Jül-4274

Institute of Solid State Research (IFF)
Microstructure Research (IFF-8)

# *Ultra-Precise Measurement of Optical Aberrations for Sub-Ångström Transmission Electron Microscopy*

*Juri Barthel*

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
FORSCHUNGSZENTRUM

# *Ultra-Precise Measurement of Optical Aberrations for Sub-Ångström Transmission Electron Microscopy*

*Juri Barthel*

**Berichte des Forschungszentrums Jülich; 4274**
ISSN 0944-2952
Institute of Solid State Research (IFF)
Microstructure Research (IFF-8) Jül-4274

D 82 (Diss., RWTH Aachen, 2007)

Vollständig frei verfügbar im Internet auf dem Jülicher Open Access Server (JUWEL)
unter http://www.fz-juelich.de/zb/juwel

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek, Verlag
D-52425 Jülich · Bundesrepublik Deutschland
Z 02461 61-5220 · Telefax: 02461 61-6103 · E-Mail: zb-publikation@fz-juelich.de

**Ultra-Precise Measurement of Optical Aberrations for Sub-Ångström Transmission Electron Microscopy**

Quantitative investigations of material structures on an atomic scale by means of high-resolution transmission electron microscopy (HRTEM) impose not only extreme demands on the mechanic and electromagnetic stability of the applied instruments but require also their precise electron-optical adjustment. Today a physical resolution well below one Ångström can be achieved with commercially available microscopes on a daily basis. However, the achieved resolution can often not be reliably exploited for the interpretation of the resulting microscopical data due to the presence of so-called higher-order lens aberrations.

At the starting time of this work, a sufficiently accurate procedure to measure higher-order aberrations was urgently missing. Since aberration measurement is a mandatory prerequisite for any technique of aberration control enabling quantitative high-resolution microscopy, the goal of this work is to develop such a measurement procedure for the Sub-Ångström regime.

The measurement procedures developed in the course of this work are based on the numerical evaluation of a series of images taken from an amorphous object under electron-beam illumination with varying tilt. New techniques have been developed for the evaluation of single images as well as for the optimized evaluation of the whole series. These procedures allow microscope users to perform quantitative HRTEM even at a resolution of 0.5 Ångström. The precision reached with the newly developed measurement procedures is unprecedented and surpasses existing solutions by at least one order of magnitude in any respect.

All the concepts and procedures for aberration measurement developed in this work have been implemented in a software package which satisfies professional demands with respect to robustness, precision, speed and user-friendliness. The new automatic aberration-measurement procedures are suitable to establish HRTEM as a quantitative technique for material science investigations in the Sub-Ångström regime.

**Ultra-genaue Messung optischer Aberrationen für Sub-Ångström Transmissionselektronenmikroskopie**

Quantitative Untersuchungen der atomaren Struktur von Materialien mittels hochauflösender Transmissionselektronenmikroskopie (HRTEM) erfordern neben der notwendigen mechanischen und elektromagnetischen Stabilität auch eine hochpräzise Justage der verwendeten Mikroskope. Mit kommerziellen Elektronenmikroskopen der neuesten Generation kann heutzutage eine physikalische Auflösung deutlich unterhalb eines Ångströms routinemäßig erreicht werden. Die erreichte Auflösung kann jedoch wegen des Vorhandenseins von Abbildungsfehlern höherer Ordnung häufig nicht zuverlässig zur Aufklärung der Objektstruktur anhand der gewonnenen elektronenmikroskopischen Daten genutzt werden.

Zu Beginn dieser Arbeit existierte keine Prozedur zur Messung von Abbildungsfehlern höherer Ordnung mit der erforderlichen Genauigkeit, um den Sub-Ångström Bereich zuverlässig zu erschließen. Eine vorhergehende Messung ist notwendige Voraussetzung für jede Art von Kontrolle über die Abbildungsfehler, welche letztendlich quantitative HRTEM ermöglicht. Es ist deswegen das Ziel dieser Arbeit, eine zuverlässige Messprozedur für Abbildungsfehler bereitzustellen, die insbesondere den Genauigkeitsanforderungen im Auflösungsbereich unterhalb eines Ångströms gerecht wird.

Die Messprozeduren, die im Rahmen dieser Arbeit erstellt wurden, basieren auf der numerischen Auswertung von Serien elektronenmikroskopischer Aufnahmen von amorphen Objekten unter variierender Kippung der einfallenden Elektronenstrahlung. Dazu wurden sowohl neue Techniken zur Auswertung der einzelnen Aufnahmen entwickelt, als auch die Auswertung der gesamten Strahlverkippungsserie optimiert. Die dabei erreichte Genauigkeit übertrifft bereits existierende Lösungen um mindestens eine Größenordnung, so dass Abbildungen sogar bei einer heute noch nicht erreichten Auflösung von 0.5 Ångström quantitativ ausgewertet werden können.

Alle in dieser Arbeit entwickelten methodischen Konzepte und numerischen Prozeduren wurden in einem Softwarepaket implementiert, das professionellen Anforderungen in Bezug auf Robustheit, Präzision, Geschwindigkeit und Nutzerfreundlichkeit genügt. Diese neuen automatischen Messprozeduren können dazu beitragen, Materialuntersuchungen auf Sub-Ångström Niveau mittels HRTEM als quantitative Technik zu etablieren.

# Contents

**Introduction** **9**

**1 Fundamentals of high-resolution transmission electron microscopy** **13**

1.1 Functionality of a high-resolution transmission electron microscope . . . . 13

1.1.1 Conventional setup of transmission electron microscopes . . . . . 14

1.1.2 Special electron-optical devices . . . . . 16

1.2 Electron diffraction for thin objects . . . . . 22

1.3 Electron-optical contrast-transfer theory . . . . . 26

1.3.1 Fundamentals of image formation in HRTEM . . . . . 26

1.3.2 Linear imaging theory for thin objects . . . . . 27

1.3.3 Fundamental aberrations: spherical aberration and defocus . . . . . 29

1.3.4 Coherent linear contrast transfer . . . . . 31

1.3.5 Partially coherent, linear contrast transfer . . . . . 34

1.3.6 Partially coherent, non-linear contrast transfer . . . . . 40

**2 Wave aberrations in HRTEM** **43**

2.1 General description of wave aberrations . . . . . 43

2.2 A new notation for the aberration coefficients . . . . . 44

2.3 Properties of individual wave aberrations . . . . . 48

2.4 Principal strategies to correct aberrations . . . . . 51

2.4.1 Treatment of aberrations in image simulations . . . . . 51

2.4.2 Aberration correction by hardware . . . . . 52

2.4.3 Aberration correction by software . . . . . 53

2.5 Tolerance limits for residual aberrations . . . . . 54

2.5.1 Historical limits . . . . . 54

2.5.2 Currently used single aberration limits . . . . . 55

2.5.3 Proposal of a new multi-aberration tolerance-limit . . . . . 55

**3 Special aspects of the contrast transfer** **59**

3.1 Linear contrast transfer with an asymmetric aberration function . . . . . 59

3.1.1 Reviewing the linear contrast transfer . . . . . 60

3.1.2 Interpretation of the resulting contrast transfer . . . . . 62

3.2 Linear contrast transfer with tilted illumination . . . . . 64

3.2.1 The effective aberration function . . . . . 64

3.2.2 The effective damping due to partial temporal coherence . . . . . 66

3.2.3 The effective damping due to partial spatial coherence . . . . . 67

3.2.4 Explicit formulation of the tilted-beam contrast transfer . . . . . 68

3.2.5 Numerical diffractograms of thin amorphous objects . . . . . 69

**4 Discussion of existing aberration-measurement methods** **73**

4.1 Why to use beam tilts for aberration measurement . . . . . 73

4.2 The tilt-induced displacement method . . . . . 74

4.3 The diffractogram method . . . . . 76

4.4 Aberration measurement based on wavefunction reconstruction . . . . . 79

4.5 Aberration measurement using caustic figures . . . . . 81

4.6 Conclusion on existing methods . . . . . 83

**5 Optimisation of the diffractogram method** **85**

5.1 Experimental data acquisition . . . . . 85

5.2 Evaluation of the experimental data . . . . . 86

5.2.1 Determination of higher-order aberrations . . . . . 87

5.2.2 Reliable error estimation . . . . . 88

5.2.3 Consideration of temporal variations . . . . . 89

5.3 Calibration of experimental parameters . . . . . 89

5.3.1 Calibration of the magnification . . . . . 91

5.3.2 Calibration of beam tilts . . . . . 94

5.4 Investigations on optimum tilt settings . . . . . 98

5.4.1 Necessary number of images . . . . . 98

5.4.2 Distribution of beam tilts . . . . . 99

5.4.3 Maximum modulus of beam tilts . . . . . 101

5.4.4 Required stability of beam tilts . . . . . 103

5.5 Layout of the final measurement procedure . . . . . 107

**6 Novel techniques for the analysis of single diffractograms 109**
6.1 Concept of the diffractogram analysis . . . 109
6.2 Detection and removal of diffractogram peaks . . . 113
6.3 Removal of the diffractogram background . . . 116
6.4 Determination of the diffractogram envelope . . . 118
6.5 Azimuthal orientation of diffractogram patterns . . . 120
6.6 Coarse pattern recognition . . . 123
6.6.1 Concept of the coarse pattern recognition . . . 123
6.6.2 Ultra-fast comparison algorithm . . . 125
6.6.3 Achievable precision . . . 126
6.7 Ultra-precise pattern recognition . . . 128
6.7.1 The principal concept for a full-area pattern recognition . . . 128
6.7.2 Generation of adaptive paths . . . 129
6.7.3 Performance and achievable precision . . . 130
6.7.4 Correlation and signal-to-noise ratio . . . 134

**7 Comprehensive evaluation of tilt tableaus 137**
7.1 Measurement of the mean defocus and of the temporal defocus drift . . . 137
7.2 Determination of the azimuthal offset of the tilt tableau . . . 138
7.3 Consistent error estimate . . . 140
7.4 Direct measurement of the defocus spread . . . 143
7.4.1 Theoretical description of damping anisotropies . . . 143
7.4.2 Experimental setup . . . 146
7.4.3 Discussion of experimental results . . . 148
7.4.4 Conclusion . . . 149
7.5 Combination of real-space and Fourier-space data . . . 149
7.5.1 Determination of the defocus sign . . . 150
7.5.2 Observation of displacement discrepancies . . . 152
7.5.3 A new method to measure the TEM-magnification . . . 153

**8 Application of the aberration measurement procedure 161**
8.1 Application of the ATLAS software package . . . 161
8.1.1 Higher-order aberration measurement . . . 162
8.1.2 Aberration-stability benchmark . . . 176
8.2 Investigations on the structure of $Cs_x(Nb,W)_5O_{14}$ . . . 182
8.2.1 Experiment . . . 183

8.2.2 Results and discussion . . . 185
8.3 Sub-Ångström HRTEM of defects in $SrTiO_3$ . . . 191
8.3.1 Experiment . . . 192
8.3.2 Results and discussion . . . 194

**Summary** **199**

**A Definitions and Theorems** **201**
A.1 Fourier transformation . . . 201
A.2 Convolution theorem . . . 201
A.3 Fourier transformation of the image intensity . . . 202
A.4 Definition of variance, covariance, and standard deviation . . . 202
A.5 The correlation coefficient . . . 203
A.6 Cross Correlation . . . 203

**B Transformation of effective second-order aberrations** **205**

**C Derivation of a novel tilt-induction formula** **209**
C.1 Explicit derivation of a general tilt-induction formula . . . 209
C.2 Application of the general tilt-induction formula . . . 213
C.2.1 Effective image displacement . . . 213
C.2.2 Effective defocus . . . 215
C.2.3 Effective twofold astigmatism . . . 216

**Bibliography** **219**

# Introduction

High-resolution transmission electron microscopy (HRTEM) is a technique which allows one to observe the structural details of solid matter on an atomic scale. In contrast to pure scattering techniques like X-ray or neutron scattering, the technique of HRTEM provides a direct insight into the local microstructure of matter in the form of an image. Modern high-resolution transmission electron microscopes resolve nowadays distances smaller than one Ångström (0.1 nm), which is sufficient to separate single atomic columns of crystals. In comparison, the technique of scanning tunnelling microscopy (STM) yields images with a similar resolution, but remains mainly sensitive to the object surface. Since HRTEM is a projection method, the signal is sensitive to the complete object-volume traversed by the incident electron beam.

Due to its ability to produce images with atomic resolution, high-resolution transmission electron microscopy is a powerful tool for materials research. The field of application ranges over prominent material classes like metals, semiconductors, and ceramics, reaching from long-term, high-end academic research to short-term emergency investigations of wafers in fabrication lines of the semiconductor industry. Therefore high-resolution microscopes can be found in academic as well as in industrial institutions.

Unfortunately, the interpretation of HRTEM images is by far not as straightforward as for traditional light-optical microscopy. The reason why HRTEM images contain the object information only in a highly encoded and non-intuitive manner is twofold: Firstly, the interaction between the incident electrons and the atoms in the object is a scattering process. The interaction follows quantum mechanical principles, where the electron is described mathematically as a complex-valued wavefunction. Secondly, the electron wavefunction is further modified during its transfer through the optical system of the electron microscope. Imperfections of the electron-optical lenses can falsify the information about the object structure contained in the wavefunction. This consecutive modification of the wavefunction is described by so-called *aberrations* in the framework of a wave-optical formalism. The recorded images finally represent the modulus-square of the resulting

electron wavefunction and are thus strongly affected by these two fundamental stages of wave-mechanical processing.

In an electron microscope electromagnetic fields act as lenses. The imaging property of an electron-optical lens is in many respects analogous to light optics. Aberrations known from light optics also occur for the electromagnetic lenses in electron optics and cause a falsification of the object information by a phase-modulation of the electron wavefunction. Amongst the many different aberrations that can be classified, the spherical aberration is by far the strongest. Just shortly after the invention of the electron microscope [1], Scherzer has proven that a spherical aberration cannot be avoided in the rotationally symmetric electromagnetic fields used for electron-optical lenses [2]. The immense strength of the unwanted image blurring caused by the spherical aberration is sometimes described by the analogous strong blurring that one observes when viewing through a bottle of glass.

During the last decades a huge amount of scientific effort has been invested into several techniques solving the interpretation problem in high-resolution transmission electron microscopy. The most prominent techniques are image simulation, high-voltage transmission electron microscopy (HVTEM), hardware correction of the spherical aberration, electron holography, and numerical wavefunction reconstruction. Except for HVTEM and image simulations most techniques provide no or only a partial solution to the first cause of the interpretation problem, i.e. the quantum mechanical scattering process, and address mainly the problem of the electron-optical contrast transfer. Different routes have been employed to avoid in first instance the dramatic blurring caused by the spherical aberration.

For a long time the standard technique to tackle the interpretation problem was the technique of image simulation [3–7]. In this approach an image, computed from a guessed atomic structure, is matched in trial-and-error procedures to the experimentally recorded image. On the basis of a frequently subjective assessment of the fit quality, the assumed structure model can be verified or rejected. For very complex or aperiodic object structures the application of the image simulation technique becomes tedious or even impossible, because of the complete lack of an initial guess to the unknown structure. Nevertheless, by fully accounting for the complex scattering process inside the object the image-simulation technique is an indispensable tool for a quantitative in-depth analysis in HRTEM.

High-voltage transmission electron microscopy [8] and electron holography [9, 10] are two techniques avoiding the optical part of the interpretation problem by a special instrumentation. The HVTEM technique takes advantage of the general physical principle that the

resolving power of an instrument increases with decreasing wavelength of the employed radiation. By increasing the accelerating voltage from the usual intermediate voltages of $200 - 400$ kV to over 1 MV, the decrease of the electron wavelength allows a higher resolution in combination with a more direct image interpretation. Due to the serious damaging of the object by the highly energetic electron beam and the high costs of such microscopes, HVTEM is only used by a handful of specialised laboratories worldwide. Applying the electron-holography technique, the complex-valued electron wavefunction can be recorded directly, avoiding artefacts caused by the imaging process. Electron holography is also employed by only a few institutions worldwide due to the very demanding requirements for the stability of the instruments and their environment.

In the recent years two approaches have attained a major breakthrough towards an intuitive direct interpretation of HRTEM images in terms of the object structure: Firstly, a hardware solution for the correction of the spherical aberration was constructed using multipole lens-systems [11–13]. Secondly, the computer-based numerical reconstruction of the electron wavefunction [14–18] allows one to correct for any unwanted phase-modulation caused by an arbitrary aberration a-posteriori [19].

Actually there is not only one aberration. A large variety of aberrations is required to describe all imperfections of an electron-optical transfer system. With improving resolution a larger amount of aberrations, for which currently no hardware corrector is available, becomes relevant for the electron-optical transfer of object information to the image. Both aforementioned aberration-correction techniques, hardware and software, require a prior measurement of the aberrations. A hardware corrector needs an accurate adjustment in order to compensate the spherical aberration of the objective lens, and to avoid the introduction of further, unwanted aberrations. Numerical a-posteriori aberration correction also requires a precise knowledge of the phase-modulation caused by the complete electron-optical system. With increasing resolution not only an increased number of different aberrations is relevant for the optical transfer of the electron wavefunction. At the same time, the precision of the aberration measurement becomes increasingly important. Uncertainties of possible residual aberrations after correction, which were negligible at resolutions slightly above one Ångström, may significantly modify the contrast of Sub-Ångström details in the image.

Just now, Sub-Ångström resolution is achieved routinely on a daily basis with the newest generation of commercially available transmission electron microscopes. The manufacturers of the novel microscopes have the serious problem that the precision of the aberration-measurement technique [20] delivered together with their instruments is insufficient for

the specified resolution well below one Ångström.
The electron-microscope manufacturers frequently put much weight to the specification of the best resolution achievable with their instruments, which is usually addressed as the information limit. In general, the specification of an information limit provides no evidence about the quality or correctness of the observable information. It testifies only of the physical resolving potential of an electron microscope. One has to distinguish between the physical resolution given by the information limit and the interpretable resolution as two similarly decisive criteria for materials-science investigations. The interpretable resolution specifies up to which detail the image contrast reflects the object structure in a straightforward and interpretable manner.
The present aberration-measurement techniques cannot cope with the nowadays routinely available Sub-Ångström information limit. It is therefore almost mandatory to improve the technique for an ultra-precise measurement of optical aberrations in high-resolution transmission electron microscopy. An improved aberration-measurement technique should satisfy the increased requirements with respect to precision, robustness and speed, demanded for the imaging of object information with Sub-Ångström resolution. In the face of the large investment in financial resources and know-how required for the purchase and operation of a Sub-Ångström transmission electron microscope, any uncertainty about the actually obtained interpretable resolution is not acceptable. By application of improved aberration-measurement techniques the so far often qualitative technique of HRTEM should become a strictly quantitative technique leaving no room for speculations about the imaging process. With an established control over the optical aberrations the microscope user should be able to rely on a straightforward imaging by the microscope, in the meaning of a standardised physical metrology providing reliable information from the Sub-Ångström regime. It is therefore the goal of this work to contribute substantially to the improvement of aberration measurement in high-resolution transmission electron microscopy.

# Chapter 1

# Fundamentals of high-resolution transmission electron microscopy

The basic principle of transmission electron microscopy follows a straightforward scheme. An electron beam illuminates the object, which is thin enough to be transparent for electrons. High-resolution work on an atomic scale requires an object thickness of less than 10 nm. Inside the object, the electrons interact with the atomic scattering potential and thus carry information about the atomic structure of the object. After the transition of the object, the electrons are guided by an electron-optical lens system in order to form a strongly magnified image. Thereby the information about the traversed three-dimensional object structure is transferred into a two-dimensional image.
The theories of electron diffraction and electron imaging describe how information of the object structure is formed inside the object and transferred thereafter to the final imaging detector. Both theories are quite complex, so that only its basic concepts can be addressed in the frame of this work. A deeper insight into selected specialised aspects of the electron-optical contrast transfer, which concern in particular the topic of aberration measurement, is given in chapter 3. For obtaining a broader insight on the topic of transmission electron microscopy, the reading of text books like refs. [21–24] is additionally recommended.

## 1.1 Functionality of a high-resolution transmission electron microscope

According to Rayleighs's diffraction-limited resolution criterion, the physical resolution of an optical instrument improves with decreasing wavelength of the radiation in use [51].

Typical intermediate voltage microscopes accelerate electrons by a high-voltage between 200 and 400 kV. The corresponding relativistically corrected electron wavelength lies then in the range between 2.5 and 1.6 pm. Due to the limited coherence of the electron source, unavoidable mechanical and electronic instabilities, and lens aberrations, the actual resolution of such instruments is roughly two orders of magnitude worse than the respective wavelength. Dedicated high-resolution instruments achieve resolutions between just below 1 Ångström up to 2 Ångströms. At these small length scales, inter-atomic distances and defects in many crystalline materials can be resolved.

### 1.1.1 Conventional setup of transmission electron microscopes

Modern transmission electron microscopes consist of many electromagnetic lenses and other optical elements. The details of these electron-optical devices are complex, but some fundamental terms suffice to describe the principal setup of a conventional transmission electron microscope (TEM) in a sequence from top to bottom. In many aspects the imaging properties of electron-optical lenses are analogous to light optics. In particular the rotationally symmetric electromagnetic field of the objective lens acts on electrons like a thick spherical glass lens on light, exhibiting a strong focusing action. The electron-optical transfer can be illustrated by a ray diagram, as shown schematically in figure 1.1.

- The electron source is the uppermost element inside the column of the electron microscope. Two types of electron sources are mostly used in transmission electron microscopes. In older instruments the electrons are extracted from from a lanthanum hexaboride ($LaB_6$) or tungsten cathode of needle-like shape by thermionic emission. Today field-emission guns (FEG) are widely implemented due to an increased brightness, which offers the possibility to improve the coherence properties of the electron beam by about one order of magnitude.

- After emission, the electrons pass the so-called electron gun for acceleration. A high tension of typically some hundred kilovolts accelerates the electrons emitted from the source cathode. For materials-science investigations, accelerating voltages between 200 to 400 kV are a compromise between higher resolution and increasing radiation damage to the object.

- Several condenser lenses and the condenser aperture are used to ensure a bright and parallel illumination of the object by the electron beam.

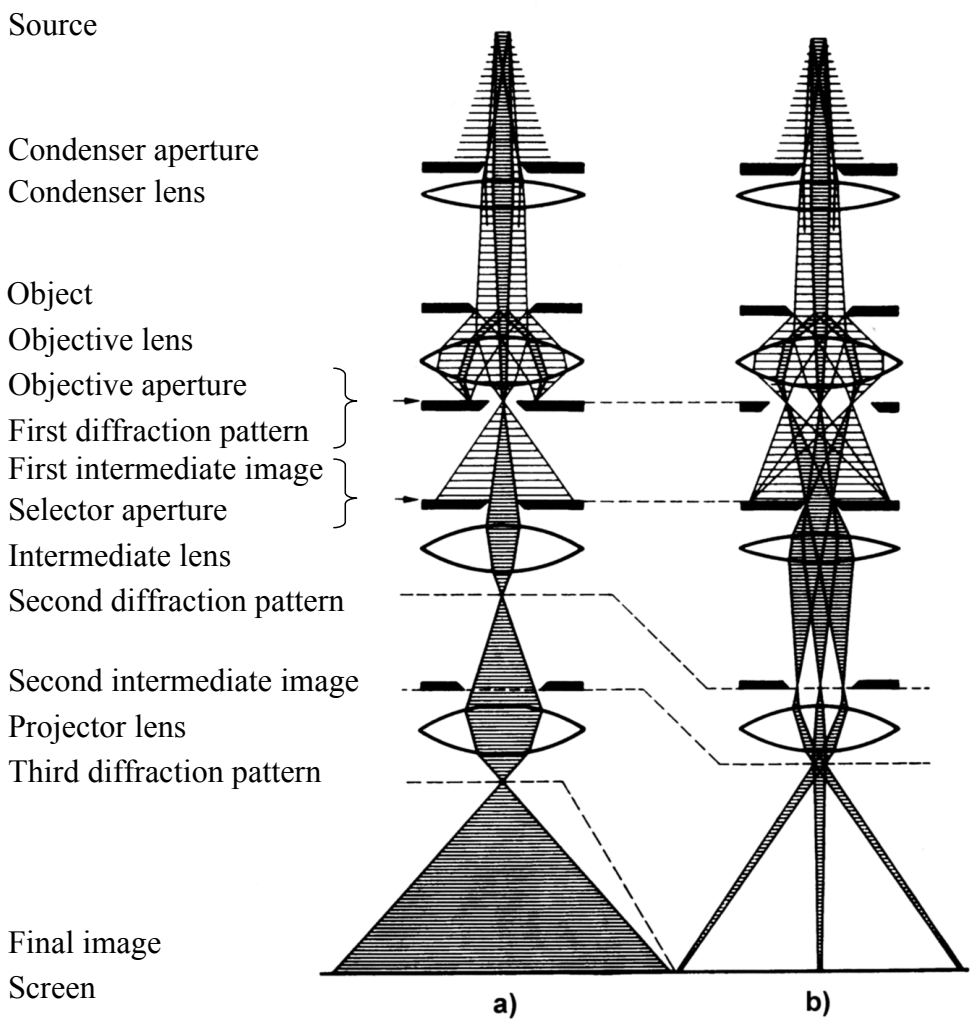



Figure 1.1: Ray-optical diagrams for a conventional transmission electron microscope in (a) the transmission imaging mode and (b) the selected-area electron-diffraction mode (reproduced from [21]).

- The object-holder is located approximately at half-height of the microscope column. It provides control over the position of the object in three dimensions and a tilt around two perpendicular axes. The incident electrons interact with the atomic scattering potentials inside the object.

- The objective lens is the most crucial component of the instrument. It has a very short focal length of approximately 1-2 mm, which requires high currents of several ampere flowing through the lens coils.

- An objective aperture can be inserted into the ray path close to the diffraction plane. It limits the amount of diffracted beams which contribute to the image.

- The selected area aperture is used in diffraction mode to select the area of the object that contributes to the diffraction pattern.

- By intermediate lenses the first intermediate image (figure 1.1a) or the first diffraction image (figure 1.1b) is post-magnified. In the last step of magnification, the projector lens generates a further magnified image. The lenses of the post-magnification system have ideally no crucial influence on the process of image formation.

- Depending on the available equipment the final image can be recorded with a photo plate, a video camera or a charge-coupled device (CCD) camera. CCD-cameras are widely used for quantitative analysis of high resolution images [25, 26]. Today typical CCD-cameras involve a CCD-chip with 2048×2048 pixels having a diameter of 15 to 25 $\mu$m.

High efforts are required in order to keep the environment of the TEM free from mechanical vibrations and electromagnetic stray fields. Passive and active damping systems including acoustic absorption, as well as electromagnetic field compensations are often installed in order to create an environment suitable for high-resolution experiments.
An example for a modern TEM is shown in figure 1.2. The Philips CM200C is a modified Philips CM200ST FEG electron microscope. The microscope is operated at 200 kV acceleration voltage and provides a resolution below 1.3 Å (0.13 nm). This instrument is equipped with the world's first prototype of a corrector for the spherical aberration. The concept of aberration correction by means of hardware is introduced in the following section.

### 1.1.2 Special electron-optical devices

The most prominent aberration in transmission electron microscopes is the spherical aberration. Just a few years after the invention of the electron microscope by Ernst Ruska and Max Knoll [1], Scherzer proved that a spherical aberration cannot be avoided in rotationally symmetric electromagnetic fields, and that a correction by an additional rotationally symmetric field is impossible. This is known as *Scherzer's theorem* [2].
In 1947 Scherzer himself made the principal suggestion to correct the spherical aberration, by means of a combination of electromagnetic quadrupole and octupole lenses [27]. Due to the complexity of the technical problems it took about 20 years from the first trials in the mid-1970s to the final realisation of a hexapole design to correct the spherical aberration of a TEM in the 1990s [11, 13, 28–30].

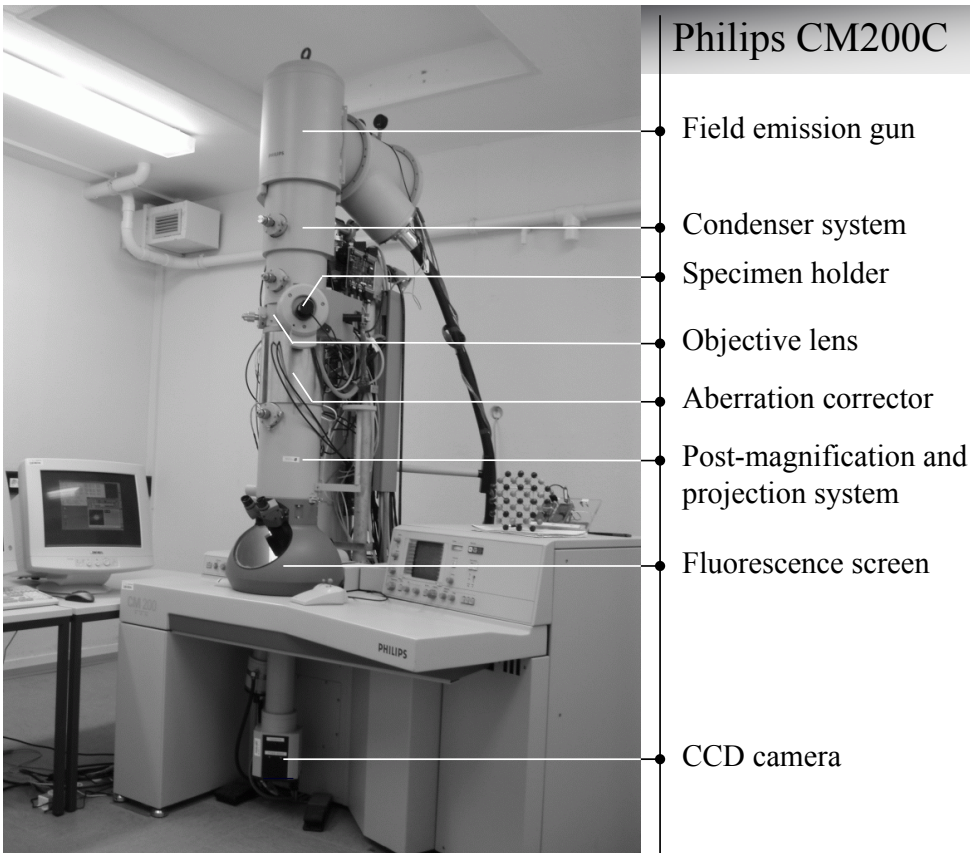



Figure 1.2: The Philips CM200C transmission electron microscope operated in Jülich is equipped with the world's first prototype of a spherical-aberration corrector.

The first prototype of the spherical-aberration corrector was developed in a joint project, funded by the Volkswagenstiftung, to M. Haider, EMBL Heidelberg, H. Rose, TU Darmstadt and K. Urban, Research Centre Jülich. This prototype is still operated in the Philips CM200C in Jülich (see figure 1.2). The construction of this prototype is based on a hexapole design as proposed by Rose [31]. Just recently the hexapole corrector has become a series product and is distributed commercially with high-end high-resolution transmission electron microscopes. An example for such a spherical-aberration corrected microscope of the newest generation is shown in figure 1.3.
The FEI Titan 80-300 transmission electron microscope is characterised by an information limit of 0.8 Å. It can be operated at a maximum accelerating voltage of 300 kV. Using a monochromator allows one to further improve the coherence of the electron beam, so that a resolution even below 0.8 Å is possible. The digital images are recorded with a Gatan 2048×2048 slow scan CCD-camera system. The FEI Titan 80-300 as well as the Philips CM200C were used for the experimental part of this work.

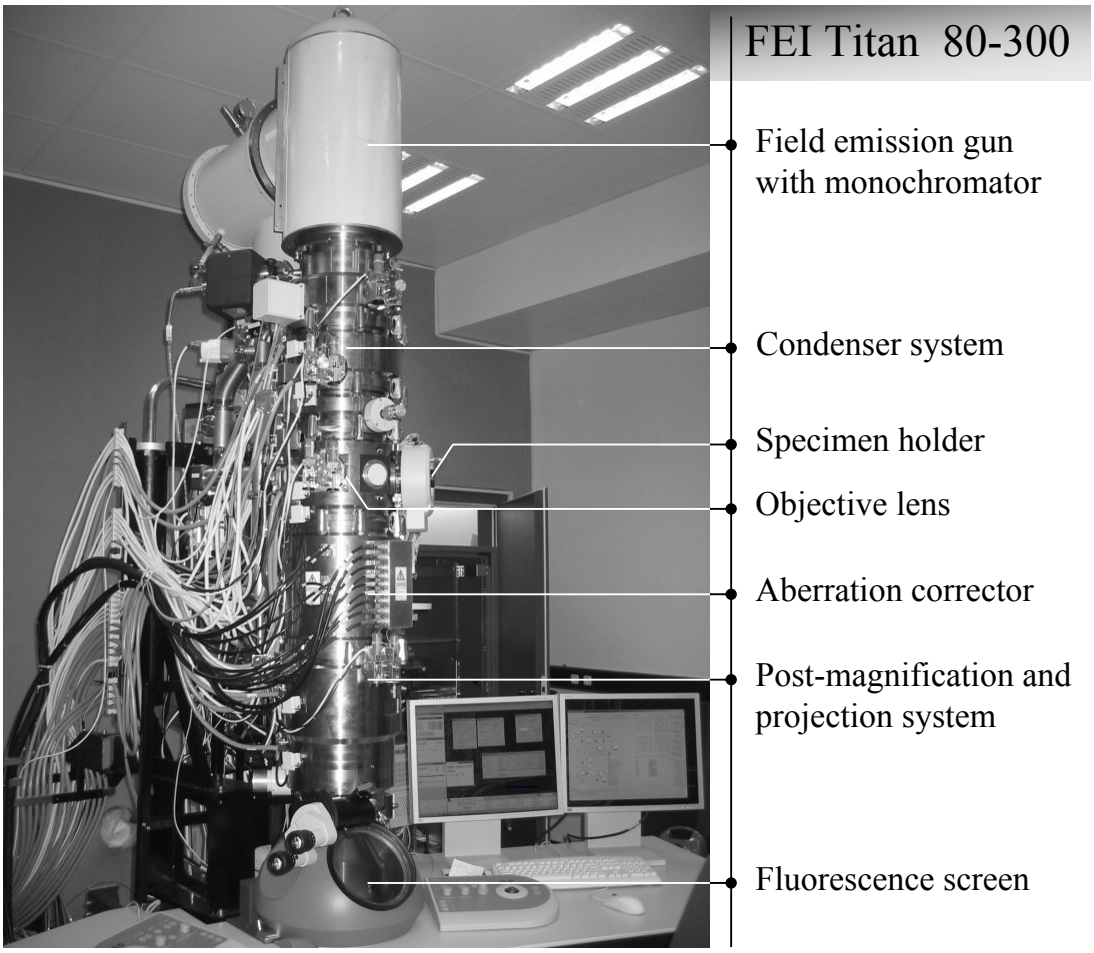



Figure 1.3: The FEI Titan 80-300 Sub-Ångström transmission electron microscope during assembly at the Ernst Ruska-Centre in Jülich, February 2006.

### Spherical-aberration corrector

The construction of the first aberration corrector for transmission electron microscopes involves two hexapole lenses and two transfer round-lens doublets, as depicted schematically in figure 1.4. Each of the two hexapole elements introduces a strong, primary, threefold astigmatism. The relative excitation and position of the two hexapoles ensure a mutual compensation of the threefold astigmatism, while a residual spherical aberration of the next higher order remains due to the non-linearity of the hexapoles' diffraction power. The resulting residual spherical aberration has the opposite sign of the spherical aberration of the objective lens and is proportional to the square of the hexapole excitation [13]. Choosing the hexapole excitation appropriately, allows one to fully eliminate or even to over-compensate the spherical aberration of the objective lens.

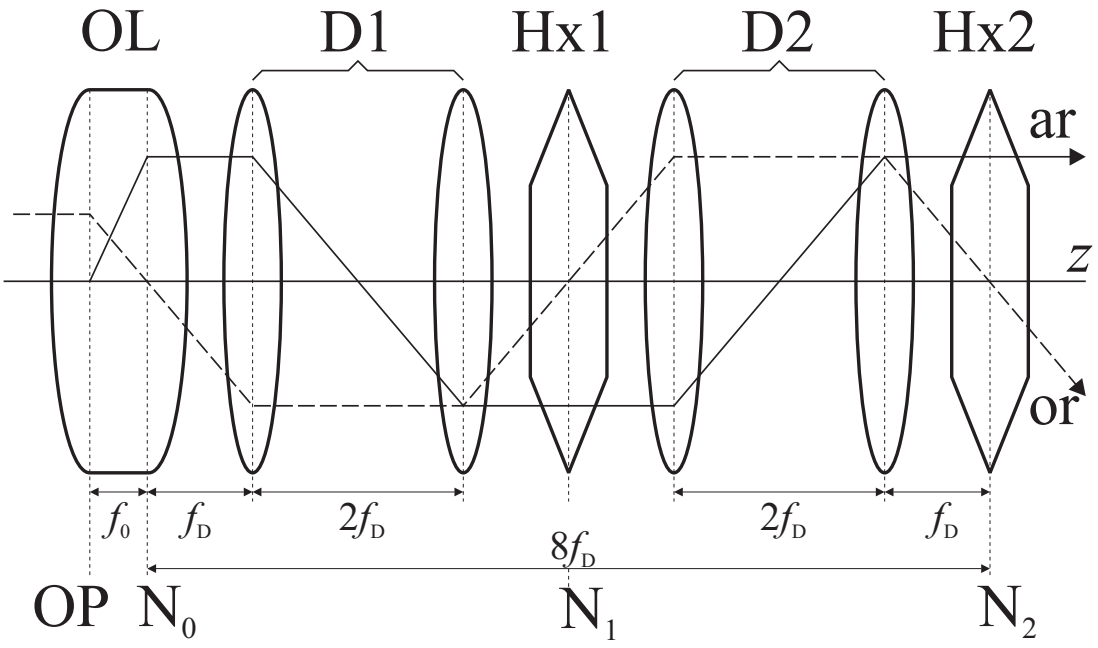



Figure 1.4: Scheme of the hexapole design to correct the spherical aberration of a compound objective lens (OL) of focal length $f_0$. The hexapole corrector consists of two transfer round-lens doublets (D1 and D2) of focal length $f_D$, and two hexapoles (Hx1 and Hx2). There are four coma-free planes, the objective plane (OP), and the outer nodal planes of the doublets $N_0$, $N_1$, and $N_2$. Any possible ray is a linear combination of the two principal rays; the axial (ar) and off-axial ray (or) [13].

### Monochromators

In order to improve the resolution of an electron microscope, much effort is invested by the manufactures to reduce influences of mechanical vibrations and electromagnetic-field variations on the imaging process. A higher resolution can in particular be achieved by an improvement of the coherence of the electron beam, for example by a reduction of the electron-energy spread. The energy spread obtained with thermionic emitters is approximately 1 eV, whereas an energy spread between 0.3 - 0.6 eV can be achieved with FEG-sources. With monochromators, the energy spread can be reduced to less than 0.2 eV [32,33]. Two different concepts have just recently been introduced: the Wien-filter monochromator and the electrostatic Omega-type (Ω-type) monochromator. A resolution well below 1 Ångström was demonstrated for transmission electron microscopes equipped with both types of monochromators [34, 35].

In a Wien-filter, the forces caused by a homogeneous electrostatic field $E$ and a perpendicular homogeneous magnetic field $B$ compensate exactly for electrons with the velocity $v = E/B$. Electrons with a different velocity, get spatially separated from the original ray path. Stronger fields result in increased spatial separation of electrons with different

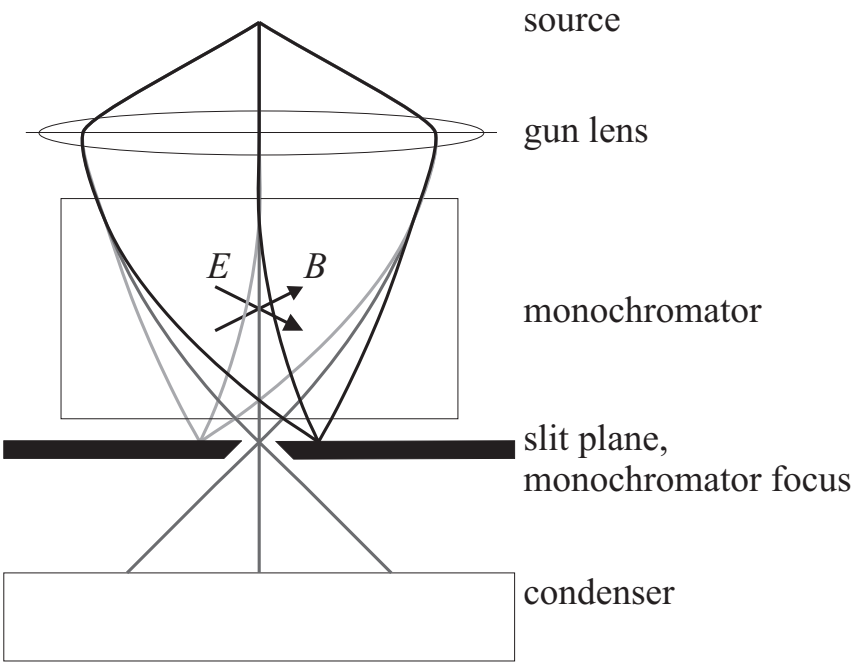



Figure 1.5: Scheme of the Wien-filter monochromator distributed by the FEI company. $E$ denotes the electrostatic field, $B$ the magnetic field.

kinetic energies, i.e. the dispersion increases at the monochromator exit. Choosing the beam-convergence at the monochromator entrance and the field excitation appropriately, each energy is focussed to a line in the slit-plane (see figure 1.5). The transmitted energy window is selected by a slit perpendicular to the line, where each energy is focussed to a different point.
The Ω-type monochromator was developed by the CEOS company together with Carl Zeiss NTS. The setup of the Ω-type monochromator was proposed by Kahl and Rose [36]. It consists of four electrostatic sector elements and a slit, as illustrated in figure 1.6. The slit selects the transmitted energy window in the dispersive plane. Due to the symmetric design, distortions of the image cancel out and the beam shape is inherently preserved [32].

For both constructions, the monochromator is placed directly below the electron gun and above the high-voltage acceleration line in the column of the electron microscope. There, the total energy of the electrons is still relatively small, so that the dispersion in the slit plane requires smaller electric and magnetic fields. Since the monochromator technique is a relatively new instrumental development in HRTEM, materials-science investigations using the monochromator have not been reported yet. It is important to note that any energy selection by a slit intrinsically reduces the current of the electron beam. As a result the mean intensity in the image is accordingly reduced.

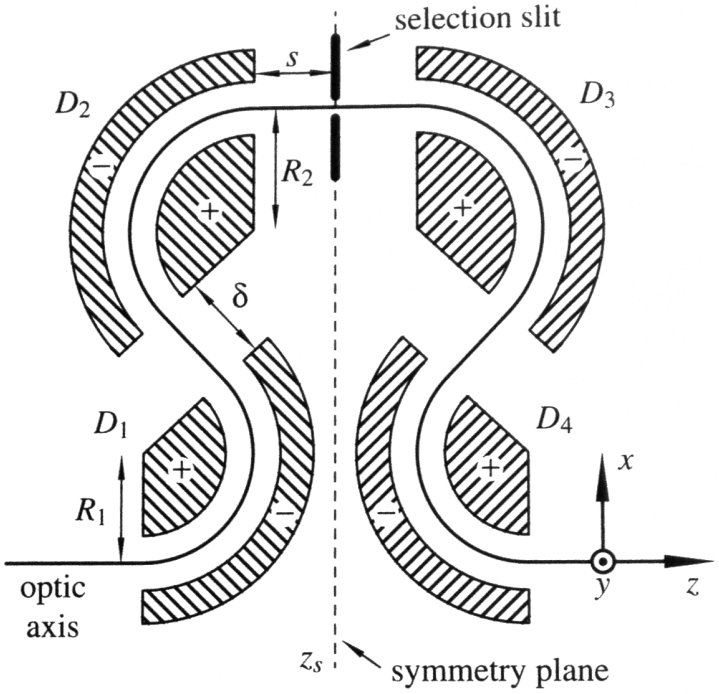



Figure 1.6: Schematic setup of the electrostatic Ω-type monochromator. Four sector elements $D_1$ – $D_4$ generate the energy dispersion in the symmetry plane, where the transmitted energy window is selected by a slit (reproduced from [36]).

### Future generation of transmission electron microscopes

The currently running TEAM project (Transmission Electron Aberration-corrected Microscope) [38] is a collaborative effort funded by the Department of Energy of the USA to construct a next-generation electron microscope. The task of the TEAM project is to create an instrument with 0.5 Å spatial resolution and less than 0.2 eV energy resolution at accelerating voltages between 100 kV and 300 kV. This goal demands the reduction of all parasitic coherent and incoherent aberrations. Future developments in hardware aberration correction have been outlined by Rose's proposal of a *superaplanator* or the more complicated *ultracorrector* [37]. Both concepts promise a correction of spherical and chromatic aberration, as well as a minimum introduction of parasitic aberrations. The two proposals combine current techniques like monochromators [32,33] and sextupole [13] or quadrupole-octupole [39, 40] spherical-aberration correctors in one single aberration-correction device.

## 1.2 Electron diffraction for thin objects

The interaction of electrons with the atoms inside the scattering object is strong compared to X-ray and neutron scattering. Consequently, the electrons are scattered multiply before leaving the object. The particle-optical picture of this multiple scattering phenomenon is not appropriate; a wave-mechanical formalism has to be used in order to describe the elastic scattering of electrons inside the object. Within this framework, the expression "scattering" is misleading. The term "diffraction" is suited better to characterise the interaction of an incident electron wave with the atoms.
The phenomenon of electron diffraction is correctly described by the solutions of the relativistically corrected Schrödinger equation for an electron wavefunction $\Psi(\vec{r})$ in a periodic crystal potential $V(\vec{r})$. This analytical approach is called the Bethe-Bloch formalism [41, 42]. The application of the Bethe-Bloch formalism is only practicable for the periodic parts of a crystal and is ill-suited for the treatment of aperiodic objects. In practice, a numerical approach called "Multislice Algorithm" is applied instead, which converges arbitrarily close to the Bethe-Bloch formalism [43, 44].
Due to the complex nature of dynamic electron diffraction, this work forbears to describe the full dynamic theory of electron diffraction. Here, the theoretical considerations are restricted to very thin objects, which are only a few atomic layers thick. Within this limitation, substantial simplifications of the diffraction formalism and the imaging formalism allow an intuitive interpretation of the HRTEM image. A comprehensive treatment of the theoretical and practical aspects of dynamic electron diffraction in HRTEM can be found in the literature [21–23].

The approximative description of the electron diffraction in thin objects starts with the assumption of an incident plane wave described by a normalised electron wavefunction $\Psi(\vec{r}, t=0) = 1$ at the entrance surface of the object, where the traversed object thickness $t$ is zero. The atoms inside the object constitute an optically thicker medium for the electron wavefunction than the surrounding vacuum. Therefore the wavelength of the electrons is shortened at the atomic positions, while it remains constant between them, as depicted symbolically for one atom in figure 1.7a. The phase of the electron wavefunction proceeds thus faster at the atom, which finally leads to a phase-modulation of the electron wavefunction. The modulation of the phase of the electron wavefunction is also symbolically illustrated in figure 1.7b by lines of equal phase.
The above phenomenological description of the phase-object approximation (POA), corresponds to a description of the electron-diffraction problem in the framework of the

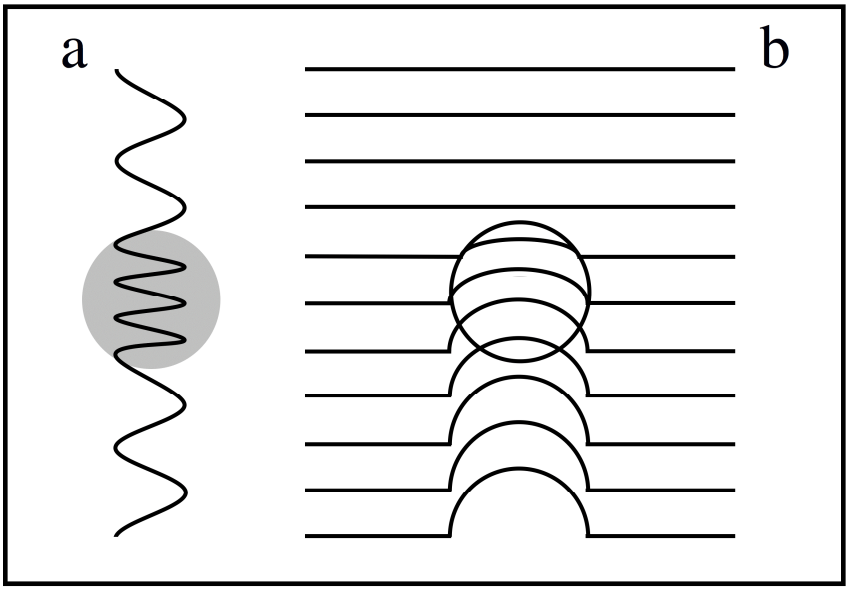



Figure 1.7: Symbolic illustration of the phase-object approximation for one atom. (a) The wavelength of the electron wavefunction is shortened while passing through an atom. (b) The lines of equal phase of the electron wavefunction reveal a cumulated phase-modulation. The illustration is taken from [24].

semi-classical WKB-approximation. For thin objects, the POA yields the electron wavefunction $\Psi(x, y, t)$ at the object thickness $t$ as

$$\Psi(x, y, t) \approx \exp\left[\mathrm{i}\,\Phi(x, y, t)\right] = \exp\left[\mathrm{i}\,\sigma V_{\mathrm{P}}(x, y, t)\right], \tag{1.2.1}$$

where $\Phi(x, y, t)$ denotes the phase of the wavefunction, $\sigma$ is the interaction constant, and $V_{\mathrm{P}}(x, y)$ is the projected crystal potential. In the following the real-space vector $\vec{r} = (x, y)$ is two-dimensional. The formalism of the POA abandons the successive propagation of the electron wavefunction through distinct atoms separated by vacuum. A continuous and constant projected potential replaces the varying object potential along the direction of the incident beam. The projected potential is positive and strongly peaked at the positions of the atomic columns.

In a further simplification of the approximate description of electron diffraction, the phase-modulation of the electron wavefunction is assumed to be small for a small object thickness. The assumption of $\Phi(\vec{r}, t) \ll 1$ is called weak-phase-object approximation (WPOA):

$$\Psi(\vec{r}, t) \approx 1 + \mathrm{i}\,\Phi(\vec{r}, t) = 1 + \mathrm{i}\,\sigma V_{\mathrm{P}}(\vec{r})\,t \tag{1.2.2}$$

The WPOA provides still the possibility to interpret the wavefunction directly. While the real part remains equal to the amplitude of the incoming beam, the imaginary part is directly proportional to the projected object potential. Imaging the imaginary part

would thus enable a direct observation of the atomic object structure by the sharply peaked maxima of the projected potential. Due to its simplicity the WPOA is applied in the description of the imaging process of the electron wavefunction in section 1.3.
Unfortunately the POA and WPOA formalism is not applicable for realistic objects in materials science. The breakdown of the WPOA can be seen directly in equation (1.2.2), which predicts a continuous and unlimited growth of the electron intensity $\Psi\Psi^*$ with the object thickness, where the raised asterix ($*$) denotes the complex conjugate. This contradicts to the conservation-rule of the intensity.
The weakness of the superior POA in equation (1.2.1) is revealed by the fact that the intensity $\Psi\Psi^*$ remains constant for all object thickness values, which is not observed in the experiment. The reason for the breakdown of the two approximations is found in the neglect of the wave-propagation in the vacuum between the successive atoms of a column. According to the Huygens-Fresnel principle, spherical waves are emitted from each point of the wave-front. The superposition of the spherical waves leads to a lateral re-distribution of the wave front on its way to the next atomic layer.
The already mentioned Multislice Algorithm [43, 44] subdivides the object into slices containing only one atomic layer perpendicular to the direction of the incident beam. Within each layer the algorithm applies the POA. In an additional step it accounts for the vacuum propagation, before the next atomic layer is considered. By taking the spatial extent of the object along the beam into account, the Multislice Algorithm yields accurate solutions of the scattering problem.
The electron diffraction in periodic objects can be investigated in framework of the so-called channelling theory also for a larger object thickness. The channelling model is an approximation in the frame of the full electron-diffraction formalism, where Schrödinger's equation is solved for the isolated atom columns individually. Thereby multiple scattering events between the neighbouring columns are neglected.
The solutions of the respective Schrödinger equation describe a pendulum like motion of the electron along the atomic column which is called *Pendellösung* in the dynamic theory. A symbolic illustration of such a Pendelösung is given in figure 1.8 for two atom types. The length of the periodicity of the electron motion along the vertical thickness axis is called *extinction length* $\xi$. At a thickness of $\xi/2$ the electron intensity is focussed on the column position, while the electron intensity at a thickness $t = \xi$ corresponds to the intensity distribution of the incoming wavefunction, i.e. the image shows no contrast for this atomic column. The extinction length depends sensitively on the charge number of the atoms inside the atomic column. With increasing atomic charge number the extinction length

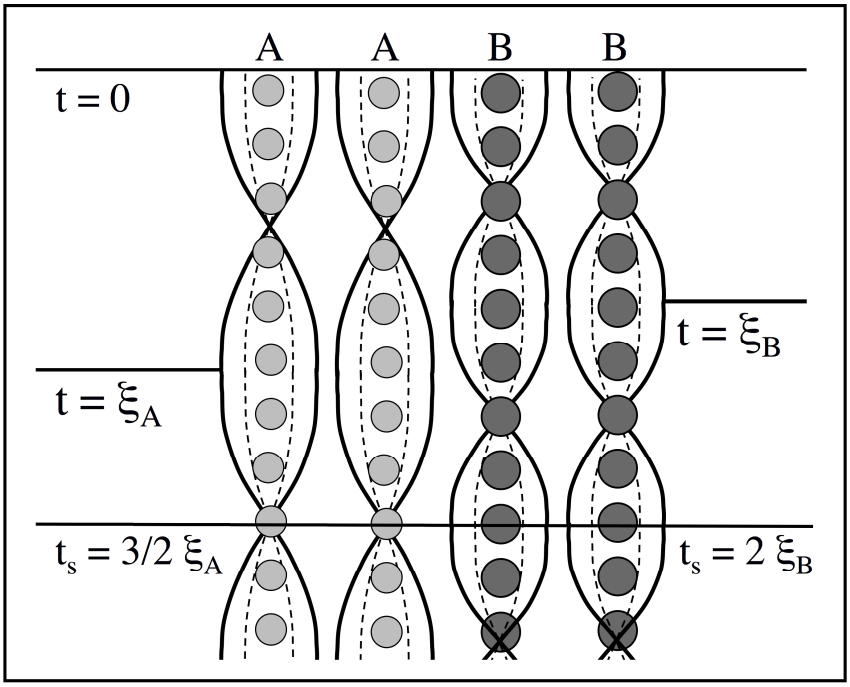



Figure 1.8: Symbolic illustration of the electron trajectories according to the channelling theory, where the electrons describe a pendulum-like path along the atom columns. Two types of atoms are drawn, representing light atoms (A) and heavy atoms (B). The length of the periodicity of the electron motion along the vertical thickness axis is called *extinction length* $\xi$. At a thickness of $\xi/2$ the electron intensity is focussed on the column position, while the electron intensity at a thickness $t = \xi$ corresponds to the intensity distribution of the incoming wavefunction. In the depicted example the type A columns are visible as intensity peaks at the special thickness value $t_s$, whereas the type B columns are completely invisible. The illustration is taken from [24].

decreases and the pendulum movement of the electron along the direction of incidence becomes faster. This corresponds to stronger phase modulation by heavier atoms in the wave formalism. As depicted in figure 1.8 thickness values may exist, where one column of an atomic species A can show up in the electron intensity, whereas the columns of a different species B are not visible in the image.

The pendulum-like behaviour for thicker objects reflects the strong Coulomb interaction between the negatively charged incident electron and the positive charges of the atom cores. Typical extinction lengths are on the order of 10 – 50 nm. The strong intensity modulation described by the channelling theory becomes most prominent at $\xi/2$. Since the POA in equation (1.2.1) yields no intensity modulation, the validity of the POA/WPOA formalism holds only up to a thickness of at maximum $\xi/4$, which corresponds to only a few nanometres.

## 1.3 Electron-optical contrast-transfer theory

In the preceding section approximations have been introduced which describe the quantum-mechanical diffraction process of an electron traversing through a thin crystalline object. This process results in an electron wavefunction at the exit plane of the object which is called the *exit-plane wavefunction* (EPW). The theory of electron-optical contrast transfer describes the consecutive imaging of the object information contained in the EPW.

The electron-optical imaging proceeds in two steps: Firstly, the electron optical lenses guide the propagation of the EPW from the exit plane to the image plane. Secondly, the image is finally represented by the modulus square of the electron wavefunction in the image plane. The essential features of HRTEM images are derived first in a linear imaging theory for thin objects using the weak-phase-object approximation. The linear theory allows to deduce an optimum adjustment of the objective-lens parameters for an almost straightforward imaging of the exit-plane wavefunction.

### 1.3.1 Fundamentals of image formation in HRTEM

The electron wavefunction in the image plane, the so-called image-plane wavefunction, is recorded by the detector system in form of an intensity distribution. The intensity $I(\vec{r})$ in the image is given by the modulus square of the image-plane wavefunction $\Psi_\mathrm{i}(\vec{r})$, with

$$I(\vec{r}) = \Psi_\mathrm{i}(\vec{r})\Psi_\mathrm{i}^*(\vec{r}), \tag{1.3.1}$$

where $\vec{r} = (x, y)$ is a two-dimensional real-space vector in the image plane. The above formulation of the image intensity merely describes the formation of the image for a given wavefunction $\Psi_\mathrm{i}$ without any consideration of previous phase- or amplitude modulations. Since the magnification $M$ is a pure scaling factor, it is usually set to one, as in equation (1.3.1). Therefore distances in the image are formally identical to distances in the object. Writing the wavefunction in amplitude-phase notation $\Psi_\mathrm{i}(\vec{r}) = A(\vec{r})\,\exp[\,\mathrm{i}\Phi(\vec{r})]$, uncovers a serious problem of the imaging process: Due to $I(\vec{r}) = A^2(\vec{r})$ the information about the object contained in the phase $\Phi(\vec{r})$ of $\Psi_\mathrm{i}(\vec{r})$ is lost. Assuming an ideal microscope, which does not alter the exit-plane wavefunction yields $\Psi_\mathrm{i} = \Psi_\mathrm{e}$. In this case, the complete information about the structure of very thin objects would be lost, when taking the wavefunction according to the POA from equation (1.2.1). Since all information about the object structure is then contained in the phase of $\Psi_\mathrm{e}$, the square of the amplitude is constant $A^2(\vec{r}) = 1$ over the whole real space and the image would show absolutely no

information about the object.
In order to be able to observe any signal from a thin object, the EPW must be modified "somehow", before taking the modulus square in the image plane. In practice this modification of the EPW is intentionally achieved by the transfer properties of the objective lens of the electron microscope.
Since the transfer properties of the objective lens are mathematically described in Fourier space, one has to apply a Fourier transform to the formulation of the image intensity in equation (1.3.1). The mathematical expression of the Fourier transform of the image intensity is derived in appendix A.3. The result of this derivation is

$$I(\vec{g}\,) = \int \Psi_{\mathrm{i}}(\vec{g} + \vec{k}\,)\, \Psi_{\mathrm{i}}^{*}(\vec{k}\,)\, \mathrm{d}^2\vec{k}\,, \tag{1.3.2}$$

where the Fourier-space vector $\vec{g}$ is conjugated to the real-space distance vector $\vec{r}$.
According to equation (1.3.2) the image is formed by integration over pairwise interference terms of two beams with wave vectors $(\vec{g} + \vec{k}\,)$ and $\vec{k}$. In a discretised formulation, every Fourier component $I(\vec{g}\,)$ of the image intensity is a sum over all interference terms of two beams with a wave-vector difference, or diffraction vector, of $\vec{g}$. In real space, each interference term adds a sinusoidal modulation with period $1/|\vec{g}\,|$, a so-called interference fringe to the image intensity. Transmission electron microscopic images are therefore interference patterns.
Since the consideration of the full non-linear image formation described by equation (1.3.2) is too complicated for a basic introduction to the theory of electron-optical contrast transfer, the non-linear contributions will be discussed at the end of this chapter in section 1.3.6.

### 1.3.2 Linear imaging theory for thin objects

A crucial simplification of the full non-linear image formation in equation (1.3.2) arises, when only interference terms between the undiffracted beam $\Psi_{\mathrm{i}}(0)$ and the diffracted beams $\Psi_{\mathrm{i}}(\vec{g}\,)$ are considered. This restriction to the so-called linear interference terms is justified for thin objects in the frame of the WPOA. In the derivation of equation (1.2.2) the real part of the EPW, which corresponds to the undiffracted beam, is assumed to be much larger than the imaginary part, which is formed by the undiffracted beam. The pairwise mutual interference of diffracted beams is thus much smaller and of second order compared to the first-order interferences involving the undiffracted beam.
The separation into linear and non-linear terms is more intuitive, when equation (1.3.2)

is written as a discrete sum over all wave vectors $\vec{k}$ :

$$I(\vec{g}) = \sum_{\vec{k}} \Psi_{\mathrm{i}}(\vec{g}+\vec{k})\,\Psi_{\mathrm{i}}^{*}(\vec{k}). \tag{1.3.3}$$

When neglecting the non-linear terms with $\vec{k} \neq 0$ and $\vec{k} \neq -\vec{g}$, the linear image $I_{\mathrm{L}}$ composes of Fourier components given by

$$I_{\mathrm{L}}(\vec{g}\neq 0) = \Psi_{\mathrm{i}}^{*}(0)\Psi_{\mathrm{i}}(\vec{g}) + \Psi_{\mathrm{i}}(0)\Psi_{\mathrm{i}}^{*}(-\vec{g}). \tag{1.3.4}$$

The total number of interference terms, which contribute to the intensity of the image is drastically reduced, since only $2(N-1)$ linear terms remain in equation (1.3.4) compared to $N^2$ linear and non-linear terms in equation (1.3.3) when considering $N$ beams. In the linear theory, the self-interference $\Psi_{\mathrm{i}}(0)\Psi_{\mathrm{i}}^{*}(0)$ of the directly transmitted, undiffracted beam determines the mean image intensity.

It has been stated by the use of the real-space formulation of the image intensity (1.3.1) that the object information contained in the phase of $\Psi_{\mathrm{i}}$ is lost upon imaging, $I(\vec{r}) = A^2(\vec{r})$. In the Fourier representation of the image intensity, this information loss is reflected in a different way. Since the image intensity is real-valued, its Fourier components have to obey to the Friedel symmetry, i.e. $I(-\vec{g}) = I^{*}(\vec{g})$. This constraint is directly implemented in equation (1.3.4) by the superposition of the information carried by a beam of diffraction vector $\vec{g}$ with that carried by a beam of diffraction vector $-\vec{g}$.

In order to continue with the linear imaging theory for thin objects, the resulting electron wavefunction obtained by the WPOA in equation (1.2.2) is likewise transformed to Fourier space as

$$\Psi_{\mathrm{e}}(\vec{g}) = \delta(\vec{g}) + \mathrm{i}\sigma V_{\mathrm{P}}(\vec{g})t, \tag{1.3.5}$$

with the interaction constant $\sigma$, and the object thickness $t$. The real part of the EPW behaves like a plane wave, which is represented by $\delta(\vec{g})$ with $\int \delta(\vec{g})\,\mathrm{d}^2\vec{g} = 1$. The Fourier transform $V_{\mathrm{P}}(\vec{g})$ of the projected crystal potential $V_{\mathrm{P}}(\vec{r})$ also obeys to the Friedel symmetry $V_{\mathrm{P}}(-\vec{g}) = V_{\mathrm{P}}^{*}(\vec{g})$ because the projected potential is real-valued. When inserting equation (1.3.5) into equation (1.3.4), the dilemma that no contrast is obtained for an ideal microscope with $\Psi_{\mathrm{i}} = \Psi_{\mathrm{e}}$, persists also in the linearised Fourier-space formulation of the image intensity within the frame of the WPOA.

Mathematically, a phase factor has to be applied, in order to transfer phase information into the finally observed image intensity using the POA, or to transfer information from the imaginary part of the wavefunction into the finally observed real part of the wavefunction using the WPOA.

Let us therefore assume a general phase shift $\chi$, which is applied by multiplication of the phase factor $\exp[-\mathrm{i}\chi]$ to the exit-plane wavefunction $\Psi_\mathrm{e}$ in its Fourier-space representation in equation (1.3.5). The resulting image-plane wavefunction $\Psi_\mathrm{i}$ is then given by

$$\Psi_\mathrm{i}(\vec{g}\,) = \Psi_\mathrm{e}(\vec{g}\,)\exp[-\mathrm{i}\chi]. \tag{1.3.6}$$

Now the modified image-plane wavefunction $\Psi_\mathrm{i}$ of equation (1.3.6) is inserted into equation (1.3.4) instead of the exit-plane wavefunction $\Psi_\mathrm{e}$ of equation (1.3.5). The obtained result is a well-known expression for the linear image intensity of a weak phase object:

$$\begin{aligned} I_\mathrm{L}(\vec{g}\neq 0) &= \mathrm{i}\sigma V_\mathrm{P}(\vec{g}\,)\,t\exp[-\mathrm{i}\chi] - \mathrm{i}\sigma V^*(-\vec{g}\,)\,t\exp[+\mathrm{i}\chi], &(1.3.7)\\ &= 2\sigma V_\mathrm{P}(\vec{g}\,)t\,\sin[\chi] &(1.3.8)\end{aligned}$$

The phase factor $\exp[-\mathrm{i}\chi]$ has been introduced first on pure mathematical reasoning without clarification of its physical origin. Nevertheless, one may easily recognise that optimum contrast for a weak phase object is obtained when $\sin[\chi] = \pm 1$, or when the phase shift $\chi$ is an odd multiple of $\pi/2$. In light optics a phase shift of $\pi/2$ or $\lambda/4$ can be achieved by inserting a so-called $\lambda/4$-phase-plate into the ray path. Unfortunately, such a device is not available for electron optics. In electron microscopy, the phase shift $\chi$ describes the coherent transfer properties of the objective lens and is called the aberration function.

### 1.3.3 Fundamental aberrations: spherical aberration and defocus

Although the aberration function can be derived within the wave-optical formalism in Fourier space, it is more illustrative to discuss the fundamental aberrations of the objective lens first by means of the more familiar ray optics in real space.

The imaging property of a lens which is affected by a spherical aberration is depicted in figure 1.9 in comparison to a perfect, aberration-free lens. For a perfect lens all optical ray paths starting from one point in the object plane intersect again in one point in the image plane. A spherical aberration causes a too strong deflection of rays passing the outer part of the lens, compared to the deflection of near-axis rays. The too strongly deflected rays are focused in front of the image plane. The spherical aberration is a fixed property of the objective lens. The spherical-aberration parameter characterising a lens will be denoted by $C_\mathrm{S}$ in the following. Values of $C_\mathrm{S} \approx 1$ mm are typical for high-resolution electron microscopes.

Due to the spherical aberration a sharp point is imaged into a disk-like intensity distribution. The disk is called error disk, or disk of confusion. The size of the error disk

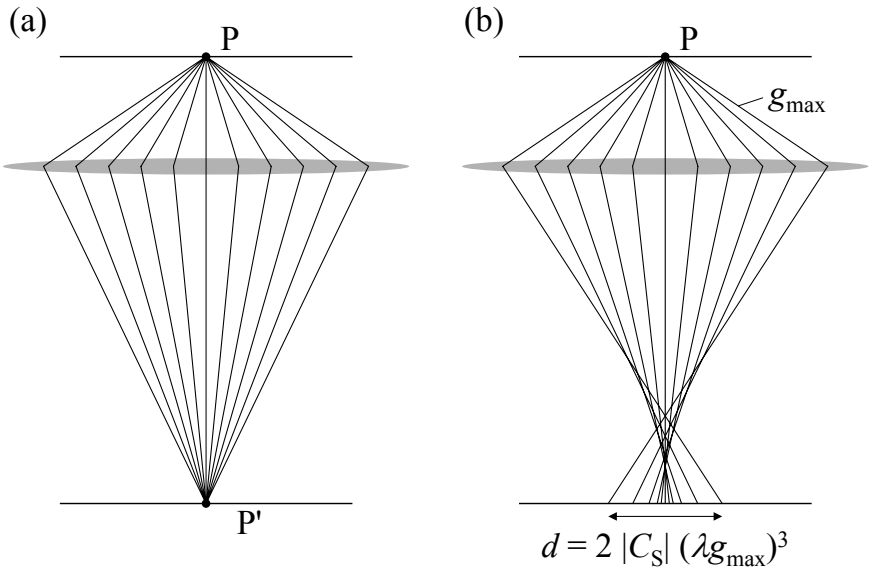



Figure 1.9: Illustration of the spherical aberration: (a) A perfect lens images one point P in the exit plane into one point P' in the image plane. (b) A lens affected by a spherical aberration. The spherical aberration causes a spread of information stemming from one point of the object over a disk-like area of diameter $d$. The diameter $d$ depends on the strength of the spherical aberration $C_S$, the electron wavelength $\lambda$, and on the highest involved diffraction-vector modulus $g_{max}$.

increases with the third power of the modulus $g_{max}$ of the highest involved diffraction vector. Therefore the smallest transferred object details suffer most from the effect due to the spherical aberration.

The focal length of the objective lens can be changed by adjusting the current, which flows through the lens coils. Changing the defocus moves the Gaussian image plane up or down by the distance $Z$, which is the defocus parameter. A defocus with $Z > 0$ is called overfocus (higher lens excitation), whereas $Z < 0$ is called underfocus (lower lens excitation). Figure 1.10 illustrates, that a defocus causes a blurring of the ideally sharp point in the image. This blurring is described by an error disk analogous to the description of the spherical aberration. The size of the error disk caused by a defocus $Z$ depends only linearly on the largest involved diffraction vector $g_{max}$.

In order to translate the ray-optical real-space description of the spherical aberration and the defocus of the objective lens to the wave-optical formalism in Fourier space, the observed ray-path differences from the object plane to the image plane can be formulated as a phase shift of the electron wavefunction [21,45]. This phase shift is expressed by the aberration function $\chi(\vec{g})$, which combines the effect of the spherical aberration $C_S$ and

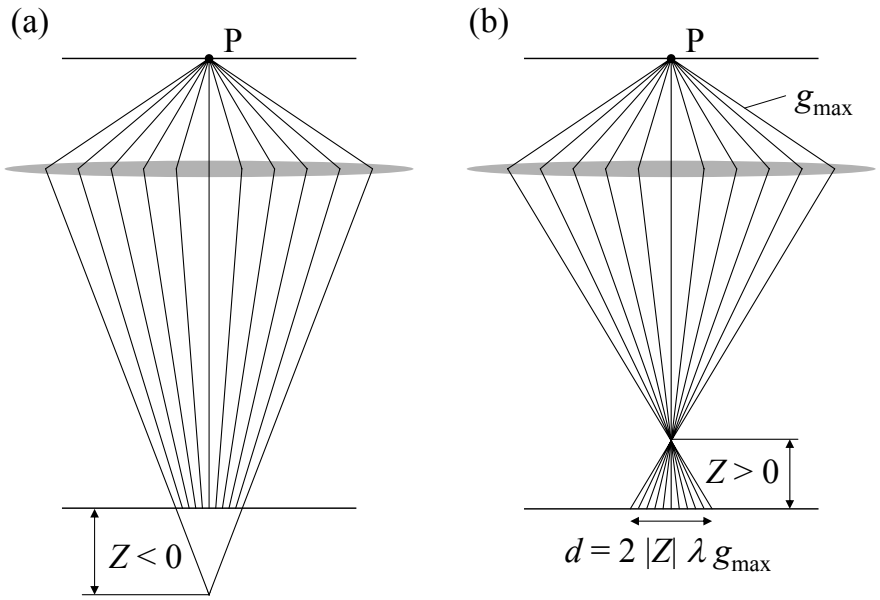



Figure 1.10: Illustration of the defocus by ray paths for (a) underfocus, and (b) overfocus. In overfocus, the rays originating from one point in the object plane intersect above, for underfocus below the actual image plane. A defocus causes a blurring of an ideally sharp point in the image. The diameter $d$ of the related error disk depends linearly on the modulus of defocus parameter $Z$ and on the modulus $g_{\max}$ of the highest involved diffraction vector.

the defocus $Z$ in one formula:

$$\chi(g) = 2\pi \left( \frac{1}{2} Z \lambda g^2 + \frac{1}{4} C_{\mathrm{S}} \lambda^3 g^4 \right). \tag{1.3.9}$$

The aberration function depends only on the modulus $g = |\vec{g}\,|$, since all aberrations, defocus and spherical aberration, considered in the present context are rotationally symmetric. The modulus $g$ of the diffraction vector $\vec{g}$ is called *spatial frequency*.

### 1.3.4 Coherent linear contrast transfer

The first factor ( $2\sigma V_{\mathrm{P}}(\vec{g})t$ ) of the linear image intensity in equation (1.3.8) depends linearly on the projected object potential. The second factor ( $\sin[\chi(g)]$ ) depends on the aberration function $\chi(g)$ as in equation (1.3.9) and thus purely on the transfer properties of the objective lens. This term constitutes the coherent *contrast transfer function* (CTF):

$$\mathrm{CTF}(g) = \sin[\chi(g)]. \tag{1.3.10}$$

The CTF determines the amount and the sign of information about the object structure (in terms of the projected potential $V_{\mathrm{P}}(\vec{g}\,)$) belonging to a certain spatial frequency $g$ which is transferred from the exit-plane wavefunction $\Psi_{\mathrm{e}}(\vec{g}\neq 0)$ to the linear image.

### Optimum contrast transfer

As discussed in section 1.3.2, ideal contrast is obtained when $\chi$ is an odd multiple of $\pi/2$ and constant for all spatial frequencies. However, the dependence of the aberration function on $g^4$ due to a spherical aberration, in competition to $g^2$ due to a defocus, does not allow to generate such a condition.
In practice, a good compromise can be found, when the two terms in the aberration function compensate each other near a value of $\pi/2$. In order to achieve such a compensation, one of the two aberrations should cause a positive phase shift, whereas the other should cause a negative phase shift. Since the spherical aberration is a fixed parameter for conventional transmission electron microscopes, only the defocus parameter can be adjusted to find an optimum condition for the contrast transfer.
From the constrictions to enable a constant contrast transfer of $\mathrm{CTF}(\vec{g}) = \pm 1$ over a maximum band of spatial frequencies, Scherzer derived an optimum defocus value

$$Z_\mathrm{S} \approx -\sqrt{\frac{4}{3} C_\mathrm{S} \lambda}, \tag{1.3.11}$$

such that the first zero-crossing of the coherent contrast transfer-function appears at the spatial frequency $g_\mathrm{S} = (3/16\ C_\mathrm{S}\lambda^3)^{-1/4}$ [46]. The *Scherzer-defocus* $Z_\mathrm{S}$ guarantees a straightforward contrast transfer up to the so-called *point resolution* $d_\mathrm{S} = 1/g_\mathrm{S}$.

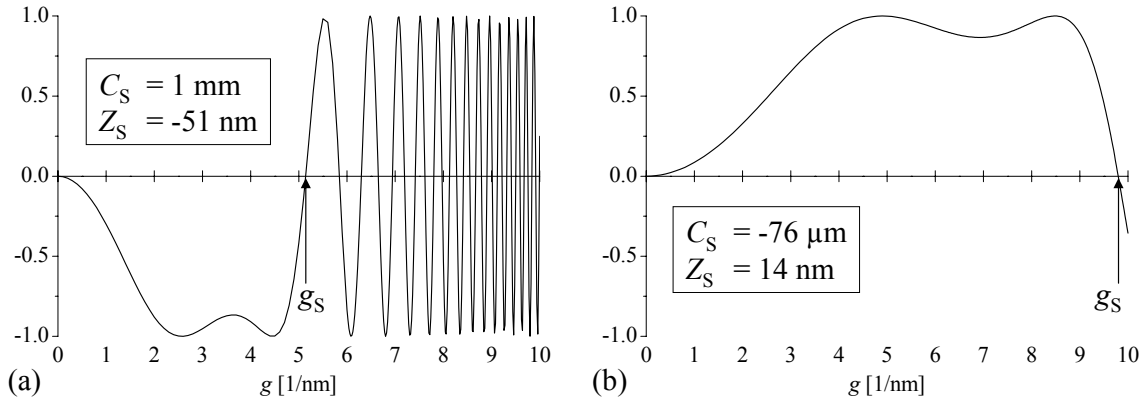



Figure 1.11: Coherent contrast-transfer functions in Scherzer-defocus $Z_\mathrm{S}$ (a) for a conventional TEM with a large spherical aberration and (b) a corrected TEM with a small negative spherical aberration. In both diagrams an accelerating voltage of 300 kV is considered.

Two examples for the coherent CTF in Scherzer-defocus are shown in figure 1.11. Subfigure 1.11a refers to the case of a conventional TEM with a large spherical aberration of $C_\mathrm{S} = 1$ mm. The respective Scherzer-focus at 300 kV accelerating voltage ($\lambda = 1.969$ pm) is $Z_\mathrm{S} = -51$ nm. The lowest spatial frequencies are transferred with almost no contrast,

which is frequently referred to as the *low-frequency gap* in HRTEM. Due to the negative defocus, the maxima of the object potential $V(\vec{g})$ are transferred with negative phase contrast for the spatial frequencies in the passband $[0, g_{\mathrm{S}}]$. Accordingly, the atom columns are imaged with dark contrast. For $g > g_{\mathrm{S}}$ the contrast transfer begins to oscillate, since the $g^4$-term in the aberration function now dominates over the defocus term. A straightforward interpretation of image details which correspond to these higher spatial frequencies is not possible.

It is obvious from the coherent CTF displayed in figure 1.11b that spherical-aberration correction bears an enormous advantage for imaging in HRTEM. With a corrector for the spherical aberration both aberrations, $C_{\mathrm{S}}$ and defocus, can be adjusted to positive and negative values. This allows one to choose the first zero-crossing $g_{\mathrm{S}}$ and the sign of the CTF freely. By this way, the passband can in principle be extended up to an arbitrary point-resolution. However, this extension also broadens the unwanted low-frequency gap. The so-called *negative spherical aberration imaging* (NCSI) technique [47–49] provides optimum bright atom contrast up to the information limit $g_{\max}$, which denotes the highest spatial frequency transferred by the microscope. Bright atom contrast is achieved by choosing a small negative spherical aberration of

$$C_{\mathrm{S}} = -\frac{64}{27}(\lambda^3 g_{\max}^4)^{-1} \tag{1.3.12}$$

and an overfocus of

$$Z = \frac{16}{9}(\lambda g_{\max}^2)^{-1}. \tag{1.3.13}$$

Besides the optimum exploitation of the information limit, the NCSI-method furthermore yields an improved contrast for the non-linear interference terms and improves the visibility of weakly scattering atom columns in the vicinity of heavier atoms [48, 50].

### Delocalisation

The blurring of object information over a larger area in the image due to the presence of strong aberrations is frequently referred to as *delocalisation*. The radius $\rho$ of the area of blurring, the so-called delocalisation radius, is related to the maximum of the gradient of the aberration function $\chi(\vec{g})$ in the spatial frequency range up to the information limit $g_{\max}$ [24, 51]. This relation is expressed by the formula

$$\rho = \max\left\{\frac{1}{2\pi}|\vec{\nabla}\chi(\vec{g})|\right\}_{g=0\ldots g_{\max}} \tag{1.3.14}$$

The delocalisation in the image becomes minimum in the so-called defocus of least confusion or *Lichte-defocus* [52] which is defined by

$$Z_{\mathrm{L}} = -\frac{3}{4} C_{\mathrm{S}} \lambda^2 g_{\max}^2. \tag{1.3.15}$$

In contrast to the Scherzer-defocus which provides optimum imaging conditions up to the point resolution $g_{\mathrm{S}}$, the Lichte-defocus is optimised for minimum delocalisation up to the information limit $g_{\max}$.

### 1.3.5 Partially coherent, linear contrast transfer

So far, a perfectly coherent electron illumination characterised by a sharp wavelength and by absolutely plane wave-fronts was assumed. According to the Rayleigh-criterion in optics, the resolution of a microscope with coherent illumination is limited by the electron wavelength and the numerical aperture of the objective lens [51]. In reality, the resolution of transmission electron microscopes is unfortunately worse by two orders of magnitude compared to the wavelength of a few picometres.

#### Resolution limiting effects

Essentially four reasons can be identified, which lead to a reduction of the resolution of a transmission electron microscope:

1. The incident electron wave-field is not monochromatic, but consists of a spectrum of waves with different wavelengths. This effect is addressed as partial temporal coherence of the electron wave-field.

2. The incident wavefunction can not be described by a single plane wave, it is instead a superposition of plane waves propagating in slightly different directions. This divergence of the electron wave-vectors is referred to as the partial spatial coherence of the electron wave-field.

3. During the exposure time of typically 1 s, drift and vibrations of the object may smear out the finest details in the image.

4. When using a CCD-camera, the detection process causes an additional damping of higher image frequencies. The resolution of a CCD is sharply limited by the number of pixels. Additionally, several charge collectors can be excited by one electron. The actual detector resolution is therefore even worse than the discretisation defined by the pixel diameter.

The reduction of the microscope's resolution due to drift and vibrations of the object, and due to the limited resolution of the detector will be addressed at the end of this section. The following descriptions of the partial temporal and partial spatial coherence are based on the notation of Frank [53] and Ishizuka [54]. Both effects cause a damping of the contrast transfer mainly at higher spatial frequencies which is mathematically accounted for by two envelopes for the image intensity in equation (1.3.8).

### Partial temporal coherence

The incident electron wave-field is not monochromatic but consists of a continuous spectrum of wavelengths. Mainly three mechanisms cause a spread of the electron energy:

- Firstly, the electron energy follows a Maxwell-distribution after the emission. The two emission principles, thermionic and field-driven, cause a spread of the electron energy of less than 1 eV. This distribution width is actually small compared to the total electron energy after acceleration by several hundred kV.

- Secondly, the acceleration voltage fluctuates with a relative width of typically $10^{-6}$ which additionally broadens the wavelength distribution.

- The third reason does not directly influence the electron energy. Temporal variations of the objective-lens current lead to a variation of its focal length. These fluctuations have also a relative width of approximately $10^{-6}$. Each electron which contributes to the image will be focussed in a slightly different plane. This variation of the focal length can formally be treated in the same way as a variation of the electron energy.

All three phenomena can be re-formulated as an effective defocus variation $\Delta$. The Maxwell-distribution of the electron energy $E$, and the temporal fluctuations of accelerating voltage $U$ and objective lens current $I$ are then approximated by Gaussian distributions. These independent events are combined by Gaussian error-propagation to one quantity, the defocus spread $\Delta$ [55], with

$$\Delta = C_{\mathrm{C}} \sqrt{\left(\frac{\Delta E}{E}\right)^2 + \left(\frac{\Delta U}{U}\right)^2 + \left(\frac{2\Delta I}{I}\right)^2}. \tag{1.3.16}$$

$\Delta E$, $\Delta U$, and $\Delta I$ denote the width of distribution for the three respective quantities, the *chromatic aberration* $C_{\mathrm{C}}$ is a constant parameter of the electron microscope. Typical values of the defocus spread range between 2 - 10 nm.

In the theory of the partial temporal coherence, the effective focal distribution function $f(Z)$ is described by a normalised Gaussian distribution, with

$$f(Z) = \frac{1}{\sqrt{\pi}\Delta}\, \mathrm{e}^{-\frac{Z^2}{\Delta^2}}. \tag{1.3.17}$$

Consequently, the final image can be seen as an incoherent superposition of coherent partial images which correspond to different defocus values. This superposition is weighted with the above distribution function $f(Z)$.

### Partial spatial coherence

The cause for a partial spatial coherence is a non-parallel illumination of the object. In practice, the illumination cannot be kept completely parallel due to a finite size of the electron source. Furthermore the beam is frequently converged intentionally during the experiment in order to increase the image intensity and thereby reduce the exposure-time. As a result, the incident electron beam is not parallel. Instead, a continuous superposition of plane waves with slightly different directions of the wave vectors has to be considered. The according wave vectors $\vec{k}$ comprise a component $\vec{k}_0$ along the optical axis and a small component $\vec{q}$ perpendicular to the optical axis.
By assuming $\vec{q} \ll \vec{k}_0$, the electron diffraction is in first-order approximation the same as for ideal, parallel illumination. It is therefore not required to calculate the exit-plane wavefunction for each of the incident plane waves separately. The tilted illumination is accounted for by superposing coherent images, based on the same EPW. The distribution of the incidence directions is described by a normalised Gaussian distribution function:

$$s(\vec{q}) = \frac{1}{\pi q_0^2}\, \mathrm{e}^{-\frac{\vec{q}^2}{q_0^2}}. \tag{1.3.18}$$

The angle of semi-convergence $\theta_0 = \lambda q_0$ denotes the half opening angle of the convergent incident ray-cone. Values of $\theta_0 \approx 1$ mrad are typical for thermionic emitters like $LaB_6$ sources. In contrast, FEG-sources allow considerably smaller angles of semi-convergence of approximately 0.2 mrad. Alternatively, a cylindrical profile of the distribution $s(\vec{q})$ is discussed in the literature [56, 57]. Nevertheless, the Gaussian distribution (1.3.18) is commonly preferred as the mathematically more comfortable formulation.

### Consideration of partial coherence in the contrast transfer theory

For narrow distributions of the electron wavelength and for small angles of semi-convergence, the effects of partial temporal and partial spatial coherence on the electron diffraction are negligible. In theory, they are thus considered to solely affect the electron-optical

contrast transfer which is formally described by a weighted superposition of coherent partial images. This superposition is expressed by integration of the coherent contrast transfer function in equation (1.3.10) over a variation $Z'$ of the mean defocus $Z$ and over a variation $\vec{q}$ of the diffraction vector $\vec{g}$, using the distribution functions $f(Z')$ and $s(\vec{q})$ as respective weighting factors. The partially coherent contrast-transfer function $\mathrm{CTF_p}(\vec{g})$ is accordingly given by

$$\mathrm{CTF_p}(\vec{g}) = \iint f(Z')s(\vec{q})\sin[\chi(\vec{g}+\vec{q}, Z+Z')]\,\mathrm{d}Z'\,\mathrm{d}^2\vec{q}, \tag{1.3.19}$$

where $\chi(\vec{g}+\vec{q}, Z+Z')$ denotes the aberration function as in equation (1.3.9).
In order to compute a more simplified expression for $\mathrm{CTF_p}$, the integrals have to be solved. Therefore the sine-function is written in form of exponential phase factors:

$$\begin{aligned}\mathrm{CTF_p}(\vec{g}) &= \iint f(Z')s(\vec{q})\sin[\chi(\vec{g}+\vec{q}, Z+Z')]\,\mathrm{d}Z'\,\mathrm{d}^2\vec{q} &(1.3.20)\\ &= \frac{1}{2\mathrm{i}}\iint f(Z')s(\vec{q})\,\{\exp[\mathrm{i}\chi(\vec{g}+\vec{q}, Z+Z')] &(1.3.21)\\ &\qquad -\exp[-\mathrm{i}\chi(\vec{g}+\vec{q}, Z+Z')]\}\,\mathrm{d}Z'\,\mathrm{d}^2\vec{q}\end{aligned}$$

Good approximations for the separate integration of the two exponential terms are found by a first-order Taylor expansion of the aberration function with respect to $\vec{q}$ and $Z'$:

$$\chi(\vec{g}+\vec{q}, Z+Z') \approx \chi(\vec{g}, Z) + \vec{q}\cdot\vec{\nabla}\chi(\vec{g}, Z) + Z'\cdot\frac{\partial\chi(\vec{g}, Z)}{\partial Z} + \mathcal{O}(Z'^2, \vec{q}^{\,2}), \tag{1.3.22}$$

where $\vec{\nabla}$ is the derivative with respect to the two spatial components of $\vec{g}$. Expansions of the aberration function containing mixed and second-order derivatives have been described by Ishizuka [54], and Bonevich and Marks [58, 59].
Substitution of $\chi$ in equation (1.3.21) by the first-order approximation (1.3.22) and insertion of the explicit expressions (1.3.17) and (1.3.18) for the two distribution functions $f(Z')$ and $s(\vec{q})$ yields a compact, approximative expression, with

$$\mathrm{CTF_p}(\vec{g}) \approx E_\mathrm{t}(\vec{g})\,E_\mathrm{s}(\vec{g})\sin[\chi(\vec{g})]. \tag{1.3.23}$$

The functions $E_\mathrm{t}(\vec{g})$ and $E_\mathrm{s}(\vec{g})$ describe a damping of the contrast transfer due to partially temporal coherent and partially spatial coherent electron beam. The partially coherent linear image of a weak phase object is therefore approximately given by

$$I_\mathrm{L}(\vec{g}\neq 0) \approx 2\sigma V_\mathrm{P}(\vec{g})t\,E_\mathrm{t}(\vec{g})\,E_\mathrm{s}(\vec{g})\sin[\chi(\vec{g})]. \tag{1.3.24}$$

The explicit expressions for the envelope functions are rotationally symmetric in the diffraction vector and depend thus merely on the spatial frequency $g$:

$$E_{\mathrm{t}}(g) = \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2 g^4\right\}, \tag{1.3.25}$$

$$E_{\mathrm{s}}(g) = \exp\left\{-\left(\frac{q_0}{2}\right)^2 \left[\frac{\partial}{\partial g}\chi(g)\right]^2\right\}. \tag{1.3.26}$$

The *temporal envelope* $E_{\mathrm{t}}$ describes a damping which becomes stronger with a larger defocus spread $\Delta$ and with the fourth power of the spatial frequency $g$. The damping due to the partial spatial coherence, as described by the *spatial envelope* $E_{\mathrm{s}}$, is stronger for a larger angle of semi-convergence $\theta_0 = \lambda q_0$ and depends on the gradient of the aberration function. Thus, larger aberrations cause, in general, a stronger damping.

For conventional microscopes with thermionic emitters both envelopes constitute a substantial limitation to the maximum transferred spatial frequency, as illustrated in figure 1.12a. For a larger defocus, the damping described by the spatial envelope may even become stronger than the respective temporal envelope. Figure 1.12b illustrates a case for a large underfocus, where $E_{\mathrm{s}}$ dampens the contrast transfer already at very low spatial frequencies but rises again for a narrow pass-band. At the respective spatial frequencies, the terms of spherical aberration and defocus compensate in the gradient of the aberration function.

The partial temporal coherence is the physically limiting factor for the contrast transfer of FEG-microscopes. Due to the small angle of semi-convergence, $E_{\mathrm{s}}$ plays no role in images recorded in Scherzer-defocus especially for spherical-aberration corrected FEG-microscopes as illustrated in figure 1.12c. Even for a larger underfocus as in figure 1.12d, the temporal envelope constitutes the dominant damping at higher spatial frequencies.

### Specification of the information limit

The information limit $d_{\mathrm{info}}$ denotes the smallest object detail which can be transferred in a linear way to the image. Due to the defocus dependence of the spatial envelope $E_{\mathrm{s}}$, a specification of the information limit by means of both envelopes in equation (1.3.24) would be ambiguous. Therefore, $d_{\mathrm{info}}$ is traditionally derived from the temporal envelope, which is independent on the defocus, by the spatial frequency $g_{\max} = 1/d_{\mathrm{info}}$, where $E_{\mathrm{t}}(\vec{g}\,)$ still yields a value of $1/\mathrm{e}^2$. This value corresponds to approximately 13.5 % of the full contrast. According to equation (1.3.25) it follows that

$$d_{\mathrm{info}} = \frac{1}{g_{\max}} = \left(\frac{\pi^2\lambda^2\Delta^2}{8}\right)^{\frac{1}{4}}. \tag{1.3.27}$$

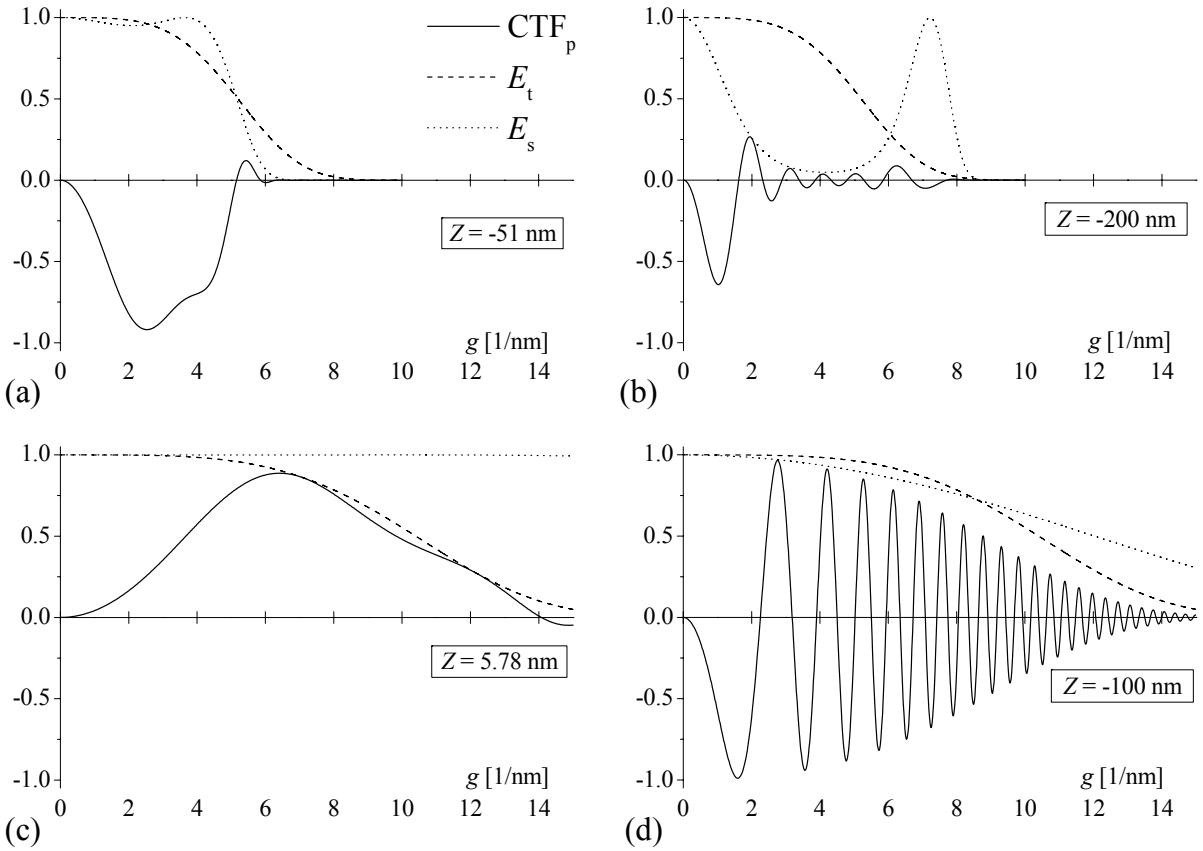



Figure 1.12: Partially coherent CTF (solid curves), temporal envelope $E_\mathrm{t}$ (dashed curves), and spatial envelope $E_\mathrm{s}$ (dotted curves). The diagrams (a) and (b) are calculated for a conventional TEM with thermionic emitter ($\theta_0 = 1$ mrad, $\Delta = 10$ nm, $C_\mathrm{S} = 1$ mm, $U = 300$ kV), whereas (c) and (d) concern a $C_\mathrm{S}$-corrected FEG-mircoscope ($\theta_0 = 0.2$ mrad, $\Delta = 2.5$ nm, $C_\mathrm{S} = -13$ $\mu$m, $U = 300$ kV). In the left diagrams (a) and (c), the defocus is equal to the respective Scherzer-defocus, whereas a large underfocus is considered in (b) and (d).

The information limit $d_\mathrm{info}$ reflects the physical limits for the resolution of a transmission electron microscope. Throughout this work, the spatial frequency $g_\mathrm{max}$ is consecutively used as a synonym for the information limit.

The information limit for a $LaB_6$-microscope with $\Delta = 10$ nm and $\lambda = 1.969$ pm as considered in the figures 1.12a and 1.12c is $g_\mathrm{max} = 6.8$ nm$^{-1}$. This value corresponds to a resolution of 1.5 Å. For the FEG-microscope with $\Delta = 2.5$ nm and $\lambda = 1.969$ pm (figs. 1.12b and 1.12d), equation (1.3.27) yields an information limit of $g_\mathrm{max} = 13.5\mathrm{nm}^{-1}$ which corresponds to a resolution of 0.7 Å.

### Damping due to drift, vibration, and detector resolution

As discussed in the beginning of this section, instabilities of the object stage and the limited resolution of the imaging detector cause an additional damping of the image contrast. In contrast to the partial coherence of the electron beam, which causes a damping in dependence on object frequencies, the damping due to instabilities of the object and due to a limited resolution of the detector depends on image frequencies. Denoting the image contrast obtained under consideration of the partial coherence in equation (1.3.24) by $I'(\vec{g})$, the additional damping is described by an additional envelope $E_{\mathrm{d}}(\vec{g})$ such that the final image intensity is given by $I(\vec{g}) = E_{\mathrm{d}}(\vec{g})\, I'(\vec{g})$. The envelope $E_{\mathrm{d}}(\vec{g})$ is a product of three envelope functions which consider the three different sources of damping separately [60–62]:

$$E_{\mathrm{d}}(\vec{g}) = E^{\mathrm{M}}(\vec{g}) \cdot E^{\mathrm{D}}(\vec{g}) \cdot E^{\mathrm{V}}(\vec{g}). \tag{1.3.28}$$

$E^{\mathrm{M}}(\vec{g})$ is called the modulation transfer function (MTF) of the detector:

$$E^{\mathrm{M}}(\vec{g}) = \exp\left\{-c_{\mathrm{MTF}}|\vec{g}|^2\right\}. \tag{1.3.29}$$

The damping constant $c_{\mathrm{MTF}}$ is a specific parameter for a CCD-camera. The image is blurred due to a point spread of the CCD scintillator, since several neighbouring pixels can be excited by one incident electron.
$E^{\mathrm{D}}(\vec{g})$ and $E^{\mathrm{V}}(\vec{g})$ denote the damping caused by drift and vibrations of the object during the exposure of the image. Both effects can be approximated by envelope functions of Gaussian shape:

$$E^{\mathrm{D}}(\vec{g}) = \exp\left\{-c_{\mathrm{D}}(\vec{u}\cdot\vec{g})^2\right\}, \tag{1.3.30}$$

$$E^{\mathrm{V}}(\vec{g}) = \exp\left\{-c_{\mathrm{V}}(\vec{v}\cdot\vec{g})^2\right\}. \tag{1.3.31}$$

The vector $\vec{u}$ denotes the mean drift velocity, while the mean amplitude and orientation of object vibrations is specified by $\vec{v}$, with the constants $c_{\mathrm{D}}$ and $c_{\mathrm{V}}$. In contrast to the MTF, the envelopes $E^{\mathrm{D}}(\vec{g})$ and $E^{\mathrm{V}}(\vec{g})$ are formally anisotropic. However, during the exposure time of 1 s, vibrations may occur in different directions and cause an isotropic damping.

## 1.3.6 Partially coherent, non-linear contrast transfer

The discussion of the electron-optical contrast transfer in this section takes also the non-linear interference terms into account which are neglected in the frame of the weak-phase-object approximation in section 1.3.2. The non-linear theory is frequently used for realistic image simulations in HRTEM [7], since the objects used in the experiments do often not

fulfil the extreme thickness-limitations demanded by the WPOA.
The image contrast including all linear and non-linear interference terms is given by equation (1.3.2) with

$$I(\vec{g}\,) = \int \Psi_{\mathrm{i}}(\vec{g}+\vec{k}\,)\,\Psi_{\mathrm{i}}^{*}(\vec{k}\,)\;\mathrm{d}^2\vec{k}\,. \tag{1.3.32}$$

Analogous to the linear contrast-transfer theory, also the non-linear theory describes the modulation of the electron wavefunction due to the aberrations, defocus $Z$ and spherical aberration $C_{\mathrm{S}}$, by a phase factor $\tau(\vec{g}\,) = \exp[-\mathrm{i}\chi(\vec{g}\,)]$, using the aberration function $\chi(\vec{g}\,)$ from equation (1.3.9). This phase factor $\tau(\vec{g}\,)$ is the so-called *coherent transfer function* and denotes the difference between the image-plane wavefunction $\Psi_{\mathrm{i}}$ and the exit-plane wavefunction $\Psi_{\mathrm{e}}$ by

$$\Psi_{\mathrm{i}}(\vec{g}\,) = \Psi_{\mathrm{e}}(\vec{g}\,)\tau(\vec{g}\,). \tag{1.3.33}$$

The coherent, non-linear image intensity can thus by expressed by

$$I(\vec{g}\,) = \int \Psi_{\mathrm{e}}(\vec{g}+\vec{k}\,)\Psi_{\mathrm{e}}^{*}(\vec{k}\,)\,\tau(\vec{g}+\vec{k}\,)\tau^{*}(\vec{k}\,)\;\mathrm{d}^2\vec{k}\,. \tag{1.3.34}$$

The partially coherent contrast transfer is derived by a weighted superposition of coherent partial images as in equation (1.3.34). The respective weighting factor $T$ for each coherent sub-image is called *transmission cross-coefficient* (TCC). Accordingly, the partially coherent, non-linear image intensity is expressed by the integral

$$I(\vec{g}\,) = \int \Psi_{\mathrm{e}}(\vec{g}+\vec{k}\,)\,\Psi_{\mathrm{e}}^{*}(\vec{k}\,)\,T(\vec{g}+\vec{k},\vec{k}\,)\,\mathrm{d}^2\vec{k}\,. \tag{1.3.35}$$

The transmission cross-coefficient $T(\vec{g}+\vec{k},\vec{k}\,)$ contains the integration of $\tau(\vec{g}+\vec{k}\,)\tau^{*}(\vec{k}\,)$ as in equation (1.3.34) over the defocus variation $Z'$ and the wave vector $\vec{q}$. By setting $\vec{g}+\vec{k} = \vec{g''}$ one obtains:

$$\begin{aligned} T(\vec{g}\,',\vec{g}\,'') &= \iint f(Z')\,s(\vec{q}\,)\,\tau(\vec{g'}+\vec{q},Z+Z')\,\tau^{*}(\vec{g''}+\vec{q},Z+Z')\,\mathrm{d}Z'\,\mathrm{d}^2\vec{q} && (1.3.36)\\ &= \iint f(Z')\,s(\vec{q}\,)\,\mathrm{e}^{-\mathrm{i}\chi(\vec{g'}+\vec{q},Z+Z')}\,\mathrm{e}^{+\mathrm{i}\chi(\vec{g''}+\vec{q},Z+Z')}\,\mathrm{d}Z'\,\mathrm{d}^2\vec{q} && (1.3.37)\end{aligned}$$

Although there is no further analytical simplification of the TCC, a good approximation is found by first-order Taylor expansion of the aberration function as used in the derivation of the partially coherent, linear contrast transfer in section 1.3.5. Replacing $\chi$ in equation (1.3.37) by the first-order approximation (1.3.22) and insertion of the explicit expressions (1.3.17) and (1.3.18) for the distribution functions $f(Z')$ and $s(\vec{q}\,)$ allows one to integrate and factorise the TCC analytically to

$$T(\vec{g'},\vec{g''}) \approx \underbrace{E_{\mathrm{t}}(\vec{g'},\vec{g''})\,E_{\mathrm{s}}(\vec{g'},\vec{g''})}_{\text{partially coherent}}\;\underbrace{\exp[\mathrm{i}\chi(\vec{g''})-\mathrm{i}\chi(\vec{g'})]}_{\text{coherent}}\,. \tag{1.3.38}$$

The coherent part of the TCC describes the mutual phase-shift of the diffracted beams by the aberration function $\chi(\vec{g})$. The amplitude of the coherent part of the transmission cross-coefficient is constant and equal to one.

Two envelope functions describe a damping of the non-linear contrast transfer. The envelope $E_\mathrm{t}(\vec{g'}, \vec{g''})$ accounts for the partial temporal coherence, whereas the envelope $E_\mathrm{s}(\vec{g'}, \vec{g''})$ considers the partial spatial coherence:

$$E_\mathrm{t}(\vec{g'}, \vec{g''}) = \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2\left[\vec{g'}^2 - \vec{g''}^2\right]^2\right\} \tag{1.3.39}$$

$$E_\mathrm{s}(\vec{g'}, \vec{g''}) = \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\vec{\nabla}_{\vec{g'}}\chi(\vec{g'}) - \vec{\nabla}_{\vec{g''}}\chi(\vec{g''})\right]^2\right\} \tag{1.3.40}$$

The gradient operators in equation (1.3.40) refer explicitly to the respective argument of the aberration function. The damping envelopes (1.3.25) and (1.3.26) derived in the linear case are readily obtained by setting one of the two diffraction vectors $\vec{g'}$ or $\vec{g''}$ to zero.

The transmission cross-coefficient describes the phase- and amplitude modulation of the interference between two diffracted beams with the diffraction vectors $\vec{g'}$ and $\vec{g''}$. The respective interference term contributes to the image Fourier coefficient $I(\vec{g})$ for the Fourier-space vector $\vec{g} = \vec{g'} - \vec{g''}$. Since each $I(\vec{g})$ is a sum over all two-beam interferences with equal difference of its wave vectors, the non-linear image intensity is in general not interpretable in a straightforward way.

Due to the mere dependence of $E_\mathrm{t}(\vec{g'}, \vec{g''})$ on the spatial frequencies $g' = |\vec{g'}|$ and $g'' = |\vec{g''}|$, the interference of beams with equal spatial frequency is not damped by a partial temporal coherence. This phenomenon is called *achromatic interference*. In contrast, $E_\mathrm{s}$ depends also on the mutual orientation of $\vec{g'}$ and $\vec{g''}$, as well as on the aberrations of the imaging system. For a given modulus of $\vec{g'}$ and $\vec{g''}$, the damping is strongest if the diffraction vectors are anti-parallel, and weakest when they are parallel.

A special non-linear artefact of the imaging process is called *half-spacing*. Half-spacings arise from the interference of two beams with opposite diffraction vectors $\vec{g}$ and $-\vec{g}$. In the extreme case of $|\vec{g}| = g_\mathrm{max}$ the non-linear image contains details of the spatial frequency $2g_\mathrm{max}$, which are definitely not related to the highest transferred frequency $g_\mathrm{max}$ of the object structure.

# Chapter 2

# Wave aberrations in HRTEM

With the nowadays routinely achievable Sub-Ångström resolution, the description of the action of the objective lens by the two prominent aberrations defocus $Z$ and spherical aberration $C_S$ alone, is no longer sufficient. A large number of additional aberrations have to be considered by the aberration function in order to allow for quantitative HRTEM measurements of structural object details in the regime around one Ångström and below one Ångström. Since it is convenient to describe the electron-optical contrast transfer by a wave-optical formalism, the extended aberration function is expressed in terms of wave aberrations.

## 2.1 General description of wave aberrations

Wave aberrations denote deviations of the wavefront emitted from a point source with respect to an ideal spherical reference wave as illustrated in figure 2.1. The aberration function $\chi$ describes these deviations by optical path differences $\Delta s$ in terms of a relative phase shift of the electron wavefunction, with $\chi = 2\pi/\lambda \cdot \Delta s$.
In general, the aberration function is a polynomial which depends on both, real-space coordinates $\vec{r} = (x, y)$ and Fourier-space coordinates $\vec{g} = (g_x, g_y)$, i.e. the diffraction vector [51, 63]. In the following, the axes of the real-space coordinate system are identical to the axes of the image. This expansion in four coordinates expresses the geometrical ray-path difference between an ideal and an aberrated ray which start both from one point of the object plane in a certain direction and arrive at the image plane at different locations and in different directions.
For the treatment of wave aberrations in HRTEM, it is well-established [51, 64, 65] to examine the general aberration function for a fixed lateral real-space origin. Due to

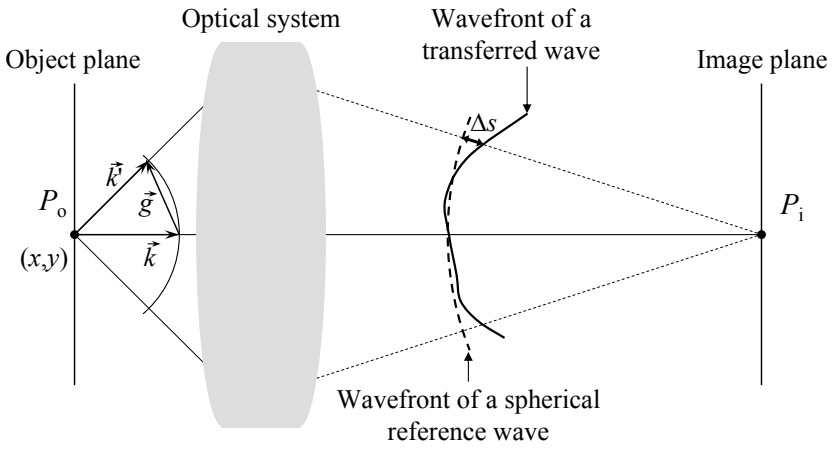



Figure 2.1: Illustration of wavefront aberrations for a cross section through the optical setup. In the absence of aberrations, the action of the optical system converts the spherical wave emitted from one point $P_o$ in the object plane to a wave with a spherical wavefront which has its centre in one point $P_i$ in the image plane. In the presence of aberrations, the wavefront is aspherical. Wave aberrations are, in general, measured by the distance $\Delta s$ between the wavefront of the ideal spherical reference wave and the aspherical wavefront of the actually transferred wave depending on the position $(x, y)$ of the point $P_o$ in the object plane and on the diffraction vector $\vec{g}$ which is the difference vector between the wave vectors $\vec{k}$ of the incident beam and the wave vector $\vec{k}'$ of a diffracted beam.

the large magnifications in HRTEM, the imaged sector of the object plane is typically small enough such that the dependence of wave aberrations on the image position can be neglected. The aberration function $\chi$ can thus be described by a polynomial expansion in the diffraction vector $\vec{g}$ only. This reduced formulation considers therefore only *axial* aberrations. In the following the adjective "axial" will be omitted. To what extent the limitation to axial terms can be justified, is investigated by a measurement of off-axis aberrations in section 8.1.1.

## 2.2 A new notation for the aberration coefficients

In addition to the most prominent aberrations, defocus $Z$ and the spherical aberration $C_S$ which have been introduced in the previous chapter, a large variety of aberrations is required to describe an arbitrary phase shift of the wavefunction during its transfer by the electron-optical system. Occasionally, non-circular aberrations like twofold astigmatism, axial coma, and threefold astigmatism are considered in the expansion of the aberration function in HRTEM [23, 66].

Higher-order aberrations influence the smallest details in the image stronger since their respective term in the aberration function depends on the spatial frequency $g = |\vec{g}|$ by

a higher power than lower order aberrations. The higher order aberrations therefore gain in importance with an increasing information limit of the electron microscopes. By utilisation of electromagnetic multipole lenses to correct the strong spherical aberration $C_{\mathrm{S}}$ especially higher-order aberrations with higher foldness of rotational symmetry are increased [20]. Until now, there is no reliable prognosis on how far the information limit can be improved by future developments in the instrumentation, and what the ultimately highest, relevant order of aberrations would be. It is therefore feasible to construct a nomenclature for the aberration coefficients which is easy to remember and easy to extend. In order to provide a straightforward identification of individual aberrations in the wave-optical formalism, a new naming convention for the aberration coefficients is proposed here. With this new nomenclature, every aberration coefficient is symbolised by a small letter $c$ together with a lower double index $(mn)$. The aberration coefficients $c_{mn}$ reflect the polynomial dependence of the respective term in the aberration function, and thus, the resulting effect of the individual aberration on the electron wavefunction.
The terms of the aberration function are in the literature [20, 67–69] frequently grouped by using a complex notation for the diffraction vector $\vec{g} = (g_x, g_y)$, with

$$g = g_x + \mathrm{i}g_y, \qquad g^* = g_x - \mathrm{i}g_y, \tag{2.2.1}$$

where $g^*$ denotes the conjugate of the complex number $g$. Non-circular aberrations $c_{mn}$ $(n \neq 0)$ are defined by a modulus $|c_{mn}|$ and an azimuth $\varphi_{mn}$ which characterise the strength and the direction of the strongest effect of an aberration respectively. The azimuth determines the angle between the direction of strongest aberration effect and the $x-$axis of the image. The aberration coefficient is given here in complex-number notation which is either expressed by two Cartesian components $c_{mnx}$ and $c_{mny}$ or in polar coordinates by modulus $|c_{mn}|$ and azimuth $\varphi_{mn}$:

$$\begin{aligned} c_{mn} &= c_{mnx} + \mathrm{i}\, c_{mny}, & (2.2.2)\\ &= |c_{mn}| \exp[\mathrm{i}\, n\varphi_{mn}], & (2.2.3)\\ \varphi_{mn} &= \frac{1}{n} \arg(c_{mn}),\ \forall\, n \neq 0. & (2.2.4) \end{aligned}$$

The notation by complex numbers allows in particular for a compact formulation of the aberration function when a large number of aberrations have to be considered. The individual terms $\chi_{mn}$ of the aberration function are thereby functions of both, $g$ and $g^*$, for $m > 0$ and $m + n = 2l$, $l \in \mathbb{N}_0$, with

$$\chi_{mn}(g, g^*) = \frac{2\pi}{\lambda} \frac{\lambda^m}{m}\ \Re\left[c_{mn} \cdot (g^*)^{(m+n)/2} \cdot g^{(m-n)/2}\right], \tag{2.2.5}$$

where $\Re[z]$ denotes the real-part of the complex number $z$. Alternatively, the terms of the aberration function can be formulated by purely real-valued quantities. The polar notation by modulus $g$ and azimuth $\varphi$ of the diffraction vector $\vec{g}$ and the aberration coefficients $(c_{mn}, \varphi_{mn})$ is given by

$$\chi_{mn}(g,\varphi) = \frac{2\pi}{\lambda}\frac{|c_{mn}|}{m}(\lambda g)^m \cdot \cos[n(\varphi - \varphi_{mn})], \tag{2.2.6}$$

The polar notation provides a more intuitive understanding of the contribution of individual aberrations to the wavefunction. Accordingly, the first index $m$ of the aberration coefficient $c_{mn}$ denotes the power-law dependence of the individual terms $\chi_{mn}$ on the spatial frequency $g$, while the foldness of the rotational symmetry is reflected by the second index $n$. For all circular aberrations ($n=0$), only the real part of the aberration coefficient is relevant, since the imaginary part is constant and equal to zero.

The complete aberration function $\chi$ is given by summation over all individual terms with

$$\chi(g,g^*) = \sum_{m,n,\ m+n=2l} \chi_{mn}(g,g^*) \tag{2.2.7}$$

$$\begin{aligned} = \frac{2\pi}{\lambda} \cdot \Re\{ & c_{00} + \frac{\lambda}{1} c_{11}\, g + \frac{\lambda^2}{2}\left[c_{20}\,(g^*)g + c_{22}\,(g^*)^2\right] \\ & + \frac{\lambda^3}{3}\left[c_{31}\,(g^*)^2 g + c_{33}\,(g^*)^3\right] \\ & + \frac{\lambda^4}{4}\left[c_{40}\,(g^*)^2 g^2 + c_{42}\,(g^*)^3 g + c_{44}\,(g^*)^4\right] \\ & + \frac{\lambda^5}{5}\left[c_{53}\,(g^*)^3 g^2 + c_{53}\,(g^*)^4 g + c_{55}\,(g^*)^5\right] \\ & + \frac{\lambda^6}{6}\left[c_{60}\,(g^*)^3 g^3 + c_{62}\,(g^*)^4 g^2 + c_{64}\,(g^*)^5 g + c_{66}\,(g^*)^6\right] + ...\} \end{aligned} \tag{2.2.8}$$

Among the multitude of individual aberrations, the spherical aberration $C_\mathrm{S}$ is equal to $c_{40}$ and the defocus $Z$ is equal to $c_{20}$.

The derivative of the aberration function is related to the displacement $\delta = \delta_x + \mathrm{i}\delta_y$ of the geometrical ray from the ideal, aberration-free ray in the Gaussian image plane [51]. The displacement $\delta$ denotes thus the total ray aberration which is related to the total wave aberration $\chi$ by

$$\delta(g,g^*) = \frac{1}{\pi}\frac{\partial}{\partial g^*}\chi(g,g^*). \tag{2.2.9}$$

Table 2.1 lists all wave aberrations up to the 6$^{\text{th}}$ order in $g$. The table also denotes two nomenclatures which are frequently used in the literature. Especially the ray-aberration notation proposed by Krivanek [39] using the symbol $C_{m,n}$ is linked to the present wave-aberration notation using the symbol $c_{mn}$ by: $C_{m-1,n} = c_{mn}$. The values of the aberration

| Aberration name | symbol here | in [20] | in [39] | magnitude |
|---|---|---|---|---|
| Constant phase offset | $c_{00}$ | - | - | |
| Image displacement | $c_{11}$ | $A_0$ | $C_{0,1}$ | |
| Defocus | $c_{20}$ | $C_1$ | $C_{1,0}$ | 1 nm – 1 $\mu$m |
| Twofold astigmatism | $c_{22}$ | $A_1$ | $C_{1,2}$ | 1 nm – 100 nm |
| Axial coma | $c_{31}$ | $\frac{1}{3}B_2$ | $C_{2,1}$ | 10 nm – 1 $\mu$m |
| Threefold astigmatism | $c_{33}$ | $A_2$ | $C_{2,3}$ | 10 nm – 1 $\mu$m |
| Spherical aberration | $c_{40}$ | $C_3$ | $C_{3,0}$ | 1 – 2 mm |
| Star aberration | $c_{42}$ | $\frac{1}{4}S_3$ | $C_{3,2}$ | 1 – 20 $\mu$m |
| Fourfold astigmatism | $c_{44}$ | $A_3$ | $C_{3,4}$ | 1 – 20 $\mu$m |
| $5^{\text{th}}$-order axial coma | $c_{51}$ | $\frac{1}{5}B_4$ | $C_{4,1}$ | $<$ 0.2 mm |
| Three-lobe aberration | $c_{53}$ | $\frac{1}{5}D_4$ | $C_{4,3}$ | $<$ 0.2 mm |
| Fivefold astigmatism | $c_{55}$ | $A_4$ | $C_{4,5}$ | $<$ 0.2 mm |
| $6^{\text{th}}$-order spherical aberr. | $c_{60}$ | $C_5$ | $C_{5,0}$ | 1-10 mm |
| $6^{\text{th}}$-order star aberr. | $c_{62}$ | $\frac{1}{6}S_5$ | $C_{5,2}$ | $<$ 10 mm |
| Rosette aberration | $c_{64}$ | $\frac{1}{6}R_5$ | $C_{5,4}$ | $<$ 10 mm |
| Sixfold astigmatism | $c_{66}$ | $A_5$ | $C_{5,6}$ | $<$ 10 mm |

Table 2.1: List of wave-aberration coefficients up to the $6^{\text{th}}$ order and the corresponding notation frequently used in the literature. The rightmost column specifies an approximate order of magnitude for the aberrations $c_{20}$ – $c_{40}$ observable with a conventional TEM. The approximate magnitudes for the aberrations appearing below the spherical aberration in the list have been observed with microscopes equipped with a hexapole $C_{\text{S}}$-corrector.

coefficients for the more historic notation of ref. [20] differ to the notation used in this work by specific factors as noted in table 2.1. Table 2.1 also gives rough estimates for the order of magnitude of the aberration coefficients typically occurring in high-resolution transmission electron microscopes.

In electron microscopy, so far, only wave aberrations with a dependence up to the fourth order of the spatial frequency $g$ were considered, and the threefold astigmatism was, for a long time, the aberration with highest foldness in rotational symmetry. In this work, the aberration function is expanded up to the sixth order of aberrations because of two reasons: Firstly, for Sub-Ångström resolution, the highest transferred spatial frequency $g_{\text{max}}$ is larger than 10 nm$^{-1}$. For this spatial frequencies and with the typical magnitudes of the aberration coefficients as listed in the right column of table 2.1, the phase shift $\chi_{6n}(g_{\text{max}})$ of the electron wavefunction due to the $6^{\text{th}}$-order aberrations comprises values

of the order of 1 which can noticeably take influence on the image contrast. The second reason, why aberrations of orders higher than the $4^{\text{th}}$ order become relevant is related to the widespread application of multipole lenses for the correction of the large spherical aberration $C_{\text{S}}$. Due to the higher foldness of rotational symmetry of the applied lens systems, also aberrations of a higher foldness of rotational symmetry have to be considered.

## 2.3 Properties of individual wave aberrations

The first aberration in the expansion of the aberration function in equation (2.2.8) is the constant phase offset $c_{00}$. It affects every beam by the same amount and does therefore not influence the beam interference. This constant phase offset corresponds to the free quantum-mechanical phase factor which is not observable in experiments.
The first-order aberration $c_{11}$ describes a directional displacement of all image details by the same amount which is a rigid image displacement. This effect is deducible from the expression of the phase shift $\chi_{11}(g,\varphi) = 2\pi(c_{11x}gx + c_{11y}g_y)$ caused by this aberration, which has to be applied to every interference term in the Fourier transform of the image. According to equation (1.3.7), each Fourier component $I_{\text{L}}(\vec{g}\,)$ of the linear image is thus multiplied with a phase factor $\mathrm{e}^{-2\pi\mathrm{i}(c_{11x}g_x+c_{11y}g_y)}$ causing an image shift by $c_{11x}$ along the $x-$axis of the image and by $c_{11y}$ along the $y-$axis of the image. The relative displacement between two images is important for the aberration-measurement method described in section 4.2.
For the discussion of the electron-optical contrast transfer with a general aberration function as in equation (2.2.8), the list of aberrations is separated into even and odd aberrations. The phase shift caused by even aberrations is invariant with respect to a sign change of the diffraction vector $\vec{g}$, which is the case for all aberrations of the even orders ($m = 2l,\ l \in \mathbb{N}_0$). Using the vectorial notation, the even part of the aberration function is given by

$$\chi^{\text{e}}(\vec{g}\,) = \frac{1}{2}[\chi(\vec{g}\,) + \chi(-\vec{g}\,)] = \chi^{\text{e}}(-\vec{g}\,) \tag{2.3.1}$$

All wave aberrations of the odd orders ($m = 2l+1,\ l \in \mathbb{N}_0$) cause an anti-symmetric phase shift as comprised by the odd part of the aberration function, with

$$\chi^{\text{o}}(\vec{g}\,) = \frac{1}{2}[\chi(\vec{g}\,) - \chi(-\vec{g}\,)] = -\chi^{\text{o}}(-\vec{g}\,). \tag{2.3.2}$$

The visualisation of wave aberrations in form of phase plates often helps to understand the basic features of a specific wave aberration. A phase plate displays the phase shift $\chi_{mn}$ of the electron wavefunction caused by the aberration $c_{mn}$ as a function of $\vec{g} = (g,\varphi)$

according to equation (2.2.6). Figure 2.2 shows phase plates for the wave aberrations $c_{22}$ to $c_{66}$, where the values of $\chi_{mn}(g,\varphi)$ are depicted in a grey scale. Grey-to-white represents a positive phase shift ($\chi(g,\varphi) > 0$), while grey-to-black represents a negative phase shift ($\chi(g,\varphi) < 0$). Contrast steps appear at multiples of $\pi/2$.

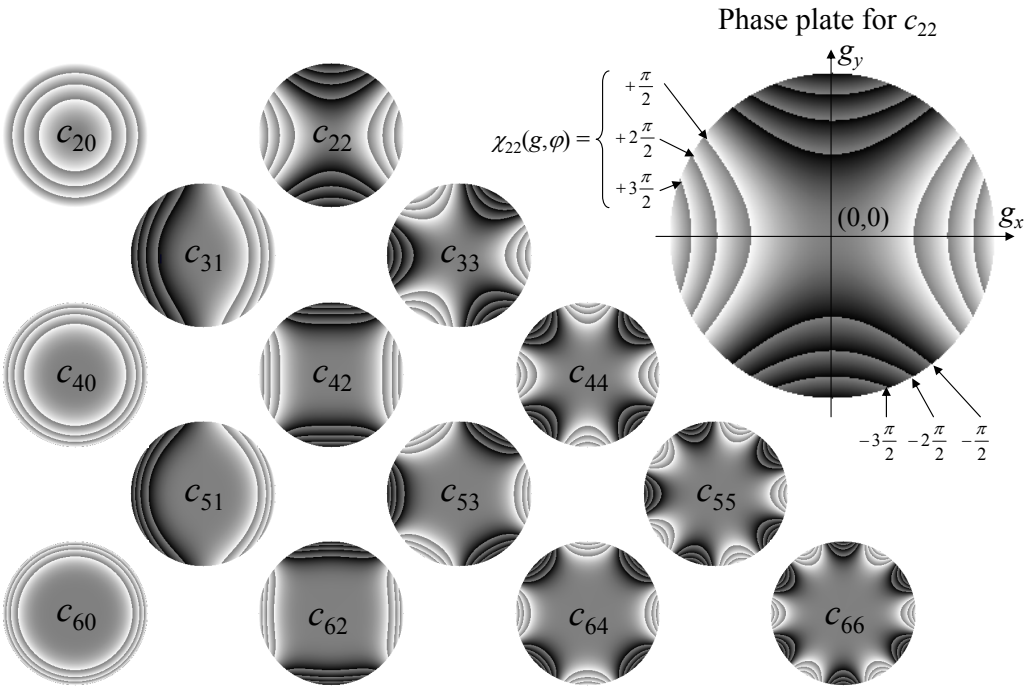



Figure 2.2: Phase plates of single wave aberrations $c_{22}$ up to $c_{66}$ arranged in rows for equal order $m$ and columns for equal rotational symmetry (foldness $n$). The moduli of the aberration coefficients are chosen such that the maximum phase shift at the outer perimeter of every phase plate is equal to $2\pi$. All azimuth parameters $\varphi_{mn}$ are equal to zero such that the non-circular aberrations have an axis of strongest effect along the $x-$axis of the phase plate. Each phase plate displays the phase shift $\chi_{mn}(g,\varphi)$ in a grey scale. Grey-to-white represents a positive phase shift, while grey-to-black represents a negative phase shift. Contrast steps appear at multiples of $\pi/2$ as denoted for the larger image on the right side which, as an example, shows the phase plate for the twofold astigmatism $c_{22}$.

For all phase plates drawn in figure 2.2, the maximum phase shift at the outer perimeter is set to the common value $2\pi$. With increasing order $m$, corresponding to the rows in figure 2.2, the first $\pi/2$-contour appears at a larger spatial frequency. Accordingly, the distance between successive $\pi/2$-contours is smaller for the higher-order aberrations. This steeper ascent of the phase shift reflects the phenomenon that higher-order aberrations affect the high-frequency image content stronger.

In order to give an impression of how wave aberrations affect the real-space image, images for an isolated silicon-atom column have been calculated for a resolution of 0.8 Ångström in the presence of different aberrations. The formalism (1.3.34) for the partially coherent,

non-linear contrast transfer was applied to calculate the images as shown in figure 2.3 for the single-column wavefunction.

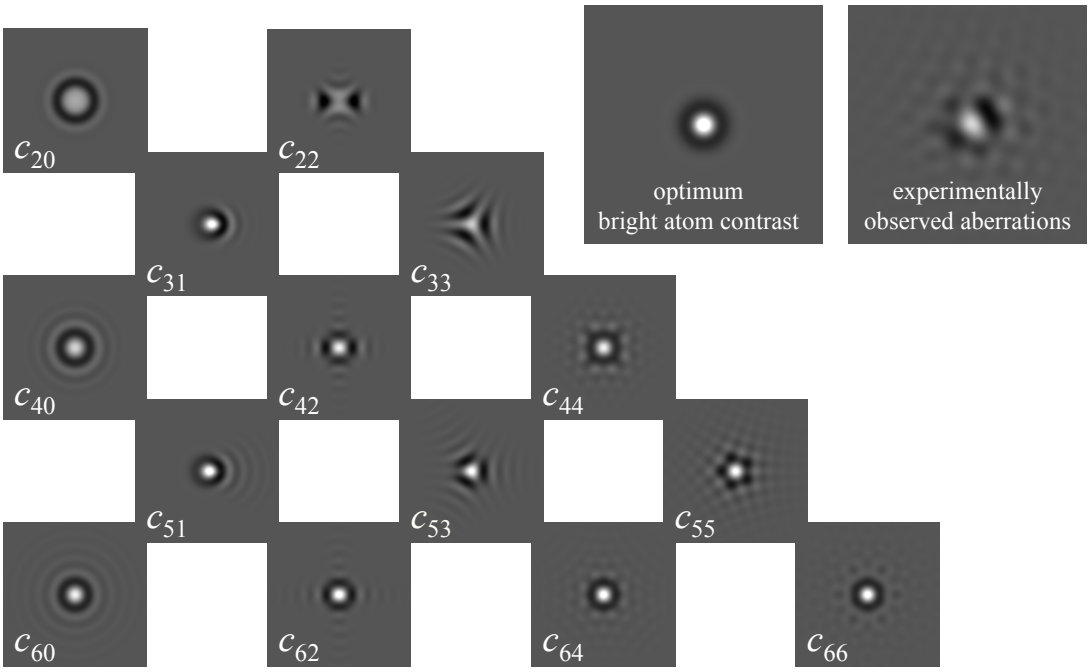



Figure 2.3: Every small image shows the calculated image intensity for an isolated Si-atom column with different aberrations added to the optimum bright-atom contrast setting. The calculations reflect the imaging by a spherical-aberration corrected transmission electron microscope with 0.8 Ångström resolution, operated at 300 kV accelerating voltage. Every patch comprises an area of 1 nm$^2$. The moduli of the additional aberrations are chosen to cause a maximum phase shift of $2\pi$ at 0.8 Ångström, while the corresponding azimuth angles are zero. The larger image on the top-right is calculated from an actually measured aberration function of a TEM.

The modulus of each aberration in the small images is chosen such that the 0.8 Ångström details experience a maximum phase deviation of $2\pi$ from the optimum setting. The combined effect of different aberrations is shown in the larger image on the top-right of figure 2.3. The image was calculated using an aberration function which comprises actually measured wave aberrations of a transmission electron microscope. The respective error figure spreads almost over the whole area of 1 nm$^2$. In projections of crystals many atom columns will be contained in an area of this size. Consequently, the contrast of several atom columns will superpose and hide much detail about the true object structure when imaging crystals. In order to avoid such complication of the experimental results, it is mandatory to reduce all unwanted aberrations and to adjust the wanted aberrations precisely. The unwanted aberrations are frequently addressed as *parasitic* or *residual* aberrations.

## 2.4 Principal strategies to correct aberrations

Any aberration, which is unintentionally introduced by the electron-optical transfer system, obviously hinders a straightforward, quantitative analysis of the image intensity in terms of the projected object structure. Since the quantitative measurement of the projected object structure is the unquestionable goal of materials-science investigations by high-resolution transmission electron microscopy, a precise correction of the undesired parasitic aberrations is required. Mainly three strategies are used in HRTEM in order to cope with the influence of unwanted aberrations on the experimental results: image simulations, a-priori aberration correction by hardware, and a-posteriori elimination of aberrations from a numerically reconstructed wavefunction.

### 2.4.1 Treatment of aberrations in image simulations

For a long time, the technique of image simulation was the only possibility to interpret and analyse experimental HRTEM images quantitatively also in the presence of aberrations. In this traditional approach, calculated images based on an assumed object-structure model are compared to an experimental image [3–6]. In order to calculate the images from an assumed structure model, electron diffraction and electron-optical transfer through the microscope are computed in the same sequence as it proceeds in the experiment.
The theoretical framework used to calculate the exit-plane wavefunction from the guessed object structure has been introduced in chapter 1. Today, comprehensive software packages exist [7], which perform the calculation of the EPW by means of the Multislice Algorithm. Such commercially distributed software packages also provide routines to calculate the full non-linear contrast transfer of the EPW to the image as described by equation (1.3.34). Thereby, all aberrations mentioned in section 2.2 can be included, in principle, by the coherent transfer function

$$\tau(\vec{g}\,) = \exp[-\mathrm{i}\,\chi(\vec{g}\,)], \tag{2.4.1}$$

using the aberration function $\chi(\vec{g}\,)$ of equation (2.2.8).
The theoretically predicted image is "fitted" to the experimental image by progressive variations of parameters describing the object structure and the imaging properties of the electron microscope. This technique requires a precise knowledge about the object structure in advance. An automatic trial-and-error variation of projected atom positions causes an immense amount of numerical operations and is thus only possible to a very limited extent [70].

In addition to a sometimes very tedious matching of the assumed object structure, the fitting of a multitude of aberrations with the technique of image simulation can not be accomplished in practice. The fact that some aberrations in the image can not be distinguished from local features of the object structure poses a particular problem of such an approach. For example, an image affected by axial coma looks similar to an aberration-free image of a slightly tilted object. The interdependence of imaging parameters and structure model can be avoided by a prior independent measurement of aberrations. The quality of the fit between calculated image and experimental image and thus the soundness of quantitative structure analysis by image simulations depends sensitively on the precision of the aberration measurement. Moreover, from a methodological point of view, it is not really satisfying to introduce aberrations into the simulations while methods exist which allow one to eliminate the aberrations from the experimental result.

### 2.4.2 Aberration correction by hardware

The application of spherical-aberration correctors yields an enormous improvement of the image quality with the nowadays available state-of-the-art high-resolution FEG-microscopes. The possibility to tune the spherical aberration to small values reduces not only the obvious delocalisation of object information in the image, but also allows to exploit the information limit fully by adjusting optimum contrast-transfer conditions. Especially the use of a small negative spherical aberration turned out to be highly advantageous [48]. The technical concept of spherical-aberration correctors, as introduced in section 1.1.2, is based on the application of hexapole lenses. A second-order compensation-effect between two hexapole lenses is exploited to achieve a compensation of the spherical aberration of the objective lens. It is therefore unavoidable that undesired second- or further higher-order effects are introduced additionally. The hexapole spherical aberration corrector, as distributed by the CEOS company, provides additional freedom for aberration corrections. By adjusting its multipole- and transfer lenses, all aberrations of the second, third and fourth order can be corrected more or less independently, if the microscope is close to an almost aberration-free state [20]. Just recently, also a correction of the spherical aberration of the sixth order, or alternatively a correction of the off-axial coma was demonstrated with a hexapole corrector [71].

In a rather uncorrected state, several successive correction-steps are required to reduce the aberrations to a tolerable small magnitude. Between each of the corrections a measurement of the residual aberrations is necessary to determine the required changes for the multiple lens currents. The precision of the applied aberration-measurement tech-

nique therefore determines the extent of an a-priori elimination of aberrations by means of hardware.

### 2.4.3 Aberration correction by software

Aberration correction after the acquisition of an image is not possible directly, since the phase of the electron wavefunction is lost. The technique of focal-series reconstruction (FSR) allows one to retrieve numerically the electron wavefunction at the exit plane of the object. The complex-valued exit-plane wavefunction, consisting of amplitude and phase, is computed from a series of 10 to 20 images recorded under different defocus settings from the same area of the object.

The first idea of a reconstruction scheme for the EPW in electron microscopy was based on the linear imaging theory and developed by Schiske in 1968 [14]. A solution using the full non-linear approach for the reconstruction was described by Kirkland in 1984 [16]. In the early 1990s, the Brite-Euram project was set up between several academic and industrial partners with the goal to reach 1 Ångström resolution with the HRTEM technique. Within the framework of this project substantial algorithmic improvements of the fully non-linear reconstruction procedure were achieved [72–74]. Owing to advanced instrumental efforts and the ongoing development of software techniques, it became possible to resolve light atoms such as carbon, nitrogen, and oxygen close to or even below 1 Ångström resolution [75]. After further methodological developments, a commercial software product called TrueImage [76] has been established, which provides access to the technique of focal-series reconstruction also for non-experts.

The computational effort required to solve the non-linear reconstruction problem is quite high due to time-expensive numerical operations applied in the Fourier-space formulation of the image intensity as in equation (1.3.34), especially when the partial coherence of the electron beam is considered. Due to the available computer power and the algorithmic improvements made with the TrueImage software, it is nowadays possible to reconstruct the exit-plane wavefunction from a digital image frame of 1024×1024 pixels in approximately five minutes on a conventional PC. Two selected experiments applying the technique of focal-series reconstruction are described in chapter 8.

Since the focal-series reconstruction is based on the evaluation of images under consideration of defocus $Z$ and spherical aberration $C_S$, the numerically reconstructed wavefunction $\Psi_r(\vec{g})$ is still affected by other residual aberrations. The aberration-free exit-plane wavefunction $\Psi_e(\vec{g})$ is obtained by multiplication of $\Psi_r(\vec{g})$ with the phase factor $\exp[i\chi_r(\vec{g})]$,

where the aberration function $\chi_{\mathrm{r}}(\vec{g})$ comprises all residual aberrations [19]:

$$\Psi_{\mathrm{e}}(\vec{g}) = \Psi_{\mathrm{r}}(\vec{g}) \exp[\mathrm{i}\chi_{\mathrm{r}}(\vec{g})]. \tag{2.4.2}$$

In contrast to the a-priori hardware-correction strategy, this numerical a-posteriori elimination of residual aberrations is technically not limited to specific orders of aberrations. All aberrations can be corrected by the software strategy, if they are known prior with sufficient precision.

## 2.5 Tolerance limits for residual aberrations

In optics, the quality of an imaging system is specified by the strength of wave aberrations at the outer diameter of the exit pupil. Analogous, the quality of results obtained with HRTEM is typically evaluated by tolerance limits for the individual values of the aberration coefficients which were present during the image acquisition. In the view of the different strategies for aberration correction, the definition of a reliable tolerance limit for the precision of aberration measurements is of immense practical relevance for a quality assessment of any applied aberration-control strategy.

### 2.5.1 Historical limits

In the presence of aberrations, the image of an isolated point of the object is distorted. In contrast to the aberration-free state, the intensity is distributed over a larger area and is thus reduced at the ideal point in the Gaussian image plane. It was shown first by Rayleigh [77] that a wavefront aberration of less than a quarter of the wavelength at the outer perimeter of the exit pupil diminishes the intensity maximum in the Gaussian focus by less than 20 %. This is an amount of intensity loss, which can usually be tolerated. As the term *usually* already implicates, *Rayleigh's quarter wavelength criterion* is based purely on subjective considerations. Many other, mostly more restrictive criteria were discussed with similar subjective reasoning in the optics community, like for example Maréchals $\lambda/14$-criterion [78].

### 2.5.2 Currently used single aberration limits

Deviating from Rayleigh's $\lambda/4$-criterion, which corresponds to a $\pi/2$ tolerance limit in terms of phase shifts, the most frequently applied criterion in HRTEM tolerates a phase shift $\chi_{mn} \leq \pi/4$ due to an individual, residual aberration $c_{mn}$ at the information limit $g_{\max}$ [19,20]. From many aberration measurements with currently existing techniques the users have learned that this $\pi/4$-limit for individual aberrations, is a little bit pessimistic when imaging crystals. This observation can be easily understood, since only a very small portion of the image contrast is transferred linearly through the passband close to $g_{\max}$. Table 2.2 lists the $\pi/4$-limit $L(m, g_{\max}) = m/(8\lambda^{m-1} g_{\max}^m)$ to the modulus of aberrations in the orders $m = 2$ up to $m = 6$ for three different information limits in the case of 300 kV accelerating voltage.

| | $L(m, g_{\max})$ | | |
|---|---|---|---|
| $m$ | $g_{\max} = 8.0$ nm$^{-1}$ | $g_{\max} = 12.5$ nm$^{-1}$ | $g_{\max} = 20.0$ nm$^{-1}$ |
| 2 | 2 nm | 0.81 nm | 0.32 nm |
| 3 | 190 nm | 49 nm | 12 nm |
| 4 | 16 $\mu$m | 2.7 $\mu$m | 0.4 $\mu$m |
| 5 | 1.3 mm | 136 $\mu$m | 13 $\mu$m |
| 6 | 97 mm | 6.6 mm | 0.4 mm |

Table 2.2: Individual tolerance limits for different orders of aberrations. The $\pi/4$-limits $L(m, g_{\max})$ are calculated for three selected information limits $g_{\max}$, with $\lambda = 1.969$ pm.

### 2.5.3 Proposal of a new multi-aberration tolerance-limit

It is of extreme importance to emphasise here that not the single contributions but the sum of all present aberrations in the aberration function is the relevant quantity determining the contrast transfer to the image. Phase shifts caused by different aberrations can add up or compensate at specific spatial frequencies. In most cases a separate, independent consideration of single aberrations results in a heavily underestimated, combined aberration effect in the image. No matter what technique is applied, be it image simulation, aberration correction by hardware, or focal-series reconstruction with aberration correction by software, the acquisition of experimental results which are free from influences of the electron microscope is only guaranteed under one condition: The total aberration function used for the correction or interpretation of the experimental result must not de-

viate by more than $\pi/4$ from the aberration function which was actually present during the experiment over the full diffraction plane up to the information limit.
This deviation is estimated by the error of the aberration measurement which is prerequisite for aberration correction. By expressing the measurement error as standard deviation[1] $\sigma_\chi(\vec{g})$ of the measured aberration function, a new multi-aberration tolerance-limit is proposed here by the following fundamental formula:

$$\sigma_\chi(\vec{g}) \leq \frac{\pi}{4}, \quad |\vec{g}| \leq g_{\max}, \tag{2.5.1}$$

The well-established formalism of error-propagation [79] provides a useful procedure to calculate the standard deviation of the aberration function based on covariances $v_{mni,\mu\nu j}$ of single aberration coefficients $c_{kli}$ and $c_{mnj}$. The aberration covariances have to be provided by the applied aberration-measurement technique. The variance $\sigma_\chi^2$ of the total aberration function is then given by

$$\sigma_\chi^2 = \sum_{kli} \sum_{mnj} \left( \frac{\partial \chi}{\partial c_{kli}} \right) \left( \frac{\partial \chi}{\partial c_{mnj}} \right) v_{kli,mnj}. \tag{2.5.2}$$

Please note that the first two indices of the triples $(kli)$ and $(mnj)$ refer to the identification of aberrations by order and rotational symmetry, whereas the third index distinguishes the two spatial components $x$ and $y$ of the complex-valued aberration coefficient as introduced by equation (2.2.2).
Since the aberration function $\chi(g)$ is linear in the coefficients $c_{mni}$ and a polynomial in the spatial frequency components $g_x$ and $g_y$, the variance $\sigma_\chi^2$ of the aberration function is also a polynomial in $g_x$ and $g_y$, which increases with higher spatial frequencies. According to the definition (2.2.8) of the aberration function, every aberration $c_{mn}$ contributes with a $g^m \cos[n(\varphi - \varphi_{mn})]$-term to the error propagation. Over the full diffraction plane up to the information limit, the error of the aberration function is therefore typically maximum at the information limit. It is consequently sufficient to apply the condition (2.5.1) at the information limit $g_{\max}$ only.
Obviously, the full error-propagation to the aberration function for all azimuthal directions yields no general figure-of-merit in form of a single value indicating that condition (2.5.1) is fulfilled. The standard deviation $\sigma_\chi(\vec{g})$ of the aberration function might already exceed the $\pi/4$-limit in one azimuthal direction, while it remains smaller in other azimuthal directions. Since a figure-of-merit in form of a single value bears an enormous practical advantage for the numerical evaluation of aberration-measurement results, the

[1] The definition of the statistical termini like standard deviation, variance and covariance is given in appendix A.4.

full error propagation is simplified by considering the mean variance $\bar{\sigma}_\chi^2$ of the aberration function at the information limit $g_{\max}$ :

$$\bar{\sigma}_\chi^2(g_{\max}) = \frac{1}{2\pi}\int_0^{2\pi} \sigma_\chi^2(g_{\max},\varphi)\,\mathrm{d}\varphi. \tag{2.5.3}$$

Due to the averaging over all azimuthal directions, the covariances between aberrations with different rotational symmetry cancel out in the analytical integration (2.5.3), while the error propagation of all variances $v_{mni,mni}$ and the covariances $v_{kni,mni}$ remain. The latter are typically negative and account for the possible compensations between pairs of aberrations showing equal rotational symmetry.

In order to have a reliable and practical condition for the tolerance of aberrations in high-resolution transmission electron microscopy, the following criterion for aberration-free imaging is proposed:

$$\bar{\sigma}_\chi(g_{\max}) \leq \frac{\pi}{4}. \tag{2.5.4}$$

By this formula, the so far unaccounted entanglement of a multitude of aberrations is resolved by a single criterion for the aberration tolerance analogous to the prior single aberration limits. An aberration-measurement result keeping the criterion (2.5.4) warrants the satisfaction of the condition (2.5.1) for the aberration-free state of experimental results in HRTEM. The aberration-free state is guaranteed up to the lowest spatial frequency $g_{\mathrm{int}}$, where the error $\bar{\sigma}_\chi$ of the measured aberration function hits the $\pi/4$-limit. One may associate $g_{\mathrm{int}}$ with the interpretable resolution. Residual aberrations are thus significant when $g_{\mathrm{int}} < g_{\max}$. In contrast, when one achieves $g_{\mathrm{int}} \geq g_{\max}$, the experiment can be considered as aberration-free.

# Chapter 3

# Special aspects of the contrast transfer

According to the discussions in the previous chapter, a multitude of aberrations has to be considered when aiming at a straightforward interpretation and quantitative analysis of experimental results in HRTEM with resolutions around one Ångström. It is therefore necessary to review the electron-optical contrast transfer as introduced in section 1.3 in the presence of an asymmetric aberration function. The development of precise aberration-measurement procedures within this work requires also a detailed elaboration of the partially coherent, linear contrast transfer under tilted-beam illumination.

## 3.1 Linear contrast transfer with an asymmetric aberration function

With the introduction of a multitude of aberrations in section 2.2, the aberration function is no longer an even and rotationally symmetric function of the diffraction vector $\vec{g}$ as treated in section 1.3. The prior centrosymmetry of the aberration function lead to the formulation of the partially coherent linear image in equation (1.3.24) with the term $\sin[\chi(\vec{g})]$, where $\chi(\vec{g})$ only accounted for two even aberrations, defocus and spherical aberration. It is obvious, that an asymmetric aberration function $\chi(\vec{g}) = \chi^{\mathrm{e}}(\vec{g}) + \chi^{\mathrm{o}}(\vec{g})$, which comprises simultaneously even aberrations $\chi^{\mathrm{e}}(\vec{g}) = \chi^{\mathrm{e}}(-\vec{g})$ and odd aberrations $\chi^{\mathrm{o}}(\vec{g}) = -\chi^{\mathrm{o}}(-\vec{g})$, breaks the symmetry between the two linear interference terms in equation (1.3.4). In the following, the influence of this symmetry break on the linear imaging process is investigated.

### 3.1.1 Reviewing the linear contrast transfer

Instead of beginning with the formulation of a coherent linear image, the derivation of the partially coherent contrast transfer in the presence of even and odd aberrations starts with the expression for a partially coherent non-linear image in equation (1.3.35). After discretisation of the integral, the linear image Fourier coefficients $I_{\mathrm{L}}(\vec{g}\neq 0)$ compose of two terms with diffraction vectors $\vec{g}$ and $-\vec{g}$ weighted by the respective transmission cross-coefficients $T(\vec{g}\,,0)$ and $T(0,-\vec{g}\,)$ :

$$I_{\mathrm{L}}(\vec{g}\neq 0) = \Psi_{\mathrm{e}}(\vec{g}\,)\Psi_{\mathrm{e}}^{*}(0)T(\vec{g}\,,0) + \Psi_{\mathrm{e}}(0)\Psi_{\mathrm{e}}^{*}(-\vec{g}\,)T(0,-\vec{g}\,). \tag{3.1.1}$$

In the frame of the weak-phase-object approximation, the undiffracted beam is normalised to 1, with $\Psi_{\mathrm{e}}(0) = \Psi_{\mathrm{e}}^{*}(0) = 1$, and the diffracted beams are given by $\Psi_{\mathrm{e}}(\vec{g}\,) = \mathrm{i}\sigma V(\vec{g}\,)t$. Analogous to the discussions in section 1.3.2, the projected potential $V_{\mathrm{P}}(\vec{g}\,)$ obeys the Friedel symmetry $V_{\mathrm{P}}(\vec{g}\,) = V_{\mathrm{P}}^{*}(-\vec{g}\,)$, so that equation (3.1.1) can be simplified to

$$I_{\mathrm{L}}(\vec{g}\neq 0) = \Psi_{\mathrm{e}}(\vec{g}\,)[T(\vec{g}\,,0) - T(0,-\vec{g}\,)]. \tag{3.1.2}$$

The transmission cross-coefficient (TCC) $T(\vec{g'},\vec{g''})$ is defined in a general form by equation (1.3.38) with two damping envelopes $E_{\mathrm{t}}(\vec{g'},\vec{g''})$ and $E_{\mathrm{s}}(\vec{g'},\vec{g''})$ in the equations (1.3.39) and (1.3.40). The respective TCC for each of the two linear terms in equation (3.1.2) is then given by

$$T(\vec{g}\,,0) = E_{\mathrm{t}}(\vec{g}\,,0)\,E_{\mathrm{s}}(\vec{g}\,,0)\,\exp[-\mathrm{i}\chi^{\mathrm{e}}(\vec{g}\,)]\,\exp[-\mathrm{i}\chi^{\mathrm{o}}(\vec{g}\,)] \tag{3.1.3}$$

$$T(0,-\vec{g}\,) = E_{\mathrm{t}}(0,-\vec{g}\,)\,E_{\mathrm{s}}(0,-\vec{g}\,)\,\exp[\mathrm{i}\chi^{\mathrm{e}}(\vec{g}\,)]\,\exp[-\mathrm{i}\chi^{\mathrm{o}}(\vec{g}\,)] \tag{3.1.4}$$

In the above equations, the asymmetric aberration function $\chi(\vec{g}\,) = \chi^{\mathrm{e}}(\vec{g}\,) + \chi^{\mathrm{o}}(\vec{g}\,)$ is already splitted into its even and odd part. The temporal envelope $E_{\mathrm{t}}(\vec{g'},\vec{g''})$ in equation (1.3.39) is invariant with respect to an exchange of the arguments and simultaneous reversal of the sign of the diffraction vectors, i.e. $E_{\mathrm{t}}(\vec{g}\,,0) = E_{\mathrm{t}}(0,-\vec{g}\,)$.
The spatial envelope $E_{\mathrm{s}}(\vec{g'},\vec{g''})$ in equation (1.3.40) shows a more complicated behaviour under exchange of the argument vectors and a sign reversal, since it involves the gradient of the aberration function:

$$E_{\mathrm{s}}(\vec{g'},\vec{g''}) = \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\vec{\nabla}_{\vec{g'}}\chi(\vec{g'}) - \vec{\nabla}_{\vec{g''}}\chi(\vec{g''})\right]^2\right\}. \tag{3.1.5}$$

The gradient of the aberration function shows the following behaviour under sign reversal of the diffraction vector $\vec{g}$:

$$\vec{\nabla}_{\vec{g}}\,[\chi(-\vec{g}\,)] = -\vec{\nabla}_{\vec{g}}\,[\chi^{\mathrm{e}}(\vec{g}\,) - \chi^{\mathrm{o}}(\vec{g}\,)]. \tag{3.1.6}$$

The two spatial envelope terms $E_{\mathrm{s}}(\vec{g}\,,0)$ in equation (3.1.3) and $E_{\mathrm{s}}(0,-\vec{g}\,)$ in equation (3.1.4) can thus be expressed by

$$E_{\mathrm{s}}(\vec{g}\,,0) = \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\vec{\nabla}_{\vec{g}}\,[\chi^{\mathrm{e}}(\vec{g}\,)+\chi^{\mathrm{o}}(\vec{g}\,)]\right]^2\right\} \tag{3.1.7}$$

$$= \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\left|\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{e}}(\vec{g}\,)\right|^2+\left|\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{o}}(\vec{g}\,)\right|^2\right]\right. \tag{3.1.8}$$

$$\left. -\,2\left(\frac{q_0}{2}\right)^2[\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{e}}(\vec{g}\,)][\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{o}}(\vec{g}\,)]\right\}$$

$$E_{\mathrm{s}}(0,-\vec{g}\,) = \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\vec{\nabla}_{\vec{g}}\,[\chi^{\mathrm{e}}(\vec{g}\,)-\chi^{\mathrm{o}}(\vec{g}\,)]\right]^2\right\} \tag{3.1.9}$$

$$= \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\left|\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{e}}(\vec{g}\,)\right|^2+\left|\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{o}}(\vec{g}\,)\right|^2\right]\right. \tag{3.1.10}$$

$$\left. +\,2\left(\frac{q_0}{2}\right)^2[\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{e}}(\vec{g}\,)][\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{o}}(\vec{g}\,)]\right\}$$

Comparing equation (3.1.8) with equation (3.1.8) reveals that the two linear interference terms in equation (3.1.1) are damped by a different exponential factor. This factor depends on a mixed product of the gradients of the even part $\chi^{\mathrm{e}}(\vec{g}\,)$ and of the odd part $\chi^{\mathrm{o}}(\vec{g}\,)$ of the aberration function.

The damping of the linear image contrast due to the partial temporal and partial spatial coherence can be described firstly, by a symmetric exponential damping factor $\exp[S_0(\vec{g}\,)]$ acting on both linear interference terms equally, and secondly, by an antisymmetric exponential damping factor $\exp[A_0(\vec{g}\,)]$, which acts differently on the two interference terms. The two abbreviations $S_0(\vec{g}\,)$ and $A_0(\vec{g}\,)$ are given by

$$S_0(\vec{g}\,) = -\left(\frac{\pi\Delta\lambda}{2}\right)^2\vec{g}^{\,4}-\left(\frac{q_0}{2}\right)^2\left[\left|\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{e}}(\vec{g}\,)\right|^2+\left|\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{o}}(\vec{g}\,)\right|^2\right]. \tag{3.1.11}$$

$$A_0(\vec{g}\,) = -2\left(\frac{q_0}{2}\right)^2[\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{e}}(\vec{g}\,)][\vec{\nabla}_{\vec{g}}\,\chi^{\mathrm{o}}(\vec{g}\,)], \tag{3.1.12}$$

with $S_0(\vec{g}\,)=S_0(-\vec{g}\,)$ and $A_0(\vec{g}\,)=-A_0(-\vec{g}\,)$. One obtains thus the following expression for the partially coherent, linear image with an asymmetric aberration function:

$$I_{\mathrm{L}}(\vec{g}\neq 0)=\Psi_{\mathrm{e}}(\vec{g}\,)\,\mathrm{e}^{S_0(\vec{g}\,)}\,\mathrm{e}^{-\mathrm{i}\chi^{\mathrm{o}}(\vec{g}\,)}\left[\mathrm{e}^{+A_0(\vec{g}\,)}\,\mathrm{e}^{-\mathrm{i}\chi^{\mathrm{e}}(\vec{g}\,)}-\mathrm{e}^{-A_0(\vec{g}\,)}\,\mathrm{e}^{+\mathrm{i}\chi^{\mathrm{e}}(\vec{g}\,)}\right] \tag{3.1.13}$$

The proportionality between the linear image contrast $I_{\mathrm{L}}(\vec{g}\neq 0)$ and the exit-plane wavefunction $\Psi_{\mathrm{e}}(\vec{g}\,)$ of a weak phase object is traditionally expressed by a contrast transfer function as used in section 1.3.4. It goes beyond the scope of the usual concept of the contrast transfer function to discern the transfer of the two diffracted beams $\Psi_{\mathrm{e}}(\vec{g}\,)$ and $\Psi_{\mathrm{e}}(-\vec{g}\,)$ contributing to a specific $I_{\mathrm{L}}(\vec{g}\,)$. Therefore, a *wave-transfer function* (WTF) is

introduced here which expresses the relation between the image Fourier coefficients and the exit-plane wavefunction by

$$I_{\mathrm{L}}(\vec{g}\neq 0) = \Psi_{\mathrm{e}}(\vec{g}) \cdot \mathrm{WTF}(\vec{g}), \tag{3.1.14}$$

$$\mathrm{WTF}(\vec{g}) = \mathrm{e}^{S_0(\vec{g})}\,\mathrm{e}^{-\mathrm{i}\chi^{\mathrm{o}}(\vec{g})}\left[\mathrm{e}^{+A_0(\vec{g})}\,\mathrm{e}^{-\mathrm{i}\chi^{\mathrm{e}}(\vec{g})} - \mathrm{e}^{-A_0(\vec{g})}\,\mathrm{e}^{+\mathrm{i}\chi^{\mathrm{e}}(\vec{g})}\right]. \tag{3.1.15}$$

The wave-transfer function can be seen as analogon to the contrast-transfer function which applies when an asymmetric aberration function has to be considered.

### 3.1.2 Interpretation of the resulting contrast transfer

The wave-transfer function in equation (3.1.15) describes four essential features of the linear contrast transfer:

1. The image contrast is damped due to the partial temporal and partial spatial coherence as described by the factor $\exp[S_0(\vec{g})]$. The argument depends on the one hand on the well-known $g^4$-term involving the defocus spread $\Delta$, and on the other hand on separate modulus squares of the gradients of the even part and of the odd part of the aberration function.

2. A phase factor $\exp[-\mathrm{i}\chi^{\mathrm{o}}(\vec{g})]$ which depends purely on the combined odd wave aberrations of the imaging system has already been mentioned in several specialised publications [66,67,80]. This phase factor expresses a $|\vec{g}|$-dispersive displacement of the $\vec{g}$- and the $-\vec{g}$-interference fringes into the same direction in the image. Details of higher spatial frequencies are shifted farther than details of lower spatial frequencies. An $n$-fold odd aberration comprises $n$ azimuthal directions of strong dispersive displacement and $n$ neutral directions where no displacement occurs. Exemplary images which illustrate the blurring due to odd aberrations are shown in figure 2.3.

3. The factor, $\exp[A_0(\vec{g})]$, accounts for an unequal damping of the two linear interference terms which contribute to the image component $I_{\mathrm{L}}(\vec{g})$ in equation (3.1.1). When both, even and odd aberrations are present simultaneously, i.e. when $A_0(\vec{g}) \neq 0$, the two linear interference terms are damped be a different factor as illustrated in figure 3.1. The interference term $\Psi_{\mathrm{e}}(\vec{g})\Psi_{\mathrm{e}}^{*}(0)$ is damped according to $\mathrm{e}^{S_0(\vec{g})+A_0(\vec{g})}$, whereas the interference term $\Psi_{\mathrm{e}}(0)\Psi_{\mathrm{e}}^{*}(-\vec{g})$ is damped according to $\mathrm{e}^{S_0(-\vec{g})+A_0(-\vec{g})} = \mathrm{e}^{S_0(\vec{g})-A_0(\vec{g})}$.

4. The phase factors $\mathrm{e}^{-\mathrm{i}\chi^{\mathrm{e}}(\vec{g})}$ and $\mathrm{e}^{+\mathrm{i}\chi^{\mathrm{e}}(\vec{g})}$ depend on the even part of the aberration function and describe a $|\vec{g}|$-dispersive displacement of the $\vec{g}$- and the $-\vec{g}$-interference

fringes into opposite directions in the image. Due to the aforementioned unequal damping, the two phase factors do not form the $\sin[\chi^{\mathrm{e}}(\vec{g})]$-term as traditionally obtained for the contrast-transfer function with its full amplitude. The "overlapping" part of a linear interference fringe is thus imaged partially without superposition by the respective other linear interference fringe. A completely separate imaging of the linear interference fringes is addressed in the literature as single-sideband imaging [81].

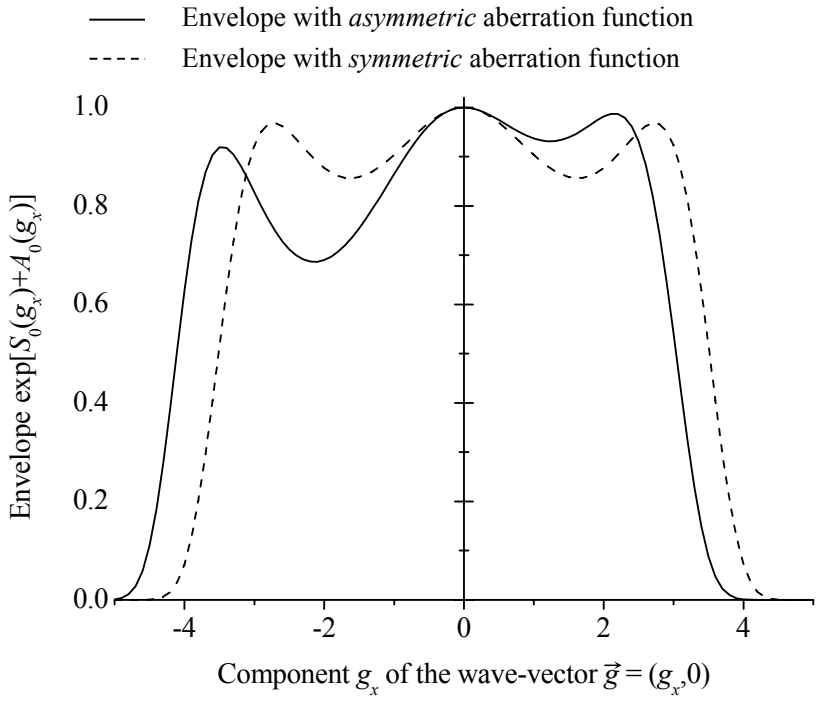



Figure 3.1: Damping $\mathrm{e}^{S_0(\vec{g})+A_0(\vec{g})}$ of the linear interference fringes due to partial coherence. The curves are calculated considering an FEG-microscope operated at 200 kV accelerating voltage ($\Delta = 6.0$ nm, $\theta_0 = \lambda q_0 = 0.2$ mrad). The solid curve illustrates an unequal damping due to an axial coma of $c_{31} = 4$ $\mu$m, a spherical aberration of $C_{\mathrm{S}} = 1.2$ mm, and a defocus of $Z = -60$ nm. The dashed curve is obtained with a symmetric aberration function comprising only the even aberrations $C_{\mathrm{S}}$ and $Z$.

## 3.2 Linear contrast transfer with tilted illumination

The electron-optical contrast transfer for tilted-beam illumination has already been investigated as a part of the hollow-cone illumination theory by Hanszen [82], among others. This investigation is repeated in this section in the frame of the linear contrast-transfer theory for a weak phase object including the new wave-aberration notation introduced in section 2.2. The obtained results will be discussed by specifically considering numerical diffractograms of thin amorphous objects as calculated from images taken under tilted-beam illumination.

In a first approximation, the projected potential $V_{\mathrm{P}}(\vec{g})$ of a weak phase object is not altered for a tilted incidence of the electron beam if the tilt angle is small. However, the exit-plane wavefunction is transferred with a different diffraction vector $\vec{g}+\vec{t}$ through the electron-optical system. The beam tilt $\vec{t}$ is expressed here as a vector in the diffraction plane and has the dimension [nm$^{-1}$] of a spatial frequency.
The derivation of the contrast transfer with tilted illumination starts with equation (3.1.2) which has already been used in section 3.1 to describe a partially coherent, linear image. The beam tilt is formally introduced by a shift of all diffraction vectors in the transmission cross-coefficients:

$$I_{\mathrm{L}}(\vec{g}\neq 0) = \Psi_{\mathrm{e}}(\vec{g}) \cdot \left[T(\vec{t}+\vec{g},\vec{t}) - T(\vec{t},\vec{t}-\vec{g})\right]. \tag{3.2.1}$$

The exit-plane wavefunction $\Psi_{\mathrm{e}}(\vec{g})$ is assumed to be unaffected by the beam tilt. In the explicit expression for the transmission cross-coefficient $T$ in equation (1.3.38), the application of a beam tilt changes the effective aberration function and the effective envelope functions $E_{\mathrm{t}}$ and $E_{\mathrm{s}}$ :

$$T(\vec{t}+\vec{g},\vec{t}) = E_{\mathrm{t}}(\vec{t}+\vec{g},\vec{t})\, E_{\mathrm{s}}(\vec{t}+\vec{g},\vec{t}) \cdot \mathrm{e}^{-\mathrm{i}[\chi(\vec{t}+\vec{g})-\chi(\vec{t})]}, \tag{3.2.2}$$

$$T(\vec{t},\vec{t}-\vec{g}) = E_{\mathrm{t}}(\vec{t},\vec{t}-\vec{g})\, E_{\mathrm{s}}(\vec{t},\vec{t}-\vec{g}) \cdot \mathrm{e}^{-\mathrm{i}[\chi(\vec{t})-\chi(\vec{t}-\vec{g})]}. \tag{3.2.3}$$

In order to clarify how the linear contrast transfer is altered in particular by the application of a beam tilt, all the factors in the two transmission cross-coefficients are analysed in detail.

### 3.2.1 The effective aberration function

Due to a tilted incident beam, the electron wavefunction is transferred differently through the electron-optical system. Therefore, the effective aberrations must be different compared to those present with untilted illumination. The aberration function $\chi(\vec{g}+\vec{t})$ is

re-formulated by an expansion with respect to a new origin of the diffraction plane. The new origin is given by the beam-tilt vector $\vec{t}$. As result of this expansion, one obtains an *effective aberration function* $\chi_t(\vec{g}) = \chi(\vec{g} + \vec{t})$ which is characterised by *effective aberration coefficients* $c'_{mn}$.
Applying the effective aberration function $\chi_t(\vec{g})$ in the respective coherent part of the transmission cross-coefficients in the equations (3.2.2) and (3.2.3) yields

$$\begin{aligned} \exp[-\mathrm{i}\chi(\vec{t}+\vec{g}) + \mathrm{i}\chi(\vec{t})] &= \exp[-\mathrm{i}\chi_t(\vec{g})] \\ &= \exp[-\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g})] \cdot \exp[-\mathrm{i}\chi_t^{\mathrm{o}}(\vec{g})] \qquad (3.2.4) \\ \exp[-\mathrm{i}\chi(\vec{t}) + \mathrm{i}\chi(\vec{t}-\vec{g})] &= \exp[\mathrm{i}\chi_t(-\vec{g})] \\ &= \exp[\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g})] \cdot \exp[-\mathrm{i}\chi_t^{\mathrm{o}}(\vec{g})] \qquad (3.2.5) \end{aligned}$$

In the above formulations, the effective aberration function $\chi_t(\vec{g})$ is separated again into an even part $\chi_t^{\mathrm{e}}(\vec{g}) = \chi_t^{\mathrm{e}}(-\vec{g})$ and an odd part $\chi_t^{\mathrm{o}}(\vec{g}) = -\chi_t^{\mathrm{o}}(-\vec{g})$, as described by the equations (2.3.1) and (2.3.2) in section 2.2. All constant terms like $\chi_t(0) = \chi(\vec{t})$ are omitted such that $\chi_t(\vec{g}{=}0) = 0$, because they do not influence the image contrast.
Throughout the literature, the dependence of effective aberrations $c'_{mn}$ on the beam tilt and on the original axial aberrations $c_{mn}$ is derived explicitly in quite a number of different notations and arbitrary terminations with respect to the highest relevant order of aberrations [19, 20, 65, 68]. As examples, effective defocus and effective twofold astigmatism are given here in dependence up to the fourth-order axial aberrations by using the new nomenclature of the aberration coefficients (see section 2.2) and the beam tilt $t = t_x + \mathrm{i}t_y$ in complex-number notation:

$$c'_{20} = c_{20} + \frac{2}{3}\,c_{31}\cdot t^* + \frac{2}{3}\,c_{31}^* t + 2\,c_{40}\cdot(t^*)t + \frac{3}{4}\,c_{42}\cdot(t^*)^2 + \frac{3}{4}\,c_{42}^*\cdot t^2 + \ldots \qquad (3.2.6)$$

$$c'_{22} = c_{22} + \frac{2}{3}c_{31}\cdot t + 2\,c_{33}\cdot t^* + c_{40}\cdot t^2 + \frac{3}{2}\,c_{42}\cdot(t^*)t + 3\,c_{44}\cdot(t^*)^2 + \ldots \qquad (3.2.7)$$

The effective aberration coefficients $c'_{mn}$ are thus polynomials in the beam tilt $t$ and $t^*$ and depend linearly on the aberration coefficients $c_{kl}$ with $k \geq m$ present for untilted illumination. Such, so-called *induction formulas* are denoted for effective image displacement, defocus and twofold astigmatism in dependence on the axial aberrations up to the $6^{\mathrm{th}}$-order in appendix C.2.
Within this work, a general formula is derived which, for the first time, describes how an arbitrary effective lower-order aberration is *induced* with tilted-beam illumination. In practice such a general induction formula allows to extend the so far fixed number of aberrations considered by numerical aberration-measurement procedures in an automated

manner. Due to its mathematical complexity, this general induction formula is derived in appendix C.1 and only its result is shown here in complex-number notation:

$$c'_{(\alpha+\gamma)(\alpha-\gamma)} = \rho[\alpha,\gamma]\cdot(\alpha+\gamma)\sum_{\beta=\alpha}^{M}\sum_{\delta=\gamma}^{M(\beta)}\left[\binom{\beta}{\alpha}\binom{\delta}{\gamma}(t^*)^{\beta-\alpha}t^{\delta-\gamma}\frac{c_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right. \tag{3.2.8}$$
$$\left. + \binom{\beta}{\gamma}\binom{\delta}{\alpha}(t^*)^{\delta-\alpha}t^{\beta-\gamma}\frac{c^*_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right]$$

The lower double index $(\alpha+\gamma)(\alpha-\gamma)$ corresponds to the known double index $mn$ of the aberration coefficients. For example $\alpha=2$, $\gamma=1$ denotes $c_{(2+1)(2-1)}=c_{31}$. The function $M(\beta)=\min(\beta,M-\beta)$ restricts the summations on the right-hand side of equation (3.2.8) to aberrations up to the order $M$. The function $\rho[\alpha,\gamma]$, with $\rho[\alpha,\alpha]=1/2$ and $\rho[\alpha,\gamma]=1$ for $\alpha\neq\gamma$, is used for normalisation in case of an effective circular aberration $c'_{(2\alpha)(0)}$.

### 3.2.2 The effective damping due to partial temporal coherence

The envelope function $E_\mathrm{t}(\vec{g'},\vec{g''})$ which describes the damping of the contrast transfer due to a partially temporal coherent electron beam is given by equation (1.3.39). The intentional beam tilt $\vec{t}$ is introduced as an additive to the diffraction-vector arguments with $E_\mathrm{t}(\vec{g'}+\vec{t},\vec{g''}+\vec{t})$. Accordingly, the respective damping $E_\mathrm{t}(\vec{t}+\vec{g},\vec{t})$ of the first linear term in equation (3.2.2) is given by

$$E_\mathrm{t}(\vec{t}+\vec{g},\vec{t}) = \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2\left[(\vec{t}+\vec{g})^2-\vec{t}^{\,2}\right]^2\right\} \tag{3.2.9}$$
$$= \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2\left[2\,\vec{g}\cdot\vec{t}+\vec{g}^{\,2}\right]^2\right\} \tag{3.2.10}$$
$$= \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2\left[4(\vec{g}\cdot\vec{t})^2+\vec{g}^{\,4}+4(\vec{g}\cdot\vec{t})\vec{g}^{\,2}\right]\right\}. \tag{3.2.11}$$

The damping $E_\mathrm{t}(\vec{t},\vec{t}-\vec{g})$ of the second linear term in equation (3.2.3) can be calculated analogously by

$$E_\mathrm{t}(\vec{t},\vec{t}-\vec{g}) = \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2\left[\vec{t}^{\,2}-(\vec{t}-\vec{g})^2\right]^2\right\} \tag{3.2.12}$$
$$= \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2\left[2\,\vec{g}\cdot\vec{t}-\vec{g}^{\,2}\right]^2\right\} \tag{3.2.13}$$
$$= \exp\left\{-\left(\frac{\pi\Delta\lambda}{2}\right)^2\left[4(\vec{g}\cdot\vec{t})^2+\vec{g}^{\,4}-4(\vec{g}\cdot\vec{t})\vec{g}^{\,2}\right]\right\}. \tag{3.2.14}$$

Except for the last term $\pm 4(\vec{g}\cdot\vec{t})\vec{g}^{\,2}$ in the arguments of the exponential terms (3.2.11) and (3.2.14), both linear terms experience the same damping. Due to the dependence of these damping terms on the scalar product $\vec{g}\cdot\vec{t}$, it is already apparent here that the rotational symmetry of the temporal coherence envelope observed with an untilted incident beam is broken in the case of a tilted-beam illumination.

### 3.2.3 The effective damping due to partial spatial coherence

The computation of the effective envelope $E_{\mathrm{s}}(\vec{g'}+\vec{t}\,,\vec{g''}+\vec{t}\,)$ which describes the damping of the contrast transfer due to the partial spatial coherence for a tilted electron beam is more complicated, since it depends on the gradient of the aberration function. Similar to the derivations in section 3.1, the specific symmetry behaviour of the gradient has to be considered

$$\vec{\nabla}_{\vec{g}}\,[\chi_t(\vec{g}\,)] = \vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g}\,) + \vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g}\,), \tag{3.2.15}$$

$$\vec{\nabla}_{\vec{g}}\,[\chi_t(-\vec{g}\,)] = -\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g}\,) + \vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g}\,). \tag{3.2.16}$$

The effective spatial coherence envelope $E_{\mathrm{s}}(\vec{t}+\vec{g}\,,\vec{t}\,)$ for the first interference term in equation (3.2.3) is then explicitly given by

$$E_{\mathrm{s}}(\vec{t}+\vec{g}\,,\vec{t}\,) = \exp\left\{-\left(\frac{q_0}{2}\right)^2\left|\vec{\nabla}_{\vec{g}}\,\chi_t(\vec{g}\,)-\vec{\nabla}_{\vec{g}}\,\chi_t(0)\right|^2\right\} \tag{3.2.17}$$

$$= \exp\left\{-\left(\frac{q_0}{2}\right)^2\left|\vec{\nabla}_{\vec{g}}\,\chi_t(\vec{g}\,)\right|^2\right\} \tag{3.2.18}$$

$$= \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g}\,)\right|^2+\left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g}\,)\right|^2\right]\right\} \tag{3.2.19}$$

$$\cdot\exp\left\{-2\left(\frac{q_0}{2}\right)^2[\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g}\,)]\cdot[\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g}\,)]\right\}$$

Analogous reformulation of the spatial envelope for the second term yields

$$E_{\mathrm{s}}(\vec{t}\,,\vec{t}-\vec{g}\,) = \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g}\,)\right|^2+\left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g}\,)\right|^2\right]\right\} \tag{3.2.20}$$

$$\cdot\exp\left\{+2\left(\frac{q_0}{2}\right)^2[\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g}\,)]\cdot[\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g}\,)]\right\}$$

The two damping terms comprise an identical, symmetric part and a different, anti-symmetric part which is given here by the mixed-gradient term of the effective, even and odd aberrations.

### 3.2.4 Explicit formulation of the tilted-beam contrast transfer

In a final step, the results of the preceding derivations have to be combined in order to formulate a linear image intensity for tilted-beam illumination. Since the explicit terms derived above are very complex, two abbreviations are introduced here. The abbreviation $S(\vec{g},\vec{t})$ combines all arguments of the exponential envelope terms which are centro-symmetric in $\vec{g}$ so that $S(\vec{g},\vec{t}) = S(-\vec{g},\vec{t})$, with

$$S(\vec{g},\vec{t}) = -\left(\frac{\pi\Delta\lambda}{2}\right)^2 [4(\vec{g}\cdot\vec{t})^2 + \vec{g}^4] - \left(\frac{q_0}{2}\right)^2 \left[\left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g})\right|^2 + \left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g})\right|^2\right]. \quad (3.2.21)$$

The other abbreviation $A(\vec{g},\vec{t})$ combines the non-centro-symmetric arguments of the exponential envelopes so that $A(\vec{g},\vec{t}) = -A(-\vec{g},\vec{t})$, with

$$A(\vec{g},\vec{t}) = -4\left(\frac{\pi\Delta\lambda}{2}\right)^2 (\vec{g}\cdot\vec{t})\vec{g}^{\,2} - 2\left(\frac{q_0}{2}\right)^2 [\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g})]\cdot[\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g})]. \quad (3.2.22)$$

By using the two abbreviations, the transmission cross-coefficients in the equations (3.2.2) and (3.2.3) can be expressed in a compact form:

$$T(\vec{t}+\vec{g},\vec{t}) = \exp[S(\vec{g},\vec{t}) + A(\vec{g},\vec{t})]\cdot\exp[-\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g}) - \mathrm{i}\chi_t^{\mathrm{o}}(\vec{g})], \quad (3.2.23)$$

$$T(\vec{t},\vec{t}-\vec{g}) = \exp[S(\vec{g},\vec{t}) - A(\vec{g},\vec{t})]\cdot\exp[\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g}) - \mathrm{i}\chi_t^{\mathrm{o}}(\vec{g})]. \quad (3.2.24)$$

Consequently, the linear image contrast as in equation (3.2.1) is expressed by

$$I_{\mathrm{L}}(\vec{g}\neq 0) = \Psi_{\mathrm{e}}(\vec{g})\,\mathrm{e}^{S(\vec{g},\vec{t})}\,\mathrm{e}^{-\mathrm{i}\chi_t^{\mathrm{o}}(\vec{g})}\left[\mathrm{e}^{+A(\vec{g},\vec{t})}\,\mathrm{e}^{-\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g})} - \mathrm{e}^{-A(\vec{g},\vec{t})}\,\mathrm{e}^{+\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g})}\right] \quad (3.2.25)$$

where $\Psi_{\mathrm{e}}(\vec{g})$ denotes the exit-plane wavefunction for a weak-phase object as used in section 1.3.2. Analogous to the discussion of the contrast transfer in section 3.1 the wave-transfer function $\mathrm{WTF}(\vec{g},\vec{t})$ is identified for the present case with tilted-beam illumination by

$$\mathrm{WTF}(\vec{g},\vec{t}) = \mathrm{e}^{S(\vec{g},\vec{t})}\,\mathrm{e}^{-\mathrm{i}\chi_t^{\mathrm{o}}(\vec{g})}\left[\mathrm{e}^{+A(\vec{g},\vec{t})}\,\mathrm{e}^{-\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g})} - \mathrm{e}^{-A(\vec{g},\vec{t})}\,\mathrm{e}^{+\mathrm{i}\chi_t^{\mathrm{e}}(\vec{g})}\right]. \quad (3.2.26)$$

The wave-transfer function defined here exhibits again the four essential features already interpreted in section 3.1.2. Indeed, the investigation in section 3.1 constitutes a special case with $\vec{t} = 0$, where the respective terms which characterise the contrast transfer become equal, i.e.

$$S(\vec{g},\vec{t}) = S_0(\vec{g}),\quad A(\vec{g},\vec{t}) = A_0(\vec{g}),\quad \chi_t(\vec{g}) = \chi(\vec{g}+\vec{t}) = \chi(\vec{g}),\quad \text{for } \vec{t} = 0. \quad (3.2.27)$$

An additional feature of the contrast transfer appears due to the tilted illumination. The symmetric damping envelope $\mathrm{e}^{S(\vec{g},\vec{t})}$ and the antisymmetric damping term $\mathrm{e}^{A(\vec{g},\vec{t})}$ comprise both a dependence on the scalar product $\vec{g}\cdot\vec{t}$. This dependence causes in particular an anisotropic damping of the linear interference fringes due to a partial temporal coherence of the electron beam.

### 3.2.5 Numerical diffractograms of thin amorphous objects

The above derived contrast transfer properties, are applied in the following to describe the intensity distribution in numerical diffractograms of thin amorphous objects.
A numerical diffractogram is computed from a digital image by digital Fourier transformation and subsequent calculation of the modulus square of the Fourier coefficients. The resulting intensity distribution is in the following denoted by $D(\vec{g}) = |I(\vec{g})|^2$, with

$$D(\vec{g}) = 4|\Psi_{\mathrm{e}}(\vec{g})|^2 \exp\left[2S(\vec{g},\vec{t})\right] \left\{\sin^2\left[\chi_t^{\mathrm{e}}(\vec{g})\right] + \sinh^2\left[A(\vec{g},\vec{t})\right]\right\}, \tag{3.2.28}$$

for $\vec{g} \neq 0$, as obtained by taking the modulus square of the right-hand side of equation (3.2.25). The term $\exp[2S(\vec{g},\vec{t})]$ is real-valued and describes a damping of the diffractogram intensity due to a partially coherent and tilted illumination. It separates into a product of $E'_{\mathrm{t}}(\vec{g},\vec{t})$ for the partial temporal coherence and $E'_{\mathrm{s}}(\vec{g},\vec{t})$ for the partial spatial coherence. The two separate envelopes are explicitly given by

$$E'(\vec{g},\vec{t}) = \exp[2S(\vec{g},\vec{t})] = E'_{\mathrm{t}}(\vec{g},\vec{t}) \cdot E'_{\mathrm{s}}(\vec{g},\vec{t}), \tag{3.2.29}$$

$$E'_{\mathrm{t}}(\vec{g},\vec{t}) = \exp\left\{-2\left(\frac{\pi\Delta\lambda}{2}\right)^2 \left[\vec{g}^{\,4} + 4(\vec{t}\cdot\vec{g})^2\right]\right\}, \tag{3.2.30}$$

$$E'_{\mathrm{s}}(\vec{g},\vec{t}) = \exp\left\{-2\left(\frac{q_0}{2}\right)^2 \left[\left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{e}}(\vec{g})\right|^2 + \left|\vec{\nabla}_{\vec{g}}\,\chi_t^{\mathrm{o}}(\vec{g})\right|^2\right]\right\}. \tag{3.2.31}$$

It is important to recognise that the rotational symmetry of the temporal envelope $E'_{\mathrm{t}}(\vec{g},\vec{t})$ is broken for tilted-beam illumination. In contrast to the well-known, rotationally symmetric damping term $E_{\mathrm{t}}^2(\vec{g},0) = E_{\mathrm{t}}^2(\vec{g})$ as in equation (1.3.25), the envelope function $E'_{\mathrm{t}}(\vec{g},\vec{t})$ decreases faster with larger spatial frequencies $|\vec{g}|$ in the azimuthal direction of the beam tilt. For an extremely large tilt angle, the intensity distribution in a diffractogram is reduced to a narrow stripe which lies perpendicular to the direction of the tilt azimuth.
The second part of equation (3.2.28) describes a sinusoidal amplitude modulation with the even part of the aberration function by $\sin^2[\chi_t^{\mathrm{e}}(\vec{g})]$. The effective odd aberrations $\chi_t^{\mathrm{o}}(\vec{g})$ do not affect the intensity-distribution of diffractograms. The sinh-term varies slowly with the argument $A(\vec{g},\vec{t})$ and is added to the sin-square term. This sinh-term describes a background motif $B'(\vec{g},\vec{t})$ in the diffractogram, with

$$B'(\vec{g},\vec{t}) = 4|\Psi_{\mathrm{e}}(\vec{g})|^2 \exp[2S(\vec{g},\vec{t})] \cdot \sinh^2\left[A(\vec{g},\vec{t})\right]. \tag{3.2.32}$$

The sinusoidal pattern can be observed straightforward by diffractograms of thin amorphous objects. In contrast to crystals, amorphous objects are characterised by the absence

of long-range order in their atomic structure and thus show no translational periodicity. Accordingly, electrons are diffracted by such materials into all azimuthal directions with equal probability. The maximum diffraction angle is limited by the amplitudes of the electron scattering factors. Approximations for the electron scattering factors can be found in refs. [83, 84]. Diffractograms of HRTEM images with thin amorphous objects show typically no intensity at spatial frequencies larger than approximately 5 $\mathrm{nm}^{-1}$. For a large modulus of the defocus, the pattern shows concentric rings, the so-called *Thon-rings* [85] as in figure 3.2.

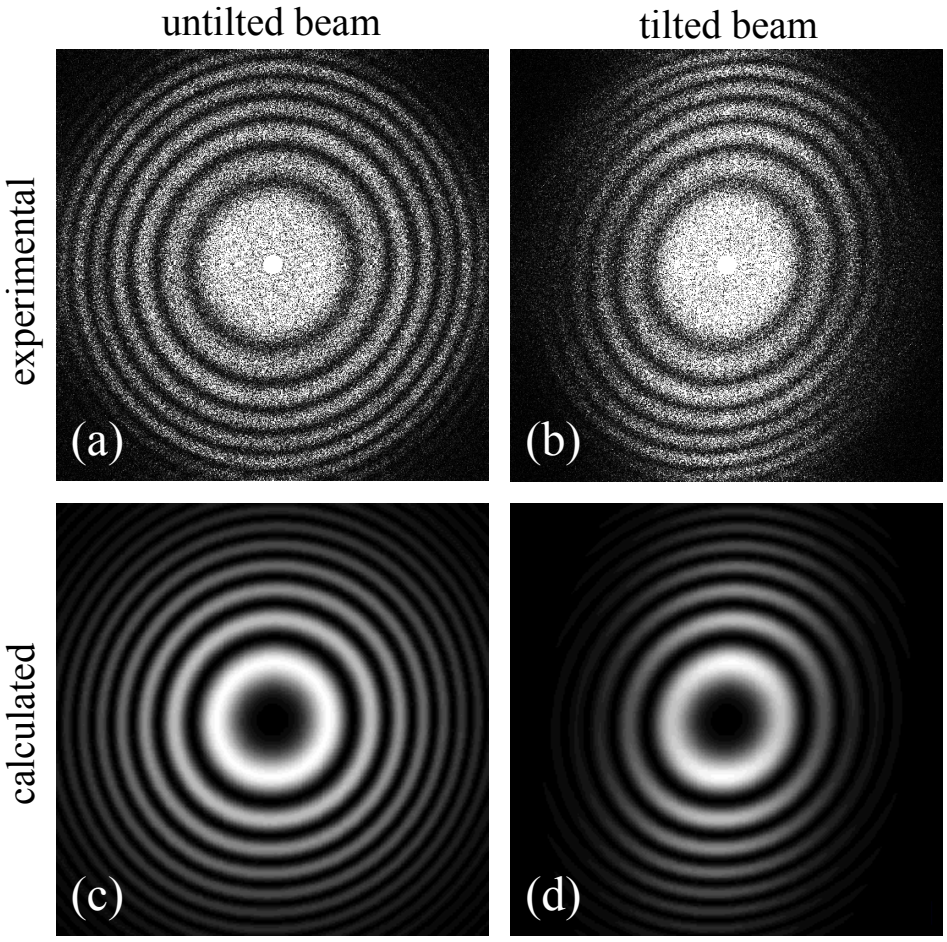



Figure 3.2: (a,b) Experimentally obtained and (c,d) calculated diffractograms of thin amorphous tantalum. The images (a,c) on the left side are obtained for untilted illumination, whereas (b,d) are obtained with tilted illumination. The applied beam tilt has a modulus of $\lambda|\vec{t}| = 35$ mrad in negative $x$-direction. (Imaging parameters: $U = 300$ kV, $\theta_0 = 0.2$ mrad, $\Delta = 2.6$ nm, $c_{20} = -187$ nm, $c_{22} = 6.4$ nm (64 °), $c_{31} = 296$ nm (0 °), $C_\mathrm{S} = -9$ $\mu$m).

In comparison, the upper, experimental diffractograms in figure 3.2 agree qualitatively to the theoretical calculations below. Except for a discrepancy at the lower spatial frequen-

cies, the examples for untilted illumination on the left, and for tilted illumination on the right exhibit similar damping and sinusoidal modulations.

In contrast to the theoretical diffractogram patterns, the experimental diffractograms in figure 3.2 do not show the low-frequency gap in the vicinity of $\vec{g} = 0$ which is expected by the linear contrast transfer theory (see section 1.3.4). The observed discrepancy must therefore be attributed to features of the electron diffraction which are not described by the WPOA, such as absorption contrast.

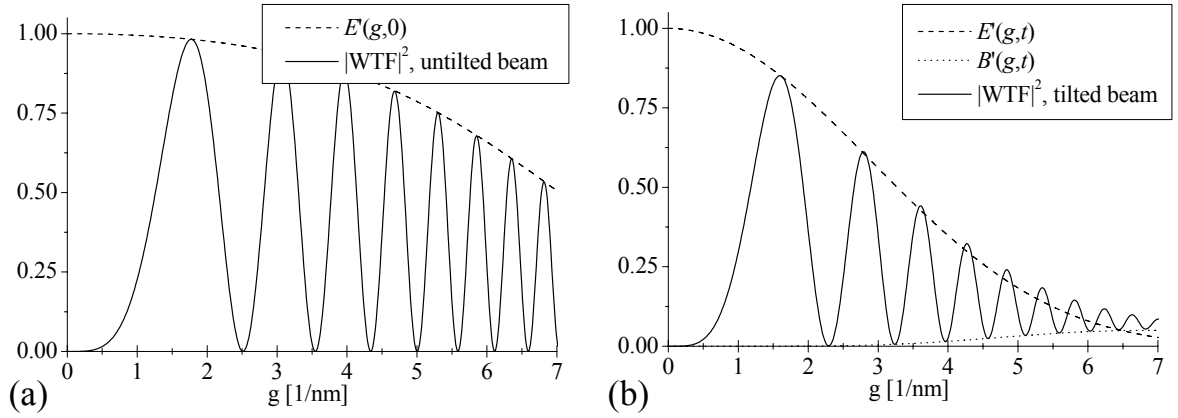



Figure 3.3: Modulus square of the wave-transfer function for (a) untilted and (b) tilted illumination in profile along the $x$-axis of the image, together with envelopes and background functions. The curves for the tilted illumination are calculated with a beam tilt of $\lambda|\vec{t}| = 18$ mrad in positive $x$-direction using the TEM-parameters: $U = 300$ kV, $\theta_0 = 0.2$ mrad, $\Delta = 3$ nm, $C_\mathrm{S} = -18$ $\mu$m, and $c_{20} = -80$ nm.

Figure 3.3 shows calculated profiles of the modulus square of the wave-transfer function $\mathrm{WTF}(\vec{g}\,,\vec{t}\,)$ of equation (3.2.26) along the $g_x$-axis. The left diagram shows the amplitudes of the Fourier components for the traditional untilted illumination, whereas the right diagram reveals stronger damping and a rising background for a beam tilt of $|\vec{t}\,| = 18$ mrad. The anisotropy of the temporal envelope $E'_\mathrm{t}(\vec{g}\,,\vec{t}\,)$ in equation (3.2.30) between the directions parallel and perpendicular to the beam tilt becomes stronger with increasing tilt modulus. As a consequence, the information transfer is reduced parallel to the azimuthal tilt direction. According to equation (3.2.30), the reduced information limit $g_{\mathrm{max},t}$ is given by

$$g_{\mathrm{max},t} = \left[\sqrt{g_\mathrm{max}^4 + 4\vec{t}^{\,4}} - 2\vec{t}^{\,2}\right]^{\frac{1}{2}}, \tag{3.2.33}$$

where $g_\mathrm{max}$ is the information limit in the unaffected, perpendicular direction. The diagram in figure 3.4 shows the dependence of $g_{\mathrm{max},t}$ on the beam tilt modulus $t$ relative to $g_\mathrm{max}$. The information limit in the direction of the tilt azimuth is already reduced to approximately $\frac{1}{2}g_\mathrm{max}$ when the tilt modulus is equal to $g_\mathrm{max}$.

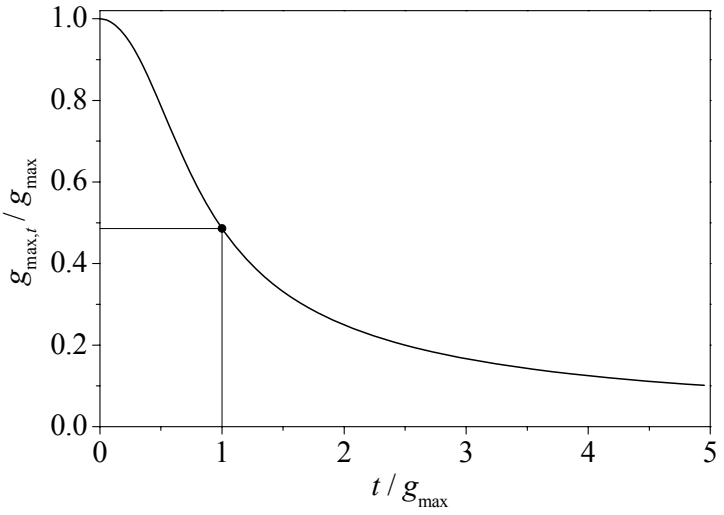



Figure 3.4: Dependence of the information limit $g_{\max,t}$ on the tilt modulus $t$ for the azimuthal direction in the image parallel to the beam-tilt. The reduced information limit $g_{\max,t}$ is depicted relative to the information limit $g_{\max}$ which is still achieved in the image direction perpendicular to the beam-tilt.

The contrast-transfer properties with tilted-beam illumination derived in this section will be used for quantitative analysis of diffractograms of thin amorphous objects in chapter 6. A new method which allows, for the first time, to measure the defocus-spread parameter $\Delta$ directly by analysing azimuthal envelope modulations due to tilted-beam illumination is described in section 7.4.

# Chapter 4

# Discussion of existing aberration-measurement methods

In high-resolution transmission electron microscopy several ways to measure aberrations are known. The basic principles behind the most frequently applied methods are introduced in this chapter. The main focus is placed here on the realisation of robust numerical procedures which can possibly achieve the required precision for aberration control in the Sub-Ångström regime. The reason why all the discussed methods apply beam tilts in order to determine the higher-order aberrations is elucidated in the beginning.

## 4.1 Why to use beam tilts for aberration measurement

There is a crucial restriction to aberration measurements for transmission electron microscopes. In contrast to methods used in traditional light optics, neither a test pattern nor a reference beam is available which would allow to quantitatively assess the imaging quality by an experiment which is independent on the microscope itself. Test-patterns have to be characterised with a better accuracy than the resolution of the investigated optical system. In the Sub-Ångström regime, no such test-pattern can be specified. A measurement of wave aberrations by an interference experiment requires a reference wave which is not influenced by the investigated optical system. Unfortunately, no such reference wave can be inserted into a TEM.
Since HRTEM images are interference patterns, the wavefront deviations between the interfering beams can be observed directly to a certain extent. This observation is typically limited to lower-order aberrations which act strongest on the image contrast. The smaller effects of higher-order aberrations are thus concealed. It is possible to dismantle

this concealment by using a tilted-beam illumination.
As discussed in section 3.2.1, the aberrations present with tilted-beam illumination can be expanded as polynomial $c'_{mn}(\vec{t}, \{c_{kl}\})$ in the beam tilt $\vec{t}$ and depend linearly on the set of axial aberrations $\{c_{kl}\}$ which are present for untilted illumination. Such *induction formulas* are derived explicitly for several lower-order aberrations in appendix C.2. The higher-order aberrations leave thus characteristic fingerprints in the effective lower-order aberrations. By measuring an effective lower-order aberration $c'_{mn}$ from images acquired with a defined set of different beam tilts, one obtains a respective set of linear equations. With more equations than unknowns, the set of linear equations can be solved by least-squares methods yielding the higher-order aberrations $\{c_{kl}\}$.
The precision of such an aberration measurement mainly depends on three effects: Firstly, on the accuracy of the direct measurement of the lower-order aberrations, secondly on temporal variations of the environmental conditions, and thirdly also on the correctness of the specified experimental parameters. The methods discussed in the following, intentionally apply beam tilts to determine higher-order aberrations, but different approaches are followed to measure the effective lower-order aberrations.

## 4.2 The tilt-induced displacement method

The combination of beam tilts and cross correlations to measure the aberrations of a TEM was investigated by several groups in recent years [64, 65, 69, 86–88]. By using the beam-tilt/cross-correlation (BTC) method, one determines the relative displacements between an image recorded without intentional beam tilt and several images recorded with a defined set of beam tilts.
Image displacements are measured by pair-wise cross correlations. The cross-correlation function (XCF) of two images is maximum at a distance from the origin of the cross-correlation image. This distance is equal to the tilt-induced image displacement, i.e. to the induced first-order aberration $\Delta c'_{11} = c'_{11} - c_{11}$. A formula which describes the dependence of the effective image displacement $c'_{11}$ on the beam tilt and on higher-order aberrations up to the $6^{\text{th}}$ order is derived in section C.2.1. From a larger set of such induction equations (C.2.3), as obtained from multiple displacement measurements with different beam tilts, the axial higher-order aberrations $c_{mn}$ with $(m > 1)$ are determined using least-squares methods.
Aberration measurements with the BTC-method can be performed with high speed, because only low electron doses are required to recognise the cross-correlation peak. Such

short-time exposures effectively reduce the radiation damage to the object [69]. The BTC-method does not require a specific type of material or structure as long as non-periodic details, like for example the edge of the object or amorphous material, remain in the field of view. This inherent independence on the material and the possibility to use low-intensity exposures bears an enormous advantage for the automated application of the BTC-method, because the aberrations can thus be measured at the object position of interest just before the actual TEM-experiment without further assumptions. Moreover, the computation of cross correlations requires relatively low numerical effort by using numerically efficient procedures like the Fast Fourier Transformation [89]. On commercial standard PCs, nowadays, cross correlations of images with a size of 1024×1024 pixels perform in less than one second.

Major disadvantages of the BTC-method are strong inconsistencies in the displacement data caused by unknown and uncontrollable drifts of the object during the experiment and by the induction of other aberrations like defocus and twofold astigmatism. According to the linear contrast-transfer theory for tilted-beam illumination (section 3.2), also the effective higher-order aberrations influence the cross-correlation function. Essentially, effective defocus $c'_{20}$ and effective twofold astigmatism $c'_{22}$ cause a sinusoidal modulation of the amplitude of the image Fourier coefficients. As a consequence, the correlation value at the position which corresponds to the "true" image displacement might become zero or, in the worst case, be actually the minimum of the XCF.

By applying a low-pass frequency filter in the cross correlation, the effect of the even aberrations can be reduced, because the aberrations affect image Fourier components of the higher spatial frequencies stronger. However, a too rigid cut-off of higher frequencies increases the width of the correlation-peak which leads to a reduced precision of the displacement measurement. A correction of the XCF for the influence of the effective even aberrations is possible by an iterative process starting with the results of an uncorrected measurement [69].

Measurement errors caused by object drift and misalignments of the tilt-stage are typically larger than the statistical error of the directly measured image displacements. In situations where the drift is small, a precision close to the requirements for one Ångström resolution can be achieved [88]. The amount of uncontrollable drift in most of the cases determines the finally achievable accuracy. However, it is not expected that an aberration measurement based on tilt-induced displacements can reliably reach the requirements for Sub-Ångström resolution.

## 4.3 The diffractogram method

The analysis of diffractograms of thin amorphous objects has a long tradition in assisting the alignment of transmission electron microscopes [85, 90–95]. A practical procedure to measure higher-order aberrations from a "tableau" of diffractograms of an amorphous object was first suggested by Zemlin *et al.* [68].

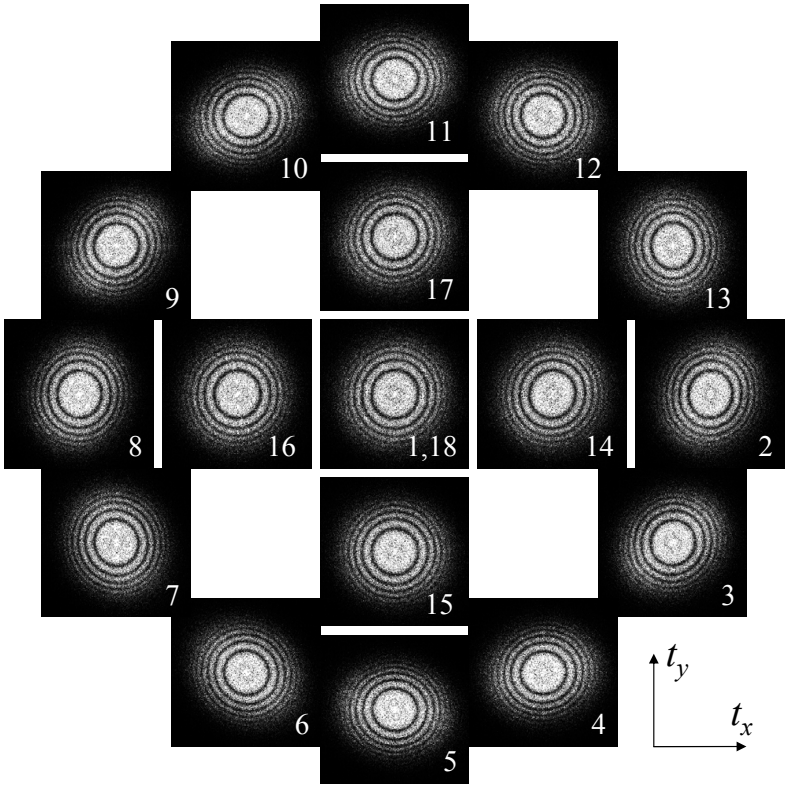



Figure 4.1: Exemplary diffractogram tableau of amorphous carbon as obtained from a series of images recorded with the Philips CM200C. The order in which the images are recorded is indicated by numbers 1 – 18. The arrangement of the diffractograms corresponds to the applied beam tilt during the exposures. The first and the last image are recorded with untilted beam. The maximum tilt modulus is used for the images 2 – 13 which are arranged equidistantly on a circle. Four images (14 – 17) are acquired with half maximum tilt modulus.

A series of images is recorded with a tilted beam of, in general, varying tilt modulus and varying azimuth and displayed in form of diffractograms as in figure 4.1. In former times, diffractograms were frequently obtained light-optically from the negatives of electron-optical micrographs. Nowadays, diffractograms are computed numerically by Fourier transformation of digital images as recorded with a CCD-camera and subsequent calculation of the modulus square. The computation of such a numerical diffractogram does not take longer than one second using conventional personal computers.

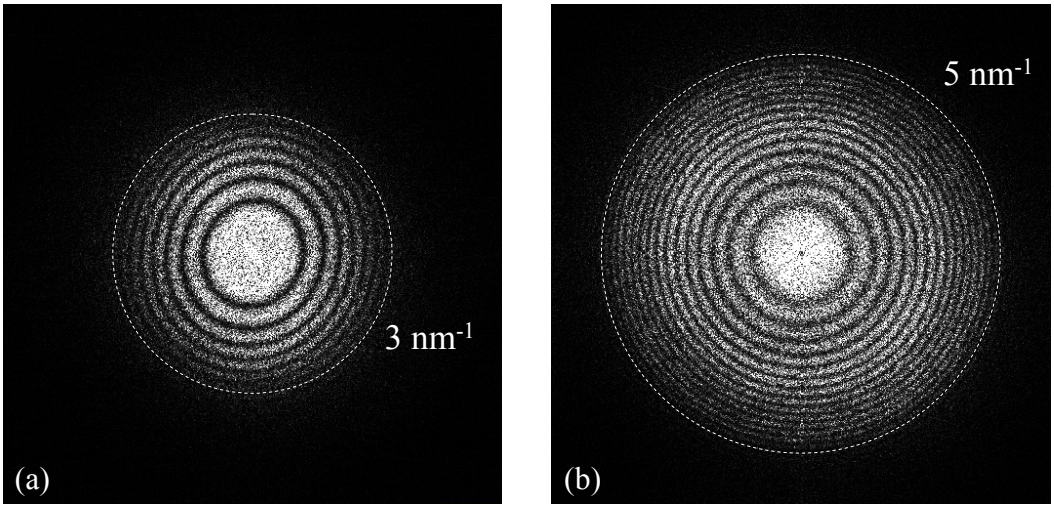



Figure 4.2: Numerical diffractograms of (a) thin amorphous carbon and (b) thin amorphous tantalum, recorded with the Philips CM200C. The circles denote the spatial frequency where the observable intensity is reduced to almost zero.

Exemplary diffractograms for a weakly and a strongly scattering atom type are shown in figure 4.2. As derived with untilted illumination in section 3.2, diffractograms of amorphous objects display modulations according to the sine square of the even aberrations. In contrast, odd aberrations do not influence the diffractogram pattern. Diffractograms of amorphous objects show typically no intensity for spatial frequencies larger than approximately 5 $\mathrm{nm}^{-1}$ due to decreasing scattering cross sections with increasing scattering angle, due to a damping of the contrast transfer by temporal and spatial incoherence, and due to the effect of a finite object thickness. Within this limited range of spatial frequencies, essentially only defocus and twofold astigmatism determine the shape of the sinusoidal diffractogram pattern. For a large defocus, the pattern yields concentric rings, so-called *Thon-rings* [85], which become elliptic or even hyperbolic depending on the strength of the twofold astigmatism.

Frequently, defocus and twofold astigmatism are measured from the positions the successive minima of the Thon-ring pattern of a single diffractogram [85, 91, 92]. Since the minima correspond to the zero crossings of the coherent contrast transfer function $\sin[\chi^{\mathrm{e}}(\vec{g})]$, the determination of the minima positions can only be erroneous due to an unknown background and the strong noise, whereas the rest of the pattern is additionally influenced by several, in practice also unknown, envelope functions.

When the diffractogram is recorded with tilted illumination, the shape of the sinusoidal pattern reflects an effective defocus $c'_{20}$ and an effective twofold astigmatism $c'_{22}$. In appendix C.2.2 and appendix C.2.3, formulas are derived which describe the induction of

$c'_{20}$ and $c'_{22}$ by the applied beam tilt. For larger higher-order aberrations, the effective second-order aberrations change more strongly, and thus also the shape of the diffractogram patterns observed with one tilt tableau exhibits strong variations. For aberration-corrected microscopes, the tilt-induced changes of defocus and twofold astigmatism are smaller. Consequently, the diffractograms of one tilt tableau exhibit then only slight mutual changes of the radius and the ellipticity of the Thon-rings. Such a case is shown in figure 4.1 for the spherical-aberration corrected Philips CM200C.

Currently, only one realisation of the beam-tilt/diffractogram (BTD) method is distributed as an alignment software for spherical-aberration correctors by the manufacturers of these devices. The accuracy of the complete procedure depends on the accuracy of the analysis of the single diffractograms, i.e. on the accuracy of the extraction of $c'_{20}$ and $c'_{22}$. In practice the existing solution achieves an accuracy of approximately 3 nm in effective defocus and effective twofold astigmatism with one diffractogram. This accuracy is sufficient to control aberrations in the resolution regime above one Ångström, but does not satisfy the demands for Sub-Ångström aberration control. Different tilt tableaus can be used to determine smaller or larger sets of higher-order aberrations automatically. The duration of the whole procedure does not take longer than the time required for the series acquisition.

The biggest advantage of the diffractogram method in contrast to the tilt-induced displacement method (section 4.2) is its reduced sensibility to mechanical and thermal instabilities which is the reason why it became the chief method for aberration measurement in HRTEM. Furthermore, the observable variation of the diffractogram patterns is not disturbed by other aberrations. The precision of the diffractogram method is mainly limited by the accuracy of the direct defocus and astigmatism measurement. With a more precise analysis of the diffractogram pattern, the precision of the diffractogram method can be improved substantially.

The crucial limit for the applicability of the diffractogram method is the requirement for amorphous material. Even small portions of crystalline material lead to peaks in the diffractogram whose local intensity can exceed the intensity of the sinusoidal pattern. Such diffractogram peaks constitute a major problem for an automatic, numerical diffractogram analysis. In typical materials-science investigations, the object contains only little amorphous material in the investigated area. A prior aberration correction with a different, amorphous test object is not practicable, since the aberrations change by an unacceptable amount after this test object is removed and the real object is inserted into the microscope. Therefore, in practice, the optical alignment of higher-order aberra-

tions is accomplished at a different area of the investigated object which contains enough amorphous material. However, defocus and twofold astigmatism have still to be adjusted manually at the actual area of interest, because a change of the object position usually changes these two aberrations.
The analysis of diffractograms requires a large defocus of some hundred nanometres, in order to obtain at least some modulations of the sinusoidal pattern. This requirement constitutes another drawback of the diffractogram method, since optimum alignments used for high-resolution imaging with small defocus values, like the optimum defocus proposed by Lentzen *et al.* [47], can thus not be analysed.

## 4.4 Aberration measurement based on wavefunction reconstruction

The complex-valued electron wavefunction, obtained for example by focal-series reconstruction (section 2.4.3), is still affected by residual aberrations. Residual aberrations, like defocus, twofold astigmatism, and coma, can be extracted from crystalline regions by numerical analysis of the reconstructed wavefunction. Thereby, only little knowledge about the object is required [96].
One approach is based on the phase-object approximation, where the amplitude of the exit-plane wavefunction should ideally be constant, i.e. the amplitude contrast is minimum. In order to determine the even aberrations from thin amorphous regions of a phase object, the contrast in the amplitude of the reconstructed wavefunction is minimised in real space [97,98]. An automated procedure to determine defocus and twofold astigmatism using the amplitude-minimisation approach is implemented in the TrueImage software for focal-series reconstruction [76].
Alternatively, the even aberrations can be obtained also on crystalline regions by minimisation of the contrast in the real part of the wavefunction within the weak-phase-object approximation. The minimisation yields reliable results even up to thickness values where the WPOA related to the crystal itself does in a strict sense not hold any more [96].
A practical scheme to measure higher-order aberrations using beam tilts and a direct measurement of even aberrations from a reconstructed wavefunction was developed by Meyer *et al.* [99, 100]. Several small focal series are recorded with a different beam tilt comprising, in total, 27 images as illustrated in figure 4.3. In the first part of the measurement procedure, a wavefunction is reconstructed from every small focal series. Effective values

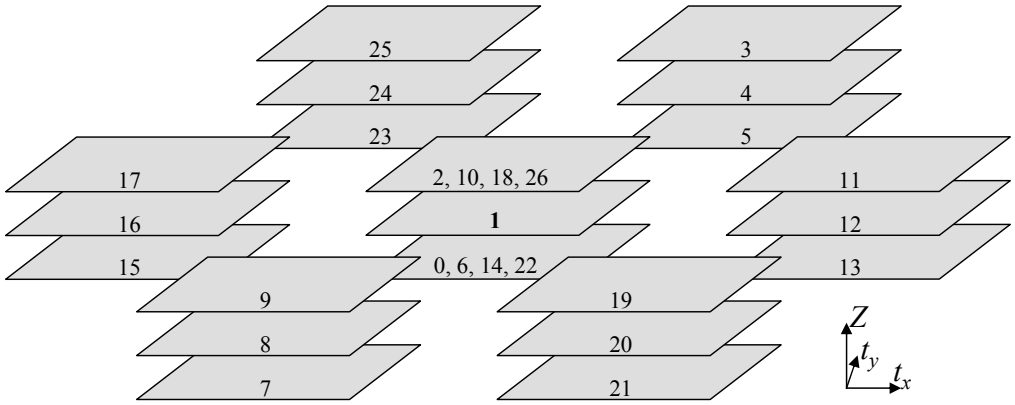



Figure 4.3: Schematic illustration of the combined defocus/tilt azimuth series with images numbered in recording sequence. A three-member focal series is recorded for each of the six different tilt azimuth angles and also for the axial illumination. Additional axial images are recorded between each pair of tilted datasets with unchanged defocus, allowing measurement of the defocus drift during acquisition. The order in which the images are recorded (indicated by numbers 0 ... 26) was chosen to minimise the influence of defocus drift and lens and deflector hysteresis. (Reproduced from [100])

of the even aberrations defocus $c'_{20}$ and twofold astigmatism $c'_{22}$ are determined from each of the reconstructed wavefunctions by a modified procedure of real-part minimisation similar to that described above. The induction formulas derived in the appendices C.2.2 and C.2.3 describe how $c'_{20}$ and $c'_{22}$ depend on the higher-order aberrations and the beam tilt. The higher-order even and odd aberrations are determined by a least squares fit of the induction formulas to the set of effective second-order aberrations. In the final step, the exit-plane wavefunction is reconstructed consistently by incorporation of all images and the measured aberrations.

Although not specified by the authors, one can estimate that the computation time for the measurement of all aberrations from the combined defocus/tilt series will be approximately equal to the duration of a focal-series reconstruction, since an iterative reconstruction procedure is applied. A focal-series reconstruction from the same amount of images lasts roughly 5 – 10 minutes on a commercial standard PC.

Meyer *et al.* demonstrate an accuracy of better than 1 nm for the absolute defocus and astigmatism coefficients in measurements with the complex oxides $Nb_{16}W_{18}O_{48}$ in [0 0 1]-projection and $Nd_4SrTi_5O_{17}$ in two different projections [100]. Thereby, temporal drifts of image position and defocus are considered in a linear model. This precision is sufficient to control the measured aberrations for Sub-Ångström resolution. That such precision could be demonstrated without specific assumptions about the material, directly on the area of interest for the HRTEM measurement bears an enormous advantage for the general

applicability of this method.
However, the combined defocus/tilt azimuth series illustrated in figure 4.3 restricts the measurable higher-order aberrations to axial coma $c_{31}$, threefold astigmatism $c_{33}$, and the spherical aberration $C_S$. More beam tilts are required to determine also fourth- and higher-order aberrations as relevant with spherical-aberration corrected microscopes with sufficient significance [100]. To what extent the tableau can be extended in practice, and whether the then achievable precision can still satisfy the requirements for Sub-Ångström resolution was not investigated.

## 4.5 Aberration measurement using caustic figures

An example of aberration measurement which applies beam tilts in a different manner than those discussed above, and a practical procedure for a coma-free alignment of transmission electron microscopes has been reported by Kimoto *et al.* [101]. A strongly defocused image of an amorphous object is recorded using a convergent electron beam. Due to the illumination by a convergent electron beam a continuous spectrum of beam tilts is rendered with one exposure. Geometrical ray aberrations as described by equation (2.2.9) are observed in such images as non-linear displacement of the diffracted electron beams forming so-called *caustic figures*.
In the centre of a caustic figure all beams for which the displacement due to the present aberrations cancel out are focussed in a so-called aberration-free point. A bright-field spot is generated by the undiffracted beam. An axial coma for example, shifts the aberration-free point relative to the bright-field spot. By the practical procedure for coma-free alignment proposed by Kimoto *et al.* [101], the incident beam is tilted until the aberration-free point coincides with the bright-field spot. The authors estimate the accuracy of the coma correction by 0.9 $\mu$m. This corresponds approximately to the individual $\pi/4$-limit of the axial coma for a resolution of 0.2 nm.
The caustic figure for a strong underfocus is delimited by a caustic curve. The caustic curve corresponds to the diffracted beams, where the beam-displacement due to higher-order aberrations compensates the displacement caused by the intentional underfocus. Due to the limited dynamic of the CCD-camera, the caustic curve can only be seen with relatively weak contrast compared to the central features of the whole caustic figure. The two images in figure 4.4 show the caustic curve recorded with the Philips CM200C. The six-foldness of the curve corresponds to the diffraction angles, where the sixfold astigmatism compensates the intentional underfocus.

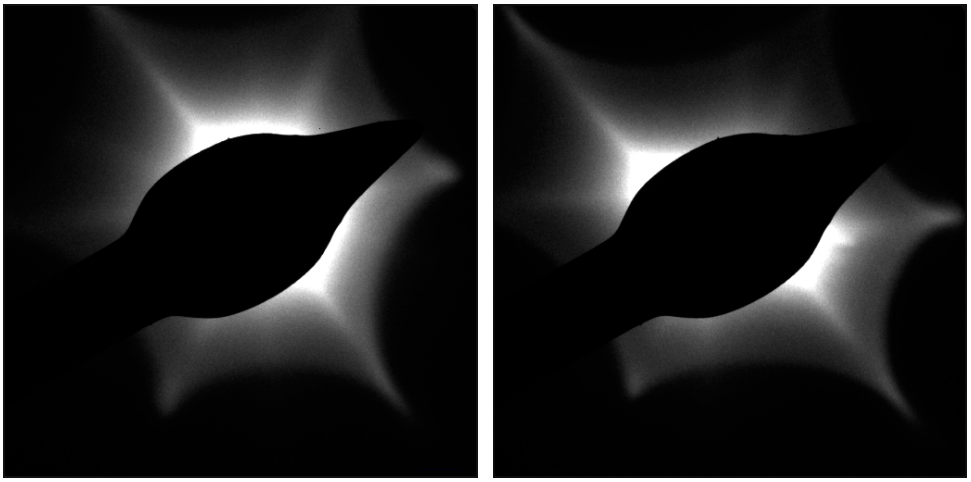


Figure 4.4: Caustic curve observed with amorphous $SrTiO_3$ using a strong underfocus of about $-6$ $\mu$m with the Philips CM200C. The insertion of the beam-stop was necessary, since the bright-field spot would have over-saturated the CCD. The sixfold symmetry of the caustic curve reflects the sixfold astigmatism introduced by the hexapole elements of the $C_S$-corrector. The right image shows a stretched caustic curve, caused by an additional strong twofold astigmatism of about 500 nm.

The difference of the caustic curves in the two images is caused by a twofold astigmatism of 500 nm. This magnitude of the twofold astigmatism is by three orders of magnitude larger than the required precision of a twofold-astigmatism measurement for Sub-Ångström resolution. A change of astigmatism values below 1 nm affects the shape of the caustic curve on a sub-pixel level. A quantitative analysis of caustic curves requires a prior precise determination of the applied defocus and a calibration of the magnification which actually changes with such large defocus values.

Separated caustic beam displacements can be observed upon imaging a crystalline object with a strong defocus of several microns and a focussed incident beam. Each diffracted beam is imaged in form of a so-called *Bragg-image* as shown in figure 4.5. Without aberrations, the Bragg-images should be placed ideally according to the spacings of the crystal lattice like in a diffraction pattern [23]. In the presence of aberrations, the Bragg-images are displaced from the ideal positions. Several groups have successfully measured the spherical aberration from displacements of multiple Bragg-images, e.g. in ref. [101].

Obviously this method is not universally applicable for an automated aberration measurement, since full knowledge about the material and its crystallographic projection is required in advance. Only then it is possible to index the experimental pattern in order to assign a specific diffraction angle to each of the observed Bragg-images. The non-uniform shape and intensity of the Bragg-images is also disadvantageous for an accurate measurement of their relative displacements.

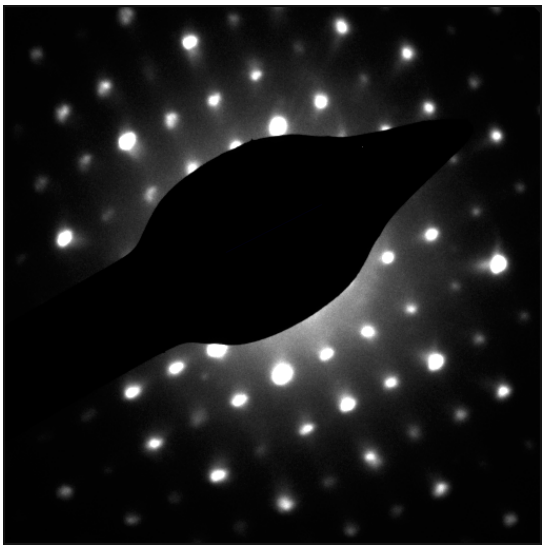


Figure 4.5: Caustic Bragg-image pattern of $SrTiO_3$ in [1 1 0] projection recorded with approximately +6 $\mu$m overfocus with the Philips CM200C. Significant non-linear deviations from the ideal diffraction pattern are not observed.

Budinger *et al.* [102] approximate the relative precision of their spherical-aberration measurement from multiple Bragg-images with 3 %. Considering a typical value of $C_S = 1$ mm, the absolute deviation would be in the order of some ten micrometres which is still an order of magnitude larger than the respective $\pi/4$-limit (2.7 $\mu$m) for Sub-Ångström resolution. Of course, the measurement error increases when a multitude of aberrations has to be determined. The analysis of the caustic Bragg-image pattern recorded with the Philips CM200C shown in figure 4.5 did not reveal significant non-linear displacements, although the present spherical aberration was measured with $-70$ $\mu$m using a different method.

## 4.6 Conclusion on existing methods

When preferring precision, the diffractogram method shows best potential to satisfy the Sub-Ångström requirements, provided the analysis of single diffractograms can be improved significantly. Unfortunately, its practical use is restricted due to the requirement of a larger defocus than commonly used in HRTEM and due to the demand for amorphous material which is rare close to scientifically interesting regions of the object. By using a different area of the object, the diffractogram method, nevertheless, allows for a determination of the higher-order aberrations which are independent on the object position. Usually, larger portions of amorphous material are present with most HRTEM specimens.

Aberration measurements based on the evaluation of tilt-induced displacements and the measurement scheme utilising reconstructed wavefunctions proposed by Meyer *et al.* seem to be least limited in the choice of the object. Unfortunately, the tilt-induced displacement method can not provide the required accuracy due to unavoidable inconsistencies caused by object drift. The wavefunction method is not favourable for in-situ aberration correction by hardware due to its long estimated computation time. In the view of radiation damage, the large number of recorded images might as well limit the quantitative significance of HRTEM investigations with this method. The caustic-figure method is effective for a rough pre-alignment of a TEM, but can not provide the precision, which is required to control the aberrations in the Sub-Ångström regime.

# Chapter 5

# Optimisation of the diffractogram method

Among the known methods discussed in the previous chapter, the diffractogram method (section 4.3) shows the best potential to develop an automatic aberration-measurement procedure which is able to fulfil the multi-aberration tolerance limit $\bar{\sigma}_\chi(g_{\max}) < \pi/4$ (see section 2.5.3) for Sub-Ångström resolution ($g_{\max} > 10$ nm$^{-1}$). Several directives for acquisition and evaluation of the necessary experimental data are proposed in this chapter in order to establish a reliable and quantitative measurement procedure. Optimum values are derived for some experimental parameters which impose as small as possible demands on the accuracy for the extraction of defocus and twofold astigmatism from single diffractograms while obtaining an as small as possible error of the measured aberration function with the whole procedure. As a result, an upper acceptable limit for the error of the single diffractogram analysis is deduced by matching the theoretically achievable precision with the complete aberration-measurement procedure against the $\pi/4$-criterion.

## 5.1 Experimental data acquisition

The experimental procedure performed for an aberration measurement using the diffractogram method starts by the acquisition of a series of images from an amorphous part of the object. Every image is recorded with a different beam tilt. The illumination conditions applied for the aberration measurement should be similar to those applied for the actual HRTEM measurement. For the subsequent analysis of the single diffractograms it is advantageous to select the magnification and the binning of the CCD-camera such that the discretisation of the acquired digital images is approximately 0.1 nm/pixel. This sampling

rate corresponds to a Nyquist frequency of 5 $nm^{-1}$ which resembles the already discussed damping of the intensity distribution in diffractograms of thin amorphous objects. The intensity distribution spans thus almost over the whole area of the diffractogram.
By choosing an underfocus of some hundred nanometres one obtains several Thon-rings in the diffractogram. The two diffractograms shown in figure 4.2 are examples for a good defocus and magnification setting. A too strong underfocus should be avoided, because this leads to a strong damping of the contrast transfer due to partial spatial coherence as discussed in section 1.3.5. As a consequence, less signal of the effective defocus and the effective twofold astigmatism is observable in the diffractograms.
Between the sequential acquisition of $N$ images, the incident beam is tilted following a predefined sequence $\vec{t}^{\,j}$, with $1 \leq j \leq N$. In this context, a raised letter $j$ denotes an index and not an exponent. Image acquisition and tilting the beam is typically controlled by prepared exposure procedures with the electron microscope. The optimum choice for the beam tilts is matter of a more detailed investigation in section 5.4.

## 5.2 Evaluation of the experimental data

The development of ultra-precise techniques for the analysis of single diffractograms of amorphous objects was one of the most elaborate tasks in this work. It is therefore treated separately in chapter 6. For a discussion of the evaluation procedures in the present chapter, it is assumed that the effective defocus $c'_{20}$ and the effective twofold astigmatism $c'_{22}$ can be measured directly from single diffractograms of amorphous objects with a given accuracy denoted by $\sigma_{20}$ and $\sigma_{22}$ respectively.
The effective second-order aberrations present with tilted illumination can be expressed by the following induction formulas as derived in the appendices C.2.2 and C.2.3:

$$\begin{aligned} c'_{20}(t) \;=\; & c_{20} + \frac{2}{3}\,c_{31}\cdot t^* + \frac{2}{3}\,c^*_{31} t + 2\,c_{40}\cdot (t^*) t + \frac{3}{4}\,c_{42}\cdot (t^*)^2 + \frac{3}{4}\,c^*_{42}\cdot t^2 \\ & + \frac{6}{5}\,c_{51}\cdot (t^*)^2 t + \frac{6}{5}\,c^*_{51}\cdot (t^*) t^2 + \frac{4}{5}\,c_{53}\cdot (t^*)^3 + \frac{4}{5}\,c^*_{53}\cdot t^3 + 3\,c_{60}\cdot (t^*)^2 t^2 \\ & + \frac{4}{3}\,c_{62}\cdot (t^*)^3 t + \frac{4}{3}\,c^*_{62}\cdot (t^*) t^3 + \frac{5}{6}\,c_{64}\cdot (t^*)^4 + \frac{5}{6}\,c^*_{64}\cdot t^4, \end{aligned} \tag{5.2.1}$$

$$
\begin{aligned}
c'_{22}(t) \;=\;& c_{22} + \frac{2}{3} c_{31} \cdot t + 2\, c_{33} \cdot t^* + c_{40} \cdot t^2 + \frac{3}{2}\, c_{42} \cdot (t^*)t + 3\, c_{44} \cdot (t^*)^2 \\
&+ \frac{6}{5}\, c_{51} \cdot (t^*)t^2 + \frac{2}{5}\, c^*_{51} \cdot t^3 + \frac{12}{5}\, c_{53} \cdot (t^*)^2 t + 4\, c_{55} \cdot (t^*)^3 \\
&+ 2\, c_{60} \cdot (t^*)t^3 + 2\, c_{62} \cdot (t^*)^2 t^2 + \frac{1}{3}\, c^*_{62} \cdot t^4 + \frac{10}{3}\, c_{64} \cdot (t^*)^3 t + 5\, c_{66} \cdot (t^*)^4.
\end{aligned}
\tag{5.2.2}
$$

where the beam tilt $t = t_x + \mathrm{i}t_y$ and the aberration coefficients $c_{mn} = c_{mnx} + \mathrm{i}c_{mny}$ are denoted as complex numbers. Obviously the effective defocus does not depend on the aberrations $c_{mm}$ which includes the three-, four-, five-, and sixfold astigmatism. In contrast, the effective twofold astigmatism depends on all aberrations of higher orders than two. For every recorded image one can thus formulate three linear equations which describe the effective second-order aberrations, one for the effective defocus $c'_{20}$ and two for the components $c'_{22x}$ and $c'_{22y}$ of the effective twofold astigmatism. Accordingly $N_\mathrm{E} = 3 \cdot N$ equations are obtained for the complete diffractogram tableau which is expressed here as vector equation

$$\vec{c'} = \mathbf{A} \cdot \vec{c}, \tag{5.2.3}$$

where the left-side vector $\vec{c'}$ is a list of all directly measured, effective aberration coefficients $c'^j_{20}$, $c'^j_{22x}$, $c'^j_{22y}$. The elements of the matrix $\mathbf{A}$ depend on the components $t^j_x$ and $t^j_y$ of the beam tilts according to the induction formulas (5.2.1) and (5.2.2). The solution vector $\vec{c}$ is a list of all higher-order aberrations $c_{mn}$ (present for $\vec{t} = 0$) which are for numerical evaluation also expressed by their real part $c_{mnx}$ and their imaginary part $c_{mny}$. The $N_\mathrm{E}$ equations have therefore to be solved for $N_\mathrm{A} = 2 \cdot A$ unknown parameters, where $A$ is the number of single aberrations which have to be measured.

### 5.2.1 Determination of higher-order aberrations

The solution vector of the set of linear equations (5.2.3) is obtained by numerical inversion of the matrix $\mathbf{A}$. With typically more linear equations than unknowns ($N_\mathrm{E} > N_\mathrm{A}$), the inverse matrix $\mathbf{A}^{-1}$ yields a least-squares solution for the higher-order aberrations by $\vec{c} = \mathbf{A}^{-1} \cdot \vec{c'}$.

In this work, a *Singular Value Decomposition* (SVD) [103] is applied for the matrix inversion. For overdetermined systems of linear equations, the SVD algorithm in practice never fails. Dependencies between effective aberrations and higher-order aberrations which are present in the induction formulas but not sampled by the experimental data can be suppressed during the inversion. The SVD algorithm contributes thus to a more robust computation of the solution, whereas for example the technique of Gaussian elimination tends to abort its computation on occurring singularities.

It is furthermore convenient that the SVD algorithm receives the so-called *design-matrix* **A** directly as input. Hence, no prior normalisation of the linear equations is required as for the Gaussian elimination. By application of the general induction formula discussed in section 3.2.1 the linear equations can in principle be extended on demand to an arbitrary set of aberrations in the left-side vector $\vec{c}'$, and even more important, also in the solution vector $\vec{c}$.

By using the SVD algorithm, the design-matrix **A** is decomposed into a product $\mathbf{U}\cdot\mathbf{W}\cdot\mathbf{V}^{\mathrm{T}}$ of an $M \times N$ column-orthogonal matrix **U**, an $N \times N$ diagonal matrix $\mathbf{W} = \mathrm{diag}(w_i)$ with positive or zero elements $w_i$ (the *singular values*), and the transpose of an $N \times N$ orthogonal matrix **V**. The decomposition of $\mathbf{A} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^{\mathrm{T}}$ allows for a simple inversion scheme with $\mathbf{A}^{-1} = \mathbf{V} \cdot [\mathrm{diag}(1/w_i)] \cdot \mathbf{U}^{\mathrm{T}}$. Singular diagonal elements $1/w_i$ are set to zero in practice which causes an exclusion of a certain subspace of the least-squares solutions [103].

### 5.2.2 Reliable error estimation

In order to obtain a reliable least-squares solution from the system of linear equations (5.2.3), each equation $j$ is weighted by the reciprocals $1/\sigma_{20}^{j}$, $1/\sigma_{22}^{j}$ of the measurement errors for effective defocus and effective twofold astigmatism from the respective diffractogram. A diffractogram analysis which exhibits a larger error thus contributes less to the solution. Parallel to the inversion of the design-matrix, a covariance matrix $\{v_{kl,mn}\}$ is computed with the SVD algorithm. Its diagonal entries $v_{mn,mn}$ denote the variances of the sample measurement from the determined mean value $c_{mn}$, i.e. the standard deviation squares of the individual aberrations. The non-diagonal entries denote covariances between the aberrations $c_{mn}$ and $c_{kl}$.

As discussed in section 2.5.3, the figure-of-merit $\bar{\sigma}_{\chi}(g_{\max})$ for the precision of aberration measurements can be calculated by straightforward error propagation from the variances and covariances at the information limit $g_{\max}$. The quantity $\bar{\sigma}_{\chi}(g_{\max})$ is in the following called *error of the measured aberration function*[1]. The error of the measured aberration function depends on the series of beam-tilt parameters $\vec{t}^{\,j}$, on the electron wavelength $\lambda$, and on the error-estimates $\sigma_{20}^{j}$, $\sigma_{22}^{j}$ of the diffractogram analysis.

[1] $\bar{\sigma}_{\chi}(g_{\max})$ is actually the square root of the mean variance of the measured aberration function at the information limit $g_{\max}$.

### 5.2.3 Consideration of temporal variations

An exclusion of the effective defocus data $c_{20}^{j}$ from the determination of higher-order aberrations yields more consistent aberration-measurement results. This observation is rather unexpected from a mathematical point of view, since approximately one third of the available data is omitted. The deviation of the effective defocus from an ideal behaviour according to the induction formula (5.2.1) is typically more than a factor of two larger compared to those observed for the effective twofold astigmatism. In order to avoid these stronger deviations from the model, the numerical evaluation of the experimental data applied in this work relies solely on the effective astigmatism.
The reason for a worse defocus consistency must be attributed mainly to a temporal drift of the object position. Although the diffractogram method is not sensitive to a drift in the object plane comprising the $x$- and $y$-directions, defocusing is identical to a movement of the object in $z$-direction. Lateral and vertical drifts are unfortunately not independent in general, for example due to a non-planar surface of the object. In the presence of a lateral object drift, one frequently observes also a defocus drift.
Temperature variations of the lens-cooling can also cause fluctuations of the twofold astigmatism [104]. A treatment of the respective time-dependencies is, however, only possible in a limited model, because the determination of additional aberration-drift parameters reduces the over-determination ratio of the linear equations system (LES) (5.2.3). The over-determination ratio is given by $N_{\mathrm{E}}/N_{\mathrm{A}}$ which is the ratio between the number $N_{\mathrm{E}}$ of linear equations and the number $N_{\mathrm{A}}$ of unknown variables. A reduced over-determination ratio of the LES leads to less significant results. The LES evaluated in this work therefore includes only two linear drift parameters $\dot{c}_{22} = \dot{c}_{22x} + \mathrm{i}\dot{c}_{22y}$ of the twofold astigmatism in the induction formula (5.2.2), with

$$c'_{22}(t,T) = c_{22} + \dot{c}_{22}(T - T_0) + \frac{2}{3}\, c_{31} \cdot t^* + \ldots \tag{5.2.4}$$

$T$ denotes the time when the respective image was recorded with respect to a given time index $T_0$ which can for example correspond to the start of the series acquisition.

## 5.3 Calibration of experimental parameters

In order to quantitatively measure effective aberrations from digital real-space images or diffractograms, the pixel vectors $\Delta\vec{p} = \Delta p_x \vec{e}_x + \Delta p_y \vec{e}_y + 0 \cdot \vec{e}_z$ in the digital arrays have to be translated reliably and accurately into real-space and Fourier-space vectors as illustrated in figure 5.1. The pixel rows and columns span a virtual inertial observer

frame which is expressed by orthonormal basis vectors $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$.

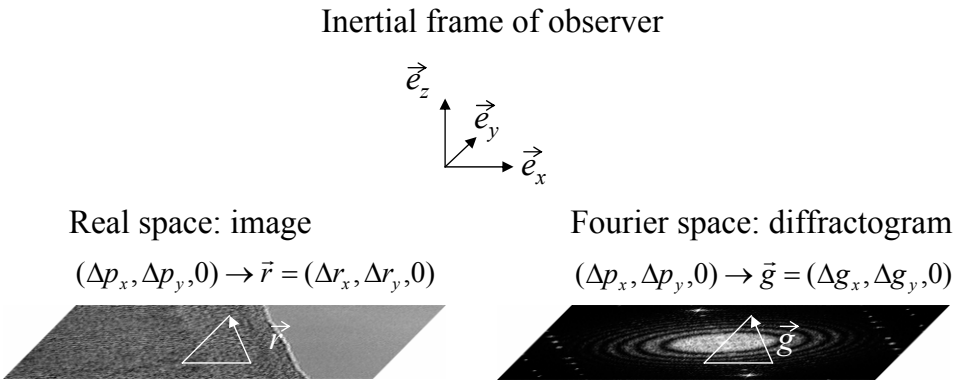



Figure 5.1: Representation of real-space vectors $\vec{r}$ in digital images and Fourier-space vectors $\vec{g}$ in numerical diffractograms. The pixel rows and columns of the recorded images span a virtual, inertial observer frame which is expressed by orthonormal basis vectors $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$.

Although image and diffractogram are two-dimensional, the observer frame is considered here as a three-dimensional system to define a rotational sense. A translation from lateral pixel coordinates in the digital data arrays to real-space vectors and Fourier-space vectors is expressed here by a linear but not necessarily orthogonal sampling rate. Thus, an anisotropic magnification can be considered. These translations are expressed by

- Real-space distance: $\Delta\vec{r} = (\Delta r_x, \Delta r_y) = \Delta p_x \vec{\alpha}_x + \Delta p_y \vec{\alpha}_y + 0 \cdot \vec{\alpha}_z$,
- Fourier-space distance: $\Delta\vec{g} = (\Delta g_x, \Delta g_y) = \Delta p_x \vec{\gamma}_x + \Delta p_y \vec{\gamma}_y + 0 \cdot \vec{\gamma}_z$, and
- Beam-tilt vector: $\vec{t} = (t_x, t_y) = t_x \vec{e}_x + t_y \vec{e}_y + 0 \cdot \vec{e}_z$,

with the basis vectors $\vec{\alpha}_i$, and $\vec{\gamma}_i$, $i \in \{x, y, z\}$. The real-space basis vectors $\vec{\alpha}_i$ have usually the dimension [nm/pixel], whereas the Fourier-space basis vectors $\vec{\gamma}_i$ have the dimension $[(\mathrm{nm} \cdot \mathrm{pixel})^{-1}]$. The beam-tilt vector $\vec{t}$ denotes the tilt angle $\vartheta = \lambda|\vec{t}|$ of the incident electron beam with respect to the normal illumination axis and an azimuth $\varphi_t$ to the $x$-axis of the observer system. Hence, $\vec{t}$ is a vector characterised by two spatial components $t_x = \vartheta/\lambda \cos(\varphi_t)$ and $t_y = \vartheta/\lambda \sin(\varphi_t)$ of the dimension $[\mathrm{nm}^{-1}]$ in the vector-space basis of the observer frame.

The rotational sense of the experimental results can vary for different microscopes depending on the readout direction of the CCD-camera, on the process of data transfer into digital arrays, and on the properties of the tilt stage. This ambiguity of the orientation is usually not considered in HRTEM measurements. For this reason, the observer system used for aberration measurements in this work is defined in connection to the $x$- and

$y$-axes of the real-space image in order to match the rotational sense of the measured aberrations to the HRTEM results. Accordingly, the basis vector $\vec{\alpha}_x$ of the physical real space is oriented parallel to the $x$-axis of the image by $\vec{\alpha}_x = \alpha_x \vec{e}_x$, whereas $\vec{\alpha}_y$ can have a different length and must not coincide with $\vec{e}_y$. The two vectors $\vec{\alpha}_x$ and $\vec{\alpha}_y$ span the image plane. In order to construct a right-hand oriented vector space, $\vec{\alpha}_z$ is defined by the normalised cross-product $(\vec{\alpha}_x \times \vec{\alpha}_y)/|\vec{\alpha}_x \times \vec{\alpha}_y|$ so that it is perpendicular to the image plane with $|\vec{\alpha}_z| = 1$.
The respective basis $\{\vec{\gamma}_x, \vec{\gamma}_y, \vec{\gamma}_z\}$ of the Fourier space follows from the equations

$$\vec{\gamma}_x = \frac{\vec{\alpha}_y \times \vec{\alpha}_z}{\vec{\alpha}_x \cdot (\vec{\alpha}_y \times \vec{\alpha}_z)}, \quad \vec{\gamma}_y = \frac{\vec{\alpha}_z \times \vec{\alpha}_x}{\vec{\alpha}_y \cdot (\vec{\alpha}_z \times \vec{\alpha}_x)}, \quad \vec{\gamma}_z = \frac{\vec{\alpha}_x \times \vec{\alpha}_y}{\vec{\alpha}_z \cdot (\vec{\alpha}_x \times \vec{\alpha}_y)} \tag{5.3.1}$$

analogous to the construction of the reciprocal crystal lattice as described in text books like ref. [23]. This definition ensures the reciprocity $|\vec{\gamma}_i| = 1/|\vec{\alpha}_i|,\ i \in \{x, y, z\}$ and a right-handed orientation of the vectors $\vec{\gamma}_i$. Since the real-space basis is not necessarily orthonormal, this can also be the case with the Fourier-space basis.
The sampling rates of real-space images and diffractograms are thus defined by 2×2-matrices

$$\mathbf{S}_\alpha = \begin{pmatrix} \alpha_{xx} & \alpha_{yx} \\ \alpha_{xy} & \alpha_{yy} \end{pmatrix}, \qquad \mathbf{S}_\gamma = \begin{pmatrix} \gamma_{xx} & \gamma_{yx} \\ \gamma_{xy} & \gamma_{yy} \end{pmatrix}. \tag{5.3.2}$$

The entries of $\mathbf{S}_\alpha$ and $\mathbf{S}_\gamma$ are related to the respective basis vectors with $\vec{\alpha}_x = (\alpha_{xx}, \alpha_{xy}, 0)$, $\vec{\alpha}_y = (\alpha_{yx}, \alpha_{yy}, 0)$, $\vec{\gamma}_x = (\gamma_{xx}, \gamma_{xy}, 0)$, and $\vec{\gamma}_y = (\gamma_{yx}, \gamma_{yy}, 0)$. Due to their mutual relation denoted in (5.3.1), both, the real-space sampling $\mathbf{S}_\alpha$ and the Fourier-space sampling $\mathbf{S}_\gamma$, can be determined by a calibration of the TEM magnification. An accurate representation of beam tilts in the observer frame requires a separate calibration procedure.

### 5.3.1 Calibration of the magnification

The magnification $M$ of a microscope is defined by the ratio between the size of a detail in the object and its respective size in the image. Since the discretisation of the image plane in pixels of the CCD-camera is constant, a calibration of the magnification is equivalent to calibration of the real-space sampling rate as defined by the matrix $\mathbf{S}_\alpha$ in equation (5.3.2). The magnification depends reciprocally on the sampling rate which relates pixel distances $(\Delta p_x, \Delta p_y)$ in a digital image to real-space distances $(\Delta r_x, \Delta r_y)$ with

$$\Delta r_x = \alpha_{xx}\, \Delta p_x + \alpha_{yx}\, \Delta p_y, \tag{5.3.3}$$

$$\Delta r_y = \alpha_{xy}\, \Delta p_x + \alpha_{yy}\, \Delta p_y. \tag{5.3.4}$$

The goal of a magnification calibration is to measure the four components $\alpha_{ij}$ with a relative error below 1 %.

The standard method described in this section determines the real-space sampling by analysing calculated diffractograms of HRTEM images recorded with an object of known crystal structure. The crystalline object has to be oriented such that the incident electron beam coincides with a zone-axis. In projection along a zone-axis, the image of the object shows a translational symmetry which is reflected by multiple peaks of high intensity in the diffractogram. According to the Bragg-condition for diffraction by crystals, such diffractogram peaks correspond to discrete diffraction angles which are related to vectors in the projected, reciprocal crystal lattice. The relation between projected reciprocal lattice and the observable scattering angles is described in text books like [23]. As an example, figure 5.2 shows a diffractogram calculated from an image of $SrTiO_3$ in [1 1 0]-projection exhibiting approximately 30 peaks.

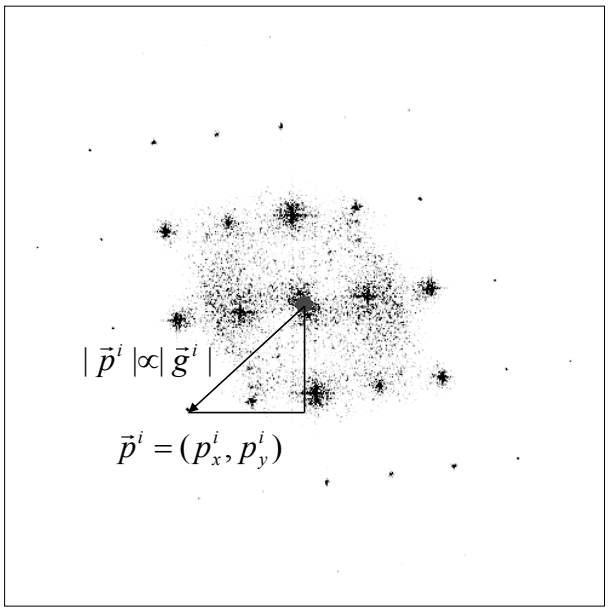



Figure 5.2: Numerical diffractogram as calculated from an image recorded of $SrTiO_3$ in [1 1 0]-projection with the Philips CM200C. In contrast to the previous display of the diffractogram intensity, here a low intensity is depicted in white and a high intensity is depicted in black. Multiple black peaks reflect the translational symmetry of the image. Their positions $\vec{p}^{\,i}$ relative to the zero peak in the centre of the diffractogram correspond to discrete diffraction angles $\vartheta^i = \lambda|\vec{g}^{\,i}|$.

In order to sketch the principle of this calibration procedure it is sufficient to consider that the diffraction angle $\vartheta^i$ for every observable peak $i$ at the pixel-position $\vec{p}^{\,i} = (p_x^i, p_y^i)$ is well-known for a known crystal structure in a specific zone-axis projection. Using this knowledge, a Fourier-space coordinate $\vec{g}^{\,i} = (g_x^i, g_y^i)$ can be assigned to each detected peak.

The modulus $g^i = |\vec{g}^{\,i}|$ is given by $g^i = \vartheta_i/\lambda$, where $\lambda$ is the electron-wavelength. The azimuthal direction of $\vec{g}^{\,i}$ is set equal to the azimuthal direction of the pixel-position.
The Fourier-space sampling is described by a linear model as in the following equations

$$g_x = \gamma_{xx}\, p_x + \gamma_{yx}\, p_y, \tag{5.3.5}$$

$$g_y = \gamma_{xy}\, p_x + \gamma_{yy}\, p_y. \tag{5.3.6}$$

The four coefficients can be determined by a least-squares fit of a linear model to the dataset $(\vec{p}^{\,i}, \vec{g}^{\,i})$. Finally, the Fourier-space sampling $\mathbf{S}_\gamma = \{\gamma_{ij}\}$ is transformed into a real-space sampling $\mathbf{S}_\alpha = \{\alpha_{ij}\}$ according to the relation between the real-space basis and the Fourier-space basis with

$$\vec{\alpha}_x = \frac{\vec{\gamma}_y \times \vec{\gamma}_z}{\vec{\gamma}_x \cdot (\vec{\gamma}_y \times \vec{\gamma}_z)}, \quad \vec{\alpha}_y = \frac{\vec{\gamma}_z \times \vec{\gamma}_x}{\vec{\gamma}_y \cdot (\vec{\gamma}_z \times \vec{\gamma}_x)}, \quad \vec{\alpha}_z = \frac{\vec{\gamma}_x \times \vec{\gamma}_y}{\vec{\gamma}_z \cdot (\vec{\gamma}_x \times \vec{\gamma}_y)}. \tag{5.3.7}$$

One obtains the following relations between the components $\alpha_{ij}$ and $\gamma_{ij}$

$$\alpha_{xx} = \gamma_{yy}/d, \qquad \alpha_{yx} = -\gamma_{xy}/d, \tag{5.3.8}$$

$$\alpha_{xy} = -\gamma_{yx}/d, \qquad \alpha_{yy} = \gamma_{xx}/d, \tag{5.3.9}$$

$$d = n(\gamma_{xx}\gamma_{yy} - \gamma_{xy}\gamma_{yx}) \tag{5.3.10}$$

where $n$ denotes the length of rows and columns of the quadratic digital images.

| | value [1/(nm·pix)] | standard deviation [1/(nm·pix)] |
|---|---|---|
| $\gamma_{xx}$ | 0.0483 | 0.0002 |
| $\gamma_{xy}$ | 0.0000 | 0.0002 |
| $\gamma_{yx}$ | −0.0001 | 0.0002 |
| $\gamma_{yy}$ | 0.0488 | 0.0002 |
| | value [nm/pix] | standard deviation [nm/pix] |
| $\alpha_{xx}$ | 0.0202 | 0.0001 |
| $\alpha_{xy}$ | 0.0000 | 0.0001 |
| $\alpha_{yx}$ | 0.0000 | 0.0006 |
| $\alpha_{yy}$ | 0.0200 | 0.0001 |

Table 5.1: Fitted Fourier-space sampling and the corresponding real-space sampling determined from figure 5.2.

Evaluation of the peak-pattern shown in figure 5.2 yields values for the coefficients of $\mathbf{S}_\gamma$ and $\mathbf{S}_\alpha$ as listed in table 5.1. The respective mean real-space sampling $\alpha = (|\vec{\alpha}_x| + |\vec{\alpha}_y|)/2$ for the applied magnification of $M = 340\,000$ and an image size of $(1024 \times 1024)$ pixels

is 20.1 (±0.1) pm per pixel, with an anisotropy $|(|\vec{\alpha}_x| - |\vec{\alpha}_y|)/(|\vec{\alpha}_x| + |\vec{\alpha}_y|)|$ of 0.5 %. The relative measurement error of 0.7 % is quite large for a magnification calibration, but can be improved by averaging over the results obtained from several images.

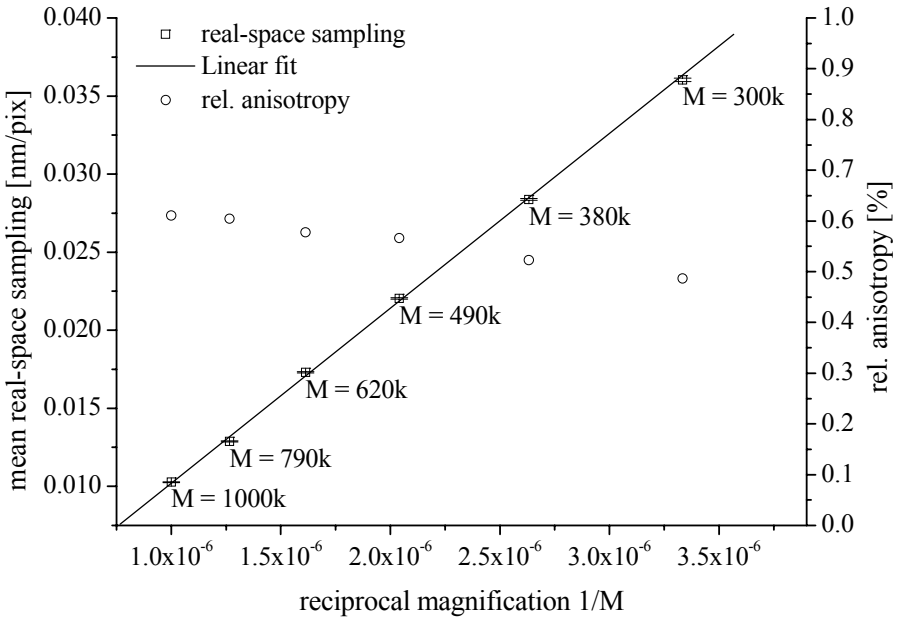



Figure 5.3: Calibrated real-space samplings of several magnification settings of the FEI Titan 80-300. The diagram displays the mean sampling with error bars and the relative anisotropy of the sampling for quadratic images containing (2048 × 2048) pixels.

Figure 5.3 plots calibrated mean real-space samplings for the FEI Titan 80-300 in dependence of the reciprocal magnification $1/M$ specified by the manufacturer. Due to cumulation of many measurements a relative error of 0.3 % was achieved.

### 5.3.2 Calibration of beam tilts

For the calibration of beam tilts, the TEM is operated in diffraction mode, where the microscope records the intensity of all beams in dependence of their angle to the optical axis. Without object, a parallel incident beam remains undiffracted and only one peak, the so-called *zero-beam*, is observed. A tilt of the incident beam causes a lateral translation of the zero-beam by an amount proportional to the tilt modulus. A sequence of different beam tilts (consecutively called *tilt tableau*) as applied for aberration measurements can thus be imaged in diffraction mode as shown in figure 5.4a.

Distinct peaks reflect the differently oriented beam tilts in a double-ring tableau which was intended to consist of 1 central tilt position, 12 equidistant positions on a circle at maximum tilt modulus, 4 equidistant positions on a circle at half of the maximum, and

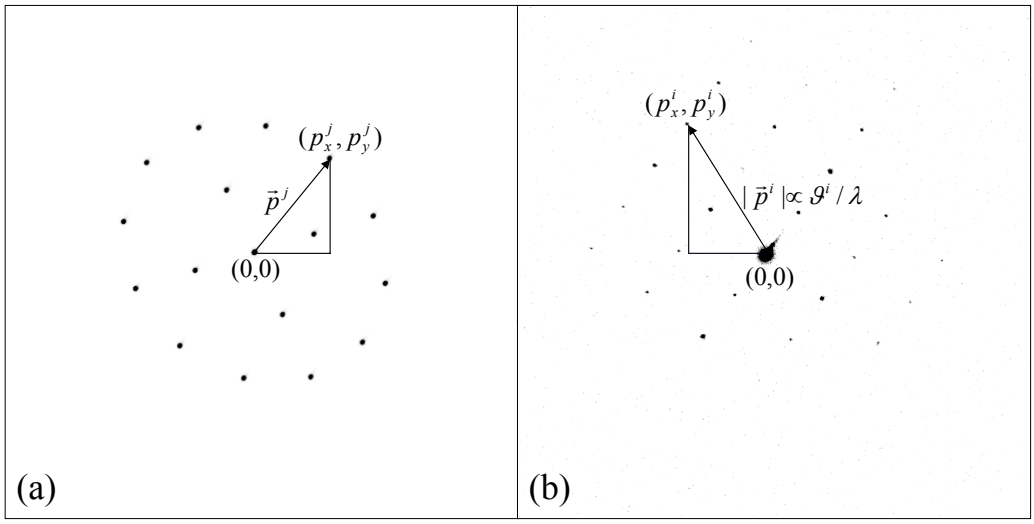



Figure 5.4: Diffraction images recorded with the Philips CM200C as used to calibrate a tilt tableau: (a) sum over a sequence of 17 images recorded without object while the electron beam is tilted according a standard sequence. The right image shows a diffraction pattern of $SrTiO_3$ in [1 1 0]-projection which is used to calibrate the diffraction images.

1 final image with again untilted beam. The two rings are obviously elongated and the azimuthal distribution of the tilt positions on these oval curves is not equidistant.
In order to determine the temporal stability of the peak-positions, several tilt series were recorded as shown in figure 5.4a. The standard deviations of the individual positions as measured by averaging over ten sequentially recorded tilt series are smaller than 0.4 % relative to the maximum tilt modulus.
The experimental procedure for the beam-tilt calibration starts with an orientation of a known crystalline object along a specific zone-axis. After acquisition of the diffraction pattern, the object is moved out of the illuminated area by using the object-holder. Then only the zero-beam remains visible. Finally, a series of images is recorded, each with a different beam tilt following a given tilt tableau. Every diffraction image of the series exhibits thus one peak of high intensity for the zero-beam which is shifted relative to the position obtained without beam tilt.
A quantitative assignment of physical tilt-angles $\vec{t}^{\,j}$ to the shifted positions $(p_x^j, p_y^j)$ of the zero-beam requires a prior calibration of the sampling of the diffraction images. For this purpose diffraction patterns were recorded with an object of known crystal structure in a known projection. These diffraction patterns are used to calibrate the Fourier-space sampling $\mathbf{S}_\gamma = \{\gamma_{ij}\}$ as described in the previous section, where specific diffraction angles $\vec{g}^{\,i} = \vartheta^i/\lambda$ are assigned to the detected peak-positions $(p_x^i, p_y^i)$ in the diffraction pattern. Figure 5.4b shows such a diffraction pattern of $SrTiO_3$ in [1 1 0]-projection. Table 5.2

lists the coefficients of the sampling $\mathbf{S}_\gamma$ of the diffraction images as determined from the $SrTiO_3$-[1 1 0] diffraction pattern.

| Coefficient | Value [1/(nm·pix)] | Standard deviation [1/(nm·pix)] |
|---|---|---|
| $\gamma_{xx}$ | 0.0493 | 0.0003 |
| $\gamma_{xy}$ | 0.0000 | 0.0003 |
| $\gamma_{yx}$ | -0.0008 | 0.0002 |
| $\gamma_{yy}$ | 0.0511 | 0.0003 |

Table 5.2: Linear coefficients of the sampling $\mathbf{S}_\gamma$ of the diffraction images in figure 5.4a, as determined from the diffraction pattern shown in figure 5.4b.

Since the diffraction patterns and the diffraction images showing the tilted zero-beam positions $\vec{p}^{\,j}$ were recorded with equal imaging conditions, the diffraction angles $\vec{g}^{\,i}$ observed in the diffraction patterns are of the same scale as the intentional beam tilts. Therefore, the following equations can be used to determine the beam tilt vectors $\vec{t}^{\,j}$ from the zero-beam positions $\vec{p}^{\,j}$ :

$$t_x = \gamma_{xx}\, p_x + \gamma_{yx}\, p_y, \tag{5.3.11}$$

$$t_y = \gamma_{xy}\, p_x + \gamma_{yy}\, p_y. \tag{5.3.12}$$

The tilt tableau shown in figure 5.5 is obtained by averaging over 10 independent calibrations from 10 diffraction patterns and 10 recorded series of the zero-beam tilts. The intended maximum tilt angle of 18 mrad is only reached by some of the positions at the outer ring (indices 2 – 13). The actual tilt angles vary between 14.8 mrad and 17.9 mrad around the mean value 16.4 mrad. A similar variation is found for the inner ring at approximately half of the maximum tilt angle. The tilt positions were measured with an accuracy of 1 % relative to the maximum tilt angle.

The described calibration procedure leaves the global azimuth offset of the beam tilts undetermined, because it is completely performed in the diffraction mode of the microscope. Switching from diffraction mode to imaging mode usually introduces an additional azimuthal rotation of the field of view. Therefore, the global azimuth of the calibrated tilt tableau was chosen such that the first beam tilt at the outer ring (index 2) coincides with the $x$-axis of the image. It is shown later in section 7.2 that the azimuthal offset of the tilt tableau can be determined by minimisation of the defocus consistency of an aberration measurement with the diffractogram method.

Besides the systematic errors caused by the over- or underestimation of the mean tilt

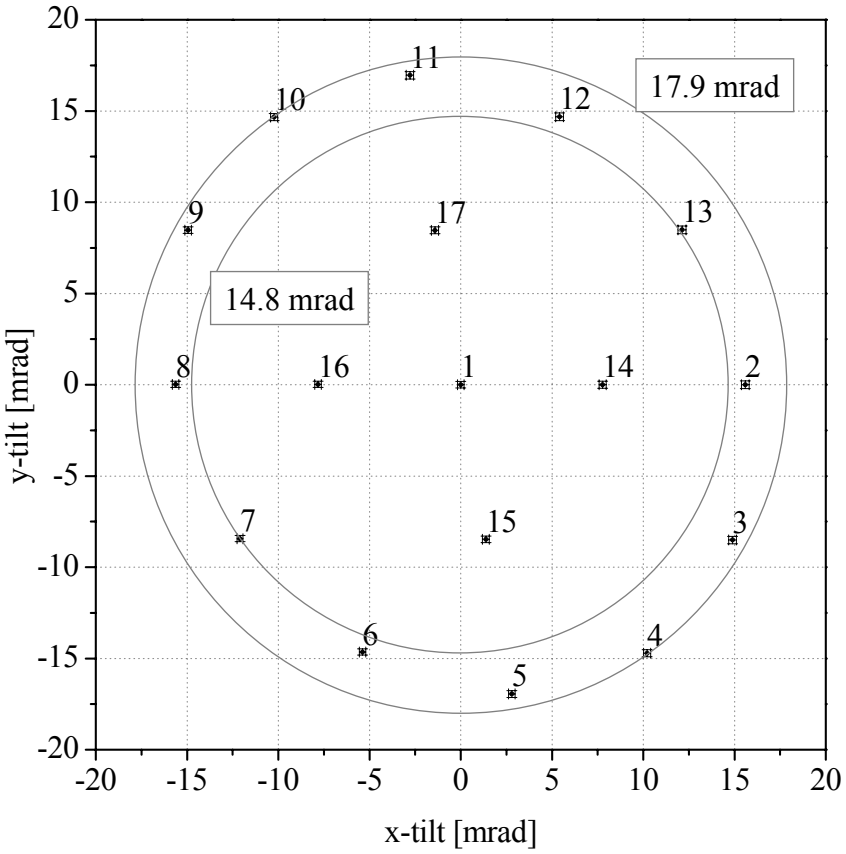



Figure 5.5: Results of a beam-tilt calibration from 10 diffraction patterns and 10 recorded series of the zero-beam tilts. Numbers beside the data points denote the sequence of recording. The standard deviations from the measured mean tilt positions are approximately 1 % compared to the maximum tilt modulus. The overall azimuth was chosen freely so that tilt position 2 coincides with the positive $x$-axis.

modulus, unconsidered elongation of the intended tilt tableau causes a misjudgement of measured non-circular aberrations especially when the aberration coefficients yield a large modulus. It is thus mandatory to apply a calibrated tilt tableau in order to obtain reliable results for the higher-order aberrations.

The results of tilt-tableau calibrations for the FEI-Titan 80-300 are shown in figure 5.6. The enhanced tableau consisting of 22 tilt positions with nominal maximum tilt angle of 24 mrad and the (1/12/4/1)-standard tableau with three different maximum tilt angles are frequently used for aberration measurements with this microscope. An elongation of the nominal tableaus is also observed here. The modulus of tilt positions which should be located on a circle vary by about 5 % around the mean modulus. The mean modulus is 7 % smaller than specified by the tilt-coil control.

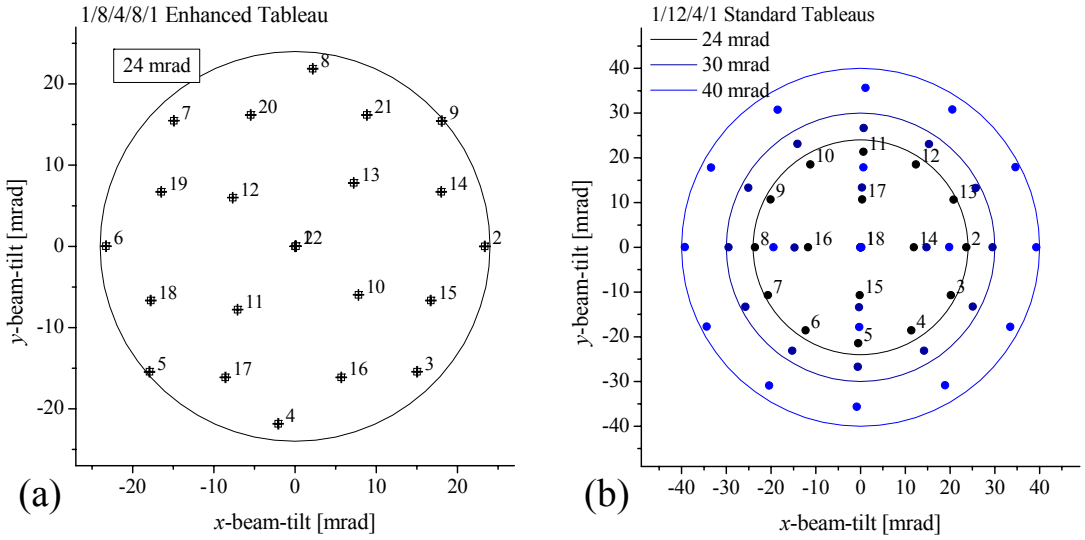



Figure 5.6: Calibration of (a) the enhanced tilt tableau with 24 mrad maximum tilt angle and (b) the standard tilt tableau with three different maximum tilt angles for the FEI Titan 80-300. Numbers beside the data points denote the sequence of recording. The mean tilt parameters were calibrated with an average accuracy of 0.5 % relative to the maximum tilt angle of the respective tableau.

## 5.4 Investigations on optimum tilt settings

The algorithm used to invert the design-matrix of the linear equation system (5.2.3) already determines the covariance matrix $\{v_{kl,mn}\}$. This allows for an investigation of the influence of the beam-tilt setting on the error $\bar{\sigma}_\chi(g_{\max})$ of the measured aberration function. Thereby, an information limit of $g_{\max} = 12.5$ nm$^{-1}$, and an electron wavelength of $\lambda = 1.969$ pm is considered. The individual measurement errors $\sigma_{22}^j$ are assumed to be equal $\sigma_{22}^j = \sigma_{22}$ for the analysis of all diffractograms. By successive evaluations of $\bar{\sigma}_\chi(g_{\max})$ with different values of $\sigma_{22}$, the required precision of the diffractogram analysis is determined which fulfils the tolerance limit $\bar{\sigma}_\chi(g_{\max}) \leq \pi/4$.
As derived in section 3.2, a tilted incident beam decreases the information limit in the direction parallel to the beam-tilt azimuth. With less available information in the diffractogram, the accuracy of the diffractogram analysis for the effective twofold astigmatism is consequently reduced. This effect will only be considered for the estimation of an optimum value of the maximum tilt modulus.

### 5.4.1 Necessary number of images

In general, least-squares solutions of a set of linear equations can be determined with better accuracy the more sample measurements are provided with significant precision.

In order to improve the precision of the aberration measurement it is consequently advantageous to record as many images with different beam tilts as possible. However, a compromise between a better statistical average and the duration of the experiment has to be found. Assuming approximately one second for the acquisition of every image and, in addition, an equal amount of time for the analysis of the diffractograms, a measurement which should not last longer than one minute can use a maximum of 30 images.
It is clear, that the required minimum number $N$ of images with different beam tilts depends on the number $N_\mathrm{A}$ of aberration parameters which have to be determined. Since, with the measurement of the effective astigmatism, two values are determined for every image, the quotient $R_\mathrm{det} = 2 \cdot N/N_\mathrm{A}$ denotes the over-determination ratio of the LES, with $2N - N_\mathrm{A} = N_\mathrm{A}(R_\mathrm{det} - 1)$ degrees of freedom. One needs at least some degrees of freedom to account for statistic and systematic measurement errors. In practice, an over-determination ratio of $R_\mathrm{det} > 1.5$ provides reliable least-squares solutions.
The polynomial expansion of the induction formula (5.2.2) in beam tilts is characterised by $N_\mathrm{A} = (M+1) \cdot (M+2)/2 - 4$ unknown parameters, when higher-order aberrations up to the maximum order $M$ have to be measured. Consequently, one needs to record a minimum of $N = 1.5 \cdot N_\mathrm{A}/2 = 18$ images with different beam tilts in order to determine all higher-order aberrations up to the $6^\mathrm{th}$ order with sufficient significance.
The frequently applied *standard tilt tableau* (see figure 4.1) contains $N = 18$ images in total with 12 equidistant positions on a circle at maximum tilt modulus $t_\mathrm{max}$, and 4 positions on the ring at half-maximum tilt. The first and the last image of the series are recorded with zero beam-tilt to easily observe a temporal drift of the twofold astigmatism and the object during the measurement. This tilt-tableau therefore provides $2 \cdot (N-1) = 34$ independent observables. However, 26 parameters have to be determined, including the two linear drift parameters $\dot{c}_{22x}$ and $\dot{c}_{22y}$. Due to an over-determination ratio of only $R_\mathrm{det} \approx 1.3$, this tableau tends to yield insignificant results for the higher-order aberrations up to the $6^\mathrm{th}$ order. In practice the five parameters $c_{60}$, $c_{62x}$, $c_{62y}$, $c_{64x}$, and $c_{64y}$ are excluded from the solution of the LES when using the standard tableau. The solution is then overdetermined sufficiently by $R_\mathrm{det} \approx 1.6$.

### 5.4.2 Distribution of beam tilts

Different concepts of tilt settings are investigated in the following. Since the measurement of higher-order aberrations involves the solution of general linear equations, there is no intrinsic necessity to restrict the tilt distribution to rings as with the already discussed standard tableau. Significant solutions for the higher-order aberrations can be obtained

as long as the independent variable, the beam tilt, is distributed more or less uniformly over a sufficiently large interval.

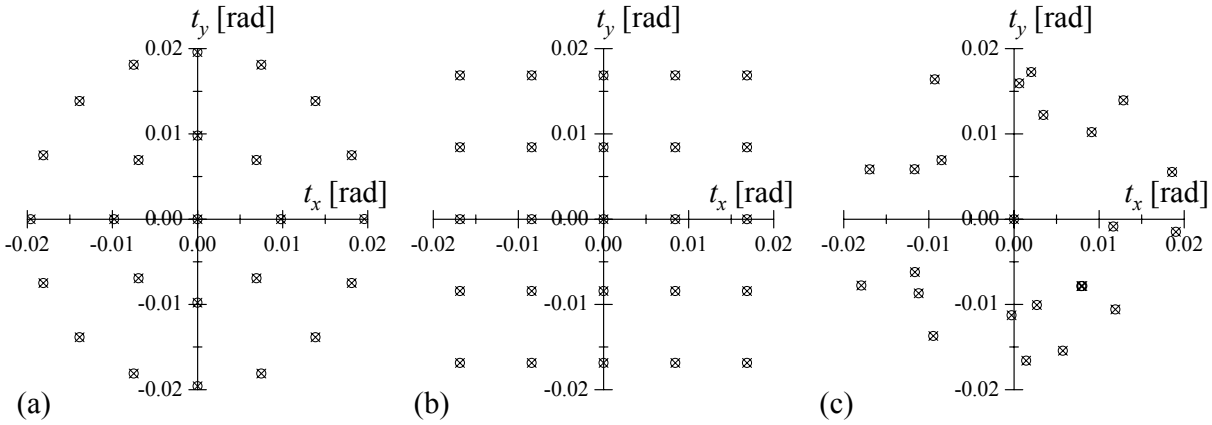



Figure 5.7: Tilt tableaus showing three different distribution models for 25 beam tilts: (a) (1/8/16)-double-ring tableau, (b) (5×5)-Cartesian tableau, and (c) a random tableau.

Three different distribution-models with 25 tilt positions are shown in figure 5.7. The achievable error $\bar{\sigma}_\chi(g_{\max})$ of the measured aberration function comprising aberrations up to the $6^{\text{th}}$ order is calculated for those tilt tableaus. The following procedure is used for the evaluation of $\bar{\sigma}_\chi(g_{\max})$ :

1. Define a tilt tableau $\vec{t}^{\,j}$ with $1 \leq j \leq N$,
2. Assume an equal precision $\sigma_{22}$ for the extraction of the effective twofold astigmatism from all diffractograms,
3. Calculate the design matrix $\mathbf{A}$ for the LES (5.2.3) by using the induction formula (5.2.2) for the twofold astigmatism with $\vec{t}^{\,j}$ and the weighting factor $1/\sigma_{22}$ for all rows $j$,
4. Calculate the inverse matrix $\mathbf{A}^{-1}$ and the covariance matrix $\{v_{kl,mn}\}$ by Singular Value Decomposition,
5. Calculate $\bar{\sigma}_\chi(g_{\max})$ to evaluate the precision of the aberration measurement which would be possible with the parameters defined in the steps (1) and (2).

The plot of the error of the measured aberration function in figure 5.8 shows only minor differences between the double-ring, the Cartesian, and the random tilt distribution. All these models satisfy the $\pi/4$-limit for $\bar{\sigma}_\chi$ at Sub-Ångström resolution ($g_{\max} = 12.5$ nm$^{-1}$) if the effective twofold astigmatism can be measured from every diffractogram with an

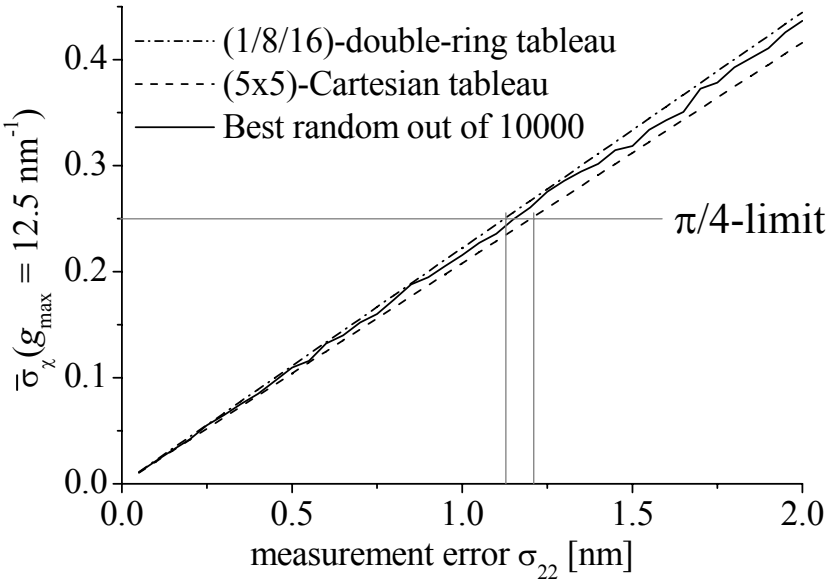



Figure 5.8: Achievable $\bar{\sigma}_\chi(g_{\max})$ in fractions of $\pi$ for $g_{\max} = 12.5\ \mathrm{nm}^{-1}$ considering three different tilt tableaus. The $\sigma_{22}$ varied to estimate the required precision for the extraction of the twofold astigmatism from single diffractograms, yielding the condition $\sigma_{22} < 1.2$ nm to satisfy the multi-aberration tolerance limit $\bar{\sigma}_\chi(g_{\max}) < \pi/4$ with the complete aberration measurement.

accuracy of $\sigma_{22} < 1.2$ nm. Obviously, the error of the aberration function depends linearly on the accuracy of the diffractogram analysis.

That also random, non-uniform distributions do not show significant disadvantages indicates a rather low sensibility of the least-squares solution on the actual beam-tilt distribution with a sufficient over-determination. It is therefore not necessary to switch to a different than the frequently used double-ring tableaus, except for a minor optimisation of the radius $t_i$ of the inner ring. The contour plot in figure 5.9 suggest a ratio of $t_\mathrm{i}/t_{\max} = 0.7$ instead the traditional ratio of 0.5 to yield the smallest error of the aberration function with the (1/8/16)-double-ring tableau for a given $\sigma_{22}$. This value for the optimum ratio for the ring-radii corresponds to the results reported in ref. [105], where also the tilt-induced defocus is considered for the measurement of higher-order aberrations with the diffractogram method.

### 5.4.3 Maximum modulus of beam tilts

With larger tilt angles the induced changes of the effective twofold astigmatism due to the higher-order aberrations become stronger. From this point of view, it was believed so far that an increase of the maximum beam-tilt modulus $t_{\max}$ reduces the error of the measured aberration function, i.e. using a larger maximum tilt should improve the

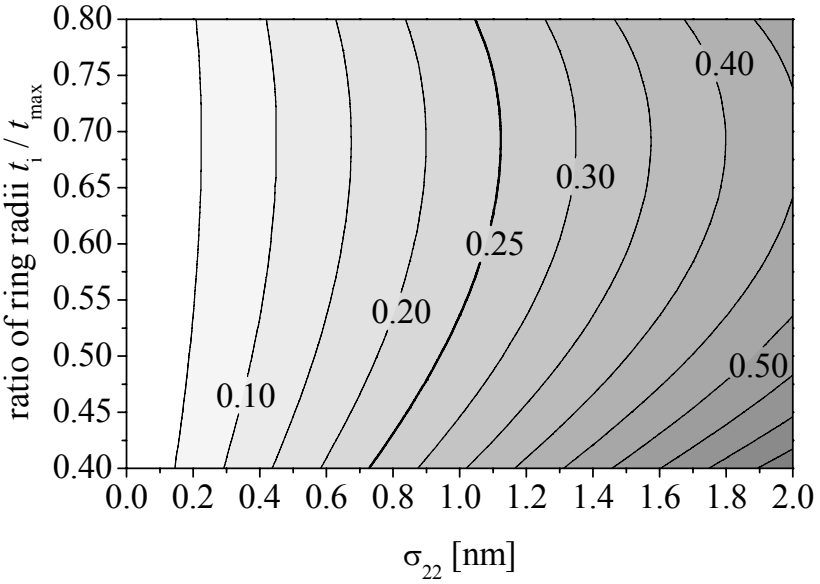



Figure 5.9: Contour plot of the error $\bar{\sigma}_\chi(g_{\max})$ of the measured aberration function in fractions of $\pi$, with $g_{\max} = 12.5$ nm$^{-1}$, in dependence of the error $\sigma_{22}$ of the analysis of single diffractograms. A (1/8/16)-double-ring tilt-tableau (fig. 5.7) is considered, where the ratio $t_\mathrm{i}/t_{\max}$ between the inner ring-radius $t_\mathrm{i}$ and the outer ring-radius $t_{\max}$ is varied. The thick contour line marks $\pi/4$-limit for the precision of aberration measurements.

precision of aberration measurement. Unfortunately, the linear interference terms are damped more strongly for larger beam tilts, as derived in section 3.2. It is demonstrated in figure 3.2 that the Thon-ring pattern disappears in the azimuthal direction parallel to beam tilt. Consequently, the effective twofold astigmatism is measured with a larger error $\sigma_{22}$ from the diffractograms with larger beam tilts.

The improvement to $\bar{\sigma}_\chi(g_{\max})$ due to an increased beam-tilt modulus and the reduction of $\sigma_{22}$ due to a non-isotropically reduced information limit compensate each other for a certain maximum tilt modulus $t_{\max,\mathrm{opt}}$. With this maximum tilt of the tableau, the precision of aberration measurements is optimum. It is assumed here that a damping of the contrast amplitude by a factor of $\mathrm{e}^{-1}$ significantly reduces the precision of the diffractogram analysis. The damping can be expressed by the temporal envelope $E'_\mathrm{t}(\vec{g}\,,\vec{t}\,)$ (3.2.30) yielding

$$E'_\mathrm{t}(\vec{g}\,,\vec{t}\,) = \frac{1}{\mathrm{e}} E'_\mathrm{t}(\vec{g}\,,0), \text{ for } \vec{g}\perp\vec{t} \tag{5.4.1}$$

$$\Rightarrow t_{\max,\mathrm{opt}} = \frac{1}{\pi\lambda\Delta\cdot g} = \frac{g_{\max}^2}{\sqrt{8}g}. \tag{5.4.2}$$

In the case of the FEI Titan 80-300 with $g_{\max} = 12.5$ nm$^{-1}$ and $\lambda = 1.969$ pm, the maximum tilt modulus should be set to $\lambda t_{\max,\mathrm{opt}} = 22$ mrad in order to provide an

isotropically damped diffractogram intensity up to $g = 5$ nm$^{-1}$. The optimum value for maximum tilt modulus for the Philips CM200C with $g_{\max} = 7.7$ nm$^{-1}$ and $\lambda = 2.508$ pm is $\lambda t_{\max,\mathrm{opt}} = 11$ mrad.

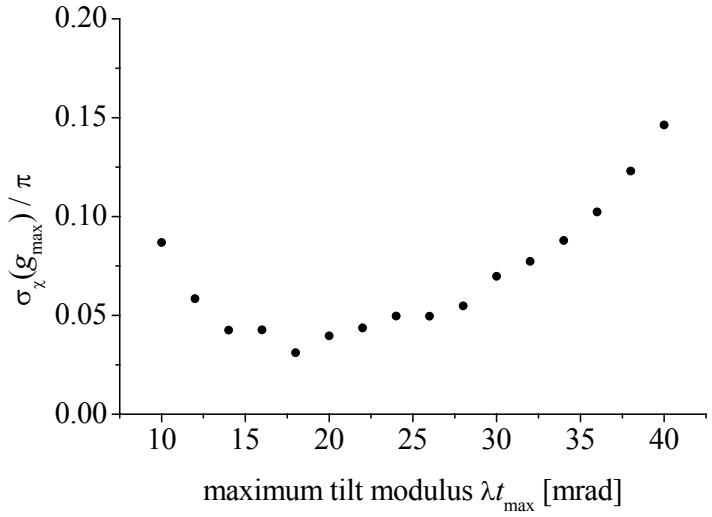



Figure 5.10: Error of the measured aberration function in fractions of $\pi$ comprising all aberrations of the orders 3 up to 6 as measured with different tilt moduli. All tilt series have been recorded from thin amorphous carbon films with the Philips CM200C.

In order to confirm the above theoretical considerations by experiments, several aberration measurements were performed with the Philips CM200C. Tilt series of thin amorphous carbon were recorded using the (1/12/4/1)-standard tableau with different maximum beam-tilt modulus $\lambda t_{\max}$ in the range from 10 - 40 mrad. The achieved error $\bar{\sigma}_\chi(g_{\max})$ of the measured aberration function is depicted in figure 5.10. It is minimum for a maximum beam-tilt modulus of approximately 18 mrad. With amorphous carbon, the Thon-ring pattern already disappears for spatial frequencies above $g = 3$ nm$^{-1}$. Inserting this value into equation (5.4.2) suggests an optimum of $\lambda t_{\max,\mathrm{opt}} = 17.5$ mrad, which agrees well to the experimental observation.

### 5.4.4 Required stability of beam tilts

Deviations between the applied beam tilts during the exposure and the tilt parameters used for the numerical evaluation will evidently cause measurement errors and systematically wrong results. Systematic deviations which correspond to a specific dependence in the induction formula (5.2.2) result in wrong aberration-coefficients. An overestimation of all tilt moduli for example leads to the underestimation of all aberration moduli. It is thus mandatory to calibrate the applied beam-tilt tableau at least once for each trans-

mission electron microscope prior to aberration measurements, for example by using the experimental procedure described in section 5.3.2.

Non-systematic deviations of the tilt positions from the calibration increase the error of the measured aberration function. A tolerance limit for temporal tilt variations and for the accuracy of beam-tilt calibrations is estimated here on the basis of numerical calculations including stochastic beam-tilt variations. The influence of random variations of the tilt positions on $\bar{\sigma}_\chi(g_{\max})$ can be evaluated by the procedure illustrated schematically in figure 5.11.

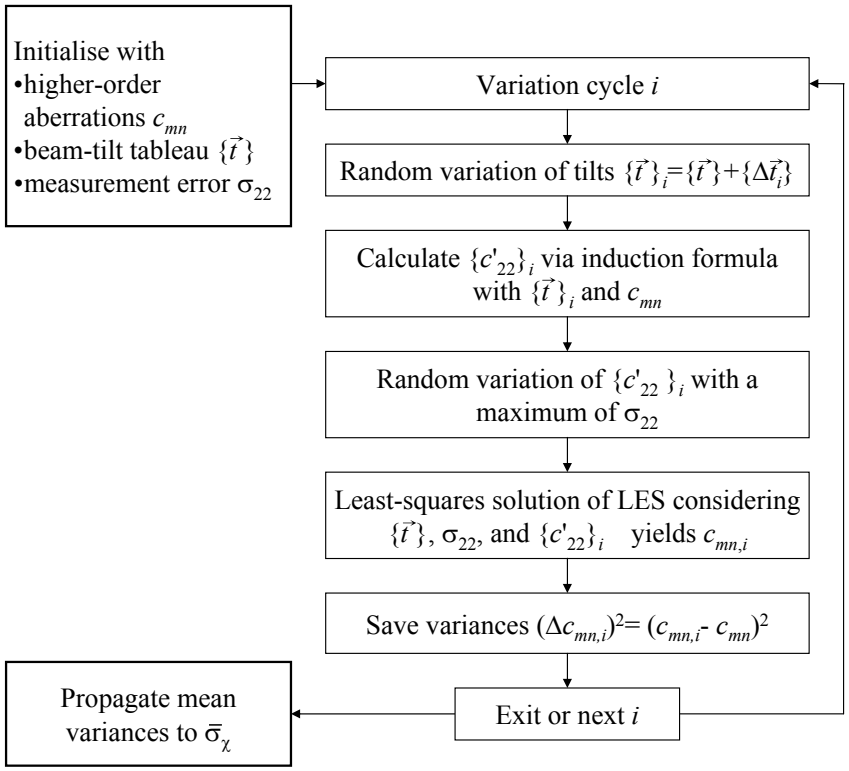



Figure 5.11: Scheme of the numerical procedure used to evaluate the influence of random beam-tilt variations on the error of the measured aberration function.

The procedure starts with a given tilt tableau $\{\vec{t}\} = \vec{t}^{(1)}, ..., \vec{t}^{j}, ..., \vec{t}^{N}$, with a set of aberrations $c_{mn}$ comprising the orders 2 up to 6, and with the measurement error $\sigma_{22}$ for the effective twofold astigmatism. For every cycle $i$ a random variation $\Delta\vec{t}_i^{\,j}$ is added to the initial beam tilts $\vec{t}^{\,j}$ yielding a varied tilt tableau $\{\vec{t}\}_i$. The maximum amount of random variations remains constant for all cycles and is denoted here by $|\Delta\vec{t}|$. The varied tilt tableau $\{\vec{t}\}_i$ represents the beam tilts realised in the microscope, while the initial tableau $\{\vec{t}\}$ stands for the beam tilts which are used for the evaluation of the experimental data. The hypothetical images comprise thus an effective astigmatism $c'_{22}(\vec{t}_i^{\,j})$ as

calculated by the induction formula (5.2.2) with the initial higher-order aberrations $c_{mn}$ and the beam tilt $\vec{t}_i^{\,j}$. A second random variation of the effective astigmatism values with a maximum of $\sigma_{22}$ simulates the error of the diffractogram analysis. Subsequent least-squares solution by assuming the initial tilt tableau and the varied $c'^{ij}_{22,i}$, evidently yields a different set of higher-order aberrations $c_{mn,i}$ than the initial values $c_{mn}$. The variances $(\Delta c_{mn,i})^2 = (c_{mn,i} - c_{mn})^2$ are saved for each cycle and averaged at the end.
The resulting mean variances are propagated to the error $\bar{\sigma}_\chi(g_{\max})$ of the aberration function. This error-propagation yields a statistical estimate for the precision of the aberration measurement based on random errors of the diffractogram analysis and random tilt variations which are not apparent to the observer.

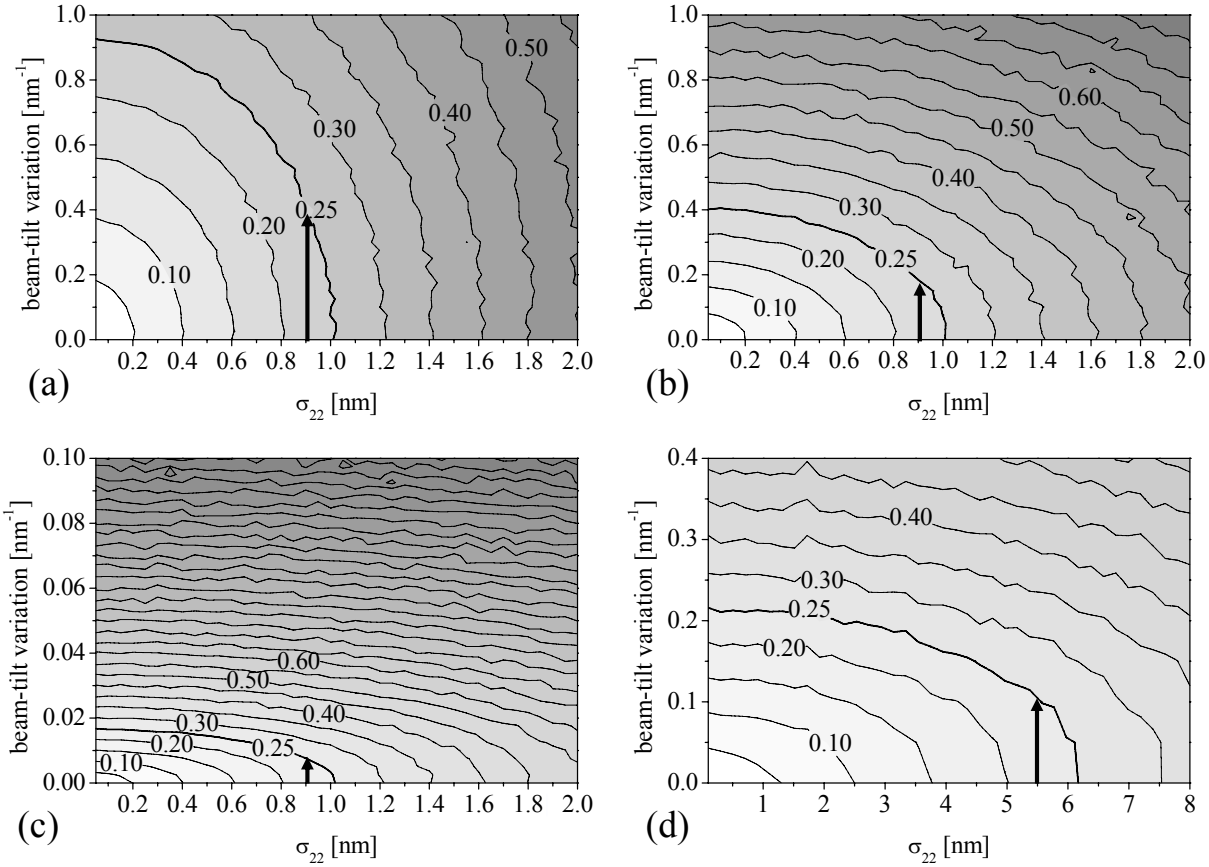



Figure 5.12: Contour plots showing the error $\bar{\sigma}_\chi(g_{\max})$ of the measured aberration function in fractions of $\pi$ as obtained by evaluation of statistical beam-tilt variations. The initial tilt tableau is a (1/8/16)-double-ring setup as in fig. 5.7a with ($t_{\max} = 10$ nm$^{-1}$). The four plots are calculated with different sets of higher-order aberrations: (a) corrected aberrations, (b) uncorrected aberrations, with a small spherical aberration, (c) and (d) a large spherical aberration of $C_S = 1$ mm. $\bar{\sigma}_\chi(g_{\max})$ is evaluated at the information limit $g_{\max} = 12.5$ nm$^{-1}$ for the plots (a, b, c) and at $g_{\max} = 5.0$ nm$^{-1}$ for the plot (d).

Figure 5.12 shows contour plots of $\bar{\sigma}_\chi(g_{\max})$ as obtained with the above described pro-

cedure. The different plots consider (a) a corrected, (b) a partially corrected, and (c,d) a completely uncorrected set of higher-order aberrations. The contour plots display the dependence of error of the measured aberration function on the measurement error $\sigma_{22}$ for the effective twofold astigmatism and on the maximum amount $|\Delta\vec{t}|$ of statistical beam-tilt variations.

A thicker contour line emphasises the tolerance limit $\bar{\sigma}_\chi(g_{\max}) \leq \pi/4$. The largest tolerable error of the diffractogram analysis $\sigma_{22}^{\max}$ is obtained without tilt variations. With increasing $|\Delta\vec{t}|$, the $\pi/4$-contour bends towards the vertical axis which means that a better accuracy demanded. The acceptance-limit for beam-tilt variations is proposed here by the amount of $|\Delta\vec{t}|$ where the error of the measured aberration function still reads $\pi/4$ for a 10 % stiffer accuracy requirement the diffractogram analysis with $\sigma_{22} = 0.9 \cdot \sigma_{22}^{\max}$. This relation is marked by the arrows in figure 5.12.

Figure 5.12a is calculated with small initial aberrations having values close to the measurement errors achievable with a completely stable tilt tableau, and thus reflect an optimum aberration correction. In this case, relative tilt variations of $|\Delta\vec{t}|/t_{\max} < 5\%$ are tolerable, where $t_{\max} = 10$ nm$^{-1}$ is the maximum tilt modulus in the initial tilt tableau. The aberration measurement is still accurate with relative tilt variations of less than 2 % in cases, where some aberrations are not completely corrected as assumed in figure 5.12b. Situations like this are frequently found in the beginning of HRTEM-experiments with $C_S$-corrected microscopes, after the object has just been inserted, or when the last correction was performed several hours ago.

The tolerance limit for beam-tilt variations decreases below 0.1 % for uncorrected microscopes ($C_S = 1$ mm, figure 5.12c) when aiming for Sub-Ångström resolution. In practice, uncorrected microscopes are applied for investigations in the resolution regime around 2 Å. For this resolution, again only a relative beam-tilt stability below 1 % is demanded.

In conclusion, the precision of the aberration measurement sustains no critical loss with relative variations $|\Delta\vec{t}|/t_{\max} < 1\%$ of the tilt positions for a wide range of aberration scenarios. The tolerance limit $|\Delta\vec{t}|/t_{\max} < 1\%$ can thus be proposed as criterion for the temporal stability of the experimentally applied beam tilts and for the required precision of beam-tilt calibrations. The tolerance limit is satisfied for both, the observed temporal stability, and the performed tilt calibrations reported in section 5.3.2.

## 5.5 Layout of the final measurement procedure

Optimum conditions for aberration measurements based on the diffractogram method are found. The complete experimental procedure used in this work is illustrated in figure 5.13. Although the effective defocus values $\{c'_{20}\}$ are measured for all diffractograms, they are excluded from the evaluation of the beam-tilt series as illustrated by the crossed-out arrow in order to avoid larger inconsistencies due to larger temporal variations.

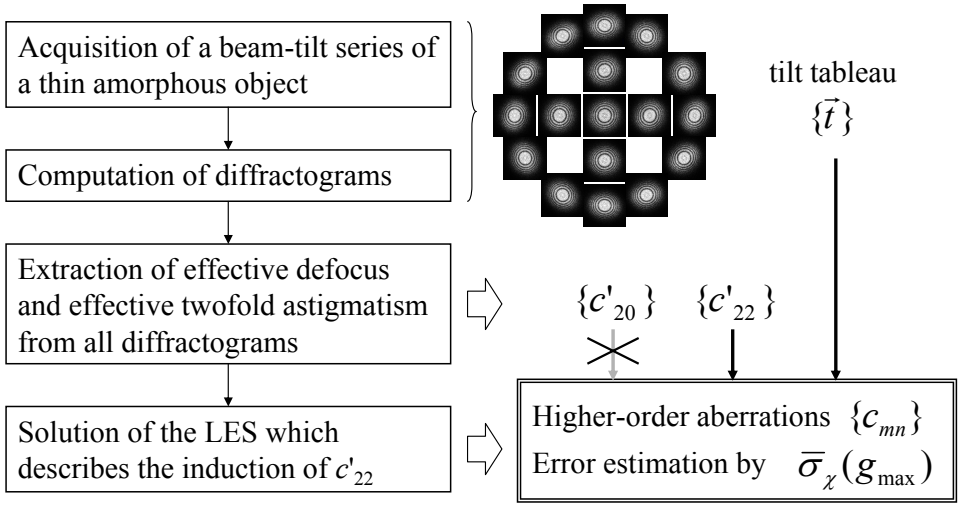



Figure 5.13: Illustration of the optimised experimental procedure based on the diffractogram method as used in this work for the measurement of higher-order aberrations in HRTEM. (LES = linear equation system)

The aberration function is determined as a least-squares solution of a set of linear equations by Singular Value Decomposition. This inversion algorithm provides a robust solution and yields reliable error estimates which can be propagated to the error of the measured aberration function.

Besides precise calibrations of the experimental parameters, optimum values are proposed leading to a smaller error of the aberration function. The respective investigations for an optimum beam-tilt tableau revealed that 16 tilts, azimuthally distributed on an outer ring at maximum modulus, plus 8 tilts on an inner ring, and 1 untilted exposure provide enough data to sufficiently over-determine the measurement of all aberrations up to the 6$^{\text{th}}$ order. Different beam-tilt distributions only result in minor improvements to the precision of the measurement. Against the common belief that the precision of the aberration measurement improves with larger tilt angle, an optimum value was derived for the maximum tilt modulus of $t_{\text{max,opt}}$ as given by equation (5.4.2). The inner ring-radius should be set to 70 % of the maximum tilt. A tilt tableau with such parameters

still requires that the effective twofold astigmatism can be measured with an error smaller than 1.2 nm from every diffractogram in order to satisfy the multi-aberration tolerance limit $\bar{\sigma}_\chi(g_{\max}) < \pi/4$ for $g_{\max} = 12.5$ nm$^{-1}$.

# Chapter 6

# Novel techniques for the analysis of single diffractograms

The main reason why existing realisations of aberration-measurement techniques with the diffractogram method cannot fulfil the tolerance limit $\bar{\sigma}_\chi(g_{\max}) \leq \pi/4$ for Sub-Ångström resolution lies in their insufficient precision of the diffractogram analysis. As discussed in the preceding section, a measurement error of $\sigma_{22} < 1.2$ nm must be attained for the effective twofold astigmatism in order to satisfy the above criterion. The error of the diffractogram analysis achieved by the existing techniques is larger by more than a factor of two. It was therefore a main challenge of this work to develop new techniques for the extraction of defocus and twofold astigmatism from a single diffractogram with a precision of less than one nanometre.

## 6.1 Concept of the diffractogram analysis

In order to depict the difficulty of the imposed task, figure 6.1 shows some examples for diffractogram patterns which can be encountered in practice with different instruments, different materials, and with a large range of aberration values.
It is a matter of robustness of the diffractogram analysis that such a large variety of patterns and apparently different signal quality can be treated successfully while achieving sub-nanometre precision. The basic concept of a new technique of a numerical diffractogram analysis is depicted in the present section. Besides the main directives to achieve the best possible accuracy and a robust work flow, the numerical procedures are developed also for high-speed processing of the experimental data. It is clear that the precision of the diffractogram analysis is best when the full amount of information about the defocus

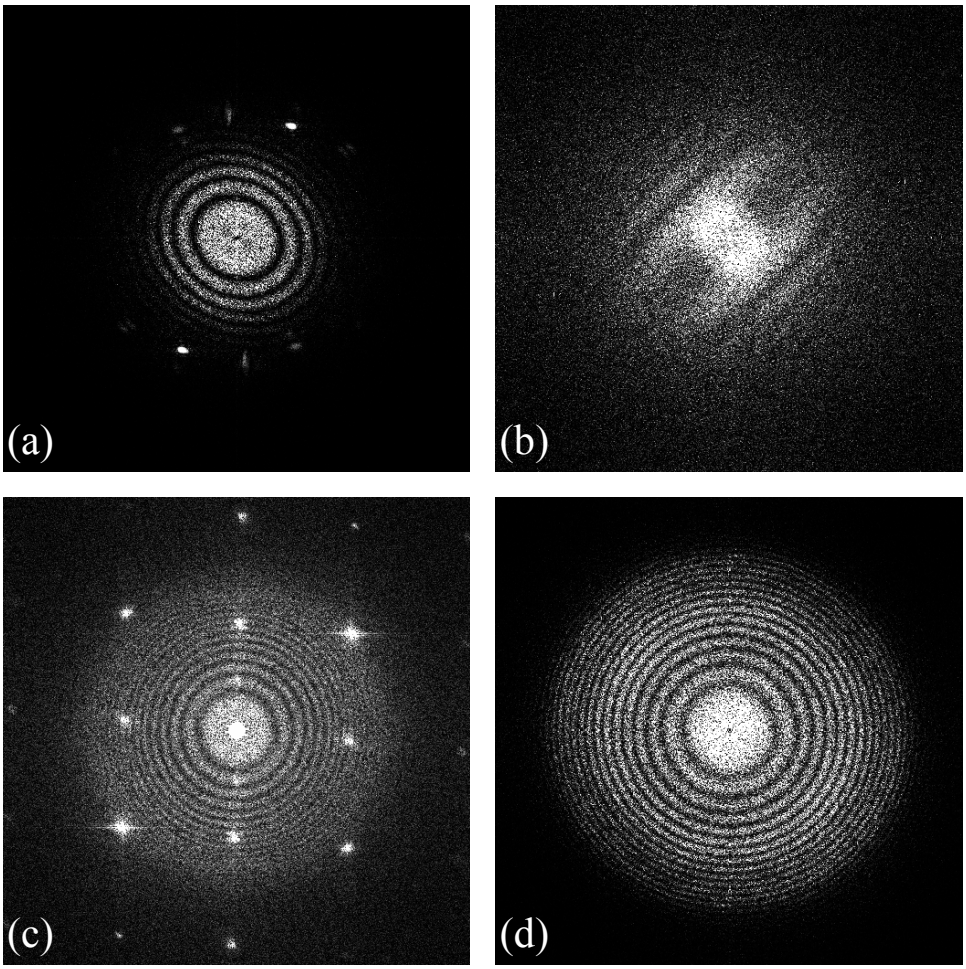



Figure 6.1: Examples of numerical diffractograms obtained with different instruments and a variety of defocus- and twofold-astigmatism values. Diffractograms of different objects are shown: (a) gold clusters on amorphous carbon, (b) amorphous carbon, (c) partially amorphous $SrTiO_3$, and (d) amorphous tantalum. (a) and (b) are recorded with microscopes having a large spherical aberration, (c) and (d) are recorded with a spherical-aberration corrected TEM.

and the twofold astigmatism contained in the diffractogram is used. The techniques developed in this work therefore consider the entirety of diffractogram pixels.

According to the contrast transfer theory discussed in section 3.2, the defocus $c_{20}$ and the twofold astigmatism $c_{22}$ influence the sinusoidal modulation of the diffractogram intensity due to the even part of the aberration function with $\sin^2[\chi^{\mathrm{e}}(\vec{g})]$. This sinusoidal pattern is directly observable with amorphous objects as in the examples in figure 6.1. Therefore, the signal of the two second-order aberrations $c_{20}$ and $c_{22}$ must be isolated from the signal of the other even aberrations contained in the diffractogram pattern.

A partial isolation of the signal of the two second-order aberrations is already achieved in a natural way, since the intensity distribution of diffractograms reaches only up to a

certain spatial frequency. In this limited range typically only a large spherical aberration $c_{40}$ additionally affects the observable pattern. The respective sinusoidal pattern $P$ is described by the formula

$$P(g_x, g_y) = \sin^2\left[\pi\lambda\left(c_{20}(g_x^2+g_y^2) + c_{22x}(g_x^2-g_y^2) + 2c_{22y}g_xg_y + \frac{\lambda^2}{2}c_{40}(g_x^2+g_y^2)^2\right)\right] \tag{6.1.1}$$

where the twofold astigmatism $c_{22} = c_{22x} + \mathrm{i}c_{22y}$ is obviously the only aberration which causes an anisotropy of the pattern. Due to the constitution of the square of the sine function, any recognition of the diffractogram pattern is ambiguous with respect to the sign of the even part of the aberration function $\chi^{\mathrm{e}}(\vec{g}\,)$. In section 7.5.1 a new method is described which allows to resolve this ambiguity of the diffractogram method.

For a precise analysis of the diffractogram pattern $P(\vec{g}\,)$, the spherical aberration must be considered. In practice, spherical-aberration corrected microscopes are typically operated with $|c_{40}| < 100$ $\mu$m for HRTEM with a resolution around one Ångström [50]. For microscopes without $C_{\mathrm{S}}$-correction, the spherical aberration is a fixed and known parameter of the objective lens.

The following robust two-step procedure has proven to yield on the one hand very precise values for the effective defocus and the effective twofold astigmatism from one diffractogram, and, on the other hand, to include the spherical aberration. In a first step, rough values for defocus and twofold astigmatism are automatically obtained on the basis of a user-specified *first-guess* value of the spherical aberration $c_{40}$ which should lie in the range of $\pm 200$ $\mu$m around the actual value. From the resulting twofold-astigmatism values, a better estimate for the spherical aberration is determined which is then used in a second step for an ultra-precise recognition of the diffractogram pattern.

The following list summarises additional features of diffractograms which can adversely influence a recognition of the pattern $P(\vec{g}\,)$ :

- Superposition with strong noise which is often of higher intensity than the signal,
- Superposition with diffractogram peaks in the presence of crystalline material,
- Superposition with a background due to
  - Single-sideband effects due to tilted-beam illumination (see section 3.2),
  - Features of electron diffraction unaccounted for by the weak-phase-object approximation, such as absorption contrast,
  - Shape, thickness and tilt of the object,

- Damping by envelope functions related to
  - Partial temporal and partial spatial coherence,
  - Electron scattering factors,
  - Vibrations and drift of the object,
  - Modulation transfer function of the detector,
  - Shape, thickness and tilt of the object.

Most of the above listed reasons for the superposition and modulation of the sinusoidal pattern can in principle be treated by physical models. However, some of these models are approximative and some of the required parameters, like for example the shape and the thickness of the object are not known in practice. In the view of an automated processing and a universal applicability of the whole aberration-measurement procedure, the analysis of single diffractograms should also not require a specification of the chemical composition of the object. Therefore, a very general phenomenological approach is used to describe the intensity distribution in diffractograms. As introduced in section 3.2.5, the intensity distribution of a diffractogram is denoted by $D(\vec{g})$, and expressed here by phenomenological terms, with

$$D(\vec{g}) = E(\vec{g}) \cdot P(\vec{g}) + N(\vec{g}) + B(\vec{g}) + R(\vec{g}). \tag{6.1.2}$$

This model accounts for a damping of the pattern $P(\vec{g})$ of equation (6.1.1) by an envelope $E(\vec{g})$. The damped pattern is superposed with three entities describing random noise $N(\vec{g})$, a slowly varying background $B(\vec{g})$, and diffractogram peaks $R(\vec{g})$. Envelope $E$, background $B$, and the diffractogram peaks $R$, are expected to be in general anisotropic. The diffractogram analysis developed in this work proceeds in the following steps:

1. Detection and removal of diffractogram peaks $R(\vec{g})$
2. Removal of additive background $B(\vec{g})$
3. Determination of multiplicative envelope $E(\vec{g})$
4. Reduction of random noise $N(\vec{g})$
5. Azimuthal orientation of the anisotropic pattern $P(\vec{g})$
6. Coarse extraction of defocus and twofold astigmatism by pattern recognition of a reduced variant of the pattern $P(\vec{g})$

7. Ultra-precise measurement of defocus and twofold astigmatism by pattern recognition of the complete pattern $P(\vec{g})$

These steps follow the principal strategy that the signal of the twofold astigmatism can be accessed more easily, if the disturbing features are known, or removed as accurately as possible from the data in advance. The analysis of the isolated pattern $P(\vec{g})$ is then less misguided by the aforementioned systematic diffractogram features. Most of these steps require the application of complicated image-processing procedures which work properly within certain limits. These procedures are described in the following sections.

## 6.2 Detection and removal of diffractogram peaks

It is actually easy for the operator, even for non-experts, to distinguish intense peaks in the diffractogram due to crystalline material from the weak oscillating pattern due to amorphous material in the diffractogram as shown in the lower left-side example of figure 6.1. However, this kind of pattern recognition poses a serious challenge to numerical procedures which often succeed only in an incomplete recognition of the respective features. Due to the high intensity of diffractogram peaks, any analysis aiming on the weaker pattern $P(\vec{g})$, be it cross correlations or least-squares fitting, will get influenced systematically or might even fail to produce results. It is therefore mandatory to exclude the respective regions of the diffractogram for the determination of defocus and twofold astigmatism.

Diffractogram peaks due to crystalline material exhibit on the one hand a much higher intensity compared to their local surrounding, and on the other hand a characteristic localisation in small regions of the Fourier space in the form of pixel clusters. The cluster-recognition procedure described in this section identifies diffractogram peaks on the basis of these two properties.

The cluster-recognition procedure starts by pre-selecting a reasonable fraction $p_{\mathrm{CR}}$ of diffractogram pixels with the highest intensity as candidate pixels. Since size and quantity of diffractogram peaks vary with the crystal structure, with the orientation, and with the shape of the imaged crystalline object, the value of $p_{\mathrm{CR}}$ can be adapted also manually by the user. $p_{\mathrm{CR}}$ should be set to values between 2 % and 10 %, which is typically more than necessary. In the latter case, also pixels belonging to the most intense parts of the damped sinusoidal pattern $E(\vec{g}) \cdot P(\vec{g})$ in equation (6.1.2) will unintentionally be selected as candidates for the cluster recognition as shown in figure 6.2b.

In order to keep the parts of the diffractogram which contain only the pattern $P(\vec{g})$ as

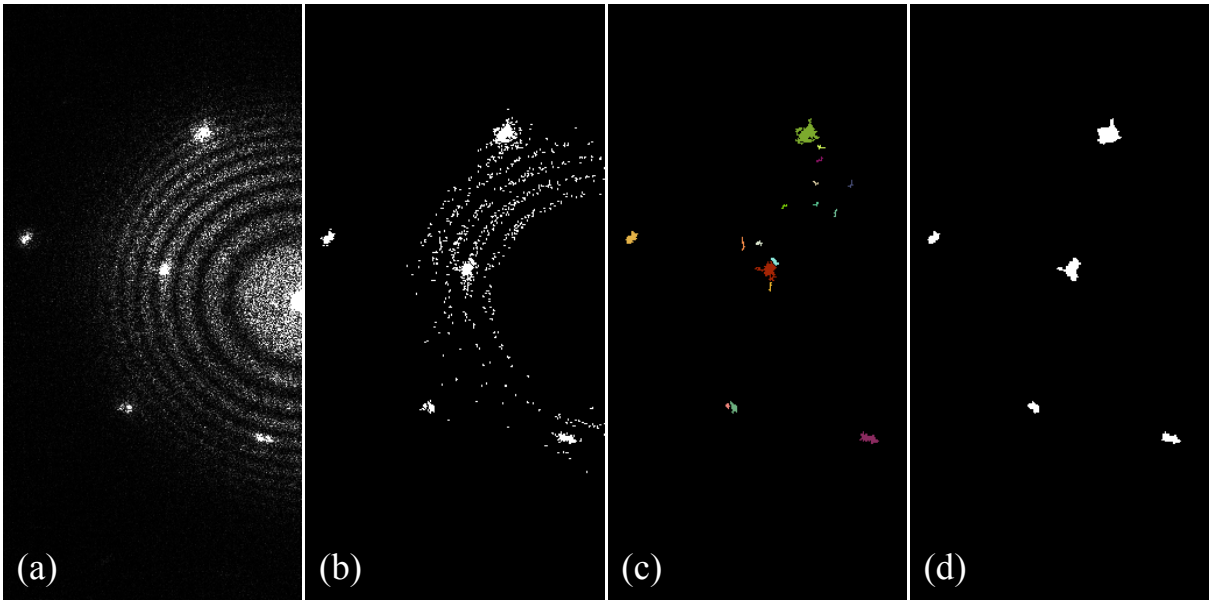



Figure 6.2: Demonstration of detection and classification of diffractogram peaks: The images show (a) a diffractogram-half obtained with an object containing crystalline and amorphous $SrTiO_3$, (b) detected candidate pixels, (c) detected larger pixel clusters, and (d) the final mask which identifies invalid areas for the pattern recognition.

unharmed as possible, a classification routine is applied to remove falsely detected candidates from the initial mask. Connected candidate pixels are assigned to clusters as indicated by the individually coloured areas in figure 6.2c. Every cluster is rated by the figure-of-merit $m_{\mathrm{CR}}$, explained below. A larger value of $m_{\mathrm{CR}}$ means that the properties of the respective cluster more likely correspond to the aforementioned characteristic properties of diffractogram peaks. The value of $m_{\mathrm{CR}}$ is calculated for every cluster as a product of the mean diffraction angle $\bar{g}$, the Fourier-space area $A$, the total intensity $D$, and the ratio $F = I_{2,\mathrm{min}}/I_{2,\mathrm{max}}$ of the smallest to largest moment of inertia of the intensity distribution, with

$$m_{\mathrm{CR}} = \bar{g}\, A\, D\, F. \tag{6.2.1}$$

Hereby, the moment of inertia $I_2$ is calculated by treating the intensity carried by a pixel analogous to a mass. Largest moment of inertia $I_{2,\mathrm{max}}$ and smallest moment of inertia $I_{2,\mathrm{min}}$ are selected from the entirety of all possible rotation axes in the diffraction plane, through the centre of gravity of a cluster. By this classification larger values of $m_{\mathrm{CR}}$ are assigned to larger clusters with higher intensity, compact shapes and positions in the diffractogram which correspond to larger diffraction angles.

From the complete list of all detected clusters, only a user-specified number of $N_{\mathrm{CR}}$ clusters

rated with the largest values of $m_{\mathrm{CR}}$ are collected in a final mask image as shown in figure 6.2d. This mask image identifies all pixels of a diffractogram which are considered as members of diffractogram peaks in the further analysis. A subsequent treatment of the additive background and multiplicative envelope of the diffractogram requires an adequate replacement of the masked peak pixels. Therefore, random data of the local surrounding is copied into the masked areas.

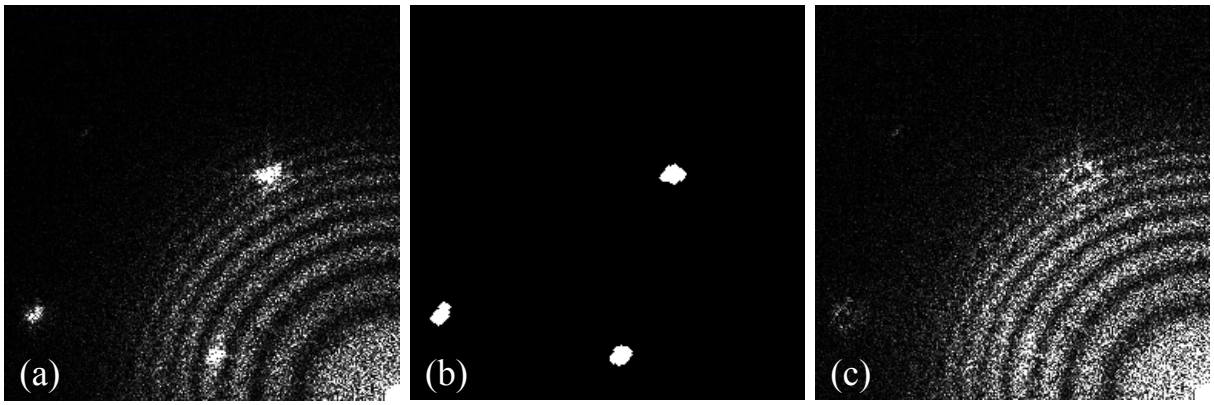



Figure 6.3: Demonstration of a detection, classification and removal of peaks from a diffractogram of amorphous and crystalline $SrTiO_3$. The images show a selected part of (a) the initial diffractogram, (b) the mask image with detected and classified clusters, and (c) the intensity distribution of the diffractogram after replacement of the masked pixels with random data from the surrounding.

As illustrated in figure 6.3, the processed intensity distribution is almost free of strong peaks. Untreated vestiges at the borders of the masked areas are small and of comparable intensity to the sinusoidal pattern. These remainders can thus be considered as harmless perturbations for the recognition of the diffractogram pattern.
The cluster-recognition procedure as described above works perfectly if the area covered by peaks is smaller than approximately 10 % of the total diffractogram area, which is the case in most routine applications. Multiple orientations like for example in nano-crystalline materials impose a serious problem to the diffractogram analysis. In such a case, the diffractogram peaks are weak and numerous, and therefore difficult to detect. They are also aligned on Debye-Scherrer rings which can be easily misinterpreted as Thon-rings of the pattern $P(\vec{g})$.

## 6.3 Removal of the diffractogram background

The background $B(\vec{g})$ typically varies considerably slower than the pattern $P(\vec{g})$ over the whole area of the diffractogram. It is assumed in the following that the frequency spectra of $B(\vec{g})$ and $P(\vec{g})$ do not overlap in the Fourier transform of the diffractogram. It is interesting to note that the Fourier transform of the diffractogram intensity $D(\vec{g})$ is the autocorrelation function of the initial real-space input image $I(\vec{r})$. Although using the terminus "frequency," it should be emphasised that functions derived from a Fourier transform of the diffractogram are real-space quantities and depend on $\vec{r}$. The assumption of non-overlapping frequency spectra holds, if the intentionally applied defocus is large enough to generate at least more than one oscillation in the undamped part of the pattern $P(\vec{g})$. In this case, the extent of the central peak in the Fourier transform of the diffractogram as shown in figure 6.4a can be used to define an adaptive cut-off radius $r_0$ for an adaptive low-pass filter which allows for an extraction of the low-frequency content $D_{\mathrm{LF}}(\vec{g})$ of the diffractogram which includes the complete background $B(\vec{g})$.

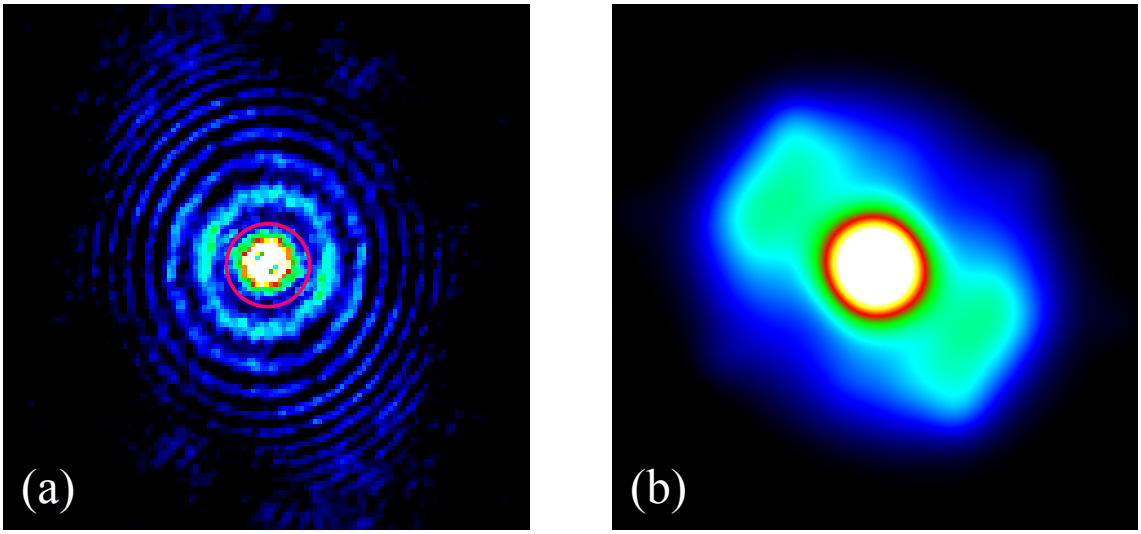



Figure 6.4: (a) Frequency spectrum of a numerical diffractogram as obtained by a discrete Fourier transformation. It is the autocorrelation function of the initial image intensity $I(\vec{r})$. The concentric rings of the autocorrelation function represent the frequency spectrum of the Thon-ring pattern, the central peak (encircled in purple) represents the frequency spectrum of the background $B(\vec{g})$ together with all other slowly varying additive features of the diffractogram intensity. (b) Filtered low-frequency content $D_{\mathrm{LF}}(\vec{g})$ after back-transformation to the diffractogram plane obtained from the Fourier coefficients inside the encircled area in (a).

The actual filter procedure applies a Fast Fourier Transform (FFT) to the diffractogram intensity and multiplies the real-space filter function $f_0(\vec{r}) = \exp[-|\vec{r}|^2/r_0^2]$ to the au-

tocorrelation function. After backwards transformation by an inverse FFT one obtains the additive low-frequency content of the diffractogram as shown in figure 6.4b. This operation is abbreviated by the expression

$$D_{\mathrm{LF}}(\vec{g}\,) = \mathrm{FFT}^{-1}\left[f_0(\vec{r}\,) \cdot \mathrm{FFT}\left[D(\vec{g}\,)\right]\right]. \tag{6.3.1}$$

The complementary high-frequency content $D_{\mathrm{HF}}(\vec{g}\,)$ is readily obtained by calculation of

$$D_{\mathrm{HF}}(\vec{g}\,) = D(\vec{g}\,) - D_{\mathrm{LF}}(\vec{g}\,). \tag{6.3.2}$$

Thereby, the background $B(\vec{g}\,)$ is removed from the initial diffractogram $D(\vec{g}\,)$. Additional to the background, $D_{\mathrm{LF}}(\vec{g}\,)$ also contains the mean value $1/2$ of the $\sin^2[\chi^{\mathrm{e}}]$-pattern multiplied by an envelope function $E(\vec{g}\,)$. In terms of the phenomenological diffractogram model of equation (6.1.2), one obtains an analytical expression for the low-frequency content, with

$$D_{\mathrm{LF}}(\vec{g}\,) = \frac{1}{2}\,E(\vec{g}\,) + N_{\mathrm{LF}}(\vec{g}\,) + B(\vec{g}\,), \tag{6.3.3}$$

where $N_{\mathrm{LF}}(\vec{g}\,)$ denotes the low-frequency components of the random noise $N(\vec{g}\,)$, and the diffractogram peaks $R(\vec{g}\,)$ are omitted. The remaining high-frequency content is thus given by

$$D_{\mathrm{HF}}(\vec{g}\,) = E(\vec{g}\,)\left(\sin^2[\chi^{\mathrm{e}}(\vec{g}\,)] - \frac{1}{2}\right) + N_{\mathrm{HF}}(\vec{g}\,), \tag{6.3.4}$$

where $N_{\mathrm{HF}}(\vec{g}\,)$ denotes the high-frequency components of the random noise $N(\vec{g}\,)$.

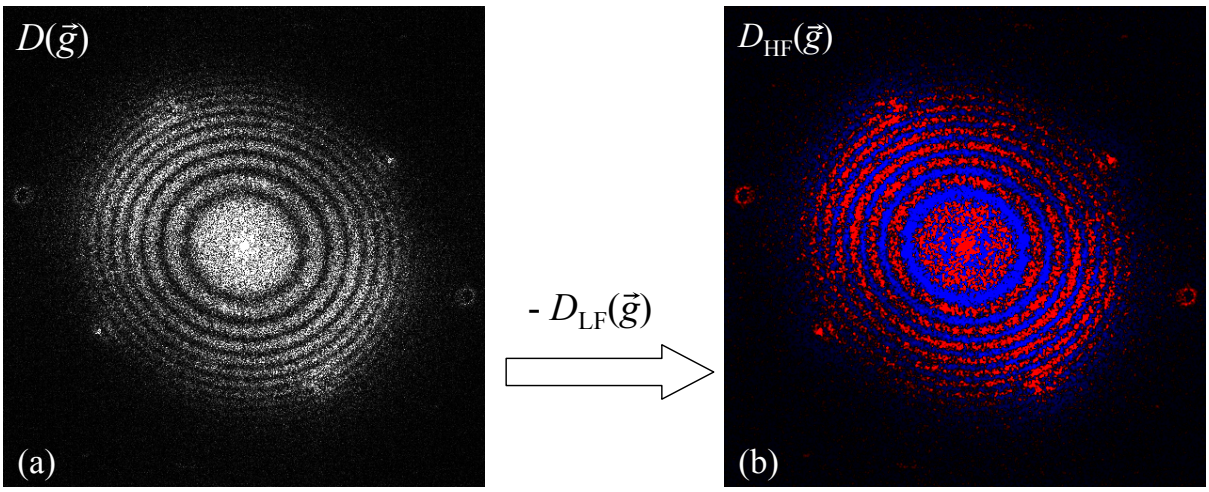



Figure 6.5: (a) Diffractogram of semi-amorphous $SrTiO_3$ after removal of the diffractogram peaks. Subtraction of the low-frequency content yields the remaining high-frequency content as shown in image (b). The high-frequency content varies around zero, where blue colours mean negative values, red colours mean positive values, and the zero level is depicted in black.

The resulting data obtained by removing the low-frequency content from an experimental diffractogram is depicted in figure 6.5. Since the mean value of the Thon-ring pattern is removed also, the remaining data varies around zero with decreasing amplitude for higher spatial frequencies.

## 6.4 Determination of the diffractogram envelope

After the background has been removed together with all low-frequency content of the diffractogram, the remaining high-frequency content as shown in figure 6.5b contains the sinusoidal pattern due to the even aberrations which is damped by an envelope $E(\vec{g})$, and high-frequency noise according to equation (6.3.4). The envelope is not isotropic when the object is illuminated by a strongly tilted electron beam as discussed in section 3.2. The described twofold anisotropy of the envelope must not necessarily coincide with the twofold anisotropy of the pattern $P(\vec{g})$ due to the effect of the twofold astigmatism. The modulation of the pattern by the envelope can thus mislead the recognition of the diffractogram pattern.

A similar adaptive low-pass filter technique as applied for background removal in the previous section is used to determine the envelope $E(\vec{g})$ which also varies slowly compared to $P(\vec{g})$. However, the product $E(\vec{g}) \cdot P(\vec{g})$ must be transformed into a summation in order to utilise a low-pass filter concept which can separate purely additive components. This transformation is achieved by taking the logarithm of the absolute value of the high-frequency content $D_{\mathrm{HF}}(\vec{g})$ in equation (6.3.4). In terms of the phenomenological model (6.1.2), one obtains the following expression for the *logarithmic high-frequency content* of the diffractogram:

$$L(\vec{g}) = \ln|D_{\mathrm{HF}}(\vec{g}) + \varepsilon|, \tag{6.4.1}$$

$$\approx \ln[E(\vec{g})] + \ln\left|\sin^2[\chi^{\mathrm{e}}(\vec{g})] - \frac{1}{2} + \frac{N_{\mathrm{HF}}(\vec{g})}{E(\vec{g}) + \varepsilon'}\right|, \tag{6.4.2}$$

where the "small" quantities $\varepsilon > 0$ and $\varepsilon' > 0$ are introduced to avoid singularities. It is again assumed that the frequency spectra of $\ln[E(\vec{g})]$ and of the second logarithmic term in equation (6.4.2) do not overlap and can thus be separated by applying a frequency filter $f_1(\vec{r})$.

The extent of the central peak in the frequency spectrum of $L(\vec{g})$ determines an adaptive cut-off radius $r_1$ of the filter function $f_1(\vec{r}) = \exp[-|\vec{r}|^2/r_1^2]$. Similar to the determination of the additive low-frequency content in the previous section, also the multiplicative low-frequency content is determined by an adaptive frequency-filter technique. This adaptive

filter technique is applied here in the discrete Fourier transform of the logarithm of the diffractogram:

$$L_{\mathrm{LF}}(\vec{g}\,) = \mathrm{FFT}^{-1}\left[f_1(\vec{r}\,) \cdot \mathrm{FFT}\left[L(\vec{g}\,)\right]\right]. \tag{6.4.3}$$

The multiplicative envelope can thus be determined by subsequent calculation of the exponential function

$$\exp[\,L_{\mathrm{LF}}(\vec{g}\,)\,] \approx \exp\{\,\ln[E(\vec{g}\,)]\} = E(\vec{g}\,). \tag{6.4.4}$$

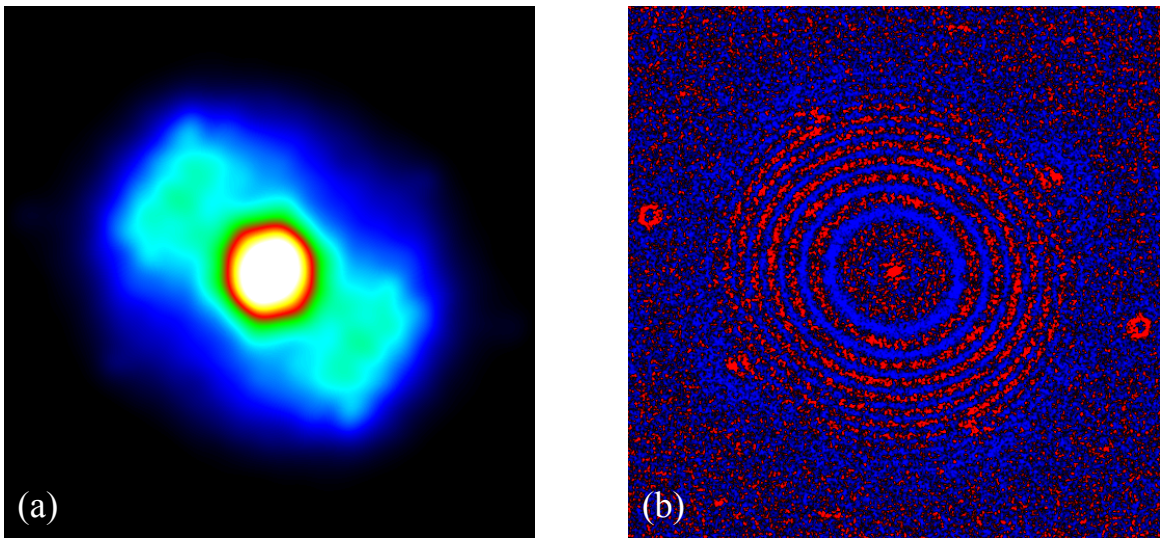



Figure 6.6: Continuation of the diffractogram analysis from figure 6.5b yields (a) the envelope $E(\vec{g}\,)$ and (b) the normalised oscillating pattern after division of the high-frequency content $D_{\mathrm{HF}}(\vec{g}\,)$ as shown in figure 6.5b by the numerically determined envelope $E(\vec{g}\,)$.

The adaptive filter procedure described above is applied to the high-frequency content $D_{\mathrm{HF}}(\vec{g}\,)$ shown in figure 6.5b and yields the envelope data $E(\vec{g}\,)$ as shown in figure 6.6a. After division of $D_{\mathrm{HF}}(\vec{g}\,)$ by the determined envelope data $E(\vec{g}\,)$, the damping at higher spatial frequencies is removed. This renormalised diffractogram content shown in figure 6.6b reveals previously damped and hardly visible parts of the Thon-ring pattern. Thon-ring pattern, noise, and remainders of diffractogram peaks in figure 6.6b are of approximately similar strength. However, both low-frequency extractions, for the additive content and the multiplicative content, are malfunctioning when intense parts of diffractogram peaks could not be successfully replaced in advance.

At spatial frequencies of typically $g < 1$ nm$^{-1}$ the diffractogram intensity exhibits high values in contradiction to the linear contrast-transfer theory discussed in section 1.3.4

which predicts a transfer gap at low spatial frequencies. The preprocessed diffractogram pattern after subtraction of the low-frequency content and after division by the envelope $E(\vec{g})$ is thus described by the following model only for $g > 1$ nm$^{-1}$

$$\frac{D_{\mathrm{HF}}(\vec{g})}{E(\vec{g})+\varepsilon'} \approx \left(\sin^2[\chi^{\mathrm{e}}(\vec{g})]-\frac{1}{2}\right)+\frac{N_{\mathrm{HF}}(\vec{g})}{E(\vec{g})+\varepsilon'}, \tag{6.4.5}$$

$$= \left(P(\vec{g})-\frac{1}{2}\right)+N'(\vec{g}), \tag{6.4.6}$$

where $N'(\vec{g})$ denotes remaining, normalised, high-frequency random noise. The inner area of the diffractogram must be excluded from the analysis for the defocus and the twofold astigmatism.

## 6.5 Azimuthal orientation of diffractogram patterns

The preparation steps of the diffractogram analysis described so far, provide two-dimensional data arrays which contain the pattern $P(\vec{g})$ almost free from unknown variations from the ideal sinusoidal shape except for noise. With the next steps of the diffractogram analysis, the shape of the preprocessed experimental data is investigated in order to obtain the three parameters for defocus and twofold astigmatism, denoted here in polar notation $c_{20}$, $c_{22}$, and $\varphi_{22}$. The azimuth parameter $\varphi_{22}$ is determined independently from the others by the procedure described in this section.

In the recorded image, $\varphi_{22}$ determines the azimuthal direction $\varphi$, where the astigmatism modulus $c_{22}$ adds positively to the isotropic defocus $c_{20}$. This is expressed by the term $\cos[2(\varphi-\varphi_{22})]$ in the aberration function, denoted here in a simplified version comprising only these two aberrations

$$\chi(g,\varphi) = \pi\lambda\left\{c_{20}g^2 + c_{22}g^2\cos[2(\varphi-\varphi_{22})]\right\}. \tag{6.5.1}$$

Along one azimuthal direction, defocus and astigmatism add with the same sign and the pattern $P(\vec{g})$ oscillates with higher frequencies than along the perpendicular direction, where defocus and astigmatism have opposite signs. The sinusoidal pattern exhibits thus two perpendicular mirror axes.

The azimuthal astigmatism parameter is determined by measuring the orientation of these principal axes with respect to the $x$-axis of the image. Rotational cross correlation between the pattern $P(\vec{g})$ and its copy $P'(\vec{g})$, which is mirrored with respect to the $x$-axis, allows for a straightforward determination of the azimuthal orientation of the principal axes, as reported by Saxton [95].

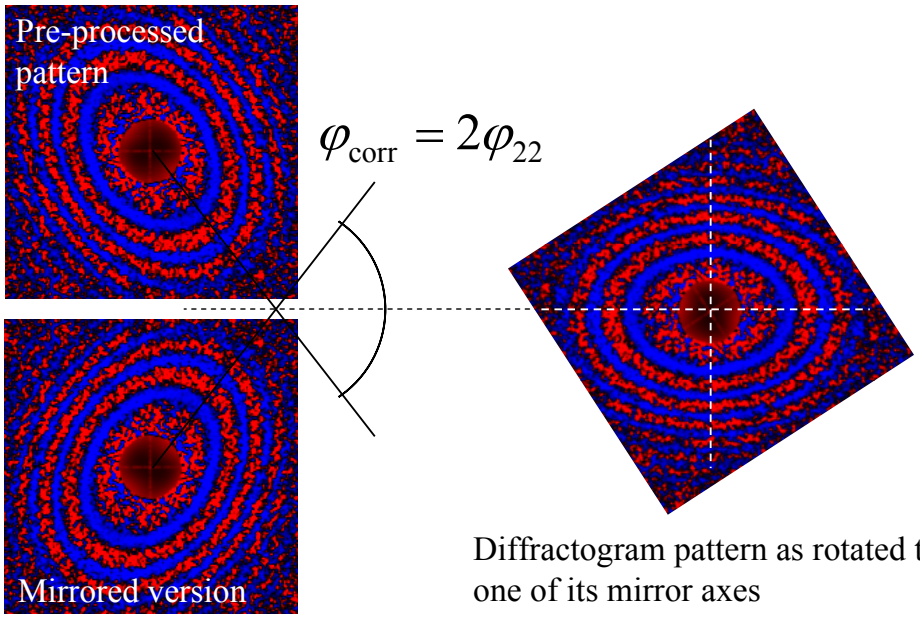



Figure 6.7: Determination of the azimuthal parameter $\varphi_{22}$ of the twofold astigmatism by rotational cross correlation between the diffractogram pattern and a second version of the same pattern which is mirrored with respect to the $x$-axis of the image.

Figure 6.7 illustrates the principle of the diffractogram orientation for the case of an astigmatic Thon-ring pattern. The correlation is maximum for the relative rotation angle $\varphi_{corr}$, which corresponds to two times the azimuthal astigmatism parameter $\varphi_{22}$.

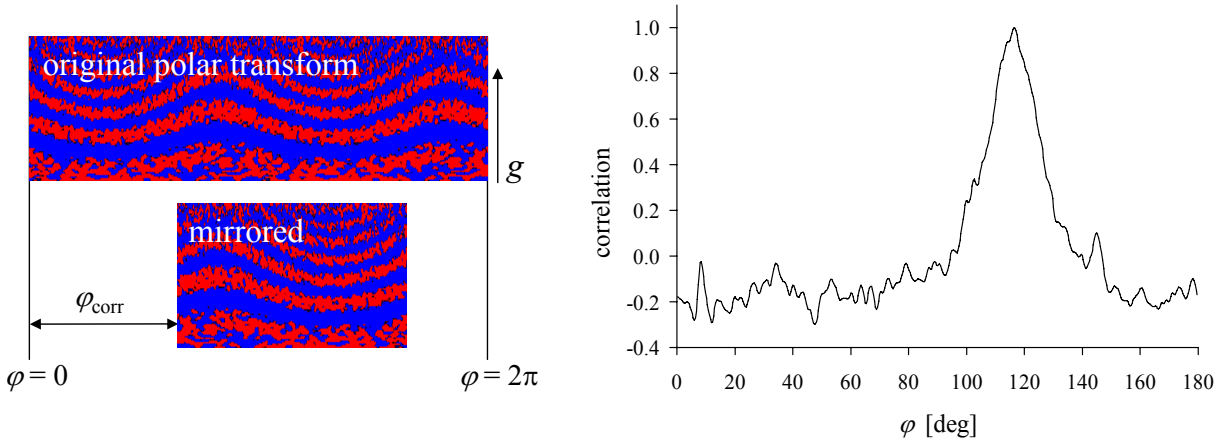



Figure 6.8: Illustration of the rotational cross correlation with polar-transformed patterns. The polar azimuth $\varphi$ varies equidistantly along the horizontal axis of the images, whereas the spatial frequency $g$ increases equidistantly from bottom to top. The rotational cross correlation is consequently represented by a correlation of the patterns with varying lateral displacement. The diagram on the right side depicts the resulting cross-correlation function $c(\varphi)$ obtained for the patterns on the left. The angle $\varphi_{corr}$ denotes the relative displacement between the patterns where the $c(\varphi)$ is maximum.

The performance of numerical rotational cross correlations simplifies drastically in polar coordinates as illustrated in figure 6.8, where a rotation in the original coordinates corresponds to a translation in the azimuthal direction. The cross-correlation function $c(\varphi)$ is calculated using the convolution theorem (A.2.6). The Fourier coefficients of the two polar-transformed versions $P(g,\varphi)$, $P'(g,\varphi)$ of the pattern are multiplied and subsequently backwards-transformed to the diffraction plane, followed by a summation over all relevant spatial frequencies $g$ :

$$c(\varphi) = \sum_g \mathrm{FFT}^{-1}\left[(\mathrm{FFT}[P(g,\varphi)])^* \cdot (\mathrm{FFT}[P'(g,\varphi)])\right]. \tag{6.5.2}$$

In order to ensure a correct orientation also for very small astigmatism values, where the pattern is almost perfectly round, the radial sampling of the polar transformation should be at least as fine as the original Cartesian sampling rate.

Owing to the existence of two mirror axes, the value found for the azimuthal parameter $\varphi_{22} = 1/2 \cdot \varphi_{\mathrm{corr}}$ is actually not unique. A rotation of the original diffractogram by $-\varphi_{22}$ turns the mirror axis showing the fastest oscillations *or* the mirror axis showing the slowest oscillations parallel to the $x$-axis of the image. Only a subsequent pattern recognition can detect which of these two cases is true.

The azimuthal orientation of the pattern and thus the determination of $\varphi_{22}$ tends to fail in the presence of diffractogram peaks, since the rotational cross correlation will lock in to these intense features of the diffractogram. This problem has been frequently observed in experiment with the manufacturer-supplied software and becomes most serious for a small astigmatism modulus, where the diffractogram pattern shows almost perfectly round Thon-rings. Unfortunately, almost round diffractograms appear regularly with spherical-aberration corrected microscopes.

Apart from the determination of the astigmatism sign, there is a more general problem to determine the sign of the full aberration function, comprising all aberrations. This general ambiguity of the diffractogram analysis is frequently avoided by advising the operator to apply a strong underfocus during the acquisition of the images and assuming, consequently, a negative defocus for the analysis of the obtained diffractogram pattern. But especially because over- and underfocussed diffractograms look the same for $C_{\mathrm{S}}$-corrected microscopes, and the operator has to control a multitude of other instrumental parameters during high-resolution experiments, the underfocus demand is overlooked easily. Section 7.5.1 describes a method which, for the first time, allows to determine the sign of the defocus automatically during the aberration measurement and to correct the result of the diffractogram analysis accordingly.

## 6.6 Coarse pattern recognition

As discussed in section 6.1, the diffractogram analysis developed in this work follows a stepwise strategy yielding the defocus and astigmatism parameters in two steps. The first of these steps determines rough estimates of the aberration coefficients which are used as initial parameters for a fine-tuning procedure in the second step. The pattern-recognition algorithm described in this section extracts defocus and twofold astigmatism from the azimuthally oriented, experimental diffractogram pattern by means of comparisons to prior calculated patterns stored in a database.

### 6.6.1 Concept of the coarse pattern recognition

The *coarse pattern recognition* (CPR) compares the experimental data to a set of prepared patterns representing different combinations of defocus and astigmatism over a large parameter range. A live calculation of such trial patterns would cost too much operating time, since the diffractogram patterns depend on the sine square of the aberration function. The calculation of such trial patterns lasts approximately 100 times longer than the time required for the comparison to the experimental pattern.
The target accuracy in defocus and twofold astigmatism of the diffractogram analysis is 1 nm. This is approximately 3 orders of magnitude smaller than the expected range of the defocus and the twofold astigmatism of approximately 1 $\mu$m, which is relevant in practice. This ratio of 1000 : 1 requires a sampling of the two-dimensional defocus-astigmatism parameter-space in up to 1 million points, 1000 for each dimension. It is therefore mandatory to reduce the data for the comparison as much as possible.
The data array obtained from the diffractogram after treatment of diffractogram peaks, removal of the low-frequency components, and azimuthal orientation of the principal axes to the image axes, as shown on the right side of figure 6.7, is used as initial data for the coarse pattern recognition. Four steps are taken to reduce rigorously the size of the data:

1. Omitting one half of the pattern which is identical to the other half due to the Friedel symmetry,

2. Averaging the two remaining quarters according to the mirror symmetry due to the second mirror axis,

3. Scaling the quarter image down to a Nyquist frequency between 3.0 $nm^{-1}$ and 4.0 $nm^{-1}$,

4. Representation of the oscillations in the pattern $P(\vec{g})$ by one bit per pixel, where zero denotes data values with negative and one denotes data values with positive variations from the mean. This last step decreases the prior data size by almost 85 % from 16 bits per pixel to 1 bit per pixel.

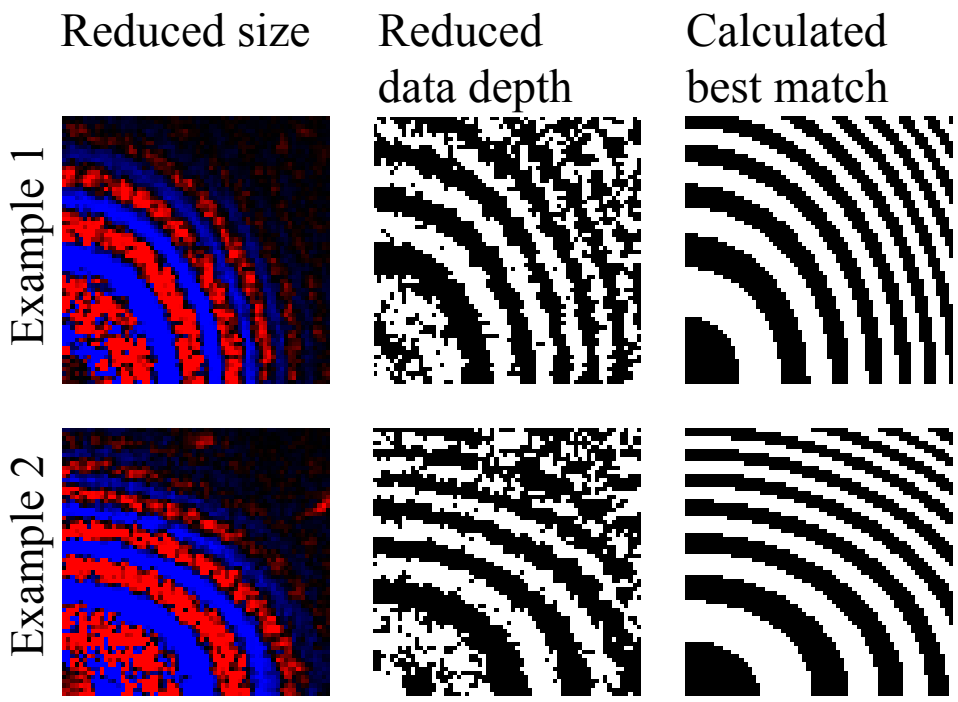



Figure 6.9: Two examples of successful searches for a predicted pattern in a large database which shows minimum difference to the experimental pattern. The images on the left are quarters of the original diffractograms after removal of the low-frequency content and rotation to one of its principal axes. The quarters are scaled to a Nyquist frequency of 3.5 $nm^{-1}$ on a 64×64-pixel array and the data is depicted in red for positive values and in blue for negative values. The colours get darker for values closer to zero. Conversion of these quarter images to 1-bit-data yields the central images, while the right ones show the best fitting pattern from a prepared database.

Figure 6.9 shows the successful identification of the diffractogram pattern from such a quarter image by a corresponding pattern from the database for two examples. The left image in each row represents a quarter of the diffractogram pattern after its rotation to the principal axis and resampling to a 64×64-pixel array with a Nyquist frequency of 3.5 $nm^{-1}$. The quarter images are converted to 1-bit data (images in the middle) and compared to the 1-bit patterns of the prepared database. The best matching database pattern is shown on the right.

The prepared database patterns $M_{db}$ are calculated on a 1-bit-valued 64×64-pixel array according to equation (6.1.1) which yields a sinusoidal oscillation around the mean value

1/2. Pixels, where the sinusoidal oscillation is larger than 1/2 are set to one, whereas for oscillations below 1/2 the respective pixels are set to zero. The sampling of the bit-level database-patterns on 64×64-pixel arrays turned out to be sufficient for a robust, coarse pattern recognition on various microscopes. Since every pixel is represented by only 1 bit in the computer memory, one such bit image occupies 512 bytes in memory. A database comprising a defocus range from –1600 nm up to 0 nm in steps of 4 nm and an astigmatism range from –500 nm up to 500 nm in 2 nm steps occupies $400 \times 500 \times 512\,\text{bytes} \approx 100$ Mb in the computer memory. The choice of a completely negative defocus range and an astigmatism range which is symmetric around zero, accounts for the defocus-sign ambiguity discussed in section 6.5. The ambiguity problem has different consequences for the $C_\text{S}$-corrected and the non-$C_\text{S}$-corrected case:

1. Diffractograms recorded with $|c_{40}| < 100$ $\mu$m are invariant to a change of the sign in the second order aberrations, because the spherical aberration is then too small to take influence on the pattern. The sign of the aberration function cannot be determined by analysing the diffractogram alone and the inclusion of positive as well as negative defoci in the database yields no advantage. If a later analysis reveals that the choice was wrong, the signs of all three parameters, $c_{20}$, $c_{22x}$, and $c_{22y}$, have to be inverted accordingly.

2. Uncorrected microscopes have a strong spherical aberration. Since the spherical aberration affects the observed diffractogram pattern, it is necessary to include the full defocus axis in the pattern database. In return, the defocus sign, and thus the sign of the full aberration function can be determined. An example where the defocus sign was determinable directly from the diffractogram with a spherical aberration of 1.2 mm is shown in figure 6.10.

### 6.6.2 Ultra-fast comparison algorithm

The fastest and most effective way to measure the difference between the database patterns $M_\text{db}(\vec{g}\,)$ and the reduced bit-level variant $M_\text{exp}(\vec{g}\,)$ of the experimental diffractogram pattern is to count the results of a bitwise XOR operation between the corresponding pixels of the data arrays. $x\,\text{XOR}\,y$ yields one, if the bits $x$ and $y$ are different and zero if they are equal. A subsequent AND operation with a mask $M_\text{mask}$ suppresses unnecessary counting of diffractogram regions containing only noise or diffractogram peaks. The mask is therefore set up by assigning the value zero firstly, to pixels which exhibit diffractogram peaks and secondly, to pixels where the experimentally determined envelope $E(\vec{g}\,)$ (section

6.4) falls below a user-defined threshold. All other pixels of the mask are set to the value one.

The core procedure of the coarse pattern recognition (CPR) thus calculates a comparison dataset $M_{\mathrm{comp}}$ by

$$M_{\mathrm{comp}}(\vec{g}\,) = M_{\mathrm{mask}}(\vec{g}\,)\,\mathrm{AND}\,\left[M_{\mathrm{db}}(\vec{g}\,)\,\mathrm{XOR}\,M_{\mathrm{exp}}(\vec{g}\,)\right], \tag{6.6.1}$$

The difference between the two patterns is expressed by the number $N_{\mathrm{diff}}$ which is calculated by summation over all bits of $M_{\mathrm{comp}}$ :

$$N_{\mathrm{diff}} = \sum_{\vec{g}} M_{\mathrm{comp}}(\vec{g}\,). \tag{6.6.2}$$

Smaller values of $N_{\mathrm{diff}}$ indicate a better similarity of the compared patterns. Owing to these basic operations, the CPR performs very fast achieving up to 200000 pattern comparisons per second on a conventional PC. A complete scan of the database for the pattern matching best a single experimental pattern requires approximately 200 milliseconds. Figure 6.10 illustrates the existence of several local minima and one global minimum of $N_{\mathrm{diff}}$ in the negative defocus range.

Another strong minimum appears in the positive defocus range, which is apparently not the global minimum. The diffractogram analysed in the shown case was recorded with a TEM having 1.2 mm spherical aberration. In contrast, for a $C_{\mathrm{S}}$-corrected microscope it is impossible to directly distinguish between over- or underfocus, since both minima have then approximately the same value $N_{\mathrm{diff}}$.

### 6.6.3 Achievable precision

The database pattern showing minimum difference to the experimental pattern identifies the defocus and astigmatism values. The precision of this result is approximately equal to the defocus and astigmatism increment of the pattern database. Since this raster does not necessarily contain the location of the "true" total minimum of the difference quantity $N_{\mathrm{diff}}$, an interpolation on the basis of the comparison results in the local surrounding is used to refine the global minimum position.

Figure 6.11 displays CPR-results obtained by analysing four different diffractograms with different databases including a subsequent interpolation. The databases sample the defocus- and astigmatism plane with an increment of 4.0 nm and 2.0 nm respectively. For each of the two samplings, 100 databases with randomly varying sampling offsets have been generated. The CPR-results obtained with these 200 databases cluster at 4

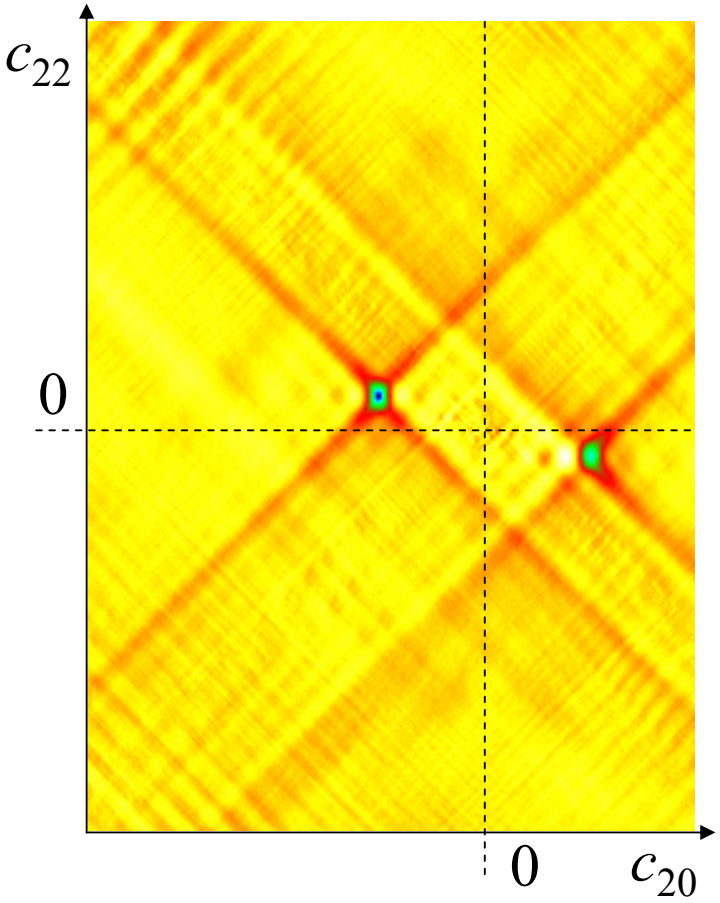



Figure 6.10: Illustration of the differences $N_{\mathrm{diff}}$ between the experimental pattern and all database patterns $M_{\mathrm{db}}$ as calculated by the CPR. $N_{\mathrm{diff}}$ is depicted in a temperature colour-scale ranging from minimum difference in blue to maximum difference in white. The prepared database comprises positive and negative defocus values and a constant spherical aberration of $c_{40} = 1.2$ mm. Due to the large spherical aberration, the defocus sign is determinable directly by the CPR in the shown case.

different positions, where each cluster belongs to one of four diffractograms. The observed deviations from the mean cluster positions is approximately 1 nm although the raster of all databases was larger. This deviation increases with diffractograms where the pattern is only visible in a smaller region of the Fourier space or where the periodicity of the pattern is close the reduced sampling rate.

A finer discretisation of the database does not decrease the deviation. However, different mean cluster positions have been obtained for some diffractograms when using a different sampling rate. One of these cases is shown on the right side of figure 6.11. In practice, a sampling of 4.0 nm is a sufficient compromise between precision and speed, since the size of the database increases quadratically with the number of sampling points in each dimension of the parameter space.

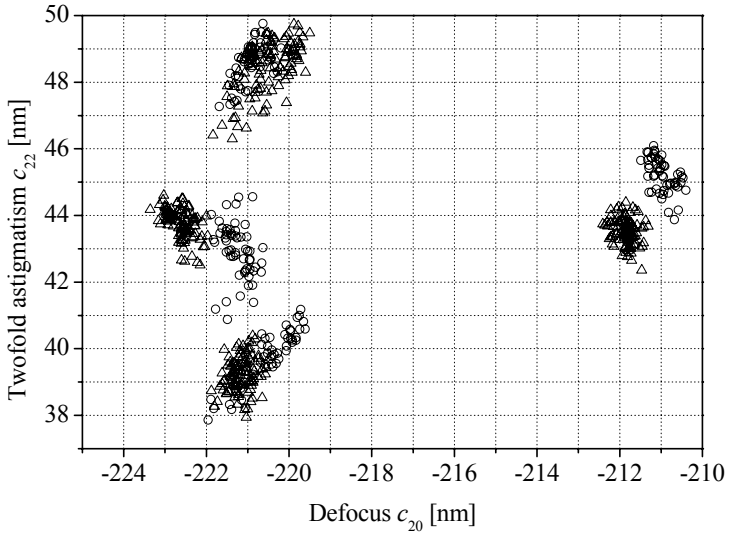



Figure 6.11: Stray diagram of CPR-results obtained by comparing four diffractograms to 200 differently sampled databases. The sampling of the defocus- and astigmatism-range in each database is defined by a constant increment of 4.0 nm (triangles) and 2.0 nm (circles). For each of these increments, 100 databases with randomly varied offsets are generated.

## 6.7 Ultra-precise pattern recognition

The *ultra-precise pattern recognition* (UPR) developed in this work exploits all relevant information about the defocus and the twofold astigmatism contained in one single diffractogram. Relevant in this sense is data at any pixel of the diffractogram pattern remaining after background subtraction, where no diffractogram peak was detected and, where the diffractogram envelope exceeds a user-defined threshold.

### 6.7.1 The principal concept for a full-area pattern recognition

Up to here, the strategy for the analysis of diffractograms was to remove quantitatively unknown, unwanted signal from the evaluated data such as a decreasing electron-scattering factors with increasing diffraction angle, or various damping envelopes, which interfere with the sinusoidal target signal of defocus and twofold astigmatism.
In contrast to this strategy, a different concept for the diffractogram analysis is followed here: A pattern-recognition procedure is used which is independent of the unwanted interfering signal and takes advantage of the complete diffractogram area. This concept is developed to an extent, where the obtained result is largely independent from all interfering signal, except for that of diffractogram peaks.
In many cases, general properties of the quantitatively unknown, unwanted signal interfer-

ences are known, such as symmetry properties. When interfering signal and target signal are independent, then an evaluation of the signal is unproblematic on areas, where the unwanted interferences are constant. Therefore, the diffractogram pattern is partitioned in the following in such a way that the unwanted interferences are extensively constant within each partition.
In the present case, the unwanted interferences comprise functions which vary strongly in the radial direction of the Fourier space but are almost constant along the azimuthal direction. Therefore, a partitioning of the diffractogram along concentric azimuthal paths around the origin of the Fourier space is used for pattern recognition.
Unfortunately, a conceptual singularity occurs when the target signal is of the same symmetry as the interfering signal. In the present case, this singularity occurs for the concentric azimuthal paths when the twofold astigmatism is zero and the target signal exhibits perfectly round Thon-rings. The singularity problem is solved by generation of an adaptive partition scheme.

### 6.7.2 Generation of adaptive paths

In the special singular case, where the target signal and the interfering signal exhibit identical symmetry properties, small adaptive deviations from the circular shape of the paths are introduced. Sinusoidal oscillations $\Delta g^j(\varphi)$ of the path $p^j$ from the mean spatial frequency $\bar{g}^j$ depending on the azimuth angle $\varphi$ are chosen here without restriction of the generality according to the following scheme:

$$\Delta g^j(\varphi) = A^j \cos\left(f_p\,\varphi\right), \tag{6.7.1}$$

where $f_p$ denotes an azimuthal frequency. The amplitude $A^j$ of the sinusoidal deviations from the ideal circular path shape is adaptively chosen according to the following constrictions:

1. The deviation amplitude $A^j$ is as small as possible so that the unwanted interfering signal variations are minimum along the adaptive path.

2. The deviation amplitude $A^j$ is large enough to ensure that the target signal exhibits at least one minimum and one maximum along the adaptive path in order to exploit the full strength of the target signal.

The mean spatial frequencies $\bar{g}^j$ of the individual paths $p^j$ through the Fourier space are distributed densely. Due to this dense radial distribution, the total path set $\{p\}$ covers

the complete relevant area of the diffractogram.
The target signal is expressed analytically in equation (6.1.1) as the sine square of the even part $\chi^{\mathrm{e}}(\vec{g}\,)$ of the aberration function. Maximum signal strength, as required according to the second of the above proposed constrictions, corresponds therefore to a minimum required variation $\Delta\chi^{\mathrm{e}}_{\mathrm{min}}$ of $\chi^{\mathrm{e}}(\vec{g}\,)$ along a given adaptive path $p^j$. Maximum variation of the target signal $\sin^2[\chi^{\mathrm{e}}(\vec{g}\,)]$ is realised with at least $\Delta\chi^{\mathrm{e}}_{\mathrm{min}} = \pi/2$. Along a circle of radius $g^j$, the twofold astigmatism $c_{22}$ changes the aberration function by $\Delta\chi_{22}(g^j) = 2\pi\lambda c_{22}(g^j)^2$. The deviation amplitude $A^j$ for the mean radius $\bar{g}^j$ which is required to realise a change of $\Delta\chi^{\mathrm{e}}_{\mathrm{min}}$ of the even part of the aberration function is thus given by

$$A^j = \begin{cases} 0, & \text{if} \quad \Delta\chi_{22}(\bar{g}^j) \geq \Delta\chi^{\mathrm{e}}_{\mathrm{min}}, \\ \frac{\Delta\chi^{\mathrm{e}}_{\mathrm{min}}}{4\pi\lambda c_{20}\bar{g}^j}, & \text{if} \quad \Delta\chi_{22}(\bar{g}^j) < \Delta\chi^{\mathrm{e}}_{\mathrm{min}}. \end{cases} \tag{6.7.2}$$

Examples for an adaptive path generation according to the above defined scheme is illustrated in figure 6.12. The sinusoidal path deviations are abandoned gradually when the maximum signal variation is already expected for given values of the defocus $c_{20}$, and the twofold astigmatism $c_{22x}$, $c_{22y}$ on a circular path $p^j$ with radius $\bar{g}^j$. For the defocus and astigmatism values considered in figure 6.12b, the paths are circles in the outer diffractogram regions, whereas strong oscillations are required on paths closer to the origin. In practice, the required variation $\Delta\chi^{\mathrm{e}}_{\mathrm{min}}$ of the even part of the aberration function is used with a larger value than the required absolute minimum of $\Delta\chi^{\mathrm{e}}_{\mathrm{min}} = \pi/2$. Thereby, the adaptive path generation ensures sufficient target-signal variations also in cases, where the assumed values $c_{20}$, $c_{22x}$, and $c_{22y}$ do not correspond to the actual aberration values which characterise the target signal.

### 6.7.3 Performance and achievable precision

The pattern recognition is realised by partitionwise correlation of an experimentally obtained pattern with a calculated trial pattern and subsequent cumulation of all partitionwise correlations to one total correlation coefficient $r$, i.e. a correlation over the full area of the diffractogram.
The experimental pattern, which is the target signal, is given by the diffractogram $D(\vec{g}\,)$ after subtraction of the low-frequency content $D_{\mathrm{LF}}(\vec{g}\,)$ as described in section 6.3. Detected diffractogram peaks and the experimentally determined envelope $E(\vec{g}\,)$ (see section 6.4) are not removed from the experimental pattern, but are used to restrict the relevant diffractogram area which is partitioned by adaptive paths and thus contributes to the correlations.

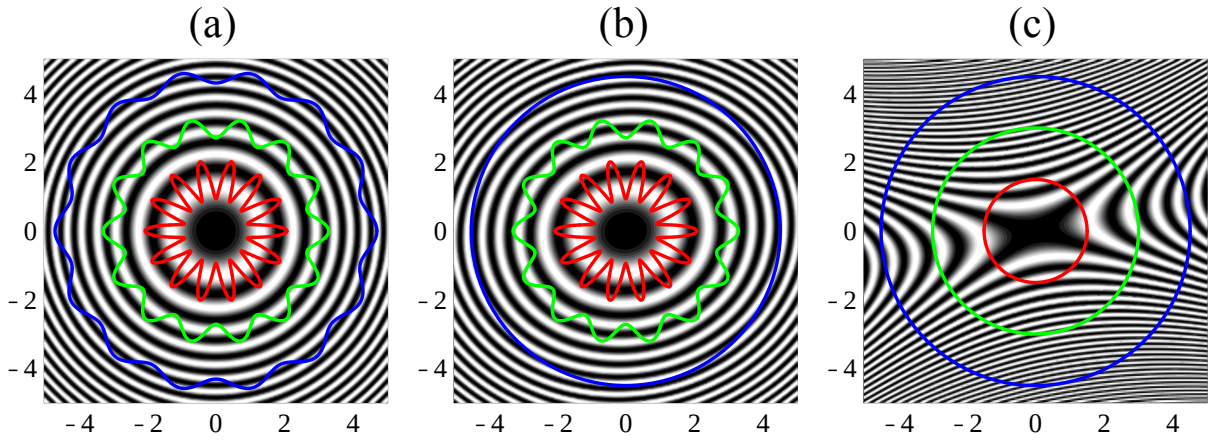



Figure 6.12: A locally adaptive path generation guarantees full signal variation on all paths with minimum radial deviations. At spatial frequencies with smaller eccentricity of the Thon-rings, larger deviations from the circular path shape are required. The paths in the given examples are adapted to comprise a minimum variation of the even part of the aberration function by $\Delta\chi^{\mathrm{e}}_{\mathrm{min}} = \pi$. The three patterns have been calculated with the same defocus $c_{20} = -150$ nm but with different astigmatism values (a) $c_{22} = 1$ nm, (b) $c_{22} = 20$ nm, and (c) $c_{22} = 250$ nm. The plot axes scale from $-5\,\mathrm{nm}^{-1}$ to $5\,\mathrm{nm}^{-1}$ for an electron wavelength of 1.969 pm ($U = 300$ kV).

The trial pattern is calculated according to equation (6.1.1) using trial values for the defocus $c_{20}$ and the twofold astigmatism parameters $c_{20x}$ and $c_{22y}$. The spherical aberration $c_{40}$ is considered by a given fixed value. For correlation with the experimental pattern, the calculated pattern is multiplied by the experimentally obtained envelope $E(\vec{g}\,)$. Thereby, the experimental pattern and the calculated pattern differ ideally only due to wrong trial aberration-values and noise.

The correlation coefficient $r$ is maximum for those values of $c_{20}$, $c_{20x}$, and $c_{22y}$ where the calculated trial pattern matches best to the experimental pattern. The maximum of the total correlation coefficient $r$ is unique over an interval of more than 40 nm for defocus and twofold astigmatism as demonstrated with four examples in figure 6.13, i.e. no local maxima appear in a parameter range of $\pm 20$ nm of defocus and astigmatism. The values of defocus and twofold astigmatism obtained with the coarse pattern recognition (CPR) are thus more than accurate enough for a reliable numerical search for the global correlation maximum.

One correlation over the full set of paths (i.e. the full area of the diffractogram) requires up to 2 milliseconds operation time on a conventional PC. It is therefore advantageous that the CPR-results are already close to the global maximum of the correlation procedure described in this section. Due to the good start values from the CPR, a Simplex Algorithm [103] typically requires approximately 30 steps with 3 trial correlations per step

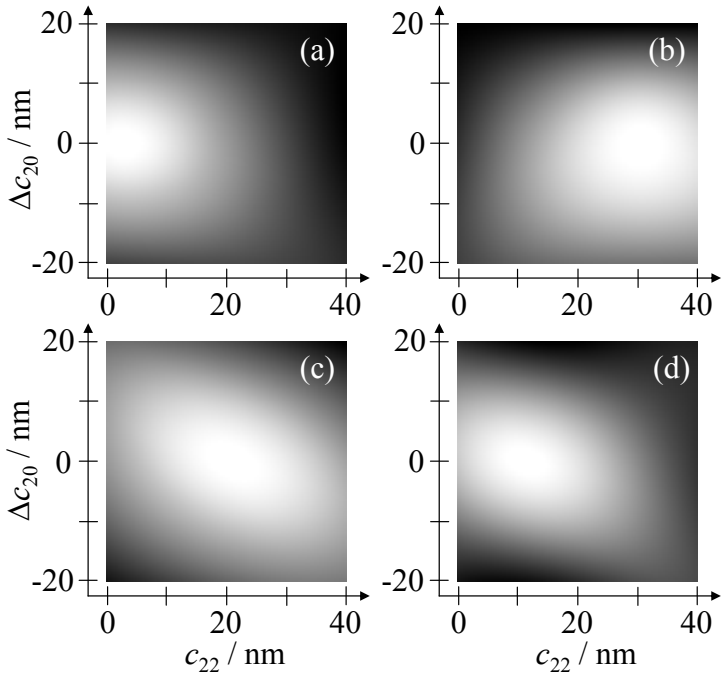



Figure 6.13: Correlation coefficients over a large range of trial parameters for exemplary images out of 4 different series, measured with different materials and microscopes: (a) semi-amorphous $SrTiO_3$, FEI Titan 80-300 (300 kV), (b) amorphous carbon with gold clusters, Philips CM300ST (300 kV), (c) amorphous Carbon, and (d) amorphous silicon, Philips CM200C (200 kV). The horizontal axes of the images cover a range from 0 nm up to 40 nm of the astigmatism modulus $c_{22}$, while the vertical axes comprise a defocus variation of $\Delta c_{22} = \pm 20$ nm around the optimum defocus value. The grey scale depicts larger correlation coefficients brighter with the maximum in white.

to converge to the maximum correlation coefficient. One run of the full UPR-algorithm lasts thus approximately 200 milliseconds for one diffractogram.

An indirect numerical method like the Simplex maximisation provides no direct error estimation. Without further restrictions, the Simplex Algorithm proceeds until the next step improves the correlation by less then a certain amount. In practice it is sufficient to stop the maximisation process when the difference of the trial parameters between the current trial and the next trial is smaller than 0.01 nm for 3 successive cycles.

In order to provide a feasible estimate for the statistical error of the pattern-recognition result, the implementation of the UPR-algorithm developed in this work maximises the correlation for two disjunct path sets. The error estimates for defocus and twofold astigmatism are then calculated from the deviations of the results obtained with the disjunct path sets to the actual mean defocus and twofold astigmatism determined by maximisation of the correlation for all paths. The disjunct sets both contain paths which are uniformly, but randomly distributed over the diffractogram. Note that this error approx-

imation tends to underestimate the actual error, when the mean path radii $\bar{g}^j$ differ by less than the original diffractogram sampling rate. It is therefore mandatory to avoid an oversampling of the radial dimension by the complete path set.
Statistical investigations on different diffractograms with about 100 randomly chosen path distributions yield standard deviations down to 0.1 nm for defocus and astigmatism. Accordingly, the estimated statistical error of the UPR is an order of magnitude better compared to the accuracy of any direct defocus and astigmatism measurement with previous realisations of a diffractogram analysis. It also provides a precision which is an order of magnitude better than the demanded $\sigma_{22} < 1.2$ nm as derived in chapter 5. An aberration measurement using the diffractogram method on the basis of this novel diffractogram analysis can thus easily satisfy the tolerance limit $\bar{\sigma}_\chi(g_{\max}) \leq \pi/4$ in the Sub-Ångström regime.

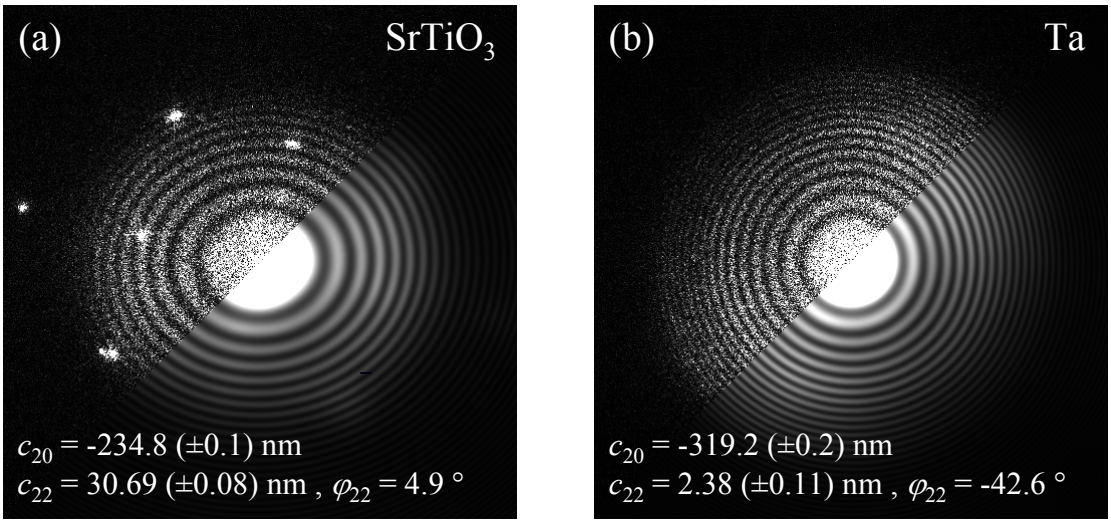



Figure 6.14: Split images of experimental diffractograms (top-left) and calculated patterns (bottom-right). The calculated patterns are determined on the basis of the defocus $c_{20}$ and the twofold astigmatism $c_{22}$ as measured with the UPR procedure including detected low-frequency content $D_{\mathrm{LF}}$ and envelope data $E$. Figure (a) shows a diffractogram of semi-amorphous $SrTiO_3$. Figure (b) shows the diffractogram of amorphous tantalum. The images were recorded with the Philips CM200C. The values for $c_{20}$, $c_{22}$ as measured by the UPR-procedure are specified with error estimates in brackets.

Figure 6.14 shows two split images comparing the experimental diffractogram in the top-left half to a calculated pattern in the bottom-right half. The latter is calculated with the measured second-order aberrations and the extracted background and envelope data. The Thon-ring pattern of the calculated pattern agrees perfectly on a visual basis with the experimental data also in figure 6.14b comprising 16 almost round Thon-rings with very small distances in the outer diffractogram regions. Figure 6.14a is an example for

a successful diffractogram analysis in the presence of several diffractogram peaks. This demonstrates that the limitation for the diffractogram method to purely amorphous material can be frequently overcome by using a detection an removal algorithm of peaks in diffractograms as described in section 6.2.

### 6.7.4 Correlation and signal-to-noise ratio

The following derivation reveals a relation between the correlation maximum $r_0 = \max(r)$ and the signal-to-noise ratio SNR of the diffractogram intensity $D(\vec{g}\,)$. First consider that in case of the maximised correlation between the experimental pattern $D_{\mathrm{exp}}(\vec{g}\,)$ and the theoretical predicted pattern $D_{\mathrm{theo}}(\vec{g}\,)$ the target signal of defocus and twofold astigmatism is equally present in both patterns and given by

$$\begin{aligned} S(\vec{g}\,) &= E(\vec{g}\,) \cdot P(\vec{g}\,), && (6.7.3)\\ &= E(\vec{g}\,) \cdot \left( \sin^2[\chi^{\mathrm{e}}(\vec{g}\,)] - \frac{1}{2} \right), && (6.7.4) \end{aligned}$$

as discussed in the preceding text. Per definition it is $D_{\mathrm{theo}}(\vec{g}\,) = S(\vec{g}\,)$ and, as background and diffractogram peaks have been removed from the experimental data, also $D_{\mathrm{exp}}(\vec{g}\,) = S(\vec{g}\,) + N'(\vec{g}\,)$, where $N'(\vec{g}\,)$ denotes the remaining random-noise (see section 6.4). Both data-sets, $D_{\mathrm{exp}}$ and $D_{\mathrm{theo}}$, exist in form of discrete arrays. Let us furthermore consider $S(\vec{g}_{\,i}) = s_i$ and $N'(\vec{g}_{\,i}) = n_i$ in respective discrete representations, where $i \rightarrow \vec{g}_{\,i}$ assigns discrete positions on the adaptive paths through the diffractogram to each data point.
The signal-to-noise ratio is consequently defined by the ratio between signal power $v_{\mathrm{ss}}$ and noise power $v_{\mathrm{nn}}$

$$\begin{aligned} \mathrm{SNR} &= \frac{v_{\mathrm{ss}}}{v_{\mathrm{nn}}}, && (6.7.5)\\ v_{\mathrm{ss}} &= \sum_i (s_i - \bar{s})^2, && (6.7.6)\\ v_{\mathrm{nn}} &= \sum_i (n_i - \bar{n})^2, && (6.7.7) \end{aligned}$$

where $\bar{s}$ and $\bar{n}$ are the mean values of the respective entities $S$ and $N$.
According to definition (A.5.1), the correlation of the two discrete datasets is given by

$$r_0 = \frac{v_{\mathrm{ss}} + v_{\mathrm{sn}}}{\sqrt{v_{\mathrm{ss}}(v_{\mathrm{ss}} + 2\, v_{\mathrm{sn}} + v_{\mathrm{nn}})}}, \qquad (6.7.8)$$

with $v_{\rm sn} = \sum_i (s_i - \bar{s}) \cdot (n_i - \bar{n})$. Further transformations yield

$$r_0 = \frac{1 + \frac{v_{\rm sn}}{v_{\rm ss}}}{\sqrt{1 + 2\,\frac{v_{\rm sn}}{v_{\rm ss}} + \frac{v_{\rm nn}}{v_{\rm ss}}}}, \tag{6.7.9}$$

$$= \frac{1 + \frac{v_{\rm sn}}{\sqrt{v_{\rm ss} \cdot v_{\rm nn}}} \cdot \sqrt{\frac{v_{\rm nn}}{v_{\rm ss}}}}{\sqrt{1 + 2\,\frac{v_{\rm sn}}{\sqrt{v_{\rm ss} \cdot v_{\rm nn}}} \cdot \sqrt{\frac{v_{\rm nn}}{v_{\rm ss}}} + \frac{v_{\rm nn}}{v_{\rm ss}}}}, \tag{6.7.10}$$

where it is easy to recognise the signal-to-noise ratio and the correlation coefficient $r_{\rm sn} = v_{\rm sn}/\sqrt{v_{\rm ss} v_{\rm nn}}$ between signal and noise:

$$r_0 = \frac{1 + \frac{r_{\rm sn}}{\sqrt{\rm SNR}}}{\sqrt{1 + 2\,\frac{r_{\rm sn}}{\sqrt{\rm SNR}} + \frac{1}{\rm SNR}}} = \frac{\sqrt{\rm SNR} + r_{\rm sn}}{\sqrt{\rm SNR + 2\,\sqrt{SNR}\, r_{\rm sn} + 1}}. \tag{6.7.11}$$

For real random noise which contains no variations similar those contained in the signal, the correlation $r_{\rm sn}$ equals zero. It follows

$$r_0 = \frac{\sqrt{\rm SNR}}{\sqrt{\rm SNR + 1}} \quad \Rightarrow {\rm SNR} = \frac{r_0^2}{1 - r_0^2}. \tag{6.7.12}$$

which expresses the relation between signal-to-noise ratio SNR and the maximum correlation coefficient $r_0$ obtained with the here developed ultra-precise pattern-recognition procedure for defocus and twofold astigmatism in single diffractograms of thin amorphous objects.

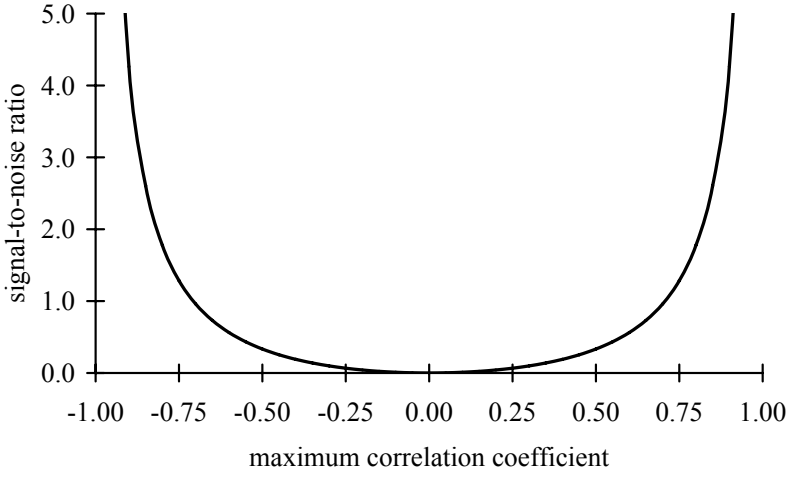



Figure 6.15: Dependence of the signal-to-noise ratio on the maximum correlation coefficient as obtained by the UPR-procedure. The signal is defined by $E(\vec{g}\,) \cdot P(\vec{g}\,)$ as in equation 6.1.2, whereas the noise represents random variations of the experimental diffractogram.

In practice, one observes maximum correlation coefficients between 0.5 to 0.75 depending on the illumination conditions and on the structure of the object. According to equation

(6.15), the signal-to-noise ratio thus varies between 0.33 and 1.3. This means that the noise power is approximately equal or up to three times larger than the signal power in numerical diffractograms of thin amorphous objects used for aberration measurements with the diffractogram method.

# Chapter 7

# Comprehensive evaluation of tilt tableaus

The novel procedures for the analysis of diffractograms as outlined in the previous chapter provides more information than actually required to determine the higher-order aberrations of a transmission electron microscope by using the evaluation scheme with the diffractogram method proposed in chapter 5. Besides effective astigmatism values, also effective defocus values and the low-frequency content of the diffractograms are extracted. Such additional data is used by the procedures described in this chapter in order to determine further properties of the electron-optical contrast transfer and to calculate reliable error estimates for aberration measurements.

## 7.1 Measurement of the mean defocus and of the temporal defocus drift

By using the evaluation scheme proposed in chapter 5 for a measurement of higher-order aberrations with the diffractogram method, the aberration function is determined on the basis of effective twofold-astigmatism values alone. This evaluation procedure does not yield a value for the mean defocus $c_{20}$. The mean defocus is determined by means of the effective defocus values which are also extracted from the diffractograms of the beam-tilt series as described in chapter 6.

The effective defocus values $c'_{20}(\vec{t}^{\,j})$ of images recorded with a tilted incident beam differ from the axial defocus $c_{20}$ present with untilted illumination according to the induction formalism of equation (5.2.1). The induced defocus $\Delta c'_{20}(\vec{t}^{\,j}) = c'_{20}(\vec{t}^{\,j}) - c_{20}$ depends on

the beam tilts $\vec{t}^{\,j}$ and the axial higher-order aberrations $c_{mn}$. Here, the latter are considered to be known from an evaluation of effective twofold astigmatism values as described in section 5.2. The measured effective defocus values are denoted here by $c_{20}^{\mathrm{M}}(\vec{t}^{\,j})$ in order to distinguish them more clearly from the calculated effective defocus values $c'_{20}(\vec{t}^{\,j})$ which one obtains by using the induction formalism of equation (5.2.1).

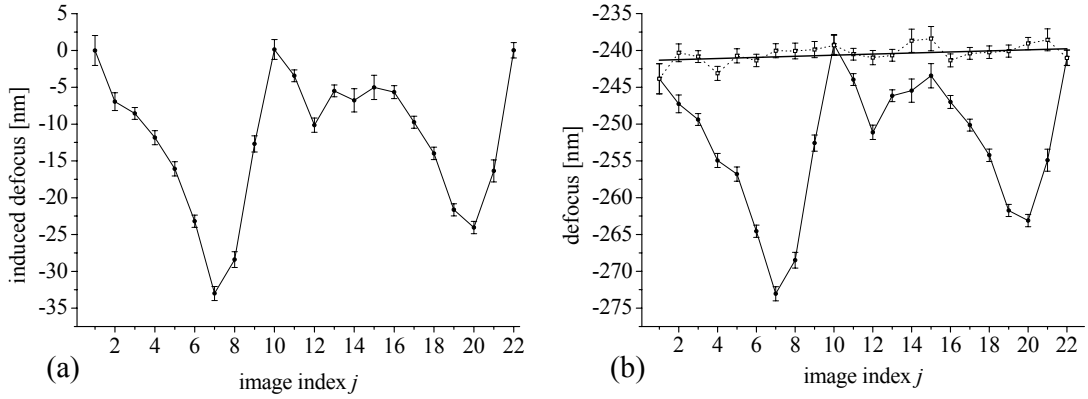



Figure 7.1: (a) Calculated induced defocus values $\Delta c'_{20}(\vec{t}^{\,j})$ and (b) effective defocus values $c_{20}^{\mathrm{M}}(\vec{t}^{\,j})$ as measured for an exemplary beam-tilt series of 22 images. The mean defocus $c_{20}$ and its linear temporal drift $\dot{c}_{20}$ are obtained from the residual defocus differences $c_{20}^{\mathrm{R}}(\vec{t}^{\,j})$ (dotted curve) between the calculated induced and the measured effective defocus values by linear regression (thick solid line).

Residual defocus values $c_{20}^{\mathrm{R}}(\vec{t}^{\,j})$ calculated by $c_{20}^{\mathrm{R}}(\vec{t}^{\,j}) = c_{20}^{\mathrm{M}}(\vec{t}^{\,j}) - \Delta c'_{20}(\vec{t}^{\,j})$ are shown in figure 7.1 for an exemplary beam-tilt series together with the measured effective defoci $c_{20}^{\mathrm{M}}(\vec{t}^{\,j})$ and the calculated induced defoci $c_{20}^{\mathrm{M}}(\vec{t}^{\,j})$. The residual defocus values are used to determine the mean defocus $c_{20}$ and its linear temporal drift $\dot{c}_{20}$ by linear regression of the dataset $(T_j, c_{20}^{\mathrm{R}}(\vec{t}^{\,j}))$ for the model

$$c_{20}(T_j) = c_{20} + \dot{c}_{20}(T_j - T_0). \tag{7.1.1}$$

$T_j$ denotes the time index of the image $j$ with respect to the time index $T_0$ which corresponds to time of acquisition in the middle of the sequentially recorded tilt series. A mean defocus of $c_{20} = -241.4$ ($\pm 0.5$) nm and a defocus drift of $\dot{c}_{20} = 0.037$ ($\pm 0.017$) nm/s was determined for the exemplary series shown in figure 7.1.

## 7.2 Determination of the azimuthal offset of the tilt tableau

The standard procedure used to calibrate beam tilts as described in section 5.3.2 is performed by operating the electron microscope in diffraction mode. By this calibration

procedure, the global azimuthal orientation of the beam tilts of a tilt tableau can not be determined, because the projection in diffraction mode and the projection in imaging mode deviate in general by a certain azimuthal offset. The tilt azimuths listed in the resulting calibrated tableau do therefore not necessarily correspond to the actual tilt azimuths measured relative to the $x$-axis in imaging mode. A practical procedure, which allows to quantify the global azimuthal offset of a tilt tableau with respect to the $x$-axis of the image is described in this section.

When rotating the whole tilt tableau around the origin in the diffraction plane by the angle $\Delta\varphi_t$ one observes a strong variation of the defocus consistency $s_{20}$ as shown in figure 7.2. The defocus consistency is defined by the square root of the mean squared differences between the residual defocus values $c_{20}^{\mathrm{R}}(\vec{t}^{\,j})$ and the time-dependent defocus $c_{20}(T_j)$ according to linear drift model in equation (7.1.1) as determined in section 7.1:

$$s_{20} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} |c_{20}^{\mathrm{R}}(\vec{t}^{\,j}) - c_{20}(T_j)|^2}. \tag{7.2.1}$$

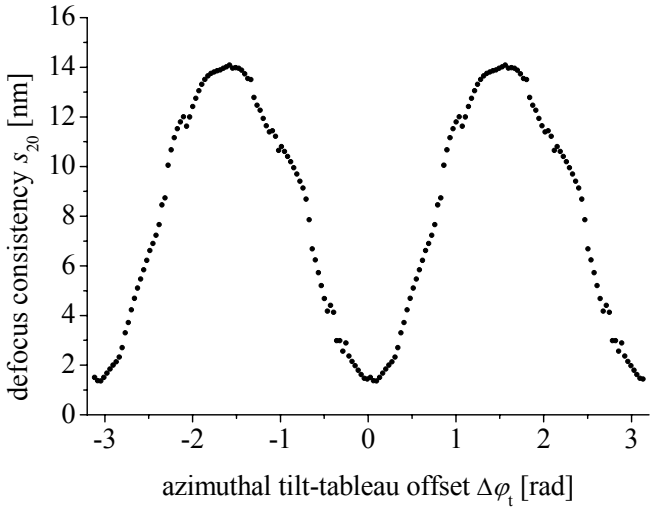



Figure 7.2: Defocus consistency $s_{20}$ depicted for variations of all azimuthal tilt parameters by the same amount $\Delta\varphi_t$. Over the whole azimuthal range (0 ... $2\pi$) one observes a strong twofold variation of the defocus consistency.

The defocus consistency $s_{20}$ is minimum when the global tilt azimuths of the tilt parameters used for the evaluation of a series of diffractograms corresponds to the actual beam tilts of the microscope. The ambiguity of the diffractogram analysis with respect to the sign of the second-order aberrations as discussed in section 6.5 is observable here as a twofold variation of the defocus consistency when changing the azimuthal tilt-tableau

offset $\Delta\varphi_t$ over the full azimuthal range. A minimisation of $s_{20}$ is therefore only unique in a range between $-\pi/2$ and $+\pi/2$.
There is a practical work around which allows to avoid the ambiguity problem in the present context. This procedure should be performed at least once for every beam-tilt calibration of tilt tableaus used for aberration measurements:

- A beam-tilt series consisting of $N$ images is acquired from amorphous material for an aberration measurement while an intentional underfocus is applied. Only if a subsequent aberration measurement including a determination of the mean defocus as described in section 7.1 yields an overfocus, i.e. $c_{20} > 0$, the whole tilt tableau should be rotated by $\pi$ (180°) in the diffraction plane.

- If the measured mean defocus is an underfocus, the considered azimuthal offset of the tilt tableau is correct within a range of $-\pi/2$ to $+\pi/2$. Within this range the defocus consistency $c_{20}$ exhibits only one absolute minimum. This minimum uniquely defines the azimuthal offset $\Delta\varphi_t$ of the tilt tableau relative to the result of the calibration procedure.

By repeating the above two steps with several beam-tilt series, a mean value $\Delta\bar{\varphi}_t$ is determined which yields the azimuthal offset of the actual beam tilts of the microscope to the result of a tilt calibration for a tilt tableau. The calibrated beam tilts considered for an aberration measurement by means of the diffractogram method have to be corrected for this azimuthal offset in order to obtain quantitatively correct values for the higher-order aberrations.

## 7.3 Consistent error estimate

As discussed in section 5.2.2, the error $\bar{\sigma}_\chi(g_{\max})$ of the measured aberration function is calculated by means of error propagation from the variances and covariance obtained numerically together with the least-squares solution of the higher-order aberrations. These variances and covariances are essentially influenced by the beam-tilt parameters and by the measurement errors $\sigma^j_{22} = \sigma^j_{22x} + \mathrm{i}\sigma^j_{22y}$ of the diffractogram analysis.
When aiming for a realistic and reliable error estimate of the measured aberration function, the error estimates of the effective twofold astigmatism should not only reflect the statistical accuracy of the diffractogram analysis as discussed in section 6.7, the error estimates should also consider systematic deviation of the experimental data from the model

describing the induction of twofold astigmatism. This inconsistency between the induction model and the measured effective twofold-astigmatism data is registered here as the *outer consistency* of the aberration measurement. In contrast, the statistical error estimates obtained with the pattern-recognition procedures for the individual diffractograms as described in chapter 6 is called *inner consistency*.

The outer consistency $s_{\text{all}}$ is defined by the square root of the mean squared differences between the effective astigmatism values $c_{22}^{\text{M}}(\vec{t}^{\,j})$, measured directly from single diffractograms, and astigmatism values $c'_{22}(\vec{t}^{\,j})$ which are calculated from the measured higher-order aberrations $c_{mn}$ and the applied beam tilts $\vec{t}^{\,j}$ by means of the induction formalism of equation (5.2.2):

$$s_{\text{all}} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} |c_{22}^{\text{M}}(\vec{t}^{\,j}) - c'_{22}(\vec{t}^{\,j})|^2} \tag{7.3.1}$$

An iterative procedure is applied to adjust the original error estimates $\sigma_{22x,j}$ and $\sigma_{22x,j}$ for the results of single diffractogram analysis, i.e. the inner consistency, in successive steps towards the observed deviations $|c_{22}^{\text{M}}(\vec{t}^{\,j}) - c'_{22}(\vec{t}^{\,j})|$ of the measurement from the mean values, i.e. to the outer consistency. Each iteration involves a new solution of the set of linear equations for the higher-order aberrations. The outer consistency is calculated on the basis of the new solution yielding updates for the error estimates $\sigma_{22x,j}$ and $\sigma_{22x,j}$ used for the next iteration. This procedure converges after typically less than 10 successive steps. The solution for the higher-order aberrations is not significantly altered, since only the weighting of the linear equations is changed.

Whereas the inner consistency values range from 0.1 nm to 0.2 nm for good-quality diffractograms (see fig. 6.14), aberration measurements on the basis of such diffractograms yield outer consistency values between 0.2 nm and 0.3 nm. According to this more reliable error estimate, the novel aberration-measurement procedures achieve still a far better accuracy than the required 1.2 nm (see section 5.4.2). It is also an order of magnitude better than the consistency achieved with current state-of-the-art aberration-measurement techniques.

In practice, several circumstances can cause larger outer-consistency values:

- Non-linear variations of residual astigmatism during the time of series acquisition,
- Falsely determined astigmatism azimuths, due to untreated diffractogram peaks,
- A too small set of higher-order aberrations,
- Falsely supplied magnification, and
- Wrong or uncalibrated beam-tilt parameters.

| | this work | | | existing solution | | |
|---|---|---|---|---|---|---|
| | $c_{mn}$ | $(\pm\sigma_{mn})$ | $\varphi_{mn}$ | $c_{mn}$ | $(\pm\sigma_{mn})$ | $\varphi_{mn}$ |
| $c_{20}$ | −236.8 | (±0.16) nm | – | −238.0 | (±2.37) nm | – |
| $\dot{c}_{20}$ | 0.11 | (±0.03) $\frac{\text{nm}}{\text{img}}$ | – | – | | – |
| $c_{22}$ | 2.46 | (±0.08) nm | 9.2° | 0.38 | (±3.18) nm | 4.5° |
| $\dot{c}_{22}$ | 0.010 | (±0.007) $\frac{\text{nm}}{\text{img}}$ | −64.5° | – | | – |
| $c_{31}$ | 388.7 | (±9.08) nm | 100.8° | 635.7 | (±94.5) nm | 83.2° |
| $c_{33}$ | 21.43 | (±3.94) nm | −11.3° | 27.91 | (±35.4) nm | 41.8° |
| $c_{40}$ | −25.58 | (±0.49) $\mu$m | – | −24.93 | (±2.21) $\mu$m | – |
| $c_{42}$ | 5.74 | (±0.17) $\mu$m | −3.2° | 7.83 | (±0.85) $\mu$m | 5.2° |
| $c_{44}$ | 1.08 | (±0.20) $\mu$m | −29.0° | 1.48 | (±0.54) $\mu$m | −27.1° |
| $c_{51}$ | 69.8 | (±10.6) $\mu$m | 176.3° | 375 | (±106) $\mu$m | −70.5° |
| $c_{53}$ | 44.3 | (±8.0) $\mu$m | −19.6° | 35 | (±56) $\mu$m | 38.6° |
| $c_{55}$ | 57.6 | (±1.5) $\mu$m | 0.5° | 73.9 | (±16.9) $\mu$m | −33.7° |
| $c_{60}$ | 10.27 | (±0.54) mm | – | 9.62 | (±2.64) mm | – |
| $c_{64}$ | 2.13 | (±0.42) mm | 8.2° | – | | – |
| $c_{66}$ | 3.50 | (±0.06) mm | −22.6° | 3.04 | (±0.59) mm | −22.6° |
| $s_{\text{all}}$ | 0.316 nm | | | 3.895 nm | | |

Table 7.1: Results of an aberration measurement using the novel techniques developed in this work and the existing software used for the alignment of spherical-aberration correctors based on the same tilt series. Both software solutions specify a value for the outer consistency $s_{\text{all}}$. The outer consistency with the novel aberration-measurement procedures is more than one order of magnitude better compared to the existing solution.

Table 7.1 lists the results and consistency values of an aberration measurement performed with the ultra-precise diffractogram analysis developed in this work. For comparison, the results obtained with the current standard technique are shown. The outer consistency obtained with the novel aberration-measurement procedures is more than one order of magnitude better compared to the existing solution. The improved consistency is reflected also by smaller variances of the single higher-order aberrations. With the example given in table 7.1, the aberration function is measured with a precision of $\sigma_\chi = 0.07\,\pi$ at the information limit $g_{\max} = 12.5\ \text{nm}^{-1}$ for 0.8 Å resolution.

The full potential of the new aberration-measurement technique can be estimated when calculating the error of the measured aberration function for different target resolutions as shown in figure 7.3. The value of $\bar{\sigma}_\chi(g)$ hits the $\pi/4$-limit for $g = g_{\text{int}} \approx 20\ \text{nm}^{-1}$.

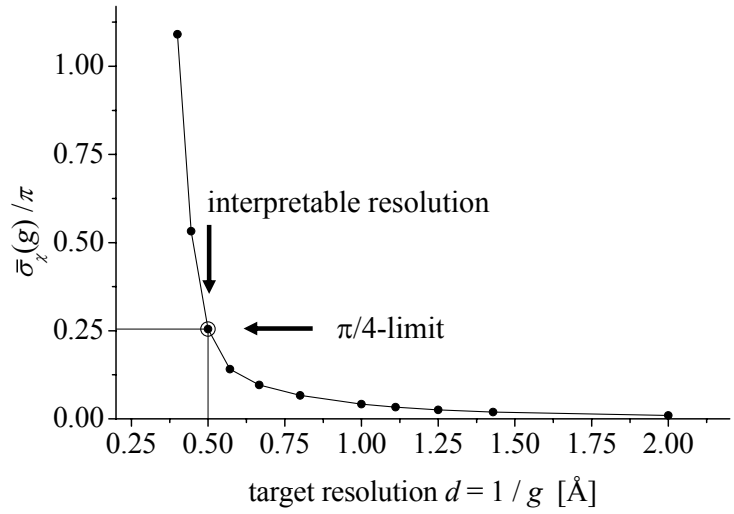



Figure 7.3: Error $\bar{\sigma}_\chi(g_{\max})$ of the measured aberration function as estimated for an exemplary aberration measurement with results as shown in tab. 7.1. The error of the measured aberration function is extrapolated here for different target resolutions $d_{\text{info}} = 1/g_{\max}$.

According to the discussions in section 2.5, the precision of this aberration measurement provides an interpretable resolution of $1/g_{\text{int}} = 0.5$ Å. Thus, the goal to develop aberration-measurement techniques for Sub-Ångström resolution around 0.8 Å is actually over-accomplished by the procedures developed in this work. The novel techniques also enable aberration control for the future generation of transmission electron microscopes with deep Sub-Ångström resolution around 0.5 Ångström.

## 7.4 Direct measurement of the defocus spread

A method which, for the first time, allows to quantitatively determine the defocus-spread parameter of a transmission electron microscope is described in this section. The defocus spread $\Delta$ characterises the partial temporal coherence of the electron beam and is used to specify the resolution limit of transmission electron microscopes. A strong anisotropic reduction of the partially coherent, linear contrast transfer can be observed when illuminating the object with a strongly tilted electron beam. This anisotropic damping is analysed quantitatively in the following.

### 7.4.1 Theoretical description of damping anisotropies

Within the contrast-transfer theory for tilted-beam illumination elaborated in section 3.2, damping envelopes of the contrast transfer are derived which exhibit an anisotropy in the

azimuthal direction of the image depending on the beam tilt $\vec{t}$. This anisotropy with tilted-beam illumination appears in particular with the damping of the contrast transfer due to a partial temporal coherence of the electron beam and becomes apparent in diffractograms of thin amorphous objects.

Within this work, numerical procedures are developed which allow to isolate an anisotropic envelope of the intensity distribution $D(\vec{g})$ in diffractograms of thin amorphous objects by means of image processing as described in section 6.3 and section 6.4. The low-frequency content $D_{\mathrm{LF}}(\vec{g})$ of a diffractogram as in equation (6.3.3) comprises three components in terms of the general phenomenological model of equation (6.1.2) for $D(\vec{g})$. These components are the half of the envelope $E(\vec{g})$, the background $B(\vec{g})$ and low-frequency components of random noise $N_{\mathrm{LF}}(\vec{g})$ :

$$D_{\mathrm{LF}}(\vec{g}) = \frac{1}{2} E(\vec{g}) + B(\vec{g}) + N_{\mathrm{LF}}(\vec{g}). \tag{7.4.1}$$

Theoretical models which describe the envelope and background of diffractograms due to the partially coherent contrast transfer for tilted-beam illumination have been derived in section 3.2 with

$$E'(\vec{g},\vec{t}) = \exp[2S(\vec{g},\vec{t})], \tag{7.4.2}$$

$$B'(\vec{g},\vec{t}) = 4|\Psi_{\mathrm{e}}(\vec{g})|^2 \exp[2S(\vec{g},\vec{t})] \cdot \sinh^2\left[A(\vec{g},\vec{t})\right], \tag{7.4.3}$$

where the abbreviations $S(\vec{g},\vec{t})$ and $A(\vec{g},\vec{t})$ are defined in equation (3.2.21) and equation (3.2.22). The term $4|\Psi_{\mathrm{e}}(\vec{g})|^2$ appears as additional factor beside the envelope $E'(\vec{g},\vec{t})$ in the diffractogram intensity $D(\vec{g})$ of equation (3.2.28), and in the background $B'(\vec{g},\vec{t})$ of equation (3.2.32). For amorphous objects, $4|\Psi_{\mathrm{e}}(\vec{g})|^2$ acts like a rotationally symmetric envelope on the terms $E(\vec{g})$ and $B(\vec{g})$ in equation (7.4.1).

Besides a partial coherence of the electron beam, other sources can dampen the intensity distribution of diffractograms. Therefore, additional rotationally symmetric but also non-rotationally symmetric contributions are contained in the low-frequency content $D_{\mathrm{LF}}(\vec{g})$ of the diffractogram as discussed in section 6.1. Phenomena which cause rotationally symmetric damping are for example limited scattering angles for the electron diffraction and a finite sample thickness. Additional non-rotationally symmetric damping can be caused for example due to object drift and object tilt.

The concept behind the evaluation scheme for the characteristic non-rotationally symmetric damping due to the partial temporal coherence applied here is to avoid an interference of the target signal by such as the aforementioned phenomena. Therefore, the low-frequency content of the diffractogram which exhibits the wanted signal is evaluated

along circles around the Fourier-space origin to avoid an interference by the various rotationally symmetric damping envelopes.

Additional non-rotationally symmetric damping envelopes are cancelled out by cumulating the low frequency content of several diffractograms obtained with an azimuthal beam-tilt series. With the cumulation, a characteristic dependence of the target signal on the tilt azimuth $\varphi_t$ is exploited, which is not existing with the unwanted interfering signal. In good approximation, only the damping due to a partially coherent contrast transfer depends on the beam-tilt azimuth $\varphi_t$.

Since the low-frequency diffractogram content $D_{\mathrm{LF}}(\vec{g}\,)$ will be analysed in terms of oscillations along circular paths around the Fourier-space origin $\vec{g}\,=0$, the target signal in equation (7.4.1) including the equations (7.4.2) and (7.4.3) is expressed in polar coordinates

$$
\begin{aligned}
D_{\mathrm{LF}}(g,\varphi) &= 4|\Psi_{\mathrm{e}}(\vec{g}\,)|^2\,\exp[2S(\vec{g}\,,\vec{t}\,)]\left\{\frac{1}{2}+\sinh^2\left[A(\vec{g}\,,\vec{t}\,)\right]\right\} && (7.4.4)\\
&= 2|\Psi_{\mathrm{e}}(\vec{g}\,)|^2\,\exp[2S(\vec{g}\,,\vec{t}\,)]\,\cosh\left[2A(\vec{g}\,,\vec{t}\,)\right] && (7.4.5)\\
&= {E'_{\mathrm{s}}}^2(\vec{g}\,,\vec{t}\,)\cdot\left\{B'_0(g)+E'_0(g)\cdot\exp\left[-\kappa(g,t)\,\cos^2(\varphi-\varphi_t)\right]\right. && (7.4.6)\\
&\qquad\left.\cdot\cosh\left[\kappa'(g,t)\cos(\varphi-\varphi_t)+2A'(\vec{g}\,,\vec{t}\,)\right]\right\}.
\end{aligned}
$$

For simplification, $B'_0(g)$ and $E'_0(g)$ represent all terms which do not depend on $(\varphi-\varphi_t)$ explicitly. The defocus spread $\Delta$ is thereby the only unknown in the respective amplitudes $\kappa(g,t)=2(\pi\Delta\lambda gt)^2$ and $\kappa'(g,t)=2(\pi\Delta\lambda gt)^2\,g/t$. The terms $E'_{\mathrm{s}}(\vec{g}\,,\vec{t}\,)$ and $A'(\vec{g}\,,\vec{t}\,)$ account for the effects due to a partial spatial coherence of the electron beam, with

$$
\begin{aligned}
E'_{\mathrm{s}}(\vec{g}\,,\vec{t}\,) &= \exp\left\{-\left(\frac{q_0}{2}\right)^2\left[\left|\vec{\nabla}\,\chi_t^{\mathrm{e}}(\vec{g}\,)\right|^2+\left|\vec{\nabla}\,\chi_t^{\mathrm{o}}(\vec{g}\,)\right|^2\right]\right\}, && (7.4.7)\\
A'(\vec{g}\,,\vec{t}\,) &= 2\left(\frac{q_0}{2}\right)^2\left[\vec{\nabla}\,\chi_t^{\mathrm{e}}(\vec{g}\,)\right]\cdot\left[\vec{\nabla}\,\chi_t^{\mathrm{o}}(\vec{g}\,)\right]. && (7.4.8)
\end{aligned}
$$

Both of the above terms depend on the one hand on the parameter $q_0$, which characterises the partial spatial coherence of the electron beam (see section 1.3.5) and on the other hand on the gradients of the even part $\chi_t^{\mathrm{e}}(\vec{g}\,)$ and the odd part $\chi_t^{\mathrm{o}}(\vec{g}\,)$ of the effective aberration function (see section 3.2.1). As depicted in figure 7.4, strong residual aberrations significantly influence the oscillations of the low-frequency content of a diffractogram described by the terms $E'_{\mathrm{s}}(\vec{g}\,,\vec{t}\,)$ and $A'(\vec{g}\,,\vec{t}\,)$. "Strong" aberrations means for example an axial coma of 500 nm which is rarely observed with aberration-corrected microscopes. The applicability of the measurement procedure for the defocus spread described here is therefore restricted to aberration-corrected microscopes, where the term $E'_{\mathrm{s}}(\vec{g}\,,\vec{t}\,)$ is considerably constant along a circular path around the Fourier-space origin and where $A'(\vec{g}\,,\vec{t}\,)\approx 0$. In

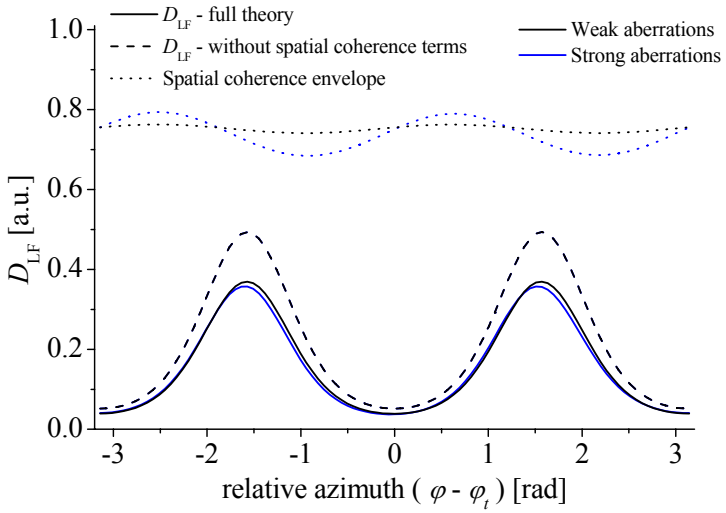



Figure 7.4: Azimuthal variations of low-frequency content $D_{\mathrm{LF}}$ of a diffractogram calculated according to equation (7.4.6) along a path of circular shape with radius $g = 3.0\ \mathrm{nm}^{-1}$ around the diffractogram origin, with a tilt modulus of $t = 40$ mrad ($\lambda = 1.969$ pm, $\theta_0 = \lambda q_0 = 0.2$ mrad, and $\Delta = 3$ nm). $D_{\mathrm{LF}}(g, \varphi)$ is plotted with all terms of eqn. (7.4.6) (solid curves), without spatial coherence terms (dashed curves). Spatial coherence envelopes (dash-dotted curves) are plotted separately. Two aberration scenarios are considered: 1. (black curves): $c_{20} = -200$ nm, $c_{22} = 5$ nm, $c_{31} = c_{33} = 100$ nm, $c_{40} = -13$ $\mu$m, and 2. (blue curves): $c_{20} = -200$ nm, $c_{22} = 25$ nm, $c_{31} = c_{33} = 500$ nm, $c_{40} = -13$ $\mu$m.

this case, the experimentally obtained low-frequency diffractogram data $D_{\mathrm{LF}}$ is described analytically in polar coordinates by

$$D_{\mathrm{LF}}(g, \varphi) = B_0(g) + E_0(g) \cdot \exp\left[-\kappa(g,t) \cdot \cos^2(\varphi - \varphi_t)\right] \cdot \cosh\left[\kappa'(g,t) \cdot \cos(\varphi - \varphi_t)\right] \tag{7.4.9}$$

with only radially varying $B_0(\vec{g}\,)$ and $E_0(\vec{g}\,)$ as discussed above. Least-squares fits of the model function (7.4.9) to the experimental $D_{\mathrm{LF}}$-data extracted from a circles of radius $g$ around the Fourier-space origin yields a value for the parameter $\kappa = \frac{t}{g}\kappa'$, and thus also for the defocus spread $\Delta$.

### 7.4.2 Experimental setup

The experimental procedure for a measurement of the defocus-spread parameter $\Delta$ is performed as a part of an aberration-measurement using a beam-tilt series with extremely large maximum tilt modulus. Completely amorphous objects with strongly scattering elements like tantalum, tungsten or strontium provide a strong target signal in the intensity distribution of diffractograms. The maximum tilt modulus should be set to approximately

$$t_{\max} = 2t_{\max,\mathrm{opt}} = \frac{2g_{\max}^2}{\sqrt{8}g}, \tag{7.4.10}$$

which is two times the optimum modulus for aberration measurements proposed in section 5.4.3. The spatial frequency $g$ in equation (7.4.10) denotes the approximate extent of the intensity distribution in the diffractograms observed with untilted illumination.

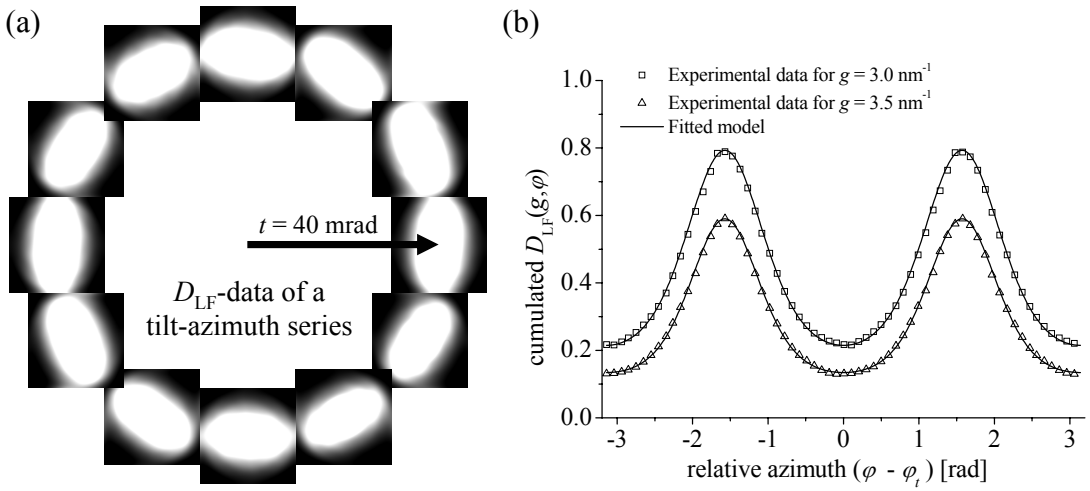



Figure 7.5: (a) Low-frequency content $D_{\mathrm{LF}}$ extracted from diffractograms of amorphous tantalum recorded with the FEI Titan 80-300. The arrangement of the images corresponds to the respective beam tilts with a maximum modulus of 40 mrad. (b) Cumulated experimental $D_{\mathrm{LF}}$-data along paths of circular shape for two exemplary spatial frequencies. The model function of eqn. (7.4.9) is drawn as lines with fitted parameters for the respective experimental data.

Figure 7.5a shows the low-frequency content $D_{\mathrm{LF}}$ as extracted from 12 diffractograms recorded under tilted-beam illumination. The arrangement of the $D_{\mathrm{LF}}$-data-images corresponds to the tilt positions 2 up to 13 of a (1/12/4/1)-standard tilt tableau which have a maximum tilt modulus of 40 mrad and sample the azimuthal tilt dimension by an average increment of 30 degrees. The low-frequency contents of all diffractograms exhibit an almost identical anisotropy. For cumulation, each $D_{\mathrm{LF}}$-dataset is rotated such that the projection of the tilt vector in the diffraction plane coincides with the $x$-axis of the cumulated data array.

The defocus-spread parameter $\Delta$ is determined quantitatively from least-squares fits of the model function (7.4.9) to the experimental data along different circles of radius $g$ around the origin of the cumulated $D_{\mathrm{LF}}$-data (e.g. in figure 7.5c). Such separate fits yield values for the amplitude parameter $\kappa$ from which the defocus spread is calculated by

$$\Delta = \sqrt{\frac{\kappa}{2}}\,\frac{1}{\pi\lambda g t_{\max}}, \tag{7.4.11}$$

at different spatial frequencies $g$.

### 7.4.3 Discussion of experimental results

According to the theory for the partially coherent linear contrast transfer, the defocus spread should not depend on the spatial frequency $g$.

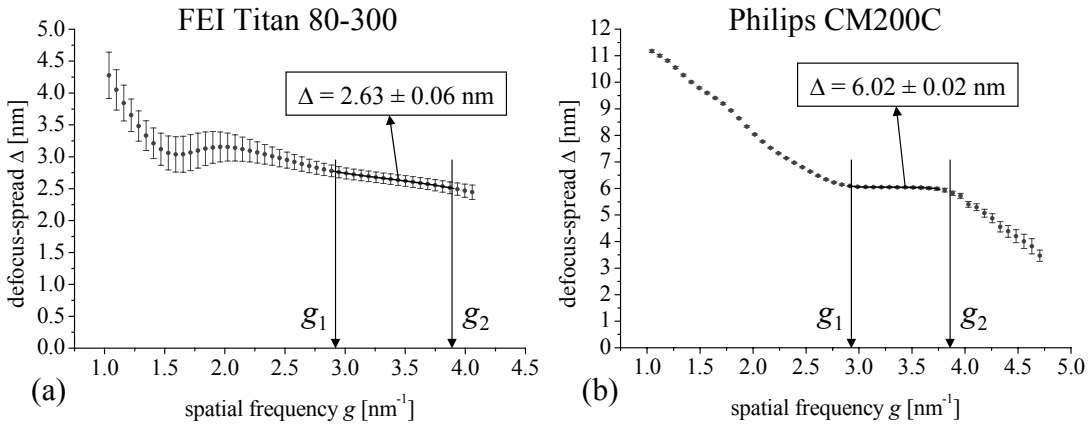



Figure 7.6: Results of defocus-spread measurements for (a) the FEI Titan 80-300 and (b) the Philips CM200C. The separate fits of the model function (7.4.9) to the cumulated low-frequency content of 12 diffractograms reveal unexpected variations of the measurement result below $g_1 \approx 3$ nm$^{-1}$ and above $g_2 \approx 4$ nm$^{-1}$. In between this interval, the variations around the defocus-spread mean-value are smaller than one Ångström.

Figure 7.6 reveals unexpected variations of the measured defocus spread for spatial frequencies below $g_1 \approx 3$ nm$^{-1}$ and for spatial frequencies above $g_2 \approx 4$ nm$^{-1}$. This deviation from the theoretical expectation can be explained qualitatively:

- For low spatial frequencies ($g < g_1$) numerical diffractograms have typically a higher intensity than expected in the linear contrast-transfer theory. The origin of this derivation of the experiment to the theory is currently not sufficiently understood. In all defocus-spread measurements carried out so far, $\Delta$ increases towards smaller spatial frequencies below approximately 3 nm$^{-1}$. Slight oscillations of $\Delta$ as in figure 7.6a can occur due to a too small underfocus. In this case, the frequency limit for the adaptive low-pass filter used for the extraction of the low-frequency diffractogram content unintentionally includes low-frequency components of the Thon-ring pattern.

- At higher spatial frequencies ($g > g_2$), the measured defocus spread typically decreases due to a reduced visibility of the Thon-ring pattern, i.e. due to a weaker target signal. This reduction is described by various rotationally symmetric envelopes. Due to the same reason, also larger error bars are obtained for $g > g_2$.

Within the interval 3 $nm^{-1} < g < 4$ $nm^{-1}$, the deviation of $\Delta$ from the mean value is smaller than one Ångström in both example results shown in figure 7.6. The mean value of the defocus spread measured for the FEI-Titan 80-300 is $\Delta = 2.63\,(\pm 0.06)$ nm which corresponds to an information limit of $g_{max} = 13.3$ $nm^{-1}$ or $d_{info} = 0.75$ Ångström. The information limit of the Philips CM200C has been determined from Young's-fringe tests with $g_{max} = 7.7$ $nm^{-1}$ [47]. According to equation (1.3.27), this result agrees with the, here directly measured, defocus-spread value of $\Delta = 6.02\,(\pm 0.02)$nm.

### 7.4.4 Conclusion

A novel method to quantitatively determine the defocus-spread parameter of a transmission electron microscope is developed. In contrast to a, so far used, optical inspection of a Young's fringe pattern, the quantitative measurement result obtained with the new method does not rely on subjective perceptions and arbitrarily chosen limits by the users. The essential idea behind this defocus-spread measurement is to analyse an anisotropy in the envelope of diffractogram intensity-distributions occurring with tilted-beam illumination.
Low-frequency data is extracted from diffractograms of a thin amorphous object and evaluated along circles around the Fourier-space origin by exploiting general symmetry properties of the involved phenomena. This approach requires no quantitative knowledge about rotationally symmetric, low-frequency features of the diffractogram intensity. Azimuthally anisotropic low-frequency features which would systematically distort a reliable quantitative defocus-spread measurement from one diffractogram are avoided by cumulation of several diffractograms of an azimuthal tilt series. The evaluation scheme described here does not account for the azimuthal anisotropy of the diffractogram due to a partially spatial coherent electron beam. A reliable application of this measurement procedure is therefore restricted to aberration-corrected microscopes.

## 7.5 Combination of real-space and Fourier-space data

Two methods which, for the first time, evaluate Fourier-space data obtained from diffractograms in combination with information obtained from the according real-space input images are described in this section. These new methods evaluate directly measured image displacements and calculated induced image displacements to determine the defocus sign and the magnification applied during aberration measurements on the basis of the

diffractogram method.

### 7.5.1 Determination of the defocus sign

An ambiguity problem of the diffractogram method with respect to the sign of the even part of the aberration function is encountered with the analysis of single diffractograms for defocus and twofold astigmatism as discussed in chapter 6. A propagation of a wrong defocus sign into a wrong sign of all coefficients of the aberration function constitutes a severe problem, since an aberration correction with a false sign of the measured aberrations would increase the aberration instead of reducing it. The ambiguity problem with respect to the sign of the aberration function is usually tackled by advising the operator to apply a negative defocus for aberration measurements with the diffractogram method. A new method is developed, which allows for the first time to correct automatically a wrong assumed defocus sign after the aberration measurement.

Aberration measurements with the diffractogram method typically apply a maximum beam-tilt modulus $t$ in the range between 10 mrad up to 30 mrad and an absolute defocus value between 150 nm up to 500 nm. Therefore, the tilt-induced displacements $\Delta\vec{c}'^{j}_{11} = c_{20} \cdot \vec{t}^{\,j}$, due to the defocus $c_{20}$ alone, range between 1.5 nm and 15 nm. This displacement is in many cases larger than typical object drifts during the time required for the acquisition of a beam-tilt series. Due to this domination of the induced image displacement over other contributions for an image displacement, a wrong assumed defocus sign is observable, when the calculated induced image displacements $\Delta\vec{c}'^{j}_{11}$ are anti-parallel to the directly measured image displacements $\vec{d}^{j}$. If these vectors are parallel, the assumed defocus sign is correct.

For a numerical determination of the defocus sign, the results obtained with the diffractogram method, i.e. the measured aberration function comprising all aberrations of the second and higher orders, is used to calculate tilt-induced image displacements $\Delta\vec{c}'^{j}_{11}$ via the induction formalism derived in appendix C.2.1. These displacement vectors are compared to the respective directly measured image displacements $\vec{d}^{j}$. The normalised inner product $p^j$ determines the angular disorientation between $\vec{d}^{j}$ and $\Delta\vec{c}'^{j}_{11}$ by

$$p^j = \frac{\vec{d}^{j} \cdot \Delta\vec{c}'^{j}_{11}}{|\vec{d}^{j}| \cdot |\Delta\vec{c}'^{j}_{11}|} = \cos\left[\angle(\vec{d}^{j}, \Delta\vec{c}'^{j}_{11})\right] \tag{7.5.1}$$

Ideally, i.e. without inconsistencies between the diffractogram method and the real-space displacement measurements, the mean value $\bar{p} = \frac{1}{N}\sum_{j=2}^{N} p_j$ for all images of a beam-tilt series yields $+1$ if the currently assumed defocus sign is correct, and $-1$ if the sign is

wrong.

Systematic deviations from the parallel or anti-parallel orientation can occur for example due to non-linear drifts of the object position. Random variations of object position and the twofold astigmatism, as well as measurement errors cause stochastic deviations of the angular orientations $p^j$ from the mean value $\bar{p}$. Both, systematic and stochastic variations, lead to a reduction of the absolute mean value $|\bar{p}|$.

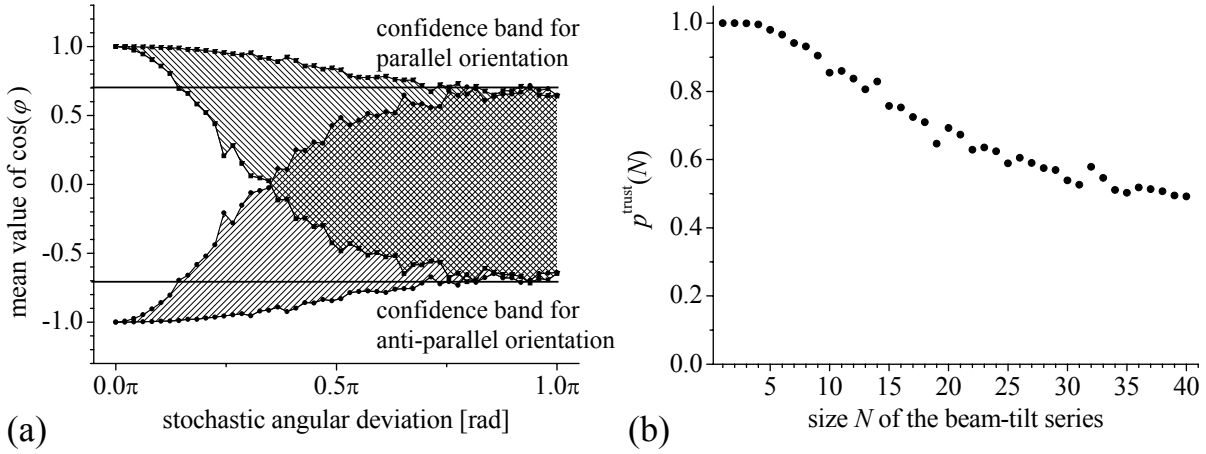



Figure 7.7: (a) Numerical estimation of confidence bands for a determination of the defocus sign by the mean value of $\cos(\varphi)$ for $N = 20$ random angles $\varphi$ deviating around $\varphi_0 = 0$ (parallel orientation) and $\varphi_0 = \pi$ (anti-parallel orientation). The minimum and maximum mean values obtained from 100000 trials are plotted in dependence on the width of the stochastic deviation which is assumed to follow a normalised Gaussian distribution. (b) Estimation of the inner limit $p^{\text{trust}}$ of the confidence bands for a reliable determination of the defocus sign in dependence of the number $N$ of images per beam-tilt series.

The confidence bands for a reliable determination of the defocus sign by means of $\bar{p}$ are estimated numerically. Figure 7.7a shows that the absolute value of $\bar{p}$ decreases with increasing stochastic angular deviation for both scenarios, parallel orientation due to a correct defocus sign and anti-parallel orientation due to a wrong defocus sign. For the shown case, a beam-tilt series of size $N = 20$ is considered. Moreover, the possible range for the value of $\bar{p}$ increases. According to this numerical estimation, the confidence band where the current defocus sign is reliably correct is $\bar{p} \in [+0.7, +1.0]$, whereas $\bar{p} \in [-1.0, -0.7]$ reliably signalises that the defocus sign is wrong. Since the amount of stochastic deviations is not known, a determination of the defocus sign for $|\bar{p}| < 0.7$ is not trustworthy. Fortunately, in most of the cases the deviations are small enough to allow for a reliable decision.

The inner limit $p^{\text{trust}}$ of the confidence bands $[+p^{\text{trust}}, +1.0]$ and $[-1.0, -p^{\text{trust}}]$ (with $p^{\text{trust}} > 0$) depends on the number $N$ of images per beam-tilt series as shown in fig-

ure 7.7b. The numerically estimated values of $p^{\text{trust}}(N)$ render the value 0.7 for $N = 20$, as observed in figure 7.7a. The confidence band broadens for larger tilt series such that, for example, $\bar{p} \in [+0.5, +1.0]$ would reliably confirm the assumed defocus sign for $N = 30$.

### 7.5.2 Observation of displacement discrepancies

Simultaneous application of the beam-tilt/diffractogram method (BTD) and the beam-tilt/cross-correlation method (BTC) with one beam-tilt series yields, in practice, different results. With the novel techniques introduced in the beginning of this chapter, the BTD-method is currently the more accurate and trustworthy solution. Nevertheless, the measurement of image displacements by application of cross correlations (see appendix A.6) provides additional data and requires comparatively short computation time.
Temporal variations of object position and of aberrations as well as incorrectly set parameters of the aberration measurements are the main reasons causing inconsistencies between the two methods. The procedure described in the following, compares the two methods on the basis of discrepancies between independently determined image displacements:

- BTC-method: The image displacements $\vec{d}^{\,j} = (d_x^j, d_y^j)$ between the first and the other images ($1 < j \leq N$) of a beam-tilt series are measured directly by cross correlations. The raised index $j$ denotes the image number in the acquired beam-tilt series.

- BTD-method: Effective defocus $c'^j_{20}$ and effective twofold astigmatism $c'^j_{22x}$, $c'^j_{22y}$, are measured directly from the diffractograms of the images. The effective twofold astigmatism data is used to determine the present higher-order aberrations $c_{mn}$. Tilt-induced displacements

$$\Delta\vec{c}'^{\,j}_{11} = (\Delta c'^j_{11x}, \Delta c'^j_{11y}) = (c'_{11x}(\vec{t}^{\,j}) - c_{11x}, c'_{11y}(\vec{t}^{\,j}) - c_{11y}) \tag{7.5.2}$$

  are then calculated via the induction formula derived in appendix C.2.1 with previously measured higher-order aberrations $c_{mn}$ and the tilt vectors $\vec{t}^{\,j}$ applied during the acquisition of the images.

- Combination: The displacement discrepancies $\vec{\delta}^j$ are finally calculated by

$$\vec{\delta}^j = (\delta_x^j, \delta_y^j) = \vec{d}^{\,j} - \Delta\vec{c}'^{\,j}_{11}. \tag{7.5.3}$$

Figure 7.8 schematically describes the procedure for the determination of displacement discrepancies between the two methods from one beam-tilt series. Provided that the microscope is well aligned and all parameters of the aberration measurement, especially

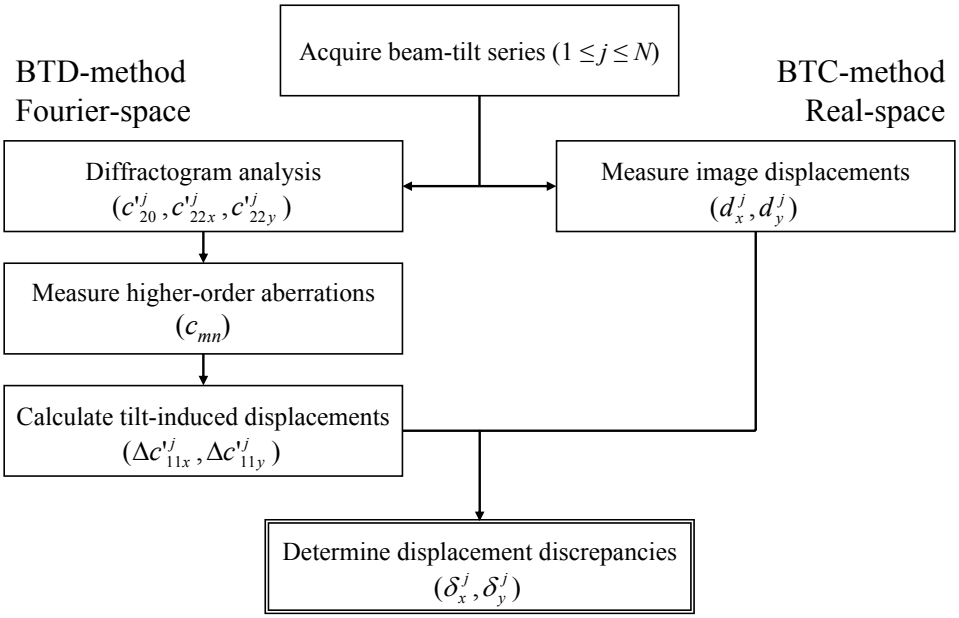



Figure 7.8: Procedure for the determination of displacement discrepancies by combination of the BTD-method and the BTC-method.

beam tilts and the real-space sampling have been calibrated precisely, the displacement discrepancies measure the drift of the object position.

### 7.5.3 A new method to measure the TEM-magnification

The magnification measurement investigated in this section is based on the unique dependence between displacement discrepancies $\vec{\delta}^{j}$ and the real-space sampling of the recorded images. It is observed that the assumption of a wrong real-space sampling by the user in both, the beam-tilt/diffractogram (BTD) method and the beam-tilt/cross-correlation (BTC) method, causes larger displacement discrepancies. The principle of the investigated method is therefore to identify the actual real-space sampling by *minimisation of displacement discrepancies* under variation of the assumed real-space sampling.

#### Explanation for the one-dimensional case

For simplification, we first consider the one-dimensional case. It is discussed in the following, how a change of the assumed real-space sampling from $\alpha$ to a different sampling rate $\alpha^{\dagger}$ affects the measurement results obtained by the BTC-method and the BTD-method and thus the displacement discrepancies.

The direct displacement measurement via cross correlation of the image intensity yields a pixel distance $\Delta p$ which corresponds to different real-space distances $d$ and $d^{\dagger}$, when two

different samplings are considered:

$$d = \alpha \cdot \Delta p, \tag{7.5.4}$$

$$d^\dagger = \alpha^\dagger \cdot \Delta p = d\,\frac{\alpha^\dagger}{\alpha}. \tag{7.5.5}$$

The measurement of the defocus $c_{20}$ from a one dimensional "diffractogram" will also yield a different value $c_{20}^\dagger$, since different Fourier-space samplings assign different spatial frequencies $g$ and $g^\dagger$ to the same diffractogram pixels:

$$g = \frac{1}{d} \;\Rightarrow\; g^\dagger = \frac{\alpha}{\alpha^\dagger} \cdot g \tag{7.5.6}$$

$$c_{20}\, g^2 = c_{20}^\dagger (g^\dagger)^2 \tag{7.5.7}$$

$$\Rightarrow c_{20}^\dagger = c_{20} \left(\frac{\alpha^\dagger}{\alpha}\right)^2 \tag{7.5.8}$$

The transformation (7.5.8) of the defocus $c_{20}$ ensures that the two respective aberration functions cause an identical diffractogram pattern. For $\alpha^\dagger < \alpha$ the pixels of one diffractogram cover a larger range of spatial frequencies. Consequently, a smaller defocus is required in order to describe the observed Thon-rings.
A change of the assumed real-space sampling from $\alpha$ to $\alpha^\dagger$ in the measurement procedures yields thus different displacement discrepancies $\delta^\dagger$. Using equations (7.5.5) and (7.5.8) to express the changed displacement discrepancy results in

$$\delta^\dagger = d^\dagger - (\Delta c_{11}')^\dagger = d^\dagger - c_{20}^\dagger t \tag{7.5.9}$$

$$= d \cdot \frac{\alpha^\dagger}{\alpha} - c_{20} \left(\frac{\alpha^\dagger}{\alpha}\right)^2 \cdot t. \tag{7.5.10}$$

Aberration measurements using the BTC-method and the BTD-method with one beam-tilt series are consistent, if the displacement discrepancies are minimum for all images. The consistent real-space sampling $\alpha^\dagger$ can be determined from the measurement results obtained with a wrong assumed real-space sampling $\alpha$ by minimisation of $|\delta^\dagger(\alpha^\dagger/\alpha)|$. This is equivalent to solving equation (7.5.10) with $\delta^\dagger(\alpha^\dagger/\alpha) = 0$. In the one-dimensional case one obtains either $\alpha^\dagger = 0$ which is rejected as a non-physical solution or $\alpha^\dagger = \alpha d/(c_{20}t)$.

### Extension to the two-dimensional case

In practice, the two-dimensional displacement discrepancies $\vec{\delta}^j$, $1 \le j \le N$ of a beam-tilt series have to be minimised simultaneously in order to achieve consistent results for aberration measurements with the BTC- and the BTD-method. The disagreement

between the two methods is therefore expressed with a displacement consistency value defined by

$$s_\delta = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\vec{\delta}^j)^2}. \tag{7.5.11}$$

The initial displacement consistency $s_\delta$ is calculated on the basis of image displacements $d_x^j$, $d_y^j$, effective defocus values $c_{20}'^j$ and effective twofold astigmatism values $c_{22x}'^j$, $c_{22y}'^j$ measured with a certain assumed real-space sampling. In order to evaluate the changed consistency for an arbitrary different sampling without a new aberration measurement, translations for all these observables have to be found analogous to equation (7.5.5) and equation (7.5.8).

The real-space sampling of two-dimensional images as introduced in section 5.3.1 in eqns. (5.3.3) and (5.3.4) is described by a $(2 \times 2)$ matrix

$$\mathbf{S}_\alpha = \begin{pmatrix} \alpha_{xx} & \alpha_{yx} \\ \alpha_{xy} & \alpha_{yy} \end{pmatrix}. \tag{7.5.12}$$

Analogous to the one-dimensional case, two different samplings $\mathbf{S}_\alpha$ and $\mathbf{S}_\alpha^\dagger$ are considered. The results $\vec{d}^j$ of displacement measurements obtained with the assumed sampling $\mathbf{S}_\alpha$ can be translated by the following formulas to displacements $\vec{d}^{\dagger j}$ which would have been measured assuming $\mathbf{S}_\alpha^\dagger$ :

$$\vec{d}^{\dagger j} = \frac{1}{D}\begin{pmatrix} \alpha_{xx}^\dagger\alpha_{yy} - \alpha_{yx}^\dagger\alpha_{xy} & \alpha_{xx}\alpha_{yx}^\dagger - \alpha_{xx}^\dagger\alpha_{yx} \\ \alpha_{xy}^\dagger\alpha_{yy} - \alpha_{xy}\alpha_{yy}^\dagger & \alpha_{xx}\alpha_{yy}^\dagger - \alpha_{yx}\alpha_{xy}^\dagger \end{pmatrix} \cdot \vec{d}^j \tag{7.5.13}$$

$$D = \det(\mathbf{S}_\alpha) = \alpha_{xx}\alpha_{yy} - \alpha_{xy}\alpha_{yx} \tag{7.5.14}$$

An adequate treatment of tilt-induced displacements in two dimensions requires a translation of all measured effective defocus values $c_{20}'^j$ and effective twofold astigmatism values $c_{22x}'^j$, $c_{22y}'^j$. According to the overview in figure 7.8, a new set of higher-order aberrations is determined from corrected effective second-order aberrations $c_{20}'^{\dagger j}$, $c_{22x}'^{\dagger j}$, and $c_{22y}'^{\dagger j}$, on the left branch via the BTD-method in order to obtain corrected tilt-induced displacements $\Delta c_{11x}'^{\dagger j}$ and $\Delta c_{11y}'^{\dagger j}$.

The derivation of the translation mechanism for the second order aberrations follows the same principle as the one-dimensional simplification above. It is obviously more complex, as two dimensions and three second-order aberration-parameters have to be considered. Accordingly, the effective aberration functions $\chi_{2n}'$ comprising the effective second-order aberrations should cause the same diffractogram pattern for two different Fourier-space

samplings:

$$\chi'_{2n} = \frac{\pi}{\lambda}\left[c'^{j}_{20}(g_x^2+g_y^2)+c'^{j}_{22x}(g_x^2-g_y^2)+2c'^{j}_{22y}g_xg_y\right] \tag{7.5.15}$$

$$= \frac{\pi}{\lambda}\left[c'^{\dagger j}_{20}\{(g_x^\dagger)^2+(g_y^\dagger)^2\}+c'^{\dagger j}_{22x}\{(g_x^\dagger)^2-(g_y^\dagger)^2\}+2c'^{\dagger j}_{22y}g_x^\dagger g_y^\dagger\right] \tag{7.5.16}$$

The transformation of the spatial frequency vector $\vec{g} \rightarrow \vec{g}^\dagger$ follows from the relations between the real-space basis vectors and the Fourier-space basis vectors denoted in (5.3.1), with

$$\vec{g}^\dagger = \frac{1}{D^\dagger}\begin{pmatrix} \alpha_{xx}\alpha^\dagger_{yy}-\alpha_{yx}\alpha^\dagger_{xy} & \alpha_{xy}\alpha^\dagger_{yy}-\alpha^\dagger_{xy}\alpha_{yy} \\ \alpha^\dagger_{xx}\alpha_{yx}-\alpha_{xx}\alpha^\dagger_{yx} & \alpha^\dagger_{xx}\alpha_{yy}-\alpha^\dagger_{yx}\alpha_{xy}\end{pmatrix}\cdot\vec{g}, \tag{7.5.17}$$

$$D^\dagger = \det(\mathbf{S}^\dagger_\alpha) = \alpha^\dagger_{xx}\alpha^\dagger_{yy}-\alpha^\dagger_{xy}\alpha^\dagger_{yx}. \tag{7.5.18}$$

Insertion of equation (7.5.17) into equation (7.5.16) and reformulation to the original $g_x$- and $g_y$-dependence of the left side of equation (7.5.15) yields three linear equations

$$\begin{pmatrix} c'^{j}_{20} \\ c'^{j}_{22x} \\ c'^{j}_{22y}\end{pmatrix} = \begin{pmatrix} M_{00} & M_{0x} & M_{0y} \\ M_{x0} & M_{xx} & M_{xy} \\ M_{y0} & M_{yx} & M_{yy}\end{pmatrix}\cdot\begin{pmatrix} c'^{\dagger j}_{20} \\ c'^{\dagger j}_{22x} \\ c'^{\dagger j}_{22y}\end{pmatrix}, \tag{7.5.19}$$

which connect the original to the transformed effective second-order aberrations. Explicit formulas for the coefficients $M_{ij}$ are denoted in appendix B.

### Realisation of the magnification measurement

According to the preliminary considerations about coordinate systems (section 5.3), the first row of the CCD-array defines the $x$-axis of the real-space coordinates. Hence, the real-space basis vector $\vec{\alpha}_x$ is fixed to the $x$-axis of the image. The other basis vector $\vec{\alpha}_y$ is linked to the physical direction of the CCD-columns which are not necessarily perpendicular to the rows. Different lengths $\alpha_x = |\vec{\alpha}_x| \neq \alpha_y = |\vec{\alpha}_y|$ account for different but constant pixel distances in both directions. The angle $\varphi_{xy} = \angle(\vec{\alpha}_y, \vec{\alpha}_y)$ describes a shear between the column and row direction as illustrated in figure 7.9.
The practical procedure for the minimisation of the displacement consistency $s_\delta$ starts with an initial aberration measurements via both methods under the assumption of a user-defined isotropic real-space sampling $\mathbf{S}_\alpha$ with

$$\alpha_{xx} = \alpha_{yy} = \alpha, \qquad \alpha_{xy} = \alpha_{yx} = 0, \tag{7.5.20}$$

where the two basis vectors are of equal length ($\alpha = \alpha_x = \alpha_y$) and orthogonal. In a subsequent iterative process, three parameters of a trial sampling-matrix $\mathbf{S}^\dagger_\alpha = (\vec{\alpha}^\dagger_x, \vec{\alpha}^\dagger_y)$

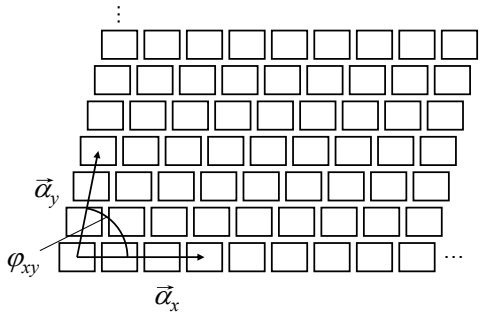



Figure 7.9: Anisotropic and sheared real-space sampling by non-equidistant distribution the CCD-camera pixels.

are varied. The trial parameters consider anisotropic pixel distances and a shear of the CCD-array along the $x$-axis with the shear angle $\pi/2 - \varphi^\dagger_{xy}$ and have thus the form

$$\alpha^\dagger_{xx} = \alpha^\dagger_x, \tag{7.5.21}$$

$$\alpha^\dagger_{yy} = \alpha^\dagger_y \cdot \cos(\pi/2 - \varphi^\dagger_{xy}), \tag{7.5.22}$$

$$\alpha^\dagger_{yx} = \alpha^\dagger_y \cdot \sin(\pi/2 - \varphi^\dagger_{xy}), \quad \text{and} \tag{7.5.23}$$

$$\alpha^\dagger_{xy} = 0, \tag{7.5.24}$$

With each iteration, new discrepancies are calculated corresponding to the trial sampling $\mathbf{S}^\dagger_\alpha$ by the following steps:

1. Transformation (7.5.13) of directly measured image displacements

$$d^j_x,\, d^j_y \rightarrow d^{\dagger j}_x, d^{\dagger j}_y,$$

2. Transformation (7.5.19) of effective second-order aberrations

$$c'^j_{20},\, c'^j_{22x},\, c'^j_{22y} \rightarrow c'^{\dagger j}_{20},\, c'^{\dagger j}_{22x},\, c'^{\dagger j}_{22y},$$

3. Determination of new higher-order aberrations $c^\dagger_{mn}$ from the above series of effective twofold-astigmatism parameters,

4. Calculation of tilt-induced displacements $\Delta c'^{\dagger j}_{11x}$, $\Delta c'^{\dagger j}_{11y}$ by the induction formula (C.2.5),(C.2.7) with the respective beam tilt $\vec{t}^{\,j}$, and

5. Calculation of displacement discrepancies $\delta^{\dagger j}_x = d^{\dagger j}_x - \Delta c'^{\dagger j}_{11x}$ and $\delta^{\dagger j}_y = d^{\dagger j}_y - \Delta c'^{\dagger j}_{11y}$.

6. Calculation of $s^\dagger_\delta$ by equation (7.5.11).

If the correct real-space sampling is found and all other involved parameters are calibrated precisely, the residual discrepancies reflect the temporal object drift. Further deviations are caused by the measurement errors of image displacement and effective twofold astigmatism.

For acquisition times in the regime of one minute, the object drift is approximately linear. Therefore, a linear drift model with $\vec{l}(j) = \vec{l}_1 \cdot (j-1)$, $(1 \leq j \leq N)$ is incorporated for the minimisation of the square root of the mean squared deviations

$$s_{\delta,l} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left|\vec{\delta}^{\dagger j} - \vec{l}(j)\right|^2}.$$

The residual displacement discrepancies $\vec{\delta}^{\dagger j} - \vec{l}(j)$ are minimised by varying the parameters $\alpha_x^\dagger$, $\alpha_y^\dagger$, and $\varphi_{xy}^\dagger$ with a Simplex Algorithm [103]. The parameter $\vec{l}_1$ measures the mean temporal object drift per image acquisition under the assumption that the time elapsing between the exposures is constant.

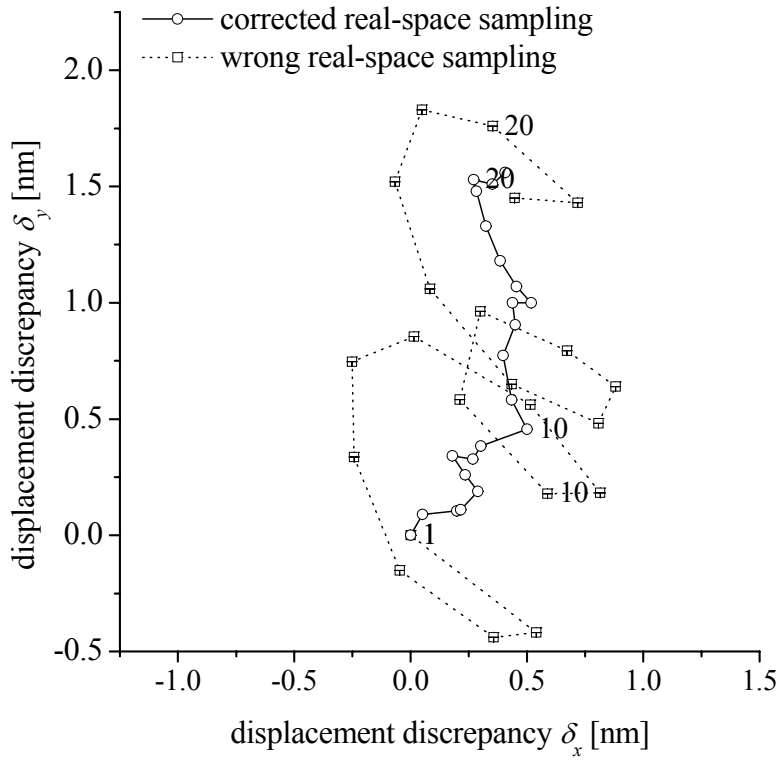



Figure 7.10: Displacement discrepancies for a wrong assumed isotropic real-space sampling of $\alpha = 0.110$ nm/pixel compared to the discrepancies obtained for a corrected sampling rate of $\alpha^\dagger = 0.119$ nm/pixel. The anisotropy of the corrected sampling is $|(\alpha_x^\dagger - \alpha_y^\dagger)/(\alpha_x^\dagger + \alpha_y^\dagger)| = 0.1$ %, with an angle of $\varphi_{xy}^\dagger = 92.6°$ between the basis vectors. The beam-tilt series used for the correction of the assumed real-space sampling was recorded with the FEI Titan 80-300 with thin amorphous carbon.

Figure 7.10 shows displacement discrepancies as measured from a beam-tilt series of amorphous carbon, recorded with the FEI Titan 80-300. Initially, an isotropic real-space sampling as in equation (7.5.20) was assumed. The resulting displacement discrepancies obviously follow a spiral path which is a superposition of a linear temporal object drift and displacements according to the applied enhanced beam-tilt tableau depicted in figure 5.6. Such a drift would be quite unnatural for an object drift alone. The residual discrepancies observed with the iteratively corrected anisotropic sampling as in the equations (7.5.21) - (7.5.24) yield a minimum consistency value of $s_{\delta,l} = 0.10$ nm in contrast to $s_{\delta,l} = 0.33$ nm for the initial discrepancies. Accordingly, the discrepancies obtained for the corrected sampling are more or less linear.

The consistency value $s_{\delta,l}$ accounts for all non-linearities in the temporal object drift, as well as for the deficiency of the applied model to describe the observed image displacements as a superposition of temporal drift and tilt-induced displacements. Consequently, the ratios between the residual deviations $s^j_{\delta,l} = \left|\vec{\delta}^{\dagger j} - \vec{l}(j)\right|$ and the modulus $\left|\Delta c'^{\dagger j}_{11}\right|$ of the induced displacements can be used to estimate the error of the corrected mean real-space sampling $\alpha^\dagger = (|\vec{\alpha}^\dagger_x| + |\vec{\alpha}^\dagger_y|)/2$. A coarse, relative error estimate $\Delta\alpha^\dagger/\alpha^\dagger$ is obtained by

$$\frac{\Delta\alpha^\dagger}{\alpha^\dagger} = \sqrt{\frac{1}{N'}\sum_{j=1,|\vec{t}_j|\neq 0}^{N}\frac{s^j_{\delta,l}}{\left|\Delta c'^{\dagger j}_{11}\right|}}, \tag{7.5.25}$$

$$N' = \sum_{j=1,|\vec{t}_j|\neq 0}^{N} 1. \tag{7.5.26}$$

The summations in the above equations exclude all images $j$ recorded without intentional beam tilt. Smaller beam tilts cause smaller tilt-induced displacements and consequently reduce the accuracy $\Delta\alpha^\dagger$ of the measurement result. Larger residual deviations $s^j_{\delta,l}$ increase the relative error of the mean real-space sampling. They are caused by statistical measurement errors, non-linear drift components or other systematic influences on the displacements which have not been considered in the model.

### Discussion of the experimental results

The displacement discrepancies shown in figure 7.10 are measured for a beam-tilt series recorded with the nominal magnification $M = 380\,000$ of the FEI Titan 80-300 and a CCD-binning factor of 4. The mean real-space sampling $\alpha^\dagger = 0.119$ ($\pm 0.003$) nm/pixel minimises $s_{\delta,l}$. This value differs from the result $\alpha = 0.1135$ ($\pm 0.0003$) nm/pixel as measured with the standard method described in section 5.3.1 by more than the estimated

error. Furthermore, the precision of the standard method is about one order of magnitude better. Similar relative errors and discrepancies have been observed for many different beam-tilt series recorded with two different instruments (see table 7.2). In most of the examples the method described in this section overestimates the real-space sampling.

| | $\alpha^\dagger$ [nm/pix] | $\Delta\alpha^\dagger$ [nm/pix] | anisotropy [%] |
|---|---|---|---|
| FEI Titan 80-300 | 0.1194 | 0.0024 | 0.50 |
| ($M = 380$k, 4 $\times$-binning) | 0.1165 | 0.0036 | -0.42 |
| | 0.1185 | 0.0034 | 0.00 |
| | 0.1176 | 0.0021 | 0.26 |
| | 0.1157 | 0.0025 | -1.20 |
| | 0.1167 | 0.0015 | -0.51 |
| standard method: | 0.1135 | 0.0003 | 0.52 |
| Philips CM200C | 0.0919 | 0.0025 | -1.06 |
| ($M = 130$k, 4 $\times$-binning) | 0.0922 | 0.0014 | -0.55 |
| | 0.0896 | 0.0019 | 1.06 |
| | 0.0919 | 0.0020 | -1.31 |
| | 0.0922 | 0.0015 | -0.25 |
| | 0.0906 | 0.0035 | 0.19 |
| | 0.0922 | 0.0016 | -0.63 |
| | 0.0880 | 0.0041 | -0.61 |
| standard method: | 0.0899 | 0.0002 | -0.43 |

Table 7.2: Measured real-space samplings from the displacement discrepancies for different beam-tilt series compared to the result of the standard method. The anisotropy measures the relative difference between the pixel distances in rows and columns of the CCD-camera.

In spite of the advantage that this new method can be applied almost for free with the results of an aberration measurement, it is by far not accurate enough for HRTEM. However, the observation of inconsistencies between the BTD- and the BTC-method, and the discrepancy to the standard method can be used to signalise a poor alignment of the microscope or a large temporal drift of the object position.

# Chapter 8

# Application of the aberration measurement procedure

Besides the main goal to achieve the demanded precision for aberration control in the Sub-Ångström regime, the aberration-measurement procedure worked out in this thesis was also designed in the view of practical aspects such as robustness, speed, and user-friendliness. The developed numerical procedures have been implemented in a software package named ATLAS (*Accurate Treatment of Lens Aberrations and Stability*). In order to demonstrate its capability in practical use with modern transmission electron microscopes, the ATLAS software package is tested in different applications.

## 8.1 Application of the ATLAS software package

The ATLAS software package realises the concepts and techniques described in the previous chapters with four separate software applications. The core procedures of these applications are based on a step-wise diffractogram analysis discussed in chapter 6. These four applications have been implemented using the programming language Fortran 90/95 in compilations for personal computers with the most widely spread 32-bit Microsoft Windows NT and XP operating systems. The separate applications are in detail dedicated to the following tasks:

1. `ATLAS` – Higher-order aberration measurement using the diffractogram method,
2. `ATLASstab` – Stability benchmark for object drift, defocus, and twofold astigmatism,
3. `mcalib` – Calibration of the magnification (real-space sampling),

4. `btcalib` – Calibration of beam tilt for tilt tableaus.

The investigations presented in this section mainly concern the applications 1 and 2 of this list which apply the novel algorithms for the extraction of defocus and twofold astigmatism from single diffractograms. The calibration tools 3 and 4 are mandatory prerequisites for accurate aberration measurements. Detailed descriptions of standard methods for the calibration of magnification and tilt tableaus have been given in section 5.3. In `mcalib` and `btcalib` these methods are realised in the form of semi-automatic and precise calibration procedures.
The ATLAS software package runs independently from the electron microscope. All necessary images have to be provided as data files. For the future it is hoped to integrate the ATLAS procedures directly into the operation of modern transmission electron microscopes. Procedures for an automatic acquisition of beam-tilt series and for a feedback to aberration correctors have still to be developed in cooperation with the manufacturers of microscopes and aberration-correction devices.

### 8.1.1 Higher-order aberration measurement

The complete procedure required to measure higher-order aberrations from a beam-tilt series of a thin amorphous object was described in detail in three previous chapters. Figure 8.1 summarises the required workflow for such aberration measurements as realised with `ATLAS`. The aberration measurement for a beam-tilt series of approximately 20 images takes about one minute.
Parameter files are used to specify the properties of the electron microscope such as accelerating voltage, real-space sampling and data-format of the recorded TEM images. The parameters define also options for the processing of the acquired series of images like for example the treatment of peaks of high intensity in the diffractograms. Separate parameter files specify beam-tilt parameters, the list of requested higher-order aberrations and directives for the output of the measurement results. Additionally, a database containing prepared diffractogram motifs has to be supplied for the coarse pattern recognition (CPR) as discussed in section 6.6.
After calculation of the diffractograms from the images, the measurement procedure begins with a pre-processing of the diffractograms. This pre-processing includes an optional detection and removal of diffractogram peaks as well as an extraction of additive and multiplicative low-frequency diffractogram content. The pre-processing provides the necessary data for a coarse pattern recognition (CPR). Any image where the pre-processing

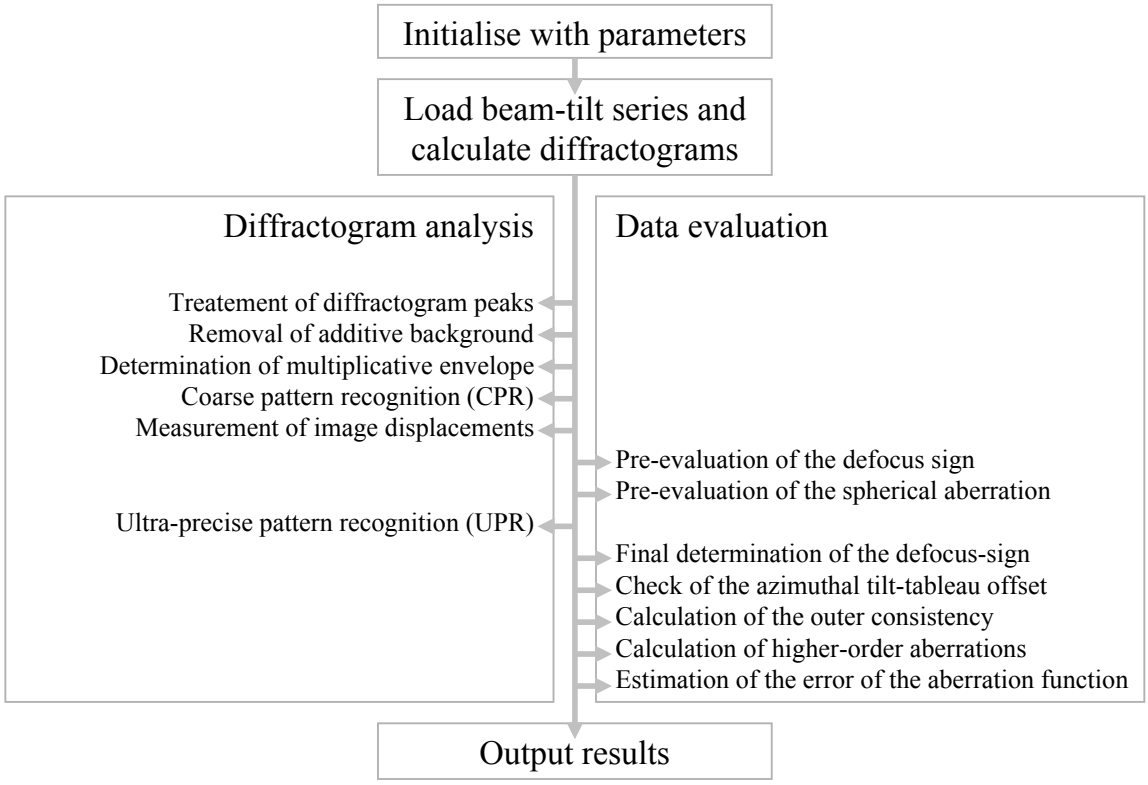



Figure 8.1: Workflow of higher-order aberration measurement realised with ATLAS.

fails or yields unexpected results are consequentially excluded from the further analysis. Between the CPR and the final ultra-precise pattern recognition (UPR), ATLAS measures image displacements by means of cross correlations between the images and determines higher-order aberrations from the effective twofold astigmatism values obtained by the CPR. On the basis of these data, the procedure described in section 7.5.1 tests for the correctness of the assumed defocus sign and corrects it if necessary. By virtue of these pre-evaluations the UPR-algorithm starts with a correct defocus sign and a rough estimate for the spherical aberration.

The results obtained by the ultra-precise pattern recognition, the measured image displacements and the beam-tilt parameters constitute the dataset for the final evaluations. After a second check for the correctness of the defocus sign, the higher-order aberrations are determined, and the outer consistency $s_{\mathrm{all}}$ is calculated as described in section 7.3. By means of the outer consistency, ATLAS computes the reliable error estimate $\bar{\sigma}_\chi(g_{\max})$ for the measured aberration function at the user-specified information limit $g_{\max}$ of the microscope. This error estimate allows for a quick quality assessment of the current measurement. All results of the diffractogram analysis and the data evaluations are saved in readable text files or visualised in images.

A beam-tilt series recorded with the FEI Titan 80-300 of thin amorphous $SrTiO_3$ is used as an example in the present and throughout the following sections. During the acquisition of the 18 images, the beam tilts have been varied according to a (1/12/4/1)-standard tableau with 24 mrad maximum tilt modulus as shown in figure 5.6.

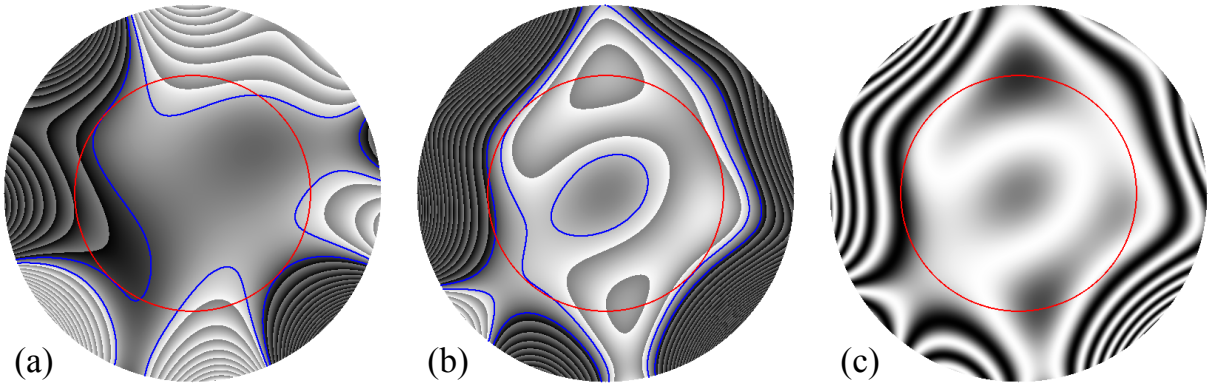



Figure 8.2: (a) Phase plate comprising all non-circular aberrations measured by ATLAS for a tilt series of amorphous $SrTiO_3$ recorded with the FEI Titan 80-300. (b) Complete (circular and non-circular) phase plate of the measured aberration function $\chi_S(\vec{g})$ at Scherzer-defocus (see section 1.3.4). (c) Coherent contrast-transfer function $\sin[\chi_S(\vec{g})]$ at Scherzer-defocus in a grey scale. The coherent contrast-transfer function depicts positive contrast transfer in white and negative contrast transfer in black. Grey denotes transfer gaps, i.e. spatial frequencies where no contrast is transferred.

*Phase plates* as shown in figure 8.2a and figure 8.2b visualise the final results of the aberration measurement in addition to the numerical values saved in the text files. A phase plate displays the values of the aberration function for a circular area of the Fourier space around the origin $\vec{g} = 0$. The maximum spatial frequency of all phase plates shown in this chapter is 20 $nm^{-1}$. The colour encoding assigns bright colours (grey to white) to positive values and dark colours (grey to black) to negative values of the aberration function. Contrast steps appear, where the aberration function is an integer multiple of $\pm\pi/2$.
A red circle in the shown phase plates marks the information limit $g_{max} = 12.5$ $nm^{-1}$ of the FEI Titan 80-300, whereas the blue curves follow the $\pm\pi/4$-contours of the aberration function. The occurrence of a $\pi/4$-contour inside the red circle in figure 8.2a indicates a too large value of the residual non-circular aberration function. In this case, the prior aberration correction by hardware was not sufficient to provide a direct interpretation of the linear image contrast at the information limit. The circular aberrations like defocus $c_{20}$ and spherical aberration $c_{40}$ are not included in the phase plate in figure 8.2a, because both aberrations are tuned by intention to specific larger values and would hide the undesired appearance of non-circular aberrations in the phase plate.
Figure 8.2b shows the phase-plate for the aberration function $\chi_S(\vec{g})$ including all mea-

sured aberrations at Scherzer-defocus as defined by equation 1.3.11. Several transfer gaps (grey) and contrast reversals (black) appear in the respective coherent contrast-transfer function $\sin[\chi_S(\vec{g})]$ shown in figure 8.2c.

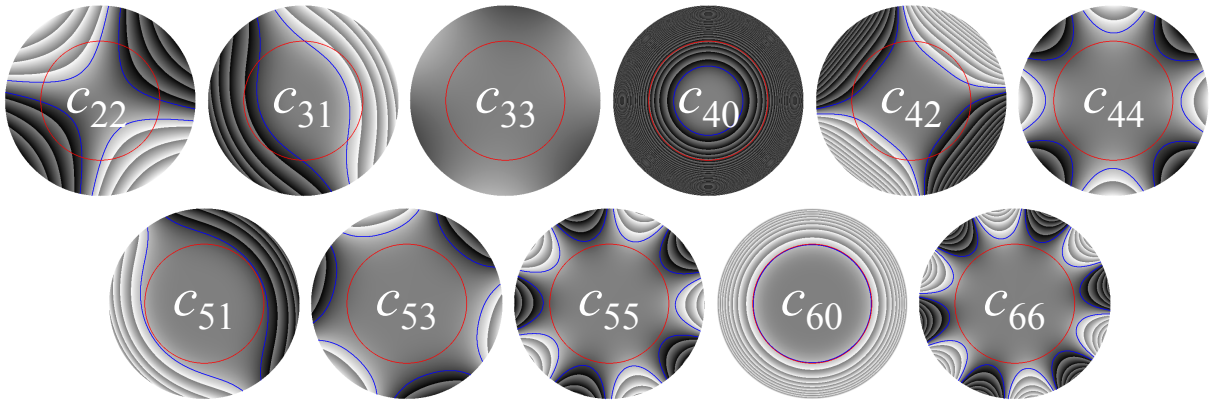



Figure 8.3: Phase-plates of all measured single aberrations, with the residual aberrations in the top row which are correctable by hardware, and the non-correctable aberrations in the bottom row.

Separate single-aberration phase plates directly signalise which of the correctable higher-order aberrations need further adjustments. From the examples shown in figure 8.3 obviously the axial coma of the third order $c_{31}$ and the star-aberration $c_{42}$ dominate the non-circular aberrations. Also some of the non-correctable aberrations of the orders 5 and 6 come alarmingly close to the $\pi/4$-limit.

### Comparison to the existing software

The example for a measurement of higher-order aberrations given in section 7.3 reveals not only a drastic improvement of the consistency compared to the existing standard software, but also obtains quite different values for the measured aberration coefficients (see table 7.1). Also for the example series analysed in this section, `ATLAS` measures different higher-order aberrations as listed in table 8.1. The aberration measurement by `ATLAS` yields an outer consistency $s_{\text{all}}$ which is more than one order of magnitude better than that of the standard software.

The aberration function was measured by `ATLAS` with an error of $\bar{\sigma}_\chi(g_{\max}) = 0.11\,\pi$ at $g_{\max} = 12.5\ \text{nm}^{-1}$ with an outer consistency of $s_{\text{all}} = 0.226$ nm. This measurement result thus over-accomplishes the multi-aberration tolerance limit $\bar{\sigma}_\chi(g_{\max}) \leq \pi/4$ required for a reliable aberration control in the Sub-Ångström regime. The improvement of the consistency in comparison to the standard software is in particular achieved by a more accurate pattern recognition. Further improvements have to be ascribed first, to the consideration

| | ATLAS | | | standard software | | |
|---|---|---|---|---|---|---|
| | $c_{mn}$ | $(\pm\sigma_{mn})$ | $\varphi_{mn}$ | $c_{mn}$ | $(\pm\sigma_{mn})$ | $\varphi_{mn}$ |
| $c_{20}$ | −308.0 | (±0.3) nm | – | −308.1 | (±2.4) nm | – |
| $\dot{c}_{20}$ | −0.18 | $(\pm 0.05)\,\frac{\text{nm}}{\text{img}}$ | – | – | | – |
| $c_{22}$ | 2.52 | (±0.10) nm | −53.6° | 3.87 | (±5.0) nm | −63.5° |
| $\dot{c}_{22}$ | 0.035 | $(\pm 0.012)\,\frac{\text{nm}}{\text{img}}$ | −65.9° | – | | – |
| $c_{31}$ | 118.8 | (±10.5) nm | 23.2° | 55.3 | (±40.1) nm | −112.6° |
| $c_{33}$ | 10.5 | (±6.3) nm | 23.27° | 42.3 | (±27.9) nm | 51.1° |
| $c_{40}$ | −23.4 | (±1.0) $\mu$m | – | −16.5 | (±0.62) $\mu$m | – |
| $c_{42}$ | 7.69 | (±0.18) $\mu$m | 51.5° | 5.91 | (±1.2) $\mu$m | 50.0° |
| $c_{44}$ | 1.61 | (±0.04) $\mu$m | −0.3° | 1.37 | (±0.81) $\mu$m | −4.9° |
| $c_{51}$ | 117.0 | (±11.6) $\mu$m | −149.8° | – | | – |
| $c_{53}$ | 46.8 | (±11.1) $\mu$m | −12.8° | – | | – |
| $c_{55}$ | 81.1 | (±1.4) $\mu$m | −0.6° | 57.9 | (±26.4) $\mu$m | −33.2° |
| $c_{60}$ | 7.1 | (±1.1) mm | – | – | | – |
| $c_{66}$ | 3.61 | (±0.05) mm | −20.1° | – | | – |
| $s_{\text{all}}$ | 0.226 nm | | | 5.603 nm | | |

Table 8.1: Higher-order aberrations measured from the same tilt series with ATLAS and the standard software used for the alignment of spherical-aberration correctors. The aberration measurement by ATLAS yields an outer consistency $s_{\text{all}}$ which is more than one order of magnitude better than that of the standard software.

of an extended set of measured higher-order aberrations, second, to the consideration of a calibrated tilt tableau, and third, to the neglect of the effective defocus data by ATLAS as discussed in chapter 5.

A severe difference between the respective measured aberration functions with ATLAS and the existing software is visualised by the *difference phase-plate* in figure 8.4. The discrepancy is so large that the $\pi/4$-contour of the aberration-function difference crosses inside the circle of the information limit. Since the consistency obtained with the procedures developed in this work is significantly better, its measurement result must be considered as more reliable.

### Choosing an appropriate set of aberrations

The consistency of an aberration measurement depends on the set of high-order aberrations contained in the measured aberration function. It is evident, that the "true" shape of

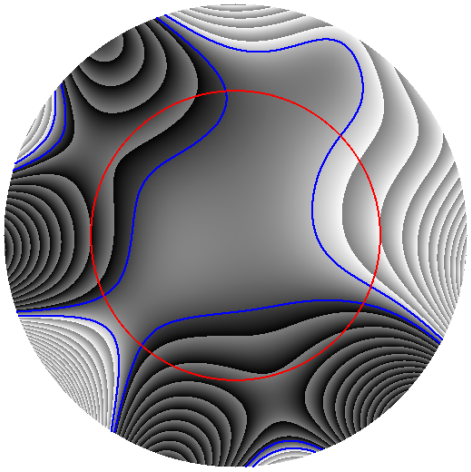


Figure 8.4: Difference $\chi_2(\vec{g}) - \chi_1(\vec{g})$ between the aberration function $\chi_1$ measured with `ATLAS` and the aberration function $\chi_2$ measured with the standard software depicted as difference phase-plate. The difference reaches absolute values larger than $\pi$ at the information limit $g_{\max} = 12.5$ nm$^{-1}$ which is marked by a red circle.

the aberration function can be represented better when more aberration terms are taken into account. However, the number of determinable aberration coefficients is limited for a given beam-tilt series. As discussed in section 5.4.1, higher-order aberrations can be excluded from the measurement in order to keep an overdetermination ratio of at least $R_{\text{det}} = 1.5$ and yield thus significant results for the remaining aberrations. A feasible rule which decides whether a specific aberration should be included in the solution vector is proposed by Saxton [67]. Hence, an aberration should be measured, if its inclusion halves the residual discrepancies, while it can be neglected if the discrepancies change by only a small fraction. The residual discrepancies with the diffractogram method are registered by the outer consistency $s_{\text{all}}$ as discussed in section 7.3.

The decision about which aberration can be neglected depends also on the aberration-correction strategy for which the measurement result is used. This decision especially concerns aberrations of equal rotational symmetry, i.e. aberrations with identical second index $n$ of the aberration coefficients $c_{mn}$. The corresponding terms in the aberration function can compensate for a certain range of spatial frequencies, similar to compensation between spherical aberration and defocus in the optimum contrast conditions discussed in section 1.3.4.

The compensation between two of such aberrations of different orders is illustrated for three different scenarios in figure 8.5 with values as listed in table 8.1. The top row shows the combined phase plates when both aberrations are measured. The phase plates in the bottom row depict only the lower-order aberrations of the above pairs, when the respective

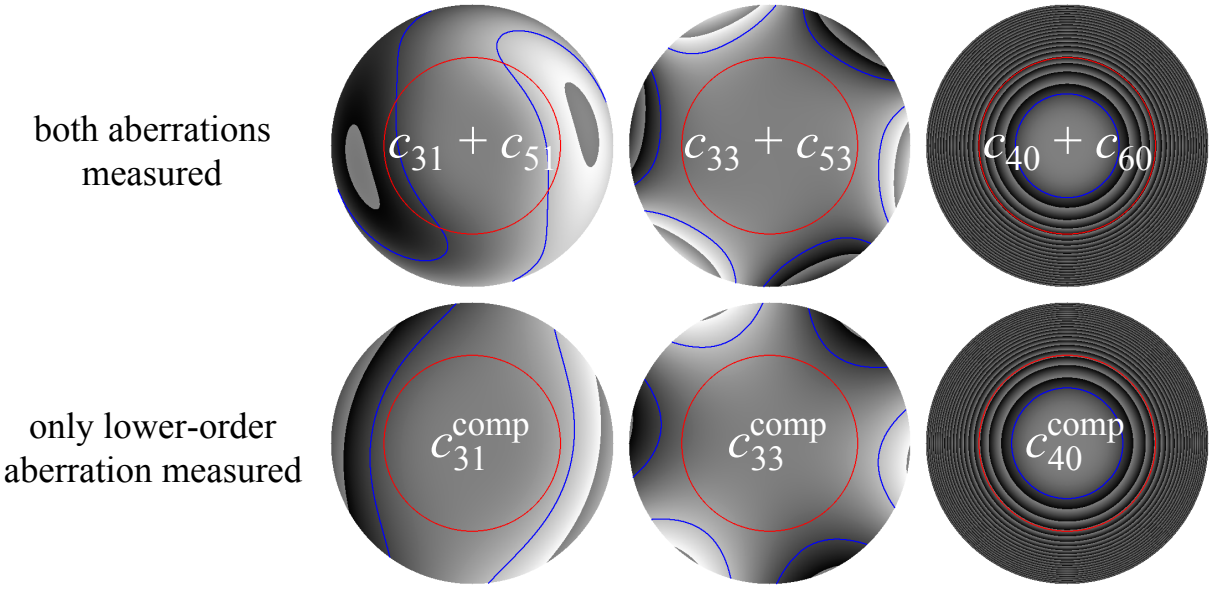



Figure 8.5: Combined phase shifts measured for two aberrations $c_{mn}$ and $c_{(m+2)n}$ of equal rotational symmetry depicted as phase plates in the top row. The phase-plates in the bottom row show the phase shift due to the lower-order aberration $c_{mn}^{\text{comp}}$ which is measured when the respective higher-order aberration is neglected.

higher-order aberration is excluded from the aberration measurement. The latter case is marked here by denoting the aberrations with $c_{mn}^{\text{comp}}$. The phase plates in the bottom row of figure 8.5 are quite similar to the respective phase plates in the top row, but of course not equal.

From the observations on how the choice of considered aberrations affects the result of an aberration measurement made in the present section, one can draw two important conclusions:

1. An aberration measurement which is used to determine the aberration function for an a-posteriori software correction should consider all relevant aberrations.

2. An aberration measurement used for a correction by hardware should neglect all not correctable aberrations which are of equal rotational symmetry as correctable lower-order aberrations.

The first conclusion can be explained straightforward: Since a-posteriori aberration correction as discussed in section 2.4.3 can account for any aberration, the aberration function should be rendered as exactly as possible. Among the examples in figure 8.5, in particular the axial-coma $c_{31}$ is misjudged when the $5^{\text{th}}$-order axial coma $c_{51}$ is neglected. Although the observed phase shift due to $c_{31}^{\text{comp}}$ is well below the $\pi/4$-limit, the summation of the

separately, and also more exactly measured $c_{31}$- and $c_{51}$-phase plates shows much higher aberrations at the information limit.

By following the second conclusion, one exploits an automatic compensation mechanism. When for example both, $c_{31}$ and $c_{51}$ are measured but only $c_{31}$ can be corrected, a correction of $c_{31}$ would remove the phase shift due to $c_{31}$ completely, whereas the phase shift due to $c_{51}$ would remain as is. On the other hand, by neglecting $c_{51}$ in the aberration measurement, the measured $c_{31}^{\text{comp}}$ describes a compromise phase-shift which fits best to the "true" combined $(c_{31}+c_{51})$-phase-shift. A hardware correction now adjusts $c_{31}$ in such a way that it partially compensates $c_{51}$. In this way, it is guaranteed that the combined effect of both aberrations, $c_{31}$ and $c_{51}$ is corrected as far as possible.

Given a restricted set of hardware-correction possibilities, it is therefore important to decide whether an aberration should be explicitly determined or not. Hereby, the inherent automatic compensation proceeds approximately according to the following mechanism: For a compensation between two non-circular aberrations $c_{mn}$ and $c_{(m+2)n}$, the respective azimuth parameters $\varphi_{mn}$ and $\varphi_{(m+2)n}$ are oriented exactly opposite, such that

$$\varphi_{mn} = \varphi_{(m+2)n} + \frac{\pi}{n}, \quad n > 0. \tag{8.1.1}$$

Analogous to the compensation principle behind Scherzer's and Lichte's defocus discussed in section 1.3.4, here the modulus of $c_{mn}$ is inherently adjusted to a certain value in order to provide a compensation of $c_{(m+2)n}$ for a maximum range of spatial frequencies. When condition (8.1.1) is fulfilled, the aberration function for the two aberration terms is described by

$$\chi_{\text{comb}}(g) = \frac{2\pi}{\lambda}\left(-\frac{1}{m}(\lambda g)^m c_{mn} + \frac{1}{m+2}(\lambda g)^{m+2} c_{(m+2)n}\right). \tag{8.1.2}$$

If only one intermediate extremum of $\chi_{\text{comb}}(g_{\text{extr}})$ is present within the spatial-frequency range range $0 < g_{\text{extr}} < g_{\max}$, the combined aberration function $\chi_{\text{comb}}(g)$ comprises minimum deviations from zero up to the information limit $g_{\max}$. The extremum $\chi_{\text{comb}}(g_{\text{extr}})$ should be of opposite sign and of equal modulus to the value $\chi_{\text{comb}}(g_{\max})$ as expressed by

$$\chi_{\text{comb}}(g_{\text{extr}}) = -\chi_{\text{comb}}(g_{\max}), \tag{8.1.3}$$

$$\text{with} \quad g_{\text{extr}} = \sqrt{\frac{c_{mn}}{\lambda^2 c_{(m+2)n}}}. \tag{8.1.4}$$

The above formulation for an optimum compensation introduces fewest aberrations below the information limit. In case of the two aberration-pairs of the third and fifth order $(c_{31}, c_{51})$ and $(c_{33}, c_{53})$, the approximation

$$c_{3n,\text{opt}} \approx 0.52 \cdot c_{5n}(\lambda g_{\max})^2 \tag{8.1.5}$$

solves equation (8.1.3) with sufficient accuracy. One obtains thus a modulus value for the third-order aberration $c_{3n}$ which provides an optimum compensation of the fifth-order aberration $c_{5n}$.

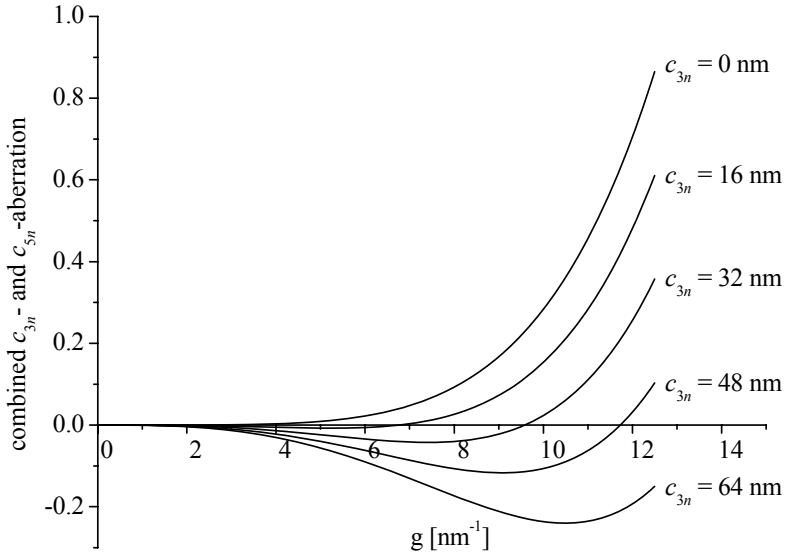



Figure 8.6: Compensation of anti-parallel aberrations of the 3$^{\text{rd}}$ order ($c_{3n}$) and the 5$^{\text{th}}$ order ($c_{5n}$) with equal rotational symmetry ($\lambda = 1.969$ pm, $c_{5n} = 150$ $\mu$m.) Optimum compensation up to the information limit $g_{\max} = 12.5$ nm$^{-1}$ as achieved with $c_{3n} = 48$ nm.

The compensation of the two terms in the aberration function of equation (8.1.2) is plotted for five different choices of $c_{3n}$ in figure (8.6). The choice of $c_{3n} = 48$ nm comes closest to the optimum solution $c_{3n,\text{opt}} = 47.25$ $\mu$m with $\lambda = 1.969$ pm, $g_{\max} = 12.5$ nm$^{-1}$, and $c_{5n} = 150$ $\mu$m.

The two conditions (8.1.1) and (8.1.3) explain how some non-correctable aberrations of higher orders can be indirectly compensated with a tunable lower-order aberration. This compensation is achieved automatically by hardware correction of the lower-order aberration when the higher-order aberration is neglected with the aberration measurement. However, it must be stated that this compensation approach makes only sense, if the higher-order aberration of the aberration pair can be measured with statistic significance.

### Discrepancy due to uncalibrated beam-tilts

Aberration measurements which consider beam-tilt parameters without prior calibration do not reflect the true lens aberrations. For example, when considering larger tilt angles than actually applied, smaller higher-order aberrations are obtained for a given set of effective astigmatism values. The numerical investigations in section 5.4.4 demonstrate

that a wrong assumption about the beam tilts results in strong deviations from the correct aberrations especially for large values of the aberration coefficients.
In case of the (1/12/4/1)-standard tableau with a nominal maximum tilt modulus of 24 mrad, which is applied for the exemplary beam-tilt series discussed here, the calibrated tilts (see fig. 5.6b) have a mean modulus of 22.1 mrad at the outer "circle" and vary between 21.0 mrad and 23.3 mrad. Separate evaluations of higher-order aberrations from identical effective astigmatism values, using the calibrated tilts on the one hand and the nominal tilts on the other hand, yield different values $c_{mn}^{\text{calib.}}$ and $c_{mn}^{\text{nominal}}$ for the aberration coefficients. The differences $\Delta c_{mn} = c_{mn}^{\text{calib.}} - c_{mn}^{\text{nominal}}$ as obtained for the exemplary tilt series mentioned in the previous sections are listed in table 8.2 and visualised as difference phase-plates in figure 8.7.

| | $\Delta c_{mn}$ | $\Delta\varphi_{mn}$ | $\lvert\Delta c_{mn}/c_{mn}\rvert$ |
|---|---|---|---|
| $\Delta c_{20}$ | 0.23 nm | – | 0.0007 |
| $\Delta c_{22}$ | 0.12 nm | 16.1° | 0.048 |
| $\Delta c_{31}$ | 21.0 nm | 116.5° | 0.177 |
| $\Delta c_{33}$ | 6.37 nm | 143.9° | 0.607 |
| $\Delta c_{40}$ | −4.21 $\mu$m | – | 0.180 |
| $\Delta c_{42}$ | 1.66 $\mu$m | 41.3° | 0.216 |
| $\Delta c_{44}$ | 0.90 $\mu$m | 3.0° | 0.559 |
| $\Delta c_{51}$ | 18.4 $\mu$m | −121.6° | 0.157 |
| $\Delta c_{53}$ | 3.00 $\mu$m | 12.0° | 0.064 |
| $\Delta c_{55}$ | 16.9 $\mu$m | 1.4° | 0.208 |
| $\Delta c_{60}$ | 2.43 mm | – | 0.342 |
| $\Delta c_{66}$ | 1.03 mm | −107.8° | 0.285 |

Table 8.2: Absolute and relative differences between higher-order aberration measurements considering the nominal tilt-tableau and the calibrated tilt-tableau. The relative differences in the right column are calculated with respect to the results of the aberration measurement with `ATLAS` listed on the left side of table 8.1

Most of the higher-order aberrations differ by more than 10 %. Especially for the spherical aberrations $c_{40}$ and $c_{60}$, the misjudgement due to uncalibrated beam tilts becomes critical for the adjustment of optimum contrast-transfer conditions. Defocus and twofold astigmatism stay constant, because their contributions to the effective astigmatism and defocus is independent of the beam tilt. The total difference phase-plate on the left side of figure 8.7 exceeds the $\pi/4$-limit along the axis in the image plane, where calibrated

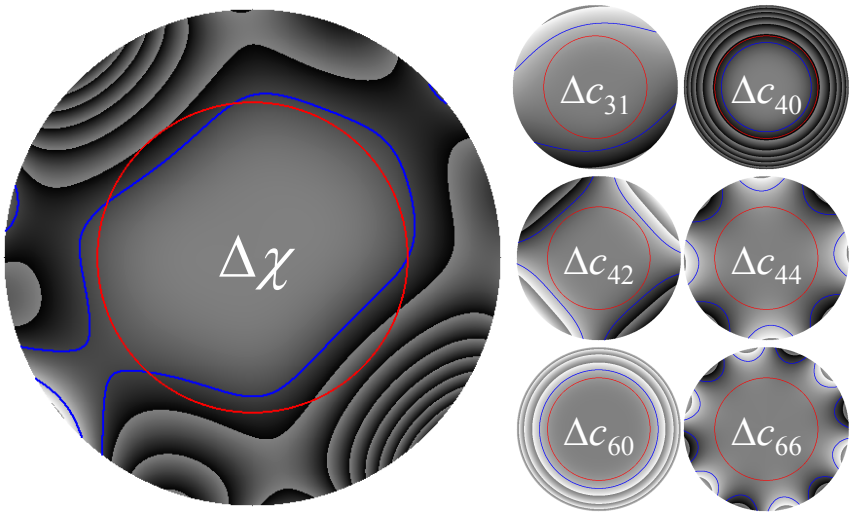



Figure 8.7: Difference phase-plates comparing aberration-measurement results obtained with and without calibrated beam tilts. The large phase plate plots the difference between two complete aberration functions as measured by `ATLAS` with a calibrated and with a nominal tilt tableau. The smaller phase plates depict the single aberrations with the most severe discrepancies.

and uncalibrated beam-tilts deviate strongest from another. These experimental observations substantiate the demand for the application of accurately calibrated tilts in order to obtain reliable aberration-measurement results.

### Planarity of aberrations

The dependence of the aberration function on the position $\vec{r} = (x, y)$ in the image plane is discussed in the literature in two different models. Based on the investigations how a non-isoplanar illumination affects the imaging conditions by Hanszen *et al.* [108], Rosenfeld describes off-axis aberrations in the frame of ray aberrations or local beam tilts [109]. In this work, an approach by Hawkes and Kasper [63] is followed by using an expansion of the total wave aberration $\chi$ in polynomial terms of the Fourier-space vector $\vec{g}$ and of the distance $\vec{r}$ from the centre in the image plane. Accordingly, the axial aberrations $c_{mn}$ in the aberration function (2.2.8) used so far, are replaced by expansions, such as

$$c_{mn} \quad \rightarrow \quad \sum_{o=0}^{N} \sum_{p=0}^{N} (r^*)^o (r)^p c_{mnop}. \tag{8.1.6}$$

A complex notation is used for the real-space vector $\vec{r}$ with $r = x + \mathrm{i}\, y$, and for the off-axis aberration coefficients with $c_{mnop} = c_{mnop,x} + \mathrm{i} c_{mnop,y}$. One obtains thus an aberration function which depends on both, Fourier-space vectors $\vec{g}$ and real-space vectors $\vec{r}$. Unfortunately, in cases, where the dependence on the real-space variables is relevant,

the imaging theory on the basis of Fourier optics discussed in chapter 1 is not valid. It is therefore of great importance that the off-axis aberrations $c_{mnop}$ for $o \neq 0 \vee p \neq 0$ are negligibly small over a large area of the image such that techniques like image simulation and focal-series reconstruction can be applied.

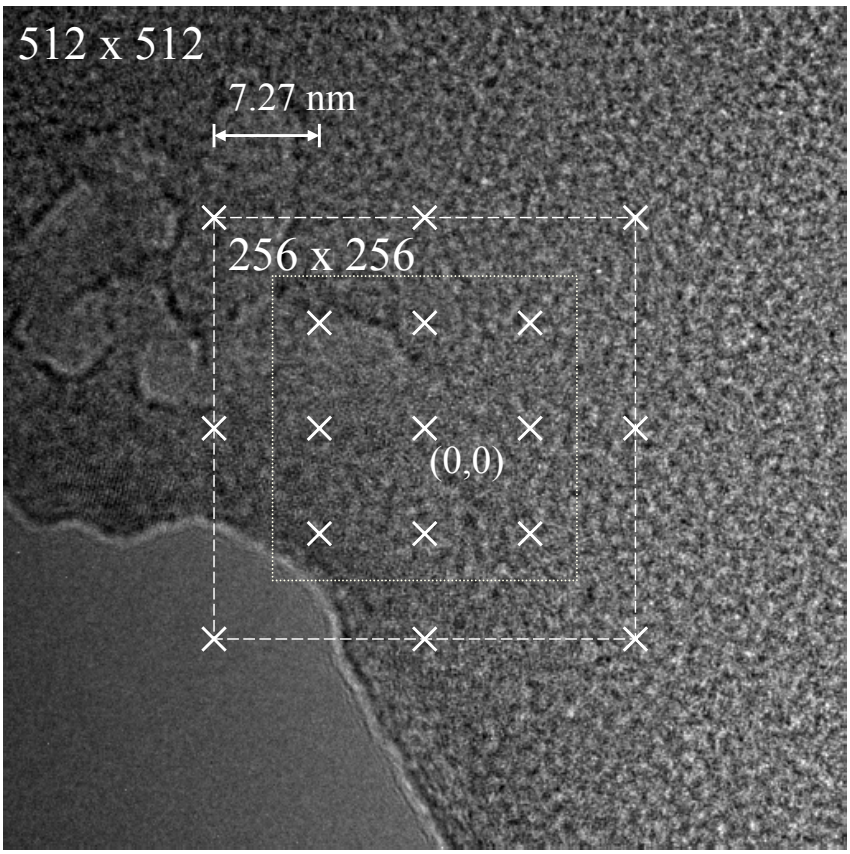



Figure 8.8: First image (size $512 \times 512$) of a beam-tilt series recorded with the FEI Titan 80-300 showing amorphous $SrTiO_3$ at the edge of the object. The crosses mark the central positions for sub-images of a smaller size ($256 \times 256$ pixel) as marked by a dashed rectangle. The dotted rectangle corresponds to the area which is imaged with a magnification of 1 million with 2048×2048 pixels as used for Sub-Ångström microscopy.

Off-axis aberrations show up in the differences between aberration measurements from several displaced sub-images cut out from the images of a beam-tilt series. For an experimental determination of the off-axis aberrations in images recorded with the FEI Titan 80-300, the already investigated exemplary beam-tilt series is divided into 17 sub-series. Each sub-series comprises again 18 images of size $256 \times 256$ pixels. In the imaging mode selected for aberration measurements (magnification $M = 380\,000$), the original image covers a total area of $58 \times 58$ nm$^2$ sampled by $512 \times 512$ pixels. Figure 8.8 illustrates the chosen distribution of the sub-image offsets $(x_i, y_i)$ from the centre $(x_0, y_0)$ of the

original image in equidistant steps of 7.27 nm. In the experiment the extreme position at the bottom-left had to be ignored, since too few amorphous material is contained in the respective area.

Only for defocus $c_{20}$, twofold astigmatism $c_{22}$, axial coma $c_{31}$, and star-aberration $c_{42}$, the differences $\Delta c_{mn,i} = c_{mn,i} - c_{mn,0}$ between the aberrations measured for the displaced sub-image $(i)$ and for the central sub-image $(0)$ are significantly larger than the respective measurement errors. The diagrams in figure 8.9b and 8.10 depict $\Delta c_{mn,i}$ for $c_{22}$, $c_{31}$, and $c_{42}$ in form of arrows beginning at the sub-image offsets. The scalar defocus differences are displayed in figure 8.9a in form of a 3d-plot where the defocus varies along the $z$-axis.

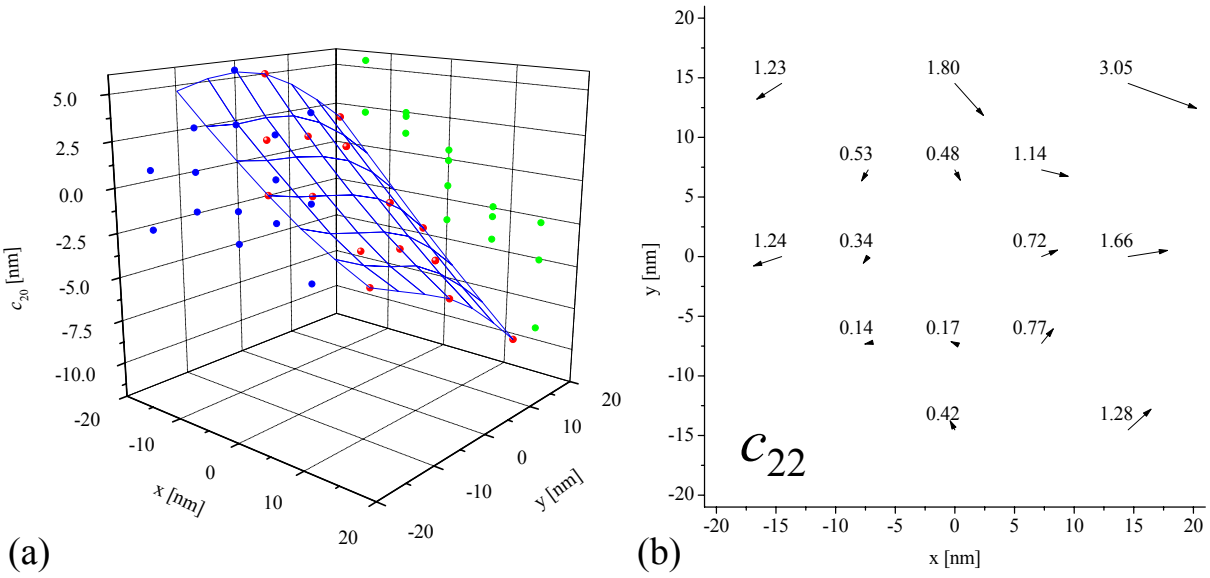



Figure 8.9: Deviations of (a) defocus and (b) twofold astigmatism measured off-axis from the image centre. The green and blue dots in (a) represent the projections of the three-dimensional defocus plot on the $xz$- and $yz$-planes. The arrows in (b) start at the centres of the off-axis sub-images, while their length (denoted next to the arrows) corresponds to the respective deviation from the axial aberration value measured at the image centre in nanometres.

Since only deviations for 15 sub-images are evaluated, the off-axis expansion of the aberration coefficients according to equation (8.1.6) are limited to the first order. This expansion includes only a constant and two linear complex coefficients $c_{mn00}$, $c_{mn01}$, and $c_{mn01}$. For numerical evaluation, the expansion in complex-number notation of equation (8.1.6) with

$$c_{mn}(r, r^*) = c_{mn00} + c_{mn01} \cdot r + c_{mn10} \cdot r^* + \ldots \tag{8.1.7}$$

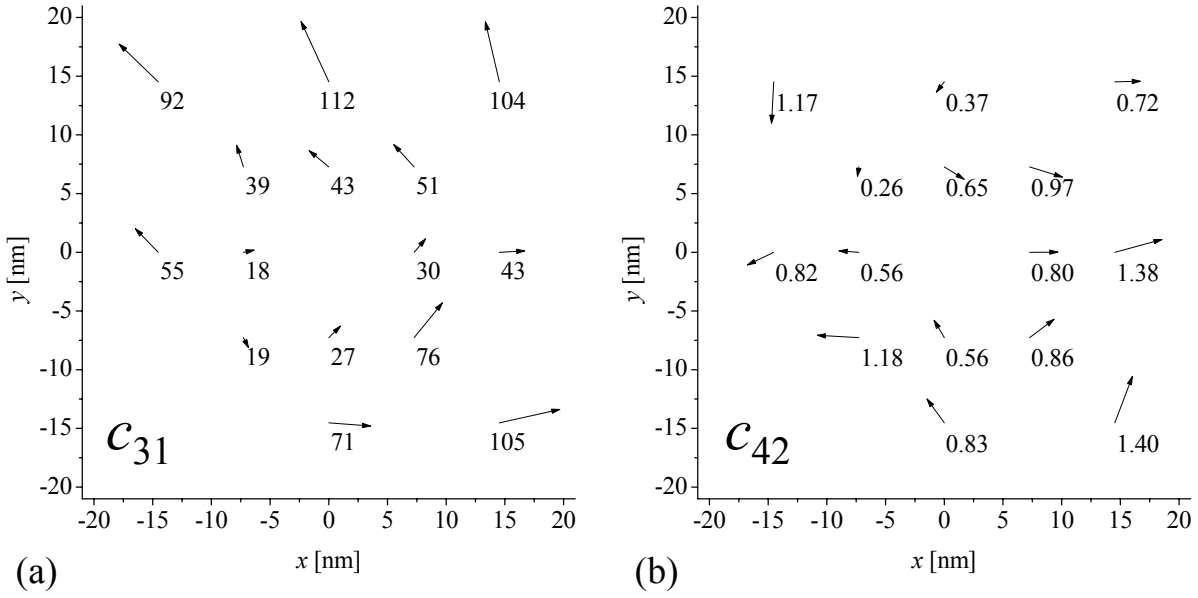



Figure 8.10: Deviations of (a) axial coma and (b) star-aberration measured off-axis from the image centre. The arrows start at the centres of the off-axis sub-images, while their length (denoted next to each arrow for (a) in nm and for (b) in $\mu$m) corresponds to the respective deviation from the axial aberration value measured at the image centre.

is re-formulated as real-valued equations with

$$c_{mn,x}(x,y) = c_{mnx0} + c_{mnxx} \cdot x + c_{mnxy} \cdot y, \tag{8.1.8}$$

$$c_{mn,y}(x,y) = c_{mny0} + c_{mnyx} \cdot x + c_{mnyy} \cdot y. \tag{8.1.9}$$

Least-squares fits of the linear models (8.1.8) and (8.1.9) to the measured off-axis differences yield the dimension-less linear coefficients as listed in table 8.3.

| | $c_{mnxx}$ | $c_{mnxy}$ | $c_{mnyx}$ | $c_{mnyy}$ |
|---|---|---|---|---|
| $c_{20}$: | $-0.37$ ($\pm 0.03$) | $-0.17$ ($\pm 0.03$) | – | – |
| $c_{22}$: | 0.10 ($\pm 0.01$) | 0.05 ($\pm 0.01$) | 0.01 ($\pm 0.01$) | $-0.05$ ($\pm 0.01$) |
| $c_{31}$: | 1.9 ($\pm 0.4$) | $-3.8$ ($\pm 0.4$) | 1.2 ($\pm 1.1$) | 3.1 ($\pm 1.1$) |
| $c_{42}$: | 62 ($\pm 10$) | 26 ($\pm 10$) | 30 ($\pm 5$) | $-42$ ($\pm 5$) |

Table 8.3: Measured linear off-axis aberration coefficients for defocus $c_{20}$, twofold astigmatism $c_{22}$, coma $c_{31}$, and star aberration $c_{42}$.

The off-axis defocus coefficients mainly reflect the three-dimensional shape of the specimen. Since the sub-images still cover a quite large area, the single off-axis defocus values basically trace the mean surface height. According to the measured values $c_{20xx} = -0.37$

and $c_{20xy} = -0.17$ the surface normal is tilted by about 22° to the $z$-axis. The interpolated blue mesh of the measured defocus distribution in figure 8.9a also indicates a slight curvature of the surface along the $y$-axis.
For experiments with Sub-Ångström resolution the Titan microscope is typically operated with a magnification such that the image covers an area of $21 \times 21\text{nm}^2$. For such images, the off-axis deviations of the twofold-astigmatism $c_{22}$ and the coma $c_{31}$ from the axial aberration values measured at the image centre exceed the individual $\pi/4$-limits denoted in table 2.2 in the image corners. The respective off-axis deviations of the star-aberration $c_{42}$ is smaller than its individual $\pi/4$-limit. The assumed planarity of the aberrations is thus not confirmed for all aberrations by this experiment.

### 8.1.2 Aberration-stability benchmark

The novel ultra-precise defocus and astigmatism-measurement procedure for single diffractograms, for the first time, allows one to observe temporal fluctuations of these aberrations with magnitudes below the nanometre scale on the time scale of single exposures. By means of sequential application of the diffractogram analysis for a series of continuously recorded TEM-images of amorphous material, `ATLASstab` provides the first benchmark for the optical stability of electron microscopes. The stability of the object position is simultaneously monitored by displacement measurements between successive images.

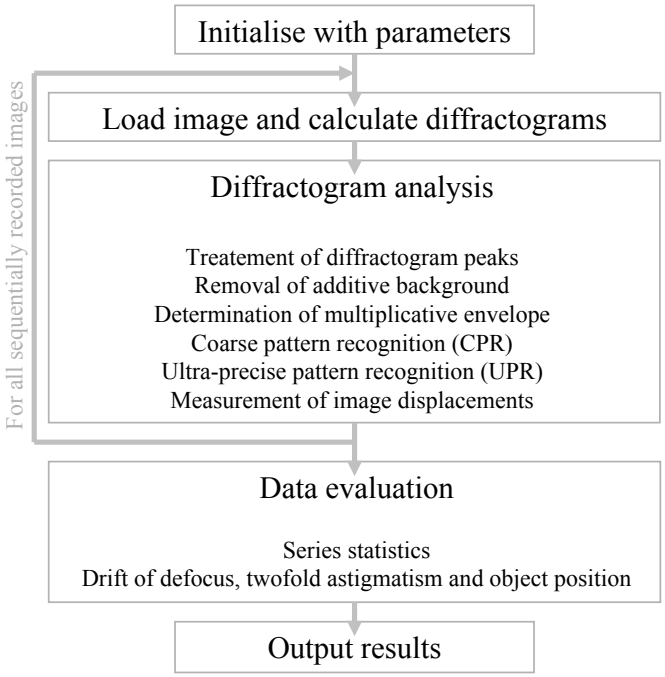



Figure 8.11: Workflow diagram for the aberration-stability benchmark `ATLASstab`.

The realisation of `ATLASstab`, as illustrated in figure 8.11, shares the modules for the pre-processing of diffractograms, for the coarse pattern recognition (CPR), for the ultra-precise pattern recognition (UPR) and for the displacement measurement with its sister-application `ATLAS`. It is therefore initialised by the same parameters. Since the spherical aberration cannot be determined by the benchmark application itself, the present $C_S$-value has to be measured in advance and specified as a parameter.
Specific results of the statistical analysis obtained from the recorded images will be discussed in the following. This discussion concerns the impact of the mean object-drift velocity $\bar{u}$ and the total long-term drift of the twofold astigmatism on the validation of stability acceptance tests.

### Observation of the object drift

The object-drift acceptance for high-resolution images is directly linked to the applied sampling rate. For a comfortable representation of the finest details, digital images should approximately sample the distance of the information limit $d_{\text{info}}$ in 8 pixels. Two pixels per resolvable distance are at minimum required due to the Shannon-Whitaker-criterion of sampling [106]. In order to adequately resolve also non-linear contrast fringes including half-spacings of $d_{\text{info}}/2$, the sampling rate has to be increased by a further factor of two. The remaining factor of two provides a comfortable over-sampling in order to avoid a damping of the image contrast due to the modulation transfer function of the CCD-camera. The desired over-sampling is retained, when the object drift does not exceed one half of the pixel distance. This constriction can be expressed by the mean drift velocity $\bar{u}$ and the exposure time $\Delta t_{\text{r}}$ with

$$\bar{u} < \bar{u}_{\max} = \frac{1}{2} \cdot \frac{d_{\text{info}}}{8\Delta t_{\text{r}}}. \tag{8.1.10}$$

In the final evaluation of the observed object drift, `ATLASstab` compares the measured mean drift velocity $\bar{u}$ to the drift acceptance-limit $\bar{u}_{\max}$ specified in equation (8.1.10). A weaker criterion is obtained when neglecting the non-linear terms in the image contrast. The tolerable mean drift velocity for purely linear image contrast is given by

$$\bar{u}_{\max,\text{lin}} = \frac{1}{2} \cdot \frac{d_{\text{info}}}{4\Delta t_{\text{r}}} = 2\,\bar{u}_{\max}. \tag{8.1.11}$$

With a typical exposure time of $\Delta t_{\text{r}} = 1$ s for HRTEM measurements, the drift acceptance-limit is given by $\bar{u}_{\max} = 0.008$ nm/s (4.8 Å/min) for the Philips CM200C with $d_{\text{info}} = 0.13$ nm, and by $\bar{u}_{\max} = 0.005$ nm/s (3.0 Å/min) for the FEI Titan 80-300 with $d_{\text{info}} = 0.08$ nm.

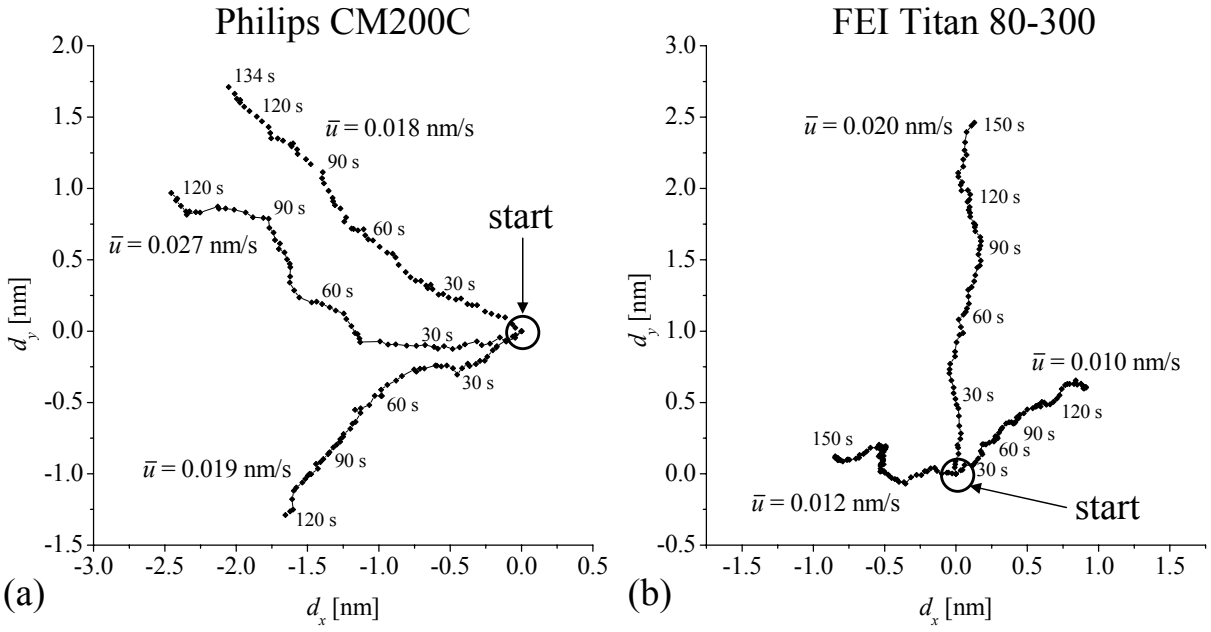



Figure 8.12: Object-drift benchmark for (a) the Philips CM200C and (b) the FEI Titan 80-300 for three, continually recorded image series. The measured relative displacements between successive images are added such that the drift path of the object is reconstructed.

All exemplary drift measurements shown in figure 8.12 do not fulfil the drift-limit $\bar{u} < \bar{u}_{\max}$ for the respective microscope. In its present state, the FEI Titan 80-300 hardly satisfies the weaker condition $\bar{u} < \bar{u}_{\max,\mathrm{lin}}$ in one of the given examples which is the smallest drift velocity observed so far with this instrument. Thus, the mechanical stability of the object-holder has to be improved for this microscope in order to fulfil the drift-tolerance criterion demanded by the advertised Sub-Ångström resolution.

### Observation of the defocus drift

Object drift and instabilities of the electron optics both contribute to defocus variations during the acquisition of a series of images. The vertical object-drift component typically dominates the defocus variations. The results of time-resolved defocus measurements as plotted in figure 8.13 display a more or less linear drift which is superposed by short-term ripples. The observed linear defocus drift is probably caused by the object drift, whereas the ripples reveal temporal optical instabilities due to variations of the objective-lens current and of the accelerating voltage. Due to the inherent object dependence, it remains impossible to define a trustworthy acceptance limit for a defocus benchmark on the basis of such measurements.

The defocus data obtained by `ATLASstab` can be used to determine the defocus step-size

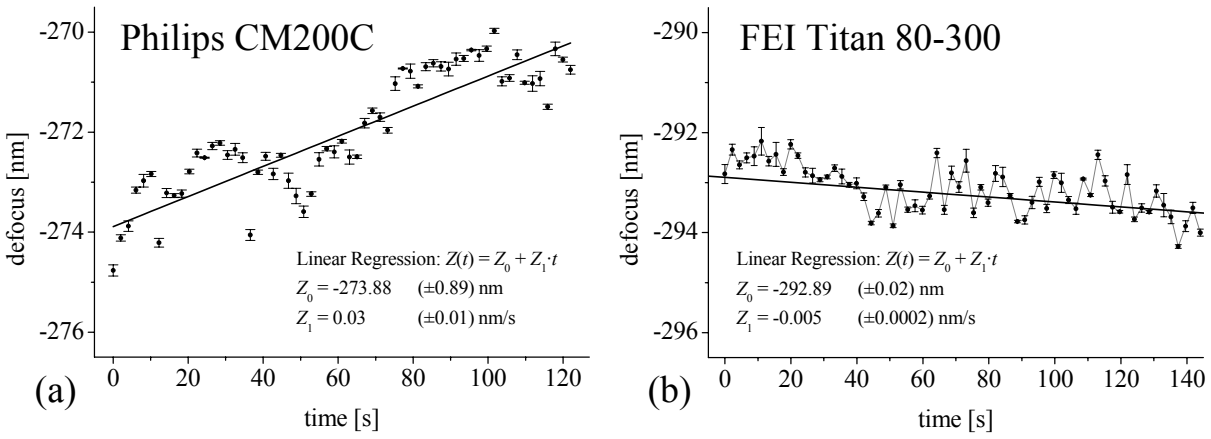



Figure 8.13: Temporal defocus variations for (a) Philips CM200C and (b) FEI Titan 80-300.

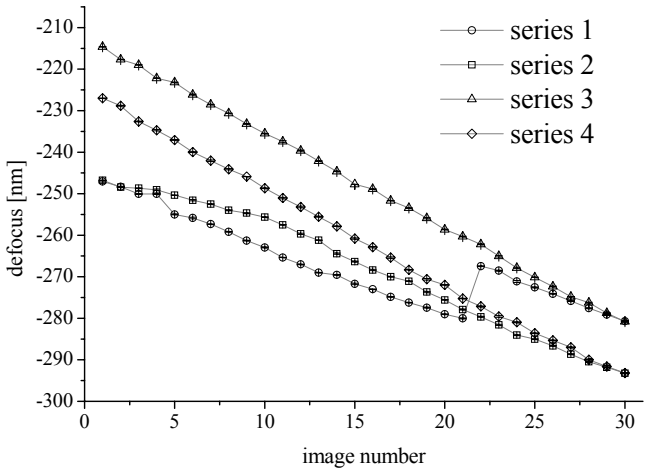



Figure 8.14: Application of `ATLASstab` for the measurement of the step-size in defocus series. All four series have been recorded for amorphous tantalum with the same series-acquisition procedure of the FEI Titan 80-300.

for a defocus series prior to the reconstruction of the exit-plane wavefunction. Figure (8.14) shows the results of defocus measurements from four focal series recorded with the FEI Titan 80-300. Separate linear regressions to series 3 and 4 yield a mean defocus step-size of $-2.32$ ($\pm 0.01$) nm. The other two defocus series are strongly disturbed due to object drift. Additionally, a large jump of more than 10 nm appears in series 1. Such jumps can be caused by acoustic or mechanical disturbances from the surrounding. The series 1 and 2 are examples which cannot be used for a focal-series reconstruction.

### Observation of the twofold-astigmatism drift

Already for very small variations $\Delta c_{2n}$ of defocus- or twofold-astigmatism parameters, the two respective terms in the aberration function exceed the $\pi/4$-limit at the information-limit $g_{\max}$. This aberration-drift tolerance limit is expressed by

$$\max[\Delta\chi_{2n}(g_{\max},)] = \pi\lambda g_{\max}^2|\Delta c_{2n}| < \frac{\pi}{4} \tag{8.1.12}$$

$$\Rightarrow |\Delta c_{2n}| < |\Delta c_{2n}|_{\max} = \frac{1}{4\,\lambda\,g_{\max}^2} \tag{8.1.13}$$

According to equation (8.1.13), a variation of the second-order aberrations by more than $|\Delta c_{2n}|_{\max} = 0.81$ nm destroys the adjusted contrast transfer at $g_{\max} = 12.5$ nm$^{-1}$ for the FEI Titan 80-300 with $\lambda = 1.969$ pm. The tolerance limit for the Philips CM200C is $|\Delta c_{2n}|_{\max} = 1.66$ nm with an information limit of $g_{\max} = 7.7$ nm$^{-1}$ and a wavelength of $\lambda = 2.508$ pm.

The variation of the twofold astigmatism can be taken as a measure for the optical stability of the electron microscope. Besides fluctuations of the multiple lens currents, also changes in the temperature of the lens cooling-water influence $c_{22}$ [104]. Similar to the random ripples observed in the defocus series (figure 8.13), the twofold astigmatism exhibits a random-walk-like behaviour away from the starting position as shown in figure 8.15.

At the end of each benchmark run, `ATLASstab` specifies the time window during which $|\Delta c_{22}|$ remains significantly smaller than $|\Delta c_{2n}|_{\max}$. This time window is just larger than the series acquisition-time of $\Delta t_{\text{ser}} = 144$ s in the shown cases. The deviation of the twofold astigmatism from the starting point almost reaches the tolerance limits $|\Delta c_{2n}|_{\max} = 1.66$ nm for the Philips CM200C and $|\Delta c_{2n}|_{\max} = 0.81$ nm for the FEI Titan 80-300 which are marked by the circles in figure 8.15.

The astigmatism-stability tests for both instruments have shown an optical stability which retains the once adjusted state of the microscope for about two minutes. The observed variations of defocus and twofold astigmatism are smaller than the precision of previous aberration-measurement techniques. A benchmark of the aberration stability is thus currently only possible with the diffractogram analysis developed in this work.

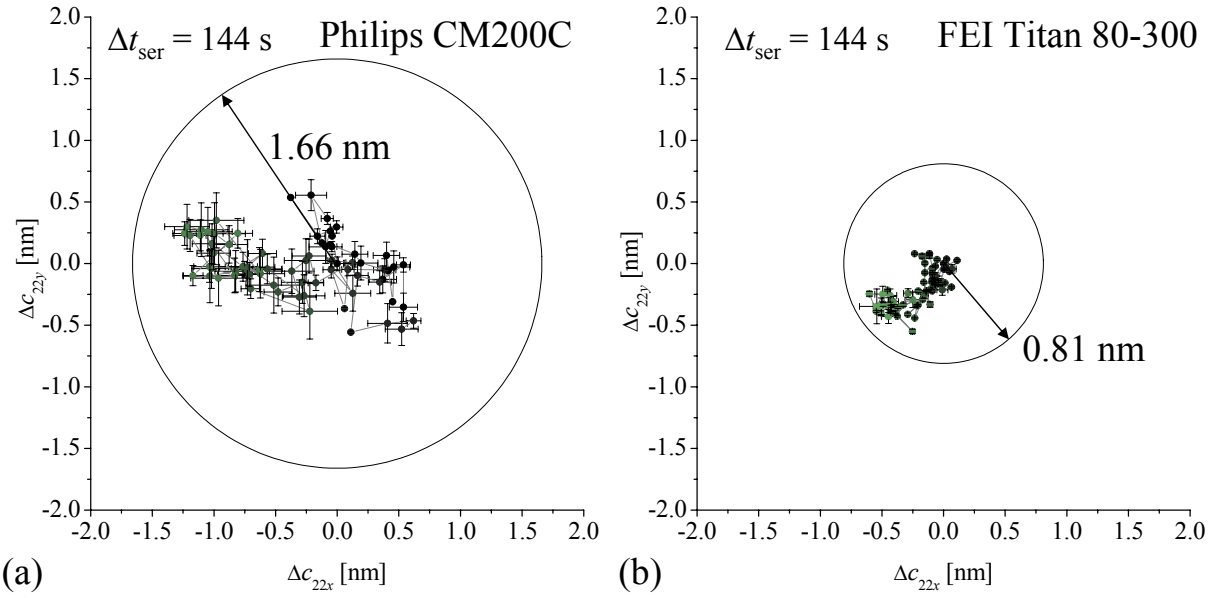



Figure 8.15: Time-resolved variations of the twofold astigmatism relative to the beginning of the series for (a) the Philips CM200C and (b) the FEI Titan 80-300. The circles, drawn around the starting point, mark the astigmatism-drift tolerances for the respective microscope. With advancing time, the positional data is plotted here in brighter colours starting with black for the first image of the series

## 8.2 Investigations on the structure of $Cs_x(Nb,W)_5O_{14}$

The structure of $Cs_x(Nb,W)_5O_{14}$ was first characterized in 1993 by Lundberg and Sundberg [110]. The $x$ in the compositional formula denotes a variable relative amount of caesium. Besides determining the approximate lattice parameters ($a$ = 2.71 nm, $b$ = 2.16 nm, $c$ = 0.394 nm) from selected area electron diffraction, Lundberg and Sundberg succeeded to uncover the basic structural features of the compound from HRTEM images recorded in projection along the short $c$-axis.

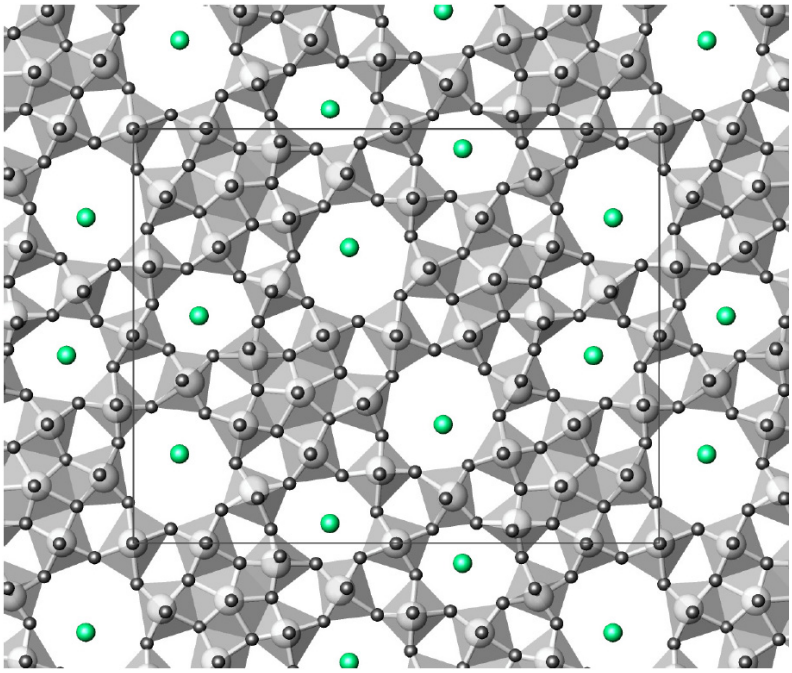


Figure 8.16: Structure model of $Cs_x(Nb,W)_5O_{14}$ in [0 0 1]-projection as obtained from Rietveld-refinement with X-ray powder data (reproduced from [111]). The heavy niobium and tungsten atoms (large grey shaded spheres) form the structural basis of $MO_6$ octahedral and $MO_7$ pentagonal coordination shells of the oxygen atoms (small dark spheres). The Cs atoms are located inside the 6- and 7-sided tunnels. The positions of the oxygen-atoms have been proposed by Lundberg and Sundberg [110]. The rectangle marks the extent of the projected unit cell with the $a$-axis in horizontal and the $b$-axis in vertical direction.

On the basis of the positions of the heavy atoms obtained from HRTEM, Lundberg and Sundberg proposed a tentative structure model in the space group $P2_12_12$ as shown in figure 8.16. This model comprises analogous structural principles as related oxides. Whereas the 6-sided tunnels in the structure were assumed to be fully occupied by caesium atoms, it remains unclear whether caesium atoms are also present in the larger 7-sided tunnels. Furthermore, the structure was described to contain a large number of local defects and disorder. A later analysis of this earlier published structure model suggests that the bulk

structure of $Cs_x(Nb,W)_5O_{14}$ matches also with higher symmetry and can be described by the space group *Pbam* [111]. In the investigated projection along the *c*-axis, the structure shows a symmetry according to the plane group *p2gg* [112].
These previous studies, based on standard X-ray and HRTEM methods for structure determination, failed to locate the positions of the oxygen atoms mainly because the investigated compound is currently only available in form of a powder sample and due to extreme difficulties to detect light atoms in the presence of heavier ones [113, 114]. The scientific interest in $Cs_x(Nb,W)_5O_{14}$ compounds specifically focusses on the heptagonal channels which are the active centres in the almost iso-structural catalyst MoVNbTeO [115]. These channels are visible in the structure model in figure 8.16 with seven niobium/tungsten columns surrounding the caesium atoms.

Due to the strong interaction of electrons with matter, HRTEM offers the possibility to investigate the projected atomic structure in objects of extremely small volumes down to the length scale of a few nanometres. HRTEM is therefore a proper means to assist in the structural analysis of very small crystals present in powder samples.
Just recently oxygen columns have been resolved and quantitatively analysed in the proximity of heavier atom columns in perovskite ceramics by HRTEM with a small negative spherical aberration ($C_S$) and the technique of focal-series reconstruction (FSR) [48, 49, 117, 118]. These techniques are also applied in the present investigation in order to confirm experimentally the oxygen positions in the structure model of $Cs_x(Nb,W)_5O_{14}$ in the projection along the short *c*-axis. In this projection all oxygen positions are revealed in pure oxygen columns and in mixed columns at the niobium-tungsten positions. In order to determine atomic positions with an as high as possible accuracy, the aberration control developed in this work is applied.

### 8.2.1 Experiment

For investigation by HRTEM, a pulverised powder sample with the average composition $Cs_{0.44}Nb_{2.54}W_{2.46}O_{14}$ was dispersed in ethanol and subsequently dried on a standard holey-carbon-coated copper-grid[1].
The object was investigated with the spherical-aberration corrected Philips CM200C. In a lower magnification mode several particles were found in orientations with the *c*-axis almost parallel to the direction of the incident electron beam. Only very few of the powder crystals were located at the edge of holes in the carbon coating. Particles fully supported

[1] The sample was provided by Dr. T.E. Weirich (GFE, Rheinisch-Westfälische Technische Hochschule, Aachen).

by the amorphous carbon can not be investigated due to the overlapping carbon signal. The magnification for the high-resolution images was chosen with a sampling rate of 20.1 pm per pixel. With a binning factor of two, 1024×1024-pixel images were acquired with the charge-coupled device (CCD) camera. This image discretisation is well below the Nyquist criterion $1/(2g_{\max}) = 65$ pm with respect to twice the information limit $g_{\max} = 7.7$ nm$^{-1}$ of the instrument [47].

Prior to recording images from the specimen region of interest, the electron-microscope's hexapole correction-elements were adjusted coarsely. The integrated standard software for the alignment of the spherical-aberration corrector was used to correct the aberrations. For the aberration measurement, a standard tilt tableau of 18 images was recorded from the amorphous carbon support film with a maximum tilt angle of 18 mrad. According to these measurements the magnitudes of the residual aberration coefficients were reduced to $c_{22} < 10$ nm for the twofold astigmatism, to $c_{31} < 100$ nm for the axial coma, and to $c_{33} < 100$ nm for the threefold astigmatism.

The measured spherical aberration of $C_S = -40$ $\mu$m is close to the optimum condition for negative spherical-aberration imaging at the maximum resolution of the instrument. The bright atom contrast achieved with the small negative spherical aberration helps to locate a properly thin specimen position and additionally increases the contrast of the oxygen columns in the vicinity of the columns containing heavier atoms [48].

A repetition of the aberration measurement with the procedures developed in this work reveals an axial coma of $c_{31} = 145$ ($\pm 9$) nm, $\varphi_{31} = 54°$, and a threefold astigmatism of $c_{33} = 106$ ($\pm 3$) nm $\varphi_{33} = 43°$. Due to the application of a calibrated beam-tilt tableau, the `ATLAS` software measures $C_S = -69.0$ ($\pm 0.4$) $\mu$m instead of $-40$ $\mu$m. This residual value is not critical, since any spherical aberration is corrected by the subsequent focal-series reconstruction of the electron wavefunction. The residual axial coma and threefold astigmatism are corrected a-posteriori in the reconstructed wavefunction.

A series of 18 images was recorded at the edge of the crystalline object to the vacuum with a sequentially decreasing defocus. According to the step-size analysis during the later FSR by TrueImage [76], the mean defocus decrement between subsequent images is $-3.52$ ($\pm 0.03$) nm over the total range from $+23.0$ nm to $-35.4$ nm. The images taken with defocus values between $+20.6$ nm and $+14.5$ nm exhibit well separated intensity maxima, with the strongest peaks belonging to Nb/W columns. Oxygen columns are only visible as weaker maxima close to the edge of the object. An example is given in figure 8.17 with the image recorded at $+14.5$ nm defocus, which is closest to the Scherzer focus of $+15.3$ nm for a spherical aberration of $C_S = -69$ $\mu$m. With further decreasing defocus

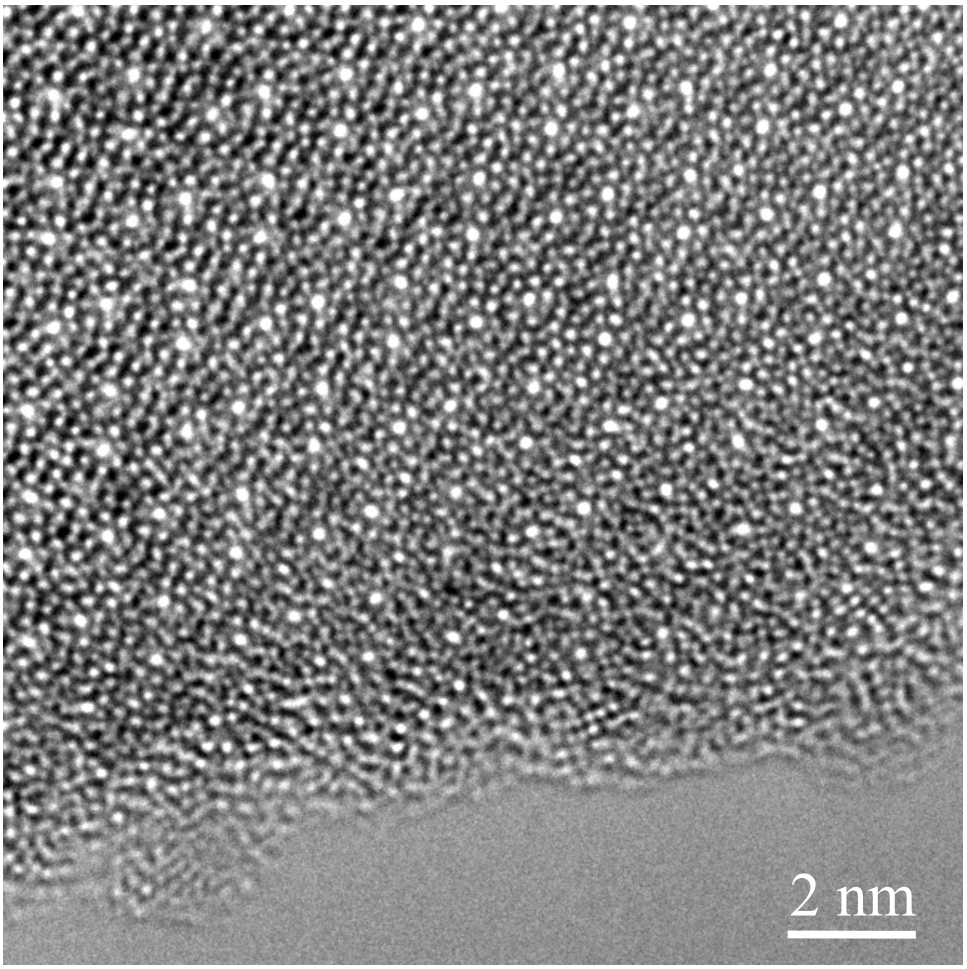



Figure 8.17: High-resolution TEM image of a $Cs_{0.44}Nb_{2.54}W_{2.46}O_{14}$ crystallite in [0 0 1]-projection. The image was recorded with the Philips CM200C as a member of a defocus series with a small negative spherical aberration of $C_S = -69$ $\mu$m and a defocus of +14.5 nanometres.

below about +10.3 nm, the blurring due to the image delocalisation increases as can be seen in the exemplary patches shown in figure 8.18. The contrast in the last images of the series is no longer directly associated with the projected structure, resembling images obtained by conventional (non-$C_S$-corrected) HRTEM.

### 8.2.2 Results and discussion

Images of the amplitude and the phase of the reconstructed wavefunction are displayed in figures 8.19a and 8.19b, respectively. Apart from a prior numerical elimination of residual defocus and residual spherical aberration, these images are still affected by non-circular residual aberrations. By minimisation of the amplitude of the reconstructed wavefunction [98], a residual twofold astigmatism of $c_{22} = 5.3$ nm was measured with an azimuth

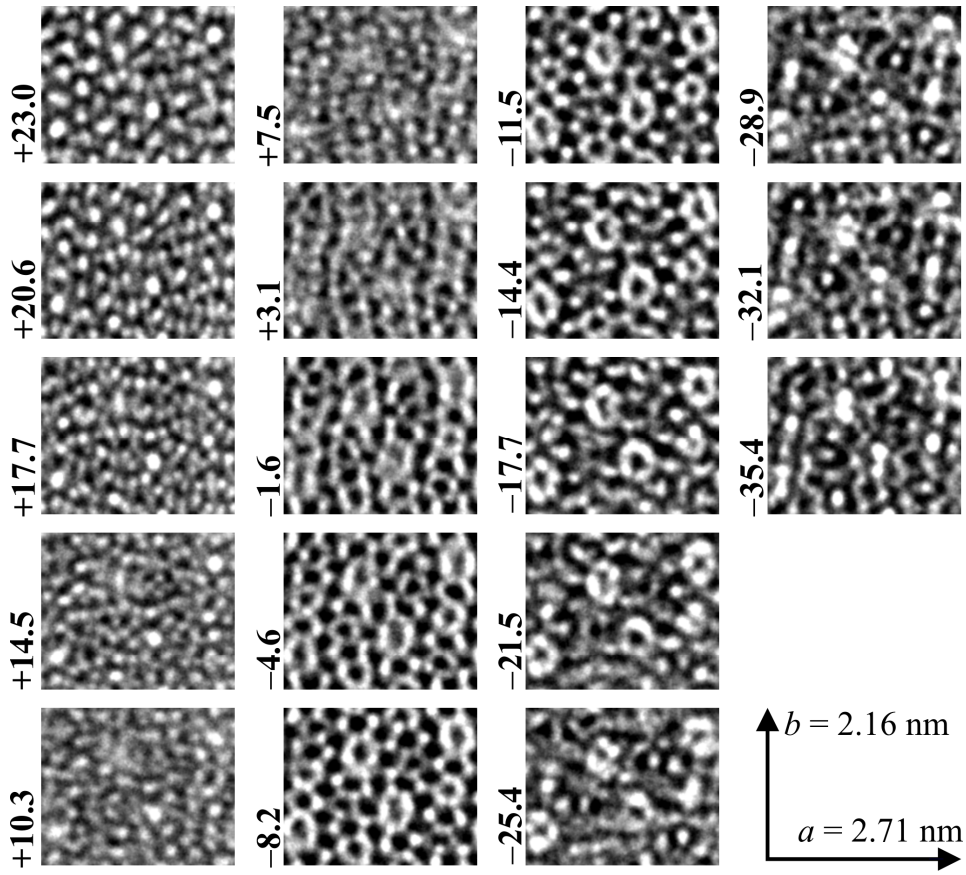



Figure 8.18: Small exemplary image patches from the experimental focus series. Each small image comprises an area of exactly one unit cell of $Cs_{0.44}Nb_{2.54}W_{2.46}O_{14}$ in [0 0 1]-projection. Numbers beside the images denote the corresponding defocus values in nanometres.

angle of 12° with respect to the $x$-axis of the recorded image. The minimisation is implemented as an automated procedure in the reconstruction software TrueImage. The other residual aberrations, axial coma and threefold astigmatism, are corrected according to the aberration measurement with `ATLAS`. The figures 8.19c and 8.19d show images of the amplitude and the phase of the reconstructed wavefunction after correction of the measured residual aberrations.

After correction for the residual aberrations, the phase reveals pronounced maxima at the positions of the strongly scattering niobium/tungsten-oxide columns. Weaker maxima relate to pure oxygen columns and appear in the most cases well separated from the metal-oxide columns. Since the crystal is thin at the edge to the vacuum, the peak-heights in the phase-image are approximately proportional to the scattering power of the respective atom types. *Thin* means in this context that the thickness of the specimen does not exceed about one quarter of the respective column extinction-lengths. Following the

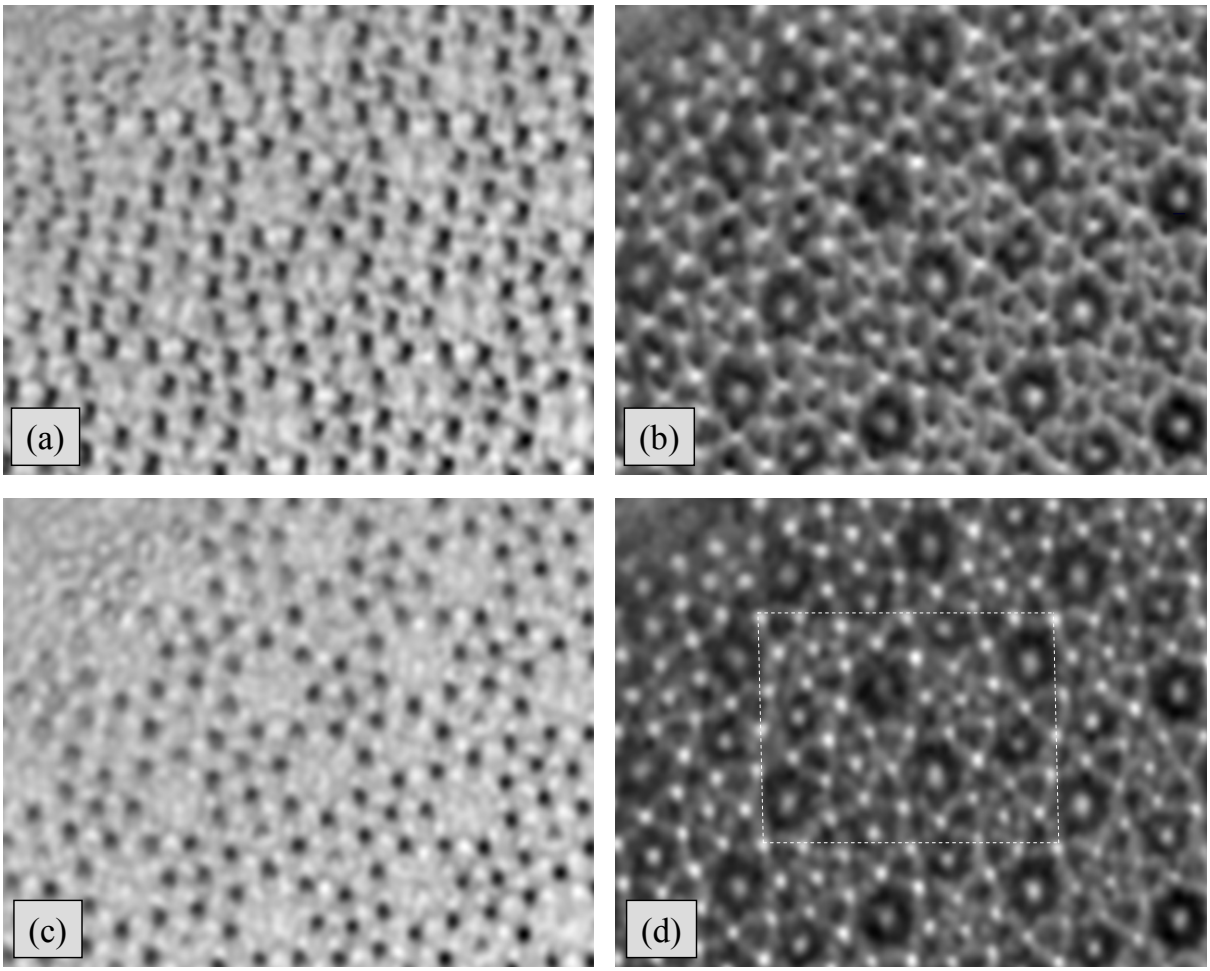



Figure 8.19: Amplitude (left) and phase (right) of the wavefunction as obtained by focal-series reconstruction close to the edge of a $Cs_{0.44}Nb_{2.54}W_{2.46}O_{14}$ crystal projected along the short $c$-axis of the unit cell. The amplitude and phase in (c) and (d) have been corrected for residual aberrations. The rectangle in (d) marks one projected unit cell.

phase-object approximation (POA), the phase of the aberration-corrected wavefunction is directly interpretable in terms of a projected crystal structure. In contrast to the ideal POA, the amplitude image in figure 8.19c exhibits minima at the metal oxide columns and maxima at pure oxygen columns. The caesium columns inside the 6- and 7-sided tunnels exhibit almost no or only very low amplitude contrast.
For the determination of the projected atom-column positions from the observed phase-maxima, a sufficient resolution of the microscope and the ability to correct for residual aberrations are the crucial prerequisites. These aspects determine whether atomic column-positions can be separated from each other in principle, and whether the positions of extremal points in the phase of the wavefunction can be evaluated in a quantitative way without systematic errors due to optical aberrations.

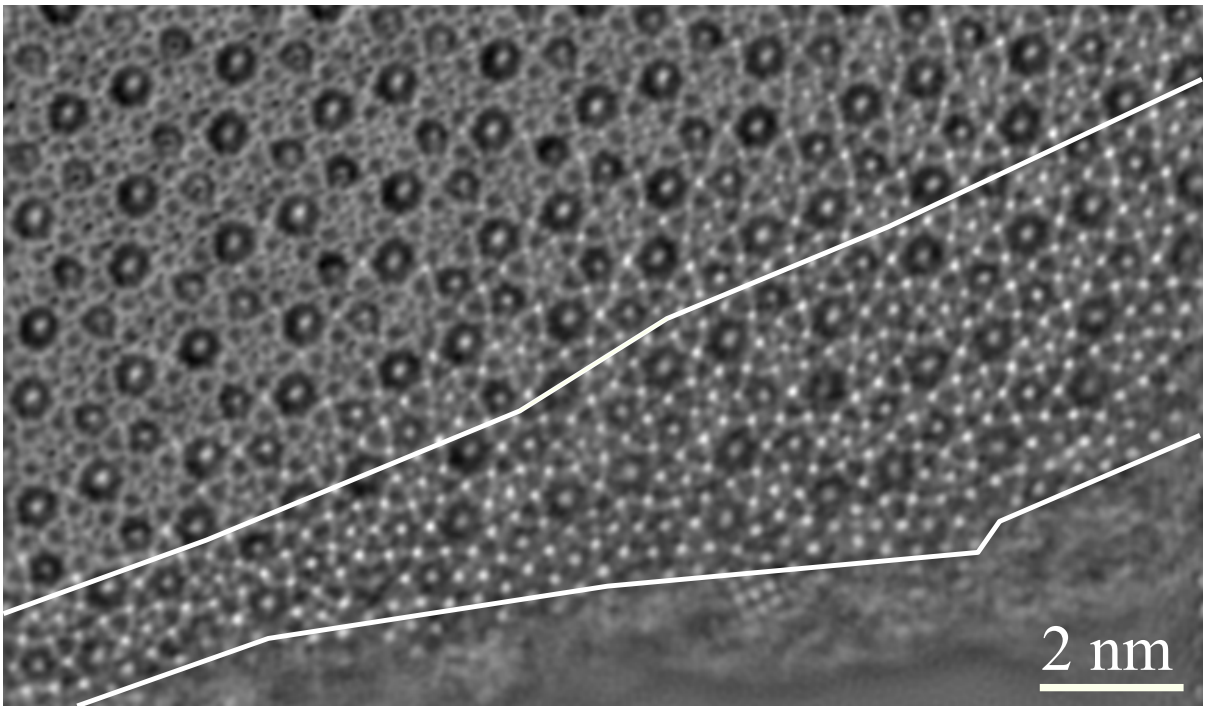



Figure 8.20: The phase of the reconstructed wavefunction after a-posteriori aberration correction close to the edge of a $Cs_{0.44}Nb_{2.54}W_{2.46}O_{14}$ crystal in [0 0 1]-projection shows clearly separated metal-oxide, caesium and oxygen columns. The positions of the atom columns are determined from the area between the white curves.

Individual least-squares fits of a Gaussian peak function to the phase-maxima in the area close to the edge of the projected particle (bounded by the white curves in figure 8.20) provide positional data of 37 Cs-columns, 191 metal oxide columns with unknown niobium/tungsten occupation, and 214 pure oxygen columns. By taking the symmetry constrictions of the plane group $p2gg$ into account, the positions of 2 caesium columns, 11

metal oxide columns, and 18 pure oxygen columns in the unit cell have to be determined.

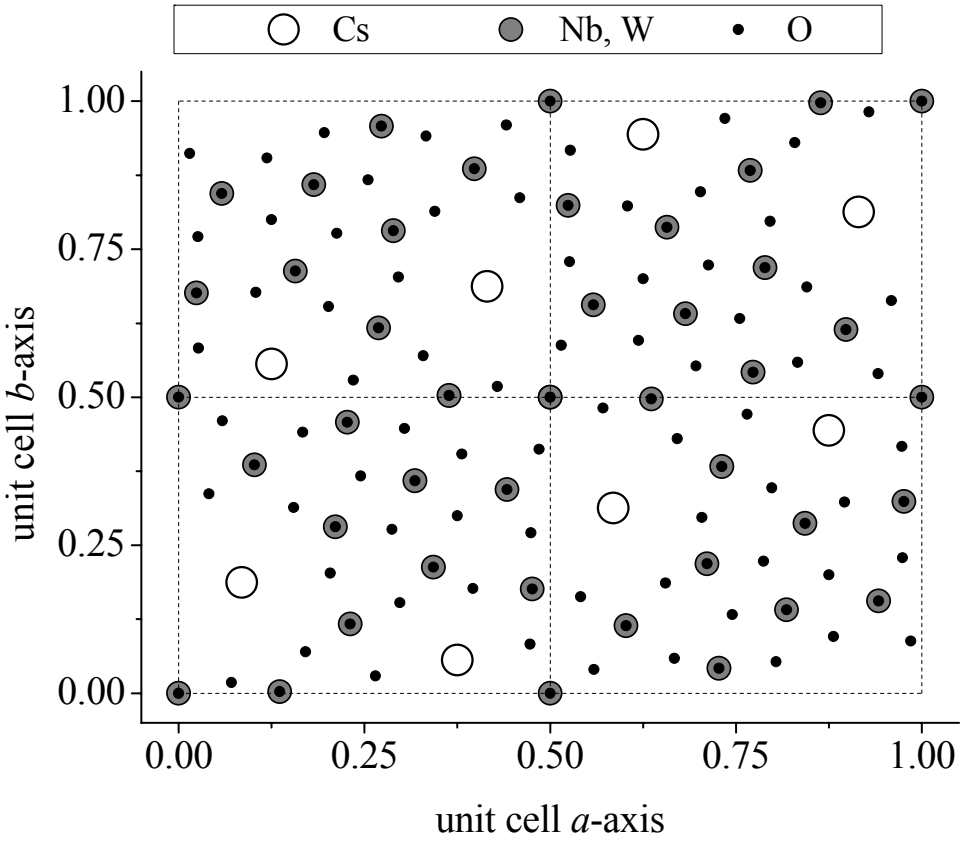



Figure 8.21: Mean projected structure of $Cs_{0.44}Nb_{2.54}W_{2.46}O_{14}$ in one unit cell, determined from the atom-column positions in the phase of the reconstructed and aberration-corrected wavefunction.

The resulting projected structure is illustrated in figure 8.21, and table 8.4 lists the corresponding numerical values in relative unit-cell coordinates. The experimental oxygen-column positions agree with the results of ab-initio calculations by Weirich *et. al.* [111] within the standard deviations of the experiment specified in table 8.4. Discrepancies larger than the standard deviations for the metal-oxide columns are probably related to local disorder frequently observed for this material [110]. Other discrepancies may arise due to the complete neglect of the variable Cs and Nb/W column-occupation in the ab-initio calculations. The average standard deviations from the mean positions of 8 pm for the metal-oxide columns and of 17 pm for the pure oxygen columns are approximately one order of magnitude larger than those achievable under optimum conditions (e.g. in ref. [116]). The larger deviations observed here may arise from the already mentioned local disorder in the structure, from a non-planarity of the surface, and from radiation damage during the acquisition of the defocus series.

| column | $x/a$ | $x/a$ in [111] | $y/b$ | $y/b$ in [111] |
|---|---|---|---|---|
| Cs-1 | 0.085 (±0.004) | | 0.187 (±0.006) | |
| Cs-2 | 0.375 (±0.003) | | 0.056 (±0.004) | |
| Nb(W)O-1 | 0.0 (±0.002) | 0.0 | 0.0 (±0.004) | 0.0 |
| Nb(W)O-2 | 0.102 (±0.002) | 0.099 | 0.386 (±0.004) | 0.381 |
| Nb(W)O-3 | 0.136 (±0.003) | 0.134 | 0.003 (±0.002) | 0.0 |
| Nb(W)O-4 | 0.231 (±0.002) | 0.227 | 0.117 (±0.003) | 0.119 |
| Nb(W)O-5 | 0.211 (±0.003) | 0.209 | 0.281 (±0.004) | 0.279 |
| Nb(W)O-6 | 0.227 (±0.002) | 0.221 | 0.458 (±0.005) | 0.456 |
| Nb(W)O-7 | 0.318 (±0.003) | 0.316 | 0.359 (±0.003) | 0.360 |
| Nb(W)O-8 | 0.343 (±0.003) | 0.343 | 0.213 (±0.005) | 0.212 |
| Nb(W)O-9 | 0.476 (±0.003) | 0.472 | 0.176 (±0.005) | 0.170 |
| Nb(W)O-10 | 0.442 (±0.003) | 0.440 | 0.344 (±0.005) | 0.342 |
| Nb(W)O-11 | 0.0 (±0.002) | 0.0 | 0.500 (±0.006) | 0.5 |
| O-1 | 0.071 (±0.010) | 0.069 | 0.018 (±0.006) | 0.024 |
| O-2 | 0.041 (±0.010) | 0.043 | 0.337 (±0.012) | 0.329 |
| O-3 | 0.059 (±0.007) | 0.063 | 0.460 (±0.008) | 0.459 |
| O-4 | 0.171 (±0.009) | 0.165 | 0.070 (±0.008) | 0.076 |
| O-5 | 0.204 (±0.004) | 0.192 | 0.203 (±0.006) | 0.196 |
| O-6 | 0.155 (±0.008) | 0.147 | 0.314 (±0.010) | 0.316 |
| O-7 | 0.167 (±0.005) | 0.153 | 0.441 (±0.007) | 0.446 |
| O-8 | 0.265 (±0.006) | 0.265 | 0.029 (±0.009) | 0.041 |
| O-9 | 0.298 (±0.007) | 0.295 | 0.153 (±0.007) | 0.151 |
| O-10 | 0.287 (±0.004) | 0.284 | 0.277 (±0.006) | 0.274 |
| O-11 | 0.245 (±0.005) | 0.240 | 0.367 (±0.009) | 0.368 |
| O-12 | 0.304 (±0.008) | 0.300 | 0.447 (±0.009) | 0.451 |
| O-13 | 0.396 (±0.008) | 0.400 | 0.177 (±0.012) | 0.165 |
| O-14 | 0.375 (±0.005) | 0.373 | 0.300 (±0.006) | 0.298 |
| O-15 | 0.381 (±0.006) | 0.383 | 0.404 (±0.005) | 0.406 |
| O-16 | 0.473 (±0.004) | 0.470 | 0.083 (±0.006) | 0.083 |
| O-17 | 0.474 (±0.008) | 0.471 | 0.271 (±0.007) | 0.265 |
| O-18 | 0.485 (±0.007) | 0.485 | 0.412 (±0.007) | 0.410 |

Table 8.4: Mean projected atom-column positions determined from the phase of the aberration-corrected, reconstructed wavefunction in comparison to results from ab-initio calculations [111]. The positions are specified in relative coordinates of the unit cell and determine one quarter of the unit cell. The positions in the three other quarters follow according to the symmetry of the plane group $p2gg$.

According to the tentative model, all pure oxygen columns contain one atom per unit cell at the relative position $z/c = 0.0$ along the $c$-axis. The caesium atoms are located at $z/c = 0.5$, if the respective position is occupied. The metal-oxide columns contain 2 atoms per unit cell, where the metal atom is located at $z/c = 0.0$ and the oxygen at $z/c = 0.5$. Owing to the relatively large thickness gradient, it is not possible to analyse the occupancy of the metal columns with respect to the niobium and tungsten concentration and the caesium occupation in the 6- and 7-sided tunnels.

Quantitative information about the oxygen sub-lattice in the structure of the complex metal oxide $Cs_x(Nb,W)_5O_{14}$ is retrieved by application of aberration-corrected HRTEM in tandem with focal-series reconstruction. By combining the experimentally obtained positions in the projected plane with results from ab-initio calculations for the metal oxide framework, the proposed structure model is now confirmed for the first time for all atom positions in the unit cell. Besides the compliance to the limitations of the weak-phase-object approximation, precise aberration measurement and a-posteriori aberration correction certifies that the observed information contained in the phase of the reconstructed wavefunction is directly linked to the projected object structure. This experiment demonstrates thus that precise aberration control enables a quantitative acquisition of structural information by high-resolution transmission electron microscopy in a straightforward procedure.

## 8.3 Sub-Ångström HRTEM of defects in $SrTiO_3$

The necessity for aberration control in high-resolution transmission electron microscopy is best demonstrated with high-resolution images of defects. The Fourier transform of aperiodic object features consist in general of a broad band of Fourier-coefficients. In order to transfer all corresponding spatial frequencies in a correct mutual phase relationship, a quantitative measurement and elimination of aberrations is mandatory.
The experiment reported in this section sketches a common situation, where a-priori aberration-control is skipped in favour of a fast acquisition of high-resolution images. A short duration of the HRTEM-experiment is important in order to avoid damage of the thin object due to the irradiation with highly energetic electrons. The imaged area in this experiment was encountered in the beginning of the microscope session. Hence, the presence of larger residual aberrations is expected. It is demonstrated that aberration measurement and a-posteriori aberration correction of the reconstructed wavefunction provides quantitative results which are interpretable in a straightforward manner in terms

of the projected object structure. The interpretability is extended below one Ångström resolution with the application of the aberration-measurement procedures developed in this work.

### 8.3.1 Experiment

A focal series consisting of 20 HRTEM-images has been recorded at the edge of a $SrTiO_3$ crystal in [1 1 0]-projection[2]. According to the analysis during the focal-series reconstruction by TrueImage [76], the mean defocus step-size is $-1.70$ ($\pm 0.02$) nm over the total range from $+7.5$ nm to $-23.6$ nm. The chosen magnification ensures a sampling of 10.3 pm per pixel with the complete CCD-array of 2048×2048 pixels. Thereby the discretisation is almost two times smaller than the Nyquist criterion $1/(2g_{\max}) = 40$ pm with respect to twice the information limit $g_{\max} = 12.5\ \mathrm{nm}^{-1}$ of the instrument, and more than a factor of ten smaller than the smallest distance of 1.38 Å between the titanium- and oxygen-columns in the projected bulk structure (figure 8.22).

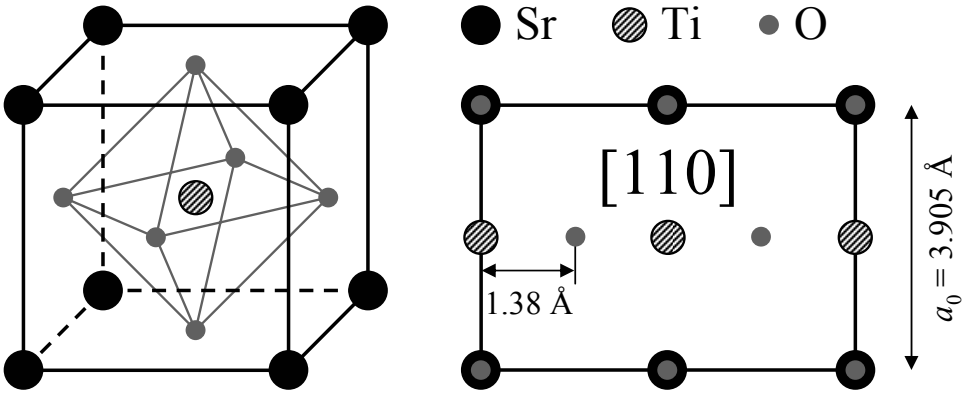



Figure 8.22: Crystal structure of $SrTiO_3$ at room temperature and its [1 1 0]-projection. The size of the spheres are chosen arbitrarily and do not resemble to the respective atomic diameters.

The full recorded area shown in figure 8.23 contains three separate defects frequently observed in $SrTiO_3$: a dislocation (D) close to the left image border, an extrinsic planar fault (PF1), and a larger curved version (PF2) of the same type at the right side. The two enlargements reveal contrast modulations between the bright SrO-columns which are not related to the projected object structure. The oxygen-columns are blurred in the bulk and more or less invisible at the defects. These deviations from the ideal projection are caused by aberrations. After the acquisition of the defocus-series, a beam-tilt series

[2]The $SrTiO_3$-sample was provided by Dr. S. Mi, Institute for Microstructure Research, Research Center Jülich.

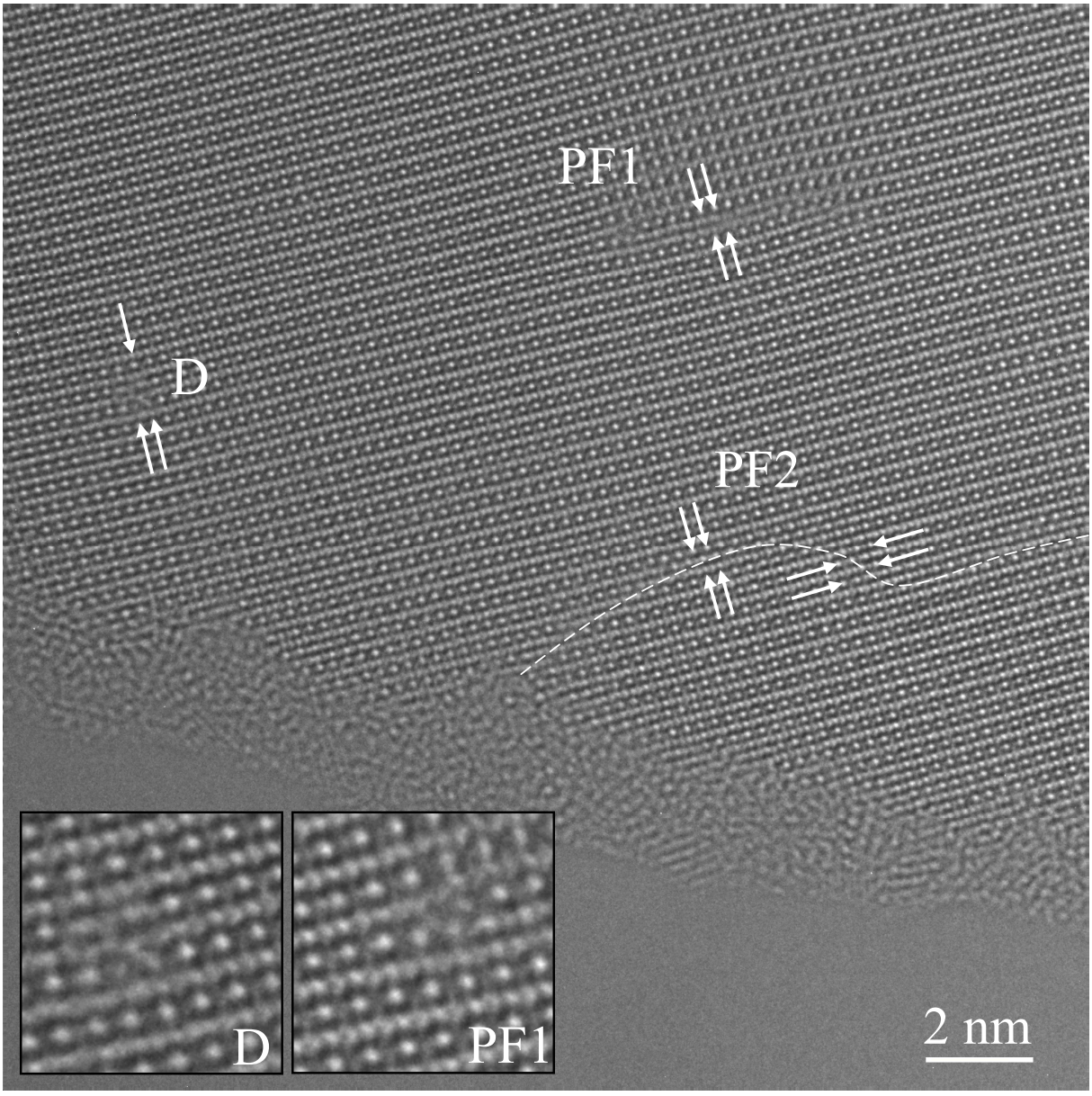
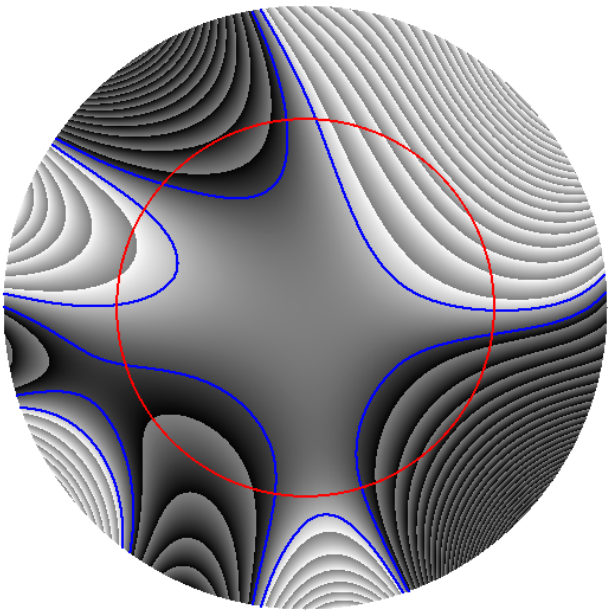



Figure 8.23: HRTEM image of $SrTiO_3$ in [1 1 0]-projection recorded as a member of a defocus series with the FEI Titan 80-300. The defocus of the image is 5.5 nm which is close to the optimum defocus for negative spherical-aberration imaging according to eqn. (1.3.13).

was recorded from an area further away from the previous location which contains mainly amorphous $SrTiO_3$. The higher-order aberrations present in the recorded defocus series were measured from this tilt series.

### 8.3.2 Results and discussion

The complex-valued electron wavefunction is reconstructed from two regions, each comprising 1024×1024 pixels of the recorded images. One region contains the defect-site (D) at the left side of the full image, the other area contains the two planar faults (PF1) and (PF2) on the right side. The resolution of the reconstruction is limited to 0.9 Å, because the inclusion of details below 0.9 Å essentially increased the noise in the reconstructed wavefunction. For the final reconstruction of the wavefunction, a reduced defocus series comprising 14 images was used which exhibits a significantly smaller unwanted variation of the defocus step-size than the complete series of 20 images.



| | $c_{mn}$ | $\sigma_{mn}$ | $\varphi_{mn}$ |
|---|---|---|---|
| $c_{22}$ | 1.3 | (± 0.10) nm | 12.6 ° |
| $c_{31}$ | 203.6 | (± 11.0) nm | 60.1 ° |
| $c_{33}$ | 175.5 | (± 4.73) nm | 38.9 ° |
| $c_{40}$ | -25.73 | (± 1.07) µm | |
| $c_{42}$ | 7.58 | (± 0.17) µm | 52.2 ° |
| $c_{44}$ | 1.66 | (± 0.04) µm | -1.1 ° |
| $c_{51}$ | 127.59 | (± 2.66) µm | -149.8 ° |
| $c_{53}$ | 47.46 | (± 8.45) µm | -26.8 ° |
| $c_{55}$ | 73.18 | (± 1.39) µm | -1.1 ° |
| $c_{60}$ | 10.00 | (± 1.09) mm | |
| $c_{66}$ | 3.60 | (± 0.04) mm | -21.6 ° |
| | Outer consistency $s_{all}$ = 0.23 nm | | |

Figure 8.24: The phase plate for the non-circular aberrations (see table on the right) measured with ATLAS from amorphous $SrTiO_3$ after the acquisition of the defocus series. The radius of the phase plate is 20 $nm^{-1}$. Grey contrast steps appear at multiples of $\pi/2$. The blue curves mark the $\pm\pi/4$-contours, whereas the red line corresponds to the information limit $g_{max} = 12.5$ $nm^{-1}$ of the microscope.

The results of the aberration measurement from the recorded tilt series of amorphous $SrTiO_3$ are summarised in figure 8.24. The error of the measured aberration function is estimated to be $\bar{\sigma}_\chi(g_{max}) = 0.08\,\pi$ with $g_{max} = 12.5$ $nm^{-1}$. However, the second order aberrations have to be determined locally from the wavefunction, since the tilt series was

recorded at a different area of the object.

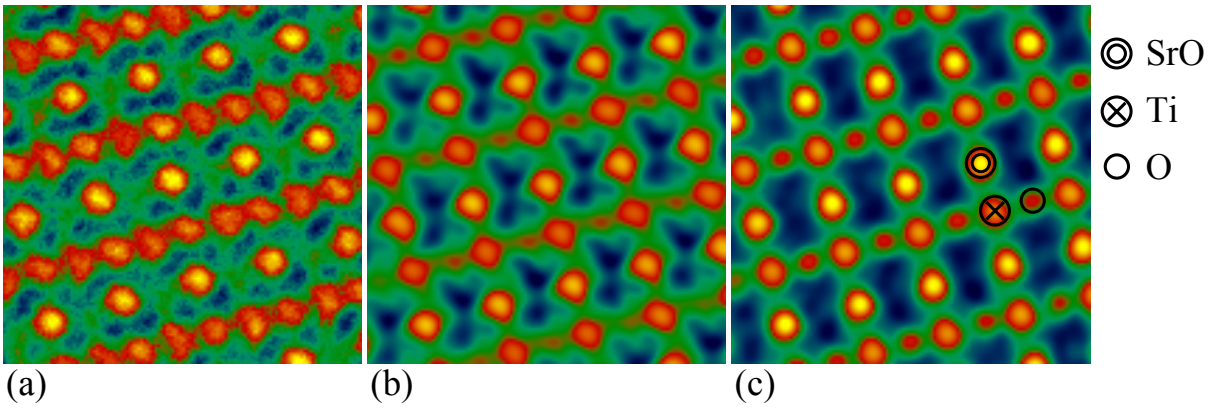



Figure 8.25: Small patches cut out from (a) the original image, (b) the phase of the wavefunction as retrieved by the focal-series reconstruction without aberration correction, and (c) the phase of the wavefunction after aberration correction. Each sub-image shows the same area from the bulk $SrTiO_3$ [1 1 0]. A colour scale is used to accentuate the differences between the image, the phase of reconstructed wavefunction, and the phase of the aberration corrected wavefunction.

The phase of the reconstructed wavefunction, as shown for a selected area in figure 8.25b, is quite strongly deteriorated by residual aberrations. A colour scale is used to emphasise the qualitative differences between the image, the phase of reconstructed wavefunction, and the phase of the aberration corrected wavefunction. In comparison to the original image in figure 8.25a, the phase of the reconstructed wavefunction exhibits a substantial improvement of the signal-to-noise ratio due to an intrinsic averaging over several images by the focal-series reconstruction. Most of the columns still appear non-round and the oxygen columns are not clearly distinguished from the neighbouring titanium columns.
After a-posteriori correction of the reconstructed wavefunction by the aberrations listed in the table of figure 8.24 only a residual defocus and twofold astigmatism remained. Depending on the geometry of the object, the defocus typically varies by several nanometres over the field of view. A tuning of second-order aberrations by hand and separately at each region of interest is therefore a standard routine in a-posteriori aberration correction. For objects which are thin in the frame of the phase-object approximation, correct aberrations of the even orders (2, 4, 6, ...) yield locally minimum amplitude contrast in the reconstructed wavefunction [98]. It should be emphasised that odd-order aberrations cannot be determined by local amplitude minimisation.
For the present experiment, a twofold astigmatism of $c_{22} = 3.8$ nm (-55° azimuth) was determined by minimisation of the amplitude contrast. The defocus differs by more than 4 nm over the full image size of 21 nm and can thus be corrected only locally. After a final

correction also for the two second-order aberrations, all atom-columns are represented by clearly separated and almost round maxima in the phase of the wavefunction as shown in 8.25c. Columns which contain different atoms have different diameters and are distinguishable by their maximum phase value. The maximum phase value is approximately proportional to the scattering power. The columns of the light oxygen atoms possess the weakest phase contrast, the heaviest atoms (strontium) show the brightest contrast, whereas the titanium columns are represented by an intermediate contrast. Only minor modulations remain in the "empty" space between the strontium columns as expected for an ideal projection of the bulk structure.

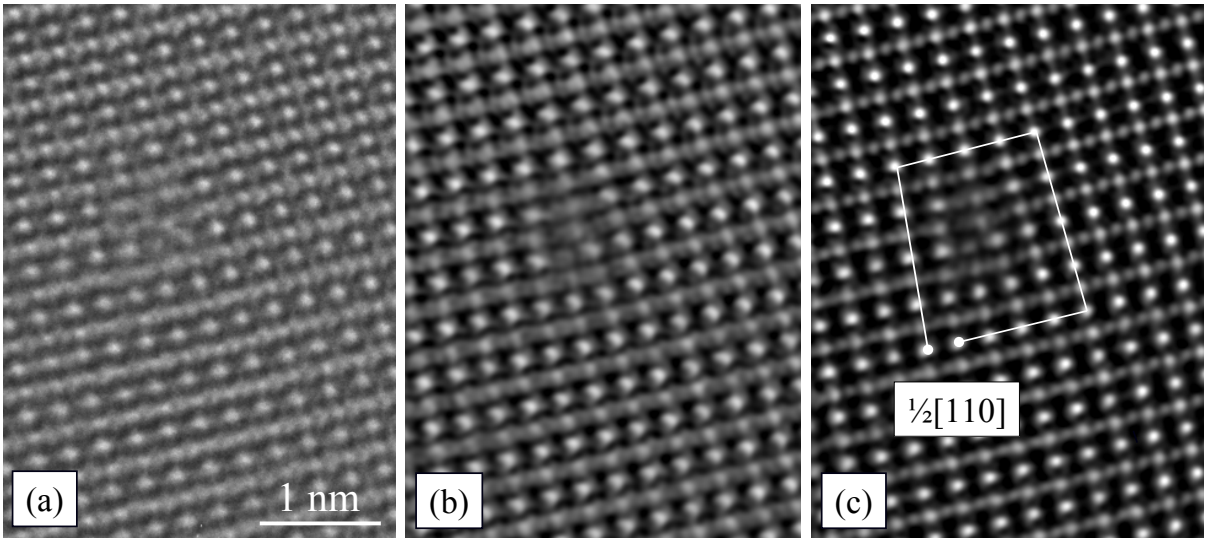



Figure 8.26: Comparison of (a) the image contrast, (b) the phase of the wavefunction as retrieved by the focal-series reconstruction without aberration correction, and (c) the phase of the wavefunction after aberration correction. The displayed images correspond to a small part on the left side of the full recorded image of fig. 8.23 showing an isolated dislocation in $SrTiO_3$ [1 1 0].

Figure 8.26 shows a selected area from the left side of the recorded HRTEM-images. In contrast to the direct image (fig. 8.26a) and the uncorrected phase of the reconstructed wavefunction (fig. 8.26b) an aberration correction with the results from the ATLAS-measurement leads to a clear resolution of atom columns at the isolated dislocation shown in figure 8.26c. The projected Burgers-vector $\frac{1}{2}[1\ 1\ 0]$ is actually no lattice translation-vector of the structure. Since there are no extra defects accompanying the dislocation, the full Burgers-vector has to be a lattice translation-vector, e.g. [1 0 0] or [0 1 0]. Therefore, a component of the vector along the viewing direction is assumed which is not visible in the image.

A larger area of the phase of the aberration-corrected, reconstructed wavefunction is shown

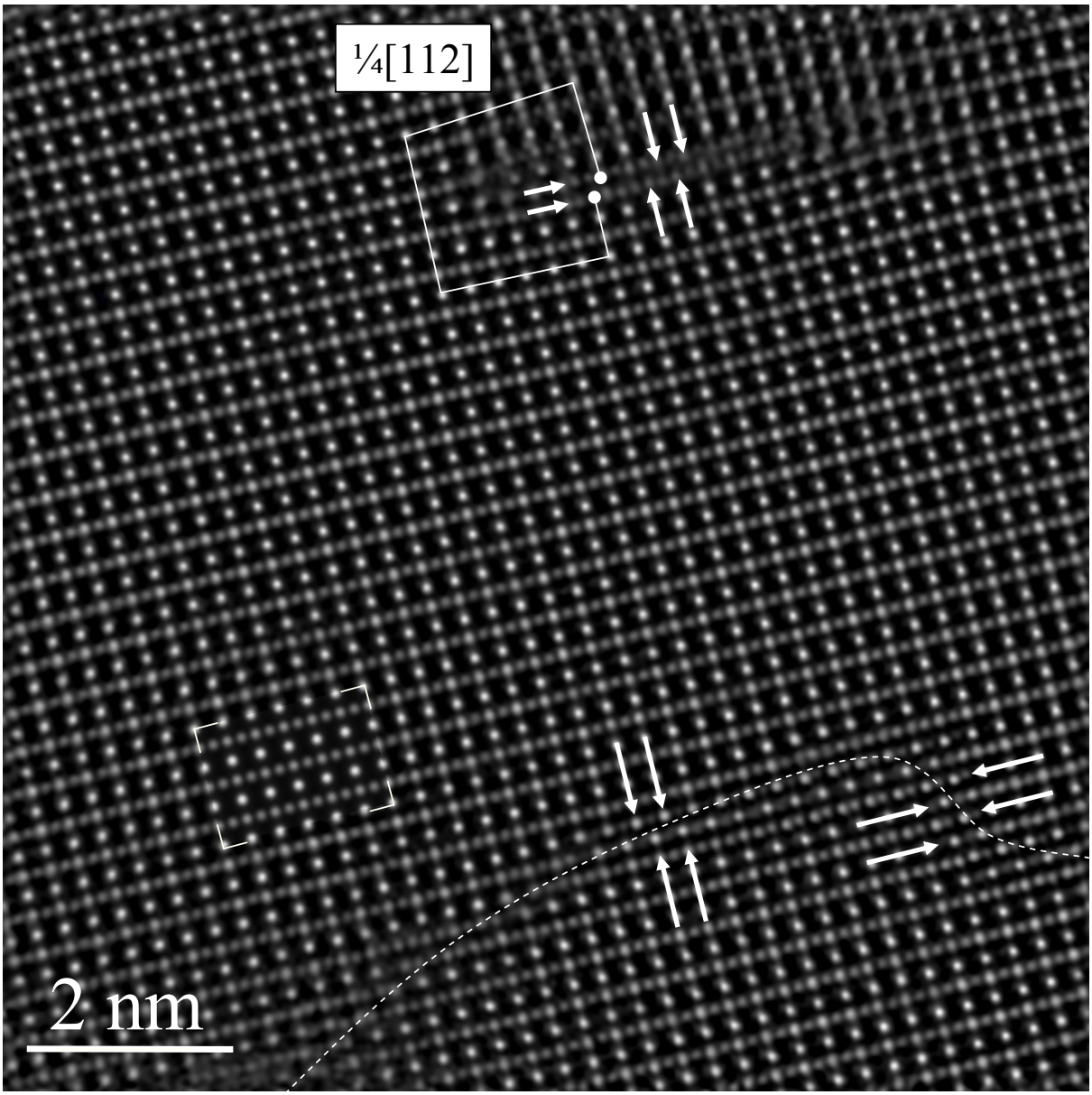



Figure 8.27: Phase of the aberration-corrected, reconstructed wavefunction of $SrTiO_3$ [1 1 0] obtained from the right side of the full recorded area. Two planar faults of the same type are resolved at atomic scale. The phase of a simulated exit-plane wavefunction is inserted to the rectangular area marked at its corners.

in figure 8.27. The phase maxima correspond directly to the projected atomic structure, which is confirmed by the insertion of the phase of a calculated exit-plane wavefunction to the rectangular area marked at its corners. The Multislice-Algorithm [43, 44] was applied to calculate the exit-plane wavefunction from stacks of four subslices per unit cell of bulk $SrTiO_3$ in [1 1 0]-projection using the EMS-software [7]. The simulated phase fits best to the experimental phase data with 3×4 subslices. This number of slices corresponds to an

object thickness of 1.68 nm. In addition to the thickness, an object tilt of 1.5° was determined with respect to the direction of the incident electron beam and with an azimuth of 105° to the $x-$axis of the images.
A planar fault with double $TiO_2$-planes stacking along the [0 0 1]-direction is clearly revealed at atomic scale in the upper part of figure 8.27. The fault is characterised by an extra $TiO_2$-plane inserted into the perfect lattice. As a result, partial dislocations appear surrounding the fault. Burgers-circuits around the ends of the fault yield a projected Burgers-vector of $\vec{b} = \frac{1}{4}[1\ 1\ 2]$. The analysis of the structural nature of the fault yields a Burgers-vector of the partials of the type $\frac{1}{2}\langle 1\ 0\ 1\rangle$. This vector is also the displacement vector of the fault which is seen in the image as a half-period displacement of the lattice across the fault.
The lattice defect in the lower part of figure 8.27 looks more complicated than the above fault which is confined in the (0 0 1)-plane. Its three-dimensional surface geometry appears in form of a bent shape in the present projection following the dashed white curve from the lower edge of the object to the right border of the image. Indeed, it has the same type of displacement vector as the above fault. This displacement vector can be recognised by the half period shift of the lattice across the defect denoted by arrow pairs. In the areas below the dislocation in figure 8.26c and above the upper fault in figure 8.27, a large number of atom columns are blurred in one direction. This distortion is mainly attributed to a tilt of the columns in the three-dimensional lattice away from the projection-axis and follows a bending of the object.

Although the experiment described in this section provides no new insights on defect structures in $SrTiO_3$, it demonstrates a substantial methodological progress towards quantitative Sub-Ångström structure analysis with high-resolution transmission electron microscopy. The electron wavefunction was reconstructed from a defocus series of HRTEM-images and corrected a-posteriori for residual aberrations. The aberration-measurement procedures developed in this work enabled a straightforward experimental procedure to retrieve reliably an aberration-clean exit-plane wavefunction ready for structure analysis. In contrast, a direct analysis of the uncorrected data would not be possible or would at least require tedious image simulations suffering from problems of parameter uniqueness.

# Summary

In the course of the present work, a significant progress towards quantitative Sub-Ångström transmission electron microscopy could be achieved. The particular contribution of this work concerns the electron-optical adjustment of high-resolution microscopes which was not satisfactorily solved in the past. Especially with the availability of new Sub-Ångström microscopes of the present generation, a precise determination of the optical transfer properties of such machines became an unsolved challenge for microscope manufacturers as well as microscope users. While a physical resolution well below one Ångström could be readily demonstrated, the achieved resolution could not be reliably exploited for the interpretation of the resulting images due to the presence of so-called higher-order aberrations. Such aberrations can deteriorate the image quality substantially, making a straightforward image interpretation impossible. At the starting time of this work, a sufficiently accurate procedure to measure higher-order aberrations was urgently missing. Such a procedure is a mandatory prerequisite for any aberration control enabling quantitative high-resolution microscopy, be it by in-situ adjustments of hardware elements during experiment, be it by image simulation, or by software phase-retrieval methods applied after experiment. The goal of this work was therefore to develop such a measurement procedure for the Sub-Ångström regime.

The measurement procedures developed in the course of this work are based on the numerical evaluation of images taken from an amorphous object under tilted electron-beam illumination. These procedures even over-accomplish the above stated goal in a certain sense: While it was initially intended to develop procedures for the nowadays feasible near Sub-Ångström regime around 0.8 Ångström, it turned out that actually the deep Sub-Ångström regime around 0.5 Ångström can also be tackled, provided such instruments would already exist. The precision reached with the newly developed measurement procedures is unprecedented and surpasses existing solutions by at least one order of magnitude in any respect, helping to render high-resolution transmission electron microscopy

a real quantitative technique for Sub-Ångström materials investigation. Major milestones during the development of the new measurement schemes were:

- Derivation of a new multi-dimensional tolerance limit for aberrations.
- Derivation of mathematically sound error-propagation formalism for aberrations.
- Elaboration of the contrast-transfer theory with partially coherent illumination by a tilted electron beam.
- Development of a novel ultra-fast numerical pattern recognition scheme.
- Development of a novel ultra-precise numerical pattern recognition scheme.
- First-time combination of real-space and Fourier-space data-evaluation.
- Development of the first quantitative procedure to measure the resolution limit.
- Implementation of a procedure for the surveillance of the optical stability.
- Materials science applications demonstrating the feasibility of the procedures.

All the concepts and procedures for aberration measurement developed in this work have been implemented in a software package named ATLAS (Accurate Treatment of Lens Aberrations and Stability). The ATLAS software satisfies professional demands with respect to robustness, precision, speed and user-friendliness. It is hoped, that ATLAS will be widely spread in the high-resolution electron microscopy community through currently negotiated licensing schemes, enabling a quantitative access to the Sub-Ångström regime to as many electron microscopists as possible.

# Appendix A

# Definitions and Theorems

## A.1 Fourier transformation

Definition of the Fourier transformation:

$$\mathcal{F}\left[f(t)\right] = f(\omega) = \int f(t)\, \mathrm{e}^{2\pi \mathrm{i}\omega t}\, \mathrm{d}t \tag{A.1.1}$$

Definition of the inverse Fourier transformation:

$$\mathcal{F}^{-1}\left[f(\omega)\right] = f(t) = \int f(\omega)\, \mathrm{e}^{-2\pi \mathrm{i}t\omega}\, \mathrm{d}\omega \tag{A.1.2}$$

## A.2 Convolution theorem

A convolution of two functions $h(t)$ and $l(t)$ in real space is given by

$$f(t) = \int h(t')\, l(t-t')\, \mathrm{d}t' = h * l(t). \tag{A.2.1}$$

The Fourier transformation of equation (A.2.1) yields

$$\begin{aligned} f(\omega) &= \iint h(t')\, l(t-t')\, \mathrm{e}^{2\pi \mathrm{i}\omega t}\, \mathrm{d}t'\, \mathrm{d}t && \text{(A.2.2)} \\ &= \iint h(t')\, \mathrm{e}^{2\pi \mathrm{i}\omega t'}\, l(t-t')\, \mathrm{e}^{2\pi \mathrm{i}\omega (t-t')}\, \mathrm{d}t'\, \mathrm{d}t && \text{(A.2.3)} \\ &= \int h(t')\, \mathrm{e}^{2\pi \mathrm{i}\omega t'}\, \mathrm{d}t' \int l(t'')\, \mathrm{e}^{2\pi \mathrm{i}\omega t''}\, \mathrm{d}t'' && \text{(A.2.4)} \\ &= h(\omega) \cdot l(\omega) && \text{(A.2.5)} \end{aligned}$$

$$\Rightarrow \mathcal{F}\left[h * l(t)\right] = h(\omega) \cdot l(\omega) \;\Leftrightarrow\; h * l(t) = \mathcal{F}^{-1}\left[h(\omega) \cdot l(\omega)\right] \tag{A.2.6}$$

In conclusion, a convolution of two functions in real space is equivalent their multiplication in Fourier space.

## A.3 Fourier transformation of the image intensity

The image intensity $I(\vec{r})$ is the modulus square of the electron wavefunction $\Psi(\vec{r})$ in the image plane, $I(\vec{r}) = \Psi(\vec{r})\Psi^*(\vec{r})$. Its Fourier transform $I(\vec{g})$ is therefore given by

$$I(\vec{g}) = \mathcal{F}[I(\vec{r})] = \int \Psi(\vec{r})\Psi^*(\vec{r})\, \mathrm{e}^{2\pi \mathrm{i}\vec{g}\,\vec{r}}\, \mathrm{d}^2\vec{r} \tag{A.3.1}$$

Replacing $\Psi^*(\vec{r})$ by its Fourier transform yields

$$I(\vec{g}) = \int \Psi(\vec{r}) \left( \int \Psi^*(\vec{k})\, \mathrm{e}^{2\pi \mathrm{i}\vec{k}\,\vec{r}}\, \mathrm{d}^2\vec{k} \right) \mathrm{e}^{2\pi \mathrm{i}\vec{g}\,\vec{r}}\, \mathrm{d}^2\vec{r}. \tag{A.3.2}$$

Exchanging the sequence of integration and subsequent application of the definition (A.1.2) of the Fourier transform leads to

$$I(\vec{g}) = \int \left( \int \Psi(\vec{r})\, \mathrm{e}^{2\pi \mathrm{i}(\vec{g}+\vec{k})\vec{r}}\, \mathrm{d}^2\vec{r} \right) \Psi^*(\vec{k})\, \mathrm{d}^2\vec{k}, \tag{A.3.3}$$

$$= \int \Psi(\vec{g}+\vec{k})\, \Psi^*(\vec{k})\, \mathrm{d}^2\vec{k}. \tag{A.3.4}$$

The Fourier coefficient $I(\vec{g})$ of the image intensity is the integral over all pairs products $\Psi(\vec{g'})\, \Psi^*(\vec{g''})$ with $\vec{g} = \vec{g'} - \vec{g''}$.

## A.4 Definition of variance, covariance, and standard deviation

Consider two data sets $X = \{x_i\}$ and $Y = \{x_i\}$ of size $N$. Then the sample variance of the data set $X$ is defined by

$$\mathrm{var}(X) = s_{xx} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{A.4.1}$$

with the mean value

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{A.4.2}$$

The covariance between two sets of data is defined by

$$\mathrm{cov}(X, Y) = s_{xy} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}). \tag{A.4.3}$$

The square root of the sample variance is called standard deviation $\sigma_x = \sqrt{s_{xx}}$ or unbiased standard deviation. The above definitions are analogous to the definitions in [119].

## A.5 The correlation coefficient

The *correlation coefficient* $r$ is a measure for the linear dependence between two sets of data $X = \{x_i\}$ and $Y = \{x_i\}$ of size $N$. For two discrete sets it is defined [119] by

$$r = \frac{s_{xy}}{\sqrt{s_{xx}\, s_{yy}}}, \tag{A.5.1}$$

where $s_{xx}$, $s_{xy}$, and $s_{yy}$ are the variances and covariances as defined in section A.4
The values of $r$ lie in the interval $[-1, 1]$. $r = 1$ means that $X$ and $Y$ are dependent on each other in a linear way. Anti-correlation for $r = -1$ denotes an antipodal variation of both sets from its mean value, while $r = 0$ signalises that $X$ and $Y$ are independent data-sets.
The correlation coefficient is often applied to measure the quality of least squares fits, where the sum of difference squares

$$\chi^2 = \sum_{i=1}^{N} (x_i - y_i)^2 = \sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N} y_i^2 - 2 \sum_{i=1}^{N} x_i y_i \tag{A.5.2}$$

between experimental data $X$ and values $Y$, calculated with an appropriate fit model, is minimised. Minimisation of $(\chi^2)$ is equivalent to maximisation of $r$.

## A.6 Cross Correlation

A cross correlation measures the similarity between two functionalities $f(t)$ and $g(t)$ with respect to a shift of the independent variable $t$. Mathematically the cross correlation between $f$ and $g$ (symbolised by $f \star g(t)$) is defined by the convolution

$$\begin{aligned} f \star g(t) &= \int_{-\infty}^{\infty} f^*(-t') \cdot g(t - t')\, \mathrm{d}t', \\ &= \int_{-\infty}^{\infty} f^*(t') \cdot g(t + t')\, \mathrm{d}t'. \end{aligned} \tag{A.6.1}$$

The cross-correlation function $c(t) = f * g(t)$ is maximum at the shift $t$ where $f$ and $g$ are most similar.
According to the convolution theorem (A.2.6) the cross correlation is equivalent to multiplication of the Fourier transforms:

$$f \star g(t) = \mathcal{F}^{-1} \left[ \mathcal{F} \left[ f^*(-t) \right] \cdot \mathcal{F} \left[ g(t) \right] \right]. \tag{A.6.2}$$

Numerically, the cross correlation between two data-sets of size $N$ requires $N^3$ operations. The application of Fast Fourier Transforms (FFT) [89] in equation (A.6.2) reduces this to $(2N + 1) \cdot N \cdot \log N$ operations.

# Appendix B

# Transformation of effective second-order aberrations

Explicit expressions for the transformation of second-order aberrations under a resampling of real-space distances are denoted in this section. Consider an effective defocus $c'_{20}$, and an effective twofold astigmatism $c'_{22x}$, $c'_{22y}$ which have been measured from a diffractogram assuming a real-space sampling $\mathbf{S}_\alpha$ for the respective HRTEM image with

$$\mathbf{S}_\alpha = \begin{pmatrix} \alpha_{xx} & \alpha_{yx} \\ \alpha_{xy} & \alpha_{yy} \end{pmatrix}. \tag{B.0.1}$$

When a different real-space sampling $\mathbf{S}^\dagger_\alpha$ would have been assumed, different second order aberrations would have been measured $c'^\dagger_{20}$, $c'^\dagger_{22x}$, and $c'^\dagger_{22y}$ with

$$\mathbf{S}^\dagger_\alpha = \begin{pmatrix} \alpha^\dagger_{xx} & \alpha^\dagger_{yx} \\ \alpha^\dagger_{xy} & \alpha^\dagger_{yy} \end{pmatrix}. \tag{B.0.2}$$

According to the discussion in section 7.5.3, three linear equations

$$\begin{pmatrix} c'_{20} \\ c'_{22x} \\ c'_{22y} \end{pmatrix} = \begin{pmatrix} M_{00} & M_{x0} & M_{y0} \\ M_{0x} & M_{xx} & M_{yx} \\ M_{0y} & M_{xy} & M_{yy} \end{pmatrix} \cdot \begin{pmatrix} c'^\dagger_{20} \\ c'^\dagger_{22x} \\ c'^\dagger_{22y} \end{pmatrix}, \tag{B.0.3}$$

can be derived, which connect the original to the transformed values of the effective second-order aberrations. Explicit formulas for the coefficients $M_{ij}$ are denoted in the

following:

$$\begin{aligned} M_{00} &= \frac{1}{2D^2}\Big[{\alpha^\dagger_{yx}}^2{\alpha_{xy}}^2+{\alpha_{yx}}^2{\alpha^\dagger_{xy}}^2+{\alpha^\dagger_{xy}}^2{\alpha_{yy}}^2 \\ &\quad -2\alpha^\dagger_{xx}\alpha^\dagger_{yx}(\alpha_{xx}\alpha_{yx}+\alpha_{xy}\alpha_{yy})+{\alpha^\dagger_{xx}}^2({\alpha_{yx}}^2+{\alpha_{yy}}^2) \\ &\quad -2\alpha_{xx}\alpha_{yx}\alpha^\dagger_{xy}\alpha^\dagger_{yy}-2\alpha_{xy}\alpha^\dagger_{xy}\alpha_{yy}\alpha^\dagger_{yy} \\ &\quad +{\alpha_{xy}}^2{\alpha^\dagger_{yy}}^2+{\alpha_{xx}}^2({\alpha^\dagger_{yx}}^2+{\alpha^\dagger_{yy}}^2)\Big] \end{aligned} \tag{B.0.4}$$

$$\begin{aligned} M_{0x} &= \frac{1}{2D^2}\Big[{\alpha^\dagger_{yx}}^2{\alpha_{xy}}^2-{\alpha_{yx}}^2{\alpha^\dagger_{xy}}^2-{\alpha^\dagger_{xy}}^2{\alpha_{yy}}^2 \\ &\quad -2\alpha^\dagger_{xx}\alpha^\dagger_{yx}(\alpha_{xx}\alpha_{yx}+\alpha_{xy}\alpha_{yy})+{\alpha^\dagger_{xx}}^2({\alpha_{yx}}^2+{\alpha_{yy}}^2) \\ &\quad +2\alpha_{xx}\alpha_{yx}\alpha^\dagger_{xy}\alpha^\dagger_{yy}+2\alpha_{xy}\alpha^\dagger_{xy}\alpha_{yy}\alpha^\dagger_{yy} \\ &\quad -{\alpha_{xy}}^2{\alpha^\dagger_{yy}}^2+{\alpha_{xx}}^2({\alpha^\dagger_{yx}}^2-{\alpha^\dagger_{yy}}^2)\Big] \end{aligned} \tag{B.0.5}$$

$$\begin{aligned} M_{0y} &= \frac{1}{2D^2}\big[\alpha^\dagger_{xx}({\alpha_{yx}}^2\alpha^\dagger_{xy}-\alpha_{xx}\alpha_{yx}\alpha^\dagger_{yy} \\ &\quad +\alpha_{yy}(\alpha^\dagger_{xy}\alpha_{yy}-\alpha_{xy}\alpha^\dagger_{yy}))+\alpha^\dagger_{yx}(\alpha_{xx}{}^2\alpha^\dagger_{yy}-\alpha_{xx}\alpha_{yx}\alpha^\dagger_{xy} \\ &\quad +\alpha_{xy}(\alpha_{xy}\alpha^\dagger_{yy}-\alpha^\dagger_{xy}\alpha_{yy}))\big] \end{aligned} \tag{B.0.6}$$

$$\begin{aligned} M_{x0} &= \frac{1}{2D^2}\Big[{\alpha^\dagger_{yx}}^2{\alpha_{xy}}^2-{\alpha_{yx}}^2{\alpha^\dagger_{xy}}^2+{\alpha^\dagger_{xy}}^2{\alpha_{yy}}^2 \\ &\quad +2\alpha^\dagger_{xx}\alpha^\dagger_{yx}(\alpha_{xx}\alpha_{yx}-\alpha_{xy}\alpha_{yy})+{\alpha^\dagger_{xx}}^2({\alpha_{yy}}^2-{\alpha_{yx}}^2) \\ &\quad +2\alpha_{xx}\alpha_{yx}\alpha^\dagger_{xy}\alpha^\dagger_{yy}-2\alpha_{xy}\alpha^\dagger_{xy}\alpha_{yy}\alpha^\dagger_{yy} \\ &\quad +{\alpha_{xy}}^2{\alpha^\dagger_{yy}}^2-{\alpha_{xx}}^2({\alpha^\dagger_{yx}}^2+{\alpha^\dagger_{yy}}^2)\Big] \end{aligned} \tag{B.0.7}$$

$$\begin{aligned} M_{xx} &= \frac{1}{2D^2}\Big[{\alpha^\dagger_{yx}}^2{\alpha_{xy}}^2+{\alpha_{yx}}^2{\alpha^\dagger_{xy}}^2-{\alpha^\dagger_{xy}}^2{\alpha_{yy}}^2 \\ &\quad +2\alpha^\dagger_{xx}\alpha^\dagger_{yx}(\alpha_{xx}\alpha_{yx}-\alpha_{xy}\alpha_{yy})+{\alpha^\dagger_{xx}}^2({\alpha_{yy}}^2-{\alpha_{yx}}^2) \\ &\quad -2\alpha_{xx}\alpha_{yx}\alpha^\dagger_{xy}\alpha^\dagger_{yy}+2\alpha_{xy}\alpha^\dagger_{xy}\alpha_{yy}\alpha^\dagger_{yy} \\ &\quad -{\alpha_{xy}}^2{\alpha^\dagger_{yy}}^2+{\alpha_{xx}}^2({\alpha^\dagger_{yy}}^2-{\alpha^\dagger_{yx}}^2)\Big] \end{aligned} \tag{B.0.8}$$

$$\begin{aligned} M_{xy} &= \frac{1}{2D^2}\big[\alpha^\dagger_{xx}(\alpha_{xx}\alpha_{yx}\alpha^\dagger_{yy}-{\alpha_{yx}}^2\alpha^\dagger_{xy} \\ &\quad +\alpha_{yy}(\alpha^\dagger_{xy}\alpha_{yy}-\alpha_{xy}\alpha^\dagger_{yy}))+\alpha^\dagger_{yx}(\alpha_{xx}\alpha_{yx}\alpha^\dagger_{xy}-\alpha_{xx}{}^2\alpha^\dagger_{yy} \\ &\quad +\alpha_{xy}(\alpha_{xy}\alpha^\dagger_{yy}-\alpha^\dagger_{xy}\alpha_{yy}))\big] \end{aligned} \tag{B.0.9}$$

$$M_{y0} = \frac{1}{2D^2}\Big[-{\alpha^\dagger_{xx}}^2\alpha_{yx}\alpha_{yy} + \alpha^\dagger_{xx}\alpha^\dagger_{yx}(\alpha_{yx}\alpha_{xy} + \alpha_{xx}\alpha_{yy}) + \alpha_{yx}\alpha^\dagger_{xy}(\alpha_{xy}\alpha^\dagger_{yy} - \alpha^\dagger_{xy}\alpha_{yy}) - \alpha_{xx}({\alpha^\dagger_{yx}}^2\alpha_{xy} - \alpha^\dagger_{xy}\alpha_{yy}\alpha^\dagger_{yy} + \alpha_{xy}{\alpha^\dagger_{yy}}^2)\Big] \tag{B.0.10}$$

$$M_{yx} = \frac{1}{2D^2}\Big[-{\alpha^\dagger_{xx}}^2\alpha_{yx}\alpha_{yy} + \alpha^\dagger_{xx}\alpha^\dagger_{yx}(\alpha_{yx}\alpha_{xy} + \alpha_{xx}\alpha_{yy}) + \alpha_{yx}\alpha^\dagger_{xy}(\alpha^\dagger_{xy}\alpha_{yy} - \alpha_{xy}\alpha^\dagger_{yy}) - \alpha_{xx}({\alpha^\dagger_{yx}}^2\alpha_{xy} + \alpha^\dagger_{xy}\alpha_{yy}\alpha^\dagger_{yy} - \alpha_{xy}{\alpha^\dagger_{yy}}^2)\Big] \tag{B.0.11}$$

$$M_{yy} = \frac{1}{2D^2}\big[\alpha_{yx}(\alpha^\dagger_{yx}\alpha_{xy}\alpha^\dagger_{xy} - 2\alpha^\dagger_{xx}\alpha^\dagger_{xy}\alpha_{yy} + \alpha^\dagger_{xx}\alpha_{xy}\alpha^\dagger_{yy}) + \alpha_{xx}(\alpha^\dagger_{yx}\alpha^\dagger_{xy}\alpha_{yy} - 2\alpha^\dagger_{yx}\alpha_{xy}\alpha^\dagger_{yy} + \alpha^\dagger_{xx}\alpha_{yy}\alpha^\dagger_{yy})\big] \tag{B.0.12}$$

In the above equations

$$D = \det(\mathbf{S}_\alpha) = \alpha_{xx}\alpha_{yy} - \alpha_{yx}\alpha_{xy} \tag{B.0.13}$$

denotes the determinant of the real-space sampling matrix in equation (B.0.1).

# Appendix C

# Derivation of a novel tilt-induction formula

## C.1 Explicit derivation of a general tilt-induction formula

The following mathematical considerations derive a general expression for the induction of lower-order aberrations $c'_{mn}$ by a tilt $\vec{t}$ of the incident beam. For this purpose the two-dimensional vectors of the beam-tilt $\vec{t}$ and the spatial frequency $\vec{g}$ will be noted as complex numbers: $g = g_x + \mathrm{i}\, g_y$, $t = t_x + \mathrm{i}\, t_y$. In terms of the dimension less diffraction angle $\omega = \lambda \cdot g$ the aberration function $\chi_M(\omega)$ up to the $M$-th polynomial order of aberrations is the summation of all relevant single aberration phases $\chi_{mn}(\omega)$ (see equation 2.2.5):

$$\chi_M(\omega) = \frac{2\pi}{\lambda} \Re \left[ \sum_{l=1}^{M} \sum_{k=0}^{M(l)} (\omega^*)^l (\omega)^k c_{(l+k)(l-k)} \frac{1}{l+k} \right], \tag{C.1.1}$$

$$M(l) = \min(l, M-l), \text{ with } l, M \in \mathcal{N}_0. \tag{C.1.2}$$

The function $M(l)$ limits the number of relevant aberrations.
The derivation consists of four steps. Firstly, the aberration function $\chi$ is transformed into an analytical and symmetric shape, then the tilt is introduced into the argument of $\chi_M$. The main part deals with the expansion of the products of $(\omega + t)$ and re-factorisation into the original form of the dependence of $\chi_M$ on $\omega$ and its complex conjugate $\omega^*$. Finally the transformations done in the first step will be reversed to obtain the expression of the new effective aberration coefficients $c'_{(l+k)(l-k)}$.

### Transformations

Firstly, the limit to the aberration-order is removed, so that the index of the second sum also runs from 0 to $M$

$$\chi_M(\omega) = \frac{2\pi}{\lambda}\Re\left[\sum_{l=0}^{M}\sum_{k=0}^{l}(\omega^*)^l(\omega)^k c_{(l+k)(l-k)}\frac{1}{l+k}\right]. \tag{C.1.3}$$

This effectively includes also aberrations out of orders higher than $M$ in the aberration function. This expansion does not change the aberration function written in equation (C.1.1) if the added aberrations all have the value 0. The correct limits to the summation-indices will be introduced again in the formulation of the result at the end of this derivation.

In a second step the real-part formulation is replaced by summing the half of the expression C.1.3 and its complex conjugate:

$$\begin{aligned}
\chi_M(\omega) &= \frac{\pi}{\lambda}\left[\sum_{l=0}^{M}\sum_{k=0}^{l}(\omega^*)^l(\omega)^k c_{(l+k)(l-k)}\frac{1}{l+k}\right. && \text{(C.1.4)}\\
&\quad \left.+\sum_{l=0}^{M}\sum_{k=0}^{l}(\omega)^l(\omega^*)^k c^*_{(l+k)(l-k)}\frac{1}{l+k}\right]\\
&= \frac{\pi}{\lambda}\left[\sum_{l=0}^{M}\sum_{k=0}^{l}(\omega^*)^l(\omega)^k c_{(l+k)(l-k)}\frac{1}{l+k}\right. && \text{(C.1.5)}\\
&\quad \left.+\sum_{k=0}^{M}\sum_{l=0}^{k}(\omega)^k(\omega^*)^l c^*_{(l+k)(k-l)}\frac{1}{l+k}\right]\\
&= \frac{\pi}{\lambda}\sum_{l=0}^{M}\sum_{k=0}^{M}(\omega^*)^l(\omega)^k\frac{b_{lk}}{l+k} && \text{(C.1.6)}
\end{aligned}$$

equation C.1.6 combines the two double-sum expressions of the previous line into one double sum, and thereby obtains a symmetric and fully analytical formulation of the aberration function. The elements $b_{lk}$ of the substituted matrix $\mathbf{B}$ contain the old aberration coefficients in the following relation

$$b_{lk} = \begin{cases} c_{(l+k)(l-k)} & \forall\ l > k \\ 2c_{(2l)(0)} & \forall\ l = k \\ c^*_{(l+k)(k-l)} & \forall\ l < k \end{cases} \tag{C.1.7}$$

### Application of beam-tilts

In the Fourier-space formulation, tilt of the beam is equivalent to a translation of the spatial frequencies. This is expressed mathematically by replacing $\omega \rightarrow \omega + t$ in the

argument of the aberration function $\chi_N(\omega + t) = \chi_{N,t}(\omega)$ :

$$\chi_{M,t}(\omega) = \frac{\pi}{\lambda} \sum_{l=0}^{M} \sum_{k=0}^{M} (\omega^* + t^*)^l (\omega + t)^k \frac{b_{lk}}{l+k} \tag{C.1.8}$$

**Expansion and re-factorisation**

Expanding the two products into binomial series yields an expression with two additional summations

$$\chi_{M,t}(\omega) = \frac{\pi}{\lambda} \sum_{l=0}^{M} \sum_{i=0}^{l} \sum_{k=0}^{M} \sum_{j=0}^{k} \binom{l}{i} \binom{k}{j} (\omega^*)^{l-i} \omega^{k-j} (t^*)^i t^j \frac{b_{lk}}{l+k} \tag{C.1.9}$$

In order to return to the original $\omega$-dependency in the formulation of the aberration function, the terms are re-factorised into products of $(\omega^*)^\alpha \omega^\gamma$ yielding

$$\begin{aligned} \chi_{M,t}(\omega) &= \frac{\pi}{\lambda} \sum_{\alpha=0}^{M} \sum_{\gamma=0}^{M} (\omega^*)^\alpha \omega^\gamma \sum_{l=\alpha}^{M} \sum_{k=\gamma}^{M} \binom{l}{\alpha} \binom{k}{\gamma} (t^*)^{l-\alpha} t^{k-\gamma} \frac{b_{lk}}{l+k}, & \text{(C.1.10)} \\ &= \frac{\pi}{\lambda} \sum_{\alpha=0}^{M} \sum_{\gamma=0}^{M} (\omega^*)^\alpha \omega^\gamma \frac{b'_{\alpha\gamma}}{\alpha+\gamma}. & \text{(C.1.11)} \end{aligned}$$

Thereby the summations over $i$ and $j$ are performed on an alternative route, which is expressed by $\alpha = l - i$ and $\beta = k - j$. The binomial coefficients are transformed applying $\binom{n}{n-k} = \binom{n}{k}$. With equation (C.1.11) the aberration function including the tilt is transformed to the same shape as in equation (C.1.6), where the new effective coefficients $b'_{\alpha\gamma}$ are given by

$$b'_{\alpha\gamma} = (\alpha + \gamma) \sum_{\beta=\alpha}^{M} \sum_{\delta=\gamma}^{M} \binom{\beta}{\alpha} \binom{\delta}{\gamma} (t^*)^{\beta-\alpha} t^{\delta-\gamma} \frac{b_{\beta\delta}}{\beta+\delta}. \tag{C.1.12}$$

**Backwards transformations**

The original aberration coefficients are encoded in the $b_{lk}$'s according to equation (C.1.7):

$$\begin{aligned} c'_{(\alpha+\gamma)(\alpha-\gamma)} &= b'_{\alpha\gamma}, \quad \forall\ \alpha > \gamma \\ &= (\alpha + \gamma) \sum_{\beta=\alpha}^{M} \sum_{\delta=\gamma}^{M} \binom{\beta}{\alpha} \binom{\delta}{\gamma} (t^*)^{\beta-\alpha} t^{\delta-\gamma} \frac{b_{\beta\delta}}{\beta+\delta} & \text{(C.1.13)} \end{aligned}$$

In order to return to an expression containing only the original aberration coefficients on the right-hand side of equation (C.1.13) too, the double sum is split into two parts in the

reverse way, as it was created from equation (C.1.5) to equation (C.1.6). Thereby it is necessary to reduce the circular aberrations $c_{m0}$ by a factor 1/2, as they occur twice in the splitted sums. The reduction-function $\rho[\alpha,\gamma] = 1 - \delta[\alpha-\gamma]/2$ yields 1 for $\alpha \neq \gamma$ and $\rho[\alpha,\alpha] = 1/2$, where the function $\delta[i]$ is the discrete delta-function which is defined to be 1 for $i = 0$ and 0 otherwise.

$$\begin{aligned} c'_{(\alpha+\gamma)(\alpha-\gamma)} &= \rho[\alpha,\gamma]\cdot(\alpha+\gamma)\left[\sum_{\beta=\alpha}^{M}\sum_{\delta=\gamma}^{\beta}\binom{\beta}{\alpha}\binom{\delta}{\gamma}(t^*)^{\beta-\alpha}t^{\delta-\gamma}\frac{c_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right. \\ &\quad \left.+\sum_{\beta=\alpha}^{M}\sum_{\delta=\gamma}^{\beta}\binom{\delta}{\alpha}\binom{\beta}{\gamma}(t^*)^{\delta-\alpha}t^{\beta-\gamma}\frac{c^*_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right] \\ c'_{(\alpha+\gamma)(\alpha-\gamma)} &= \rho[\alpha,\gamma]\cdot(\alpha+\gamma)\sum_{\beta=\alpha}^{M}\sum_{\delta=\gamma}^{\beta}\left[\binom{\beta}{\alpha}\binom{\delta}{\gamma}(t^*)^{\beta-\alpha}t^{\delta-\gamma}\frac{c_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right. \\ &\quad \left.+\binom{\beta}{\gamma}\binom{\delta}{\alpha}(t^*)^{\delta-\alpha}t^{\beta-\gamma}\frac{c^*_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right] \end{aligned} \tag{C.1.14}$$

At this point the aberrations of orders higher than $M$ can be removed from the sum again by limiting the index of the second sum to $M(\beta)$:

$$\begin{aligned} c'_{(\alpha+\gamma)(\alpha-\gamma)} &= \rho[\alpha,\gamma]\cdot(\alpha+\gamma)\sum_{\beta=\alpha}^{M}\sum_{\delta=\gamma}^{M(\beta)}\left[\binom{\beta}{\alpha}\binom{\delta}{\gamma}(t^*)^{\beta-\alpha}t^{\delta-\gamma}\frac{c_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right. \\ &\quad \left.+\binom{\beta}{\gamma}\binom{\delta}{\alpha}(t^*)^{\delta-\alpha}t^{\beta-\gamma}\frac{c^*_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right] \end{aligned} \tag{C.1.15}$$

The effective aberration calculates out of the original aberration as an offset, plus a polynomial of the tilt angle with coefficients containing aberrations of all relevant orders, which are higher than the order of the effective aberration of interest.

## C.2 Application of the general tilt-induction formula

This section contains explicit expressions for effective lower order aberrations calculated from the general formula (C.1.15) up the $6^{\text{th}}$ order of aberrations.

### C.2.1 Effective image displacement

The dependence for $c'_{11} = c'_{11y} + \mathrm{i} \cdot c'_{11y}$ on all higher aberrations up to the order $M = 6$ is calculated by setting $\alpha = 1$ and $\gamma = 0$ in equation (C.1.15):

$$
\begin{aligned}
c'_{11} &= \sum_{\beta=1}^{6} \sum_{\delta=0}^{M(\beta)} \left[ \binom{\beta}{1} \binom{\delta}{0} (t^*)^{\beta-1} t^{\delta} \frac{c_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta} \right. && \text{(C.2.1)} \\
&\quad \left. + \binom{\beta}{0} \binom{\delta}{1} (t^*)^{\delta-1} t^{\beta} \frac{c^*_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta} \right], \\
&= \sum_{\beta=1}^{6} \sum_{\delta=0}^{\min(\beta,6-\beta)} \left[ \frac{\beta}{\beta+\delta} \, c_{(\beta+\delta)(\beta-\delta)} \cdot (t^*)^{\beta-1} t^{\delta} \right. && \text{(C.2.2)} \\
&\quad \left. + \frac{\delta}{\beta+\delta} \, c^*_{(\beta+\delta)(\beta-\delta)} \cdot (t^*)^{\delta-1} t^{\beta} \right], \\
&= c_{11} + c_{20} \cdot t + c_{22} \cdot t^* + \frac{2}{3} c_{31} \cdot (t^*) t + \frac{1}{3} c^*_{31} t^2 + c_{33} \cdot (t^*)^2 && \text{(C.2.3)} \\
&\quad + c_{40} \cdot (t^*) t^2 + \frac{3}{4} \, c_{42} \cdot (t^*)^2 t + \frac{1}{4} \, c^*_{42} \cdot t^3 + c_{44} \cdot (t^*)^3 \\
&\quad + \frac{3}{5} \, c_{51} \cdot (t^*)^2 t^2 + \frac{2}{5} \, c^*_{51} \cdot (t^*) t^3 + \frac{4}{5} \, c_{53} \cdot (t^*)^3 t + \frac{1}{5} \, c^*_{53} \cdot t^4 + c_{55} \cdot (t^*)^4 \\
&\quad + c_{60} \cdot (t^*)^2 t^3 + \frac{2}{3} \, c_{62} \cdot (t^*)^3 t^2 + \frac{1}{3} \, c^*_{62} \cdot (t^*) t^4 \\
&\quad + \frac{5}{6} \, c_{64} \cdot (t^*)^4 t + \frac{1}{6} \, c^*_{64} \cdot t^5 + c_{66} \cdot (t^*)^5 .
\end{aligned}
$$

Calculation of the real part $c'_{11x}$ and the imaginary part $c'_{11y}$ of the induced images shift yields

$$
\begin{aligned}
c'_{11x} &= \Re[c'_{11}] && \text{(C.2.4)}\\
&= c_{11x} + c_{20}\cdot t_x + c_{22x}\cdot t_x + c_{22y}\cdot t_y + \frac{1}{3}\,c_{31x}\cdot(3t_x^2+t_y^2) + \frac{2}{3}\,c_{31y}t_xt_y && \text{(C.2.5)}\\
&\quad + c_{33x}\cdot(t_x^2-t_y^2) + 2\,c_{33y}\cdot t_xt_y + c_{40}\cdot t_x(t_x^2+t_y^2)\\
&\quad + c_{42x}\cdot t_x^3 + \frac{1}{2}\,c_{42y}\cdot(3t_x^2t_y+t_y^3) + c_{44x}\cdot(t_x^3-3t_xt_y^2) + c_{44y}(3t_x^2t_y-t_y^3)\\
&\quad + \frac{1}{5}\,c_{51x}\cdot(5t_x^4+6t_x^2t_y^2+t_y^4) + \frac{4}{5}\,c_{51y}\cdot(t_x^3t_y+t_xt_y^3)\\
&\quad + \frac{1}{5}\,c_{53x}\cdot(5t_x^4-6t_x^2t_y^2-3t_y^4) + \frac{4}{5}\,c_{53y}\cdot(3t_x^3t_y+t_xt_y^3)\\
&\quad + c_{55x}\cdot(t_x^4-6t_x^2t_y^2+t_y^4) + 4\,c_{55y}\cdot(t_x^3t_y-t_xt_y^3) + c_{60}\cdot(t_x^5+2t_x^3t_y^2+t_xt_y^4)\\
&\quad + \frac{1}{3}\,c_{62x}\cdot(3t_x^5+2t_x^3t_y^2-t_xt_y^4) + \frac{1}{3}\,c_{62y}\cdot(5t_x^4t_y+6t_x^2t_y^3+t_y^5)\\
&\quad + \frac{1}{3}\,c_{64x}\cdot(3t_x^5-10t_x^3t_y^2-5t_xt_y^4) + \frac{1}{3}\,c_{64y}\cdot(10t_x^4t_y-2t_y^5)\\
&\quad + c_{66x}\cdot(t_x^5-10t_x^3t_y^2+5t_xt_y^4) + c_{66y}\cdot(5t_x^4t_y-10t_x^2t_y^3+t_y^5),
\end{aligned}
$$

and

$$
\begin{aligned}
c'_{11y} &= \Im[c'_{11}] && \text{(C.2.6)}\\
&= c_{11y} + c_{20}\cdot t_y - c_{22x}\cdot t_y + c_{22y}\cdot t_x + \frac{2}{3}\,c_{31x}\cdot t_xt_y + \frac{1}{3}\,c_{31y}(t_x^2+3t_y^2) && \text{(C.2.7)}\\
&\quad - 2\,c_{33x}\cdot t_xt_y + c_{33y}\cdot(t_x^2-t_y^2) + c_{40}\cdot t_y(t_x^2+t_y^2)\\
&\quad - c_{42x}\cdot t_y^3 + \frac{1}{2}\,c_{42y}\cdot(t_x^3+3t_xt_y^2) + c_{44x}\cdot(-3t_x^2t_y+t_y^3) + c_{44y}(t_x^3-3t_xt_y^2)\\
&\quad + \frac{4}{5}\,c_{51x}\cdot(t_x^3t_y+t_xt_y^3) + \frac{1}{5}\,c_{51y}\cdot(t_x^4+6t_x^2t_y^2+5t_y^4)\\
&\quad - \frac{4}{5}\,c_{53x}\cdot(t_x^3t_y+3t_xt_y^3) + \frac{1}{5}\,c_{53y}\cdot(3t_x^4+6t_x^2t_y^2-5t_y^4)\\
&\quad - 4\,c_{55x}\cdot(t_x^3t_y-t_xt_y^3) + c_{55y}\cdot(t_x^4-6t_x^2t_y^2+t_y^4) + c_{60}\cdot(t_x^4t_y+2t_x^2t_y^3+t_y^5)\\
&\quad + \frac{1}{3}\,c_{62x}\cdot(t_x^4t_y-2t_x^2t_y^3-3t_y^5) + \frac{1}{3}\,c_{62y}\cdot(t_x^5+6t_x^3t_y^2+5t_xt_y^4)\\
&\quad - \frac{1}{3}\,c_{64x}\cdot(5t_x^4t_y+10t_x^2t_y^3-3t_y^5) + \frac{1}{3}\,c_{64y}\cdot(2t_x^5-10t_xt_y^4)\\
&\quad - c_{66x}\cdot(5t_x^4t_y-10t_x^2t_y^3+t_y^5) + c_{66y}\cdot(t_x^5-10t_x^3t_y^2+5t_xt_y^4).
\end{aligned}
$$

### C.2.2 Effective defocus

The dependence for $c'_{20}$ on all higher aberrations up to the full order $M = 6$ is calculated according to equation (3.2.8) by setting $\alpha = 1$ and $\gamma = 1$:

$$\begin{aligned} c'_{20} &= \frac{1}{2}\sum_{\beta=1}^{6}\sum_{\delta=1}^{M(\beta)}\left[\binom{\beta}{1}\binom{\delta}{1}(t^*)^{\beta-1}t^{\delta-1}\frac{c_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right. && \text{(C.2.8)}\\ &\quad \left.+\binom{\beta}{1}\binom{\delta}{1}(t^*)^{\delta-1}t^{\beta-1}\frac{c^*_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}\right], \\ &= \frac{1}{2}\sum_{\beta=1}^{6}\sum_{\delta=1}^{\min(\beta,6-\beta)}\left[\frac{\beta\delta}{\beta+\delta}\,c_{(\beta+\delta)(\beta-\delta)}\cdot(t^*)^{\beta-1}t^{\delta-1}\right. && \text{(C.2.9)}\\ &\quad \left.+\frac{\delta\beta}{\beta+\delta}\,c^*_{(\beta+\delta)(\beta-\delta)}\cdot(t^*)^{\delta-1}t^{\beta-1}\right], \\ &= c_{20}+\frac{2}{3}\,c_{31}\cdot t^*+\frac{2}{3}\,c^*_{31}t+2\,c_{40}\cdot(t^*)t+\frac{3}{4}\,c_{42}\cdot(t^*)^2+\frac{3}{4}\,c^*_{42}\cdot t^2 && \text{(C.2.10)}\\ &\quad +\frac{6}{5}\,c_{51}\cdot(t^*)^2t+\frac{6}{5}\,c^*_{51}\cdot(t^*)t^2+\frac{4}{5}\,c_{53}\cdot(t^*)^3+\frac{4}{5}\,c^*_{53}\cdot t^3+3\,c_{60}\cdot(t^*)^2t^2 \\ &\quad +\frac{4}{3}\,c_{62}\cdot(t^*)^3t+\frac{4}{3}\,c^*_{62}\cdot(t^*)t^3+\frac{5}{6}\,c_{64}\cdot(t^*)^4+\frac{5}{6}\,c^*_{64}\cdot t^4. \end{aligned}$$

The above expression for the effective defocus is a real number:

$$\begin{aligned} c'_{20} &= c_{20}+\frac{4}{3}\,c_{31x}\cdot t_x+\frac{4}{3}\,c_{31y}t_y+2\,c_{40}\cdot(t_x^2+t_y^2)+\frac{3}{2}\,c_{42x}\cdot(t_x^2-t_y^2) && \text{(C.2.11)}\\ &\quad +3\,c_{42y}\cdot t_xt_y+\frac{12}{5}\,c_{51x}\cdot(t_x^3+t_xt_y^2)+\frac{12}{5}\,c_{51y}\cdot(t_x^2t_y+t_y^3) \\ &\quad +\frac{8}{5}\,c_{53x}\cdot(t_x^3-3t_xt_y^2)+\frac{8}{5}\,c_{53y}\cdot(3t_x^2t_y-t_y^3)+3\,c_{60}\cdot(t_x^2+t_y^2)^2 \\ &\quad +\frac{8}{3}\,c_{62x}\cdot(t_x^4-t_y^4)+\frac{16}{3}\,c_{62y}\cdot(t_x^3t_y+t_xt_y^3) \\ &\quad +\frac{5}{3}\,c_{64x}\cdot(t_x^4-6t_x^2t_y^2+t_y^4)+\frac{20}{3}\,c_{64y}\cdot(t_x^3t_y-t_yt_y^3). \end{aligned}$$

### C.2.3 Effective twofold astigmatism

The dependence for $c'_{22} = c'_{22y} + \mathrm{i} \cdot c'_{22y}$ on all higher aberrations up to the order $M = 6$ is calculated by setting $\alpha = 2$ and $\gamma = 0$:

$$
\begin{aligned}
c'_{22} &= \sum_{\beta=2}^{6} \sum_{\delta=0}^{M(\beta)} \binom{\beta}{2}\binom{\delta}{0} (t^*)^{\beta-2} t^{\delta} \frac{c_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta} + \binom{\beta}{0}\binom{\delta}{2} (t^*)^{\delta-2} t^{\beta} \frac{c^*_{(\beta+\delta)(\beta-\delta)}}{\beta+\delta}, \\
&= \frac{1}{4} \sum_{\beta=2}^{6} \sum_{\delta=0}^{\min(\beta,6-\beta)} \left[ \frac{\beta(\beta-1)}{\beta+\delta}\, c_{(\beta+\delta)(\beta-\delta)} \cdot (t^*)^{\beta-2} t^{\delta} \right. \qquad \text{(C.2.12)} \\
&\quad \left. + \frac{\delta(\delta-1)}{\beta+\delta}\, c^*_{(\beta+\delta)(\beta-\delta)} \cdot (t^*)^{\delta-2} t^{\beta} \right], \\
&= c_{22} + \frac{2}{3} c_{31} \cdot t + 2\, c_{33} \cdot t^* + c_{40} \cdot t^2 + \frac{3}{2}\, c_{42} \cdot (t^*) t + 3\, c_{44} \cdot (t^*)^2 \qquad \text{(C.2.13)} \\
&\quad + \frac{6}{5}\, c_{51} \cdot (t^*) t^2 + \frac{2}{5}\, c^*_{51} \cdot t^3 + \frac{12}{5}\, c_{53} \cdot (t^*)^2 t + 4\, c_{55} \cdot (t^*)^3 \\
&\quad + 2\, c_{60} \cdot (t^*) t^3 + 2\, c_{62} \cdot (t^*)^2 t^2 + \frac{1}{3}\, c^*_{62} \cdot t^4 + \frac{10}{3}\, c_{64} \cdot (t^*)^3 t + 5\, c_{66} \cdot (t^*)^4.
\end{aligned}
$$

Calculation of the real part $c'_{22x}$ and the imaginary part $c'_{22y}$ yields

$$
\begin{aligned}
c'_{22x} &= \Re[c'_{22}] \qquad \text{(C.2.14)} \\
&= c_{22x} + \frac{2}{3}\, c_{31x} \cdot t_x - \frac{2}{3}\, c_{31y} t_y + 2\, c_{33x} \cdot t_x + 2\, c_{33y} \cdot t_y \qquad \text{(C.2.15)} \\
&\quad + c_{40} \cdot (t_x^2 - t_y^2) + \frac{3}{2}\, c_{42x} \cdot (t_x^2 + t_y^2) + 3\, c_{44x} \cdot (t_x^2 - t_y^2) + 6\, c_{44y} t_x t_y \\
&\quad + \frac{8}{5}\, c_{51x} \cdot t_x^3 - \frac{8}{5}\, c_{51y} \cdot t_y^3 + \frac{12}{5}\, c_{53x} \cdot (t_x^3 + t_x t_y^2) + \frac{12}{5}\, c_{53y} \cdot (t_x^2 t_y + t_y^3) \\
&\quad + 4\, c_{55x} \cdot (t_x^3 - 3 t_x t_y^2) + 4\, c_{55y} \cdot (3 t_x^2 t_y - t_y^3) + 2\, c_{60} \cdot (t_x^4 - t_y^4) \\
&\quad + \frac{1}{3}\, c_{62x} \cdot (7 t_x^4 + 6 t_x^2 t_y^2 + 7 t_y^4) + \frac{4}{3}\, c_{62y} \cdot (t_x^3 t_y - t_x t_y^3) \\
&\quad + \frac{10}{3}\, c_{64x} \cdot (t_x^4 - t_y^4) + \frac{20}{3}\, c_{64y} \cdot (t_x^3 t_y + t_x t_y^3) \\
&\quad + 5\, c_{66x} \cdot (t_x^4 - 6 t_x^2 t_y^2 + t_y^4) + 20\, c_{66y} \cdot (t_x^3 t_y - t_x t_y^3),
\end{aligned}
$$

and

$$
\begin{aligned}
c'_{22y} &= \Im[c'_{22}] && \text{(C.2.16)}\\
&= c_{22y} + \frac{2}{3}\,c_{31x}\cdot t_y + \frac{2}{3}\,c_{31y}t_x - 2\,c_{33x}\cdot t_y + 2\,c_{33y}\cdot t_x && \text{(C.2.17)}\\
&\quad +2\,c_{40}\cdot t_x t_y + \frac{3}{2}\,c_{42y}\cdot(t_x^2+t_y^2) - 6\,c_{44x}\cdot t_x t_y + 3\,c_{44y}(t_x^2-t_y^2)\\
&\quad +\frac{4}{5}\,c_{51x}\cdot(3t_x^2t_y+t_y^3) + \frac{4}{5}\,c_{51y}\cdot(t_x^3+3t_xt_y^2)\\
&\quad -\frac{12}{5}\,c_{53x}\cdot(t_x^2t_y+t_y^3) + \frac{12}{5}\,c_{53y}\cdot(t_x^3+t_xt_y^2)\\
&\quad -4\,c_{55x}\cdot(3t_x^2t_y-t_y^3) + 4\,c_{55y}\cdot(t_x^3-3t_xt_y^2) + 4\,c_{60}\cdot(t_x^3t_y+t_xt_y^3)\\
&\quad +\frac{4}{3}\,c_{62x}\cdot(t_x^3t_y-4t_xt_y^3) + \frac{1}{3}\,c_{62y}\cdot(5t_x^4+6t_x^2t_y^2+5t_y^4)\\
&\quad -\frac{20}{3}\,c_{64x}\cdot(t_x^3t_y+t_xt_y^3) + \frac{10}{3}\,c_{64y}\cdot(t_x^4-t_y^4)\\
&\quad -20\,c_{66x}\cdot(t_x^3t_y-t_xt_y^3) + 5\,c_{66y}\cdot(t_x^4-6t_x^2t_y^2+t_y^4).
\end{aligned}
$$

# Bibliography


[1] M. Knoll, E. Ruska, *Das Elektronenmikroskop.* Z. techn. Physik 12, 389-400 (1931)

[2] O. Scherzer, *Über einige Fehler von Elektronenlinsen.* Z. Phys 101, 593-603 (1936)

[3] D. Van Dyck, *Fast computational procedures for the simulation of structure images in complex or disordered crystals: a new approach.* J. Microscopy 119, 141-152 (1980)

[4] R. Kilaas, R. Gronsky, *Real space image simulation in High-Resolution Electron Microscopy.* Ultramicroscopy 11, 289-298 (1983)

[5] W. Coene, D. Van Dyck, *The real space method for dynamical electron diffraction calculations in high resolution transmission electron microscopy. II Critical Analysis of the dependency on the input parameters.* Ultramicroscopy 15, 41-50 (1984)

[6] M.A. O'Keefe, R. Kilaas, *Advances in High-Resolution Image Simulation.* In: *Image and Signal Processing in Electron Microscopy* (edts. P.W. Hawkes, F.P. Ottensmeyer, W.O. Saxton, A. Rosenfeld) (Scanning Microsc. Suppl. **2**, 225-244)1988

[7] P.A. Stadelmann, *EMS – a software technique for electron diffraction analysis and HREM image simulation in materials science.* Ultramicroscopy 21, 131 (1987)

[8] F. Phillipp, R. Höschen, M. Osaki, G. Möbus, M. Rühle, *New high-voltage atomic resolution microscope approaching 1 Å point resolution installed in Stuttgard.* Ultramicroscopy 51, 1-10 (1994)

[9] D. Gabor, *A New Microscopy Principle.* Nature 161, 777-778, No. 4098 (1948)

[10] H. Lichte, *Electron holography approaching atomic resolution.* Ultramicroscopy 20, 293 (1986)

[11] H. Rose, *Elektronenoptische Aplanate.* Optik 34, 287-313 (1971)

[12] M. Haider, G. Braunshausen, E. Schwan, *Correction of the spherical aberration of a 200kV TEM by means of a Hexapole-corrector.* Optik 99, 167-179 (1995)

[13] M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius, K. Urban, *A spherical-aberration-corrected 200 kV transmission electron microscope.* Ultramicroscopy 75, 53-60 (1998)

[14] P. Schiske, *Zur Frage der Bildrekonstruktion durch Fokusreihen.* In: Proc. 4th Eur. Conf. on Electron Microscopy, 145, Rom (1968)

[15] W.O. Saxton, *Computer Techniques for the Image Processing in Electron Microscopy.* Academic Press, New York, Chap. 9.7 (1978)

[16] E.J. Kirkland, *Improved high-resolution image processing for bright field electron micrographs, I. Theory.* Ultramicroscopy 15, 115-172 (1984)

[17] W.M.J. Coene, G. Janssen, M. Op de Beeck, D. Van Dyck, *Phase Retrieval through Focus Variation for Ultra-Resolution in Field-Emission Transmission Electron Microscopy.* Phys. Rev. Lett. 69, 3743-3746 (1992)

[18] D. Van Dyck, M. Op de Beeck, W.M.J. Coene, *A new approach to object wavefunction reconstruction in electron microscopy.* Optik 93, 103 (1993)

[19] A. Thust, M.H.F. Overwijk, W.M.J. Coene, M. Lentzen, *Numerical correction of lens aberrations in phase-retrieval HRTEM.* Ultramicroscopy 64, 249-264 (1996)

[20] S. Uhlemann, M. Haider, *Residual wave aberrations in the first spherical aberration corrected transmission electron microscope.* Ultramicroscopy 72, 109-119 (1998)

[21] L. Reimer, *Transmission Electron Microscopy.* Springer, Berlin (1993)

[22] J.C.H. Spence, *Experimental High Resolution Transmission Electron Microscopy.* Oxford University Press, New York, Oxford (1988)

[23] D.B. Williams, C.B. Carter, *Transmission Electron Microscopy.* Plenum Press, New York (1996)

[24] A. Thust, *Grundlagen und fortgeschrittene Methoden der hochauflösenden Elektronenmikroskopie.* Physik der Nanostrukturen, 29. Ferienkurs des Instituts für Festkörperforschung, in A.1, Jülich (1998)

[25] S. Kujawa, D. Krahl, *Performance of a low-noise CCD camera adapted to a transmission electron microscope.* Ultramicroscopy 46, 395 (1992)

[26] O.L. Krivanek, P.E. Mooney, *Applications of slow-scan CCD cameras in transmission electron microscopy.* Ultramicroscopy 49, 95 (1993)

[27] O. Scherzer, *Sphärische und chromatische Korrektur von Elektronenlinsen.* Optik 2, 114-132 (1947)

[28] H. Koops, G. Kuck, O. Scherzer, *Erprobung eines elektronenoptischen Achromators.* Optik 48, 225 (1977)

[29] H. Hely, *Messungen an einem verbesserten korrigierten Elektronenmikroskop.* Optik 60, 353-370 (1982)

[30] H. Hely, *Technologische Voraussetzungen für die Verbesserung der Korrektur von Elektronenlinsen.* Optik 60, 307-326 (1982)

[31] H. Rose, *Outline of a spherically corrected semiaplanatic medium-voltage TEM.* Optik 85, 19-24 (1990)

[32] G. Benner, M. Matijevic, A. Orchowski, P. Schlossmacher, A. Thesen, M. Haider, P. Hartel, *Sub-Ångstrom and sub-eV resolution with the analytical SATEM.* Microscopy and Microanalysis 10, (Suppl. 3) 6-7 (2004)

[33] B. Freitag, S. Kujawa, P.M. Mul, J. Ringnalda, P.C. Tiemeijer, *Breaking the spherical and chromatic aberration barrier in transmission electron microscopy.* Ultramicroscopy 102, 209-214 (2005)

[34] A.E. Thesen, M. Matijevic, G. Benner, *Design Features and Ultimate Performance of an Ultra-High Resolution Aberration-Corrected, Monochromatized 200 keV FEG-TEM.* Microscopy and Microanalysis 11, 2144-2145 (2005)

[35] B. Freitag, S. Kujawa, P.M. Mul, P.C. Tiemeijer, *First experimental proof of spatial resolution improvement in a monochromized and Cs-corrected TEM.* Microscopy and Microanalysis 10, (Suppl. 3) 4-5 (2004)

[36] F. Kahl, H. Rose, *Design of a monochromator for electron sources.* Proc. EUREM Brno/Cz III, 1458 (2000)

[37] H. Rose, *Prospects for aberration-free electron microscopy.* Ultramicroscopy 103, 1-6 (2005)

[38] *The TEAM project.* http://ncem.lbl.gov/team3.htm (July, 17, 2006)

[39] O.L. Krivanek, N. Dellby, and A.R. Lupini, *Towards sub-Å electron beams.* Ultramicroscopy 78, 1-11 (1999)

[40] O.L. Krivanek, P.D. Nellist, N. Dellby, M.F. Murfitt, Z. Szilagyi, *Towards Sub-0.5 Å electron beams.* Ultramicroscopy 96, 229-237 (2003)

[41] H. Bethe, *Vergleich der Elektronenbewegung im Heliumzustand nach verschiedenen Methoden.* Ann. Phys, 4. Folge 87, 55-129 (1928)

[42] C.J. Humphreys, *The scattering of fast electrons by crystals.* Rep. Prog. Phys. 42, 1864 (1979)

[43] J.M. Cowley, A.F. Moodie, *The scattering of electrons by atoms and crystals. I. A new theoretical approach.* Acta Cryst. 10, 609-619 (1957)

[44] P. Goodman, A.F. Moodie, *Numerical evaluation of n-beam wave functions in electron scattering by multislice method.* Acta Cryst. 30, 280-290 (1974)

[45] A. Thust, *Methodische Untersuchungen zur quantitativen Bildauswertung und zur Rekonstruktion der Austrittswellenfunktion in der hochauflösenden Elektronenmikroskopie.* Dissertation, RWTH-Aachen (1994)

[46] O. Scherzer, *The Theoretical Resolution Limit of the Electron Microscope.* J. Appl. Phys. 20, 20-29 (1949)

[47] M. Lentzen, B. Jahnen, C.L. Jia, K. Tillmann, K. Urban, *High-resolution imaging with an aberration-corrected transmission electron microscope.* Ultramicroscopy 92, 233-242 (2002)

[48] C.L. Jia, M. Lentzen, K. Urban, *Atomic resolution imaging of oxygen in perovskite ceramics.* Science 299, 870-873 (2003)

[49] C.-L. Jia, M. Lentzen, K. Urban, *High-Resolution Transmission Electron Microscopy Using Negative Spherical Aberration.* Microscopy and Microanalysis 10, 174-184 (2004)

[50] M. Lentzen, *Progress in Aberration-Corrected High-Resolution Transmission Electron Microscopy Using Hardware Aberration Correction.* Microscopy and Microanalysis 12, 191-205 (2006)

[51] M. Born, E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light.* Pergamon Press, 6$^{\text{th}}$ edition (1980)

[52] H. Lichte, *Optimum focus for taking electron holograms.* Ultramicroscopy 38, 13-22 (1991)

[53] J. Frank, *The Envelope of Electron Microscopic Transfer Functions for Partially Coherent Illumination.* Optik 38, 519-536 (1973)

[54] K. Ishizuka, *Contrast transfer of crystal images in TEM.* Ultramicroscopy 5, 55-65 (1980)

[55] J.C.H. Spence, *Experimental high-resolution electron microscopy.* Oxford University Press, New York, Oxford (1988)

[56] W.O. Saxton, *Spatial coherence in axial high resolution conventional electron microscopy.* Optik 49, 51-62 (1977)

[57] M. Troyon, *A method for determining the illumination divergence from electron micrographs.* Optik 49, 247-251 (1977)

[58] J.E. Bonevich, L.D. Marks, *Contrast transfer theory for non-linear imaging.* Ultramicroscopy 26, 313-320 (1988)

[59] J.E. Bonevich, L.D. Marks, *Contrast transfer theory for non-linear imaging (addendum).* Ultramicroscopy 32, 297 (1990)

[60] J. Frank, *Nachweis von Objektbewegungen im lichtoptischen Diffraktogramm von elektronenmikroskopischen Aufnahmen.* Optik 30, 171-180 (1969)

[61] D. Van Dyck, A.F. de Jong, *Ultimate resolution and information in electron microscopy: general principles.* Ultramicroscopy 47, 266-281 (1992)

[62] A.F. de Jong, D. Van Dyck, *Ultimate resolution and information in electron microscopy: II. The information limit of transmission electron microscopes.* Ultramicroscopy 49, 66-80 (1993)

[63] P.W. Hawkes, E. Kasper, *Principles of Electron Optics.* Ch. 31, Academic Press, London (1989)

[64] H. Rose, E. Plies, *Correction of aberrations in electron optical systems with curved axes.* In: Image rocessing and Computer-Aided Design in Electron Optics (ed. by P.W. Hawkes), pp 344-369, Academic Press, London (1973)

[65] W. Bernhard, *Erprobung eines sphärisch und chromatisch korrigierten Elektronenmikroskopes.* Optik 57, 73-94 (1980)

[66] O.L. Krivanek, P.A. Stadelmann, *Effect of threefold astigmatism on high-resolution electron micrographs.* Ultramicroscopy 60, 103-113 (1995)

[67] W.O.Saxton, *Observation of lens aberrations for very high-resolution electron microscopy. I Theory.* Journal of Microscopy 179, 201-213 (1995)

[68] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, K.H. Herrmannn, *Coma-free alignment of high resolution electron microscopes with the aid of optical diffractogramms.* Ultramicroscopy 3, 49-60 (1978)

[69] D. Typke, K. Dierksen, *Determination of image aberrations in high-resolution electron microscopy using diffractogram and cross-correlation methods.* Optik 99, 155-166 (1995)

[70] G. Möbus, *Retrieval of crystal defect structures from HREM images by simulated evolution. I. Basic technique.* Ultramicroscopy 65, 205-216 (1996)

[71] P. Hartel, H. Müller, S. Uhlemann, M. Haider, *Residual aberrations of hexapole-type Cs-Correctors.* Corrected Electron Optical Systems GmbH, Heidelberg, Germany, http://www.ceos-gmbh.de/poster/residualCEXCOR.html (August, 21, 2006)

[72] A. Thust, M. Lentzen, K. Urban, *Non-linear reconstruction of the exit plane wavefunction from periodic high-resolution electron microscopy images.* Ultramicroscopy 53, 101 (1994)

[73] A.Thust, W.M.J. Coene, M. Op de Beeck, D. Van Dyck, *Focal-series reconstruction in HRTEM: simulation studies on non-periodic objects.* Ultramicroscopy 64, 211 (1996)

[74] W.M.J. Coene, A. Thust, M. Op de Beeck, D. Van Dyck, *Maximum-likelihood method for focus-variation image reconstruction in high resolution transmission electron microscopy.* Ultramicroscopy 64, 109-135 (1996)

[75] C. Kisielowski, C.J.D Hetherington, Y.C. Wang, R. Kilaas, M.A. O'Keefe, A. Thust, *Imaging columns of the light elements carbon, nitrogen and oxygen at sub-Ångstrom resolution.* Ultramicroscopy 89, 243 (2001)

[76] TrueImage is a software for the automated reconstruction of the exit-plane wave-function from focal-series distributed by the FEI company.

[77] Lord Rayleigh, *Investigations in optics with special reference to the spectroscope.* Phil. Mag. (5) 8, 403 (1879)

[78] A. Maréchal, *Study of the combined effects of diffraction an geometrical aberrations on the image of a luminous point.* Rev. d'Optique 26, 257-277 (1947)

[79] *E.W. Weisstein,* Error Propagation. *From* MathWorld – *A Wolfram Web Resource..* http://mathworld.wolfram.com/ErrorPropagation.html (August, 20, 2006)

[80] D. Stenkamp, *Detection and quantitative assessment of image aberrations from single HRTEM lattice images.* Journal of Microscopy 190, 194-203 (1998)

[81] M. Hohenstein, *Single Sideband Imaging in High-Resolution Electron Microscopy.* Appl. Phys. A 54, 485-492 (1992)

[82] K.J. Hanszen, L. Trepte, *Die Kontrastübertragung im Elektronenmikroskop bei partiell kohärenter Beleuchtung.* Optik 33, 166 and 182 (1971)

[83] P.A. Doyle, P.S. Turner, *Relativistic Hartree-Fock X-ray and Electron Scattering Factors.* Acta Cryst. 24, 390-397 (1968)

[84] A. Weickenmeier, H. Kohl, *Computation of Absorptive Form Factors for High-Energy Electron Diffraction.* Acta Cryts. 47, 590-587 (1991)

[85] F. Thon, *Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung.* Z. Naturforsch. 21a, 476 (1966)

[86] A.J. Koster, A.F. de Jong, *Measurement of the spherical aberration coefficient of transmission electron microscopes by beam-tilt-induced image displacements.* Ultramicroscopy 38, 235-240 (1991)

[87] A.J. Koster, W.J. de Ruijter, *Practical autoalignment of transmission electron microscopes.* Ultramicroscopy 40, 89-107 (1992)

[88] A. Steinecker, W. Mader, *Measurement of lens aberrations by means of image displacements in beam-tilt series.* Ultramicroscopy 81, 149-161 (2000)

[89] J.W. Cooley, J.W. Turkey, *An Algorithm for the Machine Calculation of Complex Fourier Series.* Mathematics of Computation 19, 297-301 (1965)

[90] F. Zemlin, *A practical procedure for alignment of a high resolution electron microscope.* Ultramicroscopy 4, 241-245 (1979)

[91] O.L. Krivanek, *A method for determining the coefficient of spherical aberration from a single electron micrograph.* Optik 45, 97-101 (1976)

[92] W.M.J. Coene, T.J.J. Denteneer, *Improved methods for the determination of the spherical aberration coefficient in high-resolution electron microscopy from micrographs of an amorphous object.* Ultramicroscopy 38, 225-233 (1991)

[93] O.L. Krivanek, G.Y. Fan, *Application of slow-scan charge-coupled device (CCD) cameras to microscope control.* Scanning Microscopy, Suppl. 6, 105-114 (1992)

[94] K. Ishizuka, *Coma-free alignment of a high-resolution electron microscope with threefold astigmatism.* Ultramicroscopy 55, 407-418 (1994)

[95] W.O. Saxton, *A new way of measuring microscope aberrations.* Ultramicroscopy 81, 41-45 (2000)

[96] A. Thust, C.L. Jia, K. Urban, *Direct Determination of Imaging Parameters from Wave Functions in HRTEM.* Microscopy and Microanalysis 9, Suppl. 2, 140-141 (2003)

[97] Q. Fu, H. Lichte, *Holographic measurement of the wave aberration of an electron microscope by means of the phases in the Fourier spectrum.* Journal of Microscopy 179, (2) 112-118 (1995)

[98] A. Thust, *in: Proceedings of the 12th European congress on electron microscopy.* vol. 3, p. 107-110, edt. P. Tomanek, R. Kolarik, Brno CZ (2000)

[99] R.R. Meyer , A.I. Kirkland and W.O. Saxton, *A new method for the determination of the wave aberration function for high-resolution TEM.: 1. Measurement of the symmetric aberrations.* Ultramicroscopy 92, 89-109 (2002)

[100] R.R. Meyer , A.I. Kirkland and W.O. Saxton, *A new method for the determination of the wave aberration function for high-resolution TEM.: 2. Measurement of the antisymmetric aberrations.* Ultramicroscopy 99, 115-123 (2004)

[101] K. Kimoto, K. Ishizuka, N. Tanaka, Y. Matsui, *Practical procedure for come-free alignment using caustic figures.* Ultramicroscopy 96, 219-227 (2003)

[102] T.F. Budinger, R.M. Glaeser, *Measurement of focus and spherical aberration of an electron microscope objective lens.* Ultramicroscopy 2, 31-41 (1976)

[103] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes.* Cambridge University Press, Cambridge (1986)

[104] H. Müller, Private communications (2005)

[105] M. Lentzen, A. Thust, K. Urban, *Accuracy of Aberration Measurements for HRTEM Using Zemlin Tableaus.* Proceedings of EMC, sec. IM01.P04, Antwerpen (2004)

[106] C.E. Shannon, *A Mathematical Theory of Communication.* Bell. System Technical Journal 27, 379-423, 623-656 (1948)

[107] A.J.F. Metherell, *In: Electron Microscopy in Material Science, II, (Hrsg.: U. Valdré, E. Ruedl).* CEC, Brüssel, p. 397 (1975)

[108] K.J. Hanszen, R. Lauer, G. Ade, *The effect of non-isoplantism in micrographs taken with an electron micriscope equipped with a field emission gun.* Ultramicroscopy 16, 47-58 (1985)

[109] R. Rosenfeld, *Untersuchungen zum atomaren Aufbau aperiodischer Objekte mittels Objektwellenrekonstruktion in der hochauflösenden Elektronenmikroskopie.* Dissertation, Forschungszentrum Jülich (IFF) (2001)

[110] M. Lundberg, M. Sundberg, *New complex structures in the cesium-niobium-tungsten-oxide system revealed by HREM.* Ultramicroscopy 52, 429-435 (1993)

[111] T. E. Weirich, J. Portillo, G. Cox, H. Hibst, S. Nicolopoulos, *Ab initio determination of the framework structure of the heavy-metal oxide $Cs_xNb_{2.54}W_{2.46}O_{14}$ from 100 kV precession electron diffraction data.* Ultramicroscopy 106, 164 - 175 (2006)

[112] Th. Hahn, *International Tables for Crystallography.* Volume A, Reidel Publishing Company, Dordrecht, Boston (1996)

[113] S. Hovmöller, A. Sjögren, G. Farrants, M. Sundberg, B.-O. Marinder, *Accurate Atomic Positions from Electron Microscopy.* Nature 311, 238- 241 (1984)

[114] J.J. Hu, F.H. Li, H.F. Fan, *Crystal structure determination of $K_2O{\cdot}7Nb_2O_5$ by combining high-resolution electron microscopy and electron diffraction.* Ultramicroscopy 41, 387-397 (1992)

[115] P. DeSanto, D.J. Buttrey, R.K. Grasselli, C.G. Lugmair, A.F. Volpe, B.H. Toby, T. Vogt, *Structural aspects of the M1 and M2 phases in MoVNbTeO propane ammoxidation catalysts.* Z. Kristallogr. 219, 152-165 (2004)

[116] L. Houben, A. Thust, K. Urban, *Atomic-precision determination of the reconstruction of a* 90° *tilt boundary in $YBa_2Cu_3O_{7-\delta}$ by aberration corrected HRTEM.* Ultramicroscopy 106, 200-214 (2006)

[117] C.L. Jia, K. Urban, *Atomic-resolution measurement of oxygen concentrations in oxide materials.* Science 303, 2001-2003 (2004)

[118] C.L. Jia, A. Thust, K. Urban, *Atomic-scale analysis of the oxygen configuration at a $SrTiO_3$ dislocation core.* Phys. Rev. Lett. 95, 225506 (2005)

[119] *E.W. Weisstein,* Probability and Statistics. *From* MathWorld – *A Wolfram Web Resource.* http://mathworld.wolfram.com/ProbabilityAndStatistics.html (August, 20, 2006)

# Epilogue

This thesis was performed at the Institute for Microstructure Research in the Research Center Jülich, Germany, and the Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons in the time between September 2003 and October 2006.

The work was funded by the Dr. Isolde Dietrich-Stiftung.

The content of chapter 6 are matter to the patent application no. 10 2006 038 211.0.54 "*Verfahren und Elektronenmikroskop zur Messung der Ähnlichkeit zweidimensionaler Bilder*" deposited at the Deutschen Patent- und Markenamt on 16.08.2006.

The content of section 7.4 are matter to the patent application no. 10 2006 061 978.1 "*Elektronenmikroskop und Verfahren zur Messung der Defokusstreuung*" deposited at the Deutschen Patent- und Markenamt on 21.12.2006.

# Acknowledgements

I want to thank all those people who helped and supported me with the development and workout of this thesis.
First of all, I gratefully acknowledge Prof. Dr. K. Urban who gave me the opportunity to prepare this study in his institute. I am thankful for his constant support and interest in the forthcoming of this project.
Deepest thanks to Dr. A. Thust, supervisor, mentor, and collaborator, for his unwavering guidance, encouragement and patience throughout this work, for many fascinating and insightful discussions, for knowing when I needed direction and when I needed to find my own.
I thank especially Dr. L. Houben for his competent and patient supervision in using electron microscopes, for sharing his invaluable samples, and for the many discussions, hints and ideas to my work.
Special thanks to Dr. M. Lentzen and Dr. K. Tillmann for their constant readiness for discussions on imaging, image simulations, and to Dr. M. Boese, Dr. C. Jia, and Dr. S. Mi for sharing their wisdom on materials.
Many thanks to my room mates Marc, Martha, Norbert, Sergij, and Stephan for contributing a lot of fun and sweeties to the work.
I thank W. Pieper for his competent help with technical problems concerning microscopes and computers.
I am grateful to I. Rische-Radloff and G. Wassenhoven for their organisational support.
I thank all my current and former colleagues of the Institute for Microstructure Research for the very pleasant atmosphere during work and in between.
And, last to be named but first in my heart, deepest thanks to my beloved Gabriela for holding my hand in hard times, to my brother Thomas for his corrections and hints, and to my family for all their support and encouragement.

Aachen, September 2006

# Lebenslauf

| | |
|---|---|
| Name: | Juri Barthel |
| Anschrift: | Arndtstr. 8, 52064 Aachen |
| Geburtstag, -ort: | 20. Juli 1976, Leipzig |
| Eltern: | Dr. Rainer Barthel<br>Dr. Sigrid Barthel, geb. Wieloch |
| Familienstand: | Verheiratet mit Gabriela Swoboda |
| Schulbildung: | 1983-1986 Oberschule Zittau<br>1986-1990 Oberschule Greifswald<br>1990-1992 Oberschule Zittau<br>1992-1996 Gymnasium Aachen<br>Abschluß: Abitur |
| Wehrdienst: | 1996-1997 Luftwaffe, Budel (NL) |
| Studium: | 1997-2003 Studium der Physik an der RWTH Aachen<br>2002-2003 Diplomarbeit am Fraunhofer Institut für Lasertechnik Aachen, bei Prof. Dr. R. Poprawe<br>Apr 2003 Diplomprüfung (Gesamtnote: sehr gut) |
| Promotion: | ab September 2003 am Institut für Mikrostrukturforschung, IFF, FZ-Jülich bei Prof. Dr. K. Urban<br>Mündliche Doktorprüfung am 14.03.2007, RWTH Aachen |

Jül-4274
Juni 2008
ISSN 0944-2952

