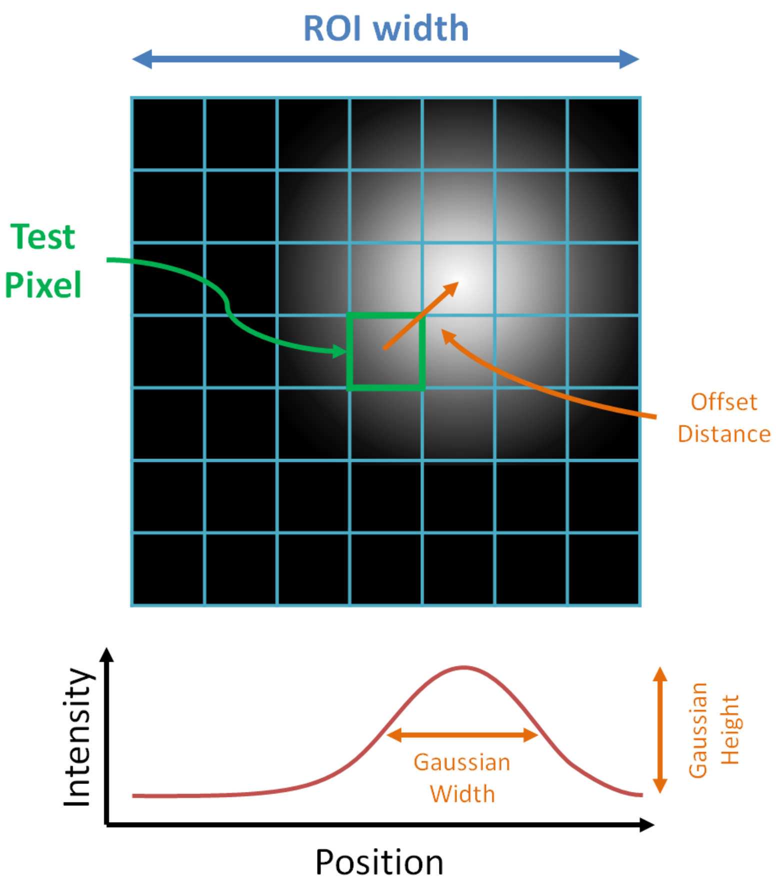


Figure 61 – Schematic of the peak-finding methodology. With an example 9x9 ROI.

The Gaussian fitted to this square ROI has three properties, all of which are useful in diagnosing genuine peak features:

1. The 'height', or magnitude, of the Gaussian,
2. The 'width', or spread, of the Gaussian, and
3. The 'offset distance', from the centre of the ROI to the centre of the 2D Gaussian.

Next we can consider the expected range of values for each of these three properties. The height of the Gaussian resulting from the fit may either be positive or negative, however as we are only interested in features that exhibit a bright-on-dark type contrast we can discount values of

[‡] Computationally, it may be faster to compute two one-dimensional Gaussians on each of the horizontally and vertically integrated values of the square region of interest. This is acceptable and yields largely identical results.

height less than zero. Equally it would be impossible for a feature to exhibit a height greater than the maximum intensity in the raw image data and so after normalising the data by the raw-image's dynamic range, any values greater than one may be discounted. Figure 62 a) shows this normalised Gaussian height evaluated at every pixel. Next, the Gaussian widths; it is again convenient to normalise these data by the width of the ROI used earlier. Any genuine features will have a width wider than single-pixel noise but narrower than the ROI width; after normalisation by the ROI size, we can then discard any values below some small value, for example 5%, or greater than one (Figure 62 b)). Finally, when the ROI is correctly located over the centre of a genuine feature we would expect the distance to the centre of the 2D fitted-Gaussian to be small. Figure 62 c) shows the offset-distance from the inspection pixel to the centre of the fitted-Gaussian; this has again been normalised by the ROI width.

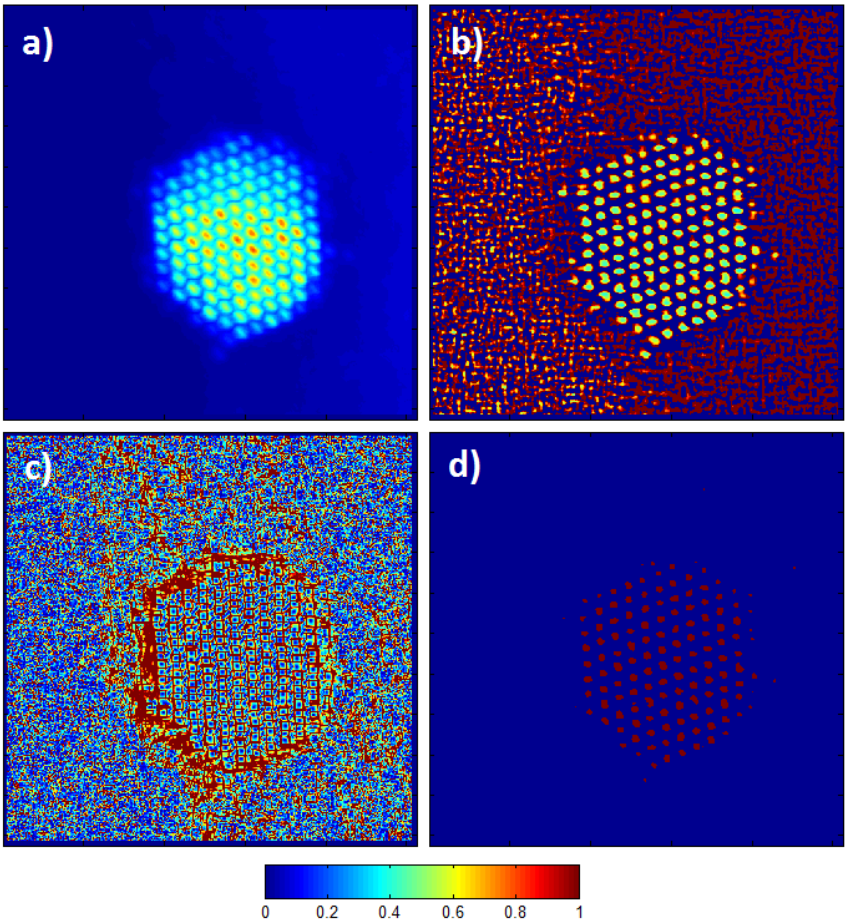


Figure 62. Peak-finding metric-figures for the HAADF data show in Figure 70. Normalised (see text) maps of the fitted Gaussian heights (a), widths (b) and peak-offsets(c). Lastly, panel d) shows the screening figure immediately before the morphological erosion to points. All figures are displayed on a false-colour scale from zero to one.

Now the time comes to combine these three metric figures. To create an overall map of pixels likely to contain genuine feature locations we can take the height-figure (Figure 62 a)), larger values more likely, and divide this (pixel-wise) by the offset-distance plot (Figure 62 c)), smaller values close to true features. This produces a new figure of merit, that is very sharply peaked, anywhere *nearby* to *positively-peaked* Gaussian fits. Screening this new figure of merit to discard locations with Gaussian widths too small or large as discussed above yields a map of regions likely to contain true features, Figure 62 d). This map contains 'blobs', usually a few pixels across centred over the image features. The final stage of the process is to perform a morphological erosion of these to reduce them to single pixel points. The coordinates of these points then serve as the coordinates of the image features (atomic columns in this case). If a precision finer than integer pixel coordinate is required an additional 2D Gaussian refinement can be performed about these coordinates. For images that exhibit a significant long-range intensity variation (sloping background or contamination) the results of this fit can creep 'uphill' with respect to intensity. To overcome this it can be useful to perform a very mild high-pass filter.

This method requires no human interaction as all the steps rely on normalisations to measureable parameters, such as the image's dynamic range. The only tuneable parameter in this search is then the width of the ROI. Whilst it is possible to estimate an optimum for this (usually just smaller than the lattice spacing) an automated method is preferable for both productivity and objectivity. The simplest means to optimise this parameter is to first try all the possible values. With an integer number of pixels either side of the central pixel of interest the width of the ROI can take the values of positive, odd, integers greater than or equal to three. Excessively large ROI sizes induce an excessively large border area and will ultimately hinder the peak finding.

If the number of features identified is plotted as a function of the ROI width, for an image of purely random noise a power-law decay would be expected. However when an image contains crystallographic structure these figures exhibit a flat plateau region over which there is little change in the number of features identified, Figure 63.

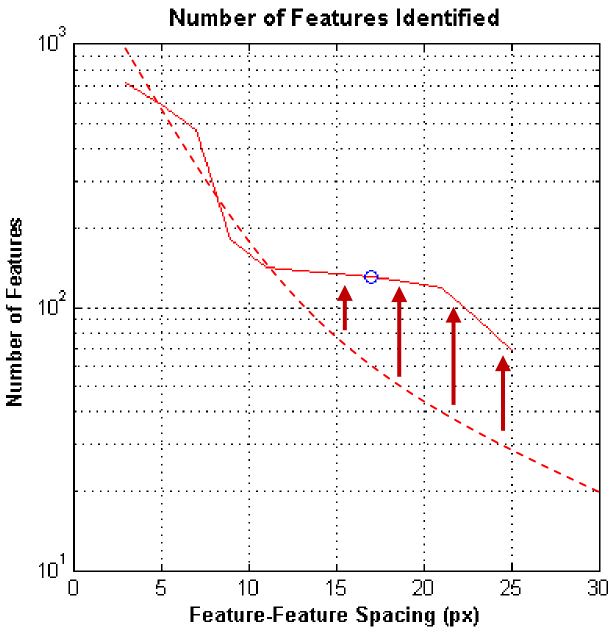


Figure 63. Number of identified features as a function of ROI width (solid line). Dashed line shows the predicted behaviour of an analysis of an image of pure noise. The persistence of the true peaks correctly identified is indicated by the red arrows. Differentiating the gradient of the solid line allows the flattest point on the plateau to be found and this is indicated by the blue circle.

The centre of this plateau (point of least negative gradient) can then be used to determine the optimum ROI width. After this the routine is run once more with this optimised condition.

3.1.2. Measuring Image Signal-Noise and Signal-Background Ratios

The most common method to inspect the signal-noise behaviour of images is by line-profiles. In the restoration code many parallel line-profiles are taken along each of several lattice-plane directions. An example intensity line-profile through the test image data shown in Figure 3 a) is presented in Figure 6.

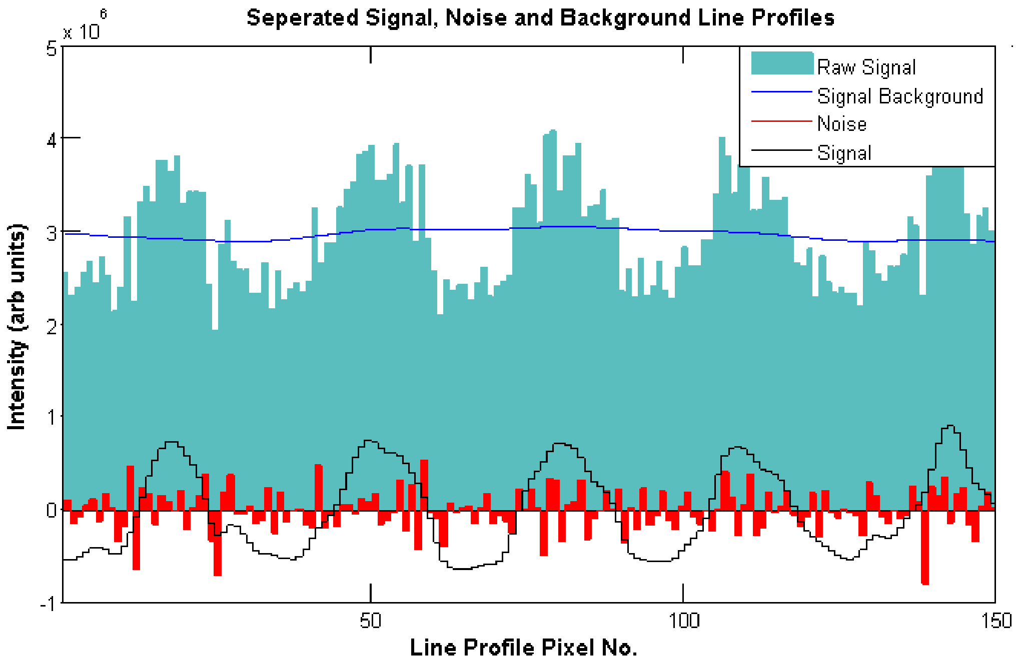


Figure 64. Example 150 pixel extract, of image line profile from the experimental test image (cyan) and its separated components; the smooth background (blue line) the structureless noise (red bars) and the remainder upon subtraction of these, the 'signal' (black).

Variations in background, the signal itself, and the high-frequency noise are found by applying a low-pass, band-pass and high-pass filter respectively. The smoothly varying background is determined by a long range smoothing that strips all local structure but retains the shape of the underlying background, (Figure 6 solid blue line through the mean of the raw signal). Here a Gaussian smoothing kernel with a full-width at half-maximum of two times the average feature spacing was found to work well, but any other suitably large value would perform similarly. This profile is then subtracted from the raw signal to leave only a combination of signal and noise. This remainder is smoothed again, this time by a far narrower Gaussian (FWHM of only three pixels) that serves only to remove single pixel noise. Subtracting this smoothed signal (Figure 64solid black line, approximately sinusoidal about the x-axis) then leaves only the noise (Figure 64 red bars either side of the x-axis). The definition only calls for the SNR to be expressed as a single ratio so the strength of these two is simply computed as the standard deviation of the two line-profiles.

Next, to evaluate the performance of this real-space SNR measurement method, simulated images were generated with additive Poisson noise, some examples of which are shown in Figure 65.

The noise was added by creating images with matched dimensions of only noise (Poisson distributed). The root-mean-squared power of these noise-images was calculated and the images added to the noise-free data in such a ratio to yield the desired target SNR.

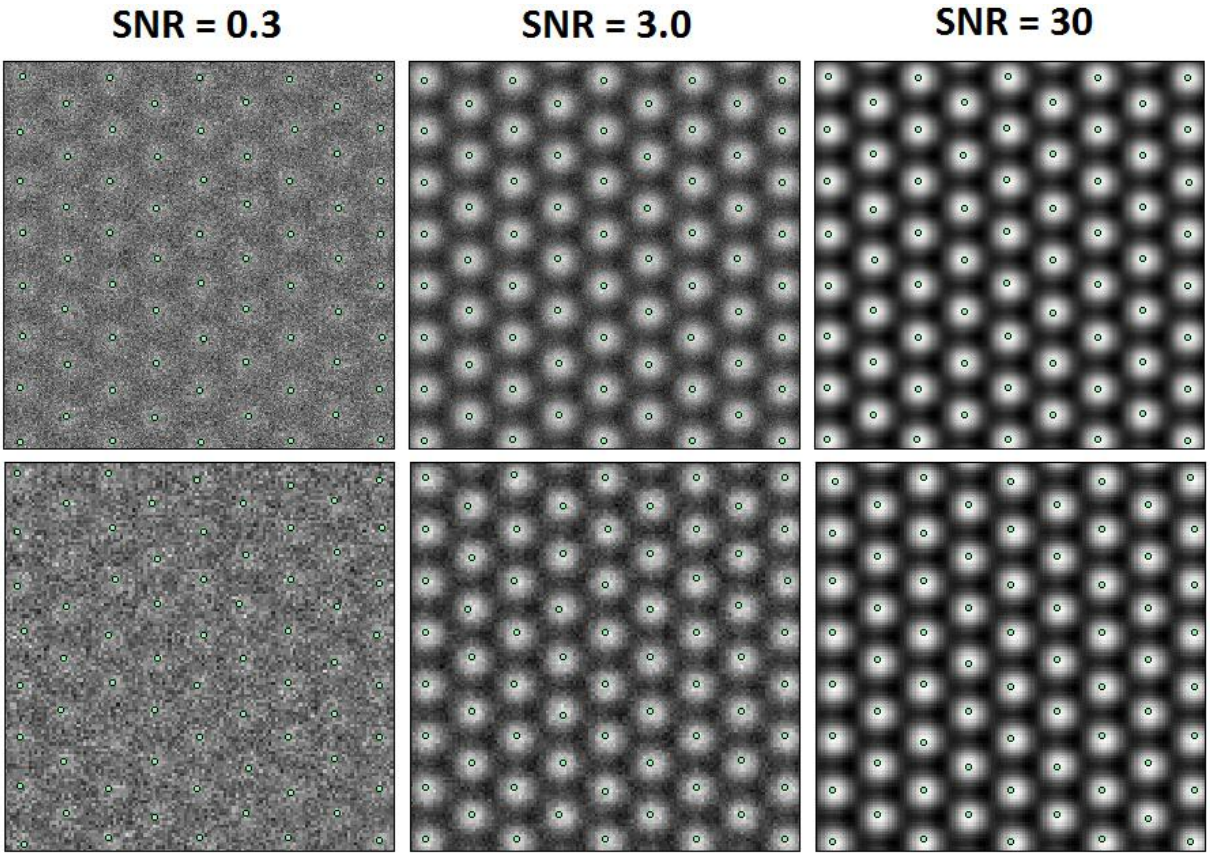


Figure 65.Simulated images with varying degrees of noise added to achieve SNRs of 0.3, 3.0 and 30. Top row shows features sampled at approximately 25 pixels per lattice spacing, bottom row at approximately 13 pixels per spacing.

In addition the sampling of the image was varied (equivalent to changing the STEM image pixel-size or magnification) to investigate pixel samplings of ≈13, ≈25 and ≈51 pixels per feature spacing. Test images with SNRs from 0.1 to 1000 were generated and the measurement results are shown in Figure 66

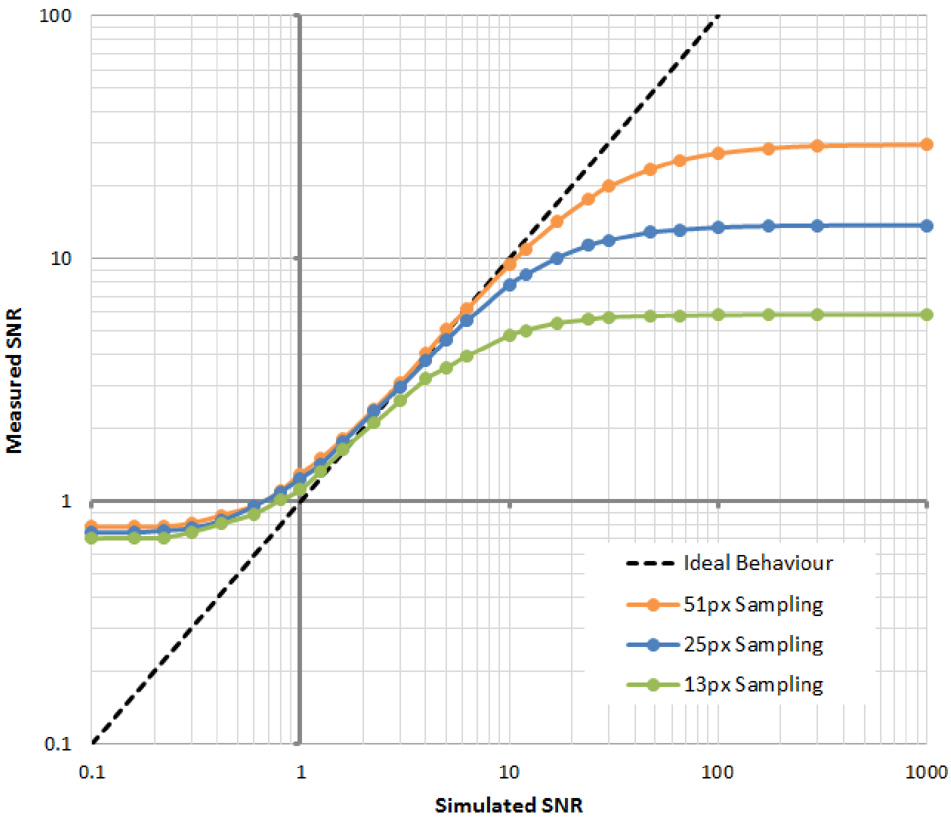


Figure 66. Simulated versus measured SNR for test images containing SNRs of 0.1 to 1000 for difference pixel samplings. The black dashed line shows the 'ideal behaviour' a perfect measurement method would yield.

Figure 66 reveals three characteristics of the performance of the SNR measurement:

- Simulated SNRs below one were not reliably measured (Figure 66 lower left plateau). The lowest SNR calculated by the analysis coded was 0.7. This is perhaps unsurprising, as at these levels the separation of true-signal from random noise becomes increasingly uncertain.
- Simulated SNRs from 1-5 (13px sampling), 1-10 (25px sampling) and 1-20 (51px sampling) are reasonably well detected (roughly linear section in the centre of each plot).
- Measured SNR values reach a ceiling that is not exceeded as the additive noise is reduced (Figure 66 top right plateau). The plateaus reached correspond to ≈5.8, ≈13.7 and ≈29.4 respectively, or around half the pixel-sampling number.

These results are unsurprising considering the nature of the real-space SNR measurement method is rooted in sets of series of band-pass filtered line-profiles. This means that single-pixel noise is more readily (and more accurately) identified the greater the difference between the Nyquist frequency of the image and the spacing of genuine features. We can also coin a general 'rule of thumb' for the experimental acquisition of data, the SNR that can be reliably measured will generally be numerically no greater than one-third of the feature spacing in pixels.

3.1.3. Measuring Image Resolution

Quantifying resolution improvement requires first an appropriate definition of resolution. This could be done by identifying spots in Fourier transforms such as Figure 76, however this would be limited to discrete jumps in performance equivalent to the lattice spacings available for Fourier analysis. Instead the approach developed here is a real-space one, based on the full-width at half-maximum (FWHM) of Gaussian profiles fitted to transects through image features. This is repeated over an angular sweep so that the shape of features is also captured (Figure 67 – red plot).

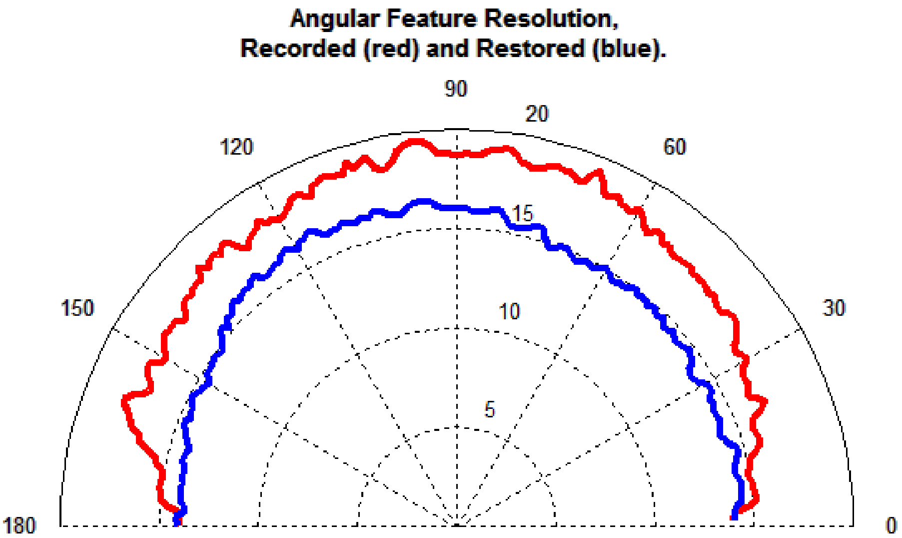


Figure 67. Example comparison of average atom widths (diameters) as a function of line-profile angle. Red line corresponds to the raw experimental data in Figure 73 a), the blue line corresponds to the resolution improvement (smaller diameter values) achieved after scan-instability and drift correction (see Section 3.2.2 and Figure 73 d)). Profiles were calculated from an average of 35 features. Shapes shown with units of pixels-FWHM, equivalent resolutions were 1.79 Å before and 1.50 Å after image restoration.

This approach has some advantages over measurement of resolution from inspecting the Fourier transform for its highest order spot. It gives an insight into the variability and improvement in the shape of features. This method is also more tolerant of noise, which can on occasion render regions of the FT unusable, and can also accommodate crystal structures with defects or any other lack of periodicity. Using this method it is still possible to average over all the angles if desired, giving a single number which can be convenient when quoting or comparing resolutions.

3.1.4. Measuring the Direction and Strength of Image Drift

It was discussed earlier that rotating the scan direction arbitrarily can lead to an underestimate of drift. Instead the method described here requires no such scan-rotation and can be applied to any recorded lattice resolution image.

Using the results from the earlier peak finding routine the lattice planes in the image can be identified and the angle between each of these measured. If we consider the test image presented earlier in Figure 3 a) we can see that {100} family of planes are not perpendicular to one another; the same is true of the $\{1\bar{1}0\}$ set. In both cases their bisection yields an angle greater than ninety degrees. These variations in lattice plane angle and spacing can also be observed in Figure 76 (top), the Fourier transforms of Figure 73 a) and d).

To calculate the direction and strength of the drift present the measured angles are compared with those expected (in this case spaced equally at forty-five degrees from one another). The difference between these angles is then plotted against the expected azimuthal angle of the planes in Figure 8.

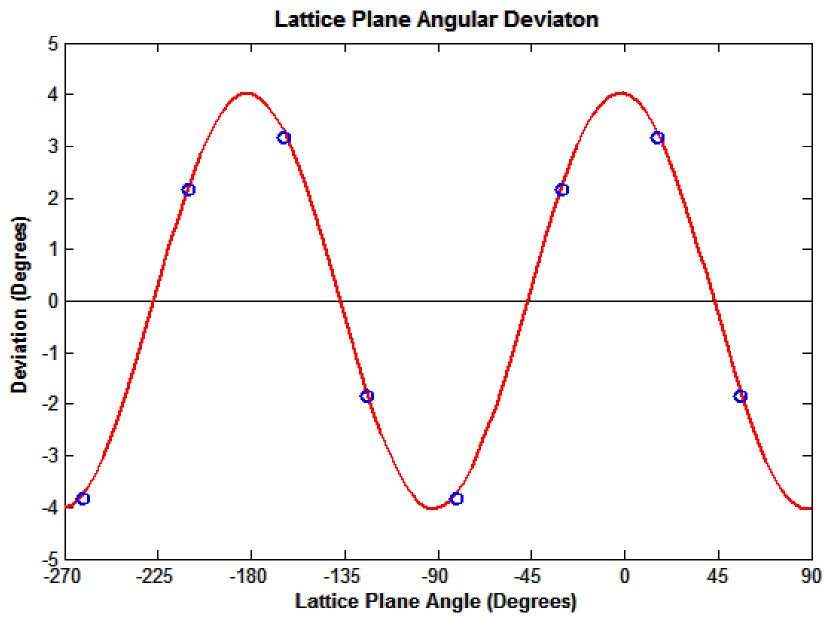


Figure 68. Lattice plane angular deviation as a function of image azimuth. Data points represent the measured deviations an expected angles of the {100} and {110} family of planes.

When this is done we find that the data points lay on a sinusoidal function with a periodicity of 180°. This periodicity necessarily arises from the 2-fold centro-symmetry of the lattice planes themselves. This fitted function then has two maxima and two minima in any given 360° cycle; these represent the directions along which the image has been purely sheared. 45° between these directions represent the angles where the image can be said to be purely stretched or compressed. The phase of the fitted function gives the direction of drift then and the magnitude gives the shear angle equivalent to the drift rate.

These measurements of drift in terms of its direction and 'equivalent pure-shear angle' are useful for image correction as they can be directly used to correct the image by a rotation and a pure-shear compensation. However it may be desirable to also quote this in terms of a physical drift rate or 'sample-speed'. We can convert to such units if we consider the example case below, Figure 69.

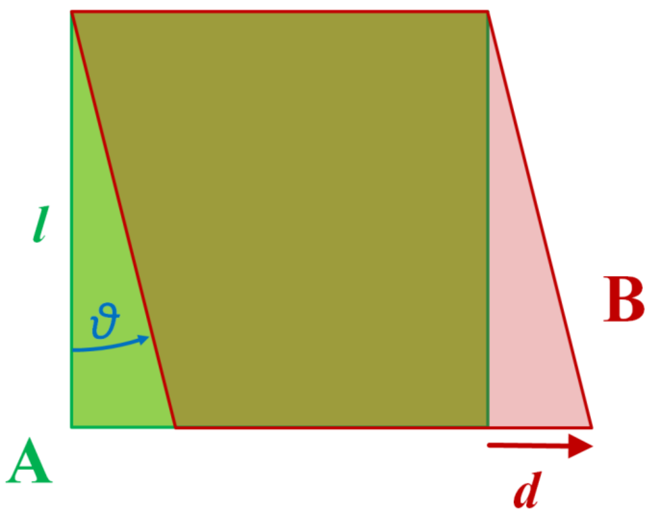


Figure 69. Example intended scan area (A) and the resulting scanned area (B) in the presence or a purely horizontal image drift of magnitude d during the acquisition.

Figure 69 shows the most simple example of an image drift purely along the horizontal direction; an experimental example of this is also shown in Figure 75. Using the method described above a shear angle of θ would be diagnosed, this can be converted into a drift rate (speed) if the image field of view (l) and acquisition time (t) are known:

$$Drift\,rate = {}^{distance}/_{time} \tag{19.1}$$

$$v = {}^{d}/_{t} \tag{19.2}$$

where:
$$d = l\,.\ \tan\vartheta \tag{19.3}$$

$$v = \frac{l\ .\ \tan\vartheta}{t} \tag{19.4}$$

For more complex drifts, such as those that change rate or direction mid acquisition, or for very severe drifts where 'small-angle' trigonometric approximations are no longer valid, caution should be taken in the calculation of absolute drifts rates; more likely any images falling into this regime will not be usable anyway.

3.1.5. Measuring the Image Pixel-size (Magnification)

Magnification in STEM is created purely digitally, that is that it is created by adjusting the field of view scanned by the probe. When recorded in to images with the same pixel dimensions, smaller scan-areas yield larger magnifications. Magnification calibration is important for precise distance measurements but also in the increasingly prevalent technique of 'quantified STEM'. In this technique the integrated fractional intensities of atomic columns must be multiplied by *pixel area* to yield a cross-section with absolute units, accurate measurement of the pixel-area is important.

It should also be noted that owing to the intrinsically serial nature of the STEM image acquisition, and to possible imprecisions in the calibration of the deflector coil pairs, there may be a small but measureable difference in the width and height of image pixels. As a result the image magnification should be calibrated in both the horizontal and vertical directions separately. To achieve this, the commonly used method of diffractogram calibration is modified to incorporate this subtlety through the use of two crystallographically equivalently but non-collinear reciprocal lattice vectors. Fortunately for most observed diffractograms two such unique points can be observed, Figure 70.

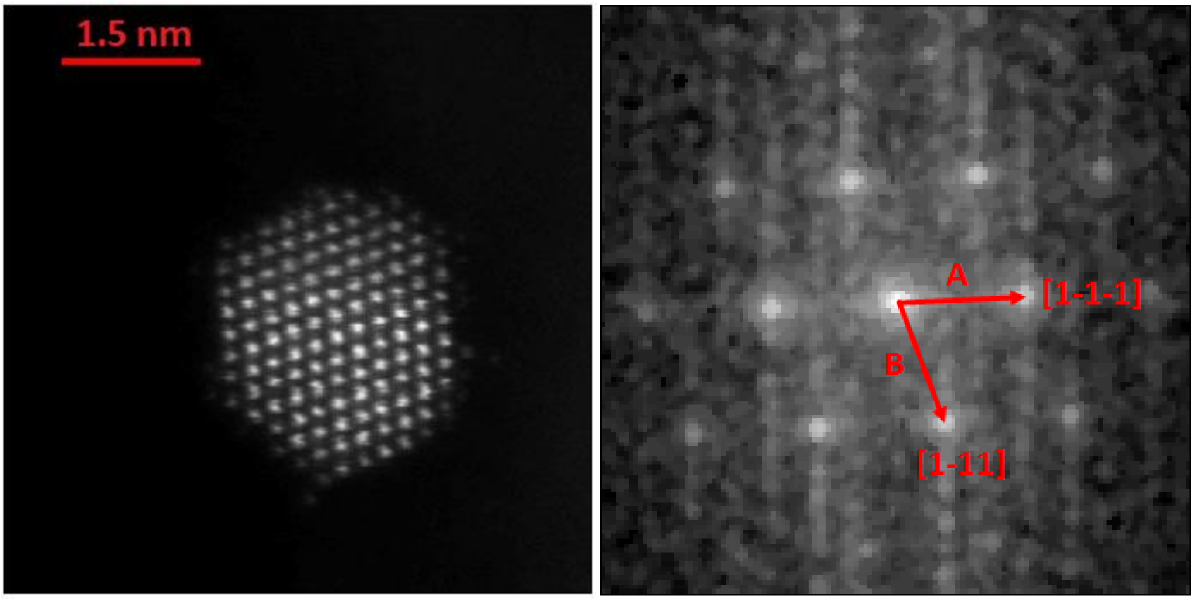


Figure 70. HAADF STEM image of a 110 oriented Pt nanoparticle (left) and its Fourier transform (right). For display purposes Fourier transform (modulus) has been enlarged by a factor of two and median filtered. The greyscale represents the logarithm of the FT modulus.

In Figure 70 two spots corresponding to the {111} family of planes are observable. Then, using each of these, we can define two vectors A and B. These two vectors must have precisely the same physical spacing, and by extension, the same reciprocal space lengths. However in the Fourier transform, being non-collinear they may contain a different number of reciprocal-X and reciprocal-Y pixels. Expressed mathematically:

$$|A| = |B| = {}^{1}/_{d_{h,k,l}} \tag{20.1}$$

$$(X_A \,.\, X')^2 + (Y_A \,.\, Y')^2 = (X_B \,.\, X')^2 + (Y_B \,.\, Y')^2 = {}^{1}/_{d_{h,k,l}} \tag{20.2}$$

where:

- X_A and X_B are the number of horizontal pixels in the vectors,
- Y_A and Y_B are the number of horizontal pixels in the vectors, and
- X' and Y' are the reciprocal width and height of a single FT pixel.

This is a system of simultaneous equations with two unknowns which can be solved for X' and Y', the inverse of which give the real-space width and height of the image pixels respectively. The product of these is then the pixel area needed in quantitative STEM calculations, and the difference between these gives again another indication of the severity of any image drift..

3.2. Scan-Noise and Drift in Recorded STEM Images

Deviation between the recorded data and the genuine sample caused by scan-noise has been reported previously, and the resulting slicing of atoms or atomic columns can be blamed on distortion effects [116]. This slicing or tearing of features is often attributed to varying AC fields [116], and is consistent with explanations derived from calibrated sensitivities of such instruments to these stray fields [117].

There are two lengths of time that neatly demark three ranges of frequency into which all distortions must fall: the pixel-dwell-time (the amount of time the probe remains stationary at a given pixel) and the frame-time (the amount of time taken to record the entire image).

3.2.1. Observing Scan-Noise and Drift

For any hope of image restoration distortions must be identifiable, and for distortions with a frequency above or near to the reciprocal of two dwell-times only a blurring would be expected. This is much the same as the parallel acquisition CTEM case and leads to an irreversible loss of resolution.

For scans synchronised to the AC mains supply typical dwell times are the reciprocal of the mains frequency divided by the image width. For the case of a 50 Hz mains power supply and for the common STEM image sizes of 512x512 and 1024x1024 this would equate to a detectability ceiling of 12.8 kHz or 25.6 kHz respectively. Acoustic or mechanical vibrations at these high frequencies are heavily damped and EM waves at this frequency are relatively easily shielded, though may still be present in the power supplies for the optics [31]. Because such high frequency distortions are not restorable from a single image, these will not be discussed further.

For distortions with frequencies between this upper limit and the reciprocal of the frame-time a localised distortion would be expected [33]. At any given moment, or equivalently for any given pixel, the distortion arises from what would physically be some probe-offset vector with both magnitude and direction.

The probe-offset vector can be projected into distortions parallel and perpendicular to the scan direction (referred to here as horizontal and vertical distortion). Because of this it is convenient to both analyse and correct for these distortions separately and this is indeed the approach followed here.

Lastly, the distortion which acts over a time greater than a single frame-time has an effect similar to that of sample or stage drift. Again this behaviour is expected to have both a magnitude and a direction and comparison of serial acquired STEM data with a CTEM image of the same material can readily confirm this behaviour [98]. In this case the area imaged on the sample becomes sheared, and the intended square area scanned by the probe instead becomes a rhombus on the sample itself.

3.2.2. Compensating for Scan-Noise and Drift

From the previous section we have three phenomena to compensate for: local horizontal distortion, local vertical distortion and stage/sample drift. This section will concentrate on the detection and correction of the two local distortions, with drift discussed in the subsequent section.

As the magnitude and direction of the probe-distortion vector is always changing (even within a single scan line) it is important too that the diagnosis and correction methodology is also able to respond locally. To achieve this, a real-space approach was developed that uses the minimum of a-priori assumptions. Compared with other (frequency-space) approaches this offers several advantages [34]:

- The image need not contain regular periodic features (such as a lattice). Atoms or atomic column features are instead inspected and restored one-by-one.
- As FT-spot streaking is of no concern, no specific sample orientation need be prescribed.
- Varying correction vectors are used within a single scan line, so no assumption regarding the distortion vector remaining constant over a scan line need be made.

- There is no specific upper limit to the amount of acceptable drift. Image drift is corrected as part of the restoration and larger drifts will only lead to increased cropping of the restored image.

The only requirement on the raw data is that there should be sufficient resolution and signal-noise ratio such that the atoms or atomic columns can be reliably distinguished. This real-space methodology is then more accommodating of point defects, dislocations and edges, etc with fewer restrictions on the microscope operator.

3.2.3. Horizontal Scan Distortion

Earlier it was noted that distortion frequencies near the reciprocal of the dwell time cannot be detected. This then implies that distortions which can be observed and detected operate at a frequency lower than this, and therefore persist over the duration of at least a few pixel dwell-times (a few horizontal nearest neighbour pixels). Within this distance then the recorded pixels reliably correlate with one another. Further afield across the whole scan line this correlation is lost confirming the distortions are indeed a localised effect.

To correct for any introduced time-varying distortions we must first detect them. As the recorded STEM data is itself a combination of genuine data and distortion it can be used to identify the distortions. For each pixel in the recorded image, a 'test-band' of pixels one row high but several pixels wide is defined centred about that pixel of interest Figure 71. The detail of determining the optimum size for this test-band is discussed in Appendix 8.2.

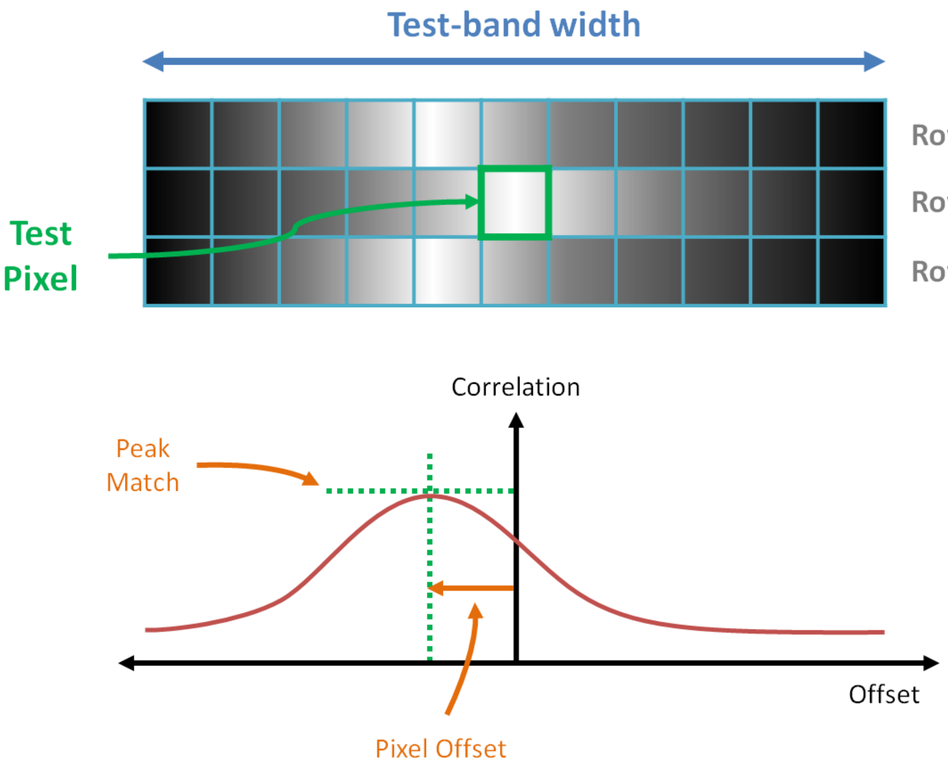


Figure 71 – Test-band definition and offset correlation schematic.

This test-band is then compared, over a range of horizontal offsets, with an equivalent area in both of the image rows above and below. At each offset the cross-correlation with these two neighbouring rows is calculated and averaged and is then used to suggest the most likely local distortion-induced offset. This offset is then equal to the displacement between the intended position of the probe and where it actually fell. The magnitude of the cross-correlation maxima also gives a measure of the certainty in the suggested offset result.

Plotting this probe-offset for each pixel as a function of time yields a view of the horizontal component of the scan distortion, Figure 72 (blue line). The offset profile is undetermined for the terminating pixels at the start and end of the data-array row as a position-sensitive cross-comparison is not possible in these locations. Figure 72 also demonstrates the benefit of using a real-space approach and not placing any restrictions on the distortion remaining constant across whole scan-lines [34].

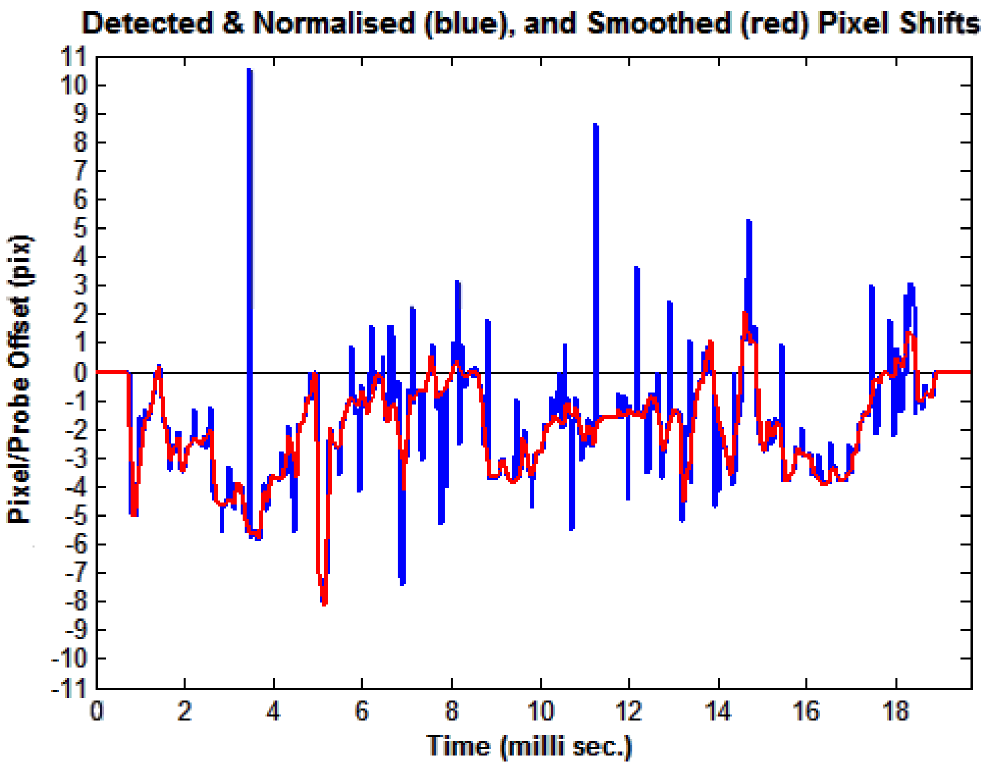


Figure 72. Probe offset along example data row both as measured (blue) and following normalisation and smoothing as described in the text (red).

Considering again the probe-sample offset profile, Figure 72, some refinement is still needed to account for both measurement reliability and plausibility. Firstly the reliability of the determined shifts; some cross-correlation matches may be of very low certainty, say less than 20% and it would be unwise to include these. Hence, any pixel-offset-reading with a cross-correlation maxima less than 0.2 is not moved in the restoration (zero offset applied). Cross-correlations between 0.2 and 1.0 are then scaled linearly so that by a confidence of 100% the whole of the indicated offset is applied. Secondly considering the plausibility of these normalised offsets; it is not physically meaningful that there can be discontinuities in the suggested offset profile because this parameter reflects the movement of the electron probe. To counteract this, the readings are screened using a method originally developed for the extraction of economic trends from financial observations, the 'Hodrick-Prescott' filter [122]. The approach just described and the parameters used were found empirically to give the most robust correction. Once the probe offset is known for each pixel within a row of the recorded image, the image row can then be resampled. This resampling has the effect of converting the recorded image (pixels recorded at regular time interval) to the restored image (pixel values recorded at physically equally spaced points over the sample). This resampling has the effect of restoring the previously 'torn' atoms or atomic columns to smooth features. A magnified example of the effect of restoring only this horizontal restoration is shown in Figure 73 b).

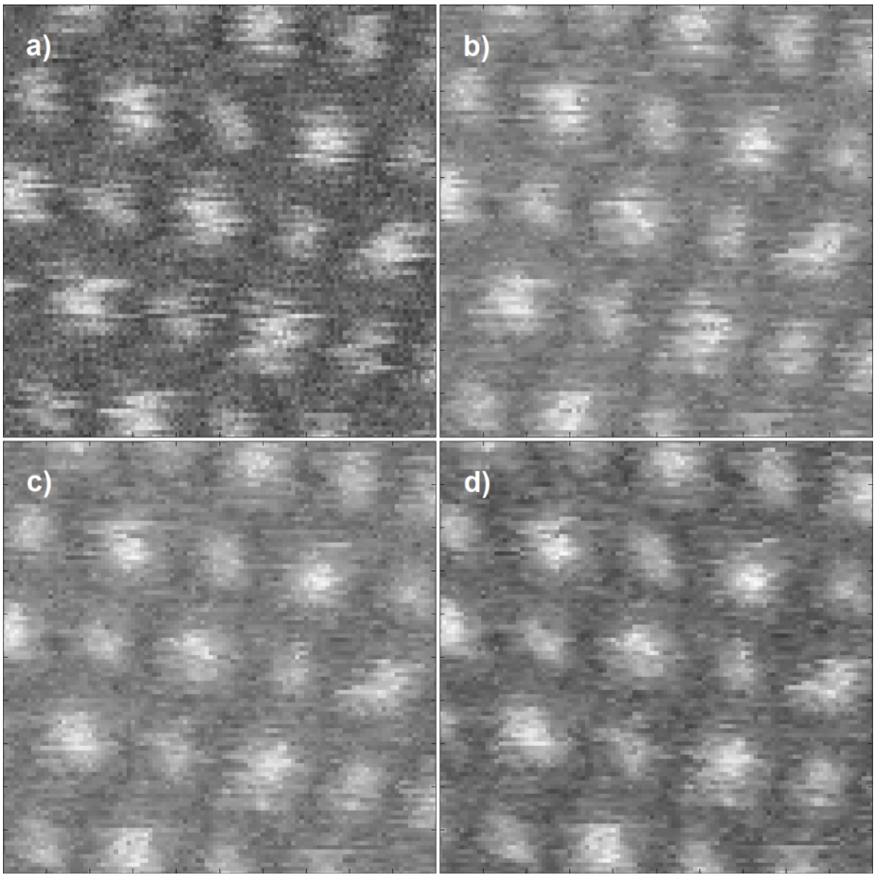


Figure 73. 100x100 pixel STEM HAADF image of [100] oriented $SrTiO_3$ and sequential reconstructions from it; a) as recorded image, b) following horizontal restoration only, c) after horizontal and vertical restoration, and d) as for c) but with drift also corrected. Note that the reconstruction has now allowed a previously un-noticed small amount of astigmatism or sample tilt to become observable as an elongation of the column peaks aligned top-left to bottom-right.

Figure 73 b) demonstrates how such horizontal resampling can largely restore the horizontal component of distortion.

3.2.4. Vertical Scan Distortion

In the previous section distortions parallel to the scan direction (horizontal) were considered, now the attention turns to distortions perpendicular to the scan direction (vertical). Whereas horizontal distortions served to shift features left or right in each image row by a number of pixels, vertical distortions have the effect of locally moving short lengths of adjacent pixels 'higher' up into the row above or 'lower' and down in to the row below. In the case that the scan line and its neighbours are entirely between atoms or atomic columns (flat background) no distortion induced row reordering will be observable, but equally none need be restored. However, often near the top and bottom of atomic features this effect can be seen as dissociated rows of bright pixels separated off of features, or equally dark rows brought within features, Figure 74 (top-left).

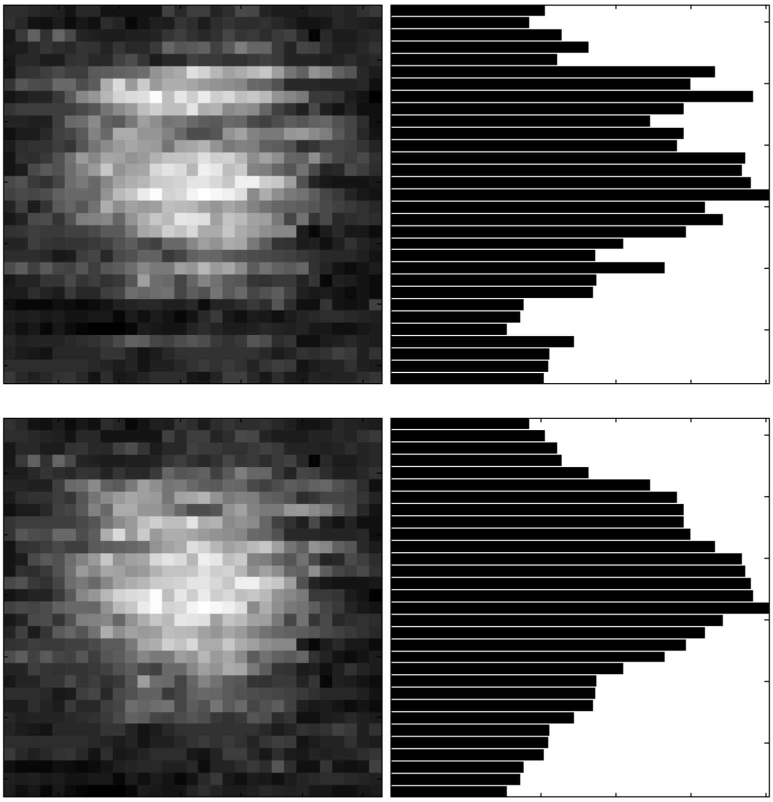


Figure 74. Enlargement from Figure 3 showing an example strontium atomic column (top-left) and its horizontally integrated intensity (top-right). The figure shows the raw data (top) and the same area after restoration (bottom). The inspection region of interest is a square defined about the brightest feature pixel with width and height equal to the test-band width.

A useful method of visualizing both the problem as well as the methodology to correct for it is to examine the intensity profile of a given feature when integrated horizontally (row-wise), and this profile is shown on the right hand side of Figure 74. In this view dissociated bright rows appear as single-row peaks and dark rows inside features as troughs. This intensity profile is used to restore the feature with the methodology being to simply reorder the rows such that the intensity decays monotonically from the brightest pixel (at the central row) out towards the top and bottom. An implicit restriction is also then imposed such that rows do not cross over from the top-half of the feature to the bottom or vice-versa.

The result of this is that in not only are image features restored to a more round appearance but also that without the dissociated bright or dark rows the misinterpretation of these as noise is inhibited and hence the effective image SNR is increased. Also, and importantly, as image rows are merely reordered the integrated intensity of any given feature is unchanged.

An important aspect of both the horizontal and vertical correction described here is that the object should be oversampled during the imaging process. For example, if a periodic lattice is imaged close to the Nyquist sampling pixel-resolution (magnification) then it will not be possible to use cross-correlations for the horizontal correction, and in the presence of vertical distortion there is a chance that the total intensity recorded in the data would not accurately reflect the structure of the sample, and so neither would the restored image. When a sample is imaged well above Nyquist then the number of rows where intensity is underestimated (dark streak, probe intended to be on atomic column but actually in between) and overestimated (bright streak) are expected to be equal, and the rows can be sorted and re-ordered. This again suggests the criterion that operating at magnifications high enough to sample well above Nyquist is likely to give the most reliable 'image' of a given feature.

The reader will note that both the horizontal and vertical scan distortions described above require comparison of pixels with their neighbours. The reader will also note that if there has been a vertical reordering of image rows during the reconstruction of an image then the neighbour rows about each pixel may have changed. If a very large amount of vertical distortion was present it may be necessary to repeat the horizontal distortion analysis with these new neighbours. As a result an iterative approach to converge towards a distortion free image was developed.

3.2.5. Sample / Stage Drift

When either the sample or stage drifts during a serial acquisition, the result is a distorted image. Assuming a constant drift it can be categorised by two parameters, a unique direction and drift rate. It should be noted that a second unique solution with a drift direction perpendicular to the first with a negative magnitude will give the same image appearance but with a small change in magnification. As magnification is an arbitrary concept without knowing the size of display / printed pixels, and is often not calibrated precisely, this second solution can be ignored.

Although the data is recorded in a square array of pixels the actual area scanned could be rectangular (drift parallel to one edge) or a rhombus (arbitrary drift). These drifts can usually be reduced, but not entirely eliminated, by the use of column thermal insulation [30] or piezo-drive based stages or holders. These drifts can be exacerbated when using heating or cooling holders or if airflow fluctuations cause temperature variations in the instrument suite.

Experimentally, drift can occur in any direction. Drift with a component parallel to the optic-axis will result in an effective drift in focus over the duration of the scan. This is outside the remit of this work and is not discussed further. The components of drift in the plane normal to the beam is however of interest here and is often associated with thermal changes in the column / holder, or with hysteresis in the stage drives. Both of these effects can be modelled by a slow exponential decay but are often approximately linear over the duration of any single image acquisition. Small drift rates of the order of 0.5 – 1.0 nm min^{-1}, or tens of picometres per second, are generally acceptable but can become significant when performing many-second 'stationary' probe spectroscopy measurements [123], [124], [125].

A sometimes observed practice for measuring drift is a one dimensional (1D) approach [125]. The user will align the scan direction such that the sample crystallography has a lattice plane horizontal and parallel with the top of the image (the fast scan direction). Then the drift is estimated from horizontal offset between complimentary lattice planes near to the top and the bottom of the image. This 1D analysis has the unintended effect of projecting the drift vector onto an arbitrary set of orthogonal axes (set by the scan rotation), and effectively discarding one component.

An analysis of all the lattice plane angles and spacings in an image however gives a far more accurate two-dimensional description of the drift present. In the following section this more reliable 2D analysis method is described, on which a subsequent correction transformation is based.

In this work it is assumed that the drift *rate* and *direction* present during the time of the STEM image acquisition is constant; no assumptions however are made about its direction, and hence there is no requirement for any particular scan orientation to be prescribed. In the presence of an arbitrary drift the actual area scanned will become a rhombus. Consequently we can say then that to convert the recorded data back to a drift-free image that two parameters are needed: the orientation and the rate of the drift. The need for two parameters such as the above is clarified if we examine the matrix transformations needed to restore the image. A general 2D transformation will require 4 parameters (usually expressed as a 2 by 2 matrix). However, as the uniform dilation is effectively a small change in magnification (which itself requires calibration) this term can be ignored. Equally as scan-rotation can be set arbitrarily we can similarly ignore image-rotation. This leaves us with only two parameters to fit rather than four.

The two remaining parameters can be represented as a combination of a rotation and a simple shear term. The distortion can then be remedied by a rotation, ϑ, followed by a simple shear, s, to restore the data to a genuine image:

$$Image = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} (Data\ Array) \tag{21}$$

Equipped now with a route to restore the recorded image we now only need the image rotation and simple shear parameters ϑ and s. From the discussion presented in above it is clear that simply comparing the angle between two planes alone risks not detecting drift in the chance case that the drift vector is aligned perpendicular to the scan direction. Instead the method deployed here requires at least three lattice plane angles and is described in detail in section 3.1.4. Applying the transformation given by equation 21 (and with suitable edge padding) the drift corrected image is returned (Figure 75).

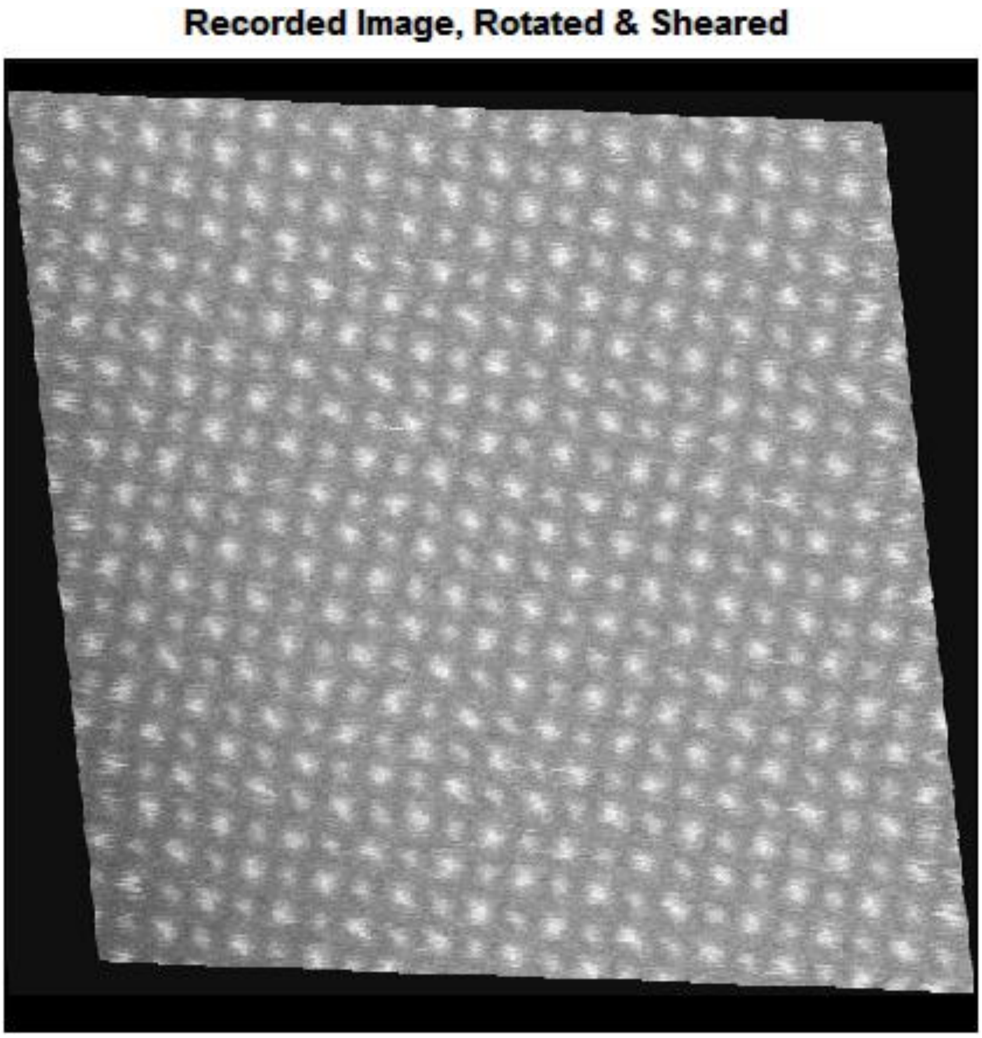


Figure 75. Drift correction of the restored image via the rotation and shear determined from angular deviation analysis. Resulting parallelogram area shows the region of the sample that was actually scanned during imaging. Black background is perfectly square for comparison.

Often one final operation that can be performed, is to rotate the image back by the same angle ϑ so that the largest possible square area can be cropped from the transformed array and this is indeed the strategy employed in the author's image restoration code. Conveniently each of these image processing operations are computationally easy at the image sizes recorded for STEM.

Finally once drift has been corrected it may be the case that a previously masked 2-fold astigmatism becomes visible (Figure 73 d). This is unsurprising as the STEM operator often performs the final stigmation on a continuously updated (live) but never the less recorded image where drift can serve to mask a slight astigmatism. If this is the case, then a non-ideal probe was used to record the data and that resolution will have been permanently lost.

Combining the steps for horizontal, vertical and drift correction for this example data (Figure 73 d)), improvements in resolution and SNR of 16% and 30% respectively were achieved. Figure 76 shows the corresponding Fourier transforms.

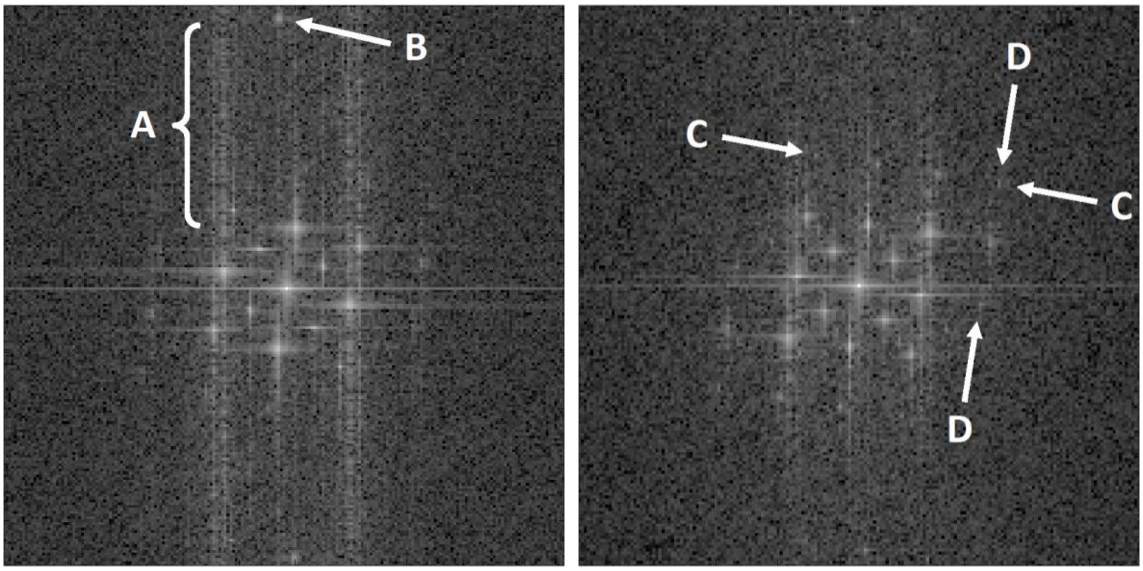


Figure 76. Enlargements from Fourier transforms of the test images corresponding to left: the raw-data (Figure 73 a)), and right: to the restored image (Figure 73 d)). Labels A and B indicated vertical streaking and satellite spots caused by image distortions. Labels C and D indicate positions of Fourier spots with enhanced clarity as a result of image restoration.

In this reciprocal-space view of the same data a general reduction in the noise (A) and strength of satellite spots was observed. Some spots at higher spatial frequencies also became more easily identified consistent with the improvement in resolution.

3.2.6. Horizontal Scan-Distortion Spectra for Suite Diagnostics

As the test-band comparison process involves examining the rows above and below that of interest, an offset profile can be calculated for all but the first and last row of the recorded data. This means that an offset-profile (like that in Figure 72) and hence frequency spectrum can be determined for all but two of the image rows. This gives a large number of spectra that can be averaged and an example of the average of such spectra is shown in Figure 77. Calculating this distortion spectrum also shows conclusively that it is unreliable to assume that distortions will always appear constant over the duration of a scan-line time [34].

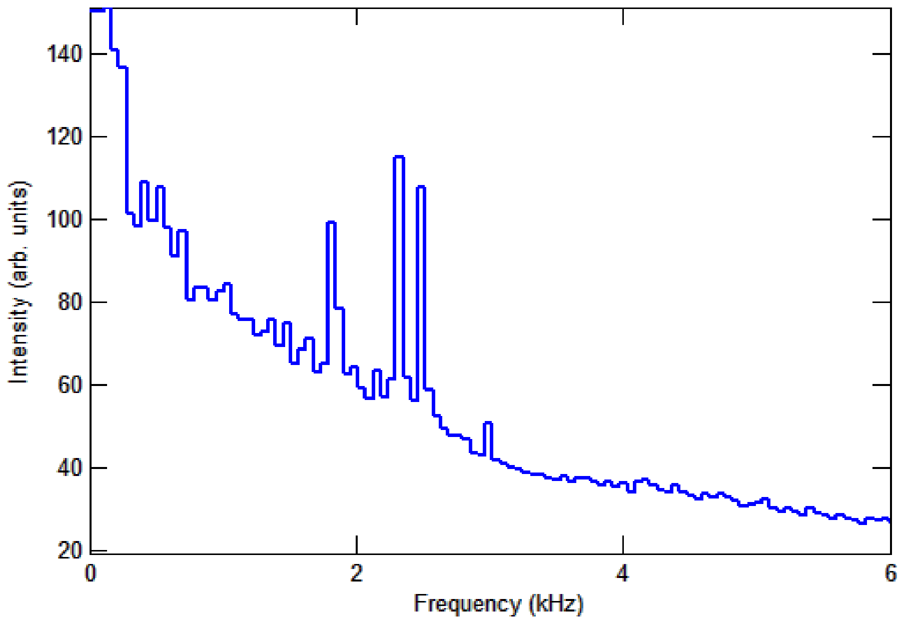


Figure 77. Average distortion spectrum from 510 individual spectra; showing peaks at 1.79kHz, 2.28kHz and 2.45 kHz. A decaying background exists and above 3kHz peaks are not readily discernible over this background.

Once this reliable spectrum is calculated, it is then possible to compare this with the environment of the microscope to identify the primary distorting culprit. In general different sources of environmental distortion, and in rough order of increasing frequency these include; thermal drift, seismic vibrations, column-resonance vibrations, AC motors & fans, sample holder resonant frequencies and finally EM field distortions [31]. Some of these phenomena are common to all facilities, such as the speed of computer fans, but other factors such as the column or sample holder resonances need to be assessed on an installation specific basis.

Studying such spectrum while making modifications to the microscope suite, such as turning on/off air handling systems or adjusting acoustic foam panels, allows the user to identify the most dominant distortion source at any one time.

4. The Utility of HAADF STEM Focal-Series Data

When a well-aligned zone-axis sample is imaged, electron channelling theory states that the intensity of the focussed STEM probe remains highly localised to each atomic column as it propagates through the sample [37]. This has the effect of making the sample appear as if a two-dimensional object, coincident with the entrant surface and is the basis of the '2D object convolution model' of image formation (Equation 10). In this model the illumination function is also assumed to be two-dimensional (as if it were only evaluated at the entrance surface of the sample); however, from Section 2.1 we have seen that the full-3D form of the STEM OTF is richly detailed. If a focal-series is recorded, we are forced to consider the full 3D STEM OTF; this raises additional important questions that must be addressed:

- What is the resulting information content of the focal-series data volume?
- Does the 2D object assumption remain valid, and over what range of sample thickness?
- What information can be deduced about the illumination from focal-series data?

In this section these topics will be investigated with the aid of some experimental examples.

4.1. The Spatial-Frequency Contents of Focal-series Volume Data

Experimental data acquired by focal-series take the form of a so-called 'data-cube'. This data cube has real-space lateral dimensions and defocus (or equivalently z-height) as its third real-space dimension. Now we are tasked with visualising the information content of this 3D data-volume.

For electron micrographs of crystalline (periodic) materials, it is common to inspect them in reciprocal-space – that is to inspect their Fourier Transform. In this case, arrays of periodic features in real-space (atomic columns) become a set of points in the Fourier transform (sometimes referred to as reflections; although not caused by diffraction, their symmetry is often similar to the diffraction pattern of the same sample). Figure 78 shows three example frames from such a focal-series dataset.

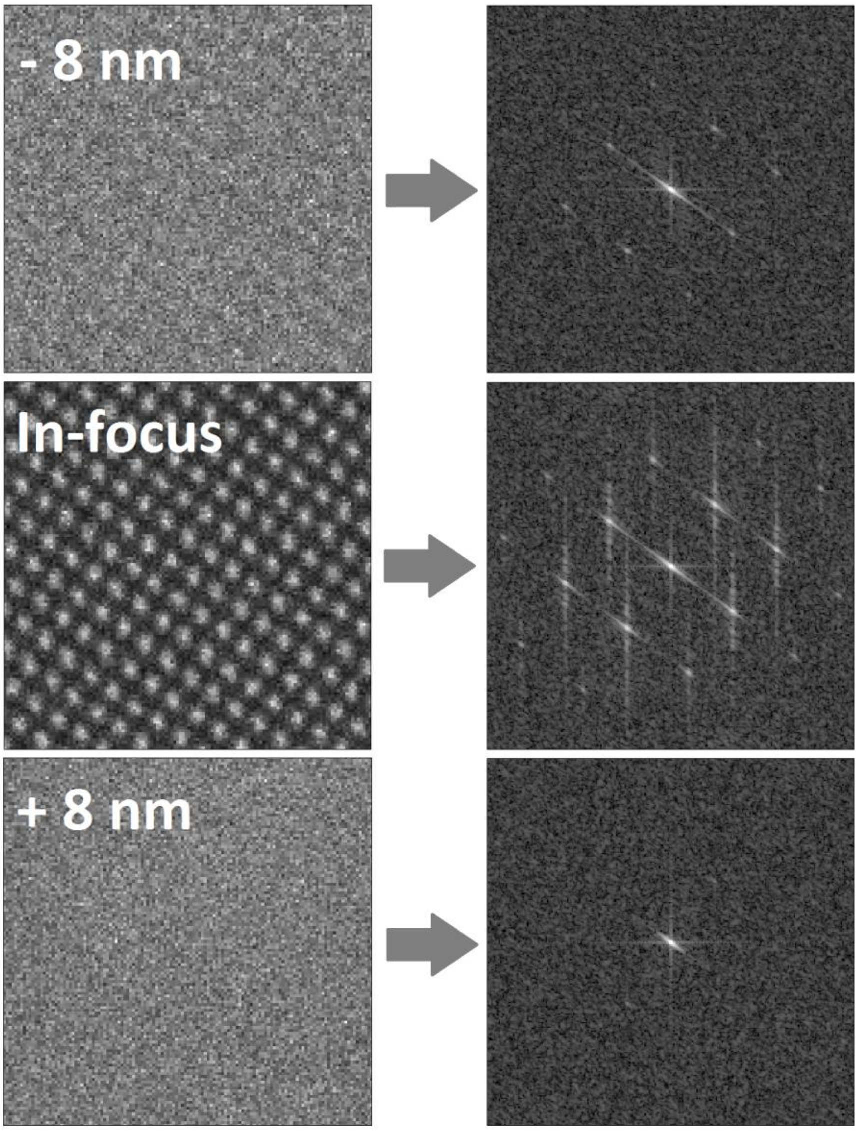


Figure 78. Enlargements from three example frames of a larger STEM focal-series dataset. Left, the real-space images and right, the corresponding reduced Fourier transforms (central 50% of FT, all with matched grey-scales). Indicated defocus corresponds to the focus relative to the operator tuned value (middle) with 8 nm under-focus (top) and 8 nm over-focus (bottom). Some scan-noise induced vertical streaking of the FT spots is also visible.

We can convert each of the real-space frames into their frequency space representations, their Fourier transforms (Figure 78 – right), and stack these into a similar data cube. This volume now has lateral dimensions of reciprocal-space (or spatial frequency) but the third dimension remains defocus. Microscopists will be familiar with the 2D Fourier transform containing an array of bright spots of interest (corresponding to the lateral spatial frequencies contained in the real-space image) and a central bright 'zero-spot', both superimposed upon a background of structureless noise.

Figure 78 shows only a subset of three frames from the entire dataset but already some important observations can be drawn from this. As would be expected the real-space in-focus image exhibits the best resolution and SNR; in its Fourier transform it also contains the greatest number of unique spots. This corresponds to the OTF at this defocus extending further out to higher spatial frequencies. We can also observe that for the lower-order FT-spots (closer to the centre) intensity persists over several nanometres. Of this persistence though we can observe a slight asymmetry with respect to defocus, with the -8 nm data showing considerably brighter FT-spots than at +8 nm.

What we have then are many spots with individual intensity profiles with respect to defocus that hint at a contrast-transfer non-uniformity to be investigated.

Next having a truly three-dimensional data-volume, the issue of visualising this information clearly must be addressed. While it is easy to view individual frames of this 2D data, usually as a logarithmically-scaled greyscale image, representing 3D data becomes more challenging. One method to represent this data is using a 3D iso-surface, that is to show the surface of the volume that represents a subset of the 3D voxels that contain intensities above a certain value. This can be thought of as a three-dimensional thresholding. When this is performed the former Fourier transform 'spots' become rods, elongated with respect to defocus. An example of such a data-volume is shown in Figure 79. The z-axis scale has units of nanometres relative defocus on an arbitrary scale where the first image sits at one, and as the images were recorded at a one nanometre focal step – this also coincides with the image number in the series.

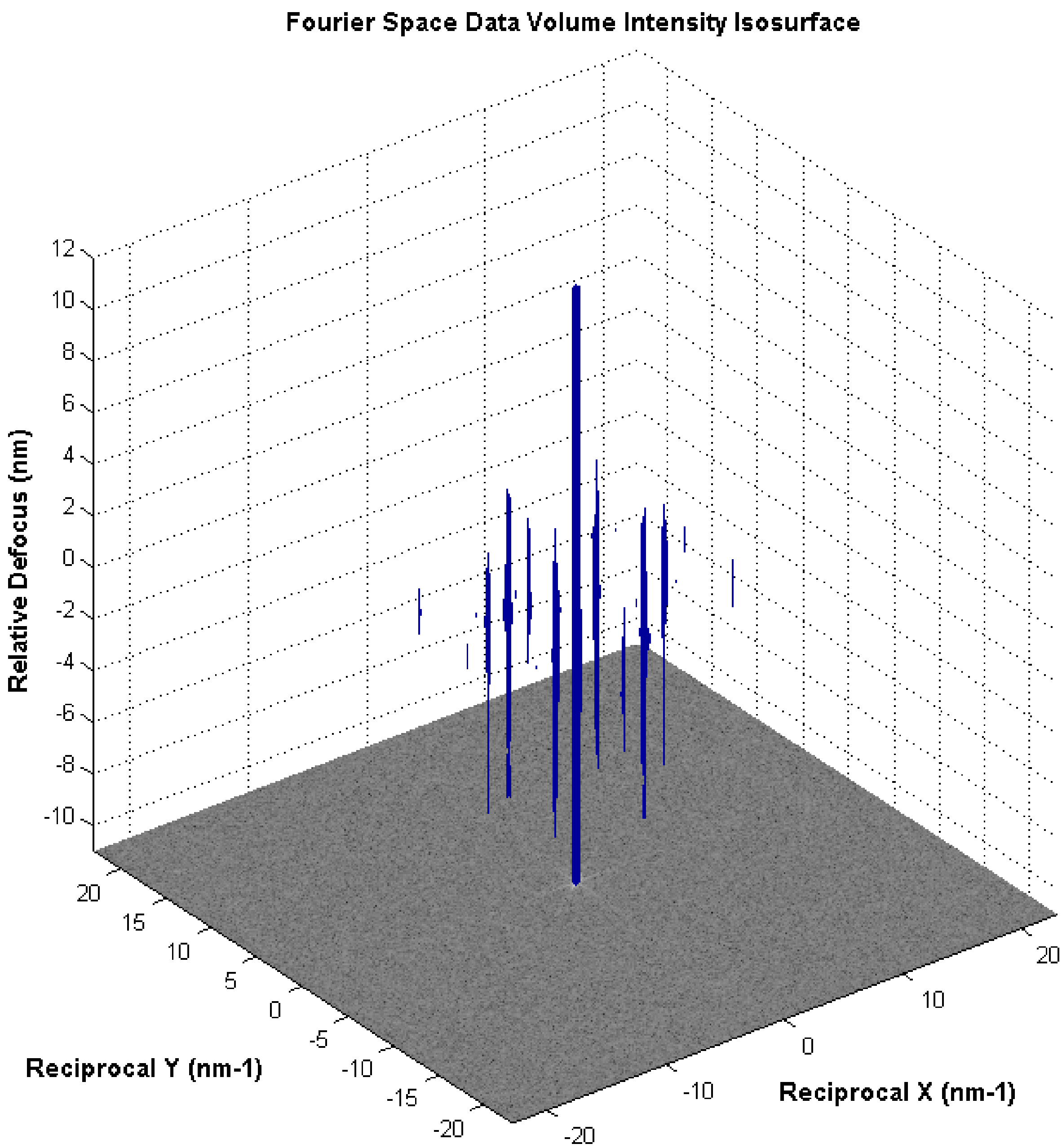


Figure 79. Example intensity iso-surface extracted from a Fourier stack data-volume of an MgO smoke cube. The data includes 24 Fourier transforms acquired with a 1 nm focal step. Fourier transform number one is used as the base of the plot. Iso-surface is drawn at 25 standard deviations above the mean background.

In the Fourier transform of a high-resolution low-noise image, it is common to observe many pairs of Fourier spots; in three-dimension this yields many contrast transfer rods (Figure 79). For simplicity we will now discuss some of the features of the above figure using a more directly interpretable 'side-view' (Figure 80).

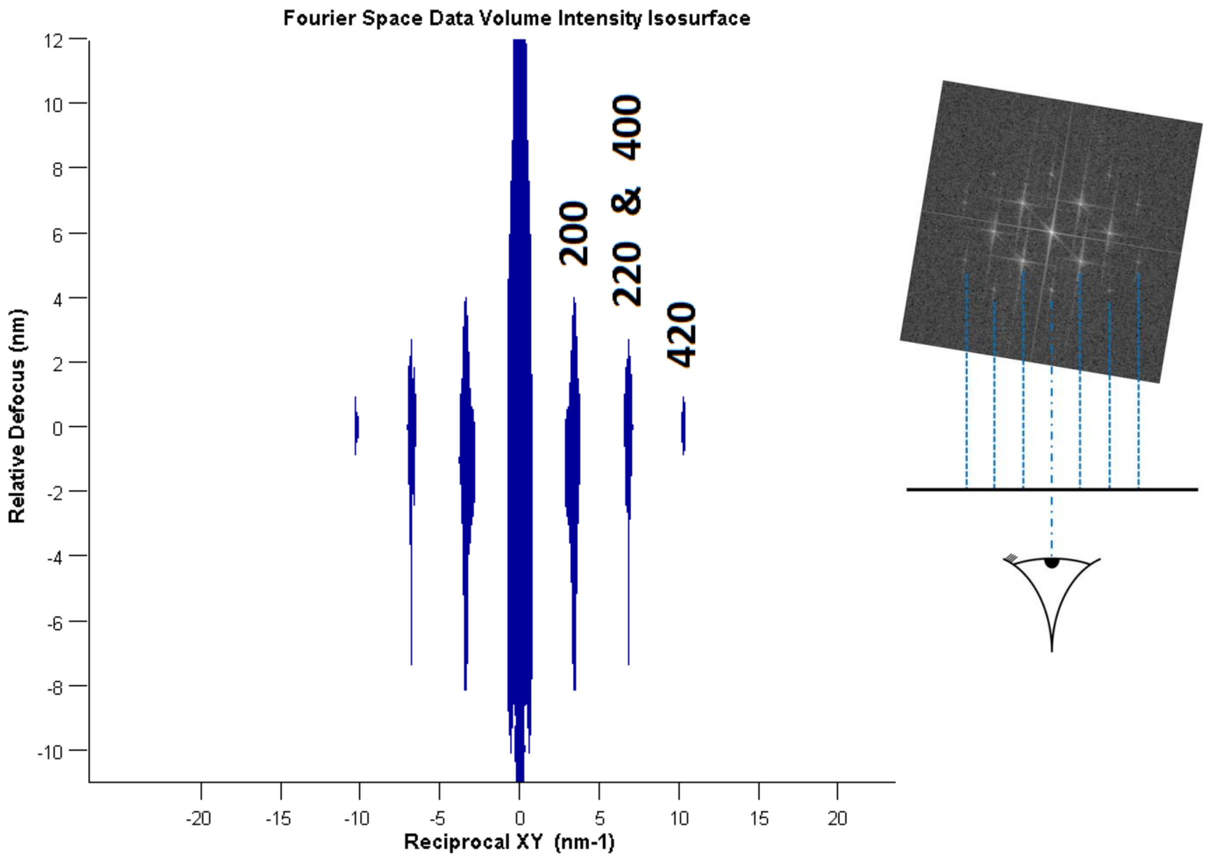


Figure 80. The same Fourier cube iso-surface shown in Figure 79 but rotated to a 'side-on' view. Inset schematic shows the Fourier transform of the medial image (0 nm relative-defocus) to define the viewing direction of the main figure.

Figure 80 shows the exact same iso-volume render as Figure 79, only it has been rotated so as to bring the 110 rod directly in front of the zero-order rod. In this orientation other crystallographic rods also appear coincident and are labelled as such. Now we have this new figure we can observe several noteworthy properties:

- The central rod runs the entire length of the data-volume. This represents the defocus-rod corresponding to the zero-order reflection (also known as the D.C. offset),
- Additional rods are present that correspond to sets of crystallographic spacings (see Figure 80 inset),
- The rods have varying defocus-extents, with the higher order rods progressively shorter and shorter,
- The intensity of the rods (more bulged areas are more intense) peaks but not necessarily in the centre. Both the [200] and [400] set are more intense at defocii slightly greater than the midpoints of the defocus-rods.

The explanation for all four of these features can be found in the form of the three-dimensional optical-transfer function (OTF) for a STEM probe with a small remnant spherical aberration – Figure 44 (right). This transfer function describes the expected performance of the illumination alone, when this is convolved with a 2D object function we obtain the 3D Fourier volume shown in Figure 79. In the example used here the sample was an MgO smoke cube, this has little material interest but gives a simple square array of atomic columns that yields the square array of reciprocal lattice vectors (Figure 80 - inset). When this square array of FT points is swept through the OTF volume shown in Figure 44 (convolved with during FS imaging), we obtain the FT stack with defocus rods shown in Figure 79. In Figure 44 we can also see the same four characteristics described above; a zero-order spot that persists across the entire defocus range, transfer that extends well into the sub-angstrom range facilitating higher order FT spots, a defocus spread that reduces with increasing spatial frequency from the origin, and a defocus asymmetry of the strength of the contrast transferred.

4.2. Testing the Validity of the 2D Object Function Assumption

For a single STEM image recorded at focus the so-called '2D-object convolution model' is often referred to (Equation 10); that is the recorded image is expected to be the two-dimensional convolution of the probe point-spread function (PSF) and the specimen, represented as an array of points corresponding to the locations of atoms / atomic-columns. This has been shown to be true for 2D data by applying the appropriate deconvolution and reconstruction [48]. However, an important question remains as to whether this remains true for the three-dimensional case and this is the focus of this section.

4.2.1. The 2D Object Function in Focal-series Data

When a focal-series data-set is recorded the result is a three-dimensional real-space data-volume. If the 2D convolution model remains valid for each acquired image, then the expected result would be the *three-dimensional convolution* of the 3D illumination (Section 2.1) with the 2D object function. To verify whether this is true experimentally, we first need to Fourier transform each image in the series to yield an FT-stack (such as Figure 80). In this representation, we would now expect every spot in the Fourier transform of the 2D object-function to have become elongated with respect to defocus (persistent over several images), with the amount of elongation being determined by the 3D-OTF at the spatial frequency of interest (such as Figure 44 - right). An example of such a persistent FT spot is shown in Figure 81.

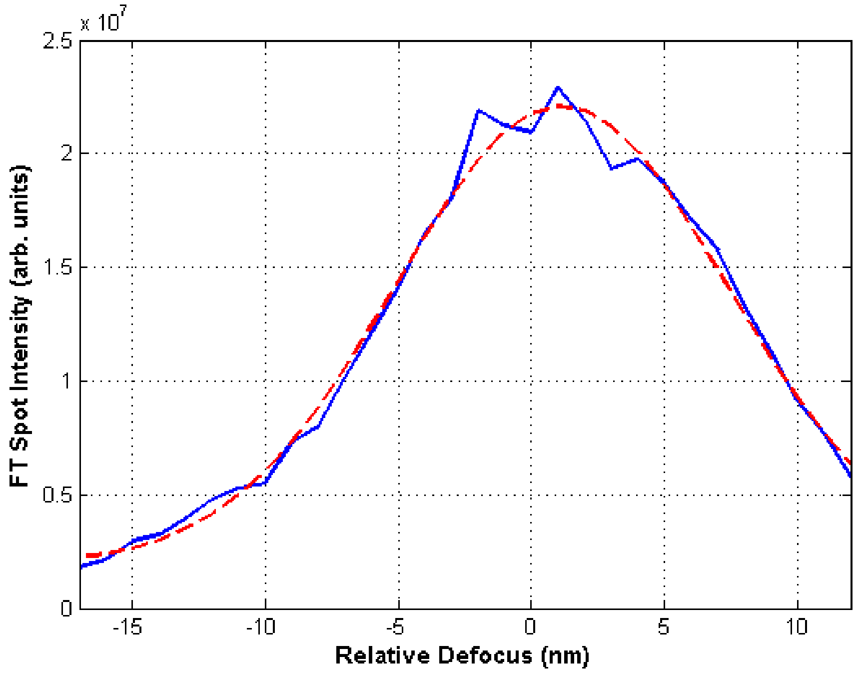


Figure 81. Modulus (in arbitrary units) of an example 200 Fourier spot (blue), tracked with respect to defocus. Profile is from the example MgO cube data shown in Figure 107 – sample was ≈ 83 nm thick. To reduce noise on the profile, intensity was integrated within a square region immediately surrounding the FT-spot with a radius of 3 pixels. To this data a Gaussian profile was fitted (red) to extract the defocus corresponding to the maxima and the defocus spread.

From this intensity profile we can determine the defocus corresponding to the maximal strength of contrast transfer at the spatial frequency of this FT-spot as well as the spread of this intensity with respect to defocus. The position of this maxima may be used for aberration diagnosis and this is discussed later in Section 4.3. For now this discussion will concentrate on the persistence of information over defocus; to evaluate this it is convenient to fit a Gaussian profile (or some similar function) to the data presented in Figure 81 and this is shown by the red dashed line. Using such a fit we can quantify this spread using the 'standard deviation' of the fitted Gaussian profile. The FWHM may also be used so long as a consistent metric is used throughout both image diagnosis and image restoration. Where the profile of interest is accurately modelled by a Gaussian, with standard deviation σ, then a constant relationship exists between these two measures:

$$Profile\,Width\;(FWHM) = 2\sqrt{2\;\ln 2}\;.\sigma \approx 2.35\,.\sigma \tag{22}$$

Figure 81 shows the transfer intensity profile for only a single spot whereas a similar profile exists for each clearly identifiable spot in the Fourier-data-volume. Repeating this analysis and determining the 'defocus-spread' of the Fourier-spots for every observable Fourier-spot yields a map of this depth-resolution as a function of lateral spatial-frequency, Figure 82 (red dots).

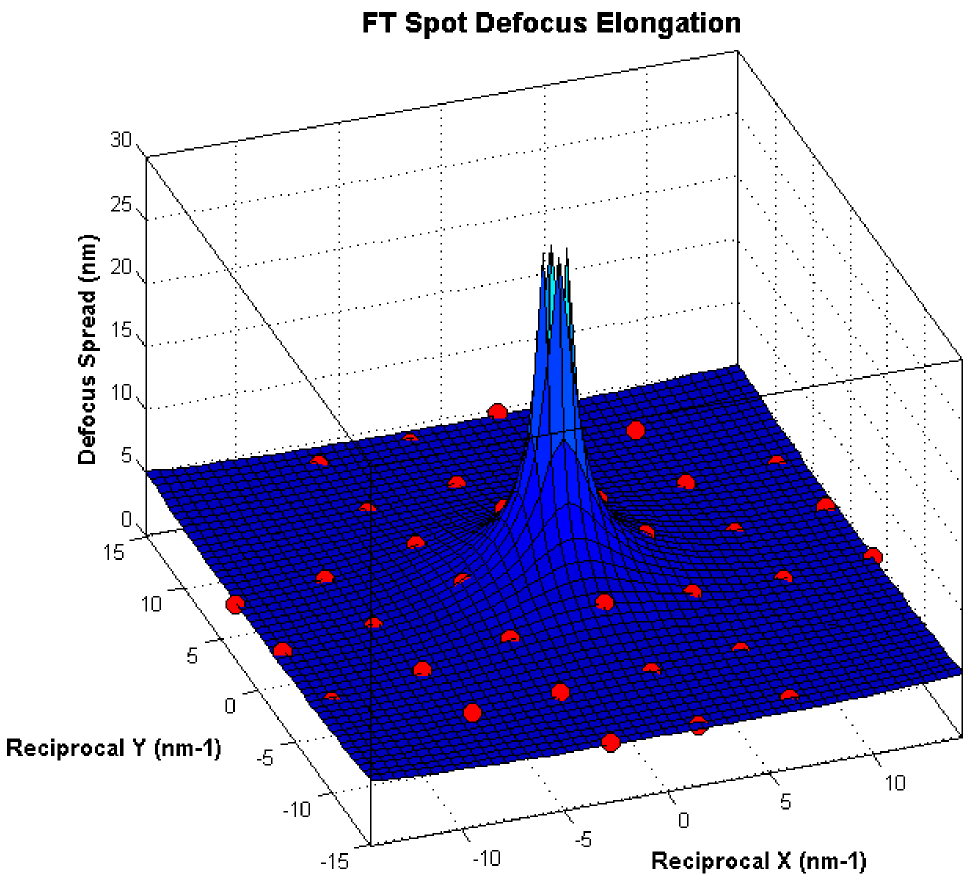


Figure 82. Experimental illumination performance results as evaluated using experimental focal-series FT-spot defocus-elongation data. Red circles at various reciprocal-xy coordinates (corresponding to the FT-spots of various spatial-frequencies) show the FWHM of fitted Gaussian profiles (z-coordinate) for each identifiable spot in the Fourier transforms. To these a smooth surface (blue) has been fitted to yield values across the whole Fourier surface.

To evaluate the depth-resolution performance of the image-stack as a whole, we could take the minimum value from all the fits on the individual spots. However, as the Fourier-spots only sample a small subset of the spatial-frequencies transferrable by the 3D-OTF a more robust approach is to fit a smooth surface through these scattered points and take the minima of this surface. Examining the shape of the OTF 'wings' from Section 2.1 and from inspecting many experimental data sets, an empirical form for such a smooth surface was developed so that it may be described parametrically; including an asymptotic decay and also an increasing power-law ramp with increasing spatial-frequency (**r***):

$$Spot\ Spread\ (\boldsymbol{r}^*) = A + {}^{B}\!/_{|\boldsymbol{r}^*|} + C\,.\,(|\boldsymbol{r}^*|^D) \tag{23}$$

This surface was evaluated for the FT-spots in Figure 82 and is superimposed as the surface fit (blue). We can see that Equation 23, and hence the fitted surface in Figure 82, is undefined at the centre **r*** = 0 point; this is no limitation as there is also no physical meaning to this point as it corresponds to the D.C. offset (or background contribution) in the experimental FTs.

Comparing the form of this surface *qualitatively* with the example OTF in, say, Figure 44 – right, we see that for low spatial-frequencies (zero to two reciprocal nm) the depth resolution is poor, this improves to a peak at (minimum-spread) at around nine reciprocal nanometres before gradually spreading again beyond this.

In quoting the depth resolution exhibited by a particular experimental OTF and sample combination it is convenient to express this performance as a single value. To yield this we can simply extract the minima from the defocus-spread surface shown in Figure 82 (not necessarily the same as parameter A above).

Figure 82 shows the spatial-frequency dependent FT-spot defocus spread measured directly from the experimental data, however we can also use the same simulation tools presented in Section 2.1 to evaluate the expected performance of the illumination in isolation. Figure 83 shows a section through the 2D surface presented in Figure 82 and also the equivalent measurement (spatial-frequency dependent depth-resolution) taken from a simulated probe with matched voltage, convergence angle, aberrations etc. (methods as in Section 2).

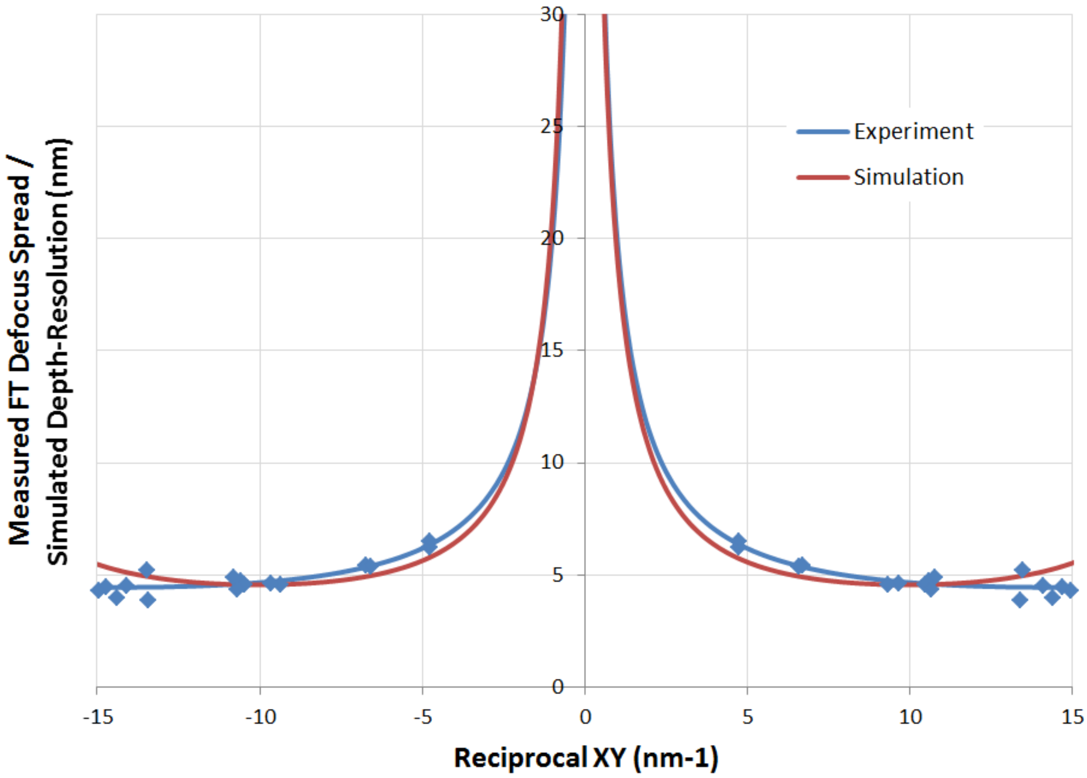


Figure 83. Line-profile through the experimentally fitted surface shown in Figure 82 (blue) showing the spatial-frequency dependent defocus-elongation of FT-spots. For comparison the spatial-frequency dependent depth-resolution (calculated from a probe simulation matched to the experimental conditions) is overlaid in red.

The most striking feature in Figure 83 is the exceptional agreement between experiment and simulation and three important conclusions can be drawn from this:

- The experimental curve in Figure 83 (blue line) very closely mirrors the form of the simulated one; this implies that the chosen form for the empirical shape of the fitted surface (Equation 23) is reasonable.
- As the raw-data for the defocus-spread surface fitting shown in Figure 82 was based on many individual FT-spot intensity-tracking measurements (such as that in Figure 81), this excellent agreement would seem to validate both the practice and accuracy of such FT-spot tracking.
- Thirdly, and perhaps most importantly, the experimental data show no additional defocus spread over the simulated profile. The only additional factor introduced experimentally was the sample itself, and this directly confirms the two-dimensional behaviour of the specimen as an object function.

This result confirms that the 2D object-function assumption remains valid at the sample thickness used in the above examples (80 nm in this case) but to verify the model in general terms the agreement should be studies over a range of thicknesses.

4.2.2. The 2D Object Function Assumption and Specimen Thickness

In the previous section we saw *qualitatively* that the fitted depth-resolution performance (Figure 82) exhibits a similar form to simulated 3D-OTFs. To more quantitatively confirm the validity of the 2D object function assumption we can repeat the analysis from Section 4.2.1 for a range of different sample thicknesses and compare the results to the depth-resolution performance from probes simulated using the techniques in Section 2.1. Furthermore, this need not be limited to only one experimental setup; in fact for a fully general discussion, in this section results will be incorporated for a range of different accelerating voltages and convergence angles to further separate the probe dependant- and sample dependant- behaviours.

Data were collected from cubes of MgO (prepared by burning magnesium ribbon in air – see reference [126]) using microscopes from three different manufacturers, Nion, JEOL & FEI, operating at voltages of 80, 200 & 300 kV respectively. Sample thicknesses were estimated from measuring the lateral extent of the MgO cubes (Figure 84).

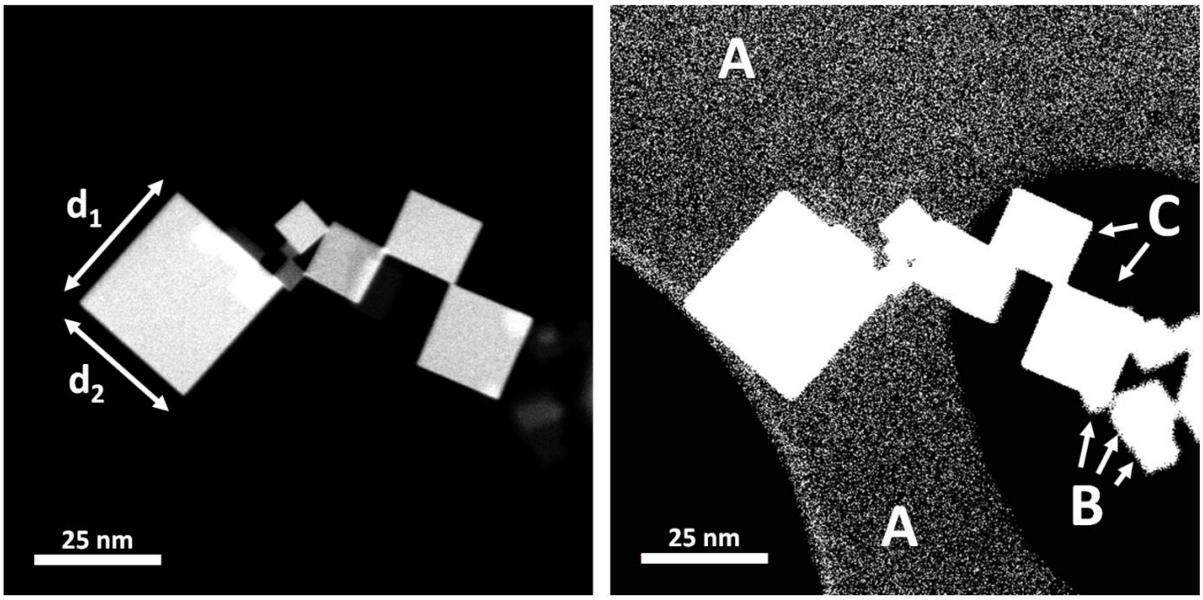


Figure 84 – Left: example low-magnification HAADF STEM image of several MgO smoke cubes used for the measurement of cube diameters. Right: the same image but with the greyscale range adjusted to reveal the position of the carbon support (A) and also several smaller cubes not otherwise visible (B). Using this method cubes entirely overhanging the carbon support (C) were selected for focal-series imaging.

Next the attainable depth-resolution of the respective STEM probes was simulated and the experimental defocus-spreads normalised by this amount. This yields a so-called 'normalised thickness', where a value of unity (100%) indicates the sample behaves as if infinitely thin (totally two-dimensional) and values larger than this indicate an additional defocus blurring arising from the finite thickness of the sample.

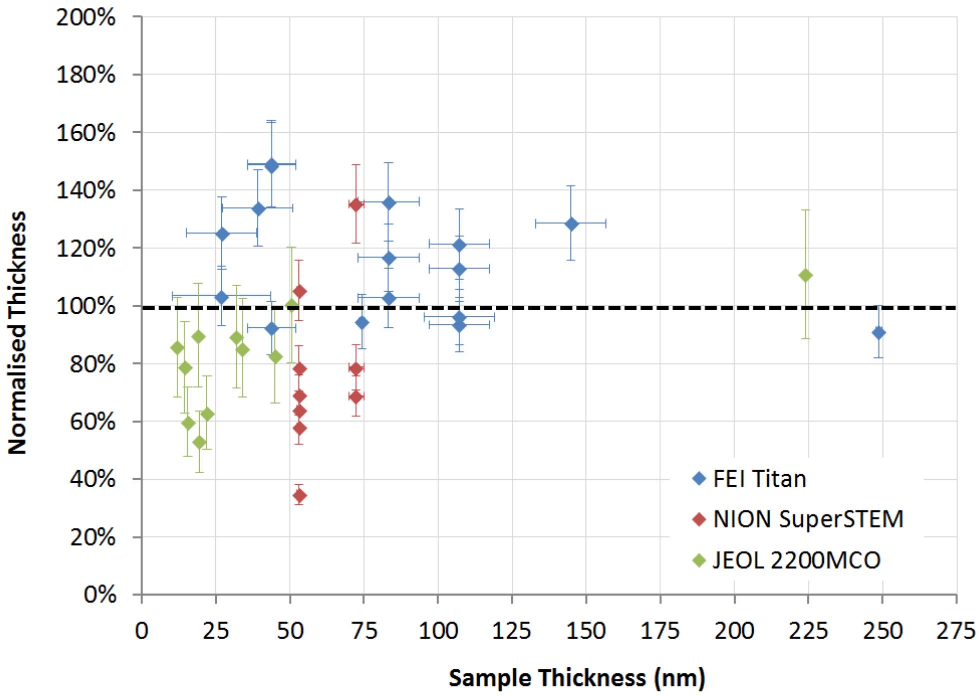


Figure 85. Aggregated 'normalised thickness' results as a function of sample thickness (see text for experimental details).

Where possible it has been attempted to display an error bar on both the thickness (where slightly cuboid particles were observed) and on the normalised thickness (resulting from a combination of depth-resolution measurement error and uncertainties in the simulation input parameters). However, this was not possible in all cases and Figure 85 should strictly be considered a semi-qualitative analysis. Never the less it is sufficient for the purposes of this discussion.

From Figure 85 we can see that the majority of the data points fall around unity ± 40%, even up to thicknesses of ≈250 nm (the thickest particle successfully imaged). This confirms that the 2D object function / convolution model remains valid even at these large thicknesses. Whilst sample thickness can explain results greater than 100% it should also be noted that improper calibration of the STEM defocus-step size (anecdotally known to vary by ≈10% between columns) can lead to underestimates of the apparent thickness. As such repeating this experiment after more careful calibration of this step-size may be advantageous.

One subtlety that should be observed though, is that these data are calculated from the depth-resolution of the laterally-spatially-resolved image-content and it is this which appears to behave as if 2D. As thickness increases the relative strength of this signal versus the constant background reduces. In this case electron de-channeling and sample-probe interactions far from the probe cross-over may be contributing to the diffuse background incorporated in the images and should not be neglected for any quantitative investigation. Unfortunately the focal-series raw-data recorded for Figure 85 were not acquired in a method suitable for quantitative STEM [102], [127].

For the very thickest sample examined at 250 nm it was also possible to observe a small secondary maxima in the intensity profile of one of the 220 spots (approximately one seventh the intensity of the main peak), Figure 86.

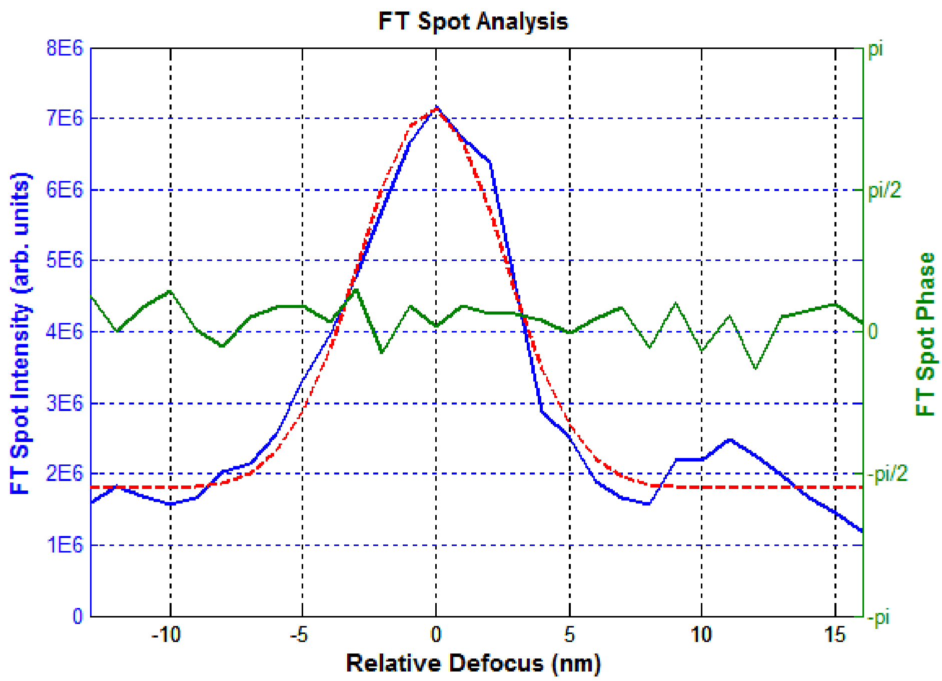


Figure 86. Intensity and phase plots of the 220 reflection from a focal-series data-volume from the 250 nm thick MgO cube from Figure 85. It should be remarked that the width of the modulus profile is approximately 35x narrower than the sample thickness.

This secondary maxima is an excellent example of convolution model between the 2D object and the 3D-OTF as it corresponds perfectly with the secondary maxima observed in the simulated OTF for the case of spherical aberration (Figure 44 - right). In this case the secondary maxima manifesting on the positive (over-focus) side indicates a small negative spherical aberration remained after hardware tuning. The defocus-width of the profile shown in Figure 86 (FWHM) is ≈7 nm, *around 35 times smaller than the thickness of the sample used*. The secondary peak, whilst difficult to measure accurately, has the same width again. This coupled with the roughly equal background intensity at large defocii on both side, shows that the 2d-object model remains valid up till at least this thickness. This is an essential property for the hope of image retrieval by focal-series in the STEM in the analogous manner that the weak-phase-object assumption must not be violated in HR-TEM FSR [108].

4.3. Determining Remnant Aberrations from STEM Focal-Series

From the form of the three-dimensional STEM OTF we expect the defocus-surface of maximal contrast transfer to be non-flat with respect to defocus. The form of the defocus-surface can be related to the strengths and symmetries of the remnant aberrations. The azimuthal symmetry can be interpreted intuitively but the radial order (upward or downward curvature of the surface) must be evaluated analytically.

Let us begin with the aberration function χ(k):

$$\chi(k) = \pi \, \lambda \, C_{1,0} \, k^2 \; + C_{3,0} \frac{\pi \, \lambda^3}{2} \; k^4 \; + \; C_{5,0} \frac{\pi \, \lambda^5}{3} \; k^6 + \; \ldots \qquad (24.1)$$

$$\chi(k) = \frac{2\,\pi}{\lambda} \; \left[\frac{C_{5,0} \, k^6 \, \lambda^6}{6} + \frac{C_{3,0} \, k^4 \, \lambda^4}{4} + \frac{C_{1,0} \, k^2 \, \lambda^2}{2} + \cdots \right] \qquad (24.2)$$

For a given spatial frequency in the image, say g, then the transferred contrast will be at its most intense when the intensity of the fringes in the overlap region of the two corresponding disks is maximised.

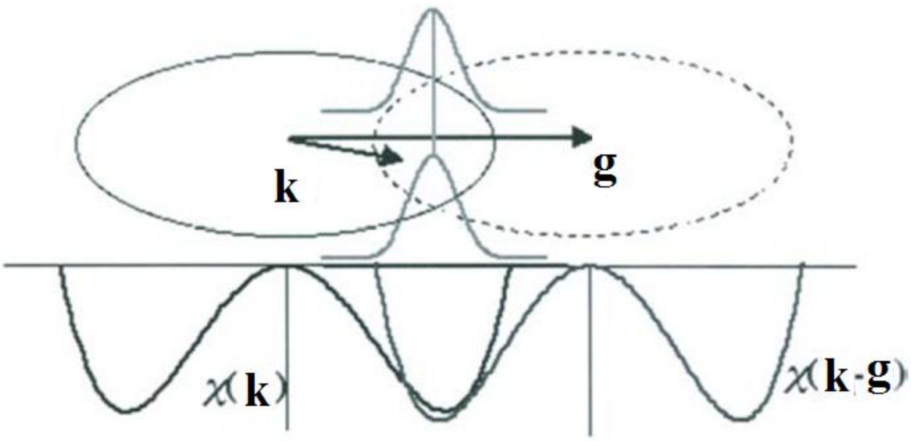


Figure 87. Aberration functions, χ, for the un-scattered beam k and the scattered beam g. At sufficiently large defocus their aberration function become similar resulting in a peak in the integral within their overlapping disks. The finite chromatic spread in the beams results in a damping envelope shown by the grey Gaussian profile over this overlapping disk. Image credit: [46]**.**

This will occur at a given value of defocus where the aberration functions of the two reflections are most similar. First then, we require an expression for the difference between the two aberration surfaces of interest, χ_{diff}.

$$\chi_{diff} = \chi(k) - \chi(k-g) \tag{25}$$

We expand this using the form given in equation 24.2. Here we will only consider the round aberrations but the same expression could be expanded out with additional terms for the even –fold azimuthal symmetry aberrations also. Odd-fold azimuthal symmetry aberrations need not be incorporated as, considering again Figure 87, the vector from k to g and g to k will lie on opposite sides of their odd-fold symmetry aberration surface, resulting in no defocus dependence of their overlapping-disk contrast.

$$\chi_{diff} = \frac{2\,\pi}{\lambda}\left[\frac{C_{5,0}\,k^6\,\lambda^6}{6} + \frac{C_{3,0}\,k^4\,\lambda^4}{4} + \frac{C_{1,0}\,k^2\,\lambda^2}{2}\right] - \frac{2\,\pi}{\lambda}\left[\frac{C_{5,0}\,(k-g)^6\,\lambda^6}{6} + \frac{C_{3,0}\,(k-g)^4\,\lambda^4}{4} + \frac{C_{1,0}\,(k-g)^2\,\lambda^2}{2}\right]$$

(26.1)

$$\chi_{diff} = \frac{2\pi}{\lambda} \left[\frac{1}{6} C_{5,0}\, k^6 \lambda^6 + \frac{1}{4} C_{3,0}\, k^4 \lambda^4 + \frac{1}{2} C_{1,0}\, k^2 \lambda^2 - \frac{1}{6} C_{5,0}\, (k-g)^6 \lambda^6 - \frac{1}{4} C_{3,0}\, (k-g)^4 \lambda^4 - \frac{1}{2} C_{1,0}\, (k-g)^2 \lambda^2 \right]$$

(26.2)

$$\chi_{diff} = 2\pi\lambda \left[\frac{1}{6} C_{5,0}\, k^6 \lambda^4 + \frac{1}{4} C_{3,0}\, k^4 \lambda^2 + \frac{1}{2} C_{1,0}\, k^2 - \frac{1}{6} C_{5,0}\, (k-g)^6 \lambda^4 - \frac{1}{4} C_{3,0}\, (k-g)^4 \lambda^2 - \frac{1}{2} C_{1,0}\, (k-g)^2 \right]$$

(26.3)

$$\chi_{diff} = 2\pi\lambda \left[\frac{1}{6} C_{5,0}\, k^6 \lambda^4 + \frac{1}{4} C_{3,0}\, k^4 \lambda^2 + \frac{1}{2} C_{1,0}\, k^2 - \frac{1}{6} C_{5,0}\, \lambda^4 [\, k^6 - 6k^5 g + 15k^4 g^2 - 20k^3 g^3 + 15k^2 g^4 - 6kg^5 + g^6] - \frac{1}{4} C_{3,0}\, \lambda^2 [\, k^4 - 4k^3 g + 6k^2 g^2 - 4kg^3 + g^4] - \frac{1}{2} C_{1,0}\, [\, k^2 - 2kg + g^2] \right]$$

(26.4)

$$\chi_{diff} = 2\pi\lambda \left[\, C_{5,0}\, \lambda^4 \left(k^5 g - \frac{5}{2} k^4 g^2 + \frac{10}{3} k^3 g^3 - \frac{5}{2} k^2 g^4 + kg^5 - \frac{1}{6} g^6\right) + C_{3,0}\, \lambda^2 \left(\, k^3 g - \frac{3}{2} k^2 g^2 + kg^3 - \frac{1}{4} g^4\right) + C_{1,0} \left(kg - \frac{1}{2} g^2\right) \right]$$

(26.5)

When the two aberration functions are most similar we expect a stationary point in this difference function. This can be solved by setting the differential of this difference function equal to zero:

$$\left(\frac{\partial \chi_{diff}}{\partial k}\right) = 2\pi\lambda \left[\, C_{5,0}\, \lambda^4 (5k^4 g - 10k^3 g^2 + 10k^2 g^3 - 5kg^4 + g^5) + C_{3,0}\, \lambda^2 (\, 3k^2 g - 3kg^2 + g^3) + C_{1,0}(g) \,\right] = 0$$

(27)

Which because of the chromatic damping envelope [46] it is reasonable to solve this expression at the centre of the overlapping disks, that is the point k = $^g/_2$:

$$2\,\pi\,\lambda\left[\;C_{5,0}\,\lambda^4\left(5\left(\frac{g}{2}\right)^4 g-\;10\left(\frac{g}{2}\right)^3 g^2+\;10\left(\frac{g}{2}\right)^2 g^3-\;5\left(\frac{g}{2}\right)g^4+\;g^5\right)\right.$$
$$\left.+\,C_{3,0}\,\lambda^2\left(\;3\left(\frac{g}{2}\right)^2 g-\;3\left(\frac{g}{2}\right)g^2+g^3\right)+C_{1,0}(g)\;\right]=0$$

(28.1)

$$2\,\pi\,\lambda\left[\;C_{5,0}\,\lambda^4\left(\frac{5}{16}g^5-\frac{10}{8}g^5+\frac{5}{2}g^5-\frac{5}{2}g^5+g^5\right)+C_{3,0}\,\lambda^2\left(\frac{3}{4}g^3-\frac{3}{2}g^3+g^3\right)+C_{1,0}(g)\right]$$
$$=0$$

(28.2)

$$2\,\pi\,\lambda\left[-\tfrac{1}{16}C_{5,0}\,\lambda^4 g^5\;+\tfrac{1}{4}C_{3,0}\,\lambda^2 g^3+C_{1,0}g\right]=0 \qquad (28.3)$$

$$C_{1,0}=\tfrac{1}{16}C_{5,0}\,\lambda^4 g^4-\tfrac{1}{4}C_{3,0}\,\lambda^2 g^2 \qquad (2.4)$$

But also (from equation 16.2) we can note that the convergence angle, θ, equals λ.g , leaving:

$$C_{1,0}=-\tfrac{1}{4}C_{3,0}\,\theta^2+\tfrac{1}{16}C_{5,0}\,\theta^4-\;\dots \qquad (29)$$

Equation 29 then leaves us with an expression describing the defocus dependence of the surface of maximal contrast transfer as a function of θ for all odd-order aberrations, parameterised by the aberration coefficients. When non-round aberrations are present these will also depend on the azimuthal angle ϕ.

It should also be noted that in Equation 29 the terms contain only 1st, 3rd, 5th etc. odd-order aberrations. Comparing this result with the table of common aberrations (Table 3) we see that even-order aberrations (those with odd-fold azimuthal symmetry) are not represented. From this result we can conclude that only odd-order, and round / even-fold symmetry, aberration are expected to give a defocus shift or curvature of the surface of maximum contrast transfer. This if further verified on inspection of the 3D OTFs presented in Figure 50 (three-fold astigmatism), and Figure 56 (coma) whose forms remain flat with respect to defocus. It should however be noted, that these aberrations would be expected to yield a phase change across the surface of maximal contrast transfer with the form given in Figure 51 and Figure 57. So far the technique developed has concentrated on the defocus shifting of intensity maxima with respect to defocus to compensate for round and even-fold symmetry aberrations. This has proven largely sufficient up until this point as small two-fold astigmatism (of first or third order) and small remnant spherical aberration appears to frequently dominate experimental data. However, in the future, a full treatment which includes the odd-fold azimuthal-symmetry phase variations across the maximal-contrast-transfer-surface would be merited.

To test the aberration measurement accuracy, two evaluation routes were explored; comparison with hardware-corrector readout, and observation of a known introduced aberration.

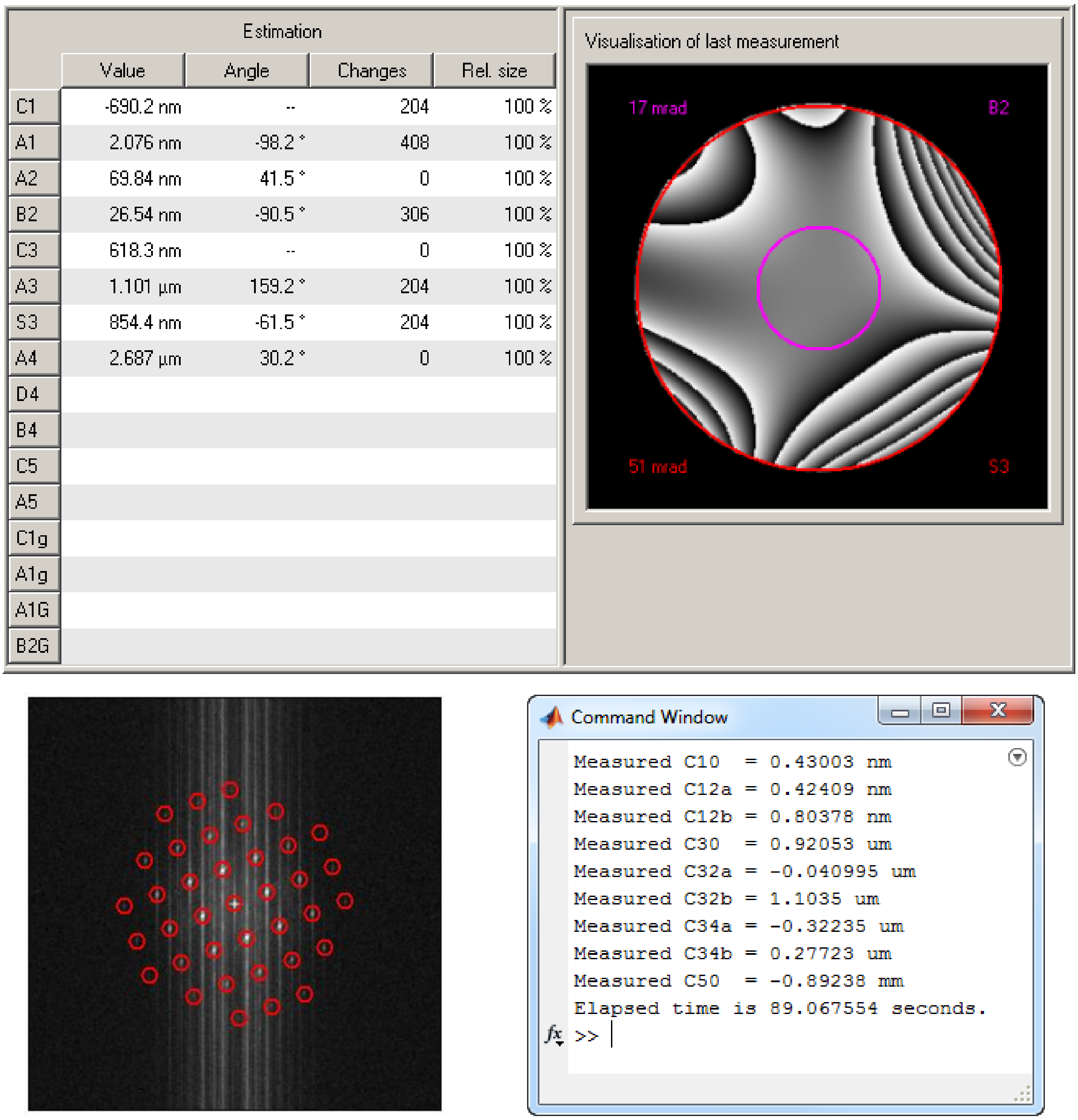


Figure 88. Comparison of aberration measurement methods. Top panel shows the CEOS hardware-corrector readout immediately following a tuning cycle. Lower-left panel shows the FT-spots tracked for the aberration measurement, and lower-right – the fitted aberrations corresponding to the maximal contrast-transfer surface up to $C_{5,0}$.

After the corrector readout was taken (Figure 88) fine-manual adjustments of two-fold astigmatism and focus were performed on a *live* experimental image before a 30-frames series of images were recorded with a 1 nm step-size (≈ -18 nm → ≈ +11 nm). The first aberration suitable for quantitative comparison, therefore, is the spherical aberration ($C_{3,0}$). The values from the two measurement methods are compared in Table 5; as scan-orientation is wholly arbitrary and could not be correlated across the two measurements only the combined magnitude of the aberrations is shown.

Table 5. Hardware readout of remnant aberrations and the diagnosis from focal-series analysis. *$C_{5,0}$ was not measured directly but had been previously measured for the instrument used.

Aberration	Hardware Readout	Focal Series Diagnosis	Relative Strength
$C_{3,0}$ / C3	618 nm	921 nm	148% (overestimate)
$C_{3,2}$ / S3	854 nm	1104 nm	129% (overestimate)
$C_{3,4}$ / A3	1.101 µm	0.425 µm	38.6% (underestimate)
$C_{5,0}$ / C5	≈ -1.5 mm *	-0.89 mm	59.4% (underestimate?)

In general reasonable agreement is found between the measurements from the hardware-readout and the focal-series diagnosis. It would also appear, that in general, the lower order aberrations are overestimated (relatively) whereas the higher orders are underestimated. As no third measurement method was available it is impossible to confirm which is more accurate, however it is entirely reasonable that such a discrepancy is observed. The discrepancy could be caused by one or more of several factors including:

- the favouring of lower order aberrations in the contrast transfer surface fitting due to the increased-relative-brightness of the lower spatial frequency FT-spots,
- a small focal drift during the acquisition or equally by poor calibration of the STEM defocus step-size (nominally 1 nm),
- outdated calibrations of the hardware corrector's tilt / defocus-step calibrations,
- sample damage affecting the intensity of the very high frequency spots (discussed further in Section 5.8), or
- insufficient parameters in the fitting model not accommodating odd-fold azimuthal symmetry aberrations (discussed further in Section 6.3).

Next to evaluate the ability of the method to detect a wilfully introduced aberration, the hardware corrector was deliberately 'mis-tuned' with respect to two-fold astigmatism introducing a known 8 nm of $C_{1,2}$. To illustrate this mis-tuning, two example frames from the recorded focal-series are presented in Figure 89.

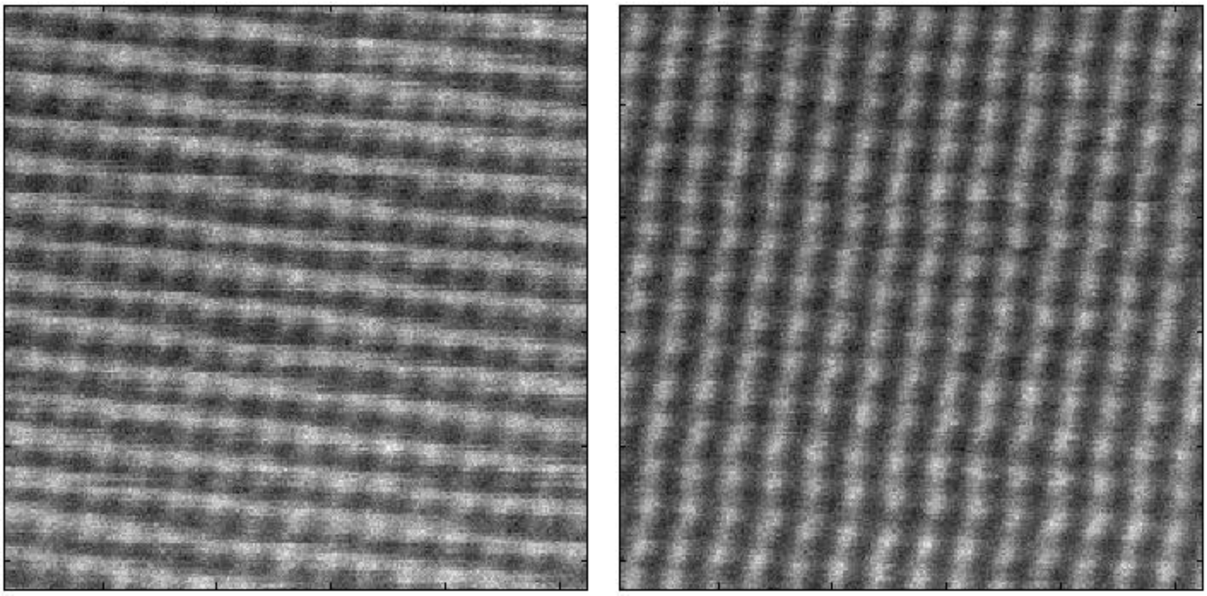


Figure 89. Enlargements from example frames of the mistuned MgO focal-series acquisition. Left, 3 nm under-focus (blurred horizontally) and right, 3 nm over-focus (blurred vertically).

As descrbed above the FT-spot intensities were tracked and the maximal-contrast-transfer surface fitted parametrically. Figure 90 (top) shows the FT-spot intensity iso-surface rods and the fitted maximal-contrast-surface throught them. In the presence of such a large two-fold astigmatism the central ‘pinch’ feature that represents the defocus discontinuity dominates so the same surface has been plotted in plan view for clarity (Figure 90 – lower left). This represetnation immdiatly shows the clear two-fold symmetry and the reader may find it interesting to compare this with the equivalent optical measurment used by opticians (Appendix 8.5).

The fitted surface gives $C_{1,2a}$ and $C_{1,2b}$ coefficients but it is convenient to convert this to an overal magnitude (again the orientation / scan-rotation relationship is arbitary). This gives a value of 8.23 nm which agrees perfectly with the 8 nm deliberate mis-alignment. The deviation of the two most likely lies in a small focal drift during the data acquisistion or the limited total focal-range used for the estimation.

As an aside, inspecting any given circuit around the surface shown in Figure 90 (varying direction but fixed spatial frequency) yields a representation simmilar to that presented by Baba et al for the amorphous specimen case (reference [89], figure 8).

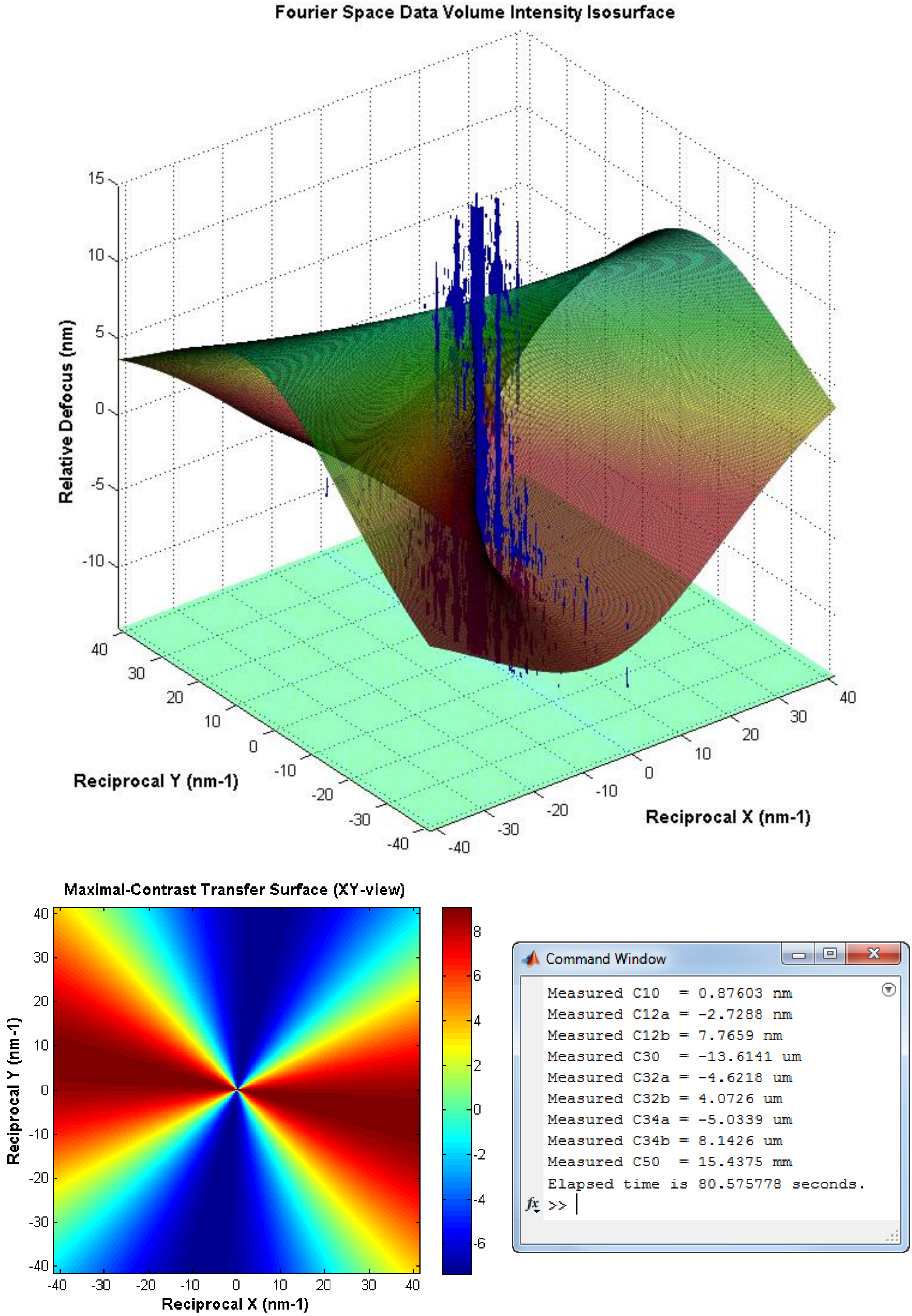


Figure 90. Example aberration measurement of the deliberately stigmated data shown in Figure 89. Top panel shows the FT-spot rods used to track contrast transfer intensity and superimposed the fitted maximal contrast-transfer surface. Lower-left panel shows this same surface but viewed from above; the colour-scale shows nanometres relative defocus. Lower-right panel shows the output of the aberration measurement.

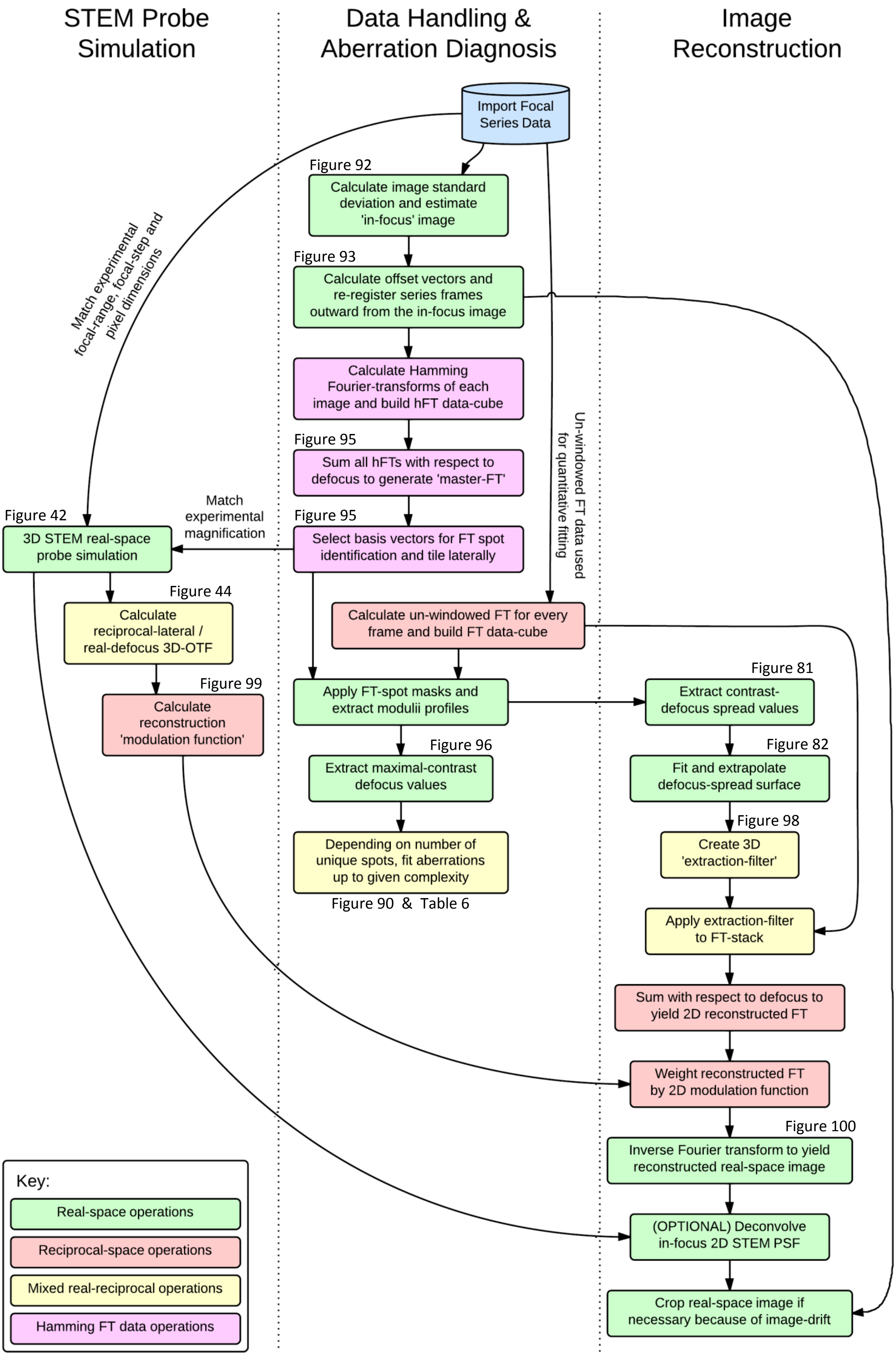
5. Compensating for Remnant Aberrations in the STEM

Before discussing the image reconstruction procedure a brief recap is merited. So far in Section 2.1 we have seen that STEM probes with small remnant aberrations have a complex and non-flat transfer behaviour with respect to defocus, and in Section 4.1 we saw that data acquired by through-focal series exhibits these same variations in the defocus dependence of its lateral spatial-frequency content; in Section 4.2 it was shown that the 2D object function model appears valid (with respect to laterally spatially resolved contrast at least) and that therefore the overall data volume must contain information about the 3D probe as well as the sample; lastly in Section 4.3, it was demonstrated how the presence of aberrations lead to the defocus surface of maximal-contrast-transfer to sweep as a function of both the spatial-frequency and azimuth (in the presence of non-round aberrations). Using these themes a new STEM focal-series reconstruction method was developed which offers the potential for both signal-noise and resolution improvements, as well as the ability to mitigate the effects of low-frequency image distortion and even fluctuations in the microscope emission current.

In the first half of this Chapter (sections 5.1 - 5.5) a new methodology to retrieve a single aberration-compensated image from a focal-series will be presented with examples (section 5.6). This is followed by section 5.7, where the performance of the new STEM FSR routine will be evaluated both relative to a single in-focus experimental image, and to the common noise-reducing practice of image averaging. Lastly in Section 5.8 some subtle effects observed will be discussed.

The content of this thesis will focus on the academic outcomes of this new method rather than on the details of the software developed. However, as the two are inextricably linked, a flow-chart schematic of the software processing steps is included that the reader may wish to refer to.

Figure 91. (Overleaf) Flow-chart of main image reconstruction steps. Operations are divided by the vertical dashed lines into three themes of probe-simulation, aberration diagnosis and image reconstruction. Colour code indicates the coordinate space of the various operations and figure / table cross-references indicate illustrative examples of various steps shown elsewhere in this work.


STEM Probe Simulation
Data Handling & Aberration Diagnosis
Image Reconstruction
Import Focal Series Data
Match experimental focal-range, focal-step and pixel dimensions
Figure 92
Calculate image standard deviation and estimate 'in-focus' image
Figure 93
Calculate offset vectors and re-register series frames outward from the in-focus image
Calculate Hamming Fourier-transforms of each image and build hFT data-cube
Figure 95
Sum all hFTs with respect to defocus to generate 'master-FT'
Un-windowed FT data used for quantitative fitting
Figure 42
3D STEM real-space probe simulation
Match experimental magnification
Figure 95
Select basis vectors for FT spot identification and tile laterally
Figure 44
Calculate reciprocal-lateral / real-defocus 3D-OTF
Calculate un-windowed FT for every frame and build FT data-cube
Figure 99
Calculate reconstruction 'modulation function'
Apply FT-spot masks and extract modulii profiles
Figure 81
Extract contrast-defocus spread values
Figure 96
Extract maximal-contrast defocus values
Figure 82
Fit and extrapolate defocus-spread surface
Figure 98
Create 3D 'extraction-filter'
Depending on number of unique spots, fit aberrations up to given complexity
Figure 90 & Table 6
Apply extraction-filter to FT-stack
Sum with respect to defocus to yield 2D reconstructed FT
Weight reconstructed FT by 2D modulation function
Figure 100
Inverse Fourier transform to yield reconstructed real-space image
Key:
Real-space operations
Reciprocal-space operations
Mixed real-reciprocal operations
Hamming FT data operations
(OPTIONAL) Deconvolve in-focus 2D STEM PSF
Crop real-space image if necessary because of image-drift

5.1. Lateral Re-registration of Focal-series Images

Experimentally it is challenging to reduce stage / sample drift to zero. In Section 3.2.5 the effects that such drifts exhibits in a single STEM image were discussed; here, in the context of an image-series acquisition the effect of drift becomes one of image mis-registration – that is features from one image are not correctly aligned with the preceding frame. To enable faithful focal-series reconstruction (or image frame averaging) it is essential to accurately re-register these images.

Image re-registration algorithms are well established but are generally optimised for photographic style images and can produce erroneous results with the levels of noise commensurate with electron micrographs. Additionally most atomic resolution STEM images comprise substantively of an array of often highly similar atomic columns. In the presence of noise it can also become a problem that automatic image registration methods find an erroneous solution that is the vector sum of the true solution plus an integer multiple of a crystallographic vector.

Furthermore, re-registration techniques are often designed for the realignment of a single pair of images and not an entire series. This can present further problems as it is no longer immediately apparent which image should be used as the 'reference' image and which the 'moving' one. Choices for the reference image could be the first image, the immediately previous or subsequent image, or the middle image from a stack. If the first (or last) image from a focal series is chosen then this will be intentionally out of focus and more likely to give uncertain or erroneous registration results. If the reference image is taken as one adjacent to the current image of interest, then global offsets can be found from the vector sum of the offsets between individual pairs preceding a given frame. Unfortunately with this method any error between an individual pair corrupts the entire summation beyond that point. For this reason it is desirable to directly reregister all images to one 'master' frame; at least in this scenario an individual mis-registration will only affect a single frame in the series and not all that follow. The challenge then becomes the choice of this 'master' frame, and while registration to a fixed frame avoids the need to sum individual reference vectors, it does rely on the chosen frame being free of any corruption or distortion.

What has been developed for this application is a hybrid solution; an initial master reference frame is selected from the series from which the offsets are calculated for the frames immediately above and below. The re-registration algorithm used is based around a discrete Fourier transform method [128] and allows for a solution to be found to sub-pixel offset precision with a high degree of computation efficiency. As the master frame and its neighbours are all close to optimum focus they should all contain clearly resolved atomic features, allowing the re-registration to produce a clear solution. Next these three frames can be averaged to produce a new reference frame with reduced random noise and greater SNR. This process can be repeated, each time re-registering frames immediately before and after those already offset, growing the registered volume from the centre outward. Every iteration also generates an improved (higher SNR) master image to which all images are realigned. The smoothing effect that this has on the reference image removes single-pixel random noise superbly and eliminates the risk of 'noise registering onto noise' and 'crystal vector hops' that can be a problem with pairs of low SNR images.

The problem is now reduced to the choice of the initial master frame for the first loop of the re-registration. It may be most direct to simply specify this manually each time the FSR is to be performed. However, automation to the general case allows for a higher throughput of data analysis and generally leads to reduced errors. One means would be to measure the image SNR or resolution for every frame (using the methods in Section 3) and to select the quantitatively 'clearest' image as the initial master frame. However, for series of 30 images or more this would take several minutes to perform. An expedient 'short-cut' is to calculate the standard deviation of all the pixel values within the images, as an inference of their contrast behaviour. Those images with greater contrast will generally correspond to those that exhibit larger SNRs. This is somewhat similar conceptually to using the 'mean phase-contrast index' in HRTEM exit-wave restoration to find a starting estimate for defocus [107]. The standard deviation of an example data set is shown below (Figure 92).

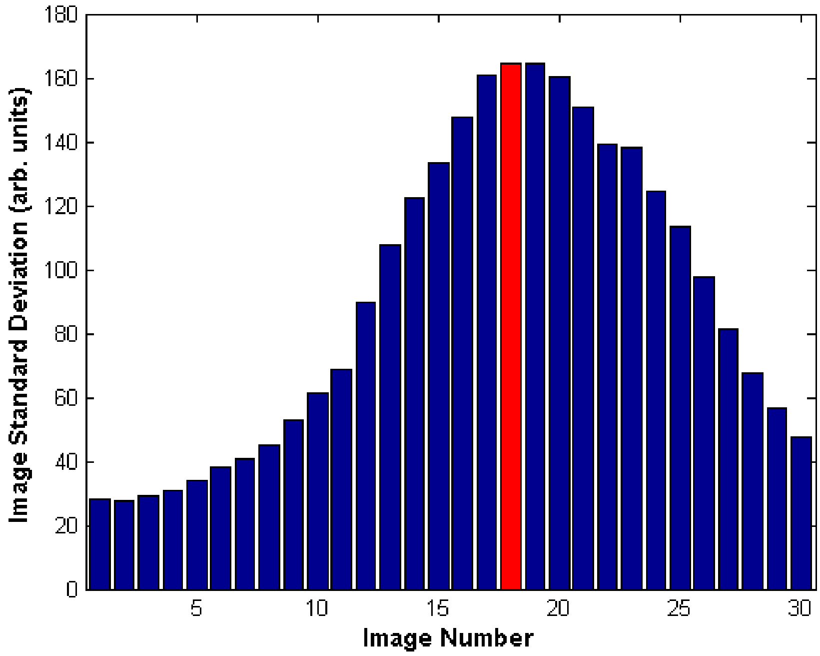


Figure 92. Standard deviations of the individual image frames from a 30 frame focal series. The peak was identified to be frame eighteen (red).

Figure 92 shows the increase in standard deviation near the centre of the focal series. The peak was identified at image eighteen of the series, although image nineteen would have served equally well. In fact, it can be convenient to manually select an image to one side of the automatic maxima should it be found to contain any temporary instability (such an EM suite door closing).

After selecting the master frame this can be designated as the centre of the focal-range and the frames re-registered outwardly from this. Figure 93 shows example re-registration offset outputs.

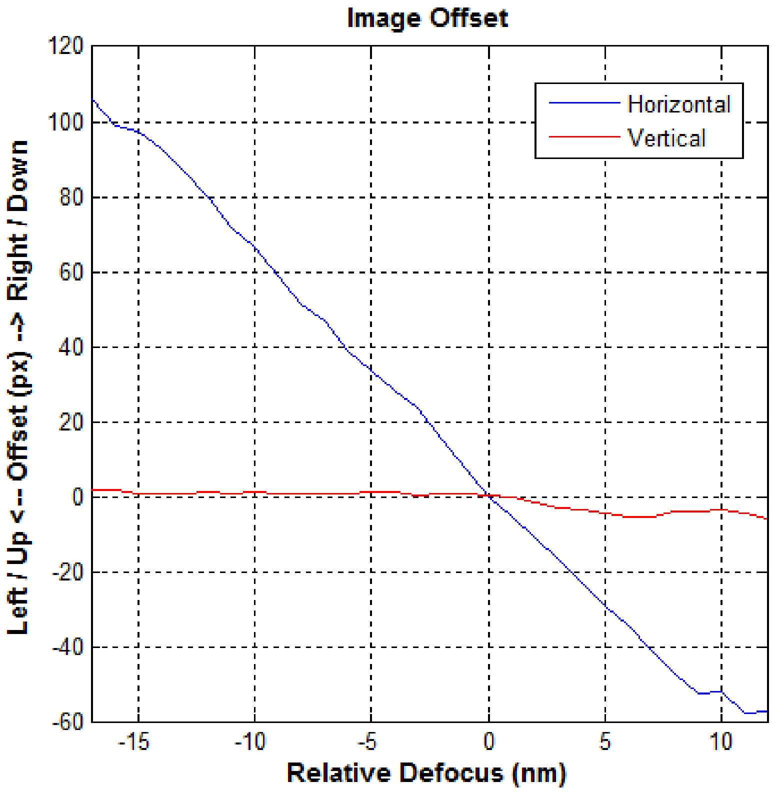


Figure 93. Example image-offset results from re-registration. From Figure 92 image number 18 was designated as the centre of the series (zero relative defocus) from which the horizontal and vertical offsets were measured.

In Figure 93 we see that the master image has, by definition, an offset of zero in both the horizontal and vertical directions.

Following re-registration it may be necessary to crop the data-volume to remove any 'overhanging' regions – that is any areas that are not present in all images due to the drift offsets. In this case very large drifts will have the effect of leaving an insufficiently small data-volume for onward processing and image restoration.

Re-registration Precision

The re-registration method used allows for the precision to be specified in any integer fraction of a pixel, from nearest-whole-pixel or finer [128]. This method is highly robust, having been tested and found suitable for images where crystal lattice fills the whole or part of the field of view, nanoparticles and grain/twin boundaries; in-fact the presence of point or line defects increases the reliability of the re-registration with the changes of 'integer crystal-cell-unit hops' being greatly reduced in their presence.

To inspect the effect of re-registration offset precision the dataset discussed in Figure 92 was re-registered to varying degrees of precision, while the phase of one of the lowest spatial-frequency spots was monitored. The results are compared against the phase of the same spot – but for the unregistered case, Figure 94.

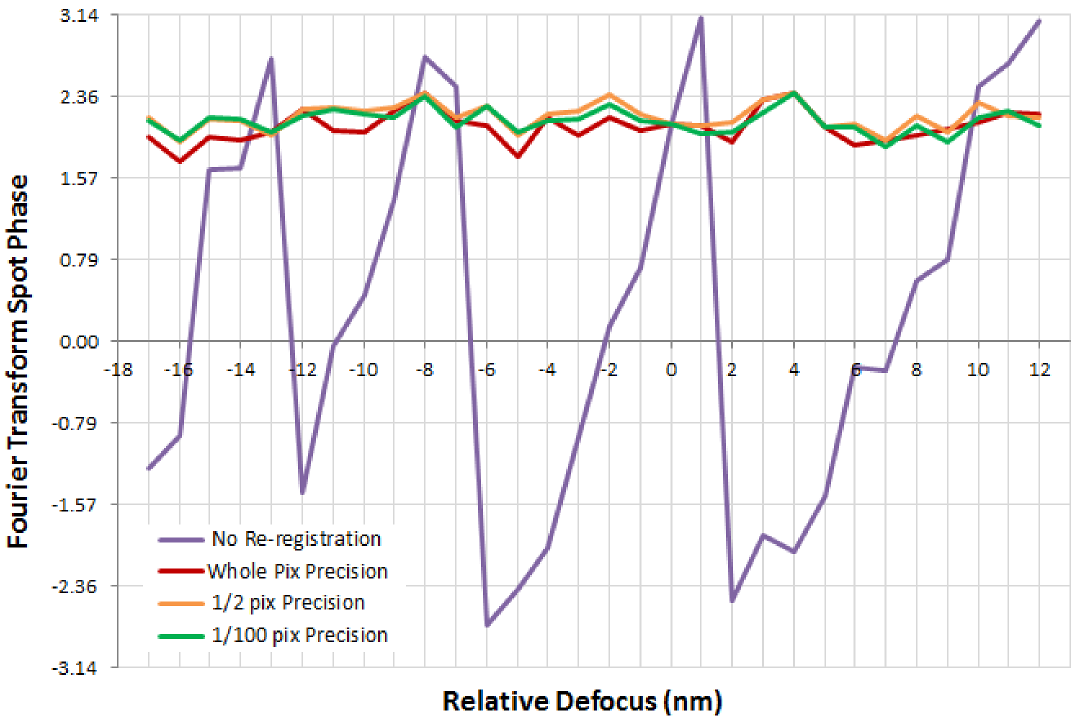


Figure 94. Fourier transform spot though-focus phase-tracking for each of no-registration, integer pixel registration $^1/_2$ pixel and $^1/_{100}$ pixel precisions. Note again, that as no image offset is applied for the zero relative-defocus image all phases are coincident at that point.

The most striking feature of Figure 94, is the saw-tooth profile of the un-registered FT-spot phase. This continually increasing phase-ramp is a clear indicator of uncorrected image-drift, and when measured from the experimental Fourier transforms appears wrapped into the $-\pi$ to $+\pi$ range. It would be impossible to attempt any reliable image averaging / image-reconstruction with such unregistered data. The further three plots on this figure show the phase after image registration. Correcting the image offsets to the nearest integer-pixel vector offset has immediately overcome the severe phase ramp of the unregistered case but does still exhibit some small fluctuation. Comparing many precisions of registration it is found that this phase becomes increasingly stable with finer and finer allowed offsets. Figure 94 shows this for the case of nearest-half and also 1/100 pixel precisions. Beyond 1/100 pixel precision, no further improvement in the phase stability was observed.

The reader will observe that there remains a small undulation in the FT-spot phase even after 1/100 pixel precision re-registration. This arises from the fact that this is only the plot from a singlepixel (spatial-frequency) from the Fourier transform. The image re-registration must include all of the real-space image globally, regardless of any real-space instabilities. Any real-space image distortions persist as phase instabilities in the Fourier transform that may prevent a perfect re-registration and yield such small fluctuations in the phase of individual FT-spots. Additionally, any change in the rate of image drift during the data-acquisition may lead to slight changes in the lattice plane angles (Section 3.1.4) and hence the position of FT-spots. This will also propagate through to this analysis as a small fluctuation in the phase of the FT-spots.

Finally, it should be noted that any re-registration by a non-integer pixel vector will necessarily result in a resampling of the image data to allow it to be recorded as an array of pixels. This resampling is unavoidably equivalent to a two-dimensional filter which highly dampens any single pixel noise. As a result this resampling can have the effect of a noise-filtering and will increase the SNR of the image frames. Where a quantitative SNR analysis is required then integer-pixel offset vector re-registration should be used for a wholly fair comparison (later in Section 5.7.2 such integer-precision re-registration were used for exactly this reason). This need for high-precision re-registration serves as an additional reason to acquire experimental data with a fine intrinsic sampling of image features; the higher the sampling of the raw data the lesser the risk of any genuine image information being lost to this resampling filtering effect.

5.2. Tracking the Intensity and Phase of Fourier Spots as a Function of Defocus

The intensity in each pixel of the STEM image, *I*, is a function of the probe position vector, $\boldsymbol{R}$, and can be expressed as $I(\boldsymbol{R})$. For incoherent imaging signals, this image is related to the real-space object-function, $O(\boldsymbol{R})$, through a convolution with the 2D form of the imaging probe. However, the probe must be treated as a three-dimensional entity with the form $P(\boldsymbol{R},z)$. As a result the STEM image formed depends on both the probe position and also its defocus, z:

$$I(\boldsymbol{R}, z) = P(\boldsymbol{R}, z) \otimes O(\boldsymbol{R}) \tag{30}$$

This can be expressed equivalently in reciprocal-space as:

$$\tilde{I}(\boldsymbol{r}^*, z) = \tilde{P}(\boldsymbol{r}^*, z) \, . \, \tilde{O}(\boldsymbol{r}^*) \tag{31}$$

Where now the Fourier transform of the image is the product of the object (expressed in frequency-space) with the probe's optical transfer function for a given defocus. As the representation of the crystalline sample in reciprocal-space is an array of Fourier spots, we expect the 3D transfer function to only be evaluated at these discrete points. However, we can infer the transfer function from analysing the intensity (and phase) of individual Fourier spots as a function of defocus.

5.2.1. Inspecting the Fourier Spots

First this involves identifying the spots to inspect; difficult for highly defocused images but more easy closer to Gaussian focus. To facilitate clear identification of the Fourier spots it is convenient to sum the modulus of all the individual Fourier transforms in the defocus stack, Figure 95.

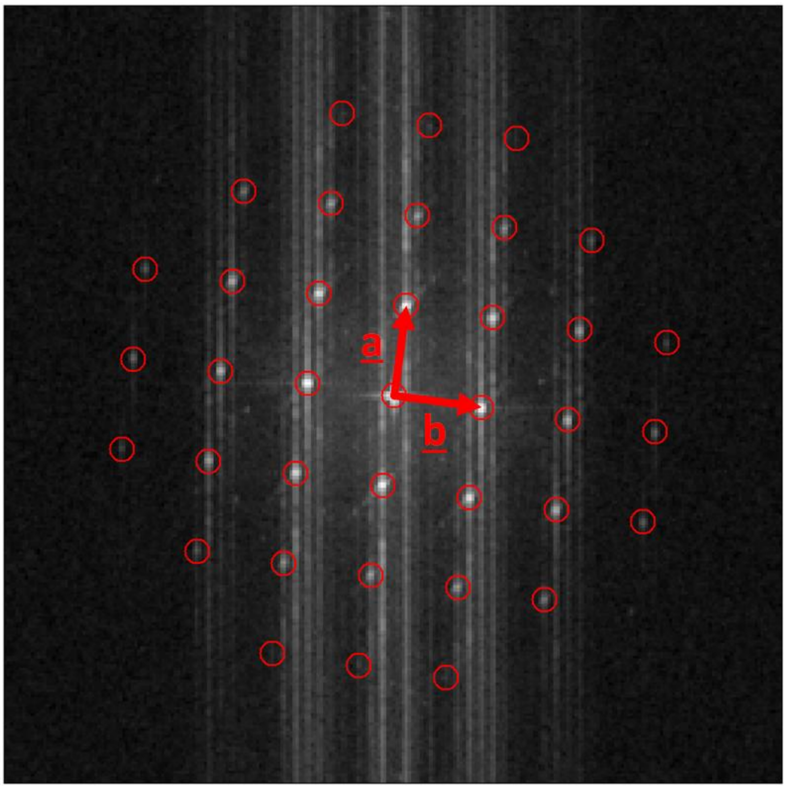


Figure 95. Sum of the moduli of the Hamming Fourier transform (HFT) frames from the example focal-series used in Figure 107. The array of bright spots (labelled with red circles) correspond to the lattice spacing's in the real-space image, vertical streaks through these spots are the result of scanning noise in the image.

Using this summed-FT it is convenient to define a pair of reciprocal space basis-vectors, here labelled **a** and **b**. These define an array that allows all the Fourier spots to be identified and their pixel-coordinates found. It is also advisable to define an outer bound to this array that is at most no larger than the expected information limit as points beyond this will represent only noise. Once these coordinates are determined from this summed FT the same coordinates are used for all individual Fourier-transforms, even if spots in those frames are too weak to observe them independently. Figure 96 shows the result of one such spot-tracking corresponding to the spot above on the end of vector **b**.

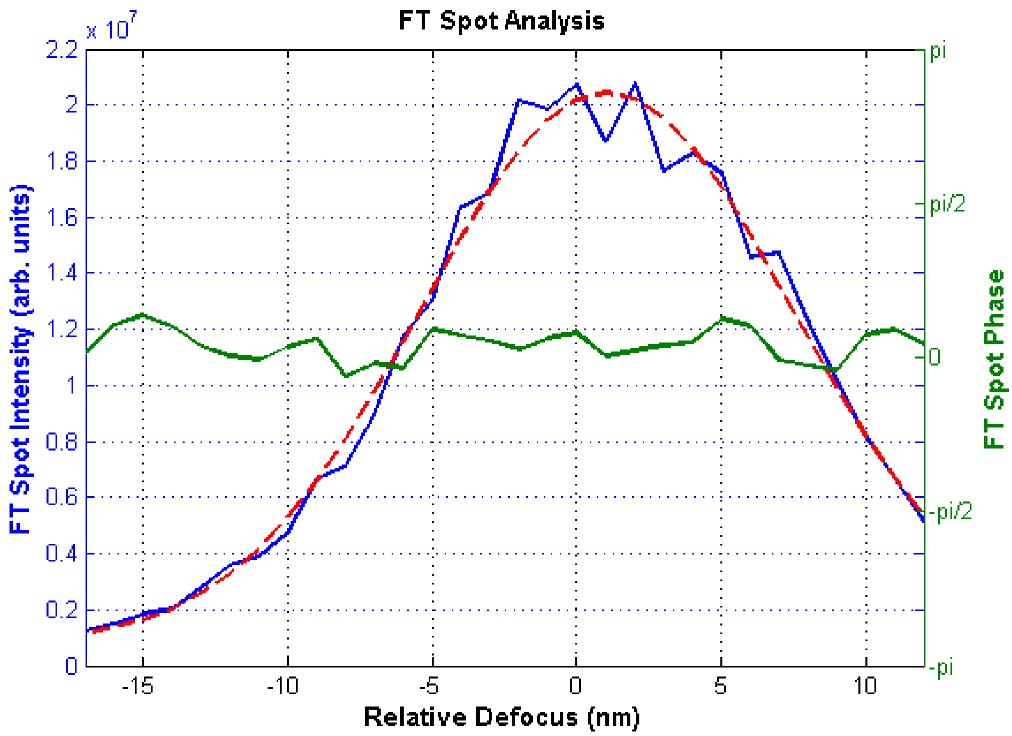


Figure 96. FT-spot tracking information for the FT spot at vector b from Figure 95, showing the modulus in blue (left axis) and phase in green (right axis). The smooth red line corresponds to a Gaussian curve fitted to the modulus data used to calculate the defocus position of the peak contrast transfer and the defocus-spread of same.

Here we see the modulus rise to a maximum at approximately +1 nm relative defocus, this is coincident with the in focus image from visual inspection but also the peak in the standard deviation plot (Figure 92) used in the image re-registration stage. Depending on the quality of the experimental data set this modulus data can exhibit varying degrees of noise; because of this it can be beneficial to fit a Gaussian or some similar form to this profile to more robustly determine the peak position (with respect to defocus). This fitting procedure also yields a measure of the defocus-spread of the contrast. This is equivalent to the vertical extent of the optical-transfer-function shown in Figure 44, and is equal to the depth-resolution at that spatial frequency.

The phase remains relatively constant throughout the series and is generally within the range $^{-\pi}/_{10}$ to $^{+\pi}/_{10}$. This tracking and fitting can be repeated for all the spots highlighted in the summed FT shown in Figure 95 to yield a set of reciprocal-X, reciprocal-Y, defocus data.

5.2.2. Unique Fourier Spots and the Number of Diagnosable Aberrations

The array of FT spots identified in plots such as Figure 95 will always number a total of the form 2n + 1, that is each spot and its conjugate plus the zero-order spot. As we cannot utilise the zero-order spot, and as conjugate spots are not a source of new information we have a finite number of unique data points. The example shown in Figure 95, contains a total of 37 spots yielding 18 *unique* data points.

In this data the resolution and stability performance of the instrument facilitated sub-angstrom performance. However, for other data-sets with more severe instabilities, or larger remnant uncorrected aberrations, fewer unique FT spots may be visible in such a summed FT. In this case may not be possible to track their modulus and phase, nor use them for aberration diagnosis.

In Section 4.3 it was presented that the surface of maximum contrast transfer follows a defocus response dependant on both spatial frequency and remnant aberrations (Equation 29). If we wish to diagnose these aberrations from the FT-spots then we must fit such a surface *parametrically* through the scattered spatial-frequency versus defocus data obtained from all the FT-spots. If this surface is defined parametrically, then coefficients can be extracted that correspond to the strengths of the various remnant aberrations. However, it is impossible to solve for more variables (aberration coefficients) than the number of unique points (FT-spots) that are available. Additionally, as for many aberrations a magnitude *and* direction must be found, in the case that the crystal symmetry of the sample matches the azimuthal symmetry of the aberration of interest (as in the case of cubic materials and two-fold astigmatism for example) an additional spot may be needed to find a unique solution.

Table 6 shows a summary of the aberrations that can be observed qualitatively and compensated for during restoration (orange ticks) and also those which can be diagnosed quantitatively where additional unique data-points exist (symmetry dependent, green ticks).

Table 6. Chart of the identifiable and quantifiable aberrations with various numbers of unique FT spots.

Unique Spots	$C_{1,0}$	$C_{1,2}$	$C_{3,0}$	$C_{3,2}$	$C_{3,4}$	$C_{5,0}$
2	✓	✓	✗	✗	✗	✗
3	✓	✓	✓	✗	✗	✗
4	✓	✓	✓	✓	✗	✗
5	✓	✓	✓	✓	✓	✗
6	✓	✓	✓	✓	✓	✓
7	✓	✓	✓	✓	✓	✓
8	✓	✓	✓	✓	✓	✓
9	✓	✓	✓	✓	✓	✓
10	✓	✓	✓	✓	✓	✓

It should be noted here that odd-fold azimuthal symmetry (even radial order) aberrations are not shown as these cannot be diagnosed from defocus shifts in the maximal contrast-transfer surface (OTFs in Section 2.1 all remain flat with respect to defocus).

5.2.3. FT Spot Streaking and Scan-Direction Rotation

To produce the Fourier transforms summed for Figure 95 the real-space frames were first Hamming window smoothed at their perimeter to reduce cross shaped streaking of the FT spots. However, even taking this precaution, significant instability induced streaking exists running vertically in the FTs. In extreme examples this could lead to mis-measurement of the aberrations and an inaccurate image reconstruction. For this reason it is advisable to slightly rotate the acquisition scan-rotation direction so that FT spot streaking does not impinge upon neighbouring spots.

5.3. Fitting the Maximal Contrast Transfer Surface

To begin the image reconstruction process we must first find the 'surface of maximal contrast-transfer' following the method presented in Section 4.3. Continuing with the same example data-set, the corresponding surface has been fitted in Figure 97.

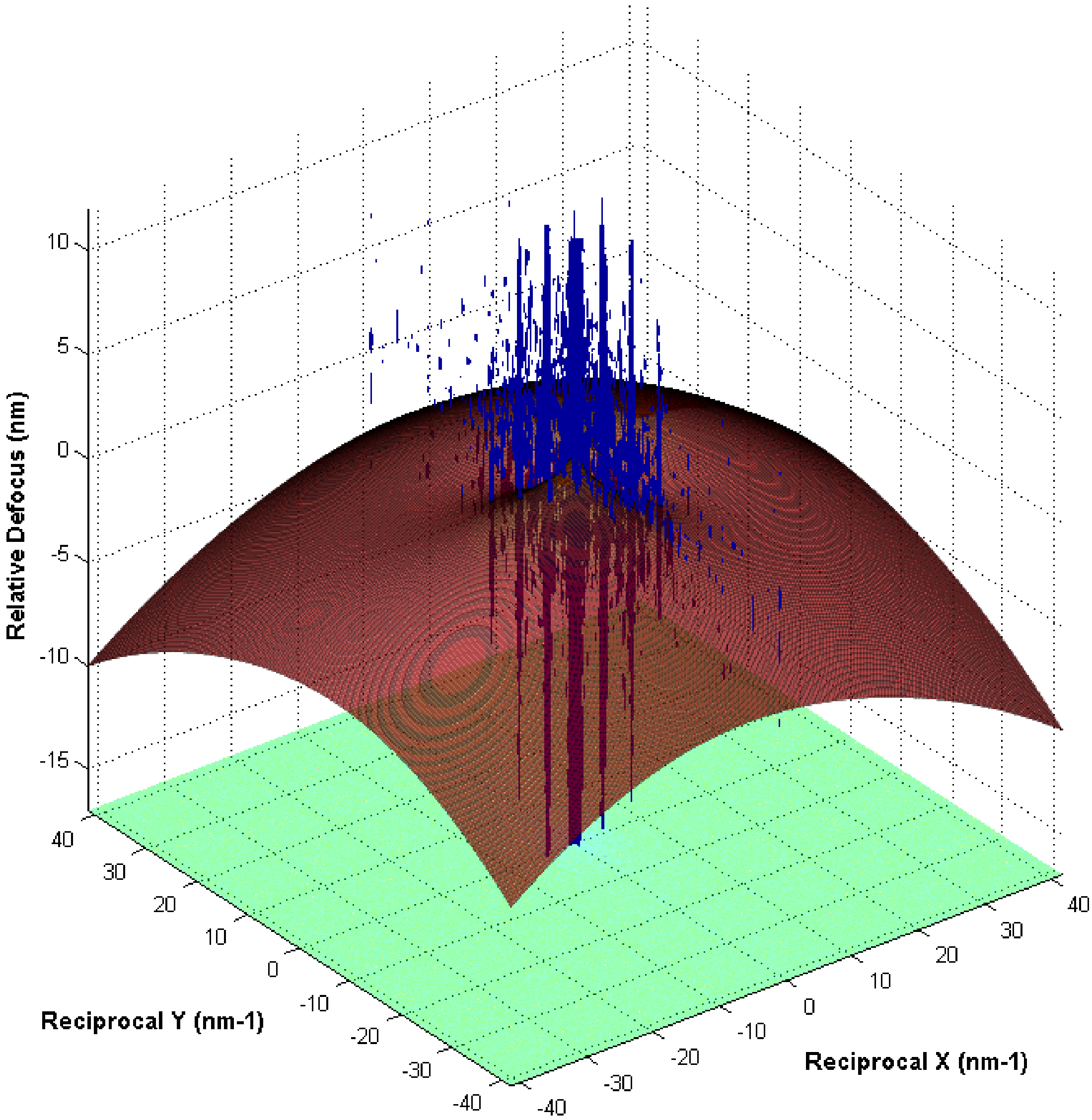


Figure 97. Example FT intensity iso-surface (blue rods), with the parametrically fitted maximal-contrast-transfer surface superimposed (red). This surface is dominated by a two-fold astigmatism (pinch at centre) and a small negative spherical aberration (downward curving bowl).

After this fitting is complete, it is advisable to verify the result by visual inspection and we can indeed see that the surface does intersect each of the iso-surface FT-spot rods at roughly their midpoints. From this fitting we can also extract estimates for the coefficients of the aberrations. For the data shown in Figure 97 (which also corresponds to the examples shown later in Figure 111) aberrations were fitted to include defocus (1.07 nm), two-fold astigmatism (0.83 nm @ 40.6°) and spherical aberration (-125µm). These three features are in turn responsible for a small offset in defocus (arbitrary as the axis was in units relative to the operator's tuning), a pinch in the surface's centre (where the stigmation discontinuity exists) and the smooth downward curve with increasing spatial frequency (square relationship).

Fitting this surface parametrically allows it to be extrapolated to cover all reciprocal-space vectors (the whole field of the Fourier transform) even though the FT-spots are only visible in roughly the central half of the FTs as well as the space between FT spots.

5.4. Filtering Focal-series Data Volumes to Compensate for Remnant Aberrations

From the discussion presented in the previous section, it would now be possible to reconstruct a 'restored Fourier transform'; one could retrieve from the 3D Fourier transform stack the complex value of each FT pixel from the nearest available defocus dictated by the fitted surface. However, such a basic approach would be unsatisfactory in two ways; firstly the fitted surface will often suggest a defocus value between two available experimental values, and secondly, this makes very inefficient use of the contrast known to persist over several nanometres (predicted in Section 2.1 and validated experimentally in Figure 81).A more efficient means to rebuild the restored Fourier transform is to extract a 'weighted average' of all of the available 3D experimental Fourier data volume. Observing that in Figure 81, we often see an approximately Gaussian profile to the intensity of contrast with respect to defocus, this seems a reasonable approach.

This can be compared to the summation of transfer-functions that gives rise to the 'parabolic filter function' in HR-TEM FSR [108], with the key difference being that in HR-TEM the exit-wave is reconstructed "back up" to the exit surface of the sample (after all elastic interactions have occurred), whereas in this newly developed incoherent method, the electron channelling theory effectively projects the sample to behave as a 2D-object incident on the entrance surface of the sample, and it is at this focus for which we reconstruct the image.

To design the 3D filter then we need only the two parameters for a Gaussian curve at every reciprocal space vector (reconstructed FT pixel). The surface presented in Figure 97 gives the defocus centre of the Gaussian profile and the FT-spot persistence from Figure 82 gives the width. Importantly to ensure a reliable, quantitative, reconstruction these Gaussian profiles should be normalised, such that the sum of all the weightings equals to one (over the defocus range corresponding to the experimental acquisition). Mathematically this can be expressed as the filter, *F*, such that:

$$F(\boldsymbol{r}^*, z) = \frac{\tilde{P}(r^*,z)}{\int \tilde{P}(r^*,z)\, dz} \tag{32}$$

where P is the reciprocal-lateral, real-defocus representation of the probe OTF.

The 3D-filter can then be generated such that it has the same dimensions (both physically and in terms of its array-size computationally) as the experimental Fourier-stack. Again being three-dimensional, this can be difficult to visualise, so an X-Z section through the filter (corresponding to the data presented in Figure 97) is shown in Figure 98.

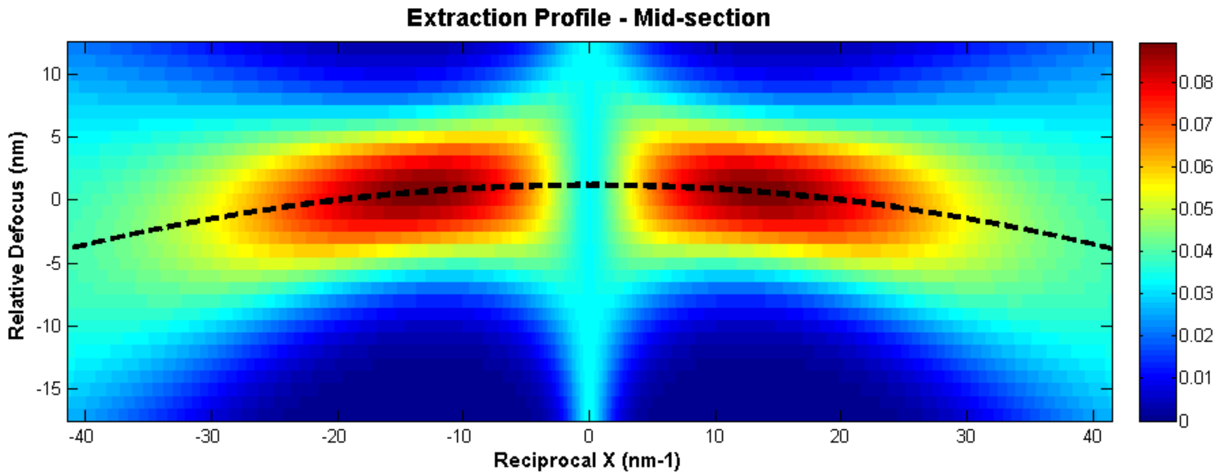


Figure 98. X-Z section through the 3D reconstruction filter. Dashed black line indicates the maximal-contrast-transfer surface from Figure 97.

At first impression the appearance of this filter may seem complex, but its form may be easily understood with reference to Figure 97 and Figure 82. First we should inspect the plot column-by-column (separate spatial frequencies). For every spatial frequency the filter profile is one of a Gaussian, centred on the surface defined by Figure 97 (equivalent to the black dashed line in Figure 98) and with a width dictated by the surface in Figure 82. The small remnant negative spherical aberration is reflected in the downward curvature of the filters 'wings' at increasing spatial-frequency, while the poor depth-resolution of the probe's OTF at very-high and very-low spatial frequencies (Figure 44) causes the filter to become more diffuse at these points. This explains the bounds of the plot; for the intensity we have simply to recall that the filter has been designed to sum to unity with respect to defocus. As a result the z-diffuse regions have a lower peak-strength, while the regions with a narrower depth-resolution are more tightly localised with respect to defocus. Again this reflects the design of the filter to be biased to regions where sample contrast is known to be localised (Figure 81).

Lastly, it should be noted that this filter is purely real in nature, that is, it has no effect on the phase of the restored data. This is both allowable, as the phase is expected to remain constant over the central several nanometres of the OTF (Figure 51), and practicable, as the phase is also observed experimentally to be approximately constant (so long as image data is re-registered appropriately – Figure 94). It is also worth considering the effect of this filter on any random noise present in the images. In the Fourier-transforms this noise will be manifest in the regions of very high spatial frequencies and / or the regions between higher-order FT-spots. If we examine the nature of this filter at such spatial frequencies of 20 nm^{-1} and greater (real-space features smaller than 0.5 Å) the filter is diffuse over perhaps 15 nm of defocus. Because this *real* filter is applied to the *complex* FT-stack data-volume, the result of this is that genuine object-contrast is preserved while the noise (random-phase natured) is progressively averaged towards zero.

Now using this filter we can perform a "pan-defocus weighted-average" image reconstruction, utilising all the available information of our data-volume. However, whilst this filter is able to compensate for remnant probe aberrations, for quantitative STEM imaging the relative weighting of all spatial frequencies must be preserved in the reconstructed image and this is the focus of the following section.

5.5. Fourier-space Weighting to Yield the Aberration Compensated Image

From Section 5.4 we are able to reconstruct the Fourier-transform of the restored image; with each reconstructed-FT pixel drawing information from a few nanometres either side of focus in a weighted-average manner. In this section we will examine how this reconstruction must be modified to preserve the quantitative strengths of the experimental OTF.

In designing the filter care was taken to ensure that the summation with respect to defocus for each lateral spatial frequency equalled unity. If we now consider how this is applied to the profile of the contrast-transfer strength (FT-spot intensity, Figure 81) which is also of an approximately Gaussian form, we see that the weighted defocus-average incorporates parts of the contrast-transfer which are weaker than at its peak. The effect of this is, whilst the total filter Gaussian sums to unity, that the weighted-average modulus is below that of the peak in the experimental contrast-transfer. What is required then is an up-weighting by a factor slightly larger than unity, to restore the transferred contrast-strength to be identical to that of the experimental image (thus preserving the quantitative integrity of the data as faithfully as possible). This is of course required not just for the spatial-frequencies (FT pixels) of the FT-spots but across the whole of the rebuilt FT, and is hence a 2D up-weighting function that must be applied to the rebuilt FT prior to its inversion back to the real-space restored-image. In this work this scaling is referred to as the modulation function, *M*.

The analytical form of this modulation function can be determined from inspecting the difference between the filtered data-cube and the expression for an ideally reconstructed image.

First let us begin with the description of an ideal image defined as the product in reciprocal-space between the 2D object and the (also 2D) OTF of a *perfect probe* at optimum focus:

$$Ideal\ Image\ \ (\boldsymbol{r}^*)\ =\ \tilde{O}(\boldsymbol{r}^*)\,.\,\tilde{P}(\boldsymbol{r}^*, z=0) \tag{33}$$

From the previous section we have our 3D filter, *F*, that is both a function of lateral spatial frequency and also defocus. In applying this filter the reconstructed FT is the sum over all defocii of the product of this with the experimental FT-stack, *I*:

$$Filtered\ Image\ \ (\boldsymbol{r}^*)\ =\ \int I(\boldsymbol{r}^*, z)\,.\,F(\boldsymbol{r}^*, z)\; dz \tag{34}$$

Next we can express the experimental FT-stack, *I*, in terms of the 3D reciprocal-space product of the *non-perfect 3D-OTF* (equivalent to the 3D real-space convolution with the probe including remnant aberrations):

$$Experimental\ FT\ stack\ \ \tilde{I}(\boldsymbol{r}^*, z)\ =\ \tilde{P}(\boldsymbol{r}^*, z)\ .\ \tilde{O}(\boldsymbol{r}^*) \tag{35}$$

Recalling the filter from the previous section, we have the estimate of the probe's OTF defocus spread, normalised by the z-integral back to unity:

$$3D\ Reconstruction\ Filter\ \ F(\boldsymbol{r}^*, z) = \frac{\tilde{P}(\boldsymbol{r}^*,z)}{\int \tilde{P}(\boldsymbol{r}^*,z)\; dz} \tag{36}$$

Substituting equations 35 & 36 into equation 34, the reconstructed FT then becomes the product of the experimental data and our 3D filter summed with respect to defocus to yield the rebuilt 2D Fourier transform:

$$Filtered\ Image(\boldsymbol{r}^*)\ =\ \int \tilde{P}(\boldsymbol{r}^*, z)\ .\ \tilde{O}(\boldsymbol{r}^*)\ .\ \frac{\tilde{P}(\boldsymbol{r}^*,z)}{\int \tilde{P}(\boldsymbol{r}^*,z)\; dz}\; dz \tag{37.1}$$

$$=\ \int\ \ \tilde{O}(\boldsymbol{r}^*)\ .\ \frac{\tilde{P}^2(\boldsymbol{r}^*,z)}{\int \tilde{P}(\boldsymbol{r}^*,z)\; dz}\; dz \tag{37.2}$$

$$=\ \tilde{O}(\boldsymbol{r}^*)\ .\ \frac{\int \tilde{P}^2(\boldsymbol{r}^*,z)\; dz}{\int \tilde{P}(\boldsymbol{r}^*,z)\; dz} \tag{37.3}$$

Comparing equation 37.3 with the earlier expression for a perfect imaging system (equation 33) then the required modulation function, *M,* as:

$$M(r^*) = \tilde{P}(r^*, z = 0) \cdot \frac{\int \tilde{P}(r^*,z)\, dz}{\int \tilde{P}^2(r^*,z)\, dz} \tag{38}$$

Where all right-hand side terms can be evaluated from simulation of aberration free probes; these are in words, the in-focus 2D perfect-probe OTF, multiplied by the ratio of the defocus-sum of the 3D simulated OTF, and divided by the defocus-sum of the square of this 3D OTF. Such simulations *must* have the same convergence angle and accelerating voltage as the experimental probe and have sufficient range in z either side of focus to give the numerical solution sufficient precision. It has been found that, in practice, if the defocus range of the simulated 3D-OTF matches that of the experimental data (which itself was required to have sufficient defocus range to encapsulate the full OTF of the probe) good agreement is achieved. It is also essential that the reciprocal-space range and sampling exactly math that of the experimental Fourier-transform stack. This calibration can be deduced from the experimental data using the techniques described in Section 3.1.5.

An example of such a 2D up-weighting modulation function is shown in, Figure 99 – left.

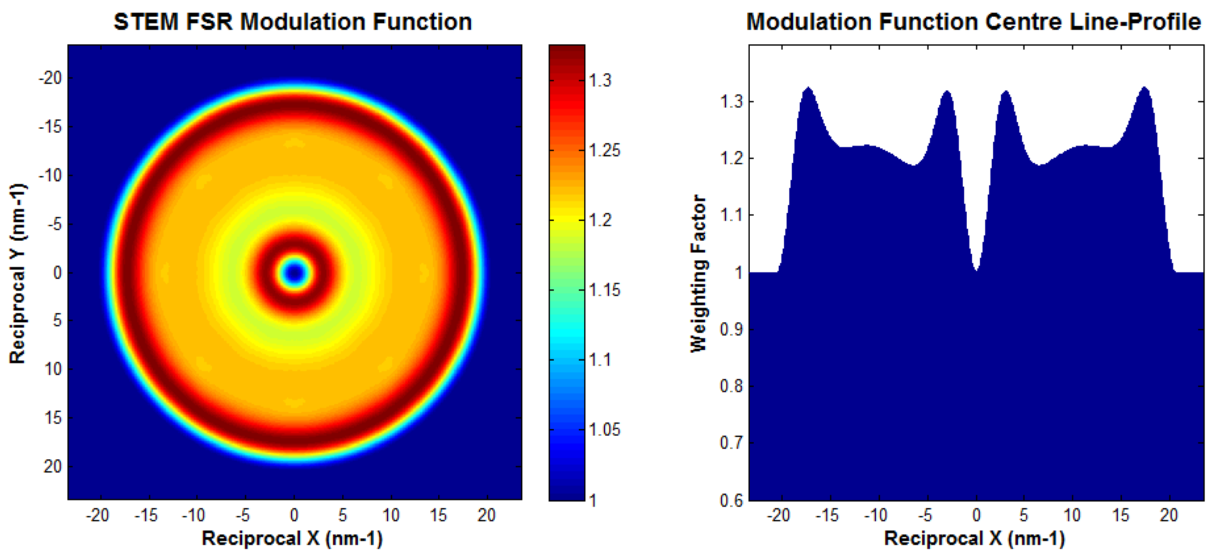


Figure 99. Example Fourier-space modulation function for the image reconstruction of the data shown in Figure 78. Note that both at the point of the D.C. spot, and beyond the Nyquist limit of the simulated OTF, the weighting factor is unity and applies no change to the FT.

To better visualise the form of this example modulation function, Figure 99 (right) shows a line-profile taken through the reciprocal-Y = 0 mid-line. The 2D modulation function contains several noteworthy features which will be discussed briefly.

Firstly, the most striking feature is the circular outer bounds of the function at ≈19 nm^{-1}. This represents the ultimate resolution capability determined from the probe simulations and corresponds to the 2α / λ limit [7], [43]. Next, at no point does the profile fall below unity; including, importantly, both the region outside the 2α / λ limit, and the zero (D.C. offset) spot at the centre. This has two important effects, the strength of pure noise spatial-frequencies is not diminished by this modulation function – the image is not smoothed; also the not affected – preserving quantitative image intensity.

It is reasonable then that with these boundary conditions, of the modulation function being equal to unit at its centre and also at its circular bounds, that a generally toroidal shape is exhibited. However, the up-weighting is greatest for spatial-frequencies that just neighbour these bounds. This can be understood by considering the 3D form of the OTF; at these points the depth resolution is poorest [24], [43] resulting in a more defocus-elongated intensity fitting (Figure 81) and hence extraction filter (Figure 98) for these frequencies. From the discussion presented earlier, the more spread with respect to defocus the 3D filter the greater the up-weighting must be to compensate.

Importantly it should be noted that this modulation function must be recalculated each time a different voltage, convergence angle or image pixel-size (sampling or magnification) is used. Fortunately this is computationally easy taking only the same amount of time as the probe simulations from Section 2.1.

5.6. Example STEM Focal-Series Restorations

To demonstrate the general applicability of the STEM FSR method a variety of test data sets were analysed; including hexagonal symmetry crystallography (Figure 100), differing column compositions (Figure 101), and a twinned free-standing nanoparticle (Figure 102).

Some further examples are also shown later in 5.7, including a single-crystal slab (Figure 111) and a free-standing crystal sample over vacuum (Figure 113).

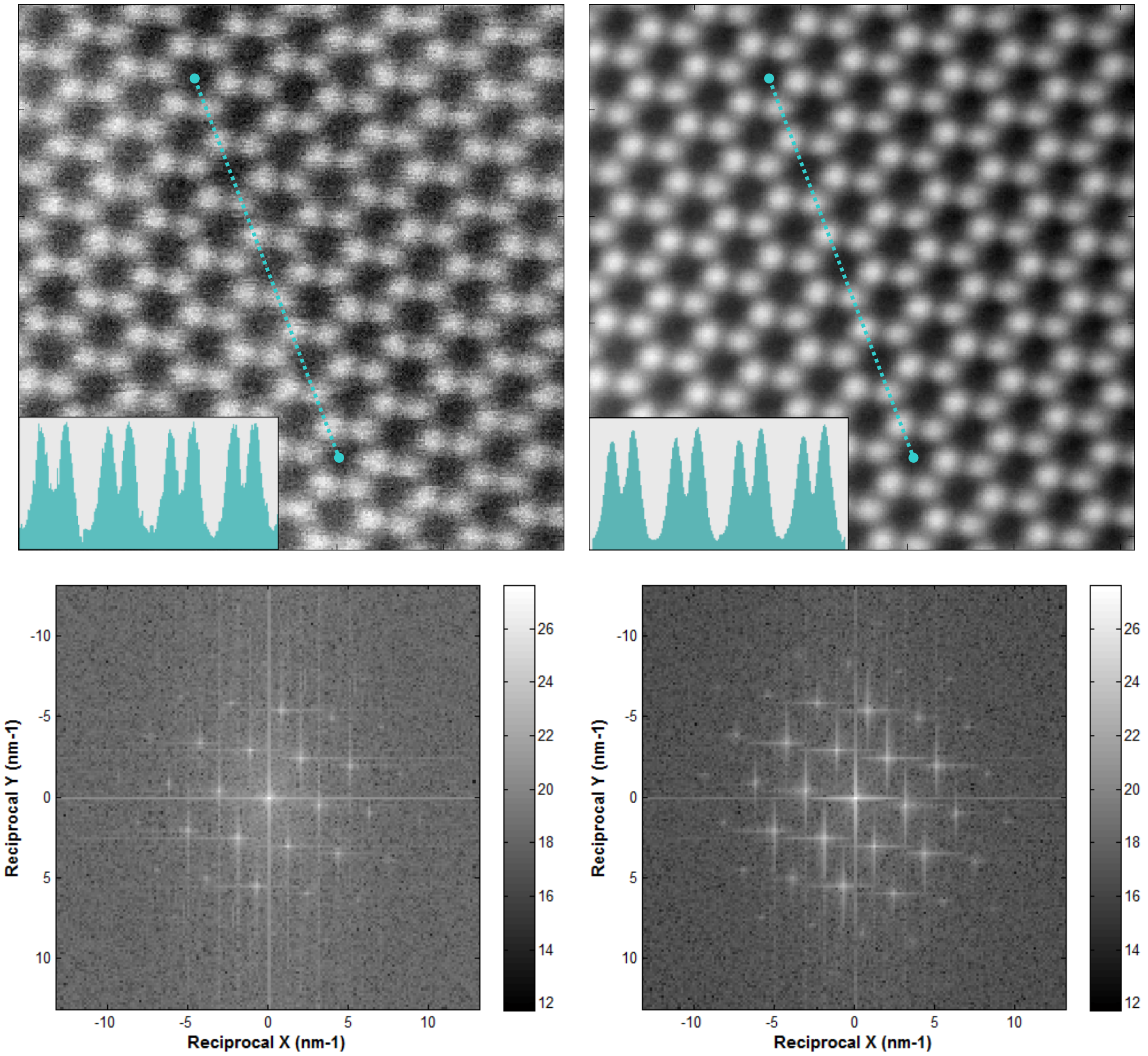


Figure 100. Enlargements from example focal-series data of MoS_2. Top-left, the operator focussed image and top-right, the reconstruction (matched greyscales). Insets show line-profiles as indicated by the dashed lines. The polarity of the hexagonal rings is more clearly visible in the reconstruction indicating an odd number of layers were present in this sample Below, enlargements of the corresponding Fourier transforms (matched logarithmic greyscales). FT-spot streaking is the result of the finite image size. In this example SNR was improved by 2.2x and resolution improved 6.3%.

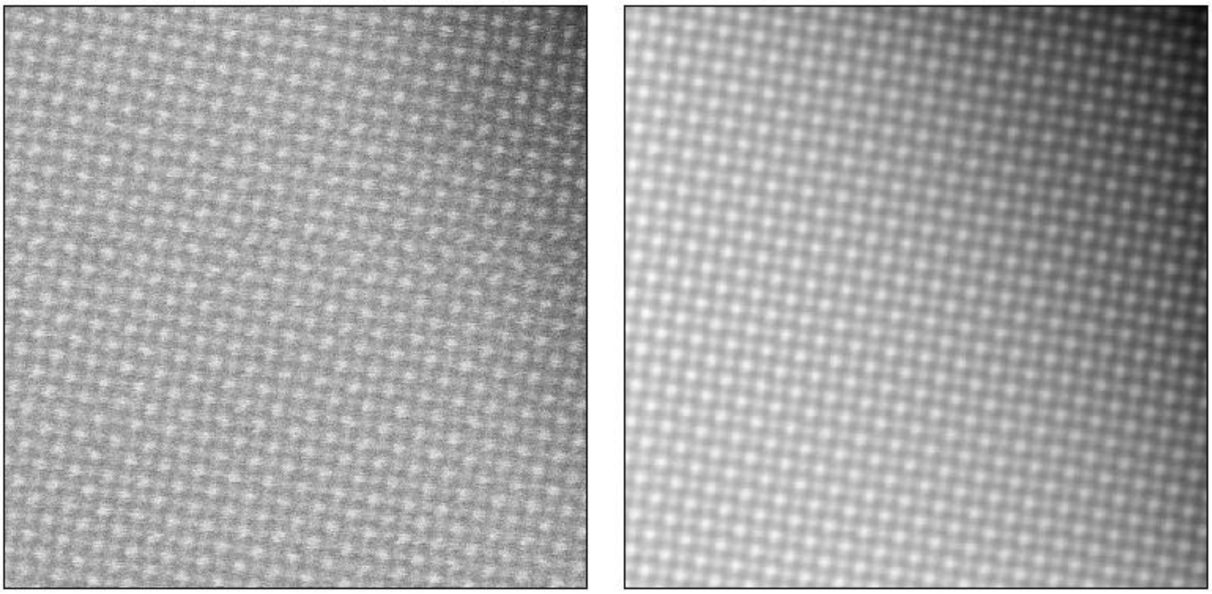


Figure 101. Example focal-series data of $SrTiO_3$; left, the operator focussed image and right, the reconstructed image. In this example SNR was improved by 3.6x and resolution improved 9.5%.

In Figure 101 both the differing column intensities of the Sr and TiO columns are preserved along with the overall thickness gradient increasing from top-right to bottom-left.

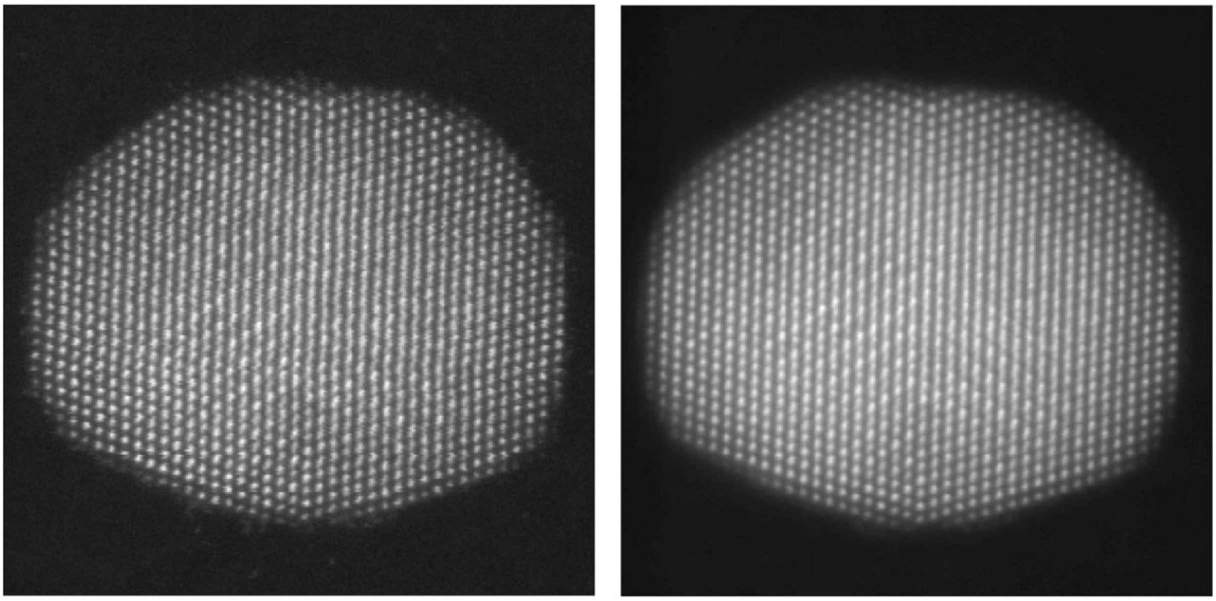


Figure 102. Example focal-series data from a Pt-Co nanoparticle; left, the operator focussed image and right, the reconstructed image. In this example SNR was improved by 43% while resolution was unchanged.

In Figure 102, the twinned nature of the nanoparticle sample yielded two superimposed spot-patters in the Fourier transform. However, only one set of spots were used for the aberration diagnosis and image restoration. This limitation only stems from the current software design and is hoped to be overcome in the future.

5.7. Quantitative Evaluation of the STEM FSR Routine

In addition to the examples shown in the previous section the data presented in this section incorporates the collective library of 42 usable data-sets acquired during the duration of this project. The majority of these were MgO smoke-cubes as used in the sample thickness experiments discussed in Section 4.2 (Figure 84). Whilst it is not possible to show here the restorations from this whole library, select example are included with enlargements where necessary to highlight certain behaviours.

In the comparisons of raw and restored data, the 'before' image in each case has not necessarily been taken as the operator focussed image but as the image-frame determined by the analysis to be closest to optimum focus (although these were often coincident with one another). This choice of comparator image from the raw focal-series gives the comparison the fairest footing possible and is reasonable given the 1 nm step-size of manual focus and the subjective nature of manually focussing a live image.

The first metric to investigate is whether the technique is genuinely compensating for aberrations. This can be examined qualitatively by inspecting the 'roundness' of atomic features in the examples presented earlier but it can also be evaluated more quantitatively by comparing the resolution and SNR changes associated with the image reconstruction as a function of the severity of the remnant aberrations present. If the degradation in the raw-data is indeed caused by these remnant aberrations and if the algorithm is correctly compensating for these, then the images should exhibit measurable improvements when restored, with the improvement being greater the larger the remnant aberration.

Most data sets appear dominated by first-order two-fold astigmatism or occasionally spherical aberration. This is reasonable as $C_{1,2}$ is known to drift and need frequent small-adjustment, and the collective effects of round aberrations can often appear as a small remnant $C_{3,0}$. As a result, the restoration performance data presented here will be expressed relative to the modulus of these two aberrations (azimuthal direction of $C_{1,2}$, and sign of the $C_{3,0}$ is not relevant to the restoration software's performance). Experimentally of course the illumination frequently contains small amounts of both aberrations so when collapsing the results to each some scatter is to be expected.

5.7.1. Image Resolution Improvement

Resolution was determined as described in Section 3.1.3 with at least 50 atomic columns inspected and averaged for each reading both before and after image reconstruction.

In general the microscopes used possessed a manual fine-focus step of around 0.5nm; observationally it was also found that, when fine-tuning astigmatism mid-session, skilled operators were able to achieve remnant astigmatisms generally less than 2 nm. As a result more than half of the data sets recorded fall within this regime, Figure 103.

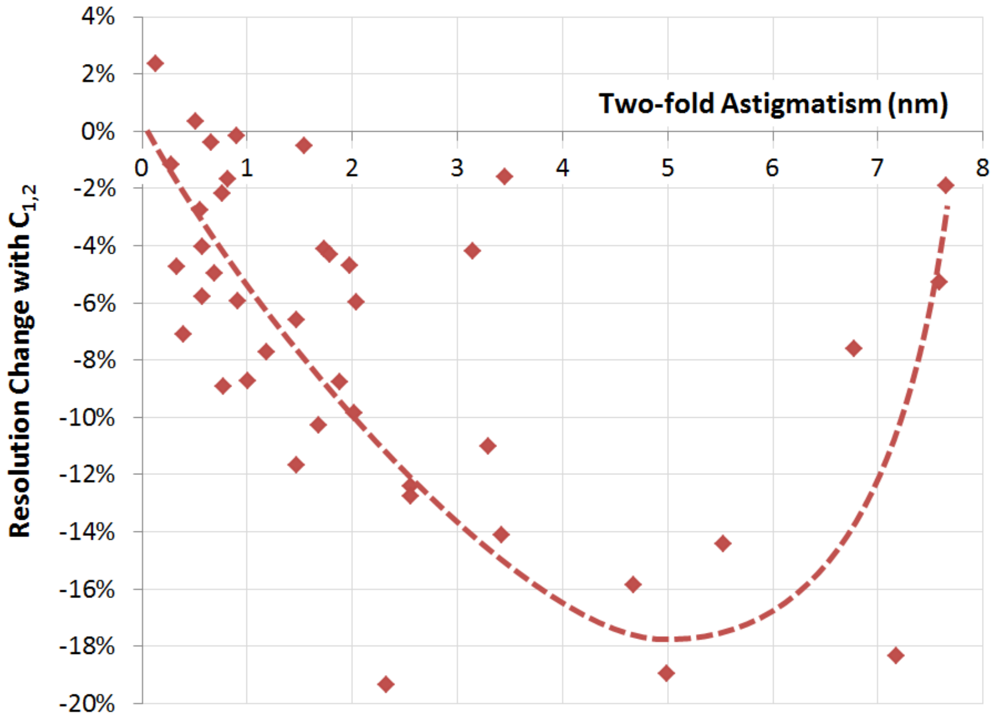


Figure 103. Resolution improvement following FSR method for 42 data-sets with increasing degrees of indicated two-fold astigmatism (lower is better). An approximately linear correspondence was found for astigmatisms of up to 3 nm. No data-sets however showed an improvement greater than 13%.

In general a near-linear improvement in resolution (reduction in feature size) was observed for remnant two-fold astigmatisms of up to 5 nm (dashed-line, Figure 103). For increasingly large astigmatisms, less improvement was achievable upon reconstruction and this was almost null by the point of 8 nm (corresponding to the deliberately stigmated data shown in Figure 89).

The same data presented dispersed with respect to remnant spherical aberration is shown in Figure 104. Here, again more resolution improvement was achieved for the data sets afflicted with larger remnant spherical-aberration suggesting that the image reconstruction had genuinely restored previously-lost image information. This trend does not continue endlessly and none of the data-sets restored exhibited resolution improvements of greater than 20%

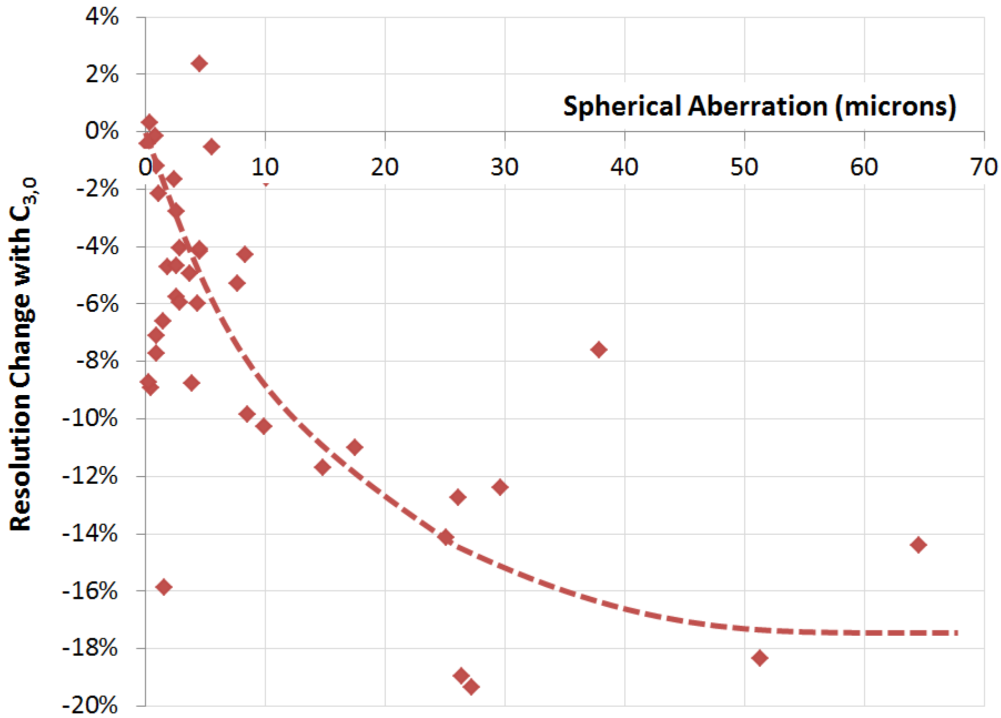


Figure 104. Resolution improvement following FSR method for increasing degrees of indicated spherical aberration. An approximately linear trend is seen for spherical aberrations up to ≈25 µm, beyond this additional improvement diminishes.

From this data it would be recommended that the operator should ensure that the astigmatism is tuned to be better than ±5 nm and the spherical aberration better than ±50 µm.

5.7.2. Image Signal-Noise Ratio (SNR) & Comparison with Image Averaging

It is well known that averaging the image data from multiple views of the same sample leads to a reduction in the appearance of random noise. The counter point to this is, as for STEM FSR, this requires more total acquisition time and electron dose on the sample. In this section the performance of the newly developed FSR routine will be evaluated quantitatively and also compared with the special case of image-averaging. Figure 113 shows the SNR improvement realised upon restoration from the whole collection of recorded focal-series.

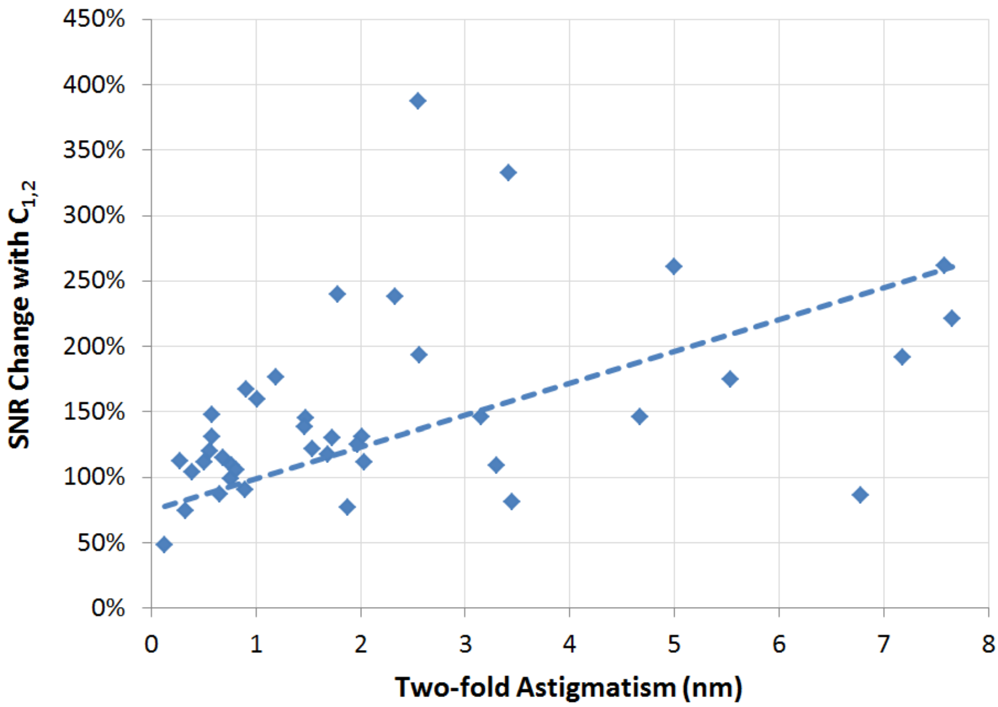


Figure 105. SNR increase realised after compensation of increasing severities of remnant two-fold astigmatism. Results indicate an improvement in SNR in all cases between ≈50% (1.5x boost) and ≈300% (4x).

A general trend is observed in Figure 105 (and is shown with a dashed line only) indicating that a greater improvement in SNR performance was observed for data sets with larger remnant two-fold astigmatisms. The 'y-intercept' of this figure is not zero as was the case in the previous section. This is because the defocus-width of the extraction function does impart a degree of image averaging for the case that remnant aberrations are near zero. From averaging of say three frames we would expect a √3 enhancement of SNR that is 1.73x or +73% approximately the value indicated at the y-intercept of Figure 105.

It is important to stress; this does not mean that the restored results are better if astigmatism remains, rather that the performance *relative to the Gaussian focus image* is relatively larger. This suggests that the restoration algorithm is in fact-compensating for the aberrations present and not simply equivalent to image averaging. SNR improvements of greater than 300% (a factor of 4x) were very rare. This may be due to the fact that no data sets were sufficiently degraded in their raw state or because of the SNR measurement ceiling arising from the finite image sampling used (see Section 3.1.2).

The data were also inspected as a function of the remnant spherical aberration present, Figure 106. Again this was less than 20µm in the vast majority of cases as a result of the microscopes auto-tuning accuracy.

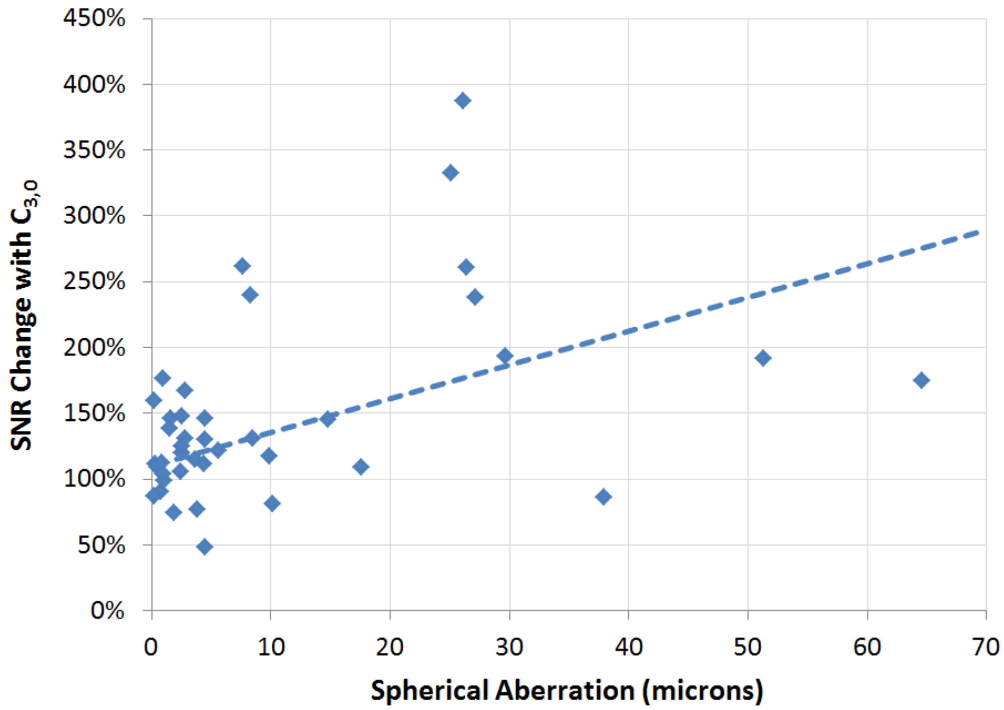


Figure 106. SNR improvement realised after compensation of increasing severities of spherical aberration

Here a weaker relationship was observed but in general an improvement in the SNR of 100 – 200 % was achieved.

One final benchmark for this technique is to compare it against image averaging. Where multiple image are recorded of the same subject, and contain only random noise (shot-noise) then this can be reduced by image averaging. To test the noise reduction performance of the FSR method we can compare the results against the simple averaging of multiple 'at-focus' frames.

To test this a 30-frame, 1 nm step focal series was recorded immediately followed by a set of 31 frames with no focus change, from an area of virgin material just to the side of the first acquisition. The frames were re-registered and averaged and can be compared qualitatively in the enlargements shown below.

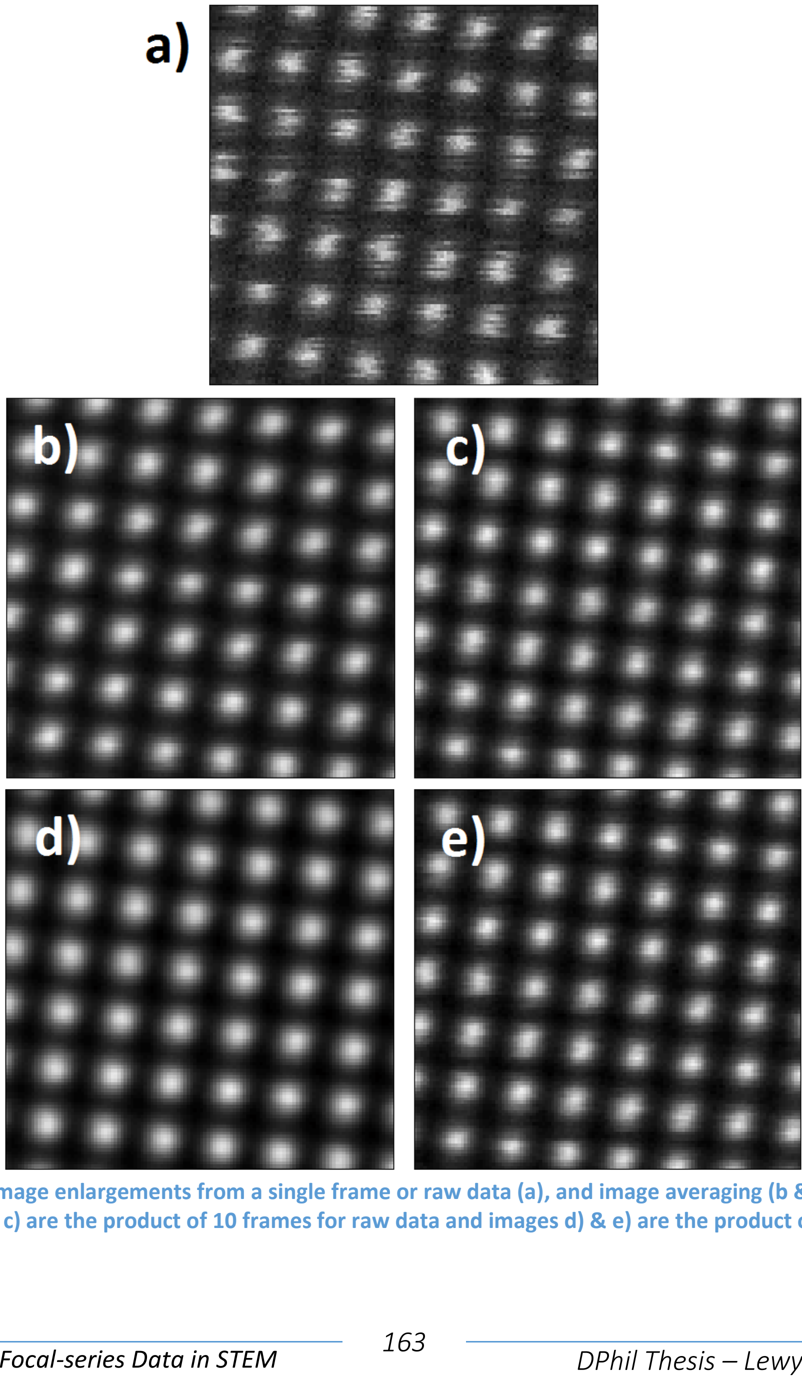


Figure 107. STEM Image enlargements from a single frame or raw data (a), and image averaging (b & d) and FSR (c & e). Images b) & c) are the product of 10 frames for raw data and images d) & e) are the product of 30 frames.

For the FSR comparison, the as-recorded data cube contained 30 frames. However we can generate smaller data-sets by truncating the original data in the defocus dimension. This data cropping was performed symmetrically such that the Gaussian focus frame remained as close as possible to the middle of the stack. Figure 107 shows (top) the first image from the series for comparison and below it, averaging (left) and FSR (right) results from 10 frames (middle row) and from 30 frames (bottom row).

Qualitatively, the images generated from 10 frames of raw data are relatively similar. Both show a significant reduction in scan-distortion and random noise. For the 30-frames results the simple averaging output certainly appears 'smoother' than the focal-series result, however, the features appear somewhat dilated (resolution degraded). To better evaluate the differences between the two methods a more quantitative comparison must be performed using the methods described earlier in Section 3.1.2.

For this comparison, 'averaged-images' were produced from a zero-focal-step series using frames 1-to-n, where n is the total summation width with respect to number of images; focal-series restorations were computed using defocus cropped data-volumes, retaining n frames of the original data (keeping Gaussian focus in the centre of the cropped volume). The results of the quantitative comparison are shown below in Figure 108.

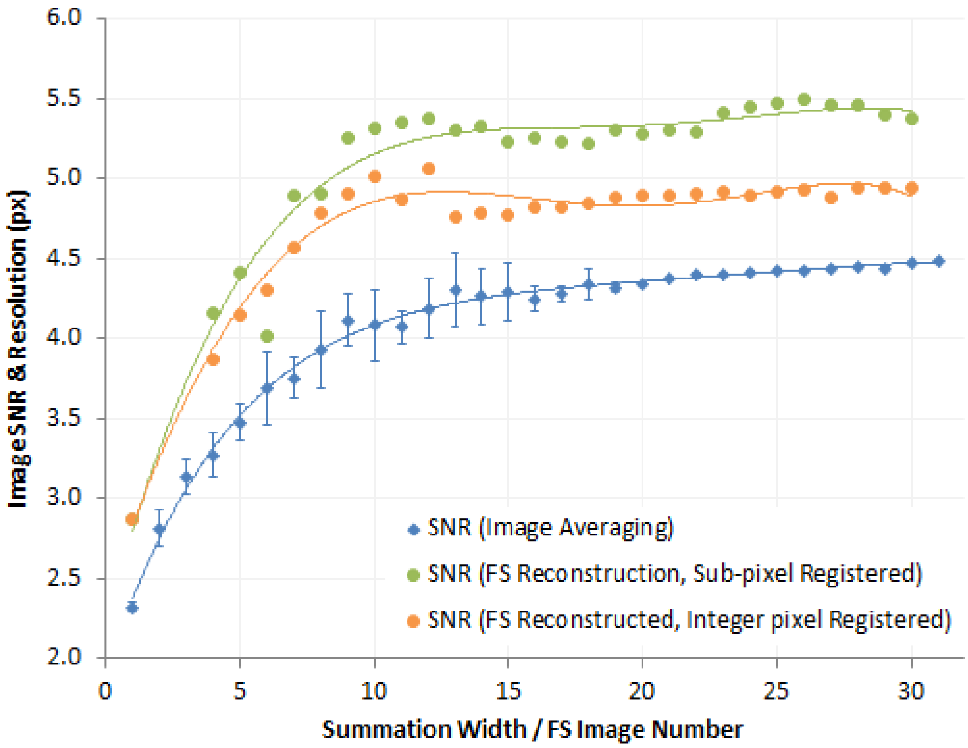


Figure 108.SNR measurements for increasing numbers of summed / restored image frames. The offset at the origin (1 frame) is the result of two different experimental series being used, one with focus change and one without.

Considering first the results of the image averaging, we see as expected an increase in the SNR with increasing number of averaging frames. It would be expected that for images containing only random shot-noise that this would increase with the square-root of the number of images. This behaviour is observed initially but does not continue indefinitely. This is thought to be due to the presence of scan-noise in the raw data, the smoothing effects of the image registration process and the difficulties in separating and measuring increasingly small amounts of noise remaining in the averaged images.

Next, the focal-series restoration results; the software was unable to achieve a OTF-lock for data sets containing fewer than 4 frames. This result is unsurprising and is a fundamental limitation of attempting to fit too many unknown parameters to insufficient raw data. For four or more input-frames an almost identical relative SNR increase (approximately 2x) was observed again plateauing at roughly the same point (10 frames). As the microscope had been exceedingly well tuned for this experiment no change in resolution was observed in either the image averaging or the focal-series reconstructed case.

While the image-averaging and FSR approaches seem to perform equally well in terms of SNR, it should be stated that simple image averaging offers no means to compensate for aberrations, and in the presence of any focal-drift during acquisition will ultimately lead to a loss of resolution.

5.7.3. SNR & Resolution Improvement versus Sample Thickness

Another equally important performance metric for the FSR method, is to determine over what range of sample thicknesses this new technique is applicable. From Section 4.2 we recall that the 2D object function does appear to remain valid for thicknesses greater than 200 nm, however, that is not to imply that the restoration performs *equally well* over the whole of this range. To investigate this the same library of focal-series was re-examined and the relative SNR and resolution performance plotted dispersed with respect to the estimated thicknesses of the samples, Figure 109.

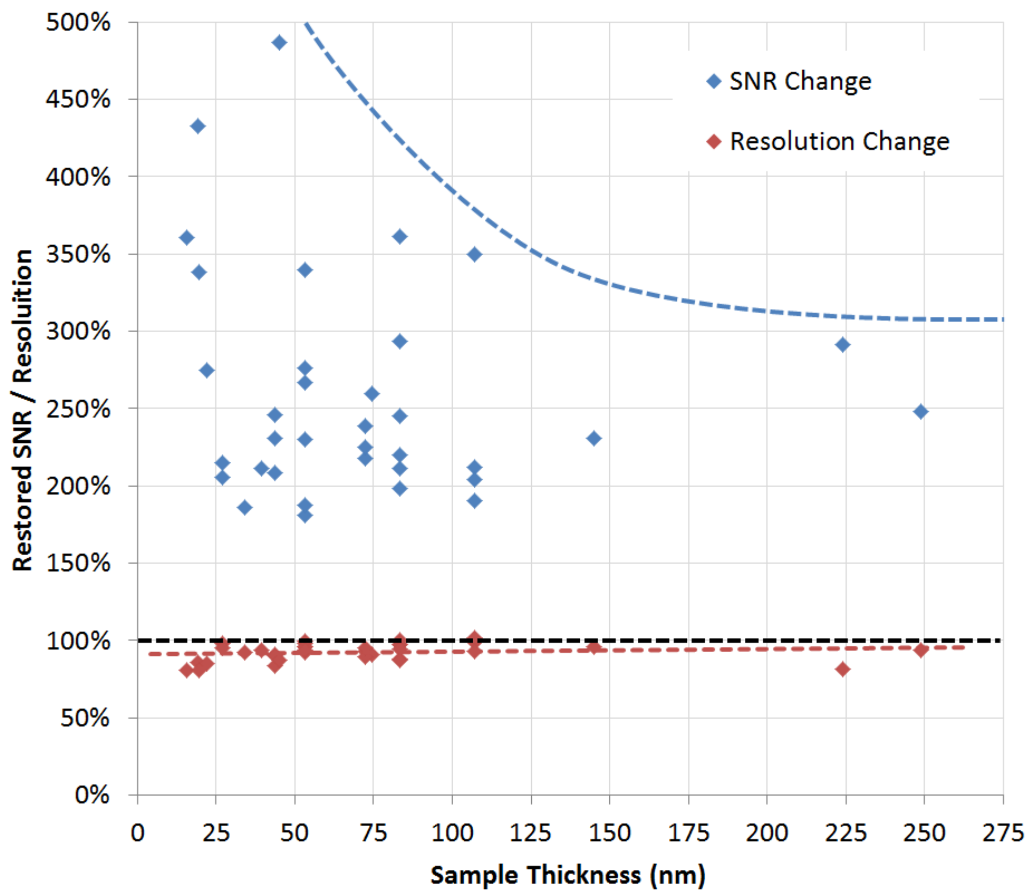


Figure 109. Relative SNR (larger is better) and resolution (smaller is better) after FSR for MgO samples of varying thicknesses. In both cases 100% indicates no change. Dashed blue line indicates the upper limit of SNR improvement that was observed suggesting the FSR method performs at its best for samples up to 150 nm in thickness. Red dashed line indicates the trend in resolution improvement – only a weak correlation was observed but generally indicates a preference for thinner samples also.

Figure 109 again shows that the FSR routine delivers SNR improvements of 150% or more (1.5x original), however the apparent upper limit to this improvement(blue dashed line) is reduced for larger and larger thicknesses. This is thought to be a result of the reduced signal-background ration (SBR) present in these thick specimens. Physically, this can be explained as a result of significant beam de-channelling resulting in an erosion of the strength of the spatially varying signal and an increase in the contribution to the structure-less background. We can consider then this figure of 125 nm as a rough guide to the upper limit of thicknesses suitable for this technique (perhaps less for heavier specimens). Fortunately this limit is relatively generous and many samples can be produced and examined in thicknesses less than this.

5.7.4. Verification of the Modulation Function and the Image Information Limit

In Section 5.5 the concept of the 'Modulation function' was introduced – the 2D up-weighting function that restores the relative weights of the moduli in the reconstructed Fourier transform. An example modulation function was shown in Figure 99 and discussed qualitatively but its importance in the image-reconstruction process merits some further evaluation of its effects.

The modulation function must meet two criteria to accurately weight the reconstructed Fourier transform; it must have the correct outer bounds (that is the reciprocal spacing corresponding to the aperture enforced information-limit beyond which it should take a value of unity), and within this bound it should weight the Fourier moduli appropriately to restore the 2D-OTF corresponding to a perfect illumination formed for the given voltage and aperture size applicable to the data-set.

One method by which the OTF responsible for an image can be inferred is to study the strength at which differing spatial-frequencies are transferred [45], [100]. Figure 110 (top) shows an example reconstructed FT, again from a [100] oriented MgO smoke cube, where for a perfect illumination system information would be transferred right up to the aperture (diffraction) limited information-limit (red circle annotation). Two 640 reflections are visible at the left side of the FT but their symmetric partners, that would be expected towards the top of the image, are lost to the scan-noise induced, vertical-streaks. The yellow arrow in Figure 110 indicates the position of a weak 800 reflection; to evaluate this more objectively a modulus line-profile has been extracted through the FT (Figure 110 - bottom). Their strength over and above the noise-floor is a measure of the 1D experimental OTF (blue dashed line).

From this figure two conclusion may be drawn; firstly, the modulation function yields an effective OTF that is smoothly varying with respect to spatial frequency (important for the avoidance of imaging artefacts in real-space), and following the reconstruction procedure, information is visible (just) almost all the way to the ceiling imposed by the information limit [32].

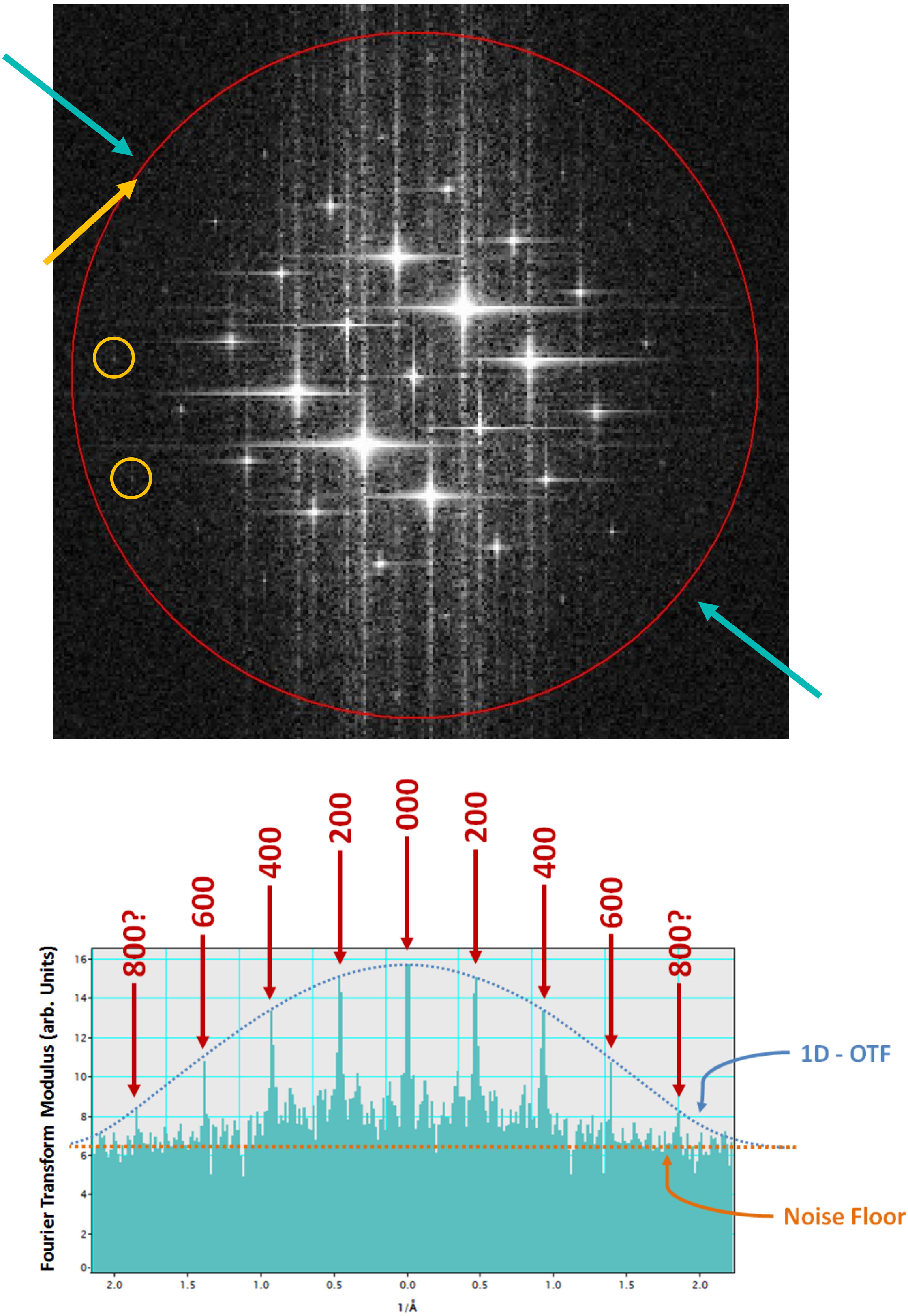


Figure 110. Top – enlargement of the centre of a reconstructed FT showing many unique FT-spots. Yellow circles indicate 640 reflections (≈58pm) and the yellow arrow indicates the location of a possible 800 reflection (≈53pm information transfer). Red circle shows the expected information limit at the voltage and aperture sizes used (20.5 nm^{-1} , ≈49 pm). Cyan arrows indicate the position of the extracted FT-modulus line-profile.

Bottom – FT modulus line-profile. Blue-dashed line shows the effective 1D-OTF deduced from the peak intensities of the FT-spots and the approximate noise floor.

The data presented in Figure 110 were recorded at 300kV and the form of the resulting deduced 1D-OTF is consistent with the work shown by Haider et al. (reference [32] – figure 8) again suggesting that the performance of the reconstructed data closely approaches the ultimate limit of the instruments capability.

5.7.5. Correction of Low-Frequency Image Distortions

It was discussed earlier that the serial-acquisition nature of the STEM leaves it vulnerable to high-frequency instabilities that can cause image distortions on the length scale of a few pixels (Section 3.2.1). However, much lower frequencies can also affect the images and cause far slower, long range image 'wobbles'. These are visible qualitatively in Figure 111.

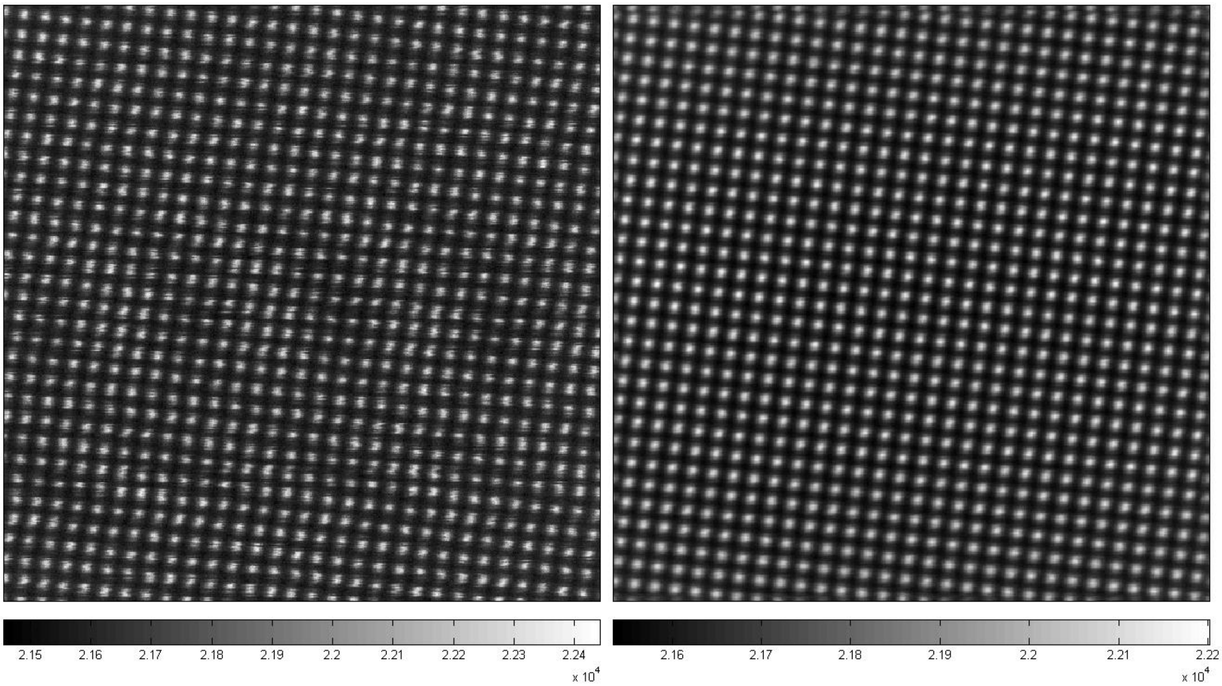


Figure 111. Gaussian focus image from a focal-series (left) and the resultant reconstruction (right). The effects of low frequency instabilities can be observed as bright and dark bands through the unprocessed image.

These small distortions in the real-space image equally manifest as small disruptions in the image Fourier transform (in both modulus and phase). As the focal-series restoration algorithm rebuilds the image in Fourier-space using information from the nearby defocus frames in *both* modulus and phase such outliers are suppressed returning an image free from such low-frequency distortions.

The low frequency distortion-induced image-wobbles can be observed more quantitatively if a geometric phase analysis (GPA) is performed on the recorded images [129]. In the case of an image of perfect crystal no strain is expected and the GPA results reveal the effect of low frequency instabilities. In the absence of these instabilities the strength of a given periodic component (modulus), and its position deviations (phase) would both be expected to be constant at zero.

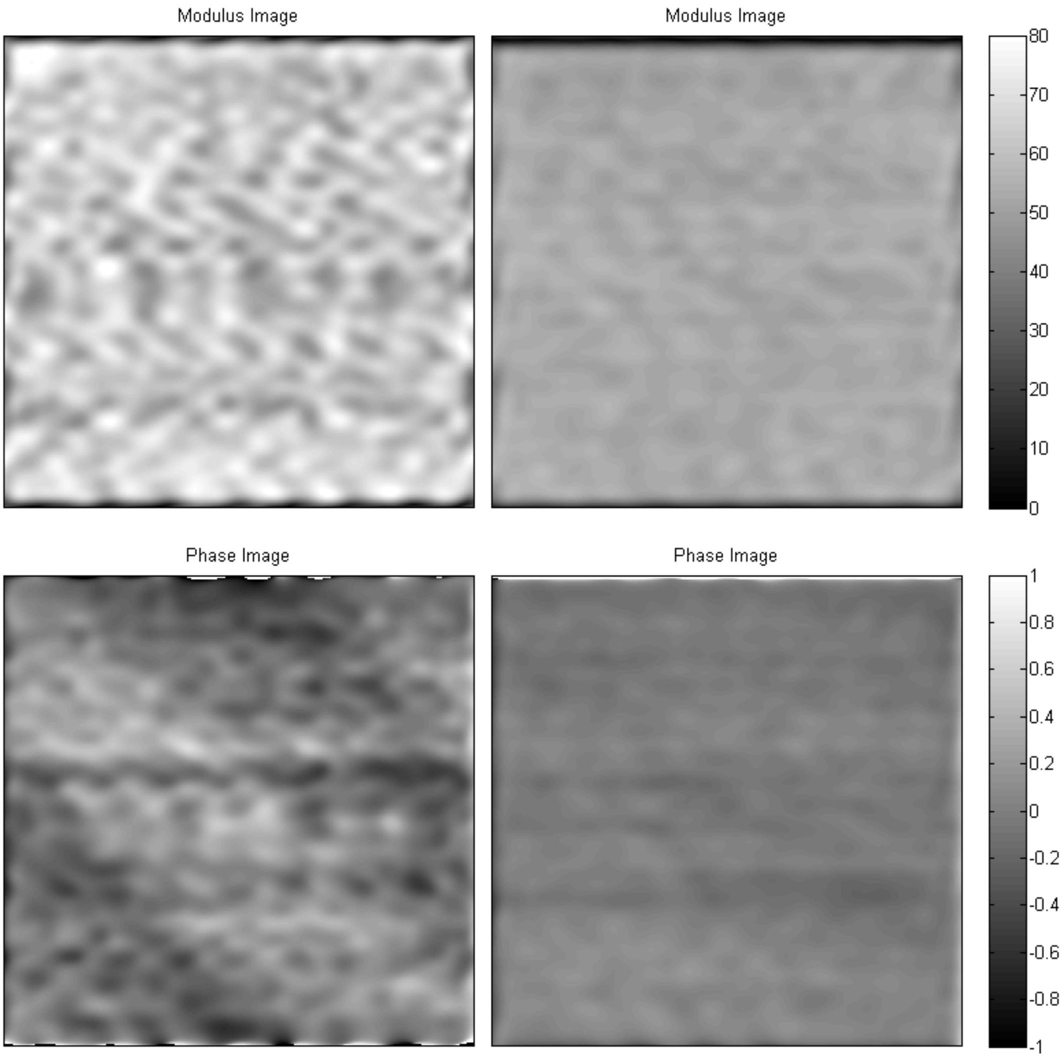


Figure 112. Geometric phase analyses for the two images presented in Figure 111 using the [2 0 0] Fourier component closest to the image 'North' (left: Gaussian focus raw image, and right: its restoration). Top row shows the modulus, lower row shows the phase deviation (greyscale shows ±1 radian deviation in both cases).

Inspecting again the data presented in Figure 111 we can see the perfect crystal in the Gaussian focus image is far from that. In fact at the time of the acquisition of this data some low-frequency instabilities were present. The modulus map of the raw frame displays a significant variation, with the range being anywhere from 30 – 80 arbitrary units; this is in contrast to the restored frame which has a far more consistent magnitude of ≈55 ±5. Examining the phase map, the instabilities are manifest as both broad bands (inclined at approximately 45° left of vertical), and horizontal striations which are more narrow but with no clearly apparent period. In the restored image the broad bands have been entirely eliminated and very little remains of the horizontal striations.

Importantly, this should be highlighted as a difference between the newly developed FSR method and image averaging. For the image-averaging case these low-frequency distortions prevent (locally) the lattice from being perfectly registered, even if the image as a whole is ideally registered. This can also be considered in the Fourier transform representation, where here the uncertainties in column positions lead to errors in the phase of the Fourier-spots (becoming increasingly problematic for higher order spots). Hence, if images are then summed (wholly equivalent to FT-summation) then this error in FT-spot phase will lead to weaker total intensity in the high-order spots which manifests as a loss of resolution in real-space; this is precisely what is observed when such low-frequency distorted images are summed.

5.7.6. Fidelity of Sample/Vacuum Interfaces

Any manipulation of imaging data in Fourier space risks the introduction of real-space artefacts and must be performed with great care. This effect is well known in the HRTEM exit-wave restoration community where errors can lead to ‘ripples’ emanating from regions of sample off into vacuum. This can make determination of edge structures very challenging or unreliable. To investigate the sample-vacuum-edge performance of the new STEM FSR methodology the sharp edges of an MgO smoke cube were investigated, Figure 113.

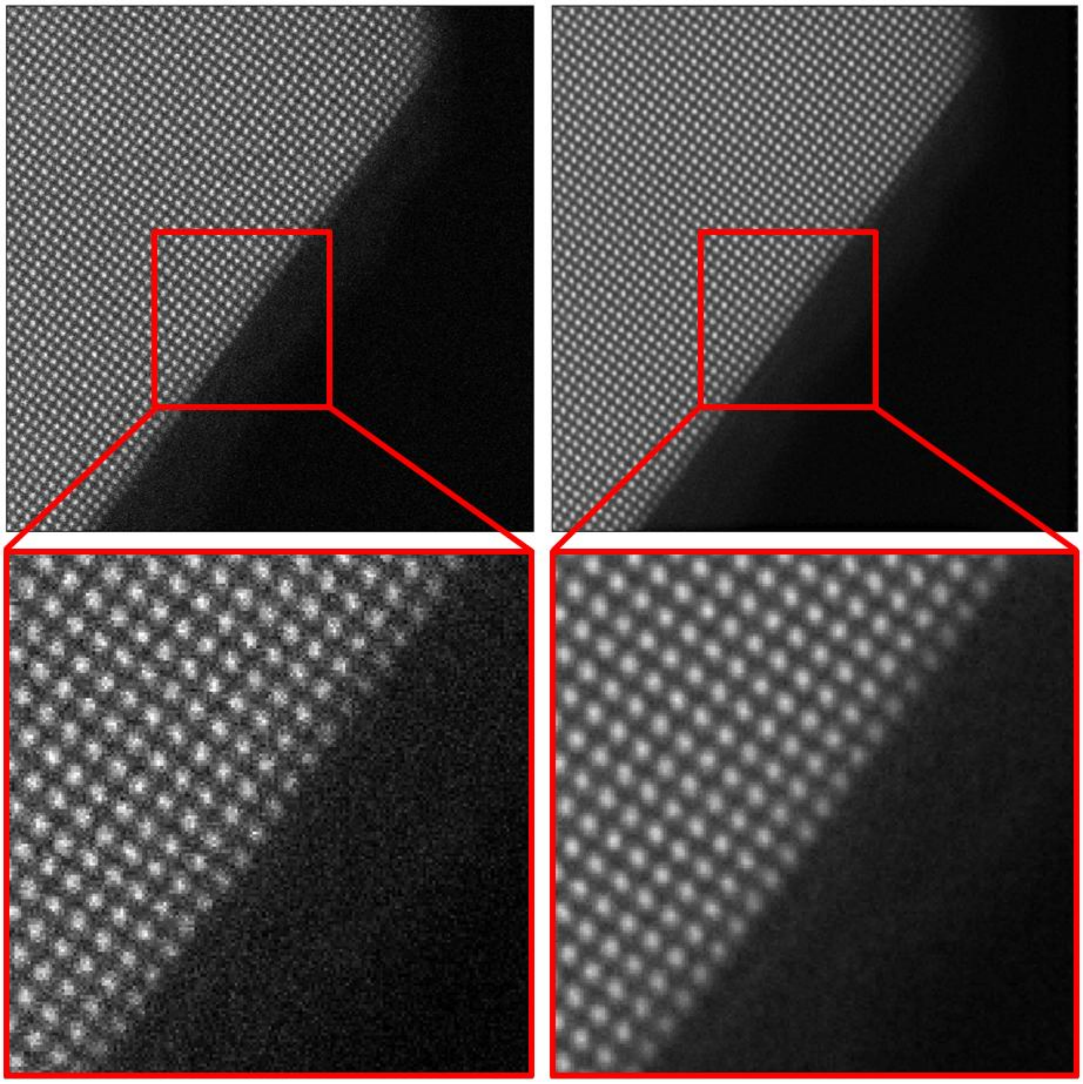


Figure 113. Raw best-focus image (top-left) and its restoration (top-right) from the sharp edge of an MgO smoke cube. Lower pair of images shows matched enlargements from the indicated areas. In both images the atomic columns are elongated in a direction just left of vertical; this is present both before and after processing and is most likely caused by a small specimen mis-tilt or residual coma. Grey-scale in both data are locked to be equal.

Figure 113 shows an appreciable improvement in the image SNR, but more importantly no image artefacts occur at the sample-vacuum interface. This behaviour verifies the utility of this method for a-periodic features within image, and whilst regular crystal lattice is required for the restoration algorithm, the method is robust to the restoration of edges, twin-boundaries, dislocations and other a-periodic features.

5.7.7. Preservation of Quantitative Image Intensity

In the literature that has recently been an increasing interest in fully *quantitative* STEM [127], [130]. In this technique careful normalisation of the data allows the results to be compared with simulation on an absolute scale. In this case then it is essential that the absolute magnitudes of the image intensities are preserved throughout the focal-series restoration procedure. To evaluate this, the two images shown in Figure 113 (top) were investigated. Automatic peak-finding was performed to identify the locations of the atomic-columns followed by integration within circular regions about each column. The results are shown in Figure 114.

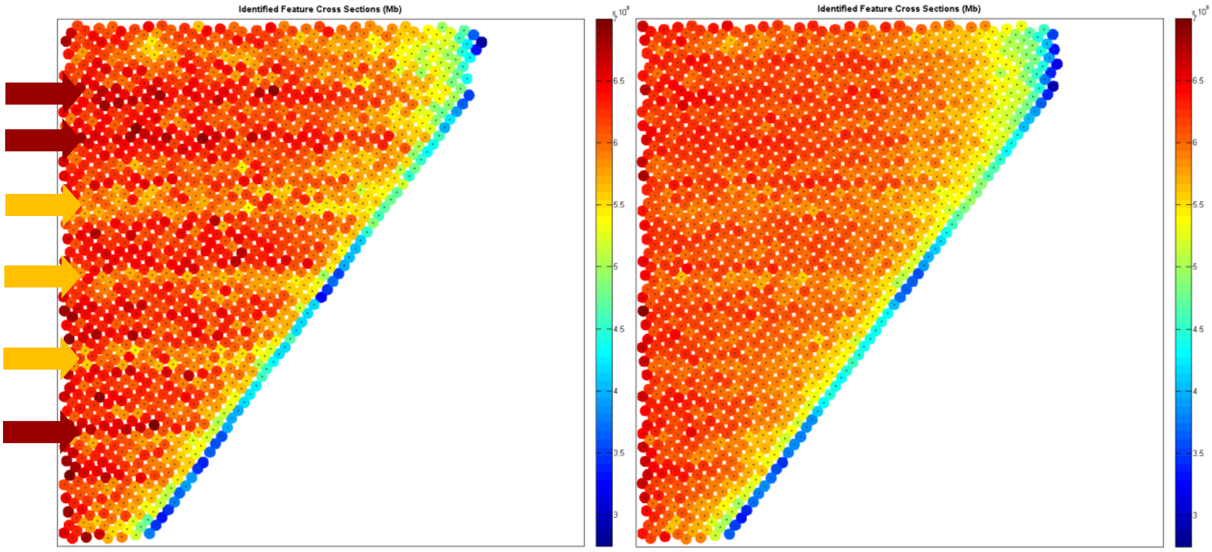


Figure 114. Column-wise integrated intensities of the images presented in Figure 113; left, the raw in-focus image and right, the restoration from the same data. Red arrow labels indicated scan-lines of above average image intensity, yellow arrows lines below average. Image colour-scales show identical ranges from 2.75×10^8 to 7.0×10^8 arbitrary units.

Qualitatively, inspecting the two analyses confirms that the quantitative intensity of the restored image closely matches the raw data. Surprisingly it was observed that certain scan-lines within the image were fractionally above or below the mean intensity; even though the sample is known to be of constant thickness. This behaviour also persists with scan-direction rotation so is not sample-related. This intensity fluctuation is thought to arise from a small fluctuation of the imaging beam-current and is itself partially restored in the image reconstruction. This can be explained within the context of the restoration algorithm as, so long as the emission current fluctuations are not synchronised in any way to the imaging then these fluctuations will have a random phase through the defocus stack and will not be incorporated favourably in the image reconstruction. In this way the effect of tip emission current fluctuations is partially mitigated in the same way that it would be by image averaging (hence why some residual effect can still be observed even in the restored image).

To evaluate the restoration methods preservation of intensity more quantitatively we can examine the histograms of the column intensities shown in Figure 114. The histograms for the raw data and the restored data are overlaid in Figure 115.

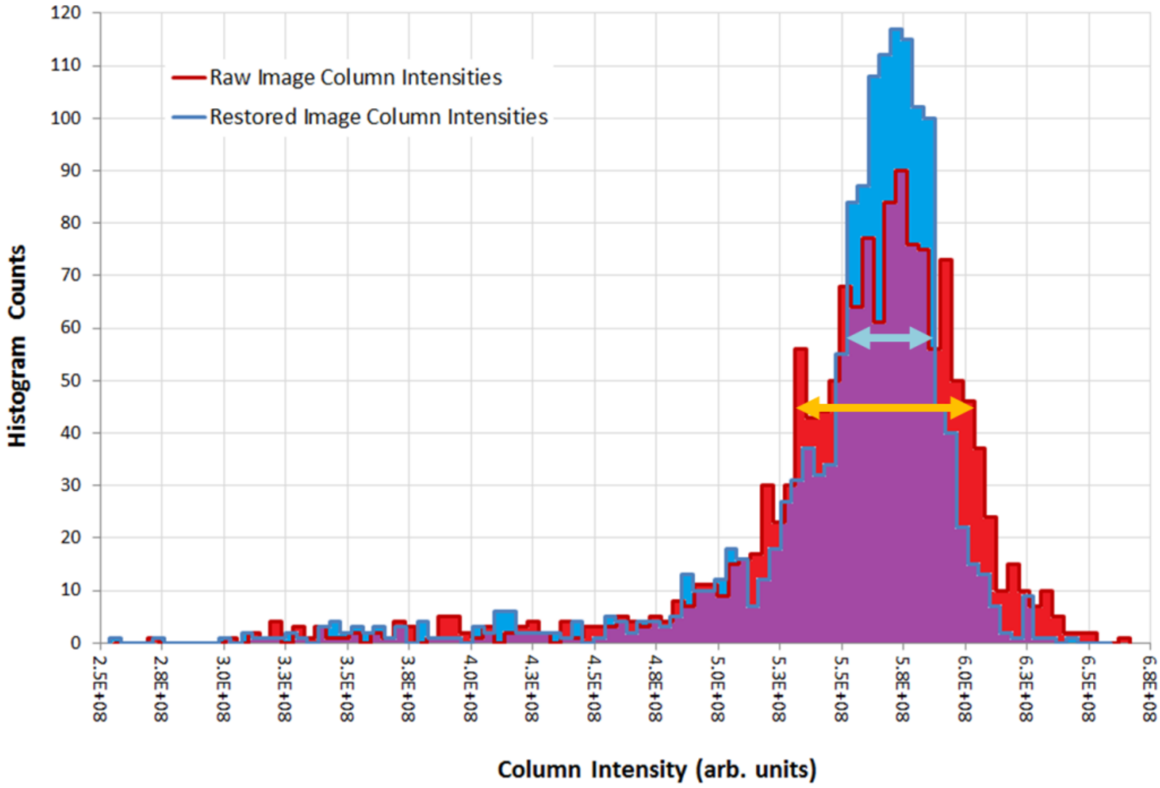


Figure 115. Overlaid histograms of the integrated column intensities shown in Figure 114. FWHM values have been labelled for each profile.

From Figure 115 two important properties can be extracted, the mean value and the FWHM (a measure of the spread in the data). The mean from the raw and restored data sets were 5.53×10^8 and 5.51×10^8 respectively; and are identical to within the precision that can be inferred. This validates that the STEM image intensity for quantitative analyses has been preserved correctly.

The FWHM of the histograms give an indication of the combined noise/instability performance and emission current fluctuation behaviour. The sample is known to be a perfect cube so all integrated intensities are expected to be identical (excluding the last atomic plane of the cube which may have endured some damage induced material loss[131]). The FWHM of the raw and restored distributions were 7.23×10^7 and 3.56×10^7 respectively and represents a reduction by ≈51%. This is a significant improvement upon the raw value and demonstrated the utility of this new FSR technique in quantitative STEM studies.

5.8. Experimental Subtleties of the Focal-Series Acquisition

Fundamentally it is impossible to separate the focal-series aspect of the imaging, from the time-series nature of serial acquisition. As an example, a typical data acquisition will consist of 30 frames, each at 512x512 pixels. For a 50 Hz line-synched acquisition this yields a frame time of approximately ten seconds and (with a small settling time after each focal change) a total acquisition time of just over five minutes. It is a result of this prolonged total time that lateral stage / sample drift can become apparent and image re-registration is often required; additionally though vertical sample drift (or defocus drift) and sample damage must be considered.

Vertical sample drift and defocus drift are observed identically and cannot be separated. As such they will be together referred to as defocus drift. The effects of this drift can be minimised by ensuring two experimental precautions are observed:

- Whenever possible focal-series work should be performed using a piezo-electric driven sample holder (often called fine or super-fine stage drive). Use of the mechanical stage drives should be avoided for some minutes before the start of the acquisition; the beam may be blanked during this time to avoid unnecessary sample contamination or damage.
- A small amount of time should be allowed after focus changes for lens stabilisation. Generally one half to one second is sufficient for this and is easily incorporated in the automated acquisition script.

With drift minimised we must turn to sample damage. Samples prone to damage will begin to amorphise to some small extent with continued exposure to the electron beam. In this context of focal series this manifests as a loss of intensity in the higher spatial frequency components used in the Fourier-stack analysis. The highest spatial frequencies will fade first and most rapidly followed more slowly by lower spatial frequencies. As a result, the higher the spatial frequency, the 'earlier' it will appear to peak with respect to time but also hence to defocus. Experimentally this is observed as an apparent spherical aberration with higher frequencies sweeping towards the defocus corresponding to the start of the acquisition.

To evaluate the severity of this effect the MgO smoke cubes were again used. These are known to damage under electron illumination [131], and at the 80kV used for this test were found to be especially vulnerable near corners or edges. To evaluate independently the effects of time-dependent damage and defocus drift, two matched data sets were acquired one immediately after the other, one in the normal fashion from under-focus to over-focus and the other from Gaussian focus outward (alternating between under and over focus with increasing magnitudes). In the former case, time (and hence damage) and focus are inseparable; however in the latter, temporal effects are equally symmetric with respect to defocus and as such any discrepancy between these two data sets can be attributed to damage related effects. Importantly, absolutely no retuning was performed between the acquisitions and only a piezo-electric drive was used to present a virgin area of crystal to the beam.

Figure 116 shows side-on views of the Fourier-stack for each of these two data-sets. In this case the focal step used was 1 nm and the defocus shown in nanometres also corresponds to the number of the image-frame in the acquisition series.

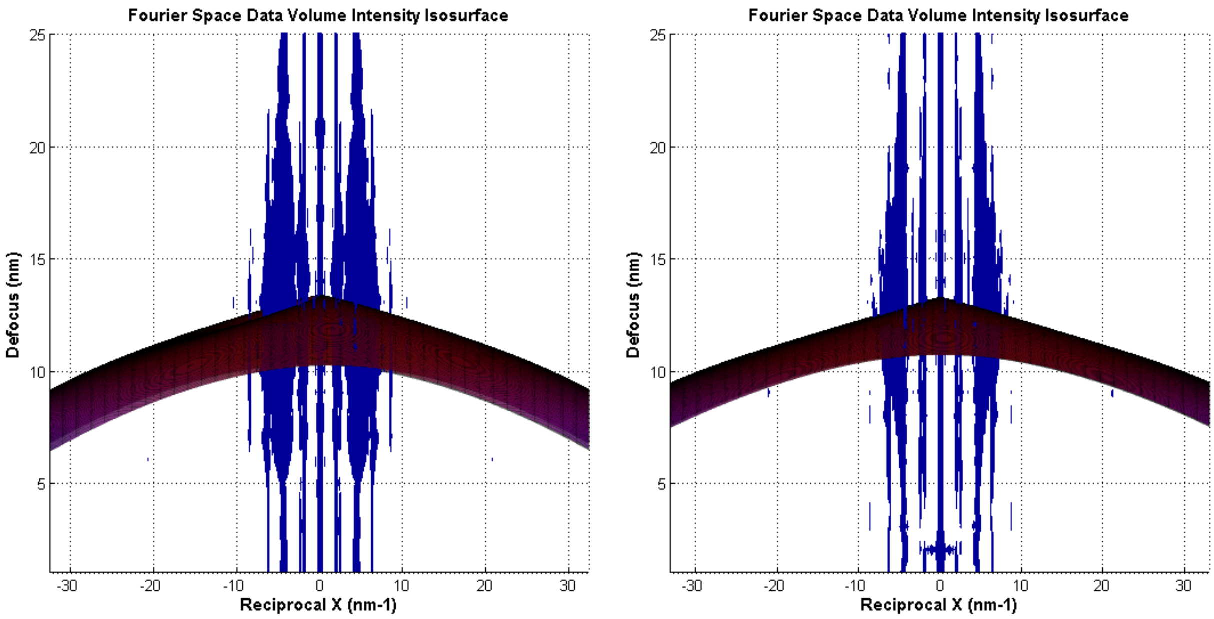


Figure 116. X'-Z views of Fourier-stack intensity iso-surfaces (blue rods) with fitted contrast-transfer surfaces (red downward curving band) for left, a standard 'bottom-up' acquisition, and right, for a 'centre-out' acquisition. Defocus scales show nano-metres relative defocus to an arbitrary zero; Gaussian focus in these terms was found to be 12.6 nm and 12.7 nm respectively.

Comparing the Gaussian defocii of the two data sets we see that a negligible amount of focus drift occurred in the several minutes between the two acquisitions, however, the increased curvature of the contrast surface towards the start of the 'bottom-up' test-case confirms that indeed some small amount of damage had occurred. Interestingly, even though this damage causes the contrast transfer surface to be misdiagnosed the reconstruction will remain centred about the surface of maximum image information – even though this may no longer faithfully represent the true state of the probe optics.

5.9. Practical Guidance for the Recording of High Quality Focal-Series Data

In the development of this new FSR technique several practical aspects have been evaluated including; the desirable pixel-sampling or magnification, scan-rotation and even upper limits for tolerable remnant aberrations. Should the reader wish to perform data-acquisition suitable for focal-series reconstruction the important parameters needed to get the most from the restoration method are summarised on a one-page pull-out sheet at the back of this thesis (Appendix 8.6).

6. Conclusions & Future Work

6.1. Conclusions

In the introduction to this thesis four main goals were outlined. It seems reasonable then to evaluate the outcomes in the same framework.

What are the information-transfer capabilities of aberration corrected STEM probes?

The simulations presented in section 2.1 revealed many of the non-round optical-transfer behaviours that can exist with the use of multi-pole hardware aberration correctors. That is not to say that aberration correction should not be used, rather that after the previously limiting spherical aberration has been corrected we must instead concern ourselves with ever-smaller (and in this case) non-round aberrations.

The hybrid reciprocal-lateral / real-defocus coordinate space that was introduced here does offer some advantages, making the non-round 3D-OTF behaviour immediately obvious. Additionally for the odd-order aberrations that cause the peak-strength of optical-transfer to be shifted or curved with respect to defocus, this representation makes this immediately visible. Unfortunately, such three-dimensional graphics are difficult to communicate on paper or by computer monitor meaning that either static-projections, x-y and x-z sections, or some form of 3D viewing method (such as the anaglyphs shown in appendix 8.4) must be used instead. This is likely to limit the adoption of this representation but this is fundamentally three-dimensional data and in the author's opinion the benefits outweigh the difficulties.

At the end of the chapter in section 2.2 it was demonstrated that in the presence of real experimental aberrations this representation could be used to easily plan focal-series experiments. Unfortunately at the moment this process requires the user to manually transcribe the values from the hardware corrector readout (which often varies by microscope model) to the probe simulation code in MatLab. This is somewhat tedious and ideally would be automated in the future.

What effects do instrumental instabilities have on recorded STEM data, and can they be counteracted by post-processing?

It became clear relatively early in the research that instrumental stability could not be ignored and was in fact often equally as performance limiting as poor aberration tuning. Furthermore, while it is often said that the STEM image is a position-series of serial acquired data, this is only true in the perfect-instrument case. In practice the only statement that can be made with absolute certainty is that the pixelated data represents a time-series, and it is small deviations between this time-series and the intended position-series that lead to the horizontal and vertical signs of scan-noise. Whilst this concept was then pursued as the basis for the restoration procedure, there remains one further discrepancy to be resolved. The restoration methodology developed involved projecting the true experimental probe-offsets in to horizontal and vertical components, whereas in fact experimentally these offset vectors can be in any direction. Never the less, the restoration method developed was able to significantly improve the appearance of the recorded STEM data both visually and quantitatively, with image resolution and SNR being improved by up to 16% and 30% respectively. The performance of the scan-noise correction algorithm was itself found to be ultimately limited by the sampling of the raw data and the presence of random noise.

It is unclear whether a full two-dimensional consideration would yield differing or better results. This would necessarily involve significantly more computation and perhaps need to draw on the maximum-entropy image restoration literature. In either case it seems likely that such a 2D approach would again ultimately yield an image which can only be considered as 'more close' to the true structure.

For the determination of peak-positions and the measurement of resolution and SNR it was found that finer pixel-sampling was preferable. However, in practice this requires longer acquisitions times that may reveal stage/sample drift more strongly.

Ideally any environmental instabilities should be solved 'at source' and the *cause* of the instability should be pursued where possible. Unfortunately this is not always practical, but where modifications to the microscope suite are possible the distortion's frequency spectrum output can be of great use.

What is the information content of STEM focal-series, to what extent do they agree with the two-dimensional object function model, and can they be used for aberration diagnosis?

It was found that experimental HAADF STEM focal-series data exhibit a focus-dependent information content that can be accurately described by the convolution of a three-dimensional illumination OTF with the object (sample). Furthermore, the depth-sensitivity of the spatially-resolved information from this was many times more localised than the sample thickness and detailed comparison with simulation showed that the object behaved as effectively two-dimensional.

The defocus at which different spatial-frequencies came to peak contrast was found to vary, both radially and azimuthally, and was successfully used to parametrically fit both round and non-round aberrations and this was verified for the case of a deliberately introduced two-fold astigmatism. However, going forward it would be useful to expand this verification method to include other aberrations (where it is possible to introduce them in a similarly controlled manner). Expanding this verification to incorporate other forms of comparator aberration diagnosis (such as analysis of experimental ronchigrams) would further validate the method.

So far the aberration fitting has made use of the FT-modulus data only and as a result is limited to the fitting of odd-order, even-fold azimuthal symmetry, aberrations. In future a more full analysis incorporating the FT-stack phase to diagnose the first few even-order aberrations would be merited.

Given an 'over-determined' STEM focal-series data-set can the illumination's effects be separated to leave only an image of the sample free from any imaging distortions?

Using knowledge of the defocus-spread of the usable contrast within an FT-stack, an 'extraction filter' was designed; then from probes simulation a Fourier-space weighting (modulation) function was calculated. These two operations together reconstruct a single two-dimensional Fourier transform that when inverted yield a reconstructed STEM image. This STEM image, drawing on information from across many original data frames, exhibited significantly improved resolution and SNR performance (up to 20% and 400% respectively) as well as partial compensation for other more slowly varying effects such as low-frequency scan-distortions and fluctuations in tip emission current (approximately 75% and 50% respectively).

In addition to the inclusion of additional aberrations as described above, there are still some improvements that can be made to the methodology both scientifically and in the software. At present the application of the extraction-filter and the modulation-function are performed sequentially. Combining these two into a single 3D weighted-filtering operation would simplify the computer code (always preferable to prevent the introduction of errors) and bring the method more in line with the single-step filtering operations used in the analogous TEM FSR methods.

Additionally the current STEM FSR method does not as yet incorporate any consideration of the finite source size and emission energy spread. The FSR method restores the data from that which would be consistent with the experimental probe, including aberrations, noise and instabilities, to that which would be recorded by a probe free from these effects. The restored data, whist its resolution has been improved, still appears blurred (convolved) with these two remaining effects.

Some sort of deconvolution of these may yield an even higher resolution image more closely matching the pure 2D object function.

6.2. Experimental Confirmation of the 2D Object Model from a Monolayer Sample

The most accurate approximation to the 2D object function comes with a genuinely 2D sample. Now with a family of atomic mono-layer specimens available such as graphene or boron-nitride it would be merited to analyse focal-series data collected from these materials. In these samples electron channelling becomes irrelevant and the image information is known to come precisely from the entrant surface of the sample. Such samples – if they can survive the imaging dose needed for the acquisition – may allow for even higher performance focal-series to be acquired with the potential to directly measure source size or coherence effects.

6.3. Utilisation of the Fourier Data Volume Phase Information

It was discussed earlier in Section 4.3, that remnant even-order aberrations including coma, may manifest themselves as a phase deviation in the Fourier-space representation of the data-volume. Whilst this has not been incorporated in the current restoration code it would be a desirable addition in the future. With that in mind, it is worth documenting some of the avenues that may bear fruit.

The phase component of image Fourier transforms (whether experimental or reconstructed) is rarely utilised, perhaps because of its daunting form (Figure 117 - left). However, much of this complexity arises from the inbuilt phase wrapping associated with viewing such data – that is representing the phase on a scale of $-\pi$ to $+\pi$. This wrapping, whilst physically meaningful and not-innacurate can make patterns and trends in the data more difficult to observe by eye. Fortunately many codes exist to 'unwrap' these phases. This process involves identifying jumps of 2π and offsetting the adjoining data, usually with a metric based on gradient smoothness. Using such a tool [132] the same phase plot can be visualised in its unwrapped form (Figure 117 – right).

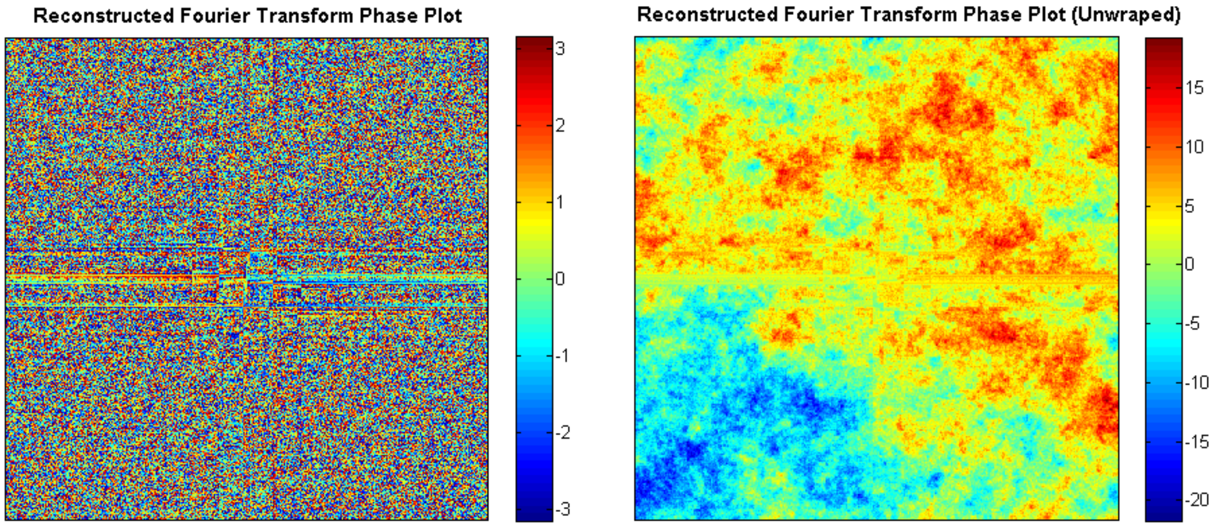


Figure 117. Phase plot of the Reconstructed Fourier Transform corresponding to the example data shown in Figure 111, expressed as phase wrapped (left) and unwrapped (right). Colour-scales show radians.

After phase unwrapping we begin to see more of a planar ramp nature appear, but there is additional structure superimposed upon that behaviour that may correspond to the phase maps simulated earlier (Figure 57).

If the expected forms of the phase-surfaces from section 2.1 were to be described parametrically, then it may be possible to diagnose odd-fold azimuthal symmetry aberrations and possibly designs a more advanced restoring filter. In this case any fitting procedure should be weighted pixel-wise by the FT modulus to avoid the fitting being skewed by noisy data.

Study of such data and fitting of the expected forms of these phase shifts may allow for these additional aberrations to be compensated for also.

6.4. Further Investigation of Secondary Maxima in the FT-spot Contrast Profiles

In Figure 86 we saw that for the case of a very thick (250 nm) sample a secondary maxima was observed in the tracked contrast transfer strength (FT-spot modulus). The fitted Gaussian used to fit the peak contrast-transfer defocus and defocus spread appear to have not been affected by the presence of this secondary maxima but ideally a fully general fitting solution should be developed.

To achieve this first a more detailed study of the phase should be conducted in the region near this secondary maxima. To improve the fit to the FT-spots it was usually the case that say a 3x3 pixel region about the spot was integrated; this has the effect of clarifying trends in the FT modulus but can mask variations in phase if the angles sum progressively to zero. In fact the phase plots shown in Figure 94 were taken from only a single FT-pixel (no lateral integration) to avoid this potential problem. From probe simulation it should be possible to predict the defocus offset of the first contrast-transfer minima (peak in modulus but with an anti-phase shift) and subsequently the secondary, in-phase, maxima. Next, once the phase relationship of the two maxima has been reliably determined the image reconstruction methodology may potential be improved. In the case that this secondary maxima in modulus remains in-phase then this represents additional contrast that can be utilised in the image reconstruction; if anti-phase then at least this region should be avoided or at best this negative-contrast may be subtracted. In either case this may squeeze an additional 5-10% image contrast from the same incident electron dose.

6.5. Restoring Weak Parallel Analytical Signals by Focal-Series

The experimental acquisition required for focal-series reconstruction necessarily requires significant time (a typical line-synched, 512x512 pixel, 30 frame date-set takes around five minutes to acquire). For this reason it may be unsuitable to beam-sensitive specimens. However, for samples able to withstand the imaging dose additional information may be possible to be collected. If parallel analytical signals such as EDX or EELS spectra are recorded with the same pixel sampling as the HAADF data set, then having the same illuminating probe, it is possible to process such data in the same way as presented in Chapter 5. The dark-field signal can be used to diagnose the remnant aberrations and to define the extraction filter (section 5.4) and the modulation function (section 5.5). As each of these are properties of the probe, and the same probe is responsible for all incoherent signals then these can be used to reconstruct the analytical signals with greatly improved signal-noise.

6.6. Investigation of the Signal-Background Ratio with Thickness

Quantitative STEM allows for the image intensity to be expressed on an absolute scale and offers the possibility to correctly calculate the signal-background ratio (SBR). Physically this is analogous to the proportion of the probe intensity that remains channelling 'on-column' compared with the amount that has de-channelled and become more diffuse. With the STEM FSR technique reliant on channelling to provide the necessary 2D object function a study of this SBR as a function of thickness would be useful as it would more precisely define an upper limit of sample thickness suitable for the method.

7. References

[1] A. V. Crewe, “High-resolution scanning transmission electron microscopy,” *Science*, vol. 221, no. 4608, pp. 325–330, 1983.

[2] A. Tonomura, “Applications of electron holography,” *Reviews of Modern Physics*, vol. 59, no. 3, pp. 639–669, Jul. 1987.

[3] SPI Supplies, “Some Comments on the Differences Between Tungsten Thermionic vs. LaB6 Sources for Electron Microscopes,” 2006. .

[4] A. Tonomura, T. Matsuda, J. Endo, H. Todokoro, and T. Komoda, “Development of a Field Emission Electron Microscope,” *Microscopy (Tokyo)*, vol. 28, no. 1, pp. 1–11, Jan. 1979.

[5] N. de Jonge, Y. Lamy, K. Schoots, and T. H. Oosterkamp, “High brightness electron beam from a multi-walled carbon nanotube.,” *Nature*, vol. 420, no. 6914, pp. 393–5, Nov. 2002.

[6] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy*, 2nd ed. Springer, 2009.

[7] P. D. Nellist, P. Hawkes, and J. C. H. Spence, “Science of microscopy,” *interscience.wiley.com*, vol. 1, pp. 65 – 132, 2007.

[8] J. M. LeBeau, S. D. Findlay, L. J. Allen, and S. Stemmer, “Standardless atom counting in scanning transmission electron microscopy.,” *Nano letters*, vol. 10, no. 11, pp. 4405–8, Nov. 2010.

[9] C. Maunders, C. Dwyer, P. C. Tiemeijer, and J. Etheridge, “Practical methods for the measurement of spatial coherence--a comparative study.,” *Ultramicroscopy*, vol. 111, no. 8, pp. 1437–46, Jul. 2011.

[10] C. Dwyer, R. Erni, and J. Etheridge, “Measurement of effective source distribution and its importance for quantitative interpretation of STEM images,” *Ultramicroscopy*, vol. 110, no. 8, pp. 952–957, Jul. 2010.

[11] J. Verbeeck, A. Béché, and W. Van den Broek, “A holographic method to measure the source size broadening in STEM.,” *Ultramicroscopy*, vol. 120, pp. 35–40, Sep. 2012.

[12] R. Klie, “Reaching a new resolution standard with electron microscopy,” *Physics*, vol. 2, p. 85, Oct. 2009.

[13] A. S. A. Alamir, “On the chromatic aberration of magnetic lenses with a field distribution in the form of an inverse power law (B(z) ∝ z–n),” *Optik - International Journal for Light and Electron Optics*, vol. 115, no. 5, pp. 227–231, 2004.

[14] P. Wang, G. Behan, M. Takeguchi, A. Hashimoto, K. Mitsuishi, M. Shimojo, A. I. Kirkland, and P. D. Nellist, “Nanoscale Energy-Filtered Scanning Confocal Electron Microscopy Using a Double-Aberration-Corrected Transmission Electron Microscope,” *Physical Review Letters*, vol. 104, no. 20, May 2010.

[15] J. Goldstein, D. Newbury, D. Joy, P. Echlin, and C. Lyman, *Scanning electron microscopy and X-ray microanalysis*. Springer, 2003, pp. 48 – 50.

[16] A. Y. Borisevich, A. R. Lupini, S. Travaglini, and S. J. Pennycook, "Depth sectioning of aligned crystals with the aberration-corrected scanning transmission electron microscope," *Journal of Electron Microscopy*, vol. 55, no. 1, pp. 7–12, Mar. 2006.

[17] J. C. H. Spence, "High-Resolution Electron Microscopy," in in *High-Resolution Electron Microscopy*, Oxford University Press, 2008, pp. 41–44.

[18] O. L. Krivanek, J. P. Ursin, N. J. Bacon, G. J. Corbin, N. Dellby, P. Hrncirik, M. Murfitt, C. S. Own, and Z. S. Szilagyi, "High-energy-resolution monochromator for aberration-corrected scanning transmission electron microscopy/electron energy-loss spectroscopy.," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 367, no. 1903, pp. 3683–97, Sep. 2009.

[19] M. Mukai and T. Kaneyama, "'MIRAI-21' Analytical Electron Microscope - Performance of the Monochromator," *Microscopy and Microanalysis*, vol. 9, no. S02, pp. 954–955, 2003.

[20] L. J. W. S. Rayleigh, *Wave Theory of Light in Scientific Papers*. Cambridge University Press, 1902, pp. 47 – 189.

[21] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge: Cambridge University Press, 1999.

[22] O. Scherzer, "The Theoretical Resolution Limit of the Electron Microscope," *Journal of Applied Physics*, vol. 20, no. 1, p. 20, 1949.

[23] C. J. Koester, "Scanning mirror microscope with optical sectioning characteristics: applications in ophthalmology," *Applied Optics*, vol. 19, no. 11, pp. 1749–1757, 1980.

[24] G. Behan, E. C. Cosgriff, A. I. Kirkland, and P. D. Nellist, "Three-dimensional imaging by optical sectioning in the aberration-corrected scanning transmission electron microscope.," *Philosophical transactions. Series A*, vol. 367, no. 1903, pp. 3825–44, Sep. 2009.

[25] J.-A. Conchello and J. W. Lichtman, "Optical sectioning microscopy.," *Nature methods*, vol. 2, no. 12, pp. 920–31, Dec. 2005.

[26] P. D. Nellist, E. C. Cosgriff, G. Behan, and A. I. Kirkland, "Imaging Modes for Scanning Confocal Electron Microscopy in a Double Aberration-Corrected Transmission Electron Microscope," *Microscopy and Microanalysis*, vol. 14, no. 01, pp. 82–88, Feb. 2008.

[27] A. Y. Borisevich, A. R. Lupini, and S. J. Pennycook, "Depth sectioning with the aberration-corrected scanning transmission electron microscope.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 9, pp. 3044–8, Feb. 2006.

[28] P. E. Batson, N. Dellby, and O. L. Krivanek, "Sub-ångstrom resolution using aberration corrected electron optics.," *Nature*, vol. 418, no. 6898, pp. 617–20, Aug. 2002.

[29] H. von Harrach, "Instrumental factors in high-resolution FEG STEM," *Ultramicroscopy*, vol. 58, no. 1, pp. 1–5, 1995.

[30] D. A. Muller and J. Grazul, “Optimizing the environment for sub-0.2 nm scanning transmission electron microscopy.,” *Journal of Electron Microscopy*, vol. 50, no. 3, pp. 219–26, Jan. 2001.

[31] D. A. Muller, E. J. Kirkland, M. G. Thomas, J. L. Grazul, L. Fitting, and M. Weyland, “Room design for high-performance electron microscopy.,” *Ultramicroscopy*, vol. 106, no. 11–12, pp. 1033–40, 2006.

[32] M. Haider, H. Müller, S. Uhlemann, J. Zach, U. Loebau, and R. Hoeschen, “Prerequisites for a Cc/Cs-corrected ultrahigh-resolution TEM.,” *Ultramicroscopy*, vol. 108, no. 3, pp. 167–78, Feb. 2008.

[33] A. Recnik, G. Möbus, and S. Sturm, “IMAGE-WARP: a real-space restoration method for high-resolution STEM images using quantitative HRTEM analysis.,” *Ultramicroscopy*, vol. 103, no. 4, pp. 285–301, Jul. 2005.

[34] N. Braidy, Y. Le Bouar, S. Lazar, and C. Ricolleau, “Correcting scanning instabilities from images of periodic structures.,” *Ultramicroscopy*, vol. 118, no. null, pp. 67–76, Jul. 2012.

[35] D. C. Joy, D. E. Newbury, and D. L. Davidson, “Electron channeling patterns in the scanning electron microscope,” *Journal of Applied Physics*, vol. 53, no. 8, pp. R81–R122, Jan. 1982.

[36] M. Berry, “Diffraction in crystals at high energies,” *Journal of Physics C: Solid State Physics*, 1971.

[37] D. Van Dyck and M. Opdebeeck, “A simple intuitive theory for electron diffraction,” *Ultramicroscopy*, vol. 64, no. 1–4, pp. 99–107, Aug. 1996.

[38] D. Van Dyck, G. Vantendeloo, and S. Amelinckx, “Direct interpretation of high resolution electron images of substitutional alloy systems with a column structure,” *Ultramicroscopy*, vol. 10, no. 3, pp. 263–280, 1982.

[39] S. E. Maccagnano-Zacher, K. A. Mkhoyan, E. J. Kirkland, and J. Silcox, “Effects of tilt on high-resolution ADF-STEM imaging.,” *Ultramicroscopy*, vol. 108, no. 8, pp. 718–26, Jul. 2008.

[40] R. Hamminger, G. Grathwohl, and F. Thümmler, “Microanalytical investigation of sintered SiC,” *Journal of Materials Science*, vol. 18, no. 2, pp. 353–364, Feb. 1983.

[41] J. Cowley, “Image contrast in a transmission scanning electron microscope,” *Applied Physics Letters*, vol. 15, p. 58, 1969.

[42] E. C. Cosgriff, P. D. Nellist, A. Alfonso, S. D. Findlay, G. Behan, and P. Wang, “Image Contrast in Aberration-Corrected Scanning Confocal Electron Microscopy,” *Imaging and Electron Physics*, vol. 162, pp. 45–76, 2010.

[43] H. L. Xin and D. A. Muller, “Aberration-corrected ADF-STEM depth sectioning and prospects for reliable 3D imaging in S/TEM.,” *Journal of Electron Microscopy*, vol. 58, no. 3, pp. 157–65, Jun. 2009.

[44] D. Jesson and S. J. Pennycook, “Incoherent Imaging of Thin Specimens Using Coherently Scattered Electrons,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 441, no. 1912, pp. 261–281, May 1993.

[45] S. J. Pennycook, A. R. Lupini, M. Varela, A. Y. Borisevich, Y. Peng, M. P. Oxley, K. van Benthem, and M. F. Chisholm, *Scanning Microscopy for Nanotechnology: Techniques And Applications*. Springer, 2007, pp. 152–191.

[46] P. D. Nellist and S. J. Pennycook, “The principles and interpretation of annular dark-field Z-contrast imaging,” *Advances in Imaging and Electron Physics*, vol. 113, pp. 147–203, 2000.

[47] S. J. Pennycook and D. Jesson, “High-resolution incoherent imaging of crystals,” *Physical Review Letters*, vol. 64, no. 8, pp. 938–941, Feb. 1990.

[48] J. R. McBride, T. C. Kippeny, S. J. Pennycook, and S. J. Rosenthal, “Aberration-Corrected Z-Contrast Scanning Transmission Electron Microscopy of CdSe Nanocrystals,” *Nano Letters*, vol. 4, no. 7, pp. 1279–1283, Jul. 2004.

[49] P. D. Nellist and S. J. Pennycook, “Accurate structure determination from image reconstruction in ADF STEM,” *Journal of Microscopy*, vol. 190, no. 1–2, pp. 159–170, Apr. 1998.

[50] M. Vijayalakshmi, S. Saroja, and R. Mythili, “Convergent beam electron diffraction—A novel technique for materials characterisation at sub-microscopic levels,” *Sadhana*, vol. 28, no. 3–4, pp. 763–782, Jun. 2003.

[51] Y. Kotaka, “Essential experimental parameters for quantitative structure analysis using spherical aberration-corrected HAADF-STEM.,” *Ultramicroscopy*, vol. 110, no. 5, pp. 555–62, Apr. 2010.

[52] P. D. Nellist, “Incoherent imaging using dynamically scattered coherent electrons,” *Ultramicroscopy*, vol. 78, no. 1–4, pp. 111–124, Jun. 1999.

[53] O. L. Krivanek, M. F. Chisholm, V. Nicolosi, T. J. Pennycook, G. J. Corbin, N. Dellby, M. Murfitt, C. S. Own, Z. S. Szilagyi, M. P. Oxley, S. T. Pantelides, and S. J. Pennycook, “Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy.,” *Nature*, vol. 464, no. 7288, pp. 571–4, Mar. 2010.

[54] Z. Wang, Z. Li, S. Park, A. Abdela, D. Tang, and R. Palmer, “Quantitative Z-contrast imaging in the scanning transmission electron microscope with size-selected clusters,” *Physical Review B*, vol. 84, no. 7, p. 073408, Aug. 2011.

[55] R. F. Loane, P. Xu, and J. Silcox, “Incoherent imaging of zone axis crystals with ADF STEM,” *Ultramicroscopy*, vol. 40, no. 2, pp. 121–138, Feb. 1992.

[56] a J. Koster, R. Grimm, D. Typke, R. Hegerl, a Stoschek, J. Walz, and W. Baumeister, “Perspectives of molecular and cellular electron tomography.,” *Journal of structural biology*, vol. 120, no. 3, pp. 276–308, Dec. 1997.

[57] J. C. H. Spence and J. Lynch, “STEM microanalysis by transmission electron energy loss spectroscopy in crystals,” *Ultramicroscopy*, vol. 9, no. 3, pp. 267–276, 1982.

[58] C. Jeanguillame and C. Colliex, “Spectrum-image: The next step in EELS digital acquisition and processing,” *Ultramicroscopy*, vol. 28, no. 1–4, pp. 252–257, Apr. 1989.

[59] O. L. Krivanek, C. Ahn, and R. Keeney, "Parallel detection electron spectrometer using quadrupole lenses," *Ultramicroscopy*, vol. 22, no. 1–4, pp. 103–115, 1987.

[60] D. A. Muller, L. F. Kourkoutis, M. Murfitt, J. H. Song, H. Y. Hwang, J. Silcox, N. Dellby, and O. L. Krivanek, "Atomic-scale chemical imaging of composition and bonding by aberration-corrected microscopy.," *Science (New York, N.Y.)*, vol. 319, no. 5866, pp. 1073–6, Feb. 2008.

[61] N. D. Browning, M. F. Chisholm, and S. J. Pennycook, "Atomic resolution chemical analysis," *Advanced Materials*, vol. 6, no. 4, pp. 328–331, Apr. 1994.

[62] M. Varela, S. D. Findlay, A. R. Lupini, H. Christen, A. Y. Borisevich, N. Dellby, O. L. Krivanek, P. D. Nellist, M. P. Oxley, L. J. Allen, and S. J. Pennycook, "Spectroscopic Imaging of Single Atoms Within a Bulk Solid," *Physical Review Letters*, vol. 92, no. 9, Mar. 2004.

[63] W. King, "An experimental technique to measure x-ray production and detection efficiencies in the analytical electron microscope," *Ultramicroscopy*, vol. 18, no. 1–4, pp. 151–154, 1985.

[64] C. Lyman, D. B. Williams, and J. Goldstein, "X-ray detectors and spectrometers," *Ultramicroscopy*, vol. 28, no. 1–4, pp. 137–149, Apr. 1989.

[65] R. D. Leapman and J. A. Hunt, "Comparison of detection limits for EELS and EDXS," *Microscopy Microanalysis Microstructures*, vol. 2, no. 2–3, pp. 231–244, 1991.

[66] O. Scherzer, "Uber einige Fehler von Elektronenlinsen (some defects of electron lenses)," *Optik*, vol. 101, no. 9–10, pp. 593–603, Sep. 1936.

[67] Z. Shao and A. V. Crewe, "Ideal lenses and the Scherzer theorem: A supplement," *Ultramicroscopy*, vol. 26, no. 4, pp. 385–388, 1988.

[68] E. Ruska, "The development of the electron microscope and of electron microscopy," *Reviews of Modern Physics*, vol. 59, no. 3, pp. 627–638, Jul. 1987.

[69] H. Sawada, T. Sannomiya, F. Hosokawa, T. Nakamichi, T. Kaneyama, T. Tomita, Y. Kondo, T. Tanaka, Y. Oshima, Y. Tanishiro, and K. Takayanagi, "Measurement method of aberration from Ronchigram by autocorrelation function.," *Ultramicroscopy*, vol. 108, no. 11, pp. 1467–75, Oct. 2008.

[70] H. Müller, S. Uhlemann, P. Hartel, and M. Haider, "Advancing the Hexapole Cs-Corrector for the Scanning Transmission Electron Microscope," *Microscopy and Microanalysis*, vol. 12, no. 06, pp. 442–455, Dec. 2006.

[71] S. Uno, "Aberration correction and its automatic control in scanning electron microscopes," *Optik - International Journal for Light and Electron Optics*, vol. 116, no. 9, pp. 438–448, Sep. 2005.

[72] F. Thon, "On the defocusing dependence of phase contrast in electron microscopical images," *Naturforsch*, vol. 21a, pp. 476–478, 1966.

[73] W. O. Saxton, "A new way of measuring microscope aberrations," *Ultramicroscopy*, vol. 81, no. 2, pp. 41–45, Mar. 2000.

[74] K. Urban, B. Kabius, M. Haider, and H. Rose, “A way to higher resolution: spherical-aberration correction in a 200 kV transmission electron microscope,” *Journal of Electron Microscopy*, vol. 48, no. 6, pp. 821–826, Jan. 1999.

[75] K. Tani, “A set of computer programs for determining defocus and astigmatism in electron images,” *Ultramicroscopy*, vol. 65, no. 1–2, pp. 31–44, Sep. 1996.

[76] S. P. Mallick, B. Carragher, C. S. Potter, and D. J. Kriegman, “ACE: automated CTF estimation.,” *Ultramicroscopy*, vol. 104, no. 1, pp. 8–29, Aug. 2005.

[77] J. Lin and J. Cowley, “Calibration of the operating parameters for an HB5 STEM instrument,” *Ultramicroscopy*, vol. 19, pp. 31–42, 1986.

[78] A. R. Lupini, P. Wang, P. D. Nellist, A. I. Kirkland, and S. J. Pennycook, “Aberration measurement using the Ronchigram contrast transfer function.,” *Ultramicroscopy*, vol. 110, no. 7, pp. 891–8, Jun. 2010.

[79] O. L. Krivanek, N. Dellby, and A. R. Lupini, “Towards sub-Å electron beams,” *Ultramicroscopy*, vol. 78, no. 1–4, pp. 1–11, Jun. 1999.

[80] A. R. Lupini and S. J. Pennycook, “Rapid autotuning for crystalline specimens from an inline hologram.,” *Journal of Electron Microscopy*, vol. 57, no. 6, pp. 195–201, Dec. 2008.

[81] M. Haider, S. Uhlemann, E. Schwan, H. Rose, and B. Kabius, “Electron microscopy image enhanced,” *Nature*, vol. 392, no. 6678, pp. 768–769, Apr. 1998.

[82] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, and K.-H. Herrmann, “Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms,” *Ultramicroscopy*, vol. 3, pp. 49–60, 1978.

[83] M. Overwijk, A. Bleeker, and A. Thust, “Correction of three-fold astigmatism for ultra-high-resolution TEM,” *Ultramicroscopy*, vol. 67, no. 1–4, pp. 163–170, Jun. 1997.

[84] R. R. Meyer, A. I. Kirkland, and W. O. Saxton, “A new method for the determination of the wave aberration function for high-resolution TEM.; 2. Measurement of the antisymmetric aberrations.,” *Ultramicroscopy*, vol. 99, no. 2–3, pp. 115–23, May 2004.

[85] R. R. Meyer, A. I. Kirkland, and W. O. Saxton, “A new method for the determination of the wave aberration function for high resolution TEM 1. Measurement of the symmetric aberrations.,” *Ultramicroscopy*, vol. 92, no. 2, pp. 89–109, Jul. 2002.

[86] P. D. Nellist, M. F. Chisholm, N. Dellby, O. L. Krivanek, M. F. Murfitt, Z. S. Szilagyi, A. R. Lupini, A. Y. Borisevich, W. H. Sides, and S. J. Pennycook, “Direct sub-angstrom imaging of a crystal lattice.,” *Science (New York, N.Y.)*, vol. 305, no. 5691, p. 1741, Sep. 2004.

[87] A. R. Lupini and S. J. Pennycook, “Rapid Methods for Dynamic Autotuning,” *Microscopy and Microanalysis*, vol. 16, no. S2, pp. 244–245, Aug. 2010.

[88] K. Wong, E. Kirkland, P. Xu, R. Loane, and J. Silcox, “Measurement of spherical aberration in STEM,” *Ultramicroscopy*, vol. 40, no. 2, pp. 139–150, Feb. 1992.

[89] N. Baba, K. Terayama, T. Yoshimizu, N. Ichise, and N. Tanaka, “An auto-tuning method for focusing and astigmatism correction in HAADF-STEM, based on the image contrast transfer function,” *Microscopy*, vol. 50, no. 3, pp. 163–176, May 2001.

[90] M. Bischoff and B. Rieger, “Method for Adjusting a STEM Equipped with an Aberration Corrector,” 2013/01056892013.

[91] M. Haider, “Upper limits for the residual aberrations of a high-resolution aberration-corrected STEM,” *Ultramicroscopy*, vol. 81, no. 3–4, pp. 163–175, Apr. 2000.

[92] P. W. Hawkes, “Some advances in electron optics since CPO-5,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 519, no. 1–2, pp. 1–6, Feb. 2004.

[93] H. Rose, “Correction system for a charged-particle beam apparatus,” 5084622Jan-1992.

[94] N. Dellby, O. L. Krivanek, M. Murfitt, and N. Co, “Optimized Quadrupole-Octupole C3/C5 Aberration Corrector for STEM,” *Physics Procedia*, vol. 1, no. 1, pp. 179–183, Aug. 2008.

[95] A. V. Crewe, “Aberration free lens system for electron microscope,” *US Patent 5,336,891*, 1994.

[96] C. S. Own, W. Sinkler, and L. D. Marks, “Prospects for aberration corrected electron precession.,” *Ultramicroscopy*, vol. 107, no. 6–7, pp. 534–42, 2007.

[97] K. van der Mast and A. De Jong, “Charged particle beam device,” *US Patent 5,221,844*, 1993.

[98] N. Nakanishi, “Retrieval process of high-resolution HAADF-STEM images,” *Journal of Electron Microscopy*, vol. 51, no. 6, pp. 383–390, Nov. 2002.

[99] S. J. Pennycook, M. F. Chisholm, A. R. Lupini, M. Varela, A. Y. Borisevich, M. P. Oxley, W. D. Luo, K. van Benthem, S.-H. Oh, D. L. Sales, S. I. Molina, J. García-Barriocanal, C. Leon, J. Santamaría, S. N. Rashkeev, and S. T. Pantelides, “Aberration-corrected scanning transmission electron microscopy: from atomic imaging and analysis to solving energy problems.,” *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 367, no. 1903, pp. 3709–33, Sep. 2009.

[100] E. . James and N. . Browning, “Practical aspects of atomic resolution imaging and analysis in STEM,” *Ultramicroscopy*, vol. 78, no. 1–4, pp. 125–139, Jun. 1999.

[101] A. R. Lupini and N. de Jonge, “The three-dimensional point spread function of aberration-corrected scanning transmission electron microscopy.,” *Microscopy and Microanalysis*, vol. 17, no. 5, pp. 817–26, Oct. 2011.

[102] J. M. LeBeau, S. D. Findlay, L. J. Allen, and S. Stemmer, “Quantitative Atomic Resolution Scanning Transmission Electron Microscopy,” *Physical Review Letters*, vol. 100, no. 20, p. 206101–, May 2008.

[103] Y. Taniguchi, Y. Takai, T. Ikuta, and R. Shimizu, “Correction of Spherical Aberration in HREM Image Using Defocus-Modulation Image Processing,” *Journal of Electron Microscopy*, vol. 41, no. 1, pp. 21–29, Feb. 1992.

[104] Y. Takai, Y. Taniguchi, T. Ikuta, and R. Shimizu, "Spherical-aberration-free observation of profile images of the Au (011) surface by defocus-modulation image processing," *Ultramicroscopy*, vol. 54, no. 2–4, pp. 250–260, 1994.

[105] C. Y. Tang, J. H. Chen, H. W. Zandbergen, and F. H. Li, "Image deconvolution in spherical aberration-corrected high-resolution transmission electron microscopy.," *Ultramicroscopy*, vol. 106, no. 6, pp. 539–46, Apr. 2006.

[106] F.-R. Chen, "Extension of HRTEM resolution by semi-blind deconvolution method and Gerchberg-Saxton algorithm: application to grain boundary and interface," *Journal of Electron Microscopy*, vol. 50, no. 6, pp. 529–540, Nov. 2001.

[107] A. I. Kirkland and R. R. Meyer, "High-Resolution Transmission Electron Microscopy: Aberration Measurement and Wavefunction Reconstruction," *Microscopy and microanalysis the official journal of Microscopy Society of America Microbeam Analysis Society Microscopical Society of Canada*, vol. 10, no. 4, pp. 401–413, 2004.

[108] M. Op de Beeck, D. Van Dyck, and W. Coene, "Wave function reconstruction in HRTEM: the parabola method," *Ultramicroscopy*, vol. 64, no. 1–4, pp. 167–183, Aug. 1996.

[109] L. C. Gontard, L.-Y. Chang, R. E. Dunin-Borkowski, a I. Kirkland, C. J. D. Hetherington, and D. Ozkaya, "The application of spherical aberration correction and focal series restoration to high-resolution images of platinum nanocatalyst particles," *Journal of Physics: Conference Series*, vol. 26, pp. 25–28, Feb. 2006.

[110] A. Steinecker and W. Mader, "Exit wave reconstruction using a conventional transmission electron microscope," *Journal of Microscopy*, vol. 190, no. 1–2, pp. 281–290, Apr. 1998.

[111] O. L. Krivanek, G. J. Corbin, N. Dellby, B. F. Elston, R. J. Keyse, M. F. Murfitt, C. S. Own, Z. S. Szilagyi, and J. W. Woodruff, "An electron microscope for the aberration-corrected era.," *Ultramicroscopy*, vol. 108, no. 3, pp. 179–95, Feb. 2008.

[112] G. Black and E. H. Linfoot, "Spherical Aberration and the Information Content of Optical Images," *Proceedings of the Royal Society A*, vol. 239, no. 1219, pp. 522–540, Apr. 1957.

[113] P. D. Nellist and S. J. Pennycook, "Subangstrom Resolution by Underfocused Incoherent Transmission Electron Microscopy," *Physical Review Letters*, vol. 81, no. 19, pp. 4156–4159, Nov. 1998.

[114] B. R. Frieden, "Optical Transfer of the Three-Dimensional Object," *Journal of the Optical Society of America*, vol. 57, no. 1, p. 56, Jan. 1967.

[115] M. Arnison, "A 3D vectorial optical transfer function suitable for arbitrary pupil functions," *Optics Communications*, vol. 211, no. 1–6, pp. 53–63, Oct. 2002.

[116] P. D. Nellist and P. Wang, "Optical Sectioning and Confocal Imaging and Analysis in the Transmission Electron Microscope," *Annual Review of Materials Research*, vol. 42, no. 1, pp. 125–143, Aug. 2012.

[117] P. M. M. Voyles, J. L. Grazul, and D. A. A. Muller, "Imaging individual atoms inside crystals with ADF-STEM.," *Ultramicroscopy*, vol. 96, no. 3–4, pp. 251–73, Sep. 2003.

[118] D. A. Blom, L. F. Allard, S. Mishina, and M. a. O'Keefe, "Early Results from an Aberration-Corrected JEOL 2200FS STEM/TEM at Oak Ridge National Laboratory," *Microscopy and Microanalysis*, vol. 12, no. 06, p. 483, Oct. 2006.

[119] K. Du, Y. Rau, N. Jin-Phillipp, and F. Phillipp, "Lattice distortion analysis directly from high resolution transmission electron microscopy images-the LADIA program package," *Cailiao Kexue Yu Jishu(Journal of …*, 2002.

[120] P. L. Galindo, S. Kret, A. M. Sanchez, J.-Y. Laval, A. Yáñez, J. Pizarro, E. Guerrero, T. Ben, and S. I. Molina, "The Peak Pairs algorithm for strain mapping from HRTEM images.," *Ultramicroscopy*, vol. 107, no. 12, pp. 1186–93, Nov. 2007.

[121] K. Watanabe, N. Nakanishi, T. Yamazaki, J. R. Yang, S. Y. Huang, K. Inoke, J. T. Hsu, R. C. Tu, and M. Shiojiri, "Atomic-scale strain field and In atom distribution in multiple quantum wells InGaN/GaN," *Applied Physics Letters*, vol. 82, no. 5, p. 715, Feb. 2003.

[122] R. J. Hodrick and E. C. Prescott, "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, vol. 29, no. 1, pp. 1–16, 1997.

[123] P. M. Voyles and D. A. Muller, "Fluctuation microscopy in the STEM," *Ultramicroscopy*, vol. 93, no. 2, pp. 147–159, Nov. 2002.

[124] M. Takeguchi, A. Hashimoto, M. Shimojo, K. Mitsuishi, and K. Furuya, "Development of a stage-scanning system for high-resolution confocal STEM.," *Journal of Electron Microscopy*, vol. 57, no. 4, pp. 123–7, Aug. 2008.

[125] R. F. Klie, C. Johnson, and Y. Zhu, "Atomic-resolution STEM in the aberration-corrected JEOL JEM2200FS," *Microscopy and Microanalysis*, vol. 14, no. 1, pp. 104–112, 2008.

[126] J. D. Nowak and C. B. Carter, "Forming contacts and grain boundaries between MgO nanoparticles," *Journal of Materials Science*, vol. 44, no. 9, pp. 2408–2418, Feb. 2009.

[127] J. M. Lebeau and S. Stemmer, "Experimental quantification of annular dark-field images in scanning transmission electron microscopy.," *Ultramicroscopy*, vol. 108, no. 12, pp. 1653–8, Nov. 2008.

[128] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33, no. 2, p. 156, Jan. 2008.

[129] M. J. Hÿtch, E. Snoeck, and R. Kilaas, "Quantitative measurement of displacement and strain fields from HREM micrographs," *Ultramicroscopy*, vol. 74, no. 3, pp. 131–146, Aug. 1998.

[130] A. Rosenauer, T. Mehrtens, K. Müller, K. Gries, M. Schowalter, P. V. Satyam, S. Bley, C. Tessarek, D. Hommel, K. Sebald, M. Seyfried, J. Gutowski, A. Avramescu, K. Engl, and S. Lutgen, "Composition mapping in InGaN by scanning transmission electron microscopy.," *Ultramicroscopy*, vol. 111, no. 8, pp. 1316–27, Jul. 2011.

[131] P. S. Turner, T. J. Bullough, R. W. Devenish, D. M. Maher, and C. J. Humphreys, "Nanometre hole formation in MgO using electron beams," *Philosophical Magazine Letters*, vol. 61, no. 4, pp. 181–193, Apr. 1990.

[132] M. Costantini, “A Novel Phase Unwrapping Method Based on Network Programming,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 813–821, 1998.

[133] R. Maloney, S. Bogan, and G. Waring, “Determination of Corneal Image-forming Properties from Corneal Topography,” *American journal of Ophthalmology*, vol. 115, pp. 31–41, 1993.

[134] T. Tanabe, A. Tomidokoro, T. Samejima, K. Miyata, M. Sato, Y. Kaji, and T. Oshika, “Corneal regular and irregular astigmatism assessed by Fourier analysis of videokeratography data in normal and pathologic eyes.,” *Ophthalmology*, vol. 111, no. 4, pp. 752–7, Apr. 2004.

[135] F. Gekeler and F. Schaeffel, “Measurement of Astigmatism by Automated Infrared Photoretinoscopy,” *Optometry & Vision Science*, vol. 74, no. 7, pp. 472–82, 1997.

[136] T. W. Raasch, “Corneal topography and irregular astigmatism.,” *Optometry and vision science : official publication of the American Academy of Optometry*, vol. 72, no. 11. pp. 809–15, Nov-1995.

8. Appendices

8.1. Optimising the Scan-Distortion Restoration Iterations Number

Performing a horizontal and vertical distortion analysis sequentially has yielded so far very good results, however the physically the probe-sample distortion offset is a 2D vector with both direction and magnitude. This means that the most robust restoration would involve a 2D comparison of the local environment about each pixel in the raw data.

A 2D analysis would be computationally far more difficult and would increase the processing time by a factor roughly equal to the test-band size, commonly around 20 or so. Even if this could be achieved, perhaps by parallel computing, then there is the question of what this comparison should be made between; when comparing test-bands the ideal comparator would be the undistorted image but this is obviously unavailable at the outset of the reconstruction.

The reconstruction approach followed contains elements that have been empirically optimised. The approach is to sequentially perform the horizontal and vertical analyses (yielding a partially-restored image that may not be totally restored but is better than the original distorted image). Next this partially-restored image is smoothed by convolving it with a Gaussian with a FWHM of one pixel. This removes some of the effects of single-pixel additive noise and give us a smoothed and 'partially-restored comparator-image'. This comparator-image is then used as the reference for a new iteration of the horizontal and vertical analyses. Importantly raw data is used as the image to which the correcting offsets are applied and the smoothed comparator-image is only used to measure the probe-instability offsets. Unless a restored image is to be used again as a comparator for another iteration no smoothing is performed. It is important to note that totally raw data is used during each iteration and that only the comparator image has received any smoothing. As no smoothing is applied to the raw data, the user can choose to perform as many iterations as necessary to generate a comparator-image.

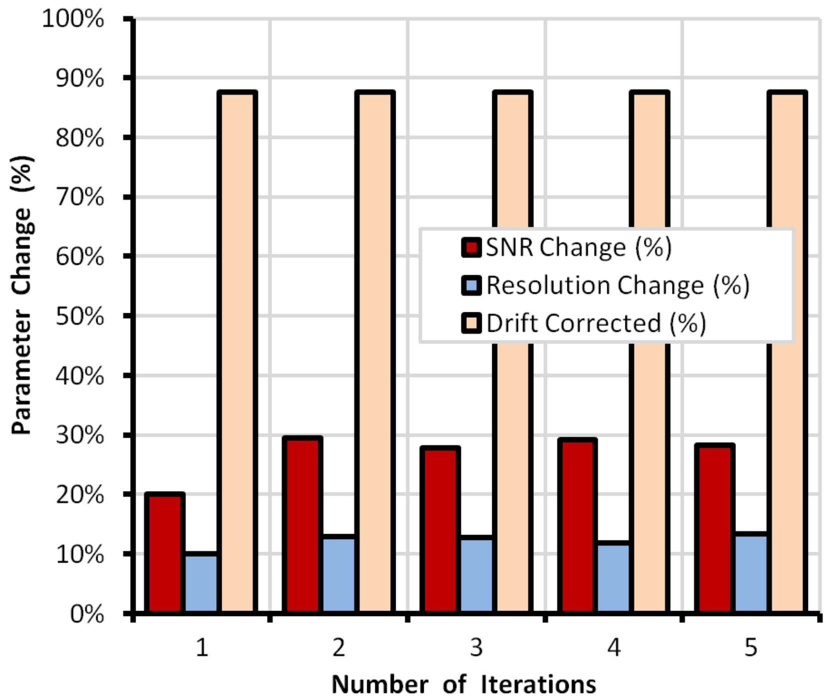


Figure 118. SNR and resolution performance change, degree of drift correction and restoration runtime for one to four iterations.

Figure 118 shows the overall improvement relative to the raw data after an increasing number of iterations. One iteration corresponds to using the raw data as the comparator image and successive values represent images that were restored against the partially-restored comparator-images as described above.

Being only affected by lattice plane angles drift was corrected roughly equally well in all cases.

For this test image both SNR and resolution performance approached a plateau after the second iteration. This point represents when the comparator image is good enough for the final iteration to retrieve the best possible image. Across all the test images inspected the number of iterations needed to reach this point was never greater than four.

We found, as expected, that runtime increased linearly with the number of iterations over and above the constant amount of time needed for quantification of the raw and the restored images. The restoration algorithm has been tested up to 500 iterations but generally no improvement is seen above the first three or four and certainly none that justifies the additional computational time. This extreme case also validates the approach of smoothing the comparator but always restoring the image from raw-data. For reference the data presented in Figure 73 was restored using only two iterations.

8.2. Optimising the Restoration Test-band Size

The 'test-band' is an essential part of the restoration process. It is the 1 pixel high strip of the image that is compared with its neighbours of equal size above and below. The length of the test-band must be optimised. Too short and it could be dominated by random noise and an unreliable comparison will be made; too long and it may attempt to find a better fit with an image row an entire period feature (lattice point) off to one side. As a result and the optimal length is related mainly to the image feature periodicity. This method of comparing test-bands also imparts a somewhat mute restriction – that the amplitude of the distortion must be not greater than the feature spacing.

The restrictions on test-band length can be probed become obvious if any of the restoration parameters, for example, SNR improvement is plotted as a function of varying the test-band size, Figure 119. Note test-band size is always an odd number of pixels as it is defined as a symmetric shoulder of an integer number of pixels either side of the central pixel of interest.

Optimising the size of the test-band to achieve a compromise between sensitivity and reliability is fortunately relatively easy. Figure 119 shows the three metrics that can be used to assess the restoration output; the SNR, resolution change and the degree to which the image drift has been corrected. Again the test-image from Figure 73 was used and had a horizontal feature spacing measured to be 19 pixels.

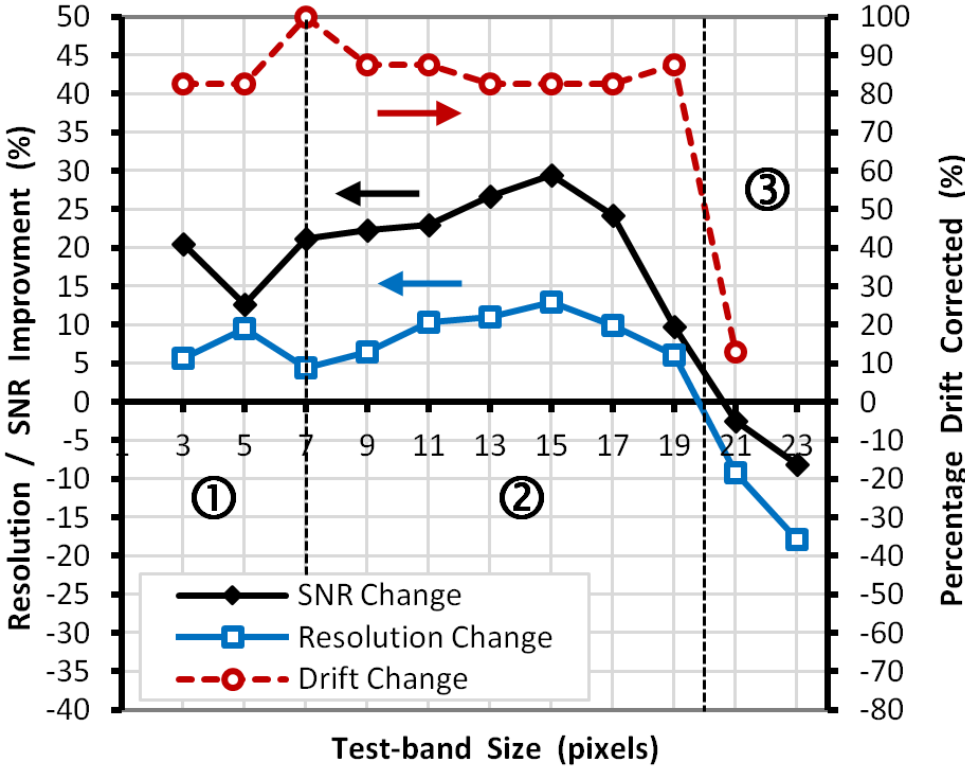


Figure 119. The effect of test-band length on the improvement in restoration SNR and resolution (left axis) and drift strength improvement.

Figure 119 can be divided into three sections based on test-band length:

1. Below seven pixels, the SNR and resolution increase is small. This restricted comparison size has two effects, both limiting the potential for restoration from pixel-shifting, and also using only a very small test band increases the likely hood of random noise coalescence rather than genuine feature realignment. A three-pixel test-band for example compares only the as-recorded neighbour pixels and only one alterative either side of this.
2. Between seven and nineteen pixels, a smoothly peaked improvement of both SNR and resolution are achieved and the drift correction is always greater than 80%. SNR and resolution improvement peaks for a test-band of fifteen pixels wide. Whilst any value within this range will yield an acceptable restoration it is clear that optimising test-band size is worthwhile.
3. Above nineteen pixels in length the test-band length exceeds the feature spacing and, the SNR and resolution improvement is lost. In fact these go negative meaning the image was

degraded by the restoration when the features risk being aligned with the next feature over, effectively scrambling the image. This region should be avoided at all times.

Data for a test-band size of one pixel test-band is not shown as this is equivalent to only comparing the target pixel with one possible set of neighbours; this results in no offset and effectively no restoration being performed.

For the test data evaluated in Figure 119 a test-band length of fifteen pixels would be advised, or more generally, equal to approximately three-quarters the horizontal feature spacing. The process of test-band size optimisation is automated in the restoration code and the code-suggested size for the above data was indeed fifteen pixels.

Drift correction does not use a comparison test-band and is corrected from an analysis of only the raw image. Drift is consistently well corrected (>80%) but when the test-band size exceeds the feature spacing the image begins to become scrambled. The apparent drop in drift-correction for a test-band size of 21 pixels is actually the result of no longer being able to determine the drift present in such a scrambled image after incorrect restoration.

8.3. Commercial Software Development

The scan-noise correction work described in Section 3.2.2 was developed into an automated software for the Matlab environment. This has since been recoded as a plug-in for the Digital Micrograph software, Figure 120.

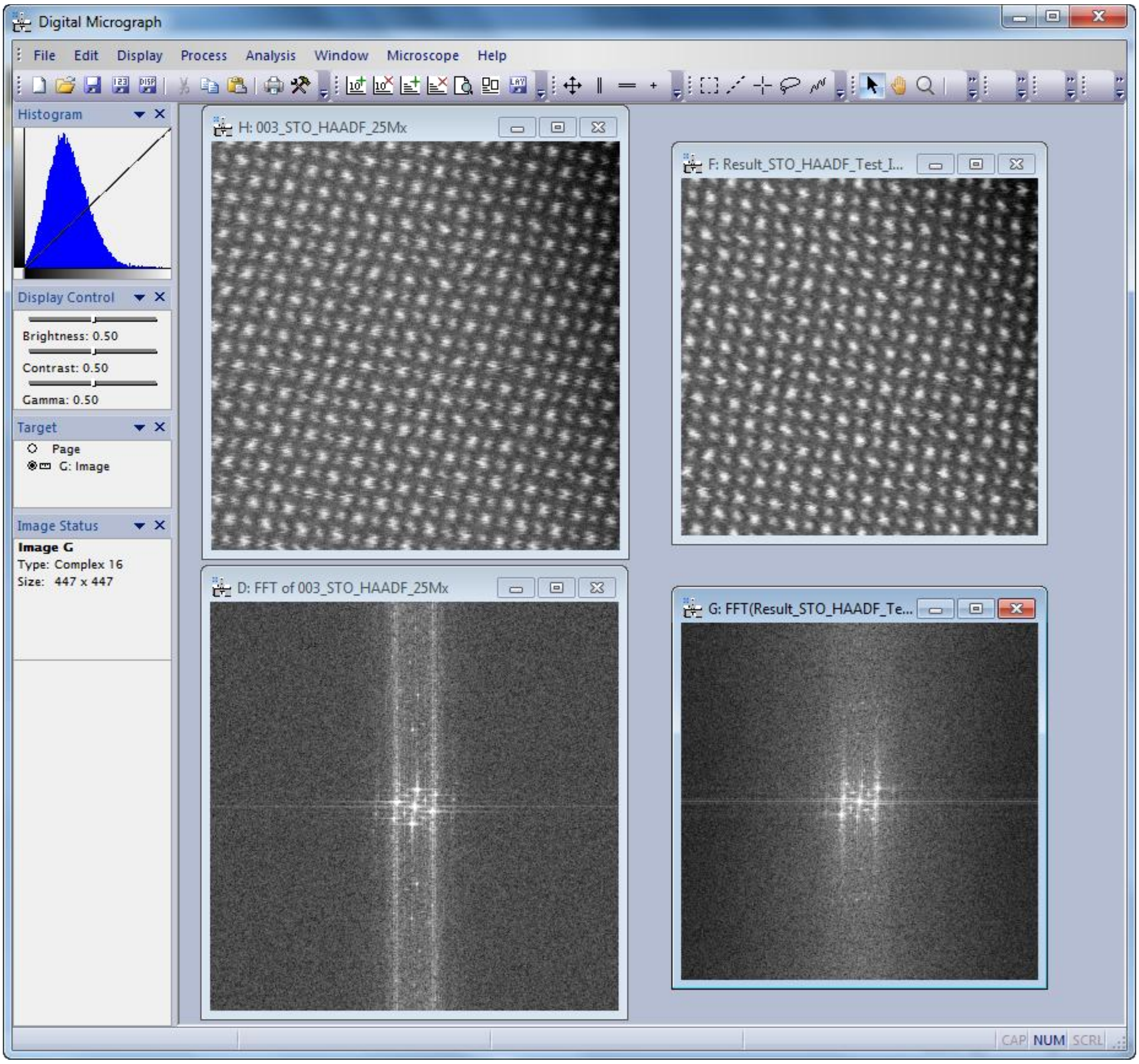


Figure 120. Screenshot of the scan-noise correction code running in Digital Micrograph.

Readers interested in knowing more about this commercial software should contact innovation@isis.ox.ac.uk quoting reference MCR/7869.

8.4. 3D OTF Modulus Iso-surfaces (Anaglyphs)

This supplementary section presents a series of red-cyan anaglyphs that duplicate the figures in the main text. These require red/blue glasses to view correctly with the red filter in front of the viewer's left eye.

Unaberrated Probe

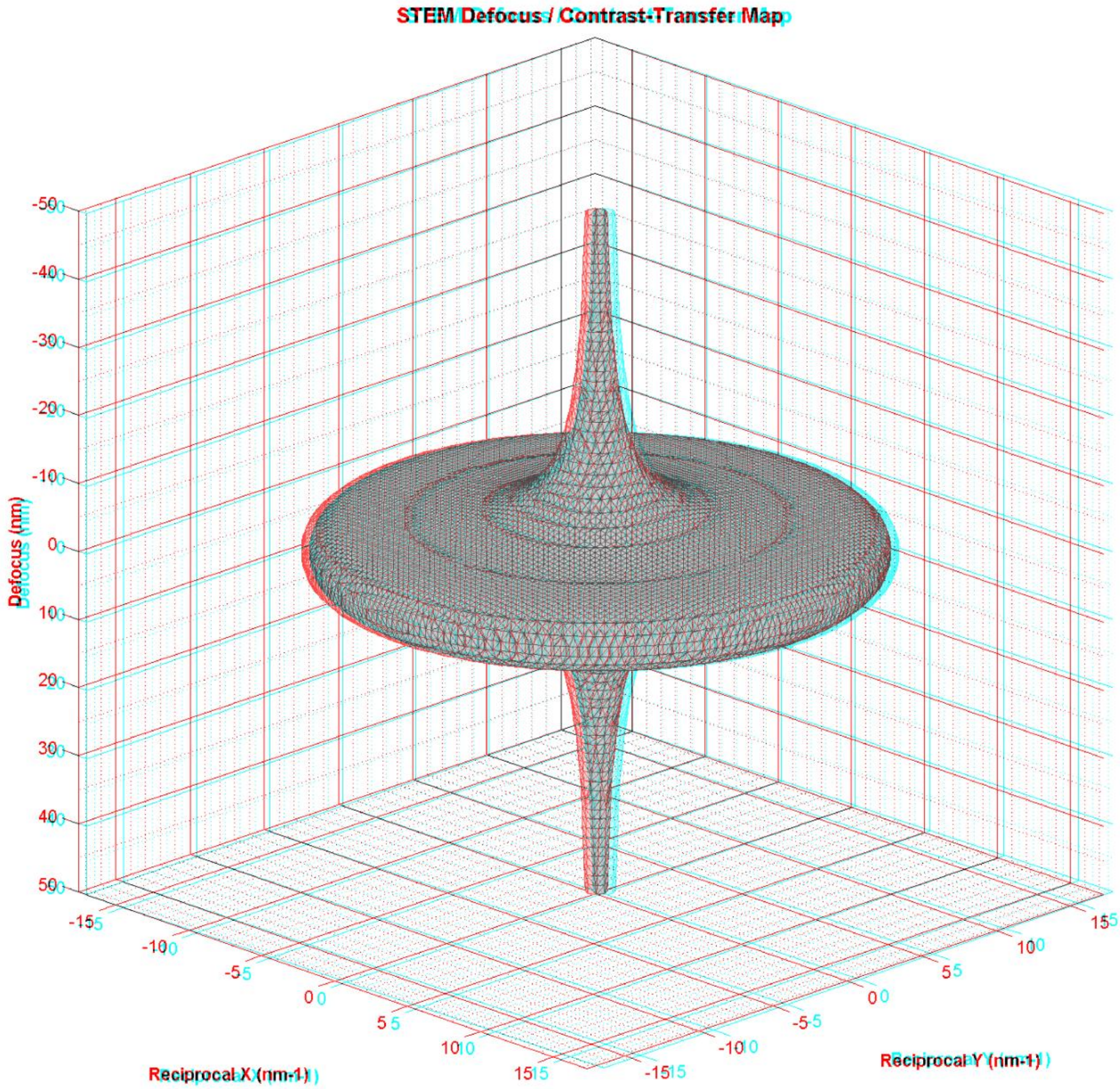


Figure 121. 3D visualisation of the spatial-frequency versus defocus optical transfer volume corresponding to Figure 4 (left). Surface shows the bounds of the 10% contrast transfer region.

3rd Order Spherical Aberration – $C_{3,0}$

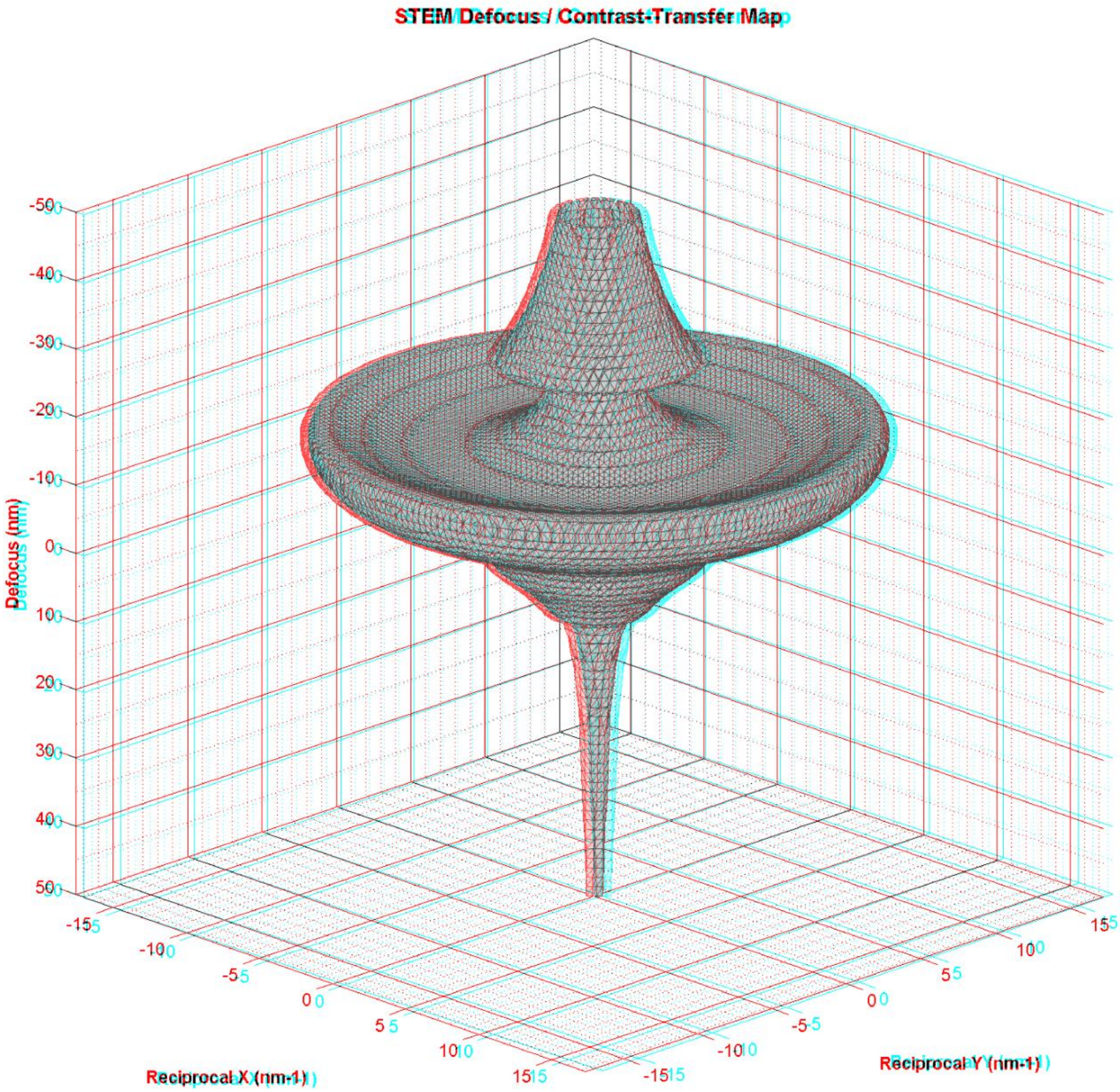


Figure 122. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +50µm $C_{3,0}$. Bounding surface shows 10% contrast level.

1^{st} Order Two-fold Astigmatism – $C_{1,2}$

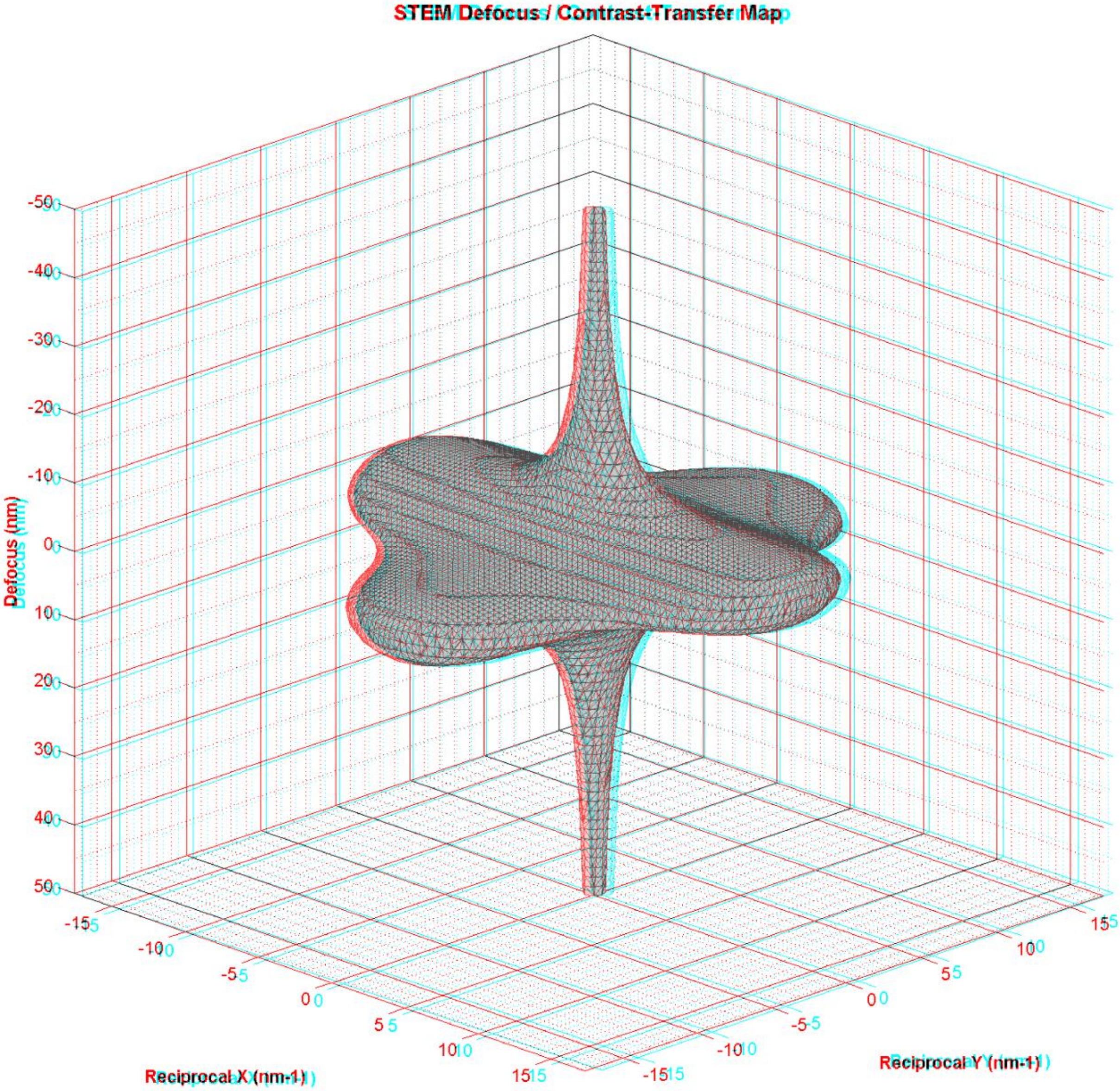


Figure 123. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +2.5nm $C_{1,2}$. Bounding surface shows 10% contrast level.

2^{nd} Order Three-fold Astigmatism – $C_{2,3}$

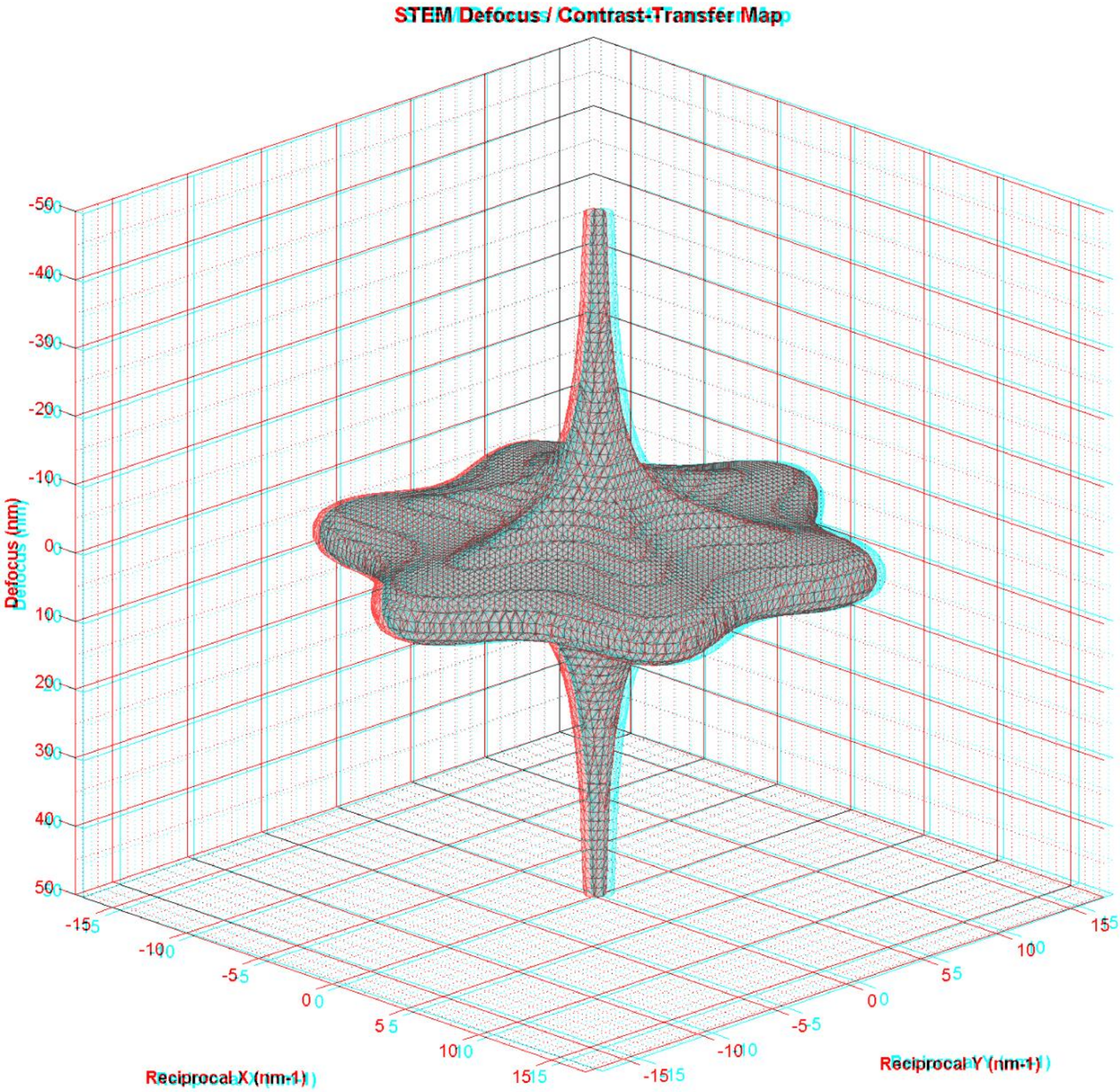


Figure 124. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +250nm $C_{2,3}$. Bounding surface shows 10% contrast level.

3rd Order Four-fold Astigmatism – $C_{3,4}$

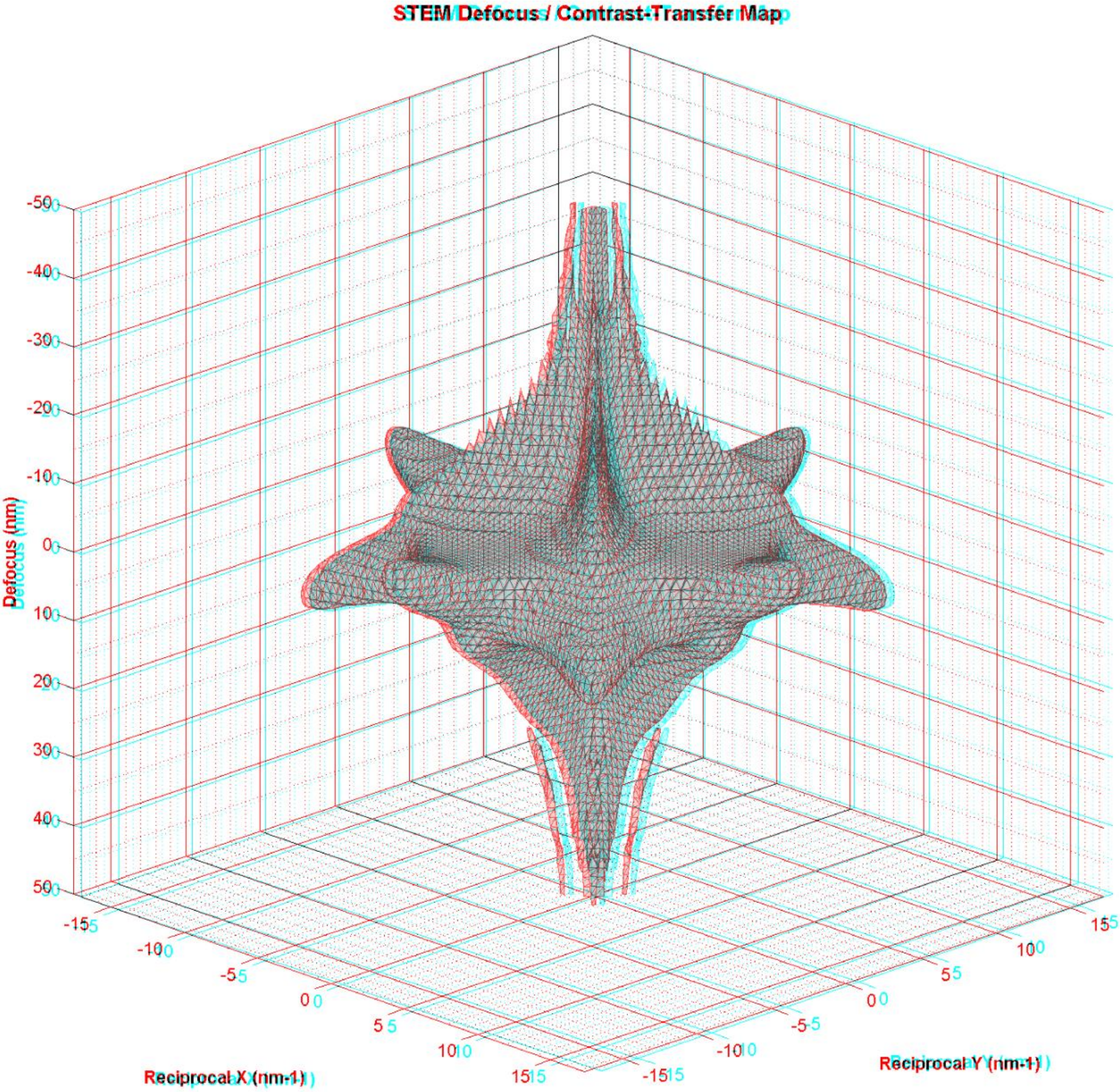


Figure 125. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +25µm $C_{3,4}$. Bounding surface shows 10% contrast level.

5^{th} Order Six-fold Astigmatism – $C_{5,6}$

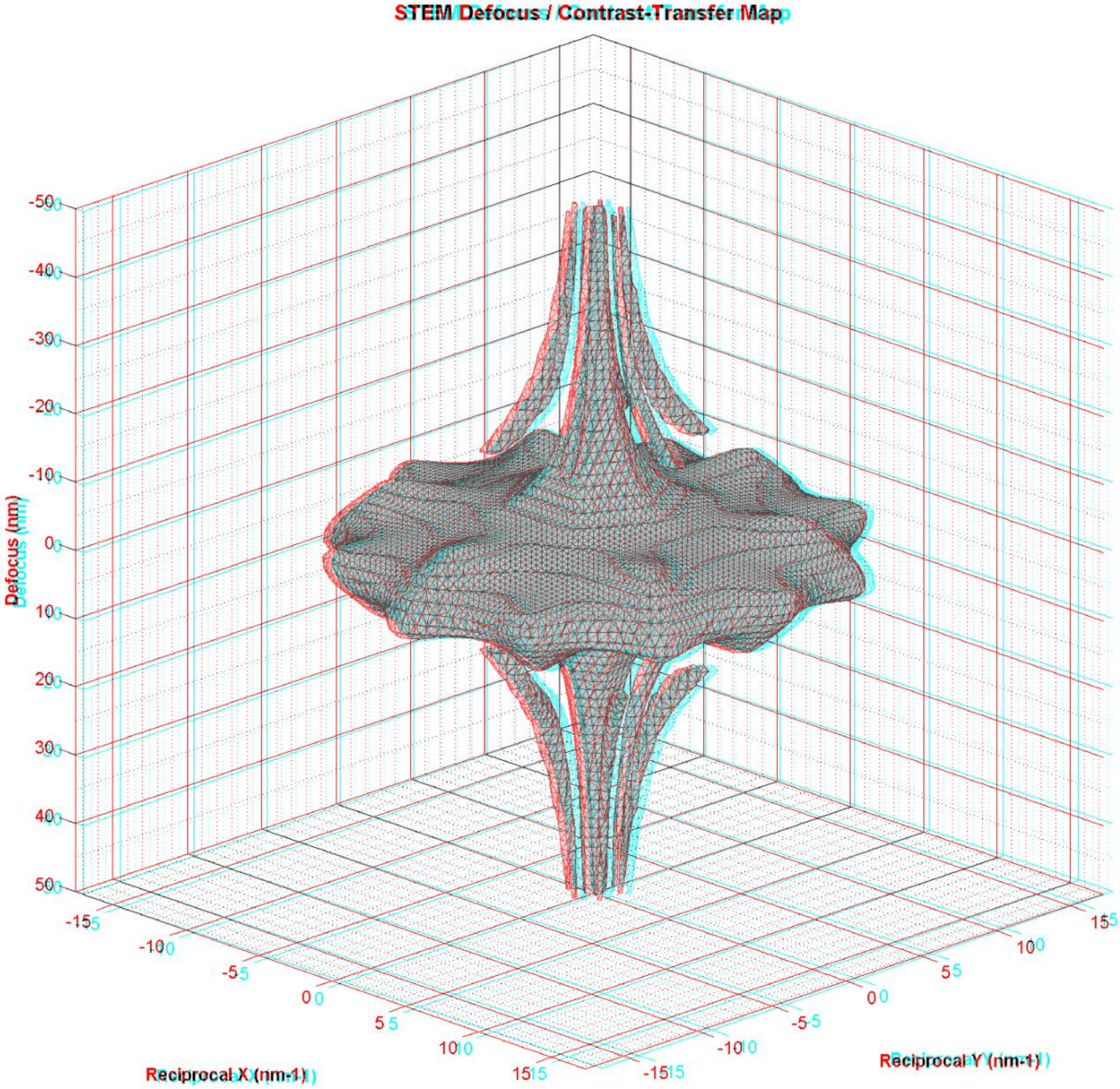


Figure 126. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +50mm $C_{5,6}$. Bounding surface shows 10% contrast level.

2^{nd} Order Axial Coma – $C_{2,1}$

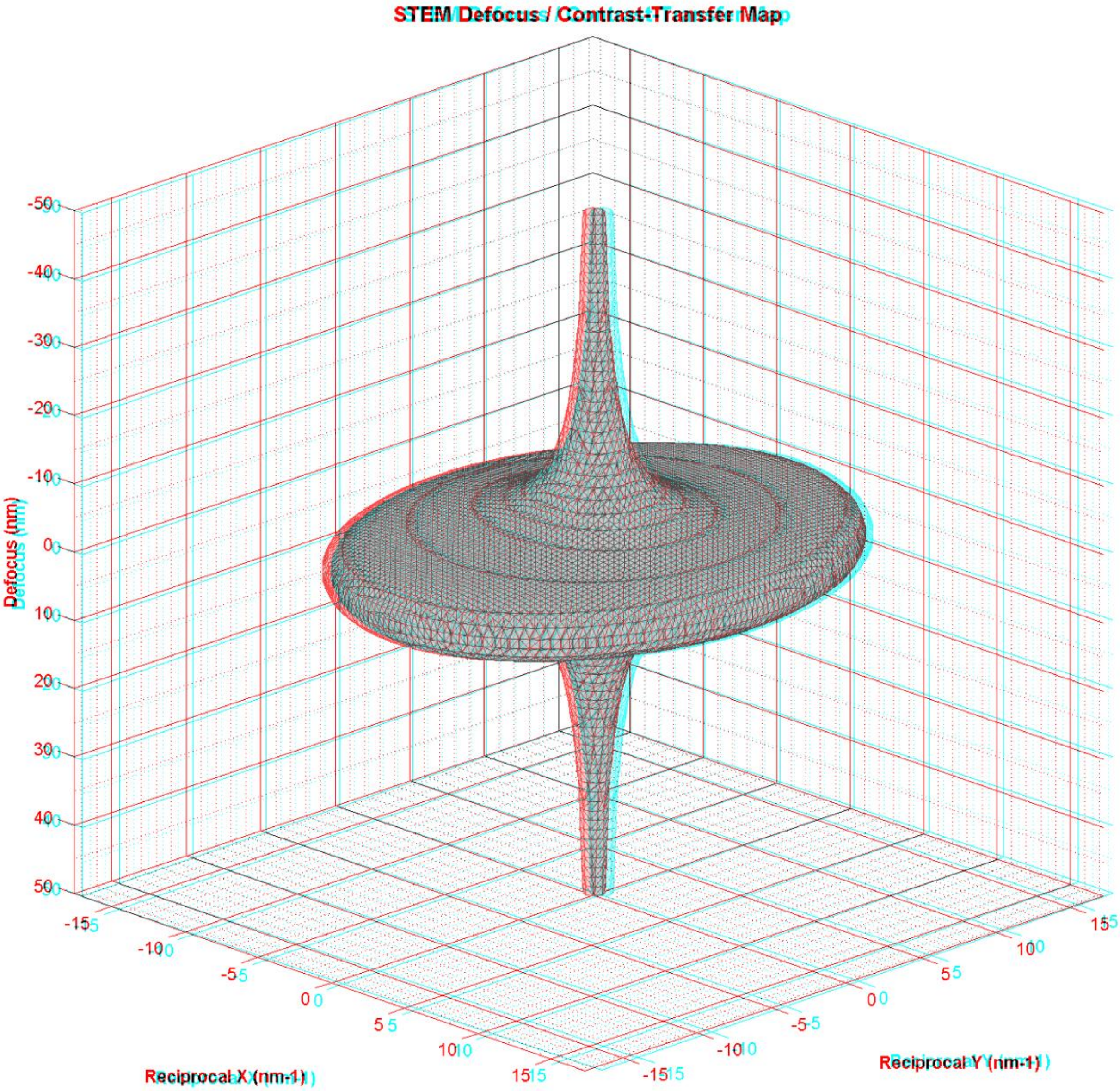


Figure 127. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +150nm $C_{2,1}$. Bounding surface shows 10% contrast level.

3^{rd} Order Two-Fold Astigmatism – $C_{3,2}$

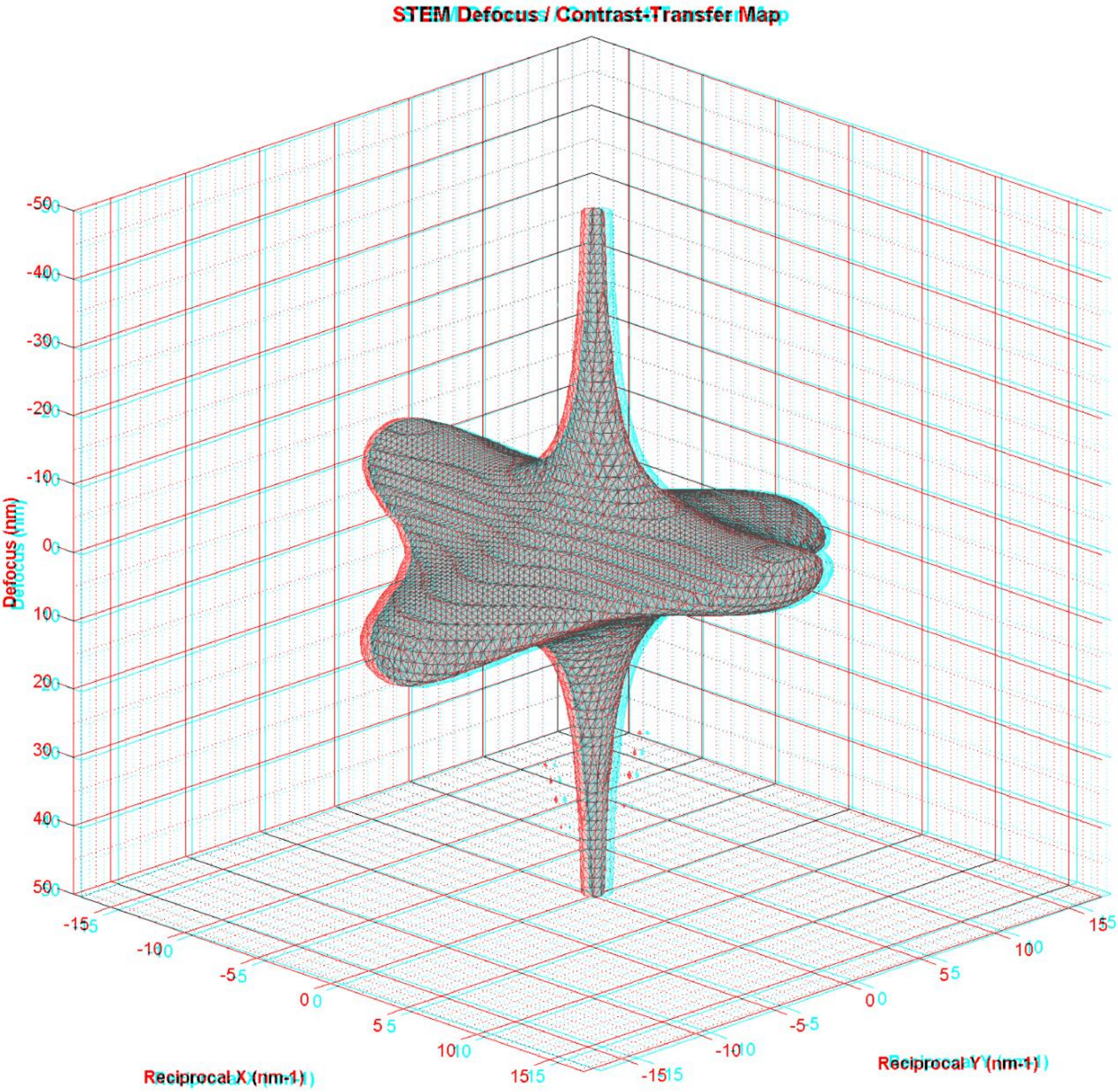


Figure 128. 3D optical-transfer versus defocus iso-surface for a simulated probe with no aberrations other than +20µm $C_{3,2}$. Bounding surface shows 10% contrast level.

A 'Real' STEM Probe with Mixed Aberrations

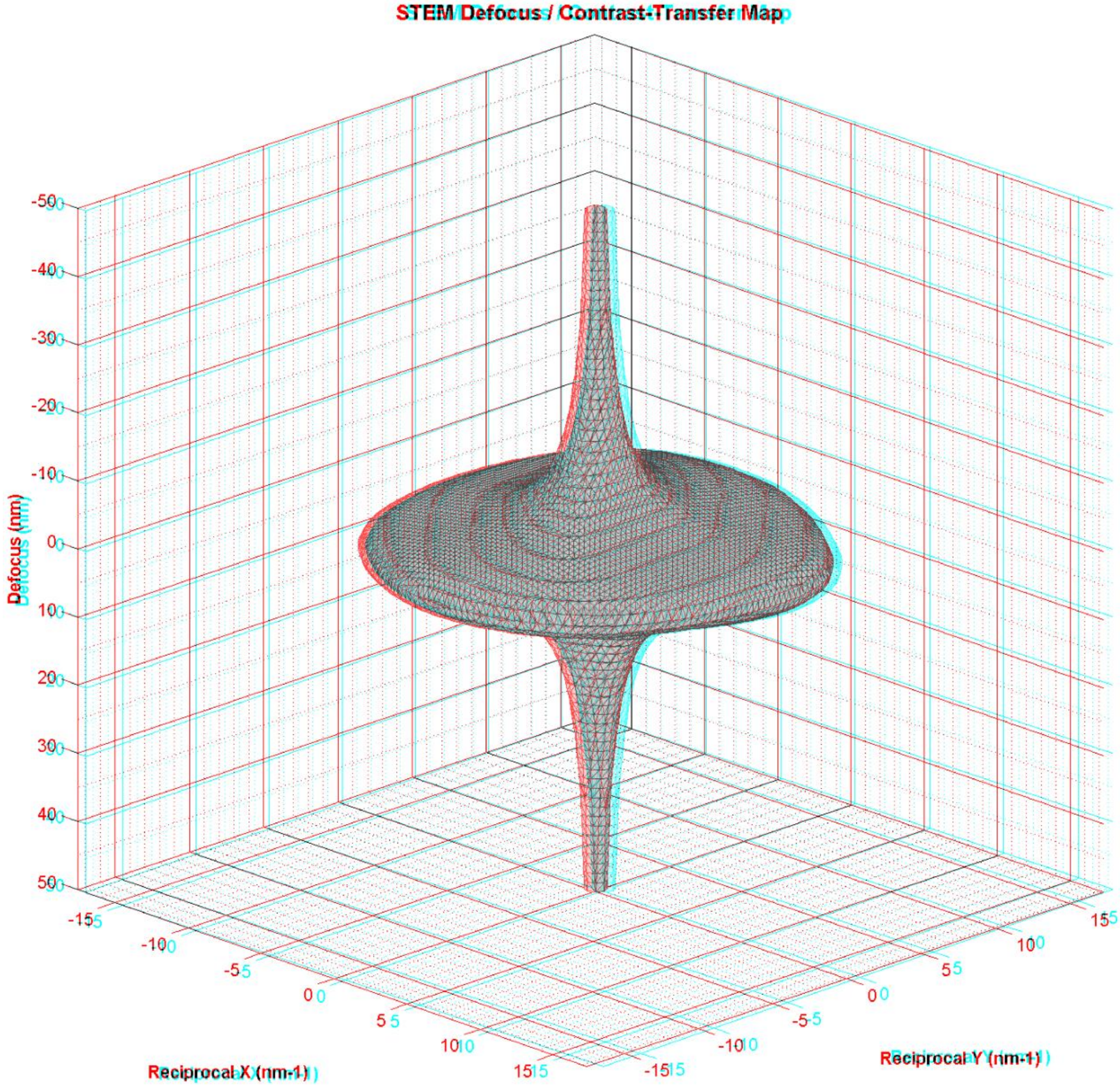


Figure 129. 3D optical-transfer versus defocus iso-surface for a simulated probe with a 'real' mixture or aberration strengths, orientations and other microscope properties. Bounding surface shows 10% contrast level.

8.5. 'Aberration Diagnosis' from Human Corneal Topography Measurements

The measurement of optical distortion will be familiar to all spectacle wearers. There the 'prescription' must be measured with precision in the same way as in electron optics. Where the manufacture of spectacles is concerned a mistake is of little consequence; however laser eye surgery is a different matter. Here a laser is used to ablate material from the front of the cornea by vaporisation to correct for imperfections in the lens curvature. Before such a procedure can be attempted precise measurements of the curvature of the eye must be made. This is done using a keratoscope, a type of non-invasive reflective interferometer used by opticians. This machine records topographic information from the cornea which after differentiation of this surface with respect to radial-angle gives a radial-curvature measurement – loosely related to the reciprocal of focal-length and the cause of poor-vision.

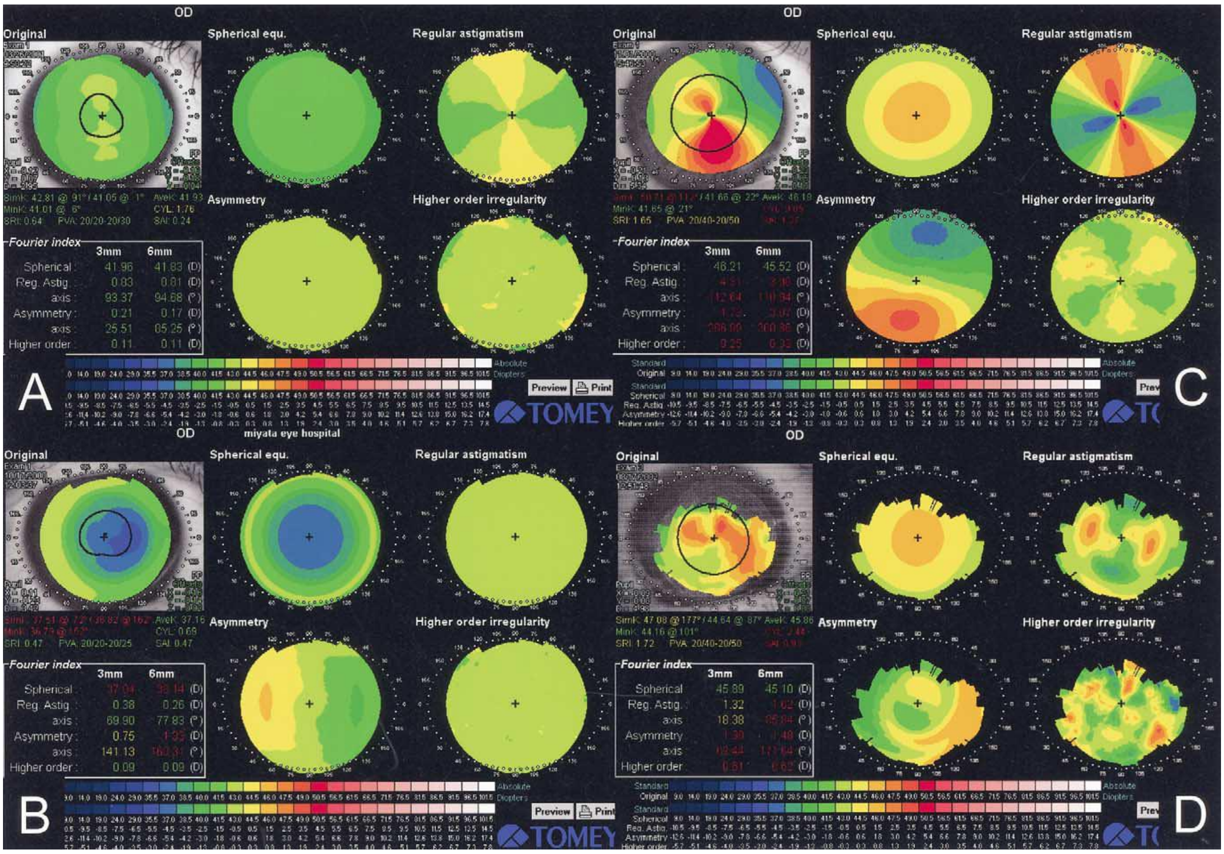


Figure 130. Examples of corneal topographic measurement. Each example shows five subpanels, top-left the black & white photograph of the eye with the raw-topographic data superimposed. The remaining four subpanels show the raw-results decomposed into four components; top-left the spherical component, top-right the two-fold astigmatism (or cylindrical prescription), lower-left the asymmetry (c.f. coma) and lower-right the remainder. The four patients A-D represent; a normal eye(A), an eye following laser-ablation surgery (B), an eye with significant irregularity as a result of keratoconus before treatment (C), and an eye treated by conventional surgical methods (D).

The raw topographic data collected by the optician is differentiated with respect to radial angle [133] to produce the radial curvature maps shown in Figure 130 [134]. This is analogous to the case in electron microscopy where the ray-order of optics is one lower than the wave order and gives the ophthalmic diagrams a very similar appearance to those produced by the STEM focal-series analysis. Compare Figure 130, patient C's regular astigmatism with Figure 90, or their asymmetry (coma) with Figure 22 b).

The optician may prescribe glasses using a conventional eye-test but from these maps the prescription (aberrations) can be directly diagnosed by parameterisation just as in the earlier discussion (see Section 4.3) [133], [135]. However we can readily see that spherical and cylindrical (astigmatic) prescriptions cannot wholly parameterise arbitrary data. The residual plot or *higher-order irregularity* for patient C (and to a lesser extent D) in Figure 130 is immediately obvious to the microscopist as a three-fold astigmatism and this is exactly mirrored in the ophthalmic literature as is third-order two-fold astigmatism [136].

8.6. Pull-out Sheet for Optimal STEM Focal-Series Data Acquisition

This pull-out sheet contains a brief summary of the optimum experimental acquisition parameters for the recording of HAADF focal-series suitable for image reconstruction.

Image Size / Magnification (pixel-sampling)

For high quality restoration at least 15 pixels per inter-atomic feature are essential; 20 or more desirable. Insufficient pixel sampling will impart a ceiling to the resolution and SNR of the restored image. For beam-sensitive materials, this may be limited by acquisition time / field-of-view considerations

Number of Unique Fourier Transform Spots

For reasonable image-restoration, 6 unique Fourier spots are required in the in-focus FT, 10 or more are desirable. Centro-symmetric conjugate spots, and the zero- or D.C. spot do not count towards this unique total.

Focal Range

The focal range should be sufficient such that the first and last images are just out of focus. This should ideally be determined in advance from OTF simulation but as a general ‘rule of thumb’ for aberration corrected instruments, 20 nm is sufficient (best-focus ± 10 nm) and 30 nm (best-focus ± 15 nm) is desirable.

Focal Steps

Focal steps should be sufficiently finely spaced such that the rapid variations in resolution either side of focus are observed by around 10 images or more. In general a focal-step of 1 nm will achieve this.

Upper Limits for Remnant Aberrations

Remnant aberrations should be tuned to better than 5 nm first-order two-fold astigmatism ($C_{1,2}$) and better than 50 µm of spherical aberration ($C_{3,0}$).

Scan Rotation

The STEM scan-rotation should be set such that no ‘cross-streaking’ from lower order FT-spots impinges upon neighbouring spots. In determining this no real-space image-windowing should be applied when calculating the Fourier transform. In general this condition is met by ensuring the scan-rotation is ≈ 10° off of a crystallographic plane.

Dwell Time

The acquisition should be line-synched, but there are no other restrictions to dwell-time.

Stage / Sample Drift

Drift should be minimised as far as practically possible. If crystal-lattice entirely fills the field of view, excessive drift may lead to crystal-unit ‘jumps’ during re-registration.

Annular Detector Angles

The inner angle selected should sufficiently large that negligible elastic scattering contributes to the recorded signal. The image should be entirely incoherent.