# Quantitative Automated Object Wave Restoration in High-Resolution Electron Microscopy
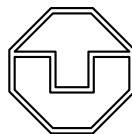
Eingereicht am: 20. Juni 2002

1. Gutachter: Prof. Dr. H. Lichte (TU Dresden)

2. Gutachter: Dr. A. I. Kirkland (University of Cambridge)

3. Gutachter: Prof. Dr. D. van Dyck (University of Antwerp - RUCA)

Verteidigt am: 25. November 2002

*To Sabine.*

*You cannot get anything for nothing, not even an observation*

Dennis Gabor, 1951

# Table of contents

# Zusammenfassung

Das zentrale in dieser Arbeit behandelte Problem ist die automatische Bestimmung der Abbildungsbedingungen im Elektronenmikroskop mit hoher Genauigkeit. Dies ermöglicht die Rekonstruktion der Objektwelle, die direkte strukturelle Information über die Probe enthält, aus einem Satz von Aufnahmen mit unterschiedlichen Abbildungsfehlern.

Nach einem Überblick über die der Bildentstehung und Objektwellenrekonstruktion zugrundeliegenden Theorie, wird der letzte Schritt in der Bildentstehung, d.h. die Aufzeichnung des Bildes mit einer CCD-Kamera, genauer betrachtet. Frühere Beschreibungen dieser Kameras beruhten oft auf der unbegründeten Annahme, daß die Modulationstransferfunktion (MTF) auch die raumfrequenzabhängige Dämpfung des Schrotrauschens der Elektronen korrekt beschreibt. Es wird daher eine neue Theorie vorgestellt, die zwischen der Übertragung von Signal und Rauschen unterscheidet. Sie ermöglicht die Berechnung beider Eigenschaften mit Hilfe eines detaillierten Monte-Carlo Simulationsmodells für die Streuung von Elektronen und Photonen im Szintillator der Kamera. Desweiteren werden Methoden zur genauen experimentellen Bestimmung der für Signal und Rauschen gültigen Übertragungsfunktionen vorgestellt. In Übereinstimmung mit den Monte-Carlo Simulationen zeigen Meßergebnisse für kommerzielle CCD-Kameras, daß die Signalübertragung deutlich schlechter ist als die des Rauschens.

Anschließend werden, nach einer Darstellung bekannter Methoden zur Aberrationsbestimmung, neue Verfahren vorgestellt, um einerseits die Bildfehlerdifferenzen in einem Satz von Aufnahmen und andererseits den absoluten Wert der symmetrischen Aberrationskoeffizienten aus der rekonstruierten Bildwelle zu bestimmen.

Bei beiden Verfahren wird lediglich die Phaseninformation im Frequenzraum genutzt, so daß alle Fourierkomponenten mit einer von ihrer Stärke unabhängigen Gewichtung in das Ergebnis eingehen. Dadurch ist die neue Methode im Gegensatz zu bekannten Verfahren wie automatischer Diffraktogrammanalyse auch für Proben anwendbar, die weitgehend kristallin sind und nur wenig amorphe Kontamination zeigen. Die Methode wird im weiteren auf die Bestimmung der asymmetrischen Aberrationen mittels kombinierter Strahlverkippungs- und Fokus-Serien ausgeweitet.

Die Anwendbarkeit der neuen Methode wird anhand von Objektwellenrekonstruktionen aus Kipp- und Fokus-Serien für Proben von komplexen anorganischen Oxyden sowie Proben von mit eindimensionalen anorganischen Kristallen gefüllten Kohlenstoff Nanoröhrchen demonstriert. Die letzteren Proben ermöglichten dabei erstmals einen direkten Vergleich der Phasenschiebung in der rekonstruierten Objektwelle einer Probe von *exakt* bekannter Dicke mit dem durch Simulationen vorhergesagten Wert.

Ein praktisches Problem für die Anwendung der neuen Aberrationsbestimmungsmethode zur Korrektur der Bildfehler oder sogar zur Rekonstruktion der Objektwelle in Echtzeit ist die langsame Kommunikation heutiger Mikroskope mit externen Computern. Zur Abhilfe wurde im Rahmen dieser Arbeit DigiTEM entwickelt: ein Microcontroller-Interface, das direkt in den digitalen Bus von dem im Mikroskop eingebauten Microcomputer zu den für die Einstellung der Spulenströme zuständigen Digital/Analog-Wandlern eingreift. Dies ermöglicht einem externen Computer, teilweise oder vollständig die Kontrolle über das Mikroskop zu übernehmen. Außerdem wird der Computer in Echtzeit über alle Justierungen informiert, die der Anwender am Mikroskop vornimmt.

Im abschließenden Ausblick wird diskutiert, wie die in dieser Arbeit beschriebenen Entwicklungen zu einem System zur Rekonstruktion, Fehlerkorrektur und Anzeige der Objektwelle in Echtzeit erweitert werden könnten.

# Abstract

The main problem addressed by this dissertation is the accurate and automated determination of electron microscope imaging conditions. This enables the restoration of the object wave, which confers direct structural information about the specimen, from sets of differently aberrated images.

Following a review of the theory of image formation and object wave restoration, the last step in image formation, *i.e.* the recording of the image with a CCD camera, is discussed in more detail. Previous characterisations of these cameras often relied on the unjustified assumption that the Modulation Transfer Function (MTF) also correctly describes the spatial frequency dependent attenuation of the electron shot noise. A new theory is therefore presented that distinguishes between signal and noise transfer. This facilitates the evaluation of both properties using a detailed Monte-Carlo simulation model for the electron and photon scattering in the scintillator of the camera. Furthermore, methods for the accurate experimental determination of the signal and noise transfer functions are presented. In agreement with the Monte-Carlo simulations, experimental results for commercially available CCD cameras show that the signal transfer is significantly poorer than the noise transfer.

Following an account of existing methods of aberration determination in high-resolution TEM, new methods for determining the relative aberrations in a set of images and the absolute symmetric aberrations in the restored wave are presented. Both are based on the analysis of the phase information in the Fourier domain and give each Fourier component a weight independent of its strength. This makes the method suitable even for largely crystalline samples with little amorphous contamination,

where conventional methods, such as automated diffractogram fitting, usually fail. The method is then extended to also cover the antisymmetric aberrations, using combined beam tilt and focal series.

The applicability of the new method is demonstrated with object wave restorations from tilt and focal series of complex inorganic block oxides and of carbon nanotubes filled with one-dimensional inorganic crystals. The latter specimens allowed for the first time a direct comparison between the phase shift in the restored object wave of a specimen with *precisely* known thickness and the value predicted by simulations.

A practical problem for the application of the new aberration determination method for the correction of aberration or even restoration of the object wave in real time is the slow communication of present generation microscopes with external computers. As part of this project, DigiTEM, a microcontroller interface that directly connects into the digital bus from the built-in microscope computer to the digital/analogue converters that control the coil currents, has been developed to alleviate this problem.

In the concluding chapter, the prospective applicability of the methods developed in this work to a system that allows the real-time restoration, aberration correction and display of the object wave is discussed.

# Acknowledgements

I am deeply indebted to my supervisors, Prof. Hannes Lichte (Dresden), for his genuine interest in my research; Dr. Angus Kirkland (Cambridge), who took the bulk of the supervising duties and always had an open ear for any problem and whose efficient microscopy provided most of the images with which the methods developed in this thesis were developed and demonstrated; and Dr Owen Saxton (Cambridge)whose mathematical rigour and precision in language has always been an inspiration for me.

I would also like to thank all present and rescent past members of the HREM group for making the working environment as stimulating and enjoyable as it was. Optimising wavefront reconstruction wouldn't have been half as interesting without fascinating samples to look at. I am deeply indepted to the carbon nanotechnology group in Oxford, expecially Jeremy Sloan and Steffi Friedrichs, for the fruitful collaboration on filled carbon nanotubes. Furthermore, I would like to express my thanks to Keith Page for his help with DigiTEM and for getting me started on microcontrollers.

Finally and foremost I would like to thank Sabine for her love and encouragement.

# Introduction

Atomic resolution is now routinely achieved in electron microscopy (Smith, 1997).
However, the contrast observed in a single electron micrograph, which is in fact an
interference pattern formed by the superposition of the beams diffracted from the
sample, is difficult to interpret. This is firstly due to the very large aberrations
of electron lenses compared to optical lenses. A particular limiting factor is the
*spherical aberration* which cannot be avoided for round lenses with static fields that
have no charges in the beam path (Scherzer, 1936). This causes contrast reversals for
image detail finer than the *point resolution*. Recently, a corrector for this spherical
aberration based on hexapole lenses has successfully been implemented (Haider et al.,
1998; Urban et al., 1999), however, this corrector is still very expensive and difficult
to use. The direct interpretation of high resolution images is furthermore impeded
by the fact that the image intensity records solely the modulus squared $|\Psi|^2$ of the
aberrated electron wave (image wave) and the phase information in the image wave is
lost. Due to this missing information, it is not possible to correct this single image for
the degrading effects of the lens aberrations and recover the full specimen information.
This phase problem can be solved in various ways. Gabor (1948) suggested recording
an interference pattern between the image wave and a reference wave that is coherent
with the image wave. As a phase shift in the object wave leads to a shift of the
interference fringes, the thus recorded *hologram* contains the full amplitude and phase
information of the image wave, and the object wave can reconstructed by light-optical
means, as in light optics, correctors for any lens aberration are readily available.

Gabor's original idea of electron holography has taken a long route to fruition,

mainly because electron sources were, for a long time, not sufficiently coherent to produce good interference patterns. Only in the last two decades has electron holography begun to be established as a useful technique for measuring, for instance, the phase shift due to magnetic fields (Tonomura, 1992) or for restoring the object wave at atomic resolution (Lichte, 1991; Orchowski et al., 1995; Lehmann et al., 1999).

Another solution to the phase problem is 'ptychography', where a focused convergent probe is scanned across the specimen and the interference between overlapping disks in the diffraction pattern is used to derive the relative phases of the diffracted beams (Rodenburg and Bates, 1992; Nellist et al., 1995).

The most straightforward way to solve the phase problem in imaging mode, however, is to combine the complementary information available in images taken at different imaging conditions to yield a restored wave as it exited from the specimen. This was first suggested for series of images at different foci in 1968 by Schiske (1968), but computing technology has only recently improved to a level where the fast restoration of the object wave from these image series is possible. Subsequent improvements described in more detail in section 1.6 of this thesis led to successful restorations from both focal series and tilt-azimuth series (Kirkland et al., 1997).

However, the successful object wave restoration from a series of images requires that the aberrations in each of the contributing images are known accurately. This is because firstly, the starting conditions at which the first image is taken are generally unknown and secondly because the actual changes in imaging conditions often do not accurately reflect the expected ones due to drift or hysteresis in the lens or deflector coils.

Furthermore, the whole process of image formation needs to be characterised accurately, especially if quantitative information is to be extracted from the restoration.

Finally, the method will only find widespread use in practice if both the acquisition and restoration are fully automated, with the ultimate goal being the real-time display of the fully corrected restored wavefunction.

In this dissertation some mileage towards this goal is covered, especially in the

formulation of a new technique for aberration determination that can be robustly automated and works reliably for nearly arbitrary specimens. Aided by the ever increasing computing power available at reasonable cost and the progress made in image detection devices, this means that the realisation of real time object wave restoration is probably only a few years away.

The dissertation is organised as follows. Chapter 1 gives an overview of the theory of image formation and object wave restoration from sets of images. Most of this theory has been understood for a long time, starting with the work of (Glaser, 1943) and (Scherzer, 1949) who were the first to treat the imaging process within a wave optic framework. The last step of the image formation chain, *i.e.* the actual recording of the image with a CCD-camera (or other electron detector) was found to be less well described in the literature, as most authors based their measurements on the unjustified assumption that shot noise and information in the image are equally affected by the Modulation Transfer Function (MTF) of the camera. Therefore, extensive work on the characterisation of CCD cameras was carried out, leading to a theoretical description that introduces a separate Noise Transfer Function (NTF) and allows the calculation of both, and thereby the spatial frequency dependent Detection Quantum Efficiency (DQE) from a Monte-Carlo model of the electron and photon scattering in the scintillator of the camera. This is summarised in chapter 2.

In chapter 3, established methods for aberration determination are reviewed, in particular diffractogram-based methods and the Tilt Induced Displacement method. A new method for accurate automated determination of the symmetric aberrations based on the analysis of phase information in the image Fourier transforms is presented in chapter 4 and additionally extended to cover the antisymmetric aberrations in chapter 5.

This new method for automated object wave restoration is demonstrated in chapter 6 with specimens including complex block oxides and carbon nanotubes filled with inorganic nano-crystals.

A practical problem that needs to be addressed when the determined aberrations

are to be corrected in real time on the microscope is the communication between the microscope and an external computer. The manufacturer of most of our microscopes (JEOL) has made provisions for this in form of a serial port to the microscope computer by which an external computer can issue commands to the microscope. However, communication via this port is very slow, with a maximum of about 10 commands per second. Furthermore, the communication is simplex and the external computer gets no notification when the operator has changed any microscope settings. A solution, presented in chapter 7, involves a small microcontroller based circuit board, called DigiTEM, which has been developed and programmed.

Finally, conclusions are drawn and suggestions for further work are given in chapter 8.

# Chapter 1

# Theory of image formation and wavefront restoration

## 1.1 The wave aberration function

In an ideal lens, a point object at position $(x, y)$ in the object plane leads to a spherical wavefront in the diffraction plane, contracting to the image point in the image plane. However, all real lenses and in particular electron lenses however suffer from aberrations causing deviations from this ideal spherical wavefront hence reducing the sharpness of the point object image much more severely than the diffraction limit alone. The *wave aberration function,* $W(x, y, u, v)$ describes the distance between the ideal and the actual wavefront in the diffraction plane as a function of both the position $(x, y)$ of the point object in the image plane and of the scattering angle $(u, v)$, *i.e.* the position in the diffraction plane. At high resolution, only a very small part of the specimen is in the field of view, hence the dependence of $W$ on $(x, y)$ is small and usually neglected (*isoplanatic* approximation). The notation in this chapter follows that given by Typke and Dierksen (1995), where a more detailed account of wave aberrations can be found.

The wave aberration function can be expanded as a power series in $x$, $y$, $u$ and $v$ about the origin of zero scattering angle. For this purpose it is convenient to

| Aberrat-ion | Order in $\omega$ | Order in $w$ | Azimuthal symmetry | Name and description |
|:---:|:---:|:---:|:---:|:---|
| $A_0$ | 1 | 0 | 1 | Image shift |
| $A_1$ | 2 | 0 | 2 | Two-fold astigmatism |
| $C_1$ | 2 | 0 | $\infty$ | Defocus, overfocus positive |
| $A_2$ | 3 | 0 | 3 | Three-fold astigmatism |
| $B_2$ | 3 | 0 | 1 | Axial coma |
| $A_3$ | 4 | 0 | 4 | Four-fold astigmatism |
| $B_3$ | 4 | 0 | 2 | Axial star aberration |
| $C_3$ | 4 | 0 | $\infty$ | Spherical aberration |
| $B_{31}$ | 3 | 1 | 1 | Seidelian coma |
| $A_{0c}$ | 1 | 0 | 1 | Dispersion $^a$ |
| $C_{1c}$ | 2 | 0 | $\infty$ | Chromatic aberration $^b$ |
| $A_{11c}$ | 1 | 1 | 1 | Chromatic magnification change |

$^a$Change of image displacement with electron energy.
$^b$Change of focus with electron energy.

Table 1.1: The aberration coefficients used in equation 1.1.

describe the scattering angle with a complex variable[1] $\omega = (u + iv)$ and the object plane position with a complex variable $w$. Table 1.1 gives a comprehensive list of the aberration coefficients important for high-resolution electron microscopy.[2] In terms of these coefficients, the expanded wave aberration function, $W_c(\omega)$ is given as:

---

[1] It is worth noting that this is not a power series in $\omega$ in the sense of complex analysis, as it contains terms in $\omega^*$ as well as terms in $\omega$. The main advantage of the complex notation is that it allows a simple transition between cartesian and polar form and that each aberration coefficient uniquely corresponds to a particular order in $\omega$ and a particular azimuthal symmetry.

[2] The seemingly counter-intuitive notation (*e.g.* $C_1$ for a **second** order term in $W(\omega)$) stems from the ray-optical theory of Seidel aberrations, which are described in terms of displacements of ray-path intersections with the image plane as a function of $(x, y)$ and $(u, v)$. These displacements are proportional to the *gradient* of the wave aberration function, hence an $n^{th}$ order Seidel aberration will correspond to a term of order $n + 1$ in the wave aberration function.

$$W_c(\omega) = \Re\left\{A_0\omega^* + \frac{1}{2}A_1\omega^{*2} + \frac{1}{2}C_1\omega^*\omega\right.$$
$$+ \frac{1}{3}A_2\omega^{*3} + \frac{1}{3}B_2\omega^{*2}\omega$$
$$+ \frac{1}{4}A_3\omega^{*4} + \frac{1}{4}B_3\omega^{*3}\omega + \frac{1}{4}C_3\omega^{*2}\omega^2 + \frac{1}{3}B_{31}\omega^{*2}\omega w$$
$$\left. + \left[A_{0c}\omega^* + \frac{1}{2}C_{1c}\omega^*\omega + A_{11c}\omega^* w\right]\frac{\Delta E}{E_0}\right\}. \tag{1.1}$$

The terms in the last line of this equation are chromatic aberrations as they depend on the variation $\Delta E$ of the electron energy $E_0$. Terms with even order in $\omega$ are referred to as 'even' or 'symmetric' aberrations, and the coefficients of the odd order terms are called 'odd' or 'antisymmetric' aberrations. The spherical and chromatic aberrations are sometimes referred to as "unavoidable aberrations" since, according to the Scherzer theorem (Scherzer, 1936), their coefficients[3] $C_3$ and $C_{1c}$ are always positive for round lenses with static fields and no charges in the beam path. The defocus $C_1$, adjusted by changing the objective lens current, is a free parameter that can be used to control the transfer of object information.

All the other aberrations listed in table 1.1 are due to lens imperfections and would not appear in a perfect round lens. They can be corrected by aligning the instrument using stigmators and beam deflectors. The importance of these aberrations generally decreases with increasing order in $k$, and the high-order aberrations are only important for the highest resolutions. The effect of the two-fold astigmatism $A_2$ is a directional dependence of the focal length with maximum and minimum focal lengths reached along the two orthogonal principal axes of the two-fold astigmatism. This aberration can be corrected by a rotatable magnetic quadrupole field, which is realised by two stationary quadrupoles at a mutual angle of 45°. The importance of the axial coma $B_2$ was first realised by Zemlin et al. (1978), and as discussed in section 1.2, axial coma can be avoided by a suitable choice of the illumination direction. The effect of the three-fold astigmatism went unnoticed until the early

---

[3]Often, the alternative notations $C_s = C_3$ and $C_c = C_{1c}$ are used for the spherical and chromatic aberration coefficients.

1990s, (Krivanek and Leber, 1993; Saxton et al., 1994; Chand, 1997), when it was shown to lead to serious image degradations at resolutions approaching 0.1 nm. The three-fold astigmatism can be corrected using the standard quadrupole stigmators, provided they are re-wired to provide independent control of the current to each pole (Overwijk et al., 1997).

The presence of the fourth order aberrations $A_3$ and $B_3$ has never been observed, hence they are ignored in the further discussion. In addition, I will concentrate on the non-chromatic isoplanatic aberrations, using a simplified form of the $W(\omega)$:

$$
\begin{aligned}
W(\omega) = \Re\Big\{ & A_0\omega^* + \frac{1}{2}A_1\omega^{*2} + \frac{1}{2}C_1\omega^*\omega \\
& + \frac{1}{3}A_2\omega^{*3} + \frac{1}{3}B_2\omega^{*2}\omega \\
& + \frac{1}{4}C_3\omega^{*2}\omega^2 \Big\} .
\end{aligned}
\tag{1.2}
$$

In polar notation, with $\omega = ke^{i\phi}$, $A_n = |A_n|e^{i\alpha_n}$ and $B_n = |B_n|e^{i\beta_n}$, this is expressed as

$$
\begin{aligned}
W(k,\phi) = & |A_0|\, k\cos(\phi - \alpha_0) \\
& + \frac{1}{2}|A_1|\, k^2\cos 2(\phi - \alpha_1) + \frac{1}{2}C_1 k^2 \\
& + \frac{1}{3}|A_2|\, k^3\cos 3(\phi - \alpha_3) + \frac{1}{3}|B_2|\, k^3\cos(\phi - \beta_2) \\
& + \frac{1}{4}C_3 k^4 \, ,
\end{aligned}
\tag{1.3}
$$

which makes the azimuthal dependence of the various aberration terms more obvious.

For the restoration of the object wave from a tilt or focal series of images, an accurate knowledge of the aberration coefficients is paramount. Table 1.2 lists the necessary accuracies for a range of target resolutions, based on the criterion that the maximum RMS wave aberration due to each of the aberrations is less than $\lambda/16$, corresponding to a phase shift by $\pi/8$ for frequencies lower than the target resolution. For symmetric aberrations this leads to a transfer of $\sin^2 \pi/8 = 14\%$ of the object

9

| Resolution | | Accuracy | | | |
|---|---|---|---|---|---|
| $k_{max}$ [nm$^{-1}$] | $d_{\min}$ [nm] | $A_1$, $C_1$ [nm] | $A_2$, $B_2$ [nm] | $A_3$, $B_3$, $C_3$ [$\mu$m] | Tilt [$\mu$rad] |
| 11 | 0.09 | 0.5 | 35 | 2.1 | 21 |
| 10 | 0.1 | 0.6 | 47 | 3.1 | 27 |
| 9 | 0.11 | 0.8 | 64 | 4.8 | 38 |
| 8 | 0.125 | 1.0 | 92 | 7.6 | 54 |
| 7 | 0.14 | 1.3 | 137 | 13 | 80 |

Table 1.2: Accuracy to which the aberration coefficients need to be determined such that each of them causes a maximum RMS wave aberration of less than $\lambda/16$, *i.e.* a phase change of less than $\pi/8$ within the target resolution. The values are calculated for an accelerating voltage of 300kV ($\lambda \approx 2$pm). The necessary accuracy for the beam tilt $\tau$ is calculated from that for $B_2$ using $\Delta\tau = \Delta B_2/(3C_3)$ with $C_3 = 0.57$mm.

power of a pure weak phase object into the amplitude image, whereas for antisymmetric aberrations such a phase shift leads to a lateral contrast displacement by at most $d_{\min}/8$.

It should be noted that in practice the accuracy requirement for $C_3$ is less than listed in the table as the residual phase shift due to the an error in $C_3$ is partially compensated by a bias in the determined value of $C_1$ (Steinecker and Mader, 2000). This implies that the maximum deviation of the wave aberration function for $0 \leq k \leq k_{\max}$ can be reduced by a factor of $3 - 2\sqrt{2} = 17.2\%$,[4] which means that for $C_3$, an accuracy five times poorer than that listed in table 1.2 is sufficient, provided that $C_1$ is determined such that the maximum phase shift due to the residual aberrations in the range of used spatial frequencies is minimized.

---

[4] For $\Delta C_1 = 0$ the maximum error in the wave aberration due to $\Delta C_3$ alone is $\Delta\gamma = \pi/2\,\Delta C_3\lambda^3 k_{\max}^4$. For $\Delta C_1 < 0$, the residual wave aberration function $\Delta\gamma(k)$ has a minimum at $k_1 = \sqrt{-\Delta C_1/\Delta C_3}/\lambda$, and choosing $\Delta C_1 = (1 - \sqrt{2})\lambda^2 k_{\max}^2\Delta C_3$ yields $\Delta\gamma = -\Delta\gamma(k_1) = \Delta\gamma(k_{\max}) = \pi/2\,(3 - 2\sqrt{2})\Delta C_3\lambda^3 k_{\max}^4$.

## 1.2   Tilting the illumination

When the illuminating beam is tilted by a tilt angle $\tau = t_x + it_y$, the origin of the diffraction plane is shifted with respect to the wave aberration function, which can be expressed as power series expanded about the new origin by

$$
\begin{aligned}
W'(\omega) &= W(\omega + \tau) - W(\tau) \\
&= \Re\Big[ A_0'\omega^* + \frac{1}{2}A_1'\omega^{*2} + \frac{1}{2}C_1'\omega^*\omega \\
&\quad + \frac{1}{3}A_2'\omega^{*3} + \frac{1}{3}B_2'\omega^{*2}\omega \\
&\quad + \frac{1}{4}C_3'\omega^{*2}\omega^2 \Big] ,
\end{aligned}
\tag{1.4}
$$

where the new aberration coefficients are given by:

$$
\begin{aligned}
A_0' &= A_0 + A_1\tau^* + C_1\tau + A_2\tau^{*2} + \frac{1}{3}B_2^*\tau^2 + \frac{2}{3}B_2\tau^*\tau + C_3\tau^*\tau^2 \\
A_1' &= A_1 + 2A_2\tau^* + \frac{2}{3}B_2\tau + C_3\tau^2 \\
C_1' &= C_1 + \Re(\frac{4}{3}B_2\tau^*) + 2C_3\tau^*\tau \\
A_2' &= A_2 \\
B_2' &= B_2 + 3C_3\tau \\
C_3' &= C_3 .
\end{aligned}
\tag{1.5}
$$

$$
\tag{1.6}
$$

It is immediately apparent from the first line of this equation that the shift $A_0' - A_0$ between two images taken at beam tilts differing by $\tau$ depends on all other aberration coefficients. Hence, measuring image shifts induced by a suitable set of beam tilts provides a method for measuring all aberration coefficients (section 3.2). Furthermore, tilting the beam changes the values of $C_1$ and $A_1$ by an amount that depends on $A_2$, $B_2$ and $C_3$. This is exploited in the tilt-induced diffractogram method (section 3.3.2) and also in the new method described in chapter 5 to determine the antisymmetric aberrations.

Equation 1.5 shows that tilting the beam by $\tau_c = -B_2/(3C_3)$ makes the axial coma $B_2'$ zero. This new axis is called the coma-free axis and the beam direction should always be aligned along to this axis for high-resolution microscopy.

In principle, it is also possible to compensate the axial coma with off-axial (Seidelian) coma ($B_{13}$ in equation 1.1), *i.e.* centring the coma-free point of the object plane for the given beam direction using the image shift coils. It is, however, even better to use both $\tau$ and $w$ to compensate the dispersion (table 1.1) as well as the axial coma (Zemlin, 1979; Kunath et al., 1987). The dispersion and coma free axis, defined in direction and position by

$$\tau = -\frac{B_2 A_{11c} - B_{31} A_{0c}}{3C_3 A_{11c} - B_{31} C_{1c}} \tag{1.7}$$

$$w = -\frac{3C_3 A_{0c} - B_{31} C_{1c}}{3C_3 A_{11c} - B_{31} C_{1c}} \tag{1.8}$$

can be designated as "the optical axis of the microscope". However, finding this axis is a cumbersome iterative process. For a given image position, however, the dispersion-free axis is found easily using the standard *voltage centre* alignment procedure. In the microscopes used in this work the coma-free axis was always within less than 1 mrad of the voltage centre, therefore for most of the image series presented here the microscope was aligned to the voltage centre axis and remaining axial coma was corrected for *a posteriori* in the object wave restoration procedure. When images are taken without the possibility of any *a posteriori* aberration correction, the coma-free axis is preferred, as discussed in section 1.5.2.

In the following, the aberrations are referred to with respect to the coma-free axis and it is assumed that coma-free and dispersion free axes coincide. Additionally, the wave aberration function is given in the dimensionless form

$$\gamma(\mathbf{k}) = \frac{2\pi}{\lambda} W(k_x \lambda, k_y \lambda) \, , \tag{1.9}$$

expressing the spatial frequency dependent phase shift of the aberrated wave.

With this wave aberration the intensity observed from a specimen with exit wave $\psi(\mathbf{x})$ under coherent illumination with beam tilt $\lambda \mathbf{k}_0$ can be expressed as

$$C(\mathbf{x}) = \left| \mathrm{FT}^{-1} \left[ \mathrm{FT}[\psi(\mathbf{x})] e^{-i\gamma(\mathbf{k}+\mathbf{k}_0)} \right] \right|^2 \, . \tag{1.10}$$

Using the Wiener-Khinchin theorem, the Fourier transform of the image intensity is

$$c(\mathbf{k}) = \int d\mathbf{k}' \left[ \psi^*(\mathbf{k}') e^{i\gamma(\mathbf{k}'+\mathbf{k}_0)} \psi(\mathbf{k}' + \mathbf{k}) e^{-i\gamma(\mathbf{k}'+\mathbf{k}+\mathbf{k}_0)} \right] . \tag{1.11}$$

## 1.3   The effects of limited coherence

The coherence of real electron sources, however, is always limited. The finite source size leads to a loss of spatial coherence as each point in the specimen is illuminated with beams from slightly diverging directions, characterised by a beam profile $I_s(\mathbf{k}_d)$ around the average incident beam direction, where $\mathbf{k}_d$ is the deviation from this direction.

Due to the presence of the chromatic aberration, the energy spread as the electrons leave the source, together with instabilities in the accelerating voltage and the lens currents, lead to a spread of the defocus value described by a distribution function $I_t(u)$ where $u$ is the deviation from the mean defocus.

The actual form of these distribution functions is far less important than their width and in general Gaussian distribution functions are used, as this makes integrals in the further derivation analytically tractable.[5]   Hence the following distribution functions are assumed:

$$I_s(\mathbf{k}_d) = \frac{\lambda^2}{\pi\beta^2} e^{-\lambda^2 k_d^2/\beta^2} \tag{1.12}$$

$$I_t(u) = \frac{1}{\sqrt{2\pi}d} e^{-u^2/2d^2} , \tag{1.13}$$

where $\beta = \lambda \langle |\mathbf{k}_d|^2 \rangle^{1/2}$ is the *beam divergence* and

$$d = C_{1c} \frac{1 + eU_a/m_0c^2}{1 + eU_a/2m_0c^2} \left[ \left( \frac{\Delta E}{eU_a} \right)^2 + \left( \frac{\Delta U_a}{U_a} \right)^2 + \left( \frac{2\Delta B_l}{B_l} \right)^2 \right]^{1/2} \tag{1.14}$$

---

[5]For the defocus spread, the terms due to voltage and lens current instabilities are well described by Gaussian function, whereas the intrinsic source energy spread would be more accurately described by a Maxwellian distribution for thermionic emitters or the Fowler-Nordheim equation (Fowler and Nordheim, 1928) for field emitters.

is the *defocus spread.* In this expression (e.g Reimer, 1989), $C_{1c}$ is the chromatic aberration (equation 1.1), $E$ is the electron energy as it leaves the source, $U_a$ is the accelerating voltage and $B_l$ is the strength of the objective lens magnetic field, which fluctuates due to objective lens current fluctuations.[6] [7]

As the electron density in the beam is low, treating them as independent quantum mechanical particles is an excellent approximation. Therefore, the contrast under this partially coherent illumination can be calculated as an incoherent superposition of the image contrasts under coherent illumination with the range of beam tilts and defocus values described above. This yields:

$$
\begin{aligned}
c(\mathbf{k}) &= \int \mathrm{d}\mathbf{k}_d \int \mathrm{d}u \, I_s(\mathbf{k}_d) I_t(u) \int \mathrm{d}\mathbf{k}' \\
&\quad \left[ \psi^*(\mathbf{k}')\psi(\mathbf{k}' + \mathbf{k}) e^{\mathrm{i}\gamma_u(\mathbf{k}'+\mathbf{k}_0+\mathbf{k}_d)} e^{-\mathrm{i}\gamma_u(\mathbf{k}'+\mathbf{k}+\mathbf{k}_0+\mathbf{k}_d)} \right] ,
\end{aligned} \tag{1.15}
$$

where the index in $\gamma_u$ means that the wave aberration has to be calculated at an defocus offset of $u$.

Changing the order of integration leads to:

$$
c(\mathbf{k}) = \int \mathrm{d}\mathbf{k}' \, \psi^*(\mathbf{k}')\psi(\mathbf{k}' + \mathbf{k}) T(\mathbf{k}' + \mathbf{k} + \mathbf{k}_0, \mathbf{k}' + \mathbf{k}_0) , \tag{1.16}
$$

with the 'Transmission Cross-Coefficient' (TCC) defined by

$$
T(\mathbf{k}_1, \mathbf{k}_2) = \int \mathrm{d}\mathbf{k}_d \int \mathrm{d}u \, I_s(\mathbf{k}_d) \, I_t(u) \, e^{-\mathrm{i}\gamma_u(\mathbf{k}_1+\mathbf{k}_d)} \, e^{\mathrm{i}\gamma_u(\mathbf{k}_2+\mathbf{k}_d)} . \tag{1.17}
$$

Making the assumption that $\mathbf{k}_d$ is sufficiently small that its influence on the chromatic term can be neglected, the integrals can be separated. Furthermore, $\gamma$ can be

---

[6] Usually, the last term is written in terms of the objective lens current $I_l$. However, this is not strictly correct, as the lens is normally operated close to saturation and the field $B_l$ is not proportional to $I_l$

[7] It may seem surprising that the objective lens current only influences the focus in the presence of chromatic aberration. However, this is due to the general scaling rule that the trajectories are identical when energy and magnetic field are changed according to $E' = k^2 E$ and $B' = kB$. Hence, in a microscope corrected for the chromatic aberration, the focus cannot be changed by changing the currents in all lens elements by the same factor. Instead, for instance, lens and corrector currents have to be change by different amounts.

replaced by its first-order Taylor expansion, leading to

$$
\begin{aligned}
T(\mathbf{k}_1, \mathbf{k}_2) \;=\; & \mathrm{e}^{-\mathrm{i}\gamma(\mathbf{k}_1)} \mathrm{e}^{\mathrm{i}\gamma(\mathbf{k}_2)} \\
& \int \mathrm{d}\mathbf{k}_d\, I_s(\mathbf{k}_d)\, \mathrm{e}^{-\mathrm{i}(\nabla\gamma(\mathbf{k}_1)-\nabla\gamma(\mathbf{k}_2))\cdot\mathbf{k}_d} \\
& \int \mathrm{d}u\, I_t(u)\, \mathrm{e}^{-\mathrm{i}\pi\lambda(k_1^2-k_2^2)u} \;.
\end{aligned}
\tag{1.18}
$$

Using the gaussian distribution functions from equations 1.12 and 1.13, these integrals can be solved analytically. Equation 1.18 then simplifies to

$$
T(\mathbf{k}_1, \mathbf{k}_2) = \mathrm{e}^{\mathrm{i}\gamma(\mathbf{k}_1)} \mathrm{e}^{-\mathrm{i}\gamma(\mathbf{k}_2)} \mathrm{e}^{-|\nabla\gamma(\mathbf{k}_1)-\nabla\gamma(\mathbf{k}_2)|^2 \beta^2/(4\lambda^2)} \mathrm{e}^{-\pi^2\lambda^2 d^2 (k_1^2-k_2^2)^2/2} \;.
\tag{1.19}
$$

The TCC describes how contributions arising due to the interference of two beams with angles $\lambda\mathbf{k}_1$ and $\lambda\mathbf{k}_2$ relative to the optic axis are transferred to the image contrast Fourier transform. Apart from the phase shifts due to the wave aberration, it contains the coherence envelope function

$$
E_c(\mathbf{k}_1, \mathbf{k}_2) = E_s(\mathbf{k}_1, \mathbf{k}_2) E_t(\mathbf{k}_1, \mathbf{k}_2) = \mathrm{e}^{-|\nabla\gamma(\mathbf{k}_1)-\nabla\gamma(\mathbf{k}_2)|^2 \beta^2/(4\lambda^2)} \mathrm{e}^{-\pi^2\lambda^2 d^2 (k_1^2-k_2^2)^2/2} \;,
\tag{1.20}
$$

that consists of the spatial and temporal coherence envelopes $E_s$ and $E_t$.

## 1.4 The linear imaging approximation

When the beams scattered by the object are weak compared to the transmitted beam (weak scattering approximation) terms due to interference between scattered beams can be neglected. In this case, using $T(\mathbf{k}_2, \mathbf{k}_1) = T^*(\mathbf{k}_1, \mathbf{k}_2)$, equation 1.16 simplifies to

$$
c(\mathbf{k}) = \psi(\mathbf{k}) T(\mathbf{k}_0 + \mathbf{k}, \mathbf{k}_0) + \psi_-^* T^*(\mathbf{k}_0 - \mathbf{k}, \mathbf{k}_0) - \mathrm{FT}[1] \;,
\tag{1.21}
$$

where $\lambda\mathbf{k}_0$ is again the tilt of the primary beam.[8] In this expression the image contrast depends linearly on the electron wavefunction therefore this approximation is also referred to as linear imaging approximation. Practical electron detectors such

---

[8]The last term is necessary because otherwise the cross term of the transmitted beam with itself would be included twice.

as photographic film or CCD cameras lead to an additional spatial frequency dependent attenuation of the image contrast, described by a Modulation Transfer Function $\text{MTF}(\mathbf{k})$ and discussed in more detail in the next chapter. For the wave restoration procedures used extensively in this work, it is convenient to include this MTF into a total wave transfer function defined by:

$$w(\mathbf{k}) = \begin{cases} \text{MTF}(\mathbf{k})T(\mathbf{k}_0 + \mathbf{k}, \mathbf{k}_0) & : \quad \mathbf{k} \neq 0 \\ 1/2 & : \quad \mathbf{k} = 0 \, , \end{cases} \tag{1.22}$$

where the exception at $\mathbf{k} = 0$ serves to eliminate the last term in equation 1.21,[9] so that the equation for the linear image contrast including the MTF finally reads

$$c(\mathbf{k}) = \psi(\mathbf{k})w(\mathbf{k}) + \psi^*(-\mathbf{k})w^*(-\mathbf{k}); \, . \tag{1.23}$$

When working within the linear imaging approximation it is convenient to drop the argument $\mathbf{k}$ and for a function $f(\mathbf{k})$ define $f = f(\mathbf{k})$ and $f_- = f(-\mathbf{k})$, simplifying the above equation to

$$c = \psi w + \psi_-^* w_-^* \, . \tag{1.24}$$

## 1.5  The phase contrast transfer function

The primary effect of thin specimens on the electron wave is an advance of its phase at the atomic positions due to the shorter electron wavelength in the attractive atomic core potential. For very thin objects (weak phase objects) the phase shift $\Phi$ is proportional to the projected atomic potential and the object wave can be described by:

$$\Psi(\mathbf{x}) = 1 + i\Phi(\mathbf{x}) \quad \text{with} \quad |\Phi(\mathbf{x})| \ll 1 \, , \tag{1.25}$$

which leads to the symmetry relation

$$\psi = -\psi_-^* \tag{1.26}$$

---

[9]It is common practice in the theory of linear image to avoid this extra term by subtracting one from both $C(\mathbf{x})$ and $\Psi(\mathbf{x})$ such that they describe the image contrast (fractional intensity variation) and the deviation from the illuminating wave, respectively. However, this makes the non-linear expression 1.16 less obvious.

between the object wave and the conjugate object wave, further simplifying the expression for the image contrast to

$$c = \psi(w - w_-^*) \tag{1.27}$$

Because the object only changes the phase of the wave, no contrast would be visible in case of aberration-free imaging. The contrast when aberrations are present is now discussed for the case of axial imaging.

## 1.5.1 Axial imaging

For axial illumination, the wave transfer function in equation 1.22 simplifies to

$$w = \mathrm{MTF}(\mathbf{k})E_c(\mathbf{k})e^{-\mathrm{i}\gamma(\mathbf{k})} \ , \tag{1.28}$$

where

$$E_c(\mathbf{k}) = E_c(\mathbf{k}, 0) = e^{-|\nabla\gamma(\mathbf{k})|^2\beta^2/(4\lambda^2)}e^{-\pi^2\lambda^2d^2k^4/2} \ , \tag{1.29}$$

following the definition in equation 1.20. Hence, the contrast is given by

$$c(\mathbf{k}) = \psi(\mathbf{k})\mathrm{MTF}(\mathbf{k})E_c(\mathbf{k})2\sin\gamma_e(\mathbf{k})e^{-\mathrm{i}\gamma_o(\mathbf{k})} \ , \tag{1.30}$$

where $\gamma$ has been split into symmetric (even order) and antisymmetric (odd order) parts according to

$$\gamma_e = (\gamma + \gamma_-)/2 \quad \text{and} \quad \gamma_o = (\gamma - \gamma_-)/2 \tag{1.31}$$

Equation 1.30 shows that the symmetric aberrations lead to a spatial frequency dependent transformation of phase into amplitude contrast described by the phase contrast transfer function

$$\mathrm{PCTF}(\mathbf{k}) = E_c(\mathbf{k})\sin\gamma_e(\mathbf{k}) \ . \tag{1.32}$$

One favourable focus setting for imaging of weak phase objects is the Scherzer-focus (Scherzer, 1949)[10]

$$C_{1,\mathrm{Scherzer}} = -1.2(C_3\lambda)^{1/2} \ , \tag{1.33}$$

---

[10]In this definition of the Scherzer focus, the passband in $\sin\gamma$ contains a local minimum with $\sin\gamma = 0.7$. The original definition, $C_{1,\mathrm{Scherzer}} = (C_3\lambda)^{1/2}$ avoids a local minimum in the passband at the cost of a slightly poorer point resolution $\rho_S = 0.707(C_3\lambda^3)^{1/4}$.

which leads to a relatively broad band of phase contrast transfer without zero-crossings up to a frequency of $k_{max} = 1.6(C_3\lambda^3)^{-1/4}$. The reciprocal of this is called the point resolution and given by

$$\rho_S = 0.625(C_3\lambda^3)^{1/4} \ . \tag{1.34}$$

The point resolution defines the maximum interpretable resolution. For higher spatial frequencies the contrast is partially reversed because the PCTF starts to oscillate as demonstrated in figure 1.1. The spatial and temporal coherence envelopes are also shown, demonstrating that the resolution can be significantly enhanced if information transferred by the oscillating part of the PCTF is made accessible. This can done with the wavefront restoration procedures described in the next section.

Fig 1.1 also shows the PCTF at $C_1 = 0$ (Gaussian focus), where very little phase contrast is transferred at medium and low spatial frequencies. This means that the Gaussian focus can be found by minimising the visible contrast of an amorphous weak phase object, which is exploited in manual focusing and the minimum variance method given in section 3.1.

More sophisticated methods of aberration determination directly exploit the zero-crossings of the PCTF, as the *diffractogram, i.e.* the power spectrum $|c|^2$ of an amorphous weak phase object will display dark bands (Thon rings) at these positions (Thon, 1966). This is exploited in diffractogram based aberration determination methods, as described in section 3.3.1.

For most crystalline objects, both the phase and the modulus carry information about the specimen, hence it is desirable to recover the full complex aberration free object wave for a large range of spatial frequencies. This can be done by combining the information from several images taken at different imaging conditions and is described in detail in section 1.6.
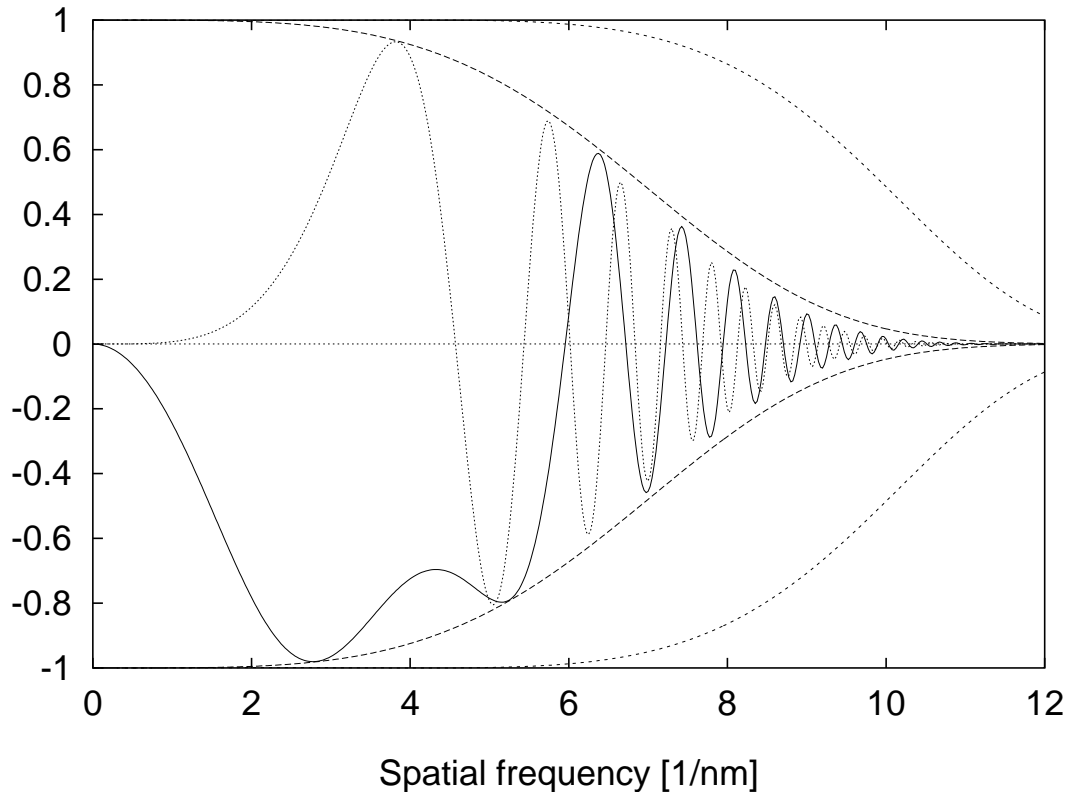
Figure 1.1: The phase contrast transfer function of the JEM3000F microscope installed in the Department of Materials in Oxford ($C_3 = 0.57$ mm, focal spread $d = 4$ nm and beam divergence $\beta = 0.1$ mrad). Solid graph: PCTF at Scherzer focus; the first zero crossing defines the point resolution of 1.678 nm. Dotted graph: PCTF at Gaussian focus. The coherence envelopes are also shown. The temporal coherence envelope $E_t$ (long dashed) is far more limiting than the spatial coherence envelope $E_s$ (short dashed, calculated at Scherzer focus) due to the extremely small source size of field emitters.

### 1.5.2 Voltage centre versus coma-free alignment

The above has been derived under the assumption that both the voltage centre and the coma-free axis are at the origin of Fourier space. In practice these two axes do not coincide, hence the effects of misalignment from each axis have to be studied separately. Misalignment from the voltage-centre axis alone only leads to an asymmetry in the temporal coherence envelope, as this is centred around the voltage centre axis. In case of the coma-free axis, the effects are more severe, as misalignment not only affects the spatial coherence envelope, but also the wave aberration function directly. Figure 1.2 shows the relative error, defined by

$$\Delta = \frac{2|\psi_\tau - \psi|}{|\psi_\tau| + |\psi|} \ , \tag{1.35}$$

in the wavefunction $\psi_\tau$ due to the presence of a misalignment $\tau$ from either axis, in which a misalignment from the coma-free axis induces more severe errors.

## 1.6 Restoring the object wave from sets of images

The possibility of restoring the complex object wave using a series of images at different foci was first discussed by Schiske (1968). However, this idea was ahead of its time and significant progress both instrumentally and theoretically was needed to put it to practical use. On the instrumental side, the coherence of the electron beam was greatly increased by introduction of stable Schottky field emission source into TEMs by Otten and Coene (1993) and Honda et al. (1994), pushing the information limit well beyond the point resolution. This made wave restoration more attractive because firstly the rapidly oscillating PCTF leads to a contrast delocalisation making single images more difficult to interpret even at Scherzer focus and secondly, such restoration now leads to a substantial gain in interpretable resolution from the point resolution to the information limit. The second key improvement came with the introduction of CCD cameras into electron microscopy (chapter 2), which provide an almost perfectly linear and sensitive electron detector recording images directly in
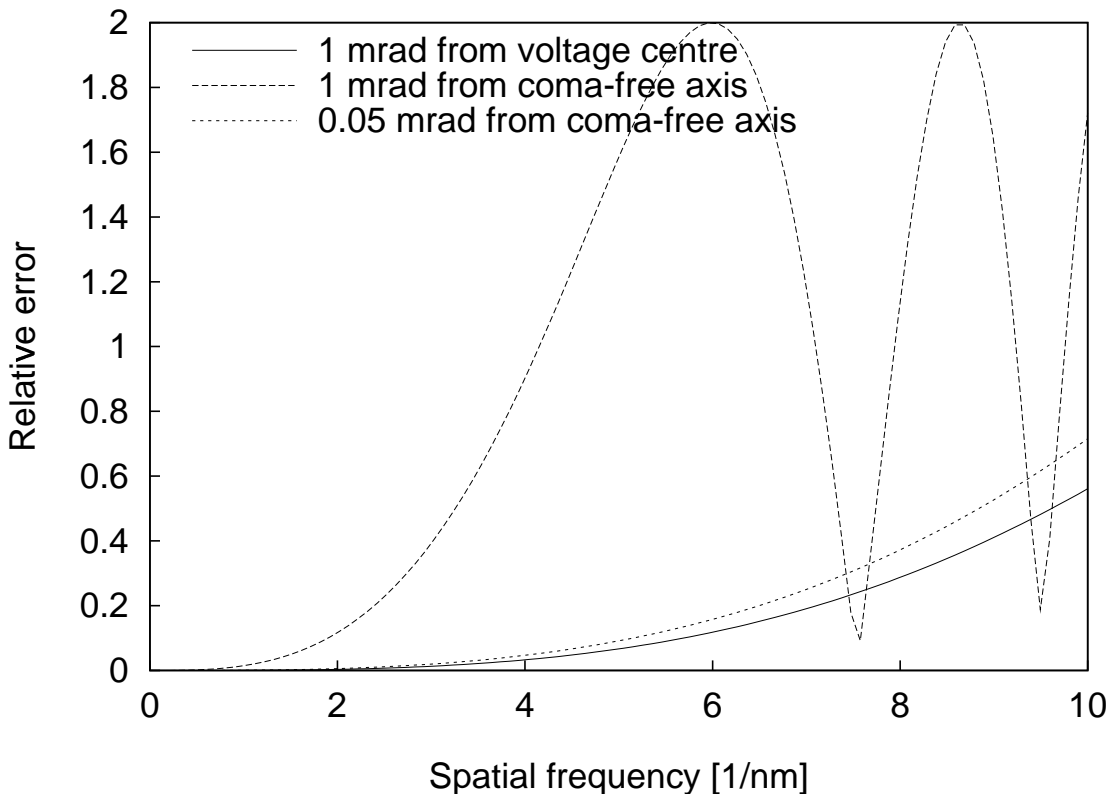
Figure 1.2: Relative error in the image wave for alignment conditions where the voltage centre and the coma-free axis do not coincide. Solid line: The microscope is aligned to the coma-free axis, which is 1 mrad off the voltage centre. Long dashed line: The microscope is aligned to the voltage centre, which is 1 mrad off the coma-free-axis. Short dashed line: The microscope is aligned to the voltage centre, which is 0.05 mrad off the coma-free-axis.

digital form. The third and probably most crucial development was the breathtaking improvement of computing power available at ever decreasing cost.

Concomitant theoretical developments include the Wiener formulation of the problem (Schiske, 1973), which allows an optimum restoration in the presence of noise, the incorporation of the effects of partial coherence (Wade and Frank, 1977) by (Saxton, 1978), followed by a generalisation of the restoration scheme to nonlinear imaging (Kirkland, 1984). Much later, an independent derivation of a linear restoration scheme called 'paraboloid method' based on 3D-Fourier transforms was given (Coene et al., 1992; van Dyck et al., 1993; Op de Beeck et al., 1996). Although this method is inferior to the ones presented earlier in the suppression of the conjugate wave (Saxton, 1994b), it provides some intuitive insights into how the information from a series of images is gathered to yield the complex object wave and for this reason the derivation is briefly discussed here.

### 1.6.1 The paraboloid method

The basic features of the paraboloid method are illustrated in figure 1.3. Taking a series of images at different focus levels can be interpreted as taking a three-dimensional sample of the image wave. In the three-dimensional Fourier transform of the image wave itself, all contributions due to elastically scattered electrons are located on the Ewald sphere with radius $k = 1/\lambda$ corresponding to the incident energy (indicated as solid circle section in figure 1.3). However, only the image intensity can be recorded, which in the linear imaging approximation contains not only the interference term between the incident beam and scattered beam, but also the conjugate thereof. The former leads to contributions on the Ewald sphere shifted to the origin (long dashed), and the latter is the complex conjugated point reflection of this (short dashed). Due to the small scattering angles involved, the spheres are excellently approximated by paraboloids, which is effectively equivalent to using the Fresnel approximation to describe the propagation between the image planes.

As only data on the paraboloid corresponding to the direct wave is used in the
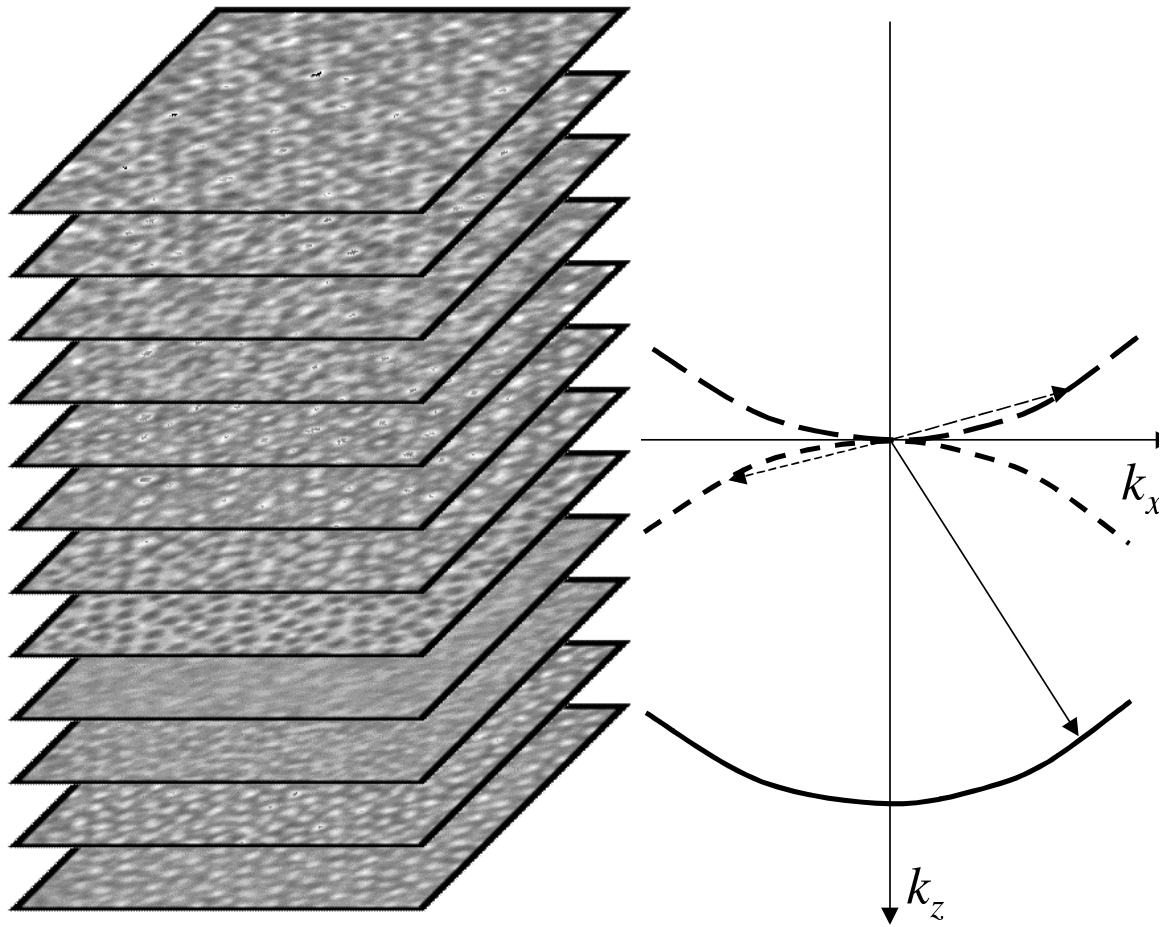
Figure 1.3: Illustration of the paraboloid method. A focal series represents a measurement of the intensity in the image waves in three dimensions (left). In the 3-D Fourier transform of the image wave, contributions from elastically scattered electrons are located on the Ewald sphere (solid curve). The 3-D Fourier transform of the recorded intensities contains contributions from both the direct wave (long dashed) and the conjugate wave (short dashed)

restoration, calculating the 3-D transform in order to extract the restored wave would be wasteful and also cumbersome due to the sparse sampling of the 3-D transform in the $z$-direction. Instead, the required Fourier component at $k_z = 0.5\lambda k^2$ can be evaluated with a single integration

$$\int dz \, c(\mathbf{k}, z) e^{-\pi i \lambda k^2 z} \; . \tag{1.36}$$

However, $c(\mathbf{k}, z)$ is only known at a finite number $N$ of equally spaced focus levels $z_n = n * d_z$, and hence the restored wave is evaluated as

$$\psi' = \frac{1}{N} \sum_{n=0}^{N-1} c(\mathbf{k}, z) e^{-\pi i \lambda k^2 z_n} \; . \tag{1.37}$$

This discretisation introduces two numerical artifacts: aliasing due to the discrete sampling and windowing due to the restriction of $z$ to the range $0 \leq z < N d_z$. The former effect leads to a periodic repetition of the Fourier spectrum in the $z$-direction, effectively copying the pair of paraboloids in figure 1.3 from the origin to positions $k_z = 1/(N d_z)$, thereby leading to additional overlaps of the direct and the conjugate wave. The latter effect is equivalent to a convolution with a sinc-function, which broadens the range of frequencies with significant transfer of the unwanted conjugate wave. Together, both effects are equivalent to a convolution with the function

$$\frac{\sin(N \pi d_z k_z)}{\sin(\pi d_z k_z)} \; , \tag{1.38}$$

which represents the wave diffracted from a $N$ slit grating with spacing $d_z$. This leads finally to the restored wave

$$\psi' = \psi + \frac{\sin(N \pi \lambda d_z k^2)}{\sin(\pi \lambda d_z k^2)} \psi_-^* \; , \tag{1.39}$$

when the effects due to limited coherence and noise are ignored.

## 1.6.2 Improved linear restoration filters

Equation 1.37 is an example of an equation describing a linear restoration filter, which have the general form:

$$\psi' = \sum_i r_i c_i \; . \tag{1.40}$$

The aim of this section is to find *restoring filters, $r_i$* that provide a better suppression of the unwanted conjugate wave, incorporate the limited coherence and take the noise present in the images into account. The approach described here follows that given by Saxton (1988), which has the further advantage that it is not restricted to restoration from equally spaced focal series but can be applied to any set of differently aberrated images with known wave transfer functions $w_i$.

The most obvious solution to finding the two unknowns $\psi$ and $\psi_-$ at each spatial frequency $\mathbf{k}$ consists of minimising the least-squares difference between actual predicted image transforms:

$$\sum_i \left( w_i \psi + w_{i-}^* \psi_-^* - c_i \right)^2 \overset{!}{=} \min . \tag{1.41}$$

The condition that the first derivative with respect to both $\psi$ and $\psi_-^*$ vanishes yields an equation for $\psi$ in the form of equation 1.40 with restoring filters

$$r_i = \frac{W_- w_i^* - V^* w_{i-}}{W_- W - |V|^2} \tag{1.42}$$

where $W = \sum_i |w_i|^2$ and $V = \sum_i w_i w_{i-}$. This solution yields the best fit to the available data. However, obvious problems arise when this expression is singular, which happens at frequencies where insufficient information is available, either because the transfer functions $w_i$ are zero for all images, or $w_i w_{i-}$ is identical for all images with non-zero $w_i$.

Furthermore, when the denominator of equation 1.42 is small, any noise present in the image is strongly amplified. Due to this amplified noise, the agreement between the restored wave and the actual wave can be poor in spite of the optimal agreement between the experimental images and predicted images (where the noise is reduced to its original value again). These problems are solved by an ingenious approach due to Wiener, where the (unknown) noise $n_i$ is directly incorporated into the equations for the observed image contrast:

$$c_i = w_i \psi + w_{i-}^* \psi_-^* + n_i \tag{1.43}$$

Restoring filters $r_i$ are now sought such that

$$\left\langle |\psi' - \psi|^2 \right\rangle = \left\langle |\sum_i r_i c_i - \psi|^2 \right\rangle \overset{!}{=} \min , \tag{1.44}$$

where $\langle \cdots \rangle$ denotes an average over an ensemble of typical restorations, each characterised by an experimental wavefunction $\psi$ and image noise $n_i$.

By differentiating by the real and imaginary parts of $z$ separately it can be shown that for an analytical function $f(z)$

$$\left\langle |f(z)|^2 \right\rangle \overset{!}{=} \min \quad \Longleftrightarrow \quad \left\langle f(z) \left( \frac{\mathrm{d}f}{\mathrm{d}z} \right)^* \right\rangle = 0 , \tag{1.45}$$

hence the minimum condition w.r.t. $r_k$ yields

$$\left\langle \left( \psi \Sigma r_i w_i + \psi_-^* \Sigma r_i w_{i-}^* + \Sigma r_i n_i - \psi \right) \left( \psi^* w_k^* + \psi_- w_{k-} + n_k^* \right) \right\rangle = 0 . \tag{1.46}$$

Using the identities

$$\begin{aligned}
\left\langle \psi_\pm n_k^* \right\rangle &= 0 \quad &\text{(no correlation between object and noise)} \tag{1.47} \\
\left\langle n_i n_k^* \right\rangle &= \left\langle n_k n_k^* \right\rangle \delta_{ik} \quad &\text{(noise in different images is not correlated)} \tag{1.48} \\
\left\langle n_k n_k^* \right\rangle &= P_N \quad &\text{(definition of mean noise power in images)} \tag{1.49} \\
\left\langle \psi^* \psi \right\rangle &= P_S \quad &\text{(definition of mean signal power in object wave)} \tag{1.50} \\
\left\langle \psi \psi_- \right\rangle &= 0 \quad &\text{(no assumption is made about the object type)} \tag{1.51}
\end{aligned}$$

equation 1.46 can be simplified to

$$\sum_i P_S (w_k^* w_i + w_{k-} w_{i-}^*) r_i + P_N r_k - P_S w_k^* = 0 . \tag{1.52}$$

This can be solved with the Ansatz

$$r_i = a w_i^* + b w_{i-} \tag{1.53}$$

yielding

$$r_i = \frac{(W_- + \nu) w_i^* - V^* w_{i-}}{(W_- + \nu)(W + \nu) - |V|^2} \tag{1.54}$$

in which $W$ denotes $\sum_i |w_i|^2$, $V$ denotes $\sum_i w_i w_{i-}$, and $\nu = P_N/P_S$ denotes an estimate of the expected noise intensity in $c_i$ as a fraction of the expected signal intensity in $\psi$.

Provided that the wave transfer functions, $w_i$ assumed in calculating the restoring filters accurately represent the experimental imaging conditions, the transfer from the object wave to the restored wavefunction can be described by

$$\text{WTF} = \sum_i r_i w_i = \frac{(W_- + \nu)W - |V|^2}{(W_- + \nu)(W + \nu) - |V|^2} \; . \tag{1.55}$$

This total wave transfer function is free of phase shifts and close to one where at least two images that provide complementary information have a substantial transfer function $w_i$.

For some (especially very low) spatial frequencies, the conjugate wave is not completely suppressed due to a lack of phase information from any image. This unwanted transfer of the conjugate wave is described by

$$\text{WTF}_C = \sum_i r_i w_{i-}^* = \frac{V^* \nu}{(W_- + \nu)(W + \nu) - |V|^2} \; . \tag{1.56}$$

The difference

$$Q = |\text{WTF}|^2 - |\text{WTF}_C|^2 \tag{1.57}$$

is hence a measure of the quality of the restoration and is used as a weighting function in the following.

At this point, a few remarks should be made on the definitions of the mean signal and noise powers $P_S$ and $P_N$. While $P_N$ can be easily characterised (as discussed in detail in chapter 2), $P_S$ depends on the choice of the "ensemble of typical wavefunctions". Obviously, restricting this ensemble supplies prior information about the object and will in general improve the signal-to-noise ratio in the restoration. In the general case where prior assumptions about the object are not desirable, $P_S$ should be chosen as constant. An alternative scheme involves the use of prior knowledge about the elements present in the sample and their scattering factors and choose $P_S$ to decay in proportion to a typical scattering factor. An attempt could also be made

to estimate the signal power from the experimental images themselves. However, a single experimental object is hardly a representative ensemble and for crystalline specimens, $P_S$ will have sharp peaks at the positions of the crystal reflections, effectively leading to a Fourier filter that is biased towards the frequencies of crystal reflections. This will reduce the noise, but it will also lead to an exaggeration of the periodic information in the crystal and obscure aperiodic information

It is also instructive to re-investigate the assumption that the ensemble average of $\psi\psi_-$ is zero. Here it is possible to incorporate prior knowledge that the object is a pure phase object satisfying $\psi_- = -\psi^*$. An ensemble containing only such objects would yield

$$\left\langle \psi\psi_- \right\rangle = \left\langle \psi\psi^* \right\rangle = P_S \;, \qquad (1.58)$$

giving restoring filters that yield a restoration with zero contrast in its modulus for any object. Again, the noise is reduced due to the prior knowledge employed, as the noise in the modulus is eliminated.

One item of prior knowledge always used by the Wiener filter is that the object spectrum is bounded and hence $\psi = 0$ is the best estimate for frequencies where no information is available. This sets it apart from the simple least-squares filter discussed at the beginning of this section, as the least-squares approach assumes that, *a priori*, all values in the complex plane are equally probable for $\psi$. In fact, the least-squares restoration filters (equation 1.42) are reproduced from the Wiener filter equation 1.54 when $P_N \to 0$ or $P_S \to \infty$, both of which are highly unrealistic scenarios.

When comparisons with simulations are made, a simulated restored wave $\psi_{\text{simrest}}$ according to

$$\psi_{\text{simrest}} = \text{WTF}(\mathbf{k})\psi_{\text{sim}}(\mathbf{k}) + \text{WTF}_C(\mathbf{k})\psi_{\text{sim}}^*(-\mathbf{k}) \;. \qquad (1.59)$$

is calculated from the simulated back-propagated exit wave $\psi_{\text{sim}}$ for direct comparison (appendix A). This means that irrespective of how $P_S$ is chosen the experimental and simulated waves are subject to identical transfer functions, and the only dependence of this comparison on $P_S$ arises from the weight given to discrepancies at different

spatial frequencies.

Computationally, the restoration process is not demanding, requiring little more than a handful of Fourier transforms. The process also lends itself to iterative inclusion of additional images to the restoration. For this, it is necessary to keep track of the sums

$$
\begin{aligned}
W &= \textstyle\sum_i |w_i|^2 \;, & V &= \textstyle\sum_i w_i w_{i-} \;, \\
T &= \textstyle\sum_i w_i^* c_i & \text{and} \quad U &= \textstyle\sum_i w_{i-} c_i \;,
\end{aligned}
\tag{1.60}
$$

whereby at any stage the restoration can be calculated as:

$$
\psi' = \frac{(W_- + \nu)T - V^* U}{(W_- + \nu)(W + \nu) - |V|^2} \;.
\tag{1.61}
$$

The object wavefunction itself is obtained simply by inverse transformation of this estimate. An image can also be removed again from the restoration by subtracting its contribution from the sums in equation 1.60.

A restoration of the object wave is only possible when the images are accurately registered with respect to a common origin and when the imaging parameters for each image, and hence the wave aberration functions $w_i$, are known.

An estimate for the aberrated *image* wave can, however, already be obtained when the relative aberrations between the images, *i.e.* in the case of a focal series the relative focus levels and the image displacements, are known. This ensures that the contributions from the individual images are summed with the correct phases.

In chapter 4, a method is presented for firstly obtaining the relative aberrations and image shifts in a focal series and subsequently using a restored image wave to determine the symmetric aberration coefficients.

## 1.6.3 Restoration from sets of tilted images

The Wiener filter restoration algorithm discussed in the previous section is sufficiently general to allow a restoration of the object wave from nearly arbitrary sets of images taken under different imaging conditions. The most obvious choice is a focus series as discussed above, which offers interpretable resolution up to the information limit

set by the imperfect temporal coherence. Limited spatial coherence does not reduce the obtainable resolution as long as images at large enough underfocus values are included so that their overlapping pass bands (occurring where the wave aberration function gradient is small) provide the high spatial frequency information.

Resolution beyond this information limit can be achieved using sets of images with tilted illumination. For light microscopes, this was suggested by Abbe (1873): when the primary beam is placed close to one side of the aperture, it can interfere with beams scattered by up to twice the aperture semiangle, which dictates the resolution in axial mode. However, beams at different azimuth are masked completely, so that information from several images recorded at different tilt azimuths must be combined.

The effect is similar in electron microscopy, even though the temporal coherence envelope does not act as a solid aperture. It follows from equation 1.20 that for a primary beam direction $\lambda \mathbf{k}_0$ the transfer of a beam scattered by an angle $\lambda \mathbf{k}$ is given by

$$\mathrm{e}^{-\pi^2 \lambda^2 d^2 (|\mathbf{k}_0|^2 - |\mathbf{k}_0 + \mathbf{k}|^2)^2 / 2} . \tag{1.62}$$

It is apparent from this equation that beams scattered into directions that subtend the same angle with the voltage centre axis as the incident beam will not be affected by the temporal coherence envelope. The circle defined by these directions is therefore called the *achromatic circle*. Provided that the tilt angle is not very large, the transfer inside the achromatic circle will also be good. As in optical microscopy, images taken at tilted illumination show a one-sided transfer function extending to scattering angles of approximately twice the tilt angle in one direction. However, the tilt angle is not limited to the maximum transferred scattering angle under axial illumination. If tilt angles larger than that corresponding to the axial information limit are chosen, the contrast transfer on the achromatic circle is not reduced by the imperfect temporal coherence, but a region of weak transfer appears within this circle and for very large tilts. For tilts larger than 2Gl/Sch[11] this transfer gap has to be filled with additional images at smaller beam tilts.

---

[11]1 Gl/Sch represents a unit tilt in reduced units, where Scherzer (1 Sch = $(C_3\lambda)^{0.5}$) and Glaser (1Gl = $C_3^{0.25}\lambda^{0.75}$) are the reduced units for longitudinal and transverse distances in the microscope.
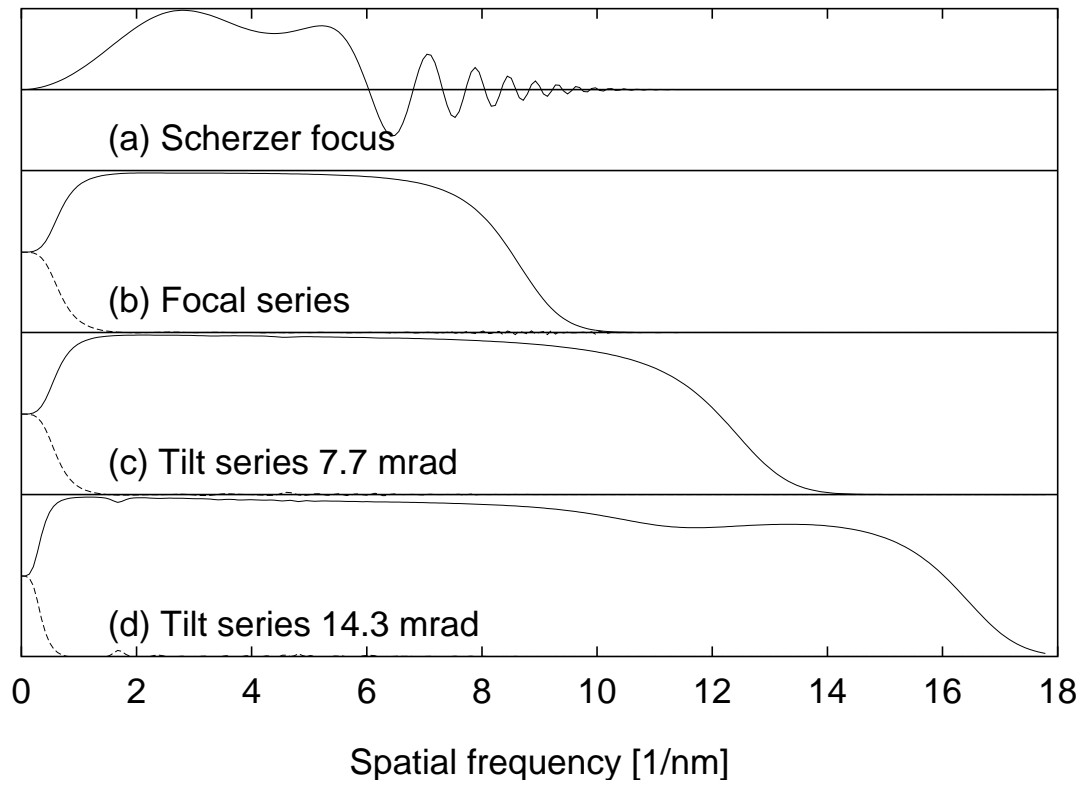
Figure 1.4: (a) The phase contrast transfer function for a single axial image at Scherzer focus (300kV, $C_3 = 0.57$, focal spread $d = 4\,\text{nm}$ and beam divergence $\beta = 0.1\,\text{mrad}$) in comparison with the effective wave transfer function for (b) a 21 member focal series with focus step 5nm; (c) a 21 member tilt/focus series with a tilt of 7.7mrad (1Gl/Sch) and (d) a 21 member tilt/focus series with a tilt of 15.3mrad (2Gl/Sch). In (b) to (d) the effective conjugate wave transfer function is also shown (dashed).
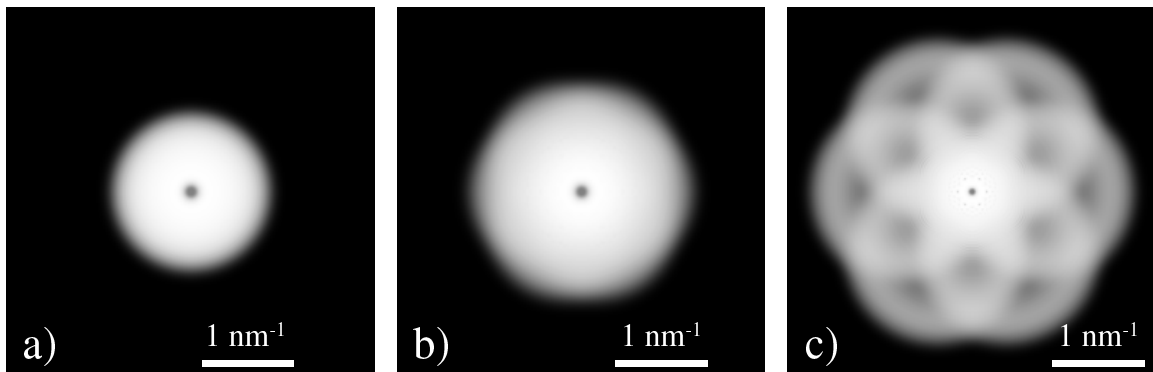
Figure 1.5: The effective wave transfer function for (a) a 21 member focal series with 5nm focus step, (b) a tilt/focus series with three focus levels at each tilt and tilts of 1 Gl/Sch =7.7 mrad with 6 different azimuths, (c) as for (b) but with tilts of 2 Gl/Sch=15.3mrad.

Figures 1.4 and 1.5 show simulations of the effective wave transfer function, demonstrating how the resolution is improved by combining images with different beam tilt directions.

### 1.6.4   The parallax problem

A potential problem of object wave restoration from images taken at different beam tilts is that, in contrast to the case of a focus series, the images are not simply different measurements on the same exit wavefunction. Instead, each image records the specimen projected along a slightly different direction. However, as illustrated in figure 1.6, this parallax surprisingly improves the resemblance of the restored phase to the projected potential, at least in the kinematic approximation. This effect can be understood both in real space and Fourier space models. For axial illumination there is a phase difference of $\delta = 2\pi(1 - \cos\theta)\Delta z/\lambda$ between the contributions to scattering by an angle $\theta$ from two atoms in the same atomic column separated by $\Delta z$. As the contributions from different specimen layers are increasingly out of phase for increasing scattering angle, the high spatial frequency information is suppressed

compared to that expected from the 2-dimensional Fourier transform of the full projected crystal potential. If the illumination is tilted by an angle $\Theta$, contributions to the beam scattered by $2\Theta$ in the opposite direction from different specimen layers are now in phase, and the phase shifts are strongly reduced for beams scattered in similar directions. In Fourier space, tilting the beam corresponds to tilting the Ewald sphere towards the crystal reflections that correspond to scattering in the opposite direction.

As the beams that are scattered in directions approximately opposite the tilting direction are those transferred most strongly into the image contrast by the microscope transfer functions, restoring a wavefunction from a tilt series has the effect of reducing the curvature of the Ewald sphere and thereby, in the kinematical approximation, making the phase of the restored wavefunction more similar to the projected potential than the phase of the exit surface wavefunction under axial illumination.

However, for thicker specimens, where the scattering is highly dynamical, the structural interpretation of the object wave is expected to be more complicated when it is restored from images with different beam directions, *i.e.* different specimen excitations.
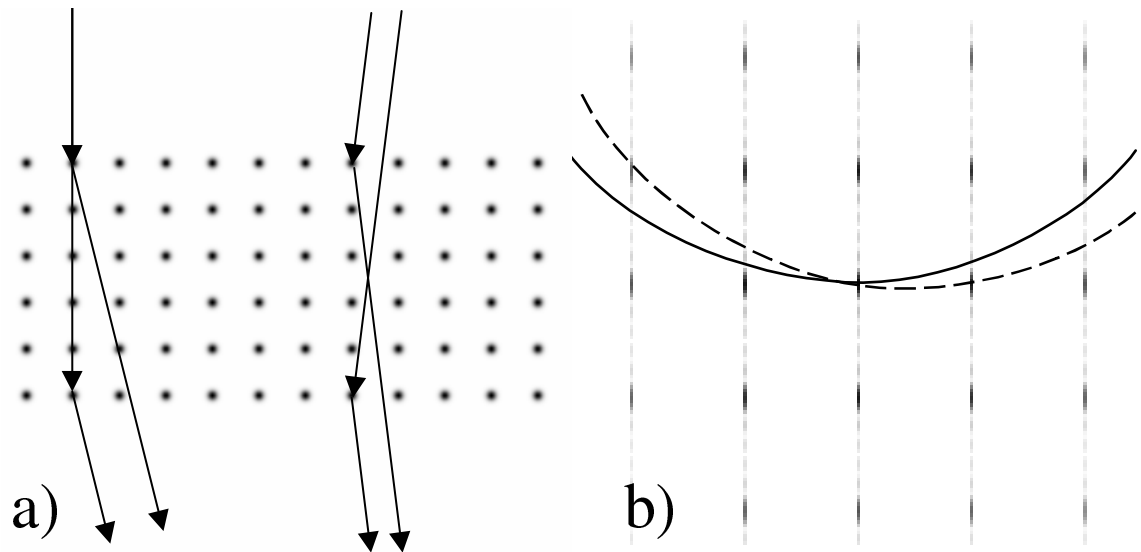
Figure 1.6: (a) Diffraction at a slab of crystal. For axial illumination (left) there is a phase shift between the beam paths into a specific scattering direction via different atoms of one column. Under tilted illumination (right), the phase shift is reduced to zero for scattering by twice the tilt angle. (b) Due to the finite thickness of the crystal, the reflections in the reciprocal lattice are elongated in $z$-direction, which allows some overlap between the $k_z = 0$ reflections and the Ewald sphere under axial illumination (solid arc). Under tilted illumination (dashed arc), the Ewald sphere is tilted towards the beams scattered in directions opposite of the tilt direction.

# Chapter 2

# Image detection: Characterisation of CCD cameras

The image detection device is an essential part of the image formation chain and therefore its properties must be accurately characterised to allow quantitative restoration of the object wavefunction as described in the last chapter. Extensive work that has been carried out on the accurate characterisation of image detectors is presented in this chapter. The work focuses on CCD cameras, the most widely used modern image detector in electron microscopy, although most of the theory can also be applied to other image detectors.

## 2.1   Introduction

### 2.1.1   Charge-coupled devices

Charge-Coupled-Devices, originally designed for early solid state memory chips, have been in use as pixelated light detecting devices for several decades (Howell, 2000). They exploit the internal photoelectric effect in semiconductors and the field effect in Metal-Oxide Semiconductor (MOS) structures. Due to the photoelectric effect a photon whose energy exceeds the semiconductor bandgap energy can excite a valence

band electron into the conduction band thereby generating an electron-hole pair. In a MOS structure, a thin oxide layer is used to insulate a metal (or polysilicon) gate from the generally (for CCD chips) p-doped semiconductor. Due to the field effect a positive voltage on this gate expels the majority charge carriers (holes) away from the semiconductor surface, leaving a negative space charge due to the filled acceptor levels.

Figure 2.1 shows the bending in the band structure that arises from this space charge. If the applied voltage is high enough, the lower edge of the conduction band will dip below the Fermi energy, and hence an inversion layer where the majority charge carriers are electrons rather than holes is created. Electron-hole pairs generated within or close to the this p-n junction[1] are separated, and the electrons are collected in the inversion layer at the surface.

With an array of independent gates, the semiconductor surface can be split into mutually insulated potential wells by applying different voltages to the gate electrodes (figure 2.2(a)). When the camera is illuminated, each of these pixels accumulates a charge that is proportional to the intensity for a given wavelength integrated over the pixel area. By clocking the gate voltages, the pattern of stored charges can be moved towards one end of the chain of pixels as illustrated in figure 2.2. This allows the sequential readout of the stored charge packets into a single readout amplifier, where the collected charge is measured and converted to a digital number. In two-dimensional arrays the most basic readout geometry is that of slow-scan CCD cameras. In these cameras, the contents of the whole array can be moved vertically, so that the charge in each pixel is shifted to the pixel directly below it. The charge from the lowest line is transferred to a special transfer register, in which it can be transferred horizontally to read a whole line of pixels. Subsequently, the next line is shifted into the transfer register until the whole array is read.

---

[1] A detailed account of the physics of the p-n junction is given in may solid state and semiconductor textbooks. Though slightly dated, the thorough description in Ashcroft and Mermin (1976) is recommended.
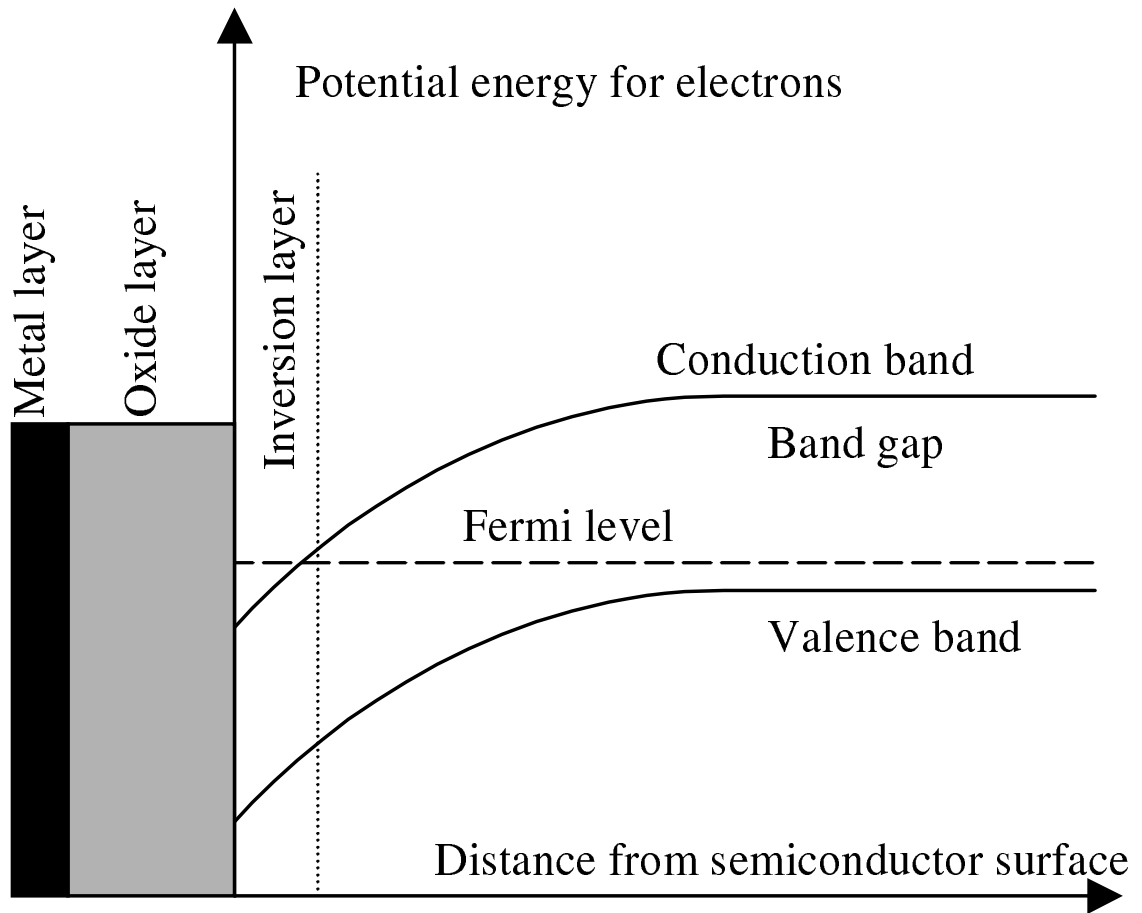
Figure 2.1: The band structure of a p-type semiconductor near a MOS interface. The positive voltage on the metal gate repels the holes from the interface leaving a region with negative space charge in which the band edges are bent. In this illustration the voltage is high enough to push the conduction band edge below the Fermi level and hence to create an n-type inversion layer.
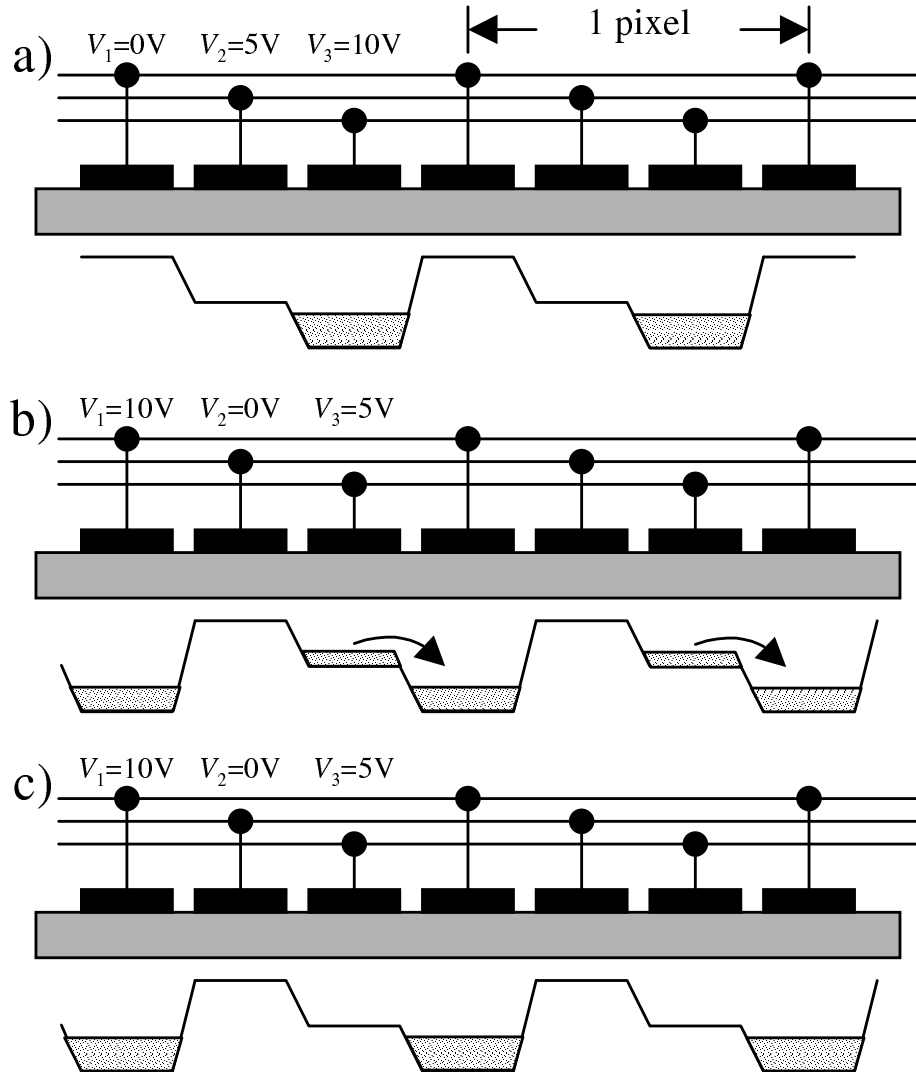
Figure 2.2: Charge transfer in a row of CCD pixels. There are three electrodes on top of each pixel with equivalent electrodes of all pixels connected. During exposure, the pattern shown in (a) is applied to the gates, yielding one deep potential well per pixel where the electrons generated in the whole pixel area are collected. After exposure, the pattern of collected charge can be shifted by clocking the voltage pattern applied to the electrodes, which moves the charge in each potential well to the right into a newly created potential well (b). In (c) the charge transfer is completed and the whole pattern of stored charges has move by 1/3 of a pixel to the right.

During readout the illumination must be blocked by a shutter, otherwise an image feature illuminating a certain pixel will also contribute briefly to the charge read out from all pixels above this pixel. Shutterless operation can be achieved with interline transfer and frame transfer CCD cameras. In the former, each column of light-sensitive pixels is interleaved with a column of pixels with opaque coating which serve as memory only. During readout, the first step is the simultaneous charge transfer from all light-sensitive pixels to the neighbouring memory pixels. The array of memory pixels is then read out as usual, while the next frame is already accumulated in the sensitive pixels. This loss of sensitive area reduces the sensitivity by a factor of two and the rectangular sensitive pixel area results in an anisotropic modulation transfer function. These disadvantages are avoided in frame-transfer CCD cameras, which function like ordinary CCD cameras, apart from separate vertical clocking for the upper and lower halves of the array and an opaque coating on the lower half, which serves as memory array. During readout, the image in the sensitive array is first transferred to the memory array by fast vertical clocking of both arrays. Then, the memory array is read out line by line as usual, while the sensitive array accumulates the next frame. The shutterless operation of frame-transfer and interline-transfer CCD cameras makes them attractive for electron microscopy as the use of a pre-specimen beam blank can lead to specimen drift during exposure due to local heating and charging effects and the use of the mechanical shutter can cause vibrations. For the recording of image series, a frame-transfer camera allows 100% of the specimen irradiation time to be used for image acquisition. Unfortunately, these cameras are currently not available with array sizes and pixel numbers[2] comparable to conventional slow-scan cameras which is why most microscopes, including those used in this work, are equipped with cameras using slow-scan CCD chips.

---

[2]Typical commercially available slow-scan cameras have 2048×2048 pixels with 30 $\mu$m spacing, while typical frame-transfer cameras feature only 1280×1024 pixels with 14 $\mu$m spacing. However, Fan et al. (2000) have reported the experimental design of a frame-transfer camera with 2560×2048 24 $\mu$m pixels.

### 2.1.2 CCD cameras in electron microscopy

CCD-cameras for electron microscopy have been commercially available for about 10 years, following developments described elsewhere (Spence and Zuo, 1988; Daberkow et al., 1991; Kujawa and Krahl, 1992; Daberkow et al., 1996; de Ruijter, 1995; Krivanek and Mooney, 1993; Faruqi and Subramaniam, 2000). Their advantages include practically linear response over an intrinsic dynamic range of approximately $10^4$, high sensitivity (making single-electron detection possible) and instant availability of the recorded images in digital form.

Though early experiments using CCD chips as direct TEM electron detectors have been performed (Roberts et al., 1982), this has not proved viable due to the sensitivity of the gate insulator to radiation damage. For low energies ($<$10 keV), this damage can be avoided using back-thinned CCDs (Ravel and Reinheimer, 1991), but higher energy electrons would penetrate through to the sensitive gate oxide on the front side. Moreover, in silicon a primary electron of energy $E$ generates $E/3.64eV$ electron-hole pairs (Fiebinger and Müller, 1972), hence the CCD well capacity of typically $400\,000$ electrons (de Ruijter, 1995) would be saturated after the detection of only a few electrons at the energies used in TEM.

Therefore, an indirect approach is employed in all cameras used in TEM applications. The electrons impinge on a YAG single crystal[3] or phosphor powder scintillator,[4] and the generated light is relayed to the CCD chip via a lens- or fibre optical coupling, as illustrated in figure 2.3.

### 2.1.3 Characterisation of CCD cameras

Scattering of both the primary electrons and the emitted photons blurs the recorded image by attenuating its high spatial frequencies. This spatial frequency dependent

---

[3]YAG stands for Yttrium Aluminium Garnet $Y_3Al_5O_{12}$, a transparent crystal which is made scintillating by doping with impurity atoms, usually europium (YAG:Eu) or cerium (YAG:Ce).

[4]A wide range of materials is available for powder scintillators. Popular choices in CCD cameras include: P22 ($Y_2O_2S : Eu$) and P43 ($Gd_2O_2S : Tb$ or short GOS).

signal transfer is described by the Modulation Transfer Function (MTF). The attenuation alone does not impair the quality of an image; as long as all frequency components are transferred, the image can, in principle, be perfectly restored by deconvolution if the MTF is known. However, the Signal to Noise Ratio (SNR) is reduced if the camera adds noise. The statistical performance of radiation detectors is most conveniently described by the Detection Quantum Efficiency (DQE) (Herrmann and Krahl, 1982), defined as the quotient of the squared SNR at the output and the input of the camera (in general as a function of spatial frequency (de Ruijter, 1995)) and this spatial frequency dependent DQE is the most important figure of merit for the performance of a CCD camera.

Two methods are used for measuring the MTF, deterministic methods and stochastic methods. In the former the detector is illuminated with spatially varying intensity and the MTF is determined by comparing the contrast between the input and the output images at different spatial frequencies. In practice the use of a knife edge where the input intensity can be assumed to be a step function is most commonly applied (Daberkow et al., 1991; de Ruijter, 1995; Weickenmeier et al., 1995). Alternative input signals include holographic fringes (de Ruijter and Weiss, 1992) and amorphous carbon (van Zwet and Zandbergen, 1996). In the stochastic or 'noise' method (de Ruijter and Weiss, 1992; de Ruijter, 1995; Zuo, 1996; Weickenmeier et al., 1995), the camera is evenly illuminated so that the input intensity is white noise. Due to blurring effects, the intensities measured in two pixels are correlated if the distance between the pixels is small and thus the noise power spectrum is no longer white, but attenuated at high spatial frequencies. This attenuation of the noise signal is usually assumed to be equal to the attenuation of a deterministic signal, and a rotationally averaged magnitude of the Fourier transform of a uniformly illuminated image is generally presented as the MTF measured by the noise method.

In general, however, the output noise of a stochastic scattering process is not simply the product of the input noise and the MTF (Rabbani et al., 1987). Therefore a new theoretical description is introduced in this work that clearly distinguishes

between the transfer of signal (*i.e.* spatial intensity variation) and the transfer of the inevitable Poisson noise into the recorded image by defining a separate Noise Transfer Function (NTF). This theoretical model makes the direct evaluation of both transfer properties with Monte-Carlo simulations possible (Meyer and Kirkland, 1998). In addition, novel techniques have been developed to measure both properties accurately (Meyer and Kirkland, 2000) and a range of commercially available cameras have been tested (Meyer and Kirkland, 2000; Meyer et al., 2000a). These experiments, in agreement with simulations, showed that at electron energies of 200 keV and above, the NTF exceeds the MTF by up to a factor of 4, which implies that the DQE at high spatial frequencies drops to 7% of its zero spatial frequency value. This problem is now being recognised by camera manufactures and referred to as the 'high voltage problem'.[5]

## 2.2   Theory of signal and noise transfer

The aim of this section is to derive expressions for the MTF and NTF from the statistical properties of the electron and photon scattering processes in the conversion chain from the scintillator surface, on which the primary electrons impinge, to the CCD chip, where electron-hole pairs are generated. Up to this step, the system is translationally invariant, which greatly simplifies the analysis. Therefore the last step, *i.e.* the collection of the electron in the pixelated CCD potential wells is treated separately and the transfer functions up to this step are denoted MTFS, NTFS and DQES to distinguish them from the overall transfer functions MTF, NTF and DQE, which include pixelation effects (figure 2.3).

In an electron microscope, the specimen scatters the incident electron beam, causing spatial variations of the electron wavefunction $\Psi$. The aim of an image detecting device is to yield an accurate representation of this spatial variation of the electron detection probability $|\Psi|^2$ in the image plane.

---

[5]Paul Mooney, Gatan, personal communication.
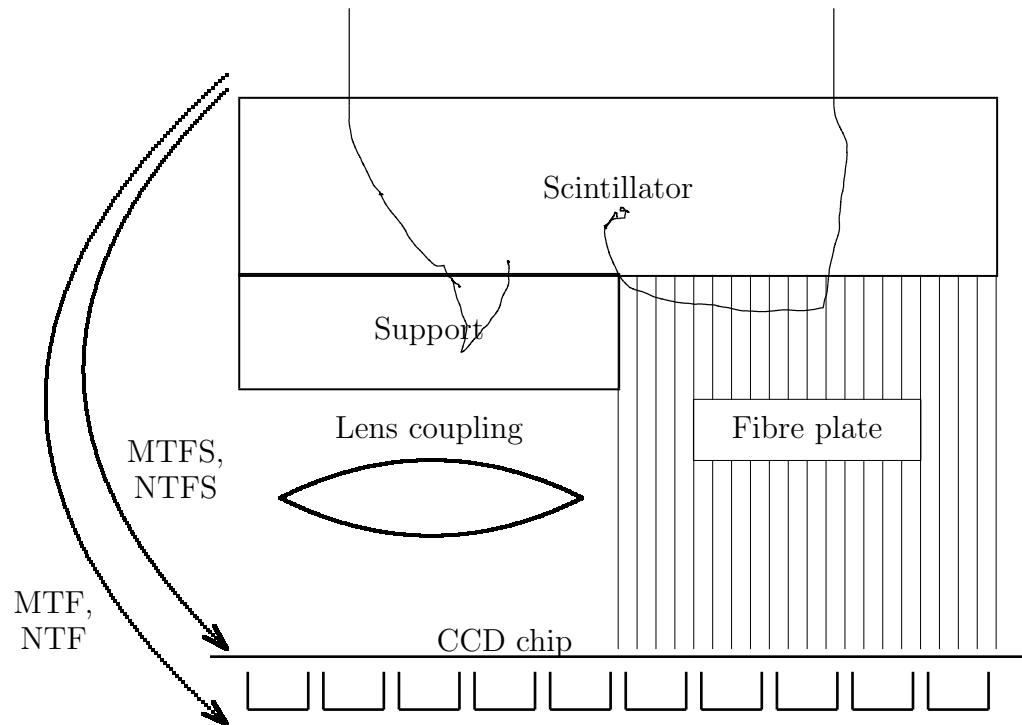
Figure 2.3: Schematic diagram of a CCD camera for electron detection. The incoming high-energy electrons are scattered and, along their trajectory within the scintillator, they give rise to photon emission. Electrons that penetrate through the scintillator may be back-scattered into the scintillator from the fibre plate or a mechanical support layer. A lens-or fibre optical coupling system conveys the generated light to the CCD chip, where the photons generate electron-hole pairs. The electrons of these pairs are collected in pixelated potential wells. This integrated sampling is treated separately here and the functions MTFS and NTFS describe the signal and noise transfer by the scintillator and the support up to the point of generation of electron-hole pairs.

In principle, electron detection in the camera is a quantum-mechanical position measurement process. However, as the scintillator is a large macroscopic object, decoherence occurs rapidly, which collapses the electron wavefunction into a localised wavepacket as soon as the electron enters the scintillator. For all practical purposes it is therefore sufficient to regard the electrons as classical particles that impinge on the scintillator, but with the proviso that the events of electron arrival are fundamentally random with the probability density for arrival at position $(x, y)$ given by $p(x, y) \propto |\Psi(x, y)|^2$.

Due to this statistical nature of quantum mechanical observations, variations in the output intensities $I_o(x, y)$ recorded in different exposures of identical duration taken of the same object are always present, even with an ideal camera. Only in the limit of averaging over an infinite number of exposures recorded by an ideal camera, the intensity converges to the probability distribution $|\Psi|^2$.

As the important properties of the camera do not depend on the position, but strongly on the spatial frequency of object detail, it is most useful to investigate the Fourier transform $\hat{I}_o(u, v)$ of the recorded intensity. For each $(u, v)$ there are variations of $\hat{I}_o(u, v)$ between different exposures, thus $\hat{I}_o(u, v)$ can be described as a complex random variable whose expectation value $E(u, v)$ describes the signal in the image and whose variance $V(u, v)$ describes the noise. The noise, i.e the random fluctuations from the expectation value, limits the amount of information that can be obtained from a single exposure.

Expressions for $E(u, v)$ and $V(u, v)$ will now be derived based on an analysis of the statistical properties of the contributions of individual electrons to the output signal $I_o$.

The trajectories of electrons hitting the scintillator at the origin can be described as members of a probability space $M$. Each element $\mu \in M$ describes the event of one incident electron, *i.e.* a particular trajectory with photons emitted at particular positions in particular directions. The probability for this event is $d\mu$, and the contribution of this event to the output signal is denoted $g_\mu(x, y)$. For an electron that

hits the scintillator $(x', y')$ rather than the origin, the trajectory is simply translated, contributing $g_\mu(x - x', y - y')$ to the output signal. Hence the contribution to $\hat{I}_o(u, v)$ is given by:

$$\Delta \hat{I}_o(u, v) = e^{-2\pi i(ux' + vy')} \hat{g}_\mu(u, v) \tag{2.1}$$

An infinitesimal scintillator area $dA$ at $(x, y)$ therefore contributes to $E(u, v)$ and $V(u, v)$ by

$$
\begin{aligned}
dE(u, v) &= (1 - p(x, y)dA) \cdot 0 \, + \, p(x, y)dA \int e^{-2\pi i(ux + vy)} \hat{g}_\mu(u, v)d\mu \\
&= p(x, y)e^{-2\pi i(ux + vy)} dA \int \hat{g}_\mu(u, v) \, d\mu
\end{aligned}
\tag{2.2}
$$

and

$$
\begin{aligned}
dV(u, v) &= (1 - p(x, y)dA) \cdot |0 - dE(u, v)|^2 \\
&+ \, p(x, y)dA \int |e^{-2\pi i(ux + vy)} \hat{g}_\mu(u, v) - dE(u, v)|^2 d\mu \\
&= p(x, y)dA \int |\hat{g}_\mu(u, v)|^2 d\mu \, + \, O(dA^2) \, .
\end{aligned}
\tag{2.3}
$$

The first and second term in each equation represent the cases of zero and one electron impinging on $dA$, respectively. The contributions of electrons impinging on mutually disjunct scintillator areas to the output signal are statistically independent, which makes both the expectation value and the variance additive, hence $E(u, v)$ and $V(u, v)$ can be evaluated by integration over the complete scintillator area:

$$E(u, v) = \hat{p}(u, v) \int \hat{g}_\mu(u, v) \, d\mu \tag{2.4}$$

$$V(u, v) = \hat{p}(0, 0) \int |\hat{g}_\mu(u, v)|^2 d\mu \, . \tag{2.5}$$

It is instructive to follow the discussion with two model examples:

1. An ideal camera: In this case, each electron generates a sharp signal exactly at its incident position, hence $g_\mu(x, y) = \delta(x, y)$, with no dependence on $\mu$. This means that $\hat{g}_\mu(u, v) \equiv 1$ and therefore

$$E(u, v) = \hat{p}(u, v) \; \text{and} \; V(u, v) = \hat{p}(0, 0) \, . \tag{2.6}$$

The variance of each Fourier component is therefore equal to the total electron dose, independent of the spatial frequency and hence the noise is pure white noise.

2. A randomly scattering camera: In this (practically useless) camera, each electron is scattered at random before detection and generates a sharp signal at a position entirely unrelated to its incident position. This is described by $g_\mu(x, y) = \delta(x - x_\mu, y - y_\mu)$ with $d\mu = dx_\mu dy_\mu / A$, where $A$ is the total scintillator area. The probability space $M$ is a simple one in this case, with each trajectory $\mu$ fully described by the displacement vector $(x_\mu, y_\mu)$ and with equal probabilities for all displacements within the scintillator. The result is $\hat{g}_\mu(u, v) = e^{-2\pi i(ux_\mu + vy_\mu)}$ and therefore $E(u, v)$ is zero for all non-zero spatial frequencies, whereas $V(u, v)$ is the same as for the ideal camera.

### 2.2.1 The modulation transfer function

The MTFS is defined as the ratio $E(u, v)/\hat{p}(u, v)$ of the output to the input signal in reciprocal space, normalised to unity at zero spatial frequency, hence

$$\text{MTFS}(u, v) = \int \hat{g}_\mu(u, v) \, d\mu / G \tag{2.7}$$

This function is real as long as the scattering is isotropic. $G$, defined by

$$G = \int \hat{g}_\mu(0, 0) \, d\mu \tag{2.8}$$

represents the total gain of the scintillator, *i.e.* the average number of detectable CCD well electrons per primary electron.

The inverse Fourier transform of the MTFS is the Point Spread Function (PSF)

$$\text{PSF}(x, y) = \int g_\mu(x, y) \, d\mu / G \tag{2.9}$$

given by the average intensity generated by many electrons hitting the scintillator at the origin.

## 2.2.2 The noise transfer function

The NTFS is defined as the square root of the ratio between output and input variance, normalised to unity at zero spatial frequency. The input variance can be calculated as the output variance of an ideal detector with unit gain, *i.e.* with $\hat{g}_\mu(u,v) \equiv 1$, hence

$$V_{in}(u,v) = \hat{p}(0,0) \tag{2.10}$$

The NTFS is therefore given by:

$$\text{NTFS}(u,v) = \sqrt{\frac{\int |\hat{g}_\mu(u,v)|^2 \mathrm{d}\mu}{\int |\hat{g}_\mu(0,0)|^2 \mathrm{d}\mu}} \ , \tag{2.11}$$

where the integral in the denominator ensures normalisation to unity at zero spatial frequency.

This expression is different from the MTFS, in that the NTFS represents an average over the power spectra of the intensity distribution generated by individual electrons. Hence, whereas the MTFS is related to the extent of the intensity distribution of many electrons with the same incident point, the NTFS is related to the average area of the intensity distribution caused by single electrons. This is exemplified by the random scattering camera, where the NTFS is one for all frequencies whereas the MTFS is zero for all non-zero frequencies.

## 2.2.3 The spatial frequency dependent DQES

The DQES is defined in terms of the signal to noise ratios of input and output signal:

$$\text{DQES}(u,v) = \frac{\text{SNR}_{out}^2(u,v)}{\text{SNR}_{in}^2(u,v)} \ . \tag{2.12}$$

The square of the output SNR is given by:

$$\text{SNR}_{out}^2(u,v) = \frac{(E(u,v))^2}{V(u,v)} \tag{2.13}$$

The input SNR is the $\text{SNR}_{out}$ of an ideal detector, hence the DQES is:

$$\text{DQES}(u,v) = \frac{|\int \hat{g}_\mu(u,v)\,\mathrm{d}\mu|^2}{\int |\hat{g}_\mu(u,v)|^2 \mathrm{d}\mu} \ . \tag{2.14}$$

This equation shows that the DQES can never be larger than one and is one for a given frequency iff $\hat{g}_\mu(u,v)$ is independent of $\mu$, *i.e.* if there is no stochastic component in the signal conversion.[6]

Equation 2.14 can also be re-written in terms of experimentally accessible quantities:

$$\text{DQES}(u,v) = \frac{\hat{p}(0,0)G^2(\text{MTFS}(u,v))^2}{V(u,v)} \; , \tag{2.15}$$

where $\hat{p}(0,0)$ is the total number of electrons recorded in each exposure, $V(u,v)$ is the variance, here in units of 'number of CCD well electrons squared', and $G$ is the gain, i.e the average number of CCD well electrons generated per incident electron. It should be noted that, unless the dose is very low, such that dose independent noise sources such as readout noise become important, the observed variance $V(u,v)$ is proportional to the electron dose and hence the DQES is dose independent. A simple rescaling of units into the digital numbers $(DN)$ that are read out from the camera finally gives the expression,

$$\text{DQES}(u,v) = \frac{I_{DN}G_{DN}(\text{MTFS}(u,v))^2}{V_{DN}(u,v)} \; , \tag{2.16}$$

where $G_{DN}$ is the gain in digital numbers per incident electron, $I_{DN} = \hat{p}(0,0)G_{DN}$ is the total intensity in digital numbers and $V_{DN}(u,v)$ is the variance in $(DN)^2$. Usually, rotational symmetry can be assumed and it is furthermore advantageous to introduce the dose independent Noise Power Spectrum $\text{NPS}(k) = V_{DN}(k)/I_{DN}$ so that the DQES becomes

$$\text{DQES}(k) = \frac{G_{DN}(\text{MTFS}(k))^2}{\text{NPS}(k)} \; . \tag{2.17}$$

Both the MTFS and the NPS can be measured experimentally and are corrected for sampling effects, as discussed in sections 2.4 and 2.5.

---

[6]Proof: with $\hat{g} = \int \hat{g}_\mu \mathrm{d}\mu$, the proposition follows from $0 \leq \int |\hat{g}_\mu - \hat{g}|^2 \mathrm{d}\mu = \int \hat{g}_\mu^* \hat{g}_\mu \mathrm{d}\mu - 2\hat{g}^*\hat{g} + \hat{g}^*\hat{g} = \int |\hat{g}_\mu|^2 \mathrm{d}\mu - |\hat{g}|^2$, where the equality holds iff $\hat{g}_\mu = \hat{g} \; \forall \mu$.

## 2.2.4 Calculating the transfer properties by Monte-Carlo integration

The output signal of the image conversion chain in the CCD-camera is a continuous distribution of CCD well electrons, hence $I_o(x, y)$ consists of $\delta$-functions at the positions where CCD well electrons are generated. These diffuse into pixelated potential wells, but this pixelation is treated separately, as it destroys the translational invariance of the problem.

A CCD well electron generated at position $(x_p, y_p)$ in the image plane, contributes $\delta(x - x_p, y - y_p)$ to the output signal collected and hence $\exp(-2\pi\mathrm{i}(x_p u + y_p v))$ to its Fourier transform $\hat{g}(u, v)$. Thus

$$\hat{g}_\mu(u, v) = \sum_p \mathrm{e}^{-2\pi\mathrm{i}(x_p u + y_p v)} \ , \tag{2.18}$$

where the summation extends over all detected photons that are generated from a primary electron with trajectory $\mu$. The integrals $\int \hat{g}(u, v)\mathrm{d}\mu$ and $\int |\hat{g}(u, v)|^2\mathrm{d}\mu$ can be approximated by taking the average of $\hat{g}(u, v)$ and $|\hat{g}(u, v)|^2$ over a large number of simulated trajectories. This can only be performed for a limited number of $(u, v)$-values. In this work, the averaging was performed at discrete and equidistant sampling points along the positive $u$-axis in Fourier space, which is sufficient because of the rotational symmetry. It should be noted that the Fourier transform $\hat{g}(u, v)$ evaluated this way is not impaired by aliasing, as $(x_p, y_p)$ are continuous co-ordinates and at this stage, no sampling in real space has taken place.

## 2.2.5 Sampling effects

The electrons of the electron–hole pairs generated in the CCD-chip are collected in relatively sharp sided, pixelated potential wells.[7] Mathematically, this can be reasonably described as a convolution with a top-hat function

$$I_p(x, y) = I_o(x, y) \otimes \left[\chi_{[-1/2, 1/2]}(x)\chi_{[-1/2, 1/2]}(y)\right] \tag{2.19}$$

---

[7]In the following, all distances and spatial frequencies are given in units of the pixel spacing and the sampling frequency, respectively

followed by a multiplication with an infinite 2-D array of $\delta$-functions

$$I_{CCD}(x,y) = I_p(x,y) \cdot \left[ \sum_{m,n} \delta(x-m)\delta(y-n) \right] . \qquad (2.20)$$

The first operation represents the integration over the pixel area, such that for any $(x,y)$, $I_p(x,y)$ is the average value of $I_o$ in the unit square centred at $x,y$. The second operation represents the sampling, discarding all information between the discrete sampling points $(m,n)$ with integer co-ordinates. In the Fourier domain, this is equivalent to a multiplication with a sinc-function

$$\hat{I}_p(u,v) = \hat{I}_o(u,v)\text{sinc}(\pi u)\text{sinc}(\pi v) \qquad (2.21)$$

followed by a convolution with an array of $\delta$-functions

$$\hat{I}_{CCD}(u,v) = \hat{I}_p(u,v) \otimes \left[ \sum_{m,n} \delta(u-m)\delta(v-n) \right] . \qquad (2.22)$$

The latter convolution leads to the well-known aliasing effect, which makes the distinction between frequencies that differ by an integer vector $(m,n)$, impossible. The applicability of the MTF concept to sampled systems has been the subject of some discussion (de Ruijter, 1995, and references therein). Clearly, the apparent amplitude of a sampled sinusoidal function is dependent on the positions of the maxima and minima with respect to the sampling points. However, the sampling theorem (Jerri, 1977) implies that as long as the input signal is bandwidth-limited to frequencies below the Nyquist limit, *i.e.* $(u,v) \in [-1/2,1/2][-1/2,1/2]$, the intensity at any point in the image can be retrieved from the values at the discrete sampling points. Equation 2.22 shows that the sampling does not alter the amplitude of isolated Fourier components, it only copies them across to different frequencies. This implies that the MTF is well defined by the output/input ratio of the Fourier transform components. The MTF concept is therefore fully applicable to sampled imaging systems (de Ruijter, 1995). The MTF can even be sensibly defined for spatial frequencies above the Nyquist limit as the amplitude of the output signal divided by the amplitude of a sinusoidal input

signal. However, it is only possible to distinguish contributions from above and below the Nyquist limit in specialised situations using prior knowledge about the image. Relevant examples are images of skewed sharp edges (section 2.4.1), undersampled images of crystals (Koeck, 1998) and electron holograms (Ishizuka, 1993). According to equation 2.21 the total MTF is given by

$$\text{MTF}(u, v) = \text{MTFS}(u, v)\text{sinc}(\pi u)\text{sinc}(\pi v), \qquad (2.23)$$

with the additional caveat that frequencies above the Nyquist limit are aliased to below the Nyquist limit.

**The effect of sampling on the noise transfer**

Distinction between noise from below and above the Nyquist limit is not possible. As the contributions from different spatial frequencies mapped together by aliasing are independent, their variances are additive. Therefore, the variance of the sampled CCD image is:

$$V_{CCD}(u, v) = \sum_{m,n} V(u + m, v + n)\text{sinc}^2(\pi(u + m))\text{sinc}^2(\pi(v + n)) \qquad (2.24)$$

This immediately yields the total noise transfer function

$$\text{NTF}(u, v) = \left\{ \sum_{m,n} \text{NTFS}^2(u + m, v + n)\text{sinc}^2(\pi(u + m))\text{sinc}^2(\pi(v + n)) \right\}^{1/2} \qquad (2.25)$$

and the DQE

$$\text{DQE}(u, v) = \text{DQE}(0, 0)\frac{\text{MTF}^2(u, v)}{\text{NTF}^2(u, v)} , \qquad (2.26)$$

where $\text{DQE}(0, 0) = \text{DQES}(0, 0)$ provided that the pixel width is equal to the pixel spacing, as assumed in this work.
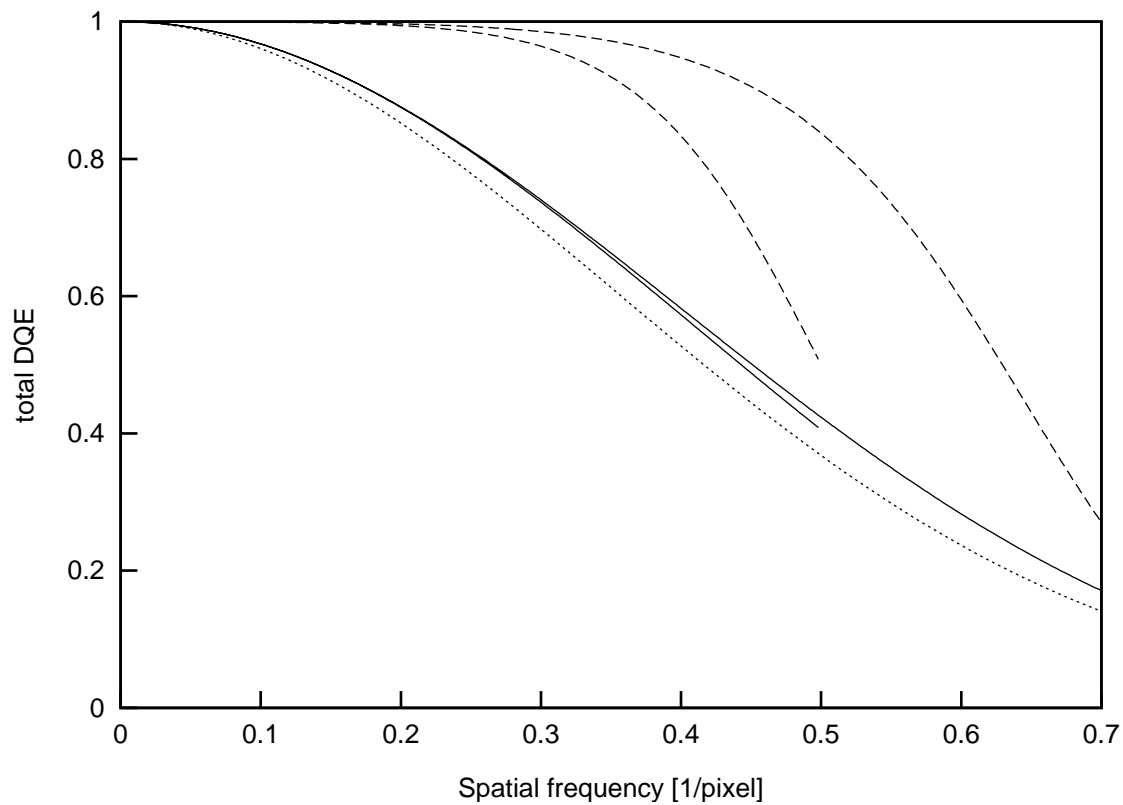
Figure 2.4: The DQE of an 'ideal' camera with independent pixels, plotted for axial and diagonal directions (solid lines), is substantially smaller than 1 due to noise aliasing. The dashed lines show how the DQE is improved when deterministic blurring (in this example: $\text{MTFS}(k) = \text{NTFS}(k) = \mathrm{e}^{-4k^2}$, dotted line) is introduced into the image chain before the sampling takes place.

**The DQE for a camera with no spreading in the scintillator**

An application of equation 2.25 to the ideal camera (MTFS $\equiv 1$ and NTFS $\equiv 1$) yields:

$$\text{NTF}(u,v) = \left[\sum_m \text{sinc}^2(\pi(u+m))\right]\left[\sum_n \text{sinc}^2(\pi(v+n))\right] \equiv 1 \;, \qquad (2.27)$$

using the identity

$$\sum_n \text{sinc}^2\pi(n+k) = \frac{1}{\pi^2}\sum_n \frac{\sin^2(\pi(n+k))}{(n+k)^2} = \frac{\sin^2\pi k}{\pi^2}\sum_n \frac{1}{(n+k)^2} = 1 \;. \qquad (2.28)$$

with the standard sum (Riley et al., 1997, p 661)

$$\sum_n \frac{1}{(n+k)^2} = \frac{\pi^2}{\sin^2\pi k} \;. \qquad (2.29)$$

The total NTF of the ideal camera is hence identical 1, in agreement with the fact that its pixels are independent of each other, giving rise to white noise in the discrete Fourier transform. However, as the signal transfer is attenuated due to pixel integration according to equation 2.23, the resultant DQE (figure 2.4) is substantially smaller than 1:

$$\text{DQE}(u,v) = \text{sinc}^2(\pi(u+m))\text{sinc}^2(\pi(v+n)) \;. \qquad (2.30)$$

The additional noise that gives rise to this lowered DQE can be understood as quantisation noise, which occurs when an analogue entity (incident position) is converted into a digital one (pixel number). This detrimental effect of the sampling on the DQE can be avoided when both MTFS and NTFS drop rapidly above the Nyquist limit. This necessitates a degree of deterministic blurring in the imaging chain. Figure 2.4 shows an example where $\text{MTFS}(k) = \text{NTFS}(k) = \text{e}^{-4k^2}$. This demonstrates that it is beneficial for pixelated systems when blurring with a point spread function width comparable to the pixel size occurs between amplification and sampling. This spreads the signal from a single detected particle over a small number of pixels and provides information about the incident position at sub-pixel accuracy, reducing the

quantisation noise. However, in CCD cameras, the blurring occurs partly before the amplification due to the random scattering of the incident electrons, leading to an overall DQE that is substantially inferior to an ideal pixelated camera.

**The effective pixel number**

As a useful single figure of merit for pixelated image detectors, an "effective pixel number" can be defined by:

$$N_{\text{eff}} = \sum_{\mathbf{k}} \text{DQE}(\mathbf{k}) \ , \tag{2.31}$$

where the sum extends over all pixels in the discrete Fourier transform. This figure of merit gives a good comparator of how much information can be obtained from a single image in a given exposure time in spite of the presence of noise. When the number of pixels is doubled by doubling the array size in one dimension, the effective pixel number appropriately doubles. If instead the pixel number is doubled by reducing the pixel size to half in one dimension, the increase in the effective pixel number is less than a factor of 2. For a camera where the MTF and hence the DQE is already zero beyond the Nyquist limit, this doubling of the pixel number would only add 'dead resolution', reflected in an effective pixel number that remains constant. This figure of merit also implies that doubling the average DQE is as valuable as doubling the camera area by doubling the pixel number. This is a sensible assertion for noise and dose limited applications, since doubling the DQE allows the use of half the exposure time to get the same signal to noise ratio. Hence the smaller camera can capture the same information as the larger camera in the same time by using two subsequent exposures of adjacent areas.

As noted in section 20, the DQE of a camera for which both the MTFS and the NTFS are identical 1, is substantially smaller than 1, and for this camera, $N_{\text{eff}} = 0.60 N_{\text{tot}}$. An effective pixel number equal to the total pixel number $N_{\text{tot}}$ could only be achieved with a camera where the MTFS and NTFS are equal and large below the Nyquist limit and vanish above the Nyquist limit. This is nearly impossible to

achieve, however, in the example with $\mathrm{MTFS}(k) = \mathrm{NTFS}(k) = \mathrm{e}^{-4k^2}$ discussed in section 20, $N_{\mathrm{eff}} = 0.84 N_{\mathrm{tot}}$, already a substantial improvement. In section 2.6, the effective pixel number is calculated for commercially available cameras, giving values that are in many cases below 10% of the actual pixel number.

## 2.3   The Monte-Carlo model

As discussed in the previous section, the signal and noise properties of the image conversion chain can be evaluated by Monte-Carlo simulations. To this end, the trajectories of a large number of electrons incident at the origin have to be calculated. For each electron, all processes up to the generation of CCD well electrons have to be simulated, with random numbers used for each stochastic process involved. For each electron,

$$\hat{g}_\mu(u, v) = \sum_p \mathrm{e}^{-2\pi\mathrm{i}(x_p u + y_p v)} \ , \tag{2.32}$$

where $(x_p, y_p)$ are the positions where CCD well electrons are generated. This $\hat{g}_\mu(u, v)$ must be calculated at suitable sampling points $(u, v)$. The averages of $\hat{g}_\mu(u, v)$ and $|\hat{g}_\mu(u, v)|^2$ at these sampling points over a large number of simulated trajectories approximate the integrals $\int \hat{g}(u, v)\mathrm{d}\mu$ and $\int |\hat{g}(u, v)|^2\mathrm{d}\mu$ that are needed to evaluate the signal and noise transfer properties.

### 2.3.1   The electron scattering model

The electron scattering model used in the simulations follows the description by Joy (1995). In this model electron trajectories are straight lines between discrete elastic Rutherford (1911) scattering events. Inelastic scattering is modelled by assuming a continuous energy loss according to the Bethe (1930) formula along these trajectories, neglecting the usually small directional change in inelastic scattering events. The modifications to Joy's original model used here and described subsequently include:

1. More accurate relativistic corrections to the Rutherford formula for elastic scattering are incorporated (section 2.3.1).

2. The relativistic Bethe formula is used instead of the non-relativistic one (section 2.3.1).

3. The generation of fast secondary electrons is modelled by including both the path of the secondary electron and the directional change of the first electron. This means that large-angle electron-electron scattering is treated as a discrete scattering event and is removed from the continuous Bethe energy loss (section 2.3.1).

**Discrete scattering events**

When the total scattering cross section per atom is $\sigma$, the mean free path $\lambda$ between two scattering events in a material with atomic weight $A$ and density $\rho$ is given by

$$\lambda = \frac{A}{N_a \rho \sigma} \ . \tag{2.33}$$

As scattering events are mutually independent, the actual path length is governed by a negative exponential distribution with mean $\lambda$, which can be implemented by choosing the path length according to

$$l = -\lambda \ln(\text{RND}) \ , \tag{2.34}$$

where RND is a random number betweeen 0 and 1. A scattering event is further characterised by the scattering angle $\theta$ and the azimuth $\phi$. Figure 2.5 illustrates how the unit direction vector $\mathbf{c}_n$ after the scattering event $n$ can be determined. The first step is to evaluate the unit vectors $\mathbf{u}$ and $\mathbf{v}$ from the initial direction $\mathbf{c} = \mathbf{c}_{n-1}$:

$$\mathbf{u} = \frac{c_z^2}{1 - c_z^2} \begin{pmatrix} c_x \\ c_y \\ -\frac{c_x^2 + c_y^2}{c_z} \end{pmatrix} \qquad \text{and} \qquad \mathbf{v} = \mathbf{c} \times \mathbf{u} \ . \tag{2.35}$$
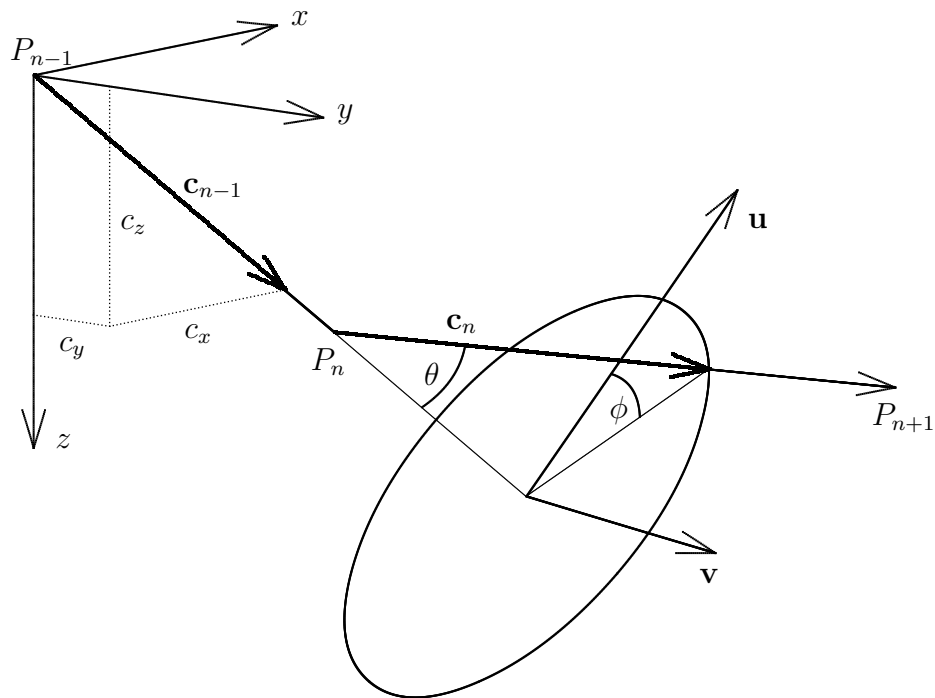
Figure 2.5: The modelled electron trajectory consists of straight lines between scattering events at the points $P_i$. The momentary direction after the $n$-th scattering event is described by a unit vector $c_n$. Each scattering event is characterised by a scattering angle $\theta$, an azimuth angle $\phi$ and the free path length $l_{n-1}$ since the last scattering event at $P_{n-1}$.

It can be shown by inspection that $\mathbf{c}$, $\mathbf{u}$ and $\mathbf{v}$ are orthonormal and that $\mathbf{u}$ is in the plane defined by $\mathbf{c}$ and the $z$-axis. In terms of these vectors, the new direction $\mathbf{c}' = \mathbf{c}_n$ can be expressed as

$$\mathbf{c}' = \mathbf{c}\cos\theta + \mathbf{u}\sin\theta\cos\phi + \mathbf{v}\sin\theta\sin\phi . \tag{2.36}$$

## Elastic scattering

The main process leading to directional changes in the path of the electron is elastic scattering by the nuclei. In the non-relativistic limit for a naked nucleus of charge $Ze$ this process is described by the Rutherford cross-section (Rutherford, 1911):

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left( \frac{Ze^2}{2 \cdot 4\pi\epsilon_0 \, mv^2} \right)^2 \frac{1}{\sin^4(\theta/2)} . \tag{2.37}$$

The cross section diverges for $\theta \to 0$ and yields an infinite total cross section, which is due to the infinite range of the coulomb interaction. When the field of the nucleus is screened by the orbiting electrons, the cross section becomes finite. Relativistic effects can be crudely incorporated by using the relativistic mass $m = m_0/\sqrt{1 - \beta^2}$ with $\beta = v/c$. These modifications yield:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left( \frac{Ze^2}{2 \cdot 4\pi\epsilon_0 \, m_0c^2\beta^2v^2} \right)^2 \frac{1 - \beta^2}{(\sin^2(\theta/2) + \alpha_s)^2} \tag{2.38}$$

$$= \left( \frac{Ze^2}{2 \cdot 4\pi\epsilon_0 \, E_k^2} \right)^2 \frac{(E_k + E_0)^2}{(E_k + 2E_0)^2} \frac{1}{(\sin^2(\theta/2) + \alpha_s)^2} , \tag{2.39}$$

where $E_0 = m_0c^2$ is the rest energy and $E_k = E_0/\sqrt{1 - \beta^2} - E_0$ is the kinetic energy of the electron. The screening parameter $\alpha_s$ can be evaluated using an analytical approximation (Bishop, 1976):

$$\alpha_s = 3.4 \times 10^{-3} Z^{0.67} \frac{\mathrm{keV}}{E} . \tag{2.40}$$

For high energies, the relativistic correction given in equation 2.38 is no longer sufficient. Instead, the Dirac equation for an electron in the Coulomb field of the nucleus has to be solved. This has been performed by Mott (1929, 1932), giving an
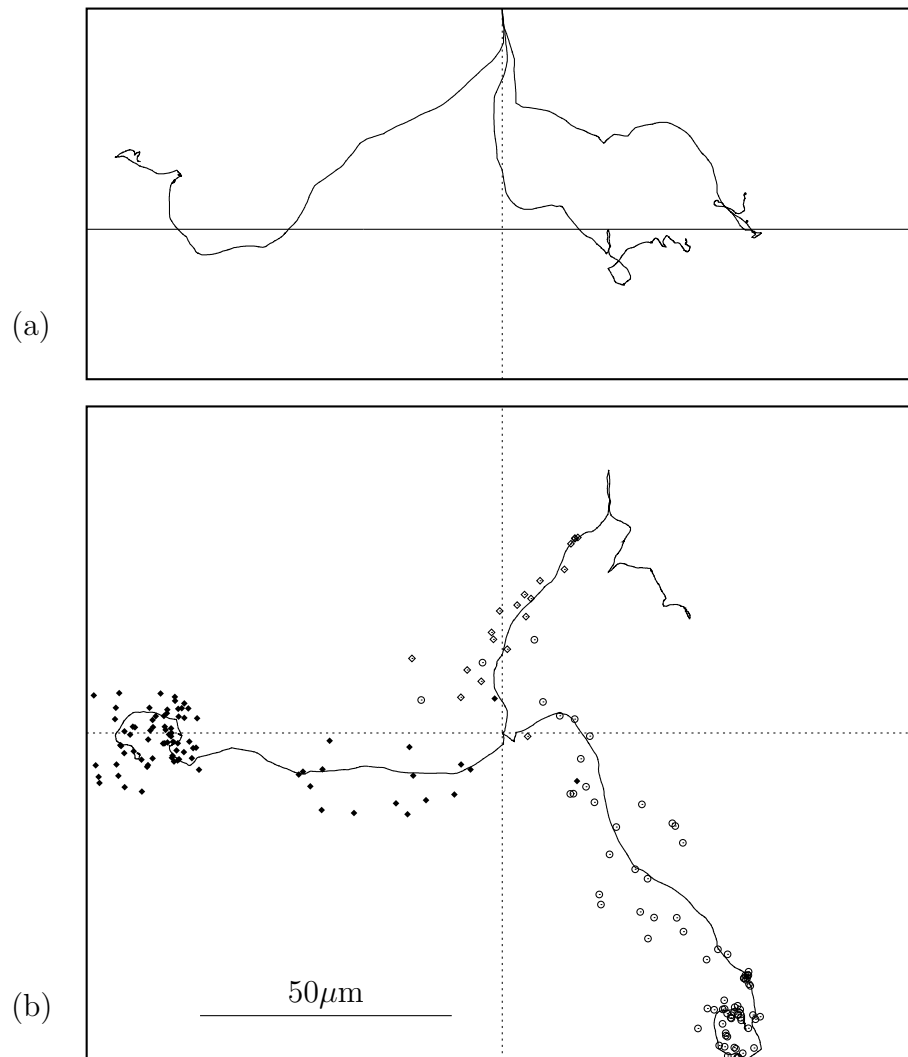
Figure 2.6: Three trajectories of 200kV electrons in a glass-supported $50\mu$m thick YAG scintillator in (a) $x-z$-projection and (b) $x-y$-projection. In (b), the positions where photons are detected on the CCD-chip are also marked, with different symbols for the three trajectories. As the stopping power increases with decreasing energy, localised spots of illumination are generated where an electron is stopped in the scintillator. The light generation can occur at a large lateral displacement from the original incident point, especially if the electron is stopped after being back-scattered from the supporting layer. This effect leads to a strong discrepancy between MTF and NTF, as the former is determined by the area illuminated by many electrons incident at the same point, while the latter is determined by the average area illuminated by a single electron.

infinite series in powers of $\alpha_Z = Z\alpha$, where $\alpha \approx 1/137$ is the fine structure constant. When terms above the second order in $\alpha_Z$ are ignored, the result can be expressed in the form of equation 2.38 multiplied by a correction factor (Birkhoff, 1958, p. 103)

$$R = 1 - \beta^2 \sin^2\left(\frac{\theta}{2}\right) + \pi\alpha\beta\sin\left(\frac{\theta}{2}\right)\left[1 - \sin\left(\frac{\theta}{2}\right)\right] . \qquad (2.41)$$

Although this correction considerably reduces the cross-section for large angle scattering (by up to 50% at 300kV), the change in the resulting MTF and NTF was found to be less than 2%, as large angle scattering is sufficiently rare that the precise cross section for large angle scattering has little impact on the overall result. Nevertheless, the relativistically correct version of the Rutherford scattering formula was used in all simulations shown.

**Energy loss: The Bethe stopping power**

The main process by which an electron loses energy in the energy range in question ($1\text{keV} \leq E \leq 400\text{keV}$) is scattering by the electrons in the solid. Radiative loss by bremsstrahlung is far less important and ignored in the following, as the ratio of these two loss mechanisms is given by (Birkhoff, 1958, p. 63)

$$\frac{(\mathrm{d}E/\mathrm{d}x)_{\text{rad}}}{(\mathrm{d}E/\mathrm{d}x)_{\text{coll}}} = \frac{EZ}{800\text{MeV}} < 2.5\% \quad \text{for } E < 400kEv, \ Z < 50. \qquad (2.42)$$

A full quantum mechanical treatment of inelastic scattering of electrons by atoms was first performed by Bethe (1930), who treated atoms as assemblies of oscillators with frequencies $\omega_i$ and oscillator strengths $f_i$ with $\sum f_i = 1$. This yields the Bethe equation for the energy loss:

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{4\pi N_A e^4 \rho}{m_0 v^2 A} \ln\frac{m_0 v^2}{J} , \qquad (2.43)$$

where

$$J = \left(\prod_i (\hbar\omega_i)^{f_i}\right)^{1/Z} \qquad (2.44)$$

is the *average ionisation potential*, which has been measured experimentally and is tabulated. In this work, the values supplied with the program by Joy (1995) were

used. At low energies, the logarithmic term in equation 2.43 diverges and the formula becomes inaccurate. This can be overcome by adding $2c_j J$ with $c_j = 0.85$ to the numerator of this term (Gauvin and L'Esprance, 1992; Joy, 1995). At higher energies, a full relativistic treatment is necessary. Bethe carried out this calculation and found

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{2\pi N_A e^4 \rho}{m_0 v^2 A} \left[ \ln \frac{(m_0 v^2 + 2c_j)(E + c_j)\gamma^2}{2J^2} - (\frac{2}{\gamma} - \frac{1}{\gamma}) \ln 2 + \frac{1}{\gamma^2} + \frac{\gamma - 1}{8\gamma} \right] ,$$
(2.45)

with the usual abbreviations $\gamma = 1/\sqrt{1 - \beta^2}$ and $\beta = v/c$. Equation 2.45 has been amended with the low-energy correction mentioned above. In the simulations, $c_j = 0.85$ is used, while the original equation is recovered for $c_j = 0$.

**Fast secondary electron generation: Möller scattering**

Electron-electron scattering processes in which the primary electron loses an appreciable fraction of its energy lead to the generation of fast secondary electrons. The two electrons that emerge from this process are indistinguishable, but by convention, the one that has more energy is called the primary electron and the other one the fast secondary electron. The cross-section for a process with a fractional energy transfer is approximately $\propto 1/\varepsilon^2$ and hence processes with relatively small energy transfer are most important. These give rise to secondary electron emitted at nearly right angles to the primary electron trajectory and hence the fast secondary electrons can lead to a substantial broadening of the interaction volume for high-energy electrons in thin films (Joy et al., 1982). Non-relativistic formulae for the Monte-Carlo modelling of this process have been given (Joy et al., 1982; Joy, 1995). In this work, the relativistic model given in Gauvin and L'Esprance (1992) is used instead. It is based on the cross-section derived by Møller (1932) for the interaction of an energetic primary electron with a free electron at rest:

$$\frac{d\sigma}{d\epsilon} = \frac{2\pi}{m_0 c^2 \beta^2 E} \left( \frac{e^2}{4\pi\varepsilon_0} \right)^2 \left[ \frac{1}{\varepsilon^2} + \frac{1}{(1 - \varepsilon)^2} + \left( \frac{\tau}{\tau + 1} \right)^2 + \frac{2\tau + 1}{(\tau + 1)^2} \cdot \frac{1}{\varepsilon(1 - \varepsilon)} \right] ,$$
(2.46)

where $\varepsilon = \Delta E/E$ is the fractional energy transfer, $\varepsilon_0$ is the permittivity of free space and $\tau = E/m_0 c^2$. The total cross section for processes with an energy transfer larger than a threshold $\varepsilon_c$ is calculated by integrating equation 2.46:

$$\sigma = \int_{\varepsilon_c}^{1/2} \frac{\mathrm{d}\sigma}{\mathrm{d}\varepsilon}(\varepsilon)\mathrm{d}\varepsilon \ . \tag{2.47}$$

Similarly, a uniformly distributed random number RND can be converted into a correctly distributed random value for $\varepsilon$ by solving

$$\mathrm{RND} = \int_{\varepsilon_c}^{\varepsilon} \frac{\mathrm{d}\sigma}{\mathrm{d}\varepsilon}(\varepsilon')\mathrm{d}\varepsilon' \Big/ \int_{\varepsilon_c}^{1/2} \frac{\mathrm{d}\sigma}{\mathrm{d}\varepsilon}(\varepsilon')\mathrm{d}\varepsilon' \tag{2.48}$$

for $\varepsilon$ (Gauvin and L'Esprance, 1992). The polar angles $\theta_p$ and $\theta_s$ of the primary and fast secondary electron can be calculated from $\varepsilon$ as (Møller, 1932):

$$\sin^2 \theta_p = \frac{2\varepsilon}{2 + \tau(1 - \varepsilon)} \tag{2.49}$$

$$\sin^2 \theta_s = \frac{2(1 - \varepsilon)}{2 + \tau\varepsilon} \ . \tag{2.50}$$

When the generation of fast secondary electrons with energies greater than $\varepsilon E$ is treated separately, the contribution of this processes to the Bethe stopping power (equation 2.45) has to be subtracted from the stopping power used to calculate the continuous energy loss.

This contribution of fast secondary electron generation to the Bethe stopping power is

$$
\begin{aligned}
-\frac{\mathrm{d}E}{\mathrm{d}x}\bigg|_{\mathrm{FSE}} &= \frac{\rho N_A Z}{A} \int_{\varepsilon_c}^{1/2} \varepsilon E \frac{\mathrm{d}\sigma}{\mathrm{d}\varepsilon}\,\mathrm{d}\varepsilon \\
&= -\frac{2\pi\rho N_A Z}{m_0 c^2 \beta^2 A}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^2 \left[3\ln 2 - \frac{1}{8}\left(\frac{\tau}{\tau+1}\right)^2 + \frac{2\tau+1}{(\tau+1)^2}\ln 2 \right. \\
&\qquad \left. + \ln\varepsilon_c + \frac{\ln(1-\varepsilon_c)}{1-\varepsilon_c} + \frac{1}{2}\varepsilon_c^2\left(\frac{\tau}{\tau+1}\right)^2 + \frac{2\tau+1}{(\tau+1)^2}\ln(1-\varepsilon_c)\right] \ . \tag{2.51}
\end{aligned}
$$

## 2.3.2 Emission and scattering of photons

Photons emitted along the electron trajectories can reach the CCD chip directly or after reflection at the top scintillator surface. As the photons are generally not emitted in the focal plane of the optical system, the position where they are detected is displaced from the projected trajectory. For YAG and other single crystal scintillators an additional effect is important and must be considered. Due to its high refractive index the crystal acts as a waveguide for photons emitted at small angles to the scintillator surface. Some of these photons will eventually be scattered into the aperture of the optical system and give rise to intensity at a large lateral distance $d$ from the place where they were produced, which leads to a characteristic tail in the PSF.

Provided that the scattering of photons is isotropic and sufficiently rare that it is only significant for light confined in the scintillator waveguide a simple model can be used to calculate the additional intensity due to this effect. For a photon emitted or scattered into a random direction, the probability of reaching the entrance aperture without further scattering is

$$p_1 = \frac{1 + r_{\mathrm{us}}}{2}(1 - \cos\alpha_{\mathrm{aper}}) \; , \tag{2.52}$$

where $r_{\mathrm{us}}$ is the upper surface reflectivity for photons emitted into the upwards cone subtended by the aperture semiangle $\alpha_{\mathrm{aper}}$. The aperture semiangle is the acceptance angle in the scintillator material with refractive index $n_{\mathrm{scint}}$. It is related to the acceptance angle $\alpha_{\mathrm{air}}$ in air and the numerical aperture $N_A$ by

$$n_{\mathrm{scint}} \sin\alpha_{\mathrm{aper}} = \sin\alpha_{\mathrm{air}} = N_A \; , \tag{2.53}$$

hence $P_1$ is

$$p_1 = \frac{1 + r_{\mathrm{us}}}{2}\left(1 - \sqrt{1 - \frac{\sin^2\alpha_{\mathrm{air}}}{n_{\mathrm{scint}}^2}}\right) = \frac{1 + r_{\mathrm{us}}}{2}\left(1 - \sqrt{1 - N_A^2/n_{\mathrm{scint}}^2}\right) \; . \tag{2.54}$$

Similarly, the probability for the photon trajectory to be in a direction that leads to multiple internal reflections is

$$p_2 = \cos\alpha_{\mathrm{lim}} \; , \tag{2.55}$$

where $\alpha_{\text{lim}}$ is the limiting angle for total reflection at the scintillator/support interface. With these notations, the probability that a photon emitted in a random direction reaches the optical entrance aperture after exactly $n$ scattering events is $p_1 p_2^n$, so the ratio of the intensities reaching the aperture with and without scattering is

$$\frac{I_{\text{indirect}}}{I_{\text{direct}}} = \left( p_1 \sum_{n=1}^{\infty} p_2^n \right) \Big/ p_1 = p_2/(1 - p_2) . \tag{2.56}$$

For a YAG/glass interface,[8] $\alpha_{\text{lim}} = \arcsin(1.5/1.83) = 55°$, thus $p_2 = \cos \alpha_{\text{lim}} = 0.57$ and hence the above ratio is 1.35 for a glass-supported YAG. The long-range blurring associated with multiple total reflection and scattering cannot be avoided with a dielectric anti-reflection coating, as this has no influence on the total reflection. A light-absorbing coating on the top surface could be used to reduce the waveguide effect, albeit at the penalty that the upper surface reflectivity $r_{\text{us}}$, and hence the conversion efficiency, is also reduced.

Amorphous phosphors have been modelled by passing the electron at random between the active phosphor grains and inactive matrix, where the path length in each material is a random number with negative exponential distribution and a mean value given by the grain size and the fill factor.

The photon scattering in phosphor scintillators is modelled as a random walk with a mean free path for scattering events (which are supposed to be isotropic) and a different mean free path for absorbtion.

### 2.3.3 The Monte Carlo simulation program

Although the Monte Carlo simulation program used to predict the performance of CCD cameras by simulations is based on an original program by Joy (1995) a significant number of changes were made over time such that almost none of the original code remained unchanged. These significant changes include:

- Conversion to Visual C++ to make the program run under Microsoft Windows.

---

[8]Refractive indices $n_{\text{YAG}} = 1.83$ (Daberkow et al., 1991) and $n_{\text{glass}} = 1.5$ (Vogel, 1995).

- Design of a graphical user interface for entering parameters.

- Implementation of a batch program interpreter. This allows calculations with different values for the simulation parameters to be run as a batch job. The batch programming language is described in appendix B.

- The simulation of electron passage through layers of different materials (coating, scintillator, support).

- The simulation of photon trajectories.

- Integration of the detected light distribution to yield the MTFS and NTFS.

- Modifications to the electron scattering model as described in section 2.3.1.

### 2.3.4 Simulation results

The Monte Carlo simulation results presented in this section illustrate the influence of some of the simulation parameters and the underlying physical effects on the transfer properties of CCD camera.

**The pulse height distribution**

The distribution $p(g)$ of the number $g$ of photons produced by individual electrons is called the *pulse height distribution*. It is directly related to the gain $G$ and to DQE(0) by

$$G = \overline{g} \tag{2.57}$$

$$\text{DQE}(0) = \frac{\overline{g}^2}{\overline{g^2}} . \tag{2.58}$$

Daberkow et al. (1996) have measured this distribution for a 50 $\mu$m YAG scintillator on a glass support by illuminating it with a low-intensity electron beam, collecting
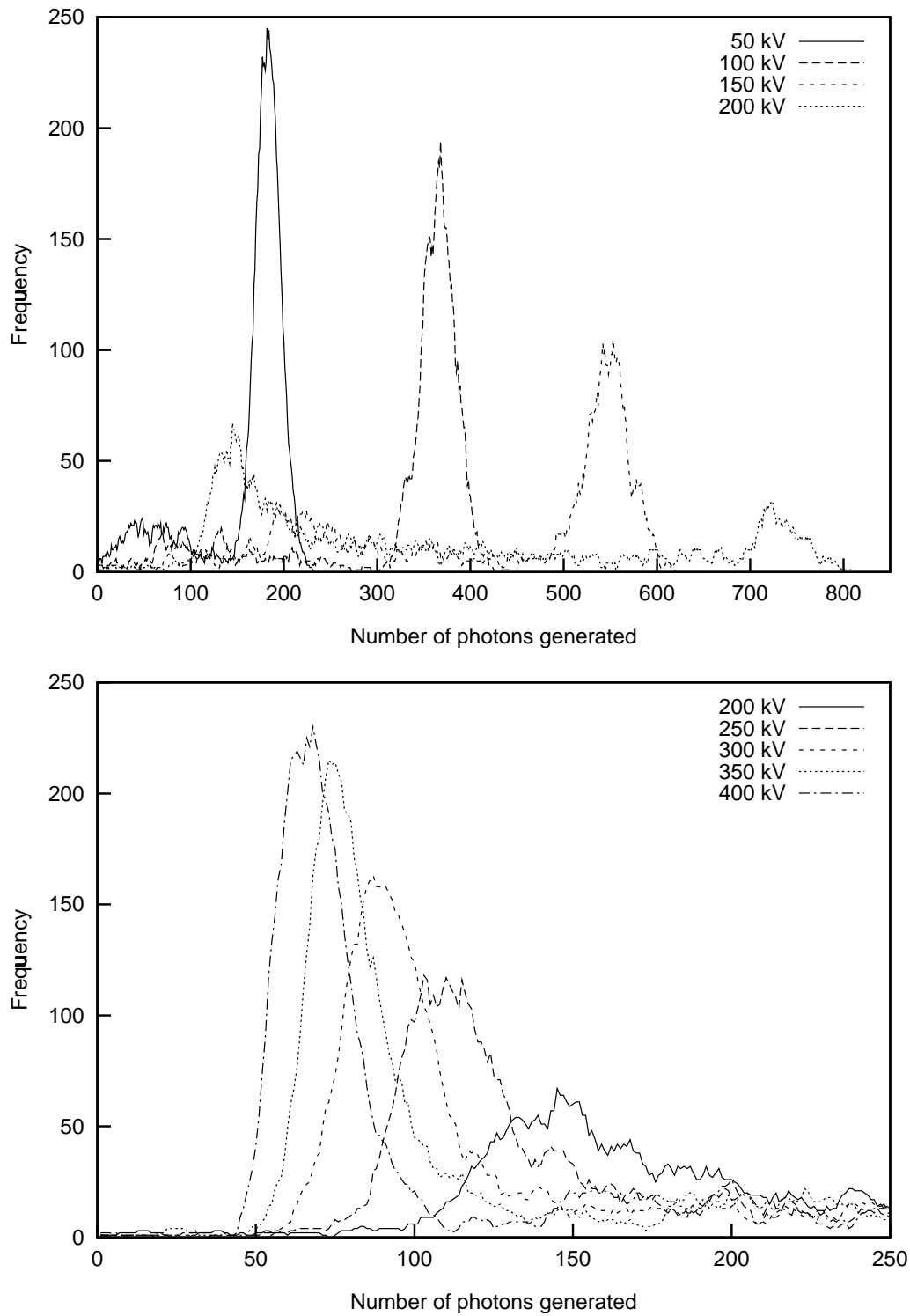
Figure 2.7: Distribution of the number of photons generated by individual electrons in an glass-supported 50 $\mu$m YAG scintillator.

the produced light with a photo multiplier and analysing the pulse heights of the multiplier output with a multi-channel analyser. Figure 2.7 shows simulation results for such a scintillator, which are in good agreement with the experimental findings in (Daberkow et al., 1996). Most electrons with low primary energy lose their entire energy in the scintillator. This gives rise to a pronounced peak in the distribution function at a number of generated photons proportional to the primary energy. Only the small fraction of electrons that are back-scattered from the scintillator give rise to events with lower photon numbers. At higher energies, electrons can penetrate through the scintillator, and as the energy increases, fewer electrons lose their entire energy in the scintillator, hence the peak at the associated photon number vanishes in figure 2.7 at energies larger than 200 kV. Instead another peak starts to emerge at much lower photon numbers, corresponding to electrons that pass through the scintillator on relatively straight path. As both the stopping power and the cross-section for large-angle scattering events decrease with increasing energy, this peak shifts towards lower photon numbers as the energy increases. There are, however, still events where more photons are produced. These are partly due to electrons that are scattered at a large angle, giving rise to a longer path length in the scintillator. The more important contribution arises from electrons that are back-scattered into the scintillator from the support layer, giving rise to a continuous spectrum of photon numbers due to the unspecified energy lost in the inactive support layer.

This broad spectrum of photon numbers greatly increases the variance of the pulse height distribution and therefore reduces the DQE, as shown in figures 2.8(a) and (b). In figure 2.8(c), the probability for different types of trajectories is given as a function of electron energy. At low energies, most electrons lose their entire energy in the scintillator and the gain increases in proportion to the incident energy, while the DQE is only limited by the Poisson spread of the photon number and by back-scattering losses. At 150 kV, electrons start to penetrate the scintillator, and the photon number starts to fall behind a linear increase. As the voltage increases

(a)



(b)



(c)
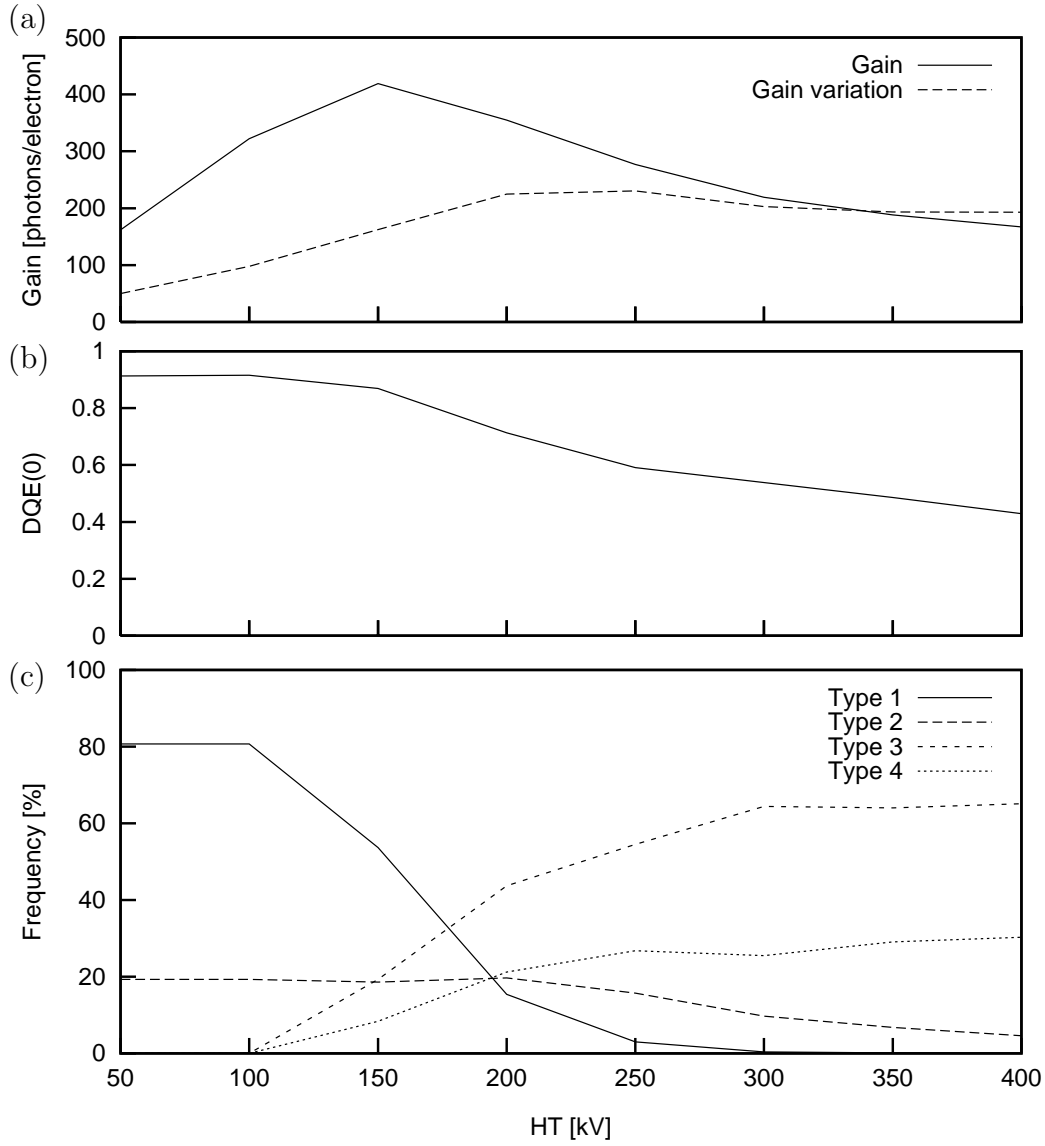


(d)  Type 1   Stopped in the scintillator without having reached the support layer.

Type 2   Back-scattered from the scintillator without having reached the support layer.

Type 3   Stopped in the support layer without going back into the scintillator.

Type 4   Re-entered the scintillator from the support layer at least once.

Figure 2.8: (a) Average $\overline{g}$ and the standard deviation $\overline{g^2} - \overline{g}^2$ of the number $g$ of photons generated by individual electrons as a function of accelerating voltage. (b) Zero spatial frequency DQE given by $\mathrm{DQE}(0) = \overline{g}^2 \big/ \overline{g^2}$. (c) Frequency of different types of trajectories, as defined in (d).

further, the gain starts to decrease due to the decreasing stopping power and cross-section for large angle scattering. At 300 kV, the probability of not reaching the support layer is vanishingly small, but about 30% of the electrons are back-scattered into the scintillator after reaching the support layer,[9] reducing the DQE(0) to ca 0.5.

**The effect of the support layer on the resolution**

The back-scattering from the support layer not only reduces the DQE(0), but also the MTF because the electron trajectories in the support layer can extend over relatively large lateral distances.

To study the implications of this effect, simulations at a range of voltages have been carried out for supported and unsupported scintillators. Results for a phosphor scintillator with a thickness of 20 $\mu$m are presented in this section.

Figure 2.9 demonstrates how back-scattering from the support layer broadens the point spread function at high energies.

The effect on the MTFS is shown in figure 2.10. At 100 kV, the support layer has no influence, as very few electrons penetrate through the scintillator. With increasing voltage, the MTFS of the supported scintillator shows an increasingly rapid decline of the MTFS at low frequencies that is not present for the unsupported scintillator.

The NTFS, however, is largely unaffected, since the back-scattered electrons generate relatively sharp spots of illumination. Therefore, at high electron energies, the DQE is strongly reduced at medium and high spatial frequencies by the presence of a supporting layer, as demonstrated in figure 2.11.

## 2.4   Experimental determination of the MTF

In general, measuring the MTF of an imaging system requires nothing more than analysing the response of the system to an input signal containing many spatial

---

[9]This may appear a high value for a back-scattering coefficient but it should be noted that most electrons enter the support layer at an angle to the plane normal.
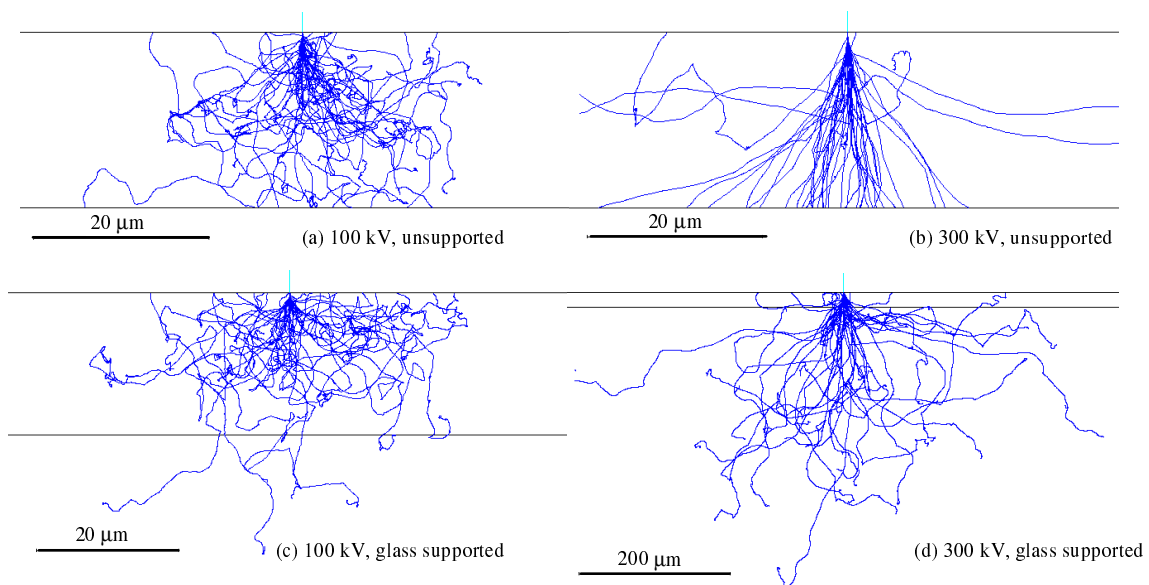
Figure 2.9: Simulated electron trajectories for a 20 μm GOS phosphor scintillator, unsupported (top) and with a glass support layer (bottom), at 100 kV and 300 kV accelerating voltage. At high electron energies, back-scattering from the supporting glass layer leads to a substantial broadening of the scintillator point spread function. Note that (d) is drawn at a different scale.
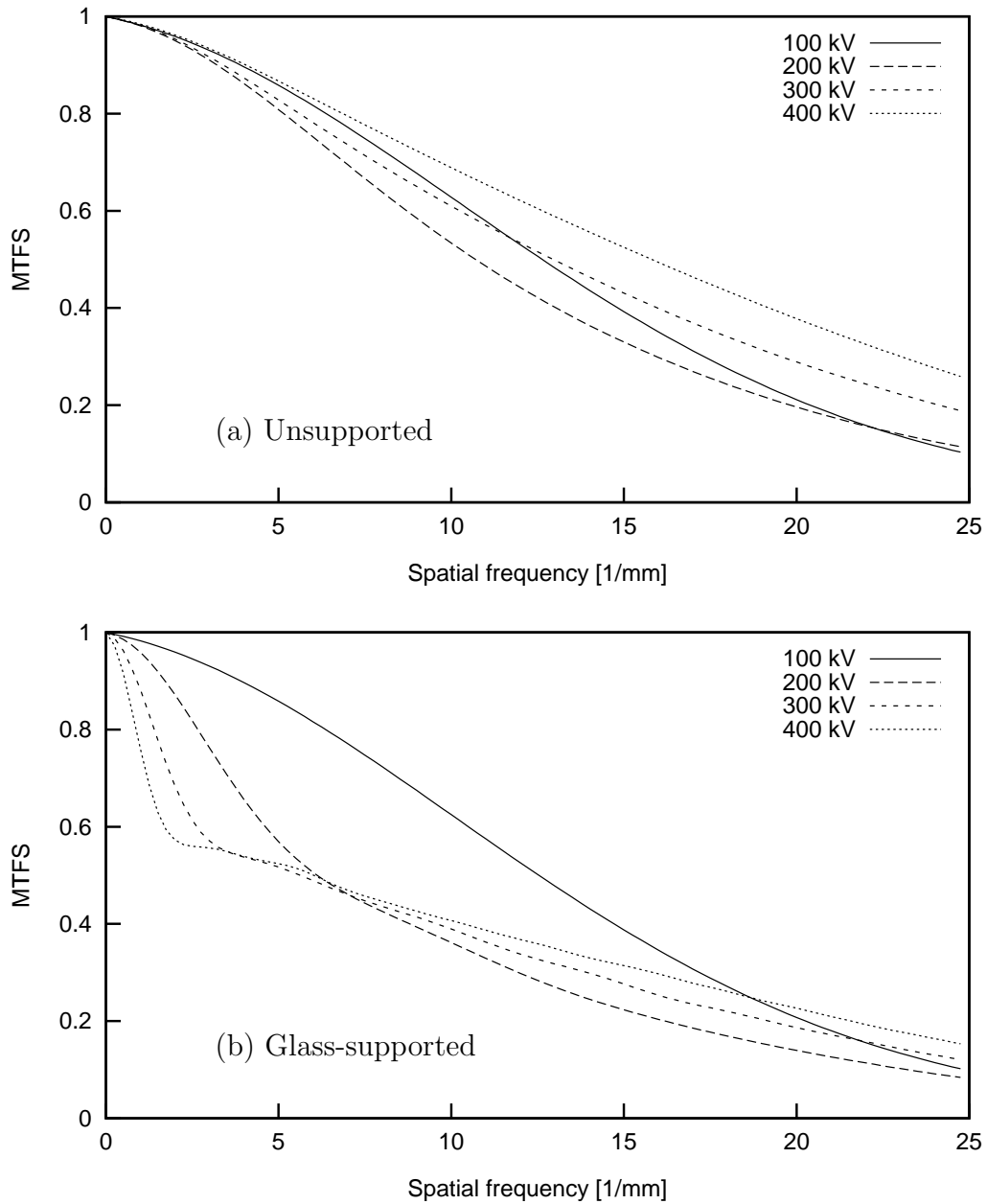
Figure 2.10: Simulation results for the MTFS of (a) unsupported and (b) glass-supported 20 $\mu$m phosphor scintillators. The presence of a support layer has no influence at 100 kV, but leads to an increasingly rapid drop of the MTFS at low spatial frequencies for higher accelerating voltages.
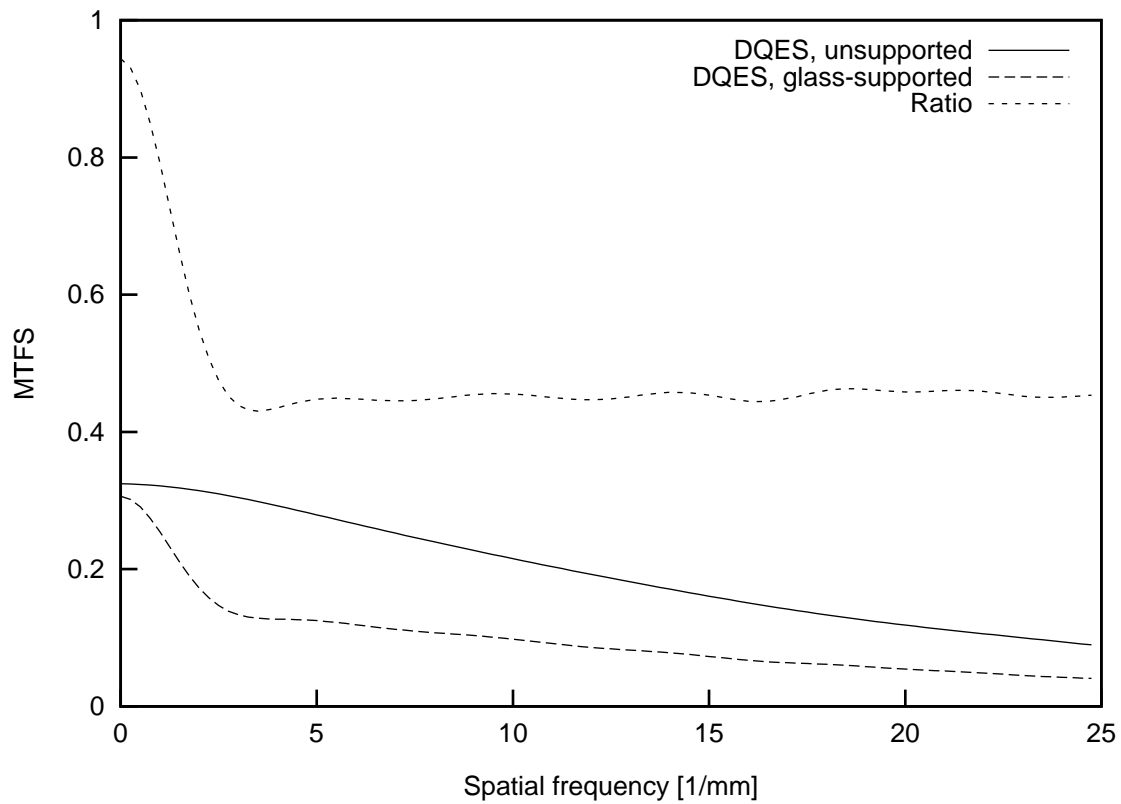
Figure 2.11: DQES for a 20 $\mu$m phosphor scintillator at 300 kV. The comparison and the ratio plot show that the DQE at all but the lowest frequencies is reduced by a factor of two in the presence of a support layer.

frequencies of known strength. The most obvious choice of input signal is a sharp step function, which can be achieved experimentally by partially covering the camera with a sharp knife edge (Daberkow et al., 1991; de Ruijter, 1995; Weickenmeier et al., 1995). As a step function contains all spatial frequencies, this allows the MTF to be determined from a single image. Care must be exercised to avoid aliasing artefacts as the spectrum of the input signal extends beyond the Nyquist limit, and effective noise filtering must be applied, as the high spatial frequency components of a step function are weak. Alternative input signals include holographic fringes (de Ruijter and Weiss, 1992) and images of amorphous carbon (van Zwet and Zandbergen, 1996), in these cases, however, the absolute contrast of the input signal is unknown and hence only ratios of the MTF at different spatial frequencies can be obtained by comparing images at different magnifications.

In the following a detailed account of the accurate MTF determination with the knife edge method is given, as reported earlier (Meyer and Kirkland, 2000).

## 2.4.1   The knife edge method

The main challenge in using the knife edge method is the image processing needed to avoid aliasing artefacts and to provide a noise-free MTF. De Ruijter [1995] describes an elegant way to solve the former problem. Their solution makes use of an edge that is slightly skewed with respect to the pixel columns, hence the contributions from above and below the Nyquist limit are separated in the 2-dimensional Fourier transform.

The approach developed in this work also uses an edge with a small skew relative to the pixel columns, which means that the edge position varies slightly from line to line in this image. This is used to extract an oversampled edge profile by combining the data from all lines shifted to a common origin with sub-pixel accuracy. In the following this process is outlined for an oversampling factor of 8, which is sufficient to safely avoid any aliasing artefacts.

1. An image, $I_1$, of a knife edge slightly skewed with respect to the pixel columns

is recorded together with an image, $I_2$, without the edge but with otherwise identical illumination conditions and an image, $I_3$, without illumination.

2. A normalised edge image $(I_1 - I_3)/(I_2 - I_3)$ is calculated.

3. The edge position in each line of this image is determined to an accuracy of $1/8$ pixel. This is achieved by multiplying the Fourier transform of the line with a Gaussian filter (with a width of approximately half the Nyquist frequency) and subsequently extending it over an 8-fold longer interval by zero padding. Inverse Fourier transform yields an 8-fold oversampled, filtered edge profile and the edge position is between the points where this function changes from below to above $1/2$.

4. An 8-fold oversampled averaged edge profile from the original linescans is obtained by shifting the edge positions to a common origin as illustrated in figure 2.12(b). Each linescan contributes to the values at every eighth sampling point of the oversampled edge profile and different linescans provide complementary information at the other sampling points because the edge is skewed.

5. The oversampled averaged edge profile already shows almost perfect point symmetry, i.e $f(x) \approx 1 - f(-x)$. However, the bright side is more affected by noise than the dark side as for Poisson noise the variance is proportional to the intensity. Therefore the accuracy of the MTF can be enhanced by a weighted symmetrisation

$$\tilde{f}(x) = \frac{w(x)f(x) + w(-x)\left(1 - f(-x)\right)}{w(x) + w(-x)} \tag{2.59}$$

where for each $x$, the weighting factor $w(x)$ is the inverse standard deviation of the individual linescan values contributing to $f(x)$. With this modification, the signal-to noise performance of the edge method is no longer inferior to the slit method as reported by Cunningham and Reid (1992).
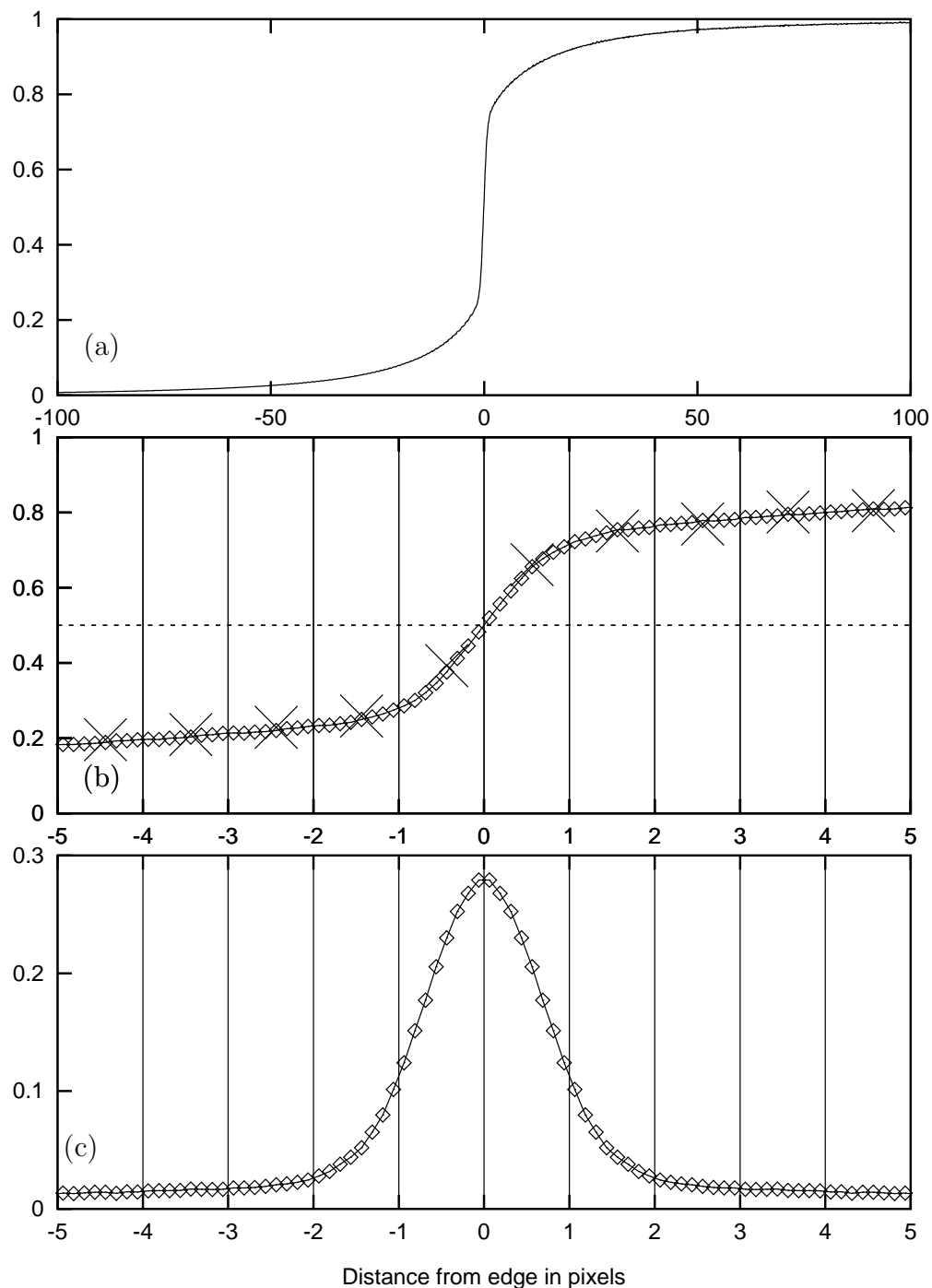
Figure 2.12: **(a)** 8-fold oversampled edge profile from the image of a skewed edge recorded at 100 kV on a camera with a YAG scintillator. **(b)** The central region of this edge profile. Each individual linescan (one example shown with crosses) contributes to the values at every eighth sampling point in the average oversampled edge profile (diamonds). **(c)** After symmetrisation, a numerical differentiation according to equation 2.59 yields the line spread function convolved with a top-hat function.

6. The MTF is obtained as the Fourier transform of the Line Spread Function, $\text{LSF}(x)$. It should be noted that the edge profile $\tilde{f}(x)$ can be expressed as $\text{LSF(x)}$ convolved with a step function:

$$\tilde{f}(x) = \text{LSF}(x) \otimes \chi_{[0,\infty]}(x) , \tag{2.60}$$

where the characteristic function $\chi_I$ is 1 within the interval $I$ and 0 elsewhere. A numerical differentiation (figure 2.12(c)) now yields:

$$
\begin{aligned}
g(x) &= -\tilde{f}(x + 1/2) - \tilde{f}(x - 1/2) & \text{(2.61)} \\
&= \text{LSF}(x) \otimes \chi_{[0,\infty[}(x) \otimes (\delta(x + 1/2) - \delta(x - 1/2)) & \text{(2.62)} \\
&= \text{LSF}(x) \otimes \chi_{[-1/2,1/2]}(x) & \text{(2.63)}
\end{aligned}
$$

This numerical differentiation involves subtraction only between values obtained from the same set of linescans, which is important in reducing the noise. Fourier transforming both sides yields

$$\text{MTF}(k) = \hat{g}(k)/\text{sinc}(\pi k) \tag{2.64}$$

7. Correction for the integration over the pixel size is effected by a further division by $\text{sinc}(\pi k)$:

$$\text{MTFS}(k) = \text{MTF}(k)/\text{sinc}(\pi k) \tag{2.65}$$

8. All the information about the MTF at high spatial frequencies is contained in the central region of the function $g$. The stochastic inaccuracy in the extracted MTF can be greatly reduced when, prior to calculating the MTF, the tails of the function $g$ are filtered with a real space averaging filter whose width increases with increasing $|x|$.

Figure 2.13 gives an example of an MTF obtained by this method. Due to the oversampling, the method yields meaningful values even above the Nyquist limit up to
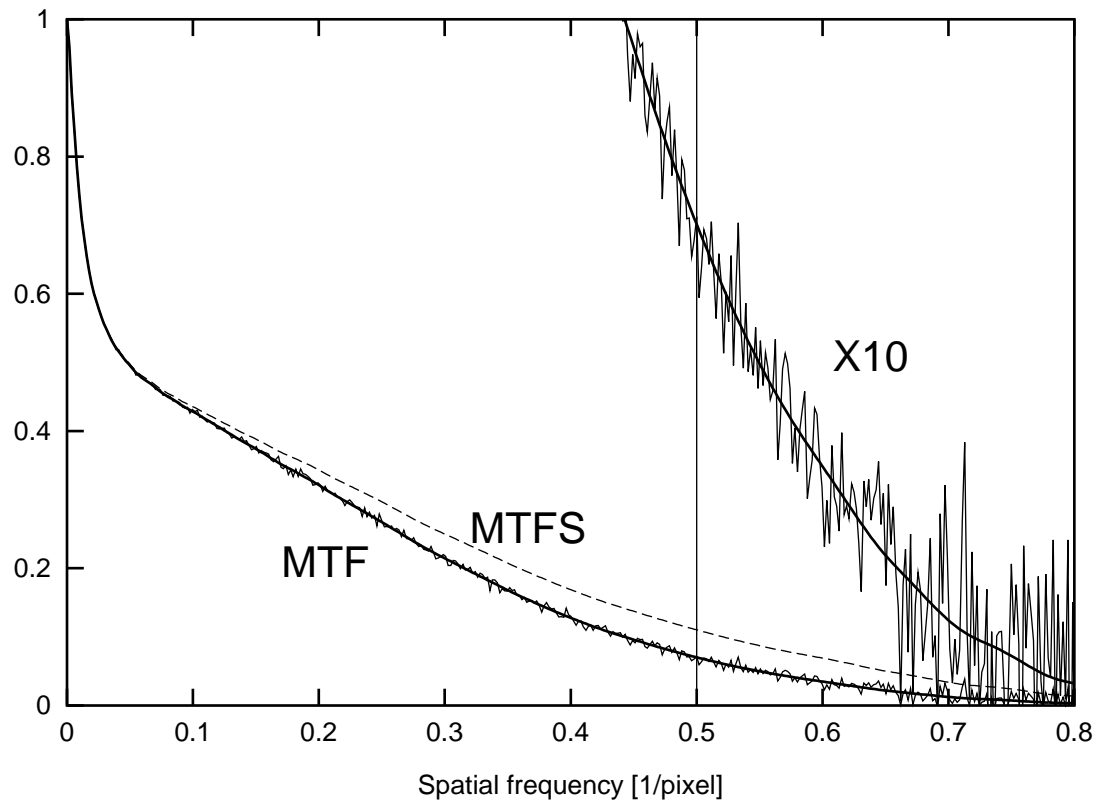
Figure 2.13: The MTF obtained from the line profile in figure 2.12. The thick lines show how the noise is reduced when the tails of the LSF are filtered with a real-space filter. The dashed line shows the MTFS evaluated from this MTF.

ca. 0.8 times the sampling frequency. Above this limit, the values become unreliable as the denominator in equation 2.64 approaches zero. The MTF extracted with this method is extremely reproducible as illustrated by the independence of the MTF on the edge position relative to the camera found in this work.

### 2.4.2 Importance of oversampling

It should be noted that the MTF cannot be obtained without oversampling from a single linescan, because in the latter case, the transferred signal at high spatial frequencies is disturbed by aliased contributions from above the Nyquist limit. When the edge coincides with a pixel centre, this interference is destructive, giving a measured MTF of zero at the Nyquist frequency. For an edge position located exactly between two pixels, contributions from above the Nyquist limit will enhance the apparent transfer, giving an MTF with twice its correct value at the Nyquist limit.

These artefacts are even more severe in the method proposed by Weickenmeier et al. (1995) in which the effect of the camera is modelled by a discrete LSF acting on a sampled sharp edge. This effectively interchanges the sequence of blurring and sampling, which is incorrect since the two operation do not commute. The effect of this on the measured MTF is investigated in the following example.

A camera with a given total MTF and corresponding (continuous) line spread function LSF is covered by an edge at $x_0$ with $0 < x_0 < 1/2$. The intensity measured at pixel $n$ (at position $x = n$) is given by:

$$F(n) = \int_{x_0}^{\infty} \text{LSF}(n - x)\mathrm{d}x \tag{2.66}$$

The effect of the camera is now modelled by a discrete LSF (DLSF) acting on an pixel-integrated sampled edge $F_S(n)$ with

$$F_S(n) = \begin{cases} 0 & : \quad n < 0 \\ 1/2 - x_0 & : \quad n = 0 \\ 1 & : \quad n > 0 \end{cases} \tag{2.67}$$

Hence, the DLSF fulfils;

$$F(n) = \sum_{m=-\infty}^{n-1} \text{DLSF}(m) + (1/2 - x_0)\text{DLSF}(n). \tag{2.68}$$

Therefore,

$$F(n) - F(n-1) = (1/2 - x_0)\text{DLSF}(n) + (1/2 + x_0)\text{DLSF}(n-1) \tag{2.69}$$

Taking the Discrete Fourier Transform (DFT) on both sides, using the shift theorem and rearranging yields:

$$\text{DMTF}(k) = \frac{1}{(1/2 - x_0) + (1/2 + x_0)e^{2\pi i k}} \text{DFT}[F(n) - F(n-1)], \tag{2.70}$$

where the DMTF is the DFT of the DLSF. Figure 2.14 demonstrates that this function differs significantly from the true MTF and is strongly dependent on the position of the edge relative to the pixel columns.

## 2.4.3 MTF correction by deconvolution

Once the MTF is determined, its effect on the image can be reversed by a simple deconvolution process. This will restore all Fourier components up to the Nyquist limit with their original strengths. However, contributions from above the Nyquist limit will not be correctly restored and therefore an MTF correction that restores sharp edges at arbitrary positions is not possible, as demonstrated in the previous section.

It nevertheless is instructive to apply the MTF correction to an image of an edge and to subsequently extract an oversampled edge profile from this image as described above.

The result is shown in figure 2.15. The edge profile shows oscillations of about 9% of the step height. This is effect is known as Gibbs' phenomenon and occurs when the Fourier transform of a sharp edge is truncated at a certain spatial frequency.

The MTF extracted from this line profile (figure 2.16) represents the complete transfer from the original object to the deconvolved image. It is apparent that all
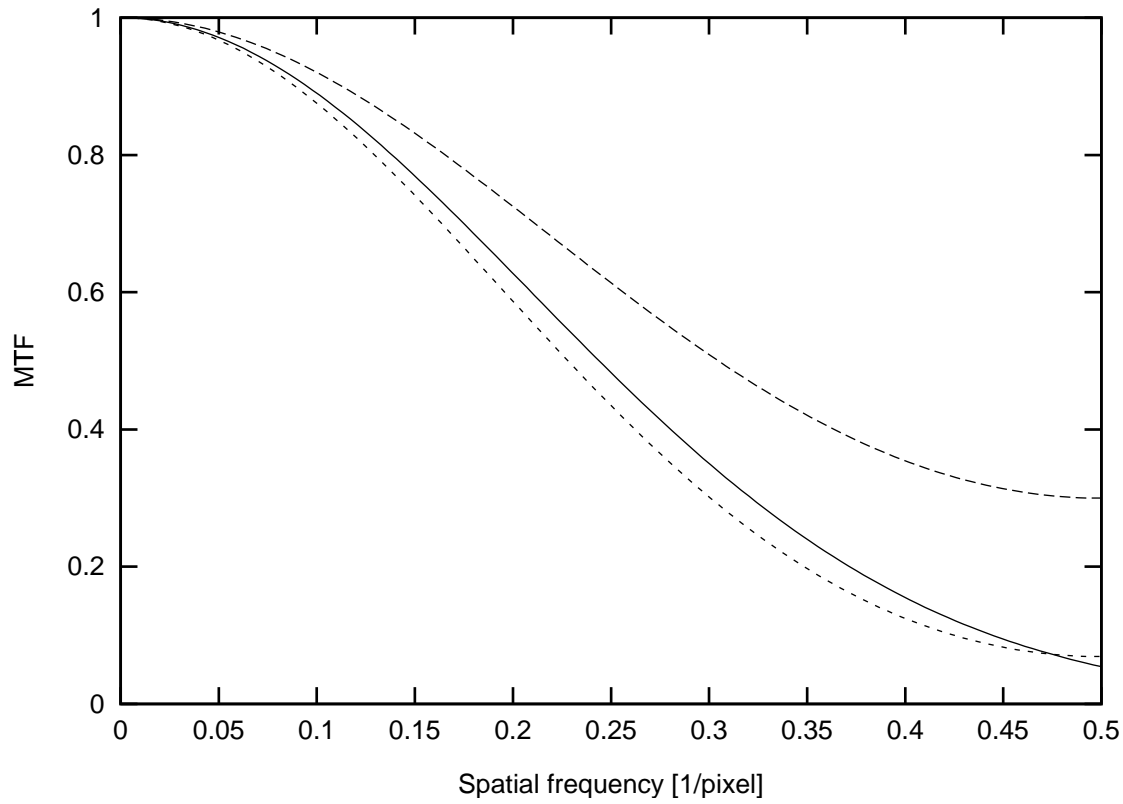
Figure 2.14: Influence of sampling artefacts on MTF measurements. The edge profile that would be obtained by a camera with a total $\mathrm{MTF}(k) = \exp(-12k^2)$ (solid graph) is calculated and sampled (see text). A discrete LSF is sought such that a convolution of the sampled sharp edge with this LSF yields the calculated intensities. The discrete FT of this LSF is shown for an edge aligned with a pixel column centre (dashed line) and for an edge aligned between two pixel columns (dotted line). This illustrates the strong dependence of this apparent MTF on the position of the edge and the significant over-estimation of the MTF when the edge is close to a pixel column centre.
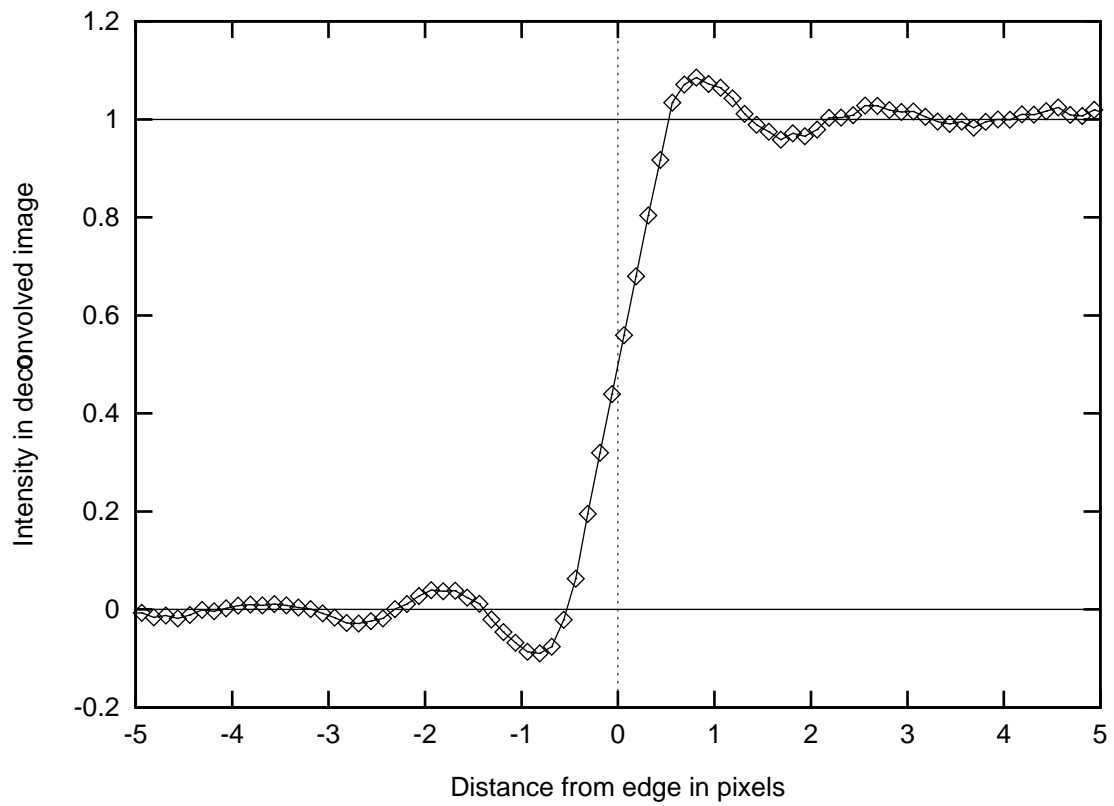
Figure 2.15: Oversampled edge profile from an edge image that is fully MTF corrected by deconvolution. Note the oscillation on both sides of the step due to the truncation of frequencies above the Nyquist limit (Gibbs' Phenomenon).
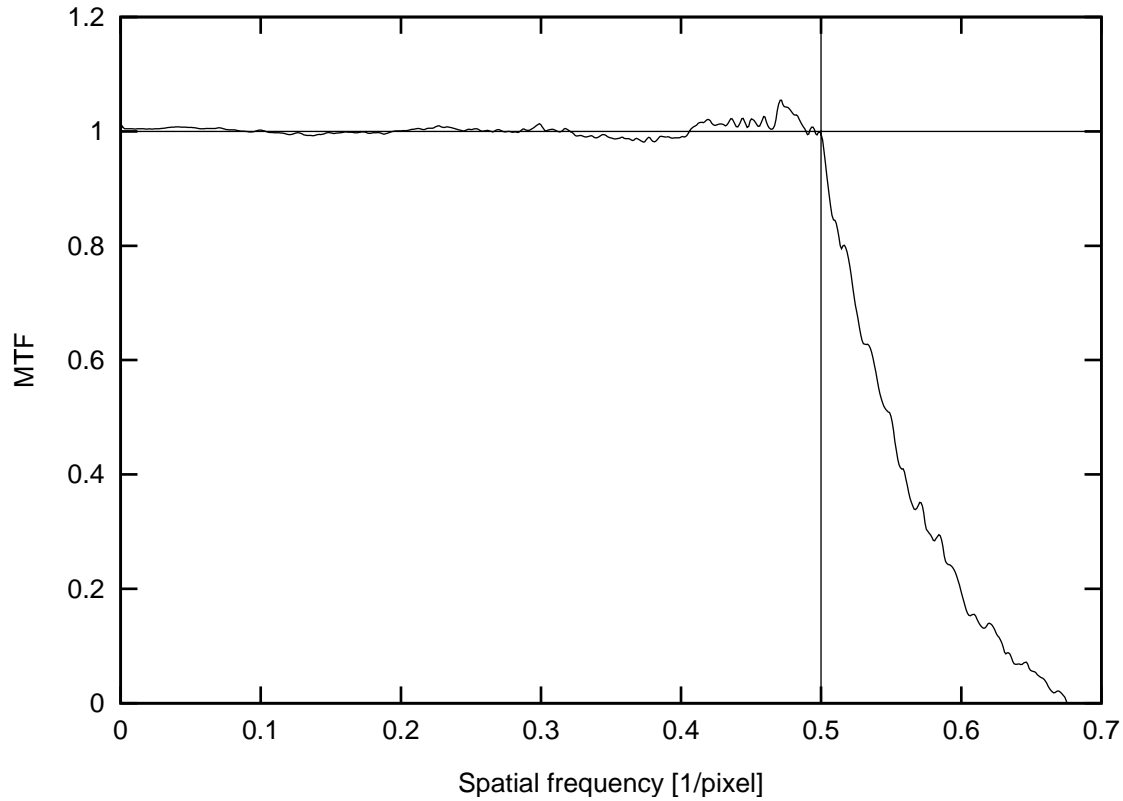
Figure 2.16: MTF extracted from the edge profile in figure 2.15, representing the complete transfer from the object to the deconvolved image. Frequencies below the Nyquist limit are restored to their original strengths, while the transfer decays rapidly above the Nyquist limit.

frequencies up to the Nyquist limit are indeed perfectly restored and that the transfer decays rapidly above the Nyquist limit.

In practice, it is rarely favourable to extend the deconvolution to the Nyquist limit. In order to avoid intolerable noise levels it is experimentally better to truncate the total signal transfer at a lower frequency with a soft edged filter. This also eliminates aliasing artefacts from components just above the Nyquist limit.

|        | $r=1$  | $r=2$  | $r=5$   | $r=10$  | $r=20$ |
|--------|--------|--------|---------|---------|--------|
| $N=1$  | 40%    | 28%    | 18%     | 13%     | 9%     |
| $N=14$ | 11 %   | 8 %    | 5 %     | 3.4%    | 2.4%   |
| $N=70$ | 5 %    | 3.4%   | 2.1 %   | 1.5 %   | 1.1%   |
| $N=294$ | 2.3 % | 1.6 %  | 1.0 %   | 0.7 %   | 0.5%   |

Table 2.1: Statistical error in the NTF measured from a rotational average at Fourier space radius $r$ from $N$ power spectra.

## 2.5 Experimental determination of the NTF

For each frequency $(u, v)$ the image intensity Fourier transform $\hat{I}(u, v)$ is a complex random variable with Gaussian distribution and variance $V(u, v)$. When the illumination is uniform,[10] the expectation value $E(u, v)$ is zero for all $(u, v) \neq (0, 0)$. The values $|\hat{I}(u, v)|^2$ in the power spectrum are themselves random variables, with expectation value $V(u, v)$ and variance $(V(u, v))^2$.[11] Therefore, an average of $M$ power spectra of images with uniform illumination provides estimates of $V(u, v)$ at each $(u, v)$ with an relative accuracy of $1/M$ corresponding to a relative error of $1/\sqrt{M}$ in the NTF. If $V(u, v)$ can be assumed to be rotationally symmetric, rotational averaging can be used concommittantly, reducing the relative error in the NTF estimated from $M$ images with $N \times N$ pixels to approximately $1/\sqrt{2\pi k N M}$, where $k$ is the spatial frequency in 1/pixel and hence $kN$ is the Fourier space radius in pixels.

Table 2.1 demonstrates that in order to achieve a high accuracy at low frequencies, averaging over many power spectra is necessary.

Therefore, several series of $M = 16$ unprocessed images with uniform illumination

---

[10]For non-uniform illumination the difference $I_1 - I_2$ of two images with identical illumination can be used instead.

[11]Proof: A complex random variable $z$ with Gaussian distribution and standard deviation $\sigma$ has the probability density $p(z) = (1/\pi\sigma^2)\,\mathrm{e}^{-|z|^2/\sigma^2}$. Integration of $|z|^2 p(z)$ and $|z|^4 p(z)$ over the full complex plane is simplified by changing to polar co-ordinates with $r = |z|$ and substituting $u = r^2/\sigma^2$. This yields $\overline{|z|^2} = (1/\pi\sigma^2)\int_0^\infty 2\pi r r^2 \mathrm{e}^{-r^2/\sigma^2}\mathrm{d}r = \int_0^\infty u\sigma^2 \mathrm{e}^{-u}\mathrm{d}u = \sigma^2$ and $\overline{|z|^4} = \int_0^\infty u^2\sigma^4 \mathrm{e}^{-u}\mathrm{d}u = 2\sigma^4$. Hence $\mathrm{var}(|z|^2) = \overline{|z|^4} - \left(\overline{|z|^2}\right)^2 = \sigma^4$.

were recorded.[12] After subtracting an average of the dark current images, the unprocessed images were then gain-normalised by division by the mean of the series. This also removes the effect of the illumination being slightly non-uniform. Subtracting the average of the previous and the next image from each processed image in the series effectively eliminated all remaining non-noise contributions to the power spectrum due to drift in the illumination conditions. Each non-zero spatial frequency Fourier component of these now truly uniform noise images has an expectation value of zero and a variance $3(M-1)/(2M\bar{I}^2)V_{\mathrm{CCD}}(u,v)$, where $\bar{I}$ is the average intensity prior to gain normalisation.

The variance can be measured by averaging many power spectra of such images. The averaged power spectrum must then be corrected for the effects of the CCD sensor, namely attenuation by pixel integration and aliasing. The difficulty of aliasing correction is that contributions from below and above the Nyquist limit cannot be distinguished in a noise image, as discussed in section 2.2.5. This is expressed in equation 2.24, for convenience re-iterated below.

$$V_{CCD}(u,v) = \sum_{m,n} V(u+m, v+n)\mathrm{sinc}^2(\pi(u+m))\mathrm{sinc}^2(\pi(v+n)) \qquad (2.71)$$

The aim of aliasing correction is to obtain an estimate for $V(u,v)$ from the measured values of $V_{CCD}(u,v)$. The method developed in this work relies on the assumption that $V(u,v)$ has rotational symmetry and that $V(k)$ is small enough to be ignored for $k \geq 3/2$.[13]

This implies that

$$
\begin{aligned}
V_{\mathrm{CCD}}(1/2, 0) &= 2\,(2/\pi)^2\,V(1/2) + 2\,(2/(3\pi))^2\,V(3/2) + \ldots \\
&= 8/\pi^2\,V(1/2) && (2.72) \\
\text{and } V_{\mathrm{CCD}}(1/2, 1/2) &= 4\,(2/\pi)^4\,V(\sqrt{2}/2) + 8\,(2/\pi)^2(2/(3\pi))^2\,V(\sqrt{10}/2) + \ldots \\
&= 64/\pi^4\,V(\sqrt{2}/2)\,. && (2.73)
\end{aligned}
$$

---

[12]It is important that many uniformly illuminated images are recorded to achieve a high accuracy at low frequencies. The number of dark images is less critical.

[13]The latter assumption allows to restrict the summation in equation 2.71 to $m, n \in \{-1, 0, 1\}$.

For a good estimate of V(1/2,0) the pixels in the near neighbourhood of $(1/2, 0)$ should be utilised as well. This can be done by finding constants $a$, $b$, $c$ such that the function[14] $c + a(u - 1/2)^2 - bv^2$ best approximates the averaged experimental power spectrum $V_{\exp}(u, v)$ in the neighbourhood of $(1/2, 0)$. $V(1/2)$ can then be obtained as an average of $c\pi^2/8$ and the value obtained in the same manner from the pixels near $(0, 1/2)$.[15]

Taking into account that aliasing increases the variance 4-fold at the corners of the power spectrum, $V(\sqrt{2}/2)$ can be obtained by fitting a paraboloid function to the values around $(1/2, 1/2)$. Fitting an exponential function through the values obtained for $k = 1/2$ and $k = \sqrt{2}/2$ gives an estimate $V_{\text{est}}(k)$ for $V(k)$ above the Nyquist limit. This is used to evaluate the expected contributions due to aliasing as

$$V_{\text{alias}}(u, v) = \sum_{m=-1}^{1} \sum_{n=-1}^{1} (1 - \delta_{m0}\delta_{n0}) V_{\text{est}}(u + m, v + n) \text{sinc}^2(\pi(u + m)) \text{sinc}^2(\pi(v + n))$$

(2.74)

and subtract them from the power spectrum. The corrected power spectrum, $V_{\text{corr}}(u, v)$ is thus obtained as

$$V_{\text{corr}}(u, v) = (V_{\text{exp}}(u, v) - V_{\text{alias}}(u, v)) / \left( \text{sinc}^2(\pi u) \text{sinc}^2(\pi v) \right) .$$

(2.75)

A rotational average of this function finally yields $V_{\text{corr}}(k)$ and the NTFS is evaluated as

$$\text{NTFS}(k) = \sqrt{V_{\text{corr}}(k)/V_{\text{corr}}(0)}$$

(2.76)

This method for aliasing correction is illustrated in figure 2.17. It should be noted that an analytical model of the NTF is not required. An analytical function is fitted only for the corrections and the consistency of this fit can be checked by comparing the fitted function $V_{\text{est}}$ and the resulting $V_{\text{corr}}$ for $1/2 \leq k \leq \sqrt{2}/2$. If the former was too high, $V_{\text{alias}}$ would be over-estimated and the resultant $V_{\text{corr}}$ would be too low.

---

[14]For reasons of symmetry, this function is the complete second order Taylor expansion of $V_{\text{CCD}}$ in $(1/2, 0)$.

[15]It should be noted that the discrete Fourier transform is continued periodically, so that the neighbouhood of $(1/2,0)$ contains values from both the left and right border of the discrete Fourier transform.
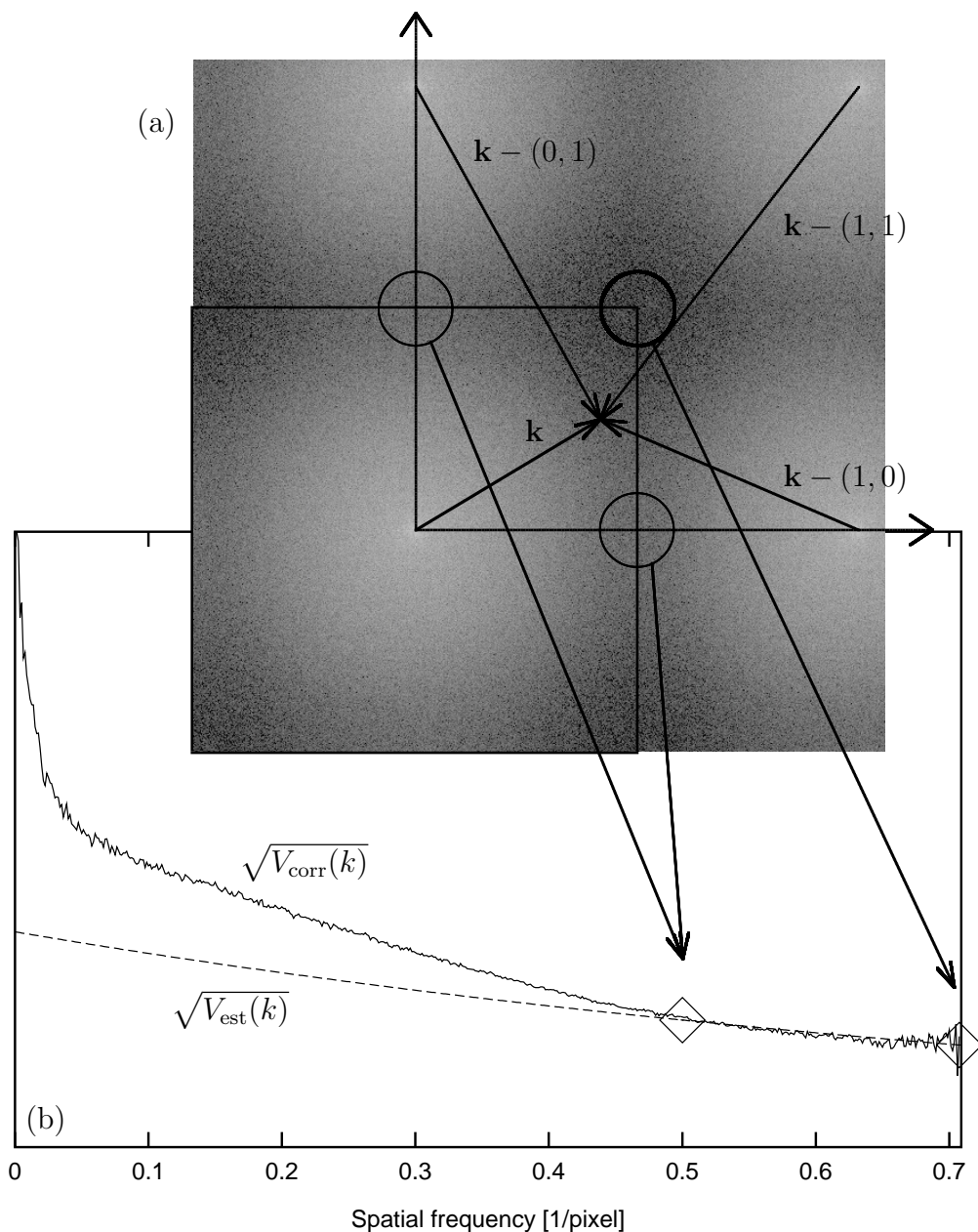
Figure 2.17: Aliasing correction for the NTFS measurement. (a) At a given spatial frequency $k$, the uncorrected experimental noise power spectrum $V_{\text{exp}}$ contains contributions from all frequencies $\mathbf{k} - (m,n)$ with integers $m,n$. Due to aliasing $V_{\text{exp}}$ is increased the 2-fold at the border centres (thin circles) and 4-fold at the corners of Fourier space (thick circle). With this knowledge, the encircled regions can be used to obtain estimates for $V(0.5)$ and $V(0.707)$. An exponential function $V_{\text{est}}(k)$ is fitted through these points and used to estimate aliasing contributions to $V_{\text{exp}}$, which can then be subtracted. After correcting for pixel integration, a rotational average is taken. (b) The root of the corrected noise power spectrum together with the exponential function (thin line) that was used above the Nyquist frequency for aliasing correction. The good match shows that the contributions from aliasing were estimated with sufficient accuracy.

An alternative aliasing correction, where an analytical function is fitted to the NTF, has previously been proposed (Hülk and Daberkow, 1998). However, this method leads to an over-estimate of the aliasing correction necessary at Nyquist when compared to the method discussed here.

Figure 2.18 shows a typical corrected power spectrum, averaged over all series. This averaging is essential for an accurate normalisation for which the root of the noise power spectrum has to be extrapolated to zero spatial frequency. Experimentally, to keep the data collection times and the total volume of data within acceptable limits, most series were recorded with pixel binning and thus only contribute to the centre regions of Fourier space, where data is most needed. In total 21 series were recorded, one without binning, 4 with $2 \times 2$ binning and 16 with $4 \times 4$ binning. Each series had 16 images and hence provided 14 power spectra.[16] The aliasing correction has to be performed separately for the 3 different binning conditions, as they have different Nyquist frequencies. No systematic difference could be observed between the binned an unbinned power spectra, even near the Nyquist limit of the binned ones, which indicates that the aliasing correction is effective.

## 2.6 Experimental results

Using the methods described above, a number of cameras with different scintillators at different voltages were investigated. Table 2.2 lists the cameras for which results are presented and discussed in this section.

### 2.6.1 A camera equipped with YAG scintillator at 100 kV

The first experiments were carried out on camera 1 in table 2.2, a Gatan 679 CCD camera with YAG scintillator fitted to a post-column imaging filter (GIF), which is installed on a 100kV VG STEM in Cambridge to record energy spectra and energy-filtered nano-diffraction patterns. The MTF was measured using the shadow image of

---

[16]For the first and the last image, the subtraction of $(I_{n-1} + I_{n+1})/2$ is not possible.
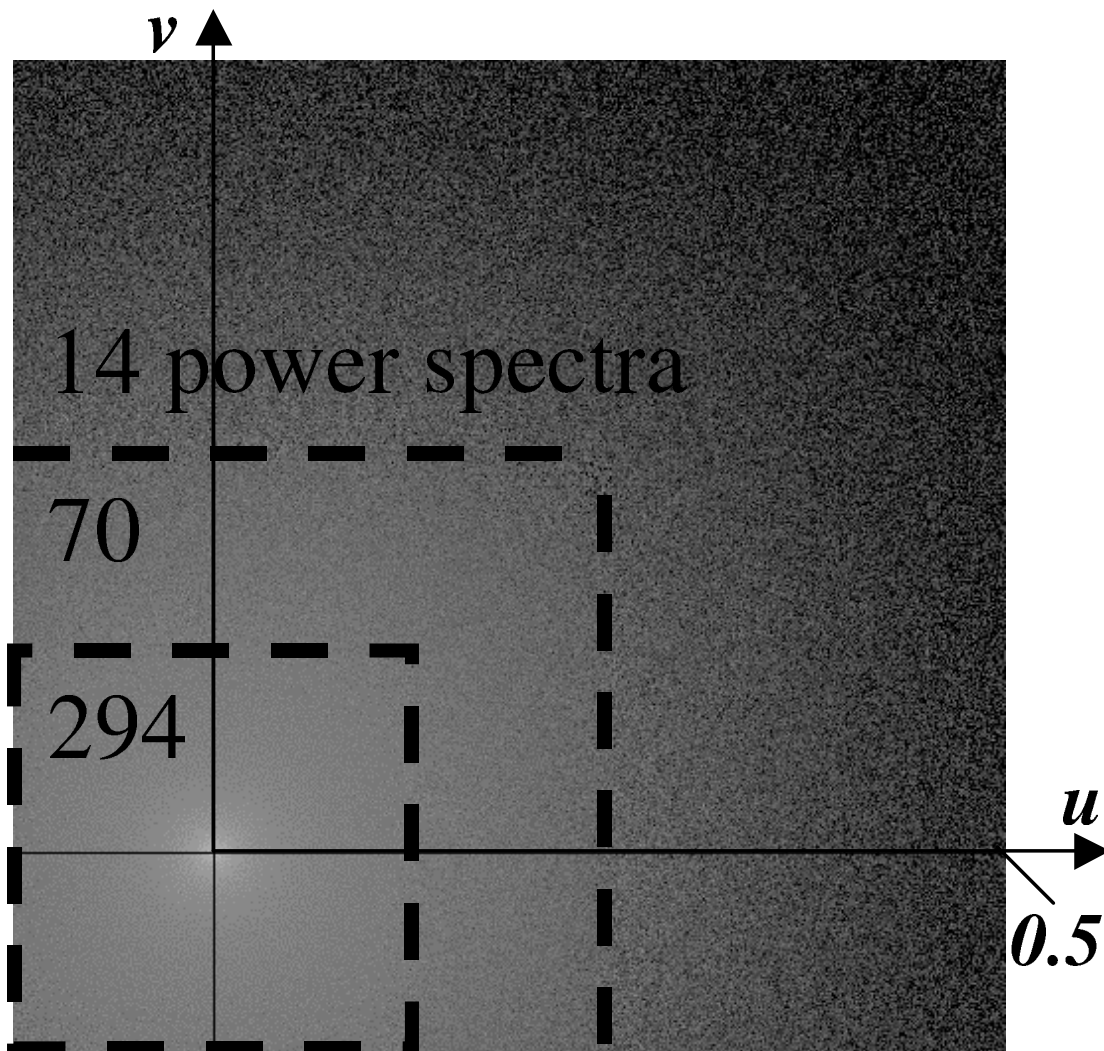
Figure 2.18: Average of a large number of aliasing corrected power spectra obtained from several series of images. The very long series were recorded with pixel binning and contribute only to the centre region of Fourier space.

| No. | Model | Scintillator | HT [kV] | Microscope | Page |
|-----|-------|--------------|---------|------------|------|
| 1 | Gatan 679 | YAG | 100 | VG STEM, Cambridge | 87 |
| 2 | SIS BioCam[a] | Phosphor | 100 | Philips CM 100 | 94 |
| 3 | Gatan 679 | Phosphor | 100-400 | JEM 4000FX, Cambridge | 95 |
| 4 | Gatan 794 Multiscan | Phosphor | 300 | JEM 3000F, Oxford | 99 |
| 5 | Gatan 794/IF20 Megascan | Phosphor | 300 | JEM 3000F, Oxford | 99 |
| 6 | SIS Megaview | Phosphor[b] | 120-200 | JEM 2010 | 102 |

[a]The raw data for this measurement was provided by I. Daberkow (Hülk and Daberkow, 1998).
[b]Lens-coupled.

Table 2.2: The key characteristics of the cameras investigated in this work.

the edge of a pneumatically retractable bright-field (BF) detector mounted between the last GIF crossover and the CCD camera. The detector was stopped between the inserted and retracted position by interrupting the airflow during retraction. To avoid additional intensity due to electrons reflected from the side of the retractable device, only images where the edge covered more than half of the field of view, as in figure 2.19, were included in the analysis.

The results were first published in (Meyer and Kirkland, 2000) and have also been used to illustrate the experimental methods described in the previous section.

Figure 2.20 shows a rapid decline of both the MTFS and the NTFS at low spatial frequencies. This can be attributed to the waveguide effect described in section 2.3.2, which affects both MTFS and NTFS as it takes place after amplification. This also means that this effect has little influence on the DQE, which only declines slowly at low spatial frequencies. Extrapolation to zero spatial frequency yields DQE(0)=0.87, which closely agrees with the value reported by Daberkow et al. (1991) and also with simulations (see below) and is close to the back-scattering limit of 0.93 discussed in section 2.3.4. The high value of the DQE helps in achieving a high effective pixel number of $N_{\mathrm{eff}} = 392000 = 0.37 N_{\mathrm{tot}}$, which is the highest percentage of the total pixel number observed in any of the cameras measured.

Figure 2.21 shows a comparison of the experimental results and a simulation with the parameters given in column 3 of table 2.3. Most of the parameters reflect known characteristics of this camera. The scintillator thickness, however, was not known, as

| Variable | Description | YAG | Phosphor |
|---|---|---:|---:|
| Aperture | Semi-angle of optical entrance aperture | 33 | 60 |
| AperWidth | Width of soft aperture edge | 0 | 0 |
| E | Incident energy [keV] | 100 | 400 |
| c | Conversion factor | 0.05 | 0.05 |
| EPhot | Average energy of generated photons | 2.2 | 2.2 |
| IncAngle | Incident angle of the electrons | 0 | 0 |
| TopRefl | Reflectivity of the top scintillator surface | 1 | 0 |
| BotRefl | Reflectivity of the bottom scintillator surface | 0 | 0 |
| Focus | Position of optical focus plane [$\mu$m] | 31 | 41 |
| DFrac | Total reflection ration | 1.35 | 0 |
| DRad | Mean displacement by multiple total reflection [$\mu$m] | 600 | 0 |
| KMax | Maximum frequency $k_{max}$ [mm$^{-1}$] | 25 | 25 |
| KSamp | Sampling step between $k$-values [mm$^{-1}$] | 0.25 | 0.25 |
| N | Number of primary electrons in simulation | 20000 | 20000 |
| TCoat | Thickness of coating [$\mu$m] | 1 | 1 |
| TScint | Thickness of scintillator [$\mu$m] | 30 | 40 |
| TSupp | Thickness of support layer [$\mu$m] | 2000 | 2000 |
| MScint | Scintillator material | 71 (YAG) | 83 (GOS) |
| MSupp | Support layer material | 73 (Glass) | 73 (Glass) |
| MFill | Filling material in phosphor scintillators | | 73 (Glass) |
| GrainSize | Phosphor grain size | — | 3 |
| FillFactor | Phosphor fill factor | 1 | 0.7 |
| LamPhAbs | Mean free path for photon absorbtion [$\mu$m] | 0 | 80 |
| LamPhSca | Mean free path for photon scattering [$\mu$m] | 0 | 80 |
| NScint | Refractive index of the scintillator | 1.83 | 1.2 |
| NSupp | Refractive index of the support | 1.5 | 1.5 |
| Rob0 | Width of optical PSF of the CCD camera | 10 | 10 |
| Rob30 | Ditto, at 30 incidence angle | 13.6 | 13.6 |
| MtfFile | File where MTF, NTF and DQE are recorded | YAG.dat | phos40.dat |

Table 2.3: List of the variable names used in the batch language of the Monte Carlo simulation program, together with the values used in the simulations that yielded the best match to experimental data for a YAG scintillator (figure 2.21) and a phosphor scintillator (figure 2.25).

Final cross-over

Retractable BF detector

Slow-scan CCD camera
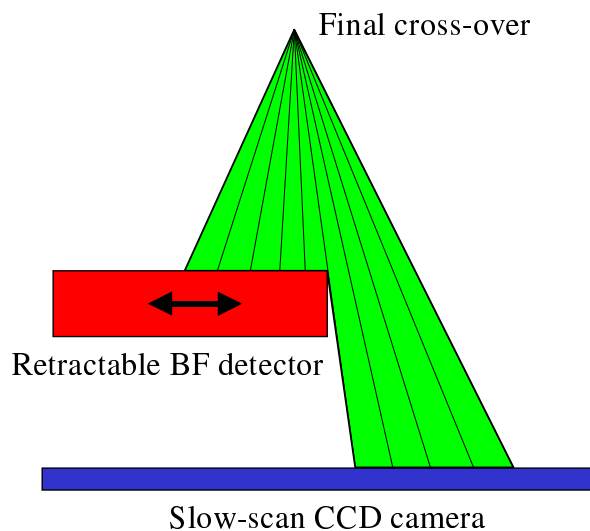
Figure 2.19: The edge of a pneumatically retractable bright-field detector in a half-retracted position in front of the CCD camera was used to measure the MTF.

it is treated as proprietary information by the manufacturer. In simulations it was found that the MTF at high spatial frequencies depends strongly on the scintillator thickness, mainly because increasing the thickness increases the distance of the incident point from the optical focus plane, which invariably is the back scintillator surface for fibre-optically coupled systems. The best match of the simulation results to the experimental data was achieved for a thickness of 30 $\mu$m, an estimate that was confirmed to be reasonable by Gatan representatives. The second parameter that was not known independently is the 'mean displacement by multiple total reflection', which was determined by fitting the low spatial frequency part of the MTFS. The good overall agreement of experiment and simulation suggests that the simulation program is a useful predictive tool for optimising parameters in camera design.
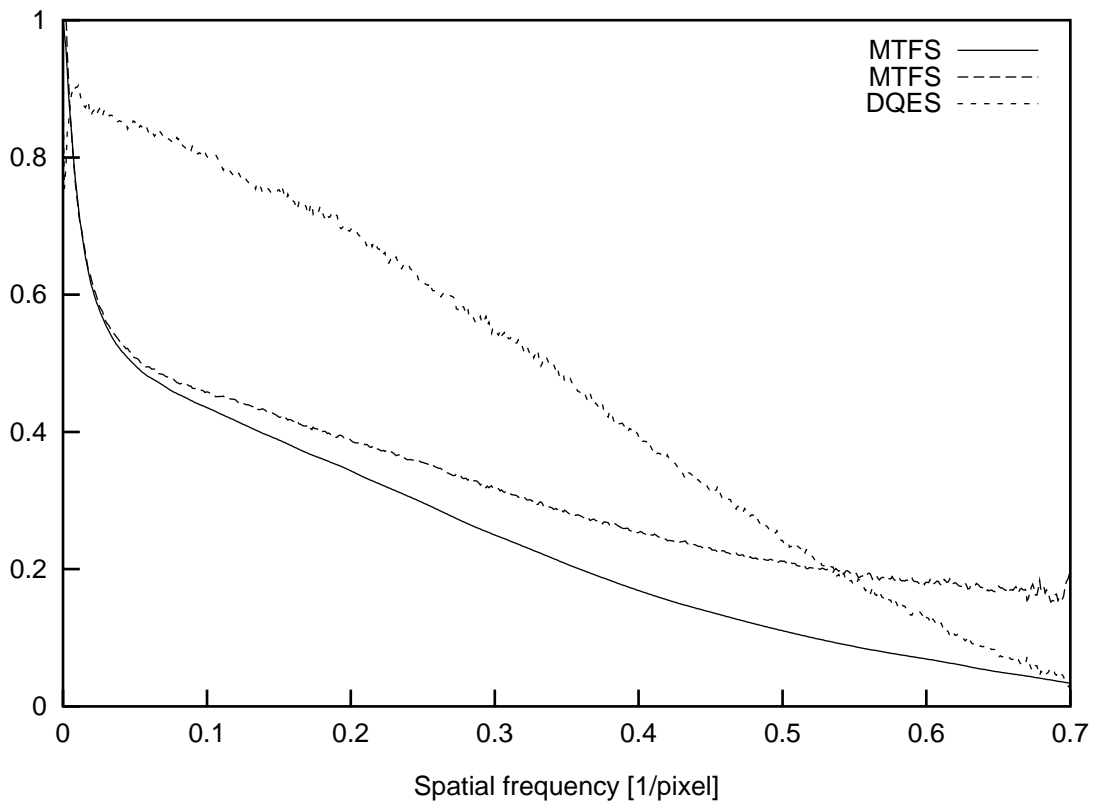
Figure 2.20: MTFS, NTFS and DQES for camera 1 at 100 kV. Both MTFS and NTFS drop rapidly at low spatial frequencies due to the waveguide effect in the YAG scintillator. The DQES, however, remains high over a wide range of frequencies.
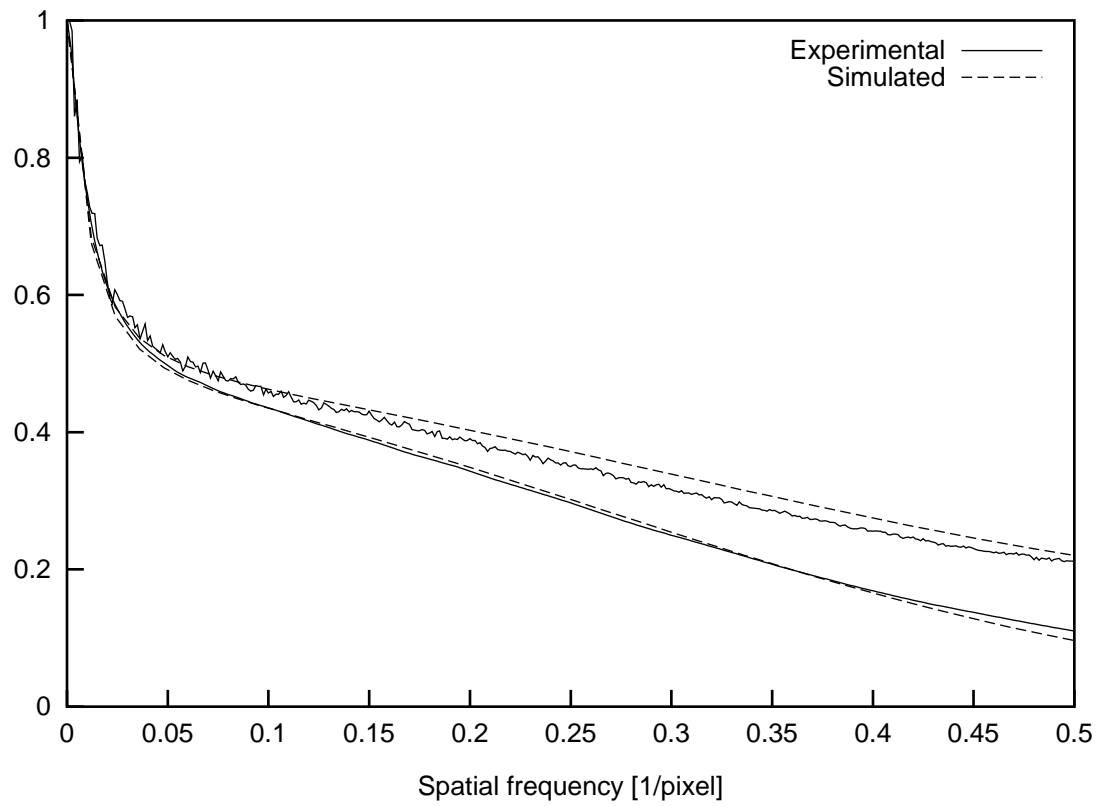
Figure 2.21: Comparison of the experimental results for camera 1 with simulations with the parameters given in column 3 of table 2.3
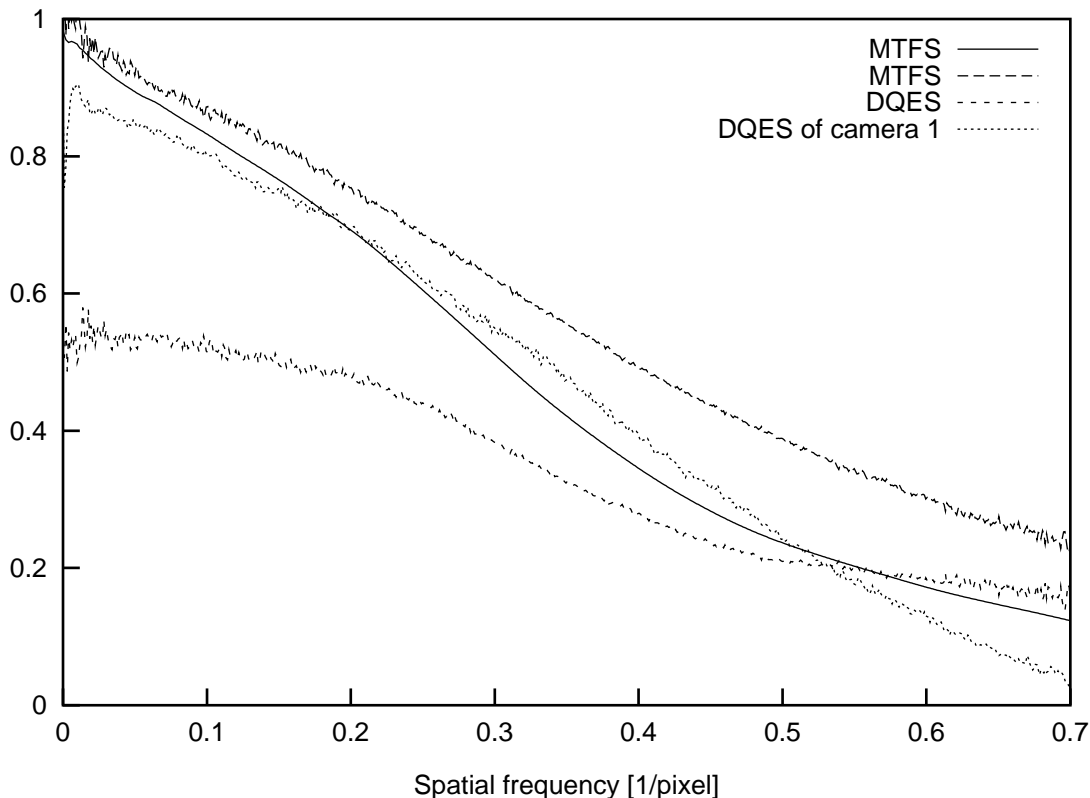
Figure 2.22: MTFS, NTFS and DQES for camera 2 at 100 kV. For comparison, the DQES of camera 1 is also shown.

## 2.6.2   A camera with thin phosphor scintillator at 100 kV

This section presents results from a re-examination of the raw data taken by Hülk and Daberkow (1998) on a SIS BioCam camera at 100 kV with the methods developed in this work. The camera has a thin[17] phosphor scintillator with is fibre-optically coupled to a 1k×1k CCD chip with 24 $\mu$m pixel size.

The results in figure 2.22 show that the NTFS is substantially higher than the MTFS at high frequencies, in contrast to the good agreement found by Hülk and Daberkow (1998). This is mainly because in the determination of the NTFS in that

---

[17]The precise value of the scintillator thickness is proprietary information, but my estimate of 10 $\mu$m was confirmed to be a good estimate by company sources.

paper aliasing was over-corrected as the correction procedure was based on the assumption that the NTFS is increased 2-fold at the Nyquist limit, whereas in fact it is the power spectrum that is increased 2-fold.

In comparison with camera 1 (figure 2.20), the MTFS is clearly superior due to the lack of the waveguide effect and because the light is generated closer to the focus plane due to the much thinner scintillator. However, this comes at the price of a reduced zero-spatial frequency DQE due to the phosphor granularity and, more importantly, because the electrons are able to penetrate through the thin scintillator. Both effects widen the distribution of path length spent in scintillating material and hence the pulse height distribution. The comparison of the DQES of both cameras in figure 2.22 shows that these disadvantages outweigh the benefits of having a thin scintillator, as the DQES of camera 2 is only superior at the highest spatial frequencies, which are of no practical importance as they only represent pixels at the corners of the 2-dimensional Fourier transform. The effective pixel number $N_{\text{eff}} = 292000 = 0.28 N_{\text{tot}}$ is therefore slightly lower than for camera 1, but the difference is sufficiently small that it might be outweighed by other benefits.

### 2.6.3   The high voltage problem

In order to investigate the long-range blurring predicted for high electron energies and supported scintillators in section 2.3.4, measurements at voltages ranging from 100 kV to 400 kV were performed on a Gatan 679 CCD camera with phosphor scintillator fitted to a post-column imaging filter (GIF), which is installed on a JEOL 4000FX in Cambridge (camera 3). The retractable TV camera mounted in front of the CCD camera in this GIF was used as an edge in the same fashion as the BF detector for camera 1.

Figure 2.23 shows that, as the voltage increases, the MTFS starts to drop increasingly rapidly at low spatial frequencies due to the electrons back scattered from the support layer. The influence of the accelerating voltage on the NTFS is much weaker and therefore, the DQES in figure 2.24 drops increasingly rapidly from its zero spatial
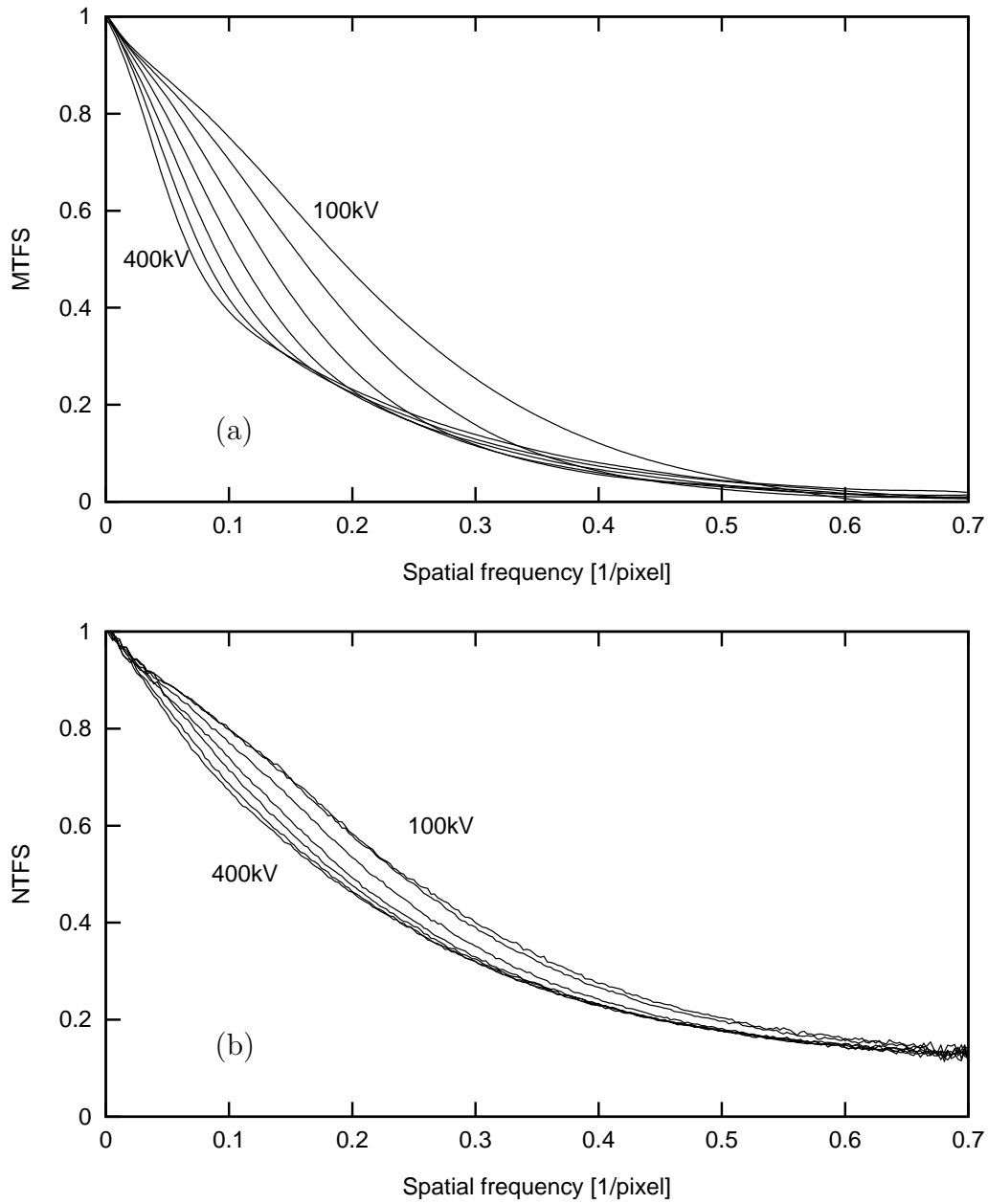
Figure 2.23: (a) MTFS and (b) NTFS of the 1k Gatan camera (camera 3) fitted to the JEM4000FX microscope in Cambridge, measured at electron energies from 100keV to 400keV in steps of 50keV
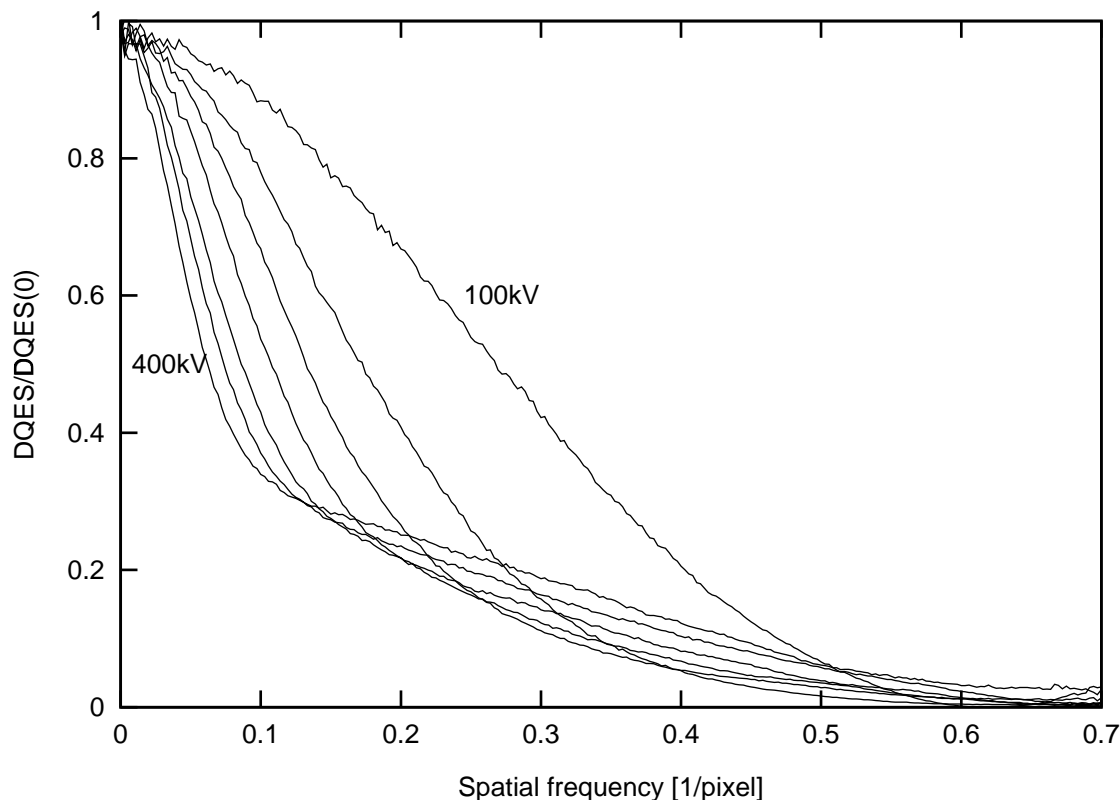
Figure 2.24: DQES of the 1k Gatan camera (camera 3) fitted to the JEOL 4000FX microscope in Cambridge, measured at electron energies from 100keV to 400keV in steps of 50keV.

frequency value as the voltage increases.

The low values of the MTFS at high frequencies indicate that the scintillator is thicker than in the simulations for 20 $\mu$m thickness presented in section 2.3.4. The best match, shown in figure 2.25, was achieved with simulation at a thickness of 40 $\mu$m. The simulation parameters are given in column 4 in table 2.3 on page 90. The figure shows that qualitatively, the voltage dependence of the MTFS and NTFS is adequately described by the Monte-Carlo model, but disagreements at some spatial frequencies and voltages show that the model is not entirely accurate.

The main problem is that many of the characteristics of the phosphor, such as the

MTFS, 100kV

NTFS, 100kV

MTFS, 200kV

NTFS, 200kV

MTFS, 300kV

NTFS, 300kV

MTFS, 400kV

NTFS, 400kV
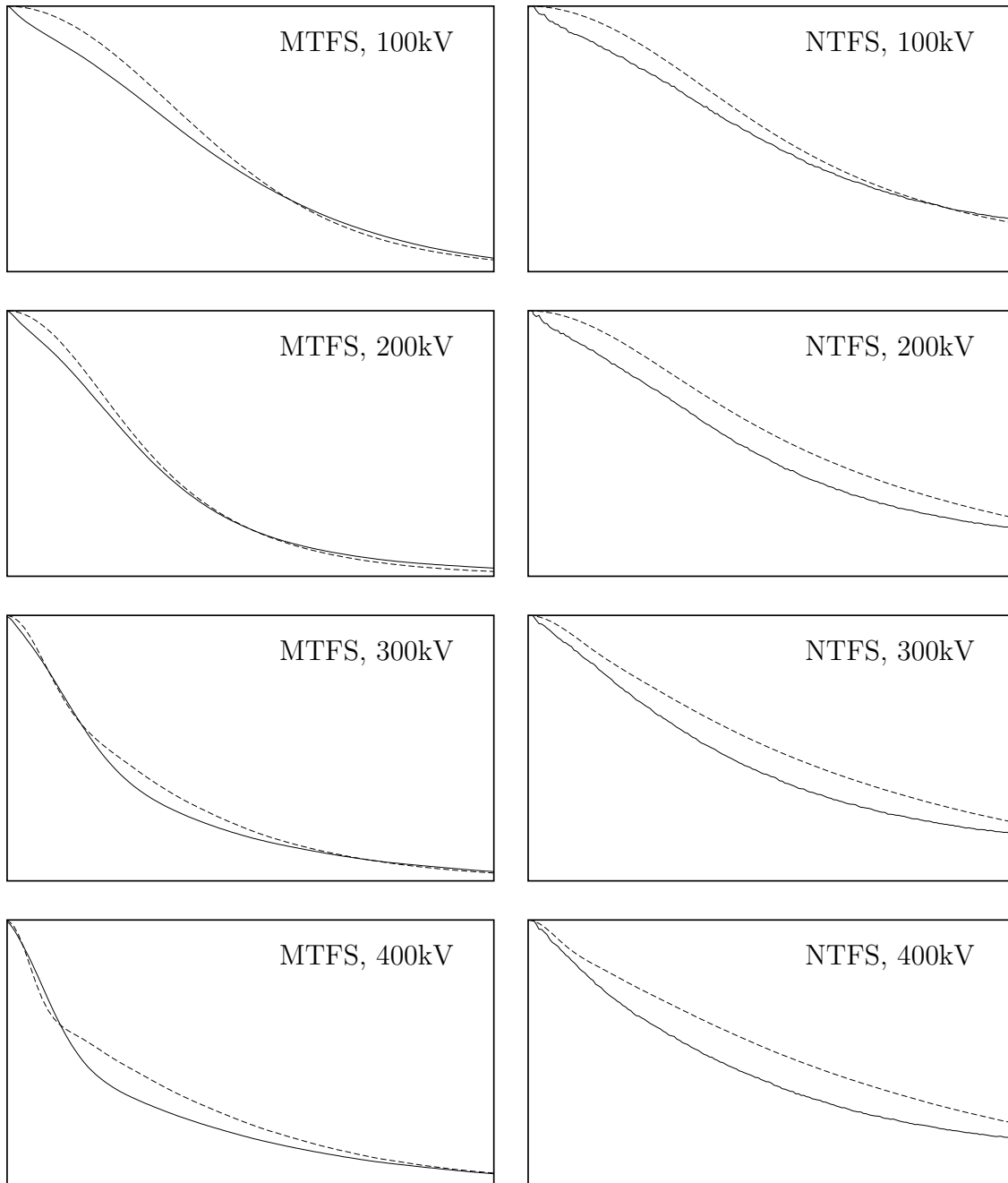
Figure 2.25: Comparison of the MTFS and NTFS measured for camera 3 (solid lines) with simulations (dashed lines). The parameters used for the simulations are given in table 2.3 (column 4).

| Camera | Gain G | DQE(0) | $N_{\mathrm{eff}}$ | $N_{\mathrm{eff}}/N_{\mathrm{tot}}$ | $\sigma_r$ |
|---|---|---|---|---|---|
| Camera 4 (1k) | 4.5 | 0.57 | 130000 | 12% | 1.8 |
| Camera 5 (2k) | 16.6 | 0.76 | 470000 | 11% | 1.1 |
| Camera 5 (2k), binned | 16.6 | 0.76 | 230000 | 22% | 1.1 |

Table 2.4: Key characteristics of camera 4 and camera 5, where $\sigma_r$ is the readout noise per (binned) pixel in digital numbers.

grain size and the mean free paths for light scattering and absorbtion are unknown proprietary information. Building a more accurate simulation model for phosphor scintillators would require an accurate characterisation of the optical properties of the phosphor first to reduce the number of unknowns in the Monte Carlo Simulation, which was beyond the scope of this work.

## 2.6.4  Comparison of two Gatan cameras at 300kV

This section summarises results obtained for two CCD cameras that are both mounted on a JEOL JEM-3000F FEG-TEM in Oxford (Meyer et al., 2000a). The first (camera 4) is a Gatan 794 Multiscan 1k×1k camera with a pixel spacing of 27 $\mu$m. The camera is retractable and mounted axially directly above the entrance aperture of a GIF 2000 imaging filter. The second device (camera 5) is a Gatan 794/IF20 Megascan 2k×2k camera with 30 $\mu$m pixel spacing located at the end of the GIF, behind a retractable TV camera. Both CCD cameras were equipped with polycrystalline phosphor scintillators fibre-optically coupled to the CCD chip.

The retractable TV camera was used as an edge for the MTFS measurement on camera 5. For the characterisation of camera 4, an aluminium insert containing an 1.2 mm thick aluminium edge was mounted on the lower flange of the microscope photographic plate camera.

The results of these measurements are shown in figure 2.26 and table 2.4. Camera 5 has a higher gain and a higher zero spatial frequency DQE, but a lower high spatial frequency MTFS than camera 4. This indicates that the scintillator of camera 5 is
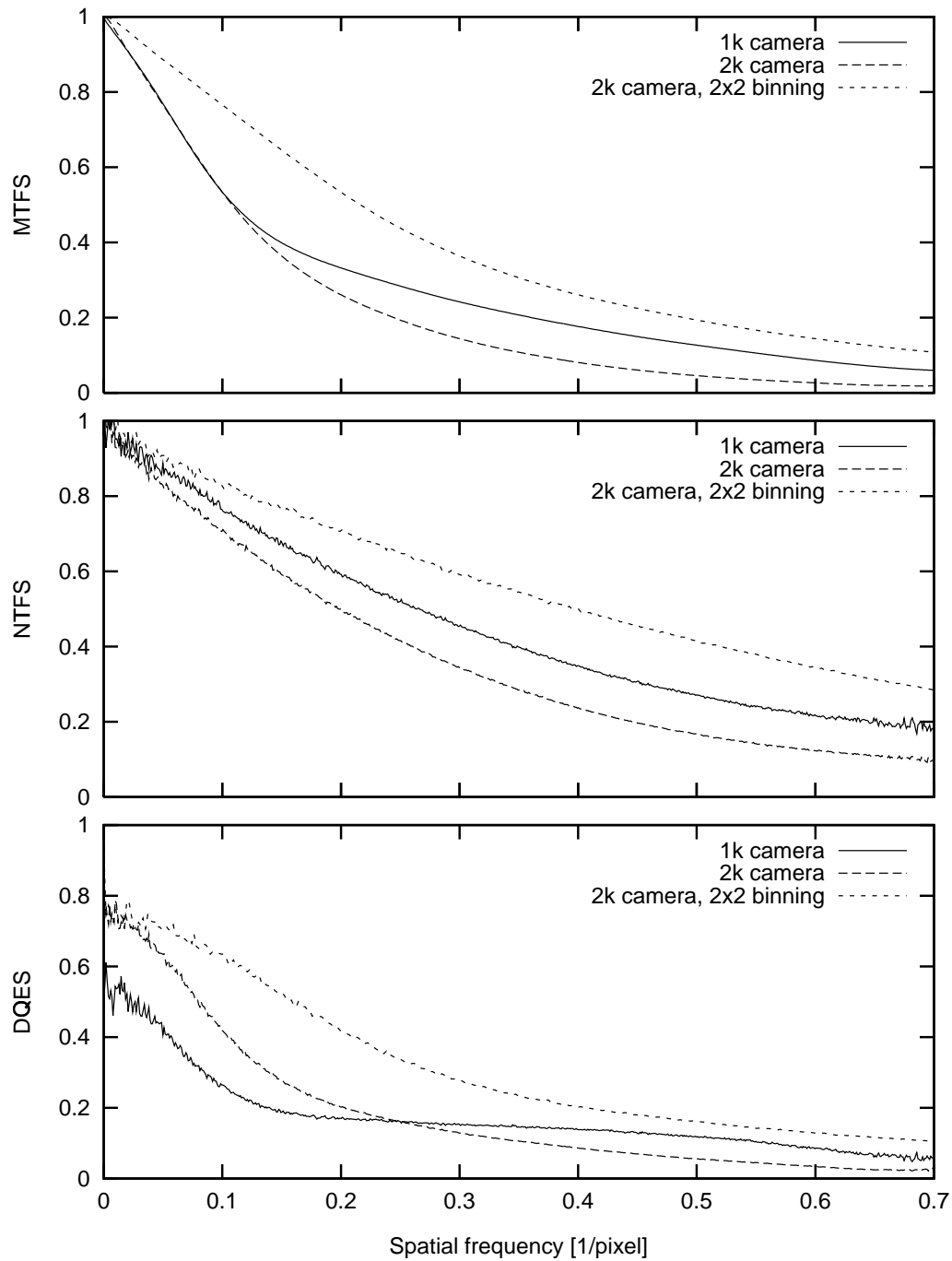
Figure 2.26: MTFS, NTFS and DQES of the post imaging filter Gatan 2k×2k camera (camera 5) and the axially mounted Gatan 1k×1k camera (camera 4) attached to the JEM3000F microscope in Oxford.
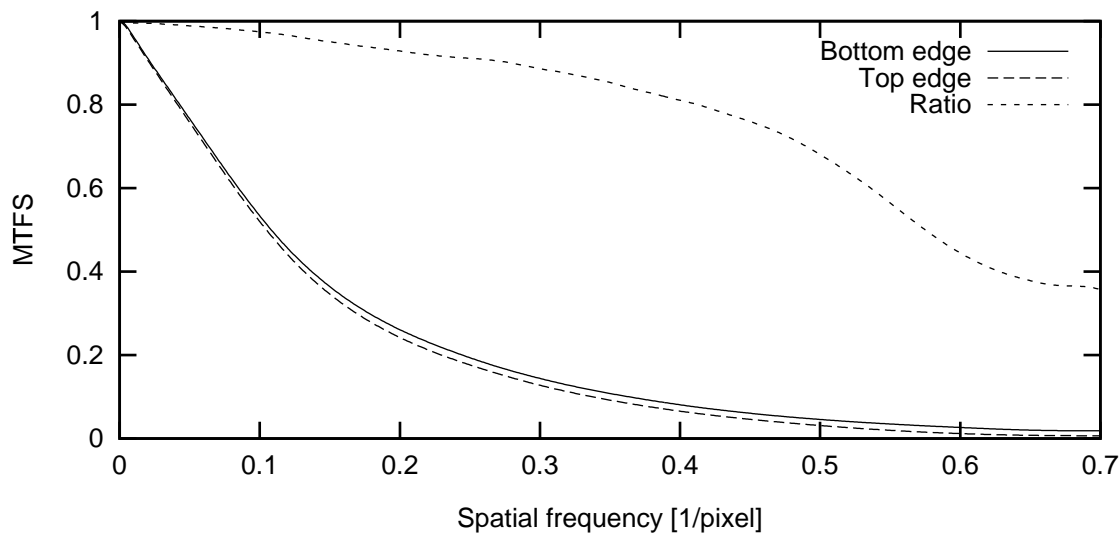
Figure 2.27: The MTFS of camera 5 is slightly lower when measured from the edge above the GIF, which indicates that the edge is not sufficiently sharp to be used with 20-fold magnification.

substantially thicker than for camera 4, which was confirmed by Gatan. Comparison with simulations gave estimated thicknesses of 20 $\mu$m and 40 $\mu$m, respectively.

In order to assess whether the measured MTFS values are underestimated because the knife-edges were insufficiently sharp, the MTFS of the lower camera was also determined using images of the upper edge. In these images, the edge is magnified 20-fold due to the electron-optical magnification of the GIF. Figure 2.27 shows that the MTFS is slightly under-estimated when determined from the upper edge, which implies that the edge is not sharp enough to be used at a 20-fold magnifications. In the shadow images used in the other measurements, the edge was at most magnified 3-fold, and as the ratio in figure 2.27 is larger than 0.97 for $k < 0.1$, this implies that the relative error due to insufficient edge sharpness in those measurements is at most 3%.

The total DQE, calculated according to equation 2.26 to include the additional noise due to aliasing is shown for both cameras in figure 2.28(a), and the effective

pixel number, calculated 2-dimensional summation over this function, is quoted in table 2.4.

At low electron dose, the readout noise[18] becomes important, reducing the DQE and the effective pixel number. The strength of the readout noise was determined by calculating the power spectrum of a difference image between two dark field images. An essentially white power spectrum was obtained,[19] from which the readout noise was measured, with the results quoted in table 2.4.

Figure 2.28(b) gives the total DQE of camera 4 at various electron doses, while figure 2.29 shows how the effective pixel number deteriorates for low electron doses. For very low electron doses, it can be beneficial to reduce the influence of the readout noise by pixel binning, as $n \times n$-binning increases the signal per binned pixel $n^2$-fold, while leaving the readout noise constant. This counterbalances the loss of information and additional noise from beyond the new Nyquist limit at very low doses, as shown in figure 2.29. It should however be noted that this applies only at extremely low doses, which are hardly used in electron microscopy and at which useful information can only be extracted using sophisticated averaging techniques.

### 2.6.5   A lens-coupled camera with phosphor scintillator

Measurements on a SIS Megaview camera (camera 4) were done while this camera was temporarily installed on a JEM2010 in the department of chemistry in Cambridge. The camera is a retractable device mounted on the 35 mm port. The phosphor scintillator is mounted on a glass prism that serves to deflect the emitted light by 90° sidewards into a standard photo objective, which focusses the phosphor screen onto the CCD chip with a magnification factor of $M = 0.7$. This demagnification is

---

[18]At the exposure times of about 1 s used, the dark current noise is small compared to the readout noise. This is confirmed by the fact that there was no measurable difference between the noise in dark images at exposure times of 0.1 s and 1 s, respectively.

[19]For most cameras this power spectrum also showed isolated peaks due to coupling of a parasitic 50 cycle signal into the readout signal. Therefore, the strength of the genuine readout noise was estimated from the power spectra, where these peaks can be eliminated, rather than directly from the standard deviation in the difference image.
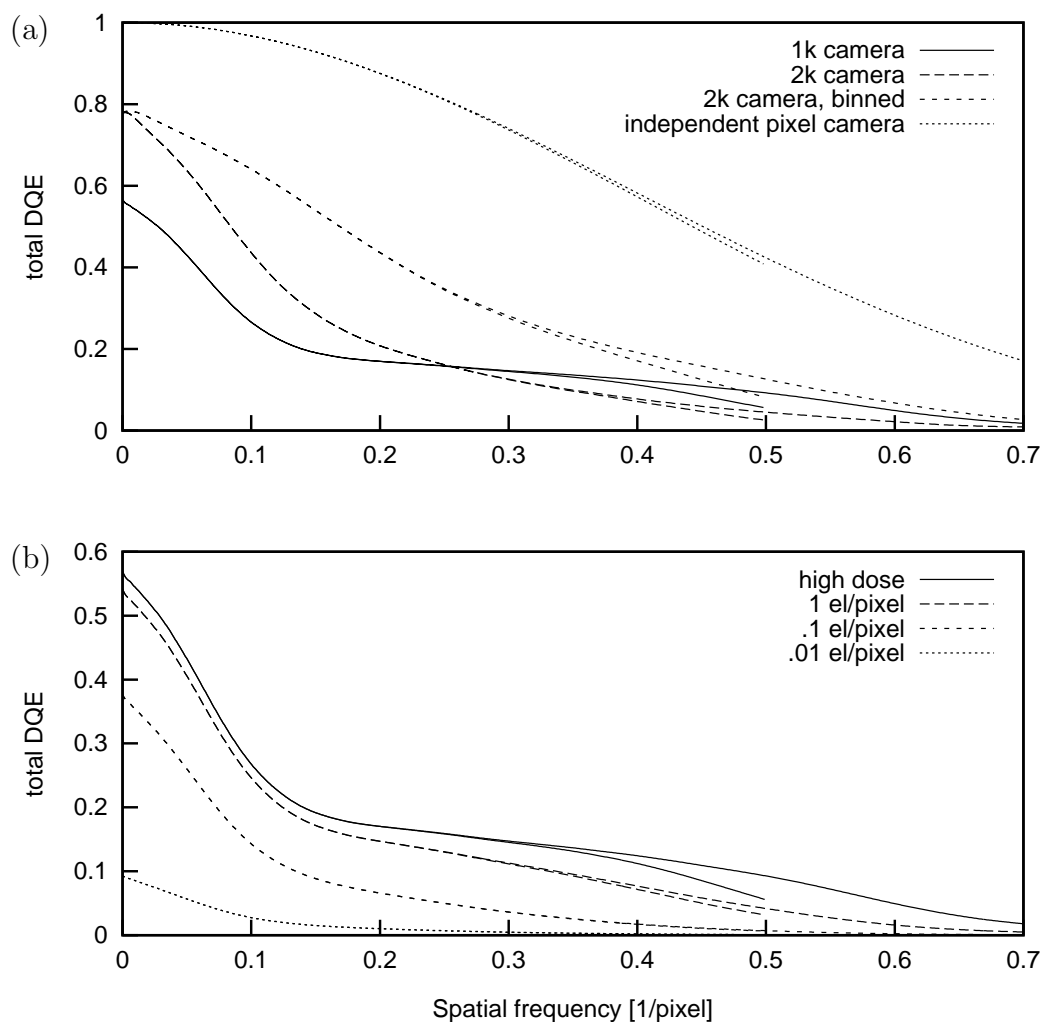
Figure 2.28: (a) Total DQE of camera 4 (1k×1k ) and camera 5 (2k×2k) including the additional noise due to aliasing. (b) The total DQE of camera 4 at different electron doses. At low dose conditions, the DQE is reduced as the readout noise becomes important.
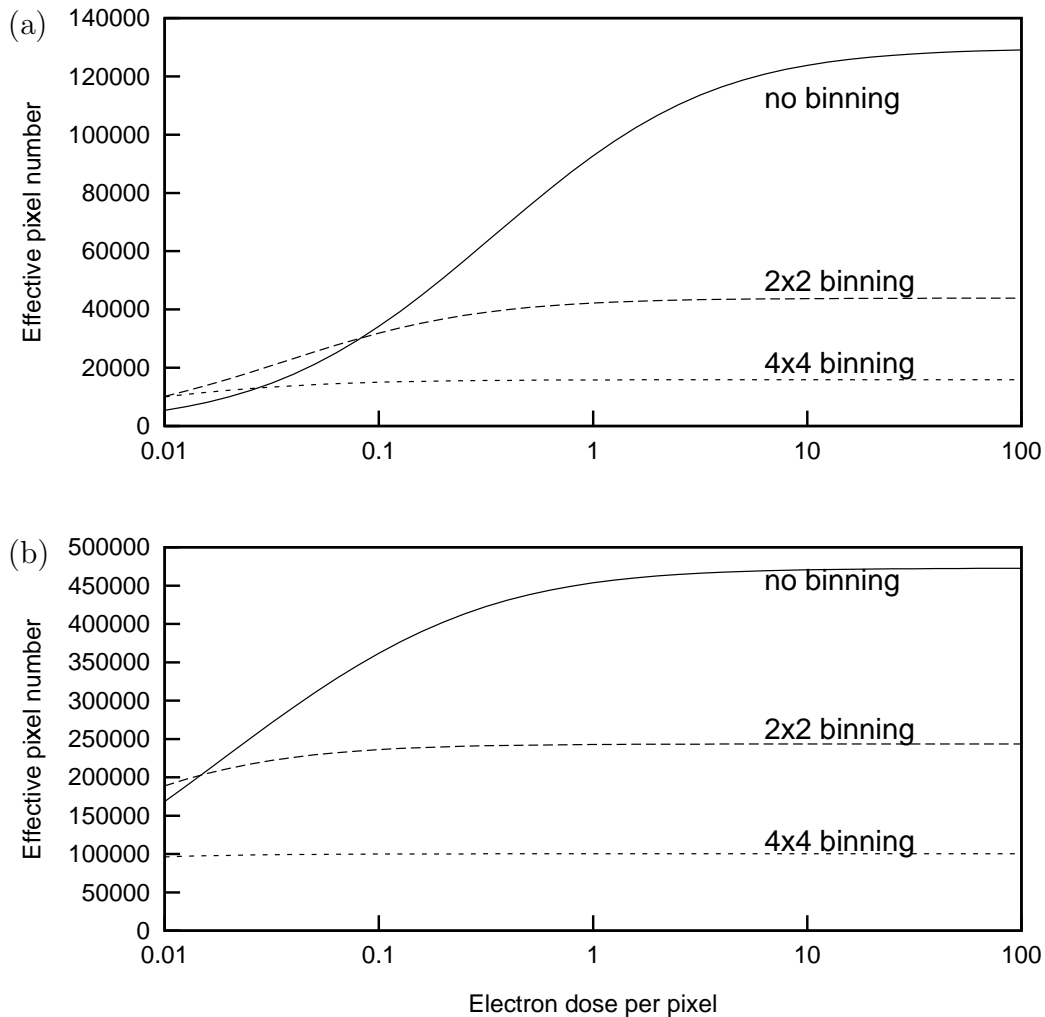
Figure 2.29: (a) Effective pixel number $N_{\text{eff}}$ as a function of electron dose for (a) camera 4 and (b) camera 5. At very low doses $N_{\text{eff}}$ is larger for $2 \times 2$ binning than without binning, as the fourfold increase in the electron dose per superpixel reduces the influence of the readout-noise and thereby compensates for the loss in resolution. Camera 5 features higher gain and lower readout noise than camera 4, therefore the deterioration of $N_{\text{eff}}$ at low dose is less pronounced.

used to improve the match between the small pixel size (14 $\mu$m) and the width of the scintillator point spread. Generally, the light collection efficiency of lens-coupled systems is inferior to fibre-optical systems. For a lens characterised by its focal length $l$ and its $f$-number[20] the lens-object distance is $l_o = l(M+1)/M$ and hence the aperture semiangle is given by

$$\tan\alpha = M/(2f(M+1)) . \tag{2.77}$$

The Megaview camera has a lens with $f = 1.4$ (at $l = 50$ mm) and hence an acceptance angle $\alpha_{\text{air}} = 8.4°$, which means, according to equation 2.54 that the collection efficiency is reduced to 1.7% of that achieved with a fibreplate with $N_A = 1$.[21]

The MTF measurements were carried with a custom-build edge mounted on top of the scintillator. The focus of the coupling lens was adjusted manually for each measurement by maximising the high frequency intensity in the power spectrum. Figure 2.30 shows the MTFS at accelerating voltages of 120, 160 and 200 kV. The graphs are very similar at high spatial frequencies, but the initial decline at low frequencies becomes stronger with increasing accelerating voltage as for camera 3.

Figure 2.31 shows that, after an initial exponential decay, the spectra approach a common asymptotic value of 0.19. This white noise background is due to the Poisson noise of the generated CCD-well electrons, which gives a constant contribution equal to $G_{DN,w}$, which is the gain in digital numbers per CCD well electron, to the noise power spectrum. From the asymptotic behaviour of the NPS graphs in figure 2.31, a value $G_{DN,w} = 0.19$ has been estimated.

This contribution can be subtracted from the NPS to yield a gain independent NPS:

$$\text{NPS}_g(k) = \text{NPS}(k) - G_{DN,w} \tag{2.78}$$

Simple exponential functions provided good fits to this $\text{NPS}_g$ for all voltages, and these fitted functions, together with $G_{DN,w}$, were used to calculate the DQES for different voltages as shown in figure 2.32. Due to the small gain and therefore the

---

[20]$f = l/d$ is defined as the ratio of focal length $l$ and pupil diameter $d$.

[21]The ratio given is based on a phosphor matrix refractive index of $n_{\text{scint}} = 1.2$. For YAG ($n_{\text{scint}} = 1.83$), this ratio is 2.0%.
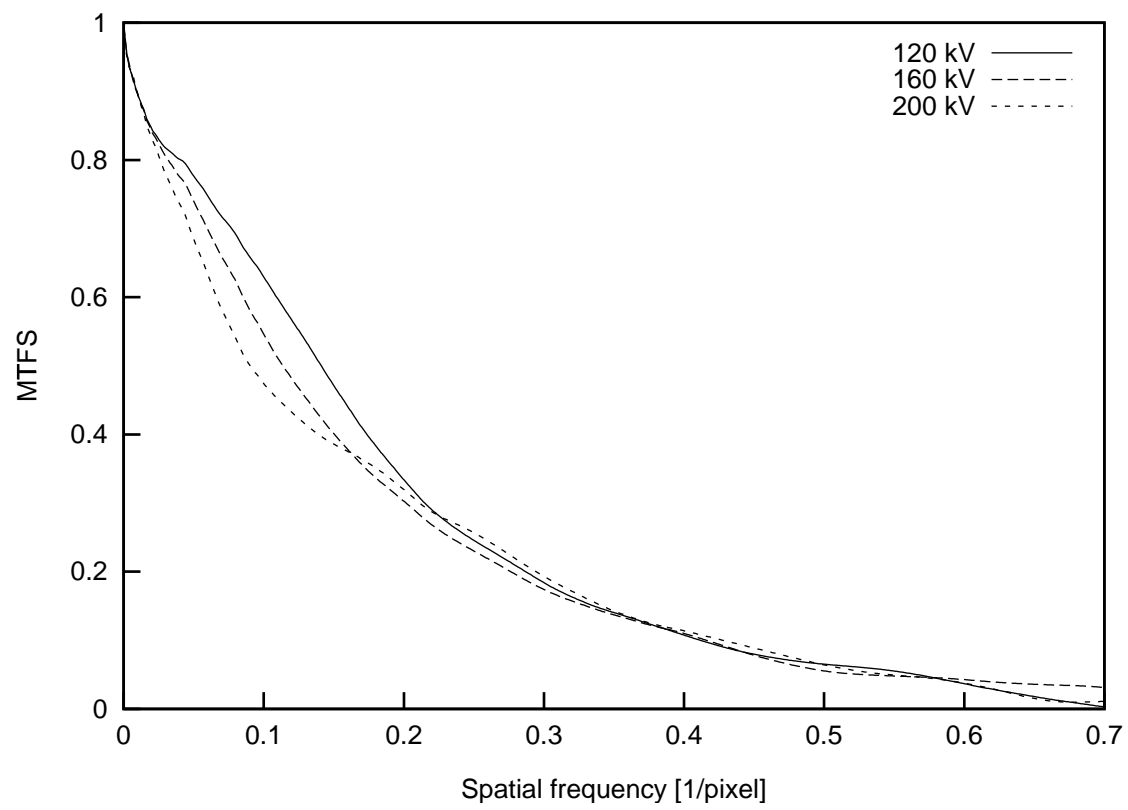
Figure 2.30: MTFS of the lens-coupled camera 6 at different voltages.
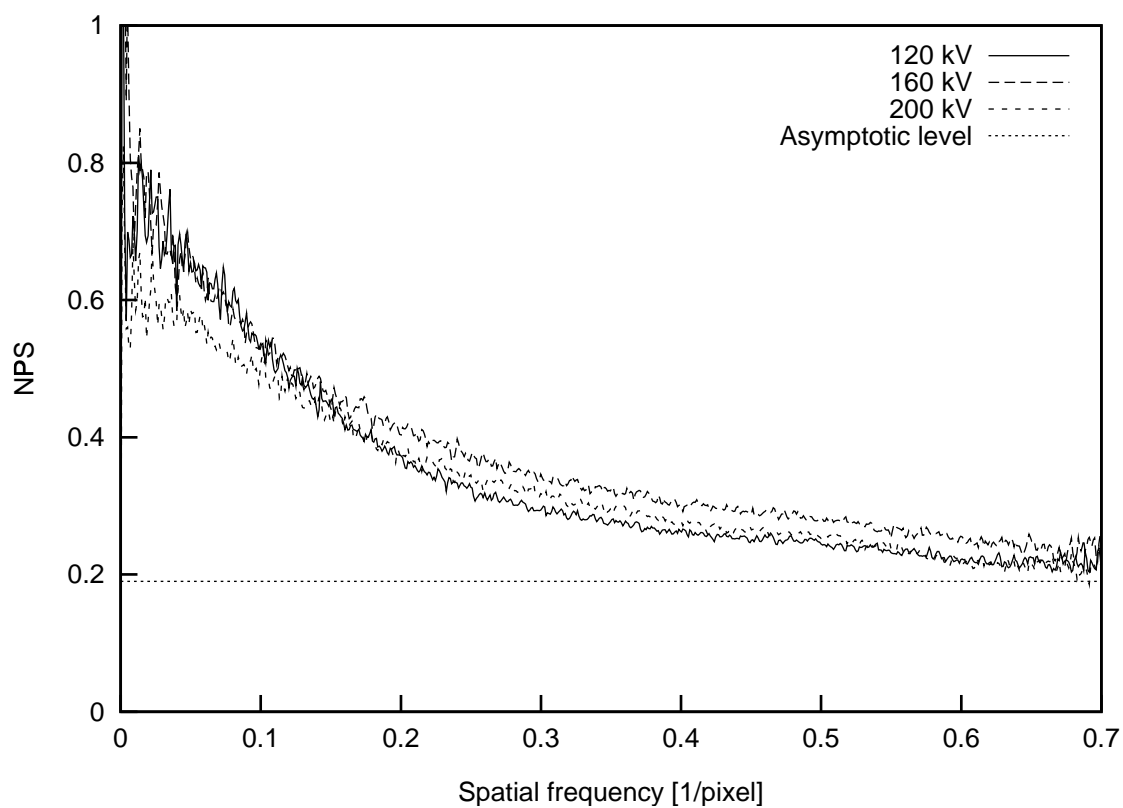
Figure 2.31: Noise power spectra at different voltages converging to a common asymptotic value due to the CCD well electron Poisson noise.
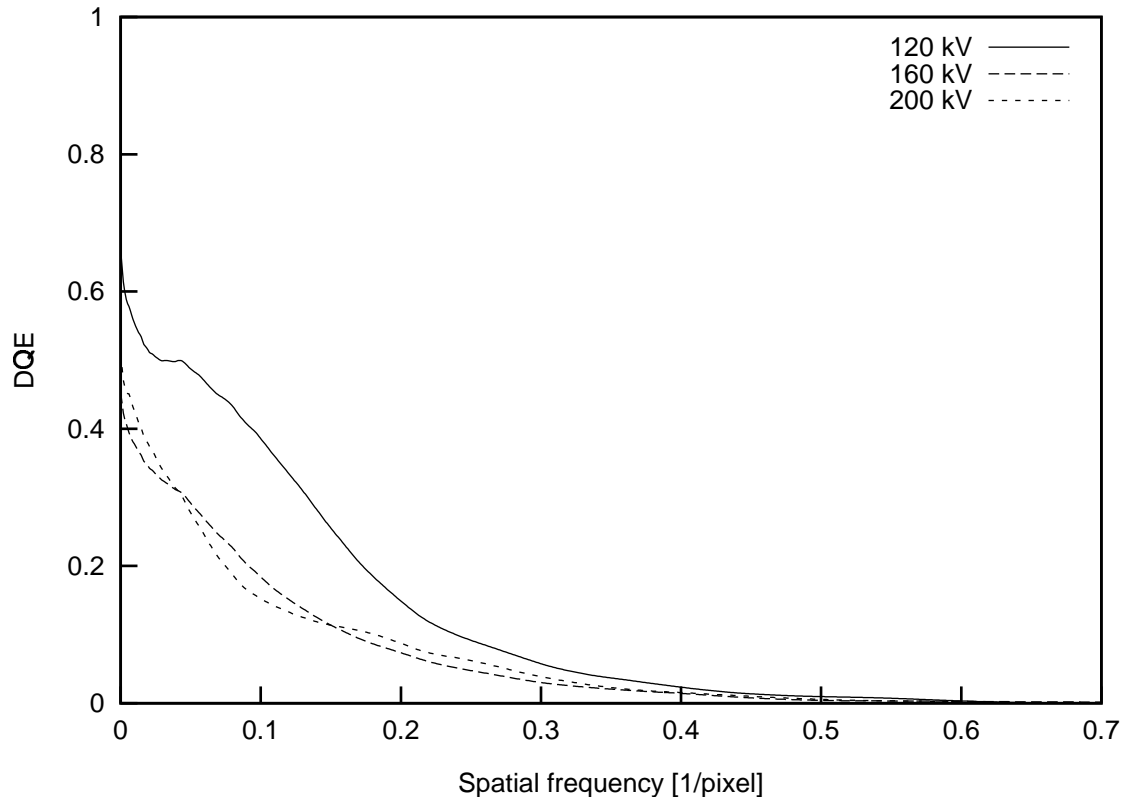
Figure 2.32: The DQES of camera 6 at high frequencies is mainly limited by the CCD well electron Poisson noise.

large CCD well electron Poisson noise, the effective pixel number, given in table 2.5, is lower than for the fibre-optically cameras.

The influence of this well electron Poisson noise on the DQE can be reduced by increasing the efficiency $G_{DN}$ of the coupling where $\text{NPS}_g(k)$ increases by the same factor, while $G_{DN,w}$ remains constant. Therefore, it is interesting to investigate if the improved resolution that is achieved by demagnification is counterbalanced by the loss in gain due to the smaller aperture angle according to equation 2.77. Figure 2.33 shows that the optimum trade-off between resolution and gain would be reached with $M = 0.35$, *i.e.* a 3-fold demagnification. This maximum in the effective pixel number would, however, move to larger values of $M$ under low-dose conditions. In addition,

| Camera | Gain G | DQE(0) | $N_{\text{eff}}$ | $N_{\text{eff}}/N_{\text{tot}}$ |
|--------|--------|--------|------|-----------|
| 120 kV | 0.52 | 0.66 | 65000 | 6% |
| 160 kV | 0.35 | 0.45 | 39000 | 3% |
| 200 kV | 0.33 | 0.51 | 41000 | 3% |

Table 2.5: Key characteristics of camera 6 at different voltages.

the choice of $M$ is restricted by other design constraints, like the maximum size of the phosphor screen and the choice of commercially available lenses.

## 2.7  Summary and discussion

The experimental results and simulations presented in this chapter demonstrate that it is very important to distinguish between signal and noise transfer for CCD cameras. It has been shown that both the modulation transfer function and the noise transfer function can be calculated by Monte-Carlo integration of the light distribution $g_\mu$ generated by single electrons over a large number of simulated trajectories $\mu$ using the expressions

$$\text{MTFS}(u,v) = \frac{\int \hat{g}_\mu(u,v)\,\mathrm{d}\mu}{\int \hat{g}_\mu(0,0)\,\mathrm{d}\mu} \tag{2.79}$$

$$\text{NTFS}(u,v) = \left[\frac{\int |\hat{g}_\mu(u,v)|^2\,\mathrm{d}\mu}{\int |\hat{g}_\mu(0,0)|^2\,\mathrm{d}\mu}\right]^{1/2}. \tag{2.80}$$

These expressions imply that the signal transfer depends on the area illuminated by many electrons with the same incident point, while the noise transfer depends on the average area illuminated by a single electron. Importantly, the difference between signal and noise transfer causes the detection quantum efficiency, given by

$$\text{DQES}(u,v) = \frac{|\int \hat{g}_\mu(u,v)\,\mathrm{d}\mu|^2}{\int |\hat{g}_\mu(u,v)|^2\mathrm{d}\mu} \tag{2.81}$$

to be strongly spatial frequency dependent.
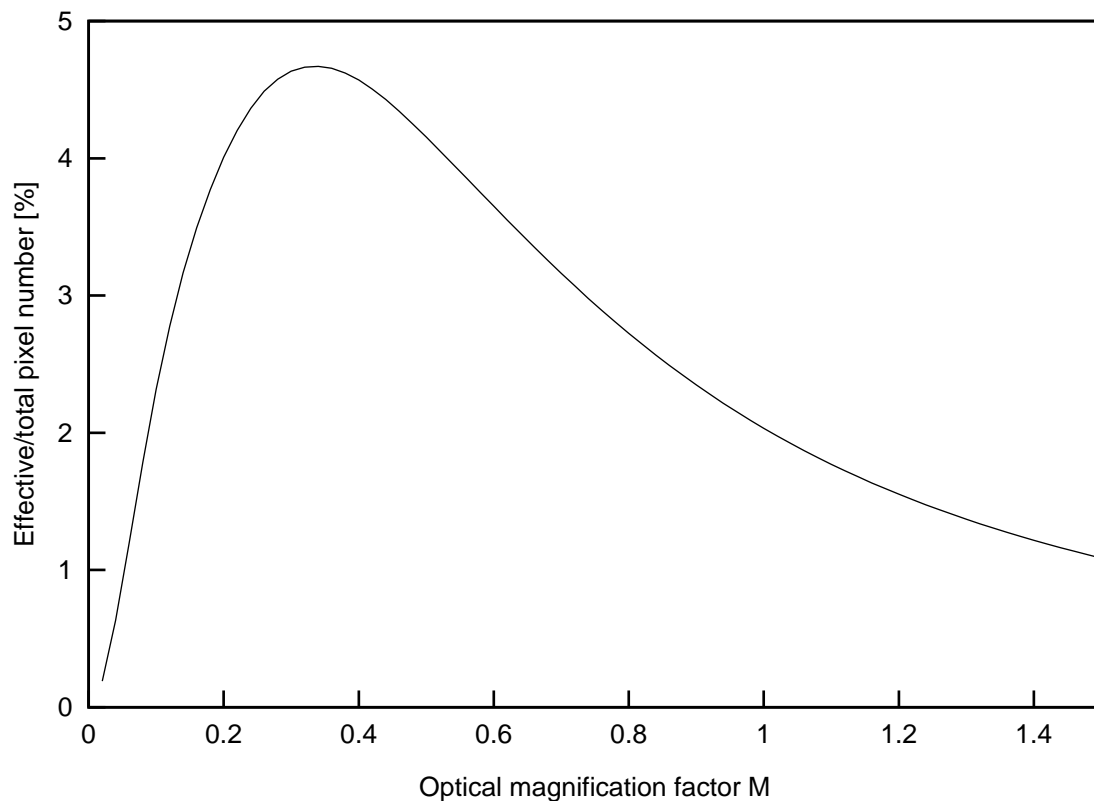
Figure 2.33: The effective pixel number $N_{\text{eff}}/N_{\text{tot}}$ as a function of the optical magnification factor $M$, predicted from the properties measured at $M = 0.7$. The maximum at $M = 0.35$ indicates an optimum trade-off between resolution (which increases with decreasing $M$) and collection efficiency (which increases with increasing $M$).

As a useful single figure of merit for pixelated image detectors, an "effective pixel number" can be defined by:

$$N_{\text{eff}} = \sum_{\mathbf{k}} \text{DQE}(\mathbf{k}) \; , \tag{2.82}$$

where the sum extends over all pixels in the discrete Fourier transform.

The MTFS, which is the MTF corrected for the sampling effects of the camera, can be measured using images of a sharp edge above the camera. Having the edge inclined with respect to the pixel columns allows the accurate extraction of an oversampled edge profile to avoid aliasing artefacts.

The measurement of the NTFS requires series of images with uniform illumination. An averaged noise power spectrum is computed from power spectra of differences between the images of the series after gain normalisation. Subsequently, the effect of aliasing is corrected and the square root of the rotational average is normalised to give the NTFS.

In agreement with the simulations, experimental results with different cameras demonstrated a strong discrepancy between MTFS and NTFS, which is increasingly severe for increasing electron energy.

This strong discrepancy at high voltages, even at moderate spatial frequencies, is due to back-scattering of electrons into the scintillator from the fibre optic coupling and causes the effective pixel number to be only about 10% of the actual pixel number for electron energies of 300 and 400 keV, whereas at 100 keV effective pixel numbers of up to $N_{\text{eff}} = .37 N_{\text{tot}}$ were measured. Independently of this work, Downing and Hendrickson (1999) also found at 400 keV that the signal transfer is much poorer than the noise transfer. This 'high-voltage problem' has previously been largely obscured since the noise method has often been used to measure the MTF (Zuo, 1996, 2000; de Ruijter, 1995; Rabbani and van Metter, 1989; Sherman et al., 1996; Fan et al., 2000).

Possible instrumental solutions to improve the DQE include the use of a tapered fibre optics to enable the use of even larger scintillator areas with available CCD chips

(Faruqi et al., 1999), or to use self-supporting lens-coupled scintillators (Fan et al., 2000).

# Chapter 3

# Automated aberration determination: existing methods

## 3.1 Minimum variance method

The earliest method used used for automated alignment is the minimum variance method (Erasmus and Smith, 1982; Saxton et al., 1983). This relies on the fact that the image variance for weak phase objects has an absolute minimum when the misalignment and astigmatism are zero and the focus is Gaussian. The method requires very little image processing and hence was the only one viable for on-line automated alignment at the time. The main disadvantage is that the method is not predictive, *i.e.* the minimum has to be found by trial and error, which is only possible in a reasonable amount of time when the unknown parameters[1] are initially close to their correct values. Furthermore the method is only applicable to weak phase objects. Within these constraints, the method still proved useful as an automated aid for the fine-tuning of the microscope alignment.

The same method has also been used to determine the residual aberrations in

---

[1]In practice, five variable parameters have to be adjusted routinely. These are the focus ($C_1$) and the two components of the two-fold astigmatism ($A_1$) and the axial coma ($B_2$).

restored wavefunctions obtained using holography (Fu et al., 1991) or from focal series (Tang et al., 1996) by minimising the variance in the restored modulus for an amorphous specimen region. The fact that the method is not predictive in this case increases the computation time rather than the exposure time. However, the computational effort required for each set of trial parameters now comprises a multiplication with a correction phase plate and an inverse Fourier transform in addition to the simple variance determination, so this disadvantage can be quite substantial, even at the computing speeds now available. Advanced multi-parameter optimisation algorithms such as the genetic algorithm are therefore necessary to find the correct aberration parameters from arbitrary starting positions (Lehmann, 2000).

## 3.2   Tilt induced displacement measurements

In this method, the required aberration parameters are derived from measurements of the image displacements induced by different injected beam tilts. The measurement of these image shifts requires the use of cross-correlations and hence more computing power than the minimum variance method. The method was first introduced in 1987 (Koster et al., 1987) and more recent accounts can be found in Koster et al. (1989, 1992) and Typke and Dierksen (1995).

According to equation 1.5, the shift $A$ of an image recorded with tilt $\tau$ relative to the axial image is given by

$$A = A_1 \tau^* + C_1 \tau + A_2 \tau^{*2} + \frac{1}{3} B_2^* \tau^2 + \frac{2}{3} B_2 \tau^* \tau + C_3 \tau^* \tau^2 \; . \tag{3.1}$$

As this depends on all aberration coefficients, these can be determined by minimising the least-squares difference between the measured and expected image shifts, provided that a suitable set of beam tilts is used.

For sets of $N$ beam tilts of the form $\tau_n = \tau_0 e^{2\pi i n/N}$, the parameters can be separated analytically by calculating Fourier sums of the observed shifts $A^{(n)}$, defined

by:

$$A^{((m))} = \frac{1}{N} \sum_{n}^{N-1} e^{-2\pi i n m/N} A^{(n)} \, , \tag{3.2}$$

which automatically separates parameters of different azimuthal symmetry as:

$$
\begin{aligned}
A^{((0))} &= \frac{2}{3} B_2 |\tau_0|^2 \\
A^{((1))} &= C_1 \tau_0 + C_3 |\tau_0|^2 \tau_0 \\
A^{((-1))} &= A_1 \tau_0^* \\
A^{((2))} &= \frac{1}{3} B_2^* \tau_0^2 \\
A^{((-2))} &= A_2 \tau_0^{*2} \\
A^{((m))} &= 0 \; : \; m \notin \{-2, -1, 0, 1, 2\} \, . \tag{3.3}
\end{aligned}
$$

The above equations apply for $N \geq 5$ but for smaller $N$, the contributions from coefficients that are equivalent modulo $N$ have to be added. This implies that

$$A^{((2))} = A^{((-2))} = \frac{1}{3} B_2^* \tau_0^2 + A_2 \tau_0^{*2} \quad \text{for} \quad N = 4 \, . \tag{3.4}$$

However, the parameters $A_2$ and $B_2$ that appear jointly in this expression can still be separated with the help of $A^{((0))}$ that contains $B_2$ only. A separation of the imaging parameters from observations of image displacements for fewer tilt azimuths $N < 4$, however, is not possible.

The paramters $C_1$ and $C_3$ cannot be separated when beam tilts with only one magnitude are used. The direction of $\tau_0$, *i.e.* the rotation between the tilt coil axes and the camera axes can determined from $A^{((1))}$, however a 2-fold ambiguity in this direction remains unless the sign of $C_1 + C_3|\tau_0|^2$ is known.

The accuracy to which the parameters can be estimated for a given accuracy $\sigma$ in the shift measurements is given in table 3.1.

Using $N > 5$ tilt azimuths not only reduces the error in proportion to $1/\sqrt{N}$, but also makes the dataset over-redundant and makes it possible to estimate the error $\sigma$ from the coefficients $A^{((m))}$ that should be zero:

$$\sigma = \frac{1}{N-5} \left[ \sum_{m=3}^{N-3} \left( A^{((m))} \right)^2 \right]^{1/2} \, . \tag{3.5}$$

| $N$ | $\Delta C_1$ | $\Delta A_1$ | $\Delta A_2$ | $\Delta B_2$ |
|---|---|---|---|---|
| $N = 4$ | $1/2\ \sigma/\tau$ | $1/2\ \sigma/\tau$ | $3/4\ \sigma/\tau^2$ | $\sqrt{5}/4\ \sigma/\tau^2$ |
| $N \geq 5$ | $1/\sqrt{N}\ \sigma/\tau$ | $1/\sqrt{N}\ \sigma/\tau$ | $3/\sqrt{5N}\ \sigma/\tau^2$ | $1/\sqrt{N}\ \sigma/\tau^2$ |

Table 3.1: Accuracy to which parameters can be estimated from $N$-azimuth tilt induced displacement measurements with tilt magnitude $\tau$ when the error in the measured image shifts is $\sigma$.

A practical difficulty with the application of the tilt induced displacement is the measurement of the image shifts using the peak position in the cross-correlation function

$$\mathrm{XCF}(\mathbf{r}) = FT^{-1}[c_1^* c_2] \tag{3.6}$$

The tilt-induced change of the parameter $A_0$ in equation 1.5 introduces a linear phase variation in the cross spectrum $c_1^* c_2$, which leads to a displacement of the XCF peak to a position given by the shift vector between the images. However, equation 1.5 also implies that the other imaging parameters also change as the beam is tilted. This causes the phase-variation to become non-linear at higher spatial frequencies and leads to distorted cross-correlation peaks, which in general do not have their highest point at the peak centre. However, when the imaging conditions in both images are approximately known, the non-linear phase shifts can be compensated and a sharp XCF peak can be restored (Saxton, 1994a). Another problem with the tilt-induced shift method is that the displacement by specimen drift is indistinguishable from the tilt-induced displacement to be measured. This is why this method is most frequently used at low resolution or for coarse alignment at high resolution (Krivanek and Leber, 1994). Recently, however, Steinecker and Mader (2000) used the tilt induced displacement method to determine the aberrations to an accuracy sufficient for 0.1 nm resolution.
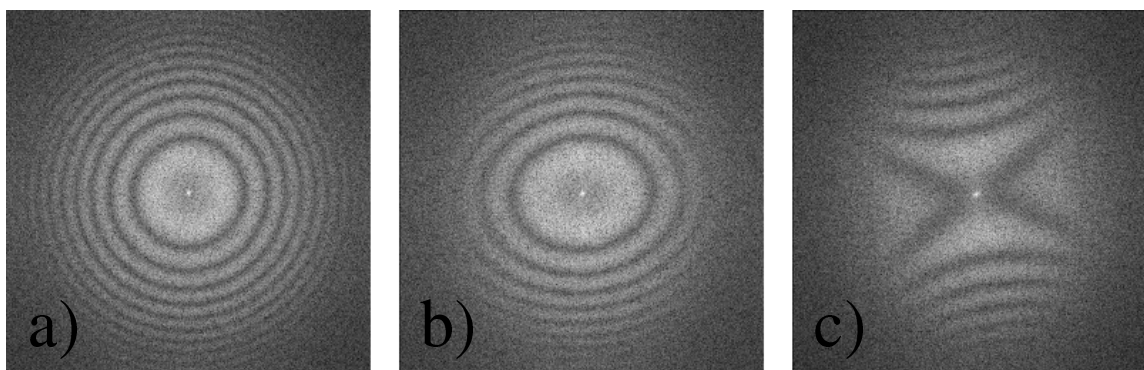
Figure 3.1: Typical diffractograms from an amorphous germanium foil. (a) Axial image with large underfocus ($C_1 = -300$nm) (b) The beam is tilted along the horizontal axis, giving rise to apparent astigmatism. (c) At larger beam tilts, the induced astigmatism is so large that the focus along the axis parallel to the tilt direction changes to overfocus, giving rise to a cross shaped diffractogram.

## 3.3 Diffractogram based methods

The *optical diffractogram* is defined as the power spectrum of the recorded image intensity. As discussed in section 1.5.1, this diffractogram displays characteristic dark rings at the zero crossing of the phase contrast transfer function for weak phase objects (figure 3.1). Traditionally, diffractograms were generated by light optical means as diffraction patterns of micrographs observed on an optical bench. This technique has now been superseded by numerical calculation of the diffractogram using fast Fourier transforms of digitally recorded data. With modern computers, it is possible to perform this calculation fast enough to display a diffractogram on the computer screen in real time at 25 frames per second, which provides a valuable tool for manual focusing, stigmation and alignment. The use of diffractograms for the determination of aberrations was first suggested by Thon (1966), who devised a method of measuring the defocus from the ring positions in the diffractogram. The method was later extended to additionally measure the spherical aberration $C_3$ by Krivanek (1976). Coene and Denteneer (1991) improved the accuracy of the $C_3$ determination using a

focus series of diffractograms under axial illumination. Zemlin et al. (1978) used a tableau of diffractograms acquired with different beam tilt azimuths to additionally access the antisymmetric aberrations, *i.e.* axial coma $B_2$ and three-fold astigmatism $A_2$. They found $A_2$ to be insignificant, but substantial axial coma was measured in a microscope aligned using the then common current reversal centre alignment procedure.[2] However, the diffractogram tableau method was too time-consuming for routine use at the time and was only used to demonstrate that the alignment achieved with the current-reversal method was inadequate and to introduce a new alignment method based on the visual comparison of the contrast observed from an amorphous specimen at equal and opposite beam tilts (Zemlin, 1979).

### 3.3.1   Automated diffractogram fitting

Fitting the parameters $C_1$ $A_1$ such that simulated diffractogram patterns match the experimental ones can be done manually to relatively high accuracy (Chand, 1997). This, however, is a lengthy and tedious process and automation of this task is highly desirable. The major challenges for implementing this are as follows:

1. The most abundant amorphous material in the microscope, carbon, is a very weak scatterer. Hence, especially at high spatial frequencies, the signal is weak compared to the noise background.

2. The strength of the observed signal depends on the object as well as the phase contrast transfer function.

3. Diffractograms taken close to Scherzer or Gaussian conditions show no rings and are difficult to fit.

---

[2]The *current reversal centre* alignment involves reversing the current of the objective lens and is no longer practicable with the strong lenses used in modern instruments. This is not to be confused with the *current centre* alignment, where the objective lens current is oscillated by a small amount. Similarly, in the *voltage centre* alignment, the high tension is oscillated. Generally, the axes found by the three methods are distinct (Chau, 1993) and usually the voltage centre is a workable approximation to the *coma-free axis*, but for lattice images at high resolution coma-free alignment should be used.

4. For some conditions, the diffractograms are cross rather than ring shaped (figure 3.1(c)), which leads to difficulties for algorithms based on the evaluation of rotationally averaged sectors of the diffractogram.

5. It is sometimes difficult to distinguish between over and underfocus.

6. Most automated algorithms fail when a significant amount of crystalline material is present, leading to strong reflections at positions unrelated to the ring pattern.

.

Automation of the diffractogram analysis was first tackled by Baba et al. (1987). Fan and Krivanek (1990) present an algorithm where the diffractogram is divided into 32 sectors and the defocus along the directions in each sector is determined by cross-correlating the rotational sector average with an array of theoretical diffractograms. A sinusoidal focus variation as a function of the azimuth angle is fitted to the measured focus values to determine the defocus and the astigmatism. Using this automated diffractogram analysis, automatic alignment is achieved using a tableau with 4 tilt azimuths (Krivanek and Fan, 1992). Since Zemlin's original paper, the resolution had meanwhile improved sufficiently for $A_2$ to become important and the accuracy of the alignment was greatly improved when the hitherto neglected $A_2$ term was included into the auto-alignment procedure (Krivanek and Leber, 1994).

Many of the practical problems associated with diffractogram fitting are avoided in an alternative method (Saxton, 2000a), which relies only on the determination of the orientation of a set of diffractograms at different tilt azimuths. However, $C_1$ cannot be determined and because less information from each diffractogram is used, more tilt azimuths are required than with conventional methods. Furthermore, the presence of crystalline material hampers the automated determination of the symmetry axis of the diffractograms.

The advent of $C_3$ corrected microscopes (Haider et al., 1998; Urban et al., 1999) has made automated aberration determination even more important because the non-round lens elements introduce a multitude of high-order aberrations (up to six-fold astigmatism) that have to be corrected in an elaborate alignment procedure. Uhlemann and Haider (1998) have developed an algorithm that can evaluate the apparent defocus and astigmatism from a diffractogram within 400ms. This is based on a comparison of the experimental diffractogram with a large database of pre-calculated diffractograms and is fairly robust to disturbances from the presence of crystalline material (Haider, 2001).

### 3.3.2 The antisymmetric aberrations

As the antisymmetric aberrations have no influence on whether the object phase contrast is transferred into an amplitude contrast in the image, their effect is not visible in a diffractogram of a single axial image. However, the change in the observed defocus and astigmatism when the beam is tilted by $\tau$ does depend on the antisymmetric aberrations. According to equation 1.5 tilting the beam by $\tau$ yields values for the apparent defocus and astigmatism given by:

$$A_1' \;=\; A_1 + 2A_2\tau^* + \frac{2}{3}B_2\tau + C_3\tau^2 \tag{3.7}$$

$$C_1' \;=\; C_1 + \Re(\frac{4}{3}B_2\tau^*) + 2C_3\tau^*\tau = C_1 + \frac{2}{3}(B_2\tau^* + B_2^*\tau) + 2C_3\tau^*\tau \; . \tag{3.8}$$

Therefore the antisymmetric aberrations can be measured using tableaus of diffractograms taken with different injected beam tilts. Equations giving estimates for these aberrations from the measured values of the apparent defocus and astigmatism have been given elsewhere (Zemlin et al., 1978; Saxton, 1995). Here, a derivation based on on Fourier sums of these parameters is given for datasets comprising $N$ images with equal tilt magnitudes at regularly spaced tilt azimuths ($\tau_n = \tau_0 e^{2\pi i n/N}$). If the observed aberrations are denoted $C_1^{(n)}$ and $A_1^{(n)}$, Fourier sums of these parameters

can be defined as:

$$A_1^{((m))} = \frac{1}{N} \sum_{n=0}^{N-1} e^{-2\pi imn/N} A_1^{(n)} \tag{3.9}$$

$$C_1^{((m))} = \frac{1}{N} \sum_{n=0}^{N-1} e^{-2\pi imn/N} C_1^{(n)} \tag{3.10}$$

Equations 3.7 and 3.8 for $N \geq 3$ then yield

$$A_1^{((0))} = A_1 \tag{3.11}$$

$$A_1^{((1))} = \frac{2}{3} B_2 \tau_0 \tag{3.12}$$

$$A_1^{((2))} = C_3 \tau_0^2 \tag{3.13}$$

$$A_1^{((-1))} = 2A_2 \tau_0^* \tag{3.14}$$

$$A_1^{((n))} = 0 : n \notin \{-1, 0, 1, 2\} \tag{3.15}$$

$$C_1^{((0))} = C_1 + 2C_3 \tau_0^* \tau_0 \tag{3.16}$$

$$C_1^{((1))} = \frac{2}{3} B_2 \tau_0 \tag{3.17}$$

$$C_1^{((-1))} = \left( C_1^{((1))} \right)^* \tag{3.18}$$

$$C_1^{((n))} = 0 : n \notin \{-1, 0, 1\} \tag{3.19}$$

Extraction of the parameters is now straightforward; $B_2$ can be evaluated from both the $C_1$ and $A_1$ measurements with an optimum estimate using both given by

$$B_2 = \frac{2}{3\tau} \frac{\sigma_C^2 A_1^{((1))} + \sigma_A^2 C_1^{((1))}}{\sigma_A^2 + \sigma_C^2} , \tag{3.20}$$

where $\sigma_C$ and $\sigma_A$ are the errors in the measured values for apparent defocus $C_1$ and complex astigmatism $A_1$, respectively. Two-fold and three-fold astigmatism are determined as $A_1 = A_1^{((0))}$ and $A_2 = A_1^{((-1))}/(2\tau_0^*)$ from the tilt-induced astigmatism measurements. The axial value of $C_1$ is best determined from an axial image, as the estimate $C_1 = C_1^{((0))} - C_3 |\tau_0|^2$ requires an accurate knowledge of $C_3$ and the tilt strength.

The angle between the tilt $x$-axis and the image $x$-axis, $\arg(\tau_0)$, can be determined from $A_1^{((2))}$, however, the sign of $\tau_0$ remains undetermined and the equations are

invariant under a simultaneous change of the signs of $\tau_0$, $A_2$ and $B_2$. Furthermore, it has been assumed in equations 3.11–3.19 that $\tau_n$ progresses counterclockwise when $n$ is increased. When the direction of the $y$ tilt coil is reversed, the opposite is the case and the sign of $n$ has to be changed in the above equations. Both cases can easily be distinguished for $N > 4$ by comparing $A_1^{((2))}$ and $A_1^{((-2))}$, but this ambiguity remains for $N = 4$. Both ambiguities can be resolved using tilt-induced shift measurements.

The measurements of $|\tau_0|$ and $C_3$ cannot be separated, hence $C_3$ can only be measured when the tilt coil strength is calibrated accurately. This calibration is achieved by measuring the shift of the diffraction pattern of a known crystal when the beam is tilted. As the apparent strength of the tilt coils for the JEOL3000F was found to be a sensitive function of the lower condensor lens settings, these settings must be identical for calibration and experiment. On the other hand, if $C_3$ is known accurately, $|\tau_0|$ can be determined along with all the other parameters. Expressions for the accuracy of the determined parameters are given in table 3.2 in terms of the accuracies $\sigma_C$, $\sigma_A$ of the estimates for the apparent $C_1$ and $A_1$ values. Finally, for over-determined datasets, the RMS difference between the experimental values of $C_1^{(n)}$, $A_1^{(n)}$ and the values predicted from the fitted parameters according to equations 3.7 and 3.8 is a measure of the quality of the fit. This can be used to estimate the accuracy of the underlying $C_1$ and $A_1$ values as

$$\sigma_C = \frac{1}{N-3} \left[ \sum_{m=2}^{N-2} \left( C_1^{((m))} \right)^2 \right]^{1/2} \tag{3.21}$$

$$\sigma_A = \frac{1}{N-4} \left[ \sum_{m=3}^{N-2} \left( A_1^{((m))} \right)^2 \right]^{1/2}. \tag{3.22}$$

## 3.4   Phase cross-spectrum analysis

According to equation 1.5, all aberration parameters change when the beam is tilted. In the cross-spectrum of two images recorded with different beam tilts the change

| $\Delta C_1$ | $\Delta A_1$ | $\Delta A_2$ | $\Delta B_2$ |
|---|---|---|---|
| $\sigma_C/\sqrt{N}$ | $\sigma_A/\sqrt{N}$ | $\sigma_A/(2\tau\sqrt{N})$ | $3\sigma_A\sigma_C/[2\tau\sqrt{N(\sigma_A^2+\sigma_C^2)}]$ |

Table 3.2: Accuracy to which the aberrations can be estimated from an $N$-azimuth ($N \geq 4$) diffractogram tableau with tilt magnitude $\tau$ when the errors for the measured values of $C_1$ and the complex $A_1$ are $\sigma_C$ and $\sigma_A$, respectively.

in $A_0$ leads to a phase variation that is linear in $\mathbf{k}$. However, due the change in the other parameters, there is also a nonlinear phase variation, which is a hindrance in the application of the tilt induced displacement method presented in section 3.2. This non-linear phase variation in the cross spectrum between two images taken at different incident beam tilt can also be exploited for aberration determination. This was first suggested by Koster et al. (1989), who derives expressions for the phase cross spectrum expected for weak phase objects. The idea was successfully tested in practice by Ichise et al. (1997), who were able to measure defocus $C_1$, astigmatism $A_1$ and coma $B_2$ using the phase cross spectra of two images with equal beam tilts in perpendicular directions with an axial image. The method was found to give accurate results with an amorphous germanium foil as a sample, however, even with this ideal sample, the features in the phase cross spectrum are difficult to recognise due to the presence of noise and it was not confirmed if the method was still applicable when smaller amounts of amorphous material were present. An additional problem with this method is that all unknown parameters have to be estimated simultaneously by fitting a predicted phase cross spectrum with the experimental one. This makes the method too slow to be considered for on-line use.

## 3.5 Holographic measurement using Fourier transform phases

Fu and Lichte (1995) present a method to measure the symmetric part $\gamma_s$ of the wave aberration $\gamma$ directly rather than indirectly through the coefficients. This method requires the holographic recording of the image wave for an amorphous weak phase object. For such an object, all spatial variation of the object wave is purely imaginary, therefore its Fourier transform fulfils the symmetry relation

$$\psi_{\mathrm{obj}-} = -\psi_{\mathrm{obj}}^* \, . \tag{3.23}$$

In terms of the image wave $\psi$, this implies that

$$\psi_- \mathrm{e}^{-\mathrm{i}\gamma_-} = -\psi^* \mathrm{e}^{\mathrm{i}\gamma} \tag{3.24}$$

$$\mathrm{e}^{2\mathrm{i}\gamma_s} = \psi\psi_- \big/ |\psi(\mathbf{k})|^2 \tag{3.25}$$

$$\gamma_s = \frac{1}{2}\arg[\psi\psi_-] \qquad \text{modulo } \pi \, . \tag{3.26}$$

Note that the last equation is only valid modulo $\pi$, whereas a correction of the symmetric aberrations requires $\gamma$ to be known modulo $2\pi$. Fu and Lichte (1995) resolve this ambiguity using an algorithm similar to phase-unwrapping, based on the fact that $\gamma_s$ is a continuous function. As the method measures $\gamma_s$ separately for each $\mathbf{k}$ and therefore attempts to determine a large number of independent parameters, its accuracy is mainly limited by the signal to noise ratio. Therefore, it is advantageous to exploit the knowledge that $\gamma_s$ can be approximated by

$$\gamma_s = \frac{2\pi}{\lambda}\Re\Big[\frac{1}{2}A_1'\omega^{*2} + \frac{1}{2}C_1'\omega^*\omega + \frac{1}{4}C_3'\omega^{*2}\omega^2\Big] \tag{3.27}$$

where $\omega = \lambda k_x + \mathrm{i}\lambda k_y$. This reduces the number of parameters to be determined to 4. A method that exploits the above symmetry relation to determine the symmetric aberration parameters is presented in section 4.2.

# 3.6 'Diffprop': propagation of crystal reflections in the diffraction plane

All the methods discussed thus far work best for amorphous objects and require at least some irregularities in the object structure. In contrast, an alternative ingenious approach (Malamidis and Lichte, 1998) allows the determination of the wave aberrations from the holographically reconstructed image wave of a crystalline object. The gradient $\nabla\gamma$ is determined at the positions of the crystal reflections by measuring how these reflections shift as the image wave is defocused from the diffraction plane. For an understanding of this method consider the image wave of a crystalline sample with crystalline reflections at frequencies $\mathbf{k}_i$:

$$\psi = e^{i\gamma}\left(\sum_i F_i(\mathbf{k} - \mathbf{k}_i)\right) \tag{3.28}$$

where $F_i(\mathbf{k})$ is the Fourier transform of the planar shape function $f_i(\mathbf{x})$. The shape function describes how the complex amplitude of a diffracted beam varies as a function of position in the object plane.[3] These variations are slow compared with the lattice spacing, hence $F_i(\mathbf{k})$ is strongly localised at the origin and therefore the wave aberration function $\gamma(\mathbf{k})$ can be replaced by its first order Taylor expansion about $\mathbf{k}_i$, yielding

$$\psi_i(\mathbf{k}) = e^{i\gamma(\mathbf{k}_i)}e^{i(\mathbf{k}-\mathbf{k}_i)\cdot\nabla\gamma(\mathbf{k}_i)}F_i(\mathbf{k} - \mathbf{k}_i) \tag{3.29}$$

for the contribution of reflection $i$. The second exponential factor in this equation represents a linear variation of the phase across the diffraction spot. This gives rise to a lateral shift of the diffraction spot when the wave is propagated to planes below and above the diffraction plane. The diffprop method relies on measuring the gradient of $\gamma$ at the positions of diffraction spots from this lateral shift.

It is evident that diffprop requires some inhomogeneity in the object, as a uniform $f_i(\mathbf{x})$ would yield a sharp diffraction spot $\delta(\mathbf{k} - \mathbf{k}_i)$ with no information on the wave

---

[3]$|f_i(\mathbf{x})|^2$ would be observed as the dark field image using reflection $i$. The phase of $f_i(\mathbf{x})$ is constant across a crystallite with no grain boundaries.

aberration gradient. A suitable object would be a small crystallite, as the spatial restriction of $f_i(\mathbf{x})$ leads to a broadening of the diffraction spot and facilitates the measurement of $\nabla\gamma(\mathbf{k})$ across the diffraction spot. For a diffraction pattern recorded directly in the microscope, the diffraction spots are also broadened by the beam divergence. This is, however, not the case for a diffraction pattern calculated from a holographically or otherwise restored image wave. The recorded hologram (or image) is an incoherent superposition of holograms (or images) recorded at different beam directions. In each of these, the zero-order beam is tilted in the same way as the diffracted beams in the object wave. Hence the beam divergence does not broaden the diffraction spots in the restored diffraction patterns but instead reduces their intensity due to the spatial coherence envelope. The necessary spatial restriction of the shape function, however, leads to another problem. If $f_i(\mathbf{x})$ is not symmetric about the origin, the shape function itself introduces a phase gradient in the diffraction spot which is indistinguishable from that due to the wave aberration.[4] This problem can be alleviated when the shape functions of the investigated diffraction spots are equal apart from constant factors,[5] as in this case the shape function leads to equal shifts of all diffraction spots and the aberration function can be evaluated from the shift differences. This makes the method equivalent to evaluating the aberrations by bringing the dark field images of a crystallite calculated from different diffracted beams into registry, which is a viable approach as in the presence of aberrations, the contributions from different diffracted beams are shifted relative to each other.[6]

It may be argued that a suitable shape function can be forced by restricting the field of view with a window function $w(\mathbf{x})$. This leads to the new wave function

$$\psi'(\mathbf{k}) = \psi(\mathbf{k}) \otimes W(\mathbf{k}) \tag{3.30}$$

---

[4]This agrees with the fact that when a diffraction pattern is defocused in the microscope, the diffraction spots widen and begin to show the patterns of dark field images. Hence for a crystalline object covering only part of the selected area the diffraction spots appear to move laterally as the diffraction pattern is defocused.

[5]This is the case for thin crystals, but does not hold for larger thicknesses since, with dynamical scattering, the strength of different beams does not vary in proportion as the thickness increases.

[6]This is similar to the aberration determination method suggested by Budinger and Glaser (1976).

and the phase gradient due to the shape function is compensated when $w(\mathbf{x})$ is centred on the shape function. However, a restricted window function is no substitute for inhomogeneities in the object because while restricting $w(\mathbf{x})$ does indeed broaden the diffraction spots, it also smears out the phase gradient that is to be measured. This can be illustrated in 1 dimension with Gaussian shape and window functions with widths $a$ and $b$, respectively, such that $F_i(k) = e^{-a^2 k^2}$ and $W(k) = e^{-b^2 k^2}$.

This yields

$$
\begin{aligned}
\psi_i'(k) &= e^{i\gamma(k_i)} \int e^{-a^2(k'-k_i)^2} e^{i(k'-k_i)\nabla\gamma(k_i)} e^{-b^2(k-k')^2} dk' \\
&= e^{i\gamma(k_i)} \sqrt{\frac{\pi}{a^2+b^2}} \exp\left[\frac{(k_i a^2 + k b^2 + i\nabla\gamma(k_i)/2)^2}{a^2+b^2} - a^2 k_1^2 - b^2 k^2 - i k_i \nabla\gamma(k_i)\right]
\end{aligned}
\tag{3.31}
$$

which implies that the phase factor due to the wave aberration gradient is now given by

$$
\exp\left[i\frac{b^2}{a^2+b^2}(k-k_i)\nabla\gamma(k_i)\right],
\tag{3.32}
$$

*i.e.* the phase gradient is reduced by a factor $b^2/(a^2+b^2)$, which is small when the window function is narrow compared to the shape function. This makes it problematic to use the method for investigating the isoplanaticity by evaluating local values for the wave aberration coefficients using small subimages of a crystal as reported in (Malamidis and Lichte, 1998).

In summary, the calculations in this section suggest that while diffprop may be useful for determining the aberrations in restored wavefunctions of relatively small crystallites, the requirements on the object are too restrictive to make it a method of choice for routine aberration determination.

# Chapter 4

# A new method for aberration determination

In this chapter, a new method for the accurate determination of the symmetric aberrations ($C_1$, $A_1$ and $C_3$) is presented, based on the acquisition of a through-focus series. The first step, described in section 4.1, involves accurate alignment of the images within the series and refinement of the relative focus levels. Subsequently ,the image wave in the plane of one of the images (reference image) is recovered using a Wiener filter restoration as described in section 1.6.2. The next step (section 4.2) relies on the assumption that the object acts as a weak phase object at most spatial frequencies. A phase contrast index that provides a measure of how well the phase modulation in the object wave is transferred to phase contrast in the image wave is evaluated for each spatial frequency, and determining the aberrations amounts to finding correction parameters $C_1$ and $A_1$ that maximise the phase contrast index integrated over a suitable range of spatial frequencies.

The method works for datasets comprising a minimum of three images at different focus levels, hence it is equally useful for fast aberration determination from a few images for autotuning purposes (chapter 8), and for accurate off-line restoration of the object wave from focus and tilt series datasets (chapter 6).

The antisymmetric aberrations cannot be determined directly using the phase

contrast index, as they do not lead to an interchange of phase and amplitude contrast. The most successful conventional method for determining the antisymmetric aberrations is based on the measurement of the tilt-induced defocus and astigmatism from diffractograms of images taken at different deliberately induced beam tilts (section 3.3.2). In order to avoid the problems associated with diffractogram fitting described in section 3.3.1, the method described in this chapter can be also used to determine the apparent defocus and astigmatism from short focus series taken at different beam tilts. This approach is presented in chapter 5.

In this chapter, the method is illustrated using a focal series of the complex oxide $Nb_{16}W_{18}O_{94}$ and figure 4.1 shows the Scherzer focus image of this series. The results and interpretation of the restoration are presented in section 6.1.

## 4.1   Image registration

Image registration (to a common origin) is required to compensate for any drift that may have occurred between the exposures. This is usually achieved by cross-correlation techniques (Saxton, 1994a). In this work, the word 'registration' is used in a wider context and refers to the determination of not only the shift, but also the change in the aberration function between the images, *i.e.* the focus difference in case of a focal series.

### 4.1.1   The phase correlation function

The Cross Correlation Function (XCF) between two images $C_i(\mathbf{x})$ can be defined in terms of their Fourier Transforms $c_i$ as:

$$\mathrm{XCF}(\mathbf{x}) = \mathrm{FT}^{-1}[c_1^* c_2] \ . \tag{4.1}$$

Two related functions can also be defined. The first is the Mutual Correlation Function (MCF) (van Heel et al., 1992), where the modulus of the cross spectrum is
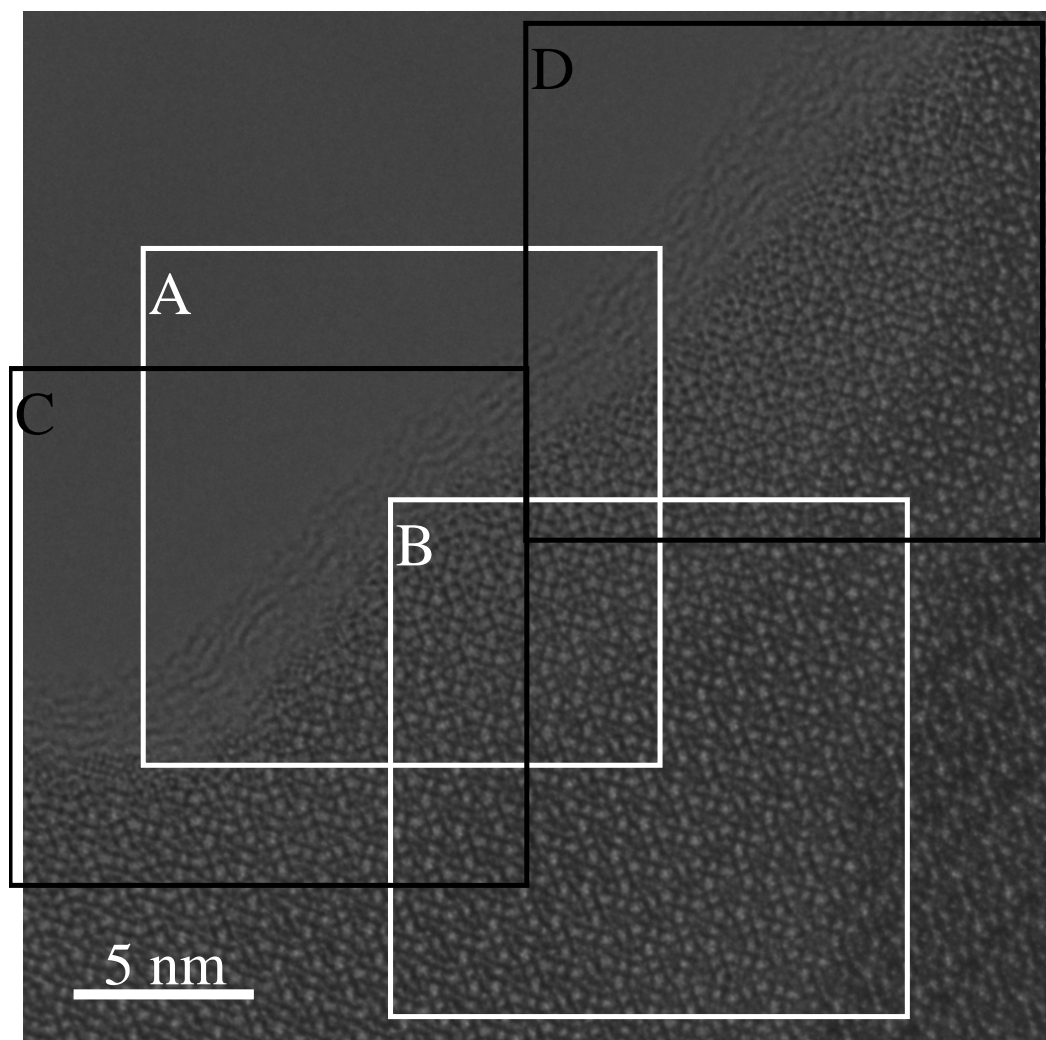
Figure 4.1: Image recorded at the Scherzer defocus taken from a focal series of images of $Nb_{16}W_{18}O_{94}$ in the [001] projection. The two regions marked A and B were used for aberration determination and regions C and D were used in the experimental accuracy assessment.

replaced by its square root and the phase is preserved:

$$\text{MCF}(\mathbf{x}) = \text{FT}^{-1}\left[\frac{c_1^* c_2}{\sqrt{|c_1^* c_2|}}\right]. \tag{4.2}$$

This new method makes extensive use of the Phase Correlation Function (PCF) (Kuglin and Hines, 1975; Horner and Gianino, 1984)),[1]

$$\text{PCF}(\mathbf{x}) = \text{FT}^{-1}\left[F\frac{c_1^* c_2}{|c_1^* c_2|}\right], \tag{4.3}$$

in which the modulus of the transforms is set to unity and therefore the weight of each Fourier component is independent of its magnitude. The real and positive weighting factor $F(\mathbf{k})$ is used to suppress the influence of the very high spatial frequencies which in most cases are only noise.

Figure 4.2 illustrates the differences between XCF, MCF and PCF for two images recorded at different focus levels. As shown in figure 4.2, the PCF between two images recorded for different focus levels under otherwise identical conditions consists of a centrosymmetric ring pattern, the form of which is determined by the focus difference between them. As described later it is possible to compensate for defocus induced phase shifts which give rise to these ring patterns under which circumstances the compensated PCF comprises a single sharp peak at a position given by the shift vector between the images. The main advantage in using the PCF is that the weight given to each Fourier component is independent of its strength. Conversely, the conventional XCF is dominated by contributions where the moduli of the image Fourier transforms are large, *i.e.* the low frequencies and crystalline reflections. The low frequencies are often due to shading patterns unrelated to the object, whereas the dominance of the crystalline reflections leads to a periodic repetition of the cross-correlation peak. This arises because the Fourier transform phases at the positions of crystalline reflections are invariant under image translation by a lattice vector. These crystalline reflections

---

[1]The PCF was originally introduced for optical correlation techniques, where it has the additional benefit that no intensity is lost in the correlator when only the phase is manipulated. Unaware of this work, I initially found the usefulness of the PCF by serendipity when limiting the modulus in image transforms in order to be able to use fast integer Fourier transforms for cross-correlation.
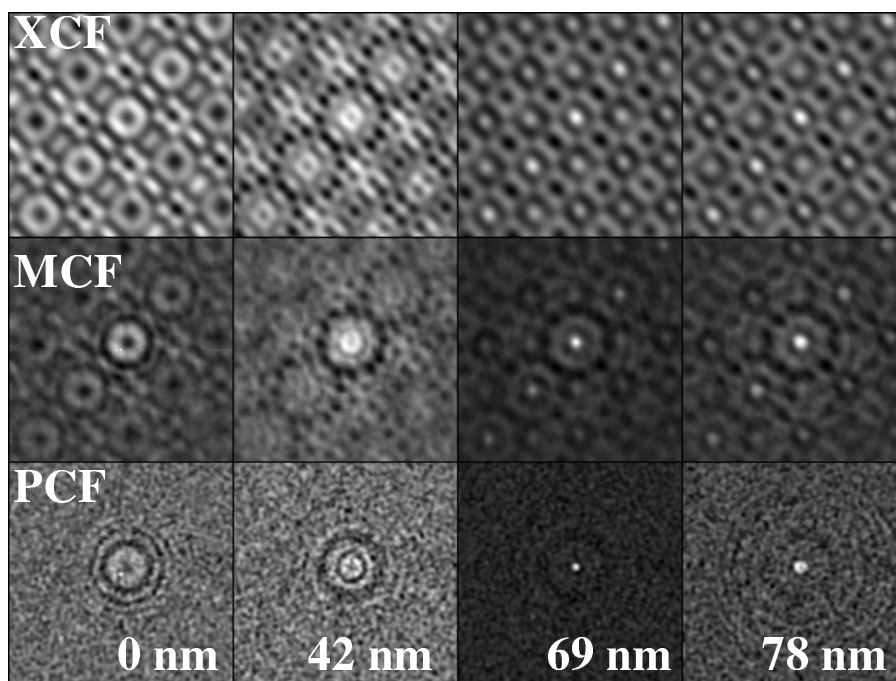
Figure 4.2: Comparison of Cross- Mutual- and Phase Correlation Functions (XCF, MCF and PCF) between two images from the focal series, using the region marked as B in figure 4.1 for the compensated focus differences indicated. Due to the periodicity of the crystalline specimen, the XCF peak repeats periodically. The PCF does not show this repetitive pattern, but without phase compensation (left column) the peak is broadened into concentric rings due to the focus difference. When the phase compensation for the actual focus difference of 69 nm is applied, the PCF collapses into a single sharp peak.

are strong, and therefore dominate the XCF, but they only cover a small proportion of Fourier space and therefore have little influence on the PCF. This feature of the PCF allows the approach described here to be used as readily with crystalline as with non-crystalline specimens representing a significant advantage over existing methods.

## 4.1.2 Phase compensation of the PCF

As shown in figure 4.2, the PCF of two images taken at different focus levels shows a concentric ring pattern rather than a single peak due to the difference in imaging conditions. For images that differ only in defocus, Saxton (1994a) gives two alternative phase compensation methods which are both based on the weak phase object approximation. The first consists in a multiplication of the cross spectrum with the factor $\sin \gamma_1 \sin \gamma_2$ and hence requires the knowledge of the wave aberrations $\gamma_1$ and $\gamma_2$ in both images. In the second approach, the phases in the cross spectrum are simply doubled, eliminating the sign changes in the cross spectrum due to contrast reversals in the phase contrast transfer functions. This yields a single sharp peak at a position $2\mathbf{d}$, where $\mathbf{d}$ is the displacement. However, phase doubling was found to give poor results for the PCF when little amorphous material was present, mainly due to the poor signal to noise ratio at frequencies not corresponding to crystal reflections. The doubling of the phase error also leads to a disproportionately poor performance of the phase-doubled PCF for low signal-to noise ratios, as demonstrated in figure 4.3. In this section a different phase compensation is introduced which

1. Does not rely on the weak phase object approximation.

2. Does not require any prior knowledge of the imaging conditions.

3. Additionally allows the determination of the focus difference.

In the linear imaging approximation the image contrast Fourier transform is given by $c = \psi + \psi_-^*$ in terms of the image wave Fourier transform $\psi$. Hence the cross spectrum between the image contrast Fourier transforms $c_i$ in the presence of a defocus
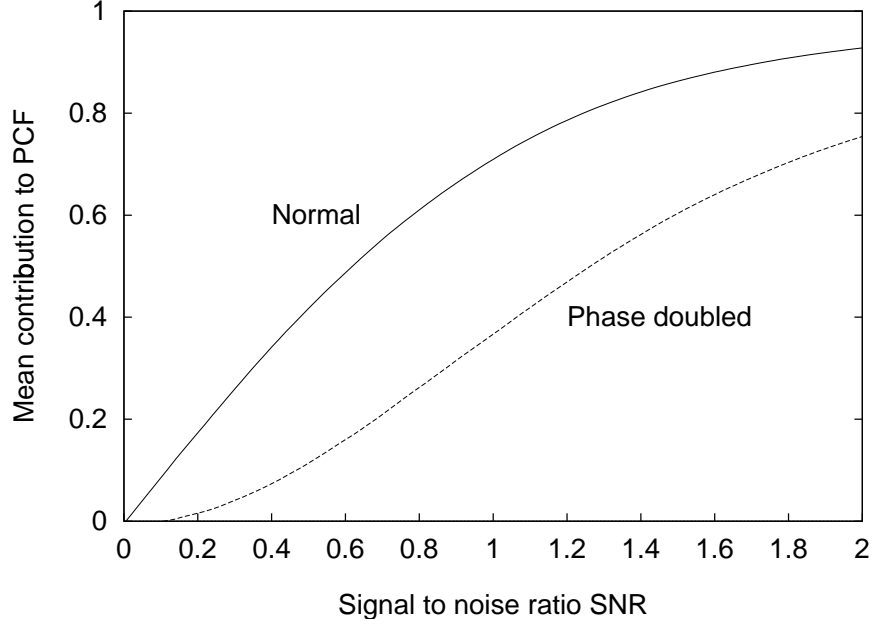
Figure 4.3: The mean contribution of individual frequencies in the phase cross spectrum to the PCF peak with and without phase doubling as a function of the signal to noise ratio. The performance when the SNR is low is strongly reduced by phase doubling. The functions plotted are $\langle\cos(\arg(x+Z))\rangle$ (solid) and $\langle\cos(2\arg(x+Z))\rangle$ (dashed), where $x$ is the SNR and $Z$ is an ensemble of Gaussian complex random numbers with $\langle Z\rangle = 0$ and $\langle|Z|^2\rangle = 1$.

difference $D$, a displacement vector $\mathbf{d}$ and otherwise identical imaging conditions is given by:

$$c_1^* c_2 = \left(\psi + \psi_-^*\right)^* \left(\mathrm{e}^{-2\pi\mathrm{i}\mathbf{k}\mathbf{d}}e^{-\mathrm{i}\gamma_D}\psi + \mathrm{e}^{-2\pi\mathrm{i}\mathbf{k}\mathbf{d}}\mathrm{e}^{\mathrm{i}\gamma_D}\psi_-^*\right) = \mathrm{e}^{-2\pi\mathrm{i}\mathbf{k}\mathbf{d}}\times$$
$$\left\{\cos\gamma_D\left[|\psi|^2 + |\psi_-|^2\right] - \mathrm{i}\sin\gamma_D\left[|\psi|^2 - |\psi_-|^2\right] + 2\Re\left[\mathrm{e}^{\mathrm{i}\gamma_D}\psi\psi_-\right]\right\} , \quad (4.4)$$

where the wave aberration

$$\gamma_D = \pi D\lambda|\mathbf{k}|^2 \qquad (4.5)$$

describes the propagation from the first to the second image and $\psi_\pm = \psi(\pm\mathbf{k})$ is the image wave Fourier transform *in the plane of the first image*.

The first term in curly brackets in equation 4.4 is of particular interest as its sign is independent of the wavefunction. When the sign changes of $\cos \gamma_D$ are compensated, the first term therefore gives positive contributions at all spatial frequencies.

This is exploited in the phase-compensated PCF, defined as

$$\mathrm{PCF}_\gamma(\mathbf{x}) = \mathrm{FT}^{-1}\left[F(k)\frac{\cos\gamma \; c_1^* c_2}{|\cos\gamma \; c_1^* c_2 + h|}\right] , \tag{4.6}$$

where the very small positive number $h$ prevents a zero denominator. When the focus difference used in the phase compensation factor $\cos\gamma$ matches the actual focus difference, the first term in the curly brackets of equation 4.4 yields positive contributions to the $\mathrm{PCF}_\gamma$ at a position $\mathbf{x} = \mathbf{d}$ for all $\mathbf{k}$ *for any image wave,* giving rise to a strongly localised correlation peak at this position.

For arbitrary objects, $\psi$ and $\psi_-$ are unrelated and therefore the contributions from both the second and the third term in equation 4.4 have oscillating signs and hence largely cancel when the inverse Fourier transform in equation 4.6 is evaluated. However, this may not be the case for a more restricted class of objects. For weak phase objects, the image wave fulfils the condition $\psi \mathrm{e}^{-\mathrm{i}\gamma_s} = -\psi_-^* \mathrm{e}^{\mathrm{i}\gamma_s}$, where $\gamma_s$ is the symmetric wave aberration for image 1. Hence for weak phase objects, the second term in equation 4.4 vanishes for all $\mathbf{k}$. However, at a given $\mathbf{k}$, the third term cancels the positive first term under the condition

$$0 \overset{!}{=} \cos\gamma_D - \cos(\gamma_D + 2\gamma_s) = 2\sin\gamma_s \sin(\gamma_s + \gamma_D) , \tag{4.7}$$

where the equivalence on the right hand side can be inferred from the formula $2\sin x \sin y = \cos(x-y) - \cos(x+y)$. Hence for weak phase objects the PCF peak will only disappear if, at most spatial frequencies, the phase contrast vanishes in either the first image ($\sin\gamma_s = 0$) or the second image ($\sin(\gamma_s + \gamma_D) = 0$).

As an illustration of this phase compensation, figure 4.2 shows how the concentric ring pattern of the PCF collapses into a sharp peak when phase compensation for the correct defocus difference is applied. The plot shown in figure 4.4 demonstrates that the relative defocus difference between the images can be accurately determined by maximising the PCF peak height as a function of the compensated defocus difference.
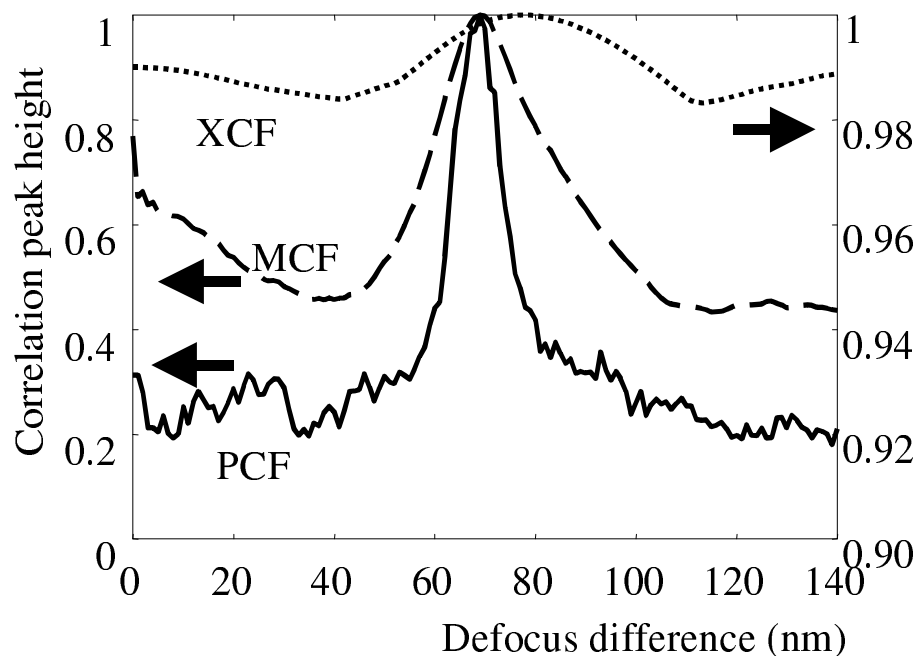
Figure 4.4: The peak height of the XCF, MCF and PCF between two images for region B in figure 4.1 as a function of the compensated focus difference. The PCF (solid) shows a sharp maximum at the correct focus difference of 69 nm. When the MCF is used (dashed), the maximum is less pronounced and the XCF (dotted) shows little variation in the peak height with the maximum peak height occurring at a different focus difference. This arises because the XCF is dominated by low spatial frequencies where the defocus difference has little impact.

The position of this maximised peak in the PCF therefore measures the displacement between the two images with high accuracy.

It is important to note that the phase compensation necessary to produce the highest PCF peak depends only on the modulus of the focus difference (or in general on the difference in the wave aberration function) and is not sensitive to the absolute aberrations in each image.

### 4.1.3 PCF with predicted images

Once at least two images within the focal series have been registered with the reference image as described above, an initial restoration of the image wave in the plane of this reference image can be calculated as described in section 1.6.2. From this restored image wave $\psi_i$, the image contrast $c_D$ predicted for an image taken at any focus difference $D$ from the reference image can be evaluated as

$$c_D = \psi_i e^{i\gamma_D} + \psi_{i-}^* e^{-i\gamma_D} \tag{4.8}$$

The registration of a further image can now be achieved by calculating the PCF with images predicted from the restored image wave. The relative focus level of a particular experimental image is then determined by maximising the PCF peak height with respect to the defocus of an image predicted from the initial restoration without a need to apply further phase compensation as illustrated in figure 4.5. Hence, an initial restoration from a few images can be extended iteratively, where the displacement and relative focus level of each new image is determined by calculating the PCF with images predicted from the current restoration.

This successive addition of images and the use of images predicted from the current restoration is more robust than simply using the phase compensated PCF with a reference image for several reasons:

1. When the restoration already contains $N$ images, the signal to noise ratio in a predicted image is improved by a factor of $\sqrt{N}$ compared to a single experimental image.

2. The phase-compensated PCF cannot utilise information from spatial frequencies near the zero crossings of $\cos \gamma_D$ (equation 4.4) because at these frequencies the two images carry independent and complementary information as their wave aberrations differ by $(2n+1)\pi/2$. This problem does not arise when the experimental image is correlated with one predicted for the correct focus value.

3. The iterative method proceeds towards either end of the focal series, therefore,
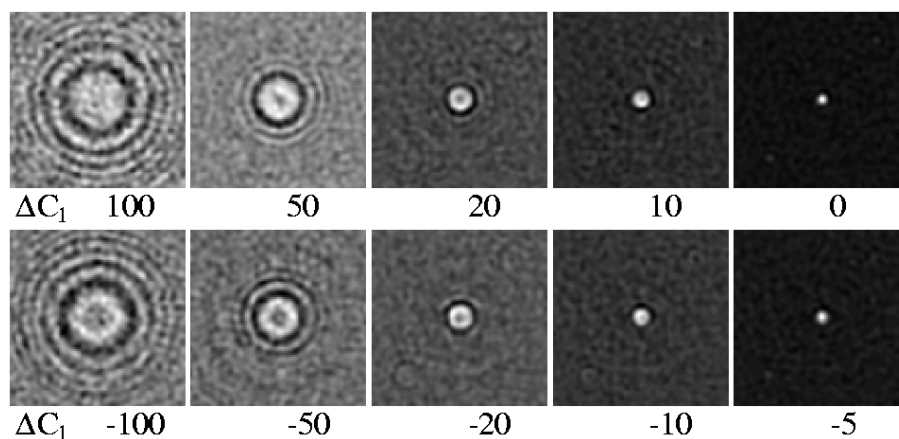
Figure 4.5: PCF between one experimental image and images predicted for various focus levels from a restoration that includes all other images. The values $\Delta C_1$ indicate the focus difference between predicted and experimental image. The PCF shows concentric rings for a mismatched defocus and collapses into a sharp peak at the correct value of 0. However, the peak quality is significantly improved compared to figure 4.4 due to the better signal-to-noise ratio in the predicted image compared to that in a single experimental image.

for each image to be added, the restoration already contains the images acquired immediately before or afterwards, yielding useable correlations even if the specimen changes significantly during the total series acquisition.

Correlation with predicted images can also be used to refine previously determined focus levels. In this case, the image to be refined must be temporarily removed from the restoration as described in section 1.6.2. If this is not done, there is a bias toward the original registration, since in this case even the *noise* in this image will contribute to the correlation. Figure 4.6 shows the variation in the PCF peak height as a function of predicted image focus levels, clearly demonstrating the improved peak quality compared to that of the phase-compensated PCF.

## 4.2 Defocus and astigmatism determination

Once the complex image wave in the plane of the reference image is restored, the object wave can be calculated by correcting for the wave aberration with an appropriate phase plate in the diffraction plane. However, this requires accurate knowledge of the aberration parameters. In this section a new method to determine the defocus $C_1$ and astigmatism $A_1$ in a restored image wave is described.

The determination of the aberrations from a restored image wave alone inevitably requires some prior knowledge about the object. This is also the case for most of the methods described in the previous chapter, notably the minimum variance method (section 3.1) and diffractogram based methods (section 3.3), which rely on the weak-phase approximation. This states that the object wave $\psi_o$ satisfies

$$\psi_{o-} = -\psi_o^*. \tag{4.9}$$

The new method only requires that equation 4.9 is valid for the majority of spatial frequencies $\mathbf{k}$. As in the previous sections, contributions from crystalline areas are suppressed by giving equal weights to all spatial frequencies. Therefore, even if the specimen is crystalline with a thin amorphous contamination layer, the weak phase
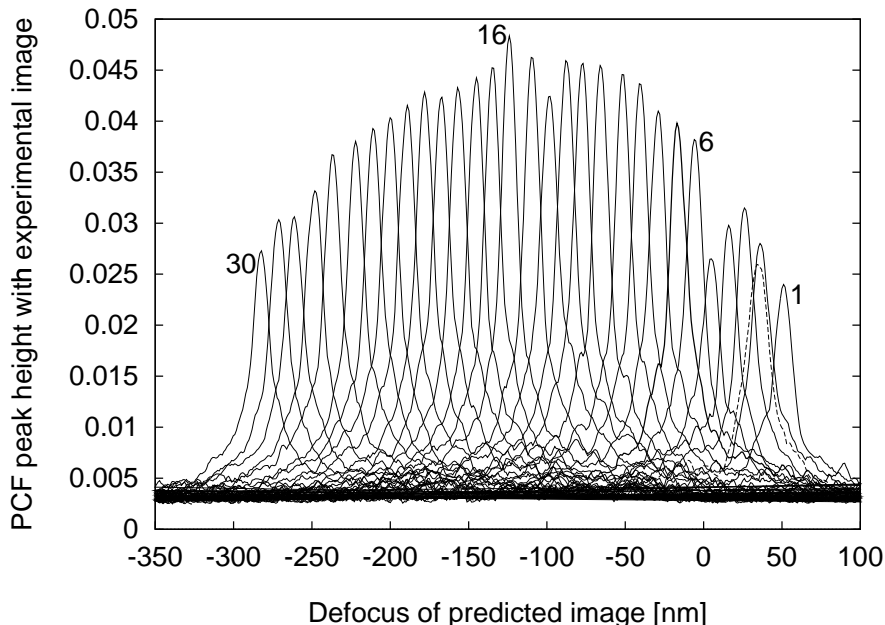
Figure 4.6: The variation in the PCF peak height of successive experimental images with an image predicted from the restoration using all other images as a function of the predicted image focus. Values on the x-axis are absolute focus values as determined later. Note that the peaks are not evenly spaced, indicating small but detectable deviations from the nominal focus levels. The dotted graph with its peak under that for image 2 belongs to the final image taken with the nominal focus reset to the initial value used for the first image providing an estimate of the total focal drift across the entire series. The strongest peaks are observed for images in the middle of the series indicating that the amorphous layer changes during acquisition.

approximation will still hold for all $\mathbf{k}$ that do not correspond to crystal reflection, *i.e.* the vast majority, provided that multiple scattering involving both amorphous and crystalline specimen layers can be ignored.

## 4.2.1 The phase contrast index

As a measure of the quality of aberration correction with a trial phase plate given by:

$$\exp\left(-i\gamma_c(C_1, A_1)\right) \tag{4.10}$$

(with the spherical aberration $C_3$ fixed), a Phase Contrast Index function

$$f_{\text{PCI}}(C_1, A_1) = -\cos\left[\arg(\psi_i) + \arg(\psi_{i-}) + 2\gamma_c(C_1, A_1)\right] \tag{4.11}$$

is introduced. This function is 1 (-1) where the phase modulation in the object wave is transferred into phase (amplitude) contrast in the corrected image wave, given by

$$\psi_c = \psi_i \exp\left(-i\gamma_c(C_1, A_1)\right). \tag{4.12}$$

The function is hence a measure of the conjugate symmetry of $\psi_c$, being $+1$ where $\psi_c$ is conjugate antisymmetric and -1 where it is conjugate symmetric. When the trial parameters $C_1, A_1$ match the actual aberration parameters, the $f_{\text{PCI}}$ will tend to be 1 (or at least close to 1 when noise is present) for all spatial frequencies, as shown in figure 4.7(d). For mismatched parameters, the $f_{\text{PCI}}$ shows dark rings or bands where the difference between actual and corrected wave aberration is $\pi/2$. Hence, the function

$$F_{\text{PCI}}(C_1, A_1) = \left\langle f_{\text{PCI}}(C_1, A_1) \right\rangle_{\mathbf{k}}, \tag{4.13}$$

is a good measure of the quality of the trial parameters $C_1, A_1$. Here the notation $\left\langle \cdots \right\rangle_{\mathbf{k}}$ has been introduced for averaging over $\mathbf{k}$ with the function $Q(\mathbf{k})$ defined in equation 1.57 as weighting function:

$$\left\langle f \right\rangle_{\mathbf{k}} = \frac{\int d\mathbf{k} f(\mathbf{k}) Q(\mathbf{k})}{\int d\mathbf{k} Q(\mathbf{k})}. \tag{4.14}$$

## 4.2.2 Defocus determination

When only the defocus ($C_1$) is mismatched, the $f_{\text{PCI}}$ shows a pattern of concentric circular rings (figure 4.7). The dark rings vanish for the correct defocus value, making $F_{\text{PCI}}$ maximal. Figure 4.7(e) shows that the maximum is unique and very sharp, hence the correct defocus value can be found automatically with high precision and confidence. It can be estimated that a misfit in the wave aberration by $\pi/4$ at the highest spatial frequency used is still easily detectable, as it reduces the $f_{\text{PCI}}$ from positive values to zero. For the experimental dataset discussed here, this gives a
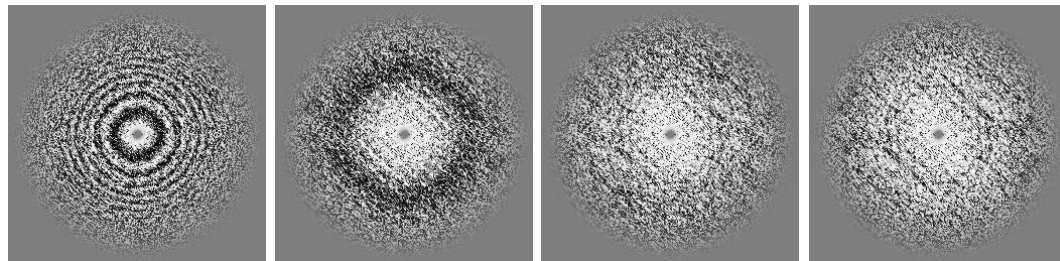
(a) $\Delta C_1 = -50$nm  (b) $\Delta C_1 = -10$nm  (c) $\Delta C_1 = -3$nm  (d) $\Delta C_1 = 0$
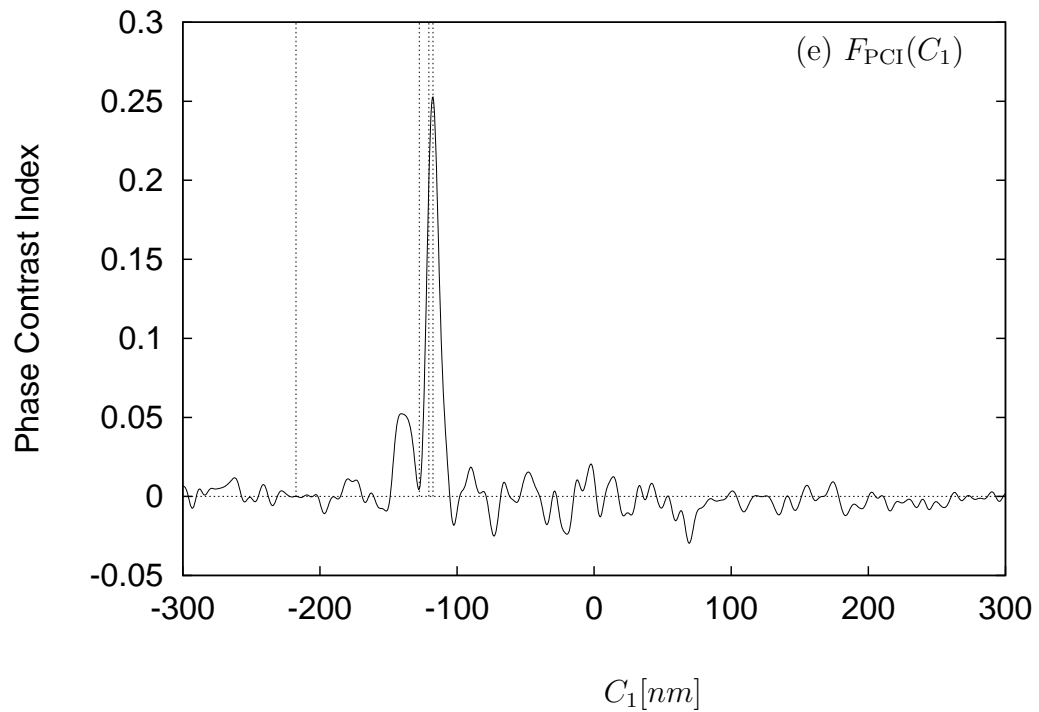


Figure 4.7: The phase contrast index function $f_{\mathrm{PCI}}$. For mismatched values of $C_1$ $f_{\mathrm{PCI}}$ shows dark rings ((a) to (c)), whereas at the correct value of $C_1$, $f_{\mathrm{PCI}}$ is close to one (white) at all spatial frequencies (d). (e) $f_{\mathrm{PCI}}$ plotted as a function of $C_1$. From the position of the sharp maximum, the focus value of the reference image is determined as -128 nm. The values $\Delta C_1$ given in (a)-(d) are relative to this focus.

conservative estimate of 2 nm for the accuracy in the defocus determination. A more comprehensive account is given in section 4.3, showing that the actual accuracy achieved is substantially better.

### 4.2.3 Astigmatism determination

In general, the twofold astigmatism is also unknown and the maximum of $F_{\mathrm{PCI}}$ w.r.t. three parameters has to be found. This subsection describes a method for the fast and efficient location of this maximum.

Figure 4.8 shows the $f_{\mathrm{PCI}}$ for various values of $C_1$ and with $A_1 = (10 + 6\,\mathrm{i})\mathrm{nm}$, deliberately mismatched from its true value close to zero. A large defocus mismatch now yields an elliptical ring pattern, as shown in figure 4.8(a). For a mismatch of $C_1 = \pm|A_1|$ the $f_{\mathrm{PCI}}$ is close to 1 in extended regions close to one of the principal astigmatism axes.

The solid graph in figure 4.8(e) shows maxima at these values of $C_1$, but these maxima are far less pronounced than that in figure 4.7. When $\phi_0$ is the direction of one of the principal astigmatism axes and $\phi$ is the polar angle of $\mathbf{k}$,

$$\left\langle f_{\mathrm{PCI}} \cdot (\cos 2(\phi - \phi_0) + 1) \right\rangle_{\mathbf{k}} \tag{4.15}$$

has a more pronounced maximum as a function of $C_1$. However, the astigmatism axes are unknown, therefore the maximum of expression 4.15 w.r.t. $C_1$ and $\phi_0$ needs to be found.

For a fixed value of $C_1$, the maximum w.r.t. $\phi_0$ can be found analytically. With the abbreviations

$$F_{\mathrm{PCIC}} = \left\langle f_{\mathrm{PCI}} \cos 2\phi \right\rangle_{\mathbf{k}} \tag{4.16}$$

$$F_{\mathrm{PCIS}} = \left\langle f_{\mathrm{PCI}} \sin 2\phi \right\rangle_{\mathbf{k}} \tag{4.17}$$

the maximum condition can be expressed as

$$\begin{aligned} \max \; &\overset{!}{=}\; \left\langle f_{\mathrm{PCI}} \cdot (1 + \cos 2(\phi - \phi_0)) \right\rangle_{\mathbf{k}} \\ &=\; F_{\mathrm{PCI}} + F_{\mathrm{PCIC}} \cos 2\phi_0 + F_{\mathrm{PCIS}} \sin 2\phi_0 \;. \end{aligned} \tag{4.18}$$

(a) $\Delta C_1 = -30$nm (b) $\Delta C_1 = -10.5$nm (c) $\Delta C_1 = 0$ (d) $\Delta C_1 = 9$nm
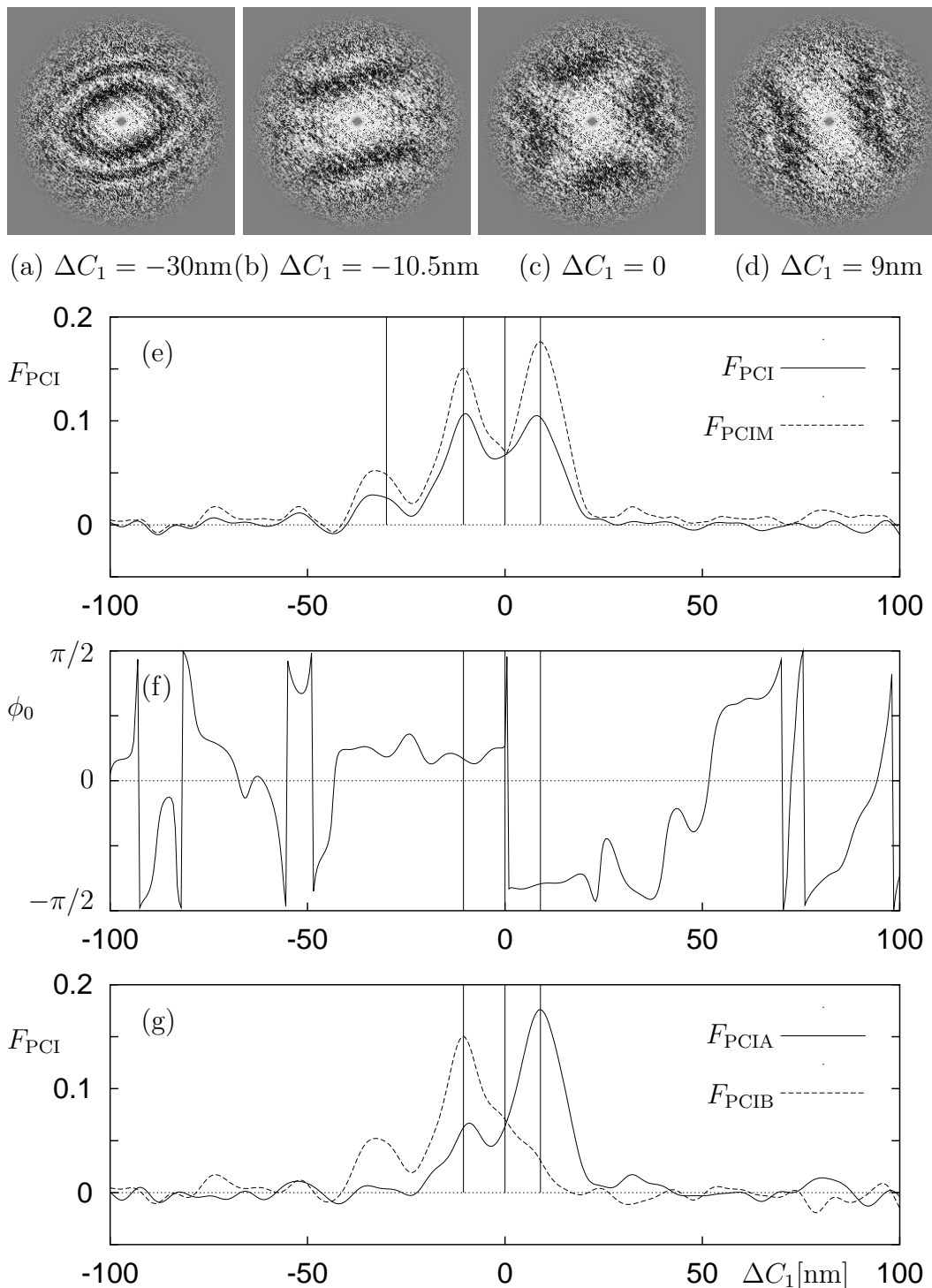


Figure 4.8: The $f_{\mathrm{PCI}}$ with deliberately mismatched astigmatism $A_1 = 10 + 6i$. (a-d) The $f_{\mathrm{PCI}}$ at 4 different values for the defocus mismatch $\Delta C_1$. (e) The maximum directionally weighted phase contrast index $F_{\mathrm{PCIM}}$ exhibits stronger maxima than the $F_{\mathrm{PCI}}$. (f) The angle $\phi_0$ that maximises this directionally weighted PCI is relatively constant in vicinity of each maximum, and abruptly changes by $\pi/2$ between them. (g) The directionally weighted PCI along either astigmatism axes has a single pronounced maximum.

This yields

$$2\phi_0 = \arg(F_{\mathrm{PCIC}} + iF_{\mathrm{PCIS}}) \tag{4.19}$$

with the maximum value

$$F_{\mathrm{PCIM}} = F_{\mathrm{PCI}} + \sqrt{F_{\mathrm{PCIC}}^2 + F_{\mathrm{PCIS}}^2} \ . \tag{4.20}$$

Using equations 4.15–4.20, the following procedure has been developed to determine both $C_1$ and $A_1$:

1. For a range of values for $C_1$, $F_{\mathrm{PCI}}(C_1)$, $F_{\mathrm{PCIC}}(C_1)$ and $F_{\mathrm{PCIS}}(C_1)$ are determined.

2. $F_{\mathrm{PCIM}}$ is calculated according to equation 4.20.

3. The value $C_{1a}$ for $C_1$ that maximises $F_{\mathrm{PCIM}}$ is located

4. $\phi_0(C_{1a})$ is calculated according to equation 4.19.

5. With this $\phi_0$, the value $C_{1b}$ that maximises $F_{\mathrm{PCI}} - F_{\mathrm{PCIC}} \cos 2\phi_0 - F_{\mathrm{PCIS}} \sin 2\phi_0$ is found. This is the focus value in the direction $\phi_0 + \pi/2$.

6. At this stage of the process, the determined parameters are $C_1 = (C_{1a} + C_{1b})/2$ and $A_1 = A'_1 + 1.5 \exp(2i\phi_0) \cdot (C_{1a} - C_{1b})/2$, where $A'_1$ is the previous value of $A_1$.

7. Steps 1–6 are repeated with the new value of $A_1$ and a small defocus range around $C_1$ until $A_1$ and $C_1$ are stable.

The factor 1.5 that appears in step 6 merits further explanation. If $C_{1a}$ and $C_{1b}$ were the focus values along the two principal astigmatism directions, a factor 1 would be needed. However, $C_{1a}$ and $C_{1b}$ are measured from large sectors around the principal direction, hence the astigmatic focus difference is underestimated. A factor of 1.5 was found to give the fastest convergence, and with this value, the procedure only has to be repeated 2 or 3 times to yield stable parameters.

The reliability and speed of the method is increased by using a weighting function $Q$ with a cut-off at a lower spatial frequency for the first iteration of the astigmatism

determination. This is particularly beneficial when a large value of astigmatism is present, as in this case, when the focus is matched along one of the principal axes, the $f_{\mathrm{PCI}}$ will have high values along a strip of width $1/|A_1|$ directed along this axis. As this strip represents an increasingly narrow sector for increasing spatial frequency, the method becomes prone to spurious maxima due to noise when large spatial frequencies are used while a large astigmatism is still present.

### 4.2.4 Optimisation of the algorithm

Even with the directed search described in the previous section to find the astigmatism, integrating the $f_{\mathrm{PCI}}$ for each of a large number $N$ of trial values for $C_1$ is a lengthy process and can be the rate-limiting step for a fast initial aberration determination where the *relative* focus levels are assumed to be correct. A reduction of the problem to a single 2-D integration and a 1-D Fourier transform can be achieved by defining the 1-D function

$$d_{\mathrm{PCI}}(n) = \sum_{\{\mathbf{k}\,:\,n-1/2\,\leq\,k^2/h\,<\,n+1/2\}} Q\frac{\psi + \psi_-}{|\psi + \psi_-|} \; . \qquad (4.21)$$

For $|C_1| \ll \lambda/h$, the $F_{\mathrm{PCI}}$ is given to an excellent approximation by

$$F_{\mathrm{PCI}}(C_1) = -\frac{1}{2}\Re\left[\sum_n \mathrm{e}^{2\pi i n C_1 h/\lambda} d_{\mathrm{PCI}}(n)\right] \; , \qquad (4.22)$$

where the definition of the $d_{\mathrm{PCI}}$ is extended to negative $n$ by $d_{\mathrm{PCI}}(n) = d_{\mathrm{PCI}}(-n)$.

Hence, the real part of the inverse discrete Fourier transform of $-d_{\mathrm{PCI}}(n)/2$ over the interval $-N/2 \leq n < N/2$ yields the $F_{\mathrm{PCI}}(C_1)$ for $-\lambda/(2h) \leq C_1 < \lambda/(2h)$ with a sampling interval of $\Delta C_1 = \lambda/(Nh)$. The functions $F_{\mathrm{PCIC}}$ and $F_{\mathrm{PCIS}}$ can be evaluated in the same way when the argument of the sum in 4.21 is multiplied by $(k_x^2 - k_y^2)/k^2$ or $2k_x k_y/k^2$, respectively.

### 4.2.5 Spherical aberration determination

The spherical aberration $C_3$ is generally a constant of the microscope, and hence does not need to be determined within a routine parameter determination method.

However, it is still of interest to examine how accurately $C_3$ can be estimated using the phase contrast index. Figure 4.9(a) shows the $F_{\mathrm{PCI}}$ as a function of $C_1$ and $C_3$. As for all methods that determine $C_3$ from axial images alone (Krivanek, 1976; Saxton, 1995; Fu and Lichte, 1995), the method suffers from a strong interdependence of the estimated parameters $C_3$ and $C_1$, which causes the maximum in figure 4.9(b) to be elongated in diagonal direction. Nevertheless, $C_3$ can be determined as $(0.55 \pm 0.05)$mm, which is close to the value of 0.57mm measured previously by tilt-induced displacement and tilt-induced diffractogram measurements (Hutchison et al., 1999). This is remarkably accurate considering that the specimen had only a small amount of amorphous material. This good agreement with the value determined by other methods also shows that the weak phase approximation is valid, which is discussed in more detail in section 4.3.5.

## 4.2.6   Systematic errors in the PCI due to noise

Under certain circumstances, the noise present in the images may also contribute to the conjugate antisymmetry measured by the PCI and thereby lead to artefacts. The restoration filters

$$r'_i = r_i \mathrm{e}^{\mathrm{i}\gamma_s(C_1, A_1)} \tag{4.23}$$

restore the corrected image wave $\psi_c$ for a particular set of trial parameters $C_1, A_1$. The corrected image wave can then be written as

$$\psi_c = \sum r'_i c_i \tag{4.24}$$

The measured intensities $C_i(\mathbf{r})$ are real, hence the Fourier transforms $c_i$ are conjugate symmetric and the noise contributions $n_i$ also have this symmetry. Therefore, the contribution of noise to the expectation value $\langle -\psi_c \psi_{c-} \rangle$, which determines the PCI, can be written as

$$\langle -\psi_c \psi_{c-} \rangle_n = -\sum r'^2_i \langle n_i n_i^* \rangle = -P_N \sum r'^2_i \ , \tag{4.25}$$

where $P_N$ is the mean noise power in the images as a function of spatial frequency as defined in section 1.6.2. This noise term can lead to systematic errors in the

(a)



(b)
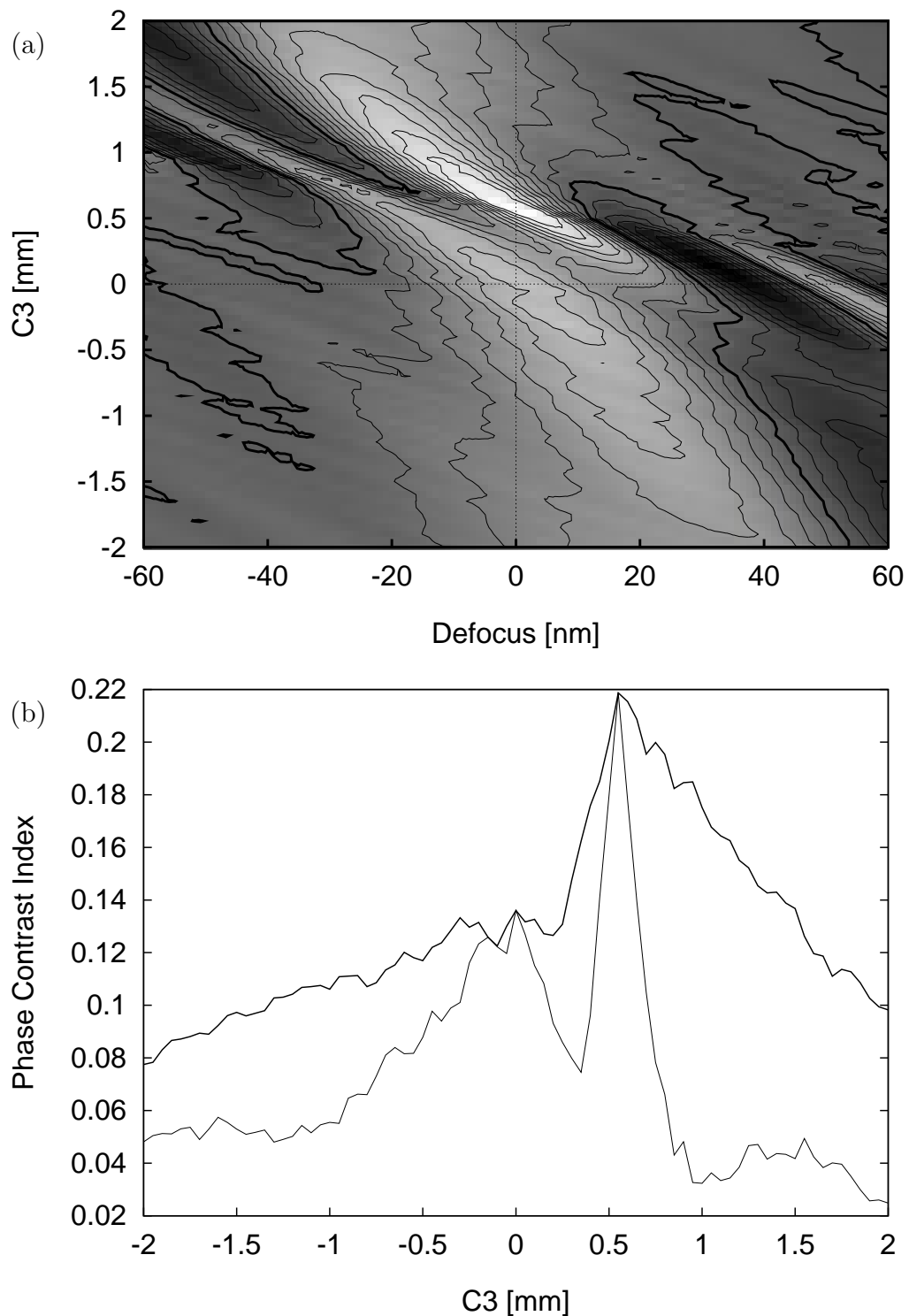


Figure 4.9: (a) Contour plot of $F_{\mathrm{PCI}}(C_1, C_3)$ showing an elongated maximum at $(C_1, C_3) = (2\mathrm{nm}, 0.55\mathrm{mm})$. The thick contour line indicates $F_{\mathrm{PCI}} = 0$. (b) The $F_{\mathrm{PCI}}$ for a fixed value of $C_1$, $F_{\mathrm{PCI}}(2\mathrm{nm}, C_3)$, shows a very sharp maximum at a particular value of $C_3$ (thin line). However, $C_1$ cannot be determined independently, therefore the accuracy of the measured $C_3$ is determined by the peak sharpness of the function plotted with a thick line, which shows the maximum of $F_{\mathrm{PCI}}(C_1, C_3)$ w.r.t. $C_1$ for each value of $C_3$.

determined parameters if it is comparable in strength to the signal component. This is only the case when both of the following conditions are fulfilled:

1. The image signal spectrum is weak compared to the noise spectrum. Note that this would also restrict visibility of diffractogram rings.

2. The number of images used in the restoration is small, since for longer series $|\sum r_i'^2| \ll 1$, at most frequencies, as contributions from different images cancel.

In practice, systematic errors due to this noise term can be avoided even for short focus series by choosing unequal and sufficiently large focus steps.

## 4.3   Experimental accuracy

### 4.3.1   Relative focus level determination

The focus difference between any pair of images measured independently from different subregions of the images should in general be equal and therefore the accuracy of the relative focus levels can be determined by comparing the results obtained from independent non-overlapping subregions of the image. Figure 4.10 shows the differences in the focus levels determined from the two subregions marked as C and D in figure 4.1. The RMS difference between the measured relative focus values is 1.8 nm. However, it is apparent from figure 4.10 that there is a systematic increase in this focus difference by about 4 nm across the whole series. This implies that the total focus change between the first and last image is 4 nm (1%) larger for region A. It might be argued that this discrepancy is due to an accumulation of systematic errors in the iterative relative focus determination. However, the discrepancy is confirmed to be real by examination of the correlation of the last image with one predicted from a restoration using only the first five images, which yields the same value. Therefore, this 1% discrepancy in the total focus difference is attributed to a 0.5% difference between the average sampling interval of the two regions caused by variations in the
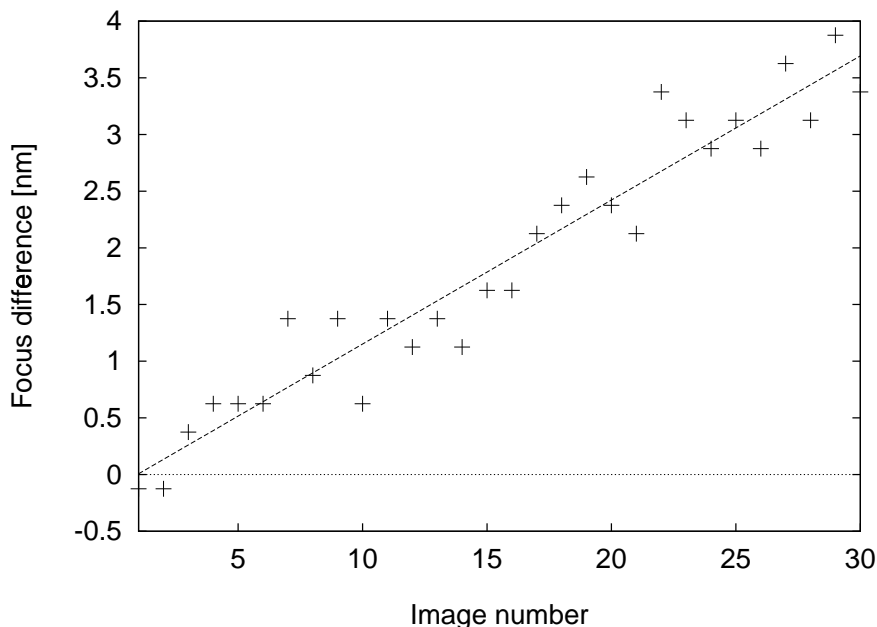
Figure 4.10: The difference between focus levels determined from the independent regions C and D (figure 4.1) shows a small systematic increase across the series. The remaining RMS difference from the fitted straight line is only 0.3 nm, demonstrating the high accuracy of the relative focus determination.

local magnification due to image distortion by the projector and intermediate lenses. This explanation is supported by the measured peak positions of equivalent crystal reflections in the diffractograms from both regions.

When this linear increase in focus difference is corrected, the RMS difference between the relative focus values measured from both regions is reduced to only 0.3 nm. However, this is atypically low and more typical values from other data sets give RMS differences of 0.8 nm leading to a typical accuracy in the relative focus values determined from each region of approximately $0.8\,\mathrm{nm}/\sqrt{2} = 0.6\,\mathrm{nm}$.

### 4.3.2 Absolute focus levels

The accuracy of the absolute focus level and astigmatism determination cannot be assessed using the above method, as the absolute defocus may genuinely be different for the two subregions due to tilt or distortion of the sample across the image field.

Instead, the accuracy has been investigated using restorations of small subseries of images. For each image $n$, the focus and astigmatism were calculated using a restoration from the three images ($n - 1$, $n$ and $n + 2$), which gives an unequally spaced set of focus levels in order to avoid transfer gaps in the restoration. The RMS differences between the $C_1$ values established from the subseries and from the full series was 0.5 nm for region C and 0.6 nm for region D and the difference between the total focus change in these two regions was also confirmed (table 4.1). This table also shows that although the RMS deviations in the measured values of the complex astigmatism from a constant value is already small (1.5 nm), the data are more consistent (RMS 0.8 nm) if a linear change in the astigmatism by 4 nm across the series is incorporated. This change in astigmatism is equivalent to a 1% difference in the total change in focus along two principal axes and can be attributed to a distortion that leads to a 0.5% difference in the magnification along these axes.

It can be concluded that both $C_1$ and $A_1$ can be measured to an accuracy of better than 1 nm from a focus series of only three images. It is worth noting that this accuracy is independent of the starting condition and was also achieved for subsets close to Scherzer and Gaussian focus where diffractogram based methods tend to fail. A similar accuracy was also achieved from subsets of images that were all more than 300 nm underfocus, in this case representing a relative accuracy of 0.3%.

### 4.3.3 Comparisons with diffractogram fitting

The focus levels of the individual images determined using the new method described here can be directly compared with results from automated or manual diffractogram fitting procedures (Chand et al., 1995; Krivanek and Fan, 1992). In this work the

| Parameter | Std. dev. $\sigma$ | fitted change | $\sigma$ from fitted line |
|---|---|---|---|
| $C_1^{FC}$ | N/A | -334.9 | 1.3 |
| $C_1^{FD}$ | N/A | -331.2 | 1.4 |
| $C_1^{FD} - C_1^{FC}$ | 1.2 | 3.7 | 0.3 |
| $C_1^{C} - C_1^{FC}$ | 0.6 | 1.0 | 0.5 |
| $C_1^{D} - C_1^{FD}$ | 0.6 | 1.0 | 0.6 |
| $\Re(A_1^C)$ | 1.2 | 3.9 | 0.6 |
| $\Im(A_1^C)$ | 0.7 | 1.8 | 0.5 |
| $\Re(A_1^D)$ | 1.2 | 3.7 | 0.7 |
| $\Im(A_1^D)$ | 0.7 | 1.7 | 0.5 |

Table 4.1: Accuracy of the absolute focus determination. The absolute aberrations of individual images were determined from image regions C and D, either from the full series or from 3-member subseries. For instance, $C_1^{FC}$ represents the value obtained from the **full** series from region **C**. Column 1: $\sigma_{n-1}$ variation from a constant value. Column 2: fitted linear change. Column 3: remaining $\sigma_{n-2}$ variation from the fitted straight line.

astigmatism and defocus present in diffractograms were calculated both automatically using the Semper program 'dpfit' (Saxton, 2000b) and also manually by visual comparison of simulated and experimental diffractograms displayed as a merged image. The 'dpfit' program first establishes the mirror axes of the diffractogram by locating the maximum value in an orientation correlation function (OCF)[2] between the diffractogram and its mirror image. This reliably determines the principal axes of astigmatism and reduces the number of unknown parameters to 2. Subsequently, the agreement (measured as the correlation coefficient) between the experimental diffractogram pattern and patterns simulated for a range of focus values varied along the two principal astigmatism axes is tabulated in order to find the pair of focus values that yields the best agreement. Since this method is not applicable when crystalline material is in the field of view, diffractograms were calculated from images where only the thin amorphous specimen edge was selected using a manually created mask.

---

[2]The OCF between two images is defined as the correlation coefficient between the first image and a rotated version of the second image as a function of the rotation angle.
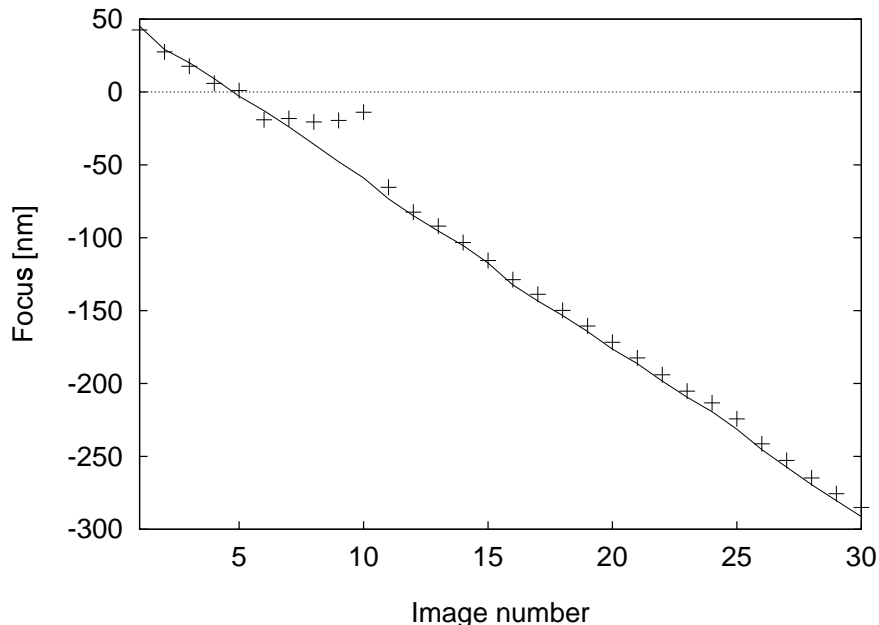
Figure 4.11: Focus levels determined from region A (figure 4.1) using the new method (line) compared to results from automated diffractogram fitting with the program 'dpfit' (crosses), using diffractograms calculated from the entire image fields with a mask selecting only the amorphous specimen edge.

Figure 4.11 illustrates the typically good agreement between values calculated using the new method and automated diffractogram fits for large underfocus values. However the automated diffractogram fitting routine fails for conditions close to Gaussian and Scherzer defoci, which do not show clear rings. Furthermore, some prior knowledge was required to restrict the automated search range to either over- or underfocus values for each of the images.

Using the alternative process of manual (visual) diffractogram fitting, the heavy streaking that results from masking out a narrow specimen region is usually more disturbing than the regular pattern of crystalline reflections. Therefore, the diffractogram patterns were in this case evaluated from the full specimen region marked A in figure 4.1.
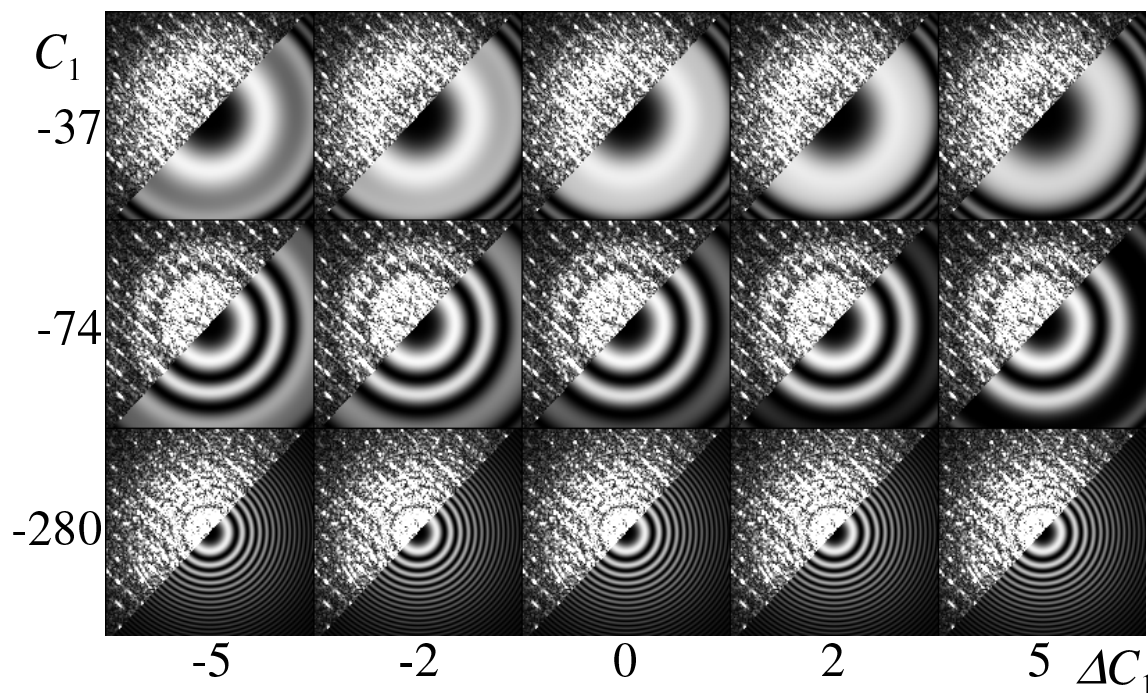
Figure 4.12: Comparison of the experimental diffractograms of images 8, 11 and 29 from the focal series with simulated diffractograms for the manually measured focus levels and for additional focus offsets of -5,-2,0,2 and 5 nm. It is apparent that a defocus mismatch of 2 nm would be difficult to detect in manual diffractogram fitting, whereas a mismatch of 5 nm is detectable, at least for the moderately underfocused image (middle row).

Figure 4.12 shows an excellent match between these diffractogram patterns and the ones expected from the manually determined defocus values. However deviations of $\pm 2$ nm in the measured values give an equally good visual match, whereas a mismatch is discernable for deviations of $\pm 5$ nm. It can therefore be concluded that the accuracy in values determined manually is clearly no better than 2 nm, with additional difficulties when no rings are visible, *i.e.* near Scherzer or Gaussian defoci.

### 4.3.4   The specimen focal plane

As shown in the previous section the absolute focus level can now be determined with an accuracy better than 1 nm, which is less than the typical TEM specimen thickness. Therefore an important question arises as to which plane in the specimen the focus level is determined relative to. This is of particular interest when the specimen consists of crystalline material with amorphous contamination layers or damage on both upper and lower surfaces.

To investigate this, multislice simulations (appendix A) were carried out for various sandwich structures comprising surface layers of amorphous carbon above and below crystalline $Nb_{16}W_{18}O_{94}$ in the [001] projection, matching the experimental data discussed in section 6.1. For this particular investigation the phase contrast index was used to determine the focus level in the calculated exit surface wavefunction according to the procedures described in section 1.6.2.

Figure 4.13 summarises the results of these simulations. With only a single amorphous layer, the determined focus level that yields the maximum $F_{PCI}$ is located accurately in the centre of that layer and is unaffected by the presence of a crystalline layer. In case of amorphous contamination on both crystal surfaces, the determined focus level deviates systematically from the centre of mass of the two amorphous layers according to the following trends:

1. The focus level is biased towards the thicker amorphous layer.

2. The contamination layer on the specimen exit surface has a stronger influence.

Although the underlying reasons behind these trends remain at present unclear, this systematic deviation in the location of the focal plane determined is noted here due to its potential importance in the measurement of defocus values from thick specimens. Figure 4.13 demonstrates that, depending on the relative thickness of the two contamination layers, the determined focus level can be above or below as well as within the crystal. However, for the important case where the crystal edge is within the field of view, any contamination on this crystal edge is nearly at the same height
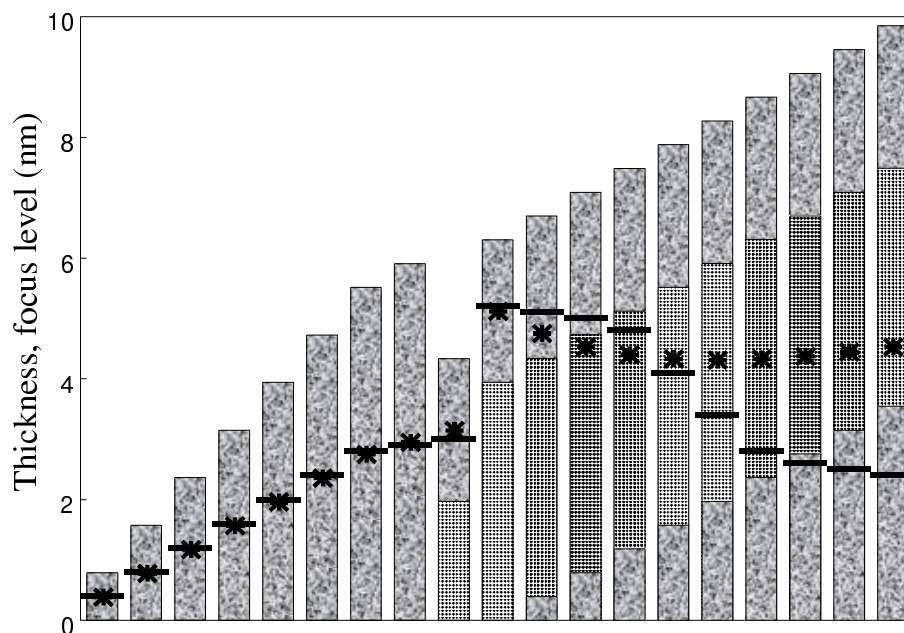
Figure 4.13: Focus levels (thick horizontal lines) determined for different simulated structures with amorphous layers (bars with a disordered pattern) and a crystalline layer (bars with a regular pattern). In case of crystalline material contaminated on both surfaces, the determined focus levels deviate systematically from the centre of mass of the amorphous layers (indicated by asterisks).

as the crystal and the determined focus value is expected to be close to the centre plane of the crystal.

## 4.3.5 Validity of the weak phase object approximation

The absolute aberration determination method discussed in the previous sections relies on the weak phase object (WPO) approximation (section 1.5) to be valid at most spatial frequencies, or, more precisely, it requires that $-\psi\psi_-$ has a positive real part for a large majority of spatial frequencies, which is a weaker requirement. The WPO approximation has been predicted to break down at high spatial frequencies for all but very thin amorphous specimens due to a purely kinematic projection effect,

with serious consequences for the determination of $C_3$ from single diffractograms at large underfocus (Gibson, 1994).

In the kinematic approximation, each slice of a thick specimen can be treated independently as a WPO out of focus by a distance $z$. Hence, the expectation value

$$f = \langle -\psi\psi_- \rangle = \int \mathrm{d}z \, \rho(z) c e^{-2\pi i z \lambda k^2} \tag{4.26}$$

for the contributions to the PCI can be evaluated, where $c$ is real and positive and incorporates scattering and structure factors, and $\rho(z)$ is the density distribution of the specimen.

For a uniform sample with thickness $t$ and $\rho(z) = \rho_0$ for $|z| < t/2$, the function

$$f = c\rho_0 t \operatorname{sinc}(\pi \lambda t k^2) \tag{4.27}$$

is real and positive for $\lambda t k^2 < 1$, a requirement that is 4 times less stringent than that for the validity of the WPO approximation given in Gibson (1994). This means that for a typical $k_{\max} = 7\,\mathrm{nm}^{-1}$ used for the PCI, sample thicknesses of up to 10 nm are still acceptable with 300 keV electrons.

## 4.4   Experimental Restoration Procedures.

With the theoretical tools and estimates of accuracy presented in the previous sections, various experimental imaging parameter estimation and wavefunction restoration schemes can now be assembled.

Two typical possibilities are illustrated schematically in figures 4.14 and 4.15. The first example (figure 4.14) shows a procedure for fast on-line focusing and astigmatism correction. A series of only three images is used and the relative focus values are assumed to be at their nominal levels,[3] so that the registration for each image requires the calculation of only one PCF, and the total computational effort is limited to 5 Fourier transforms.

---

[3]This assumption ultimately limits the accuracy of this fast on-line method when focus drift is present.

| Acquisition and control | Image processing |
|---|---|
| Acquire image 1 | |
| Change focus by 26 nm | FT image 1 |
| Acquire image 2 | |
| Change focus by 16 nm | FT image 2 |
| Acquire image 3 | PCF between 1,2 |
| | FT image 3 |
| | PCF between 1,3 |
| | Restore FT of image wave |
| | Use PCI to determine $C_1$, $A_1$ |
| Change focus to target, correct $A_1$ | |
| *Acquire image 4* | |
| | *FT image 4* |
| | *PCF image 4 with predicted image* |
| *Final correction of focus* | |
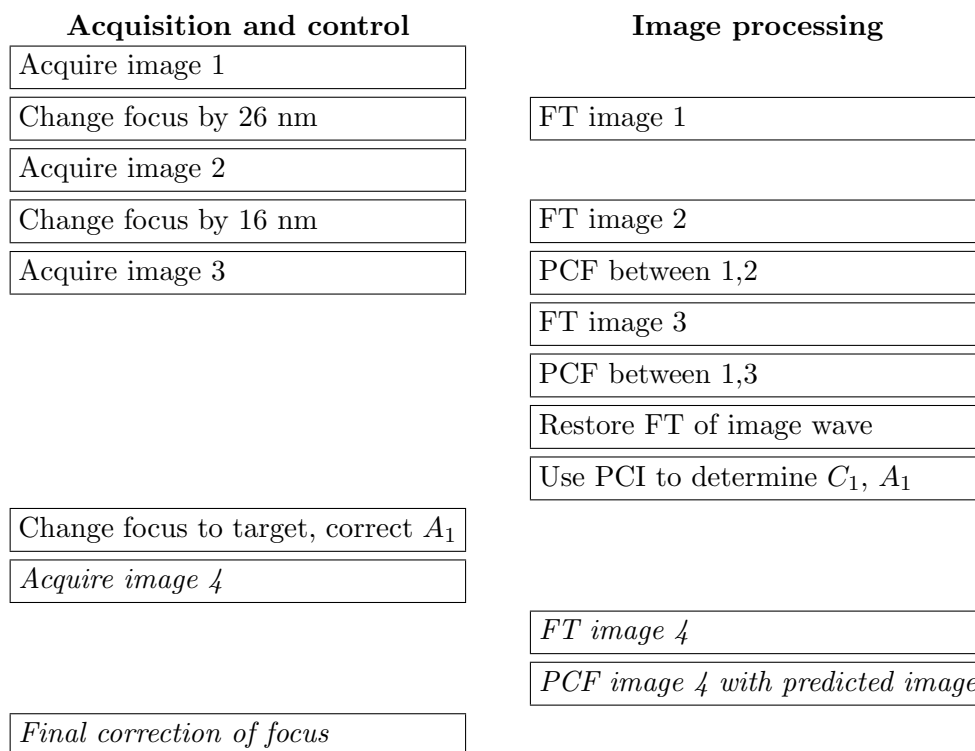
Figure 4.14: A fast on-line focusing and stigmation routine, where part of the image processing can be executed in parallel with the image acquisition. The last steps (in italics) are only necessary if the required focus change is large. The specific focus steps of 16 and 26 nm are suitable for the 3000F at high resolution.

The second example in figure 4.15 shows a flowchart for the off-line restoration of a focus series where the aberrations have to be determined to a higher precision. In this procedure the initial step is the determination of the relative defocus levels of two images relative to one chosen as a reference image[4] by maximising the peak height of the phase-compensated PCF. From these three images the image wave in the reference image plane is then restored. Additional images are subsequently included into the restoration with relative foci and displacements established by maximising the peak height of the PCF with images predicted from the current restored image wave. This

---

[4]The choice of the reference image is arbitrary as long as care is taken not to choose a low-contrast image taken near Gaussian focus.

160

| Choose reference image $n_1$. |
|---|

| Determine focus and registration of $n_2$ and $n_3$ relative to $n_1$ using the phase-compensated PCI. |
|---|

| For all other images $n$ | Determine focus level of image $n$ using the PCF with images predicted from the current restoration. |
|---|---|
| | Add image $n$ to the restoration. |

| Determine absolute $C_1$ and $A_1$ using the PCI. |
|---|

| Determine the antisymmetric aberrations. |
|---|

| Restore the object wave using the established absolute aberrations. |
|---|

Figure 4.15: High precision off-line aberration determination for image restoration. The determination of the antisymmetric aberrations requires additional data which is discussed in detail in chapter 5.

initial phase therefore yields the registration of all images, *i.e.* their relative shifts and focus levels to high accuracy.

In the second phase of the process the absolute focus and astigmatism are established using the PCI of the restored image wave, and finally, the restoration is repeated with the correct absolute imaging parameters. Repeating the restoration is preferred over simply correcting the restored image wave for the absolute focus and astigmatism, because in the latter case the focus-dependent spatial coherence envelope (section 1.3) for the individual images is not accommodated correctly.

# Chapter 5

# Antisymmetric aberration determination

The methods described in the previous chapter can also be used to measure the apparent defocus and astigmatism from a short focal series taken under tilted illumination conditions. The measurements of the apparent values of $C_1$ and $A_1$ for a number of different deliberately injected tilts can then be used to determine the antisymmetric aberrations as described in section 3.3.2 using a least squares fit or analytically for symmetrical tilt arrangements. Apart from $C_3$, only aberrations up to the third order in $\mathbf{k}$ (*i.e.* up to $A_2$ and $B_2$) were taken into account, since this provided excellent fits to the experimental data, indicating that a model that ignores higher order aberrations is sufficiently accurate.

## 5.1   Acquistion of tilt-focus series datasets

The datasets used in this chapter (*tiltf6* datasets) comprise 3-member focus series at 6 different tilt azimuth angles together with a 3-member series with axial beam direction and additional axial images between each pair of tilted datasets, as illustrated in figure 5.1.

The standard *tiltf6* dataset comprises 27 images, although for routine aberration
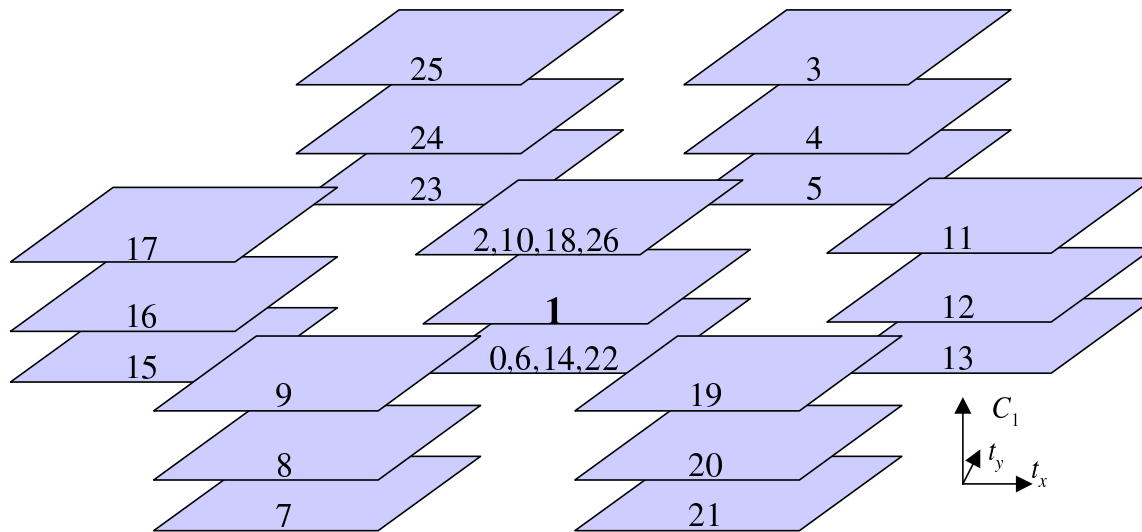
Figure 5.1: A schematic illustration of a combined defocus/tilt azimuth dataset *tiltf6* with images numbered in recording sequence. A three-member focal series is recorded for each of 6 different tilt azimuth angles and also for axial illumination. In addition, an axial image is recorded between each pair of tilted datasets with the focus level unchanged. This allows correction of the focus drift during acquisition. The order in which the images are recorded (indicated by numbers $0 \ldots 26$) was chosen to minimize the influence of focus drift and lens and deflector hysteresis by ensuring that focus and tilt are not changed together between subsequent images and by recording series with opposite tilts in immediate succession.

determination, smaller datasets can be used. The determination of both beam tilt and three-fold astigmatism $A_2$ requires 4 tilt azimuths, leading to a 'tiltf4' dataset comprising 19 images, and if the usually constant value of $A_2$ is known, a tiltf2 dataset (11 images) with 2 orthogonal tilts can be used.

In this work only *tiltf6* datasets were used, since the fitted parameters are over-determined by data from 6 tilt azimuths, which allowed an investigation of the accuracy of both the apparent $C_1$ and $A_1$ values and the underlying aberration model. In addition these datasets yield a uniform coverage of Fourier space when a tilt-series restoration of the aberration-corrected object wave is performed (*c.f.* section 1.6.3). Figure 5.2 illustrates the complete process of aberration determination and tilt series

| | |
|---|---|
| (1) Register images 0 and 2 with image 1 using phase-compensated PCFs. | |
| (2) Restore image wave in plane of image 1. | |
| (3) For all other axial images | (a) Register image to current restoration using the PCF with predicted images. |
| | (b) Add image to restoration |
| (4) Determine absolute focus and 2-fold astigmatism using the PCI. | |
| (5) For all tilted focus series. | (a) Coarsely register the position of the central tilted image |
| | (b) Register first and last image to middle image using PCFs. |
| | (c) Restore aberrated wavefunction from the three tilted images. |
| | (d) Determine apparent absolute defocus and astigmatism using the PCI. |
| At this point, the apparent absolute defoci $C_{1n}$ and astigmatisms $A_{1n}$ ($n \in \{0 \ldots 26\}$) of all images are known. | |
| (7) Determine drift-corrected aberrations for the tilted sets ($n \in \{0 \ldots 5\}$) as $A_1^{(n)} = A_{1(4n+4)}$ and $C_1^{(n)} = C_{11} + (C_{1(4n+3)} + C_{1(4n+5)} - C_{1(4n+2)} - C_{1(4n+6)})/2$. | |
| (8) Determine the antisymmetric aberrations, including the beam tilt for the axial images and the magnitude and direction of the injected tilt, analytically or numerically. | |
| (9) Re-calculate the restoration from the axial images | |
| (10) For all tilted focus series. | (a) Calculate restoration from tilted images with beam tilt corrected. |
| | (b) Register with restoration from axial images using PCF. |
| (11) Calculate the restoration from all images | |

Figure 5.2: Procedure for the determination of all aberration coefficients (steps (1) to (8)) and subsequent tilt series restoration from a *tiltf6* dataset (steps (9) to (11)).

| Set | Specimen | Mag. | tilt [DAC] | $k_t$ | Mode | Date | Section |
|---|---|---|---|---|---|---|---|
| Ge1 | amorphous Ge | 250 kX | 800–3200 | 2.95 | ima. | 23-08-2000 | 5.2 |
| Ge2 | amorphous Ge | 250 kX | 800–3200 | 2.89 | ima. | 23-08-2000 | 5.2 |
| Ge3 | amorphous Ge | 250 kX | 800–3200 | 2.72 | ima. | 23-08-2000 | 5.2 |
| Ge4–Ge6 | amorphous Ge | 500 kX | 800 | 1.69 | ima. | 19-11-2001 | 5.2.2 |
| Ge7 | amorphous Ge | 300 kX | 1600 | 1.53 | ima. | 19-11-2001 | 5.2.2 |
| Ge8–Ge10 | amorphous Ge | 500 kX | 1600 | 1.78 | ima. | 19-11-2001 | 5.2.2 |
| dp1–dp4 | Au foil | 500 kX | 800–4000 | 1.75 | diffr. | 19-11-2001 | 5.2.2 |
| dp5 | Au foil | 40 kX | 800–4000 | 5.79 | diffr. | 19-11-2001 | 5.2.2 |
| dp6 | Au foil | 100 kX | 800–4000 | 5.86 | diffr. | 19-11-2001 | 5.2.2 |
| dp7 | Au foil | 120 kX | 800–4000 | 5.99 | diffr. | 19-11-2001 | 5.2.2 |
| NdSr100 | $Nd_4SrTi_5O_{17}$ [100] | 600 kX | 1600 | 2.56 | ima. | 15-03-2000 | 5.3 |
| NdSr010 | $Nd_4SrTi_5O_{17}$ [010] | 600 kX | 800 | 2.34 | ima. | 15-03-2000 | 5.3 |

Table 5.1: Key parameters of the image series discussed in this chapter. The series recorded in imaging mode are *tiltf6* series while the 'dp' series, recorded in diffraction mode, comprise images at 4 tilt azimuths with 5 different tilt magnitudes (*tilt4m5* series). Sets Ge1–Ge3 each comprise 4 *tiltf6* series with different tilt magnitudes. The nominal injected tilts (in DAC units) have to be multiplied with the tilt calibration factor $k_t$ (section 5.2.2) to yield the actual tilt (in $\mu$rad).

restoration from a *tiltf6* dataset.

In the following, exemplary results from *tiltf6* datasets of both crystalline specimen with little amorphous contamination and amorphous germanium film are presented. Table 5.1 lists key parameters of the series used as examples in this chapter.

## 5.2 Amorphous germanium sample

For a thorough investigation of the new method, 3 sets (Ge1–Ge3) of independent *tiltf6* series of a thin amorphous germanium foil were acquired. Each set comprised 4 series with injected tilts of 800, 1600, 2400 and 3200 DAC units taken from the same specimen region. Between sets, a new specimen region was sought and the illumination was recentred.

In order to observe any change in the aberration parameters over the field of view and hence deviations from the isoplanatic approximation (*c.f.* section 1.1),

the complete aberration determination procedure was repeated on 49 independent $128 \times 128$ pixel subregions centred at $128(n, m)$ with $n, m \in -3 \ldots 3$.[1]

In the datasets with large beam tilts, the astigmatism determination algorithm failed to find the global maximum of the $F_{\text{PCI}}$ for some subregions. Instead, a local maximum with a small value of $A_1$ was found. This is due to the artefact described in section 4.2.6. For the initial trial value $A_1 = 0$, the noise in the images gives rise to a false weak circular ring pattern in the $f_{\text{PCI}}$ in addition to the strongly elliptical ring pattern, leading to spurious maxima in the $F_{\text{PCI}}$. However, this effect can easily be avoided by starting from a nonzero $A_1$ trial value.

Table 5.2 summarises the results for the parameters that did not display any systematic change over the field of view. The statistical errors given in brackets are determined from the spread of the values measured from all regions. This does not include any systematic errors that affect all subregions equally, *e.g.* deviations in the injected tilts due to tilt coil hysteresis. Therefore, the variation of the parameters from series to series is larger than the statistical errors given. Ignoring the values from series with an injected tilt of 800 DAC units, which are the least accurate, the best estimate for $A_2$ is $A_2 = (805 - 205i)$nm with an accuracy of 20nm.

The variation of the tilt coil calibration is significantly larger from set to set than within each set, suggesting that these calibration values depend slightly on the illumination settings.

## 5.2.1 Illumination convergence

The beam misalignment was found to vary significantly across the field of view for two of the datasets. Figure 5.4 shows that the variation of the beam tilt across the field of view can be fitted linearly by $t_x(x) = t_x + m_x x$ and $t_y(y) = t_y + m_y y$ with nearly equal negative slopes ($m_x \approx m_y$). This implies that the beam was convergent for sets Ge1 and Ge2 while a nearly parallel illumination condition was achieved for

---

[1]This leaves an unused 64 pixel wide strip around the reference image to ensure that all subregions are fully contained in each image in spite of small specimen drift.

| set | tilt | $C_1$ [nm] | $A_1$ [nm] | | $A_2$ [nm] | | $k_t$ | $\theta$ |
|-----|------|-----------|-----------|------------|-----------|-----------|-----------|-----------|
| Ge1 | 800 | -55.61(21) | -0.60(6) | -1.94(10) | 782(13) | -169(12) | 2.869(20) | 1.431(8) |
| Ge1 | 1600 | -44.31(22) | -1.07(9) | -1.75(8) | 831(9) | -188(10) | 2.966(8) | 1.435(2) |
| Ge1 | 2400 | -41.10(17) | -1.12(10) | -1.91(8) | 789(6) | -218(5) | 2.940(4) | 1.431(1) |
| Ge1 | 3200 | -33.75(28) | -1.69(14) | -1.39(12) | 785(7) | -210(6) | 2.948(2) | 1.433(1) |
| Ge2 | 800 | -91.06(14) | 0.09(8) | -1.23(7) | 763(15) | -186(12) | 2.928(17) | 1.447(10) |
| Ge2 | 1600 | -104.50(18) | -0.35(7) | -1.54(7) | 781(7) | -190(8) | 2.886(6) | 1.429(2) |
| Ge2 | 2400 | -100.06(17) | -0.32(8) | -2.01(6) | 801(5) | -183(5) | 2.889(2) | 1.427(1) |
| Ge2 | 3200 | -101.58(16) | -0.50(8) | -1.83(8) | 802(5) | -211(4) | 2.893(2) | 1.432(1) |
| Ge3 | 800 | -133.61(17) | 0.91(9) | -0.83(9) | 801(15) | -236(19) | 2.766(21) | 1.410(8) |
| Ge3 | 1600 | -136.37(19) | 1.60(7) | -1.22(7) | 842(6) | -211(6) | 2.711(4) | 1.406(2) |
| Ge3 | 2400 | -134.46(16) | 0.79(6) | -1.51(6) | 809(4) | -224(4) | 2.723(2) | 1.411(1) |
| Ge3 | 3200 | -138.08(17) | -0.12(6) | -1.34(7) | 810(4) | -214(3) | 2.727(1) | 1.410(1) |

Table 5.2: Aberration parameters determined from 12 *tiltf6* series. The values given are averages obtained from 49 independent subregions with the statistical error in the last digits given in brackets. The tilt coil calibration factor $k_t$ is given in mrad/(1000DAC), and $\theta$ gives the direction of the tilt $x$ axis in radians. The accuracy of the $A_2$ measurement and the tilt calibration increases with increasing injected tilt.

set Ge3. To avoid confusion, it is important to distinguish between three different convergence angles as illustrated in figure 5.3.

Table 5.3 shows that this linear fit gives an excellent match to the experimental data and, as with the three-fold astigmatism, the accuracy increases with increasing injected tilt angle. The variation of the measured central beam tilt $(t_x, t_y)$ within each set is very small, indicating an accuracy of 0.02mrad, which is sufficient for a resolution of $11\text{nm}^{-1}$ according to table 1.2. However, if the illumination is not parallel, a large beam misalignment remains near the borders of the field of view. With a beam convergence as in set Ge2 ($m_x = m_y = 1.9\,\text{mrad}/1024\,\text{pixel}$) and the beam tilt corrected in the image centre, only a central region with radius $t_{\text{max}}/m_x = 43\,\text{pixel}$ is aligned to an accuracy better than $t_{\text{max}} = 0.08\,\text{mrad}$ necessary for a resolution of $7\,\text{nm}^{-1}$. It is possible to correct for this spatial variation of the axial coma *a posteriori* by applying a parabolic phase plate on the real space restored image wave followed by the usual correction phase plate in Fourier space and finally the conjugate of the initial parabolic phase in real space (Lehmann, 2001).
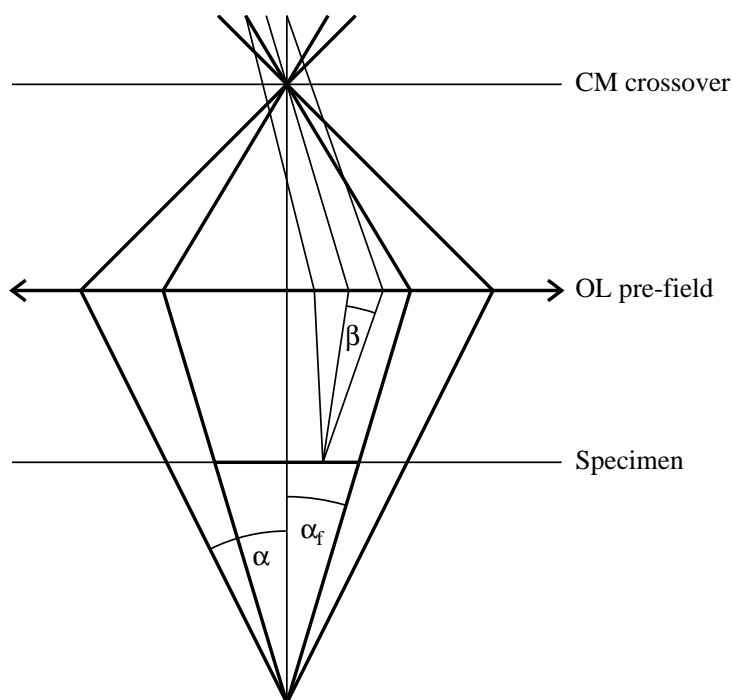
Figure 5.3: Ray diagram to illustrate convergence angles. Due to the finite source size, each point on the specimen is illuminated from a range of directions, characterised by the beam divergence $\beta$, which gives the RMS spread of the beam direction at a given specimen point. It is thereby defined by the angle subtended by the source at the specimen and depends on the physical source size for a field emission source and on the illuminated condensor lens aperture for a thermionic source. When the illumination is not parallel, the mean beam direction varies with position on the specimen. This is described by the convergence angle $\alpha$ between central and peripheral rays in the illuminated patch. The field of view convergence angle $\alpha_f < \alpha$ is measured between central and peripheral rays in the field of view. In this drawing, the objective lens post-field is ignored and all angles are exaggerated.
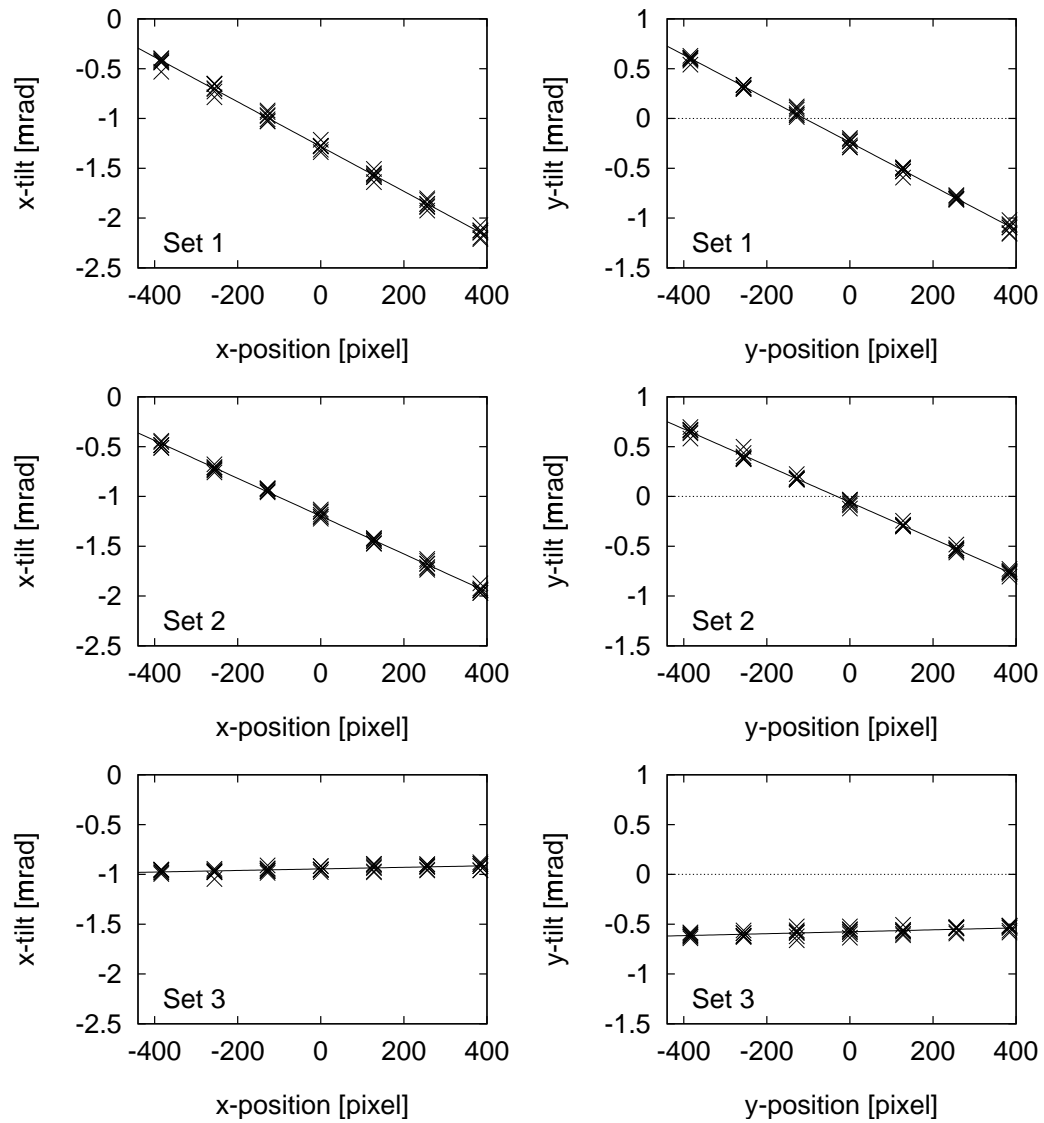
Figure 5.4: Measured beam tilt components as a function of position in the field of view. The linear dependence with negative slope shows that the beam was convergent for sets 1 and 2. In set 3, nearly parallel illumination has been achieved.

| set | tilt | $t_x$ | $t_y$ | $m_x$ | $m_y$ | $\sigma_{t1}$ | $\sigma_{t2}$ |
|-----|------|-------|-------|-------|-------|---------------|---------------|
| Ge1 | 800 | -1.358(17) | -0.327(12) | -2.213(67) | -2.276(50) | 0.815 | 0.146 |
| Ge1 | 1600 | -1.263(6) | -0.250(7) | -2.388(26) | -2.244(27) | 0.830 | 0.066 |
| Ge1 | 2400 | -1.297(5) | -0.212(5) | -2.294(22) | -2.206(21) | 0.806 | 0.053 |
| Ge1 | 3200 | -1.282(6) | -0.240(5) | -2.301(24) | -2.249(20) | 0.815 | 0.055 |
| Ge2 | 800 | -1.278(10) | -0.033(15) | -1.812(40) | -1.872(62) | 0.670 | 0.128 |
| Ge2 | 1600 | -1.209(7) | -0.152(9) | -1.975(27) | -1.881(35) | 0.693 | 0.077 |
| Ge2 | 2400 | -1.207(6) | -0.092(6) | -1.963(23) | -1.884(23) | 0.689 | 0.057 |
| Ge2 | 3200 | -1.198(4) | -0.058(4) | -1.944(17) | -1.881(17) | 0.684 | 0.042 |
| Ge3 | 800 | -0.963(15) | -0.533(15) | 0.133(58) | 0.096(62) | 0.153 | 0.149 |
| Ge3 | 1600 | -0.973(6) | -0.582(7) | 0.097(23) | 0.113(29) | 0.075 | 0.065 |
| Ge3 | 2400 | -0.956(5) | -0.576(4) | 0.048(19) | 0.118(17) | 0.055 | 0.045 |
| Ge3 | 3200 | -0.945(4) | -0.576(4) | 0.081(15) | 0.099(17) | 0.051 | 0.039 |

Table 5.3: Linear fits $t_x(x) = t_x + m_x x$ and $t_y(y) = t_y + m_y y$ to the variation of the beam misalignment over the field of view. The fitted parameters $t_x, t_y$ and $m_x, m_y$ are given in mrad and mrad/(1024pixel), respectively, with the statistical error in the last digit given in brackets. $\sigma_{t1}$ is the standard deviation of the measured misalignment vectors and $\sigma_{t2}$ is the residual deviation from the fitted linear tilt variation.

It is however preferable to set up parallel illumination in the microscope, but the presence of this strong beam convergence in two of the datasets even though the microscope was carefully aligned shows that it is difficult to achieve parallel illumination. An automated procedure for measuring and correcting beam convergence would involve measuring the beam tilt at the image corners. A faster alternative requiring just two images exploits the fact that convergent or divergent illumination causes a small change in magnification as the focus is changed. This can be detected by calculating the phase compensated PCF on subregions near the image corners between 2 images with a large focus difference, providing a more rapid method for setting up parallel illumination conditions.

## 5.2.2   Tilt calibration

The tilt magnitude calibration value $k_t = 2.80$mrad/1000DAC established from the amorphous germanium *tiltf6* datasets discussed above is much smaller than the value

| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$-tilt | 0 | 1 | -1 | 2 | -2 | 3 | -3 | 4 | -4 | 5 | -5 | |
| $y$-tilt | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| $N$ | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| $x$-tilt | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $y$-tilt | 0 | 1 | -1 | 2 | -2 | 3 | -3 | 4 | -4 | 5 | -5 | 0 |

Table 5.4: Order in which the diffraction patterns in a *tilt4m5* dataset are recorded. The injected tilts are in units of 800 DAC units.

| Set | CL2 setting | illumination | $\alpha$ [mrad] | $k_t$ | $\theta$ |
|---|---|---|---|---|---|
| dp1 | 42200 | parallel | <0.1 | 1.761 | -0.950 |
| dp2 | 42088 | divergent | 0.73 | 1.751 | -0.950 |
| dp3 | 42028 | divergent | 1.16 | 1.747 | -0.952 |
| dp4 | 42385 | convergent | 1.14 | 1.761 | -0.950 |

Table 5.5: Tilt coil calibration at a magnification of $500\,000$ with an $\alpha$-angle of 1. The calibration factor $k_t$ varies by less than 1% as the CL2 setting is changed to achieve parallel, divergent and convergent illumination. The convergence angle $\alpha$ is determined from the diffraction disc radius. The tilt orientation $\theta$ varies by less than 2 mrad, but there is a rotation by -2.36 rad relative to the angle measured in imaging mode using *tiltf6* series.

4.66mrad/1000DAC measured earlier using diffraction patterns (Dunin-Borkowski, 1999). In order to investigate whether this is due to a systematic discrepancy between measurements from diffraction patterns and from *tiltf6* Zemlin tableaus, the experiment was repeated using both methods in succession. For the diffraction pattern measurements, a gold $\langle 100 \rangle$ single crystal film was used and diffraction patterns were recorded in *tilt4m5* datasets, which comprise tilts in 4 directions with 5 tilt magnitudes (table 5.4). The peak positions were analysed to measure the shift between the diffraction patterns, with the camera length determined using the known spacing of the gold crystal. To improve the accuracy, the shift vectors obtained from different peaks were averaged, using only peaks that did not saturate the CCD camera.

| Set | Magnification | $k_t$ | $\theta$ |
|-----|---------------|-------|----------|
| dp5 | 40 000 | 5.789 | -1.374 |
| dp6 | 100 000 | 5.861 | -1.376 |
| dp7 | 120 000 | 5.992 | -1.377 |

Table 5.6: Tilt coil calibration at magnifications within the medium magnification range with an $\alpha$-angle of 3. The calibration factor $k_t$ is larger by a factor of 3.3 compared with $\alpha$-angle 1 and increases by 5% as the magnification is increased from 40 000 X to 120 000 X.
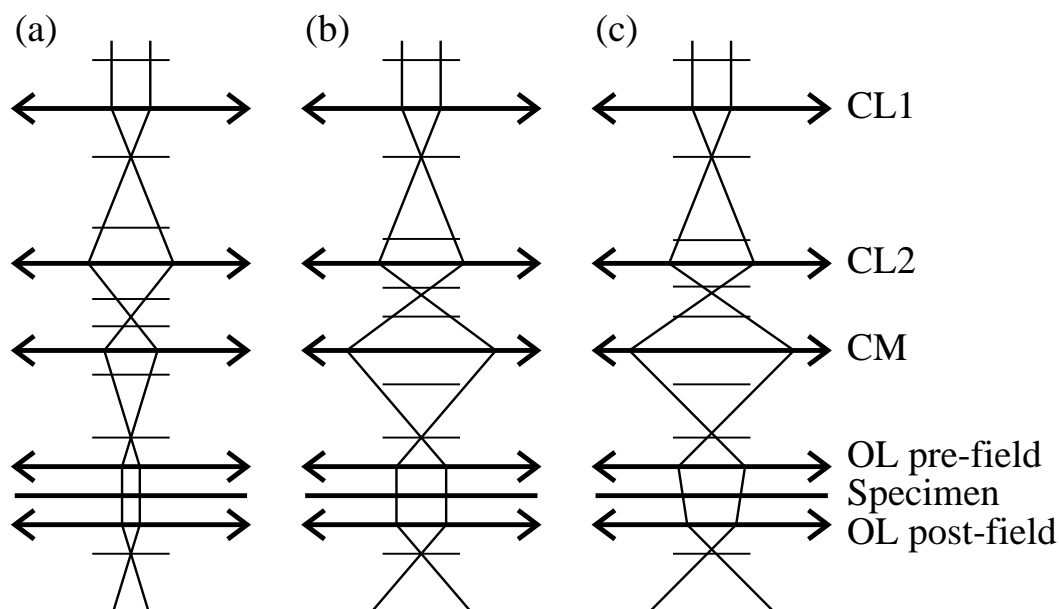


Figure 5.5: Illumination system of a typical field-emission TEM in imaging mode. For each lens, the focal planes are indicated by short horizontal lines. For a given condensor mini-lens (CM) excitation, the brightness control (condensor 2, CL2) can be adjusted such that the illumination is parallel. (a) For a large CM excitation, the parallel illumination is concentrated on a small patch, suitable for high magnification. (b) For lower magnification, a smaller CM excitation is used to yield a larger illuminated patch. A stronger CL2 excitation is then needed for parallel illumination. (c) The CL2 excitation is too strong for the given CM excitation, hence the CM crossover is slightly above the objective lens pre-field focal plane, giving rise to a convergent rather than parallel illumination. Similarly, divergent illumination results when the CL2 excitation is too weak.

In order to investigate the dependence of the tilt calibration on the illumination conditions, series were recorded with different illumination settings. It is known that the calibration factor depends strongly on the condensor mini-lens (CM) excitation which can be controlled by the "$\alpha$-angle setting" on JEOL microscopes.[2] Figure 5.5 illustrates how the CM lens excitation determines the size of the specimen region illuminated by a parallel beam. Table 5.5 lists the tilt calibration results obtained with *tilt4m5* sets of diffraction patterns at high magnification with an $\alpha$ angle setting of 1 appropriate to high magnification imaging. The tilt calibration factor $k_t = 1.75$ agrees well with the result $k_t = 1.71$ obtained using *tiltf6* series recorded the same day (figure 5.6) and shows that $k_t$ does not change significantly when the illumination is made convergent or divergent using CL2.

For comparison, data were also obtained with an $\alpha$-angle of 3 at lower magnification. Table 5.6 shows that $k_t$ increases by a factor of 3 at this setting and that it varies by 5% as the magnification is increased across the range from 40 kX to 120 kX.

Figure 5.6 shows the good agreement of the calibration factors measured with images and diffraction patterns. However, comparing all measurements shows a significant change in the calibration factor over time. The reason for this is not known although it may be due to movement or erosion of the emission tip. Since this suggests that the tilt coil strength cannot be regarded as a constant for the microscope, it should be included as free parameter in routine aberration determination. However, when measurements of apparent $C_1$ and $A_1$ values in a *tiltf6* or other Zemlin tableaus are used to measure $C_3$, it is necessary to independently calibrate the tilt coil strength using diffraction patterns immediately before or afterwards.

## 5.3   Crystalline samples

The antisymmetric aberration determination method was also tested using largely crystalline samples and two examples of *tiltf6* series of the complex oxide $Nd_4SrTi_5O_{17}$

---

[2]This $\alpha$-angle setting should not be confused with the convergence angle $\alpha$ defined in figure 5.3.
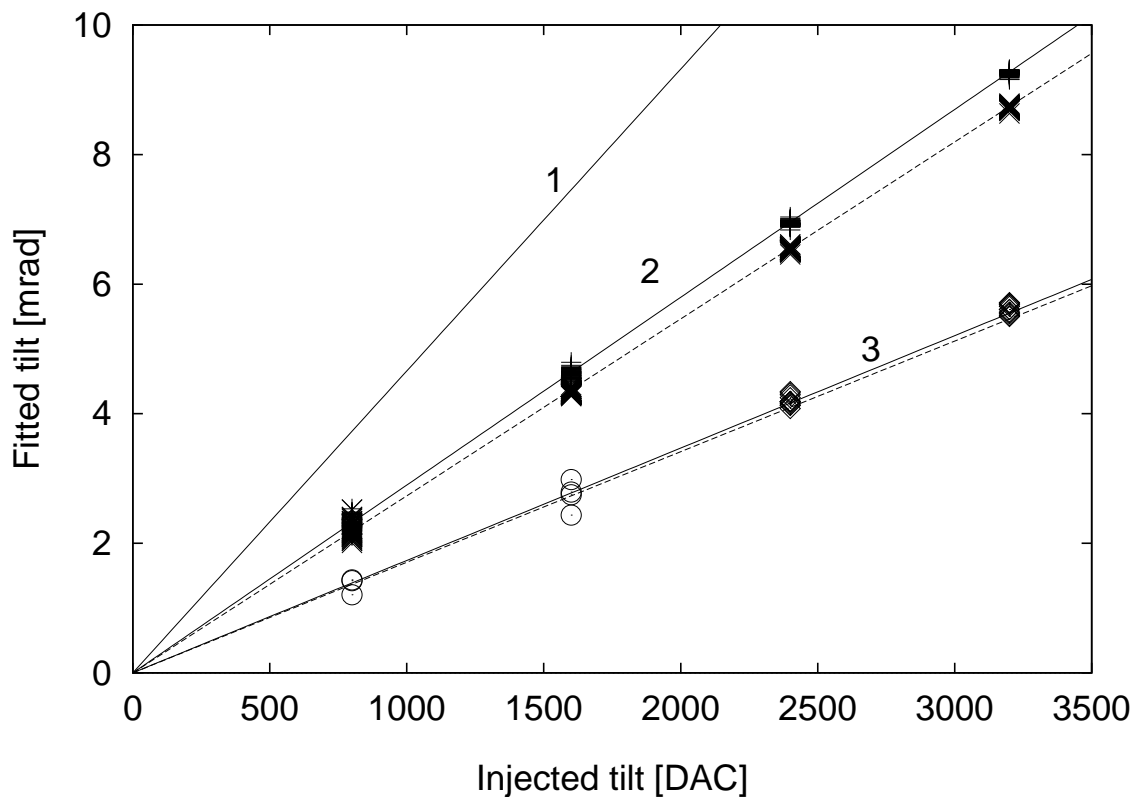
Figure 5.6: Calibration of the tilt coils of the JEM3000F microscope in Oxford. 1) Calibration obtained in September 1999 using diffraction patterns (Dunin-Borkowski, 1999) 2) Data obtained from sets of *tiltf6* series of amorphous germanium recorded in August 2000. Plus signs and fitted solid line: results from 49 independent image regions of a set Ge2. Crosses and fitted dashed line: the same for set Ge3. 3) Data recorded in November 2001. Diamonds and solid line: calibration using diffraction patterns (800 and 1600 DAC results not shown for clarity). Circles and dashed line: results from *tiltf6* series.

| set | tilt | $C_1$ [nm] | $A_1$ [nm] | | $A_2$ [nm] | | $t_x$ | $t_y$ | $k_t$ | $\theta$ |
|---|---|---|---|---|---|---|---|---|---|---|
| NdSr100 | 1600 | 8.1 | -3.6 | 1.1 | 863 | -160 | .68 | -.52 | 2.56 | 1.39 |
| NdSr010 | 800 | -7.5 | -5.3 | 5.7 | 944 | -240 | .43 | -3.37 | 2.34 | 1.38 |

Table 5.7: Measured aberration coefficients for two *tiltf6* datasets of $Nd_4SrTi_5O_{17}$.

| No | tilt azimuth | | Defocus $C_1$ | | | | | Astigmatism $A_1$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coils | image | obs. | drift | corr. | fitted | $|Err|$ | observed | | fitted | | $|Err|$ |
| 0 | 0 | 0 | 9.0 | 0.0 | 9.0 | 9.3 | 0.3 | -0.5 | 1.4 | -1.6 | 1.4 | 1.1 |
| 1 | 0 | 80 | 29.4 | 3.5 | 25.9 | 25.3 | 0.6 | -7.5 | -0.2 | -7.9 | 0.7 | 0.9 |
| 2 | 180 | -100 | 40.0 | 7.0 | 33.0 | 33.0 | 0.0 | -14.2 | 8.4 | -14.1 | 8.7 | 0.3 |
| 3 | -60 | 20 | 41.8 | 9.0 | 32.8 | 33.1 | 0.4 | 14.9 | 3.6 | 15.4 | 3.3 | 0.5 |
| 4 | 120 | -160 | 33.8 | 9.5 | 24.3 | 25.1 | 0.8 | -3.3 | 12.8 | -3.6 | 12.2 | 0.7 |
| 5 | -120 | -40 | 48.6 | 11.0 | 37.6 | 37.0 | 0.6 | 5.8 | -8.7 | 6.5 | -8.8 | 0.7 |
| 6 | 60 | 140 | 33.0 | 11.5 | 21.5 | 21.3 | 0.2 | -6.7 | -7.9 | -6.2 | -7.9 | 0.5 |

Table 5.8: Comparison of observed and fitted apparent aberration parameters for dataset NdSr100. The apparent $C_1$ values are corrected for the focus drift measured using the intermediate axial images. The mean RMS misfit is only 0.5 nm.

recorded in March 2000 are analysed in this chapter. A detailed discussion of this specimen is given in section 6.2 and only two series with injected beam tilts of 1600 and 800 DAC units, respectively, will be discussed here. Table 5.7 lists the aberration parameters determined and table 5.8 shows the agreement between observed and fitted parameters for datset NdSr100, with an RMS misfit of 0.5 nm. It should be noted that a relatively large focus drift was detected using the intermediate axial images and without focus drift compensation, the quality of the fit deteriorates (table 5.9). For dataset NdSr010, the RMS misfit is 0.7 nm.

The deviations $|\Delta A_2|$ of 73 nm and 143 nm from the value $A_2 = 805 - 205i$ established from datasets Ge1-Ge3 indicated a lower accuracy than for the Ge datasets, possibly due to the small injected tilt. However, values for $A_2$ are in satisfactory agreement with those measured using amorphous germanium. The tilt calibration factors $k_t$ are slightly smaller than the value $k_t = 2.72$mrad/1000DAC established from dataset GGe3.

| No | tilt azimuth | | Defocus $C_1$ | | $|Err|$ | Astigmatism $A_1$ | | | | $|Err|$ |
|----|------|-------|------|--------|------|------|------|------|-------|------|
| | coils | image | obs. | fitted | | observed | | fitted | | |
| 0 | 0 | 0 | 9.0 | 14.0 | 5.0 | -0.5 | 1.4 | -1.6 | 1.4 | 1.1 |
| 1 | 0 | 81 | 29.4 | 32.3 | 2.9 | -7.5 | -0.2 | -9.1 | 0.9 | 1.9 |
| 2 | 180 | -99 | 40.0 | 41.6 | 1.6 | -14.2 | 8.4 | -15.9 | 9.2 | 1.9 |
| 3 | -60 | 21 | 41.8 | 40.4 | 1.4 | 14.9 | 3.6 | 16.5 | 4.1 | 1.6 |
| 4 | 120 | -159 | 33.8 | 33.5 | 0.3 | -3.3 | 12.8 | -2.5 | 13.7 | 1.2 |
| 5 | -120 | -39 | 48.6 | 45.0 | 3.6 | 5.8 | -8.7 | 6.6 | -10.5 | 2.0 |
| 6 | 60 | 141 | 33.0 | 28.9 | 4.1 | -6.7 | -7.9 | -5.5 | -9.3 | 1.9 |

Table 5.9: Raw data as in table 5.8, but fitted without focus drift compensation. The RMS misfit is increased fourfold to 2nm, demonstrating the necessity of focus drift correction.

An elegant way of demonstrating the beam-tilt induced change in imaging parameters is the Zemlin-tableau (Zemlin et al., 1978), a composite picture where the diffractograms are displayed at a position appropriate to the beam tilt direction. Figure 5.7 displays such a tableau of diffractograms merged with simulated diffractograms calculated using the automatically determined parameters. Automated fitting of the diffractograms would have posed great difficulties due to the crystalline reflections.

Figure 5.8 shows a similar tableau of the uncorrected image wave PCI's, where no disturbing influence of the crystalline reflections is visible. Even in cases where the beams diffracted by the crystal do not obey the symmetry relations expected for weak phase objects, their influence on the integrated PCI used in the parameters estimation is small, since all Fourier space pixels are used with equal weight.

The examples discussed in this section demonstrate that the method developed in chapter 4 can successfully be applied to short focus series under tilted illumination. This makes it possible to determine both symmetric and antisymmetric aberration coefficients from a combined tilt/focus series, using the $f_{\mathrm{PCI}}$ rather than diffractogram fitting to determine the apparent aberrations at each beam tilt. The method can be applied to small subregions, allowing investigations of the change in imaging parameters across the field of view. In two of the examples shown, a large unintentional beam convergence was detected, which has implications for high resolution imaging.

For crystalline samples, it has been demonstrated that the method is accurate even if there is very little amorphous contamination under which conditions diffractogram based methods would fail. The low RMS misfit between observed and fitted parameters in the highly overdetermined *tiltf6* datasets confirmed that the accuracy of the apparent $C_1$ and $A_1$ values at each beam tilt measured from 3-member focus series using the method described in chapter 4 is better than 1 nm.
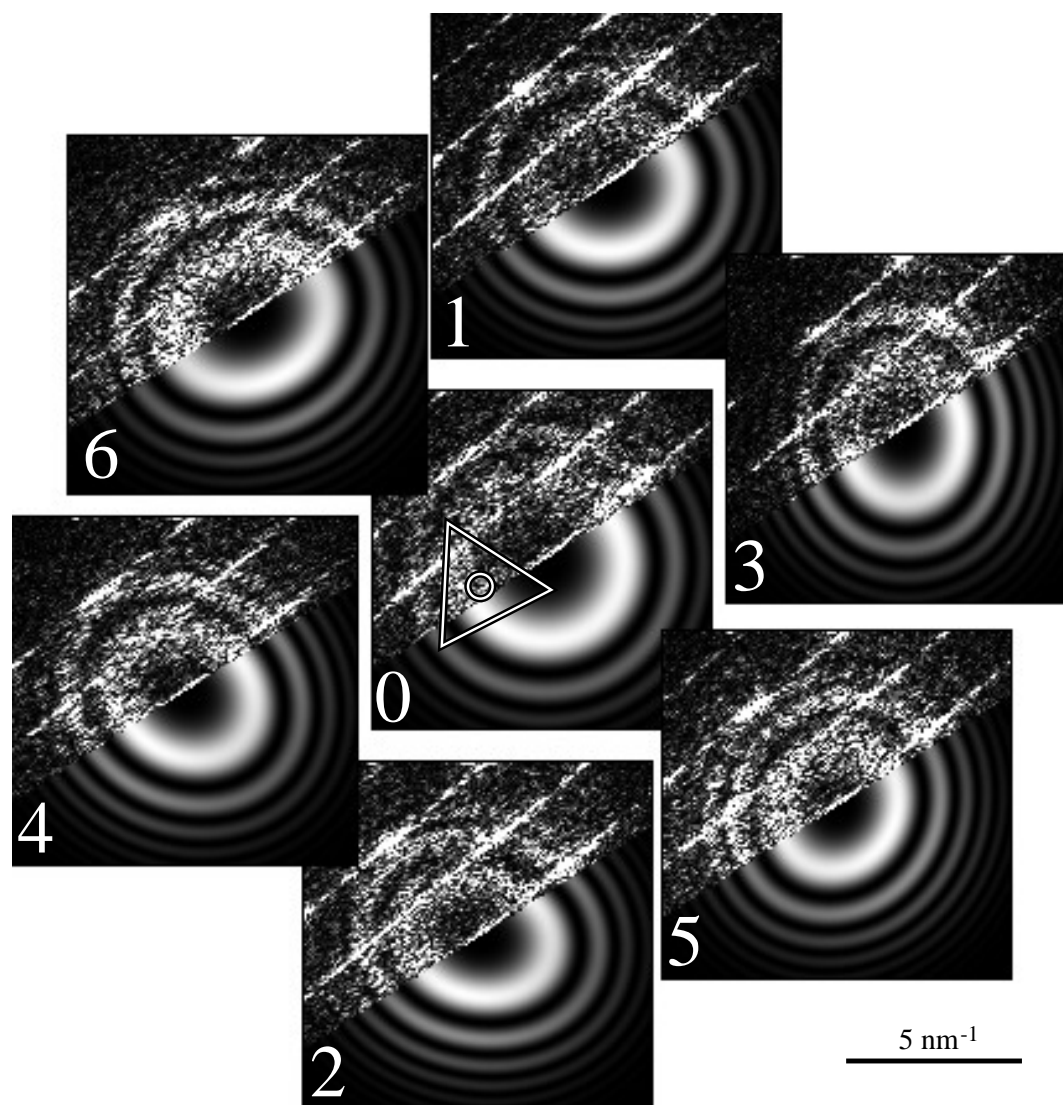
Figure 5.7: Zemlin Tableau of dataset NdSr100, showing the diffractogram calculated from the middle image of the short focal series at each tilt. These patterns are in good agreement with diffractograms simulated using the automatically determined apparent $C_1$ and $A_1$ values. The tilt magnitude is 1600 DAC units, corresponding to 4.1 mrad. The equilateral triangle marks the coma-free axis and indicates the direction of the three-fold astigmatism.
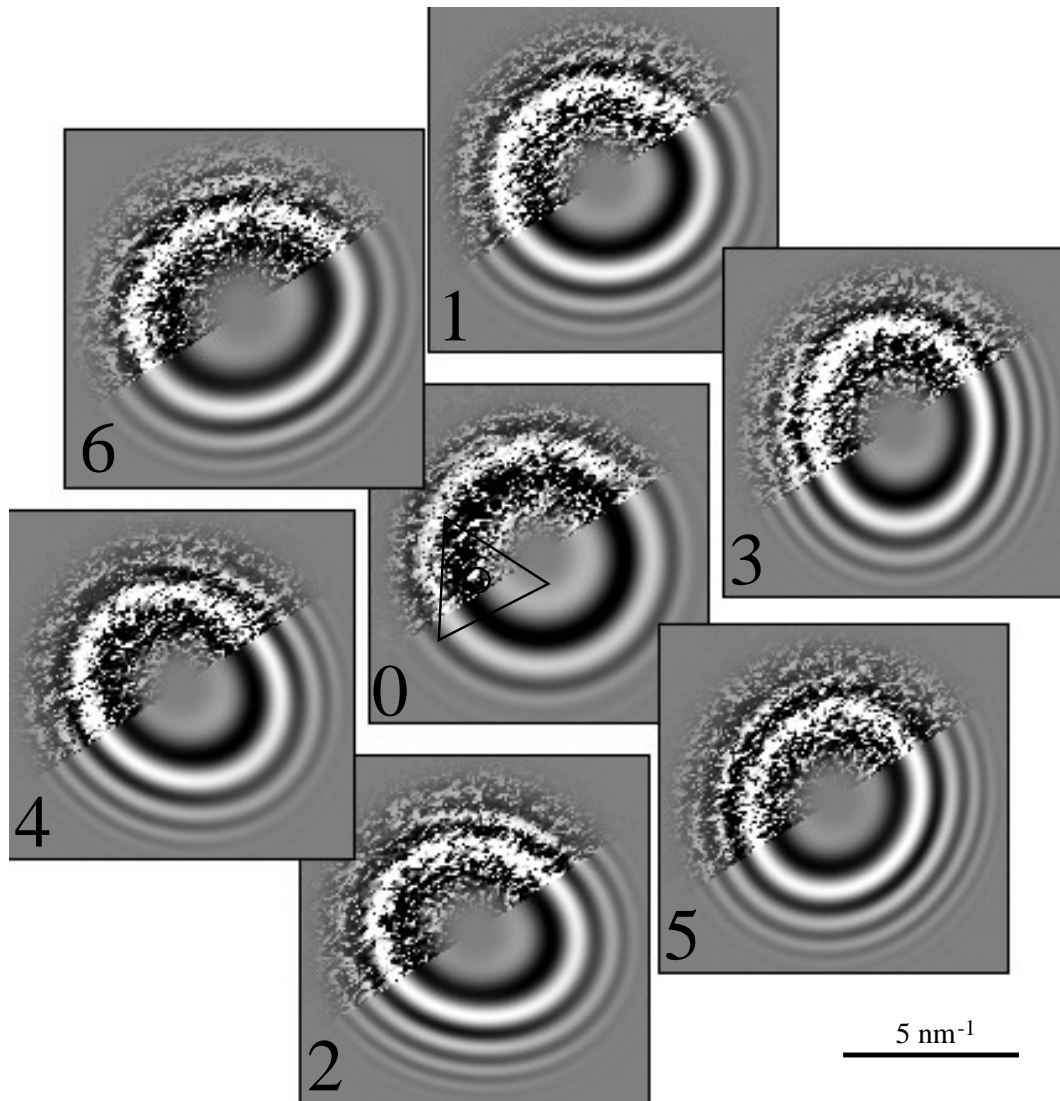
Figure 5.8: Tableau in which the uncorrected image wave $f_{\mathrm{PCI}}$ calculated from a three-member focal series is displayed for each beam tilt of the *tiltf6* dataset NdSr100. The simulated patterns merged with the experimental ones are calculated using the automatically determined aberration coefficients. The tilt magnitude is 1600 DAC units, corresponding to 4.1 mrad. The equilateral triangle marks the coma-free axis and indicates the direction of the three-fold astigmatism.
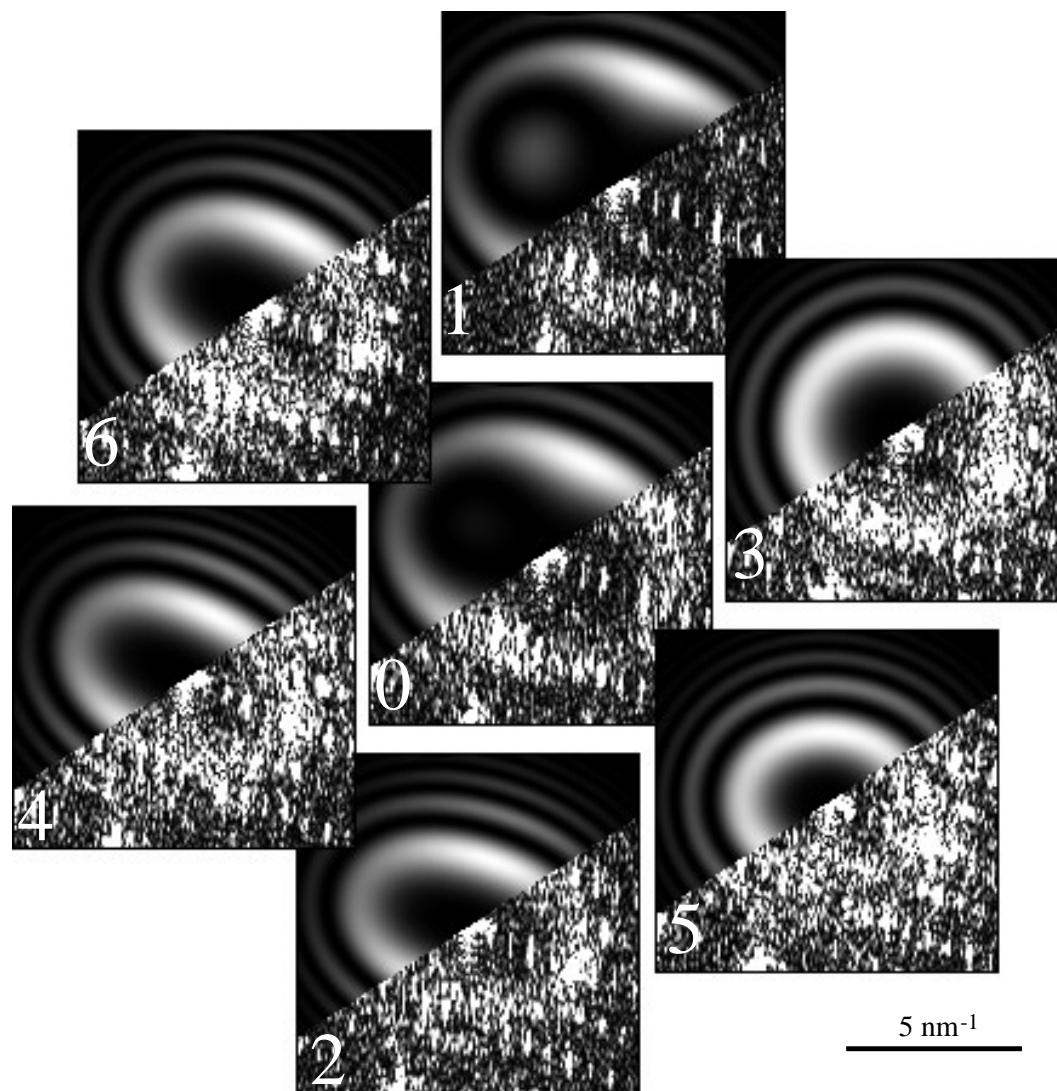
Figure 5.9: Zemlin Tableau of dataset NdSr010, showing the diffractogram calculated from the middle image of the short focal series at each tilt. This specimen was nearly free of contamination. Although the eye is capable of recognising some of the ring structures when guided by the simulated diffractograms, manual fitting would have been virtually impossible for this dataset. The tilt magnitude is 800 DAC units, corresponding to 1.9 mrad. In this dataset the axial coma is large and the coma-free axis lies outside the tableau, close the upper page margin.
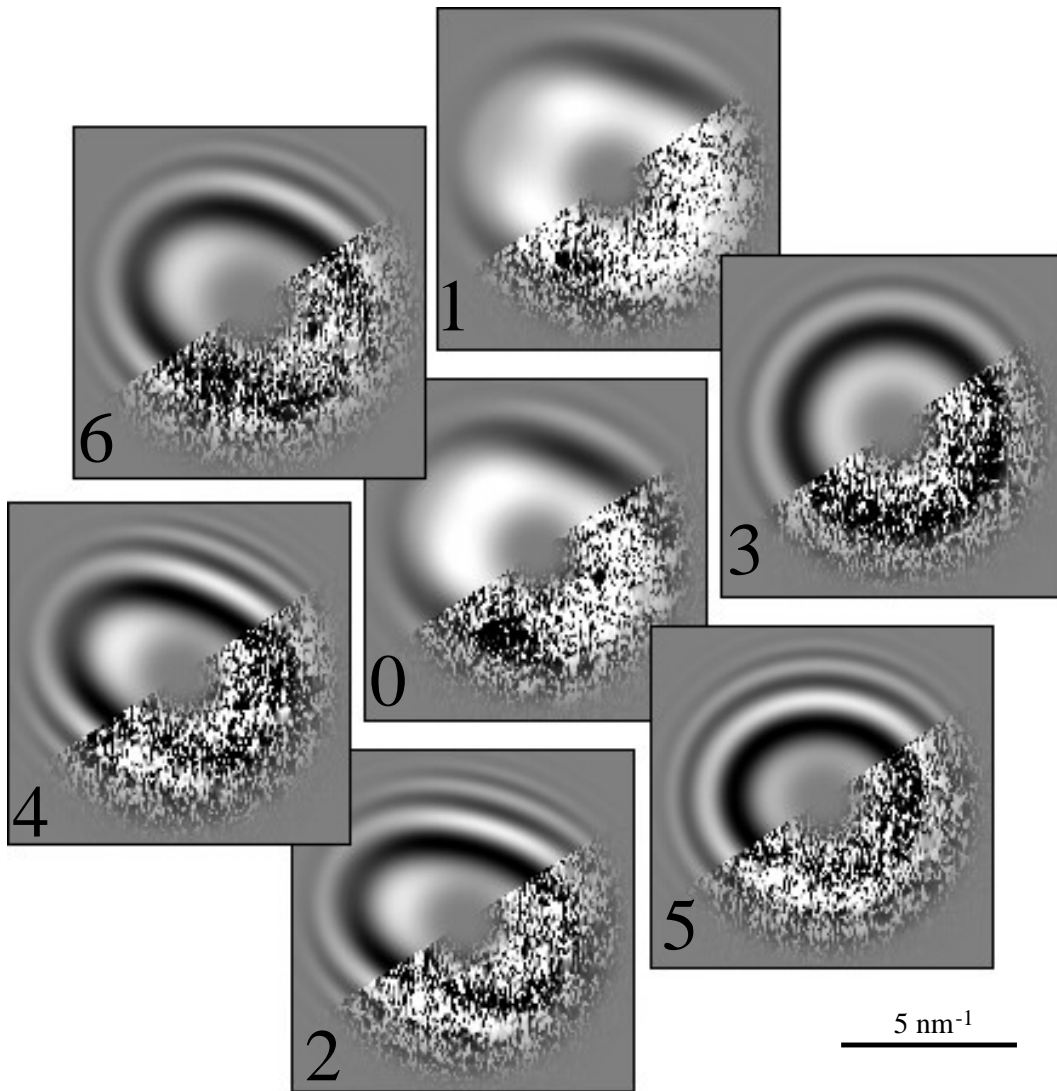
Figure 5.10: Tableau in which the uncorrected image wave $f_{\mathrm{PCI}}$ calculated from a three-member focal series is displayed for each beam tilt of the *tiltf6* dataset NdSr010. The simulated patterns merged with the experimental ones are calculated using the automatically determined aberration coefficients. The tilt magnitude is 800 DAC units, corresponding to 1.9 mrad. Note that the coma-free axis lies outside the tableau, close the upper page margin.

# Chapter 6

# Experimental examples of automated object wave restoration

This chapter details examples of results obtained using the aberration determination and image series restoration methods discussed in the previous chapters. All the restorations presented in this chapter are based on image series recorded on the JEM 3000F microscope in the Department of Materials in Oxford using a $1024 \times 1024$ pixel CCD camera mounted axially with primary microscope magnifications between 300 kX and 600 kX. The calibration of this instrument has been given in chapter 5 and the transfer properties of the CCD camera have also been accurately measured and are given in section 2.6.4 (camera 4).

## 6.1 Restoration from a focus series of the complex oxide $Nb_{16}W_{18}O_{94}$

This section describes the results of the focal series restoration of $Nb_{16}W_{18}O_{94}$ that served as an illustration in the discussion of the new method throughout chapter 4.

TEM samples were prepared by crushing and dispersion of the as-synthesised material in propanol onto a holey carbon film, and individual crystals were oriented

along the [001] direction using selected area diffraction patterns.

The microscope was manually aligned to the coma free axis using on-line diffractograms of the amorphous carbon support film and the two-fold astigmatism was also corrected manually. A focal series of 30 images was recorded of a thin crystal edge at an initially overfocus condition and with a nominal focal increment of 12nm towards underfocus. In addition a final image at the nominal starting defocus was also recorded in order to assess the overall focal drift across the entire data set. Immediately following the acquisition of each focal series a tilt azimuth data set of images of an adjacent area of amorphous carbon was recorded comprising an initial and final axial image together with four images recorded with orthogonal beam tilts of approximately 1 Gl/Sch (7.7 mrad) in order to determine the residual antisymmetric objective lens aberrations. The acquisition of both focal and tilt series was fully automated using a series of Digital Micrograph[1] scripts with the microscope under automatic control via the RS232 interface. All subsequent image processing was carried out using programs written in the Semper image processing language (Saxton et al., 1979).

Figure 6.1 shows a structural model of the complex oxide $Nb_{16}W_{18}O_{94}$ projected along the [001] direction. In this projection, the metal cation lattice comprises a complex sequence of corner and edge sharing pentagonal columns built from $MO_6$ octahedra some of which are filled with a central metal atom and some of which have the central tunnel site vacant (Sleight, 1966). The exact distribution of Nb and W within the various sites has not been unambiguously determined although High Angle Annular Dark Field (HAADF) imaging in a TEM-STEM has provided evidence for a degree of non-site specific cation segregation (Kirkland and Saxton, 2002).

This particular dataset was recorded before the aberration determination methods described in this thesis were fully developed and hence, only a focal series rather than a combined tilt/focus series was collected and it was necessary to determine the antisymmetric aberrations via a conventional fitting of diffractograms taken at four

---

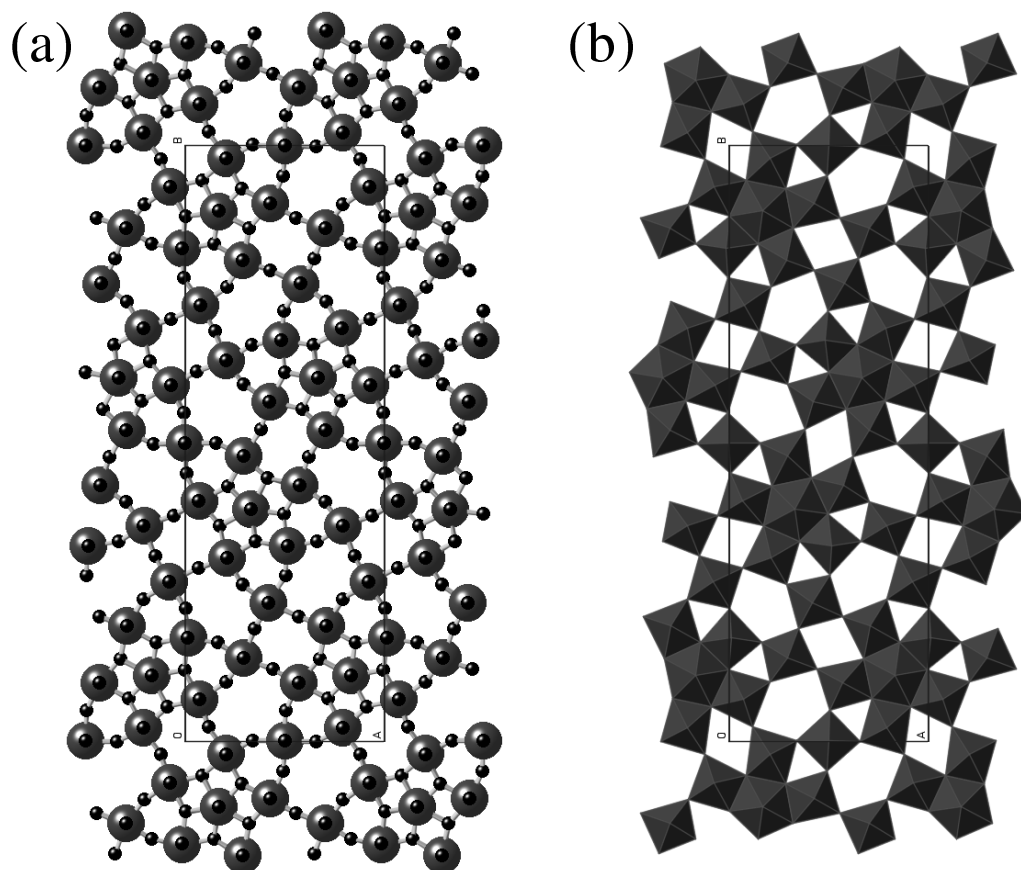[1]Digital Micrograph, Gatan Inc., Pleasanton, CA, USA.

Figure 6.1: Structural Model of $Nb_{16}W_{18}O_{94}$ projected along [001] with the unit cell marked. (a) Ball and stick model with the cations displayed as large gray spheres and the oxygen anions as small black spheres. (b) Polyhedral model showing a complex arrangement of 30 corner sharing $MO_6$ octahedra and 4 $MO_7$ decahedra that share edges with neighbouring octahedra.

184

different tilts from a nearby region of amorphous carbon as discussed in sections 3.3 and 3.3.2. The results of this determination of the antisymmetric aberrations give values of (880,-60) nm for the three-fold astigmatism $A_2$ and (0.5,0.3) mrad for the beam misalignment.

Figures 6.2 and 6.3 show the restoration of the exit plane wavefunction as both modulus (figure 6.2) and phase (figure 6.3), recovered to a transfer limit[2] of $0.91\text{nm}^{-1}$. It is immediately apparent that whilst the basic cation lattice can be determined from the axial image recorded close to the Scherzer defocus (figure 4.1) the restored modulus shows the positions of the cation columns in projection at substantially higher resolution due to the wider transfer in the total WTF compared with the conventional Scherzer limit. Moreover the modulus remains directly interpretable to a greater specimen thickness than the axial image, a feature that has also been noted in earlier restorations and simulations (Kirkland et al., 1995, 1997; Saxton, 1988; Foschepoth and Kohl, 1998). Thus the restored modulus alone provides a convenient route to directly determining the cation lattice in complex materials though in this sample, a distinction between different cations is not possible. For this purpose, the complementary technique of High Angle Annular Dark Field Imaging in the STEM (Pennycook and Boatner, 1988; Pennycook and Jesson, 1990, 1991; Nellist and Pennycook, 2000) provides a higher degree of chemical sensitivity (Kirkland and Saxton, 2002). Within the restored modulus a stacking fault comprising a 1/3 unit cell translation along [010] is clearly visible (figure 6.2).

The restored phase (figure 6.3) shows a more complex contrast than the modulus, changing rapidly with specimen thickness as has been previously observed (Saxton, 1988). In the thinnest areas of the crystal in addition to the strong positive contrast (white, corresponding to a phase advance) located at the cation sites and corresponding directly to the strong negative black contrast in the modulus there is weak positive contrast at positions between the cations. The origin of this weak contrast has been

---

[2]Throughout this chapter the *transfer limit* of a restoration is defined as the spatial frequency where the total wave transfer function falls below 10%.
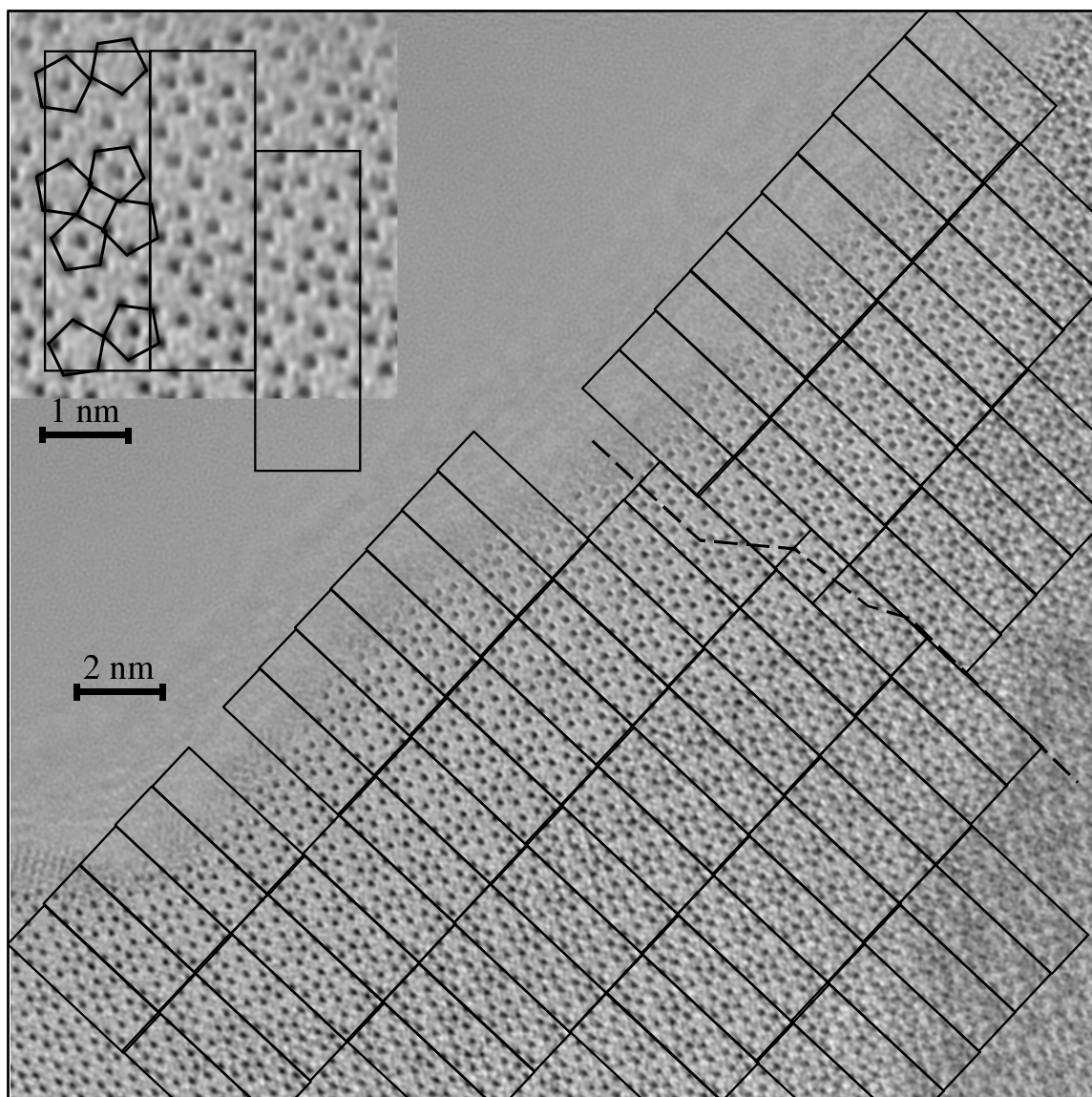
Figure 6.2: Restored modulus of the exit plane wavefunction, which directly shows the cation positions (black) with improved resolution compared to the axial image. Unit cell perimeters are marked in the full image and in the enlarged and rotated inset top left. The dashed line indicates a stacking fault with a shift of a third of a unit cell along [010].
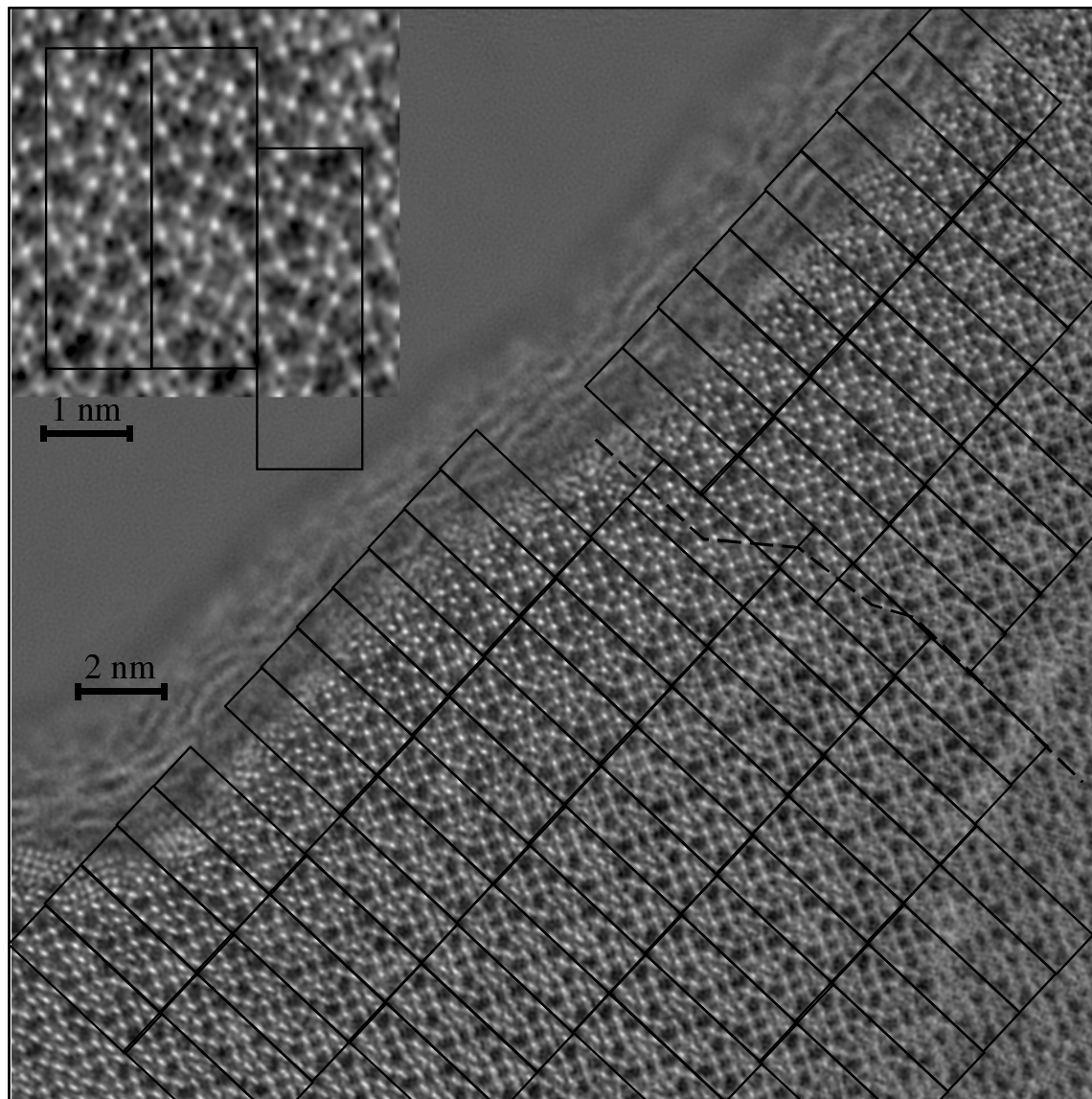
Figure 6.3: Restored phase of the exit plane wavefunction. The appearance of the phase is much more thickness dependent than that of the restored modulus. The cation sites are recovered in strong positive (white) contrast. The inset shows clearly distinctive features present between the cation atomic columns which provide information about the anion sublattice. In the main image, a surface reconstruction close to the crystal edge is also clearly visible.

investigated further by means of multislice image simulations (appendix A) to generate the exit surface wavefunction for various specimen thicknesses, and calculate a simulated restored wave according to appendix A.3. Two structural models were simulated, the first consisting of the bulk structure with atomic coordinates taken from an earlier X-ray determination (Sleight, 1966) in which the structure is fully oxygenated and the second consisting of a lattice fraction in which the entire oxygen sublattice is absent.

The results of these simulations are shown in figure 6.4 for thicknesses of 2 nm and 3.5 nm. It is immediately apparent that in the hypothetical deoxygenated model the weak contrast visible in the experimental phase is absent and moreover could not be reproduced for any specimen thickness considered. However, for the oxygenated model, contrast at the anion sites is clearly present in figure 6.4(b), closely matching that observed experimentally suggesting that some information due to the oxygen anion lattice is recovered in the restoration. Finally it should be noted that the restored phase shows evidence for a surface reconstruction wherein the parent material is locally reduced to cubic $WO_3$ at the surface (Smith and Bursill, 1985).

This example demonstrates the resolution enhancement in a focal series restoration compared to a single axial image. Furthermore, it exemplifies the complementary nature of the information provided by the modulus and phase. While the former shows the cation column positions at high resolution, the latter provides information on the weakly scattering oxygen columns.
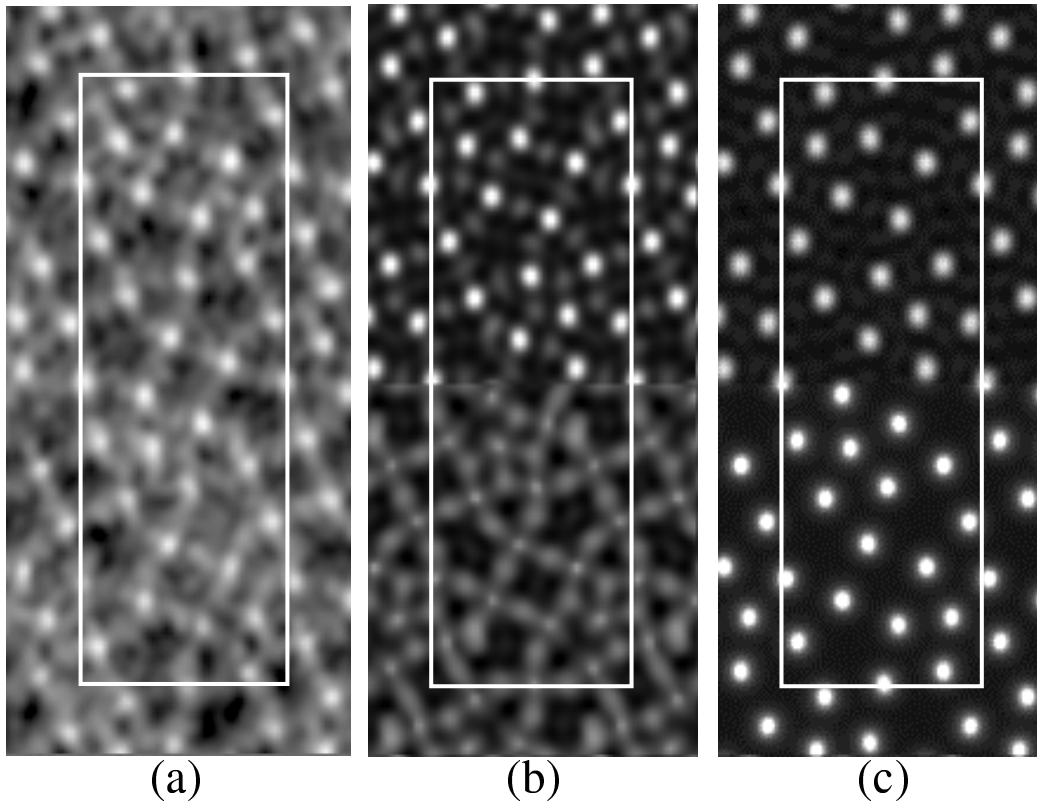
Figure 6.4: (a) Single unit cell section of the experimental restored phase, with the thickness increasing from top to bottom. (b) Phase obtained from a multislice simulation of the complete structure with a thickness of 2 nm (top half) and 3.5 nm (bottom half). (c) As in (b), but for a deoxygenated lattice fraction. Comparing (b) and (c) to the experimental phase shows that contrast between the cation sites due to the oxygen sublattice is clearly resolved.

## 6.2  Focal and tilt series restorations of perovskite-related layered titanates

This section describes the application of both tilt and focal series restorations with automated aberration determination to refine an inorganic perovskite structure. Perovskites $ABO_3$, are known for a large variety of metallic cations $A$, $B$ and often form more complicated layered structures when the oxygen content is increased (Lichtenberg et al., 2001). The composition dependence of these structures allows the engineering of materials with desired properties and therefore a good understanding of the structure/composition relationship is important. Perovskites find applications as dielectric materials[3] and the perovskite $(La, Ba)_2CuO_4$ was the first high $T_c$ superconductor to be discovered (Bednorz and Müller, 1986). Another striking property of some perovskites is an up to 1000 fold increase of the resistivity in strong magnetic fields (colossal magneto resistance) discovered by Jin et al. (1994) for thin films of $La_{0.67}Ca_{0.33}MnO_x$.

In this work, two structures were investigated; the ternary structure $Nd_5Ti_5O_{17}$ and the quaternary structure $Nd_4SrTi_5O_{17}$, both prepared by J. Sloan and first described in (Sloan and Tilley, 1994; Connolly et al., 1996). The effect of the partial substitution of Nd by Sr manifests itself in the diffraction pattern in figure 6.5, where the systematic $(h0l)$ rows with odd $h$ are replaced by continuous oscillating streaks, an effect not present in the ternary structure or any other known perovskite structure (Langley et al., 2002).

In this section, the wavefunction restoration methods developed in this thesis are applied to deduce a structural model, starting from the known structure of the parent perovskite $NdTiO_3$, and in particular to understand the origin of the peculiar diffraction phenomenon displayed by the quaternary phase.

---

[3]One classical example for this is the ferroelectric perovskite $BaTiO_3$. Ferroelectricity up to extremely high temperatures is displayed by some layered perovskites such as $La_2Ti_2O_7$ (Nanamatsu et al., 1974).
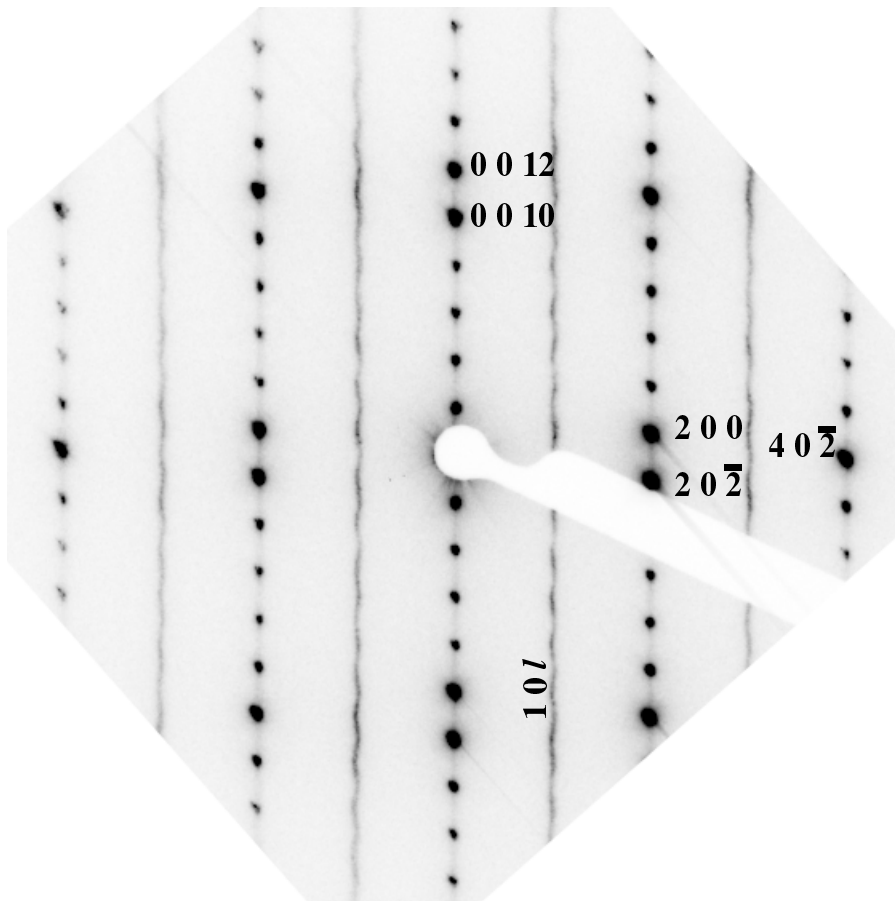
Figure 6.5: Diffraction pattern of the quaternary $Nd_4SrTi_5O_{17}$ in the [010] projection. The systematic rows with odd $h$ are replaced by wavy streaks.

## 6.2.1 The structure of layered perovskites

The titanates investigated belong to the family of $A_nB_nO_{3n+2}$ compounds (Lichtenberg et al., 2001), which are based on the $ABO_3$ perovskite structure. Perovskites can be described as a cubic array of corner-sharing $BO_6$ octahedra with A cations at the cube centres, as shown in figure 6.6(a). When the oxygen content is increased, layered $A_nB_nO_{3n+2}$ structures are formed with layers $n$ octahedra wide.[4] The $A_nB_nO_{3n+2}$ structure type was established by Galy and Carpy (1974) and studied in detail by Nanot et al. (1981). It has some similarity with the more well-known Ruddlesden-Popper phases (Ruddlesden and Popper, 1957, 1958) $A_{m+1}B_mO_{3m+2}$ that are also derived from perovskite but with the stacking direction parallel to the [001] rather than the [011] perovskite direction. Ruddlesden-Popper phases have been observed for many different $B$ cations, whereas $A_nB_nO_{3n+2}$ type structures are only known for $B \in \{\text{Ti}, \text{Nb}, \text{Ta}\}$ and $n \in \{4, 4.\overline{3}, 4.5, 5, 6\}$ (Lichtenberg et al., 2001). The members with fractional $n$ have ordered stacking sequences of $n = 5, 4, 4, 5, 4, 4, \ldots$ and $n = 5, 4, 5, 4, \ldots$, respectively. Figure 6.6(b) illustrates the $n = 5$ structure studied in this work.

## 6.2.2 Experimental data

Figure 6.7 shows the modulus and phase for the quaternary compound in the [100] projection, restored from the *tiltf6* dataset (NdSr100) that was used as example in section 5.3. Both modulus and phase reveal a 'checkerboard' pattern of stronger scattering neodymium columns and weaker scattering titanium columns, which also contain oxygen (*c.f.* figure 6.6(b)). However, neither in the modulus nor in the phase is this contrast difference between these types of columns apparent at all thicknesses. The restoration also reveals that within each perovskite layer the atomic columns are slanted by a small angle away from the [001] direction in opposite directions for neighbouring slabs.

---

[4]Note that $n = \infty$ corresponds to the parent perovskite $ABO_3$ and that $n$ decreases as the oxygen content is increased.

(a) $ABO_3$ perovskite　　(b) $A_5B_5O_{17}$ [100]　　(c) $A_5B_5O_{17}$ [010]

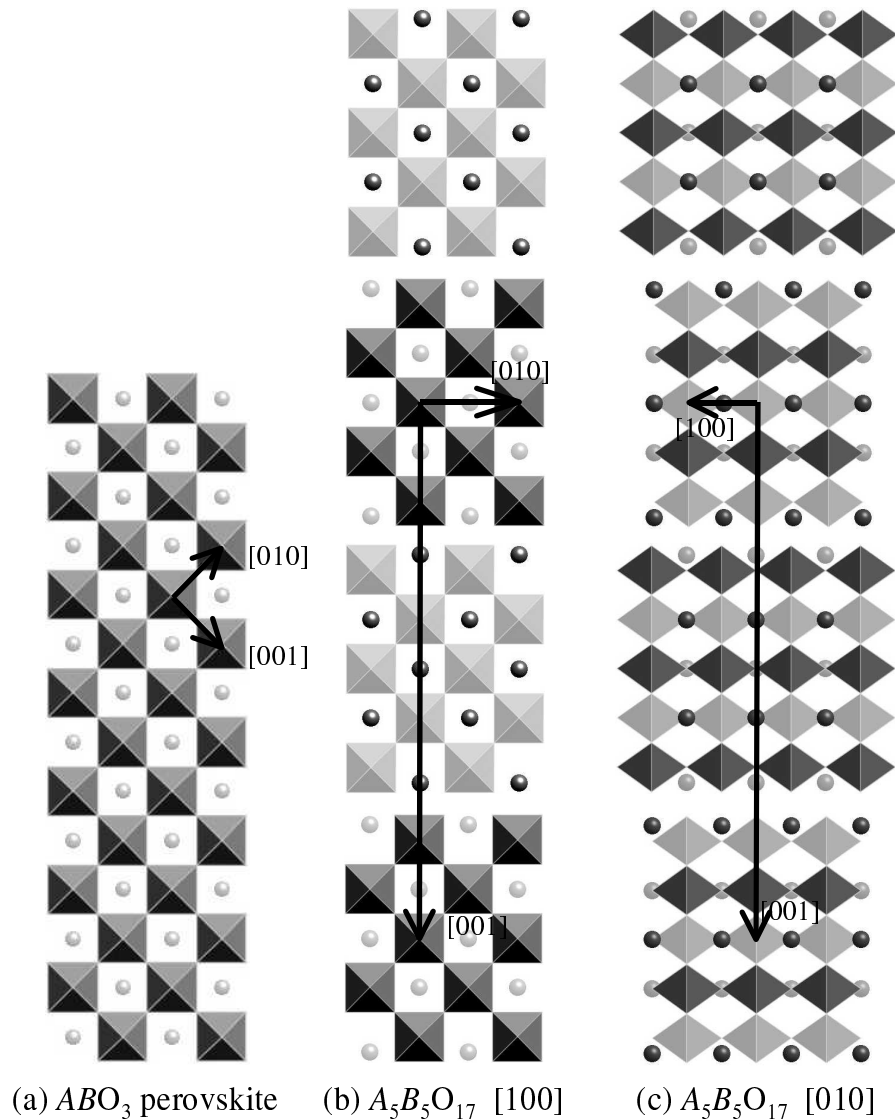Figure 6.6: Simplified structural models of $A_nB_nO_{3n+2}$ compounds represented by corner-sharing $BO_6$ octahedra and isolated $A$ cations. Octahedra and cations drawn with light grey shading are half a lattice plane below those with darker shading. (a) The perovskite $ABO_3$. (b) and (c) The $n = 5$ structure $A_5B_5O_{17}$ with layers 5 octahedra wide, projected in the [100] and [010] directions.
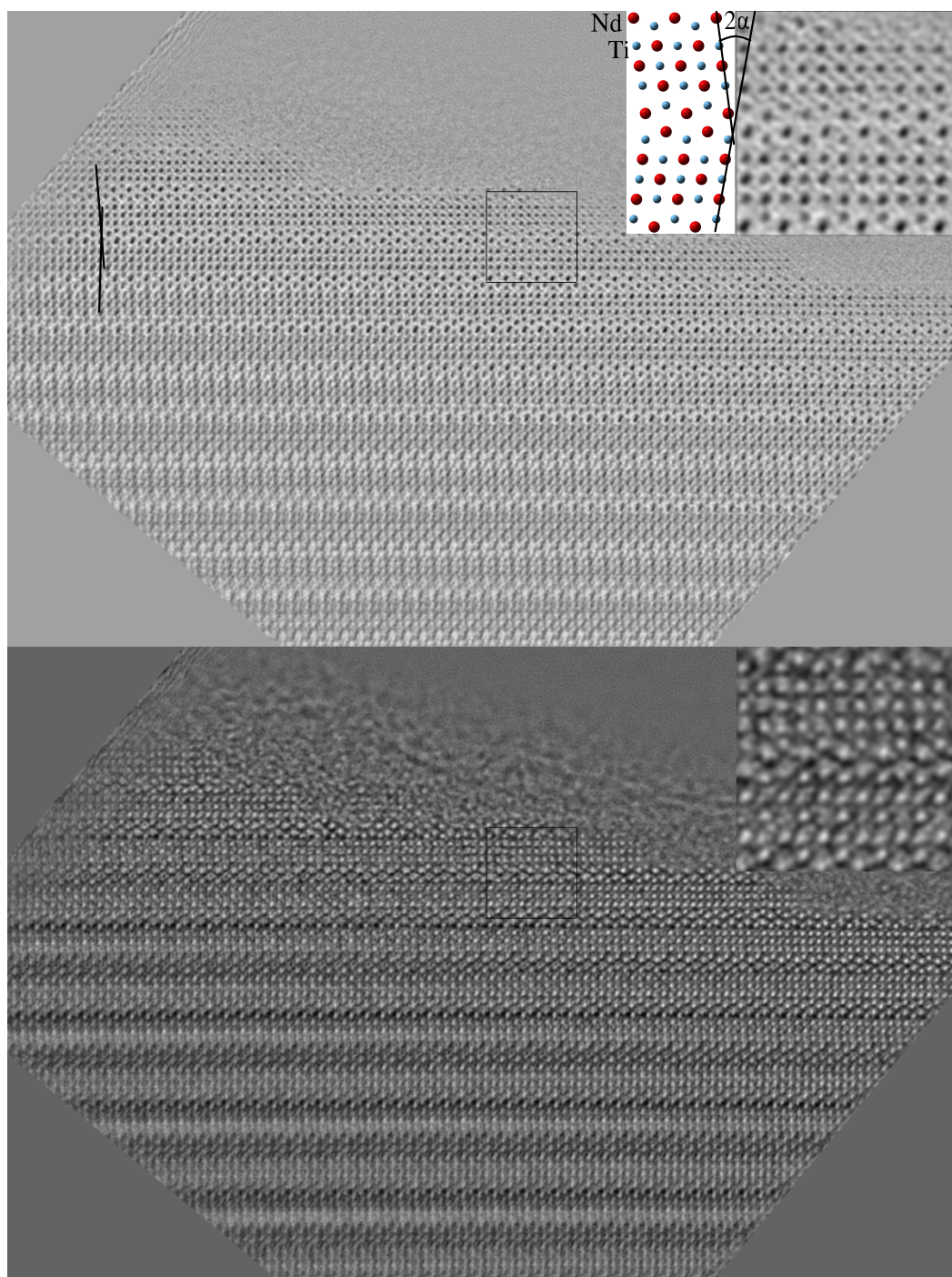
Figure 6.7: Restored modulus and phase for the quaternary compound in [100] direction. The images have been rotated to show the perovskite slabs in a horizontal direction to simplify the comparison with the structural models in figures 6.6 and 6.13. The inset shows an enlargement of the marked region together with a structural model where the oxygen is omitted and the slanting angle $\alpha$ is exaggerated. In the enlarged region, the difference between Nd and Ti columns is visible in both modulus and phase at certain thicknesses.

Figure 6.8: Square root of the power spectrum of the complex restored exit wave for the quaternary compound in the [100] projection. The transfer limit of this tilt series restoration is $9.2\,\mathrm{nm}^{-1}$.
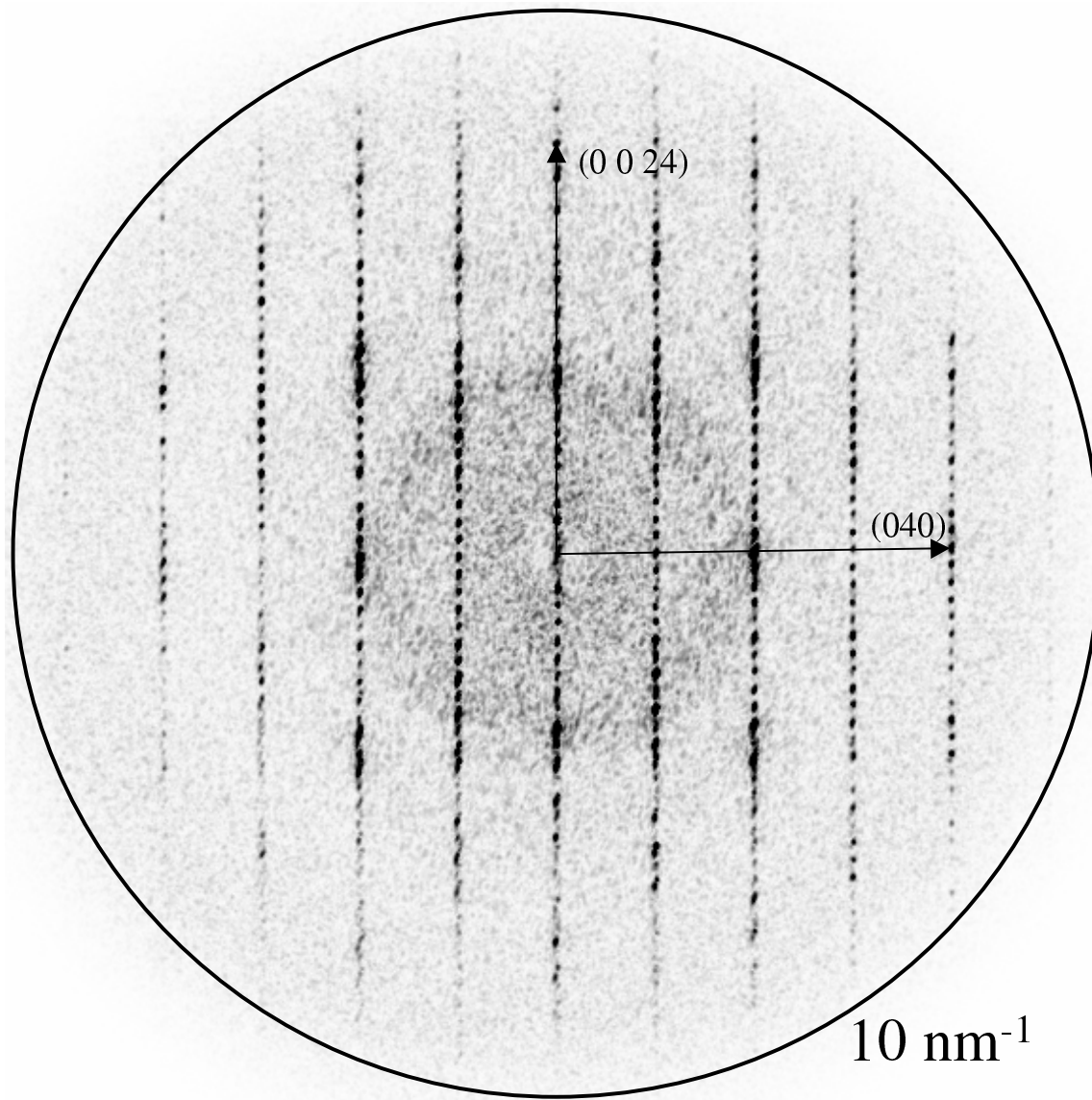
This slant angle can be measured accurately by extracting a single slab centred at an atomic site and then determining the angle by which the slab has to be slanted to maximise the cross-correlation coefficient between the resulting image and its mirror image. Such an analysis yields a slant angle of $2.5° \pm 0.5°$ for both the ternary and the quaternary phases.

The transfer limit for this restoration is $9.2\,\text{nm}^{-1}$, and the diffraction pattern calculated from the restored wave (figure 6.8) shows that reflections up to this limit are included in the restored wave, demonstrating the resolution enhancement that can be achieved using tilt series restorations. From this diffraction pattern the lattice constant $b = 0.59\,\text{nm}$ and $d_{001} = c\cos\beta = 31.3\,\text{nm}$ were measured. In the [100] projection, both diffraction patterns and images are indiscernibly similar for the ternary and quaternary compounds, therefore no example for the ternary material is shown.

Figure 6.9 shows the modulus of the Fourier transform for two [010] exit wave restorations. The restoration for the ternary phase was obtained from a 30-member focal series and that for the quaternary phase was obtained from the *tiltf6* dataset NdSr010 that was used as example in section 5.3. The transfer limit of $6.7\,\text{nm}^{-1}$ is lower than in figure 6.8, mainly because of the large initial mistilt of 3.4 mrad (table 5.7), which also causes an asymmetry in the total transfer function.

The difference between the ternary and quaternary structure can be clearly seen in this projection since reflections in systematic rows with odd $h$ are replaced by a straight streaked line for the quaternary compound. However, these streaks do not oscillate as in the electron optical diffraction pattern (figure 6.5).

The diffraction pattern can be consistently indexed assuming a monoclinic unit cell with $a = 0.78$ *nm* $c = 3.16$ *nm* $\beta = 97.1°$, in agreement with the findings of Williams et al. (1991) for the analogous compound, $La_5Ti_5O_{17}$. The lattice constant $a$ corresponds to twice the Ti-Ti distance in the simplified model in figure 6.6(c). However, this reduction in symmetry only manifests itself in the weak reflections with odd $h$. The space group was determined by Langley et al. (2002) as $P2_1/c$ (14 in the international tables).
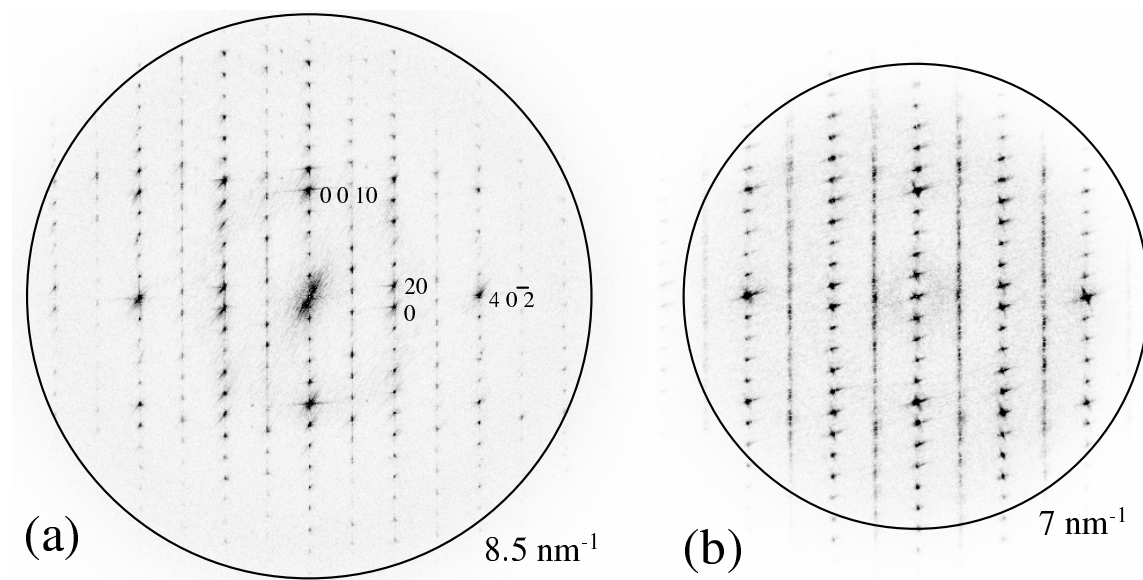
Figure 6.9: Modulus of the Fourier transform of restored exit waves. (a) Ternary compound in the [010] projection, where the exit wave is restored from a 30 member focal series with a transfer limit of $8.3\,\mathrm{nm}^{-1}$. (b) Quaternary compound in the [010] projection, where the exit wave is restored from a *tiltf6* series. The transfer limit in this particular *tiltf6* series is only $6.7\,\mathrm{nm}^{-1}$ due to a large unintentional initial beam tilt, which also makes the total wave transfer function asymmetric.

Figure 6.10 shows the restoration for the quaternary material in the [010] projection using a *tiltf6* dataset (NdSr010), demonstrating that with accurate aberration measurement and correction, a high quality restoration can be achieved even when a large initial beam misalignment is present.

The cation sites, which are characterised by minima in the modulus and maxima in the phase, are shown annotated in figure 6.11. Where the crystal is very thin (bottom left corner), the Nd sites are easily identified as their contrast is much higher than that of the Ti sites. The difference in the $z$-coordinate between the Nd4 and Nd5 sites is a manifestation of the doubling of the $a$ lattice constant compared to the simplified model in figure 6.6(c).

### 6.2.3  Structural model

The simplified structural models in figure 6.6 assume that in the $BO_6$ octahedra are aligned with the perovskite unit cell axes. However this is generally not the case and the octahedra are rotated to increase the packing density with a rotation angle that depends on the ionic radii of the cations (Thomas, 1996). The structural model developed here is based on the bulk $NdTiO_3$ X-ray structure determined by Amow and Greedan (1996) with the atomic positions given in table 6.1, and shown in figure 6.12. Tilting of the octahedra strongly reduces the symmetry of the perovskite. Since neighbouring octahedra are tilted in opposite directions, translational symmetry is reduced and the new unit cell comprises 4 formula units and has base vectors [200], [011] and [01$\bar{1}$] in terms of the original cubic cell. A distortion reduces the resulting tetragonal symmetry to orthorhombic ($b \neq c$ as determined for the X-ray structure). The doubling of the $x$ base vector agrees with the observed diffraction pattern for the layered structure and it is also likely that the concerted rotation is responsible for the difference in $z$-position between the Nd4 and Nd5 sites.

This bulk $NdTiO_3$ structure is used as a starting point for a model of the layered compounds. The fractional coordinates of a model consisting of undistorted slabs 5 octahedra wide of the bulk structure are listed in table 6.2. These were calculated
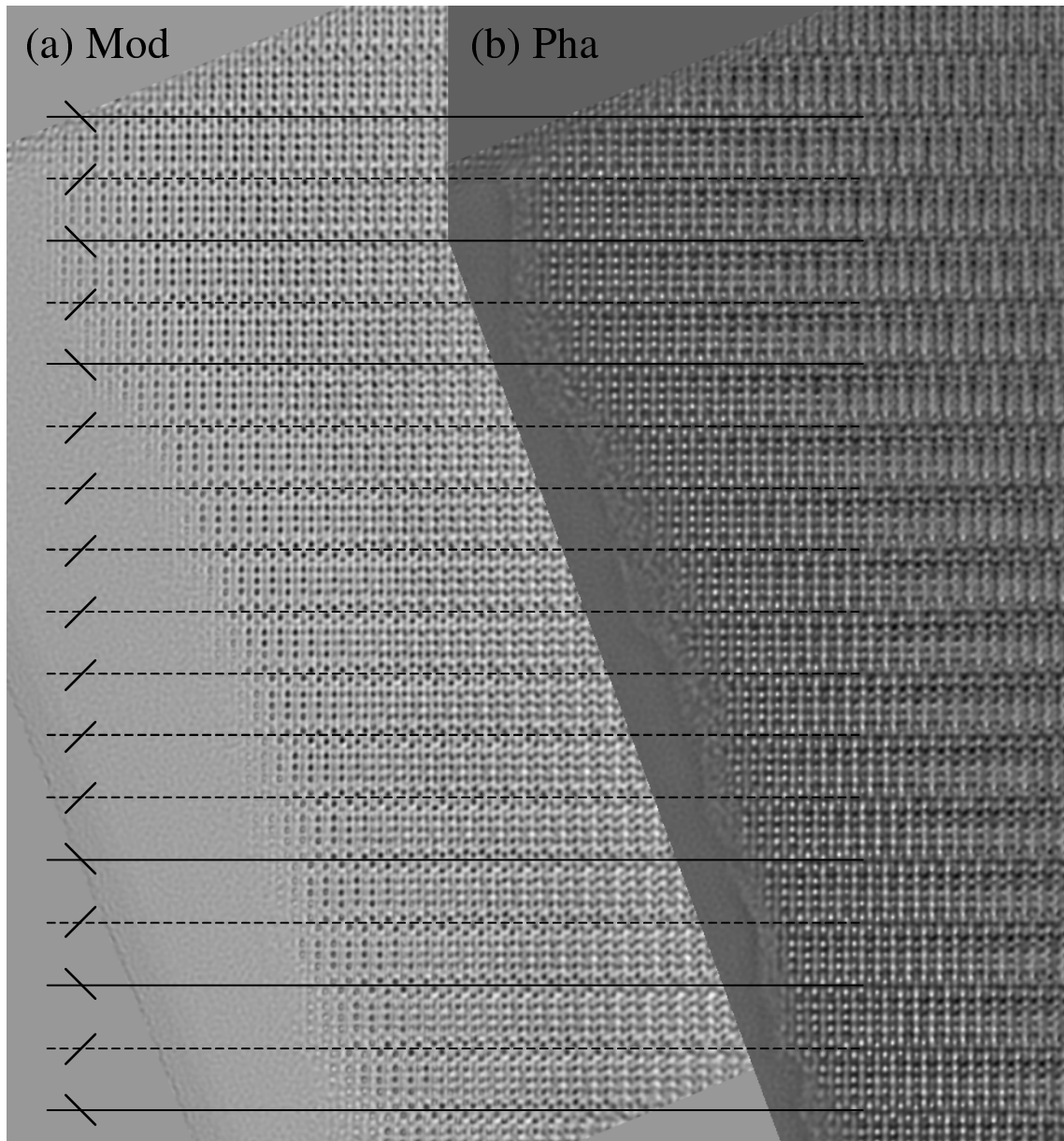
Figure 6.10: Modulus and phase for the quaternary compound in the [010] projection, restored from the *tiltf6* dataset NdSr010. The horizontal lines separate the perovskite slabs. The line style relates to the stacking order which is explained in section 6.2.4.
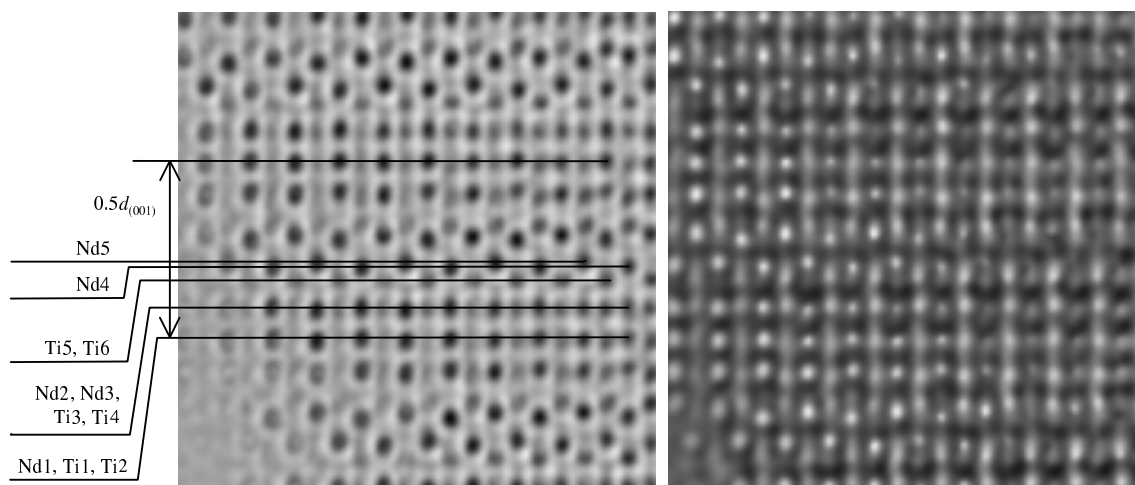
Figure 6.11: Detail of the restored modulus and phase for the quaternary compound in the [010] projection.

| a=7.785(2), $b$=5.647(1), $c$=5.519(1) | | | | |
|---|---|---|---|---|
| $\alpha = \beta = \gamma = 90°$ | | | | |
| Spacegroup: P n m a (62) | | | | |
| Atom | Ox | $x$ | $y$ | $z$ |
| Nd | +3 | 0.250 | 0.0598(8) | 0.9927(13) |
| Ti | +3 | 0.0 | 0.5 | 0.0 |
| O1 | -2 | 0.250 | 0.4770(11) | 0.0900(13) |
| O2 | -2 | 0.0501(6) | 0.3005(8) | 0.6992(7) |

Table 6.1: Structure of the bulk perovskite $NdTiO_3$, determined using X-ray diffraction by Amow and Greedan (1996). Compared to that reference, the $a$ and $b$ axis have here been interchanged for consistency with the usual definition for the layered structure.
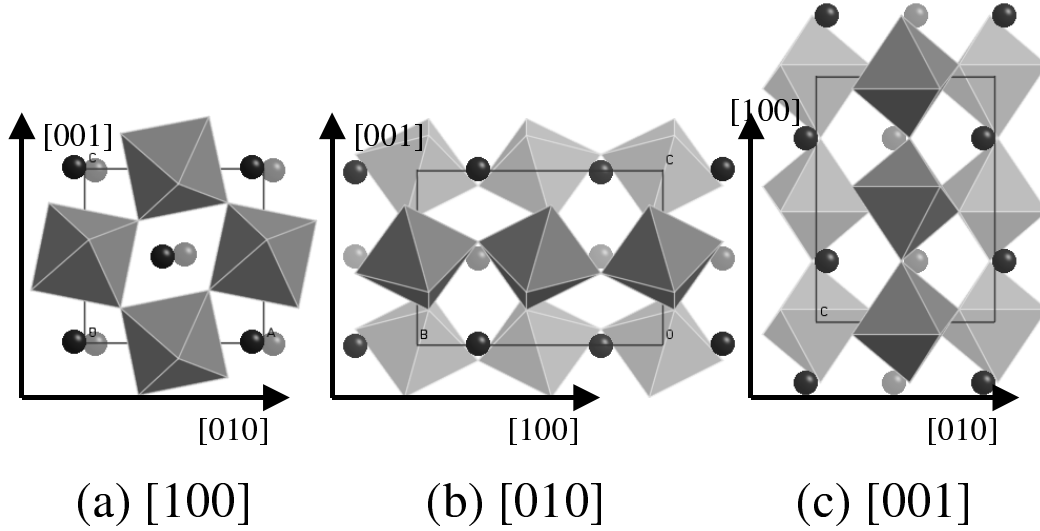
Figure 6.12: Structural model of bulk $NdTiO_3$ according to Amow and Greedan (1996), clearly showing the concerted rotation of the $NdTiO_3$ octahedra and its effect on the Nd cation positions.

using the bulk atomic fractional coordinates within the volume $[0, 1[\times[0, 1[\times[0, 1.36[$ and expressing the positions in terms of the monoclinic unit cell vectors using the transformation

$$x' = x + z'/2 \tag{6.1}$$

$$y' = y \tag{6.2}$$

$$z' = zc/(c' \cos \beta') , \tag{6.3}$$

giving the positions of the sites within a half slab. The positions in the other half are generated by the $P_2 1/c$ symmetry operator $(x, y, z) \rightarrow (-x, -y, -z)$ and the symmetry operator $(x, y, z) \rightarrow (x, 0.5 - y, 0.5 + z)$ generates the second slab in the unit cell.

This model is generally consistent with the experimental data except in two areas:

1. The slanting observed in the [100] projection is not taken into account.

2. The experimental data show that the Nd cations closest to the layer interface are displaced towards the interface and that the displacement is different for the Nd4 and Nd5 sites.

. This basic model can therefore be modified to give the refined model in table 6.2, which takes the above into account by

1. A further transformation $y'' = y' + z'(c' \cos \beta'/b') \tan \alpha_s$ where $\alpha_s = 2.5°$ is the measured slant angle.

2. Substitution of the $z$-coordinates for Nd4 and Nd5 by the values measured from the restored modulus.

All other coordinates are in good agreement with the experimental data and are therefore not modified. The refined model is shown in figure 6.13(a) and (b), and figure 6.13(c) illustrates an alternative perovskite slab stacking sequence, which is only realised in the quaternary structure and which is discussed in detail in the next section.

## 6.2.4   The stacking order of the perovskite slabs.

The model in figure 6.13 illustrates that there are two alternative positions for the stacking of a perovskite slab onto the previous one and that they differ by an $a/2$ shift. This allows for a variety of stacking sequences apart from the regular monoclinic structure shown by the ternary material. Due to the shift in the Nd positions closest to the boundary induced by the octahedra tilting, it is possible to study the stacking sequence by investigating the cation positions in the restored modulus or phase. In the monoclinic ternary structure, the arrows linking the closest Nd sites across the layer boundary are parallel for all boundaries (figure 6.13(b)). An alternative arrangement where these arrows point in different directions in alternate layer boundaries leads to the structure illustrated in figure 6.13(c) with an orthorhombic unit cell. However, a completely random alternation of the two arrow directions is also possible.

| $a$=7.785(2), $b$=5.647(1), $c$=5.519(1) | | | | | |
|---|---|---|---|---|---|
| $\alpha = \gamma = 90°$, $\beta = 97.1°$ | | | | | |
| Spacegroup: P2$_1$/c (14). | | | | | |
| Atom | Bulk structure (1) | | | Refined structure (2) | | |
| | $x$ | $y$ | $z$ | $x$ | $y$ | $z$ |
| Nd1 | 0.2506 | 0.5598 | 0.0013 | 0.250 | 0.5601 | 0.0013 |
| Nd2 | 0.2930 | 0.0598 | 0.0861 | 0.2930 | 0.0813 | 0.0861 |
| Nd3 | 0.7943 | 0.9402 | 0.0886 | 0.7943 | 0.9623 | 0.0886 |
| Nd4 | 0.8367 | 0.4402 | 0.1734 | 0.8367 | 0.4835 | 0.2160 |
| Nd5 | 0.3380 | 0.5598 | 0.1759 | 0.3380 | 0.6038 | 0.2030 |
| O1 | 0.7579 | 0.0230 | 0.0157 | 0.7579 | 0.0269 | 0.0157 |
| O2 | 0.7858 | 0.5230 | 0.0716 | 0.7858 | 0.5409 | 0.0716 |
| O3 | 0.3015 | 0.4770 | 0.1030 | 0.3015 | 0.5028 | 0.1030 |
| O4 | 0.3295 | 0.9770 | 0.1589 | 0.3295 | 0.0167 | 0.1589 |
| O5 | 0.8452 | 0.0230 | 0.1904 | 0.8452 | 0.0706 | 0.1904 |
| O6 | 0.0675 | 0.3005 | 0.0348 | 0.0675 | 0.3092 | 0.0348 |
| O7 | 0.4673 | 0.3005 | 0.0348 | 0.4673 | 0.3092 | 0.0348 |
| O8 | 0.4762 | 0.8005 | 0.0525 | 0.4762 | 0.8136 | 0.0525 |
| O9 | 0.0764 | 0.8005 | 0.0525 | 0.0764 | 0.8136 | 0.0525 |
| O10 | 0.6112 | 0.1995 | 0.1221 | 0.6112 | 0.2300 | 0.1221 |
| O11 | 0.0110 | 0.1995 | 0.1221 | 0.0110 | 0.2300 | 0.1221 |
| O12 | 0.0198 | 0.6995 | 0.1399 | 0.0198 | 0.7345 | 0.1399 |
| O13 | 0.6200 | 0.6995 | 0.1399 | 0.6200 | 0.7345 | 0.1399 |
| O14 | 0.1548 | 0.3005 | 0.2094 | 0.1548 | 0.3528 | 0.2094 |
| O15 | 0.5546 | 0.3005 | 0.2094 | 0.5546 | 0.3528 | 0.2094 |
| O16 | 0.5635 | 0.8005 | 0.2272 | 0.5635 | 0.8573 | 0.2272 |
| O17 | 0.1637 | 0.8005 | 0.2272 | 0.1637 | 0.8573 | 0.2272 |
| Ti1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Ti2 | 0.5000 | 0.0000 | 0.0000 | 0.5000 | 0.0000 | 0.0000 |
| Ti3 | 0.0437 | 0.5000 | 0.0873 | 0.0437 | 0.5218 | 0.0873 |
| Ti4 | 0.5437 | 0.5000 | 0.0873 | 0.5437 | 0.5218 | 0.0873 |
| Ti5 | 0.0873 | 0.0000 | 0.1747 | 0.0873 | 0.0436 | 0.1747 |
| Ti6 | 0.5873 | 0.0000 | 0.1747 | 0.5873 | 0.0436 | 0.1747 |

Table 6.2: Structure of Nd$_5$Ti$_5$O$_{17}$. Model (1) consists of slabs 5 octahedra wide with the undistorted bulk perovskite NdTiO$_3$ structure. Model (2) has been refined taking into account the 2.5° slant observed in the [100] projection and substituting the $z$ positions of sites Nd4 and Nd5 with the values measured from the restored modulus.
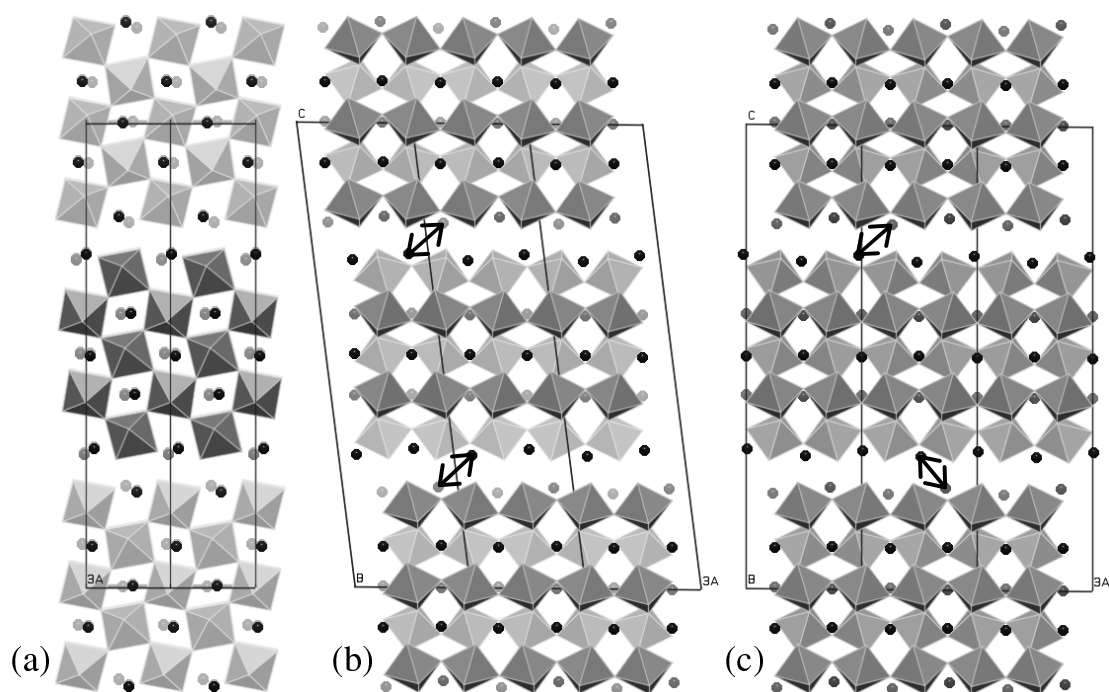
Figure 6.13: Refined model for $Nd_5Ti_5O_{17}$ based on slabs of the bulk $NdTiO_3$ structure. (a)[100] projection. (b)[010] projection. (c) An alternative stacking sequence where the bottom perovskite slab is shifted by $a/2$, leading to an orthorhombic unit cell. The closest Nd sites across the layer boundary are marked with arrows, and these point in the same direction in (b) but in alternate directions in the two boundaries in (c).

Using the cation positions near the layer boundaries, the stacking sequence has been studied for the quaternary compound using the restored wavefunction shown in figure 6.10. For each layer boundary, the direction linking the Nd sites closest to the boundary is marked. In the top and bottom regions, the two stacking directions alternate, while a block of 8 layers at the centre is stacked in only one direction as in the ternary phase. A diffraction pattern calculated from this central block only shows the same separated maxima in the odd $h$ systematic rows as the ternary diffraction pattern in figure 6.9(a). Figure 6.14 shows diffraction patterns calculated with the multislice method using the two regular models of figures 6.13(b) and (c). While the reflections with even $h$ are identical, the reflections with odd $h$ appear at different positions for the two models. For a sufficiently large region of a completely random stacking sequence, these reflections would therefore form structureless straight streaks.

More experimental data is required to investigate the origin of the wave-like oscillations in the streaks observed in the electron diffraction patterns, as such oscillations are not predicted by a random disposition of the stacking sequence alone and are also not observed in the Fourier transform of the present restoration. This is probably because the field of view is only 28 nm wide, while a selected area diffraction pattern covers a far larger specimen area. The form of the wavy diffraction pattern suggests that the $a$ repeat distance is slightly altered as a function of the local stacking order, with associated distortions at the layer boundaries. In order to investigate this, it will be necessary to restore wavefunctions from larger fields of view at lower magnification to give an area comparable to that sampled in the selected area diffraction pattern.

## 6.2.5   Multislice simulations

In order to validate the structural model established in the previous sections, multislice simulations (appendix A) were carried out and the simulated restored wavefunctions (appendix A.3) calculated from the exit surface wavefunctions were compared with the experimental restored waves.
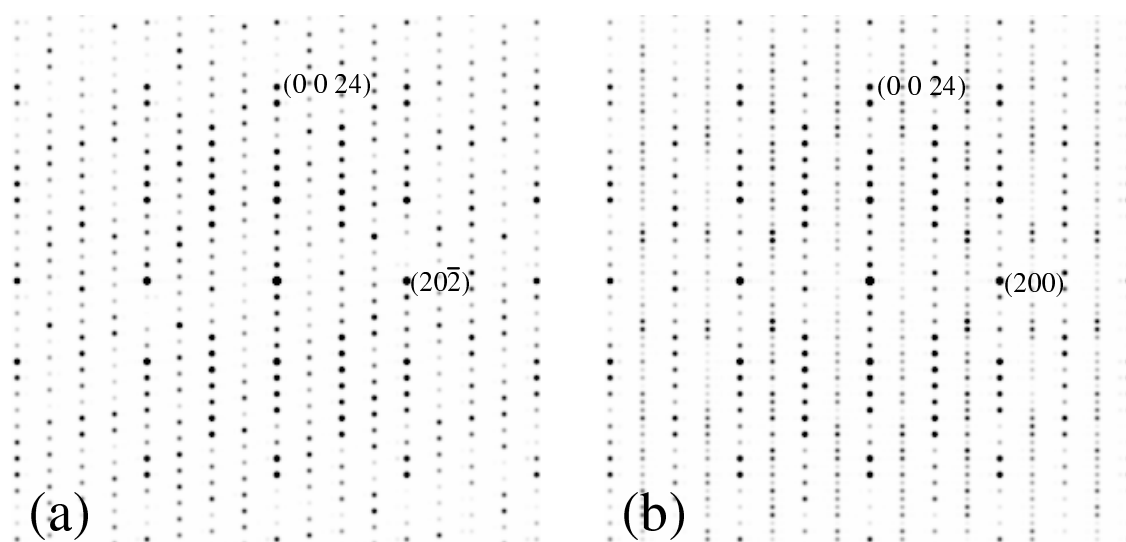
Figure 6.14: Simulated diffraction patterns calculated using the multislice method for two alternative models in the [010] projection. (a) For the model in figure 6.13(b), where the stacking direction is the same in all layer boundaries. (b) For the model in figure 6.13(c) with alternating stacking directions and an orthorhombic rather than monoclinic unit cell. The positions of the reflections with odd $h$ differ between the two models, and a completely random stacking sequence would yield uniform streaks in the systematic rows with odd $h$. The reflections with even $h$, however, occur at identical positions for both models, although they are indexed differently due to the different unit cells.
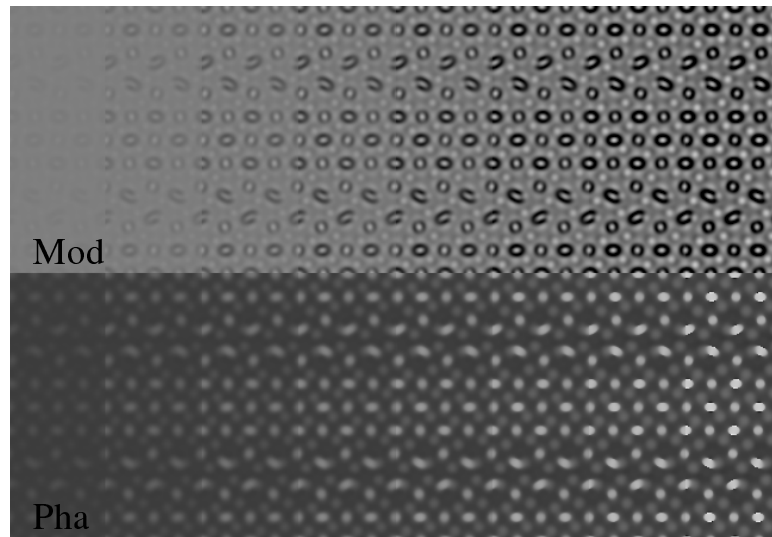
(a) [100]
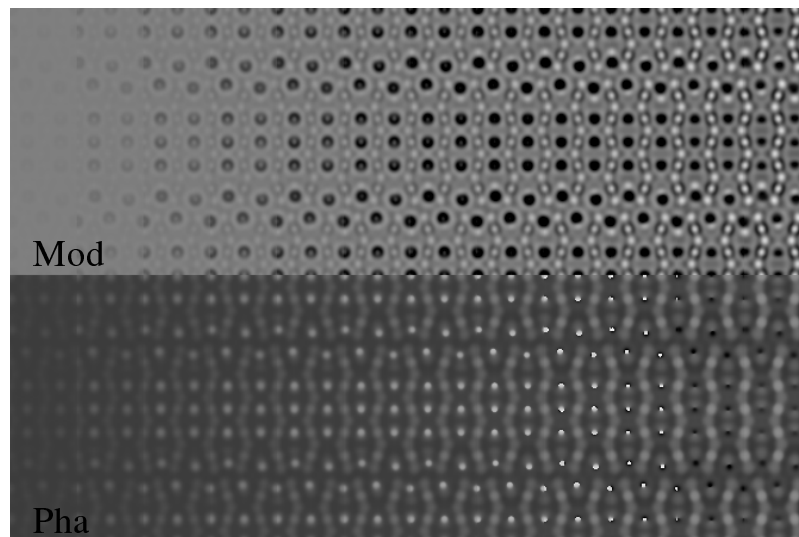


(b) [010]

Figure 6.15: Modulus (top) and phase (bottom) of the simulated restored wavefunction calculated from the model shown in figures 6.13(a) and (b) using the multislice method. (a) [100] projection, where the thickness increases from 1 to 12 slices (0.57 to 6.84 nm) left to right. (b) [010] projection with thicknesses from 1 to 8 slices (0.78 to 6.24 nm).

In the [100] projection, a good qualitative match is achieved, and the similarity of the Nd and the Ti and O columns is confirmed, making a distinction between them only possible at certain thicknesses.

In the [010] projection, the observed and predicted contrast are in good agreement for the Nd sites. For the Ti and O sites, however, the simulations predict positive contrast in the modulus, which is not observed. Furthermore, the simulated phase contrast due to the O columns is nearly as strong as for the Ti columns, such that the effect of the octahedra tilting is clearly visible as a alternating lateral displacement of the O sites. This is not observed in the experimental phase image, where the presence of the oxygen columns gives rise to a nearly straight streaking between the Ti columns in the restored phase. Overall therefore, these simulations suggest that further refinement of the structural model is necessary, particularly with regard to the anion positions.

# 6.3 Carbon nanotubes filled with inorganic crystals

This section describes focal series restorations for inorganic crystals encapsulated in Single-Walled carbon Nanotubes (SWNTs). These novel materials provide interesting specimens for a quantitative comparison with simulations, since their very small thickness is precisely known and they contain a number of different elements. The application of the automated restoration methods to these specimen was particularly challenging for several reasons:

1. The tubes are virtually free of contamination and disorder, making the focus determination difficult.

2. A substantial movement of the tubes under the influence of the electron beam could be observed between the exposures. This movement included bending and rotation, hence the registration is only valid locally, and even small regions of the tube have to be aligned rotationally as well as translationally.

3. Some of the investigated tubes were inclined with respect to the image plane so that the absolute focus varies along the tube.

Following a brief review of carbon nanotubes, this section describes solutions to these difficulties, enabling an exhaustive characterisation of both nanotube and inclusion crystals.

## 6.3.1 Carbon nanotubes

Carbon nanotubes were first discovered by Iijima (1991) in the carbon soot produced on the negative electrode by an arc-discharge evaporation method similar to that used for fullerene synthesis. The discovery triggered an explosion of active research, fuelled by the prospect of unique mechanical and electronic properties of these materials,

namely phenomenal tensile strength and electronic properties ranging from metallic to semiconducting (Saito et al., 1998).

However, the early arc-synthesis experiments provided only multi-walled nanotubes, with 2-50 (Iijima, 1991) concentric cylinders of rolled-up graphitic layers, while theoretical studies predicted semiconducting behaviour only for single-walled tubes. The first SWNTs were discovered by Iijima and Ichihashi (1993) and subsequently, techniques to synthesise these have been refined to yield larger quantities (Journet et al., 1997; Journet and Bernier, 1998). The nanotubes characterised in this work were produced using the modified electric arc technique developed by Journet et al. (1997).

**The chiral vector**

Apart from the hemispherical caps at both ends, a single-walled nanotube can be defined as a strip of graphene sheet with an interatomic distance $d_{C-C} = 0.144\,\text{nm}$ and lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$ mapped onto a cylindrical surface.[5]

The **chiral vector**

$$\mathbf{C}_h = n\mathbf{a}_1 + m\mathbf{a}_2 \equiv (n, m), \tag{6.4}$$

connects two lattice sites that coincide when the planar graphene strip is rolled into a tube as shown in figures 6.16 and 6.17(a)-(c).

The integers $n$ and $m$ uniquely define the structure of the nanotube and, due to the hexagonal symmetry, they can be chosen to satisfy the conventions $n > 0$ and $-n/2 < m \leq n$.[6] The chiral angle is defined as the angle between $\mathbf{C}_h$ and $\mathbf{a}_1$, given by

$$\alpha_c = \arctan\left(\frac{\sqrt{3}m}{2n + m}\right). \tag{6.5}$$

---

[5]The interatomic distance in nanotubes is found to be slightly larger than in graphite, where $d_{C-C} = 0.142nm$ (Saito et al., 1998).

[6]The lower limit for m differs from the definition given in Saito et al. (1998), but ensures that the chiral angle is in the range $-30° < \alpha_c \leq 30°$.
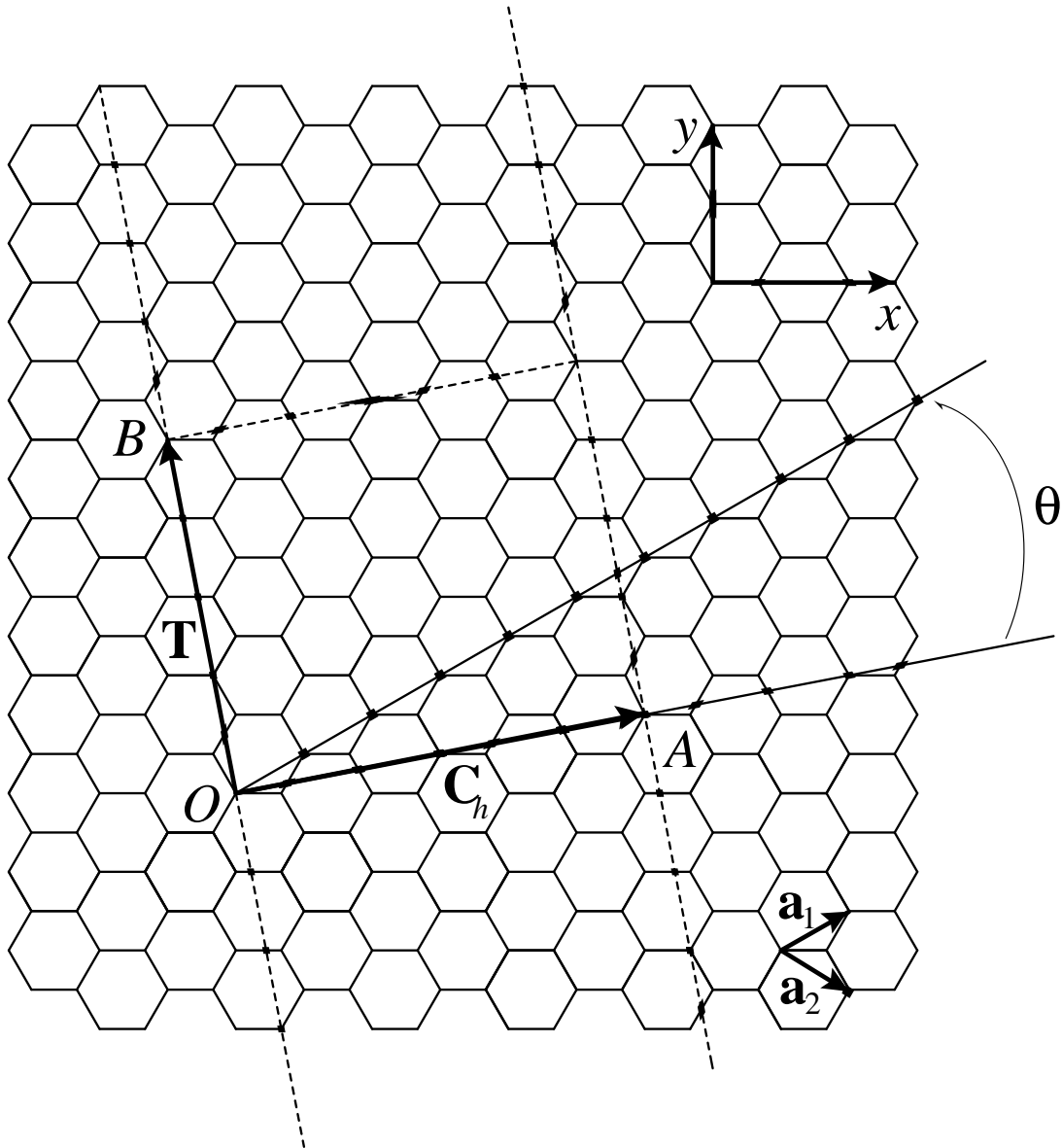
Figure 6.16: Definition of the chiral vector $\mathbf{C}_h$. A nanotube can be constructed when the strip between the lines perpendicular to $\mathbf{C}_h$ through its endpoints is rolled into a cylinder such that these endpoints coincide. The translational vector $\mathbf{T}$ is the shortest lattice vector perpendicular on $\mathbf{C}_h$. These two vectors define the unit cell of the nanotube. The figure corresponds to $\mathbf{C}_h = (4, 2)$ and $\mathbf{T} = (4, -5)$
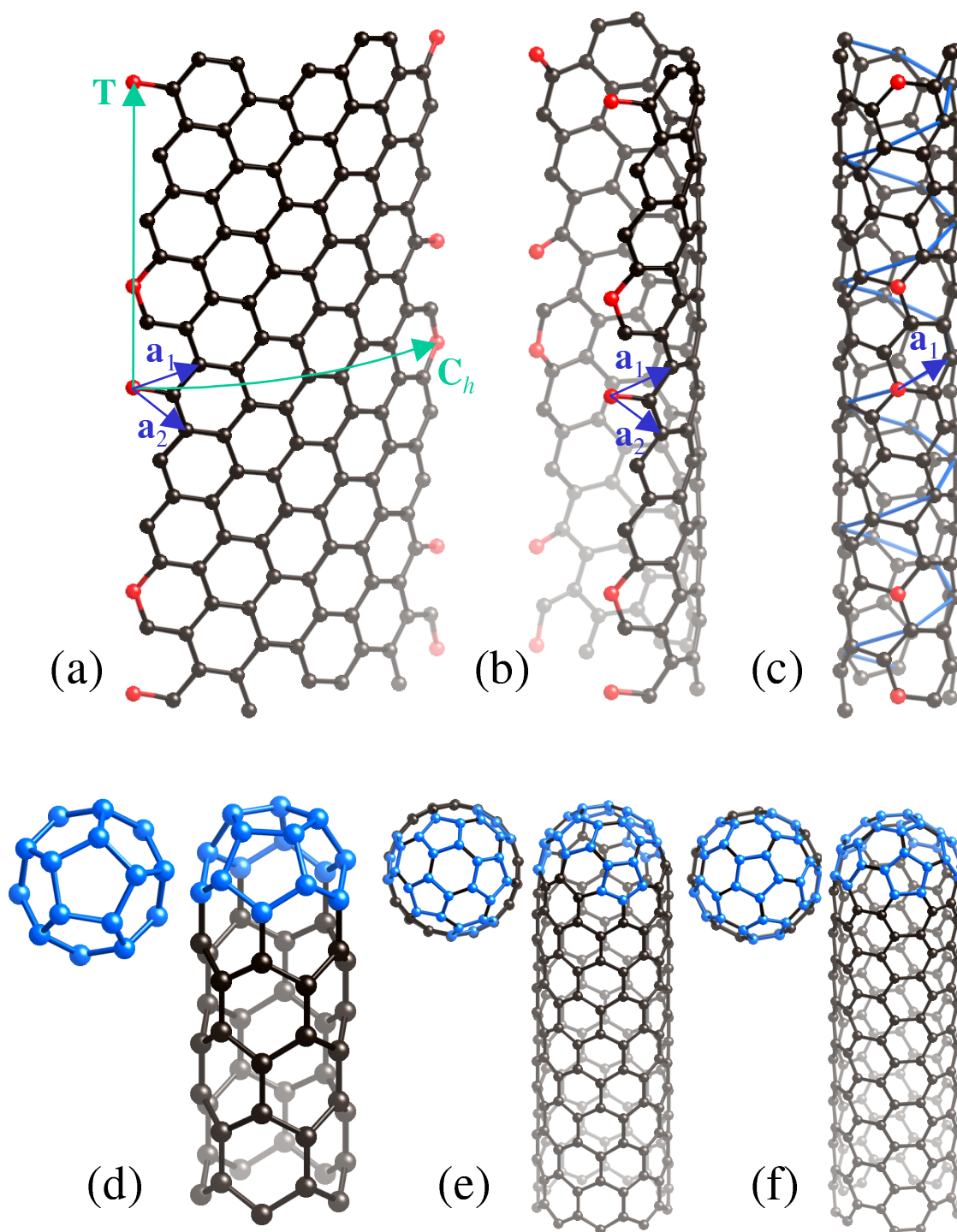
Figure 6.17: (a)-(c) Strip of a graphene sheet (*c.f.* figure 6.16) rolled into a (4,2) nanotube. The atoms on the strip borders that are merged as the graphene sheet is rolled up are marked red. The zigzag line parallel to $\mathbf{a}_1$ forms a right-handed helix, marked blue in (c). (d) (5,0) zigzag nanotube capped with half of a dodecahedron. (e) (9,0) zigzag nanotube capped with half a $C_{60}$ molecule with a hexagon at the apex. (f) (5,5) armchair nanotube capped with half a $C_{60}$ molecule with a pentagon

The special cases $(n, 0)$ and $(n, n)$ represent the non-chiral *zigzag* and *armchair*[7] configurations (figures 6.17(e) and (f)), while all other nanotubes are chiral with $(n + m, -m)$ being the mirror image of $(n, m)$. Here the convention that $\mathbf{a}_2$ is 60° clockwise from $\mathbf{a}_1$ is adopted. Though at odds with the usual crystallographic notation where $\mathbf{a}_2$ is 120° counterclockwise from $\mathbf{a}_1$, this convention is widely used in the nanotube community because it ensures that the tube diameter increases with increasing $m$ and that a tube where the zigzag line along the $\mathbf{a}_1$ direction describes a right-handed screw has positive values for $m$ and the chiral angle $\alpha_c$ (figure 6.17(c)).

The chirality determines the orientation of the graphite hexagons w.r.t. the tube axis and has a critical influence on the electronic properties of the nanotube. Armchair $(n, n)$ nanotubes are predicted to be metallic, all others are semiconducting, with a narrow bandgap when $n - m$ is a multiple of 3 and a wider bandgap otherwise (Saito et al., 1998). This is due to the peculiar band structure of the graphene sheet, where the conduction and the valence band coincide at isolated points, corresponding to the corners K of the Brillouin zone (Elliot, 1998).

As synthesised, nanotubes are terminated with hemispherical caps, each containing 6 pentagonal rings and a number of hexagonal rings that depends on the tube diameter (figures 6.17(d)-(f)). The smallest possible nanotube, shown in figure 6.17(d), is a (5,0) nanotube with diameter $d = 0.4$ nm capped by half a $C_{20}$ dodecahedron (Qin et al., 2000).

### 6.3.2   Filling of carbon nanotubes

It is possible to fill the hollow cavity of carbon nanotubes with a range of materials. For a multi-walled nanotube this was first achieved by Ajayan and Iijima (1993), and for SWNTs by Sloan et al. (1998). This filling process offers interesting possibilities, as the one-dimensional confinement is expected to change the properties of the filling material compared to that of the bulk. Additionally the filling material, acting as an electron donor or acceptor, is also expected to alter the electronic properties of

---

[7]These names arise from5the shape of the cross-sectional ring.

the surrounding nanotube (Terrones et al., 1998). It has also been suggested to use nanotubes as a mold to grow one-dimensional nanowires (Kiang et al., 1999).

Prior to the filling process, the terminating end caps of the nanotube have to be opened. This is usually achieved by oxidation in the gas phase (Ajayan et al., 1993; Tsang et al., 1993) or in solution (Tsang et al., 1994; Sloan et al., 1998), which preferentially etches the 5-membered rings in the end caps. The filling material is then drawn into the opened nanotube from a solution or melt aided by capillary action. The nanotubes characterised in this work were prepared by the nanotube group in the Oxford Inorganic Chemistry Laboratory using the capillary filling technique from molten material.

### 6.3.3   A Nanotube filled with Potassium Iodide

**Exit wave restoration**

The techniques outlined in the previous chapters had to be refined in several ways to deal with the particular difficulties of carbon nanotube samples. As they were supported only on one end, the observed nanotubes were susceptible to bending, which caused slight sample rotations between subsequent exposures, rendering the relative focus determination unreliable. Therefore, the relative focus levels were determined from a neighbouring small area of amorphous carbon that was also in the field of view. The rotation of the tube between subsequent exposures also needed to be taken into account in the registration process. Therefore, the displacements and rotation angles of the images were determined separately for 12 subregions of size $200 \times 200$ pixel by maximising the PCF peak height as a function of rotation angle with the focus differences fixed. From the restoration subsequently calculated, the residual defocus was determined and found to vary little along the tube, indicating that the tube is oriented nearly parallel to the image plane.

Figure 6.18(b) shows the phase of the restored wavefunction constructed from the restorations of the individual subregions. The structure of the encapsulated crystal is more easily visible in figure 6.18(d), where the nanotube has been straightened
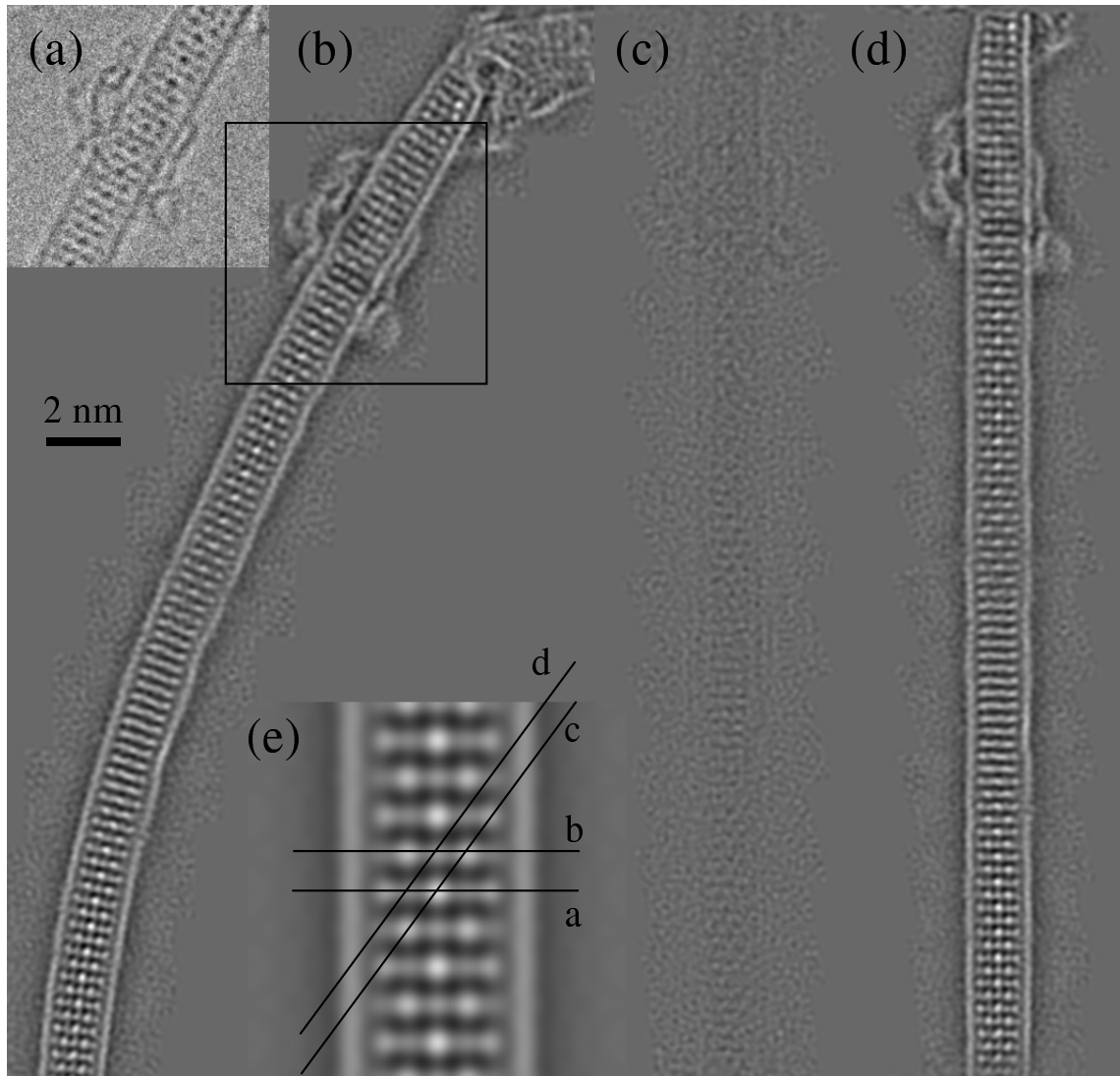
Figure 6.18: Restoration of a nanotube filled with potassium iodide. (a) Subregion of the Scherzer focus image. (b) Restored phase, constructed from 12 overlapping subregions. For each subregion, the registration, absolute focus determination and restoration was carried out individually. The area shown in (a) is marked. (c) Modulus and (d) phase after resampling to compensate for the bending of the tube. In (c), $|\Psi| - 1$ is displayed with the same contrast range as the phase $\arg(\Psi)$ in (d). The modulus shows very little contrast, as the tube and the crystal are sufficiently thin to form a nearly pure phase object. (e) Restored phase, averaged from 15 equivalent subregions in 4 equivalent orientations. The four linescans displayed in figure 6.21 are marked.

as discussed subsequently. The contrast in the restored modulus (c) is very low, indicating that the filled nanotube is nearly a pure phase object. Figure 6.18(e) shows the restored phase where the signal to noise ratio has been improved using unit cell averaging over 15 suitable unit cells and averaging the result over the 4 orientations that are equivalent by symmetry.

**Structural analysis**

A careful examination of the atomic positions and the phase shifts at the measured positions allowed the crystal to be identified as a $3 \times 3$ atom crystal fragment of KI in a [110] projection, with the tube axis pointing in a [001] direction. This implies that the alternate rows of the crystal perpendicular to the tube show in projection arrangements of I-2K-3I-2K-I and K-2I-3K-2I-K, respectively. Since iodine is a stronger scatterer than potassium, the former type of row is easily identified by the strong maximum in the centre. In the restored phase, the crystal is clearly resolved in the lower and upper part of the tube, in the middle, however, the atomic rows are blurred, probably because the crystal is twisted. Interestingly, the I-2K-3I-2K-I layers in the upper part of the crystal are an odd number of rows above their counterparts in the lower part, indicating that there is a stacking fault in the twisted region of the crystal.

The lattice spacings, measured from the peak positions in the restored phase, are given in table 6.3. Significant lattice distortions are observed compared to the bulk structure; most importantly, the 1I column is pushed inwards by the tube wall, reducing the 3I-1I spacing by 5%. The resolution in the restoration is not sufficient to fully separate the peaks for neighbouring atomic columns and the apparent position of the weak 1K peak is shifted because it lies on the shoulder of the much stronger 2I peak. Therefore, this position had to be refined by comparison with simulations, as discussed later.

In this focal series, no information about the tube conformation was recovered, since no periodic pattern is resolved on the tube wall (*c.f.* section 6.3.5). The diameter of the (12,12) armchair nanotube used in the model in figure 6.19 matches closely to
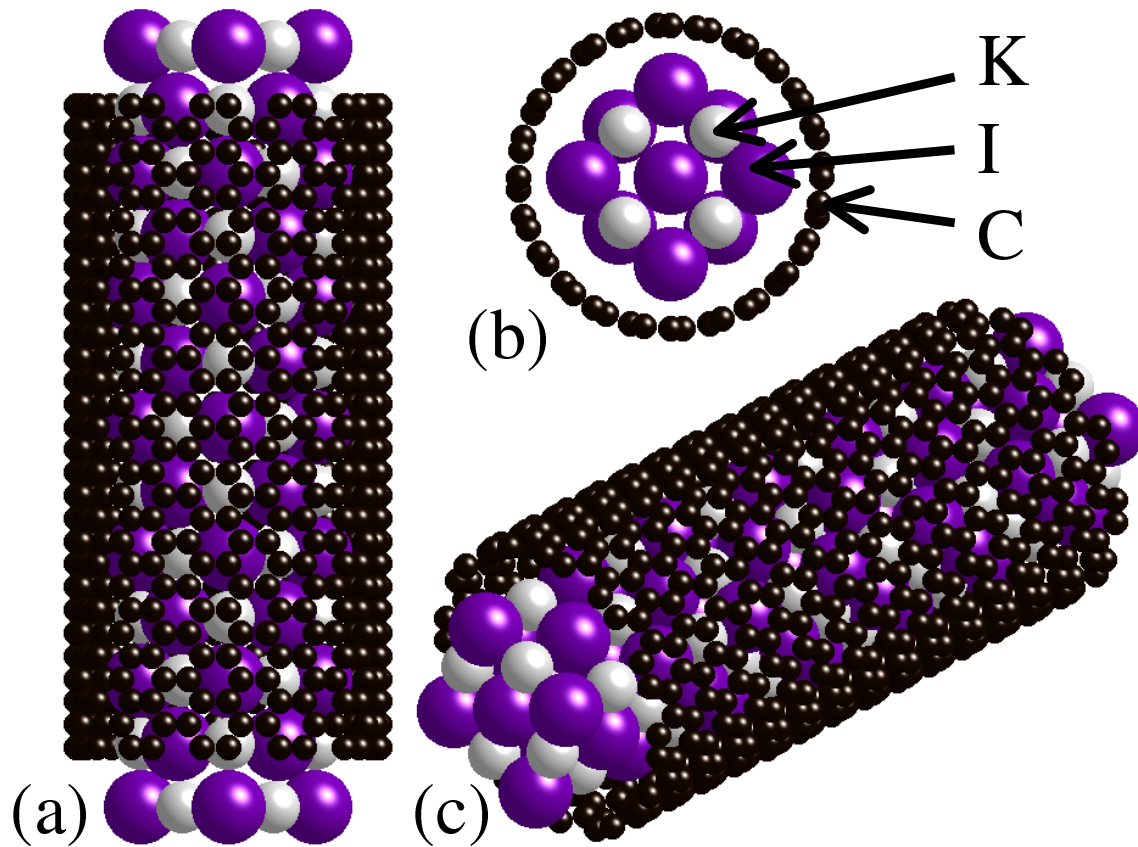
Figure 6.19: Model for the potassium iodide filled (12,12) armchair carbon nanotube with the refined inclusion crystal atomic positions listed in table 6.3. (a) Shown in the experimental viewing direction, with the crystal in a [110] projection. (b) End on view of the tube, showing the inward displacement of the outermost iodine ions. (c) 3-dimensional view of the composite.

| Spacing | Bulk | Measured | Refined |
|---------|------|----------|---------|
| 3I-3I | 0.705 | 0.690 | 0.690 |
| 3I-2K | 0.249 | 0.251 | 0.251 |
| 3I-1I | 0.498 | 0.474 | 0.474 |
| 3K-2I | 0.249 | 0.273 | 0.273 |
| 3K-1K | 0.498 | 0.510 | 0.474 |

Table 6.3: Measured interatomic spacings for the encapsulated KI crystal, compared to those in bulk KI. The outermost iodine ions (1I) are displaced inwards, while the 2I columns are displaced outwards. The 1K columns are also displaced outwards, albeit by a lesser amount. A small contraction along the tube axis (3I-3I) is also observed. When the values measured directly from the linescan peak positions are used in a multislice simulation, the resultant simulated phase shows a significant outward shift of the peaks for the single potassium ion compared to the experimental peak positions. This is accommodated in the 3I-1K spacing in the refined model, which yields simulated peak positions in good agreement with the experimental ones.

the experimentally determined diameter of 1.6 nm, but other conformations with similar diameters could also be chosen.

Figure 6.19 shows a model of the nanotube/KI composite using the refined KI lattice spacings listed in table 6.3.

**Comparison with simulations**

Using the above model, multislice simulations (appendix A) were calculated and the simulated restored wavefunctions (appendix A.3) were compared with the experimental one. Initial comparisons (Meyer et al., 2000b) led to the conclusion that the contrast in the experimental restored phase is lower by a factor of 2 than that in the simulated one. However, a nearly Gaussian dependence of this discrepancy on spatial frequency is observed when the moduli of the Fourier transforms of both waves at the positions of prominent crystalline reflections are compared. This indicates that the specimen vibration was underestimated in the original comparison.[8]

---

[8]Another possible explanation is that the focal spread or beam divergence have been underestimated. However, this is unlikely, since a much better transfer was observed for different datasets
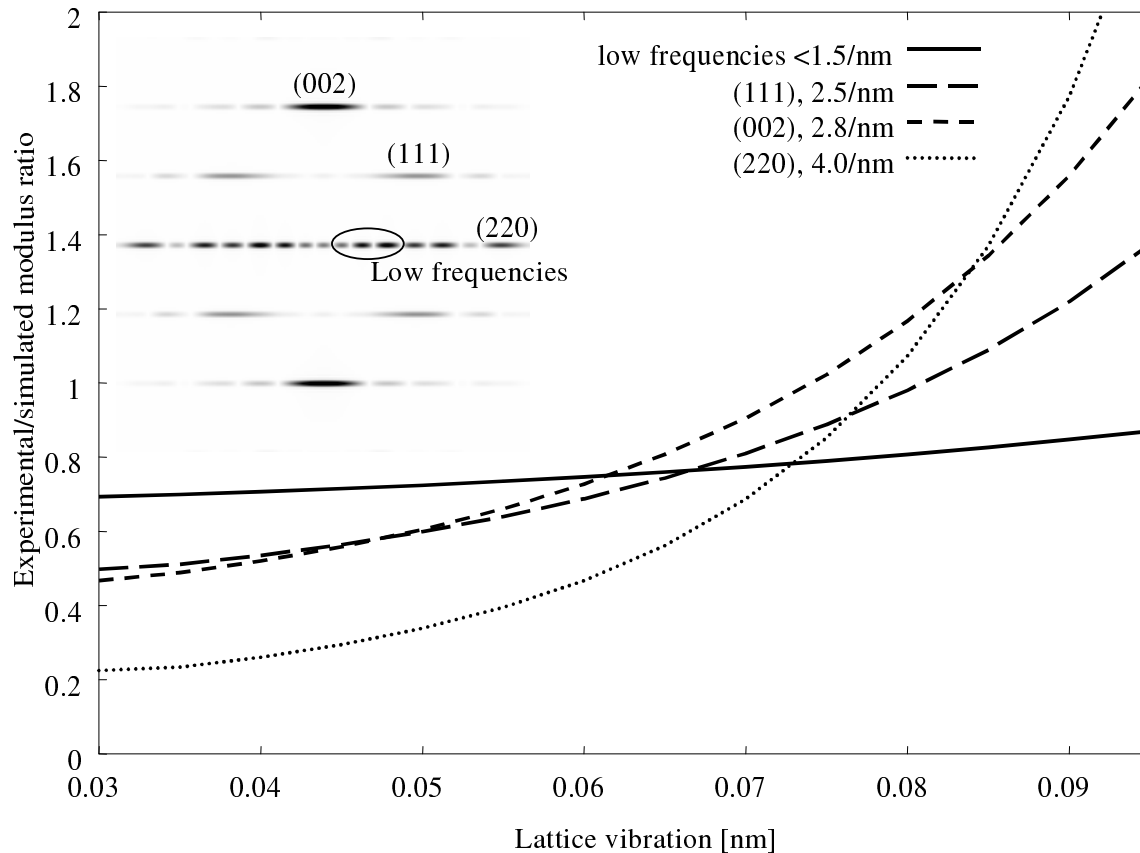
Figure 6.20: Comparison of the restored exit wave with multislice simulation results. The experimental/simulated modulus ratios of prominent reflections for the encapsulated $3 \times 3$ KI crystal are plotted as a function of the assumed lattice vibration. The low frequency values are taken from the streak perpendicular to the tube axis. The inset shows the Fourier transform modulus of the unit cell averaged, restored wavefunction with the lattice reflections indexed.
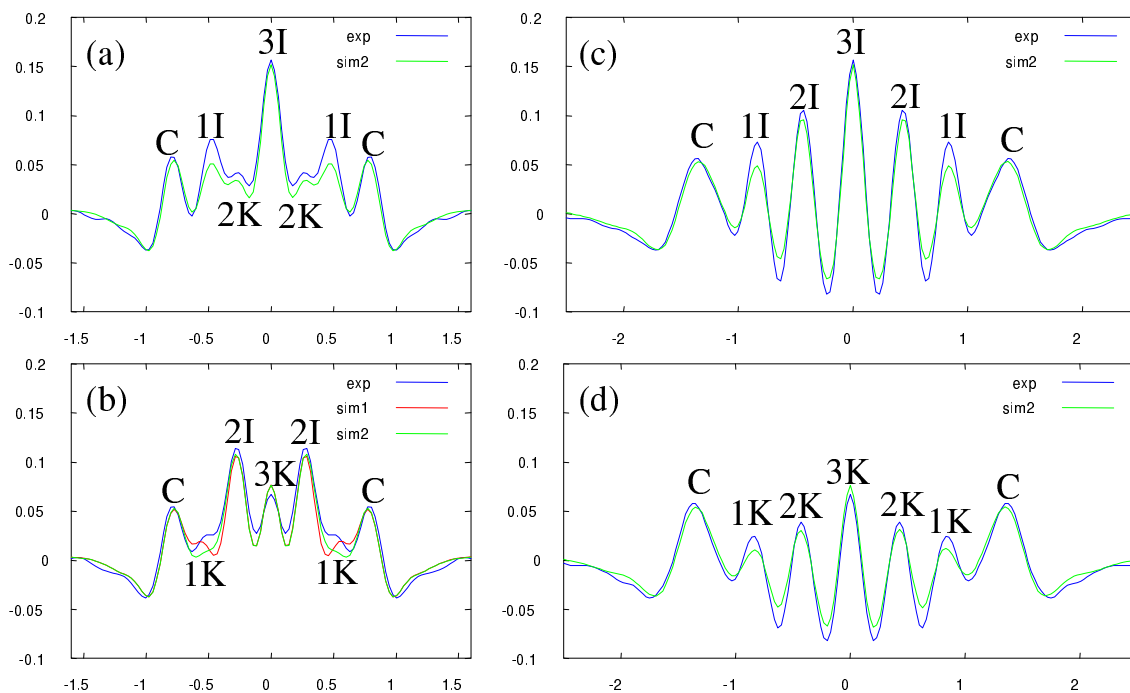
Figure 6.21: Line traces ((a) to (d) marked correspondingly in figure 6.18(e)) of the unit cell averaged experimental restored phase compared with the simulations sim1 and sim2, based on the measured and refined spacings given in table 6.3, respectively. Both restoration and simulations are calculated under the assumption of a specimen vibration of 0.07 nm and in addition, the contrast in the simulation is reduced by a frequency independent factor of 0.8.

Figure 6.20 shows how the ratio between the experimental and simulated Fourier transform modulus changes as a function of the of RMS specimen vibration included in the restoration and the simulation. The modulus at low frequencies is nearly unaffected by the lattice vibration, and was lower by a factor of 0.8 in the experiment compared to the simulation. This contrast reduction for all frequencies can possibly be attributed to inelastic scattering.[9]

Taking into account this spatial frequency independent contrast reduction, the best match between experiment and simulation is achieved for an assumed lattice vibration of 0.07 nm.

In the original comparison (Meyer et al., 2000b) a specimen vibration of 0.05 nm was assumed, and figure 6.20 shows that under this assumption, the highest order beam is weaker by a factor of 3 in the experimental restoration compared to the simulated restored wave.

Figure 6.21 shows that a good match between experiment and simulation can be achieved by assuming a lattice vibration of 0.07 nm together with an additional frequency independent contrast reduction by a factor of 0.8.

**Straightening of the tube**

From the restoration, the straight nanotube shown in figures 6.18(c) and (d) was extracted as follows as follows:

1. Two small templates were extracted from the restored phase, one centred at a strong (3I column) and one at a weak spot (3K column) on the line in the centre of the tube. Using orientational correlation with the mirror image, both were rotated to display the tube vertically and then symmetrised by averaging over 4 orientations. Figures 6.22(a) and (c) show these templates.

recorded with the same microscope under the same operating conditions, *e.g.* the datasets discussed in the previous section. This indicates that the contrast reduction is specimen rather than microscope related.

[9]For thicker specimens of amorphous carbon, Boothroyd (2000) has found spatial-frequency independent reduction by a factor of 0.33 in the contrast of experimental images compared to simulations.
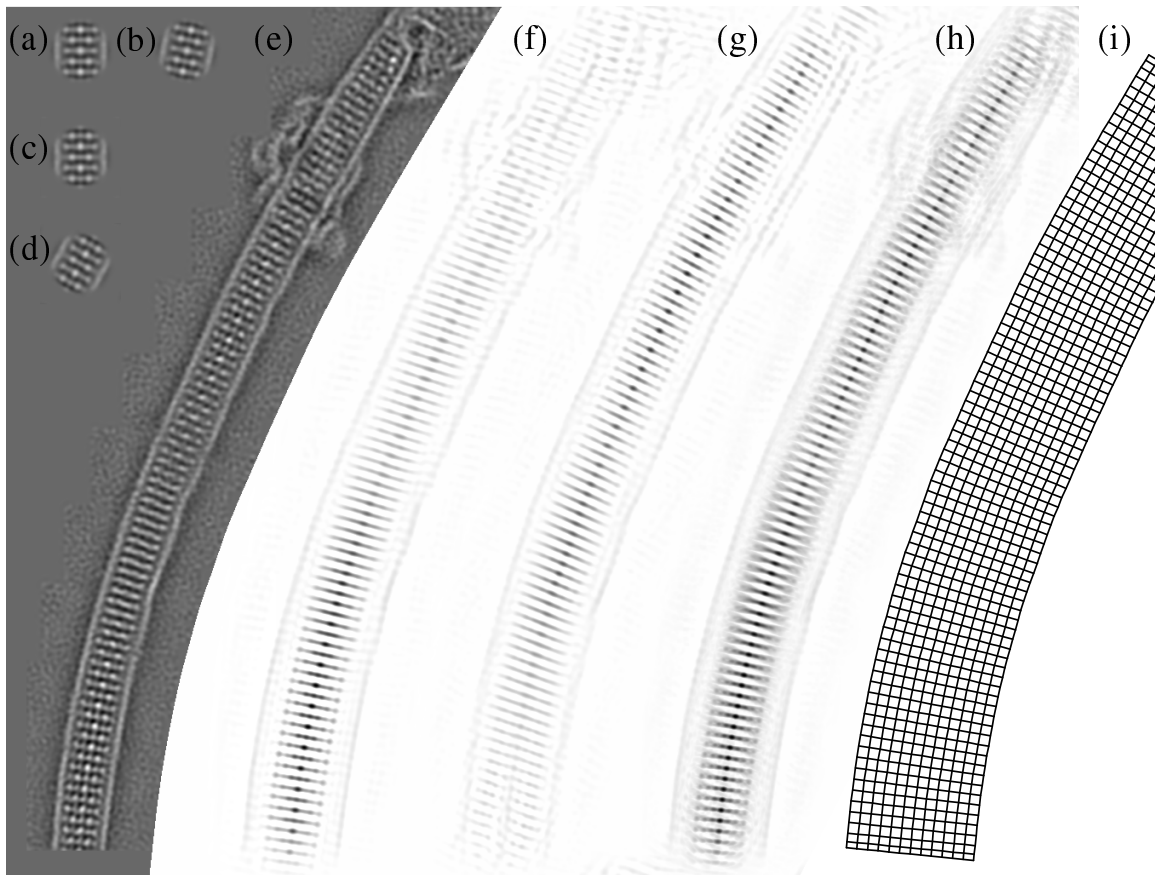
Figure 6.22: Image processing to compensate for the tube bending. (a) Template centred at a 3I column, cut out from the restored phase, (e) rotated and symmetrised. (b) Template (a) rotated by $\theta = 0.2\,\text{rad}$ clockwise. (c) Template centred at a 3K column. (d) Template (c) rotated by $\theta = 0.5\,\text{rad}$ clockwise. (e) Restored phase. (f) Cross-correlation of (e) with (b). Strong correlation peaks (black) appear at the 3I sites at the lower end of the tube. g) Cross-correlation of (e) with (b). Strong correlation peaks appear at the 3K sites at the upper end of the tube, where the rotation angle is close to 0.5 rad. (h) Maximum of all correlation functions of (e) with (a) and (b) rotated by angles in the range $0 \leq \theta \leq 0.7\,\text{rad}$, yielding strong peaks at all 3I and 3K sites. (i) Remapping grid calculated using the local rotation angle determined at each of these sites.

2. Images $C(x,y)$, $D(x,y)$ and $T(x,y)$ were created and initialize with 0 values for all pixels.

3. For both templates T$\in \{1,2\}$ and a range of rotation angles ($0 \leq \theta \leq 0.7$ rad, step 0.01 rad) the following steps were repeated;

   (a) The XCF $C_{T,\theta}(x,y)$ was calculated between the restored phase and the template T rotated clockwise by $\theta$ (figure 6.22(f) and (g)).

   (b) $D(x,y) := \theta$ was updated for all $(x,y)$ with $C_{T,\theta}(x,y) > C(x,y)$.

   (c) $T(x,y) := T$ was updated for all $(x,y)$ with $C_{T,\theta}(x,y) > C(x,y)$.

   (d) $C(x,y) := \max(C(x,y), C_{T,\theta}(x,y))$ was updated.

4. $C(x,y)$ now shows a line of peaks along the tube centre (figure 6.22(h)). At the peak positions $(x,y)$, $D(x,y)$ gives the local rotation angle and $T(x,y)$ is 1 or 2, respectively, if the peak corresponds to a 3I or 3K column.

5. The peak positions and local rotation angles were used to construct a grid (figure 6.22(i)) to extract a straightened version of the tube.

### 6.3.4 Encapsulated antimony oxide

In this section a focal series restoration of a single-walled nanotube filled with a antimony oxide ($Sb_2O_3$) is discussed. This particular restoration has enabled a detailed study of the nanotube conformation and the inclusion crystallography simultaneously. Since the latter was determined in collaboration,[10] I will focus here on the nanotube conformation, highlighting the use of accurate focus determination to derive the inclination of the specimen with respect to the image plane by measuring the focus in small subregions. A full account of this work has been published elsewhere (Friedrichs et al., 2001).

As in the previous example, the lateral movement of the nanotube between exposures made it necessary to perform the registration separately in small subregions. The restored subregions were then combined to give the final restoration of the complete nanotube in the same orientation and position as in the reference image (figure 6.23(a)). The measured focus levels plotted in figure 6.23(b) increase systematically by 4 nm towards the top end of the tube, indicating that the tube is inclined with respect to the image plane, with the top end in figure 6.23(a) higher in the beam path. Fitting a straight line to these values gives a standard deviation $\sigma_{n-2} = 1.1\,\mathrm{nm}$ for the 6 individual focus values and yields an inclination angle of $\beta = 17° \pm 5°$.

### 6.3.5 Tube conformation

Careful inspection of the restored phase in figure 6.23(a) reveals a periodic contrast in the right tube wall that is absent in the left tube wall. This is confirmed by power spectra taken from linescans along the tube wall and plotted in figure 6.24, which show a strong peak corresponding to a spacing of 0.224 nm for the right tube wall only. This corresponds to the centre-to-centre spacing $1.5\,d_{C-C} = 0.216\,\mathrm{nm}$ between neighbouring 'zigzag' rows of carbon atoms in the SWNT wall lattice when viewed in projection.

---

[10]with Steffi Friedrichs, Inorganic Chemistry Department, Oxford University
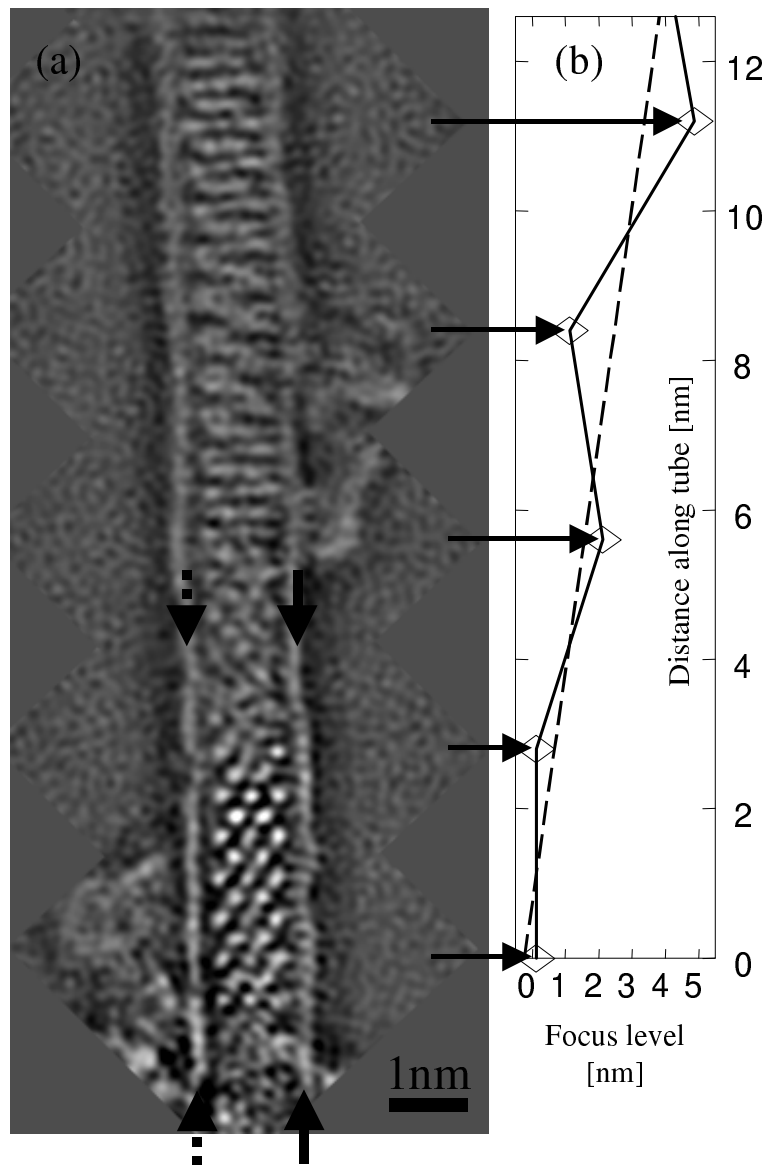
Figure 6.23: (a) Phase of the $Sb_2O_3$-filled carbon nanotube restored from a 20 member focal series. The image is composed of 6 individually restored overlapping subregions along the tube axis (uppermost subregion not shown). The lower part shows the crystal clearly resolved and also regular fringe contrast along the right tube wall. (b) The focus levels measured for the individual subregions show a systematic increase along the tube, indicating that it is inclined w.r.t. the image plane. The gradient of the straight line fitted to the 6 measured values (dashed) gives an estimate for the inclination angle of $17° \pm 5°$.
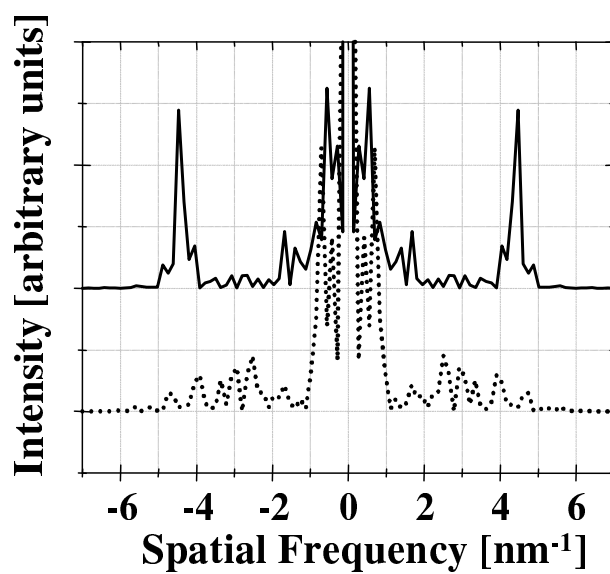
Figure 6.24: Power spectra of one-dimensional linescans along the tube walls as indicated in figure 6.23(a) The spectrum for the right wall (solid graph) shows a strong peak corresponding to spacing of 0.224 nm. This peak is absent for the left wall (dotted graph). Overall this indicates that the tube is chiral and inclined to the image plane (see later).

The extent to which this periodic contrast along the tube wall is visible is determined by the image resolution, the nanotube conformation and the tilt of the tube relative to the electron beam. In figure 6.25 this is illustrated with structure models and corresponding simulations of three SWNTs that closely match the experimental tube diameter of 1.45 nm.

For achiral tubes, the zigzag directions are parallel on both tube walls, leading to equal contrast on both walls for all inclination angles.

For a 'zigzag' tube in the image plane (figure 6.25(a)), strong contrast is observed on both tube walls, as the zigzag rows are aligned parallel to the beam direction. This contrast gets weaker as the tube is tilted out of the image plane because the rows start to stagger in projection.

For an 'armchair' tube, no wall contrast is observed when the tube lies in the image plane, as the projected distance between carbon columns is now only $d_{C-C}/\sqrt{3} = 0.125$ nm, beyond the experimental resolution in the restoration. Some contrast along the wall with a larger spacing becomes visible as the tube is tilted out of the image plane, but a full alignment of the 'zigzag' columns with the projection direction requires a tilt angle of $\beta = \pm 30°$.

With its axis in the image plane, the chiral (21,-8) tube in figure 6.25(c) shows little contrast on both tube walls. However, when the tube is tilted out of the image plane, the alignment of the zigzag lines with the projection direction improves for one tube wall and worsens for the other, giving rise to the asymmetric contrast observed experimentally.

In general, the visibility of contrast in the left (right) tube wall depends on the angle $\delta = |\alpha_c - \beta|$ ($\delta = |\alpha_c + \beta|$) between the projection direction and the 'zigzag' lines, where $\alpha_c$ is the chiral angle (equation 6.5) and the angle $\beta$ between the tube axis and the image plane is defined as positive if the top end of the tube is above the image plane (as for this specimen). The observed contrast on the right wall of the imaged tube in figure 6.23(a) therefore means that the tube is chiral with a negative chiral angle $\alpha$ and hence a negative integer $m$. When the zigzag lines are

| Tilt Angle $\beta = 0°$ | Vector Model | Tilt Angle $\beta = 15°$ |
|---|---|---|
| (a) | (18,0) zigzag SWNT | |
| (b) | (10,10) armchair SWNT | |
| (c) | (21,-8) chiral SWNT | |
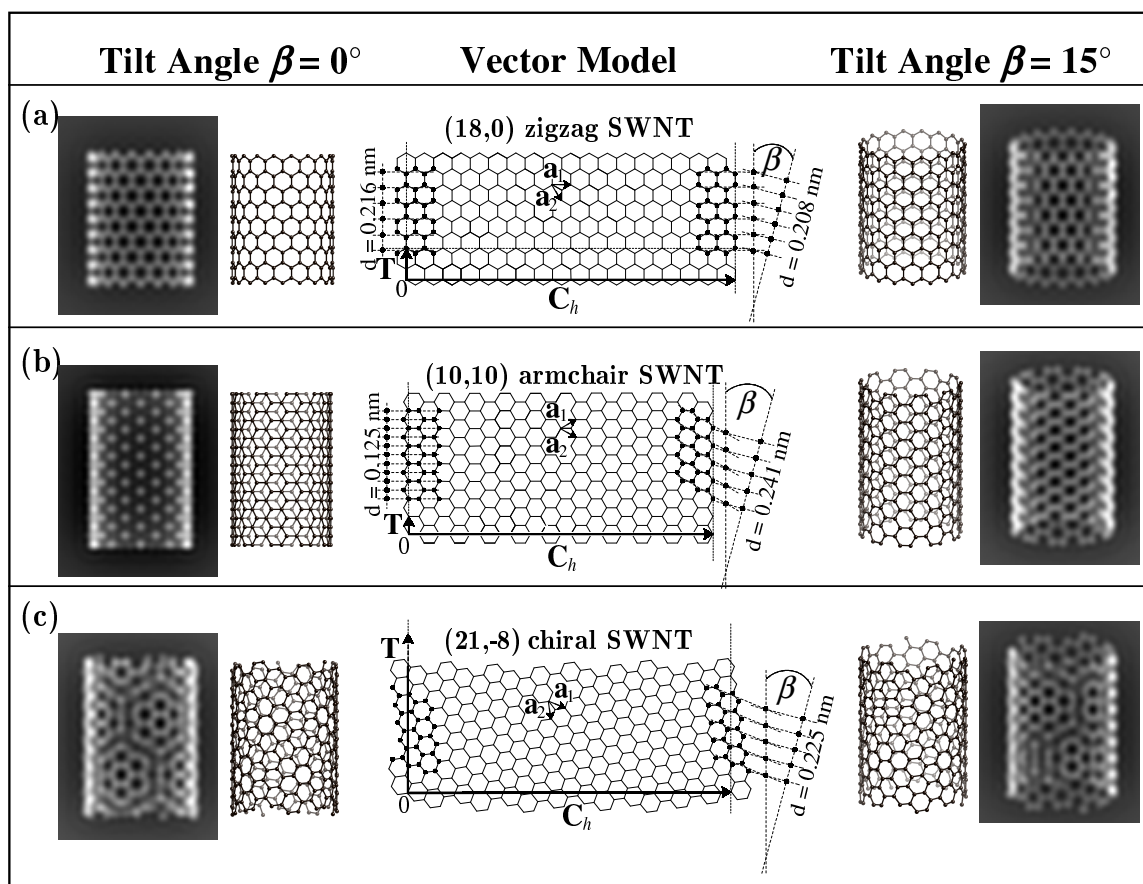
Figure 6.25: Schematic illustration of the observable resolution of fringes within SWNT walls, depending on the conformation and the tilt angle $\beta$. The columns adjacent to the illustrated graphene sheet display modelled fractions of the SWNTs, where the models on the right are tilted by $\beta = 15°$ out of the image plane. The outermost columns show the simulated phase of the displayed models. Asymmetric wall contrast is only observed when the tube is chiral and tilted out of the image plane ((c) right).
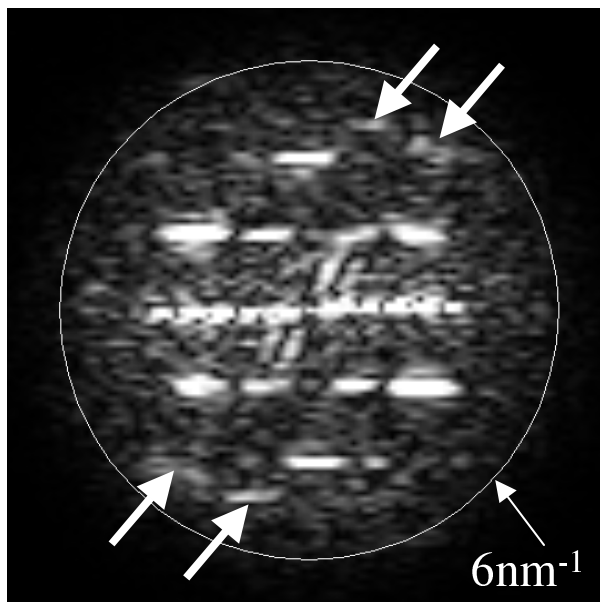
Figure 6.26: Square root of the 2-dimensional power spectrum calculated from the tube in the lower half of figure 6.23(a). The reflections marked by arrows are incommensurate to the crystal reflections and arise from the tube itself.

not exactly aligned with the projection direction contrast along the tube wall remains visible provided that the angle $\delta$ is smaller than *ca.* 10°, because even though the atoms start to stagger in projection, the contrast from neighbouring zigzag rows does not overlap since the number of atoms in projection on the tube wall is small. The observed spacing in this case is given by $1.5\,d_{\mathrm{C-C}}\cos\beta/\cos\alpha$ as illustrated with the model for $\mathbf{C}_h = (21, -8)$ in figure 6.25(c).

Further information on the tube chirality can be obtained from the Fourier transform of the restored wave. The reflections marked by arrows in figure 6.26 are not commensurate with the reflections due to the encapsulated crystal and further confirmation that these reflections arise from the nanotube itself is that the $k_y$-component of the more intense maxima corresponds to the spacing observed in the linescan along the right tube wall. In order to determine the tube conformation multislice simulations for all tubes with diameters $d$ in the range $1.37\,\mathrm{nm} \leq d \leq 1.5\,\mathrm{nm}$, tilted by

(a) (18,0)     (b) (10,10)     (c) (21,-8)

β=0°     β=0°

β=15°     β=15°
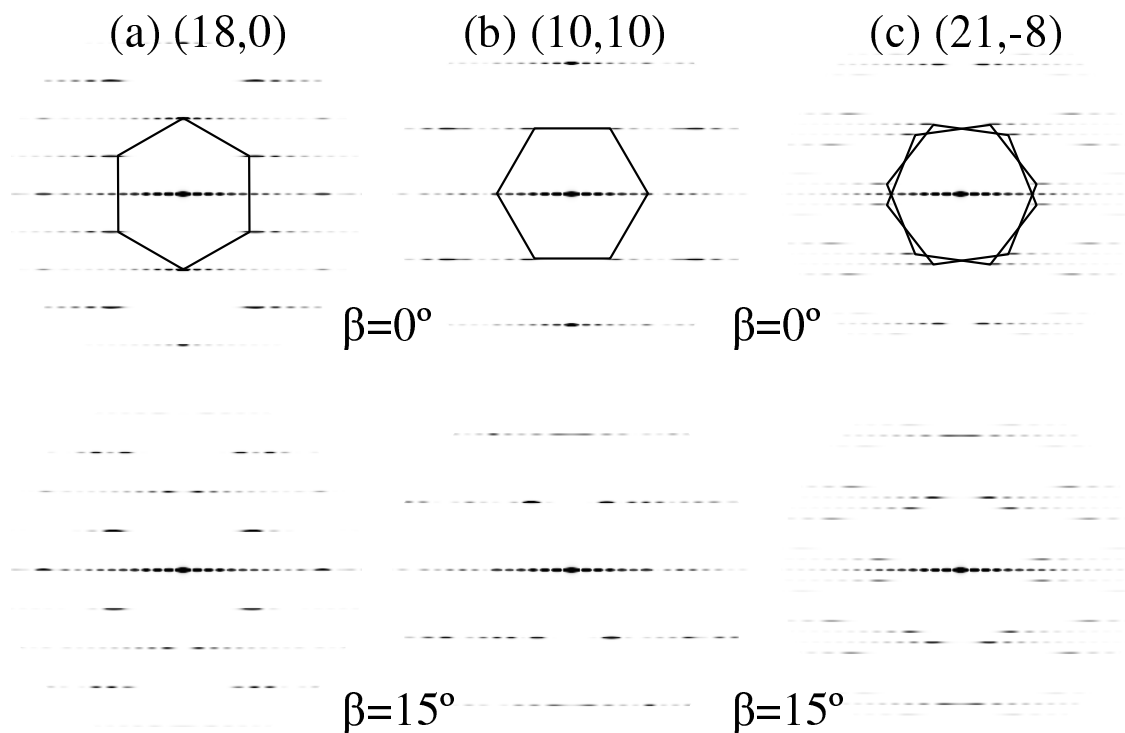
Figure 6.27: Simulated object wave power spectra for (a) zigzag (b) armchair and (c) chiral tube conformations. The first-order reflections are situated at the vertices of two hexagons (due to the contributions from the upper and lower tube wall, respectively) which are rotated by $\pm\alpha$ w.r.t. the tube axis (Amelinckx et al., 1999). In case of the achiral zigzag (a) and armchair (b) conformations, both hexagons coincide. In all figures, the maximum spatial frequency shown is $11\,\mathrm{nm}^{-1}$. Since the projected spacings perpendicular to the tube near the tube wall are reduced compared to a flat graphene sheet, the reflections are elongated and extend from the hexagon vertices towards higher frequencies.

angles of $0°$, $5°$, $10°$, $15°$ and $20°$ were calculated. The only tubes with negative chirality that reproduced the observed peaks in the power spectrum were $(20, -7)$, $(20, -8)$ and $(21, -8)$ with tilt angles of $15°$ and $20°$. As the diameter of the $(21, -8)$ tube (which is the mirror image of a $(13, 8)$ tube) provided the best match to the experimental diameter, this tube was used for all subsequent simulations. The simulated power spectra in figure 6.27 demonstrate this dependence of the first order peak positions on the chiral angle of the tube.[11]

Overall, the conformation of the tube is hence determined using a combination of real space information (tube diameter, tube wall contrast) and Fourier space information (diffraction peaks) obtained from the restored image wave. In addition, the accurate measurement of the change in focus along the tube made it possible to determine the chirality sense. Figure 6.28 illustrates how the combination of these three methods is used to derive the tube conformation.

## 6.3.6 The encapsulated crystal

The restored phase shows contrast due to the filling material in the SWNT over the entire field of view. However only within the lower region of the tube is the crystalline $Sb_2O_3$ sufficiently well resolved to provide useful structural information (Figure 6.29(d)). The uninterpretable crystalline contrast in the remainder of the tube is in part due to movement and rotation of the tube during data acquisition and to misorientation and possible irregularity of the filling in some regions of the tube.

The regular structural motif of the imaged encapsulated $Sb_2O_3$ crystal can be described as a repeating sequence along the tube of a row containing three columns of antimony atoms followed by a row containing two columns of antimony atoms. This pattern resembles the arrangement of Sb columns for an approximate $\langle 1\,0\,\bar{1}\rangle$ projection through a fragment of $Sb_2O_3$ derived from the orthorhombic polymorph

---

[11]The smallest angle between the tube axis and a diffraction maximum on either hexagon is given by $|\alpha_c|$, giving a direct estimate for the chiral angle. However, as the diffraction spots are elongated (figure 6.27) and their position also depends on the tube inclination, this estimate is not as accurate as a comparison with simulations of tubes with different conformation and inclinations.
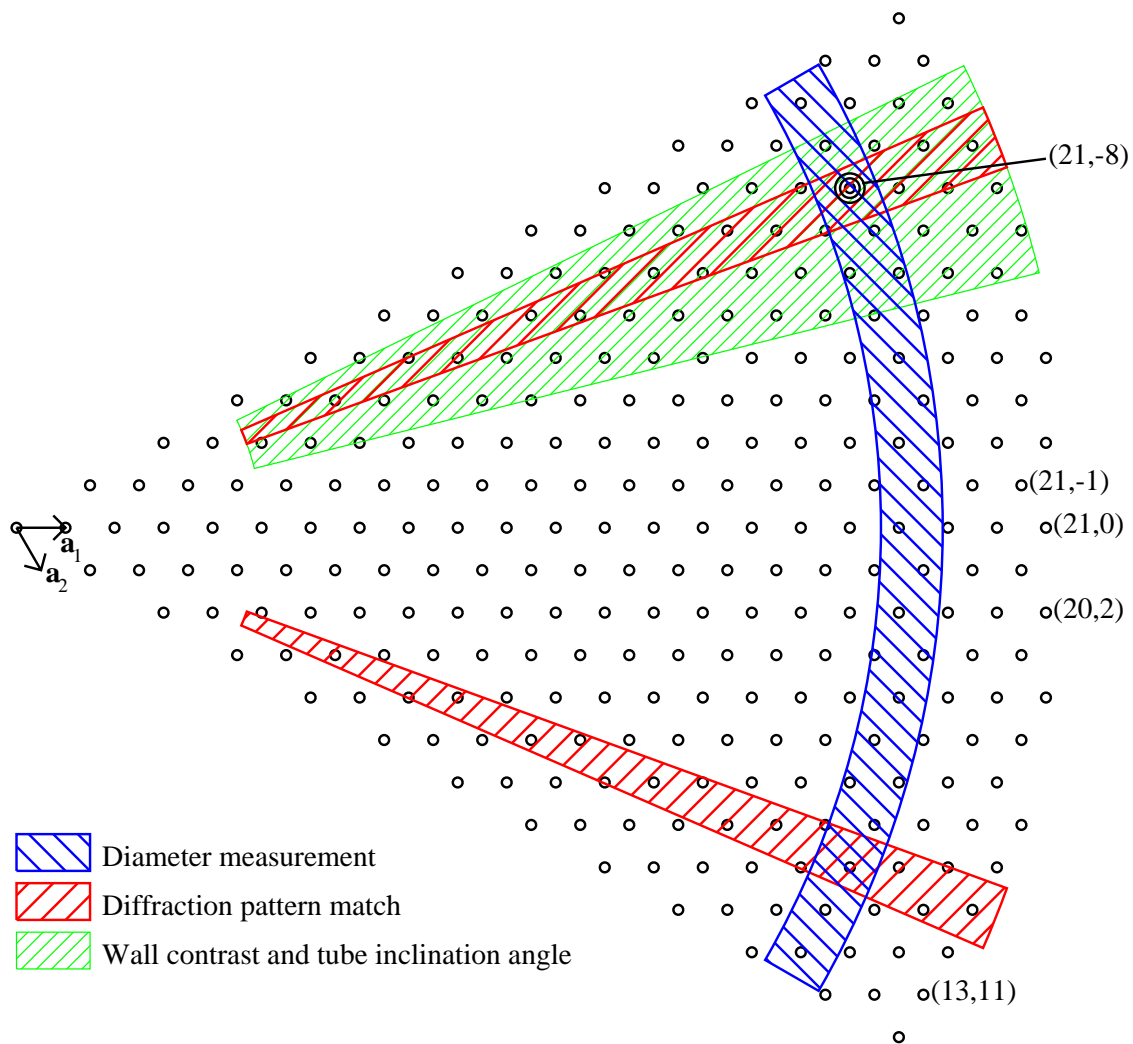
Figure 6.28: The tube conformation is determined by a combination of methods. Measuring the tube diameter yields the length of the chiral vector. Conversely, the diffraction pattern (calculated from the restored wave), is sensitive to the chiral angle modulus $|\alpha_c|$, but not to the tube diameter. The diffraction pattern is also insensitive to the chirality sense (the sign of $\alpha_c$), which can, however, be determined from the asymmetry in the tube wall contrast when the tube is inclined with respect to the image plane and the inclination angle is measured. This investigation of the inclination angle and wall contrast also yields an estimate for $|\alpha_c|$, albeit less accurate than diffraction pattern matching.
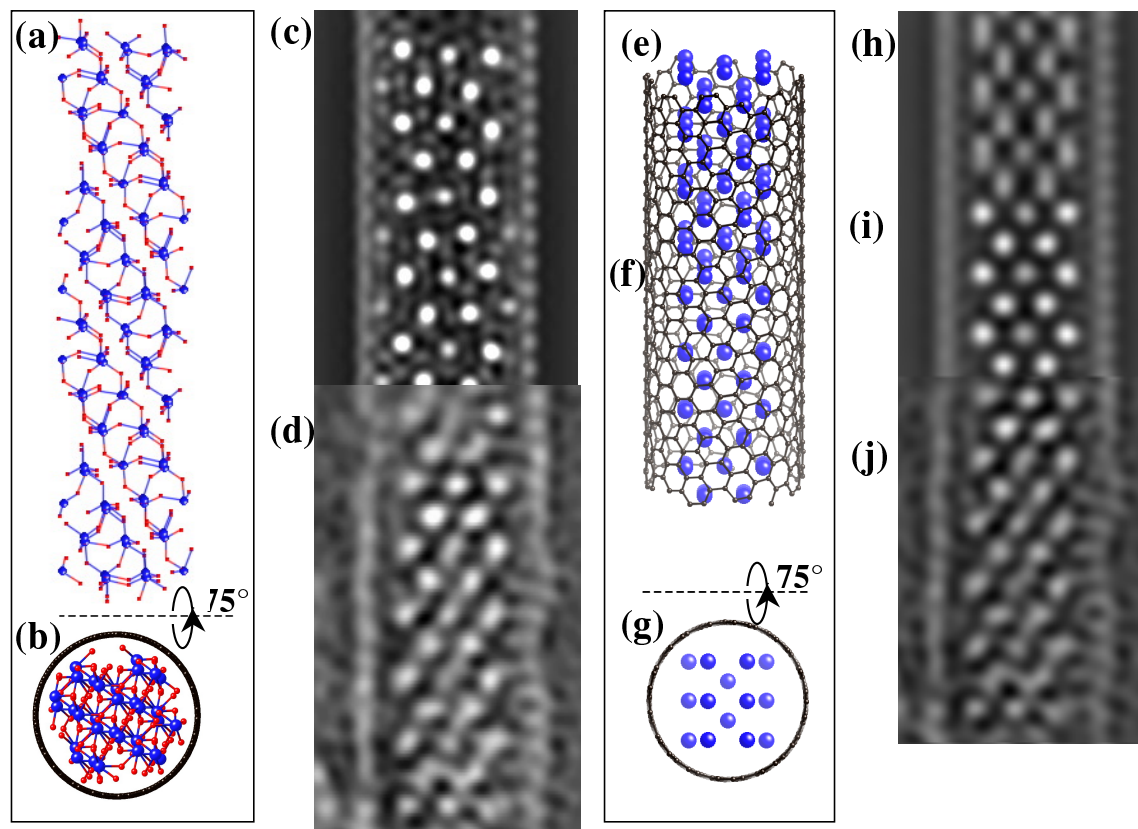
Figure 6.29: (a)-(d) Structural model of the encapsulated $Sb_2O3$ crystal fragment based on the bulk valentinite structure (a) Ball and stick model in the experimental projection close to $\langle 10\bar{1}\rangle$. The antimony atoms are represented by the larger balls. (b) End-on view of the tube, corresponding to the $\langle 4\,\bar{1}\,2\rangle$ direction of the encapsulated $Sb_2O_3$ crystal. (c) Phase of the multi-slice simulated restored wave, neglecting lattice vibrations (d) Experimental restored phase, assuming zero lattice vibrations. (e)-(j) Simplified model for the cation sublattice. (e) In a model where the atomic columns are perpendicular to the tube axis, tilting of the tube by 15° leads to a visible staggering in projection. (f) Model with a monoclinic unit cell, where the atomic columns are nearly parallel to the projection direction (g) End-on view of both models. (h), (i) Simulated restored phase from models (e) and (f) with a lattice vibration of 0.05 nm. (j) Experimental restored wave assuming a lattice vibration of 0.05 nm.

valentinite form. This compound (Svensson, 1974)) is considered to be the high-pressure form and consists of infinite double chains of $Sb_2O_3$ units, as opposed to the cubic senarmonite form (Svensson, 1975), which consists of molecular units of $Sb_4O_6$. In the bulk valentinite structure, the $Sb_2O_3$ double chains run parallel to $\langle 0\,0\,1 \rangle$ with the two chains linked at each antimony site by a shared oxygen, such that each antimony atom is coordinated by three oxygen atoms.(Svensson, 1974)

The proposed structure model and an end-on view of the $Sb_2O_3$/SWNT composite are shown in figure 6.29. In this model, the tube axis is aligned with the $\langle 4\,\bar{1}\,2 \rangle$ direction of the crystal, as this was the only possible tube axis that reproduced the good alignment of Sb atomic columns along the centre line of the tube and was consistent with the observed tube tilt angle. The angle between $\langle 4\,\bar{1}\,2 \rangle$ and the optimum $\langle 1\,0\,\bar{1} \rangle$ viewing direction is 78.3°. Small deviations from this viewing direction make an insignificant difference to the observed contrast and hence a tube inclination of 15° (rather than 11.7°), as used in the simulation in figure 6.29(c) in order to account for the observed contrast within the tube wall, is plausible.

Significant lattice distortions were observed in the $Sb_2O_3$. In comparison with the bulk structure of valentinite, the encapsulated crystal shows a longitudinal contraction of 13% along the $\langle 4\,\bar{1}\,2 \rangle$ direction (*i.e.* along the tube axis), a much larger distortion than observed for the encapsulated KI crystal discussed in the previous section.

The repeat distance along the tube in the restored phase corresponds to 2 Sb rows, whereas the model predicts a repeat distance of 8 Sb rows, indicating that the crystal is forced into a state of higher symmetry due to the confinement by the tube walls. This model of the inclusion crystal is discussed in more detail elsewhere (Friedrichs et al., 2001).

Due to the apparent mismatch between the model based on the bulk valentinite structure and the experimental images, an alternative simplified structural model was also constructed where the cation positions are directly measured from the restored phase and the anion sublattice is ignored, since it is likely to make no difference to the contrast (figure 6.29(c)). Figures 6.29(i) and (j) demonstrate that a better match

between experiment and simulation is achieved using this model and an increased lattice vibration. A comparison with figure 6.29(h) shows that it is necessary to include a tilt of the atomic columns with respect to the tube axis in order to account for the sharp appearance of the atomic columns in spite of the measured tilt of the nanotube out of the image plane.

The restoration in this section demonstrated that

1. The object wave restored from a focal series allowed a characterisation of both the SWNT and the encapsulated crystal with an accuracy that could not be achieved from a single HRTEM image.

2. The chiral tube conformation could be inferred from the measured tube diameter and from peak positions in the power spectrum of the restored wave.

3. For the first time, the chiral sense of a carbon nanotube has been directly determined using HRTEM. This has been achieved by an analysis of the asymmetric tube wall fringe contrast in combination with an accurate measurement of the change in defocus along the tube.

4. For the encapsulated crystal, a model based on the valentinite form of $Sb_2O_3$ has been devised. However, compared to that model, a significant contraction along the tube was observed and the translational symmetry along the tube axis was found to be higher than in the bulk structure. As a starting point for an alternative model, the cation positions measured from the restored phase can be used, and a model using these positions and ignoring the anion sublattice yields a good match to the experimental data.

# Chapter 7

# DigiTEM: A dedicated system for external control of electron microscopes

The extension of the automated aberration determination methods developed in the previous chapters to fully automated microscope alignment requires direct access by an external computer to all the microscope settings. The only facility provided on most JEOL microscopes for such access is a relatively slow serial port, which suffers from the following disadvantages:

1. Only *ca.* 3-5 commands per second can be transmitted to the microscope.

2. The external computer is not notified about local changes in the microscope settings made directly by the operator. This makes it difficult for an external auto alignment system to adapt to changes, for instance, in the magnification.

3. It is not possible for the external computer to take over complete control of selected microscope settings.

In order to overcome these difficulties, a piece of dedicated hardware, called 'Digi-TEM' was developed, which is connected into the digital bus that links the microscope's internal computer (MC) with the deflector unit[1] as shown in figure 7.1. In order to also control the lens currents a second identical[2] unit has to be connected between the MC and the lens unit. Figure 7.2 demonstrates that the installation of DigiTEM only requires minimal modification to the microscope.[3] The external PC is connected via an optically isolated fast (115200 baud) standard RS422 serial port and hence only a standard dual RS422 board is required in the PC to control both the deflector and the lens DigiTEM units.[4]

## 7.1  Firmware

The PIC17C43 microcontroller was programmed in assembler code using the microchip MPLAB integrated development environment.[5]

Figure 7.3 illustrates the program and dataflow for the DigiTEM firmware. Four queues served by interrupt routines have been implemented to allow a quick reaction to external events. Reaction to write commands from the MC is particularly time critical, since the write pulse is only 12 $\mu$s long and hence, this interrupt is given the highest priority and can be nested with other interrupts.

---

[1]The deflector unit contains the digital to analogue (D/A) converters that control the deflector coil currents.

[2]Though the busses leading to the deflector and lens units are identical, the pin assignment of the connectors is entirely different. These differences can, however, be fully accommodated by the firmware, so that identical boards can be used for both purposes.

[3]Although DigiTEM was initially designed for general control purposes, its scanning features have made it attractive for other research groups. At present, units are installed in the JEM 4000EX and JEM4000FX microscopes in the Department of Materials Science in Cambridge, in a JEM 2010 in the Department of Chemistry in Cambridge and in a JEM 2010 in the electron microscopy unit of the Lincoln University in Nebraska. These latter installations are mainly used by the respective research groups for hollow cone (Kunath et al., 1985) and related work.

[4]In the RS422 standard, data are transmitted via a differential line pair, which makes it more reliable than the single ended RS232 standard used in the built in serial ports on PCs. Macintosh computers use the RS422 standard by default.

[5]MPLAB, as well as data sheets of all PIC microcontrollers are available free of charge at www.microchip.com.
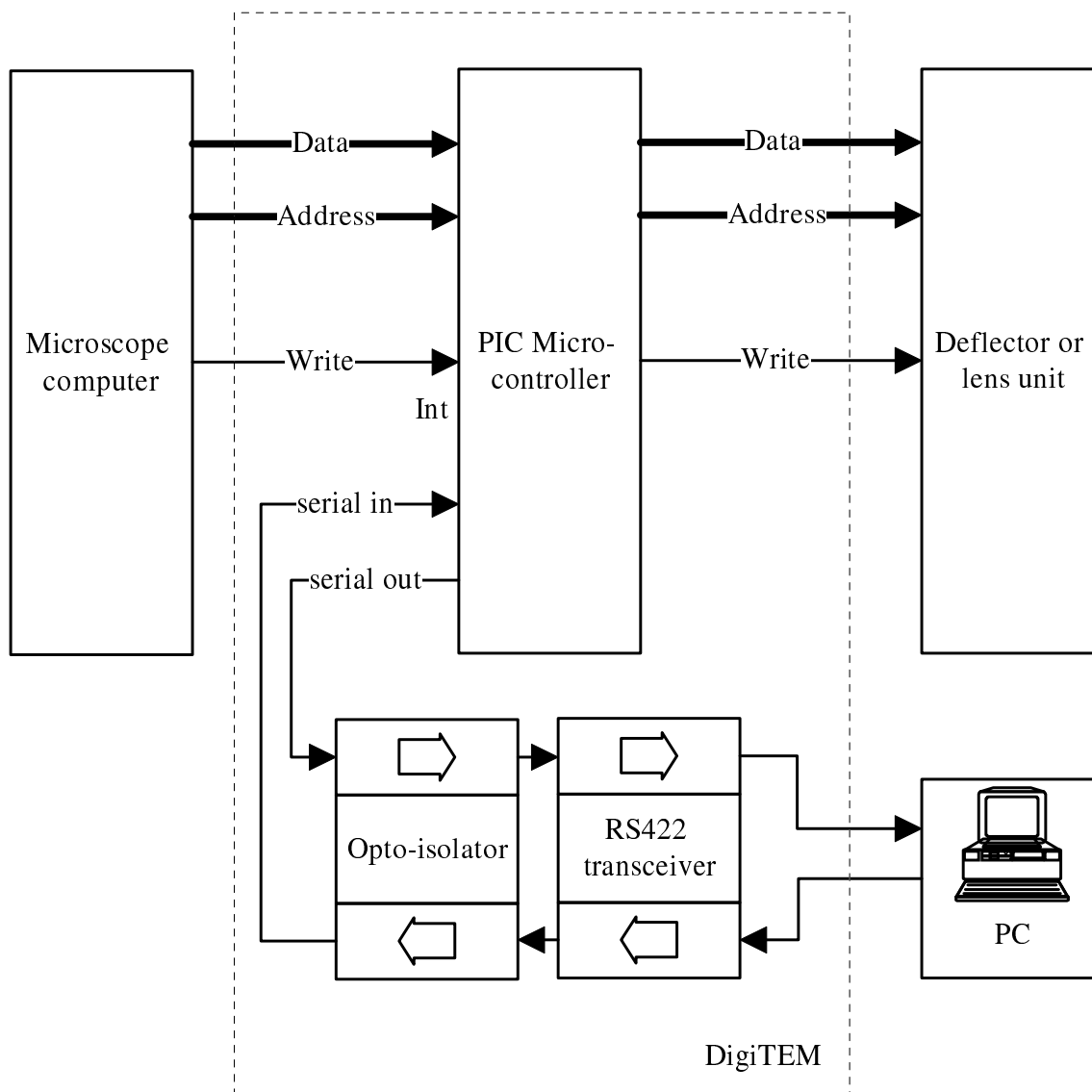
Figure 7.1: Schematic diagram of the DigiTEM interface. The DigiTEM intercepts write commands on the digital bus from the microscope computer to the deflector or lens unit.
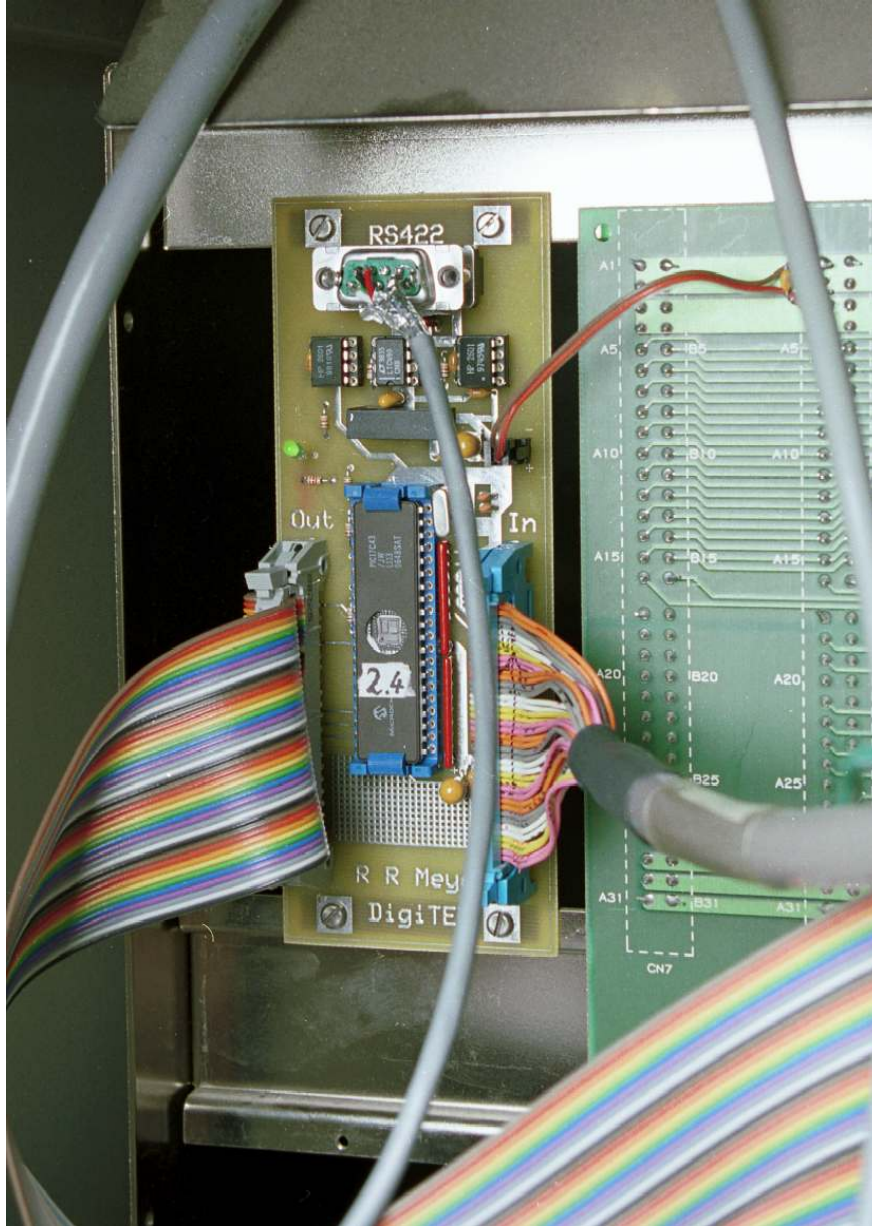
Figure 7.2: A DigiTEM board installed in the deflector unit of a JEOL JEM4000 microscope. The standard cable from the MC to the DEF unit is connected to the DigiTEM input, and a short ribbon cable connects the DigiTEM output to the DEF unit.

# DigiTEM
## Data Flow Diagram

Serial input buffer

Serial receive interrupt

PC command queue

Serial transmit buffer empty interrupt

Serial output buffer

PC message queue

Process command from PC

DAC value memory

MC | PC

Process write instruction from microscope

**Main Loop**

MC queue

Prepare new sample point when scanning

MC write interrupt

Input from MC

Scan timer interrupt

Scan queue

Output to Def unit

Figure 7.3: DigiTEM dataflow diagram. The microcontroller firmware is based on interrupt routines that take prepared data from first in first out buffers (queues) or put received data into queues. These queues are served at the other end by the main loop of the program. In the diagram, the program flow is illustrated with thick arrows, while thin arrows describe the data flow.

In the main program loop, tasks from the different queues are executed in turn, whilst DigiTEM maintains current information for each coil in the internal microcontroller memory. This information includes the mode, the MC value and the offset or PC value, and the PC can read this information at any time. DigiTEM also supports the scanning of up to 4 coils with sinusoidal, triangular or saw-tooth shaped waveforms. Scanning parameters such as the amplitude and the phase offset can be changed from the PC in real time without interrupting the scan, which is important for setting up complicated scan conditions.

In practice, DigiTEM has two basic modes of operation, the passive mode and the active mode. In the passive mode, all commands from the MC are fed through to the deflector or lens unit unaltered and DigiTEM only interprets these commands to update its internal memory of the current microscope state. In the active mode, each deflector coil can be operated in either MC, dual or PC submodes. In the MC submode, the coil current is determined by the microscope only, whereas in the dual submode, the PC can add an offset to the MC value and in the PC submode, the value is given by the PC only and the respective control knob on the microscope has no influenc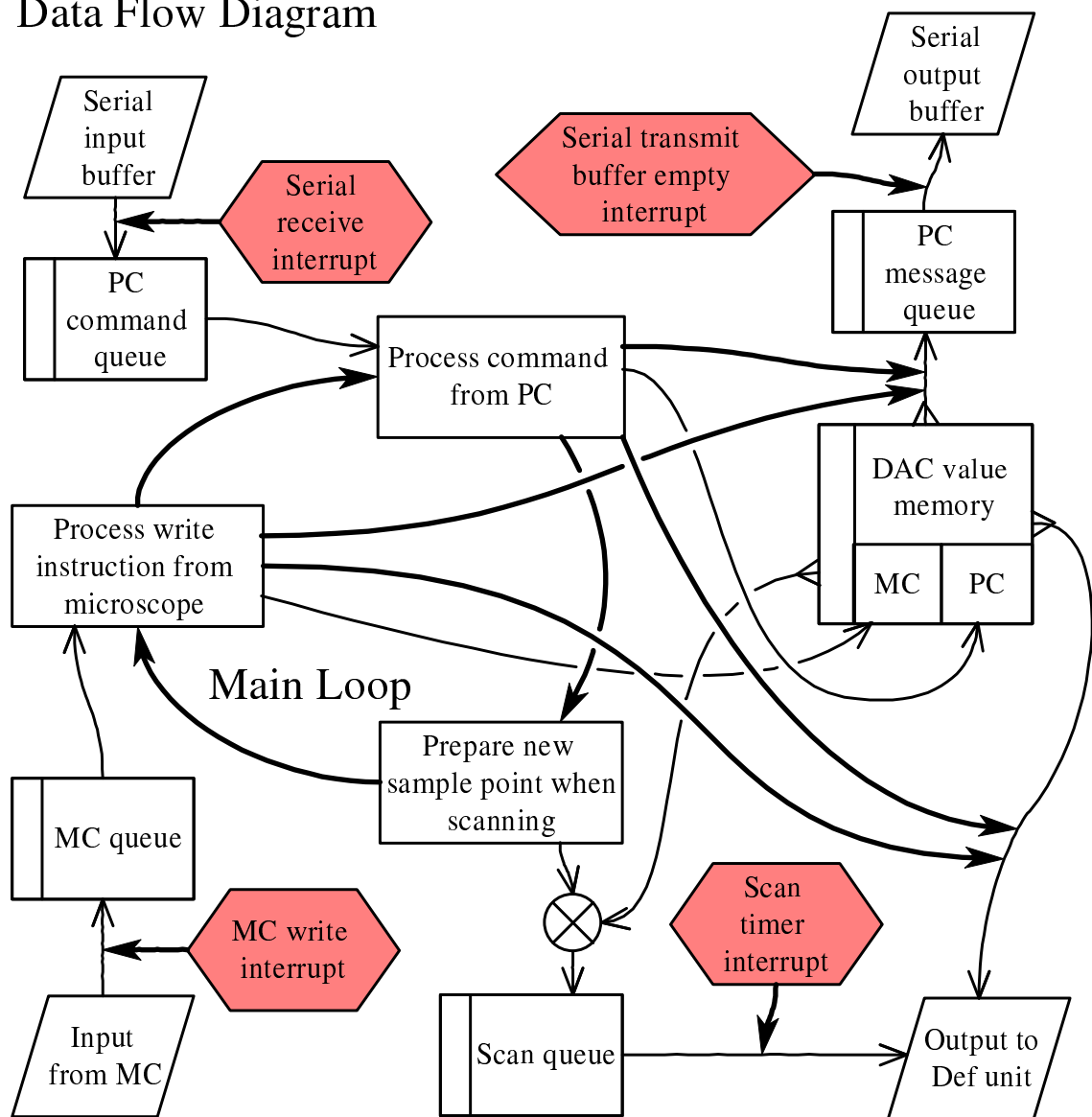e. In active mode, DigiTEM also sends messages to the PC whenever a MC setting is changed, even if this change has no effect because the respective coil is in PC mode. As a failsafe to avoid the microscope being left in a locked state when the external PC is switched off or fails, a timeout mechanism is implemented such that DigiTEM reverts back to passive mode when the PC fails to respond to a special message sent by DigiTEM at one second intervals.

## 7.2 DigiControl software

The Windows based application DigiControl, written in Visual Basic, allows the control of all deflector and lens currents of the microscope using DigiTEM using an external PC. The deflector page, shown in figure 7.4 lists the current digital values for all deflector coils, updated in real time. For each coil pair the mode of operation

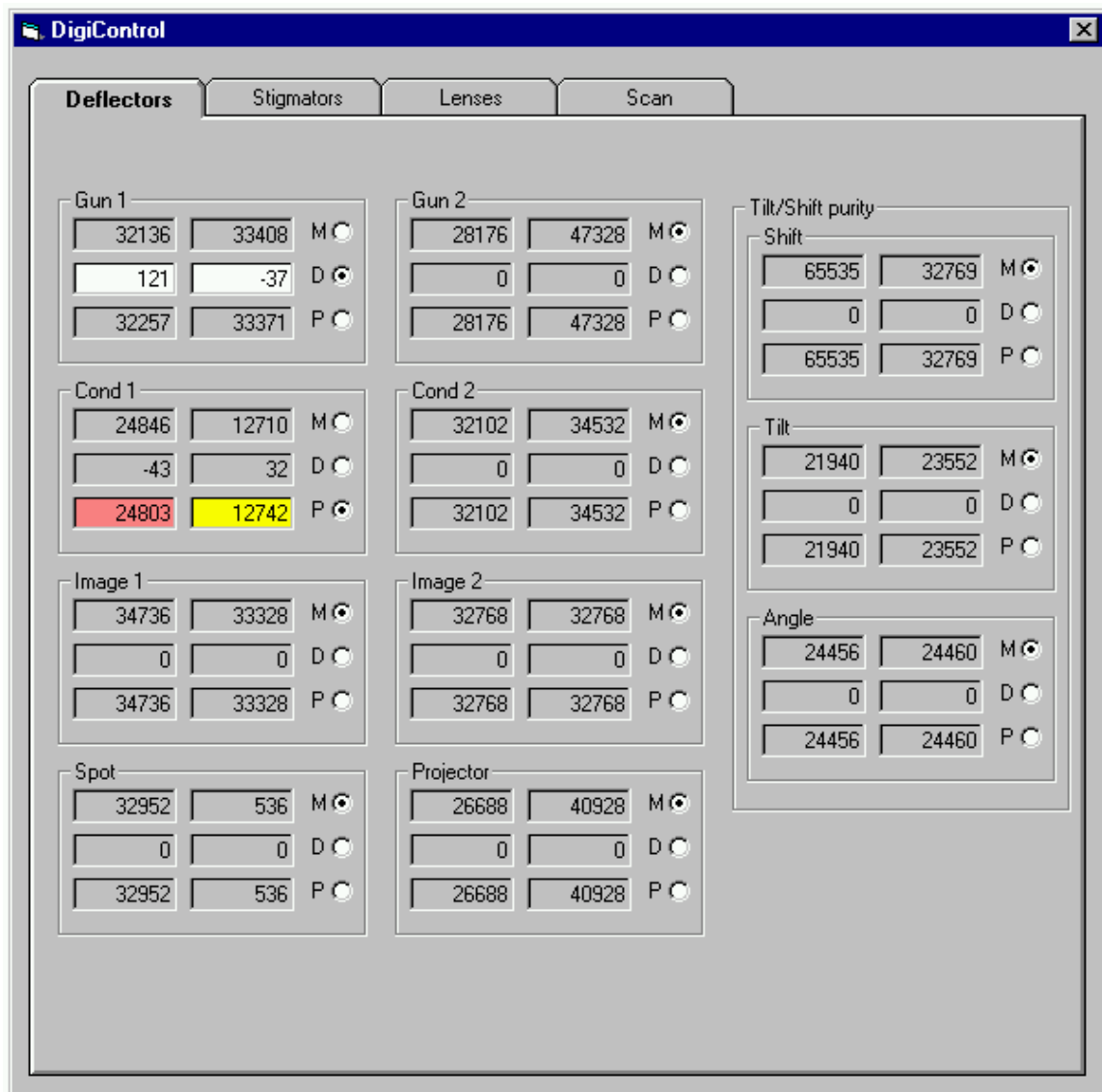Figure 7.4: The deflector page of the DigiControl software. Each coil pair can be operated in MC, dual and PC submodes. In dual mode, the user can specify an offset to the MC value, whereas in PC mode the coil setting is exclusively under PC control. These values can be changed using the mouse wheel after clicking the appropriate field. The numbers in the field change in real time when any setting is altered at the microscope.

can be selected, and in dual and PC mode, the value for the offset or the total coil current can be changed by clicking the appropriate field and using the mouse wheel for adjustment. Left and right mouse clicks then toggle between the $x$ and the $y$ coil. This functionality is also provided for the stigmators and lenses on separate pages.

## 7.3   Scanning features

The scan page of the DigiControl software (figure 7.5) provides an interface to Digi-TEM's scanning features. Up to 4 coils can be scanned simultaneously with a choice of 4 waveforms (figure 7.6). Scanning is performed in discrete steps, where the time between steps is given by $t_{\text{step}} = n_{\text{step}} * 0.2713\mu$s and is set globally for all coils in the field 'step time' as an integer multiple, $n_{\text{step}}$ of the instruction cycle. The waveforms have a period of 1024 steps (sin and saw) and 512 steps (saw up and down). For faster scans, it is possible to advance the phase by $k$ in each scanning step, where $k$ is given in the field 'step' for each scan coil. For slow scans, a number $j > 1$ can be entered in the field 'interval' with the effect that the phase is advanced only every $j$ scanning steps, so that the total period of the scan is given by

$$T = \begin{cases} 1024\, t_{\text{step}}\, j/k & \text{for sin and saw} \\ 512\, t_{\text{step}}\, j/k & \text{for saw up and down.} \end{cases} \tag{7.1}$$

Table 7.1 lists the parameters required for various scanning patterns. The maximum sampling frequency is limited by the speed of the microcontroller, which imposes the limit $n_{\text{step}} \geq 850$ (or $n_{\text{step}} \geq 750$ when only two scan coils are used). For fast scans, $e.g.$ entry (a) in table 7.1 with $f = 50\text{s}^{-1}$, the phase is advanced in larger steps (13 in this example) so that in each scanning period only 79 points around the circle are sampled. However, as 13 has no common factor with 1024, all 1024 points around the circle are sampled after 13 periods.

Tools for the setting up of circular scans are implemented at the bottom of the 'scan' page. The controls on the left side simply group the coils together in pairs, with a phase shift of 256 steps ($\pi/2$) between the $x$ and $y$ directions. The amplitude of
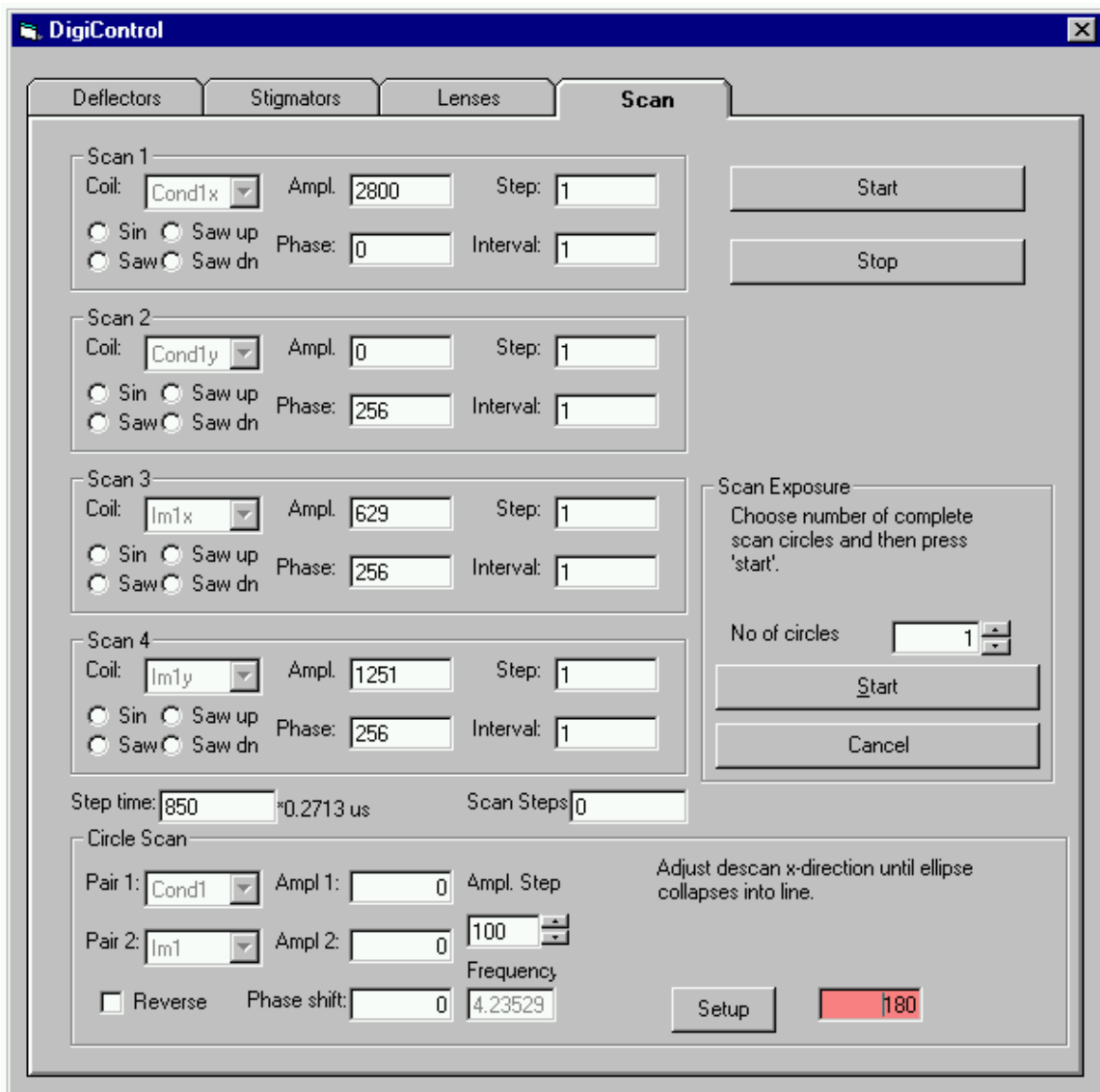
Figure 7.5: The scan page of the DigiControl software. Up to 4 scan coils can be set up individually with various scanning functions. The scan setup feature at the bottom right allows a compensated scan/descan condition to be set up. In this particular screen shot, the user is performing step 4 in table 7.2.
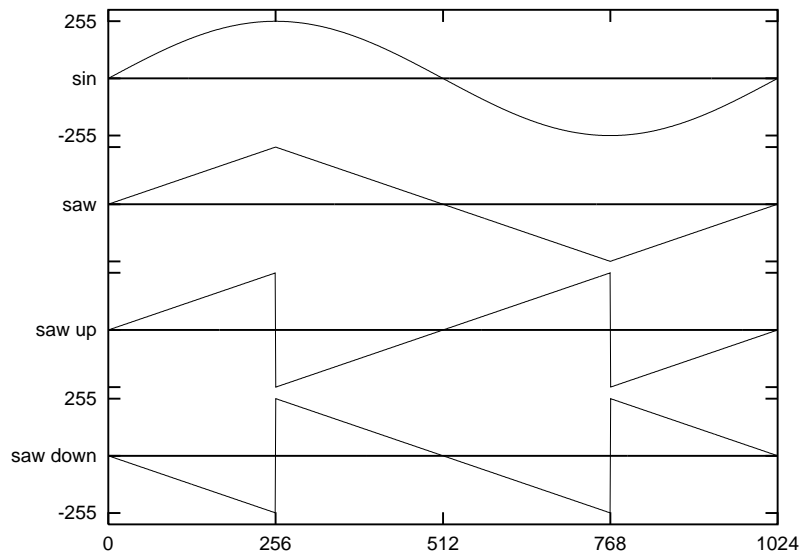
Figure 7.6: Four different scanning waveforms can be selected to facilitate circular and elliptical scans as well as raster scans.

| Scanning pattern | $n_{\text{step}}$ | $x$ | | | | $y$ | | | | $f[s^{-1}]$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Ampl. | Phase | $k$ | $j$ | Ampl. | Phase | $k$ | $j$ | |
| (a) Circular scan | 936 | 1000 | 256 | 13 | 1 | 1000 | 0 | 13 | 1 | 50 |
| (b) Elliptical scan | 900 | 1000 | 128 | 1 | 1 | 1000 | 0 | 1 | 1 | 4 |
| (c) Lissajoux scan | 900 | 1000 | 256 | 5 | 1 | 1000 | 0 | 15 | 1 | 20 |
| (d) Raster scan | 937 | 1000 | 256 | 2 | 1 | 1000 | 256 | 2 | 256 | 0.06 |
| (e) Raster scan | 900 | 1000 | 256 | 8 | 1 | 1000 | 256 | 8 | 64 | 1 |

Table 7.1: Examples for scanning patterns using 'sin' ((a)-(c)) and 'saw up' ((d) and (e)) waveforms. (a) Fast circular scan. (b) Slow elliptical scan with the long axis along the $x = y$ coil direction. (c) When the frequencies on both coils are different a Lissajoux figure is produced (in this example described by $(\cos(\omega t), \sin(3\omega t))$). (d) Raster scan with $256 \times 256$ sampling points. (e) Fast raster scan with $64 \times 64$ sampling points.

each pair and the phase shift between the pairs (*e.g.* to allow for a rotation between a scanning and descanning coil pair) can be adjusted using the wheel mouse while scanning.

Software to facilitate the setting up of compensated scan/descan conditions is also provided. When the 'setup' button at the right hand side of the 'circle scan' frame is clicked, the user is guided through a procedure in which only a single variable has to be adjusted at each stage until an easily recognisable condition in the visible scanning pattern is met (table 7.2). This is a useful simplification since for exact compensation of the scanning by the descanning coil pair it is not only necessary to match the amplitudes of the two pairs and compensate for the rotation between them by a phase shift, but it is also necessary to take into account that the coils of each pair are generally neither of equal strength nor exactly orthogonal. The setting up procedure is based on first applying linear (one-dimensional) scanning patterns with a 256 step phase shift to scanning and descanning coils. This leads to an elliptical scanning pattern unless the directions of the linear scan and descan are parallel, in which case the ellipse collapses into a line. This is used in step 3 to find the direction of the $x$ scanning axis with respect to the descanning coils and again in step 5 for the $y$ scanning axis. Once the direction is found, matching the amplitudes is simply achieved by varying the descanning amplitude with fixed (now known) direction and zero phase shift until the linear scanning pattern collapses into a point (steps 4 and 6).

| What the user needs to do | Variable changed | Pair 1 $x$ | Pair 1 $y$ | Pair 2 $x$ | Pair 2 $y$ |
|---|---|---|---|---|---|
| 1. Adjust scan amplitude | $A$ | $A\sin(\omega t)$ | $A\cos(\omega t)$ | 0 | 0 |
| 2. Adjust scan ellipticity to yield circular scan | $B$ from $A$ | $A\sin(\omega t)$ | $B\cos(\omega t)$ | 0 | 0 |
| 3. Adjust descan amplitude for same deflection as scan amplitude | $C$(only coarse) | 0 | 0 | $C\sin(\omega t)$ | $C\cos(\omega t)$ |
| 4. Adjust descan $x$-direction until ellipse collapses to line | $p$ | $A\sin\omega t$ | 0 | $C\cos p\cos\omega t$ | $C\sin p\cos\omega t$ |
| 5. Adjust descan $x$-amplitude until line collapses to point | $C$ from 0 (positive or negative) | $A\sin\omega t$ | 0 | $C\cos p\sin\omega t$ | $C\sin p\sin\omega t$ |
| 6. Adjust descan $y$-direction until ellipse collapses to line | $q$ from $p+\pi/2$ | 0 | $B\sin\omega t$ | $C\cos q\cos\omega t$ | $C\sin q\cos\omega t$ |
| 7. Adjust descan $y$-amplitude until line collapses to point | $D$ from 0 (positive or negative) | 0 | $B\sin\omega t$ | $D\cos q\sin\omega t$ | $D\sin q\sin\omega t$ |
| 8. Finished | none | $A\sin\omega t$ | $B\cos\omega t$ | $C\cos p\sin\omega t+$ $D\cos q\sin\omega t$ | $C\sin p\sin\omega t+$ $D\sin q\sin\omega t$ |

Table 7.2: The manual adjustment procedure through which the user is guided in the 'scan adjust' feature of DigiControl. This feature greatly simplifies the setting up of a compensated scan/descan condition, as at each stage of the adjustment, only a single parameter is altered using the wheel mouse until the visible scanning pattern fulfills an easily recognisable criterion.

# Chapter 8

# Conclusions and future work

## 8.1 Conclusions

The work in this thesis can be broadly divided into three areas and the main results from within each area are summarised in this chapter.

### 8.1.1 Characterisation of CCD cameras

- It has been demonstrated that it is critical to distinguish between signal and noise transfer for CCD cameras (2.1.3).

- The noise method that has frequently been used in the past to measure the Modulation Transfer Function (MTF) was found to be inadequate for this purpose, as it in fact measures the Noise Transfer Function NTF (2.1.3).

- A mathematical framework that allows calculating both transfer functions using Monte-Carlo simulations has been developed (2.2).

- The signal transfer depends on the area illuminated by many electrons with the same incident point, while the noise transfer depends on the average area illuminated by a single electron (2.2.2).

- A detailed Monte-Carlo model for the conversion process in the scintillator has been developed, including elastic and inelastic electron scattering, fast secondary electron generation and photon scattering, reflection and absorbtion (2.3).

- A method has been developed for the accurate experimental measurement of the MTF using images of a sharp edge, which is skewed with respect to the pixel columns so that an oversampled edge profile can be extracted to avoid aliasing artefacts (2.4).

- Without oversampling, aliasing can lead to a significant overestimation of the MTF measured using the edge method (2.4.2).

- The noise method can be used to measure the NTF and an approach to aliasing correction has been developed (2.5).

- The MTF, NTF and DQE have been measured for a number of commercially available cameras at electron energies in the range 100-400 keV (2.6).

- A strong discrepancy between MTF and NTF has been found both experimentally and in simulations. At high voltages, the high frequency MTF was up to a factor of 4 times lower than the NTF (2.6.3).

- The DQE was found to be strongly spatial frequency dependent and often drops below 10% at high frequencies (2.6.3, 2.6.4).

- The deterioration of the DQE is most pronounced at voltages above 200 kV (high voltage problem). This is due to the increasing lateral range of electrons that re-enter the scintillator after being back-scattered from the support layer (22, 2.6.3).

- As a simple figure of merit, an effective pixel number has been introduced (20). At high voltages, this was found to be typically only 10% of the actual pixel number (2.6).

- For YAG scintillators, the results of Monte-Carlo simulation were found to be in good agreement with the experimental results (2.6.1).

- Due to the photon scattering and absorbtion, phosphor powder scintillators are more difficult to simulate, and the agreement between experiment and simulation is less perfect (2.6.3).

### 8.1.2   Automated aberration determination

- Existing methods for automated aberration determination have been reviewed (3).

- The accuracy of the tilt induced shift method at high resolution suffers from specimen drift (3.2).

- Automated diffractogram fitting requires extended regions of amorphous material and the field of view must be free of crystalline material (3.3.1).

- The "DiffProp" method, which exploits the shift of crystalline reflections when a holographically restored image wave is defocused in diffraction space, only works for small crystallites (3.6).

- A new method for determining the symmetric wave aberrations from a focal series of an arbitrary object has been developed (4).

- The first step in this new method is the accurate registration of the images, including the determination of the relative focus levels using Phase Correlation Functions PCF (4.1).

- In the PCF all spatial frequencies are given equal weight, whereas the conventional cross-correlation function XCF is dominated by the strong Fourier components, *i.e.* low frequencies and crystal reflections, leading to a periodic repetition of the cross-correlation peaks for crystalline objects (4.1.1).

- The focus difference between two images is determined by exploiting that the PCF peak hight is maximised when a phase compensation for the correct focus difference is applied (4.1.2).

- The accuracy of the relative focus determination is better than 1 nm (4.3.1).[1]

- The image wave in the plane of a reference image can be restored when at least two images have been registered to this reference image (4.1.3).

- The symmetric aberration coefficients are determined by maximising Phase Contrast Index (PCI) integrated over $\mathbf{k}$. The PCI is a measure for the conjugate asymmetry of the image wave corrected with trial parameters (4.2.1).

- This approach gives equal weight to all pixels in Fourier space, and hence only requires that the object wave is conjugate antisymmetric (weak phase) for a majority of spatial frequencies. The presence of crystalline material has little influence, since the crystalline reflections only cover a small fraction of the Fourier space area (4.2, 5.3).

- Using directionally weighted integration of the PCI, the defocus $C_1$ and the two-fold astigmatism $A_1$ can be determined simultaneously by only varying the trial parameter $C_1$ (4.2.3).

- The process of integrating the PCI over $\mathbf{k}$ for each trial value of $C_1$ can be replaced by 3 integrations and a 1-dimensional Fourier transform, which makes the algorithm very fast (4.2.4).

- The accuracy of the determined parameters $C_1$, $A_1$ is better than 1 nm (4.3.2, 5.3).

- The spherical aberration $C_3$ can also be determined with an accuracy of 0.05 mm (4.2.5).

---

[1]All accuracies quoted in this section have been determined at magnifications between 250 kX and 800 kX on a JEOL 3000F 300 kV microscope with field-emission gun.

- Determining the asymmetric aberrations with the new method requires recording a combined tilt-focus series, *i.e.* a short (3 member) focal series at a number (*e.g.* 6) of tilt azimuth angles (5.1). For each tilt, the apparent defocus and two-fold astigmatism are measured using the new method, and from these values, the complete set of aberration coefficients is determined analytically or using a least squares fit (3.3.2).

- The three-fold astigmatism can be measured with an accuracy of 20 nm (5.2), and the accuracy of the measured beam tilt is 0.02 mrad (5.2.1). Both accuracies are sufficient for a resolution of 11 $nm^{-1}$ (table 1.2).

- The complete aberration determination can be carried out using small sub-regions (*e.g.* 128×128 pixel). This allows the investigation of the position dependent variation in the aberration parameters (5.2).

- For some datasets, a significant unintentional beam convergence was found, resulting in a variation of the beam tilt by up to 2 mrad across the field of view (5.2.1).

## 8.1.3 Experimental examples of automated object wave restoration.

- An application of the new aberration determination method is the restoration of the aberration corrected object wave from a focus or tilt/focus series (6).

- Since the relative focus levels are accurately determined, the restoration is not affected by focus drift or deviations from the nominal focus step size (4.1.3).

- The automated aberration determination allows a reliable restoration even if the microscope is not well aligned during data acquisition (6.2.2).

- Due to oscillations of the phase contrast transfer function, high-resolution images recorded with field emission gun electron microscopes suffer from contrast

delocalisation, which makes interpretation of conventional individual images difficult, especially near grain boundaries or other aperiodic features. This delocalisation is not present in the restored object wave (1.6).

- The restored modulus shows the positions of atomic columns with heavier atoms as sharp and pronounced minima up to relatively large thicknesses (6.1, 6.2).

- The weakly scattering oxygen columns in complex oxides were clearly identifiable in the restored phase (6.1, 6.2.2).

- A restoration of the layered perovskite $Nd_4Srti_5O_{17}$ revealed the structural origin for streaks observed in diffraction patterns of this compound (6.2.4).

- In the restored phase of a carbon nanotube filled with a $3\times3$ crystal of potassium iodide, atomic columns containing only a single potassium atom were visible (6.3.3).

- For the $3\times3$ KI specimen the restored phase was found to be in good agreement with multislice simulations, provided that a specimen vibration of 0.07 nm and a general contrast reduction by 20% were taken into account (6.3.3).

- For a carbon nanotube filled with antimony oxide, the object wave restored from a focal series was used for a simultaneous determination of the tube conformation and the structure of the filling material (6.3.4).

- For the determination of the chirality sense, the orientation of the nanotube with respect to the image plane was determined by observing the change in focus measured from small subregions along the tube using the new aberration determination approach (6.3.5).

## 8.2 Suggestions for future work

### 8.2.1 New digital electron detectors

The work presented in chapter 2 on the characterisation of CCD cameras for electron detection shows that at intermediate voltages, the DQE of current CCD cameras at medium and high spatial frequencies is low, leading to an effective pixel number that is only 12% of the actual pixel number, and lower for very low electron doses. However, it is also shown that this problem is caused mainly by the back-scattering of electrons from the supporting fibre plate into the scintillator and is therefore common to all fibre-optically coupled systems and cannot be solved by optimising the scintillator material or thickness. While the problem can in principle be avoided with lens-coupled cameras using self-supported scintillators (Fan et al., 2000), the low sensitivity of lens-coupled systems (section 2.6.5) also severely reduces the DQE due to the photon shot noise.

In order to overcome these intrinsic problems of indirect electron detection using scintillators, alternative direct solid state detectors will be investigated.

### 8.2.2 Real time aberration correction and restoration

The automated aberration determination developed in this work is fast and robust and therefore ideally suited for real time applications. This requires a frame-transfer CCD camera (section 2.1.1) and a powerful computer to carry out the realtime restoration.

The aim of such a system is the display of modulus and phase of the aberration corrected object wave. The fast restoration scheme presented in figure 4.14 is ideally suited to this purpose, and a restoration comprising the last three images could be continually displayed. The computational effort for each new frame is little more the 3 2-dimensional Fourier transforms (a forward transform of the image, then a reverse transform of the product with the predicted image Fourier transform, and finally a reverse transform to obtain the real space object wave).

Table 8.1 demonstrates that this can be achieved with a $512 \times 512$ pixel field

| Processor | MHz | $t/\mu$s | Reference |
|---|---|---|---|
| SHARC | 40 | 457 | `www.alacron.com/Products/boards/PB_FT2106xPCI.htm` |
| 8 x SHARC | 40 | 72 | `www.alacron.com/Products/boards/PB_FT2106xPCI.htm` |
| DSP-24 | 100 | 21 | `www.valleytech.com/vti_in_the_news.asp` |
| FFT Engine | 128 | 10 | `www.doublebw.com/FFTEngine.html` |
| Pentium II | 300 | 250 | `www.fftw.org/benchfft/results/pii-300-single.html` |
| Pentium 4 | 1500 | 20 | `developer.intel.com/software/products/mkl/mklspecs_new.htm` |

Table 8.1: Time $t$ required for a 1k 1-dimensional complex single precision floating point FFT for various dedicated Fourier processors and 2 standard Pentium processors. The time required for a 2-dimensional $512 \times 512$ pixel Fourier transform is 512 times the quoted time.

of view using a standard Intel Pentium 4 processor with 1.5GHz clock frequency, predicting that this processor computes the 3 Fourier transforms required in 30 ms, hence allowing a frame rate of 25 frames per second.

A practical problem for the implementation of this system is the rapid change of focus required between frames in order to record the focal series. In order to improve this response, the acceleration voltage rather than the objective lens current can be changed to effect the focus change (Kimura et al., 1994).

## 8.2.3 Automated object wave restoration beyond the linear imaging approximation

The restorations presented in this work have been calculated using linear Wiener filters and ignoring the non-linear contributions to the image intensity. While the artifacts in the restoration from $N$ images caused by the neglected non-linear terms is expected to be reduced by $1/\sqrt{N}$ compared to the effects in a single image (Saxton, 1994b), it is likely that the non-linear terms are significant for thick crystalline objects. A non-linear algorithm for restoration from an equally spaced focal series has previously been developed by Coene et al. (1996), however, in order to be usable with the automated aberration determination developed in this work, an extension to the more general case of a tilt/focal series will be necessary.

# Appendix A

# The multislice method

The multislice method is commonly used to compute the electron wave at the exit plane of a specimen with known atomic structure. This approach was first suggested by Cowley and Moodie (1957) and has found widespread use, especially since the re-invention of the Fast Fourier Transform (FFT) algorithm (Cooley and Tukey, 1965; Brigham, 1974) and the general increase in available computing power has made earlier constraints on its accuracy due to limitations in the number of slices or the number of diffracted beams immaterial. More recent accounts of the multislice algorithm can be found in (Goodman and Moodie, 1974; Kirkland, 1998). The derivation presented here is largely taken from (Kirkland, 1998).

The basis of the multislice method is the division of the specimen into a number of thin slices perpendicular to the direction of the incident beam. The effects of the specimen potential (transmission) and of Fresnel diffraction (propagation) are then treated separately for each slice. Each slice used in the simulation must be thin enough to be a weak phase object. All the atoms within the slice (*i.e.* between $z$ and $z + \Delta z$) are compressed into a flat plane at $z$.

An alternative to the multislice method is the Bloch wave approach, first introduced by Bethe (1928) and described in detail for instance in Buseck et al. (1988). Analogous to the Bloch theorem in solid state physics (Ashcroft and Mermin, 1976), the solution of the Schrödinger equation in the periodic crystal potential is written

as a product of a plane wave and a function that has the same periodicity as the crystal. The latter function is then expanded into its Fourier components and the Schrödinger equation reduces to a matrix equation for these Fourier coefficients. For simple crystals, relatively accurate solutions can be obtained using only a few of these Bloch waves, in some cases only two (two beam approximation). This approach provides valuable insight into the working of dynamical electron diffraction and explains phenomena such as thickness fringes. However, for more complex crystals with larger unit cells, the method becomes unpracticable, as a large number of beams has to be used in the calculation and the computation time for the matrix solution scales with $N^3$, where $N$ is the number of beams included.

## A.1   Derivation of the multislice equations

The equations of the multislice algorithm can be derived starting from the Schrödinger equation for the wavefunction $\psi_f$ an electron in the electrostatic potential $V(x, y, z)$ of the specimen:

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 - eV(x, y, z) \right] \psi_f(x, y, z) = E\psi_f(x, y, z) \,, \tag{A.1}$$

where $m$ is the relativistic electron mass of the electron. For the high-energy electrons in the electron beam, the motion is predominantly in $z$-direction, hence it is convenient to separate the full wavefunction $\psi_f$ into a product of the solution of the free Schrödinger equation (with $V \equiv 0$), which is a plane wave propagating in the $z$-direction, and a wavefunction $\psi$ that represents the small effects due to the specimen and varies much more slowly with $z$:

$$\psi_f(x, y, z) = \mathrm{e}^{2\pi \mathrm{i} k z} \cdot \psi(x, y, z) \,, \tag{A.2}$$

where $k = 1/\lambda$ is the inverse wavelength of the free electron. Substituting this into equation A.1 and using $E = h^2 k^2/2m$ yields[1]

$$-\frac{\hbar^2}{2m}\left[\nabla_{xy}^2 + \frac{\partial^2}{\partial z^2} + 4\pi\mathrm{i}k\frac{\partial}{\partial z} + \frac{2meV(x,y,z)}{\hbar^2}\right]\psi(x,y,z) = 0 \ . \tag{A.3}$$

In an elastic scattering process away from the $z$-direction, $k_x$ and $k_y$ are proportional to the scattering angle, while the change $\Delta k_z$ is proportional to its square. For small scattering angles therefore the term $\left|\frac{\partial^2 \psi}{\partial z^2}\right|$ is much smaller than $\ll \left|\nabla_{xy}^2\psi\right|$ and can be neglected in the paraxial approximation. What remains is a first-order differential equation in $z$

$$\frac{\partial\psi(x,y,z)}{\partial z} = [A + B]\psi(x,y,z) \tag{A.4}$$

with the operators

$$A = \frac{\mathrm{i}\lambda}{4\pi}\nabla_{xy}^2 \tag{A.5}$$

$$B = \mathrm{i}\sigma V(x,y,z) \tag{A.6}$$

where the interaction parameter $\sigma$ is defined by

$$\sigma = \frac{2\pi me\lambda}{h^2} \ . \tag{A.7}$$

This differential equation has the formal solution

$$\psi(x,y,z+\Delta z) = \exp\left[\int_z^{z+\Delta z}[A(z') + B(z')]\,\mathrm{d}z'\right]\psi(x,y,z) \ . \tag{A.8}$$

When $\Delta z$ is small, this can be simplified to

$$\psi(x,y,z+\Delta z) = \exp\left[\frac{\mathrm{i}\lambda}{4\pi}\Delta z\nabla_{xy}^2 + \mathrm{i}\sigma V_{\Delta z}(x,y,z)\right]\psi(x,y,z) \ , \tag{A.9}$$

where $V_{\Delta z}$ is the projected specimen potential between $z$ and $z + \Delta z$:

$$V_{\Delta z}(x,y,z) = \int_z^{z+\Delta z} V(x,y,z')\mathrm{d}z' \ . \tag{A.10}$$

---

[1]It should be noted that $k$ is defined as $1/\lambda$ rather than $2\pi/\lambda$ and therefore $h$ rather than $\hbar$ must be used in the expression for $E$.

The two operators in the exponent in equation A.9 do not commute and therefore this function cannot be re-written as a product of two exponential functions. However, as both operators are small (to the order of $\Delta z$), the approximation

$$e^{\epsilon A + \epsilon B} = e^{\epsilon A} e^{\epsilon B} + \frac{\epsilon^2}{2}[B, A] + \mathcal{O}(\epsilon^3) = e^{\epsilon A} e^{\epsilon B} + \mathcal{O}(\epsilon^2) \tag{A.11}$$

can be used. This leaves the expression

$$\psi(x, y, z + \Delta z) = \exp\left(\frac{i\lambda\Delta z}{4\pi}\nabla^2_{xy}\right) t(x, y, z)\psi(x, y, z) , \tag{A.12}$$

where $t(x, y, z)$ is the transmission function for the specimen slice between $z$ and $z + \Delta z$:

$$t(x, y, z) = \exp\left(i\sigma V_{\Delta z}(x, y, z)\right) . \tag{A.13}$$

The first operator can be separated into components for $x$ and $y$ coordinates by

$$\exp\left(\frac{i\lambda\Delta z}{4\pi}\nabla^2_{xy}\right) = \exp\left(i\alpha\frac{\partial^2}{\partial x^2}\right) \cdot \exp\left(i\alpha\frac{\partial^2}{\partial y^2}\right) , \tag{A.14}$$

where the abbreviation $\alpha = \lambda\Delta z/4\pi$ is introduced.

Considering the Fourier transform of the operator for the $x$-direction applied to a function $f(x)$:

$$\begin{aligned}
FT\left[\exp\left(i\alpha\frac{\partial^2}{\partial x^2}\right) f(x)\right] &= \int e^{-2\pi ikx} \exp\left(i\alpha\frac{\partial^2}{\partial x^2}\right) f(x)\mathrm{d}x && \text{(A.15)} \\
&= \sum \frac{(i\alpha)^n}{n!} \int e^{-2\pi ikx} f^{(2n)}(x)\mathrm{d}x && \text{(A.16)} \\
&= \sum \frac{(i\alpha)^n}{n!}(-2\pi ik)^{2n} \int e^{-2\pi ikx} f(x)\mathrm{d}x && \text{(A.17)} \\
&= e^{-4i\pi^2 k^2 \alpha} FT[f(x)] && \text{(A.18)}
\end{aligned}$$

The same operation can be repeated for the $y$-direction, yielding

$$FT\left[\exp\left(\frac{i\lambda\Delta z}{4\pi}\nabla^2_{xy}\right) f(x, y)\right] = \exp\left(-i\pi\lambda\Delta z(k_x^2 + k_y^2)\right) FT\left[f(x, y)\right] . \tag{A.19}$$

Using this result the equation for a single multislice step (A.12) can be written as

$$\boxed{\psi(x, y, z + \Delta z) = FT^{-1} \left\{ P(k_x, k_y) FT \left[ t(x, y, z)\psi(x, y, z) \right] \right\}} ,$$
(A.20)

where the real space transmission function $t(x, y, z)$ and the Fourier space propagation function $P(k_x, k_y)$ are given by

$$
\begin{aligned}
t(x, y, z) &= \exp\left(i\sigma V_{\Delta z}(x, y, z)\right) && \text{(A.21)} \\
P(k_x, k_y) &= \exp\left(-i\pi\lambda\Delta z(k_x^2 + k_y^2)\right) . && \text{(A.22)}
\end{aligned}
$$

It should be noted that the propagator has the same form as the wave aberration due to a defocus of $\Delta z$. The algorithm can be implemented very efficiently using fast Fourier transforms, and is particularly suitable for simulations of crystals, as only one unit cell has to be calculated and the periodic boundary conditions necessary for the Fourier transforms are automatically fulfilled. The density of sampling points in the unit cell determines the maximum spatial frequency in the Fourier transform and therefore the maximum number of diffracted beams included in the simulation. To avoid aliasing artifacts, it is necessary to exclude all beams above a limit less than the Nyquist frequency in the propagation step.

## A.2   Limitations and approximations

The multislice method is a very accurate method for calculating the exit plane wavefunction due to elastic scattering of electrons in a known specimen. The approximations made in the above derivation can, under certain conditions limit the accuracy achieved with this method. Therefore the impact of these approximations is briefly discussed here.

1. **Limited number of beams.** The number of beams included in the simulation is limited due to the discrete sampling of the unit cell. Unlike the Bloch wave approach (Bethe, 1928), which gives self-consistent results even with a relatively small number of beams, the multislice method is only applicable when

scattering into beams that are not included in the simulation is negligible. For a given simulation, it can be tested if the number of beams included is sufficient by verifying that the total intensity of the wavefunction does not decrease significantly as the wave propagates through the specimen. A decrease by 5% over the complete specimen is considered acceptable (Kirkland, 1998), but with modern computers it is possible to choose the number of beams sufficiently high to give a loss of intensity smaller than 0.1%.

2. **Paraxial approximation.** The paraxial approximation implies neglecting the term

$$\frac{i\lambda}{4\pi} \frac{\partial^2 \psi}{\partial z^2} \tag{A.23}$$

from the RHS of equation A.4. A wave scattered by an angle $\theta$ towards the $x$-direction can be written as:

$$\psi = e^{2\pi i k(\sin\theta\, x + (\cos\theta - 1)z)} \tag{A.24}$$

with $k = 1/\lambda$, hence

$$\left|\frac{\partial^2 \psi}{\partial z^2}\right| < \pi^2 k^2 \theta_{max}^4 \ . \tag{A.25}$$

Therefore, the error due to the paraxial approximation accumulated over a specimen thickness $t$ can be estimated as

$$\Delta\psi < \frac{\lambda t}{4\pi} \left|\frac{\partial^2 \psi}{\partial z^2}\right|_{max} < \frac{\pi t}{4\lambda} \theta_{max}^4 \ . \tag{A.26}$$

As this term is very small for typical situations in high resolution electron microscopy (*e.g.*: $E = 300$ keV, $\lambda = 2$ pm, $t = 20$ nm, $\theta_{max} = 20$ mrad yields $\Delta\psi = 0.16\%$), the error introduced by the paraxial approximation can generally be neglected.

3. **Weak phase approximation for individual slices.** In principle, the error due to this approximation can be made arbitrarily small by making the slices sufficiently thin. However, as the full potential of each atom is usually projected

into the slice that contains the atom centre, the accuracy is still limited when propagation over the range of the atomic potential has a noticeable effect. In this case, the atomic potential has to be divided across several slices, each containing the potential projected over the slice thickness.

4. **Relativistic effects.** The derivation of the multislice algorithm is based on the Schrödinger equation, whereas a fully relativistic treatment would require the use of the Dirac equation. However, it has been shown (Fujiwara, 1962; Moodie et al., 2001) that, provided relativistically correct expressions are used for both electron mass and wavelength, the expressions derived from the Schrödinger equation are an extremely good approximation those derived from the Dirac equation.

Overall therefore, provided that the number of beams is large enough and the slices sufficiently thin, the exit wave due to elastic scattering can be calculated to high precision.

The most serious limitation in the accuracy of the multislice results, however, is due to the fact that it only takes elastic interaction effects into account (Stobbs and Saxton, 1988).

## A.3 The simulated restored wave

The multislice simulations in this work were conducted for comparison with restorations from focal and tilt series. In order to generate a simulated restored wave for direct comparison with the experimental restored wave, the simulated exit surface wavefunction is first back-propagated to the centre plane of the specimen, since this is the focal plane determined by the automated focal determination. From this simulated wave $\psi_{\mathrm{sim}}$, the simulated restored wave is calculated as

$$\psi_{\mathrm{simrest}} = \mathrm{WTF}(\mathbf{k})\psi_{\mathrm{sim}}(\mathbf{k}) + \mathrm{WTF}_C(\mathbf{k})\psi_{\mathrm{sim}}^*(-\mathbf{k}) \tag{A.27}$$

to take into account the attenuation by the total wave transfer function WTF and the unwanted transfer of the conjugate wave present in the experimental restored wave.

# Appendix B

# The Monte Carlo batch programming language

A batch programming language has been implemented as part of the Monte Carlo simulation program (section 2.3.3), which allows a batch of simulations with different parameters to be conveniently run. The batch program has access to all the parameters defined in the "Init Parameters" dialog box. The results of the simulations are written into a datafile called "MTF file" containing the data listed in table B.1 in columns separated by spaces. Optionally, datafiles describing the electron trajectories, the positions where photons are detected or the pulse height distribution can also be produced. The datafiles can be used to generate plots using standard plotting programs like "gnuplot", which can be downloaded from `www.gnuplot.org`.

The key elements of the batch language are as follows:

**Variables:** Variable names are case-sensitive and may contain letters, numbers and the underscore `"_"`, but must start with a letter. Some variable names (see below) are reserved for commands and constants and must not be used as variables. Variables may be assigned numerical values with the statement

```
variable=expression
```

| column | value |
|--------|-------|
| 1 | Spatial frequency $u$ [1/nm] |
| 2 | $\int \hat{g}_\mu(u,0)\,\mathrm{d}\mu$ |
| 3 | $\int |\hat{g}_\mu(u,0)|^2\,\mathrm{d}\mu$ |
| 4 | $\mathrm{MTFS}(u)$ |
| 5 | $\mathrm{NTFS}(u)$ |
| 6 | $\mathrm{MTFS}(u)$ with photon shot noise subtracted |
| 7 | $\mathrm{DQES}(u)$ |

Table B.1: The columns of data written to the MTF data file.

where `expression` is a numerical expression that may contain variables and the operators `+`, `-`, `*`, `/`, `(`, `)` with their usual meaning. String values can also be assigned to variables using the string assignment operator `$=`. This is mainly used to define the filenames of the output files. String constants must be enclosed in double quotes as in `variable $= "text"`. Strings can be concatenated using the comma operator, and numerical values can be converted into strings using C-type formatting strings. Example:

```
E = 400
partfilename $= "gnuplot\MTF"
filename $= partfilename,$02d E/100,".dat"
```

will assign the string value `"gnuplot\MTF04.dat"` to the variable `filename`. The batch language interpreter was written based on the expression evaluator and variable database in the very elegantly programmed calculator example given by (Stroustrup, 1997).

**Commands:** Commands have uppercase names. Commands are: `FOR`, `WITH`, `LOOP`, `START`, `MTF_COLS`, `MTF_NEW`, and `MTF_APPEND`. In detail, they have the following meanings:

```
FOR variable=expr1,expr2[ , expr3]
```

```
        statements
  LOOP
```

This sets variable to `expr1`, executes the statements, adds `expr3` (default value 1) to the variable and repeats the statements until `variable>expr2` (or `variable<expr2` if `expr3<0`). The expressions are only evaluated once and then used as constants while the loop is repeated.

When the values for a particular parameter are not evenly spaced, the `WITH` statement can be used:

```
  WITH var = expr1[, expr2[, expr3[,...]]]]
      statements
  LOOP
```

This evaluates all the expressions and then repeats the statements with all the evaluated values assigned subsequently to `var` as in the example below.

`START`: Starts the simulation for the current parameters.

`MTF_APPEND`: Switches to append mode. MTF and DQE calculated in subsequent calculations are added to an existing gnuplot data file.

`MTF_NEW`: switches to overwrite mode. If a file with the current MtfFile name already exists, it is overwritten by the results of the next simulation.

`MTF_COLS var1[, var2[, var3[,...]]]`: The values of the listed variables are written as additional columns in the MTF data file. This distinguishes between the results of different simulations concatenated into one file by `MTF_APPEND`. The first added column has number 8. For documentation, the names of these variables are also listed as a comment in the header of the output data file.

Example:

```
  MTF_COLS E
  MTF_NEW
  MtfFile $= "AllEnergies.dat"
  WITH E = 100, 120, 150, 200, 250, 300, 400
```

```
    START
    MTF_APPEND ! Set to append mode after the first simulation
  LOOP
```

Hence, multi-dimensional datasets can be created; in the above example the MTFS, NTFS and DQES are given as functions of spatial frequency and energy. The gnuplot command

```
    plot [][0:1] 'AllEnergies.dat' u 1:($8==150 ? $5 : 1/0) with lines
```

plots the NTFS as function of spatial frequency for electron energy `E=150keV`, while

```
    plot [][0:1] 'AllEnergies.dat' u 8:($1==5 ? $7 :1/0)
```

plots the DQES at a spatial frequency of 5 linepairs/mm as a function of the electron energy. Note that the `with lines` option would not work in the latter case, as points can only be connected with lines when they are defined in consecutive rows in the datafile.

**Reserved Variables:**  Some variable names are reserved and allow access to the simulation parameters. The incident electron energy E used in the previous example is one of these reserved variables. They are listed in table B.2. A non-integer value assigned to an integer variables is rounded to the nearest integer.

All element and material names that can be found in the list boxes for choosing the materials in the parameters dialog box (including Define1, Define2, Define3) are reserved integer constants which give each material a unique index number.

The syntax of the batch job file is checked in an initial first pass before any simulations are performed. This allows the debugging of a batch file quickly prior to subsequent unattended running.

The choice of the output produced must be made manually in the "view" menu. For batch processing this means in most cases that all trajectory and photon plotting is disabled (as this is very time consuming) and all MTF and DQE plotting (including "GnuPlot of MTFs") is enabled.

| Variable | Description |
|---|---|
| Floating point reserved variables: | |
| Aperture | Aperture semiangle of the optical system [degrees]. |
| AperWidth | Width of the soft edge for the aperture |
| E | Incident electron energy [keV] |
| c | Conversion factor (fraction of energy converted to photons |
| EPhot | Mean photon energy |
| IncAngle | Incident angle of the electron beam, in degrees |
| TopRefl | Photon reflection ratio on the top scintillator surface |
| BotRefl | Photon reflection ratio on the bottom scintillator surface |
| Focus | Distance upper surface of the first layer to the optical focus plane. |
| DFrac | Number ratio between the photons reaching the CCD indirecly/directly |
| DRad | mean lateral displacement of indirect photons |
| KMax | maximum k-value (in 1/mm) for MTF plots |
| KSamp | sampling distance in k-space (in 1/mm) |
| TCoat | Thickness of the coating layer |
| TScint | Thickness of the scintillator |
| TSupp | Thickness of the support layer |
| Z1 | Mean atomic number of material Define1 |
| ZS1 | Mean square atomic number for material Define1 |
| A1 | Mean atomic weight of material Define1 |
| Ro1 | Density of material Define1 |
| Z2,ZS2,A2,Ro2 | Z,ZS A, Ro for material Define2 |
| Z3,ZS3,A3,Ro3 | Z,ZS A, Ro for material Define3 |
| Integer reserved variables: | |
| N | Number of trajectories |
| MCoat | Material of the coating layer (*e.g.* MCoat = Al) |
| MScint | Material of scintillator (*e.g.* MScint = YAG) |
| MSupp | Material of the support layer, (*e.g.* MSupp = Define1) |
| String reserved variables: | |
| MtfFile | Filename of the MTF datafile |
| TrajFile | Filename of the trajectory datafile |
| PhotFile | Filename of the photon datafile |
| PulseFile | Filename of the pulse height distribution datafile |

Table B.2: Reserved variables in the batch programming language. Typical values for these parameters can be found in the table on page 90.

# Bibliography

ABBE E. (1873) *Beiträge zur Theorie des Mikrokops und der mikroskopischen Wahrnehmung*. Schultze, Archiv f. mikrosk. Anatomie **9**, 413–480.

AJAYAN P.M., EBBESEN T.W., ICHIHASHI T., IIJIMA S., TANIGAKI K. and HIURA H. (1993) *Opening carbon nanotubes with oxygen and implications for filling*. Nature **362**(6420), 522–525.

AJAYAN P.M. and IIJIMA S. (1993) *Capillarity-induced filling of carbon nanotubes*. Nature **361**(6410), 333–334.

AMELINCKX S., LUCAS A. and LAMBIN P. (1999) *Electron diffraction and microscopy of nanotubes*. Reports On Progress in Physics **62**(11), 1471–1524.

AMOW G. and GREEDAN J. (1996) *The structural and magnetic properties of* $Nd_{1-x}TiO_3$ *for x = 0, 0.05, and 0.10*. J. Solid State Chem. **121**, 443–450.

ASHCROFT N. and MERMIN N. (1976) *Solid State Physics*. Saunders College Publishing, Fort Worth.

BABA N., OHO E. and KANAYA K. (1987) *An algorithm for online digital image-processing for assisting automatic focusing and astigmatism correction in electron-microscopy*. Scanning Microscopy **1**(4), 1507–1514.

BEDNORZ J.G. and MÜLLER K.A. (1986) *Possible high-Tc superconductivity in the* Ba-La-Cu-O *system*. Zeitschrift Fur Physik B-Condensed Matter **64**(2), 189–193.

OP DE BEECK M., VAN DYCK D. and COENE W. (1996) *Wave function reconstruction in HRTEM: The parabola method*. Ultramicroscopy **64**(1-4), 167–183.

BETHE H. (1928) *Theorie der Beugung von Elektronen an Kristallen*. Annalen der Physik **87**, 55–129.

BETHE H. (1930) *Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie*. Ann. Physik **5**, 325.

BIRKHOFF R.D. (1958) *The passage of fast electrons through matter*. In *Corpuscles and radiation in matter II*, edited by FLÜGGE S., vol. 34 of *Encyclopedia of Physics*, pp. 49–131, Springer-Verlag, Berlin.

BISHOP H.E. (1976) *The history and development of Monte Carlo methods for use in X-ray microanalysis*. In *Use of Monte Carlo calculations in electron probe microanalysis and scanning electron microscopy*, edited by HEINRICH K., NEWBURY D.E. and YAKOWITZ H., vol. 460 of *NBS special publication*, pp. 5–13, National Bureau of Standards.

BOOTHROYD C.B. (2000) *Quantification of high-resolution electron microscope images of amorphous carbon*. Ultramicroscopy **83**(3-4), 159–168.

BRIGHAM E.O. (1974) *The Fast Fourier Transform*. Prentice Hall, Englewood Cliffs, NJ.

BUDINGER T. and GLASER R. (1976) *Measurement of focus and spherical aberration of an electron microscope objective lens*. Ultramicroscopy **2**, 31–41.

BUSECK P., COWLEY J. and EYRING L. (1988) *High-Resolution Transmission Electron Microscopy*. Oxford University Press, Oxford.

CHAND G. (1997) *Aberration determination and compensation in high resolution transmission electron microscopy*. PhD, University of Cambridge.

CHAND G., SAXTON W.O. and KIRKLAND A.I. (1995) *Aberration measurement and automated alignment of the TEM*. In *Electron Microscopy and Analysis 1995*, edited by CHERNS D., vol. 147 of *Institute of Physics Conference Series*, pp. 297–300, Birmingham.

CHAU K. (1993) *Automated control in high resolution electron microscopy*. PhD, University of Cambridge.

COENE W., JANSSEN G., OP DE BEECK M. and VAN DYCK D. (1992) *Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy*. Phys. Rev. Letts. **69**, 3743–3747.

Coene W.M.J. and Denteneer T.J.J. (1991) *Improved methods for the determination of the spherical- aberration coefficient in high-resolution electron-microscopy from micrographs of an amorphous object*. Ultramicroscopy **38**(3-4), 225–233.

Coene W.M.J., Thust A., Op de Beeck M. and van Dyck D. (1996) *Maximum-likelihood method for focus-variation image reconstruction in high resolution transmission electron microscopy*. Ultramicroscopy **64**(1-4), 109–135.

Connolly E., Sloan J. and Tilley R.J.D. (1996) *Perovskite related phases in the $Nd_4Ti_4O_{14}$-$NdTiO_3$ system*. European Journal of Solid State and Inorganic Chemistry **33**(4), 371–383.

Cooley J.W. and Tukey J.W. (1965) *An algorithm for the machine calculation of complex fourier series*. Mathematics of Computation **19**, 297–301.

Cowley J.M. and Moodie A.F. (1957) *The scattering of electrons by atoms and crystals. I. A new theoretical approach*. Acta Cryst. **10**, 609.

Cunningham I.A. and Reid B.K. (1992) *Signal and noise in modulation transfer function determination using the slit, wire and edge techniques*. Med. Phys **19**, 1037–1044.

Daberkow I., Herrmann K.H., Liu L. and Rau W.D. (1991) *Performance of electron image converters with YAG single crystal screen and CCD sensor*. Ultramicroscopy **38**, 215–223.

Daberkow I., Herrmann K.H., Liu L., Rau W.D. and Tietz H. (1996) *Development and performance of a fast fibre-plate coupled CCD camera at medium energy and image processing system for electron holography*. Ultramicroscopy **64**(1-4), 35–48.

Downing K.H. and Hendrickson F.M. (1999) *Performance of a 2k CCD camera designed for electron crystallography at 400 kV*. Ultramicroscopy **75**, 215–233.

Dunin-Borkowski R. (1999) *personal communication*.

van Dyck D., Op de Beeck M. and Coene W. (1993) *A new approach to object wave-function reconstruction in electron-microscopy*. Optik **93**(3), 103–107.

Elliot S. (1998) *The Physics and Chemistry of Solids*. Wiley, Chichester.

ERASMUS S. and SMITH K. (1982) *An automatic focusing and astigmatism corrrection system for the SEM and CTEM*. J. Microsc. **127**, 185–199.

FAN G. and KRIVANEK O. (1990) *Computer-controlled HREM alignment using automated diffractogram analysis*. In *12th ICEM*, edited by PEACHEY L. and WILLIAMS D., vol. 1, pp. 332–333, San Francisco Press, Seattle.

FAN G.Y., PELTIER S., LAMONT S., DUNKELBERGER D.G., BURKE B.E. and ELLISMAN M.H. (2000) *Multiport-readout frame-transfer 5 megapixel CCD imaging system for TEM applications*. Ultramicroscopy **84**(1-2), 75–84.

FARUQI A. and SUBRAMANIAM S. (2000) *CCD detectors in high-resolution biological electron microscopy*. Quaterly Rev Biophysics **33**(1), 1–27.

FARUQI A.R., HENDERSON R. and SUBRAMANIAM S. (1999) *Cooled CCD detector with tapered fibre optics for recording electron diffraction patterns*. Ultramicroscopy **75**, 235–250.

FIEBINGER J.R. and MÜLLER R. (1972) *Pair-production energies in silicon and germanium bombarded with low energy electrons*. J. Appl. Phys. **43**, 3202.

FOSCHEPOTH M. and KOHL H. (1998) *Amplitude contrast - a way to obtain directly interpretable high-resolution images in a spherical aberration corrected transmission electron microscope*. Phys. Status Solidi A-Appl. Res. **166**(1), 357–366.

FOWLER R. and NORDHEIM L. (1928) *Electron emission in intense electric fields*. Proc. R. Soc. London **A 119**, 173.

FRIEDRICHS S., SLOAN J., HUTCHISON J.L., GREEN M.L.H., MEYER R.R. and KIRKLAND A.I. (2001) *Simultaneous determination of inclusion crystallography and nanotube conformation for a* $Sb_2O_3$/*single walled nanotube composite*. Phys. Rev. B **64**, 45 406–45 413.

FU Q. and LICHTE H. (1995) *Holographic measurement of the wave-aberration of an electron- microscope by means of the phases in the fourier spectrum*. J. Microsc. **179**(2), 112–118.

FU Q., LICHTE H. and VÖLKL E. (1991) *Correction of aberrations of an electron microscope by means of electron holography*. Phys. Rev. Letters **67**, 2319.

Fujiwara K. (1962) *Relativistic dynamical theory of electron diffraction*. J. Phys. Soc. Jpn. **17**(Suppl. B2), 118–123.

Gabor D. (1948) *A new microscope principle*. Nature **161**, 777–778.

Galy J. and Carpy A. (1974) *Defects and diffusion paths in perovskite-like structures $AB O_{3+x}$*. Phil. Mag. **29**, 1207–1211.

Gauvin R. and L'Esprance G. (1992) *A monte carlo code to simulate the effect of fast secondary electrons on $k_{AB}$ factors and spatial resolution in the TEM*. J. Microc. **168**(pt 2), 153–167.

Gibson J.M. (1994) *Breakdown of the weak-phase object approximation in amorphous objects and measurement of high-resolution electron optical parameters*. Ultramicroscopy **56**, 26–31.

Glaser W. (1943) *Bildentstehung und Auflösungsvermögen des Elektronenmikroskops vom Standpunkt der Wellenmechanik*. Zeits. f. Physik **121**, 647–666.

Goodman P. and Moodie A.F. (1974) *Numerical evaluation of n-beam wave functions in electron scattering by the multislice method*. Acta Cryst. **A30**, 280.

Haider M. (2001) *personal communication*.

Haider M., Rose H., Uhlemann S., Schwan E., Kabius B. and Urban K. (1998) *A spherical-aberration-corrected 200 kV transmission electron microscope*. Ultramicroscopy **75**(1), 53–60.

van Heel M., Schatz M. and Orlova E. (1992) *Correlation-functions revisited*. Ultramicroscopy **46**, 307–316.

Herrmann K.H. and Krahl D. (1982) *The detection quantum efficiency of electronic image recording systems*. J. Microsc. **127**, 17–28.

Honda T., Tomita T., Kaneyama T. and Ishida Y. (1994) *Field-emission ultrahigh-resolution analytical electron microscope*. Ultramicroscopy **54**(2-4), 132–144.

Horner J.L. and Gianino P.D. (1984) *Phase-only matched filtering*. Applied Optics **23**(6), 812–816.

274

Howell S.B. (2000) *Handbook of CCD astronomy*. Cambridge observing handbooks for research astronomers; 2, Cambridge University Press, Cambridge.

Hülk C. and Daberkow I. (1998) *Easy methods for accurate characterization of CCD cameras*. Proc. 14th ICEM, Cancun **1**, 189.

Hutchison J.L., Doole R.C., Dunin-Borkowski R.E., Sloan J. and Green M.L.H. (1999) *The development and assessment of a high performance field-emission-gun analytical HREM for materials science applications*. JEOL News **34E**, 10.

Ichise N., Baba N. and Nagashima H. (1997) *The phase-spectrum based measurement of the tem parameters*. Ultramicroscopy **68**, 181–200.

Iijima S. (1991) *Helical microtubules of graphitic carbon*. Nature **354**(6348), 56–58.

Iijima S. and Ichihashi T. (1993) *Single-shell carbon nanotubes of 1nm diameter*. Nature **363**(6430), 603–605.

Ishizuka K. (1993) *Optimized sampling schemes for off-axis holography*. Ultramicroscopy **52**, 1–5.

Jerri A. (1977) *The Shannon sampling theorem – Its various extensions and applications: A tutorial review*. Proc. IEEE **65**(11), 1565–1598.

Jin S., Tiefel T.H., McCormack M., Fastnacht R.A., Ramesh R. and Chen L.H. (1994) *Thousandfold change in resistivity in magnetoresistive* La-Ca-Mn-O *films*. Science **264**(5157), 413–415.

Journet C. and Bernier P. (1998) *Production of carbon nanotubes*. Applied Physics A-Materials Science and Processing **67**(1), 1–9.

Journet C., Maser W.K., Bernier P., Loiseau A., de la Chapelle M.L., Lefrant S., Deniard P., Lee R. and Fischer J.E. (1997) *Large-scale production of single-walled carbon nanotubes by the electric-arc technique*. Nature **388**(6644), 756–758.

Joy D.C. (1995) *Monte Carlo Modeling for Electron Microscopy and Microanalysis*. University Press, Oxford.

JOY D.C., NEWBURY D.E. and MYKLEBUST R.L. (1982) *The role of fast secondary electrons in degrading spatial resolution in the analytical electron microscope*. J. Microsc. **128**(pt 2), RP1–RP2.

KIANG C.H., CHOI J.S., TRAN T.T. and BACHER A.D. (1999) *Molecular nanowires of 1 nm diameter from capillary filling of single-walled carbon nanotubes*. Journal of Physical Chemistry B **103**(35), 7449–7451.

KIMURA Y., TANIGUCHI Y., ANDO T., TAKAI Y., SHIMIZU R. and IKUTA T. (1994) *A practical method for modulation of accelerating voltage for active modulation image-processing electron-microscopy*. Optik **95**(4), 185–186.

KIRKLAND A.I. and SAXTON W.O. (2002) *Cation segregation in* $Nb_{16}W_{18}O_{94}$ *using high angle annular dark field scanning transmission electron microscopy image processing*. J. Microsc. **206**, 1–6.

KIRKLAND A.I., SAXTON W.O. and CHAND G. (1997) *Multiple beam tilt microscopy for super resolved imaging*. J. Electron Microsc. **1**, 11–22.

KIRKLAND A.I., SAXTON W.O., CHAU K.L., TSUNO K. and KAWASAKI M. (1995) *Super resolution by aperture synthesis - Tilt series reconstruction in CTEM*. Ultramicroscopy **57**(4), 355–374.

KIRKLAND E.J. (1984) *Improved high resolution image processing of bright field electron micrographs*. Ultramicroscopy **15**, 151–172.

KIRKLAND E.J. (1998) *Advanced computing in electron microscopy*. Plenum Press, New York.

KOECK P.B. (1998) *Superresolution by aliasing correction of undersampled crystal images*. In *50th annual meeting of the scandinavian society for electron microscopy*, vol. 101.

KOSTER A.J., VAN DEN BOS A. and VAN DER MAST K.D. (1987) *An autofocus method for a TEM*. Ultramicroscopy **21**, 209–222.

KOSTER A.J., CHEN H., SEDAT J.W. and AGARD D.A. (1992) *Automated microscopy for electron tomography*. Ultramicroscopy **46**(1-4), 207–227.

276

Koster A.J., de Ruijter W.J., van den Bos A. and van der Mast K.D. (1989) *Autotuning of a tem using minimum electron dose*. Ultramicroscopy **27**, 251–272.

Krivanek O.L. (1976) *A method for determining the coefficient of spherical aberration from a single micrgraph*. Optik **45**, 97–101.

Krivanek O.L. and Fan G.Y. (1992) *Application of slow-scan charge-coupled device (CCD) cameras to on-line microscope control*. Scanning Microscopy Supplement **6**, 105–114.

Krivanek O.L. and Leber M.L. (1993) *Three-fold astigmatsm: an important TEM aberration*. In *51st MSA meeting*, edited by Bailey G.W. and Rieder C.L., pp. 972–973, San Francisco Press, Cincinatti, Ohio.

Krivanek O.L. and Leber M.L. (1994) *Autotuning for 1Å resolution*. In *13th ICEM*, p. 157, Paris.

Krivanek O.L. and Mooney R.E. (1993) *Applications of slow-scan CCD cameras in transmission electron microscopy*. Ultramicroscopy **49**, 95–108.

Kuglin C.D. and Hines D.C. (1975) *The phase correlation image alignment method*. Proc. IEEE Int. Conf. on Cybernetics and Society p. 163.

Kujawa S. and Krahl D. (1992) *Performance of a low-noise CCD camera adapted to a transmission electron-microscope*. Ultramicroscopy **46**(1-4), 395–403.

Kunath W., Zemlin F. and Weiss K. (1985) *Apodization in phase-contrast electron-microscopy realized with hollow-cone illumination*. Ultramicroscopy **16**(2), 123–138.

Kunath W., Zemlin F. and Weiss K. (1987) *Refinement procedures for high-resolution electron microscopy*. Optik **76**, 122–131.

Langley K., Meyer R., Sayagués M., Sloan J., Kirkland A., Hutchison J. and Tilley R. (2002) *Structural characterisation of $Nd_5Ti_5O_{17}$ by HRTEM*. J. Solid State Chem. in preparation.

Lehmann M. (2000) *Determination and correction of the coherent wave aberration from a single off-axis electron hologram by means of a genetic algorithm*. Ultramicroscopy **85**(3), 165–182.

LEHMANN M. (2001) *personal communication.*

LEHMANN M., LICHTE H., GEIGER D., LANG G. and SCHWEDA E. (1999) *Electron holography at atomic dimensions - present state.* Materials Characterization **42**(4-5), 249–263.

LICHTE H. (1991) *Electron image off-axis holography of atomic structures.* In *Advances in Optical and Electron Microscopy*, vol. 12, pp. 25–91, Academic press.

LICHTENBERG F., HERRNBERGER A., WIEDENMANN K. and MANNHART J. (2001) *Synthesis of perovskite-related layered $A_nB_nO_{3n+2}=ABO_X$ type niobates and titianates and study of their structural, electric and magnetic properties.* Progress in Solid State Chemistry **29**, 1–70.

MALAMIDIS D. and LICHTE H. (1998) *Determination of the wave aberration from holograms of crystalline objects.* In *ICEM14*, edited by BENAVIDES H. and YACAMAN M., vol. 1, pp. 561–562, IoP, Cancun.

MEYER R.R. and KIRKLAND A.I. (1998) *The effects of electron and photon scattering on signal and noise transfer properties of scintillators in CCD cameras used for electron detection.* Ultramicroscopy **75**(1), 23–33.

MEYER R.R. and KIRKLAND A.I. (2000) *Characterisation of the signal and noise transfer of CCD cameras for electron detection.* Microscopy Res. and Tech. **49**(3), 269–280.

MEYER R.R., KIRKLAND A.I., DUNIN-BORKOWSKI R.E. and HUTCHISON J.L. (2000a) *Experimental characterisation of CCD cameras for HREM at 300 kV.* Ultramicroscopy **85**(1), 9–13.

MEYER R.R., SLOAN J., DUNIN-BORKOWSKI R.E., KIRKLAND A.I., NOVOTNY M.C., BAILEY S.R., HUTCHISON J.L. and GREEN M.L.H. (2000b) *Discrete atom imaging of one-dimensional crystals formed within single-walled carbon nanotubes.* Science **289**(5483), 1324–1326.

MØLLER C. (1932) *Zur Theorie des Durchgangs schneller Elektronen durch Materie.* Ann. Physik **14**, 531.

278

Moodie A., Cowley J. and Goodman P. (2001) *Dynamical theory of electron diffraction*. In *International Tables of Crystallography*, edited by Shmueli U., vol. B: Reciprocal Space, p. sect. 5.2, Kluwer academic publishers, Dordrecht, 2nd edn.

Mott N.F. (1929) *The scattering of fast electrons by atomic nuclei*. Proc. Roy. Soc. Lond. A **124**, 425–442.

Mott N.F. (1932) *The polarisation of electrons by double scattering*. Proc. Roy. Soc. Lond. A **135**, 429–458.

Nanamatsu S., Kimura M., Doi K., Matsushita S. and Yamada N. (1974) Ferroelectrics **8**, 511.

Nanot M., Queyroux F., Gilles J.C. and Portier R. (1981) $A_nB_nO_{3n+2}$ *phases of the systems* $La_2Ti_2O_7$ - $LaTiO_3$, $Nd_2Ti_2O_7$ - $NdTiO_3$, *and* $Ca_2Ti_2O_7$ - $CaTiO_3$ *.2. New crystallographic characteristics and an interpretation of the twinning revealed by the electron microscope*. Journal of Solid State Chemistry **38**(1), 74–81.

Nellist P.D., McCallum B.C. and Rodenburg J.M. (1995) *Resolution beyond the information limit in transmission electron- microscopy*. Nature **374**(6523), 630–632.

Nellist P.D. and Pennycook S.J. (2000) *The principles and interpretation of annular dark-field Z-contrast imaging*. Advances in Imaging and Electron Physics **113**, 147–203.

Orchowski A., Rau W.D. and Lichte H. (1995) *Electron holography surmounts resolution limit of electron microscopy*. Phys. Rev. Letts. **74**, 399–402.

Otten M.T. and Coene W.M.J. (1993) *High-resolution imaging on a field-emission TEM*. Ultramicroscopy **48**(1-2), 77–91.

Overwijk M.H.F., Bleeker A.J. and Thust A. (1997) *Correction of three-fold astigmatism for ultra-high-resolution TEM*. Ultramicroscopy **67**(1-4), 163–170.

Pennycook S.J. and Boatner L.A. (1988) *Chemically sensitive structure-imaging with a scanning transmission electron-microscope*. Nature **336**(6199), 565–567.

Pennycook S.J. and Jesson D.E. (1990) *High-resolution incoherent imaging of crystals*. Physical Review Letters **64**(8), 938–941.

PENNYCOOK S.J. and JESSON D.E. (1991) *High-resolution Z-contrast imaging of crystals*. Ultramicroscopy **37**(1-4), 14–38.

QIN L.C., ZHAO X.L., HIRAHARA K., MIYAMOTO Y., ANDO Y. and IIJIMA S. (2000) *The smallest carbon nanotube*. Nature **408**(6808), 50–50.

RABBANI M. and VAN METTER R. (1989) *Analysis of signal and noise propagation for several imaging mechanisms*. J. Optical Soc Am. A **6**, 1156–1164.

RABBANI M., SHAW R. and VAN METTER R. (1987) *Detective quantum efficiency of imaging systems with amplifying and scattering mechanisms*. J. Optical Soc. Am. A **4**, 895–901.

RAVEL M. and REINHEIMER A.L. (1991) *Backside-thinned CCDs for keV electron detection*. SPIE **1447**, 109–122.

REIMER L. (1989) *Transmission Electron Microscopy*. Springer, New York, 2nd edn.

RILEY K., HOBSON M. and BENCE S. (1997) *Mathematical methods for physics and engineering*. University Press, Cambridge.

ROBERTS P.T.E., CHAPMAN J.N. and MACLEOD A.M. (1982) *A CCD-based image recording-system for the CTEM*. Ultramicroscopy **8**(4), 385–396.

RODENBURG J.M. and BATES R.H.T. (1992) *The theory of superresolution electron-microscopy via Wigner-distribution deconvolution*. Phil. Trans. R. Soc. Lond. A **339**(1655), 521–553.

RUDDLESDEN S. and POPPER P. (1957) *New compounds of the $K_2NiF_4$ type*. Acta Cryst. **10**, 538–540.

RUDDLESDEN S. and POPPER P. (1958) *The compound $Sr_3Ti_2O_7$ and its structure*. Acta Cryst. **11**, 54–58.

DE RUIJTER W.J. (1995) *Imaging properties and application of slow-scan charge-coupled device cameras suitable for electron microscopy*. Micron **26**, 247–275.

DE RUIJTER W.J. and WEISS J.K. (1992) *Methods to measure properties of slow-scan CCD cameras for electron detection*. Rev Sci Instrum. **63**, 4314–4321.

RUTHERFORD E. (1911) *The scattering of $\alpha$ and $\beta$ particles by matter and the structure of the atom*. Phil. Mag, **21**, 669–688.

SAITO R., DRESSELHAUS G. and DRESSELHAUS M. (1998) *Physical properties of Carbon Nanotubes*. Imperial College Press, London.

SAXTON W. (1978) In *Computer techniques for image processing in electron microscopy*, p. Section 9.7, Academic Press, New York.

SAXTON W. (2000a) *A new way of measuring microscope aberrations*. Ultramicroscopy **81**, 41–45.

SAXTON W.O. (1988) *Accurate atom positions from focal and tilted beam series of high resolution electron micrographs*. Scanning Microscopy Supplement **2**, 213–224.

SAXTON W.O. (1994a) *Accurate alignment of sets of images*. J. Microsc. **174**, 61–68.

SAXTON W.O. (1994b) *What is the focus variation method? Is it new? Is it direct?*. Ultramicroscopy **55**, 171–181.

SAXTON W.O. (1995) *Observation of lens aberrations for very high-resolution electron microscopy. I. Theory*. J. Microsc. **179**, 201–214.

SAXTON W.O. (2000b) *DpFit library for semper, personal communication* .

SAXTON W.O., CHAND G. and KIRKLAND A.I. (1994) *Accurate determination and compensation of lens aberration in high resolution EM*. In *13th ICEM*, edited by JOUFFREY B., COLLIEX C., CHEVALIER J.P., GLAS F., HAWKES P.W., HERNANDEZ D. VERDUN, SCHREVEL J. and THOMAS D., vol. 1, pp. 203–204, Les Ulis Cedex A, Paris.

SAXTON W.O., PITT T.J. and HORNER M. (1979) *Digital image processing: the SEMPER system*. Ultramicroscopy **4**, 343–354.

SAXTON W.O., SMITH D.J. and HORNER M. (1983) *Procedures for focusing, stigmating and alignment in high resolution electron microscopy*. J. Microsc. **130**, 187–201.

SCHERZER O. (1936) *Über einige Fehler von Elektronenlinsen*. Z. Phys. **101**, 593.

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

SCHISKE P. (1968) *Zur Frage der Bildrekonstruktion durch Fokusreihen*. In *4th Eur. Conf on Electron Microscopy*, p. 145, Rome.

SCHISKE P. (1973) *Image processing using additional statistical information about the object*. In *Image processing and computer-aided design in electron optics*, edited by HAWKES P., p. 82, Academic Press, New York.

SHERMAN M.B., BRINK J. and CHIU W. (1996) *Performance of a slow-scan CCD camera for macromolecular imaging in a 400 kV electron cryomicroscope*. Micron **27**, 129–139.

SLEIGHT A.W. (1966) *The crystal structure of $Nb_{16}W_{18}O_{94}$, a member of a $(MeO)_xMeO_3$ family of compounds*. Acta Chem. Scand. **20**, 1102–1112.

SLOAN J., HAMMER J., ZWIEFKA-SIBLEY M. and GREEN M.L.H. (1998) *The opening and filling of single walled carbon nanotubes (SWNTs)*. Chemical Communications (3), 347–348.

SLOAN J. and TILLEY R. (1994) *Layered perovskite phases in the $Nd_2Ti_2O_7$-$SrTiO_3$ system*. Eur. J. Solid State Inorg. Chem. **31**, 673–682.

SMITH D.J. (1997) *The realization of atomic resolution with the electron microscope*. Reports On Progress in Physics **60**(12), 1513– 1580.

SMITH D.J. and BURSILL L.A. (1985) *Metallization of oxide surfaces observed by in situ high- resolution electron-microscopy*. Ultramicroscopy **17**(4), 387–391.

SPENCE J.C.H. and ZUO J.M. (1988) *Large dynamic-range, parallel detection system for electron diffraction and imaging*. Review of Scientific Instruments **59**(9), 2102– 2105.

STEINECKER A. and MADER W. (2000) *Measurement of lens aberrations by means of image displacements in beam-tilt series*. Ultramicroscopy **81**(3-4), 149–161.

STOBBS M. and SAXTON W. (1988) *Quantitative high resolution transmission electron microscopy: the need for energy filtering and the advantages of energy-loss imaging*. J. Microsc. **151**, 171.

STROUSTRUP B. (1997) *The C++ Programming Language*. Addison Wesley, New York, 3rd edn.

SVENSSON C. (1974) *The crystal structure of orthorombic antimony trioxide,* $Sb_2O_3$. Acta Cryst. B **30**, 458–461.

SVENSSON C. (1975) *Refinement of the crystal structure of cubic antimony trioxide,* $Sb_2O_3$. Acta Cryst. B **31**, 2016.

TANG D., ZANDBERGEN H.W., JANSEN J., DE BEECK M.O. and VAN DYCK D. (1996) *Fine-tuning of the focal residue in exit-wave reconstruction*. Ultramicroscopy **64**(1-4), 265–276.

TERRONES M., HSU W.K., SCHILDER A., TERRONES H., GROBERT N., HARE J.P., ZHU Y.Q., SCHWOERER M., PRASSIDES K., KROTO H.W. and WALTON D.R.M. (1998) *Novel nanotubes and encapsulated nanowires*. Appl. Phys. A-Mater. Sci. Process. **66**(3), 307–317.

THOMAS N.W. (1996) *The compositional dependence of octahedral tilting in orthorhombic and tetragonal perovskites*. Acta Crystallographica **B52**, 16–31.

THON F. (1966) *The defocussing of phase contrast electron microscope images*. Z. Naturforsch. **21a**, 476–478.

TONOMURA A. (1992) *Electron holography of magnetic materials and observation of flux-line dynamics*. Ultramicroscopy **47**(419-424).

TSANG S.C., CHEN Y.K., HARRIS P.J.F. and GREEN M.L.H. (1994) *A simple chemical method of opening and filling carbon nanotubes*. Nature **372**(6502), 159–162.

TSANG S.C., HARRIS P.J.F. and GREEN M.L.H. (1993) *Thinning and opening of carbon nanotubes by oxidation using carbon-dioxide*. Nature **362**(6420), 520–522.

TYPKE D. and DIERKSEN K. (1995) *Determination of image aberrations in high-resolution electron- microscopy using diffractogram and cross-correlation methods*. Optik **99**(4), 155–166.

UHLEMANN S. and HAIDER M. (1998) *Residual wave aberrations in the first spherical aberration corrected transmission electron microscope*. Ultramicroscopy **72**(3-4), 109–119.

URBAN K., KABIUS B., HAIDER M. and ROSE H. (1999) *A way to higher resolution: spherical-aberration correction in a 200 kV transmission electron microscope.* Journal of Electron Microscopy **48**(6), 821–826.

VOGEL H. (1995) *Gerthsen Physik.* Springer, Berlin, 18th edn.

WADE R.H. and FRANK J. (1977) *Electron microscope transfer functions for partially coherent axial illumination.* Optik **49**, 81–92.

WEICKENMEIER A.L., NÜCHTER W. and MAYER J. (1995) *Quantitative characterization of point spread function and detection quantum efficiency for a YAG scintillator slow scan CCD camera.* Optik **99**, 147–154.

WILLIAMS T., SCHMALLE H., RELLER A., LICHTENBERG F., WIDMER D. and BEDNORZ G. (1991) *The crystal structures of* $La_2Ti_2O_7$ *and* $La_5Ti_5O_{17}$: *High-resolution electron microscopy.* J. Solid State Chem. **93**, 534–548.

ZEMLIN F. (1979) *A practical procedure for alignment of a high resolution electron microscope.* Ultramicroscopy **4**, 241–245.

ZEMLIN F., WEISS K., SCHISKE P., KUNATH W. and HERRMANN K.H. (1978) *Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms.* Ultramicroscopy **3**, 49–60.

ZUO J.M. (1996) *Electron detection characteristics of slow-scan CCD camera.* Ultramicroscopy **66**, 21–33.

ZUO J.M. (2000) *Electron detection characteristics of a slow-scan CCD camera, imaging plates and film, and electron image restoration.* Microscopy Research and Technique **49**(3), 245–268.

VAN ZWET E.J. and ZANDBERGEN H.W. (1996) *Measurement of the modulation transfer function of a slow-scan CCD camera on a TEM using a thin amorphous film as test signal.* Ultramicroscopy **64**, 49–55.

# Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Diese Arbeit wurde in den Jahren 1997–2002 am Department of Materials Science and Metallurgy der University of Cambridge, England angefertigt. Die wissenschaftlichen Betreuer waren Dr Angus Kirkland und Dr. Owen Saxton, beide Cambridge University, sowie Prof. Dr. Hannes Lichte, Technische Universität Dresden.

Cambridge, im Juni 2002,