DETERMINATION OF CORE STRUCTURE PERIODICITY AND POINT DEFECT DENSITY ALONG DISLOCATIONS

by

Christoph Koch

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY
May 2002

DETERMINATION OF CORE STRUCTURE PERIODICITY AND POINT DEFECT DENSITY ALONG DISLOCATIONS

by

Christoph Koch

has been approved

April 2002

APPROVED:

_____, Chair

_____

_____

_____

_____

Supervisory Committee

ACCEPTED:

_____

Department Chair

_____

Dean, Graduate College

ABSTRACT

Understanding the structure of defects in crystalline materials is essential for modern materials science. While most defect characterization involves the inverse and non-unique problem of fitting possible structural models to experimental data, which are in many cases average spectra from a variety of defects (Electron Paramagnetic Resonance (EPR) Spectroscopy, Deep Level Transient Spectroscopy (DLTS), Photoluminescense Spectroscopy (PL), etc.), being able to directly probe the atomic structure of single defects would also provide their electronic as well as mechanical properties, since those can be determined computationally, once the structure is known. This work will report on a new electron diffraction technique to directly determine the periodicity of dislocation core structures as well as a way to greatly enhance the accuracy of the forbidden reflection lattice imaging (FRLI) technique to image individual structural point defects along partial dislocations.

Electron microdiffraction experiments with Silicon samples oriented along the [110] direction will be described, which will give direct experimental evidence for the double period reconstruction of the 90° and 30° partial dislocations. Also, Silicon and β-Silicon Carbide samples with atomically flat (111) surfaces have been prepared. Perfectly smooth surfaces are shown to be essential for imaging point defects such as kinks along partial dislocations in these materials.

In addition to these experiments, advances in the theory of electron diffraction will be reported. A new imaging technique called "Atomic String Holography" will be introduced, as well as a solution to the inversion problem in dynamical scattering theory based on a new expansion of the matrix exponential of two non-commuting matrices, one of which is diagonal.

ACKNOWLEDGMENTS

Most of all I thank God for being my creator, saviour, and best friend, and for bringing me to ASU. The 5 years in Arizona, first as an exchange student and later as a graduate student have in many ways been the most important and defining years of my life.

I am very grateful for having had the opportunity to be a student under the supervision of Prof. John Spence, who has been a very inspiring mentor to me. His overflowing wealth of new and sometimes in a sense adventurous ideas have made my time as a graduate student fun, exciting, and also challenging.

I would also like to thank the other members of my committee: Prof. Peter Rez, Prof. Kevin Schmidt, Prof. Fernando Ponce, Prof. John Venables, and Prof. Raghunath Acharya for their time and friendly advice. In particular I want to thank Prof. Peter Rez for his very committed engagement and very helpful discussions.

Much of the experimental work described in the following chapters has been possible because of Dr. Matha (Molly) McCartney who spent many hours with me on the microscope, John Weatley and Karl Weiss from the Center for Solid State Science for their very practical help in countless ways, Dr. Frances Ross at the T.J. Watson Research Center in Yorktown Heights, NY who committed a whole week to do experiments with us, and Dr. Gary Hembree for his efforts in making a continuation of the experimental work possible. In particular I want to thank Dr. Christian Kisielowski at the National Center of Electron Microscopy in Berkley, CA for kindly providing the samples and also performing some of the experiments. I am also very grateful for the very fruitful collaboration with Prof. João Francisco Justo Filho at the Instituto de Fisica da Universidade de São Paulo, Brazil, whose excellent theoretical dislocation core structure calculations made the conclusions drawn in this work possible.

As much as my studies, if not more, has my life been blessed by my parents and many dear friends who made my stay in Arizona possible. My mentors and good friends Jack Oppenhuizen, Kirk Eland, and Ben Sanders have shaped my life in many different ways during this time. I want to especially thank Jack for his caring support and help in times when I needed it most.

I praise Jesus for his miraculous operating and guiding in my life.

"Seek first the kingdom of God and his righteousness, and all these things shall be added to you." (Matt. 6:33)

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

In the 20th century, which has also been labeled the information age, an avalanche in the development of new methods and instruments for the scientific experimental verification of until then only theoretical hypotheses and ideas has broken loose, and at the same time led to a host of new theories founded on the newly won knowledge. Quantum theory has quickly proven an extremely useful tool to describe very diverse phenomena with a fairly simple "basis set" of physical laws, and the solution of many physical problems today depends not so much on the discovery of new physical laws as it depends on the ability to accurately describe the problem in the quantum-mechanical framework, making them "computable" by machines that are themselves a product of this rapidly increasing scientific knowledge. It is for example possible to calculate "real world" mechanical and electronic properties of "virtual" materials, as long as we can come up with a model of atomic positions for them. Predicting macroscopic properties of existing materials is therefore also possible, if we can find atomic models for them. This is especially difficult, when extended defects are involved, since defects by definition do not occur in a self-repeating crystal structure making it almost impossible to examine them by crystallographic methods. But even the determination of the atomic structure of perfect crystals is still a very difficult problem, especially for crystals that cannot be examined by X-ray crystallographic methods because of their size or radiation sensitivity. While X-ray diffraction patterns recorded under certain conditions can be inverted using single scattering theory, the inversion problem of electron diffraction patterns is much more complicated because of almost unavoidable multiple scattering due to the much shorter elastic mean free path of electrons.

In this work I will discuss the application of selected transmission electron microscopy techniques to find structural information of perfect and defective crystal structures, providing a direct link between experiment and currently computable atomic models. After giving a short introduction into the field of electron microscopy and its theoretical background in chapter 2 I will provide some necessary background about the theory of dislocations in chapter 3, along with a list of methods used to examine their properties, showing the importance of the work described in the following chapters.

In chapter 4 I will discuss the application of the Forbidden Reflections Lattice Imaging (FRLI) method to image partial dislocation kinks with a resolution capable of distinguishing between different energetically possible kink structures and show that rough

surfaces have limited previous work [1,2], while giving experimental proof that this problem can be overcome. In chapter 5 I will show how an electron beam focused to an only a few nm big spot can be used to determine the periodicity of the core of a partial dislocation along the beam direction. Applying a new expansion for the matrix exponential of the sum of two non-commuting matrices a focused electron beam can also be used to determine the projected potential from convergend beam electron diffraction patterns, as shown in appendix A. Finally I will present a new electron-holographic technique, which can be used to image the local environment of atoms in a crystal structure in appendix B.

The work described in the following chapters has in part already been published or submitted for publication:

- J. Spence, C. Koch "Atomic String Holography", Physical Review Letters **86** (24), 5510 (2001)
- C.Koch, J. Spence, C. Zorman, M. Mehregany and J. Chung "Modelling of HREM and nanodiffraction for dislocation kinks and core reconstruction", Journal of Physics: Condensed Matter **12**, 10175 (2000)
- J. Spence, C. Koch "Experimental evidence for dislocation core structures in silicon", Scripta Materialia, **45** (11), 1273-1278 (2001)
- J. Spence, C. Koch "On the measurement of dislocation core periods by nanodiffraction", Philosophical Magazine B **81** (11), 1701-1711 (2001)
- C. Koch and J.M. Zuo "Comparison of multislice programs for HRTEM simulations and the bloch wave method", Microsc. & Microanal. **6**, Suppl. 2: Proceedings (MSA 2000) Springer New York, p. 126-127
- C. Koch, J. Spence "Reconstruction of the projected potential from a through voltage series of dynamical electron diffraction patterns including absorption", Microsc. & Microanal. **7**, Suppl. 2: Proceedings (MSA 2001) Springer New York, p. 914-915
- J. Spence, C.Koch "Alchemi as Holography", Microsc. & Microanal. **7**, Suppl. 2: Proceedings (MSA 2001) Springer New York, p. 350-351
- L.J. Allen, C. Koch, M.P. Oxley and J.C.H. Spence "Inversion of dynamical electron scattering to obtain the crystal potential using data from two thicknesses.", Acta Cryst. A**57** (4), 473 (2001)
- C. Koch and J. Spence "A useful disentanglement of the exponential of the sum of two non-commuting matrices, one of which is diagonal", *submitted* (2002)

CHAPTER 2

HIGH RESOLUTION ELECTRON MICROSCOPY AND MICRODIFFRACTION

1. Introduction

In this chapter some of the general principles of electron microscopy will be discussed as well as the theoretical background of electron diffraction theory and the two different methods of dynamic electron diffraction simulation used in this work. Several different Transmission Electron Microscopes (TEM) have been used in this study (For detailed information specific to a particular microscope the reader is referred to its operation manual):

- *Akashi Topcon 002B:* 200kV High Resolution TEM with a maximum resolution of 1.8Å, LaB_6 electron-source.
- *Phillips CM200-FEG:* 200kV High Resolution TEM with a maximum resolution of 2.2Å, coherent illumination due to the field emission gun (FEG). This microscope is retrofitted with a scanning unit and two annular dark-field (ADF) and one bright-field (BF) detector for use as a scanning transmission electron microscope (STEM).
- *Phillips CM300-FEG-Ultra-twin:* 300kV High Resolution TEM at Lawrence-Berkeley National Labatory in Berkeley, CA. Coherent illumination due to the field emission gun (FEG).
- *LEO 912:* 120kV TEM mainly used for convergent beam and selected area diffraction work. This microscope is equipped with an Ω-shaped energy filter, which makes quantitative electron diffraction possible. This microscope was used in this work for the quantification of inelastic scattering.
- *MIDAS:* 100kV UHV VG-STEM with UHV sample preparation chamber and cold field emission gun. Because of its ultra-high vacuum well suited for surface science applications, but also microdiffraction due to its comparatively stable scanning electronics and low risk of contamination. This microscope has a host of analytical tools (secondary electron detector, Auger spectroscopy, EELS, etc.) available, as well as an extremely efficient CCD camera system and Angular Dark-field (ADF) imaging [3].

- *Hitachi:* 300kV UHV-TEM at IBM T. Watson Research Center in Yorktown Heigths, NY with a max. resolution of about 2.8Å. Its good vacuum makes the creation and examination of atomically flat Si surfaces possible.

- *JEOL4000:* 400kV HRTEM with a maximum resolution of 1.7Å.

2. Experimental Apparatus and Modes of Operation

Given the wavelength of high energy electrons of just a few picometer and the possibility of controlling their path by electric and/or magnetic fields makes this radiation ideal for microscopy at atomic and even subatomic dimensions. Today electron microscopes are with sub-Ångström resolution still the most powerful microscopes.

In this section we will introduce the basic principle of the (S)TEM along with the different modes of operation used in this work.

2.1. Transmission Electron Microscope (TEM). Figure 1 shows the basic configuration of a transmission electron microscope and ray diagrams for imaging and diffraction mode [4–6].

The electrons are being produced by either a Tungsten-, LaB_6-, or a heated/non-heated field emission gun (FEG), pass in some microscopes through a monochromator, and are then accelerated to their final energy. It is of importance that the tip of the electron source is as sharp as possible, ideally a single atom. The condenser lens system controls the beam divergence angle α, which is ideally zero for imaging mode, but non-zero for convergent beam electron diffraction (CBED) or scanning transmission electron microscopy (STEM). In practice we will need to have some beam divergence even in imaging mode, and especially when working at high magnification in order to keep the beam current reasonably high.

The image or diffraction pattern is recorded on either film, imaging plates (IP), or a CCD camera. Film is comparatively difficult to process and has a low dynamic range (about 8 bit). Imaging plates are a digital film with high dynamic range, recording on a logarithmic scale, which makes them ideal for diffraction work. They are similarly inconvenient as film, but are more quantitative and can be reused. CCD cameras record on a linear scale with a dynamic range of normally 12 to 16 bit, and transfer recorded images very conveniently directly into the computer where they can be processed and stored.

2.2. Microdiffraction (CBED and LACBED). As can be seen from the ray diagram in figure 1b only beams that pass through the electron lenses at exactly the same radius interfere with each other. Since the detector (film, IP, or CCD) can only detect the intensity, but not the phase of the incident radiation, cylindrical lens aberrations have no effect on the diffraction pattern and hence do not limit the resolution of the information contained in diffraction patterns. In X-ray crystallography, several methods have been developed using the information contained in one-dimensional rocking curves for finding phases of structure factors from which then the charge density distribution of the scattering

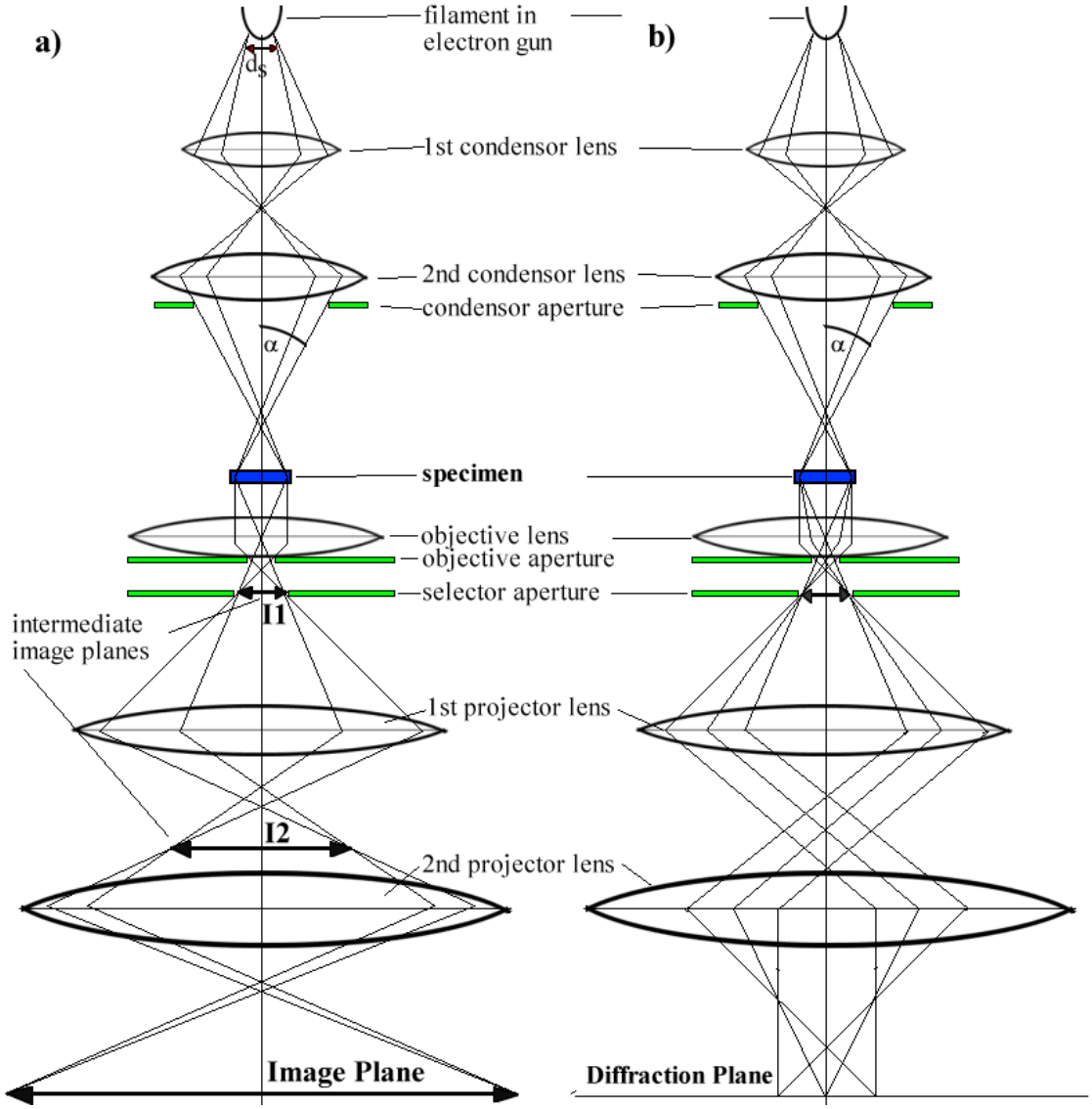

Figure 1. Ray diagram of a basic electron microscope. Depending on the type, some microscopes have three (MIDAS) or even more projector lenses. The ray diagram is shown for the microscope operated a) in imaging mode b) in diffraction mode. α is the convergence angle of the electron beam incident on the sample and d_s is the source size.

crystal can be reconstructed [7–9]. These rocking curves are being obtained by mechanically tilting the crystal.

High-energy electrons have a wavelength of just a few pm - much smaller than that of even hard X-rays, which makes the Ewald-sphere very large, exciting many reflections simultaneously, even if the Bragg condition is not exactly met. This feature allows us to record all the data of a 2-dimensional rocking curve simultaneously with a single crystal orientation. Instead of illuminating the specimen with a plane wave, the electron beam is focused in a single spot in real-space, requiring a certain angular range in reciprocal space according to the uncertainty principle. The geometry is illustrated by figure 2.

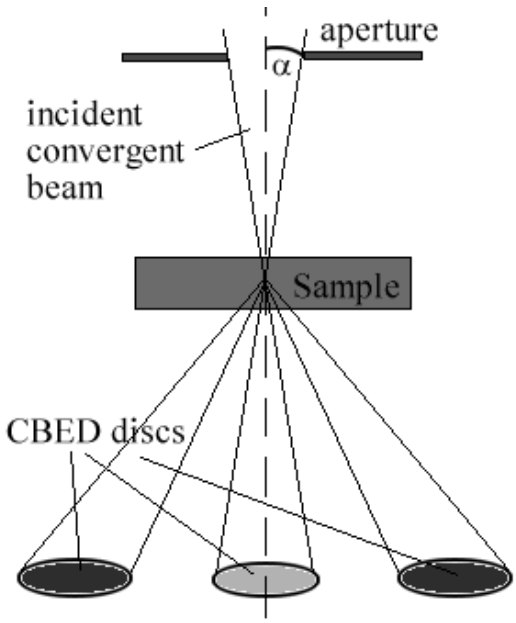

Figure 2. Illustration of the convergent beam electron diffraction (CBED) principle. The sample is illuminated by a focused beam of convergence angle α defined by the condenser aperture and its distance between to the spot the beam is focused at. The resulting diffraction pattern consists of disks, instead of spots, each representing a 2-dimensional rocking -curve of that particular reflection. As shown in figure 3 the beam does not have to be focused on the sample surface.

Because CBED patterns are insensitive to lens aberrations they can provide accurate [10] information about the thickness of a specimen of known structure, Debye Waller factors [11, 12], (low order) structure factors [13, 14], and even help to solve only partially known structures [15, 16]. In appendix A a new method for the direct inversion of CBED patterns to the projected crystal potential will be introduced, making the determination of unknown structures possible.

Because of the large Ewald sphere and the large number of excited reflections the

maximum convergence angle α for which the diffraction disks do not overlap is fairly small, especially for crystals with a large unit cell. Several methods have been proposed to extend the angular range of CBED disks, such as combining CBED patterns from different crystal orientations [17] as well as others [18–22].

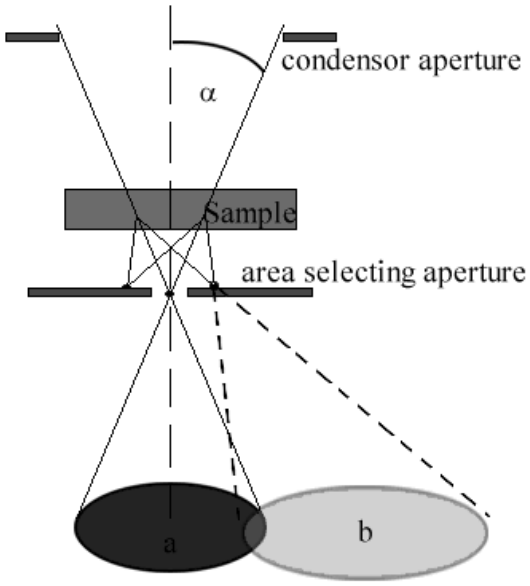

Figure 3. Illustration of the large-angle convergent beam electron diffraction (LACBED) or "Tanaka"-method principle. By shifting the area selector aperture any one disk can be recorded by itself without overlap of neighboring disks. The convergence angle can therefore be larger than 1/2 the smallest $\vec{g}$.

The geometry of the large-angle convergent beam electron diffraction (LACBED), or "Tanaka" [21,22] method is shown in figure 3. By focusing the probe below the specimen, a spot-diffraction pattern can be obtained in the plane of the area selector aperture. Using an aperture of the proper size (less than the distance between the spots, which is determined by the sample-aperture distance and the lattice constant of the specimen) a single CBED disk can be selected. The convergence angle α of this disk can be as large, as the condenser aperture allows. In fact, the condenser aperture can be removed to utilize the full angular range of the illumination system of the microscope. The LACBED method has been used to determine low-order structure factors [23], the Burgers vector of dislocations [24, 25], study the atomic displacement field around dislocations [26], the lattice mismatch in heterostructures [27], etc.. A method for the direct inversion of CBED and in particular LACBED patterns to the projected crystal potential of the specimen will be given in chapter A.

2.3. Annular Dark-Field Scanning Transmission Electron Microscope (ADF-STEM). The STEM can in principle be viewed as an inverted TEM [5, 28, 29], also because the critical electron optics responsible for image formation (in particular the objective lens) are before the sample. The image is formed by the number of electrons scattered with an angle greater than a certain θ_{min} defined by the hole in the dark-field detector as a function of the probe's position on the sample. See figure 4 for the basic geometry of the STEM.

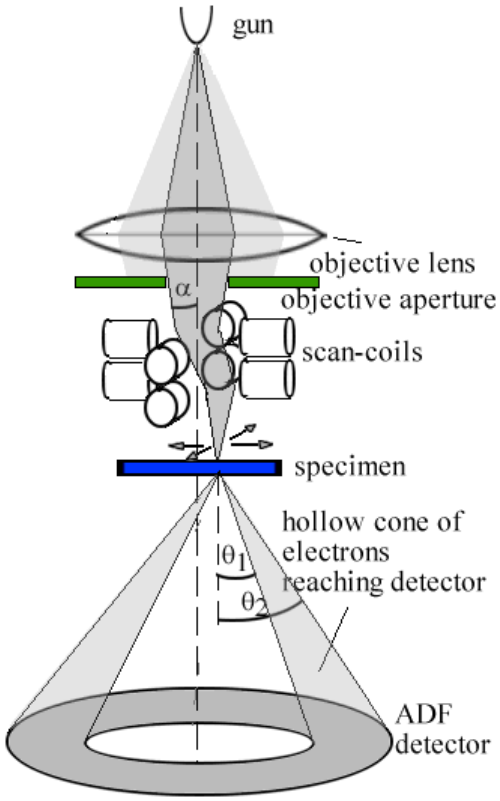

Figure 4. Diagram of a basic scanning transmission electron microscope (STEM). The electron beam is focused on the sample surface and scanned across it by a set of scan-coils. The objective aperture determines the convergence angle α. Electrons that scatter with an angle between θ_1 and θ_2 are being collected by the ADF detector and contribute to the image

To illustrate the reciprocity of TEM and STEM let us consider the case of bright-field (BF) STEM. Here the image is formed by the electrons hitting a disk-shaped detector with a radius usually of the size of the hole in the ADF detector. If we invert the path of the electrons by replacing the BF (point-like) detector with an imaginary electron source and sample the exit face wave function at the top surface of the specimen by scanning a

point detector across it we obtain in principle a conventional TEM (CTEM).

While the process of image formation in BF-STEM is analogous to that of BF-TEM (if the BF-detector size is equal to the size of the objective aperture in the TEM) and well understood, the debate whether ADF-STEM imaging is purely inelastic and therefore incoherent and linear in thickness and atomic potential squared [30] or non-linear [31], is not yet settled. In the linear imaging model according to Pennycook the image intensity at any given probe position (x, y) is given by

$$I(x,y) \quad = \quad |\Psi(x,y)|^2 \otimes V(x,y)^2, \tag{2.1}$$

which is the convolution of the incident electron probe wave function with the 2-dimensional projected potential of the object, also called object function. If this model is correct the object function can be determined from the image by deconvolution with the (known) probe function. It is assumed that the ADF-detector only collects incoherent high-angle scattering produced by the 1s-electronic states of the atoms, and is therefore independent of bonding effects, and with a small enough probe only scattering from atoms in a single column will contribute to the image.

Hillyard, Loane and Silcox [32, 33] have shown that channeling effects in the crystal are very important, and that the assumption of negligible horizontal cross-talk between atomic columns is invalid in most cases. A more realistic model taking into account channeling has to consider the variation of the probe function with thickness [34], giving the image intensity at thickness t as

$$I(x,y,t) = \int_0^t |\Psi(x,y,z)|^2 \otimes V(x,y,z)^2 dz \tag{2.2}$$

which becomes

$$I(x,y,t) = \int_0^t |\Psi(x,y,z)|^2 \otimes V(x,y)^2 dz$$

for a z-independent potential, as in perfect crystals.

The image calculations in this work avoid this problem by computing the path of both elastic and phonon-scattered electrons using a computationally very expensive, but accurate algorithm based on Kirkland's frozen phonon multislice program [35].

Many modern TEMs can also function as a STEM by running the microscope in microdiffraction (see section 2.2 of this chapter) mode, focusing the electron beam on the sample using the condenser lens system, and scanning the beam across the sample using either the beam-alignment coils (as done on the CM200 microscope used in this work), or a special set of scan-coils. The CM200 used for the experiments described in chapter 5 is equipped with two ADF- and one BF-detector which becomes exposed to the electron beam, when the CCD camera is retracted.

3. Theory of High-Energy Electron Diffraction

3.1. Notation. Reading books on diffraction theory by two different authors, and sometimes even different editions by the same author may be slightly confusing due to

the different definitions of the wave-vector and reciprocal lattice vectors. The following notations are being used throughout the literature:

$$\begin{aligned} |k| &= \frac{2\pi}{\lambda} \quad \text{and} \quad |g| = \frac{2\pi}{d} \\ |k| &= \frac{2\pi}{\lambda} \quad \text{and} \quad |g| = \frac{1}{d} \\ |k| &= \frac{1}{\lambda} \quad \text{and} \quad |g| = \frac{1}{d} \end{aligned} \tag{2.3}$$

It is therefore important to define the conventions used throughout this work. We will use equations (2.3), i.e. true reciprocal values, as is also done in [36].

Also we will define

$$\Psi_{\vec{k}} = e^{+2\pi i \vec{k}\cdot\vec{r}} \tag{2.4}$$

as the plane wave propagating in the direction defined by the wave vector $\vec{k}$. Table 1 defines additional symbols which will be used throughout this work.

3.2. The Schrödinger Equation. The collision of electrons with atoms in a crystal is in principle a quantum mechanical many-body problem with the Hamiltonian

$$H = \sum_{j=1}^{N_{electrons}} \frac{e^2}{|r - r_j|} - \sum_{j=k}^{N_{nuclei}} \frac{Z_k e^2}{|r - R_k|}$$

(r_j = position of all $N_{electrons}$ electrons in the crystal, R_k = position of all N_{nuclei} nuclei each with the charge $Z_k e$), but because of the tiny mass of the electrons as compared to any atom, and especially, since we are looking at high energy electrons and atoms in a crystal, we may neglect the momentum transfer of the electrons to the crystal, and treat the scattering process to a very good approximation as the scattering of the electron by the static crystal Coulomb potential $V(\vec{r})$, which means that the wave function of the electron $\Psi(\vec{r})$ has to be the solution of the time-independent Schrödinger equation

$$\nabla^2\Psi(\vec{r}) + \frac{2m|e|}{\hbar^2}\left[E + V(\vec{r})\right]\Psi(\vec{r}) = 0 \tag{2.5}$$

Figure 5 shows the velocity of an electron versus the voltage it has been accelerated by

$$\beta = \sqrt{1 - \left(\frac{m_0 c^2}{m_0 c^2 + |e|E}\right)^2} \tag{2.6}$$

Electrons accelerated by 50kV are traveling already at 41% of the speed of light and should be treated relativistically. We therefore need to replace the rest mass of the electron by its relativistic mass $m = \gamma_r m_0$ and adjust the energy $E \to E(1 + \gamma_r)/2$, where $\gamma_r = m/m_0 = 1 + |e|E/(m_0 c^2)$.

There are a variety of approaches to find $\Psi(\vec{r})$, given the crystal potential $V(\vec{r})$. Spence ([5] p. 148ff) gives an overview of the variety of these methods along with references to all of them. Most computer programs apply either the Bloch wave method or the

Symbol	Description
c	speed of light
v	velocity of electrons
$h, \hbar = \frac{h}{2\pi}$	Planck's constant
a_0	Bohr radius
E	acceleration voltage for electron beam
λ_0	electron wavelength in vacuum
$\lambda = (\lambda_0^{-2} + U_0)^{-1/2} \approx \lambda_0$	electron wavelength in crystal
∇_t	2-dimensional component of ∇ parallel to the surface
$\vec{k}$	electron wave-vector of magnitude λ^{-1}
k_z	component of $\vec{k}$ perpendicular to crystal surface
$\vec{k}_t$	2-dim. component of $\vec{k}$ parallel to crystal surface
$\vec{g}$	a reciprocal lattice vector
e	charge of the electron
m_0	rest mass of the electron
$\gamma_r = 1 + \lvert e\rvert v/(m_0 c^2)$	relativistic correction factor
$m = \gamma_r m_0$	relativistic mass of the electron
Ψ	a complex-valued wave function.
$V^e(\vec{r})$	real-valued real-space Coulomb potential.
$V^i(\vec{r})$	imaginary absorptive potential.
$V(\vec{r}) = V^e(\vec{r}) + iV^i(\vec{r})$	complex crystal potential.
$V_{\vec{g}} = V^e_{\vec{g}} + iV^i_{\vec{g}}$	Fourier coefficient of $V(\vec{r})$ at rec. lattice vector $\vec{g}$
$U_{\vec{g}} = \frac{2m_0\lvert e\rvert}{h^2} V_{\vec{g}}$	(non-relativistic) electron structure factor

Table 1. Definition of the most common symbols used in this work.

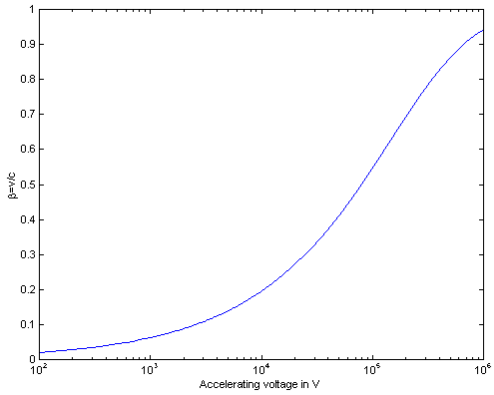

Figure 5. Relative velocity $\beta = v/c$ of electrons vs. accelerating voltage

multislice algorithm (and its variations of real-space, reciprocal space, or a hybrid method of both). The choice of algorithm depends on the problem to be calculated. In this work both of the above mentioned methods have been used, and will be discussed in further detail in this section. Because appendix A introduces an invertable expansion based on the Bloch wave formalism, this method will receive a more detailed discussion than the multislice method.

3.3. Bloch Wave Method. The Bloch wave method was first developed in 1928 by Bethe [37] and is based on direct solution of (2.5) using Bloch's theorem

$$\Phi(\vec{r}) = C(\vec{r})\exp(2\pi i\vec{k}\cdot\vec{r}) \tag{2.7}$$

and since $C(\vec{r})$ has the periodicity of the potential, we can expand it in its Fourier-components at the reciprocal lattice vectors of the potential

$$C(\vec{r}) = \sum_{\vec{g}} C_g \exp(2\pi i\vec{g}\cdot\vec{r}) \tag{2.8}$$

Expanding the crystal potential we get

$$V(\vec{r}) = \sum_{\vec{g}} V_g \exp(2\pi i\vec{g}\cdot\vec{r}) \tag{2.9}$$

We can expand the real space electron wave-function within the crystal as a sum of Bloch-waves [38]

$$\Psi(\vec{r}) = \sum_j \Psi_j \Phi^{(j)}(\vec{r}) = \sum_j \Psi_j \sum_{\vec{g}} C_{\vec{g}}^{(j)} \exp(2\pi i(\vec{k}^{(j)} + \vec{g})\cdot\vec{r}) \tag{2.10}$$

With

$$\nabla^2\Psi = -4\pi^2 \sum_j \Psi_j \sum_{\vec{g}} C_{\vec{g}}^{(j)} (\vec{k}^{(j)} + \vec{g})^2 \exp(2\pi i(\vec{k}^{(j)} + \vec{g})\cdot\vec{r}) \tag{2.11}$$

The boundary condition that the momentum parallel to the surface of the crystal has to be preserved fixes the component $\vec{k}_t$ of every $\vec{k}^{(j)}$ parallel to the crystal surface and makes it equal to that of the incident beam, as long as all the $\vec{g}$-vectors are parallel to the surface, i.e. in the zero-order Laue zone.

We can now place expressions (2.10) and (2.11) for Ψ and $\nabla^2\Psi$ back into the Schroedinger equation (2.5) using the Fourier expansion for the potential (2.9):

$$\begin{aligned} \nabla^2\Psi(\vec{r}) + \frac{2m|e|}{\hbar^2}\left[E + V(\vec{r})\right]\Psi(\vec{r}) &= 0 \\ -4\pi^2 \sum_j \Psi_j \sum_{\vec{g}} C_{\vec{g}}^{(j)} (\vec{k}^{(j)} + \vec{g})^2 \exp(2\pi i[\vec{k}^{(j)} + \vec{g}] \cdot \vec{r}) + & \\ \frac{2m|e|}{\hbar^2} \sum_j \Psi_j \sum_{\vec{g}} \left[E C_{\vec{g}}^{(j)} + \sum_{\vec{g}'} V_{g-\vec{g}'} C_{\vec{g}'}^{(j)} \right] \exp(2\pi i[\vec{k}^{(j)} + \vec{g}] \cdot \vec{r}) &= 0 \end{aligned} \tag{2.12}$$

where we use the fact that

$$V_{g-\vec{g}'} \exp(2\pi i[\vec{g} - \vec{g}'] \cdot \vec{r}) C_{\vec{g}'}^{(j)} \exp(2\pi i[\vec{k}^{(j)} + \vec{g}'] \cdot \vec{r}) = V_{g-\vec{g}'} C_{\vec{g}'}^{(j)} \exp(2\pi i[\vec{k}^{(j)} + \vec{g}] \cdot \vec{r})$$

In practice we only need to sum over a finite set of $\vec{g}$-vectors (beams). Since this equation must hold for any value of $\vec{r}$ we must require that it also holds for every $\vec{k}^{(j)}$ and $\vec{g}$ individually. Simplifying expression (2.12) and including only N beams we get therefore for $j = 0 \ldots N-1$ and $\vec{g} = \vec{g}_0 \ldots \vec{g}_{N-1}$

$$\left[\frac{2m|e|}{h^2}(E + V_0) - (\vec{k}^{(j)} + \vec{g})^2 \right] C_{\vec{g}}^{(j)} + \frac{2m|e|}{h^2} \sum_{\vec{g}'} V_{\vec{g}-\vec{g}'} C_{\vec{g}'}^{(j)} = 0 \tag{2.13}$$

Also at high enough energies the incident beam energy (E) and momentum ($|\vec{k}|$) is much greater than V_0 or $|\vec{g}|$ respectively so that

$$\begin{aligned} \sqrt{\frac{2m|e|}{h^2}[E + V_0]} &= K \approx k_z \\ \Rightarrow K^2 - |\vec{k} + \vec{g}|^2 &= K^2 - k_z^2 - |\vec{k}_t + \vec{g}|^2 \\ &= (K + k_z)(K - k_z) - |\vec{k}_t + \vec{g}|^2 \\ &\approx 2K(K - k_z) - |\vec{k}_t + \vec{g}|^2 \end{aligned}$$

where k_z is the component of the incident electron beam wave-vector perpendicular to the crystal surface. Therefore (2.13) becomes, when divided by the relativistic correction factor γ_r

$$\left[\frac{2K}{\gamma_r}\left(K - k_z^{(j)}\right) - \frac{|\vec{k}_t + \vec{g}|^2}{\gamma_r} \right] C_{\vec{g}}^{(j)} + \sum_{\vec{g}'} U_{\vec{g}-\vec{g}'} C_{\vec{g}'}^{(j)} = 0 \tag{2.14}$$

which is an eigenvalue problem and can be written in matrix form

$$\frac{\gamma_r}{2K}\begin{pmatrix} \frac{|\vec{k}_t+\vec{g_0}|^2}{\gamma_r}+U_0 & U_{\vec{g}_0-\vec{g}_1} & \cdots & U_{\vec{g}_0-\vec{g}_{N-1}} \\ U_{\vec{g}_1-\vec{g}_0} & \frac{|\vec{k}_t+\vec{g_1}|^2}{\gamma_r}+U_0 & \cdots & U_{\vec{g}_0-\vec{g}_{N-1}} \\ \vdots & \vdots & & \vdots \\ U_{\vec{g}_{N-1}-\vec{g}_0} & U_{\vec{g}_{N-1}-\vec{g}_1} & \cdots & \frac{|\vec{k}_t+\vec{g_{N-1}}|^2}{\gamma_r}+U_0 \end{pmatrix}\begin{pmatrix} C_0^{(j)} \\ C_1^{(j)} \\ \vdots \\ C_{N-1}^{(j)} \end{pmatrix}$$
$$= \left(k_z^{(j)}-K\right)\begin{pmatrix} C_0^{(j)} \\ C_1^{(j)} \\ \vdots \\ C_{N-1}^{(j)} \end{pmatrix}$$

or

$$\frac{\gamma_r}{2K}A(\vec{k}_t,\gamma_r)C^{(j)} = \left(k_z^{(j)}-K\right)C^{(j)} = \gamma^{(j)}C^{(j)}$$

where A is called the structure factor matrix.

Remembering from equation (2.10) that

$$\begin{aligned}\Psi(\vec{r}_t,t) &= \sum_j \Psi_j \sum_{\vec{g}} C_{\vec{g}}^{(j)} \exp(2\pi i(k_z^{(j)}t + [\vec{k}_t+\vec{g}]\cdot\vec{r}_t)) \\ &= \sum_{\vec{g}}\left[\sum_j \Psi_j C_{\vec{g}}^{(j)}\exp(2\pi i k_z^{(j)}t)\right]\exp(2\pi i[\vec{k}_t+\vec{g}]\cdot\vec{r}_t)\end{aligned}$$

which is the 2-dimensional Fourier transform of

$$\tilde{\Psi}(\vec{k}_t+\vec{g},t) = \sum_j \Psi_j C_{\vec{g}}^{(j)}\exp(2\pi i k_z^{(j)}t) \tag{2.15}$$

the whole vector of reciprocal space wave-function amplitudes $\vec{\tilde{\Psi}}(\vec{k}_t,t)$ consisting of the amplitudes for all the reciprocal lattice vectors $\vec{g}$ can be written in matrix form

$$\vec{\tilde{\Psi}}(\vec{k}_t,t) = C\left[\exp(2\pi i k_z^{(j)}t)\right]_D\vec{\Psi} \tag{2.16}$$

where

$$C = \begin{pmatrix} C_{\vec{g}_0}^{(0)} & C_{\vec{g}_0}^{(1)} & \cdots & C_{\vec{g}_0}^{(N-1)} \\ C_{\vec{g}_1}^{(0)} & C_{\vec{g}_1}^{(1)} & \cdots & C_{\vec{g}_1}^{(N-1)} \\ \vdots & \vdots & & \vdots \\ C_{\vec{g}_{N-1}}^{(0)} & C_{\vec{g}_{N-1}}^{(1)} & \cdots & C_{\vec{g}_{N-1}}^{(N-1)} \end{pmatrix}$$

$$\left[\exp(2\pi i k_z^{(j)}t)\right]_D = \begin{pmatrix} \exp(2\pi i k_z^{(0)}t) & 0 & \cdots & 0 \\ 0 & \exp(2\pi i k_z^{(1)}t) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \exp(2\pi i k_z^{(N-1)}t) \end{pmatrix}$$

$$\vec{\Psi} = \begin{pmatrix} \Psi_0 \\ \Psi_1 \\ \vdots \\ \Psi_{N-1} \end{pmatrix}$$

At $t = 0$ the diagonal matrix $[\exp(2\pi i k_z^{(j)} t)]_D$ becomes the identity matrix, and since no electrons have scattered yet $\vec{\tilde{\Psi}}(\vec{k}_t, 0) = \delta_{\vec{g},0}$, which means that we have the following boundary condition

$$C\vec{\Psi} = \delta_{\vec{g},0} \tag{2.17}$$

This equation is being satisfied, if $\vec{\Psi}$ is the m^{th} column of C^{-1}, where m is the index of the reciprocal lattice vector for which $\vec{g}_m = 0$.

In a more general way we can define the incident electron wave-function $\vec{\tilde{\Psi}}(\vec{k}_t, 0)$ as a column vector of several incident plane wave amplitudes. For a single incident plane wave with $\vec{g} = 0$ this vector will consist of mainly zeros, and only a single non-zero value in its m^{th} row. The exit face wave function is then given by

$$\begin{aligned} \vec{\tilde{\Psi}}(\vec{k}_t, t) &= C\left[\exp(2\pi i k_z^{(j)} t)\right]_D C^{-1} \vec{\tilde{\Psi}}(\vec{k}_t, 0) \\ &= \exp(2\pi i K t) C\left[\exp(2\pi i \gamma^{(j)} t)\right]_D C^{-1} \vec{\tilde{\Psi}}(\vec{k}_t, 0) \\ &= \exp(2\pi i K t) \exp\left(\frac{\pi i \gamma_r t}{K} A(\vec{k_t})\right) \vec{\tilde{\Psi}}(\vec{k}_t, 0) \\ &= \exp(2\pi i K t) S \vec{\tilde{\Psi}}(\vec{k}_t, 0) \end{aligned} \tag{2.18}$$

The intensity of spots in the diffraction pattern is given by

$$\begin{aligned} I_{\vec{g}_n}(\vec{k}_t, t) &= \left|\vec{\tilde{\Psi}}(\vec{k}_t, t)_n\right|^2 \\ &= \left|\sum_{j=0}^{N-1} S_{n,j} \vec{\tilde{\Psi}}(\vec{k}_t, 0)_j\right|^2 \end{aligned} \tag{2.19}$$

For an incident plane wave at $\vec{g}_m = 0$ (2.19) becomes

$$I_{\vec{g}_n}(\vec{k}_t, t) = |S_{n,m}|^2 = \left|\exp\left(\frac{\pi i \gamma_r t}{K} A(\vec{k_t}, \gamma_r)\right)\right|^2 \tag{2.20}$$

S is called the scattering matrix. The structure factor matrix A depends on the Fourier coefficients of the projected potential of the specimen, the transverse component of the incident electron wave vector $\vec{k}_t$ and the relativistic correction factor γ_r. In order to calculate the intensity of diffracted beams using the Bloch-wave method, one has to first construct the structure factor matrix A, diagonalize it to obtain its eigenvalues $\gamma^{(j)}$ and eigenvector matrix C, multiply the eigenvalues $\gamma^{(j)}$ by $\pi i \gamma_r t/K$, and take the modulus squared of the m^{th} column of the resulting S matrix.

3.4. Multislice Algorithm. Based on the physical optics theory of Cowley and Moodie [39] the multislice method is a numerical integration method to solve the Schrödinger equation (2.5) by assuming that the wave function $\Psi(\vec{r})$ can be represented by a modulated plane wave

$$\Psi(\vec{r}) = \exp(2\pi i\vec{k}\cdot\vec{r})\Phi(\vec{r}) \tag{2.21}$$

and validity of the high energy approximation ($k_z \gg |\vec{g}| \Rightarrow \delta^2\Phi(\vec{r})/\delta z^2 \ll 4\pi i k_z \delta\Phi(\vec{r})/\delta z$) so that we end up with a modified version of the Schrödinger equation

$$\frac{\delta\Phi(\vec{r})}{\delta z} = \left(\frac{i\nabla_t^2}{4\pi k_z} - \frac{\vec{k}_t}{k_z}\cdot\nabla_t\right)\Phi(\vec{r}) + \frac{2m|e|i}{4\pi k_z\hbar^2}V(\vec{r})\Phi(\vec{r}) \tag{2.22}$$

If the electron beam is normal to the crystal surface ($\vec{k}_t = 0$) then (2.22) reduces to

$$\frac{\delta\Phi(\vec{r})}{\delta z} = \frac{i\nabla_t^2}{4\pi k_z}\Phi(\vec{r}) + \frac{2m|e|i}{4\pi k_z\hbar^2}V(\vec{r})\Phi(\vec{r}) \tag{2.23}$$

This first order differential equation in t can now be integrated numerically by "cutting" the sample in many thin slices parallel to its surface, reducing the potential in each slice to a 2-dimensional one through integration over the slice thickness which separates the scattering process into a kinematic scattering process and a Fresnel propagation through "vacuum" over the thickness of the slice, relating the wave function $\Phi(t+\Delta t)$ to $\Phi(t)$. Doing this for every slice will yield the wave function at the exit surface of the crystal, which becomes exact for $\Delta t \to 0$. Figure 6 illustrates this procedure.

The propagation through a single slice is done in two steps (hybrid real- and reciprocal space method):

1. *Transmission:* multiplication of incident wave function with phase grating (in real-space) or convolution in reciprocal space.

$$\Phi(\vec{r}) \to \exp\left(\frac{2m|e|i}{4\pi k_z\hbar^2}\int_t^{t+\Delta t} V(\vec{r})\delta z\right)\Phi(\vec{r}) \tag{2.24}$$

2. *Fresnel propagation in reciprocal space:* Fourier transformation $\tilde{\Phi}(\vec{q}) = \mathrm{FT}(\Phi(\vec{r}))$, multiplication with the Fourier transform of the Fresnel propagator

$$p(\vec{q},\Delta t) = \mathrm{FT}[\exp(i\nabla_t^2\Delta t/4\pi k_z)] = \exp(-i\pi\Delta t|\vec{q}|^2/k_z) \tag{2.25}$$

$$\tilde{\Phi}(\vec{q}) \to \exp(-i\pi\lambda\Delta t|\vec{q}|^2)\tilde{\Phi}(\vec{q}) \tag{2.26}$$

and inverse Fourier transformation $\Phi(\vec{r}) = \mathrm{FT}^{-1}(\tilde{\Phi}(\vec{q}))$

The slice thickness is thin enough if the total electron current ($\int|\Phi(\vec{r})|^2\delta^2\vec{r}$) does not decrease significantly as the electrons propagate through the crystal.

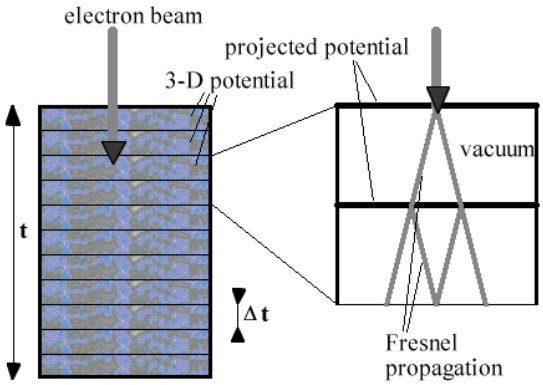

Figure 6. Diagram illustrating the multislice method. The 3-dimensional atomic potential within each slice is integrated in z-direction to make a 2-dimensional phase-grating. The real-space wave function of the incident electron beam is first multiplied by the phase grating and then propagated through a vacuum slab of the thickness of the slice, since the analytical solution of the Schrödinger equation for $V(\vec{r}) = 0$ is just the classical expression for Fresnel propagation.

The detailed derivation using this approach can be found in [35, 39–43]. Self et al. [38] use Huygen's principle as the basis for their derivation. The multislice solution can also be derived from equation (2.18) by use of the Zassenhaus theorem [40]

$$\exp(A\epsilon + B\epsilon) \quad = \quad \exp(A\epsilon)\exp(B\epsilon) + \frac{1}{2}[B, A]\epsilon^2 + O(\epsilon^3)$$

which becomes exact for small ϵ.

One advantage of the multislice method is that the crystal potential does not need to be periodic along the z-axis. This makes it possible to make the slices thinner than one unit cell. In fact for the multislice simulations in chapter 5 a slice thickness of 0.48Å has been used, which is less than the range of the atomic Coulomb potential, i.e. the "size" of an atom.

3.5. Image Formation. In the TEM operated in image mode the electrons diffracted by the crystal are being focused by the successive lenses such as the objective lens and projector lenses (see figure 1) to form an image. Aberration effects of electron-optical lenses can be described as the convolution of the real-space exit face wave function with the microscope transfer function, or as the product of the wave function in reciprocal space with the Fourier transform ($\chi(\vec{q})$) of the microscope transfer function, where $\vec{q}$ is the 2-dimensional coordinate in reciprocal space, conjugate to $\vec{r}$. This same formalism also applies to the formation of the probe wave function in scanning transmission electron microscopy (STEM), where the resolution of the scanned image depends mainly on the size

of the electron probe. The image formed by an objective lens having axial aberrations up to third order and neglecting temporal and spatial partial coherence effects is then given by [44]

$$I(\vec{r}) = |\Psi_I(\vec{r})|^2 = \left|\mathrm{FT}\left(\tilde{\Psi}(\vec{q})\exp(i\chi(q))\right)\right|^2 \tag{2.27}$$

$$\begin{aligned}\chi(\vec{q}) = \frac{2\pi}{\lambda}\Bigg[&\frac{\lambda^2(q_x^2+q_y^2)}{2}C_1 + \frac{\lambda^2(q_x^2-q_y^2)}{2}C_{1,2a} + \\ &\lambda^2(q_xq_y)C_{1,2b} + \frac{\lambda^3q_x(q_x^2+q_y^2)}{3}C_{2,1a} + \\ &\frac{\lambda^3q_y(q_x^2+q_y^2)}{3}C_{2,1b} + \frac{\lambda^3q_x(q_x^2-3q_y^2)}{3}C_{2,3a} + \\ &\frac{\lambda^3q_y(3q_x^2-q_y^2)}{3}C_{2,3b} + \frac{\lambda^4(q_x^2+q_y^2)^2}{4}C_3 + \\ &\frac{\lambda^4(q_x^4-q_y^4)}{4}C_{3,2a} + \frac{\lambda^4q_xq_y(q_x^2+q_y^2)}{2}C_{3,2b} + \\ &\frac{\lambda^4(q_x^4-6q_x^2q_y^2+q_y^4)}{4}C_{3,4a} + \lambda^4(q_x^3q_y-q_xq_y^3)C_{3,4b}\Bigg]\end{aligned} \tag{2.28}$$

where for a certain aberration coefficient $C_{n,m}$ m denotes the order of aberration, $2\pi/n$ denotes the smallest rotation angle that results in the same phase shift, and the suffixes a and b mark mutually orthogonal contributions to the same non-cylindrically symmetric aberration [44]. Table 2 gives the commonly used names for aberrations and how they translate into the $C_{n,m}$-coefficients scheme.

$C_{n,m}$ coefficient	descriptive name
C_1	defocus
$C_{1,2}$	regular astigmatism
$C_{2,1}$	axial coma
$C_{2,3}$	3-fold astigmatism
C_3	spherical aberration C_s
$C_{3,2}$	2-fold astigmatism of C_s
$C_{3,4}$	4-fold astigmatism

Table 2. Aberration coefficients and their descriptive names.

In addition to the aberrations of the objective lens the energy spread of electrons leaving the gun ΔE, and instabilities in the in the lens currents ΔI cause a change in defocus $f = C_1$, which is proportional to the chromatic aberration coefficient C_c [5].

$$\Delta f = C_c\left(\frac{\Delta E}{E} - \frac{2\Delta I}{I}\right) \tag{2.29}$$

The effect of chromatic aberration on the image can be calculated by averaging over a set of images produced by different values of incident beam energy E spread with a Gaussian distribution of width ΔE around the average value E and the resulting change in focus f. In the TEM case this effect, as well as effects due to spatial partial coherence (instead of perfectly parallel incident plane waves the illumination spans an angular range of θ_c) can be included analytically by introducing an exponential damping factor to the microscope transfer function $CTF(k)$ in reciprocal space [5].

$$CTF(k) = \exp(i\chi(k))exp\left(\frac{1}{2}\pi^2\Delta f^2\lambda^2k^2 - \pi^2u_0^2q\right) \tag{2.30}$$

where

$$q = (Cs\lambda^3k^3 + \Delta f\lambda k)^2 + \pi^2\lambda^4\Delta f^4k^6 - 2\pi i\lambda^3\Delta f^2k^3 \tag{2.31}$$

$$u_0 = \sqrt{2}\frac{\theta_c}{\lambda} \tag{2.32}$$

In STEM we need to average over a set of diffraction patterns produced by a set of incident electron probes with the same uncertainty in f.

Astigmatism and Coma are relatively easy to correct by combinations of magnetic quadrupoles. Within the last decade also spherical aberration correctors for SEM [45], TEM [46, 47] and STEM [44] have been implemented successfully, reducing C_s to zero and even negative values (there is still the problem of measuring C_s, if $C_s \to 0$).

The resolution of a TEM image is determined by the width of the point-spread function (PSF), or impulse-response function

$$PSF(\vec{r}) = \mathrm{FT}\left(\sin(\chi(\vec{q}))P(\vec{q})\right) \tag{2.33}$$

where $P(\vec{q})$ is the pupil- or aperture function, which is 1 for wave vectors that pass through the aperture, and zero otherwise.

$$P(\vec{q}) = \begin{cases} 1 & \text{if } |q|\lambda \leq \sin(\alpha) \\ 0 & \text{if } |q|\lambda > \sin(\alpha) \end{cases} \tag{2.34}$$

Equations (2.2) and (2.1) both have in common that the STEM image contrast is related to the probe intensity distribution, which is given by

$$I^{probe}(\vec{r}) = \left|\mathrm{FT}\left(\exp(i\chi(\vec{q}))P(\vec{q})\right)\right|^2 \tag{2.35}$$

Figure 7 shows a 2-dimensional plot of the impulse response function and the probe intensity for $E = 200$kV, $C_s = 0.7$mm, $\alpha = 12.5$mrad, $df = -513$Å. Comparing the two functions it becomes obvious that the STEM probe is narrower, because it is the square of the wave function amplitude. This is one of the reasons why it is generally believed that ADF-STEM images have an inherently higher resolution than BF-TEM images with the same objective aperture size.

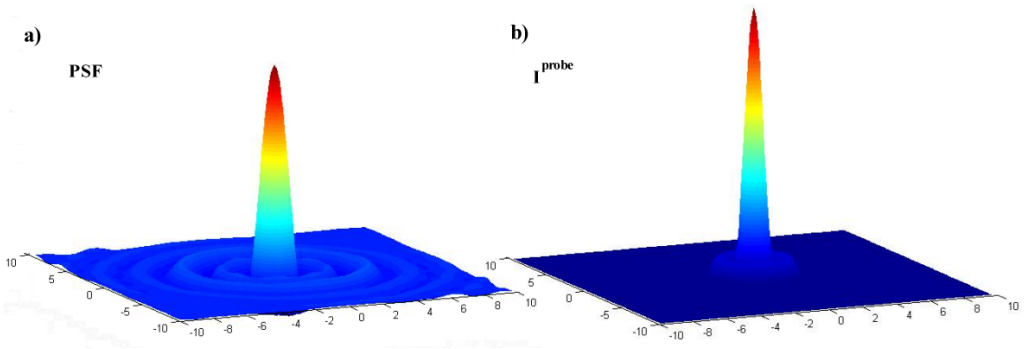

Figure 7. (color) 2-dimensional impulse response function ($PSF(\vec{r})$) and STEM probe function $I^{probe}(\vec{r})$ for the same microscope parameters (E = 200kV, C_s = 0.7mm, α = 12.5mrad, $df = -513$Å). The scale in x- and y-direction is in Å.

The optical transfer function ($OTF(\vec{q})$) for incoherent imaging is the Fourier transform of the STEM-probe intensity $I^{probe}(\vec{r})$. It characterizes the strength with which certain spacial frequencies of the object function are being represented in the image, just like the contrast transfer function ($CTF(\vec{q}) = \sin(\chi(\vec{q}))$) does for the coherent phase contrast image. Aside from the spherical aberration of the objective lens, parameters like the beam divergence angle α, and defocus f determine the shape of the STEM probe intensity distribution. Most commonly used values for the defocus and objective aperture size are those suggested by Scherzer ($f = -\sqrt{1.5 C_s \lambda}$, $\alpha = 1.5(\lambda/C_s)^{1/4}$), or Mory [48] which provide the most "compact" STEM probe:

$$f = -\frac{3}{4}\sqrt{C_s \lambda} \tag{2.36}$$

$$\alpha = 1.27(\lambda/C_s)^{1/4} \tag{2.37}$$

Figure 8 show a 1-dimensional cross-section through the center of the 2-dimensional plots shown in figure 7 as well as the radial incoherent and coherent transfer functions ($OTF(q)$ and $CTF(q)$ respectively). Because of nonlinearities due to the zero-crossings in the CTF an objective aperture has to be inserted, ideally at the first zero of the CTF, providing interpretability of the image, but limiting the resolution. Recently [49–51] this limitation in TEM imaging has been overcome by computationally correcting for the effects of the CTF using images recorded at several different foci, and resolutions of 0.8Åat 300kV have been achieved [52].

4. Inelastic Scattering

Although the elastic scattering model described above provides a very good description of Bragg scattering, quantitative electron diffraction requires the treatment of inelastic

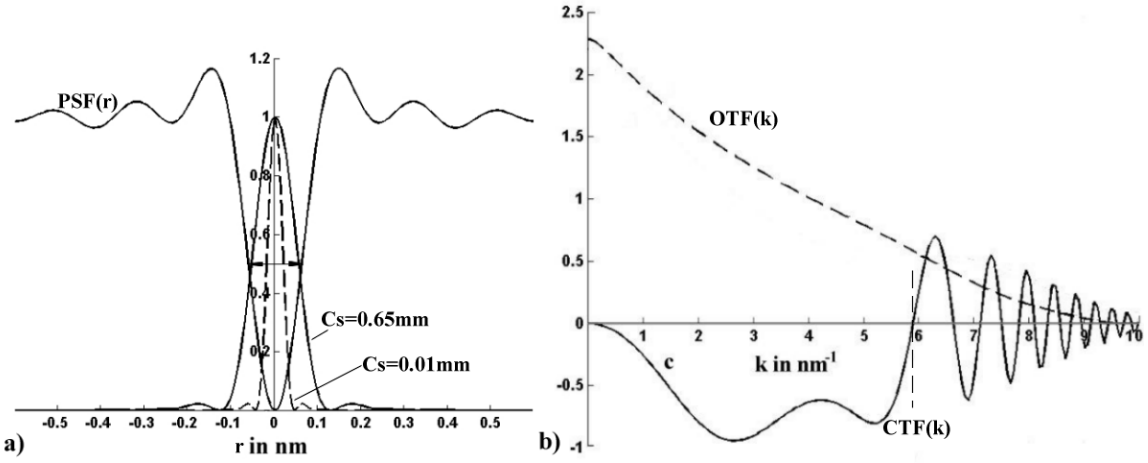

Figure 8. a) 1-dimensional cross-section of impulse response function $PSF(r)$ and STEM probe intensity distribution, b) $OTF(k)$ and $CTF(k)$. The following microscope parameters have been used for the plots: $E = 300$kV, $C_s = 0.65$mm. CTF(k) and PSF(r) were calculated using the Scherzer focus of $df = -513$Å. Damping due to temporal incoherence has been included in the CTF(k) and PSF(r) plots with $\Delta f = 20$Å (see expr. (2.29)) while the spatial partial coherence due to a beam divergence angle of $\theta_c = 0.25$mrad was assumed. The dashed vertical line indicates the size of the objective aperture which would be used in this case. The STEM probe $I^{probe}(r)$ and OTF(k) have been plotted for a beam convergence angle of $\alpha = 9.4$mrad and defocus of $f = -268$Å, giving the most "compact" probe [48]. For comparison the probe function of an aberration corrected STEM with $C_s = 0.01$mm, $\alpha = 26.8$mrad, and $f = -33$Å (dashed curve) has been included.

scattering processes, which include the transfer of energy from the incident electrons to the crystal and therefore alter its state. Such inelastic scattering processes can be described by the scattering due to quasi-particles such as phonons (lattice vibrational modes), plasmons (collective electron waves), and single electron excitations producing electron-hole pairs.

Inelastic scattering effects were initially accounted for phenomenologically by the introduction of a complex periodic potential $V^i(\vec{r})$ [53] and later justified theoretically by Yoshioka [54], who provided a quantum mechanical treatment of the problem using the tight binding approximation. Because absorption effects are non-local and their calculation requires the computationally expensive evaluation of 3-dimensional integrals the absorptive part $V^i(\vec{r})$ of the crystal potential $V(\vec{r})$ was initially approximated as 10% of its real part $V^r(\vec{r})$, which gives in Fourier space

$$V_{\vec{g}} = V^e_{\vec{g}} + iV^i_{\vec{g}} \approx (1 + 0.1i)V^e_{\vec{g}} \tag{2.38}$$

based on experimental observations for simple crystal structures.

Howie [55] has shown that, to a good approximation inelastic scattering by plasmons contributes only to V^i_0. Single electron excitations also contribute to V^i_0 and low order absorptive structure factors [56], but only of the order of a few percent [57]. For low thickness the main contribution of high angle inelastic scattering is due to phonons. For an incident electron with wave-vector k_e and mass m_e scattering with a phonon of wave-vector k_{ph} and frequency $\omega_{ph}(k_{ph})$ the equations for conservation of energy and momentum are

$$\begin{aligned} \frac{h^2k_e^2}{2m_e} &= \frac{h^2k_e'^2}{2m_e} + \omega_{ph}(k_{ph}) \\ \vec{k_e} &= \vec{k_e'} + \vec{k}_{ph} \end{aligned}$$

The optical phonon energy in Si is 0.063eV which means that mainly acoustic phonon modes will be excited, since kT =0.025eV at room temperature. The energy loss experienced by high energy electrons scattering with phonons is usually very small (less than 20meV [58]) while the scattering angle $\Delta\vec{k} = \vec{k}' - \vec{k}$ can be quite large (see [59] for the angular distribution of single phonon scattering events).

4.1. Measuring the Contribution of Inelastically scattered electrons. The post-specimen Omega energy filter on the LEO 912 electron microscope provides the opportunity to select electrons within a well-defined energy-window and measure their contribution to the diffraction pattern or image. However, since the energy loss of the incident radiation due to phonon scattering is usually comparable to the energy spread of the electrons leaving the LaB_6 electron gun, single phonon scattering cannot be separated from the purely elastic signal using this instrument. Figure 10 shows the distribution of scattered intensity as a function of scattering angle in the elastic+phonon scattered channel as well the unfiltered signal, measured from diffraction patterns as those shown in figure 9.

The graphs in figure 10 show that for thin enough specimen the highly inelastic signal is comparable to, or even less than the sum of the elastic signal and quasi-elastic signal (with

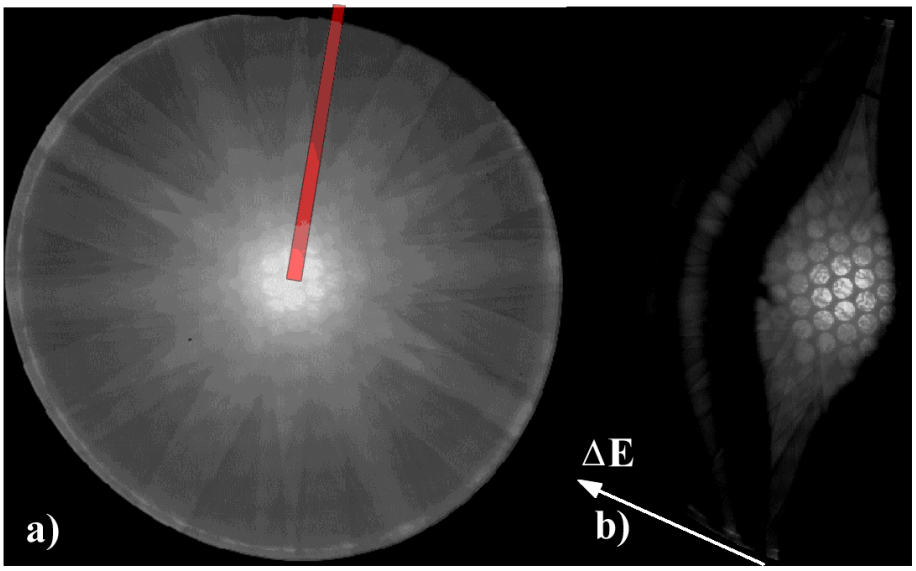

Figure 9. a) unfiltered and b) zero-loss energy filtered CBED patterns of the same specimen area. The shaded area shows the portion of the pattern used for producing the graphs shown in figure 10. The exact same area has been used for integration in both patterns in order to make a comparison between them possible. The shape of the filtered pattern is due to the Omega filter, which can only filter in one dimension and transmits patterns produced by electrons of different energies in the second dimension as indicated by the arrow. The recorded intensity is displayed logarithmically in the range 4000-18000 counts as recorded by Fuji image plates.

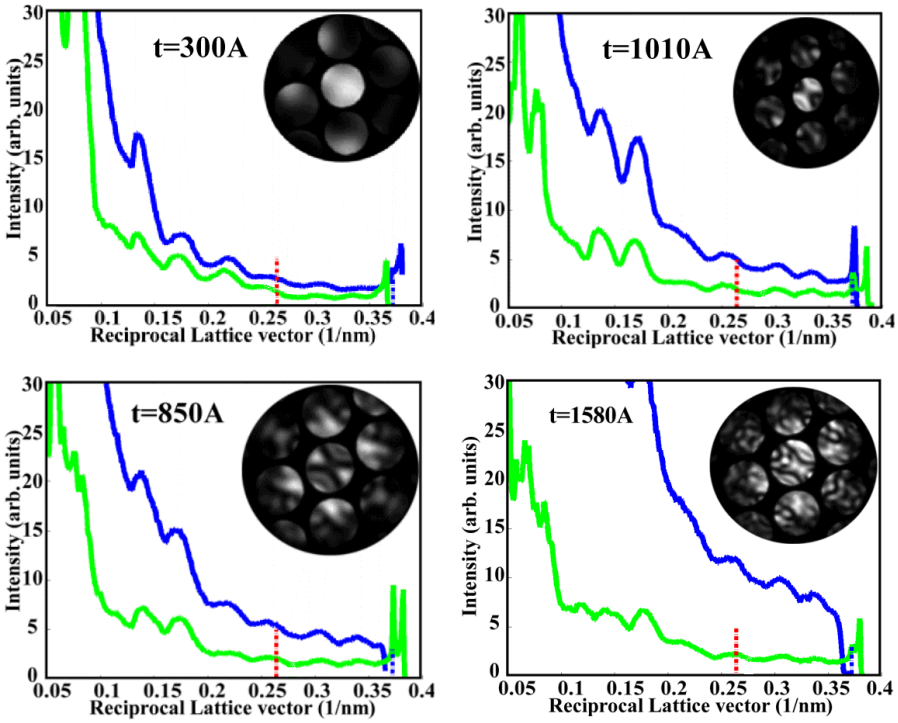

Figure 10. Averaged scattering intensity taken from diffraction patterns as those in figure 9 as function of scattering angle with and without the Omega energy filter. The CBED patterns serve for thickness estimation. The intensity peaks at high angles are due to the HOLZ ring and/or scattering off the bore of one of the projector lenses, which changes between exposures due to beam drift. The dash-dotted lines indicate the position of the regular HOLZ-ring as well as that of the ring expected for regions of double period reconstructed dislocation cores, as described in chapter 5. Recorded on LEO912 at a temperature of -164°C .

energy losses less than the width of the energy filter, mainly thermal diffuse scattering) at high angles. Assuming that scattering due to plasmons and single electron excitations is in the forward direction, we can recover the elastic and TDS signal from unfiltered data by simply dividing by a constant for thin (<400Å) specimen. The experiments described in chapter 5 will make use of this fact.

4.2. Thermal diffuse scattering (TDS). In 1965 Hall and Hirsch [60] developed a theory to calculate the contribution of thermal diffuse scattering to the absorptive part $V^i(\vec{r})$ of the scattering potential $V(\vec{r})$. Since its numerical evaluation, involving a 3-dimensional convolution of rapidly varying functions, is computationally very expensive, and in the past computers were still too slow for it to be used very much, several authors have calculated and tabulated absorptive scattering factors for most atomic species and given isotropic Debye-Waller factors [57, 61, 62], provided approximate analytical expressions based on the Doyle and Turner [63] expansion of the atomic scattering factors [64], or wrote computer programs allowing any isotropic Debye-Waller factor to be used [65]. Alternative models for describing phonon scattering have been proposed by Rez [66], Wang and Cowley [67], Wang [68], Dinges and Rose [69], Anstis [70], Amali and Rez [59], and others.

While absorptive scattering factors (equation 2.38) can account for scattering of electrons out of the elastic channel, i.e. the missing electron intensity in directions of elastic Bragg scattering, calculating the angular distribution of the inelastically scattered electrons requires a more complicated model. Phonon scattering is of particular interest, since it is responsible for most of the high-angle scattering forming the image in high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM).

The scattering amplitude for the wave vector $\vec{q}$ due to collective lattice displacements (phonons) can be described simply as the difference between scattering off a crystal in which every atom at position $\vec{r}_j$ is displaced by a small vector $\vec{u}_j$ and the crystal with every atom at its equilibrium position $\vec{r}_j$

$$f^{TDS}(\vec{q}) = \sum_j f_j^{el}(q) \left[\exp(2\pi i \vec{q} \cdot [\vec{r}_j + \vec{u}_j]) - \exp(2\pi i \vec{q} \cdot \vec{r}_j)\right] \tag{2.39}$$

Squaring this expression in order to obtain the number of electrons scattered with the wave-vector $\vec{q}$ and averaging over time and the number of atoms in the crystal gives us for a single atom per unit cell [59]

$$\begin{aligned} \left[f^{TDS}(q)\right]^2 &= \left[f^{el}(q)\right]^2 \left[1 - \exp\left(-8\pi^2\overline{u^2}\left(\frac{q}{2}\right)^2\right)\right] \\ &= \left[f^{el}(q)\right]^2 \left[1 - \exp\left(-M\left(\frac{q}{2}\right)^2\right)\right] \end{aligned} \tag{2.40}$$

where $\overline{u^2}$ is the mean square displacement of the atoms, and the term $M = 8\pi^2\overline{u^2}$ is called the Debye-Waller factor. The general expression for any number of atoms per unit cell is

given by Hall and Hirsch [60]. In the harmonic approximation $\overline{u^2}$ is given by [71, 72]

$$\overline{u^2}(T) = \frac{\hbar}{2m} \int_0^{\omega_{max}} \coth\left(\frac{\hbar\omega}{2k_B T}\right) \frac{g(\omega)}{\omega} d\omega \tag{2.41}$$

where m is the atomic mass, $g(\omega)$ the phonon density of states and ω_{max} the maximum phonon frequency. Gao and Peng [73] have parameterized expression (2.41) for 68 elemental crystals using experimental measurements of phonon density of states obtained from neutron scattering data, where available. Figure 11 shows a plot of $\overline{u^2}$ vs. the temperature.

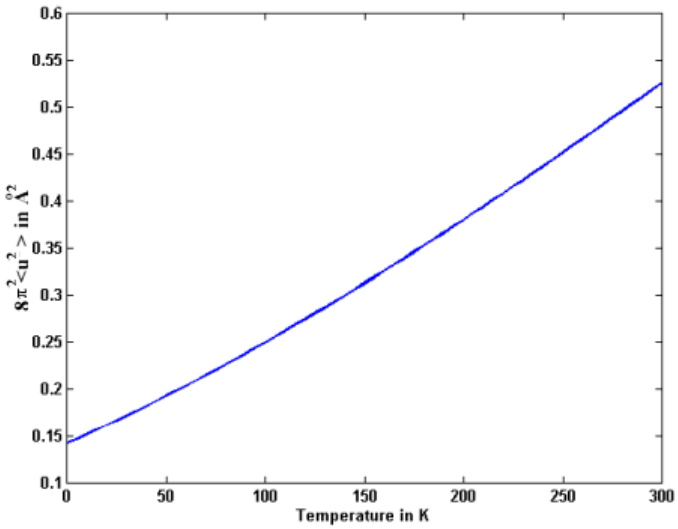

Figure 11. Mean square atomic displacement of atoms in a Si crystal as a function of temperature. The product of the atomic displacement $\overline{u^2}$ and the scaling factor $8\pi^2$ is called the Debye-Waller factor.

A more transparent, but rather "brute force" method, proposed by Loane Xu and Silcox [74], the frozen phonon approximation, which starts with expression (2.39) and integrates it numerically is based on the assumption that a single high-energy electron passing through the specimen at about half the speed of light can only probe a single "snap-shot" of the vibrating crystal, instead of a time averaged crystal potential as assumed for the derivation of equation (2.40). The image or diffraction pattern is then produced by averaging over many such "snap-shots", in each of which the atoms are slightly displaced from their equilibrium positions. These displacements can be calculated very accurately using molecular dynamics [75], or experimental phonon dispersion curves [76] and corresponding atomic displacements according to the dynamical theory of crystal lattices developed by Born and Huang [77]. Figure 12 shows simulated CBED patterns for a Si (110) specimen of 960Å thickness after averaging over an increasing number of different frozen phonon configurations (Einstein model) at room temperature.

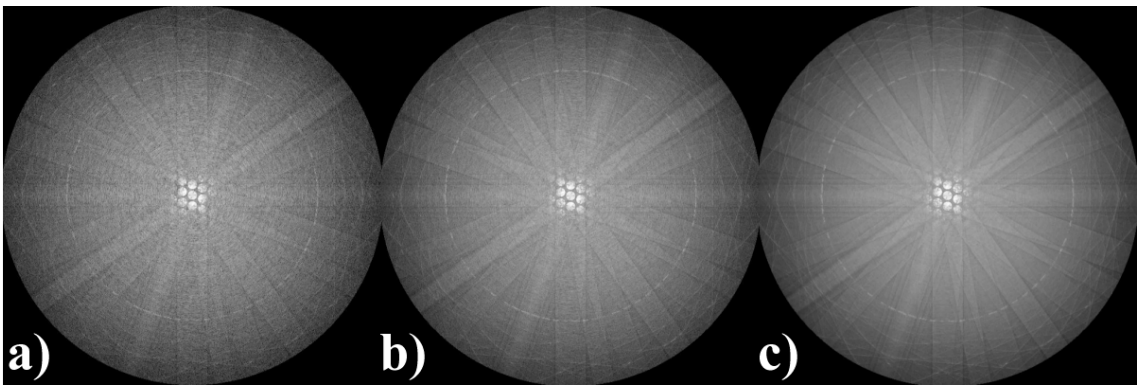

Figure 12. Multislice CBED simulations of Si (110) including thermal diffuse scattering (TDS) according to the Einstein model. The sample thickness is 960Åthickness, beam energy 100kV, and the beam divergence angle α is 6 mrad. The pattern is plotted after a) 1, b) 4, and c) 20 iterations and becomes less and less noisy and finally converges.

For HRTEM and STEM imaging it has been shown [75,76] that the difference of the phonon contribution calculated according to the exact phonon distribution model or a simple Einstein model, in which atoms are displaced, as if they were independent, uncorrelated harmonic oscillators, to the image is insignificant. However Muller et al [76] were able to show that the fine structure in the Kikuchi bands of CBED patterns can be reproduced very accurately using the parameter-free detailed phonon model. These features in the diffration pattern cannot be calculated using the Einstein model, which involves a free parameter, the average vibrational amplitude per atom species present in the crystal. The vibrational amplitude used for the simulations based on the Einstein model can be fitted to the scattered intensity distribution at high angles (as plotted in figure 10), which is very sensitive to TDS. Comparing the ratio of the integrated intensity in the FOLZ-ring and the interpolated background signal for the same scattering angle with that of simulated patterns a RMS vibrational amplitude of 0.064Å has been determined for Si at 95K. Further details of this experiment are given in chapter 5.

The fact that high angle electron scattering is dominated by TDS, i.e. a inelastic and therefore incoherent scattering process, and the scattering from a single atom potential is proportional to the square of the atomic number Z, as given by the Mott formula

$$\begin{aligned} \left(f^{el}(q/2)\right)^2 &= \left(\frac{2}{a_0}\frac{[Z - f^x(q/2)]}{(q/2)^2}\right)^2 \\ &\propto Z^2 \quad \text{for large q} \end{aligned} \tag{2.42}$$

has given the HAADF-STEM technique also the name Z-contrast imaging. Choosing the inner angle α_i of the ADF detector is a trade-off between efficiency (not enough scattering to very high angles) and interpretability (non-linear imaging due to interference effects of

coherently elastically scattered electrons at low angles). Assuming equal probability for Bragg-scattered and non-scattered electrons to be scattered to high angles by phonons [59] and a value for α_i that lets us collect mainly the TDS signal, we can assign a certain probability p^{TDS} with which electrons traveling at angles less than α_i scatter with phonons to some angle greater than α_i. The probability that these electrons scatter back to angles less than α_i is assumed to be negligible, and no matter whether these electrons are again Bragg- or TDS scattered, they will contribute to the ADF signal. This means that the current I^{BF} of electrons not potentially contributing to the ADF signal is reduced at a constant rate as the electrons pass through the crystal.

$$\begin{aligned} \frac{\delta I^{BF}}{\delta t} &= -p^{TDS} I^{BF}(t) && (2.43) \\ \Rightarrow I^{BF}(t) &= I^{BF}(0)\exp(-p^{TDS}t) \\ I^{ADF}(t) &= I^{BF}(0) - I^{BF}(t) = I^{BF}(0)\left[1-\exp(-p^{TDS}t)\right] && (2.44) \end{aligned}$$

This thickness dependence of the intensity collected by the ADF detector has been tested for Si(110) using multislice simulations and the frozen phonon approximation, as shown in figure 13 for different inner cutoff angles α_i and two different temperatures. Figure 14 shows the derivative dI/dt of the intensity shown in figure 13 in order to make oscillations in the signal more visible. The simulations were done for an aberration corrected 100kV STEM with $C_s = 0.05mm$, $\alpha = 22$mrad, and $df = -122$Å. As already established experimentally by Howie [78] and later computationally by Hillyard and Silcox [33] 40mrad seems to be a good choice for Si, avoiding most of the coherent signal, but still very efficient in collecting the TDS signal. Figure 14 also shows that assuming an exponential decay of I^{BF} with thickness (equation 2.43) can only be true in a first order approximation. The following form of I^{BF} gives a more accurate description

$$I^{BF}(t) = I^{BF}(0)\exp(-p^{TDS}t)\left[1+\alpha\cos(2\pi t/\xi)\right], \qquad (2.45)$$

where α and ξ are free parameters that depend on the inner angle of the detector and the material.

However, this result is in contrast to the findings of Nellist and Pennycook [79] who assume that TDS requires prior coherent Bragg scattering, and the "source current", out of which electrons will scatter into high angles, oscillates with thickness t as

$$I^{coh}(t) \propto e^{-\sigma t}\left[1-\cos(\xi t)\right] \qquad (2.46)$$

Again, assuming a probability of TDS scattering p^{TDS} we can integrate the current of phonon-scattered electrons in thickness and obtain

$$\begin{aligned} I^{TDS}(t) &= \int_0^t p^{TDS} I^{coh}(t')dt' \\ &\propto p^{TDS}\left\{(1-e^{-\sigma t}) - \left(1+\frac{\sigma^2}{\xi^2}\right)^{-1}\left[\frac{\sigma}{\xi}e^{-\sigma t}\sin(\xi t) + \frac{\sigma^2}{\xi^2}(1-e^{-\sigma t}\cos(\xi t))\right]\right\} \end{aligned}$$

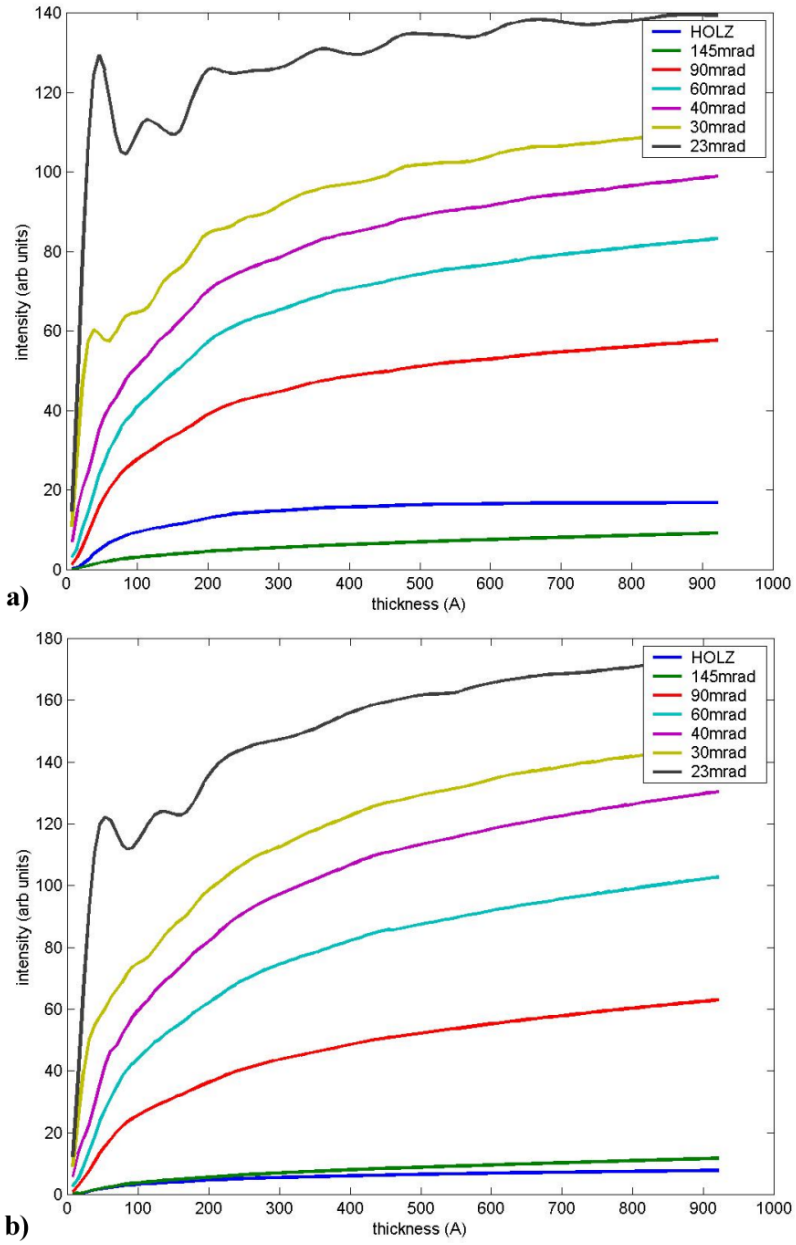

Figure 13. (color) Multislice simulation of the intensity collected by different ADF detectors as a function of thickness. The ADF detectors collect scattering up to an outer angle of 200mrad, and inner angles between 23 and 145 mrad. One of the detectors (HOLZ) only collects scattering into a narrow annulus around the first order Laue-ring. The mean square atomic displacement is a) 0.044 Å and b) 0.076 Å (room temperature)

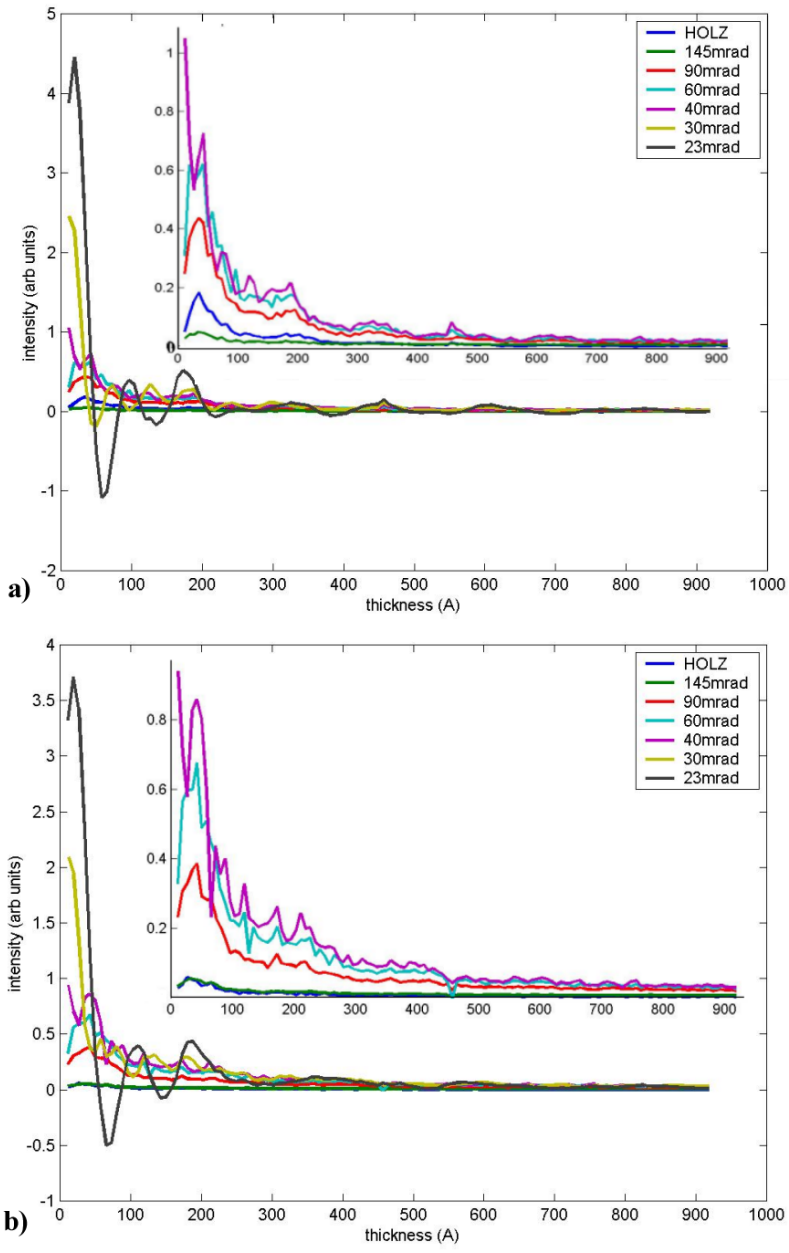

Figure 14. (color) Multislice simulation of the differential intensity dI/dt collected by different ADF detectors as a function of thickness. The mean square atomic displacement is a) 0.044 Å and b) 0.076 Å (room temperature)

This results implies thickness-oscillations in the TDS signal, which cannot be confirmed by the data shown in figure 13. The oscillations seen in figure 13 are most likely due to contribution of elastically scattered electrons reaching the ADF detector, because of their dependence on the size of the hole in the detector. Nellist and Pennycook [79] do not give any estimate for the magnitude of parameters in expression (2.46) ξ and σ, but propose to fit their expression to simulations, such as those shown in figure 13.

Conventional electron scattering calculations using Debye-Waller factors, which assume that each electron "sees" an average potential of the vibrating crystal, are not able to reproduce the correct distribution of wave-vectors of electrons that have scattered with phonons, in particular, these calculations can never produce features like Kikuchi bands. This leads to the question, how accurate quantitative CBED simulations based on this theory, and the low order structure factor measurements derived from such, are. Since the distribution of phonon inelastically scattered electrons can be simulated very accurately using phonon dispersion data obtained by neutron scattering [76], or by ab-initio molecular dynamics simulations, the phonon background can be calculated as the difference between frozen phonon multislice simulations and the standard multislice method using Debye-Waller factors and an optical potential. The high-angle peak of phonon-scattering suggests a parameterization of the strength of the Kikuchi bands with atomic scattering factors, especially high-order ones. That will allow Bloch wave calculations to accurately incorporate the phonon background. This method will also allow the exact treatment of HOLZ effects. The accurate calculation of HOLZ lines requires the precise knowledge of high angle scattering factors, which can be calculated with neutral atom Hartree-Fock methods, and the Debye-Waller factor, which is often times not known very accurately. Calculating the phonon background with the frozen phonon multislice method and the exact phonon modes (obtained from phonon dispersion curves, or ab-initio simulations) can give us very accurate results for high angle thermal diffuse scattering (TDS) and therefore also HOLZ effects. Comparing frozen phonon simulations with conventional multislice simulations using the same atomic scattering factors allows us to identify high angle scattering solely due to TDS, and a Debye-Waller like parameterization of temperature effects for the atomic scattering factors can be found, which would be based on true dynamical scattering theory.

5. The Computational Tool: *S-TEM*

Due to dynamic scattering effects which are almost impossible to avoid in electron diffraction the imaging process is non-linear, and TEM images can only be interpreted accurately by matching them with simulated images, especially if dynamic effects like forbidden reflections (chapter 4) are used to form the image. The ADF-STEM technique was initially developed to allow easy and straight-forward image interpretation according to a linear imaging model [80–82], in which the (known) probe intensity distribution is being convoluted with the object function, which can then be retrieved by deconvolution. However, as already shown in the previous section, this is an over-simplified model of the imaging process. This means that accurate image simulations require fully dynamic calculations,

including thermal diffuse scattering, especially, if the annular dark-field detector is tuned to one of the HOLZ rings as in chapter 5.

There exist a number of free as well as commercially available software based on the multislice algorithm for conventional TEM image and diffraction pattern simulations (for a comparison of a few of them see [83]). Kirkland [35] published the source code for a program that can be used for STEM calculations, but this program is rather slow and unable to calculate effects such as those essential for the work described in chapter 5. Starting with the source code of the TEM image simulation program *autoslic* published by Kirkland [35] I developed the program *S-TEM* , which is designed to perform conventional TEM as well as STEM image calculations and CBED simulations. In this section I will first explain some of the numerical details important for STEM image simulations, which have been incorporated in the program *S-TEM* , that I wrote and performed most of the image simulations shown in subsequent chapters with, and later give an overview of the user-friendly features and versatility of this code.

5.1. Details of STEM image simulation. Very much attention has been paid to produce STEM image simulations as accurate (but yet fast) as possible, by trying to avoid as many approximations as possible. The following list gives an overview of some of the details of this implementation:

- *Frozen phonon approximation for TDS simulation.*

  The frozen phonon approximation has been shown to reproduce the full diffraction pattern, including its thermal diffuse background, and even the phonon-mode related fine-structure within the Kikuchi bands, very accurately [76]. Since ADF-STEM image contrast is mainly, if not purely produced by TDS, it is important to incorporate phonon scattering as accurately as possible. However, for calculations, which do not require this degree of accuracy inclusion of temperature effects using Debye-Waller factors is also available.

- *Electron propagation is always normal to potential slices.*

  Most codes use the following modified version of the propagator function (2.25) to account for specimen tilt [84].

  $$P(\vec{q}, \Delta t, \theta) = \exp(-i\pi\Delta t|\vec{q}|^2/k_z + 2\pi i\Delta t(q_x \tan\theta_x + q_y \tan\theta_y))$$

  which is basically a translation of the slices against each other, and is therefore only valid for very small angles. This approximation is being avoided by always cutting the potential slices perfectly perpendicular to the direction of the incident beam, which is only possible because of the approximation avoided by the next item.

- *Slices can be extremely thin.*

  The multislice algorithm becomes exact only in the limit of very thin slices. However, most codes (e.g. [35]) project the full range of the Coulomb potential of an atom

into a single slice, which means that slices cannot be made thinner than about 1-2Å, and they should always coincide with lattice planes, if these slices shall truly represent the crystal potential. Sacrificing execution speed, *S-TEM* calculates the true 3-dimensional potential distribution and numerically integrates the potential from the bottom to the top surface of every slice, which can now have any desired orientation. This allows us to reduce the slice thickness in the multislice calculation without limits until convergence is reached. It is also essential for simulating TDS diffraction effects caused by the z-components of lattice vibrational modes, or periodicities in atomic displacements along the z-direction, as is the topic of chapter 5.

- *Correct treatment of Chromatic aberration.*

  Chromatic aberration C_c incoherently affects the defocus and cannot be included in the incident probe wave function for STEM simulations, instead the result from incident electrons with slightly different energies needs to be integrated over in the detector plane. Figure 15 shows how the incident probe intensity distribution varies with only slightly different values for the electron energy (dE is the deviation from the accelerating voltage, which is 100keV in this case).

- *Very fast and general Fourier transform algorithm.*

  FFTW, which stands for "Fastest Fourier Transform of the West" [85] is a public domain fft library which can handle arrays of any size, i.e. it is not bound to powers of 2 only, but has special routines for array lengths being products of powers of 2,3,5,7,11, and 13, and a general purpose routine for any other number. This makes the choice of an optimum array size for the wave function propagation possible that is big enough to avoid effects due to a finite super-cell, and at the same time small enough to keep the calculation time short.

- *Efficient sequence of multislice steps.*

  In order to calculate the intensity integrated over the area of the ADF detector for a specific beam position, i.e. the value of a single pixel in the STEM image, one has to propagate the electron wave function through an area of the sample about 30 Å$\times$ 30 Å. Calculating the projected potential is very time consuming, and doing it for every pixel anew would be a waste of time. Instead *S-TEM* calculates the projected potential for the whole area of the sample, and propagates the electron wave-function only through a small portion of it. In fact a frozen phonon configuration is generated for the whole sample at once, and the intensity reaching the ADF detector is computed with the same configuration for every pixel, before a new random frozen phonon configuration is generated. This is a valid approach, as long as the final intensity is determined by enough of these iterations.

- *Efficient implementation of electron source size effects*

For a fixed beam position the intensity collected at the STEM detector is the sum of scattered electrons stemming from different locations around the electron emitting area of the electron source. The diameter of this area is defined as the source size and is in the Å-range, depending on the microscope and condition and alignment of the electron gun. For a finite sized objective aperture (or condensor aperture in a TEM/STEM) this integration over different source positions is equivalent to an integration over different beam positions over the same area. Since we need to calculate the scattered intensity for every beam position anyway, source size effects are incorporated by convoluting the final image with a round top-hat like function of the same diameter as the source.

- *Scalable to any size problem and computer.*

  The program is designed to utilize as much memory as the user wants to allow for it. If *S-TEM* is run on computers with less memory, it will be less efficient and therefore slower, but not less accurate.

5.2. *S-TEM* **output file.** Along with image files in Tagged Image File Format (TIFF) *S-TEM* also produces an hypertext meta-language (HTML) output file like the one in figure 16. It shows simulated high-resolution annular dark-field (ADF) and bright-field (BF) STEM images of the double period reconstructed 90° partial dislocation core in Si along the [110] direction at 0.5 × 0.5Åpixel size. With the proper command in the input file these results can automatically be posted on a website, after every iteration of the frozen phonon approximation process, which enables the user to monitor its progress from anywhere in the world, and also the results to be shared with people around the globe, as has been necessary in this work.

S-TEM accepts a large number of parameters that can be adjusted and are being passed to the program via an input file in ASCII-format. Table 3 explains those parameters shown in figure 16.

An unlimited number of ADF and BF detectors can be configured. The inner and outer radius of the detector is displayed to the left of each image. The 2 black wedges in the diffraction pattern on the right (displayed on logarithmic scale, because Gamma=0) indicate the size and position of the first detector, which, of course covers the full circle and not just two 45° arcs. Positioning the cursor over any of the pixels in the STEM images will cause that particular pixel to change to the color red, and the diffraction pattern corresponding to that beam position will be displayed on the right. In figure 16 the cursor has been placed close to the core of the 90° partial dislocation which produces a bright half-order Laue ring in the diffraction pattern (see chapter 5 for further details).

The energy deviation histogram (bottom right in figure 16) can be seen again in figure 15, which shows that portion of the HTML output file produced by *S-TEM* for three different cursor positions. This histogram is also "active", in that moving the cursor across it will cause the blue square to follow its movement and the probe intensity plot to be

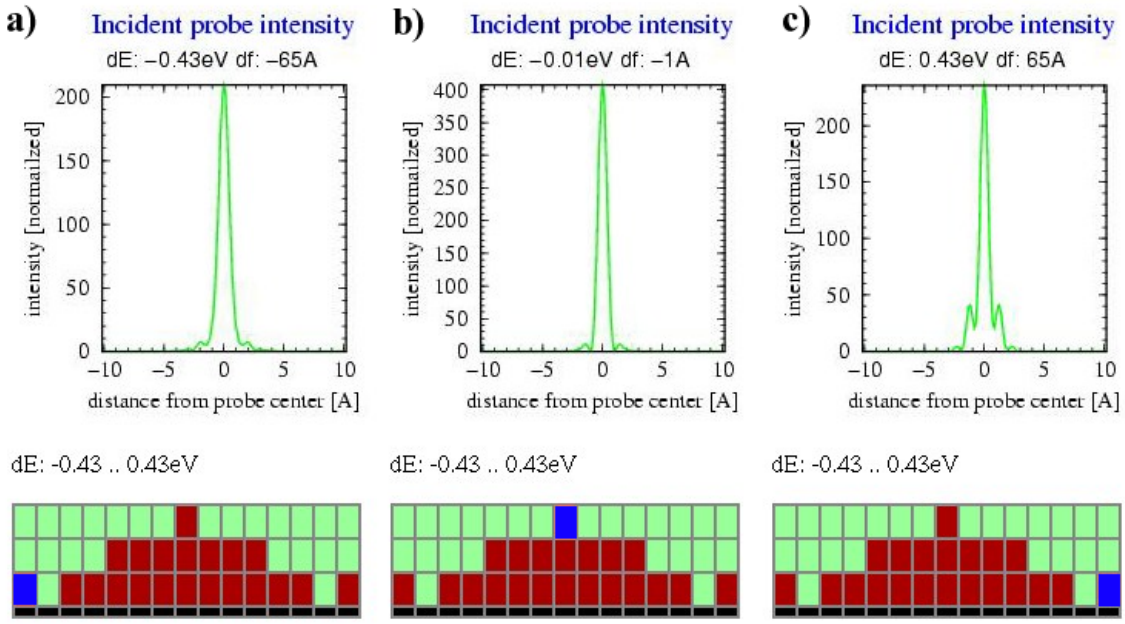

Figure 15. (color) Electron beam probe intensity distribution (top) as function of energy deviation. The spread in energy produces a spread in defocus, caused by the Chromatic Aberration C_c according to equation (2.29). The energy deviation follows a Gaussian distribution with full width at half max defined in the input file. Each square in the bottom plots represents a value of dE and corresponding defocus, which will be averaged over in the simulation. The blue square indicates the value of dE, for which the incident probe intensity is plotted on the top

updated for the different values of dE used in the simulation. Small energy deviations (central column of histogram) will be included first in the calculation, and the largest ones (far left and right columns) last.

parameter	description
thickness	thickness of the simulated sample in Å
v_0	electron probe energy in keV
runs averaged over	number of different frozen phonon configurations used
Gamma	Image scaling coefficient for display of diffraction patterns $I_{displ} = \log(I)$, if $\gamma = 0$, and $I_{displ} = I^{1/\gamma}$ otherwise.
C_S	objective lens spherical aberration in mm
C_5	second order spherical aberration in mm
C_c	chromatic aberration in mm
$defocus$	main defocus in Å(for $dE = 0eV$).
aperture	condenser aperture in mrad
T	sample temperature in K
TDS	TDS included in simulation (yes/no) if TDS="no", a Debye-Waller factor is used.
scan region	area of the sample over which the probe will be scanned in Å
resolution	true size of pixels shown in the images in Å
wave function size	array used for calculating intensity for every image-pixel

Table 3. Simulation parameters used by the program *S-TEM* and their description.

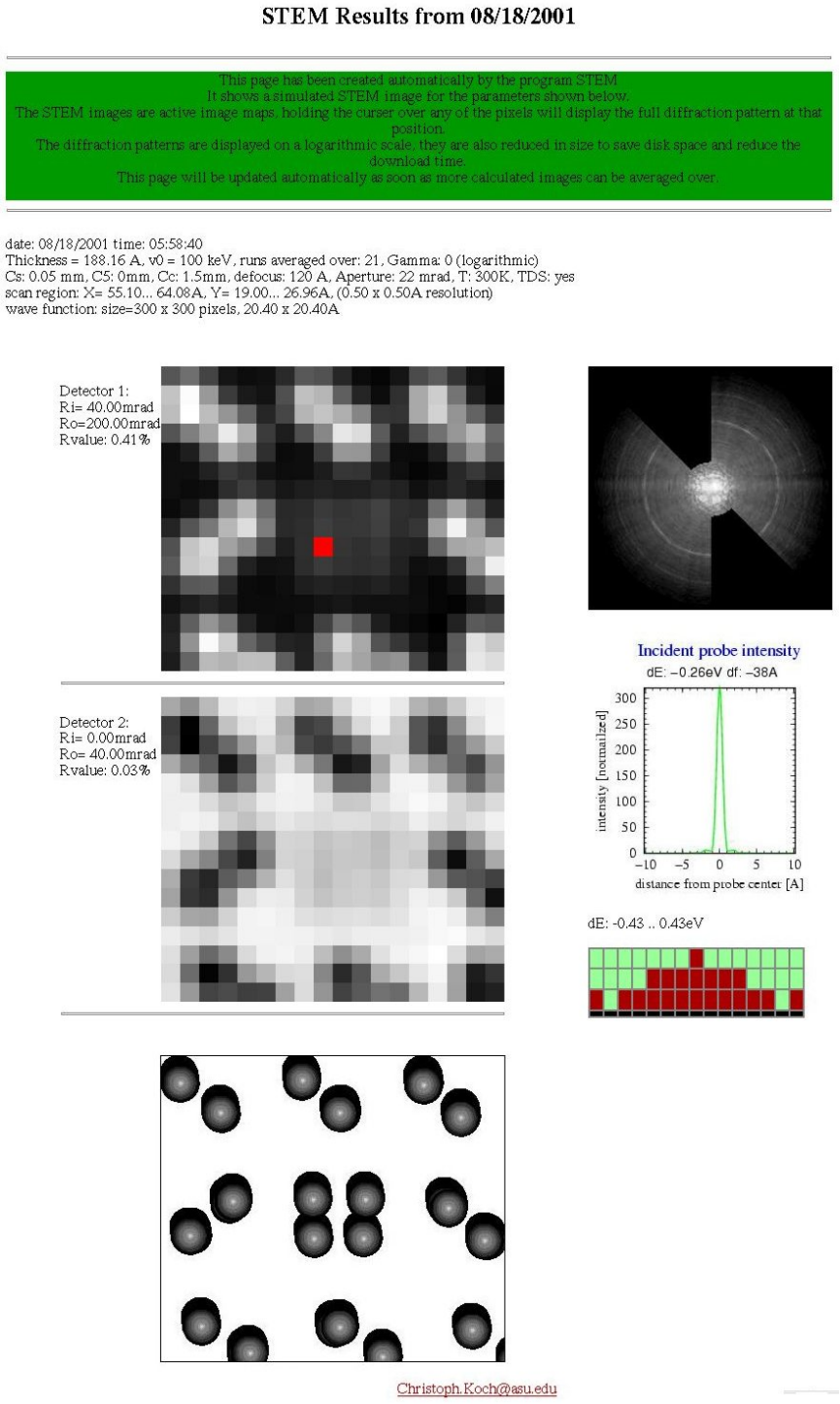

Figure 16. (color) Screen shot of the output of the *S-TEM* program for a STEM calculation of the 90° partial dislocation core for an aberration corrected 100kV VG-STEM.

CHAPTER 3

DISLOCATIONS AND PARTIAL DISLOCATIONS IN SILICON

1. Introduction

Although atomic models for dislocations in crystals of diamond-like structures (the most important semiconductors, like Si, Ge, GaAs, etc. fall into this class) have been proposed since the 1950's [86], a recent (2000) review article by Robert Jones [87] is entitled: "Do we really understand dislocations" and concludes with the statement that "it is clear that dislocations in Si, in spite of almost 50 years of effort, still possess many unexplained features." [87].

Probably the first suggestion of dislocations as the mechanism of plastic deformation in metals was made by Mügge [88] and Ewing and Rosenhain [89] in the late 19th century, even before knowledge of the crystalline structure of metals had been established [90]. The connection between the postulation of dislocations as crystalline defects and the developing field of elasticity theory [91,92] has not been made until the late 1930's. The development of X-ray diffraction techniques established the crystalline nature of many materials and gave the first experimental technique to directly observe dislocations.

2. Direct Observation of Dislocations

Direct observation of dislocations has already been possible in the 1950s, first only in transparent crystals, like AgBr and NaCl [93, 94], or by a very clever method of letting copper atoms form precipitates along dislocations in Si by a special etching and annealing procedure and imaging them with infrared light [95]. Hirsch, Horne, and Whelan were the first to observe dislocations directly by high resolution electron microscopy (HRTEM) and produce a motion picture of moving dislocations in Al [96]. Ray and Cockayne [97] showed, that 60° edge dislocations in Silicon dissociate into Shockley-partials, and recently Kolar, Spence, and Alexander [1, 2, 98] recorded the movement of kinks along a partial dislocation at a resolution of 0.33nm using HRTEM and deduct kink formation F_k and migration energy W_m from their velocity and density (see also [99–101]). Probably the highest resolution images obtained from dislocations have been recorded by 100kV (Batson [102]) and 300kV (Pennycook [103]) scanning transmission electron microscopy and through focus reconstruction of images recorded with a 300kV TEM (Kisielowski [104]).

A possible drawback of the TEM is that specimen need to be very thin because of the comparatively short inelastic mean free path of high energy electrons. X-ray topography [105] is therefore a complementary method, having much lower resolution and magnification, but allows dislocation phenomena to be observed in their "natural" bulk crystal environment. Using electron radiation for the observation of dislocations in semiconductors can also lead to effects like radiation enhanced dislocation vibration (REDV) [106] in addition to the enhancement of dislocation glide due to radiation in general [98, 106].

In chapters 4 and 5 I will discuss two different methods of direct observation of certain properties of dislocations. First in chapter 4 a method of accurately tracing the boundary of a stacking fault, formed by a 60° dislocation dissociated into two partial dislocations will be introduced, and in chapter 5 a method for the direct measurement of dislocation core periodicities will be discussed.

3. Geometry

Without extended defects such as dislocations every crystalline material would be infinitely brittle at all temperatures, up to their melting point. The theory of dislocations describes the processes necessary on the atomic scale for microscopic (e.g. misfit dislocations) and macroscopic (e.g. shear, dilatation as seen in figure 17) displacements from the perfect crystal structure. In this work we will concentrate on edge dislocations in diamond cubic materials (see [90] for a more extensive treatment of this subject).

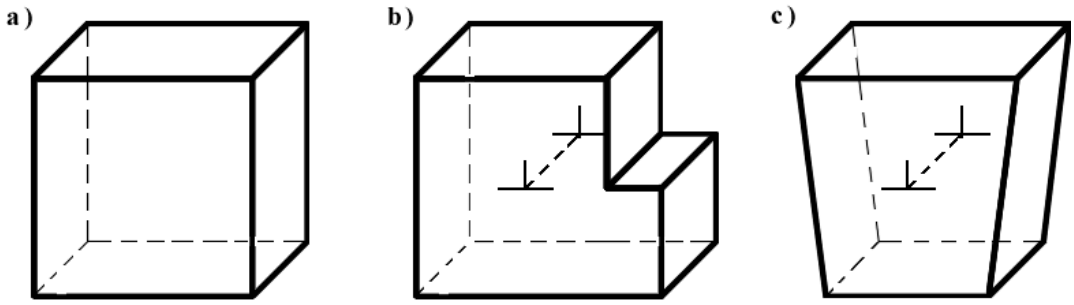

Figure 17. Operations producing pure edge dislocations a) perfect crystal cube b) shear: the horizontal lattice constant of the top part of the crystal due to shear stress differs from the one on the bottom. Misfit dislocations account for the displacement on the atomic level. c) dilatation: edge dislocations moving through the perfect crystal structure account for material shift in ductile processes.

F. Frank [107] suggested a simple procedure to define the dislocation displacement, the Burgers vector $\vec{b}$ as seen in figure 18. If the direction of the dislocation, which is not always unambiguously definable, is into the plane of the paper, the Burgers vector is

defined as the vector closing the (clockwise) Burgers circuit, which is a rectangle of an integer number of unit cells wide and high around the dislocation core. Figure 18 shows a 60° dislocation in Si running along $[\bar{1}\bar{1}0]$. The Burgers vector $\vec{b} = \frac{a}{2}[10\bar{1}]$ subtends an angle of 60° with the dislocation line.

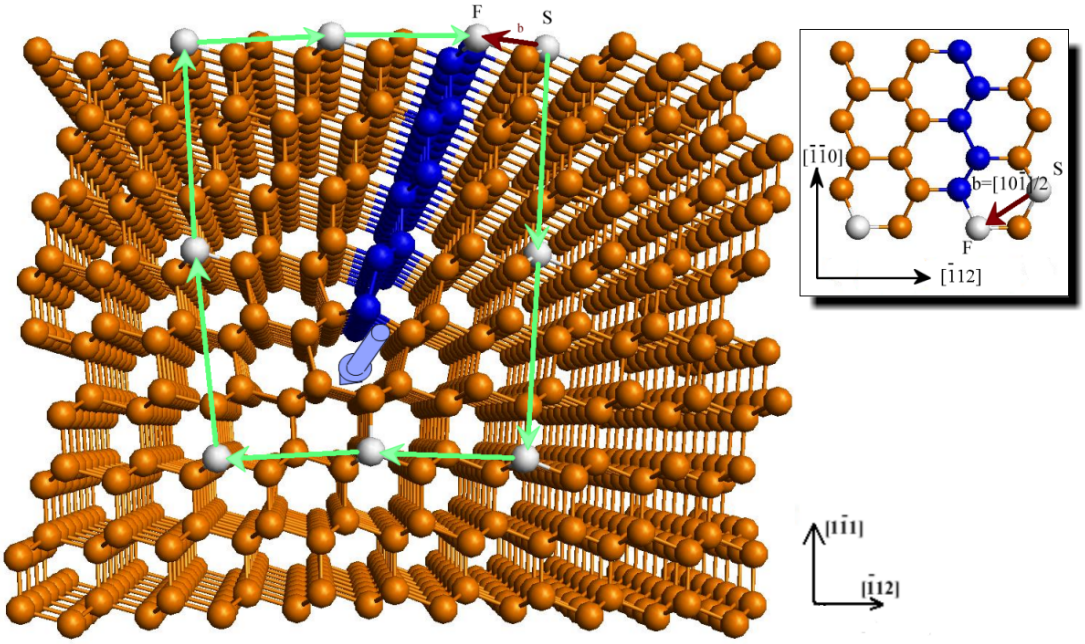

Figure 18. Burgers circuit in diamond cubic structure containing a dislocation running into the plane of the paper as indicated by the round (blue) arrow. The length and direction of the Burgers vector is the difference between the points S and F. The inset shows a portion of the top plane, showing that the Burgers vector $b = \frac{a}{2}[10\bar{1}]$ is not in the plane of the paper, but at a 60° angle with respect to the dislocation line.

In face centered cubic (fcc) crystal structures the Burgers vector has the form $\frac{a}{2}\langle 110\rangle$ and the dislocation lines are along $\langle 011\rangle$ directions. This puts the glide planes, which are normal to both, the Burgers vector and the dislocation line, in $\{111\}$ planes. Since the diamond cubic, as well as the sphalerite crystal structure, consist of two fcc sub-lattices, displaced by the vector $a(\frac{1}{4}\frac{1}{4}\frac{1}{4})$ from each other, we also have two possible types of $\{111\}$ planes, i.e. shuffle- and glide-plane, as shown in figure 19.

It was originally thought [86] and for a long time accepted that dislocations in diamond cubic structures belong to the shuffle set. Although Shockley already showed in 1953 [108] that only glide-set dislocations in diamond cubic structures can dissociate into (Shockley) partials, as in fcc metals, it took until in 1971 when Ray and Cockayne [97] showed, that 60° edge dislocations in Silicon dissociate into (Shockley) partials by direct observation of dissociated dislocations using the newly established TEM weak-beam

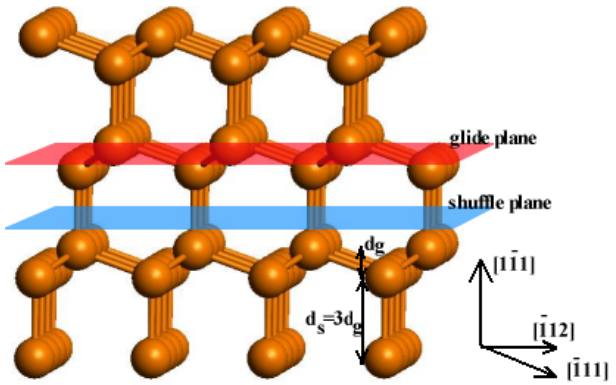

Figure 19. Projection of the diamond structure onto a (110) plane, showing glide- and shuffle-plane. d_g and d_s indicate the interplanar distances for the glide- and shuffle mechanisms.

technique, for Shockley's proposal that dislocations in the diamond-cubic structure are of the glide-set, to find wide acceptance.

60° dislocations dissociate into two Shockley partial dislocations, one with a Burgers vector of 90° to the dislocation line, and another with a 30° angle, preserving the total Burgers vector and producing a stacking fault between them. For a dislocation running along $[\bar{1}\bar{1}0]$ in the $(1\bar{1}1)$ plane we get

$$\begin{array}{rcl} \vec{b}_{60^\circ} & \rightarrow & \vec{b}_{30^\circ} + \vec{b}_{90^\circ} \\ \frac{a}{2}[10\bar{1}] & \rightarrow & \frac{a}{6}[21\bar{1}] + \frac{a}{6}[1\bar{1}\bar{2}] \end{array} \tag{3.1}$$

As a perfect 60° dislocation dissociates it produces a fault in the diamond cubic stacking sequence in the area between the two partial dislocations. Figure 21 shows the stacking sequence of the (111) layers to be *A-a-B-b-C-c*. Since the Burgers vectors of the partials are equal to the horizontal displacement vectors between the three different layers (e.g. $r_B = r_A + \frac{1}{3}[1\bar{1}1] + \frac{1}{6}[1\bar{1}\bar{2}] = r_A + \frac{1}{3}[1\bar{1}1] + \vec{b}_{90^\circ}$ in the coordinate system of figures 21 and 18) the stacking fault looks like a missing double layer. A possible stacking sequence for such an intrinsic stacking fault could for example be: ... *A-a-B-b-C-c-A-a-B-b-A-a-B-b-C-c* ..., which can be obtained by removing a *C-c* double-layer from the perfect (unfaulted) stacking sequence.

4. Dislocation Dynamics in Silicon

4.1. Energy Minimization by Dissociation. The equation for the force between any two parallel dislocations was first developed by Nabarro [109]. The radial component of the interaction force per unit length between the two partials with a distance R between

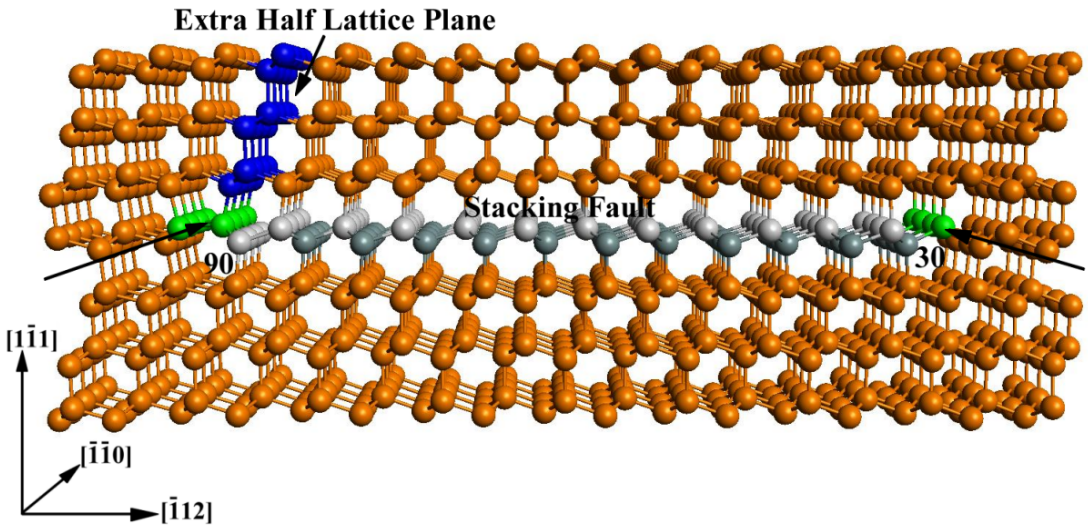

Figure 20. (color) Dissociated 60° glide-set dislocation. A stacking fault (SF) ribbon (plane of gray and white atoms) is bound by the two partials. The cores (green) of the 30° and the 90° partial are reconstructed. The extra half plane at the 90° partial is indicated by the blue atoms.

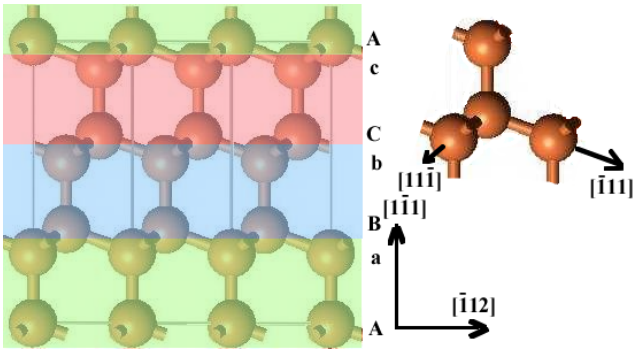

Figure 21. (color) Stacking sequence of the (111) planes in an unfaulted Si crystal.

them is

$$F_R = \frac{\mu}{2\pi R}\left(\vec{b}_1 \cdot \vec{\xi}\right)\left(\vec{b}_2 \cdot \vec{\xi}\right) + \frac{\mu}{2\pi(1-\nu)R}\left[\left(\vec{b}_1 \times \vec{\xi}\right) \cdot \left(\vec{b}_2 \times \vec{\xi}\right)\right] \tag{3.2}$$

where $\vec{b}_1$ and $\vec{b}_2$ are the Burgers vectors of the two partials, $\vec{\xi}$ the direction of the dislocation line, ν the Poisson ratio, and μ the shear modulus. For Silicon $\mu = 6.81 \times 10^4$MPa and $\nu = 0.218$ and the force between the two partials is repulsive, which means that they will separate from each other until the repulsive force is balanced by the attractive force produced by the increasing stacking fault energy. Since one of the partials has a Burgers vector perpendicular to the dislocation line, the first term in equation (3.2) vanishes, so that per unit length of the dislocation line the force between the two partials is

$$F_R^{(partials)} = \frac{a\mu}{24\pi(1-\nu)R} \tag{3.3}$$

The energy for the faulted area between the partials is composed of three kinds of terms: an energy associated with the bonds across the fault plane which are sheared by the fault, an energy term caused by the dilatation normal to the closed packed layers, and a term arising from dilatation within a close-packed layer near the fault plane [90]. The computation of the stacking fault energy is rather complicated, but it can be determined experimentally by measuring the equilibrium distance of the partial dislocations, which is of the order of $R^{equilib} \approx 50 \ldots 100$Å in Si. The force due to the stacking fault energy must counter-balance the the force due to the separation of the partials in equation (3.3) at $R = R^{equilib}$.

$$F_R^{(sf)} = -F_R^{(partials)}(R^{equilib}) = -\frac{a\mu}{24\pi(1-\nu)R^{equilib}} \tag{3.4}$$

4.2. The Peierls-Nabarro Dislocation Model and Dislocation Kinks. Equation (3.2) is based on a continuum theory, which does not take into consideration the periodic nature of the crystal lattice. However in a periodic lattice the moving of dislocation lines is associated with the breaking and forming of atomic bonds, and small displacements of atoms from their lattice positions. Figure 22 shows the self-energy of a dislocation in the diamond cubic structure as it moves from one low-energy position to the next, having to break atomic bonds (A→B) and form new ones (B→C) on the way. This produces a periodic potential "landscape", known as Peierls potential, first described by Peierls in 1940 [110] and later refined by Nabarro [111], which affects the movement of dislocations in a lattice.

The energy of a kink-pair can be broken down into the self energy W_f of a single kink and the interaction energy W_{int} between the two kinks in the pair [90]

$$W = 2W_f + W_{int} \tag{3.5}$$

For a pure edge dislocation

$$W_f = \frac{\mu b^2 h}{4\pi}\left[ln\left(\frac{h}{e\rho}\right) - \frac{1}{(1-\nu)}\right] \tag{3.6}$$

$$W_{int} = -\frac{\mu b^2 h^2}{8\pi x}\frac{1-2\nu}{1-\nu} \tag{3.7}$$

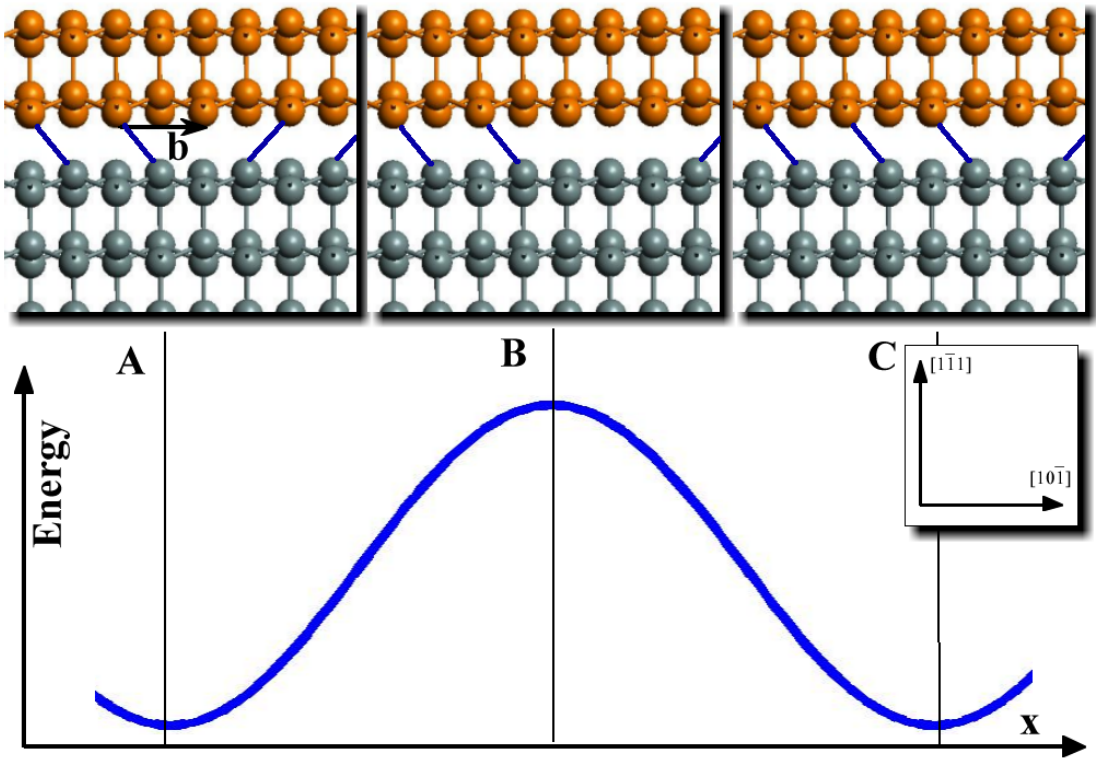

Figure 22. Approximate self-energy of a system of two semi-infinite crystals of the diamond cubic structure with dis-registry $\vec{b}/2$ in the $[1\bar{1}1]$-plane, where $\vec{b} = \frac{a}{2}[10\bar{1}]$ is the Burgers vector for a 60° dislocation. The energy (Peierls potential) as a function of dislocation position x is sinusoidal with the periodicity of the lattice and is due to small atomic displacements (not shown in the figures) and the breaking of atomic bonds (as shown in B).

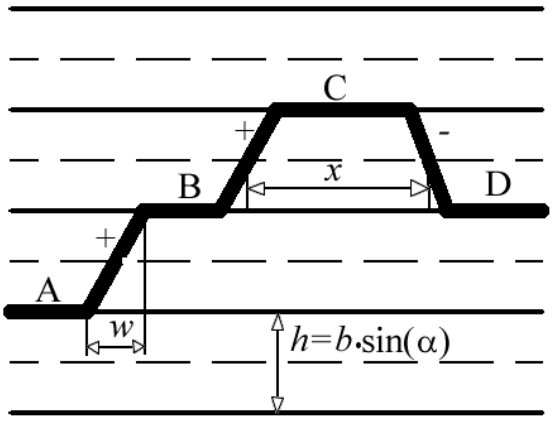

Figure 23. Segments (A-D) of a dislocation that lie in different parallel "valleys" (solid lines) of the periodic Peierls potential are connected by kinks, which traverse the Peierls potential "hills". The period of the potential is $h = |\vec{b}|\sin(\alpha)$, where $\vec{b}$ is the Burgers vector and α the angle it subtends with the direction of the dislocation line (h=3.3Åin Si). Successive kinks of opposite sign causing only a "temporary" shift of the dislocation line are called a "kink pair" of width x. w is the width of a single kink.

where b is the magnitude of the Burgers vector, h the height of the kinks, which is equal to the separation of the Peierls "valleys", e is Euler's number, and ρ is the smallest distance between the dislocation segments at which they still interact, according to the model. This cutoff-parameter ρ is a phenomenological parameter introduced to avoid the divergence of the dislocation interaction energy at very small distances by setting the interaction energy to zero for dislocations separated by less than ρ. For a kink of height $h = b\sin(\alpha)$ (α = angle between dislocation line and Burgers vector) h/ρ is of the order of 1, and the estimate of W_f of the same order as the uncertainty associated with the effects that determine the value of ρ, i.e. corner-effects, detailed atomic kink-configurations.

4.3. Kink Concentration and Mobility. Among the attempts to explain dislocation motion in undoped material are the kink-diffusion model, developed by Hirth and Lothe [90], the "weak-obstacle" theory of Celli, Kabler, Ninomiya, and Thomson [112], and a modification thereof proposed by Möller [113]. The latter theory accounts for dissociation of dislocations into partials, whereas the first two theories mentioned consider complete dislocations [114]. Because of its applicability to the work discussed in chapter 4 some of the details of the kink-diffusion model will be given.

A dislocation whose endpoints are fixed by either jogs, obstacles, or because it penetrates the crystal surface has to have a certain net amount of *geometrical kinks* if its endpoints lie in different valleys of the Peierls potential. In figure 23, for example, a net-amount of one positive kink is required. In addition to geometrical kinks there exists a thermal equilibrium of double-kinks, which account for the free energy of the system. The Helmholtz free energy of a pair of these *thermal kinks* ($2F_K$) is given by

$$2F_K = 2U_K - 2TS_k \tag{3.8}$$

where U_K and S_K are the internal energy and entropy of a single kink and T is the temperature. Since the entropy is very small compared to U_K, it is most of the time neglected. In a stress-free environment $2U_K = W = 2W_f + W_{int}$ (equation (3.5)), but more generally the energy for the constrained equilibrium is (according to Seger and Schiller [115])

$$F(x) = 2F_K - \frac{\mu b^2 h^2}{8\pi x} - \tau bhx \tag{3.9}$$

where the third term is the stress correction for an applied stress of $F_S = \tau bh$ on a kink of height h with a glide-plane component of τ [115]. This means that according to classical statistical mechanics the equilibrium concentration of thermal kink-pairs of a certain kink-kink separation x at a certain temperature T is given by

$$c(x) = \frac{1}{a^2}\exp\left(-\frac{F(x)}{k_B T}\right) \tag{3.10}$$

where a is now the spacing between possible kink sites, i.e. the periodicity of the *second order Peierls potential* along the dislocation line.

5. Partial Dislocation Core Reconstruction

Generally, dislocations in semiconductors act as electrically active defects for a number of possible reasons (deep states along the dislocation core due to dangling bonds, impurities, point defects, etc.). I will first give an overview of the experimental techniques available to give some insight into this matter and then discuss some of the recent theoretical results concerning the reconstruction of partial dislocation cores in Si. In chapter 5 a new diffraction technique for the direct observation of periodicities in the dislocation core reconstruction will be discussed.

5.1. Observation of Electronic Properties of Dislocations. Electronic properties of materials, in particular at low temperatures, are defined by their band structure in the vicinity of the Fermi-level. For metals the Fermi-level is within the conduction band, which allows charge carriers to travel freely without the need of thermal excitation into a conducting state. Semiconductors on the other hand are defined by the existence of a band-gap, a band of inaccessible states between the fully filled (in the low temperature limit) valence band and the empty conduction band, and the Fermi-level between the two.

Since the band structure depends on the atomic structure and charge-distribution (its Fourier transform) one expects it to be different in the vicinity of defects, like dislocations. If extended defects such as dislocations introduce charge carrier states within the band-gap then these states could be accessed easier by electrons in the valence band or holes in the conduction band and turn dislocations into 1-dimensional wires. The matter of reconstruction of dislocation cores is therefore very important, because it is generally believed that unreconstructed "dangling" bonds produce deep levels within the band-gap. But even if the cores are reconstructed (i.e. all valence electrons of every atom along the core are involved in covalent bonds) there may still exist shallow states within the gap, locally altering the electronic characteristics of the material.

- *Hall Effect Measurements:* From Hall effect measurements the carrier density as a function of temperature $n(T)$, $p(T)$ can be determined. Comparing crystals before and after deformation the change of carrier density can be measured and, if the dislocation density is known, an average occupation ratio f of carriers along the dislocation core can be estimated, as long as we neglect the contribution of point defects and impurities introduced by the deformation. Two possible models for the charge distribution along dislocation cores were proposed by Read [116] (single, localized acceptor levels in the band-gap) and Schröter and Labusch [117] (a one-dimensional band of states, which is neutral, when half filled, and can therefore accept as well as give up electrons). Hall measurements [118] have strongly been in favor of the latter model [114, 118]. However Ono and Sumino [119, 120], (as well as Kisielowski and Alexander using EPR spectroscopy) came to the conclusion that at least in Silicon the consequences of plastic deformation for the electrical properties are mainly due to point defects not located along dislocation lines.

It must be noted that neither one of the models [116, 117] accounts for dissociated dislocations and the stacking fault between them, which can possibly also carry states in the gap [114]. This is especially important, since theoretical calculations show that partial dislocations are most likely to be reconstructed, forming covalent bonds without deep states in the band-gap. It is therefore important to perform more localized experiments on single (partial) dislocations.

- *Electron Paramagnetic Resonance (EPR) Spectroscopy:*

  Assuming that unreconstructed dangling bonds along dislocation cores are paramagnetic centers, i.e. they have a non-zero net-electronic spin, EPR spectroscopy can reveal them, and with proper calibration also give the number of defects in the specimen. The structure of the spectrum can even give information about the symmetry of the defects and the chemical species of atoms involved. Despite two decades of experimental work applying EPR to deformed Si (see [114] for a review) the debate over the interpretation of the results is still not settled. Ossypian et al [121] propose a magnetic phase transition between defects that a) are and b) are not related to the dislocation anisotropy at about 60K, while Alexander and Kisielowski-Kemmerich [122, 123] propose the coexistence of both types of defects, because the spectra of both types can be detected at any temperature. Pohoryles [124] gives an alternative explanation of the EPR spectra based on photoluminescence data of deformed Si and Ge in a high pressure He environment concluding that dislocation cores are unreconstructed.

- *Deep Level Transient Spectroscopy (DLTS):*

  Using Shottkey diodes deposited on the crystal surface one can measure the voltage- and frequency-dependent capacitance of bulk-and defect structures within the depletion region of the diode. DLTS has become the most effective technique to establish the density and position of deep levels in the gap. In application to dislocations in Silicon this technique has been able to identify three [125], or at least one [114] peak in the DLTS spectrum with deep levels along the dislocation. Similar to EPR this method is also unable to investigate single point- or extended defects, but has to fit the sum of responses of all defects in the region investigated to some kind of model.

- *Electron Beam Induced Current (EBIC):*

  This technique employs a scanning electron microscope (SEM) beam on a sample with a thin electron-transparent Schottky contact (usually evaporated Al) on top of it. The Schottky contact is reverse biased. An image at the typical SEM resolution is formed by displaying the amplified leakage current as a function of (2-dimensional) electron beam position. The electron beam induced (minority) carriers either recombine at defects or are collected at the Schottky contact thus carrying information about the effective minority carrier life time, which will be reduced by electrically active defects. Higgs et al. [126] and Wilshaw [127] were able to ascribe EBIC contrast (as well as

PL contrast [128]) to contamination by metal (Cu, Fe, Ni) atoms during the deformation process by comparing measurements of Si samples deformed under ultra-pure conditions before and after contamination by back-plating with Cu, Fe, and Ni, even for contamination levels too low to be detected as precipitates at the dislocation core by TEM (below 1 monolayer of metal deposition on the surface). These results and others raise the question, whether clean dislocations are electrically active at all.

- *Photoluminescence (PL):*

  In 1976 Drozdov et al. [129] were able to attribute 4 PL lines (D1 ... D4) to the presence of dislocations in silicon. Since then the D1 and D2 bands have been attributed to intrinsic properties of dislocations, such as electronic transitions at stacking faults, kinks, point defects trapped in the strain field, whereas the D3 and D4 bands are considered to be associated with electronic transitions at the dislocation core [130]. This method usually samples volumes of hundreds of μm^3 at a time, but recently a high resolution PL technique for Si and SiGe structures (SiPHER [131,132]) has been developed capable of a resolution of 1 μm. PL clearly proves the existence of shallow states near or in the dislocation [114].

- *Microwave Conductivity (MWC):*

  Brohl and Alexander [133, 134] have shown that microwave conductivity (MWC) at 9GHz dominates by several orders of magnitude in the direction of dislocations at low temperatures (20K). Brohl [134] ascribes this result to band conductivity in shallow bands near the edges of the valence and conduction bands.

- *Electric Dipole Spin Resonance (EDSR):*

  Electrically induced spin-flip of electrons moving in a shallow band split from one of the conduction band edge minima by the dislocation strain-field can be observed [114] by this technique first discovered as a possible application for the study of dislocations in semiconductors by Kveder [135, 136].

- *Electron Energy Loss Spectroscopy (EELS):*

  Using a high energy (100kV) electron beam with 0.3eV energy spread focused to a spot size of 0.2nm and an EELS spectrometer with 160meV resolution Batson [102] has recently shown spatially resolved EELS measurements with numerically enhanced 0.25eV energy resolution of a dissociated 60° dislocation in a Si_xGe_{1-x}-alloy. Fitting linear combinations of calculated spectra of possible local atomic environments he was able to decide in favor of a modified version of the double period structure for the 90° partial dislocation [137]. However, he also notes that while his model structures were periodic along the core there would be very many other possible structures involving kinks explaining his results.

As this list shows there exists a variety of experimental techniques to probe the electronic properties of defects in semiconductors. But all of them are indirect methods,

relying on the interpretation of their data according to some model. As shown by Higgs et al. [126,128] and Wilshaw [127] it is possible that most or even all of the "observed" electronic activity of dislocations in semiconductors is simply due to impurities in the specimen, which has not been taken into account by the models used for the interpretation of experimental data. The availability of *direct* measurement of dislocation core properties is therefore of great importance.

5.2. Total Energy Calculations. All physical properties of materials are related to the total energy or differences between total energies of the system of electrons and nuclei [138]. Total energy calculations have been successfully used to predict lattice constants, bulk moduli, phonons, piezoelectric constants, phase transition pressures and temperatures, etc. (for a review see [139–141]). Despite the extremely rapid development and availability of fast computers most ab-initio methods for solving the Schroedinger equation for a multi-body system of atoms can currently only handle up to about 100 atoms. However, because of the long-range deformation of the crystal structure, surrounding defects, total energy calculations of systems containing defects require a very large number of atoms.

Because of the large difference in mass between the electrons and nuclei, and the fact that due to the equal charge of electrons and protons the forces acting on those particles are the same, the electrons respond essentially instantaneously to the motion of the nuclei, leading to a separation of electronic and nuclear coordinates in the many body wave-function, the so-called Born-Oppenheimer approximation. This reduces the many-body problem to one, in which the electrons move in the Coulomb potential created by the "frozen" configuration of nuclei. Parameterizing the electronic wave functions subject to the true ion-electron interaction potential using plane waves (Bloch's theorem, eqn. (2.7)) requires a rather large basis set of plane waves. Due to the fact that most physical properties of a material mainly depend on its valence electrons, the strong ion-electron interaction potential can be replaced by a much weaker *pseudo-potential*, which approximates the ion potential screened by the core electrons, as "seen" by the valence electrons [142, 143]. The following list mentions some of the pseudo-potential methods used most frequently:

- *Stillinger Weber Potential (SW):* Developed by Stillinger and Weber [144] this potential consists of two- and three-body interaction terms fitted to experimental properties of the diamond-cubic and molten form of silicon. It has been used for the study of lattice dynamics [145], point defects [146, 147], surfaces [148], and the liquid and amorphous phases. [144, 149–152]

- *Tersoff Potential:* Developed by Tersoff (3 versions: T1 [153], T2 [154], and T3 [155]) this potential consists of many-body interactions included in a bond order term and was fitted to ab initio results for several Si polytypes. It has been used to study lattice dynamics [145], thermo-mechanical properties [156], point defects [153, 154] and the liquid and amorphous phases [152, 154, 157].

- *Environment Dependent Inter-atomic Potential (EDIP):* In trying to overcome some of the drawbacks of the SW- and Tersoff-potentials, like their lack of transferability to a wide variety of structures, especially those far from equilibrium, Justo, Bazant, Kaxiras, Bulatov, and Yip [158] introduced a new empirical potential for silicon using a theoretically motivated functional form which emphasizes chemical and physical trends, using only a fairly small number (13) of parameters. This potential represents a considerable improvement over existing models in describing local structures and extended defects [158]. It provides a good description of point defects in the bulk, the exchange path for self-diffusion, and elastic properties of bulk silicon. Its ability to predict core structures of partial dislocations in the glide set (111) in excellent agreement with ab initio results were one of the main reasons for its application in this work.

The calculations of structural models used in chapter 5 as well as for figures 18 and 20 in this chapter were done by João Francisco Justo Filho at the Instituto de Fisica da Universidade de São Paulo, Brazil using the EDIP method developed by him and co-workers [158]. João Justo's description of how he did the calculation have been included in appendix C.

5.3. Periodicity of the Partial Dislocation Cores in Silicon. Experimental results and theoretical calculations, as mentioned above all agree on the fact that partial dislocations in silicon are reconstructed. In fact theory predicts only a single energetically favorable structural model for the core of the 30° partial (shown in figure 24), which has a periodicity twice that of the lattice along the direction of the core.

For the 90° partial however two possible core structure models exist, both breaking the symmetry along the direction of the core, and therefore producing at least two energetically identical solutions. The first, proposed independently by Hirsch [159] and Jones [160] in 1979, has the same periodicity along the core as the perfect crystal structure and is called the single period (SP) model. The second one proposed by Bennetto et al. [137] has a periodicity twice as long along the core and is called the double period (DP) model. The difference in total energy $E_{SP} - E_{DP}$ between the two models at 0K has been calculated by different authors to be somewhere between -0.021eV and 0.079eV [137, 161, 162], depending on the model and numerical method used. This difference in energy is comparable to $k_B T = 0.025$eV at room temperature and it may therefore be possible that both reconstructions are present at high enough temperatures.

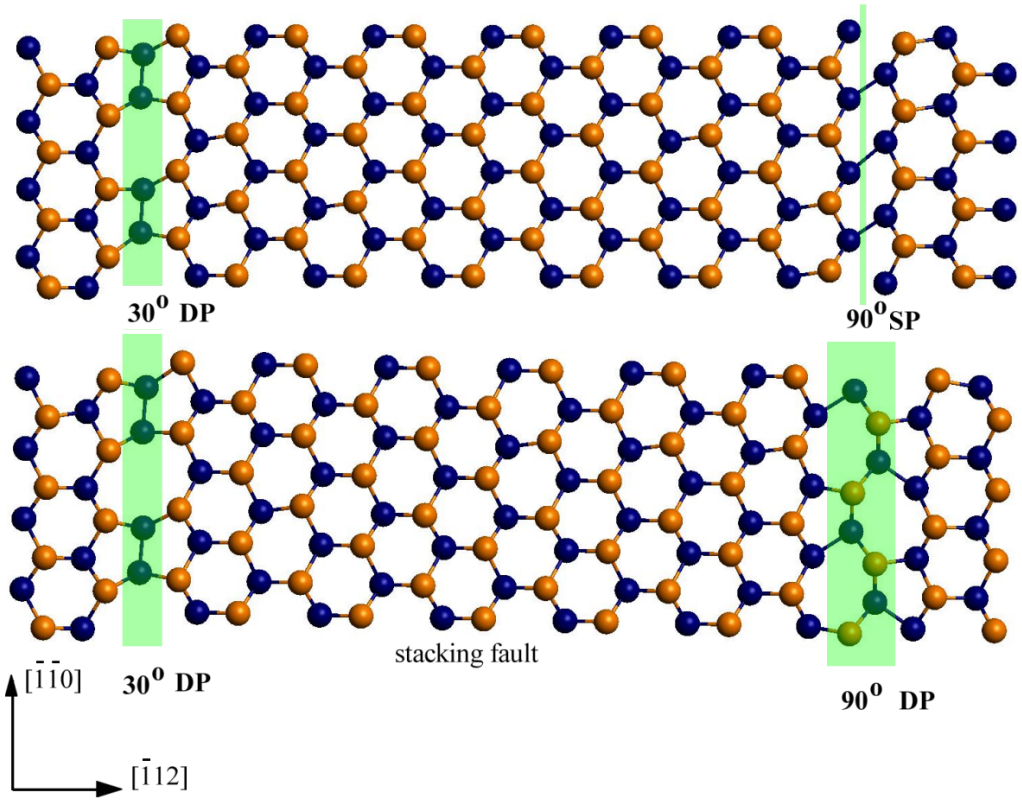

Figure 24. (1$\bar{1}$1) projection of the layer of atoms in the stacking fault plane of the dissociated 60° dislocation. Dark blue atoms are below and orange atoms above the stacking fault plane. Regions of period doubling are shaded in light green. Note that the 90° partial of the model on top (90° SP) has no period doubling along its core at all, the narrow shaded area only marks its location.

CHAPTER 4

IMAGING OF PARTIAL DISLOCATION KINKS

1. Introduction

It has been shown in chapter 3 that we can determine the kink formation energy F_K from a measurement of the concentration of kinks along partial dislocations. Several groups have already obtained experimental results (Kolar, Spence, and Alexander [1, 98] using forbidden reflections lattice imaging [FRLI] [98, 163] and Suzuki and Maeda et al [99, 100] using a Fourier filtering technique [101]), but their quality has generally not been good enough for an unambiguous determination of kink positions. In this chapter I will give one of the the reasons for the insufficient quality of the experimental results given in [1, 98] and discuss ways to improve them along with some experimental results.

Parts of the work described in this chapter have been presented at the "Extended Defects in Semiconductors 2000" conference in Brighton, UK in Juli 2000. The proceedings of this conference are contained in a special issue of "Journal of Physics: Condensed Matter **12**", where our paper can be found on page 10175 [164].

2. Forbidden Reflections Lattice Imaging FRLI

In 1974 Cherns [163] published astonishing dark-field micrographs of a 400 Å thin Au film showing surface steps of atomic height by using one of the weak, kinematically forbidden $\frac{1}{3}[422]$ reflections. Forbidden reflections are vectors in reciprocal space for which the structure factor of the full unit cell is zero, but may be non-zero for one or several slices with a surface normal parallel to the incident electron beam taken from the unit cell.

Figure 25a shows the unit cell of the zinc-blende structure in its commonly used cubic representation. If the crystal is oriented with the electron beam normal to one of the {111} surfaces, then it is more convenient to work with hexagonal coordinates. The transformation of lattice vectors

$$\begin{aligned} \vec{a'}_x &= \frac{1}{2}\vec{a}_x - \frac{1}{2}\vec{b}_y \\ \vec{b'}_y &= \frac{1}{2}\vec{a}_x - \frac{1}{2}\vec{c}_z \\ \vec{c'}_z &= \vec{a}_x + \vec{b}_y + \vec{c}_z \end{aligned}$$

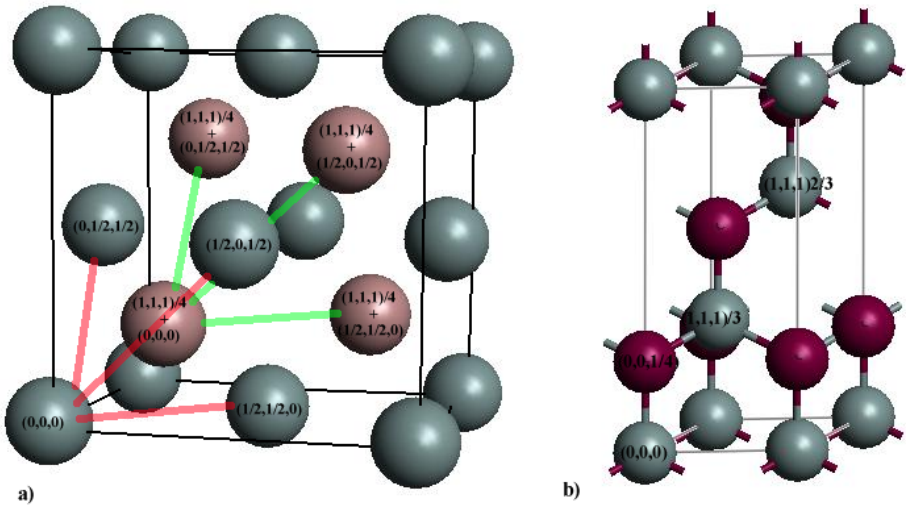

Figure 25. Model of the zinc-blende structure with fractional atomic coordinates (diamond cubic if all are the same type of atom). a) Orthogonal unit cell: The two fcc sub-lattices are emphasized. b) Hexagonal unit cell

produces a hexagonal unit cell, as shown in figure 25b, with a unit cell volume of

$$\Omega' \quad = \quad \vec{a'}_x \cdot [\vec{b'}_y \times \vec{c'}_z] = \frac{3}{4}\Omega$$

where $\Omega = \vec{a}_x \cdot [\vec{b}_y \times \vec{c}_z]$ is the volume of the cubic unit cell. The reciprocal lattice for the hexagonal unit cell relates to that of the cubic cell in the following way:

$$\begin{aligned}
\vec{g'}_x &= \frac{1}{\Omega'}\vec{b'}_y \times \vec{c'}_z = \frac{1}{3}\left(2\vec{g}_x - 4\vec{g}_y + 2\vec{g}_z\right) \\
\vec{g'}_y &= \frac{1}{\Omega'}\vec{c'}_z \times \vec{a'}_x = \frac{1}{3}\left(-4\vec{g}_x + 2\vec{g}_y + 2\vec{g}_z\right) \\
\vec{g'}_z &= \frac{1}{\Omega'}\vec{a'}_x \times \vec{b'}_y = \frac{1}{3}\left(\vec{g}_x + \vec{g}_y + \vec{g}_z\right)
\end{aligned}$$

Using the hexagonal unit cell and denoting the 2 atom types in the zinc-blende structure by the letters A and B the structure factor for any reflection $\vec{g'} = h'\vec{g'}_x + k'\vec{g'}_y + l'\vec{g'}_z$ is computed by the following formula

$$\begin{aligned}
f_{h',k',l'} &= \sum_{j=1}^{6} f_j^e(\theta_{h',k',l'}) \exp\left(2\pi i\left[h'x'_j + k'y'_j + l'z'_j\right]\right) \\
&= \left[f_A(\theta_{h',k',l'}) + f_B(\theta_{h',k',l'}) \exp\left(\frac{2\pi i l'}{4}\right)\right] \\
&\quad \left[1 + \exp\left(\frac{2\pi i}{3}(h'+k'+l')\right) + \left(\frac{4\pi i}{3}(h'+k'+l')\right)\right] \qquad (4.1) \\
&= 0, \text{ if } h'+k'+l' = 3m \pm 1 \qquad (4.2)
\end{aligned}$$

where $\sin(\theta_{h',k',l'}) = \lambda|\vec{g'}_{h',k',l'}|$ is the scattering angle associated with the reciprocal lattice vector defined by (h', k', l'), x'_j, y'_j, and z'_j are the atomic positions in the in fractional coordinates in terms of the hexagonal unit cell lattice vectors, and m is an integer.

The structure factor for any reflection (h', k', l') that satisfies (4.2) vanishes, if we sum over the atoms of the full unit cell, but each term in the sum, which corresponds to the structure factor for a single double layer (as defined in figure 27), by itself is non-zero. This means that the total (kinematic) structure factors of these forbidden reflections are only zero for a crystal with $3n$ ($n = 0, 1, \ldots$) such layers, but finite for crystals with $3n + 1$, or $3n + 2$ layers, or a fault somewhere in the stacking sequence of the layers.

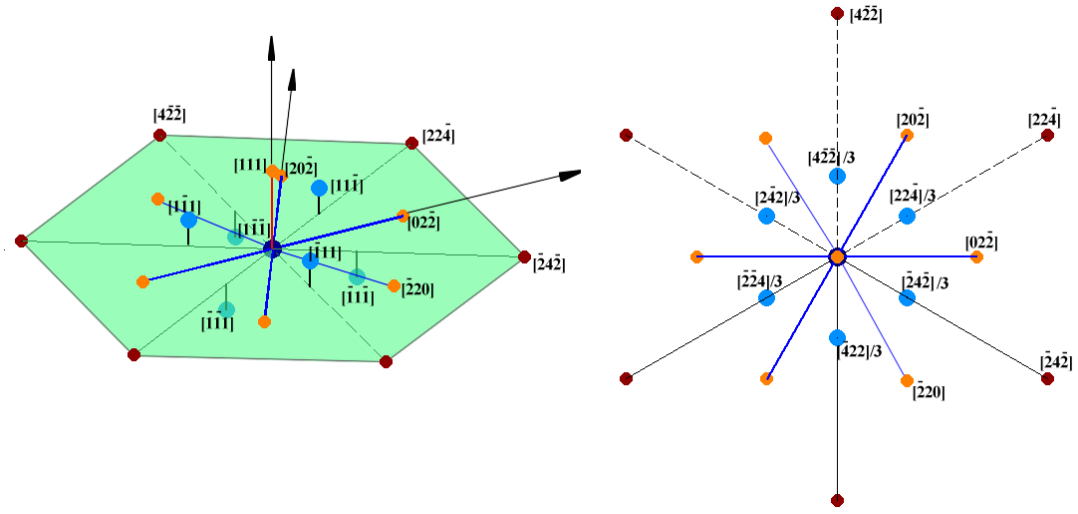

Figure 26. (color) Diagram in reciprocal space showing forbidden reflection. a) 3-dimensional plot of the points in reciprocal space. The $\langle 11\bar{1}\rangle$ reflections are not in the zero-order Laue plane but above and below it. The vertical lines extending from the $\langle 11\bar{1}\rangle$ rec. lattice points show the excitation errors. Since the Ewald sphere has a finite curvature the excitation errors for the points below the zero order Laue plane (green hexagon) are slightly bigger than the ones for spots above it. b) 2-dimensional projection onto the [111] plane of a). The $\langle 11\bar{1}\rangle$ reflections appear at $\frac{1}{3}\langle 4\bar{2}\bar{2}\rangle$ in projection. All reciprocal lattice vectors are given in the cubic notation. The arrows in b) define an orthogonal 2-dimensional set of reciprocal lattice vectors (q_x, q_y) used later in the text.

The shortest g-vectors in the [001] zone axis satisfying condition (4.2) are [010], [0$\bar{1}$0], [100], [$\bar{1}$00], [110], and [$\bar{1}\bar{1}$0], which corresponds to [$\bar{4}$22]/3, [4$\bar{2}\bar{2}$]/3, [2$\bar{4}$2]/3, [$\bar{2}\bar{4}$2]/3, [$\bar{2}\bar{2}$4]/3, and [22$\bar{4}$]/3 respectively in the cubic notation. Figure 26 shows that the presence of these $\langle 4\bar{2}\bar{2}\rangle/3$ reflections can also be explained by the existence of $\langle 11\bar{1}\rangle$-type reciprocal lattice points just above and below the zero order Laue plane. The excitation error of these

reciprocal lattice points is

$$s_g = \frac{1}{\sqrt{3}a}, \tag{4.3}$$

assuming a perfectly flat Ewald sphere, which is a reasonably valid approximation for such short g-vectors, at least to first order. The finite radius of the Ewald sphere requires a second order correction of

$$\Delta s_g = \frac{1}{\lambda}\left[1 - \cos\left(\operatorname{atan}\left(\lambda\left|\vec{g}_{\frac{1}{3}\langle\bar{4}22\rangle}\right|\right)\right)\right] \approx \lambda\left|\vec{g}_{\frac{1}{3}\langle\bar{4}22\rangle}\right|^2 = \frac{8\lambda}{3a^2} \tag{4.4}$$

for s_g, which must be added/subtracted, depending whether the reciprocal lattice point is below or above the zero order Laue zone. This slightly different excitation error will split the forbidden reflections into 2 groups of 3, i.e. diffraction amplitude and phase will be equal for only 3 of the forbidden reflections in the zone axis case. Figure 26a shows the 3-dimensional reciprocal lattice indicating the $\langle 11\bar{1}\rangle$ points and the direction of their excitation errors. Figure 26b shows how these reciprocal lattice points appear as $\frac{1}{3}\langle 4\bar{2}\bar{2}\rangle$ in the 2-dimensional projection. An experimental diffraction pattern of β-Silicon Carbide with forbidden reflections is shown in figure 41.

Figure 27 shows a dynamical calculation of the amplitude of the $\frac{1}{3}[4\bar{2}\bar{2}]$ forbidden reflection as a function of thickness. It can be noticed that the scattering amplitude for an integer multiple of whole unit cells is close to zero before we reach the stacking fault, but finite in between. The stacking fault then offsets the scattering amplitude by its value at the fault, and it does not vanish anymore. If we use the forbidden reflections to form an image (either by dark-field with a small objective aperture around one of the spots, or bright-field where we choose the aperture big enough for the first-order forbidden but not the bulk reflections to pass through), we will be able to see atomic height surface steps and stacking faults. Also after the electrons have passed through the stacking fault, the peak value of the scattering amplitude is about twice that of the perfect crystal, which enables us to distinguish stacking fault contrast from pure surface step contrast.

The d-spacing for the $\frac{1}{3}\langle 422\rangle$ reflections is

$$d_{\langle 4,2,2\rangle/3} = \frac{a}{\sqrt{h^2+k^2+l^2}} = 3.31\text{Å} \tag{4.5}$$

where $a = 5.4$Åis the lattice constant for Silicon. This means that a bright-field image with an objective aperture just big enough to let the $\frac{1}{3}\langle 422\rangle$ reflections pass would give us an image with about 3.3Åresolution. Since we can identify the boundary of the stacking fault with the partial dislocations, we are able to image the course of the partials and locate kinks along it with high resolution.

3. HRTEM and STEM Image Simulations

In order to interpret experimental partial dislocation kink images and find ways to increase their quality, a number of multislice calculations were performed using the program

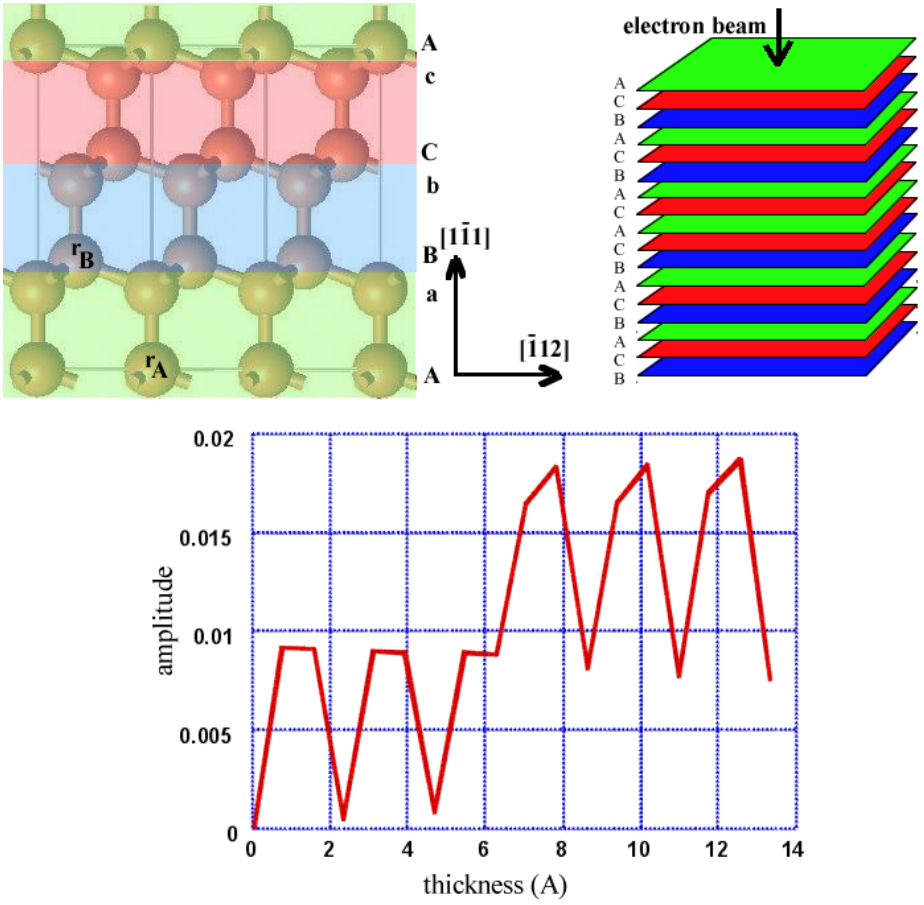

Figure 27. (color) Pendellösung plot of $\frac{1}{3}[4\bar{2}\bar{2}]$ forbidden reflection. The amplitude is plotted as a function of thickness recorded at each one of the layers, as defined on top left. The sequence of double layers is shown on the top right (thickness increases in direction of arrow, i.e. the direction of the electron beam).

'mulslice' [35], assuming a 200kV TEM. Temperature effects were accounted for by applying a Debye-Waller factor of 0.44Å^2 for Si. All multislice calculations were done using 1024 x 1024 beams.

Figures 28 and 29 show a horizontal slice through the models used for the image simulation for a dipole of two 30° partials and two 90° partials respectively. The atomic coordinates for the different core reconstruction and kink models were taken from tight-binding-total-energy calculations performed by Nunes et al [165]. A variety of point defects, such as different proposed structures of kinks, direction switching defects, etc. have also been included, in order to test whether the resolution of this imaging method is potentially high enough to resolve differences between certain kink structures. The super-cells can be repeated periodically, which avoids artifacts in the Fourier transform based multislice algorithm due to abrupt changes in the atomic structure along the boundaries. In order to make the simulations as realistic as possible, a thickness of 92Å to 100Å was chosen.

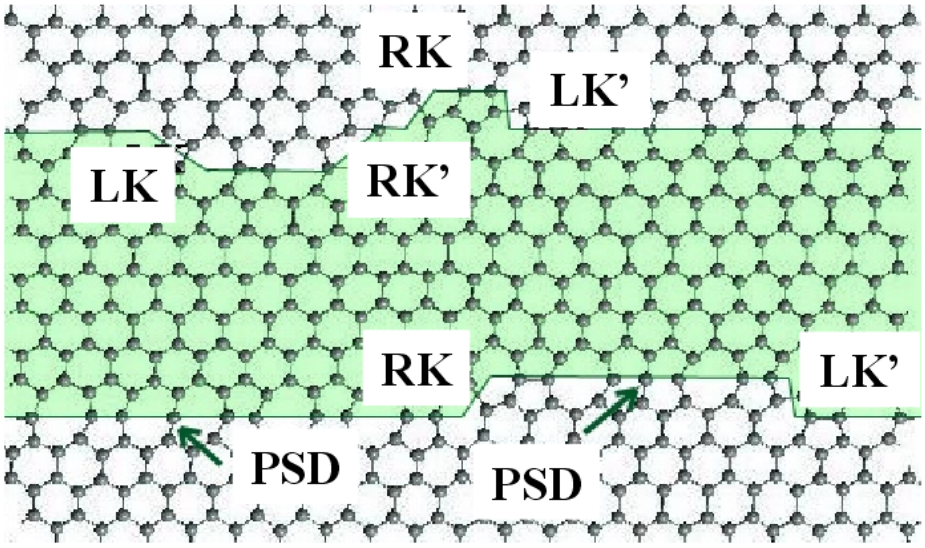

Figure 28. (color) Structural model of the atomic layer in the stacking fault plane of a dipole of two 30° partials. A variety of defects [165] is also represented: PSD = phase switching defect, LK, LK' = different left kinks, RK, RK' = different right kinks

3.1. New application of spherical aberration corrected TEM. The development of several ultra-high resolution schemes (see [166] for a summary of seven new approaches) including the use of aberration correctors, high voltage microscopes and through-focus schemes, promise that a resolution of about one Angstrom will be routinely available for HREM in the near future (site permitting). This raises the question of whether one might be able to distinguish kink models from this type of image, viewed normal to the dislocation line, using forbidden reflections. Earlier simulations for kink images at 0.3nm

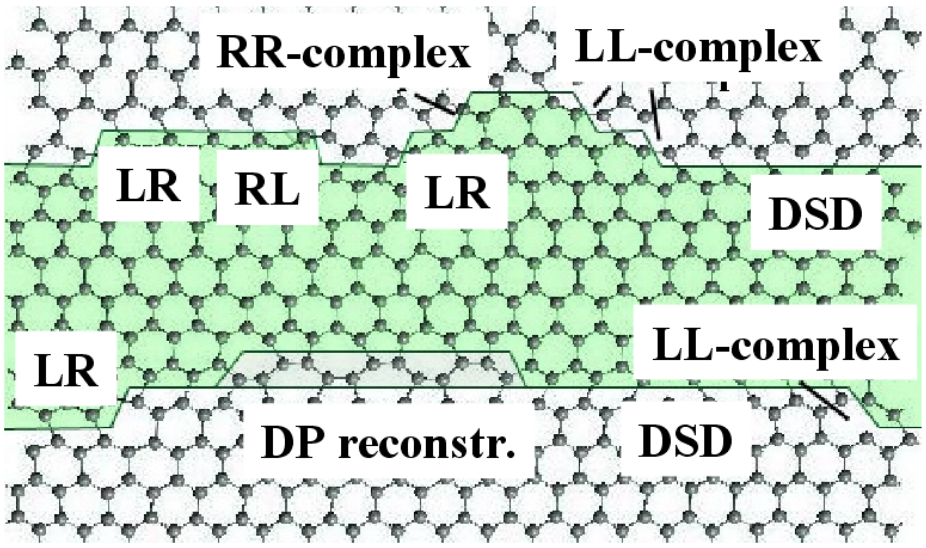

Figure 29. (color) Model of a 90° partials dipole similar to figure 28. LR/RL = kinks switching direction of reconstruction, DSD = direction switching. defect, RR-complex = LR(RL)+DSD, LL-complex = LR+DSD, DP = segment of double period reconstruction (see also chapter 5)

resolution [167] were unable to do this. Since forbidden reflections in Si are very weak compared to the bulk reflections, the only way to obtain high image contrast from the stacking fault is to allow only the forbidden reflections to contribute to the image and block the bulk reflections by an appropriate objective aperture. Recording images at a resolution comparable, or even better than that needed for resolving the Si(220) lattice fringes would require a ring aperture that blocks the bulk reflections, but lets higher order forbidden reflections pass.

The first spherical aberration C_s corrected TEM has recently been demonstrated [47]. Being able to control C_s of the objective lens makes it possible for it to be used as a tunable ring aperture as can be seen from equation (2.28) and figure 8. As already described in section 3.5 of chapter 2 the exit face electron wave-function $\Psi(\vec{r})$ is convoluted with the objective lens transfer function (equation (2.27)) giving the wave-function in the image plane (Ψ_I).

$$\Psi_I(\vec{r}) \quad = \quad \Psi(\vec{r}) \otimes \mathrm{FT}^{-1}\left[e^{-i\chi(\vec{q})A(\vec{q})}\right] \tag{4.6}$$

Using the weak phase object approximation for the exit face wave function

$$\Psi(\vec{r}) \quad = \quad \exp(i\phi(x)) \approx 1 + i\phi(x) + O\left(\phi(x)^2\right)$$

the image intensity can be approximated as

$$I(\vec{r}) \quad = \quad |\Psi_I(\vec{r})|^2 = \left|\mathrm{FT}^{-1}\left[\Psi(\vec{q})\exp(i\chi(\vec{q}))A(\vec{q})\right]\right|^2$$

$$\begin{aligned} &\approx \quad \mathrm{FT}^{-1}\left[\delta(\vec{q}) + 2\phi(\vec{q})\sin(\chi(\vec{q}))A(\vec{q})\right] \\ &= \quad 1 + 2\mathrm{FT}\left[\phi(\vec{q})\sin(\chi(\vec{q}))A(\vec{q})\right] \end{aligned} \tag{4.7}$$

where $\chi(\vec{q})$ is given by expression (2.28) and $A(q)$ is a top-hat function describing the objective aperture. Assuming that the non-cylindrically symmetric aberrations, like astigmatism, have been corrected, the remainder of (2.28) is dominated by C_s and the defocus df.

$$\chi(q) \quad = \quad \pi\lambda df q^2 + \frac{1}{2}\pi\lambda^3 C_s q^4 \tag{4.8}$$

where $q = |\vec{q}|$.

According to equations (4.7) and (4.8) one can find a combination of C_s and df analytically for any given λ which let the MTF $\sin(\chi(q))$ go to zero for the lowest bulk material spatial frequency q_B but negative for the first and second order forbidden reflections at $q_F = q_B/\sqrt{3}$ and $2q_F = 2q_B/\sqrt{3}$.

$$\begin{aligned} \chi(q_B) = 0 \quad &\Rightarrow \quad df = -\frac{1}{2}C_s\lambda^2 q_B^2 \\ \sin(\chi(q_F)) = -1 \quad &\Rightarrow \quad df = \frac{1}{\pi\lambda q_F^2} - \frac{1}{2}\lambda^2 q_F^2 C_s \\ &\Rightarrow \quad C_s = \frac{1}{(q_B^2 - q_F^2)\pi\lambda^3 q_F^2} = \frac{9a_x^4}{128\pi\lambda^3} \qquad (4.9) \\ &\qquad\quad df = -\frac{q_B^2}{(q_B^2 - q_F^2)\pi\lambda q_F^2} = -\frac{9a_x^2}{16\pi\lambda} \qquad (4.10) \end{aligned}$$

Figure 30b shows such a MTF ($\sin[\chi(q)]$) as the dash-dotted curve. Since (4.7) is only valid for weak phase objects, but starts to break down for 10nm thick samples, the parameters C_s and df have been optimized with respect to the ratio of I_F/I_B, the ratio of intensities of the spatial frequency components q_F and q_B in the Fourier transform of the image $\tilde{I}(\vec{q}) = \mathrm{FT}[I(\vec{r})] = \mathrm{FT}[|\Psi_I(\vec{r})|^2]$ for fully dynamical multislice calculations. For a 200keV TEM the optimum parameters have been found to be $C_s = 0.225$mm and $df = -341$Å, which result in the ratio $I_F/I_B = 1.2 * 10^4$ (see figure 31). This means that the contribution of the first and second order forbidden reflections is much stronger than the contribution of the bulk reflections. The optimized values for C_s and df allow us to record an image with an objective aperture about twice as large as the one used for the conventional FRLI method [98]. While a fixed spherical aberration can produce similar shapes of the MTF, being able to adjust C_s makes it possible to tune the maximum with $MTF(q) = 0$ to almost any desired spatial frequency q.

It should be noted here that the increase in objective aperture size means an increase in resolution by a factor of two, but only in certain crystallographic directions. Since the stacking fault also produces bulk-like reflections, we will be filtering away some periodic information from the stacking fault image as well as part of the diffuse scattering due to non-periodic features like the stacking fault boundary. However, this method still means a considerable increase in resolution in comparison with the conventional FRLI method, as seen in figure 32.

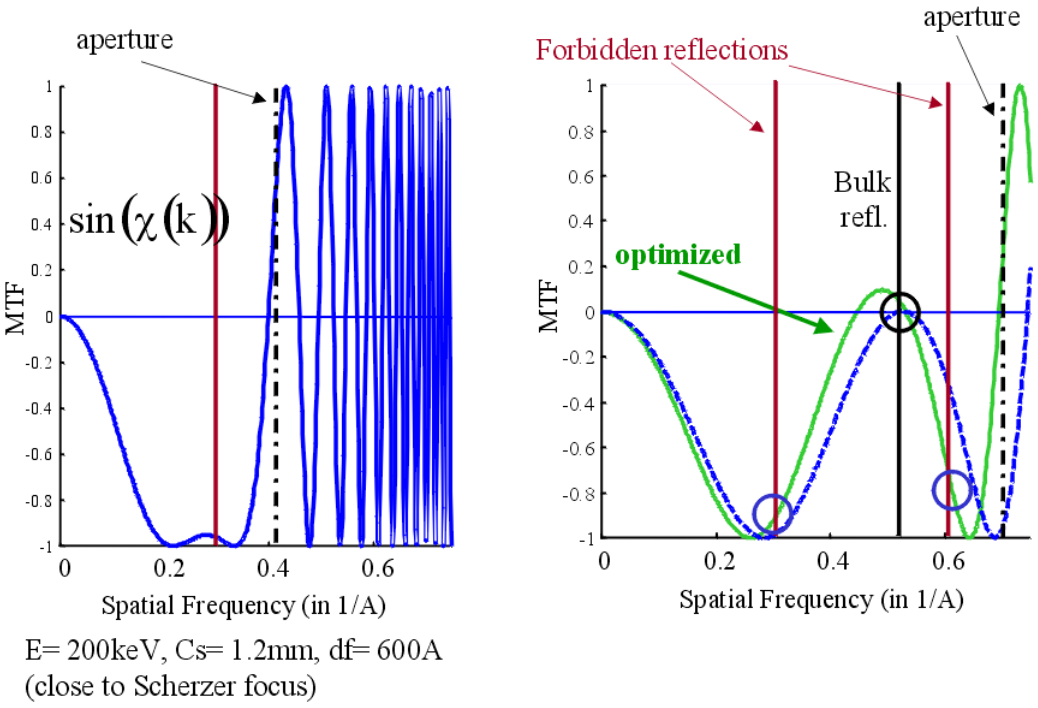

Figure 30. MTFs for a 200kV TEM a) MTF of a conventional TEM without aberration correction, and b) a possible MTF of an C_S-corrected TEM. The dash-dotted lines indicate the radius of the aperture. The dashed curve shows the MTF with C_S and according to expressions (4.9) and (4.10). The solid curve shows the MTF for $C_s = 0.225$mm and $df = -341$Å, the parameters giving optimum forbidden reflections image contrast.

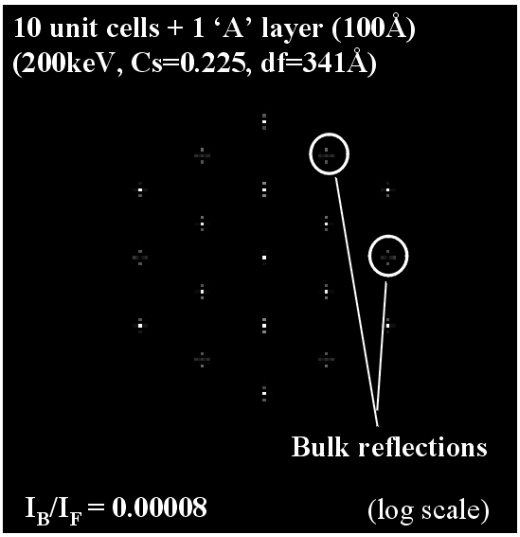

Figure 31. Fourier Transform of a simulated image at optimized imaging conditions of a model without any stacking fault but an extra surface layer, which produces forbidden reflections. Even though the pattern is displayed on a logarithmic scale, the bulk-reflections are barely visible. The ratio of intensities of the Fourier Transform of the image $I_B/I_F = 8\times10^{-5}$ shows that hardly any bulk-crystal information is being transferred by the objective lens under these imaging conditions. The stacking sequence for this model was 10(ABC)A (97Å thick).

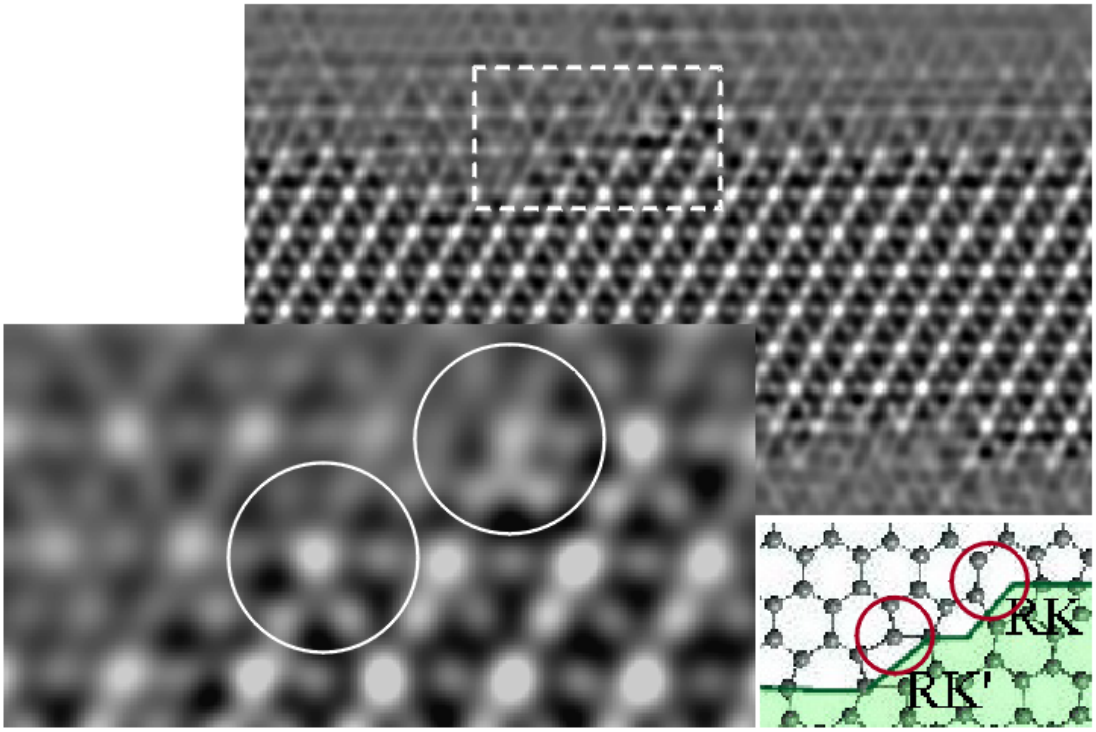

Figure 32. Fourier filtered multislice simulation of 30° partial dislocation dipole model using optimized image parameters $E = 200$kV, $C_s = 0.225$mm, and $df = -341$Å. Different kink models (see diagram on bottom right) proposed by Nunes [165] can potentially be distinguished.

In order to eliminate any residual bulk contribution to the image the image of an unfaulted crystal has been subtracted from the one containing the SF. This can also be realized with experimental data by making sure that the exposure contains a region of perfect crystal, which can then be subtracted from the region containing the SF. In addition to this subtraction the images have also been Fourier filtered. The single steps of this image manipulation are shown in figure 33 for the 30° partial dislocation dipole model.

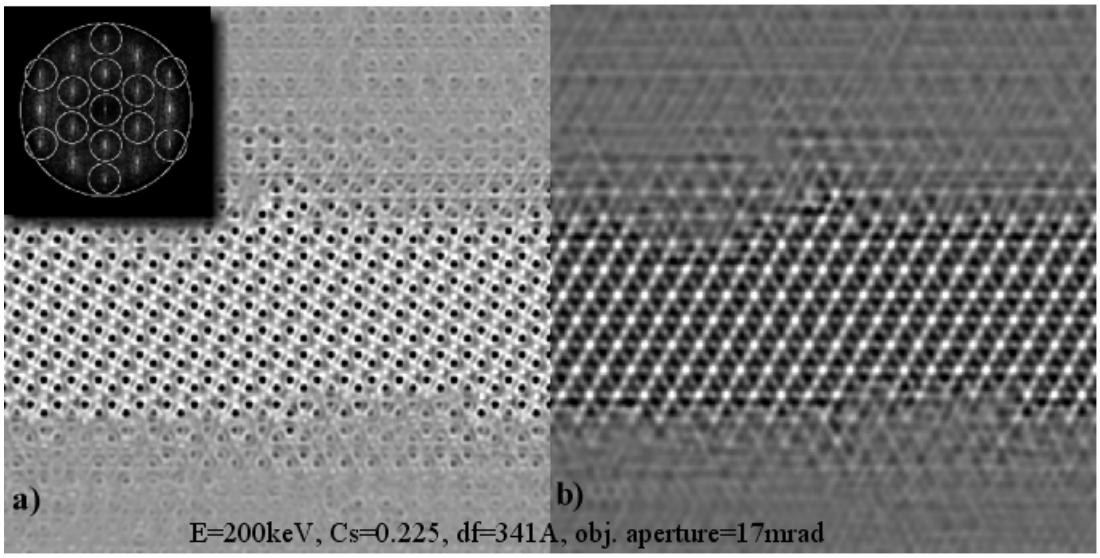

Figure 33. Image processing steps. a) HREM image of 30° partial dislocation dipole and the stacking fault in between with image of perfect lattice subtracted. b) HREM image after Fourier filtering with a mask around only the forbidden reflections of first and second order, as shown in the inset. The resolution of the image is now orientation-dependent.

Figure 34 shows the multislice simulation for the model of a Silicon crystal containing a 90° partial dislocation dipole with the dislocations running along $[\bar{1}10]$, viewed along [111] with the (111) SF in the middle of a 100Å thick slab. A TEM similar to the one used by Haider et al [47] is assumed at 200keV. The resolution is limited to 0.15nm by an objective aperture transmitting beams up to $q\lambda = \alpha_{max} = 17$mrad. The inset in figure 34 shows the spatial filtering of this image that has been applied in order to remove all the bulk allowed reflections from the image. Only the spatial frequencies contained within the white circles are present in the image. We see that the position of the kinks can be clearly identified to within an accuracy of less than 0.3nm. This resolution improvement, combined with the advantage of atomically flat surfaces (discussed in the next sub-section) will make accurate measurements of the kink density possible, and will also allow the identification of obstacles along the line. A measurement of the density of these would allow one to distinguish between the Hirth and Lothe theory [90] of dislocation motion (as limited by kink formation

or migration energies) and the Celli and Thomson theory [112], in which obstacles provide the rate-limiting step. Apart from the exact outline of the stacking fault, different kinds of defects can be detected, such as direction switching defects along the SP reconstructed 90° partial and also whether we have SP or DP reconstruction. The simulations for the 30° partial (figure 32) even show some differences between different kink models proposed by Nunes et al. [165].

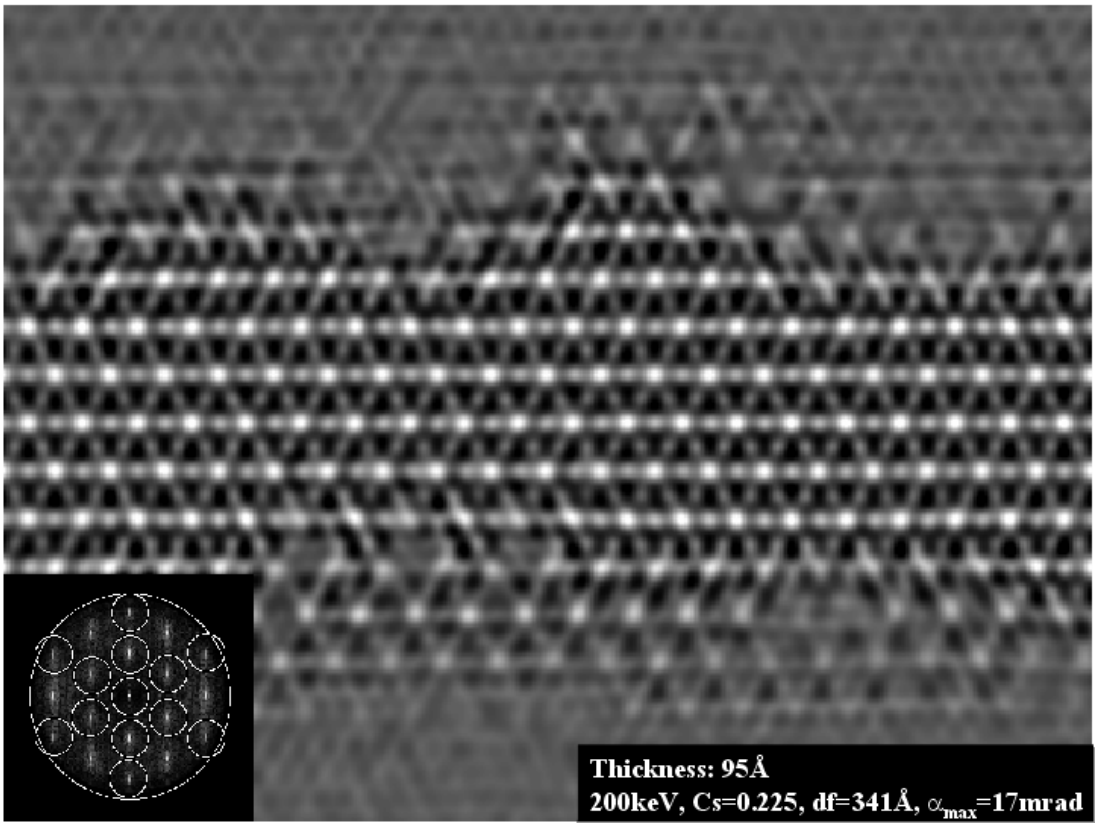

Figure 34. High resolution image of the 90° partial dislocation dipole model (figure 29). The image processing has been done in the same way as in figure 33b).

3.2. Effect of Surface Roughness on Image Contrast. Since the image produced by forbidden reflections is only due to a single double layer in in the plane of the stacking fault, we must also expect the contribution of the two surfaces to be rather strong, especially since forbidden reflections can be produced or annihilated by additional surface layers. A rough surface consisting of incomplete layers of Si atoms will cancel the image contrast in some regions and possibly enhance it in others.

Ourmazd, Anstis and Hirsch [168] thoroughly investigated the dark-field contrast produced by forbidden reflections in regions of perfect and faulted silicon crystal. They came

to the conclusion that contrast produced by additional surface layers and/or a stacking fault can be equal. I therefore decided to simulate the rough crystal surface and perform a fully dynamic simulation of the BF-TEM image in order to quantify effects of surface roughness. The simulation of the rough surfaces was done according to a very simple algorithm:

1. Distribute a few atoms randomly over the atomically flat surface, but only in legitimate lattice sites, as "seeds" for additional atoms to gather around.
2. Add more atoms and let them perform a random walk, until they "bond" to one or two other atoms already on the surface
3. Add atoms in a second layer, in some of the places where they can form 3 bonds to atoms below.

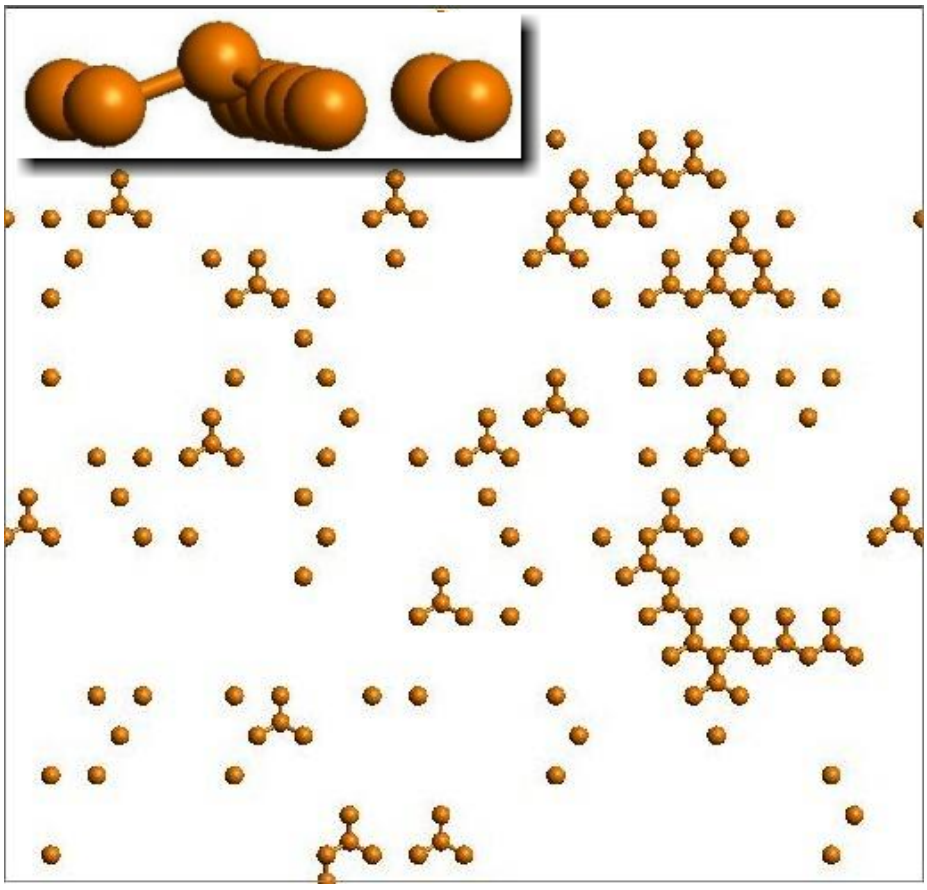

Figure 35. Simulated rough surface layer. The inset shows a cross-section in a place where an atom could find 3 other atoms to bond to in order to create a second layer.

Figure 35 shows the simulated rough surface layer which has been added to the top

of the model containing a 90° partial dislocation dipole. Performing the same multislice calculation as for figures 34 and 32, but with a conventional TEM (without aberration corrector), an image with a much fuzzier outline of the stacking fault is obtained. Figure 36 compares images computed from models with and without atomically smooth surfaces. It can be seen very clearly that the rough surface does not totally destroy the stacking fault contrast, but it has the effect of "smearing" the otherwise sharp boundary of the stacking fault, making it almost impossible to exactly locate and estimate the density of partial dislocation kinks. This effect can be observed in the images recorded by Kolar [1, 98].

Having established the fact that atomically flat surfaces very much enhance the visibility of partial dislocation kinks the question of the experimental implementation arises. It is a well know fact that Si oxidizes very easily and it is therefore very complicated to produce atomically flat surfaces on Si. In principle there exist only two possible approaches to this problem:

1. *Produce Si samples with atomically flat surfaces.*

   Figure 37 shows an experimental TEM micrograph taken under UHV conditions of a Si sample with atomically flat surfaces (see section 4 for experimental details). Areas of constant thickness form terrace like contrast, differing from sinusoidal Pendellösung fringes observed for wedge shaped samples without atomically smooth surface steps. In figure 38 the subsection of figure 37 outlined by the dashed rectangle has been analyzed. In order to estimate the average intensity value of the terraces a Matlab script has been written that enables one to outline each terrace with the mouse and fill the polygon defined in this way with the color corresponding to the average intensity of that particular terrace. The average intensities and lengths of terraces have then been plotted as a function of horizontal distance. Since the overall thickness of the specimen increases from the left to the right within this image the overall intensity changes due to Pendellösung oscillation of the direct beam, as indicated by the broken lines in the plot. The distribution of intensity levels apart from the central beam Pendellösung oscillations will be discussed in section 4.2.

2. *Use a different material with the zinc-blende crystal structure, but chemically much more inert surfaces, enabling us to prepare atomically flat surfaces more easily.*

   Striking bright-field TEM images of the surfaces of ceramics have recently been obtained by Suznitsky [169] and Ndbubuisi [170] using a high temperature anneal process which results in atomically smooth terraced surfaces which are preserved at atmospheric pressure. β-SiC has almost the same structure (zinc-blende instead of diamond-cubic) as Si and therefore the same needed forbidden reflections. Figure 39 shows a Pendellösung plot for the amplitude of the $1/3[\bar{4}22]$ Bragg reflection as a function of thickness in β-SiC for a [111] beam direction. This amplitude is zero for slabs containing 3n double layers and also for the second order $2/3[\bar{4}22]$ forbidden reflection. A stacking fault has been included. The forbidden reflection in β-SiC oscillate with about twice the amplitude (hence four times the intensity) of the forbidden

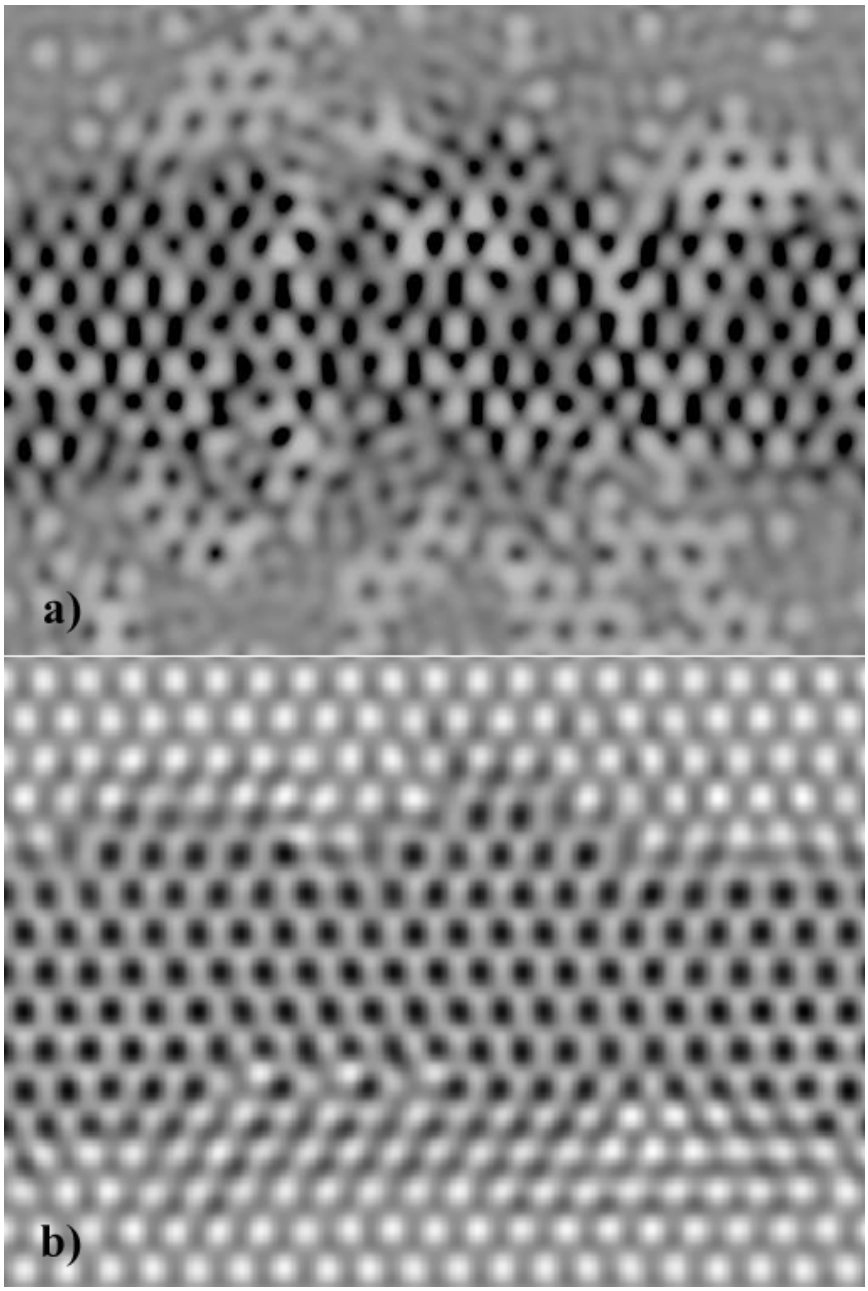

Figure 36. Simulated image of the 90° partial dislocation dipole (figure 29) model a) with b) without rough surface layers. $E = 200$keV, $C_s = 1.2$mm, $df = 600$Å, $\alpha = 9.8$mrad, thickness of atomically flat model: 10 unit cells (92Å). Subtracted perfect crystal image from both to enhance contrast. The simulation in a) shows very similar washing out of details along the stacking fault boundary as the experimental ones [1, 98]

reflections in Si and should therefore give even stronger image contrast. Samples of β-SiC with atomically smooth surfaces have been prepared and experimental results will be described in section 4.

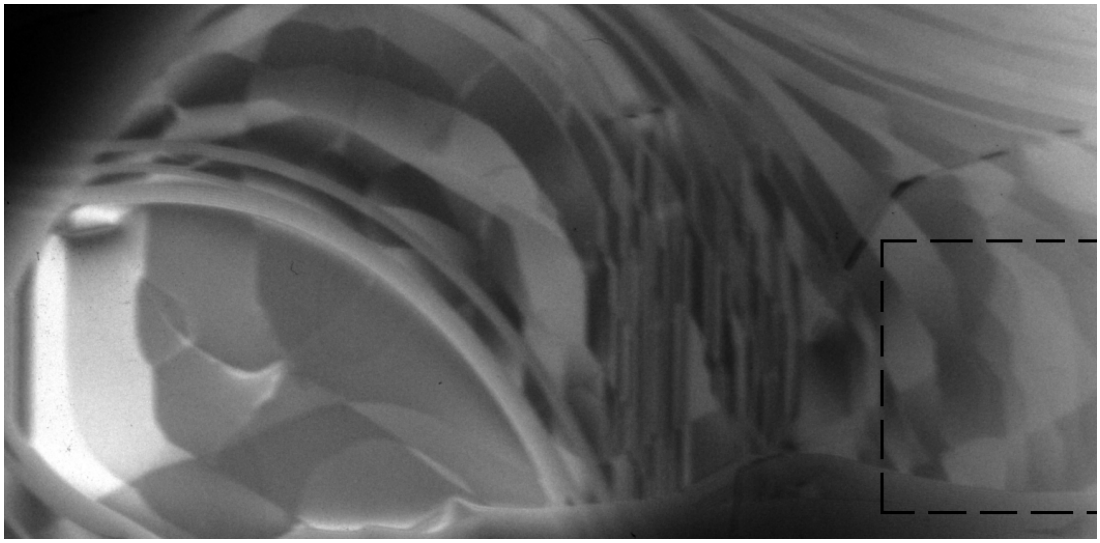

Figure 37. BF image of atomically flat Si (111) near a hole produced by the etchand during sample thinning. The aperture was chosen large enough to include the forbidden reflections. Atomically flat terraces on bottom and top of the sample intersect in the projection and produce areas of n, $n+\frac{1}{3}$, and $n+\frac{2}{3}$ unit cells thickness causing the visible contrast. Recorded by Dr. Frances Ross on 300kV UHV Hitachi TEM at the IBM T. Watson Research Center in Yorktown Heights (unpublished).

3.3. STEM simulation. Figure 40a shows a simulated STEM image recorded on a specially designed detector. The beam energy is 200kV, C_s = 1.8mm, df = −400Å, temperature = 300K. The objective aperture used for this simulation produces a beam divergence of 6.5mrad, for which the bulk diffraction disks almost touch each other and causes a probe intensity distribution shown in figure 40b. Non-overlapping diffraction disks cannot produce a lattice image in BF-STEM, and we therefore expect a constant, probe position independent signal on the detector in the absence of forbidden reflections. However, the presence of forbidden reflections, which appear at a diffraction angle of 7.6mrad, overlap with the central beam and the $\langle 220 \rangle$ diffraction disks (see figure 40d), producing probe position dependent interference fringes. The detector (shaded area in figure 40d) has been chosen to collect most of the area of overlap between the forbidden and bulk reflections. The center of the detector is cut out in order to avoid an unnecessary high offset signal produced by the part of the central disk that does not overlap with the forbidden reflection disks.

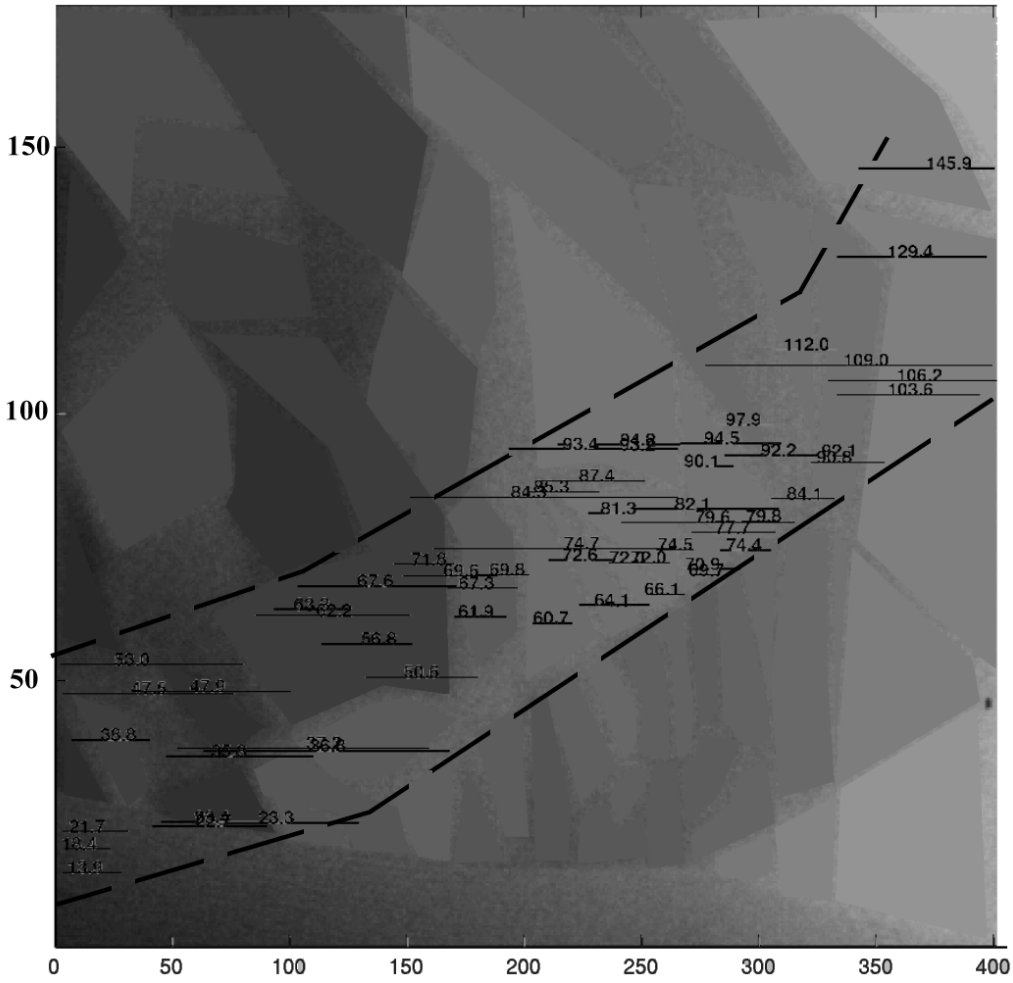

Figure 38. Distribution of intensity levels (arb. units) and size of atomically flat terraces as function of distance (in units of image pixels) in the region marked by the box in figure 37. The area from figure 37 used for this analysis is shown in the background of the plot. Terraces of constant intensity level apart from statistical noise have been replaced with "patches" of constant value showing the size and position of the terraces whose average intensity and length is shown in the plot. See text for further details.

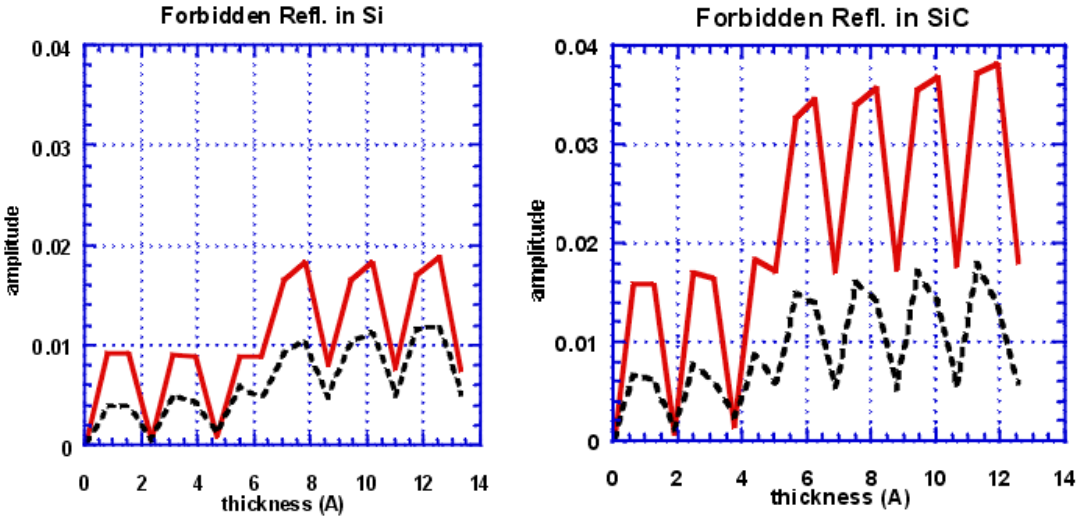

Figure 39. Pendellösung plot of the amplitude of the $\frac{1}{3}[\bar{4}22]$ (solid) and $\frac{2}{3}[\bar{4}22]$ (dashed) forbidden reflections for Si (left) and β-SiC (right). Notice that the amplitude for SiC is about twice that for Si.

Figure 40c shows the structural model of the plane of the stacking fault between two 30° partial dislocations of opposite Burgers vectors as used for this simulation. This model is a subsection of the one shown in figure 28. The abbreviations are explained in the caption of figure 28. The white arrow in figure 40 points to a feature in the image that looks like a kink pair of very small separation. Comparing that feature with the model in figure 40c we see that it is produced by a phase switching defect (PSD). Nunes [165] gives 2 possible versions of this kind of defect of which both are shown in figure 40. The PSD involving an isolated atom in the core of the dislocation does not produce forbidden reflections because this atom is only very slightly displaced from the position that it would take in the perfect crystal structure. It is therefore important to simulate the effect of certain core structures on the image in order to be able to interpret experimental images correctly.

Batson [171] has performed this experiment on a VG HB-501 dedicated STEM and obtained results very similar to those described in [98], also with similar (insufficient) image quality. However, I do not know the exact detector configuration used in this particular experiment, because the results have not been published.

4. Experiments, Results and Discussion

Experiments pursuing both approaches mentioned in section 3.2 to obtain atomically smooth surfaces have been performed. This section will describe experiments using β-SiC as well as Si samples.

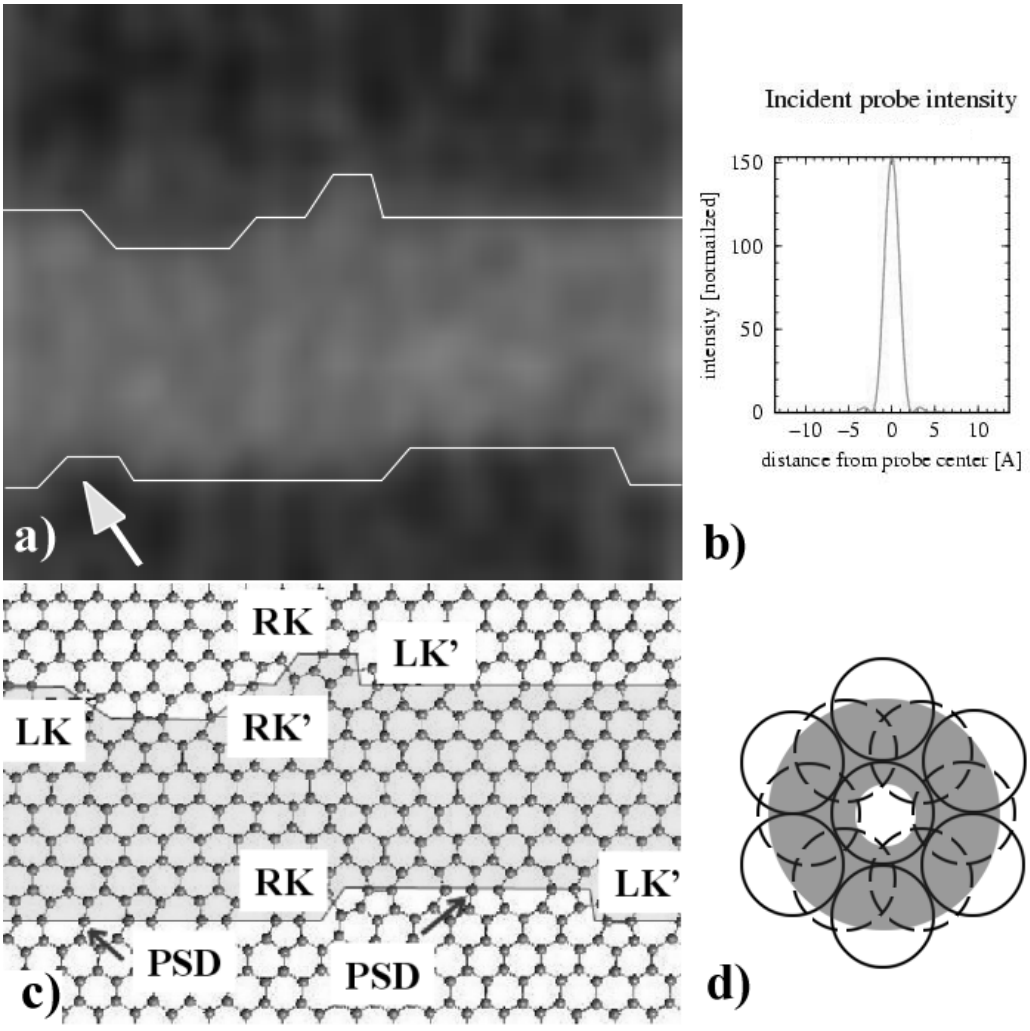

Figure 40. STEM simulation of 30° partial dislocation dipole. a) Simulated STEM image. The white lines are a guide for the eye. b) Plot of the intensity of the electron probe used to form this image. c) Diagram of the atomic positions in the plane of the stacking fault (see caption of figure 28 explanation of abbreviations). d) Diagram showing the shape of the detector and its size relative to the diffraction pattern. The dashed circles represent the forbidden reflection disks. See text for further explanation.

4.1. Experiments with 3C-SiC. Samples of 0.75μm thick β-SiC (also called 3C-SiC) film grown on the (111) surface of a 1mm thick Si wafer have been provided by C. Zorman and M. Mehregany at Case Western Reserve University in Cleveland, OH. From this wafer disks with a diameter of 3mm have been cut using an ultrasonic drill.

The disks have then been wedge-polished with a wedge-angle of 2°. The final polishing step was done with a soft cloth and a chemo-mechanical polishing agent. Residual Si from the bulk-substrate was then removed by selective etching using a 33% solution of KOH. The need for further ion-beam milling was not given, because the samples were already very thin. Wedge shaped samples were chosen over dimpled and ion-milled ones, because of the much larger thin area available for electron microscopy, and also because SiC is a strong enough material to produce such self-supporting thin films.

The wedge shaped samples have then been annealed at temperatures between 1000°C and 1300°C in order to produce the atomically flat surface [169]. The sample has been placed in a small crucible with a lid made from a rod of ultra-pure graphite. The crucible was then placed into a heating coil made from Tungsten wire, which was installed in a small vacuum chamber designed especially for this purpose. Heat was then generated by passing a high current through the Tungsten wire holding the crucible. The temperature was monitored through a window in the vacuum chamber using an infrared spectral analyzer. Great care was taken to keep the pressure inside the vacuum chamber below 10^{-6} torr while heating the sample. Degasing of the sample and crucible therefore required a very slow increase in temperature. Figure 41 shows the diffraction pattern of a sample prepared in the way described above (except for the indentation), containing forbidden reflections.

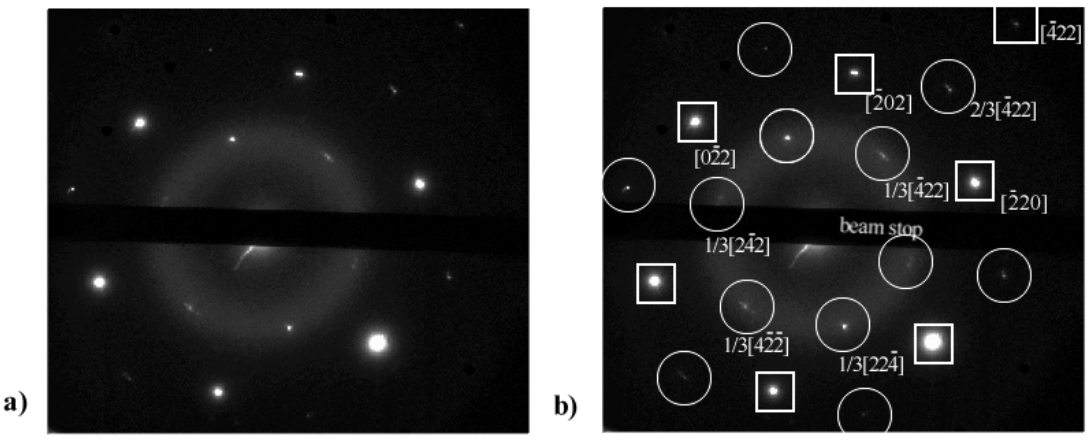

Figure 41. Selected area diffraction pattern of SiC (111) with strong forbidden reflections a) pure pattern b) indexed pattern. The forbidden reflections are marked by circles and the bulk reflections by squares. Notice that some of the forbidden reflections are as strong as the [$\bar{4}$22] bulk reflection. Recorded on LEO 912 at 120kV.

Figure 42 shows a BF-TEM image recorded at 400kV using an aperture which was big enough for the forbidden reflection to pass through. The sample has been annealed at 1250°C and a pressure of 7.5×10^{-7}torr for about 12 hours using the vacuum chamber described above. One can start to see steps in the contrast levels, especially in the upper part of the image. However, the areas of constant contrast are still fairly small. Further experiments are needed to perfect this method.

Figure 42. BF-TEM image of SiC (111) demonstrating partial success in creating atomically flat surfaces. Recorded on JEOL4000 at 400kV and a magnification of 60000×.

Some of the unpolished 3mm disks of thin SiC(111) film grown on a Si(111) substrate have been sent back to Case Western Reserve University where J. Chung (post-doc under P. Pirouz) indented them at 800°C and 1000°C in air at a loading of 50g for 5sec using a Vickers indenter apparatus [164]. Plastic deformation like that caused by indenting produces dislocations in the crystal. These samples will be used in future experiments to image dislocations and their dissociation into partial dislocations, and in particular kinks along these partials.

4.2. Experiments with Si. In parallel to the experiments with β-SiC described in the previous subsection experiments with Si have also been done. Si samples in which dislocations had previously been introduced as described in [1, 98] have been prepared for the sample holder of the 300kV Hitachi TEM, modified for ultra-high vacuum (UHV) experiments, which required the samples to be 2 × 4mm and 100μm thick. The samples have

first been cut to 2 × 4mm big rectangles and then parallel polished until the right thickness was reached.

The samples were then laid on a glass slide and coated with bees wax, which covered the whole sample and fixed it to the glass slide at the same time. While the wax was still soft, a small amount of it was removed, exposing the center of the sample. The glass slides with the samples glued to them were then held into a solution of Hydrofluoric Acid (HF), Nitric Acid (HNO_3), and Acetic Acid (CO_3COOH) in the volume ratio (3:5:3) for approximately 5 minutes, until the exposed region of the sample became transparent red when shining light through it. After rinsing the samples in deionized water they were exposed to a slower etch, comprised of a (3:5:6) solution of Hydrofluoric Acid, Nitric Acid, and Acetic Acid for less than a minute, until very small holes appeared. The samples were then rinsed in deionized water again. The wax was removed using a bath of Trichloroethylene which was kept at a high enough temperature for it to slowly boil. Residual wax was then removed using the following procedure:

1. Rinse sample in deionized water
2. Bathe sample in Acetone
3. Rinse sample in deionized water
4. Bathe sample in Propanol
5. Rinse sample in deionized water
6. Dry sample
7. Bathe sample in a 1:10 solution of Nitric and Sulfuric Acid (H_2SO_4).
8. Rinse sample in deionized water

This method proved to be much faster than mechanical dimpling and ion-beam milling, and it also produces very smooth surfaces. After rinsing the samples again in deionized water the oxide layer was removed by a quick bath in a 5% HF solution.

Since the work of Chabal et al. [172], producing chemically inert Hydrogen terminated Si surfaces by treatment with an aqueous solution of ammonium fluoride has become a current topic in semiconductor physics and chemistry. In our experiments a hydrogen terminated surface was created by the following the procedure proposed by Fukidome et al. [173]:

1. Bathe sample in a 5% Hydroflouric acid solution for 5 minutes
2. Rinse sample in deionized water
3. Bathe sample in a 40% solution of Ammonium Flouride (NH_4F) and deionized water, to which 0.05 mol/l Ammonium Sulfite (NH_4SO_3) have been added, for 6 minutes.

After this procedure the samples were visibly hydrophobic, indicating a successful hydrogen termination. All glassware used in the sample preparation was cleaned using the Huang cleaning procedure which consists of a bath in a solution of deionized water, Hydrogen Peroxide, and Ammonium.

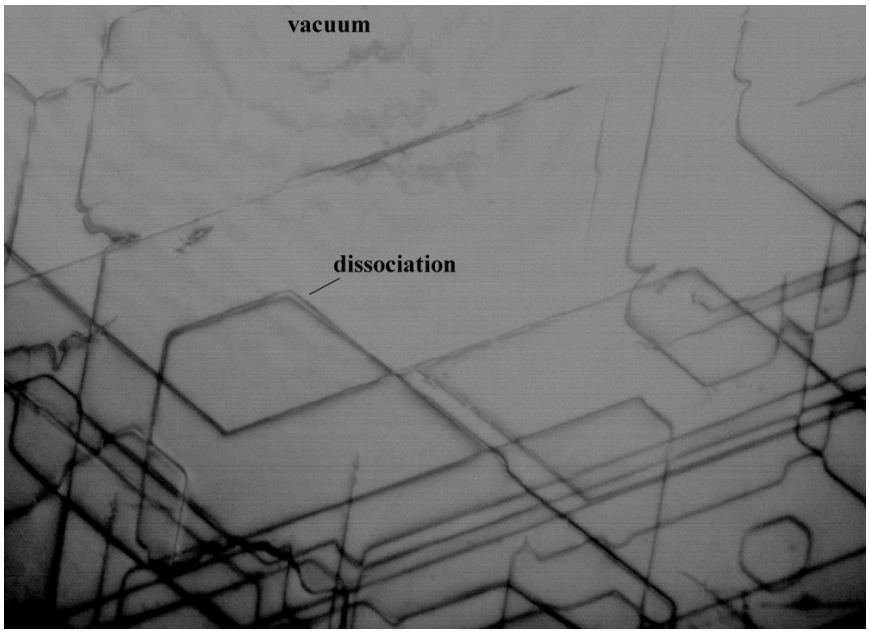

Figure 43. BF image of dislocations in Si(111). The sample has been oriented perpendicular to its main slip plane. The dissociation of dislocations into partials can be seen very clearly. The very weak contrast features concentric around the small area of vacuum at the top of the image are Pendellösung fringes. This image has been recorded using the UHV Hitachi 300kV TEM on film at IBM Thomas Watson research center in Yorktown Heights.

Figure 43 shows a low magnification TEM image of a Si sample prepared in the way described above. The dislocation density is high enough to find dislocations in material that is thin enough for HRTEM imaging. However, the sample does not show terrace shaped thickness fringes yet. The surface does not have large enough areas of constant thickness. Different tests have been done to see how heating the sample to temperatures up to 1000 °C will affect the dislocation density, especially in thin areas of the sample. This was done because dislocations in Si become mobile at temperatures above 450°C [1, 98] and might relax and disappear from the thin areas of the specimen. Our experiments showed that at least some of the dislocations were still present after O_2-etching at 800°C for several minutes, which was done by slowly leaking oxygen into the TEM chamber while heating the sample by passing a current of 350mA through the sample. This result indicates that it is in principle possible to obtain atomically flat surfaces on these samples without removing all the dislocations.

While it was possible to obtain low-resolution images with apparently atomically flat surfaces, a good quality image of the stacking fault boundary could not be recorded, although resolution tests using a graphite test sample were able to prove that this instrument was able to resolve 3.4Å lattice fringes.

Figures 37 and 44 show earlier results obtained by F. Ross using a similar sample preparation method, but undeformed Si. These samples also have not been immersed in a 40% Ammonium Fluoride solution. Instead they have been heated to about 1000°C in ultra-high vacuum, which is known to produce very flat surfaces. One of the undeformed samples examined by F. Ross contained a dislocation that had not been purposely introduced. Figure 44 shows part of the BF-TEM image which has been recorded at a magnification too low for resolving single partial dislocation kinks. However this image very nicely demonstrates the image contrast produced by forbidden reflections due to different surface terminations in the presence of a stacking fault and in perfect material. Figure 45 plots the intensity level for every numbered segment of the dislocation in figure 44. The image formation process will now be discussed in more detail.

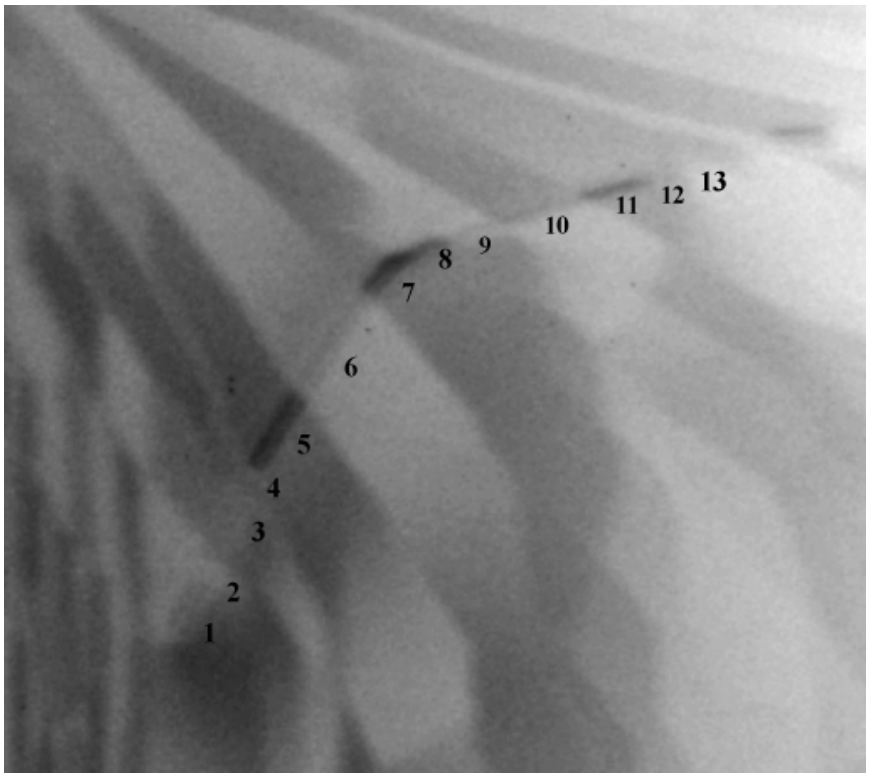

Figure 44. BF TEM image of dislocation running parallel to the Si (111) atomically flat surface, enlarged from figure 37.

We will now consider the mechanism that produces the image contrast seen in figures

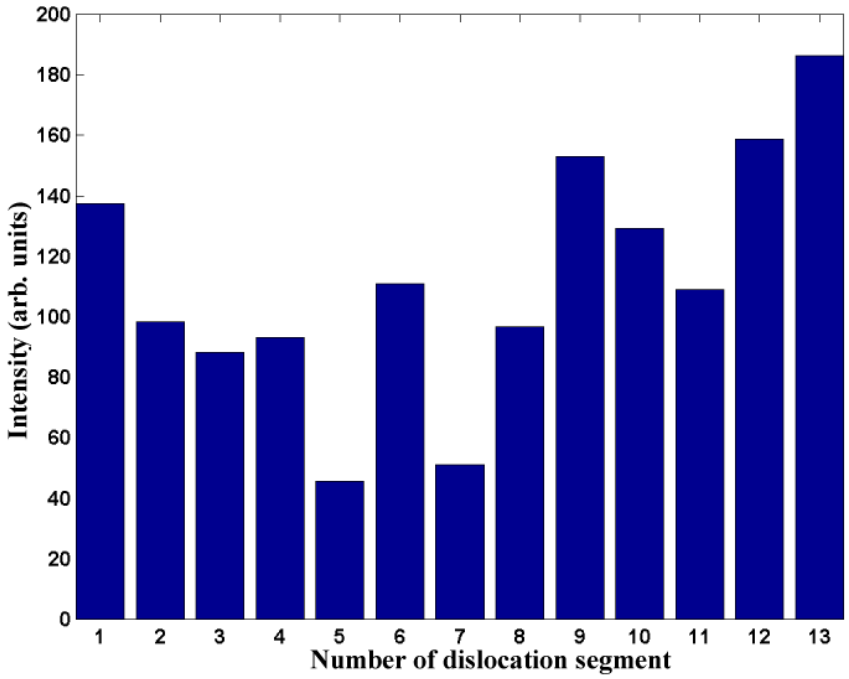

Figure 45. Average image intensity of dislocation segments for every numbered segment in figure 44.

44 and 37. In section 2 of this chapter the origin of forbidden reflections has been discussed for the general case of the zinc-blende structure. For the diamond cubic structure the 2 atoms A and B are the same, which means that $f_A(\theta) = f_B(\theta)$. In the hexagonal notation the third index l' is zero for all six forbidden reflections. The sum $h' + k'$ =-2, -1, 1, or 2. Using the layer numbering scheme from figure 27, the structure factor for the 6 layers are:

$$\begin{aligned}
F_{h',k',l'}(A) = F_{h',k',l'}(a) &= f(\theta_{h',k',0}) = f \\
F_{h',k',l'}(B) = F_{h',k',l'}(b) &= f \exp\left(\frac{2\pi i}{3}[h'+k']\right) \\
&= \begin{cases} \exp\left(\frac{2\pi i}{3}\right) & \text{for } [h',k'] = [01],[10],[\bar{1}\bar{1}] \\ \exp\left(-\frac{2\pi i}{3}\right) & \text{for } [h',k'] = [0\bar{1}],[\bar{1}0],[11] \end{cases} \\
F_{h',k',l'}(C) = F_{h',k'}(c) &= f \exp\left(\frac{4\pi i}{3}[h'+k']\right) \\
&= \begin{cases} \exp\left(-\frac{2\pi i}{3}\right) & \text{for } [h',k'] = [01],[10],[\bar{1}\bar{1}] \\ \exp\left(\frac{2\pi i}{3}\right) & \text{for } [h',k'] = [0\bar{1}],[\bar{1}0],[11] \end{cases}
\end{aligned}$$

There are two groups of 3 reflections which always have the same structure factor. The reflections $[h',k'] = [01],[10],[\bar{1}\bar{1}]$ correspond to the projection of $[h,k,l] = [\bar{1}11],[1\bar{1}1],[11\bar{1}]$ respectively, which are those reciprocal lattice points above the zero-order Laue plane. $[h',k'] = [0\bar{1}],[\bar{1}0],[11]$ correspond to the projection of $[h,k,l] = [1\bar{1}\bar{1}],[\bar{1}1\bar{1}],[\bar{1}\bar{1}1]$, which are the points below the zero-order Laue plane (see figure 26 for a graphical representation).

Assuming phase and amplitude of the forbidden reflections are exactly zero for a crystal of an integer number of unit cells thickness, and the validity of the kinematic approximation, i.e. a central beam that is much stronger than any other beam (valid for thin crystals, but see figure 46 for a calculation using fully dynamic scattering theory), one can write down the image intensity for a crystal with an additional A layer, formed by using an aperture that lets only the central beam and the forbidden reflections pass, as

$$\begin{aligned}
I^{Aa}(\vec{r}) &= \left| \int \tilde{\Psi}(\vec{q}) \exp(2\pi i \vec{q}\cdot\vec{r}) d^2\vec{q} \right|^2 \\
&\propto \left| 1 + f\left[\exp\left(2\pi i \vec{q}_{[22\bar{4}]/3}\cdot\vec{r}\right) + \exp\left(2\pi i \vec{q}_{[2\bar{4}2]/3}\cdot\vec{r}\right) + \exp\left(2\pi i \vec{q}_{[\bar{4}22]/3}\cdot\vec{r}\right) + \right.\right. \\
&\quad \left.\left. \exp\left(2i\pi \vec{q}_{[4\bar{2}\bar{2}]/3}\cdot\vec{r}\right) + \exp\left(2i\pi \vec{q}_{[\bar{2}\bar{2}4]/3}\cdot\vec{r}\right) + \exp\left(2i\pi \vec{q}_{[\bar{2}4\bar{2}]/3}\cdot\vec{r}\right) \right] \right|^2
\end{aligned}$$

Using the 2-dimensional $\vec{q}$-vector defined in figure 26 the position dependent phase factor can be determined for both groups of forbidden reflections as

$$\begin{aligned}
\Phi^{(1)}(x,y) &= \exp\left(\frac{2\pi i}{a}\left[\sqrt{2}x + \sqrt{\frac{2}{3}}y\right]\right) + \exp\left(\frac{2\pi i}{a}\left[-\sqrt{2}x + \sqrt{\frac{2}{3}}y\right]\right) + \\
&\quad \exp\left(\frac{2\pi i}{a}\left[-\sqrt{\frac{8}{3}}y\right]\right) \qquad (4.11)
\end{aligned}$$

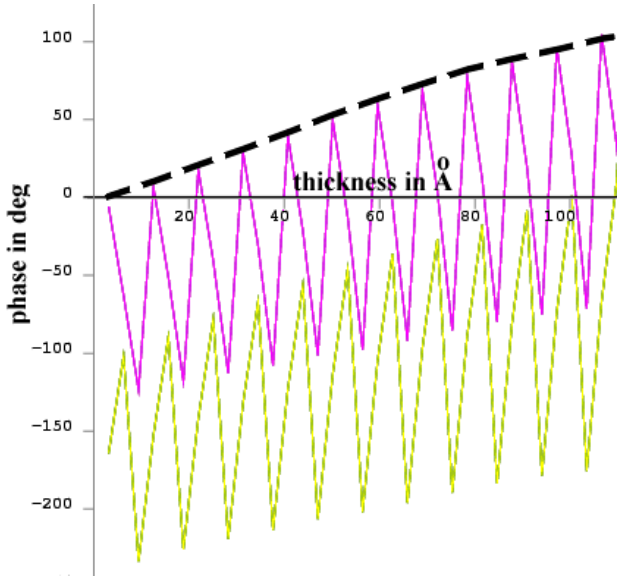

Figure 46. Pendellösung plot of the phase difference of the two types of forbidden reflections with the central beam in an unfaulted Si crystal with the 200kV electron beam normal to the (111) plane. The dashed line indicates the phase shift in addition to the kinematically derived phase oscillations due to dynamic effects. The Multislice calculation has been done using the HRTEM module of Cerius2.

$$\begin{aligned}\Phi^{(-1)}(x,y) &= \exp\left(\frac{2\pi i}{a}\left[\sqrt{\frac{8}{3}}y\right]\right) + \exp\left(\frac{2\pi i}{a}\left[-\sqrt{2}x - \sqrt{\frac{2}{3}}y\right]\right) + \\ &\quad \exp\left(\frac{2\pi i}{a}\left[\sqrt{2}x - \sqrt{\frac{2}{3}}y\right]\right) \\ &= \Phi^{(1)}(x,y)^* \end{aligned} \tag{4.12}$$

where a is the lattice constant for Si. $\Phi^{(1)}(x,y)$ and $\Phi^{(-1)}(x,y)$ are the phase factor for those reflections whose reciprocal lattice points are above and below the zero-order Laue plane respectively, so that

$$\begin{aligned} I^A(x,y) &\propto \left|1 + f\left[\Phi^{(1)}(x,y) + \Phi^{(-1)}(x,y)\right]\right|^2 \\ &= \left|1 + 2f\mathrm{Re}\left(\Phi^{(1)}(x,y)\right)\right|^2 \\ &= 1 + 6f^2 + 4f\mathrm{Re}\left(\Phi^{(1)}(x,y)\right) \end{aligned} \tag{4.13}$$

$$\begin{aligned} I^{Aa}(x,y) &\propto \left|1 + 4f\mathrm{Re}\left(\Phi^{(1)}(x,y)\right)\right|^2 \\ &= 1 + 24f^2 + 16f\mathrm{Re}\left(\Phi^{(1)}(x,y)\right) \end{aligned} \tag{4.14}$$

The last term in expressions (4.13) and (4.14) oscillates around an average value of 0. The image shown in figure 44 is not taken at a high enough magnification to resolve these oscillations, only the average intensity as determined by the first two terms in both expressions can be seen. Ourmazd et al. [168] analyzed in great detail the contrast levels of dark-field images of Si (111) produced by one of the forbidden reflections. Their result indicates that the crystal can terminate with any of the 6 layers (*A, a, B, b, C,* or *c*). Comparing the equations derived by Ourmazd with expressions (4.13) and (4.14) it becomes obvious that the bright-field intensity is just

$$I^{BF} = 1 + 6I^{DF} \tag{4.15}$$

where I^{BF} is the bright-field intensity observed in this work and I^{DF} is the dark-field intensity reported by Ourmazd et al..

In general the amplitude of any of the forbidden reflections can be approximated kinematically by simply adding the forbidden reflections structure factors for the given stacking sequence. Figure 47 gives a graphic representation of such sums. Comparing figures 47a and 47b, one can see that the kinematic forbidden reflection scattering amplitude for a crystal containing a stacking fault can be twice as high, as the maximum scattering amplitude for an unfaulted crystal. However, multiple scattering effects make this kinematic approximation only a rough estimate, and fully dynamical simulations have to be performed, in order to get an accurate estimate of the amplitude of forbidden reflections.

Using (4.15) and the results obtained by Ourmazd et al. we can conclude that the image due to surface steps in the absence of a stacking fault can take the intensity levels 1, $1+6f^2$, $1+18f^2$, and $1+24f^2$. If the atomic layers only appear in pairs, then only the image

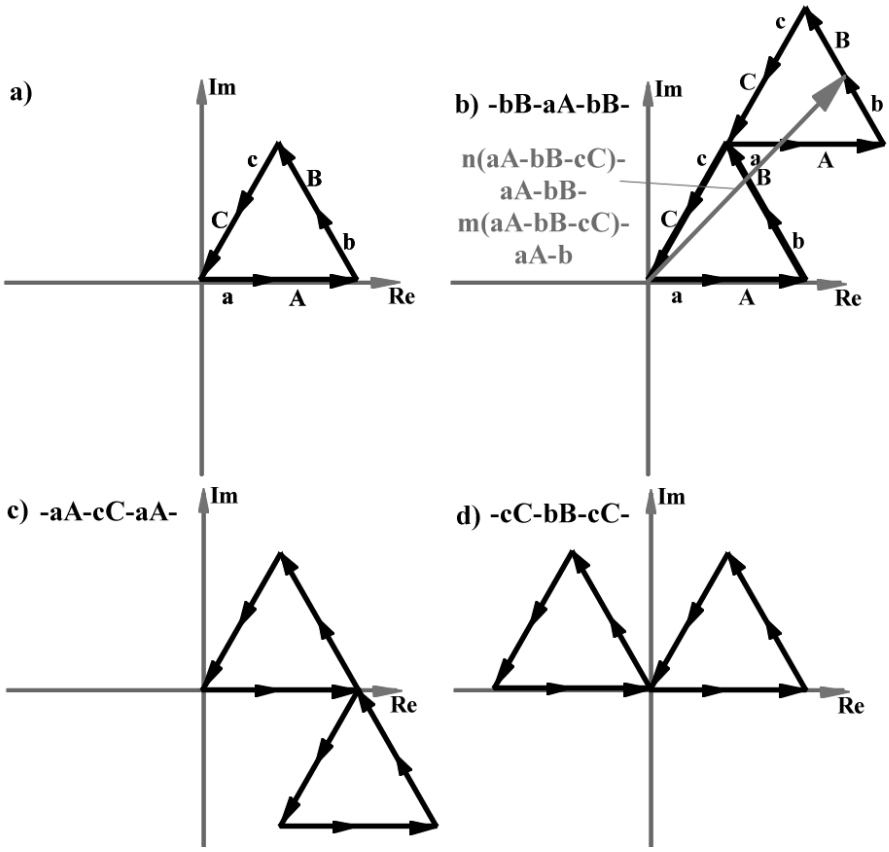

Figure 47. Argand diagrams illustrating kinematic scattering factors for different stacking sequences: a) Unfaulted crystal. After a complete set of layers (aAbBcC) the forbidden reflection scattering factor is always zero. b)-d) forbidden relection scattering factor for stacking faults at different depth in the stacking sequence, assuming that the first double layer is defined as aA. The gray arrow in b) shows the scattering factor for the stacking sequence $n(aA-bB-cC)-aA-bB-aA-m(bB-cC-aA)-b$, where n, m are arbitrary integers. Forbidden reflections produced by stacking sequence c) have always the same strength as those due to additional surface layers only.

intensities 1 and $1+24f^2$ are observable. Although the distribution of intensity levels shown in figure 38 for an unfaulted region of the crystal does not only show 4 distinct intensity levels, it clearly demonstrates that there must definitely be more than just 2 intensity levels, which is in agreement with the findings by Ourmazd et al.. The fact that there are many more than just 4 levels at any given position along the horizontal axis is due to the fact that the thickness gradient is not perfectly parallel to the horizontal axis of the image, and also dynamical effects in addition to the Pendellösung oscillation of the central beam.

Figure 48 shows the intensity levels of every terrace of constant thickness as well as the different contrast levels due to the stacking fault. The stacking fault seems to be rather close to one of the surfaces, because boundaries in the terraced surface structure coincide with the line of the dislocation. It is therefore very difficult to exactly determine the origin of the contrast levels. Further experiments are necessary investigating the stacking fault contrast levels for the case where the course of surface steps is clearly independent of the stacking fault.

5. Conclusion and Future Work

A new mode of operation for a spherically aberration corrected TEM has been proposed, which may find other applications than that of producing forbidden reflections bright-field images of stacking faults in Si at higher resolution, as has been described here. Being able to adjust the spherical aberration of the objective lens opens the door to many new experiments, not possible before. The increase in resolution of the partial dislocation kink images shows differences between different kink structures in the simulation, which might finally be detected using HRTEM imaging. This might also open the possibility of finding new kink structures by comparing experimental images with image simulations like the ones shown here.

Also, the importance of atomically flat surfaces for imaging kinks along partial dislocations in zinc-blende and diamond-cubic materials has been demonstrated by multislice image simulations. Successful preparation of samples with atomically flat surfaces have been described. An interpretation of intensity levels in images produced by forbidden reflections has been provided as far as was possible.

The results described in this chapter show that it is possible to image dislocation kinks with better resolution and determine their density with higher accuracy than has been achieved up to now. Future work should therefore pursue further experiments for both Si and β-SiC samples. Accurate measurements of dislocation kink densities will give further insight in the principles of dislocation motion, while higher resolution images may help to confirm atomic models of single structural defects along partial dislocations in materials with fcc, diamond, or zinc-blende crystal structures.

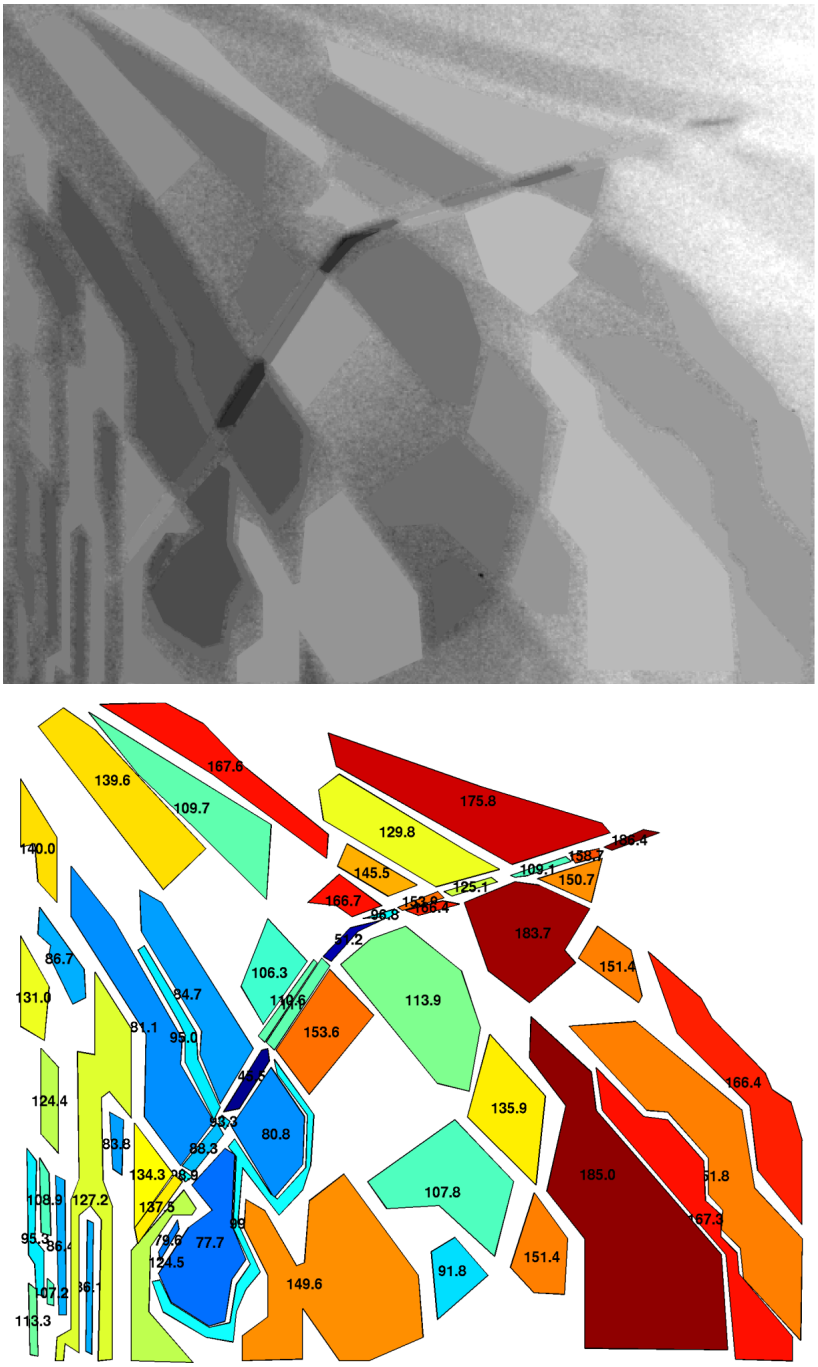

Figure 48. (color) Same image of dislocation as figure 44, but noise in the intensity levels has been removed by the procedure described in section 3.2. Top: gray scale image with the "patches" superimposed in it, bottom: "patches" of constant intensity along with their intensity value (contrast enhancing color-map).

CHAPTER 5

MICRODIFFRACTION ALONG DISLOCATION CORES

1. Introduction

Direct evidence for the atomic structures present in dislocation cores is extremely difficult to obtain; yet these structures may exert a controlling influence on the electrical and mechanical properties of semiconductors. High-resolution electron microscopy together with electronic structure calculations, as well as other experimental techniques (see chapter 3 for an overview), have until now provided the main (indirect) tools for understanding dislocation core structures. On this basis, a consensus has developed that reconstruction of dislocation cores clears the band-gap of deep states for both the 30° and the 90° partial dislocations in silicon, although some experimental results indicate otherwise [124].

First direct experimental evidence for the reconstruction of both of these cores will be presented here, and it will be shown that the periodicity of their structure along the dislocation line is twice that of the surrounding bulk material. The experimental data also shows that certain theoretically proposed structures for defects along the dislocation core cannot be present in the examined dislocations, and may therefore not exist at all.

1.1. Geometry. The Fourier coefficients

$$\tilde{V}(\vec{q}) = 2\pi \int_{-\infty}^{\infty} V(\vec{r}) \exp(2\pi i \vec{r} \cdot \vec{q}) d^3\vec{r} \tag{5.1}$$

of the periodic function $V(\vec{r})$ with periodicities of a_x in the x-direction, b_y in the y-direction and c_z in the z-direction are non-zero only for reciprocal space vectors $\vec{q} = (q_x, q_y, q_z) = (h/a_z, k/b_y, l/c_z)$, where h, k, and l are integers. This means that we can write expression (5.1) for periodic functions as a discrete Fourier series

$$\tilde{V}_{h,k,l} = 2\pi \sum_{h,k,l=-\infty}^{\infty} V(\vec{r}) \exp\left(2\pi i \left[h\frac{x}{a_x} + k\frac{y}{b_y} + l\frac{z}{c_z}\right]\right) \tag{5.2}$$

Figure 49 shows the reciprocal lattice of Silicon in the [001] projection. We will call a the real-space distance that corresponds to the spacing between the Laue zones in reciprocal space. It is equal to the vertical distance between adjacent lattice planes of the Si structure in the (110) orientation.

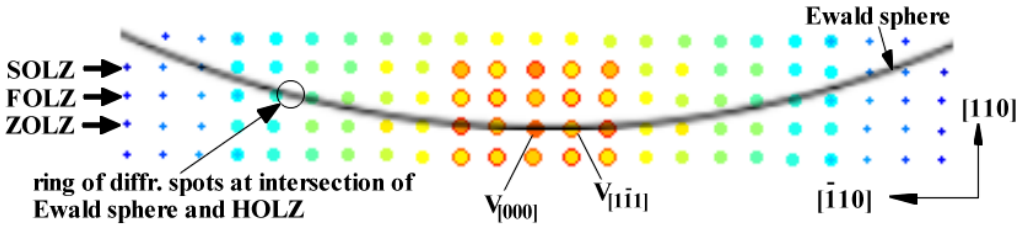

Figure 49. (color) Reciprocal lattice of Si in the [001] projection. The Ewald sphere (sphere of radius $1/\lambda$ passing through [000]) for an electron energy of about 30kV is superimposed in it. Color and size of the dots indicate the magnitude of that Fourier coefficient on a logarithmic scale. Zero order, first order, and second order Laue Zone (ZOLZ, FOLZ, SOLZ) are indicated by the arrows on the left. Reciprocal lattice points that intersect the Ewald sphere will produce diffraction spots.

The discrete Fourier Transform (5.2) can only be used in the case of a periodic function, and the values of a_x, b_y, and c_z have to correspond to the longest periodicities in those directions, which implies that for a non-periodic function the discrete Fourier Transform becomes exact in the limit $a_x, b_y, c_z \to \infty$. A 1-dimensional function $V'(\vec{r})$ which is confined to a point in the x-and y-direction but extends along the z-direction and oscillates along that axis with a periodicity of ξ can be described as

$$\begin{aligned} V'(\vec{r}) &= \delta(x,y) \sum_{n=-\infty}^{\infty} \tilde{V}_n \exp\left(2\pi i n \frac{z}{\xi}\right) \\ &\approx \frac{1}{\pi\epsilon^2} \exp\left(-(x^2+y^2)/\epsilon^2\right) \sum_{n=0}^{\infty} \tilde{V}_n \exp\left(2\pi i n \frac{z}{\xi}\right) \end{aligned} \tag{5.3}$$

approximating the delta-function with a sharply peaked Gaussian. In the limit $\epsilon \to 0$ this approximation becomes exact. The Fourier Transform of (5.3) is

$$\begin{aligned} \tilde{V}(\vec{q}) &= \sum_{n=0}^{\infty} \frac{\tilde{V}_n}{\pi\epsilon^2} 2\pi \int_{-\infty}^{\infty} \exp\left(-(x^2+y^2)/\epsilon^2 + 2\pi i \left[x q_x + y q_y + z\left(q_z + \frac{n}{\xi}\right)\right]\right) d^3\vec{r} \\ &= \exp\left(-\epsilon^2\pi^2(q_x^2+q_y^2)\right) \sum_{n=0}^{\infty} \tilde{V}_n \delta\left(q_z + \frac{n}{\xi}\right) \end{aligned} \tag{5.4}$$

Equation (5.4) is visualized in figure 50 for two different values of ϵ and a Gaussian dependence of $\tilde{V}_n$ on n.

Figure 50 shows that for a function $V(\vec{r})$ that is almost perfectly 1-dimensional the representation in reciprocal space is a set of parallel diffuse sheets in the (x,y)-plane, having a reciprocal distance of $1/xi$ between them. As the width in the x- and y-direction of the function grows, its representation in reciprocal space becomes more that of a 3-dimensional

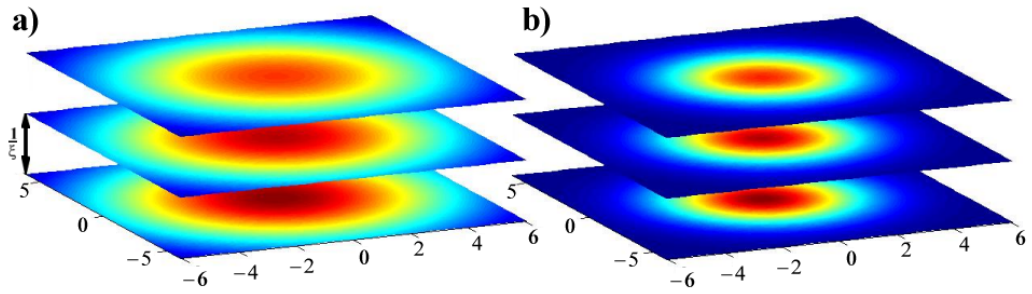

Figure 50. (color) Visualization of equation (5.4) for two different values of ϵ: a) ϵ=0.1Å, b) ϵ=0.05Å. The scale of the x- and y-axes are in $Å^{-1}$. The $\tilde{V}_n$-coefficients are assumed to be of the form $\tilde{V}_n \propto \exp(-\alpha n^2)$, but not according to any particular scale.

δ-function. Dislocation cores are such one-dimensional objects. Ideally they have a self-repeating structure, which is different from that of the surrounding bulk-material, very confined in 2 dimensions, but stretching several μm in the $3^r d$ dimension.

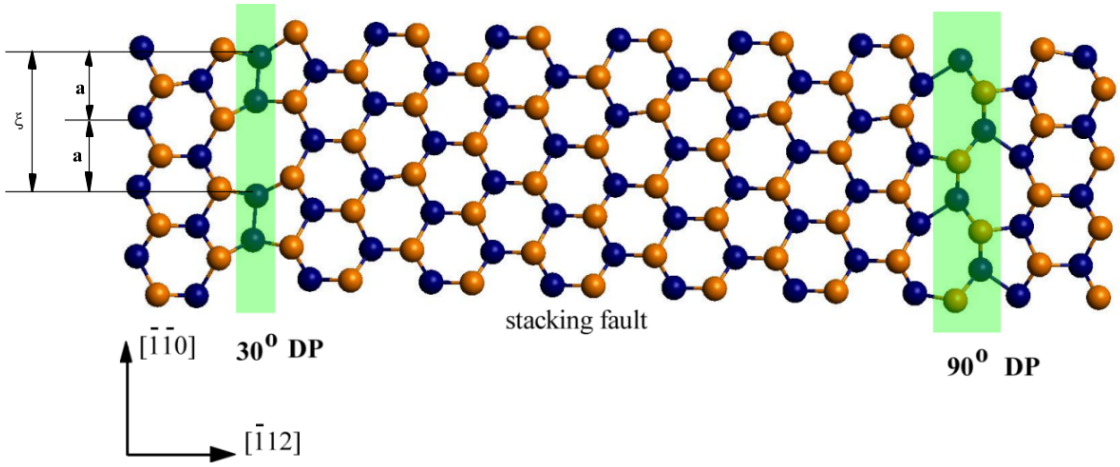

Figure 51. Single $(1\bar{1}1)$ lattice plane of a Si crystal containing a dissociated edge dislocation and the stacking fault between them. The partial dislocations, as indicated by the shaded areas are quasi 1-dimensional objects with a periodicity of $\xi = 2a$.

Figure 51 shows the model of a single layer from a faulted Silicon crystal. The dislocations at the edge of the stacking fault form one-dimensional objects with a periodicity ξ along the [110] direction of twice that of the bulk material. Since $\xi = 2a$ the potential produced by the atomic structure of the dislocation cores will form diffuse sheets in reciprocal space with a distance of $1/(2a)$ between them, which is half as much as the vertical

spacing of the layers of reciprocal lattice points of the bulk structure (see figure 49).

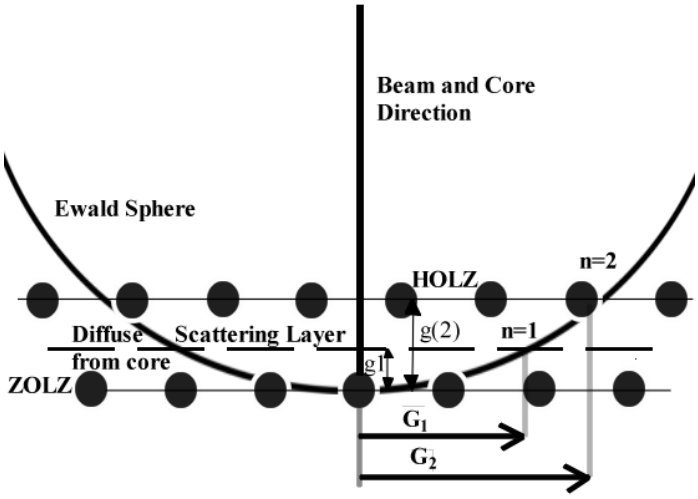

Figure 52. p

rojection with [110] layers of diffuse potential (horizontal lines) due to a dislocation core.] Reciprocal lattice of Si in the $[\bar{1}10]$ projection with [110] layers of diffuse potential (horizontal lines) due to a dislocation core. Just like in figure 49 the Ewald sphere is superimposed on it. The intersection of the Ewald sphere with the layers of diffuse scattering will produce rings in the diffraction pattern, which are in addition to the regular HOLZ-rings (labeled $n = 1, 2, \ldots$ with radii G_1, G_2, ...). This diagram is not to scale.

Using a modern field-emission STEM with cold stage and imaging energy-filter tuned to elastic scattering, it is possible to form a coherent electron probe of sub-nanometer lateral dimensions. This may be positioned over a dislocation core in a thin slab of silicon, with the beam parallel to the core. By utilizing Bragg diffraction effects along the beam direction, we may compare the periodicities in this direction in the perfect crystal with that along the core. The diffraction pattern which results from locating such a probe over a dislocation core [174] with the beam aligned along the $[\bar{1}\bar{1}0]$ core direction, will show High-Order Laue Zone (HOLZ) rings n for every plane of allowed reflections normal to the beam, not passing through the origin. As indicated in figure 52, these planes intersect the Ewald sphere at radius G_n. Lattice points in plane n satisfy $n = \vec{g} \cdot \vec{H}$ where $\vec{g}$ is any reciprocal lattice point in HOLZ plane n, and $\vec{H}$ is the zone axes, taken anti-parallel to the beam direction. The height (along $\vec{H}$) of the HOLZ is $g(n)_z \approx (\lambda/2)G_n^2$. For a beam along $[\bar{1}\bar{1}0]$ in silicon, ($\vec{H} = [110]$) the first HOLZ (FOLZ) plane containing allowed structure factors (such as $\vec{g} = [022]$) must be labeled $n = 2$, and occurs at height $g(2)_z = g(110) = 1/a$, corresponding to $a = d(110) = a_x/\sqrt{2} = 3.8$Å.

As shown in figure 52, for a probe located over the string of atoms at the core of

a partial dislocation whose period is doubled to $\xi = 2a$ (as in figure 51), a "half-order" HOLZ ring $G_1(n = 1)$ will appear. Since $g(1)_z = \frac{1}{2}g(2)_z$, the radius of the ring due to the reconstructed core is

$$G_1 = \frac{G_2}{2}\sqrt{\frac{2-\lambda/\xi}{1-\lambda/\xi}} \approx \frac{G_2}{\sqrt{2}} \tag{5.5}$$

at high energies. Figure 53 shows such a focused electron probe superimposed on the silicon structure containing a 90° partial dislocation which is double-period reconstructed, as in figure 51. Subsection 5.3 in chapter 3 explains further the structure of this partial dislocation.

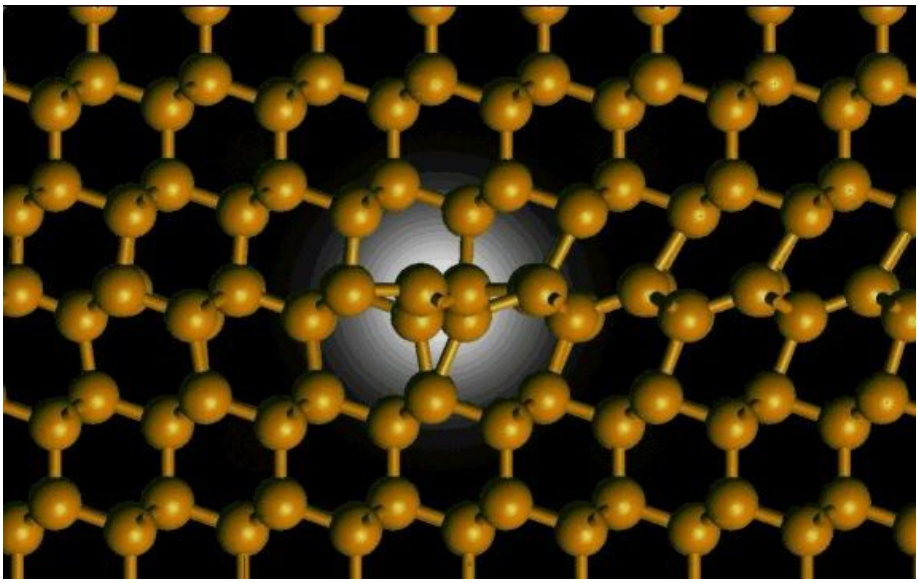

Figure 53. $(\bar{1}\bar{1}0)$ projection of the core of the double period reconstructed 90° partial dislocation with the intensity distribution of an electron beam probe focused on the dislocation core superimposed on it. The stacking fault is to the right of the core.

2. CBED Pattern and Image Simulations

Using initially the multislice program *autoslic* written by Kirkland [35] and later *S-TEM* (see chapter 2, section 5 for a description of this code), diffraction patterns and STEM images resulting from the configuration shown in figure 53 and others have been calculated. Calculations have been performed for several different microscopes, operating between 100kV and 200kV.

Atomic scattering factors are usually parameterized using Gaussians [63] or a combination of Gaussians and Lorentzians [35, 65]. Such parameterizations allow analytic integration of the atomic potential perpendicular to the plane of the potential slices used in the

multislice algorithm. Most codes therefore project the full atomic potential $\int_{-\infty}^{\infty} V(\vec{r})dz$ for all the atoms whose z-position falls into a particular slice in that slice only, which is a valid approximation, if the slices are more than 1 Å apart and are centered around the atomic layers. For structures, in which atomic layers cannot be clearly defined, and especially, if the atomic positions are not perfectly periodic because of small displacements used in the frozen phonon approximation, this approximation may lead to serious artifacts. The 30° DP partial dislocation core with a slice configuration as shown in figure 54 would not produce an extra Laue ring, if the full potential of the core atoms were projected into the slice they are located in. Using the frozen phonon approximation some of the atoms would even appear in different slices, causing serious artifacts at high scattering angles, especially, if the slices are as thin as 0.48Å as used for the calculation of the CBED patterns in figure 55.

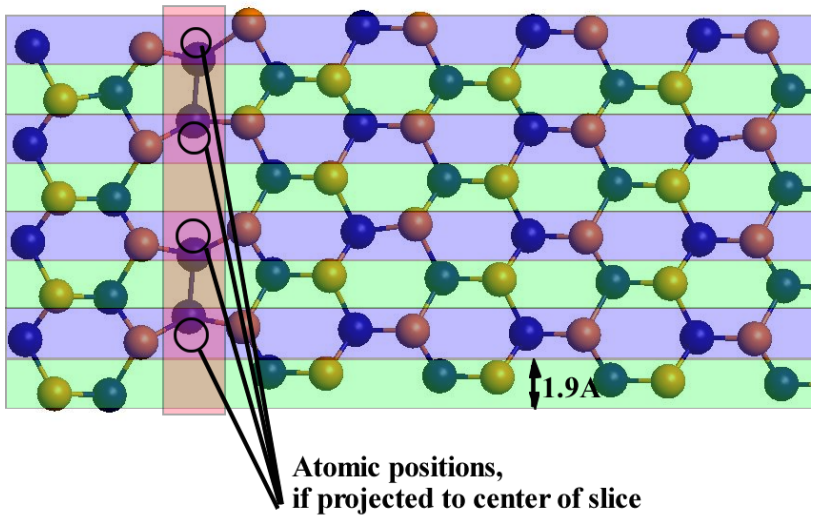

Figure 54. Potential slices of 1.9Å thickness for a multislice simulation of the 30 degree partial dislocation core. If the full range of the atomic potential is projected into a single slice, the core atoms would have exactly the same periodicity as the bulk structure and would not produce double period structure.

Figure 55 shows microdiffraction patterns, which have been simulated for an approximately 1nm small electron probe positioned along the 90° dislocation core. Two different atomic structure models have been used: the SP-reconstructed model (figure 55a) and the DP-reconstructed model (figure 55b). Scattering due to the diffuse potential layer in the ZOLZ is visible in both patterns as strong diffuse background between the Bragg disks in the center of the patterns. For the SP model these diffuse layers have the same vertical spacing as the allowed bulk-crystal reciprocal lattice points, so that rings produced by their intersection with the Ewald sphere coincide with the bulk-crystal Laue rings. The pat-

tern produced by the DP structure shows the additional Laue ring produced by a diffuse potential layer between the bulk crystal layers.

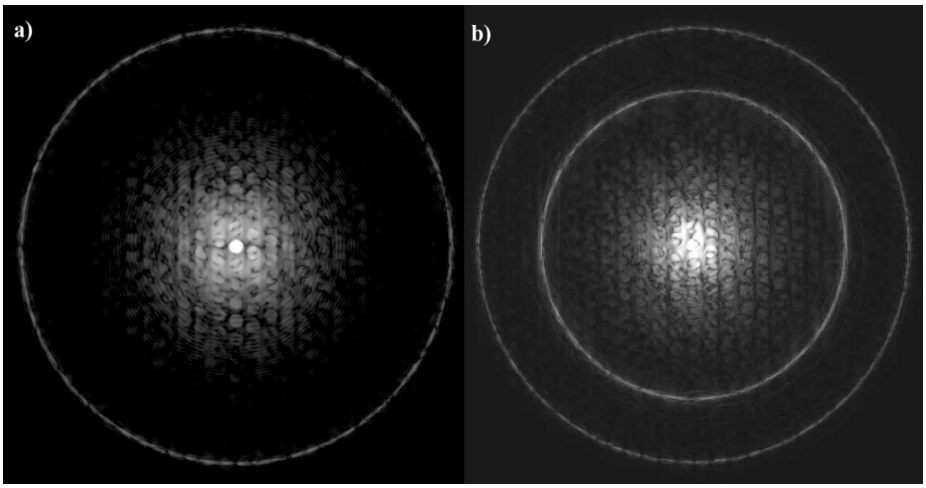

Figure 55. Simulated microdiffraction patterns for two different 46nm thick models of the 90° partial dislocation core: a) SP structure, b) DP structure. $E = 100$kV, $C_s = 1.2$mm, $df = -670$Å, probe size ≈ 1Å. Temperature effects have been included using a Debye-Waller factor of 0.44Å^2.

2.1. Effects of temperature. The inelastic mean free path for plasmon scattering, is about 1000Å. This means that most of the inelastic scattering at thicknesses lower than that is due to phonon scattering, especially at high scattering angles. Plasmon scattering also produces a shift in energy of the scattered electrons large enough to be filtered out by most energy filters. Electrons scattered by phonons do not loose very much energy, and will therefore contribute even to energy-filtered diffraction patterns.

Using the frozen phonon approximation we can calculate the TDS distribution very accurately and determine the visibility of the Laue ring produced by the DP dislocation core, i.e. the intensity of the ring compared to that of the background. Using $\overline{u^2}(T)$ given by expression (2.41) (also plotted in figure 11) and the program *S-TEM* the intensity of the Laue ring as well as that of the phonon background for the same scattering angle has been calculated for the 90° DP partial dislocation core. The background has been estimated by linear interpolation of the scattered intensity just inside and outside the Laue ring. Figure 56 shows the averaged intensity of the Laue ring and the interpolated background as a function of thickness and temperature.

2.2. Effect of core length, phase switching defects, and beam coherence. As already described in section 3 of chapter 2 there are several ways to calculate the scattering amplitude of multiply scattered electrons. If we expand the Bloch wave method in

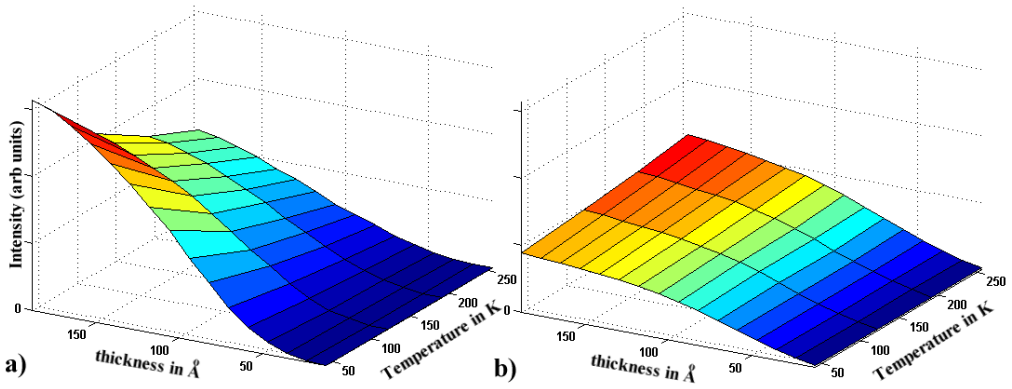

Figure 56. (color) Averaged intensity of a) the extra Laue ring and b) the phonon background as function of temperature and sample thickness. The phonon background has been interpolated from the intensity just inside and outside the Laue ring. This background has then been subtracted from the total intensity in the ring to give the elastic signal only, which has been plotted in a). The vertical scale is the same for both figures, however the color scheme has been adjusted for maximum contrast.

scattering orders, using the result derived in appendix A we get for the amplitude of any scattered beam $\vec{g}_n$ in the zone-axis case

$$\tilde{\Psi}_{g_n} = \left| \sum_{q=1}^{\infty} \sum_{l_1=0}^{N} \sum_{l_2=0}^{N} \cdots \sum_{l_{q-1}=0}^{N} \underbrace{U_{\vec{g}_n - \vec{g}_{l_1}} U_{\vec{g}_{l_1} - \vec{g}_{l_2}} \cdots U_{\vec{g}_{l_{q-1}}}}_{q} C^q_{n,l_1,\ldots l_{q-1},0} \right|^2 , \tag{5.6}$$

where the summation is over all the N reciprocal lattice vectors with non-vanishing scattering strength, and the coefficients $C^q_{n,l_1,\ldots l_{q-1},0}$ are given by expression (A.19), but not really relevant in the following non-quantitative discussion.

The electron structure factors $U_{\vec{g}}$ are defined as

$$U_{\vec{g}} = \frac{2m_0|e|}{h^2} \sum_j f_j(\theta_g) \exp(2\pi i \vec{g} \cdot \vec{r}_j) \tag{5.7}$$

where $\sin(\theta_g/2) = |\vec{g}|\lambda$ and $\vec{r}_j$ are the atomic positions. The summation has to be done over all the atoms illuminated by the electron beam. Any multiple scattering process which ultimately scatters into the $n = 1$ Laue ring, must involve scattering by reciprocal lattice vectors with a total z-component of $\sum g_z = 1/2a$, i.e. the product of U_g's representing such scattering processes must have the following form:

$$U_{\vec{g}_{l_1} - \vec{g}_{l_2}} U_{\vec{g}_{l_2} - \vec{g}_{l_3}} \cdots U_{\vec{g}_{l_q}} = \sum_j \exp\left(\frac{2\pi i z_j}{2a}\right) F(g^{x,y}_{l_1}, \ldots, g^{x,y}_{l_q}, x_j, y_j)$$

We will therefore only treat those multiple scattering events that contain only a single event with a reciprocal lattice vector of non-zero z-component (i.e. $g_z = 1/(2a)$) without loss of generality. The only restriction on the combination of the other scattering events is that their total scattering vector in the g_x,g_y-plane has the magnitude G_1, i.e. it intersects the Ewald sphere at the radius of the $n = 1$ HOLZ ring.

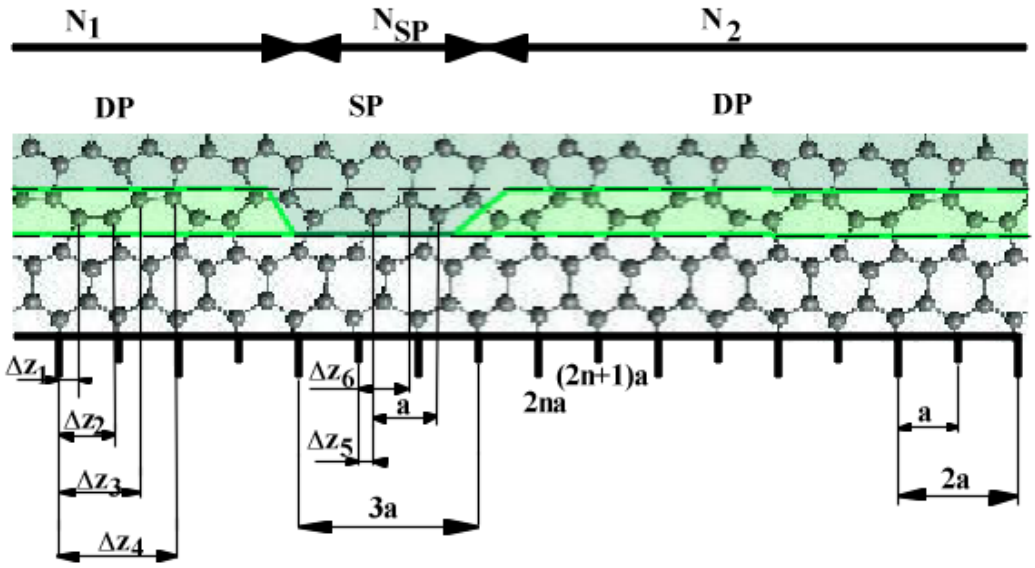

Figure 57. Model of a 90° partial dislocation segment which is partially SP and DP reconstructed. The SP segment has a length of $N_{SP} = 3$, which brings the 2 DP segments out of phase. $\Delta z_1 \ldots \Delta z_4$ define the z-component of the atomic positions within each DP segment of length $2a$. $\Delta z_5, \Delta z_6$ define the position of the atoms within each SP segment of length a. The core region illuminated by the electron beam is enclosed by the 2 dashed lines.

Figure 57 shows the model of a 90° partial dislocation core containing segments of DP and SP reconstruction. The DP segments can either start at $z = 2na$, or $z = (2n+1)a$. Again, without loss of generality we will define a coordinate system so that the dislocation core lies along the z-axis. In order to further simplify the illustration we will assume that the atoms along the core are perfectly aligned with the z-axis (as shown in figure 64a), i.e. their x,y-coordinates are zero. In any multiple scattering process which scatters into the $n = 1$ Laue ring the structure factor responsible for the z-component of this scattering process is then

$$U_g = \frac{2m_0|e|}{h^2} \sum_j f_j(\theta_g) \exp\left(\frac{2\pi i z_j}{2a}\right)$$

For the case of Silicon we can replace the atom-dependent scattering factor $f_j(\theta_g)$ with

$f_{Si}(\theta_g)$, since all atoms are the same. If there are N_1 DP segments of length $2a$ at a z-position of $z = 2na$ (n is an integer), N_2 such segments at positions $z = (2n+1)a$, and N_{SP} SP segments of length a, the structure factor for becomes

$$\begin{aligned} U_g &= \frac{2m_0|e|}{h^2} f_{Si}(\theta_g) \left[\sum_{j=0}^{N_1-1} \exp\left(\frac{2\pi i 2ja}{2a}\right) \sum_{l=1}^{4} \exp\left(\frac{2\pi i \Delta z_l}{2a}\right) + \right. \\ &\quad \sum_{j=0}^{N_2-1} \exp\left(\frac{2\pi i (2j+1)a}{2a}\right) \sum_{l=1}^{4} \exp\left(\frac{2\pi i \Delta z_l}{2a}\right) + \\ &\quad \left. \sum_{j=0}^{N_{SP}} \exp\left(\frac{2\pi i j a}{2a}\right) \sum_{l=5}^{6} \exp\left(\frac{2\pi i \Delta z_l}{2a}\right) \right] \\ &= \frac{2m_0|e|}{h^2} f_{Si}(\theta_g) \left[(N_1 - N_2) \sum_{l=1}^{4} \exp\left(\frac{2\pi i \Delta z_l}{2a}\right) + \right. \\ &\quad \left. \begin{cases} \sum_{l=5}^{6} \exp\left(\frac{2\pi i \Delta z_l}{2a}\right) & \text{if } N_{SP} \text{ odd} \\ 0 & \text{if } N_{SP} \text{ even} \end{cases} \right] \end{aligned}$$

where Δz_l are the relative z-positions of atoms within each segment as illustrated in figure 57. In order to separate 2 DP segments by a distance of $(2n+1)a$ the SP segment between them must have an odd length (see figure 57), but the total number of SP segments along the whole dislocation may still be odd. For materials involving more than one kind of atom $f_j(\theta_g)$ needs to remain inside the sum, which makes U_g dependent on the periodicity of each atomic type along the dislocation core.

If the probability for the presence of odd length SP segments is non-zero, the scattering into the $n = 1$ Laue ring will statistically cancel. An example to demonstrate this would be the following sequence: DP[4]-SP[3]-DP[4]-SP[4]-DP[6]-SP[1]-DP[6] (length of each segment in units of a given in square brackets). The sequence of these segments does not matter, but in this arrangement it becomes quite clear that the scattering from the two DP[4] segments as well as that from the two DP[6] segments is destructive. In this case any structure factor with a z-component $U(g_z = 1/[2a])$ vanishes.

Interestingly the kink structure for the DP-reconstructed 90° partial dislocation core proposed by Bennetto et al. [137] does not introduce a phase shift, nor does it introduce segments of SP structure, which means that aside from a lateral modulation of the extra Laue ring intensity due to a horizontal displacement of the dislocation core the scattering strength of the DP reconstructed 90° dislocation will not be affected by the presence of kinks, as long as the line of the dislocation is not shifted outside the illuminating electron beam by the kink. For the 30° partial dislocation, for which the DP reconstruction has found wide acceptance there have been proposed a few phase switching defects as shown in figure 64b - 64f. If a 30° partial dislocations produce an extra Laue ring, we have to conclude that such defect structures cannot be present in that dislocation core.

Knowing that in the presence of odd length SP reconstructed dislocation core segments scattering into the extra Laue ring disappears, the question arises if it would be

possible to construct experimental conditions under which the scattering into the extra Laue ring does occur. If we could shorten the length of the electron beam probe in the z-direction to less than the distance between two adjacent SP segments or other phase switching defects (PSD) we expect the intensity in the Laue ring to be proportional to the square of the length of the probe with a certain proportionality constant C. Figure 58 shows a multislice simulation for a model of the dislocation core similar to the one shown in figure 57. The SP segment joining the two DP segments has been chosen to always have a length of $2na$, which is even. As soon as the electron beam probe is much longer than the average distance between two PSDs the extra Laue ring will disappear. The proportionality constant C could be determined by varying the length of the electron beam probe.

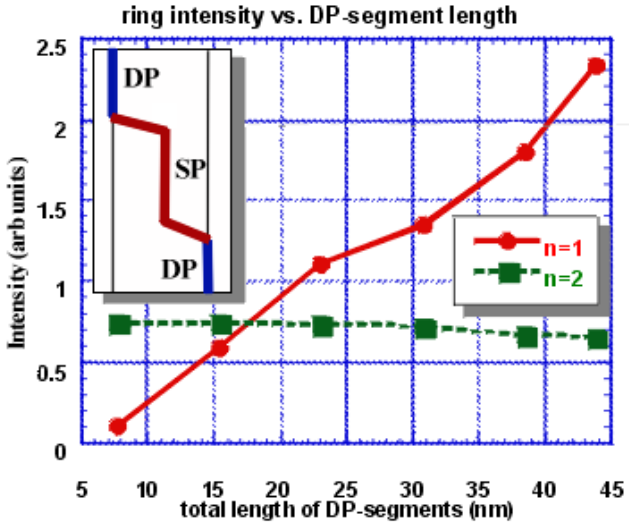

Figure 58. Calculated intensity of Laue rings averaged over the whole ring and integrated over the width of the ring (with background subtracted) as a function of the length of the DP segments in a model of total thickness of 46 nm. The remaining length of the dislocation core is SP reconstructed, connected by partial kinks, as sketched in the inset. The curves for the $n = 1$ and $n = 2$ rings are shown. This calculation was done for a 83Å $\times$ 79Å super cell using 1024 $\times$ 1024 beams with a Debye-Waller factor accounting for temperature effects (room temperature; no TDS).

There are two ways to only get coherent scattering from short segments of the sample in z-direction:

1. An increase in the convergence angle will reduce the probe length. However, an aberration corrected objective lens is necessary for angles larger than about 12mrad.

2. An increased beam energy spread will shorten the coherence length of the beam and

thereby the coherent scattering of atoms along the core. This could be accomplished by wobbling the accelerating voltage.

Both of these options will be discussed in the following paragraphs.

A focused electron beam is produced by a cone of reciprocal space vectors as shown in figure 59a. Since the length of all the k-vectors on the surface shown in figure 59a have the same length we must perform a 2-dimensional Fourier transform to obtain the 3-dimensional shape of the incident electron probe in real space

$$\Psi(\vec{r}) \quad = \quad \int_0^{\alpha} d\theta \int_0^{2\pi} d\phi \exp\left(2\pi i k \left[\sin(\theta) cos(\phi) x + \sin(\theta)\sin(\phi) y + \cos(\theta) z\right]\right) \qquad (5.8)$$

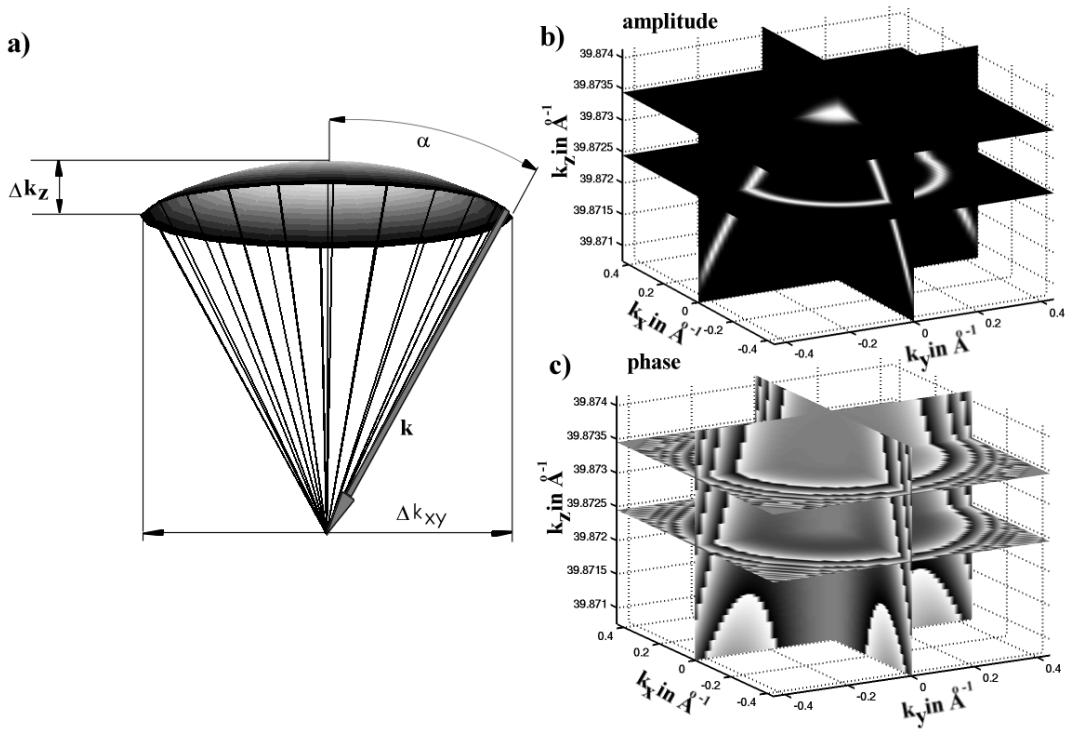

Figure 59. Reciprocal space diagram of convergent beam probe. a) An electron beam that is focused to a small spot laterally can be represented by the sum of a range of plane waves in reciprocal space which span a certain angular range with maximum convergence angle α. b) Amplitude and c) phase of a possible focused electron beam in the CM200 TEM in reciprocal space (E = 200kV, α = 10mrad, C_s = 1.2mm, $df = -700$Å, C_c = 1.0mm, and ΔE = 1eV).

Using Scherzer conditions the CBED/STEM probe is formed using an objective aperture that produces an illuminating cone of half-angle $\alpha = 1.5(\lambda/C_s)^{1/4}$ (see figure 2 for the geometry), which is about 10mrad for the CM200 200kV ($\lambda = 0.0251$Å) TEM used

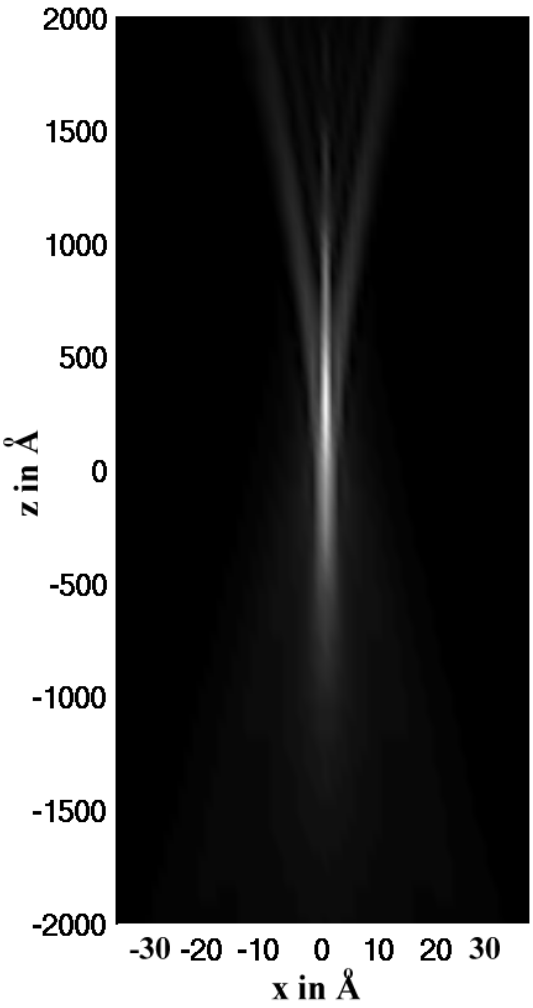

Figure 60. Gray-scale plot of amplitude of wave packet produced by the focused electron beam described in figure 59. The electron beam direction is along $\bar{z}$, i.e. down-wards. Note that the axes are not on the same scale. The wave packet has the typical cigar shape. The full width at half maximum of of a slice along the z-direction at $x = 0$ is 1167Å.

for most of the experiments and simulations in this chapter. This means a range in k_z of $\Delta k_z = 4\sin(\alpha/2)^2/\lambda = 4\times10^{-3}\text{Å}^{-1}$ for the horizontal component and $\Delta k_{xy} = 2sin(\alpha)/\lambda = 0.8\text{Å}^{-1}$ for the vertical component of the incident wave vector (see figure 59a). A energy variation of $\Delta E = 1\text{eV}$ corresponds to a change of $\Delta k 1.1 \times 10^{-4}\text{Å}^{-1}$, which is about 1 order of magnitude less than the range in k_z due to the curvature of the illuminating cone of k-vectors.

A rough estimate of the size of the focused electron beam probe can be obtained by replacing the true shape of the wave packet in reciprocal space with a 3-dimensional Gaussian with full width half maxima (FWHM) of about 2/3 the size of the box defined by Δk_z and Δk_{xy} in figure 59a). Using the uncertainty principle that $\Delta k_z \Delta z = \pi$, we obtain a probe length in z-direction of about 1180Å. This estimate does not take into account the true reciprocal space shape (i.e. the sharp aperture boundary), nor effects due to finite values of C_s, df, C_c, and ΔE. Performing the 3-dimensional Fourier transform of the true reciprocal space representation of the incident electron beam wave function whose 3-dimensional amplitude and phase shape is shown in figures 59b and 59c respectively, yields a real space probe amplitude distribution with a FWHM of 1167Å in the beam direction and a shape like the one shown in figure 60 ($E = 200\text{kV}$, $\alpha = 10\text{mrad}$, $C_s = 1.2\text{mm}$, $df = -700\text{Å}$, $C_c = 1.0\text{mm}$, and $\Delta E = 1\text{eV}$).

Figures 61 and 62 show the vertical and horizontal electron wave-function amplitude distribution as well as the FWHM length and width of the probe as a function of the convergence angle α for a possible aberration corrected microscope at 200kV. The scattering intensity is proportional to the wave function amplitude squared, which reduces the FWHM values by a factor of $1/\sqrt{2}$. Focusing the probe at different depths of the sample and changing α will reveal whether there are any PSDs within range of the focused probe at that particular depth.

The second method to reduce the length over which we probe for the presence of DP reconstruction is to shorten the coherence length of the electron beam, as already mentioned. As already stated earlier for the present application the 3-dimensional problem can be reduced to a 1-dimensional one by assuming that for a probe that is localized exactly over the dislocation core, the 3-dimensional wave packet can be replaced by a 1-dimensional wave packet of the same length in the z-direction, and the atoms of the dislocation core are perfectly aligned along the z-axis. This simplified geometry is illustrated in figure 64a.

In the single scattering approximation, and assuming a Gaussian shaped wave packet of real space FWHM ΔZ in the z-direction (i.e. a Gaussian with a spread of FWHM of $\Delta K = 4\pi\ln(2)/\Delta Z$ around a mean incident wave vector $\vec{k}_0$) the scattering intensity for scattering into the extra Laue ring (with a z-component of the scattering wave vector $[\vec{k}_0 - \vec{k}']_z = g(1)_z = 1/(2a)$, being the difference between the incident wave vector $\vec{k}$ and the scattered wave vector $\vec{k}'$) is given by

$$I(g(1)_z) = \int_{-\infty}^{\infty} dk \left| \frac{2m_0|e|}{h^2} f_j(\theta_{g(1)}) \sum_j \exp\left(2\pi i z_j [\vec{k} - \vec{k}']_z\right) \right|^2 \exp\left(-\frac{(\vec{k} - \vec{k_0})^2}{\Delta K^2/[4\ln(2)]}\right)$$

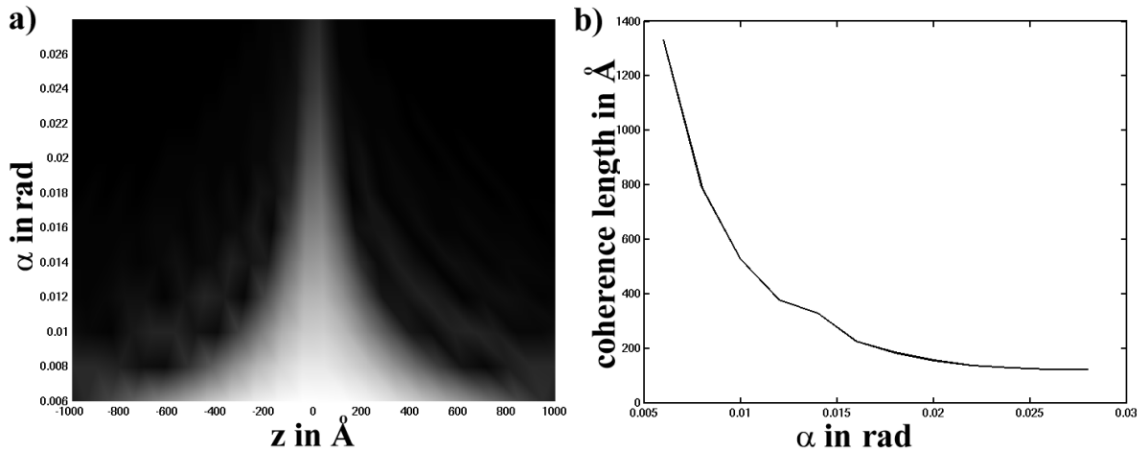

Figure 61. a) 1-dimensional amplitude distribution of wave-packet in the z-direction for different convergence angles, b) FWHM of wave packet in the z-direction, or coherence length as function of convergence angle α. An aberration corrected STEM with $C_s = 10\mu$m, $C_c = 1$mm, $E = 200$kV, $\Delta E = 1$eV, and $df = -37$Å has been assumed.

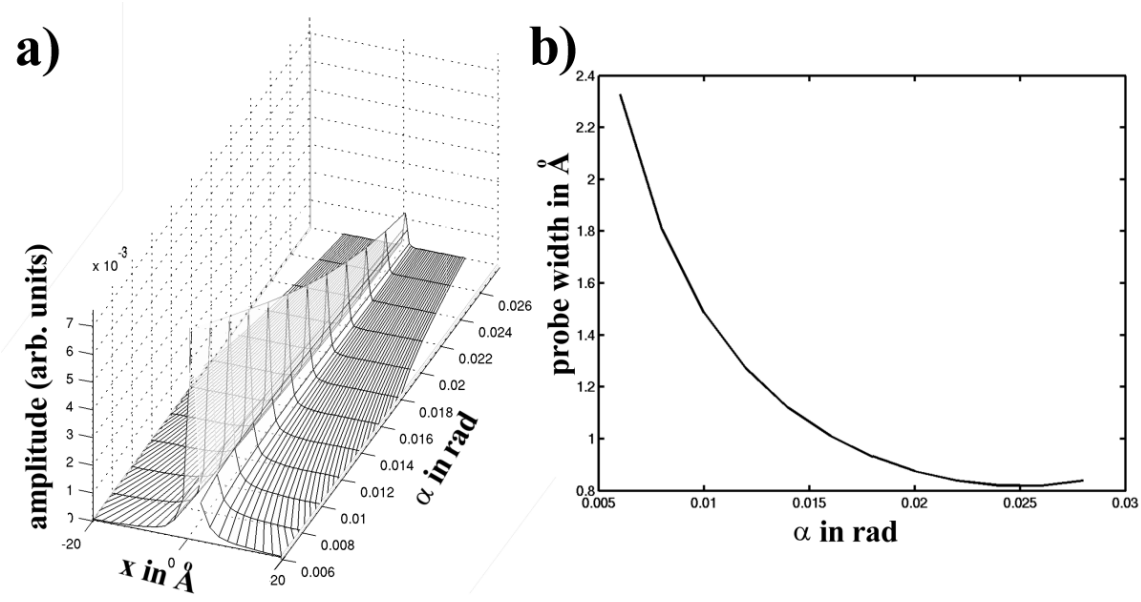

Figure 62. a) Cross-section through the electron beam amplitude distribution for different convergence angles, b) width of the probe as function of convergence angle α. An aberration corrected STEM with $C_s = 10\mu$m, $C_c = 1$mm, $E = 200$kV, $\Delta E = 1$eV, and $df = -37$Å has been assumed (same parameters as for figure 61.

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} dk \left(\frac{2m_0|e|}{h^2} f_j(\theta_{g(1)})\right)^2 \sum_j \sum_{l\neq j} \exp\left(2\pi i[z_j - z_l][\vec{k} - \vec{k'}]_z\right) \\
&\quad \exp\left(-\frac{(\vec{k} - \vec{k_0})^2}{\Delta K^2/[4\ln(2)]}\right) \\
&= \left(\frac{2m_0|e|}{h^2} f_j(\theta_{g(1)})\right)^2 \sum_j \sum_{l\neq j} \exp\left(2\pi i[z_j - z_l][\vec{k_0} - \vec{k'}]_z\right) \\
&\quad \int_{-\infty}^{\infty} dk \exp\left(2\pi i[z_j - z_l]\Delta k\right) \exp\left(-\frac{(\Delta k)^2}{\Delta K^2/[4\ln(2)]}\right) \\
&= \left(\frac{2m_0|e|}{h^2} f_j(\theta_{g(1)})\right)^2 4 \sum_j \sum_{l>j} \cos\left(2\pi i[z_j - z_l][\vec{k_0} - \vec{k'}]_z\right) \\
&\quad \sqrt{\frac{\pi}{\ln(2)}} \frac{\Delta K}{2} \exp\left(-\frac{\pi^2[z_j - z_l]^2 \Delta K^2}{4\ln(2)}\right) \\
&= \sqrt{\frac{\pi}{\ln(2)}} \frac{\Delta K}{2} \left(\frac{2m_0|e|}{h^2} f_j(\theta_{g(1)})\right)^2 4 \sum_j \sum_{l>j} \cos\left(\frac{2\pi[z_j - z_l]}{2a}\right) \\
&\quad \exp\left(-\frac{\pi^4[z_j - z_l]^2}{4\ln(2)\Delta Z^2}\right) \qquad (5.9)
\end{aligned}
$$

The shape and length of a wave packet for microscope parameters of the CM200 microscope as a function of energy spread are shown in figure 63.

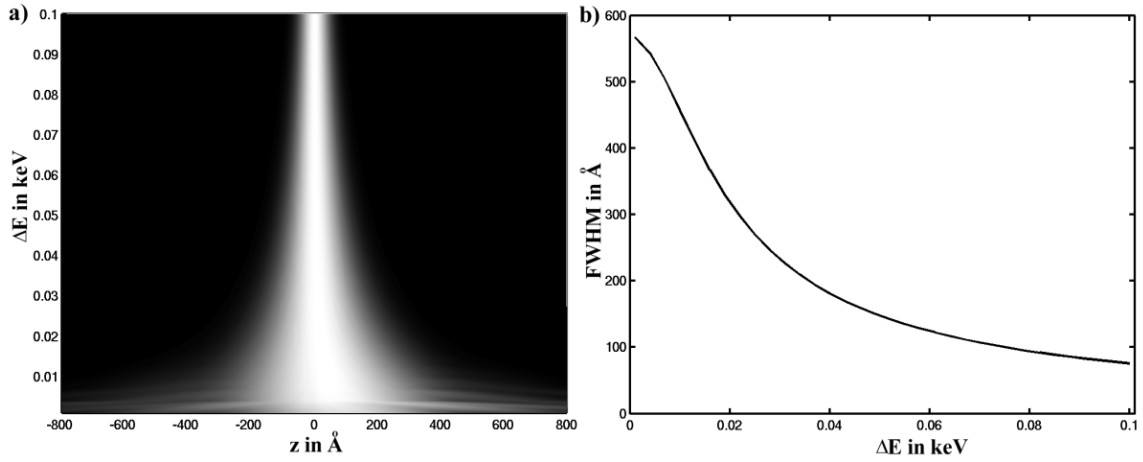

Figure 63. a) 1-dimensional amplitude distribution of wave-packet in the z-direction for different energy spreads, b) FWHM of wave packet in the z-direction, or coherence length as function of energy spread ΔE. A STEM with $C_s = 0.6$mm, $C_c = 0$mm, $E = 200$kV, and $df = -290$Å, and $\alpha = 10$mrad has been assumed.

Figure 64 illustrates one such 1-dimensional wave packet passing over a phase switch-

ing defect (PSD). Although the top and bottom DP segment are not of the same length their contribution to the extra Laue ring will mostly cancel, because the (solid) Gaussian wave packet is centered over the PSD, and the effective DP segment lengths are therefore equal, but out of phase. However, that same wave packet has also passed over the upper DP segment only (dashed Gaussian), in which all the atoms interfere constructively at scattering wave vectors on the extra Laue ring.

2.3. STEM imaging. The ideal instrument for the experiments described in this chapter is a HR-TEM/STEM which allows very exact correlation of a BF-TEM image with the STEM probe position. Due to the BF-TEM strain contrast produced by dislocations and their visibility in high-resolution lattice images the dislocation cores can be identified easily in good quality TEM images. However focusing the electron beam to a 1nm size spot will distort the image and one cannot verify the core position anymore. Most microscopes also use different gun lens settings for microdiffraction mode to produce a sufficiently small probe.

If the dislocation core can be identified in STEM images, its microdiffraction pattern can be recorded by simply stopping the beam scan when the STEM probe is over the dislocation core and record the microdiffraction pattern, assuming that the microscope is also equipped with a suitable recording device (CCD, image plate, film, etc.). This method also requires a probe stability of at least better than about 1Å per second. While for many dedicated STEM and some modern TEM/STEM instruments this is no problem, the probe drift in the CM200 used for the experiments presented here was too large to allow this.

The unique feature in the diffraction pattern of any double period reconstructed dislocation core is the presence of the extra Laue ring. By designing an annular detector which only detects scattering into angles corresponding to that of the $n = 1$ Laue ring we expect to see a higher signal in the STEM image at beam positions corresponding to regions of double period structure, like those of DP reconstructed dislocation cores. In order to quantify the expected effect STEM images have been simulated using the program *S-TEM* . Structural models of dissociated dislocations have been calculated by João Francisco Justo Filho at the Instituto de Fisica da Universidade de São Paulo, Brazil using the EDIP method developed by him and co-workers [158]. João Justo's more detailed description of how he did the calculation have been included in appendix C.

Figure 65 shows two simulated microdiffraction patterns on which portions of the annular detector have been superimposed in order to show its width and position with respect to the pattern. The detector is meant to mainly detect scattering into the extra Laue ring. This ADF-detector matches the one used for the experiments on the CM200 TEM described in section 3. STEM image simulations will be shown in section 3 in comparison with experimental results.

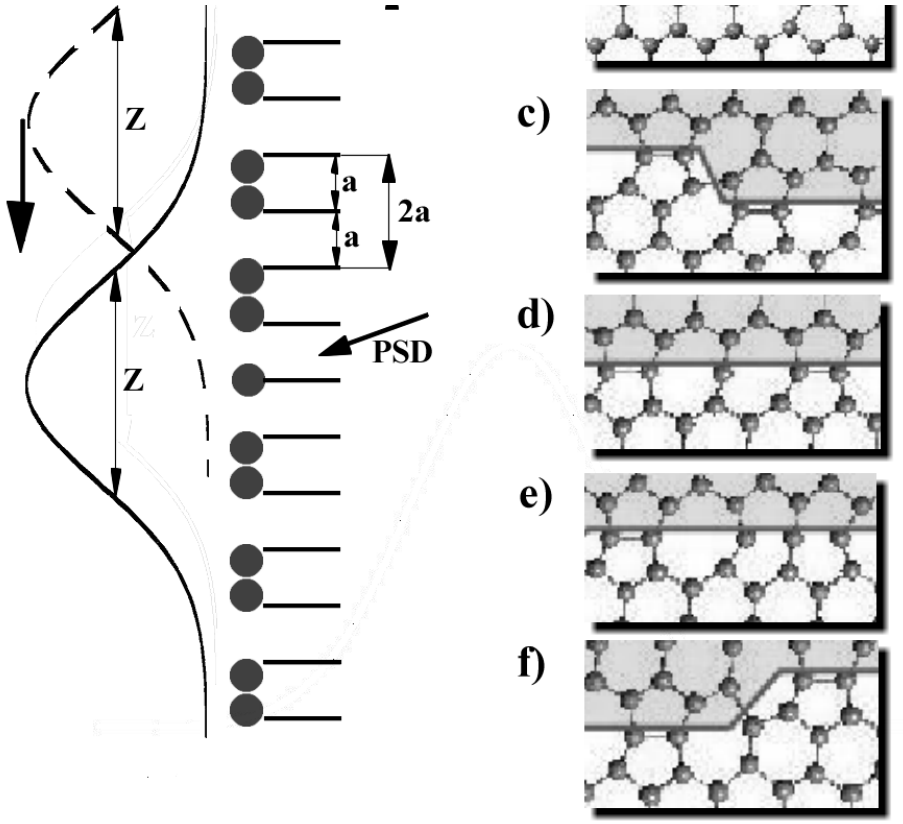

Figure 64. Illustration of a scattering of a 1-dimensional wave packet from the dislocation core. a) 1-D diagram of the wave packet as it travels along the dislocation core. The dislocation core contains a phase switching defect (PSD) at the position pointed to by the arrow. a) - e) Models of possible defect structures along a 30° partial dislocation core as published by Nunes et al. [165]. b) - c) Defect structures that do not switch the phase of the reconstruction, d) - f) phase switching defects (PSDs). The different defects were given the following names a) left kink (LK), b) alternative structure for left kink (LK') c), d) phase switching defects e) right kink (RK).

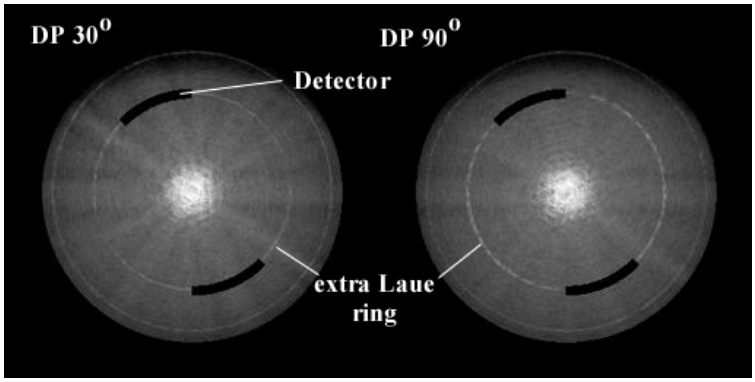

Figure 65. Simulated diffraction patterns produced by a focused 200kV electron probe positioned over the 30° and 90° DP dislocation cores. The calculation has been done for a temperature of about 20K. The simulated sample thickness was 188Å. The black arcs over the inner Laue ring indicate the width and position of the ADF detector, which covers the whole circle. The patterns are displayed on a logarithmic scale.

3. Experimental Results

Experiments have been done using two different microscopes and a third microscope has been modified to perform further experiments:

1. *Phillips CM300-FEG-Ultra-twin:* 300kV High Resolution TEM at Lawrence-Berkeley National Labatory in Berkeley, CA. Coherent illumination due to the field emission gun. This microscope is equipped with a Gatan Imaging Filter (GIF) and a 2048 × 2048 pixel CCD camera.

2. *Phillips CM200-FEG:* 200kV High Resolution TEM with a maximum resolution of 2.2Å, coherent illumination due to the Field Emission Gun. This microscope is retrofitted with a scanning unit and two annular dark-field (ADF) and one bright-field (BF) detector. The scanning electronics by Emispec drives the beam alignment coils which have not initially been designed for high resolution STEM imaging. Their large drift makes high-resolution STEM imaging a difficult task and recording the microdiffraction pattern of an area defined with Angstrom accuracy virtually impossible.

3. *MIDAS:* 100kV UHV VG-STEM with UHV sample preparation chamber. Because of its ultra-high vacuum well suited for surface science applications, but also microdiffraction due to its comparatively stable scanning electronics and low risk of contamination.

3.1. Energy Filtered Experiments at 300kV. The first experiments to test for the presence of the $n = 1$ Laue ring have been done using the One-Angström Microscope (OÅM) at the National Center for Electron Microscopy (NCEM) at the Lawrence Berkley

National Lab in Berkley, CA. The OÅM is a FEI CM300 high-resolution TEM operating at 300kV, equipped with a Gatan imaging filter (GIF), and has been used to record very high-quality data used for through-focus reconstructions [49, 50], obtaining exit face wavefunctions with a resolution down to 0.8Å [52, 175]. For these experiments the microscope has been aligned by Christian Kisielowski and Chris Nelson. Special lens settings had to be installed to allow large diffraction angles to pass through the GIF.

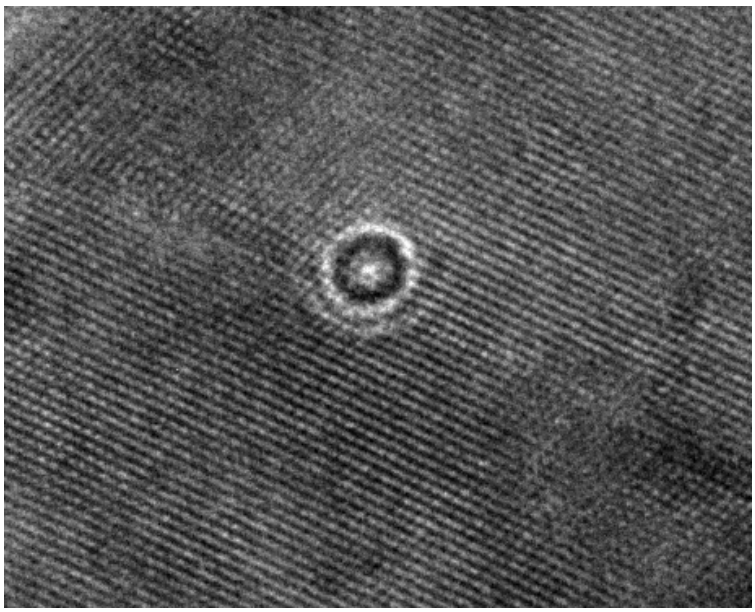

Figure 66. HRTEM image of radiation damage effects produced by a focused 300kV electron probe after short exposure time. The probe had been positioned over this thin ($<$ 100Å thick) region thought to be a dislocation core and left there for only several seconds. Recorded on 2K x 2K CCD camera through GIF on CM300 at Lawrence Berkeley National Lab.

The strategy for these experiments was to find a dislocation core in the high-resolution image, and then focus the probe on it, in order to record its microdiffraction pattern. Since the strain field induced by the dislocations warps the sample, lattice images needed to be recorded from very thin areas, where errors in the alignment are less critical.

Figures 66 and 67 show the damage produced by the electron beam in these thin areas, making experiments using a 300kV focused electron probe impossible. It was therefore decided to continue with experiments at a lower voltage, and with a cold stage, which further reduces the risk of damage, and also reduces the background signal due to TDS.

3.2. Experiments at 200kV and Low Temperature. One of the results obtained from experiments using the CM300 TEM was that an accelerating voltage of less

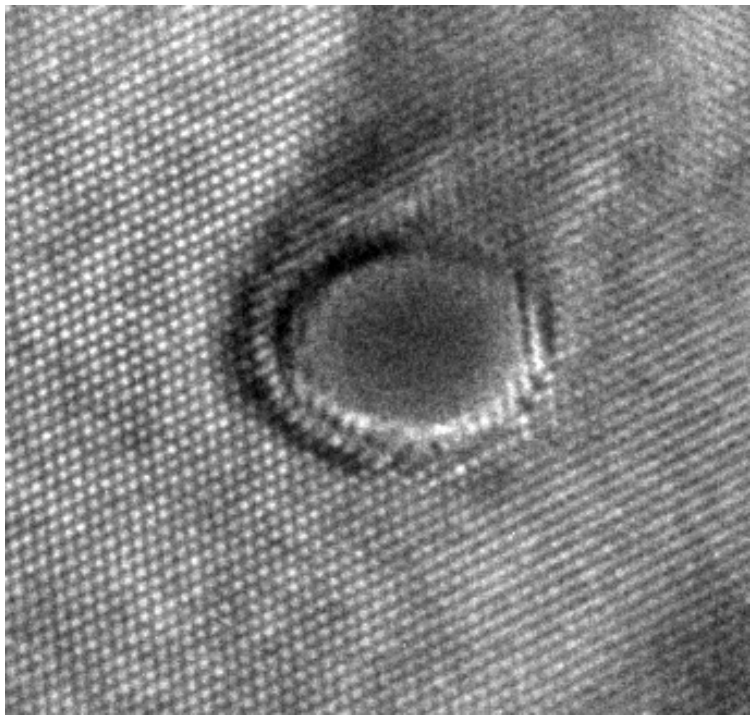

Figure 67. HRTEM image of radiation damage effects produced by a focused 300kV electron probe after short exposure time. This image was recorded under similar conditions as figure 66, only that the probe was positioned over a perfect crystal region, exposing it to the focused electron beam for about 15 sec. Recorded on 2K x 2K CCD camera through GIF on CM300 at Lawrence Berkeley National Lab.

than 300kV was required. The CM200 at ASU is equipped with an EMISPEC scan generator, a liquid Nitrogen cold stage, but no energy filter. Since plasmon scattering occurs with rather small scattering angles and for thin specimen multiple inelastic scattering events are comparatively unlikely, most of the high angle scattering is due to HOLZ effects and phonons. Plasmon scattering produces a background proportional to the elastic signal in diffraction patterns of thin specimen and does not alter the angular distribution of the elastic and phonon scattered electrons. Cooling the sample to -176°C , as is possible with the cold stage available for the CM200, drastically reduces phonon scattering, and enhances elastic scattering due to a smaller Debye-Waller factor. Calculations have shown that for thin specimen the increase in measurable signal is much higher than that due to energy filtering.

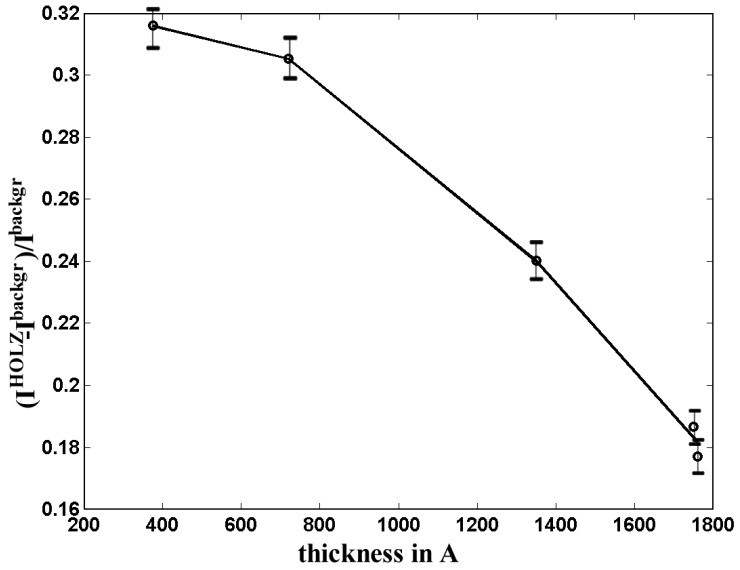

Figure 68. Plot of the ratio of experimentally measured intensities of the elastic scattering in the bulk-Si(110) HOLZ-ring and background at 95K. I^{HOLZ} is the total integrated intensity in the HOLZ ring, and I^{backgr} the background of the HOLZ ring, interpolated from the signal just inside and outside the ring.

Figure 68 shows a plot of the ratio of

$$\begin{aligned} r &= \int_0^{2\pi} d\phi \frac{I^{HOLZ}(\phi) - I^{backgr}(\phi)}{I^{backgr}(\phi)} \\ &= \frac{I^{HOLZ} - I^{backgr}}{I^{backgr}} \end{aligned}$$

where $I^{HOLZ}(\phi) - I^{backgr}(\phi)$ is the difference between the intensity in the diffraction pattern at a particular angle ϕ on the Laue-ring and the background signal, which has been

interpolated from the diffracted intensity just inside and outside the ring at the same angle. The diffraction patterns have been recorded on the CM200 microscope at 200kV with a beam convergence angle of 3.5mrad using a liquid Nitrogen cold stage, which cooled the sample to 95K. The same sample as that used for dislocation STEM imaging described in this section has been used. The thickness of the area of the sample from which the patterns have been recorded has been determined from the intensity distribution in the low order CBED disks.

The ratio r depends strongly on temperature. Using the Einstein model for calculating TDS by the frozen phonon method in multislice calculations, the root mean square (RMS) vibrational amplitude $\langle u \rangle = \sqrt{\overline{u^2}}$ needs to be known (see section 4 of chapter 2 for further details). Multislice calculations have been done for different values of $\langle u \rangle$ for the same microscope parameters used in the experiment. The ratio r in those simulated diffraction patterns has been plotted as a function of $\langle u \rangle$ in figure 69, averaged over a a range in thickness from 100Å to 600Å. The simulated patterns in this thickness range showed no trend in $r(t)$, which is in agreement with the experimental data shown in figure 68. The value of $\langle u \rangle$ for which simulated and experimental value of r agree, is approximately 0.065Å, which has been used for the multislice simulations described in the remainder of this section.

Most inelastic scattering events other than TDS occur mainly in the forward direction. One therefore expects their contribution to be roughly proportional to the total intensity, in which case it does not contribute to the value of r. This means that using non-filtered diffraction patterns for the estimation of $\langle u \rangle$ is valid, at least for thin specimen up to about 600Å. In fact, the decreasing value of r observed in the experimental data at higher thicknesses is probably due to the fact that this approximation starts to break down.

Since the probe stability in the CM200 is rather poor, it was impossible to record the actual microdiffraction pattern of a single dislocation core with a sufficiently small probe, because it would require the probe to remain at the same position with respect to the sample for several seconds at Å-accuracy. Instead of trying to record a single pattern, an ADF-detector has been designed to collect all the scattering into the angular range which would potentially be filled with the extra Laue-ring, if the probe were positioned over a DP-reconstructed core.

Since the existing ADF-detector has a very large angular range, an adapter has been designed with the help of Karl Weiss at the Center for Solid State Science at ASU and machined by the ASU machine shop, which reduces the detector to a ring that has a width of about 10% of its radius. The shape of the adapter and the way it modifies the active area of the existing ADF-detector is illustrated in figure 70. The hole in the adapter still allows a parallel detection of ADF- and EELS-signal.

Figure 71 shows an experimental microdiffraction pattern with the shape and position of the ADF-detector superimposed on it. The diffraction patterns have been aligned very carefully to position them with respect to the detector. The microscope remained in diffraction mode for the whole alignment procedure. This is important, because switching from diffraction to image mode, or TEM to nano-diffraction mode affects the current

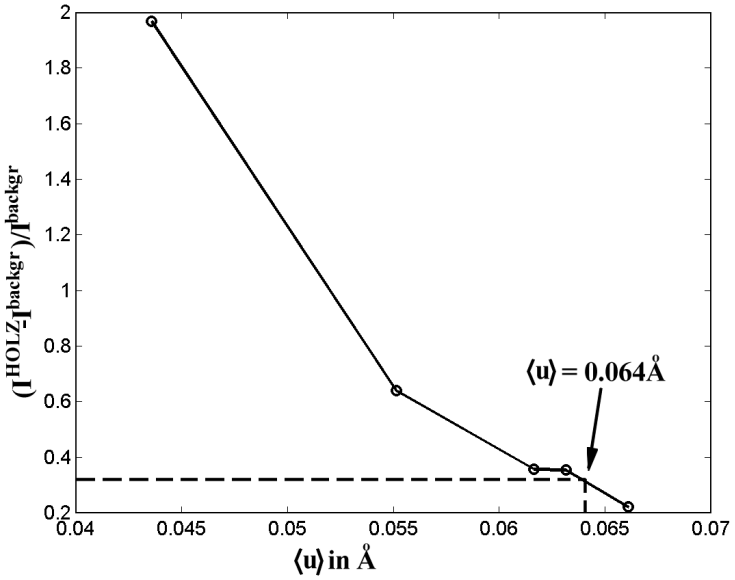

Figure 69. Plot of the ratio of simulated intensities of the elastic scattering in the bulk-Si(110) HOLZ-ring and background as in figure 68 for different values of the mean square vibrational amplitude $\langle u \rangle$. The ratio $r = (I^{HOLZ} - I^{backgr})/I^{backgr}$ has been averaged over a range in thickness from 100Å to 600Å.

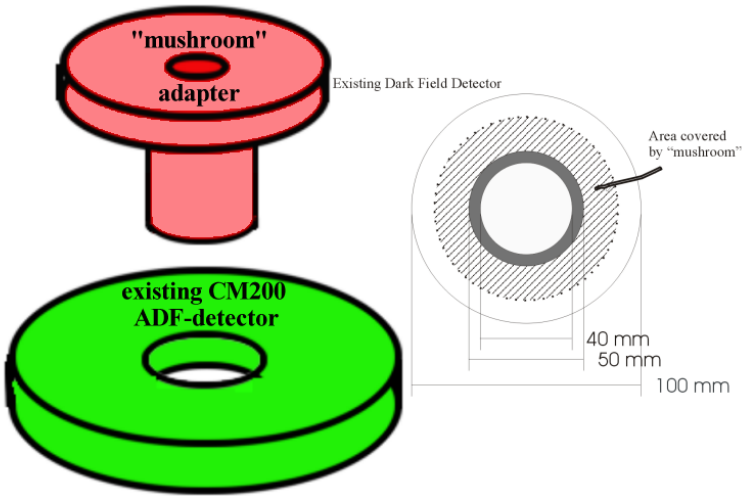

Figure 70. This "mushroom" adapter converts the large ADF-detector installed on the CM200 into a thin annular detector. The adapter is made of aluminum covering most of the ADF-detector, leaving a ring with a width of 10% its radius exposed to scattered electrons. The hole in the center of the adapter allows the direct beam to pass through on the EELS spectrometer.

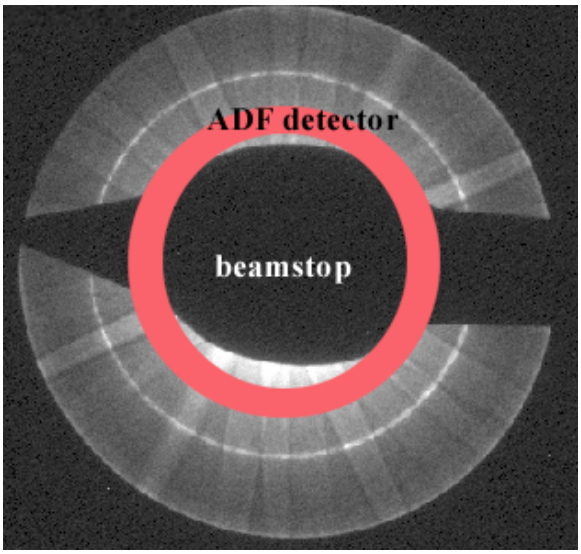

Figure 71. Si(110) microdiffraction pattern with ADF detector superimposed on it. This pattern has been recorded on a 1024 × 1024 pixel CCD camera. The sample has been cooled to -176°C .

through several lenses, which may produce errors in the alignment. The position of the ADF-detector with respect to the CCD has been determined by positioning the central beam of a diffraction pattern produced by a thin area of the specimen, or no specimen at all, over different parts of the detector while using the detected intensity as a indication whether the probe was actually focused on the detector. By replacing the detector with the CCD camera and taking a very short exposure, images with bright spots in different positions were recorded. The superposition of 5 or more such images defines a ring, which is equal to the projection of the detector annulus on the plane of the CCD.

For a fixed vertical electron beam crossover position an increase in size of the diffraction pattern in the CCD-camera plane by a factor of $\sqrt{2}$ produces an increase in diffraction pattern size in the ADF-detector plane of the same ratio. By adjusting the camera length (using a camera length of 40mm, and fine tune it by adjusting the P3 lens current in free lens control mode) to let the $n = 2$ Laue ring in a microdiffraction pattern recorded on the CCD coincide with a circle concentric with the ADF-detector annulus in the CCD camera plane (determined in the way described earlier), but a $\sqrt{2}$ times larger radius, the proper alignment of the ADF-detector to detect scattering into the $n = 1$ Laue ring was completed.

Diffraction patterns have been recorded on the CCD camera while the beam was scanning over the sample to test the lateral movement of diffraction patterns while scanning. These tests have shown that the detector alignment was good for every beam position during a scan, especially at high magnification.

Figure 72 shows an experimental ADF-STEM image recorded with the detector configuration described above. Although we only collect radiation scattered to rather high angles, contrast due to the dislocation strain field is visible.

Figure 73 shows an experimental ADF-STEM image recorded under the conditions described above. Slightly changing the size of the diffraction pattern, so that the Laue ring is not aligned with the detector anymore, will bring the contrast to zero. This means that the signal in the image is due to the presence of an extra Laue ring, indicating a locally different structure than that of the bulk material, in that it has twice its periodicity along the electron beam direction.

STEM simulations using the code *S-TEM* and models calculated by João Justo (see section 2.3 in this chapter) are shown in figure 74. The models included the full strain-field, which was required to make sure that strain contrast cannot account for the experimentally observed image contrast. Simulations have been done for both proposed core structural models of the 90° partial. Comparing figures 73 and 74 It can be seen very clearly that both of the partial dislocations are reconstructed with a core-structure that has twice the periodicity along the electron beam direction as the bulk material surrounding it.

Comparing weak-beam TEM images of dislocations with ADF-STEM images shows that the observed contrast in figure 73 is indeed due to the presence of a dislocation. Also the wide-ranging strain contrast observed in weak-beam and BF TEM images seems to be much less in the ADF-STEM image, where the bright dislocation contrast is much more localized. One of the reasons for using a fully relaxed model of a dissociated edge dislocation

for the STEM simulation was to differentiate between image contrast due to the dislocation strain field and its core reconstruction periodicity. Since total energy calculations involving a super-cell large enough to contain the full strain-field are computationally very expensive, a smaller cell had to be used. However figure 74a already shows some of the strain contrast, and it can be seen that it is much less intense than the signal due to the presence of the extra Laue ring.

3.3. Modification of MIDAS for recording the extra Laue ring. In order to record microdiffraction patterns containing the extra Laue ring, a very stable, yet small electron beam probe is required. Being able to quickly switch from imaging to diffraction mode is also essential, while maintaining the probe position with high accuracy.

A VG HB501 UHV 100kV STEM has been modified to fulfill all these requirements. A very sensitive CCD camera with direct fiber-optic coupling to a very sensitive phosphor screen on the top of the column of the microscope has been installed. The absence of a mechanical shutter avoids vibrations of the column and thereby the sample stage. A thin annular dark field (ADF) detector has been designed, consisting of an oval ring of metal, which has been coated with phosphor. The ring has been installed at a 45° angle, reflecting most of the fluorescent light produced by electrons reaching the phospor on the ring into a photo multiplier tube, where it is converted into an electrical signal, which is then plotted as a function of beam position to form an image. The oval ADF detector is round in projection, as indicated by the "virtual" ADF detector shown in figure 76.

While this setup can be used for the same experimental procedure as described in the previous subsection (except for the low temperature), it also provides the opportunity to very easily record a full microdiffraction pattern for any position in the STEM image. The size of the diffraction pattern can be selected over a fairly large range with just the last projector lens. By simply stopping the beam at the desired position in the image and reducing the size of the diffraction pattern, a microdiffraction pattern for that particular pixel in the image can be recorded using the CCD camera. This procedure does not require a change between modes of operation, nor does it affect the electron beam before it passes through the sample. Experiments using this setup are currently being done.

4. Conclusion and Future Work

A new method for measuring the structural periodicity along dislocation cores using a focused electron beam has been introduced, which is able to probe single dislocations. This method works for any crystal structure, in which dislocation core structures have a longer periodicity along the dislocation line than the surrounding bulk-material. The method has been tested for the structure of both partial dislocation cores in Silicon, giving direct evidence for the double period reconstruction of both, the 30° and the 90° partial dislocation cores at a temperature of 95K.

The experimental data has been compared with very exact multiple scattering calculations which also include thermal diffuse scattering by the frozen phonon approximation.

A Einstein-model vibrational amplitude of $\langle u \rangle = 0.064$Å has been determined for the atoms in the perfect Silicon lattice at 95K. The models, which were calculated by João Justo for these simulations were fully relaxed structures containing the two partial dislocations, the stacking fault, and the strain field around these defects. Detailed STEM simulations for the single period and double period reconstructed model for the 90° partial dislocation core in Silicon show a very clear difference in the resulting images, allowing the experimental data to be clearly identified as steming from a double period structure of both, the 30° and the 90° partial dislcoation core.

Since this method is not limited to dislocation cores the diamond cubic structure, future work may include its application to many other materials, such as inter-metallics. One requirement of this method is that the periodicity of the dislocation core structure is greater than that of the surrounding bulk material. Also, materials with a rather large Debye-temperature (e.g. Diamond) may provide even better experimental results, since the background signal due to phonon scattering will be reduced.

Repeating the experiments (STEM and microdiffraction on MIDAS) described in this chapter at different temperatures will allow one to measure potential changes in the length of dislocation core which is DP reconstructed as a function of temperature.

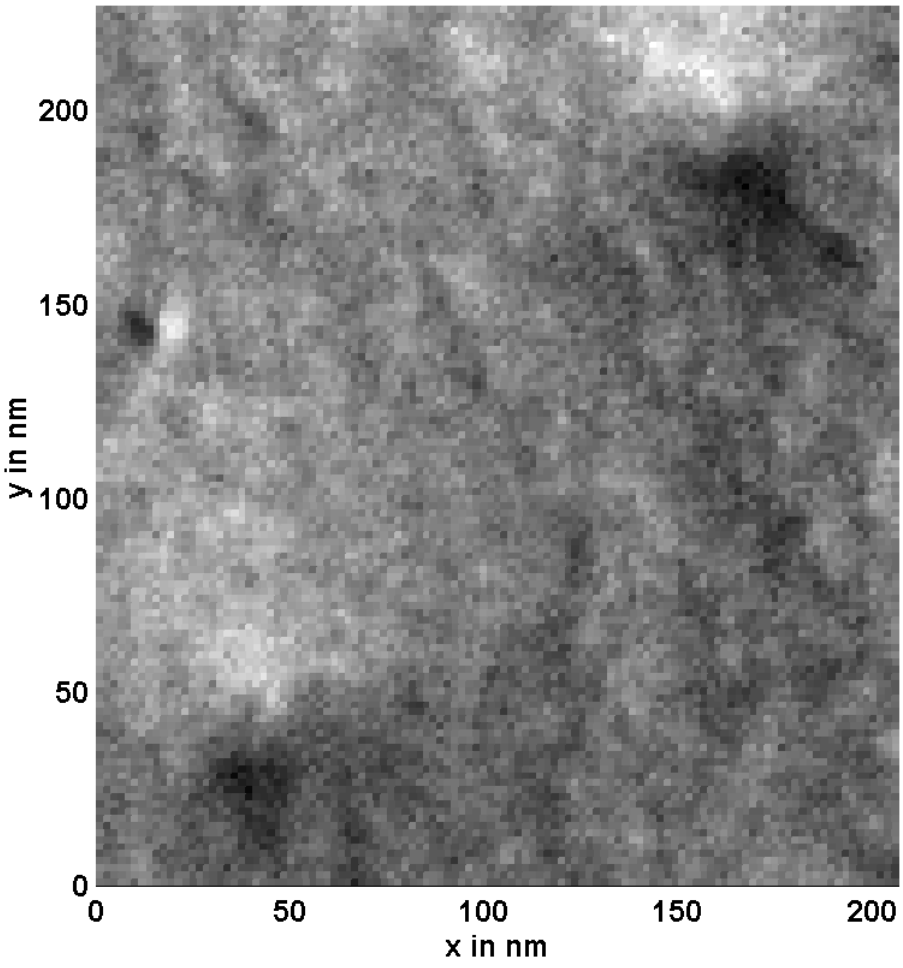

Figure 72. Low-resolution ADF-STEM image recorded with the detector tuned to the extra Laue ring. The image shows 3 dislocations running parallel to the beam direction with their associated strain contrast.

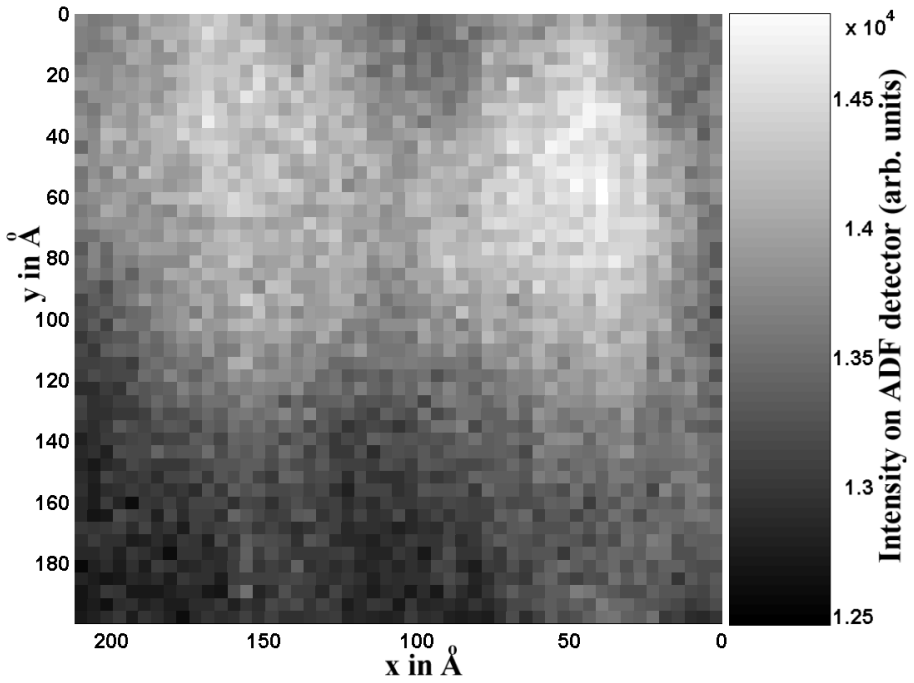

Figure 73. ADF-STEM image of end-on dissociated dislocation. The camera length has been adjusted to match the size of the ADF detector to the size of the additional Laue ring expected indicating DP reconstruction of the dislocation cores. Recorded on CM200 at 200kV using a liquid N stage to cool the sample to 95K in order to enhance contrast of the elastic signal.

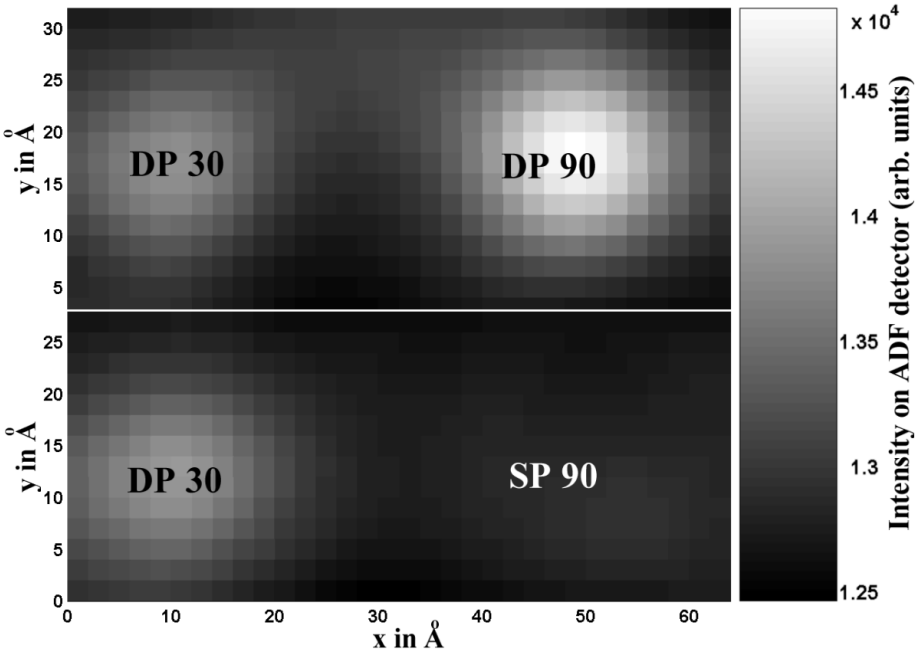

Figure 74. Simulated ADF-STEM images using the same microscope parameters, detector geometry and intensity scale as figure 73 $E = 200$kV, $C_s = 1.2$mm, $df = -400$Å, $C_c = 1.0$mm, $dE/E = 5 \times 10^{-6}$, $T = 95$K. Part of the model used for simulating the lower image is shown in figure 20 A RMS vibrational amplitude of 0.064Å has been used for including TDS, as measured by fitting the TDS background of experimental diffraction patterns. Effects due to a finite source size and probe vibration have been included by convoluting the image with Gaussian of 8Å width. The ADF detector covered the angular range of 76 ... 84mrad. The size of each image pixel is 2Å × 2Å . Top: 30° and a DP-reconstructed 90° partial dislocation. Bottom: 30° and a SP-reconstructed 90° partial dislocation. The 30° partial is always DP reconstructed.

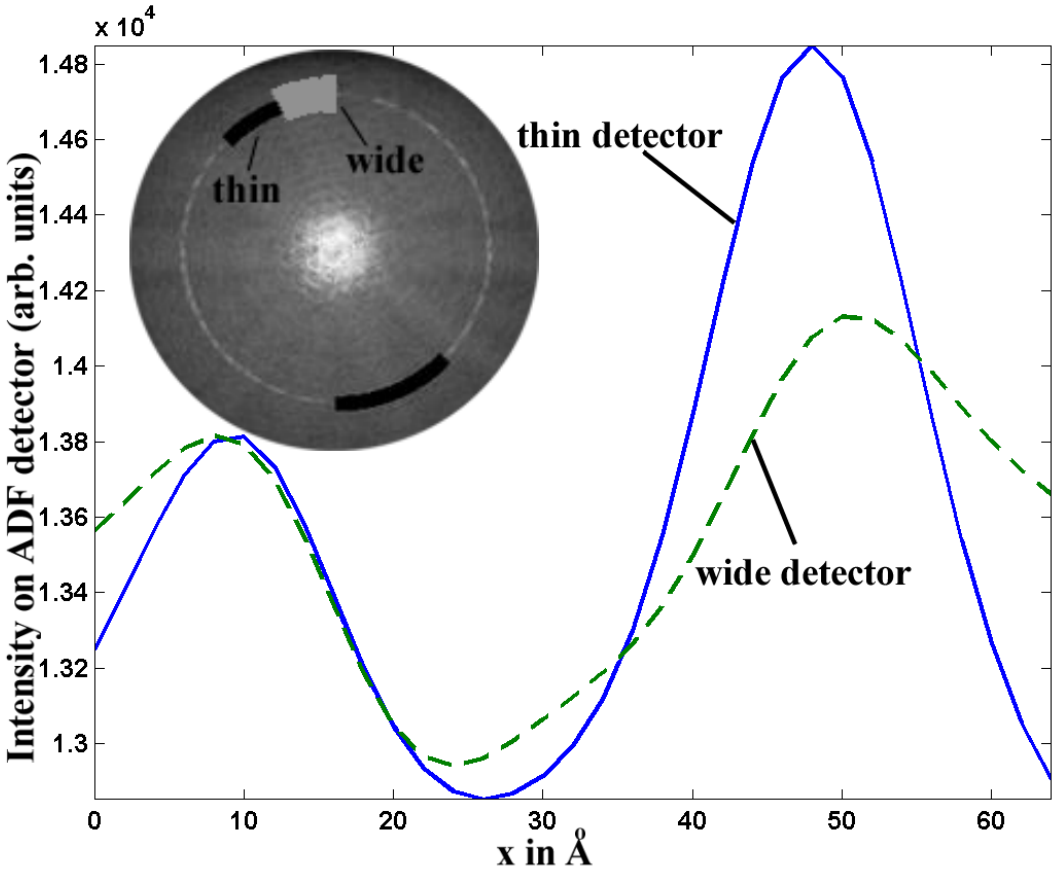

Figure 75. Linescan across simulated STEM images for different detector widths. Solid line: 76 . . . 84mrad, dashed line: 71 . . . 91mrad. The intensity for the wide detector (dashed line) has been scaled to be equal to that of the thin detector (solid line) at the 30° partial peak. The size and position of both detectors is also shown in the inset.

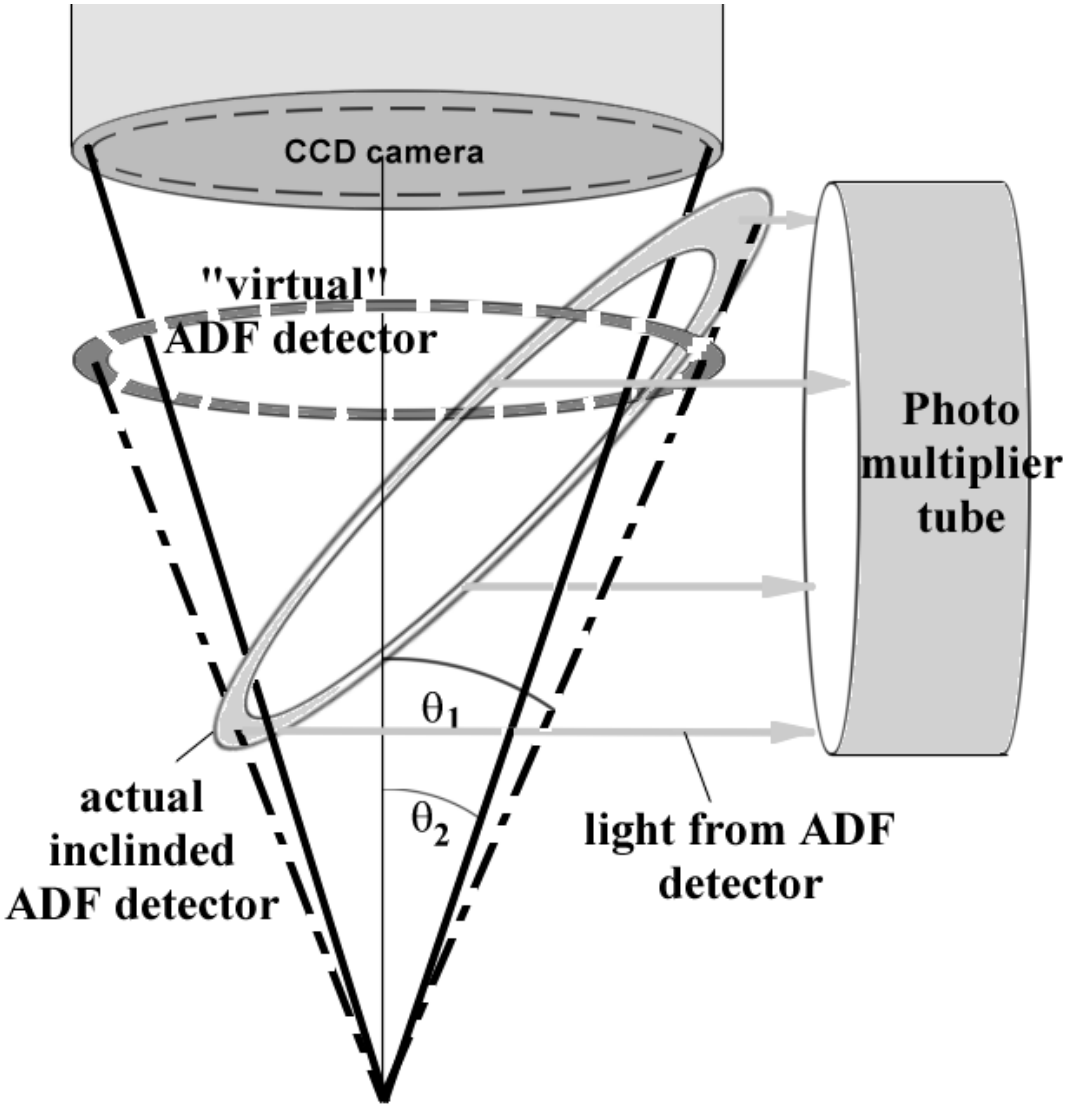

Figure 76. Microdiffraction and ADF detector system in MIDAS.

APPENDIX A
A NEW EXPANSION OF THE MATRIX EXPONENTIAL OF TWO NON-COMMUTING MATRICES AND ITS APPLICATION OF TO SOLVE THE MULTIPLE SCATTERING INVERSION PROBLEM

1. Introduction

Mathematical models of many problems in biology, physics, and economics involve systems of linear, constant coefficient equations.

$$\frac{\delta x}{\delta \lambda}(\lambda) \quad = \quad Ax(\lambda)$$

Such systems of equations have the general solution

$$x(\lambda) = e^{\lambda A} x(0)$$

As an example the solution of probably the most fundamental equation in modern physics, the Schrödinger equation,

$$i\frac{\delta \Psi(\vec{r},t)}{\delta t} = H\Psi(\vec{r},t) \tag{A.1}$$

falls into this class, where the Hamiltonian $H = \hbar/(2m|e|)\nabla^2 + V$ is a square matrix that can be split into a potential energy term, the matrix V, and a diagonal matrix containing the kinetic energy terms. Assuming that the wave function $\Psi(\vec{r},t) = \Phi(\vec{r})\exp(2\pi i[k_z + Et/h])$ has the form of a modulated plane wave and the second order derivative of $\Phi(\vec{r})$ in the direction of its propagation (z) is negligible, the time-independent Schrödinger equation can also be written as a system of linear differential equations in z

$$\frac{\delta \Phi(\vec{r})}{\delta z} = \left[\left(\frac{i\nabla_{xy}^2}{4\pi k_z} - \frac{\vec{k}_{xy}}{k_z}\cdot\vec{\nabla}_{xy}\right) + \frac{i}{4\pi k_z}V(\vec{r})\right]\Phi(\vec{r}) = 0, \tag{A.2}$$

where we have the general matrix of potential energy terms $iV/(4\pi k_z)$ and a diagonal matrix for the kinetic energy. In reciprocal space the differential operators ∇_{xy}^2 and ∇_{xy} turn into linear operators, making this also a problem of the class treated in this paper. A prominent example would be multiple scattering of high energy particles in matter, or rather its inversion, where the problem is that of determining the (periodic) scattering potential $V(\vec{r})$ from the scattered radiation, which is equivalent to finding A from the moduli of the entries of a single column of the matrix $S = \exp(\lambda[A+B])$, where B is a diagonal matrix with known elements. Varying the elements in B lets us set up many of these systems of equations, but unless we can find some linear approximation for them, they will still remain hard to solve directly. The purpose of this paper is to provide a general expansion which allows us to express the matrix exponential $S = \exp(\lambda[A+B])$ as

$$S_{n,m} \quad = \quad e^{\lambda b_n}\delta_{n,m} + \sum_{q=1}^{\infty}\sum_{l_1=1}^{N}\cdots\sum_{l_{q-1}=1}^{N} a_{n,l_1}\cdots a_{l_{q-1},m} C^{(q)}_{n,l_1,\ldots l_{q-1},m}(B,\lambda)$$

where $a_{n,m}$ are elements of A, and $C^{(q)}_{n,l_1,\ldots l_{q-1},m}(B,\lambda)$ is an analytical expressions in elements of B and λ.

The convergence of this expansion will be discussed and is shown to be superior to the Taylor expansion in terms of $(\lambda[A+B])$, especially if the largest elements of B are larger than the largest element of A.

This expansion provides a disentanglement of the matrix exponential into a sum of terms containing different length products of elements of A with known coefficients. The inversion of convergent beam electron diffraction (CBED) patterns is a problem of exactly this class. CBED patterns, whose scattering intensity, expressed in the matrix exponential Bloch wave formalism, provide experimental data for the moduli of certain elements in the matrix S for fixed structure factors in the matrix A, but many different matrices B. The application of this expansion to the inversion of CBED patterns will therefore be discussed.

2. Expansion of the Matrix Exponential

There exist many different ways to compute the matrix exponential $S = \exp(M)$. Nineteen of them are listed in [176]. Computing the matrix exponential from its Taylor expansion is probably one of the least efficient ways. However this method involves only integer powers of matrix elements, allowing every elements of S to be expressed as a polynomial of elements in M in a straightforward way. The matrix exponential is defined as

$$S = e^{\lambda(A+B)} = \lim_{p\to\infty} \sum_{j=0}^{p} \frac{\lambda^j}{j!}(A+B)^j \tag{A.3}$$

where

$$\begin{aligned}
\left[(A+B)^j\right]_{n,m} &= \sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{j-1}=1}^{N}(a_{n,l_1}+b_{n,l_1})(a_{l_1,l_2}+b_{l_1,l_2})\cdots \\
&\quad (a_{l_{j-1},m}+b_{l_{j-1},m}) \\
&= b_n^j\delta_{n,m} + \sum_{q=1}^{j}\underbrace{\sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{q-1}=1}^{N}}_{q-1} a_{n,l_1}a_{l_1,l_2}\cdots a_{l_{q-1},m} \\
&\quad \sum_{j_0=0}^{j-q}\sum_{j_1=0}^{j-q-j_1}\cdots\sum_{j_{q-1}=0}^{j-q-\sum_{i=0}^{q-2}j_i} b_n^{j_0}b_{l_1}^{j_1}\cdots b_{l_{q-1}}^{j_{q-1}}b_m^{j-q-\sum_{i=0}^{q-1}j_i} \\
&= b_n^j\delta_{n,m} + \sum_{q=1}^{j}\sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{q-1}=1}^{N} a_{n,l_1}a_{l_1,l_2}\cdots \\
&\quad a_{l_{q-1},m}C^{(j-q,q)}_{n,l_1,\ldots l_{q-1},m}
\end{aligned} \tag{A.4}$$

and (setting $l_0 = n$, $l_q = m$)

$$C^{(j,q)}_{l_0,l_1,\ldots l_{q-1},l_q} = b_{l_q}^j \underbrace{\sum_{j_0=0}^{j}\left(\frac{b_{l_0}}{b_{l_q}}\right)^{j_0}\sum_{j_1=0}^{j-j_0}\left(\frac{b_{l_1}}{b_{l_q}}\right)^{j_1}\cdots\sum_{j_{q-1}=0}^{j-\sum_{i=0}^{q-2}j_i}\left(\frac{b_{l_{q-1}}}{b_{l_q}}\right)^{j_{q-1}}}_{q}.$$

If $b_{l_{q-1}} = b_{l_q}$ then the last sum is just $j + 1 - \sum_{i=0}^{q-2} j_i$ which gives

$$C^{(j,q)}_{l_0,l_1,\ldots l_{q-1},l_q} = \frac{\delta}{\delta b_{l_{q-1}}} \left(b^{j+1}_{l_{q-1}} \sum_{j_0=0}^{j} \left(\frac{b_{l_0}}{b_{l_{q-1}}} \right)^{j_0} \cdots \sum_{j_{q-2}=0}^{j-\sum_{i=0}^{q-3} j_i} \left(\frac{b_{l_{q-2}}}{b_{l_{q-1}}} \right)^{j_{q-2}} \right)$$

In general we may have $b_{l_{q-d}} = \ldots = b_{l_{q-2}} = b_{l_{q-1}} = b_{l_q}$, $0 \leq d \leq q$. Using

$$\sum_{j_1=0}^{N} \sum_{j_2=0}^{N-j_1} \cdots \sum_{j_d=0}^{N-\sum_{i=1}^{d} j_i} 1 = \frac{(N+1)(N+2)\cdots(N+d)}{d!} = \binom{N+d}{d} \tag{A.5}$$

we get

$$\begin{aligned} C^{(j,q)}_{l_0,l_1,\ldots l_{q-1},l_q} &= \frac{1}{d!} \left(\frac{\delta}{\delta b_{l_{q-d}}} \right)^d \left(b^{j+d}_{l_{q-d}} \sum_{j_0=0}^{j} \left(\frac{b_{l_0}}{b_{l_{q-d}}} \right)^{j_0} \cdots \right. \\ &\quad \left. \sum_{j_{q-d-1}=0}^{j-\sum_{i=0}^{q-d-2} j_i} \left(\frac{b_{l_{q-d-1}}}{b_{l_{q-d}}} \right)^{j_{q-d-1}} \right) \\ &= \frac{1}{d!} \left(\frac{\delta}{\delta b_{l_{q-d}}} \right)^d b^d_{l_{q-d}} C^{(j,q-d)}_{l_0,l_1,\ldots l_{q-d}} \end{aligned} \tag{A.6}$$

An equivalent expression is given in [177]. It can be shown that we arrive at the same result by treating all the b_{l_i} at first as non-degenerate and treat the degenerate case as $\lim_{\epsilon_1,\ldots\epsilon_d \to 0} C^{(j,q)}_{l_0,l_1,\ldots l_{q-1},l_q}$, where $b_{l_{q-1}} = b_{l_q}(1+\epsilon_1) \ldots b_{l_{q-d}} = b_{l_q}(1+\epsilon_d)$. We will therefore now first concentrate on the non-degenerate case.

If $b_{l_{q-1}} \neq b_{l_q}$ the last sum is a finite geometric series:

$$\begin{aligned} C^{(j,q)}_{l_0,l_1,\ldots l_{q-1},l_q} &= b^j_{l_q} \sum_{j_0=0}^{j} \left(\frac{b_{l_0}}{b_{l_q}} \right)^{j_0} \cdots \sum_{j_{q-2}=0}^{j-\sum_{i=0}^{q-3} j_i} \left(\frac{b_{l_{q-2}}}{b_{l_q}} \right)^{j_{q-2}} \\ &\quad \frac{1 - \left(\frac{b_{l_{q-1}}}{b_{l_q}} \right)^{j+1-\sum_{i=0}^{q-2} j_i}}{1 - \left(\frac{b_{l_{q-1}}}{b_{l_q}} \right)} \\ &= \frac{1}{1 - \left(\frac{b_{l_{q-1}}}{b_{l_q}} \right)} b^j_{l_q} \sum_{j_0=0}^{j} \left(\frac{b_{l_0}}{b_{l_q}} \right)^{j_0} \cdots \sum_{j_{q-2}=0}^{j-\sum_{i=0}^{q-3} j_i} \left(\frac{b_{l_{q-2}}}{b_{l_q}} \right)^{j_{q-2}} - \\ &\quad \frac{\left(\frac{b_{l_{q-1}}}{b_{l_q}} \right)}{1 - \left(\frac{b_{l_{q-1}}}{b_{l_q}} \right)} b^j_{l_{q-1}} \sum_{j_0=0}^{j} \left(\frac{b_{l_0}}{b_{l_{q-1}}} \right)^{j_0} \cdots \sum_{j_{q-2}=0}^{j-\sum_{i=0}^{q-3} j_i} \left(\frac{b_{l_{q-2}}}{b_{l_{q-1}}} \right)^{j_{q-2}} \\ &= \frac{b_{l_{q-1}}}{b_{l_{q-1}} - b_{l_q}} C^{(j,q-1)}_{l_0,l_1,\ldots l_{q-1}} + \frac{b_{l_q}}{b_{l_q} - b_{l_{q-1}}} C^{(j,q-1)}_{l_0,l_1,\ldots l_{q-2},l_q}, \end{aligned} \tag{A.7}$$

and in particular for $q = 0, 1$:

$$\begin{aligned} C^{(j,0)}_{l_0} &= b^j_{l_0} \\ C^{(j,1)}_{l_0,l_1} &= \begin{cases} \frac{b^{j+1}_{l_0}}{b_{l_0}-b_{l_1}} + \frac{b^{j+1}_{l_1}}{b_{l_1}-b_{l_0}} & \text{if } b_{l_0} \neq b_{l_1} \\ \frac{\delta}{\delta b_0} b^{j+1}_{l_0} = \binom{j+1}{1} \frac{b^{j+1}_{l_0}}{b_{l_0}} & \text{if } b_{l_0} = b_{l_1} \end{cases} \end{aligned}$$

In general, for non-degenerate b_{l_k} $(k = 0, \dots q)$ we get the following result

$$C^{(j,q)}_{l_0,l_1,\dots l_q} = \sum_{k=0}^{q} b^{j+q}_{l_k} \prod_{\substack{r=0 \\ r\neq k}}^{q} (b_{l_k} - b_{l_r})^{-1} \tag{A.8}$$

We will give the proof by induction.

$$\begin{aligned} C^{(j,q+1)}_{l_0,\dots l_q,l_{q+1}} &= \frac{b_{l_q}}{b_{l_q} - b_{l_{q+1}}} C^{(j,q)}_{l_0,\dots l_q} + \frac{b_{l_{q+1}}}{b_{l_{q+1}} - b_{l_q}} C^{(j,q)}_{l_0,\dots l_{q-1},l_{q+1}} \\ &= \sum_{k=0}^{q-1} b^{j+q}_{l_k} \prod_{\substack{r=0 \\ r\neq k}}^{q-1} (b_{l_k} - b_{l_r})^{-1} \\ &\quad \left(\frac{b_{l_q}}{(b_{l_q} - b_{l_{q+1}})(b_{l_k} - b_{l_q})} + \frac{b_{l_{q+1}}}{(b_{l_{q+1}} - b_{l_q})(b_{l_k} - b_{l_{q+1}})} \right) + \\ &\quad \frac{b^{j+q+1}_{l_q}}{b_{l_q} - b_{l_{q+1}}} \prod_{r=0}^{q-1} (b_{l_q} - b_{l_r})^{-1} + \frac{b^{j+q+1}_{l_{q+1}}}{b_{l_{q+1}} - b_{l_q}} \prod_{r=0}^{q-1} (b_{l_{q+1}} - b_{l_r})^{-1} \\ &= \sum_{k=0}^{q-1} b^{j+q}_{l_k} \prod_{\substack{r=0 \\ r\neq k}}^{q-1} (b_{l_k} - b_{l_r})^{-1} \frac{b_{l_k}}{(b_{l_k} - b_{l_{q+1}})(b_{l_k} - b_{l_q})} + \\ &\quad b^{j+q+1}_{l_q} \prod_{\substack{r=0 \\ r\neq q}}^{q+1} (b_{l_q} - b_{l_r})^{-1} + b^{j+q+1}_{l_{q+1}} \prod_{\substack{r=0 \\ r\neq q+1}}^{q+1} (b_{l_{q+1}} - b_{l_r})^{-1} \\ &= \sum_{k=0}^{q+1} b^{j+q+1}_{l_k} \prod_{\substack{r=0 \\ r\neq k}}^{q+1} (b_{l_k} - b_{l_r})^{-1} = C^{(j,q+1)}_{l_0,\dots l_q,l_{q+1}} \end{aligned}$$

Let us now consider the degenerate case. Let us assume that the values $b_{l_0}, \dots b_{l_d}$, $(d \geq 0)$ are almost degenerate, i.e. $b_{l_0} = b'_{l_0} + \epsilon_0$, $\dots b_{l_d} = b'_{l_0} + \epsilon_d$. Then the degenerate case is defined by the limit $\epsilon_0, \dots \epsilon_d \to 0$.

$$\begin{aligned} C^{(j,q)}_{l_0,\dots l_q} = \lim_{\epsilon_1,\dots \epsilon_d \to 0} \Bigg[& \sum_{k=0}^{d} (b'_{l_0} + \epsilon_k)^{j+q} \prod_{\substack{r=0 \\ r\neq k}}^{d} \frac{1}{\epsilon_k - \epsilon_r} \prod_{r=d+1}^{q} \frac{1}{((b'_{l_0} + \epsilon_k) - b_{l_r})} + \\ & \sum_{k=d+1}^{q} b^{j+q}_{l_k} \prod_{r=0}^{d} \frac{1}{b_{l_k} - (b'_{l_0} + \epsilon_k)} \prod_{\substack{r=d+1 \\ r\neq k}}^{q} (b_{l_k} - b_{l_r})^{-1} \Bigg] \end{aligned}$$

Since the limit is independent of the direction from which we approach it, we choose $\epsilon_k = \epsilon e^{ik\phi_0}$, with $\phi_0 = 2\pi/(d+1)$, i.e. $\epsilon_0 = \epsilon$, $\epsilon_1 = \epsilon e^{i\phi_0}$, etc.. With this choice we can make use of the following equalities:

$$\sum_{k=0}^{d} e^{ik\phi_0} = \sum_{k=0}^{d} (e^{ik\phi_0})^s = \left(\sum_{k=0}^{d} e^{ik\phi_0} \right)^s = 0, \quad s = 1, 2, \ldots \tag{A.9}$$

$$\sum_{\substack{k_1=0 \\ k_1 \neq k}}^{d} (e^{ik_1\phi_0})^s = -e^{isk\phi_0} \tag{A.10}$$

$$\sum_{\substack{k_1=0 \\ k_1 \neq k}}^{d} \sum_{\substack{k_2=0 \\ k_2 \neq k, k_1}}^{d} \cdots \sum_{\substack{k_s=0 \\ k_s \neq k, k_1, \ldots k_{s-1}}}^{d} e^{ik_1\phi_0} e^{ik_2\phi_0} \cdots e^{ik_s\phi_0} = (-1)^s s! e^{isk\phi_0} \tag{A.11}$$

$$\prod_{\substack{k'=0 \\ k' \neq k}}^{d} \frac{1}{e^{ik\phi_0} - e^{ik'\phi_0}} = \frac{e^{ik\phi_0}}{d+1} \tag{A.12}$$

We will also define a new variable name: $b_{0,r} = b'_{l_0} - b_{l_r}$.

$$\begin{aligned}
C^{(j,q)}_{l_0,\ldots l_q} &= \lim_{\epsilon \to 0} \left[\sum_{k=0}^{d} \left\{ \left(\sum_{j'=0}^{j+q} \binom{j+q}{j'} b_{l_0}^{j+q-j'} \epsilon^{j'} e^{ij'k\phi_0} \right) \epsilon^{-d} \right. \right. \\
&\quad \left. \left. \prod_{\substack{r=0 \\ r \neq k}}^{d} \frac{1}{e^{ik\phi_0} - e^{ir\phi_0}} \prod_{\substack{r=d+1 \\ r \neq k}}^{q} \frac{1}{(b_{0,r} + \epsilon e^{ik\phi_0})} \right\} \right] + \\
&\quad \sum_{k=d+1}^{q} \frac{b_{l_k}^{j+q}}{(b_{l_k} - b_{l_0})^{d+1}} \prod_{\substack{r=d+1 \\ r \neq k}}^{q} (b_{l_k} - b_{l_r})^{-1} \\
&= \sum_{j'=0}^{j+q} \binom{j+q}{j'} b_{l_0}^{j+q-j'} \frac{1}{(d+1)} \\
&\quad \lim_{\epsilon \to 0} \left[\frac{\sum_{k=0}^{d} e^{ik\phi_0} e^{ij'k\phi_0} \epsilon^{j'} \prod_{\substack{k'=0 \\ k' \neq k}}^{d} \prod_{r=d+1}^{q} (b_{0,r} + \epsilon e^{ik'\phi_0})}{\epsilon^d \prod_{k'=0}^{d} \prod_{r=d+1}^{q} (b_{0,r} + \epsilon e^{ik'\phi_0})} \right] + \\
&\quad \sum_{k=d+1}^{q} \frac{b_{l_k}^{j+q}}{(b_{l_k} - b_{l_0})^{d+1}} \prod_{\substack{r=d+1 \\ r \neq k}}^{q} (b_{l_k} - b_{l_r})^{-1}
\end{aligned} \tag{A.13}$$

using (A.12). For $i > d$ the limit of the expression in square brackets is zero. For $i \leq d$ we need to evaluate that expression explicitly. Using (A.9) and some thought we get for the denominator

$$\epsilon^d \prod_{k'=0}^{d} \prod_{r=d+1}^{q} (b_{0,r} + e^{ik'\phi_0}\epsilon) = \epsilon^d \prod_{k'=0}^{d} \prod_{r=d+1}^{q} b_{0,r} = \epsilon^d \left(\prod_{r=d+1}^{q} b_{0,r} \right)^{d+1} \tag{A.14}$$

The numerator needs to be expanded in powers of ϵ. The product

$$\prod_{\substack{k'=0\\k'\neq k}}^{d}\prod_{r=d+1}^{q}(b_{0,r}+\epsilon e^{ik'\phi_0}) \;=\; \sum_{s=0}^{d(q-d)}\epsilon^s P^{(s)}(k,d,q,\phi_0)$$

over all possible combinations of r and $k' \neq k$ can be expressed in a sum of terms, each involving a certain power of ϵ. The first one, of course is just the product over all the first terms in the sum $(b_{0,r}+\epsilon e^{ik'\phi_0})$, i.e.

$$P^{(0)}(d,q) \;=\; \prod_{r=d+1}^{q} b_{0,r}^d, \tag{A.15}$$

where the power d comes from the fact that there are d different values of $k' = 0 \ldots d$, $k' \neq k$. The next few terms are

$$\begin{aligned}
\epsilon P^{(1)}(k,d,q,\phi_0) &= \epsilon \prod_{r=d+1}^{q} b_{0,r}^d \left[\sum_{r=d+1}^{q} \frac{\sum_{\substack{k_1=0\\k_1\neq k}}^{d} e^{ik_1\phi_0}}{b_{0,r}}\right] \\
&= -\epsilon e^{ik\phi_0}\prod_{r=d+1}^{q} b_{0,r}^d \left[\sum_{r=d+1}^{q}\frac{1}{b_{0,r}}\right] \qquad (A.16)\\
\epsilon^2 P^{(2)}(k,d,q,\phi_0) &= \frac{1}{2!}\epsilon^2 \prod_{r=d+1}^{q} b_{0,r}^d \left[\sum_{r_1=d+1}^{q}\sum_{\substack{r_2=d+1\\r_2\neq r_1}}^{q} \frac{\left(\sum_{\substack{k_1=0\\k_1\neq k}}^{d} e^{ik_1\phi_0}\right)^2}{b_{0,r_1}b_{0,r_2}} + \right.\\
&\qquad \left. \sum_{r=d+1}^{q} \frac{\sum_{\substack{k_1=0\\k_1\neq k}}^{d}\sum_{\substack{k_2=0\\k_2\neq k,k_1}}^{d} e^{ik_1\phi_0}e^{ik_2\phi_0}}{b_{0,r}^2}\right] \\
&= \frac{1}{2!}\epsilon^2 2! e^{i2k\phi_0}\prod_{r=d+1}^{q} b_{0,r}^d \\
&\qquad \left[\sum_{r_1=d+2}^{q}\sum_{r_2=d+1}^{r_1-1}\frac{1}{b_{0,r_1}b_{0,r_2}} + \sum_{r=d+1}^{q}\frac{1}{b_{0,r}^2}\right] \qquad (A.17)
\end{aligned}$$

where we used (A.10) in (A.16) and (A.10) and (A.11) in (A.17). The 1/2! term accounts for the times we double counted terms when performing the double sum. The left hand side of equation (A.17) demonstrates that we have to carefully separate products containing the same index r twice from those that don't. If the same index r appears twice we need to make sure that we don't also have the same k_1 index twice in the same term. This is because we are trying to find all possible combinations of sums $(b_{0,r}+\epsilon e^{ik'\phi_0})$ out of a given pool without multiplying one with itself. This method can be extended to any power $s > 0$

of ϵ:

$$
\begin{aligned}
P^{(s)}(k,d,q,\phi_0) &= \frac{1}{s!}\prod_{r=d+1}^{q} b_{0,r}^{d}\left[s!\sum_{r_1=d+s}^{q}\cdots\sum_{r_s=d+1}^{r_{s-1}-1}\frac{\left(-e^{ik\phi_0}\right)^s}{b_{0,r_1}\cdots b_{0,r_s}}+\right.\\
&\quad \binom{s}{2}\sum_{r=d+1}^{q}\frac{(-1)^2 2! e^{i2k\phi_0}}{b_{0,r}^2}\left((s-2)!\sum_{\substack{r_1=d+s-2\\ r_1\neq r}}^{q}\cdots\right.\\
&\quad \sum_{\substack{r_{s-2}=d+1\\ r_{s-2}\neq r}}^{r_{s-3}-1}\frac{\left(-e^{ik\phi_0}\right)^{s-2}}{b_{0,r_1}\cdots b_{0,r_{s-2}}}+\\
&\quad \binom{s-2}{2}\sum_{r_1=d+1}^{r}-1\frac{(-1)^2 2! e^{i2k\phi_0}}{b_{0,r_1}^2}(s-4)!\sum_{\substack{r_2=d+s-2\\ r_2\neq r,r_1}}^{q}\cdots\\
&\quad \left.\left.\sum_{\substack{r_{s-3}=d+1\\ r_{s-4}\neq r,r_1}}^{r_{s-4}-1}\frac{\left(-e^{ik\phi_0}\right)^{s-2}}{b_{0,r_2}\cdots b_{0,r_{s-3}}}+\ldots\right)+\ldots\right.\\
&\quad \left.+\binom{s}{s}\sum_{r=d+1}^{q}\frac{(-1)^s s! e^{isk\phi_0}}{b_{0,r}^s}(s-s)!\right]
\end{aligned}
$$

Writing down a few of those terms one will soon realize that the binomial coefficients due to the terms with products of multiple equal $b_{0,r}$ always cancel with the factorials that stem from the products with different $b_{0,r}$. We can therefore combine all those different sums into a single s-dimensional sum:

$$
\begin{aligned}
P^{(s)}(k,d,q,\phi_0) &= \prod_{r=d+1}^{q} b_{0,r}^{d}(-1)^s e^{isk\phi_0}\\
&\quad \left[\sum_{r_1=d+1}^{q}\frac{1}{b_{0,r_1}}\sum_{r_2=d+1}^{r_1}\frac{1}{b_{0,r_2}}\cdots\sum_{r_s=d+1}^{r_{s-1}}\frac{1}{b_{0,r_s}}\right] \qquad \text{(A.18)}
\end{aligned}
$$

If we perform the summation over k in

$$
\begin{aligned}
\sum_{k=0}^{d} e^{ik\phi_0}e^{ij'k\phi_0}e^{isk\phi_0} &= \sum_{k=0}^{d} e^{2\pi ik(j'+1+s)/(d+1)}\\
&= \begin{cases} (d+1) & \text{if } (j'+1+s)/(d+1) = 1,2,\ldots \\ 0 & \text{else} \end{cases}
\end{aligned}
$$

because of (A.9). Since we have the upper limit of $s \leq d - j'$, the only non-vanishing term is the one with $s = d - j'$, for which also the powers of ϵ cancel. This allows us to write

down the limit of the expression in square brackets in (A.13)

$$\begin{aligned}
&\lim_{\epsilon\to 0}\left[\frac{\sum_{k=0}^{d} e^{ik\phi_0}e^{ij'k\phi_0}\epsilon^{j'}\prod_{\substack{k'=0\\k'\neq k}}^{d}\prod_{r=d+1}^{q}(b_{0,r}+\epsilon e^{ik'\phi_0})}{\epsilon^{d}\prod_{k'=0}^{d}\prod_{r=d+1}^{q}(b_{0,r}+\epsilon e^{ik'\phi_0})}\right]\\
=\;&\lim_{\epsilon\to 0}\left[\sum_{k=0}^{d}\epsilon^{j'-d}e^{i(j'+1)k\phi_0}\prod_{\substack{k'=0\\k'\neq k}}^{d}\prod_{r=d+1}^{q}(b_{0,r}+\epsilon e^{ik'\phi_0})\right]\\
=\;&(d+1)(-1)^{d-j'}\prod_{r=d+1}^{q}b_{0,r}^{d}\left[\sum_{r_1=d+1}^{q}\frac{1}{b_{0,r_1}}\sum_{r_2=d+1}^{r_1}\frac{1}{b_{0,r_2}}\cdots\sum_{r_{d-j'}=d+1}^{r_{d-j'-1}}\frac{1}{b_{0,r_{d-j'}}}\right]
\end{aligned}$$

If we define $\{b'_{l_0},\ldots,b'_{l_u}\}$ as the set of unique b_{l_k} from the whole set $\{b_{l_0},\ldots,b_{l_q}\}$, each associated with a degeneracy d_k, we can write (A.13) in a more general form ($d_k=0$, if b'_{l_k} is unique, $d_k=1$, if b'_{l_k} is singly degenerate, i.e. $b'_{l_k}=b_{l_{k1}}=b_{l_{k2}}$, etc.).

$$\begin{aligned}
C^{(j,q)}_{l_0,l_1,\ldots l_q} &= \sum_{k=0}^{u}\sum_{j'=0}^{d_k}\binom{j+q}{j'}b'^{j+q-j'}_{l_k}(-1)^{d_k-j'}\frac{\prod_{\substack{r=0\\r\neq k}}^{u}(b'_{l_k}-b'_{l_r})^{d_k(d_r+1)}}{\prod_{\substack{r=0\\r\neq k}}^{u}(b'_{l_k}-b'_{l_r})^{(d_k+1)(d_r+1)}}\\
&\quad\left[\sum_{\substack{r_1=0\\b_{l_{r_1}}\neq b'_{l_k}}}^{q}\frac{1}{(b'_{l_k}-b_{l_{r_1}})}\sum_{\substack{r_2=0\\b_{l_{r_2}}\neq b'_{l_k}}}^{r_1}\frac{1}{(b'_{l_k}-b_{l_{r_2}})}\cdots\right.\\
&\quad\left.\sum_{\substack{r_{d_k-j'}=0\\b_{l_{r_{d_k-j'}}}\neq b'_{l_k}}}^{r_{d_k-j'-1}}\frac{1}{(b'_{l_k}-b_{l_{r_{d_k-j'}}})}\right]\\
&= \sum_{k=0}^{u}\sum_{j'=0}^{d_k}\binom{j+q}{j'}b'^{j+q-j'}_{l_k}D^{(j',q)}_{l_0,l_1,\ldots l_q}
\end{aligned}\tag{A.19}$$

where

$$\begin{aligned}
D^{(j',q)}_{l_0,l_1,\ldots l_q} &= (-1)^{d_k-j'}\prod_{\substack{r=0\\r\neq k}}^{u}\frac{1}{(b'_{l_k}-b'_{l_r})^{(d_r+1)}}\\
&\left[\underbrace{\sum_{\substack{r_1=0\\b_{l_{r_1}}\neq b'_{l_k}}}^{q}\frac{1}{(b'_{l_k}-b_{l_{r_1}})}\sum_{\substack{r_2=0\\b_{l_{r_2}}\neq b'_{l_k}}}^{r_1}\frac{1}{(b'_{l_k}-b_{l_{r_2}})}\cdots\sum_{\substack{r_{d_k-j'}=0\\b_{l_{r_{d_k-j'}}}\neq b'_{l_k}}}^{r_{d_k-j'-1}}\frac{1}{(b'_{l_k}-b_{l_{r_{d_k-j'}}})}}_{d_k-j'}\right]
\end{aligned}$$

If $d_k = q$, which means that $b_{l_0} = b_{l_1} = \ldots = b_{l_q} = b'_{l_0}$, then $D^{(j',q)}_{l_0,l_1,\ldots l_q} = \delta_{d_k,j'} = \delta_{q,j'}$.

This result can now be used in (A.4).

$$\begin{aligned}
S_{n,m} &= \lim_{p\to\infty}\sum_{j=0}^{p}\frac{\lambda^j}{j!}\left(b_n^j\delta_{n,m} + \sum_{q=1}^{j}\sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{q-1}=1}^{N} a_{n,l_1}\cdots a_{l_{q-1},m}\right.\\
&\quad \left.\sum_{k=0}^{u}\sum_{j'=0}^{d_k}\binom{j}{j'}b_{l_k}'^{j-j'}D^{(j',q)}_{l_0,l_1,\ldots l_q}\right)\\
&= e^{\lambda b_n}\delta_{n,m} + \lim_{p\to\infty}\sum_{q=1}^{p}\sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{q-1}=1}^{N} a_{n,l_1}\cdots a_{l_{q-1},m}\\
&\quad \sum_{j=q}^{p}\frac{\lambda^j}{j!}\sum_{k=0}^{u}\sum_{j'=0}^{d_k}\binom{j}{j'}b_{l_k}'^{j-j'}D^{(j',q)}_{l_0,l_1,\ldots l_q}\\
&\overset{s=j-q}{=} e^{\lambda b_n}\delta_{n,m} + \sum_{q\geq 1}\sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{q-1}=1}^{N} a_{n,l_1}\cdots a_{l_{q-1},m}\\
&\quad \sum_{k=0}^{u}\sum_{j'=0}^{d_k}\frac{\lambda^{j'}}{j'!}D^{(j',q)}_{l_0,l_1,\ldots l_q}\lim_{p\to\infty}\sum_{s=0}^{p-q}\frac{\lambda^{s+q-j'}}{(s+q-j')!}b_{l_k}'^{s+q-j'}\\
&= e^{\lambda b_n}\delta_{n,m} + \sum_{q\geq 1}\sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{q-1}=1}^{N} a_{n,l_1}\cdots a_{l_{q-1},m}\\
&\quad \sum_{k=0}^{u}\sum_{j'=0}^{d_k}\frac{\lambda^{j'}}{j'!}D^{(j',q)}_{l_0,l_1,\ldots l_q}\left[e^{\lambda b'_{l_k}} - \sum_{r=0}^{q-j'-1}\frac{(\lambda b'_{l_k})^r}{r!}\right] \qquad \text{(A.20)}
\end{aligned}$$

For a finite maximum value of q this expression has a finite number of terms. For the case that all $b_{l_k} = b_n$ are equal, because of $D^{(j',q)}_{l_0,l_1,\ldots l_q} = \delta_{q,j'}$ we get

$$S_{n,m} = e^{\lambda b_n}\left(\delta_{n,m} + \sum_{q\geq 1}\frac{\lambda^q}{q!}\sum_{l_1=1}^{N}\sum_{l_2=1}^{N}\cdots\sum_{l_{q-1}=1}^{N} a_{n,l_1}\cdots a_{l_{q-1},m}\right) \qquad \text{(A.21)}$$

If $B_{n,m} = b\delta_{n,m}$, i.e. a diagonal matrix with the same entry for every element along the diagonal, the exponential term is independent of the combination of $l_1 \ldots l_{q-1}$ and we get

$$\begin{aligned}
S_{n,m} &= e^{\lambda b}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}\left(A^i\right)_{n,m}\\
\Rightarrow S &= e^{\lambda B}e^{\lambda A}
\end{aligned}$$

This is simply the result for two commuting matrices A and B, where $e^{\lambda(B+A)} = e^{\lambda B}e^{\lambda A}$, as one would expect, because A commutes with any diagonal matrix that has a constant value along the diagonal.

3. Convergence

The convergence of expression (A.20) strongly depends on the form of the matrices A and B and the value of the parameter λ and must therefore be examined for the particular application of interest. Since there exist many potential applications of the matrix potential we will design a very abstract test case with a form similar to that of the solution to the multiple scattering Schrödinger equation. We will choose $N = 5$, $a_{n,m} = exp(-|n-m|) - \delta_{n,m}$, $b_n = [n - (N+1)/2]^2$, for $n, m = 1, 2, \ldots N$, and $\lambda = i$.

$$A = \begin{pmatrix} 0 & 0.367 & 0.135 & 0.049 & 0.018 \\ 0.367 & 0 & 0.367 & 0.135 & 0.049 \\ 0.135 & 0.367 & 0 & 0.367 & 0.135 \\ 0.049 & 0.135 & 0.367 & 0 & 0.367 \\ 0.018 & 0.049 & 0.135 & 0.367 & 0 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}$$

We will compare the convergence properties of the moduli of

$$S(\alpha) = \exp(\lambda[\alpha A + B]) \tag{A.22}$$

calculated according to expression (A.20) with that calculated according to (A.3), which is the fundamental expression defining the matrix exponential. The parameter α is used to scale the matrices with respect to each other. Our figure of merit will be

$$\chi(\alpha, p) = \sum_{n=1}^{N} \sum_{m=1}^{N} \left(|S_{n,m}^{(exact)}| - |S_{n,m}^{(approx)}(\alpha, p)| \right)^2 \tag{A.23}$$

where p indicates the highest order term used in the expansions (A.3) and (A.20).

Figure 77 shows $\chi(\alpha, p)$ vs. p, the order of expansion for the two expressions. Expression (A.20) converges always faster than (A.3), because we are performing the infinite sum over all products of elements of B associated with a given product of elements of A, and not only finite sized products of order $\leq p$, as in expansion (A.3). If the largest elements of A and B are comparable in size, as for the case $\alpha = 10$, both expansions will converge at almost equal rates. However, in the case where all the elements of A are smaller than those of B, the much higher convergence rate becomes obvious.

4. Inversion of CBED Patterns

The S-matrix notation is a fundamental mathematical tool in quantum scattering theory. For a review of inverse problems in quantum scattering the reader is referred to [178],

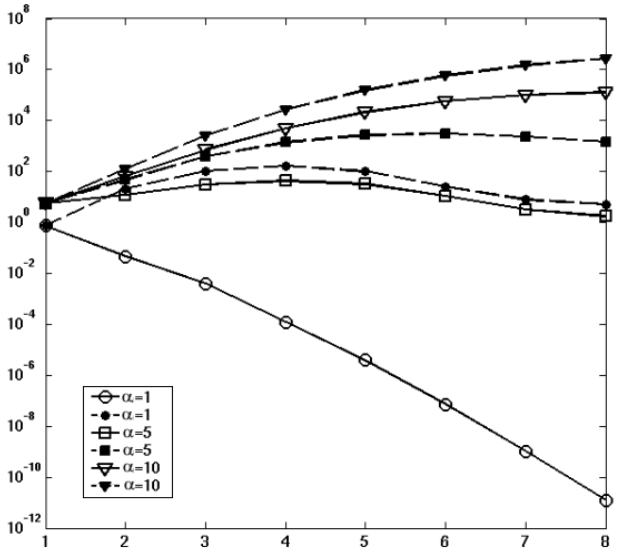

Figure 77. Comparison of convergence of expressions (A.3) (dashed lines, filled and small markers), and (A.20) (solid lines, open and large markers) with increasing number of terms in the expansion for different values of α on a logarithmic scale. For $\alpha = 10$ the largest elements in αA and B are comparable in size, producing very similar convergence behavior. If the elements of αA are small compared to those in B, expansion (A.20) converges much faster.

where its theory, application and inverse problems are discussed in detail. In high energy electron scattering the condition that large elements of B are larger than those of A is usually satisfied, making this problem a practical application of this expansion [39, 177, 179].

Based on the Bloch wave solution to the multiple scattering Schrödinger equation of high energy scattering of electrons by the crystal potential [36,180] the intensity of any point in an electron diffraction pattern defined by the reciprocal lattice vector $\vec{g}_n$ of the particular diffraction spot and the tangential component of the incident electron wavevector $\vec{k}_t$ is given by the modulus squared of the element of the scattering matrix $S(\vec{k}_t)$ in the n^{th} row and the column m of the $G(\vec{k}_t)$ matrix that contains the kinetic energy term for the $\vec{g}_m = 0$ ("central beam") reciprocal lattice vector.

$$|S(\vec{k}_t)_{n,m}|^2 = [e^{iT(A+G(\vec{k}_t))} \cdot e^{-iT(A^*+G(\vec{k}_t))}]_{n,m} \tag{A.24}$$

where A is a square matrix with the potential energy terms $A_{n,m} = U_{\vec{g}_n - \vec{g}_m}$ as its off-diagonal elements and zeros along its diagonal, G is a diagonal matrix with (relativistically corrected) terms $G_{n,n} = \xi_n = -(|\vec{g}_n|^2 + \vec{g}_n \cdot \vec{k}_t)/\gamma$ related to the kinetic energy part of the Schrödinger equation along its diagonal, and $T = \pi\gamma\lambda t$, where λ is the incident electron wavelength corrected by the mean potential U_0 of the crystal, t the thickness of the sample, and γ the relativistic correction factor (see table 1 for definitions). Note that the electron structure factors $U_{\vec{g}_n - \vec{g}_m}$ in A do not contain the usual relativistic correction factor γ, which has instead been included included in T and G instead, in order to separate true material constants ($U_{\vec{g}}$) from variable experimental parameters ($\lambda(v), \gamma(v), t, \vec{k}_t$). We also ommit the additive constants U_0 and $|k_t|^2$ for every element along the diagonal of G, because it only produces a general attenuation of the diffraction pattern due to the imaginary part of U_0 and a phase factor due to the real part of U_0 and $|k_t|^2$, which is immeasurable in a conventional diffraction CBED experiment with non-overlapping discs. The crystal potential $V(\vec{r}) = V^r(\vec{r}) + iV^i(\vec{r})$ consists of a real part describing the elastic scattering and an imaginary part accounting for inelastic scattering processes which produce a non-isotropic attenuation in the scattered signal, if one collects the zero-loss scattered electrons only, as is routinely done in today's quantitative (energy filtered) electron diffraction experiments. This means that even for centro-symmetric crystals the $U_{\vec{g}}$ are in general complex. For the symmetric 5-beam case with refelctions $\vec{g}_1$, $\vec{g}_2$, $\vec{g}_3 = 0$, $-\vec{g}_1$, and $-\vec{g}_2$ the A-matrix looks as follows:

$$A = \begin{pmatrix} 0 & U_{-\vec{g}_1+\vec{g}_2} & U_{\vec{g}_2} & U_{\vec{g}_1+\vec{g}_2} & U_{2\vec{g}_2} \\ U_{\vec{g}_1-\vec{g}_2} & 0 & U_{\vec{g}_1} & U_{2\vec{g}_1} & U_{\vec{g}_1+\vec{g}_2} \\ U_{-\vec{g}_2} & U_{-\vec{g}_1} & 0 & U_{\vec{g}_1} & U_{\vec{g}_2} \\ U_{-\vec{g}_1-\vec{g}_2} & U_{-2\vec{g}_1} & U_{-\vec{g}_1} & 0 & U_{-\vec{g}_1+\vec{g}_2} \\ U_{-2\vec{g}_2} & U_{-\vec{g}_1-\vec{g}_2} & U_{-\vec{g}_2} & U_{\vec{g}_1-\vec{g}_2} & 0 \end{pmatrix}$$

and

$$G(\vec{k}_t) = \begin{pmatrix} -|\vec{g}_2|^2 - \vec{g}_2\cdot\vec{k}_t & 0 & 0 & 0 & 0 \\ 0 & -|\vec{g}_1|^2 - \vec{g}_1\cdot\vec{k}_t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -|\vec{g}_1|^2 + \vec{g}_1\cdot\vec{k}_t & 0 \\ 0 & 0 & 0 & 0 & -|\vec{g}_2|^2 + \vec{g}_2\cdot\vec{k}_t \end{pmatrix}$$

Using the expansion for the matrix exponential shown in the previous sections the scattering matrix S can be expanded in a Born series-like expansion as

$$S_{n,m} = e^{T\xi_n}\delta_{n,m} + \sum_{q=1}^{\infty}\sum_{l_1,l_2,\ldots l_{q-1}=0}^{N} \underbrace{U_{\vec{g}_n-\vec{g}_{l_1}}U_{\vec{g}_{l_1}-\vec{g}_{l_2}}\cdots U_{\vec{g}_{l_{q-1}}-\vec{g}_m}}_{q} C^q_{n,l_1,\ldots l_{q-1},m} \tag{A.25}$$

where

$$C^q_{n,l_1,\ldots l_{q-1},m} = \sum_{k=0}^{u}\sum_{j=0}^{d_k}\frac{(iT)^j}{j!}D^{(j,k,q)}_{n,l_1,\ldots l_{q-1},m}\left[e^{iT\xi'_{l_k}} - \sum_{s=0}^{q-j-1}\frac{(iT\xi'_{l_k})^s}{s!}\right] \tag{A.26}$$

$$\begin{aligned} D^{(j,k,q)}_{l_0,l_1,\ldots l_q} &= (-1)^{d_k-j}\prod_{\substack{r=0\\ r\neq k}}^{u}\frac{1}{(\xi'_{l_k}-\xi'_{l_r})^{(d_r+1)}} \\ &\quad \underbrace{\sum_{\substack{r_1=0\\ \xi_{l_{r_1}}\neq\xi'_{l_k}}}^{q}\frac{1}{(\xi'_{l_k}-\xi_{l_{r_1}})}\sum_{\substack{r_2=0\\ \xi_{l_{r_2}}\neq\xi'_{l_k}}}^{r_1}\frac{1}{(\xi'_{l_k}-\xi_{l_{r_2}})}\cdots\sum_{\substack{r_{d_k-j}=0\\ \xi_{l_{r_{d_k-j}}}\neq\xi'_{l_k}}}^{r_{d_k-j-1}}\frac{1}{(\xi'_{l_k}-\xi_{l_{r_{d_k-j}}})}}_{d_k-j} \end{aligned}$$

where $\{\xi'_{l_0},\ldots\xi'_{l_u}\}$ is the subset of $u+1$ unique elements of the set $Q=\{\xi_n,\xi_{l_1},\ldots\ldots\xi_m\}$ of elements of $G(\vec{k}_t)$ present in that particular $C^q_{n,l_1,\ldots l_{q-1},m}$ coefficient and d_k is the degeneracy of each one of them. Again, we used the notation $l_0=n$ and $l_q=m$. For non-degenerate ξ_{l_k} this result is equivalent to the one obtained by Cowley and Moodie [39] in 1957. Fujiwara [179] also gives expressions for the degenerate case, but only for $q\leq 3$. However, in order to approximate the diffraction intensity of CBED patterns with any desired accuracy, being able to calculate higher order terms with any degeneracy is essential. Since $\xi_m=0$ the first few C^q-coefficients are

$$C^1_{n,m} = \begin{cases} \frac{e^{iT\xi_n}-1}{\xi_n} & \text{for } \xi_n\neq 0 \\ iT & \text{for } \xi_n = 0 \end{cases}$$

$$C^2_{n,l,m} = \begin{cases} \frac{e^{iT\xi_n}-1}{\xi_n(\xi_n-\xi_l)}+\frac{e^{iT\xi_l}-1}{\xi_l(\xi_l-\xi_n)} & \text{for } \xi_n\neq\xi_l \text{ and } \xi_n,\xi_l\neq 0 \\ e^{iT\xi_n}\left(\frac{iT}{\xi_n}-\frac{1}{\xi_n^2}\right)+\frac{1}{\xi_n^2} & \text{for } \xi_n=\xi_l\neq 0 \\ \frac{e^{iT\xi_n}-iT\xi_n-1}{\xi_n^2} & \text{for } \xi_n\neq\xi_l=0 \\ -\frac{T^2}{2} & \text{for } \xi_n=\xi_l=0 \end{cases}$$

From this it becomes obvious that there is only a single $C^q_{n\ldots m}$-coefficient for $q = 1$, while there are $N - 2$ such terms for $q = 2$ (coefficients with $l = n$ and $l = m$ are zero, because $A_{n,n} = 0$). In general the number of terms in the expansion for a given order q and matrix size N is

$$\sum_{j=0}^{q-1}(-1)^{q-j-1}(N-1)^j$$

However, the number of distinguishable terms is less, because this estimate does not consider the fact that for example

$$C^q_{n,l_1,l_2,m} = C^q_{n,l_2,l_1,m} \tag{A.27}$$

The convergence of the Born approximation (A.25) depends on the value of $T = \gamma\lambda t$ and the modulus of the largest structure factor. The thicker the specimen or the lower the accelerating voltage, the more terms we need in the expansion. The $C^q_{n,l,\ldots l_{q-1},m}$ coefficients depend on several variable experimental parameters, like thickness t, accelerating voltage v, and $\vec{k}_t$. Measuring diffraction intensities for a set of different values of any of these parameters will allow us to set up a set of linear equations and solve for the different products of $U_{\vec{g}_l-\vec{g}_{l_2}}$. For example, CBED patterns provide experimental data for a whole range of 2-dimensional $\vec{k}_t$-vectors, depending on the shape and size of the condensor aperture. This is a direct, though approximate inversion of the multiple scattering equation, potentially able to reconstruct the projected potential $V(\vec{r})$ of the scattering crystal and by application of the inverse Bethe-Mott formula its charge density $\rho(\vec{r})$ with a resolution comparable to that of the scattered electrons, which is just a few pm.

Since only the diffraction intensity can be measured directly, the modulus squared of (A.25) has to be taken

$$\begin{aligned}|S_{n,m}|^2 &= \delta_{n,m} + \sum_{q=2}^{\infty}\sum_{l,l_2,\ldots l_{q-2}=1}^{N}\sum_{p=1}^{q-1} U_{\vec{g}_n-\vec{g}_l}\cdots U_{\vec{g}_{l_{p-1}}-\vec{g}_m}U^*_{\vec{g}_n-\vec{g}_{l_p}}\cdots U^*_{\vec{g}_{l_{q-2}}-\vec{g}_m}\\ &\quad C^p_{n,l,\ldots l_{p-1},m}\left(C^{q-p}_{n,l_p,\ldots l_{q-2},m}\right)^* + \\ &\quad 2\mathrm{Re}\left(\sum_{q=1}^{\infty}\sum_{l,l_2,\ldots l_{q-1}=1}^{N} U_{\vec{g}_n-\vec{g}_l}U_{\vec{g}_l-\vec{g}_{l_2}}\cdots U_{\vec{g}_{l_{q-1}}-\vec{g}_m}C^q_{n,l,\ldots l_{q-1},m}\right)\delta_{n,m}\end{aligned} \tag{A.28}$$

For $\xi_n \neq \xi_m$ the first term in the expansion of $|S_{n,m}|^2$ (with $q = 2$) is

$$\left|S^{(2)}_{n,m}\right|^2 = \left|U_{\vec{g}_n-\vec{g}_m}\frac{e^{iT\xi_n}-1}{\xi_n}\right|^2 = |U_{\vec{g}_n}|^2\,\frac{\sin^2(\frac{1}{2}\pi\lambda t\xi_n)}{(\xi_n/2)^2},$$

which is the kinematic approximation for the scattering intensity. Higher order terms involve products of different structure factors. For example, for $\xi_n \neq 0$ the next higher order term

in $|S_{n,m}|^2$ is

$$
\begin{aligned}
|S^{(3)}_{n,m}|^2 &= \sum_{l=1}^{N} \begin{cases} 2\mathrm{Re}\left(U_{\vec{g}_n-\vec{g}_l} U_{\vec{g}_l} U^*_{\vec{g}_n} \left[\frac{e^{iT\xi_n}-1}{\xi_n(\xi_n-\xi_l)} + \frac{e^{iT\xi_l}-1}{\xi_l(\xi_l-\xi_n)}\right] \frac{e^{-iT\xi_n}-1}{\xi_n}\right) & \text{for } \xi_n \neq \xi_l \neq 0 \\ 2\mathrm{Re}\left(U_0 U_{\vec{g}_n} U^*_{\vec{g}_n} \left[e^{iT\xi_n}\left(\frac{iT}{\xi_n} - \frac{1}{\xi_n^2}\right) + \frac{1}{\xi_n^2}\right] \frac{e^{-iT\xi_n}-1}{\xi_n}\right) & \text{for } \xi_n = \xi_l \\ 2\mathrm{Re}\left(U_{\vec{g}_n-\vec{g}_l} U_0 U^*_{\vec{g}_n} \frac{e^{iT\xi_n}-iT\xi_n-1}{\xi_n^2} \frac{e^{-iT\xi_n}-1}{\xi_n}\right) & \text{for } \xi_l = 0 \end{cases} \\
&= \sum_{\substack{l=1 \\ l\neq m,n}}^{N} \mathrm{Re}\left(U_{\vec{g}_n-\vec{g}_l} U_{\vec{g}_l} U^*_{\vec{g}_n} \frac{2}{\xi_n(\xi_n-\xi_l)} \right. \\
&\quad \left. \left\{ \frac{\sin^2\left(\frac{T\xi_n}{2}\right)}{\xi_n} + \frac{e^{iT\xi_l} + e^{-iT\xi_n} - e^{iT[\xi_l-\xi_n]} - 1}{\xi_l} \right\} \right) \\
&= \sum_{\substack{l=1 \\ l\neq m,n}}^{N} \mathrm{Re}\left(U_{\vec{g}_n-\vec{g}_l} U_{\vec{g}_l} U^*_{\vec{g}_n}\right) \frac{2}{\xi_n(\xi_n-\xi_l)} \\
&\quad \left\{ \frac{\sin^2\left(\frac{T\xi_n}{2}\right)}{\xi_n} + \frac{\cos(T\xi_l) + \cos(T\xi_n) - \cos(T[\xi_l-\xi_n]) - 1}{\xi_l} \right\} + \\
&\quad \mathrm{Im}\left(U_{\vec{g}_n-\vec{g}_l} U_{\vec{g}_l} U^*_{\vec{g}_n}\right) \frac{2}{\xi_n(\xi_n-\xi_l)} \frac{\sin(T\xi_l) - \sin(T\xi_n) - \sin(T[\xi_l-\xi_n])}{\xi_l} \qquad \text{(A.29)}
\end{aligned}
$$

The terms with $l = n$ and $l = m$ are zero, because we set $U_0 = 0$. Here we assumed that always $\xi_l \neq \xi_n$, if $l \neq m$, which is only the case for non-symmetric orientations, i.e. not the zone-axis case. However, in the computational implementation the proper "branch" in expression (A.29) can be chosen, based on the numerical degeneracy of the $\xi_l \propto -|\vec{g}_l|^2 - \vec{g}_l \cdot \vec{k}_t$, so that this expansion works for any value of $\vec{k}_t$.

In order to use experimental CBED data for this inversion scheme, the proper $\vec{k}_t$-vector has to be assigned to every pixel in the pattern, which is easily possible, if the microscope and its recording system is properly calibrated. The intensity of every pixel in this pattern can be descibed by (A.28), which, when split into real and imaginary part, as done in expression (A.29), becomes a set of linear equations with the constant coefficients $c_r = \mathrm{Re}(U_{\vec{g}_n-\vec{g}_{l_1}} \dots U_{\vec{g}_{l_{p-1}}} U^*_{\vec{g}_n-\vec{g}_{l_p}} \dots U^*_{\vec{g}_{l_{q-2}}})$ and $c_i = \mathrm{Im}(U_{\vec{g}_n-\vec{g}_{l_1}} \dots U_{\vec{g}_{l_{p-1}}} U^*_{\vec{g}_n-\vec{g}_{l_p}} \dots U^*_{\vec{g}_{l_{q-2}}})$. The amplitude of this product of structure factors is then given by $\sqrt{c_r^2 + c_i^2}$, and its phase by $\mathrm{atan}(c_i/c_r)$. Dividing one such product of $U_{\vec{g}}$'s by another can yield the complex values of single structure factors, e.g. $U_{\vec{g}_1} = (U_{\vec{g}_1} U_{\vec{g}_2} U^*_{\vec{g}_3})/(U_{\vec{g}_2} U^*_{\vec{g}_3})$, however, when implementing this method, it turns out that only products of structure factors with unit cell origin independent phases can be determined by this method. In order to fix the origin of the unit cell, the phases of certain structure factors must be defined "by hand". If the moduli of the structure factors are already known (e.g. by proper normalization of kinematic diffraction patterns of very thin specimen), the linear equations can also be set up to solve for the structure factor phase-invariants only. The fact that the moduli of pure phase factors are 1 can then be used as an additional, though non-linear constraint.

The maximum value of q in (A.28) depends on the accuracy, with which the data shall be fitted, and the sample thickness, accelerating voltage, and scattering strength of the material. Once a set of $U_{\vec{g}}$ is determined, the scattering matrix $S(\vec{k}_t)$ can be calculated exactly, using the full matrix exponential (A.24). A comparison of the exact calculation with the experimental data shows, whether the maximum value of q must be increased. Once a close enough estimate of the $U_{\vec{g}}$'s has been found, their values can be refined using expression (A.24) and standard methods for solving non-linear equations.

Although expression (A.25) converges faster than the standard expansion (A.3) for the matrix exponential, even for reasonably thin specimen many terms in the expansion are needed. Since the intensity distribution across a single CBED disc at low thickness varies only slightly, an increase in range for $\vec{k}_t$ is desireable. The angular range for conventional CBED patterns is limited by the size of the unit cell of the crystal being examined ($4 \ldots 7$mrad for most structures at 200kV), and can be insufficient. In section 2.2 the "Tanaka"-method, or large-angle CBED (LACBED) method has been reviewed, which allows any diffraction disc to be recorded individually with an angular range in $\vec{k}_t$ up to 80mrad, or even more, as far as the microscope geometry allows. While the system of linear equations produced by expression (A.28) may be singular for the limited range in $\vec{k}_t$ of conventional CBED patterns, the additional range in $\vec{k}_t$ parameter space removes at least some of those (numerical) degeneracies in the linear equations.

5. Conclusion and Future Work

A new expansion for the matrix exponential of two non-commuting matrices, one of which is diagonal, has been derived. Many problems involving the solution of the time-dependent or time independent Schrödinger equation can be expressed in the form of such matrix exponentials by writing the potential and kinetic energy terms in separate matrices. It has been shown how this expansion can be applied to the multiple scattering inversion problem by expanding the set of non-linear equations with a finite system of linear equations which can be solved by matrix inversion or other standard linear equation solvers. In particular, the inversion from 2-dimensional rocking curves, which can be recorded as (LA)CBED patterns with a single exposure in most TEMs to the projected crystal potential has been discussed.

An experimental computer program has been written in order to test the suitability of this expansion for the solution of the multiple scattering inversion problem, providing very promising first results. More work will need to be done in exploring the analytic properties of this expansion, making its numerical implementation more efficient.

APPENDIX B
ATOMIC STRING HOLOGRAPHY

A new diffraction-channelling effect has been discovered, in which Kikuchi or channelling line-patterns formed by high energy electrons, neutrons and positrons are shown to break up into a series of annular disks if the crystal thickness traversed by the beam is small. The disks may be interpreted as Gabor in-line holograms of strings of atoms projected along the beam path. For electrons or positrons the patterns may be detected with little background by detecting characteristic X-ray emission from a thin film as a function of the diffraction conditions of a collimated, ionizing, high energy beam. Uses of the effect for structure determination and atomic-resolution lensless imaging are suggested, and experimental conditions discussed. The content of this chapter has been published by John Spence and Christoph Koch in *Physical Review Letters* **86** (24), p. 5510 (2001).

A considerable scientific payoff could be expected from the development of a general method of obtaining near-field images from the intensity of scattered radiation. By avoiding the aberrations and resolution limits introduced by lenses, and by allowing imaging with radiations for which no lenses exist, such a method could be expected to contribute greatly to our understanding of phenomena as diverse as protein folding, nano and mesoporous structure analysis, catalysis and nucleation and growth processes in surface science. The use of new radiations, such as coherent atom beams and neutrons, for imaging [181] may then open up new possibilities for minimizing radiation damage to biomolecules [182]. The associated inversion problem from scattering to image was therefore attacked by many of the leading scientists of the last century, from Rutherford and Gabor to those who solved the phase problem of X-ray crystallography by direct and other methods. Recently a variety of schemes which allow a holographic interpretation of fluorescent X-rays (or photoelectrons) have appeared [183]. These depend on the use of atoms in similar local environments as internal "point" emitters (or detectors) of radiation, each producing identical in-line holograms of the local environment on a distant detector [184, 185]. Translational symmetry between source or detector atoms is not required. These schemes may be classified according to the type of radiation and beam energy used, which determine the elastic and inelastic scattering lengths and hence the surface or bulk sensitivity. Interpretation is simplified, and intensity maximized, if either the incident or emitted radiation is non-diffracting - this may be achieved using a large illumination/detector angle, a wide range of energies, or a layer of material thinner than the elastic scattering length. The possibilities may be written, eg $(\gamma, \underline{e})$, with incident particle first and diffracting particle underlined. Application of reciprocity to the diffracting channel reverses the order of symbols and underline. Of the resulting eighteen possibilities (including neutrons), several remain untried, while the background of non-diffracting processes and kinematic constraints otherwise limit possibilities. In related techniques such as Holographic LEED [186] and the Heavy Atom Method of X-ray crystallography, there is negligible energy change on scattering. Twin images (see DeVelis et al. [187] for the first solution to this problem) and multiple scattering are minimized by integration over beam energy [183, 185], however, for electrons, variations in the phase of the reference wave with angle, and forward-focussing may introduce artifacts (but see [188]). X-ray methods do not suffer from these difficulties, but scatter weakly, however

the $(\underline{\gamma},\gamma)$ (IXFH) method, using a pink beam, greatly reduces recording time. The relevant history of ideas includes the Borrman effect, [189] and the structural information it contains [190] in addition to the theory of Kossel and Kikuchi lines [191]. Gabor's in-line holography is also relevant, however it was intended for systems without symmetry. The first observation of the dependance of fluorescence on diffraction conditions may be that of Knowles [192], using diffracting neutrons $(\underline{n},\gamma)$. Holographic reconstruction from internal sources was demonstrated with gas diffraction by Bartell [193]. Atomic resolution has recently been achieved using external-source, off-axis electron holography [194], in fulfillment of Gabor's original aim, to correct the aberrations of an electron microscope holographically.

The purpose of this letter is to describe a new $(\underline{e},\gamma)$ holographic diffraction method which uses high energy (300 kV) electrons traversing a thin crystalline slab to produce lensless, three-dimensional imaging of the internal atomic structure of the film. Characteristic X-rays are detected as a function of the angle of a coherent electron beam, which excites the X-rays and is weakly refracted by atomic strings in the crystal. The X-rays are not diffracted due to the small film thickness.

Figure 78a) shows our geometry. A thin crystal, about 10 nm thick, is used, as for atomic-resolution TEM . A high energy (e.g. 300 keV) collimated electron beam traverses the crystal slab, exciting X-ray fluorescence from the Cl atoms of the bcc CsCl structure shown. The thickness is less than an extinction distance for Bragg scattering, so that electron standing waves do not build up inside the crystal to modulate fluorescence, as in the "Alchemi" technique used to locate foreign atoms in thin crystals [195]. The X-ray intensity emitted by atom D is proportional to the intensity I_D of the diffracted electron intensity at D due to the external source S_1 [196], if localization effects are negligible (X-ray energies above about 500 eV [197]). Applying reciprocity to this arrangement, we see in figure 78b) that I_D equals the intensity at infinity from a point-source of 300 kV electrons within the sample at D. This intensity variation with angle is an in-line hologram of atom C as formed by a point source at D with "defocus" (source to sample distance) DC, shown in figure 78a). There is one hologram for every atomic column or string along the optical path between source and detector. Every string parallel to the one shown (eg at E in figure 78a)) generates an identical hologram, whose intensity distributions must be added together, since the X-ray "detection" process is incoherent. For sufficiently thin films the individual holograms from non-parallel atomic columns are spatially separated on the detector, since the angles between strings are relatively large. Where a mixture of species occurs along such a path, a sum must be taken over all the reciprocal "sources" in figure 78b) defined by the X-ray energy, and over the atoms imaged along the path. Unlike low energy electron holography, the electron scattering process is highly forward-peaked, with a strong concentration of energy along atomic strings. For thicknesses t small enough to avoid overlap of holograms, the atoms on a string act as phase objects, and the contrast at the detector arises from the "out-of-focus" propagation over the distance $\Delta z = DC$. This distance creates contrast in the image of the phase-object atom at C, which appears as a variation in X-ray emission with angle. For thicker samples, progressing from two-

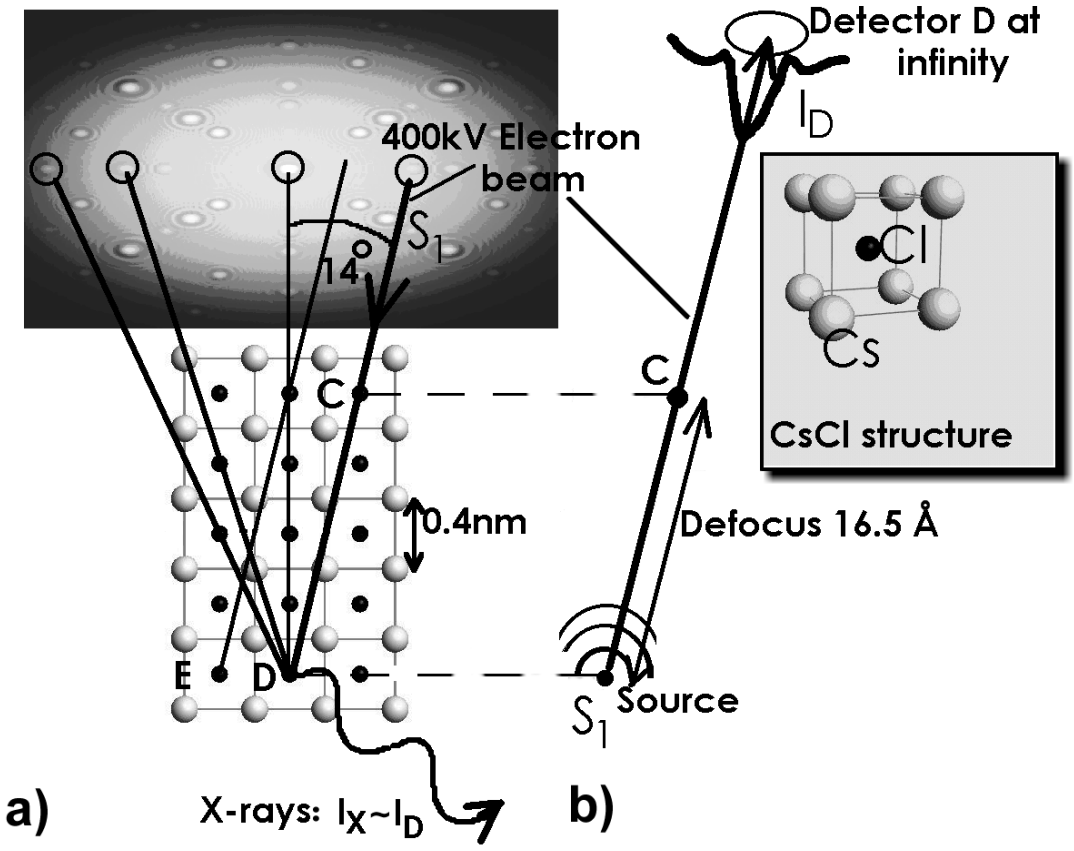

Figure 78. Atomic-string holography at high electron energies. a) Incident beam running along atomic columns excites fluorescence, whose total yield is mapped above as a function of incident beam direction. Atomic columns corresponding to spots in the pattern can easily be identified. b) Time-reversed point source of high energy electrons produces similar pattern. Inset shows CsCl bcc structure

dimensional to three-dimensional diffraction and associated momentum conservation laws, the holograms overlap and eventually form a Kikuchi pattern. In thin samples the method generates the holograms of atomic strings which would be produced by a 300 kV internal point-source of electrons, on a distant detector. This energy is sufficiently high to ensure that there are no multiple scattering artifacts. The wide angular range of illumination directions possible under computer control in TEM provides projections from many angles, overcoming the limitation of off-axis electron holography to two-dimensional reconstruction [194]. Because HEED is strongly forward-scattered, reconstruction from a single hologram cannot distinguish different planes along the beam path, as in the Barton algorithm, due to the rapid fall-off in angle for HEED scattering factors.

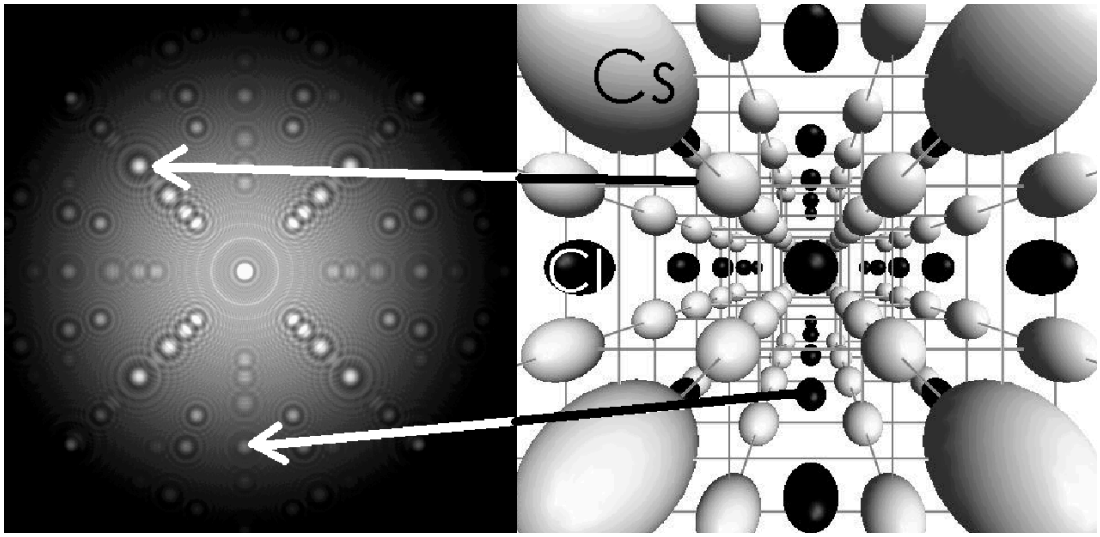

Figure 79. Diagram explaining the interpretation of atomic string holograms. Left: Multiple scattering produced by point sources of 1MeV electrons on Cl atoms in a thin film of CsCl (7 monolayers thick). Each blob in the figure lies over an atomic position in the perspective diagram of the crystal structure shown at right. The source atom lies above the plane of the paper and is not shown

To test these ideas we have performed detailed multiple-scattering calculations for 1MeV electrons traversing a thin crystal of CsCl . The Cl X-ray fluorescence which results is taken to be proportional to the intensity of the electron beam wavefield on the Cl site (all such sites are equivalent). For computational convenience, we use the reciprocal arrangement, and launch a spherical wave within the crystal from one Cl atom. A relativistically corrected one-electron Schroedinger equation is solved for the propagation of this wave through the crystal, using the multislice superlattice method [198]. The scattering potential is synthesized from relativistic Hartree-Fock calculations for atoms, and all orders of multiple scattering are included, with Debye-Waller factors for room temperature. Figure

79 shows the resulting elastic scattering for 1MeV electrons on a screen at infinity, plotted as a function of angle, and a model of the CsCl crystal as seen by the source atom (atom C in figure 78), which is at the center of the image, above the plane of the page. The intensity in the hologram is proportional to the total Cl characteristic X-ray emission, as a function of the incident beam direction. We note that the number of atomic holograms along each radial line is just equal to the number of atoms seen in perspective. The figure confirms that separate holograms are obtained for every atomic string, in the form of a central maximum surrounded by a Fresnel fringe. These holograms are seen to fall in one-to-one angular correspondance with rays drawn along the prominent atomic columns or strings in the structure from the source atom, as shown in figure 78.

A simple interpretation of these holograms is possible if the approximation is made that the potential is zero between the emitter atom and the imaged atom. Then, for atoms spaced by Δz along the string, each atomic string hologram can be shown to be identical (apart from magnification) to that which would be produced on a screen placed a distance Δz beyond an atom illuminated by a plane-wave [198]. (This follows from the equivalence of the point-projection and plane-wave geometries). In this sense the hologram is an image, out-of-focus by Δz. Hence these holograms have the form

$$I(r) = |q(r) * *t(r)|^2$$

where $q(r) = \exp(-i\sigma\phi_p(r))$, $\phi_p(r) = \int \phi(r)dz$ and $t(r) = exp(-i\pi r^2/(\Delta z \cdot \lambda))$ is the free-space propagator or Green's function. The $**$ denotes convolution, $\phi(r)$ is the electrostatic or Coulomb potential for the atom, and $\sigma = \pi/\lambda V$ is the interaction constant for an electron beam of energy eV. The integral is taken antiparallel to the beam path from an origin on the X-ray emitting atom. Since these are not Fraunhoffer holograms ($\Delta z < d^2/\lambda$, d the diameter of the atom [187]), reconstruction of the holograms would encounter a twin image problem, which is soluble in principle by varying the energy of the beam over a range which changes the Fresnel number $N = r^2/(\Delta z \cdot \lambda \cdot 2)$ by 0.5. In practice, the direct resemblance of the patterns to a projection of the real-space crystal structure from an atom identified by its characteristic radiation will be sufficient to solve it. From thicker regions, conventional electron diffraction methods can be used to determine the periodicity along the atomic strings, using high-order Laue zone lines, the space group, and lattice spacings [198].

Since the spatial resolution of these images greatly exceeds that of current TEM imaging techniques, their sensitivity to atomic number and ionic state is of interest. Consistent with the first-order expansion of q(r), we find that neighbors in the atomic table can be distinguished in the absence of background, and that, for example, the intensity at the center of Cs (55) is three times that at the center of the Cl (17) hologram. Br (35) is about twice Cl. This axial intensity is the most sensitive point to the charge state of the ions, since it involves forward electron scattering, which tends to infinity for an unscreened isolated ion [198].

Figure 80 shows how the contrast of the string holograms is inverted for positrons as incident particles and how the patterns vary with increasing thickness, first to a projection

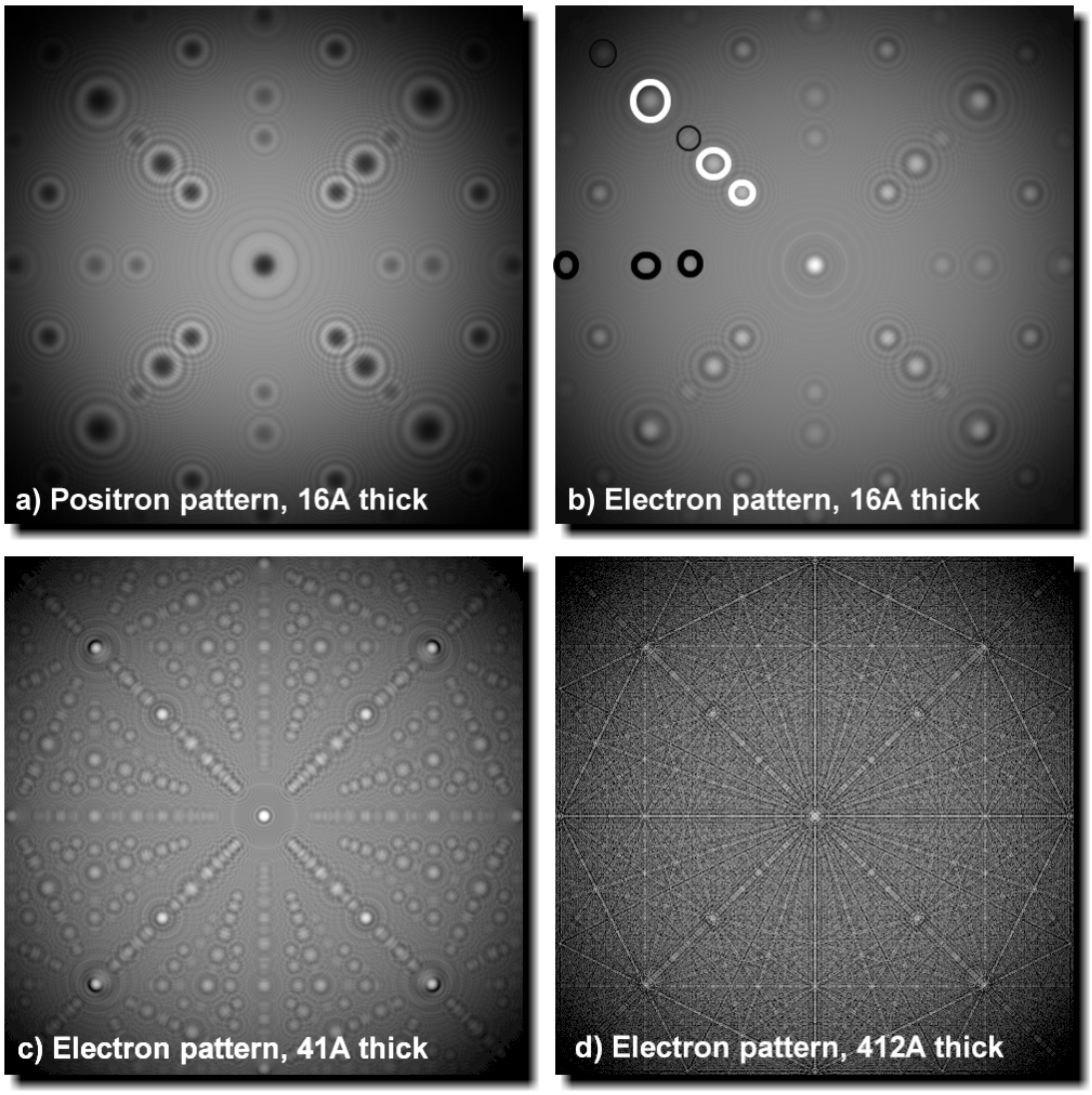

Figure 80. a) 1MeV positron hologram. b)-d) 1MeV electron holograms for samples of different thicknesses. The displayed intensity is the sum of holograms due to sources at every depth up to the sample thickness. In b) the different holograms of Cs and Cl atoms are indicated by white and black circles respectively (compare with figure 79 for position of all the holograms.

of the structure at moderate thickness, and finally to the familiar Kikuchi patterns at large thickness, due to localised phonon (quasi-elastic) scattering. The transition to a spotty pattern at small thickness was evidently overlooked by the discoverers of Kikuchi and Kossel patterns, which result from three-dimensional diffraction satisfying Bragg's law on the surfaces of cones [191]. In figure 80, however, both radial and non-radial K-lines remain. Radial lines are seen in channelling star patterns at much higher energies, due to a lack of angular resolution, but are not seen in the high angular resolution patterns obtained at sub MeV energies. Evidently the forward focussing effect, which produces the radial lines, becomes ineffective for longer strings, as the focussing distance gets out of step with the interatomic spacing [199]. These atomic focusing effects have also been observed at high energy [200]. Thus an atom in the path of a kilovolt electron beam acts as a lens, with a focal length of a few nm, depending on accelerating voltage. Hence special accelerating voltages may be chosen to enhance the intensity or vary the magnification of the string holograms.

The spatial resolution of the method depends on several factors, including the vibrational amplitude of the source atom Δr, the inelastic localisation $L \sim \lambda/\Theta_E$ [78, 201], and the divergence of the electron beam (which limits intensity). Here Θ_E is the mean inelastic scattering angle, and L is a quantum equivalent of an impact parameter. Since no static displacement is possible between source and imaged atom (as in FIM), mechanical vibration does not limit resolution, as in other atomic-resolution microscopies. For X-ray energies greater than 1 kV, $L < 0.1$ Angstrom. By reciprocity, the effect of a finite electron beam divergence $\Delta\theta$ is to smear out detail in the hologram over this angular range. Typical divergences of milliradians are very much smaller than the width of the inner Fresnel fringes. The resolution in the final reconstruction is about equal to the width of the finest Fresnel fringe observed. The angular smearing due to thermal vibration is $\Delta r/2\Delta z \approx 2.5$ mRad at room temperature, which sets an upper limit on beam divergence and hence exposure time for hard X-rays. Cooling thus increases resolution only if divergence is reduced commensurately, and exposure time increased.

Experimental observation of atomic string holograms requires the same TEM apparatus used for "Alchemi", but thinner crystals. Using a 200nm diam. electron beam (LaB6 source), with $\Delta\theta = 1.8$ mRad, in samples 200 nm thick, Rossouw et al [202] collected statistically significant standing-wave fluorescence (Alchemi) data on a 79 X 58 pixel array in 2 hrs. (0.3 sterad X-ray detector, 30 mm^2). For the 10nm thick films required for string holography, this time becomes 5 hrs using four of the new 60mm^2 detectors, or 13 mins for one hologram. Current TEM/STEM instruments are not optimized for these experiments and allow many improvements.

In summary, we have discovered a new channelling effect for very thin crystals, overlooked within previous theories of Kikuchi lines. The patterns, interpreted as holograms, solve the multiple scattering problems of earlier methods, while ensuring adequate signal, and providing true three-dimensional information. For bulk samples, the possibility arises of applying this method to a beam whose energy is chosen slightly above the X-ray ion-

ization threshold, thereby limiting the information depth to the inelastic mean free path. This requires the use of a localised excitation with short (< 10nm) inelastic mean free path. The coherent probe formed over very large angles by an aberration-corrected STEM may also be expected to form these patterns. A two-dimensional array of orientationally ordered organic molecules [203] containing a single fluorescing species, or a semiconductor overlayer would also produce string holograms. Similar string holograms can be expected from incoherent neutron Kikuchi patterns [181, 204] in thin samples, since the scattering is isotropic. Radiation damage effects on biological samples can then be small [182].

Thanks to J.M.Cowley and M. van Hove for discussions. Supported by NSF award DMR9814055.

APPENDIX C
GENERATING DISSOCIATED DISLOCATIONS

This chapter contains a copy of the letter written by João Francisco Justo Filho at the Instituto de Fisica da Universidade de São Paulo, Brazil, describing the method he used to generate the dislocation model used for image simulations in chapter 5. The reference and image numbers have been adjusted to fit in with the rest of this document.

Dear Christoph,

Here I am sending you a first description of my calculations. I have already generated a few configurations. It was a little trick to get the dissociated dislocation, but I have full control over the configurations. Now we should discuss the number of atoms and the number of layers in the dislocation line so that your simulations are not too expensive.

Methodology: Let's consider a configuration which I usually generate. Since dislocations in silicon belong to the {111} glide planes, in the ⟨110⟩ directions, my simulation cells should take advantage of this. The simulations were performed using periodically replicated super-cells, as shown in figure 81, with X, Y , and Z axes respectively parallel to the $[11\bar{2}]$, [111], and $[\bar{1}10]$ directions of the zinc-blende lattice. After generating the configuration, I relax the atomic positions using a conjugated gradient (CG) algorithm based on an interatomic potential.

I use periodic boundary conditions in my calculations to simulate an infinite crystalline system. Therefore, my simulation cells must have a dislocation dipole. For example, when studying a 30° partial dislocation, we must generate a 30° and a -30° partial with a stacking fault between them (as shown in figure 81). To study a 60° dislocation the simulation is a little more complicated.

I generated a dipole of two 60° dislocations, and then split one of them into a 30° plus a 90° partial, with a (40 Å) stacking fault between the partials. I first generated a 60° dislocation dipole, as shown in figure 82. This was the tricky part. Before dissociating one of the 60° dislocations, I relaxed the system (fig. 82) with my conjugated gradient algorithm. Then, I dissociated one of the dislocations (fig. 83), and relaxed the system again with that CG algorithm. This multiple-step procedure was really necessary, otherwise the system would relax to some ẅeirdc̈onfiguration. This dissociated dislocation has a stacking fault 40 Åwide, as you requested. Then, to perform your simulations I guess we should take only one part of the simulation cell, and remove the 60° undissociated dislocation (see the dashed line in the figure 84).

Now some details of my calculations:

I used a conjugated gradient minimization code to relax the system to an equilibrium configurations. The minimization is based on an interatomic potential which I have recently worked on. This interatomic potential is called EDIP [158, 205], and it has been shown to give a better description of dislocation core properties as compared to either Tersoff or Stillinger-Weber potentials.

The minimization is also performed in the parameters of the simulation cell [206].

This means that the lattice vectors are also allowed to in order to remove the stress generated by the dislocations. Therefore, the relaxed simulation cell is not orthorhombic anymore, it deforms a little to remove the internal stress, and the final cell is triclinic. I mention that because if you decide to replicate the cell in the dislocation direction (here in Z direction), then you should be careful with this detail. The relaxed vectors which define the lattice are not orthogonal.

The simulation cell has the following dimensions:

I generated configurations with 8 b in the dislocation line direction (Z direction), where $b = 3.84$Å.

$*DX \approx 80$Å, $[11\bar{2}]$ (the partial dislocations are 40 Åapart in the X direction)
(and the partial dislocations are at least 40 Åapart from the other undissociated 60° dislocation, in the X direction)

$*DY \approx 28$Å[111] There are 18 planes in the [111] direction.

$*DZ \approx 31$Å, $8b$, $[\bar{1}10]$ (you mentioned that you would prefer a narrow cell and then you would just replicate that. As I mentioned earlier, you should be careful here).

I generated two dissociated configurations (30° + 90°): for the 90° in a single period and double period reconstructions. (see figures 85 and 86, respectively). We can look to configuration in the Y plane in fig. 87, $(\bar{1}10)$ plane.

I will not send you the configurations just yet. I want your feedback first, so that I can send you exactly the configurations you want. Now that I overcame all the problems, I can generate those configurations in less than a day. Please let me know if you need any additional information concerning my simulations.

Best regards.
João Justo.

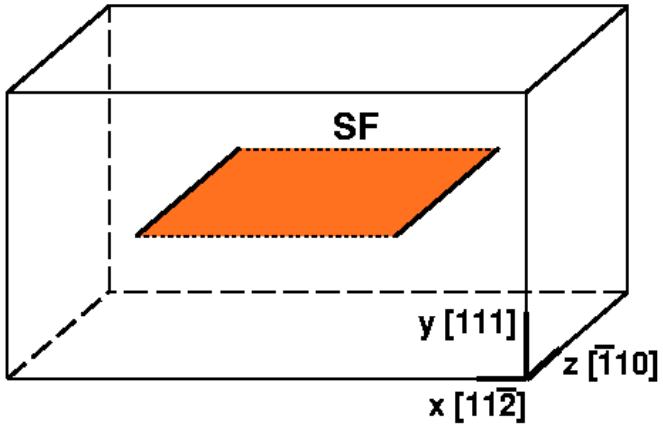

Figure 81. Simulation cell used in the simulations. The shadowed region represents the stacking fault resulting from the introduction of a dipole of partial dislocations.

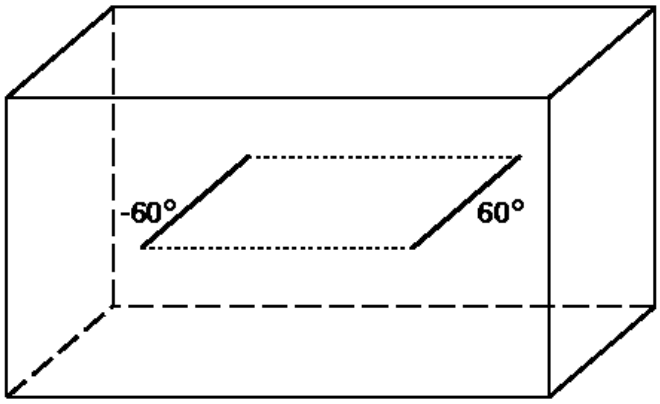

Figure 82. A dipole with two 60° dislocations.

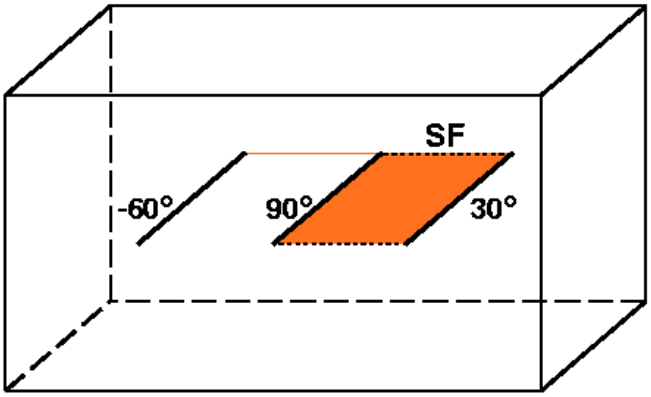

Figure 83. Simulation cell with one of the 60 degree dislocations dissociated into a 30 degree plus a 90 degree partial.

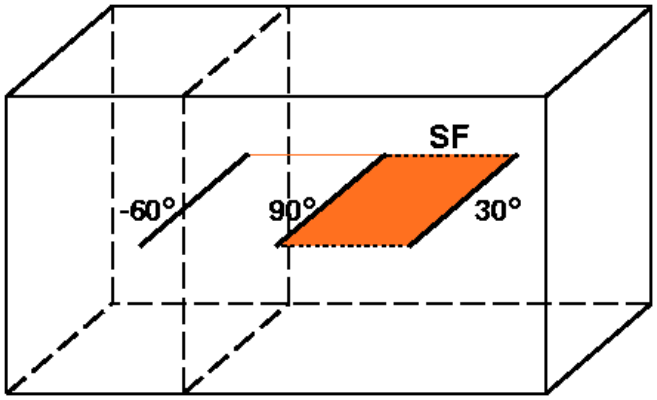

Figure 84. Whole simulation cell showing the part that can be removed (left of the dashed line) which has an undissociated 60° dislocation.

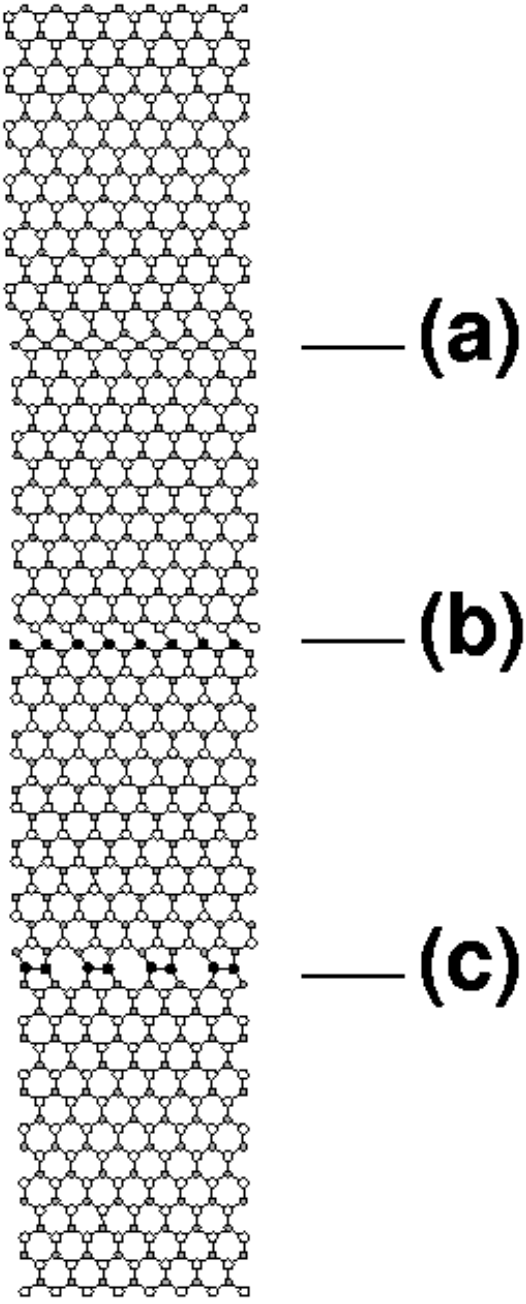

Figure 85. The {111} glide plane where the dislocations belong. Here (a) represents the core of the undissociated 60° dislocation, which we do not need to bother. (b) is the core of the 90° partial, and (c) is the core of the 30° partial. The 90° partial is in its single-period core reconstruction.

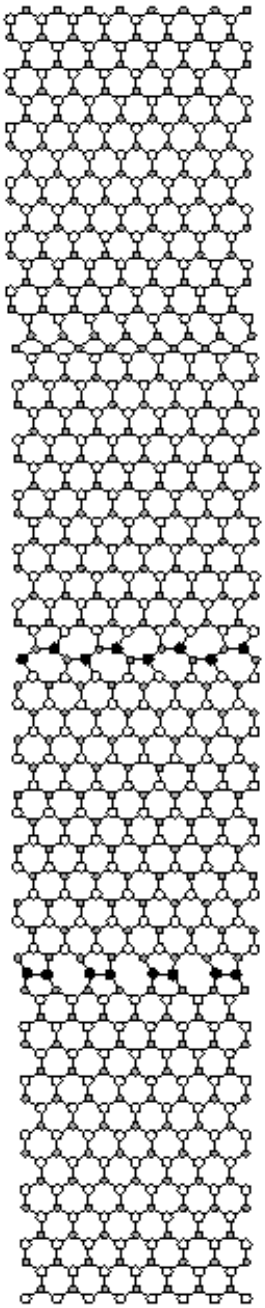

Figure 86. Double-period core reconstruction for the 90 degree partial.

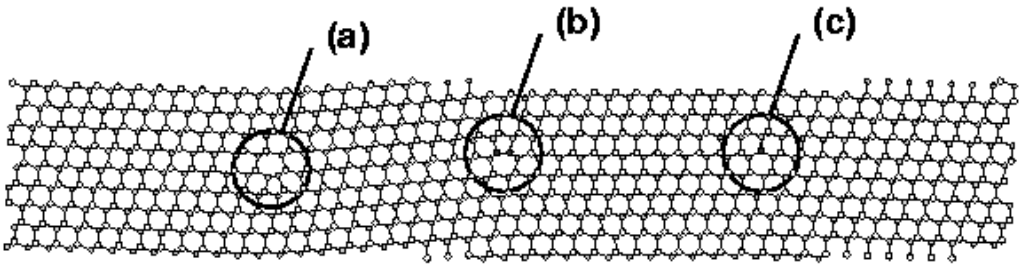

Figure 87. The $(\bar{1}10)$ plane, or Y plane. It is the same configuration as in figure 85. Here (a) represents the core of the undissociated 60° dislocation. (b) is the core of the 90° partial, and (c) is the core of the 30° partial.

References

[1] H. Kolar, *Atomic Resolution In-Situ Dynamics of Dislocation Kinks in Silicon*, Ph.D. thesis, Arizona State University (1995).

[2] J. C. H. Spence, H. R. Kolar, and H. Alexander, J. Phys. III **7**, 2325 (1997).

[3] G. Hembree, C. Koch, and J. C. H. Spence, to be published (2002).

[4] P. Hirsch, A. Howie, R. Nicholson, D. W. Pashley, and M. J. Whelan, *Electron Microscopy of Thin Crystals* (Krieger, Malabar, 1977).

[5] J. C. H. Spence, *Experimental High-Resolution Electron Microscopy* (Oxford, New York, 1988).

[6] A. W. Agar, R. H. Alderson, and D. Chescoe, *Principles and Practice of Electron Microscope Operation* (Oxford, New York, 1988).

[7] S. L. Chang, Phys. Rev. Lett. **48**, 163 (1982).

[8] H. Juretschke, Phys. Rev. Lett. **48**, 1487 (1982).

[9] Q. Chen, Phys. Rev. Lett. **80**, 3268 (1997).

[10] D. M. Bird and M. Saunders, Ultramicr. **45**, 241 (1992).

[11] M. Saunders, A. G. Fox, and P. A. Midgley, Acta Cryst. A **55**, 480 (1999).

[12] W. Nüchter, A. L. Weickenmeier, and J. Mayer, Acta Cryst. A **54**, 147 (1998).

[13] J. M. Zuo and J. C. H. Spence, Ultramicr. **35**, 185 (1995).

[14] J. M. Zuo, M. Kim, M. O'Keefe, and J. C. H. Spence, Nature **401**, 49 (1999).

[15] R. Vincent and D. R. Exelby, Acta Cryst. A **51**, 801 (1995).

[16] Y. F. Cheng, W. Nüchter, J. Mayer, A. Weickenmeier, and J. Gjønnes, Acta Cryst. A **52**, 923 (1996).

[17] G. M. Rackham and J. A. Eades, Optik **47**, 227 (1977).

[18] J. W. Steeds, G. J. Tatlock, and J. Hampson, Nature **241**, 435 (1973).

[19] M. Brunner, H. J. Kohl, and N. Niedrig, Optik **58**, 37 (1981).

[20] J. A. Eades, Ultramicr. **5**, 71 (1980).

[21] M. Tanaka, R. Saito, X. Ueno, and Y. Harada, J. Electr. Microsc. **29**, 408 (1980).

[22] J. A. Eades, J. Electr. Microsc. Techn. **1**, 279 (1984).

[23] Y. Tomokiyo, T. Kuroiwa, and Y. Hayashi, Materials Transactions JIM **36**, 1344 (1995).

[24] D. Cherns and A. R. Preston, in *Proceedings of the Xith International Congress of Electron Microscopy*, edited by T. Imura (The Japanese Society of Electron Microscopy, Tokyo, 1986), p. 721.

[25] M. Tanaka, M. Terauchi, and T. Kaneyama, *Convergent-Beam Electron Diffraction II* (JEOL, Tokyo, 1988).

[26] A. Toda, N. Ikarashi, and H. Ono, Micron **203**, 239 (2001).

[27] F. Wu, A. Armigliato, R. Balboni, and S. Frabboni, Micron **31**, 211 (2000).

[28] J. M. Cowley, Appl. Phys. Lett. **15**, 58 (1969).

[29] E. Zeitler and M. G. R. Thomson, Optik **31**, 258 (1970).

[30] D. E. Jesson and S. J. Pennycook, Proc. Roy. Soc. (London) A **449**, 273 (1995).

[31] M. M. J. Treacy and J. M. Gibson, Ultramicr. **52**, 31 (1993).

[32] S. Hillyard, R. F. Loane, and J. Silcox, Ultramicr. **49**, 14 (1993).

[33] S. Hillyard and J. Silcox, Ultramicr. **58**, 6 (1995).

[34] D. Muller (2002), personal communication.

[35] E. J. Kirkland, *Advanced Computing in Electron Microscopy* (Plenum, New York, 1998).

[36] C. Humphreys, Rep. Prog. Phys. **42**, 122 (1979).

[37] H. A. Bethe, Ann. Physik (Leipzig) **87**, 55 (1928).

[38] P. G. Self, M. A. O'Keefe, P. R. Buseck, and A. E. C. Spargo, Ultramicroscopy **11**, 35 (1983).

[39] J. M. Cowley and A. F. Moodie, Acta Cryst. **10**, 609 (1957).

[40] P. Goodman and A. F. Moodie, Acta Cryst. A **30**, 280 (1974).

[41] D. van Dyck and W. Coene, Ultramicr. **15**, 29 (1984).

[42] W. Coene and D. van Dyck, Ultramicr. **15**, 41 (1984).

[43] J. H. Chen, D. van Dyck, M. O. Beek, and J. van Landuyt, Ultramicr. **69**, 219 (1997).

[44] O. L. Krivanek, N. Dellby, and A. R. Lupini, Ultramicr. **78**, 1 (1999).

[45] J. Zach and M. Haider, Optik **98**, 112 (1995).

[46] M. Haider, G. Braunhausen, and E. Schwan, Optik **99**, 167 (1995).

[47] M. Haider, H. Rose, S. Uhleman, E. Schwan, B. Kabius, and K. Urban, Ultramicroscopy **75**, 53 (1998).

[48] C. Mory and C. C. et al., Ultrmicr. **21**, 171 (1987).

[49] W. M. J. Coene, A. Thust, M. Op de Beeck, and D. van Dyck, Ultramicr. **64**, 109 (1996).

[50] A. Thust, W. M. J. Coene, M. Op de Beeck, and D. van Dyck, Ultramicr. **64**, 211 (1996).

[51] M. A. O'Keefe, E. C. Nelson, J. H. Turner, and A. Thust, in *Microsc. Microanal. 7, Suppl. 2: Proceedings*, MSA (Springer New York, 2001), pp. 898–899.

[52] C. K. et al., in *Microsc. Microanal. 6, Suppl. 2: Proceedings*, MSA (Springer New York, 2000), pp. 16–17.

[53] K. Moliere, Ann. d. Phys. Leipzig **34**, 461 (1939).

[54] H. Yoshioka, J. Phys. Soc. Japan **12**, 618 (1957).

[55] A. Howie, Proc. Roy. Soc. A **271**, 268 (1963).

[56] C. J. Humphreys and P. B. Hirsch, Philosophical Magazine **18**, 115 (1968).

[57] G. Radi, Acta Cryst. A **26**, 41 (1970).

[58] P. Rez, in *50 Years of Electron Diffraction*, edited by P. J. Dobson, C. J. Humphreys, and J. B. Pendry (Institute of Physics, London, 1977), pp. 61–67.

[59] A. Amali and P. Rez, Microsc. and Microanal. **3**, 28 (1997).

[60] C. R. Hall and P. B. Hirsch, Proc. Roy. Soc. London A **286**, 158 (1964).

[61] D. M. Bird and Q. A. King, Acta Cryst. A **46**, 202 (1990).

[62] L. M. Peng, G. Ren, S. L. Dudarev, and M. J. Whelan, Acta Cryst. A **52**, 257 (1996).

[63] P. A. Doyle and P. S. Turner, Acta Cryst. A **24**, 390 (1968).

[64] B. F. Buxton and J. E. Loveluck, J. Phys. C: Solid State Phys. **10**, 3941 (1968).

[65] A. Weickenmeier and H. Kohl, Acta Cryst. A **47**, 590 (1991).

[66] P. Rez, C. J. Humphreys, and M. J. Whelan, Phil. Mag. **35**, 81 (1977).

[67] Z. L. Wang and J. M. Cowley, Ultramicroscopy **31**, 437 (1989).

[68] Z. L. Wang, Phil. Mag. B **65**, 559 (1992).

[69] C. Dinges and H. Rose, Phys. Status Solidi A **150**, 23 (1995).

[70] G. R. Anstis, Acta Cryst. A **52**, 450 (1996).

[71] S. W. Lovesey, *Theory of Neutron Scattering* (Clarendon Press, Oxford, 1984).

[72] P. Bruüesch, *Phonons: Theory and Experiment II - Experiments and Interpretation of Experimental Results* (Springer, Heidelberg, 1987).

[73] H. X. Gao and L. M. Peng, Acta Cryst. A **55**, 926 (1999).

[74] R. F. Loane, P. Xu, and J. Silcox, Acta Cryst. A **47**, 267 (1991).

[75] G. Möbus, T. Gemming, and P. Gumbsch, Acta Cryst. A **54**, 83 (1998).

[76] D. A. Muller, B. Edwards, E. J.Kirkland, and J. Silcox, Ultramicr. **86**, 371 (2001).

[77] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford University Press, London, 1954).

[78] A. Howie, Journal of Microscopy **117**, 11 (1979).

[79] P. D. Nellist and S. J. Pennycook, Advances in Imaging and Electron Physics **113**, 147 (2000).

[80] J. M. Cowley, Ultramicr. **2**, 3 (1976).

[81] J. C. H. Spence and J. M. Cowley, Optik **50**, 129 (1978).

[82] D. E. Jesson and S. J. Pennycook, Proc. Roy. Soc. London A **441**, 261 (1993).

[83] C. Koch and J. M. Zuo, in *Microsc. Microanal. 7, Suppl. 2: Proceedings*, MSA (Springer New York, 2000), p. 126.

[84] J. M. Cowley, *Diffraction Physics* (North Holland, Amsterdam, 1975), 2nd ed.

[85] URL `http://www.fftw.org`.

[86] F. Seitz, Phys. Rev. **88**, 722 (1952).

[87] R. Jones, Mat. Sci. Eng. B **71** (2000).

[88] O. Mügge, Neues Jahrb. Min. p. 13 (1883).

[89] A. Ewing and W. Rosenhain, Phil. Trans. Roy. Soc. A **193**, 353 (1899).

[90] J. P. Hirth and J. Lothe, *Theory of Dislocations* (Krieger, Malabar, 1992).

[91] V. Volterra, Ann. Ecole Norm. Super. **24**, 400 (1907).

[92] A. E. H. Love, *The Mathematical Theory of Elasticity* (Cambridge Univ. Press, Cambridge, 1927).

[93] J. M. Hedges and J. W. Mitchell, Phil. Mag. **44**, 223 (1953).

[94] S. Amelinckx, Phil. Mag. **1**, 269 (1956).

[95] W. C. Dash, in *Dislocations and Mechanical Properties of Crystals*, edited by J. C. Fisher, W. G. Johnston, R. Thomson, and T. Vreeland (Wiley, New York, 1957), p. 57.

[96] P. B. Hirsch, R. W. Horne, and M. J. Whelan, in *Dislocations and Mechanical Properties of Crystals*, edited by J. Fisher, W. Johnston, R. Thomson, and T. Vreeland, Jr. (Wiley, New York, 1957), p. 92.

[97] I. L. F. Ray and D. J. H. Cockayne, Proc. Roy. Soc. A **325**, 534 (1971).

[98] H. Kolar, J. C. H. Spence, and H. Alexander, Phys. Rev. Lett. **77**, 4031 (1996).

[99] K. Suzuki, N. Maeda, and S. Takeuchi, Phil. Mag. A **73**, 431 (1996).

[100] K. Maeda, M. Inoue, K. Suzuki, H. Amasuga, M. Nakamura, and E. Kanematsu, J. Phys. III **7**, 1451 (1997).

[101] M. Inoue, K. Suzuki, H. Amasuga, Y. Mera, and K. Maeda, J. Appl. Phys. **83**, 1953 (1998).

[102] P. E. Batson, Phys. Rev. Lett. **83**, 4409 (1999).

[103] M. F. Chisholm, R. Buczko, M. Mostoller, T. Kaplan, A. Maiti, S. T. Pantelides, and S. J. Pennycook, Solid State Phenomena **67-68**, 3 (1999).

[104] C. Kiesielowski (2001), personal communication.

[105] K. Sumino, Il Nuovo Cimento D **19**, 137 (1997).

[106] K. Maeda, K. Suzuki, Y. Yamashita, and Y. Mera, J. Phys. Condens. Matter **12**, 10079 (2000).

[107] F. C. Frank, Phil. Mag. **42**, 809 (1951).

[108] W. Shockley, Phys. Rev. **91**, 228 (1953).

[109] F. R. N. Nabarro, Adv. Phys. **1**, 269 (1952).

[110] R. E. Peierls, Proc. Phys. Soc. **52**, 23 (1940).

[111] F. R. N. Nabarro, Proc. Phys. Soc. **59**, 256 (1947).

[112] V. Celli, M. Kabler, T. Nimomiya, and R. Thomson, Phys. Rev. **131**, 58 (1963).

[113] H. J. Möller, Acta Metall. **26**, 963 (1978).

[114] H. Alexander and H. Teichler, in *Handbook of Semiconductor Technology Volume I*, edited by K. A. Jackson and W. Schroeter (Wiley, Weinheim, 2000), p. 291.

[115] A. Seger and P. Schiller, Acta Metall. **10**, 348 (1962).

[116] W. T. Read, Jr., Phil. Mag. **45**, 775 (1954).

[117] W. Schröter and R. Labusch, Phys. Stat. Sol. **36**, 539 (1969).

[118] Labusch and Schröter, in *Dislocations in Solids*, edited by F. R. N. Nabarro (North Holland, Amsterdam, 1980), p. 129.

[119] K. Ono and K. Sumino, Jap. J. Appl. Phys. **19**, L629 (1980).

[120] K. Ono and K. Sumino, J. Appl. Phys. **54**, 4426 (1983).

[121] Y. A. Ossypian, in *Sov. Scient. Review Sect. 4a*, edited by J. M. Khalatnikov (1982), p. 219.

[122] C. Kisielowski-Kemmerich and H. Alexander, in *Defects in Crystals*, edited by E. Mizera (World Scientific Publications, Singapore, 1988), p. 290.

[123] C. Kisielowski-Kemmerich, J. Czaschke, and H. Alexander, Mater. Sci. Forum **10-12**, 745 (1986).

[124] B. Pohoryles, Inst. Phys. Conf. Ser. **104**, 175 (1989).

[125] P. Weber, E. R. Weber, L. Montelius, H. Alexander, and J. Michael, Phys. Rev. B **32**, 6571 (1985).

[126] V. Higgs, C. E. Norman, E. C. Lightowlers, and P. K. Kightley, in *Microscopic Semiconducting Materials* (Oxford University Press, Oxford, 1991).

[127] P. R. Wilshaw and T. S. Fell, in *Polycrystalline Semiconductors*, edited by J. H. Werner and H. P. Strunk (Springer Verlag, Heidelberg, 1990), p. 77.

[128] V. Higgs, E. C. Lightowlers, and P. K. Kightley, Mat. Res. Soc. Symp. Proc. **163**, 57 (1990).

[129] N. A. Drozdov, A. A. Patrin, and V. D. Tkachev, Sov. Phys. - JETP Lett. **23**, 597 (1976).

[130] R. Sauer, J. Weber, J. Stolz, E. R. Weber, K. H. Küster, and H. Alexander, Appl. Phys. A **36**, 1 (1985).

[131] V. Higgs, F. Chin, X. Wang, Y. Kitagawara, and Y. Yoshida, in *Semiconductor Silicon ECS PV 98-1*, edited by H. R. Huff (Electrochemical Society, Pennington, 1998), p. 1564.

[132] V. Higgs, F. Chin, X. Wang, J. Mosalski, and R. Beanland, J. Phys. : Cond. Matter **12**, 10105 (2000).

[133] M. Brohl and H. Alexander, Inst. Phys. Conf. Ser. **104**, 163 (1989).

[134] M. Brohl, M. Dressel, H. W. Hellberg, and H. Alexander, Phil. Mag. B **61**, 97 (1990).

[135] V. V. Kveder, Y. A. Ossypian, and A. I. Shalynin, Sov. Phys. JETP Lett. **40**, 729 (1984).

[136] V. V. Kveder, A. E. Koshelev, T. R. Mchedlidze, Y. A. Ossypian, and A. I. Shalynin, Sov. Phys. JETP Lett. **68**, 1041 (1989).

[137] J. Bennetto, R. Nunes, and D. Vanderbilt, Phys. Rev. Letts. **79**, 245 (1997).

[138] M. C. Payne, M. P. Teter, D. C. Allen, T. A. Arias, and J. D. Joannopoulos, Rev. of Mod. Phys. **64**, 1045 (1992).

[139] M. L. Cohen, Phys. Rep. **110**, 293 (1984).

[140] J. D. Joannopoulos, in *Physics of disordered Materials*, edited by D. Adler, H. Fritzsche, and S. R. Ovshinsky (Plenum, New York, 1985), p. 19.

[141] W. Picket, Comp. Phys. Rep. **9**, 115 (1989).

[142] J. C. Phillips, Phys. Rev. **112**, 685 (1958).

[143] M. T. Yin and M. L. Cohen, Phys. Rev. B **25**, 7403 (1982).

[144] F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).

[145] X. P. Li, G. Chen, P. B. Allen, and J. Q. Broughton, Phys. Rev. B **38**, 3331 (1988).

[146] I. P. Batra, F. F. Abraham, and S. Ciraci, Phys. Rev. B **35**, 9552 (1987).

[147] M. Nastar, V. V. Bulatov, and S. Yip, Phys. Rev. B **53**, 13521 (1995).

[148] T. W. Poon, S. Yip, P. S. Ho, , and F. F. Abraham, Phys. Rev. Lett. **65**, 2161 (1990).

[149] J. Q. Broughton and X. P. Li, Phys. Rev. B **35**, 9120 (1987).

[150] W. D. Lüdtke and U. Landman, Phys. Rev. B **37**, 4656 (1988).

[151] M. D. Kluge, J. R. Ray, and A. Rahman, Phys. Rev. B **36**, 4234 (1987).

[152] M. Ishimaru, K. Yoshida, and T. Motooka, Phys. Rev. B **53**, 7176 (1996).

[153] J. Tersoff, Phys. Rev. Lett. **56**, 632 (1986).

[154] J. Tersoff, Phys. Rev. B **37**, 6991 (1988).

[155] J. Tersoff, Phys. Rev. B **38**, 9902 (1988).

[156] L. J. Porter, S. Yip, M. Yamaguchi, H. Kaburaki, and M. Tang, J. Appl. Phys. **81**, 96 (1997).

[157] P. C. Kelires and J. Tersoff, Phys. Rev. Lett. **61**, 562 (1988).

[158] J. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov, and S. Yip, Phys. Rev. B **58**, 2540 (1998).

[159] P. B. Hirsch, J. Phys. Paris **40**, C6 (1979).

[160] R. Jones, J. Phys. Paris **40**, C6 (1979).

[161] A. Valladares, A. K. Petford-Long, and A. P. Sutton, Phil. Mag. Lett. **79**, 9 (1999).

[162] N. Letho and S. Öberg, Phys. Rev. Lett. **80**, 5568 (1998).

[163] D. Cherns, Phil. Mag. **30**, 549 (1974).

[164] C. Koch, J. Spence, C. Zorman, M. Mehregany, and J. Chung, Journal of Physics: Condensed Matter **12**, 10175 (2000).

[165] R. W. Nunes, J. Benetto, and D. Vanderbilt, Phys. Rev. B **57**, 10388 (1998).

[166] J. C. H. Spence, Mat. Sci. and Eng. R. **26**, 1 (1999).

[167] J. C. H. Spence, in *Proc. 39th EMSA Mtng.*, edited by Claitors and G. Bailey (1981), p. 120, check this!!!

[168] A. Ourmazd, G. R. Anstis, and P. B. Hirsch, Phil. Mag. A **48**, 139 (1983).

[169] D. Suznitsky and C. Carter, J. Mater Res. **6**, 2403 (1991).

[170] G. Ndbubuisi, J. Liu, and J. M. Cowley, Micros. Res. and Tech. **21**, 10 (1992).

[171] P. Batson (2001), unpublished, personal communication.

[172] Y. J. Chabal, Phys. Rev. Lett. **50**, 1850 (1983).

[173] H. Fukidome and M. Matsumura, Applied Surface Science **130-132**, 146 (1998).

[174] J. C. H. Spence, Proc. Ninth Int. Congr. Electr. Micros. (1978).

[175] M. A. O'Keefe, in *Microsc. Microanal. 6, Suppl. 2: Proceedings*, MSA (Springer New York, 2000), p. 1192.

[176] C. Moler and C. van Loan, SIAM Review **20**, 801 (1978).

[177] A. F. Moodie, Z. Naturforsch. **27**, 437 (1972).

[178] K. Chadan and P. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, Berlin, Heidelberg, New York, 1989), 2nd ed.

[179] K. Fujiwara, Journ. Phys. Soc. Japan **14**, 1513 (1959).

[180] F. Fujimoto, J. Phys. Soc. Japan **14**, 1558 (1959).

[181] B. E. Allman, P. J. McMahon, K. A. Nugent, D. Paganin, D. L. Jacobson, M. Arif, and S. A. Werner, Nature **408**, 158 (2000).

[182] R. Henderson, Quarterly Reviews of Biophysics **28**, 171 (1995).

[183] G. Feigel and M. Tegze, Rep. Prog. Phys. **62**, 355 (1999).

[184] A. Szoke, *Short wavelength coherent radiations, AIP Conf Series* (AIP, New York, 1986), vol. 147.

[185] J. J. Barton, Physical Review Letters **67**, 3106 (1991).

[186] D. Saldin, Surface Review and Letters **4**, 991 (1997).

[187] J. B. DeVelis, G. Parrent, and B. Thompson, Journal of the Optical Society of America **56**, 423 (1966).

[188] J. Wider, F. Baumberger, M. Sambi, R. Gotter, A. Verdini, F. Bruno, D. Cvetko, A. Morgante, T. Greber, and J. Osterwalder, Phys. Rev. Letts. **86**, 2337 (2001).

[189] V. Laue, Acta Crystalographica **2**, 106 (1949).

[190] J. Cowley, Acta Crystalographica **17**, 33 (1964).

[191] S. Kikuchi, Japanese Journal of Physics **5**, 83 (1928).

[192] J. W. Knowles, Acta Crystalographica **9**, 61 (1956).

[193] L. S. Bartell and C. L. Ritz, Science **185**, 1163 (1974).

[194] A. Ochowski, W. D. Rau, and H. Lichte, Physical Review Letters **74**, 399 (1995).

[195] J. C. H. Spence and J. Tafto, J. Micros. **130**, 147 (1982).

[196] D. Cherns, A. Howie, and M. Jacobs, Zeitschr. fuer Naturforsch. A **28**, 565 (1973).

[197] W. Qian, B. Totdal, R. Hoier, and J. C. H. Spence, Ultramicroscopy **41**, 147 (1992).

[198] J. C. H. Spence and J. M. Zuo, *Electron Microdiffraction* (Plenum, New York, 1994).

[199] M. V. Hove (2000), personal communication.

[200] J. Cowley, J. C. H. Spence, and V. V. Smirnov, Ultramicroscopy **68**, 135 (1997).

[201] J. C. H. Spence, in *High Resolution Electron Microscopy*, edited by C. Buseck and Eyring (Oxford University Press, 1988).

[202] C. Rossouw *et al.*, Phil. Mag. A **74**, 57 (1996).

[203] J. Larson *et al.*, Phys Rev. Letts. **85**, 2470 (2000).

[204] S. L. Dudarev, Kristallografiya **35**, 596 (1990).

[205] M. Z. Bazant, E. Kaxiras, and J. F. Justo, Phys. Rev. B **56**, 8542 (1997).

[206] M. Parrinello and A. Rahman, J. Chem. Phys. **76**, 2662 (1982).