Ultra High Resolution Imaging of Radiation-Sensitive Materials

Chen Huang

St Peter's College

University of Oxford

Supervisor: Angus I. Kirkland

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2016

To my parents

And

In memory of grandpa.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof Angus I. Kirkland, whose guidance has always been both enlightening and inspiring. It has been a great pleasure for me to have had the chance to work with you.

I would also like to express my sincere gratitude to all the members of the Electron Image Analysis group and the microscopy community at Oxford as a whole for forming such a great research environment where I truly have enjoyed all the engaging discussions on topics within and beyond the realm of science.

My thanks also go to my collaborators, Dr Benjamin Berkels at Aachen University for the work on non-rigid registration, and Dr Tan Sui from the Department of Engineering at Oxford for providing the human tooth sample.

Moreover, life at Oxford would not have been half as complete without my wonderful friends. Your friendships have added colours to the black-and-white world of a young electron microscopist, for which I feel enormously fortunate.

Finally, to my parents, Mrs Xiangyan Kong and Dr Jingxiong Huang, I can never thank you two enough for your constant encouragement and love, without which I could not imagine to have come this far on the way to becoming a scientist. And to my beloved partner, Ms Shengnan Dong, thank you for being there with me at the end of this long but enlightening journey.

Disclaimer

I hereby declare that this thesis is my own original work and has not been submitted before to any institution for assessment purposes.

Further, I have acknowledged all sources used and have cited these in the reference section.

The cerium dioxide and silicon nitride focal series data used in Chapter 5 was provided by Prof Angus I. Kirkland and Dr Hidetaka Sawada. The non-rigid registration stategy in Chapter 6 is a collaboration with Dr Benjamin Berkels from Aachen University.

Abstract

Electron radiation damage is an important topic in electron microscopy. A large proportion of materials or biological structures are still unable to be directly imaged by electron microscopes at high resolutions due to their radiation sensitivity, instead of by the instrumental resolution capacity of the microscopes. In fact, as the availability of aberration correctors was rapidly increased in the past decade, as well as the emergence of the next generation of chromatic aberration correctors, attention has been again focused on the control and reduction of radiation damage.

This thesis proposes a complete low-dose high resolution imaging strategy for characterising radiation-sensitive materials using an aberration-corrected transmission electron microscope. The microscope was quantitatively calibrated for electron dose and was operated under a strict low-dose condition to ensure maximum protection for the radiation-sensitive samples. Time series and focal series imaging were employed to allow other data processing techniques to be applied.

During the course of pursuing higher resolutions using low-dose imaging and data processing, several side problems have been explored. Image registration of low-dose image series was first tested using a variable dose time series of human tooth tissue. The impact of excessive image noise was reduced by modification of the registration algorithm. In the case of focal series image registration, a simulation-assisted registration procedure was developed. This demonstrated the ability to register focal series of cerium dioxide and silicon nitride at various doses, despite the contrast reversal

problem when the series defoci moved from under focus to over focus. A quantitative evaluation metric, the IQ factor, was implemented to assess the signal-to-noise ratio in the power spectrum of an image and has proved to be a useful indirect criterion for measuring the quality of registration and that of exit wave reconstruction.

Exit wave reconstruction from low-dose HRTEM was investigated using this improved image registration and the resultant exit waves were quantitatively compared using the IQ factor. The influence of noise on exit wave reconstruction was an intertwined problem with the low electron dose and has also been addressed in this work.

By combining improvements in the various aspects of low-dose imaging and data processing, time/focal series of ZSM-5, an important type of zeolitic catalyst, was acquired. The low-dose time series of ZSM-5 was aligned by both the traditional rigid XCF image registration and a more sophisticated non-rigid image registration method. The low-dose focal series of ZSM-5 was used to restore the electron wave function at the exit plane of the sample. The restored exit wave was able to resolve the fine structure inside the ZSM-5 framework, which was not clearly resolved in the individual images of the focal series.

Contents

List of Figures

List of Tables

Glossary

CCD charge coupled device.

Cryo-EM cryo-electron microscopy.

DAC digital-to-analogue converter.

DDD direct detection device.

DQE detective quantum efficiency.

ED electron diffraction.

EELS electron energy-loss spectroscopy.

FFT fast Fourier transform.

FT Fourier transform.

FWHM full width at half maximum.

HAADF high-angle annular dark-field.

HAP hydroxyapatite.

HREM high resolution electron microscopy.

HRTEM high resolution transmission electron microscope.

MAL maximum-likelihood.

MDS minimum dose system.

MTF modulation transfer function.

ROI region of interest.

SNR signal-to-noise ratio.

STEM scanning transmission electron microscope.

TEM transmission electron microscope.

WPOA weak phase object approximation.

XCF cross-correlation function.

Chapter 1

Introduction

1.1 Motivation for research

Seeking higher spatial resolution and gaining more knowledge about the sample are the main goals of microscopy. Transmission electron microscope (TEM) has been the leading microscopic technique in the exploration of the smallest structure details in materials since its development. Nowadays, sub-angstrom resolution has become routine for electron microscopists for a range of materials. Nevertheless, such high resolution structural information is not accessible in some materials that are susceptible to radiation damage caused by the high energy electron beam used in the electron microscopes. As a result, researchers who are interested in the structures of these *soft* materials have to compromise between the theoretically attainable spatial resolution which requires intense electron beams and the delicate sample structure which cannot survive the high electron dose. The control and reduction of electron radiation through low-dose imaging is limited by the fundamental need for the minimum amount of electrons required to generate a good representation of the sample in the images. With a very low amount of electrons used, new problems other than radiation damage arise, from poor image contrast to various kinds of noise corrupting

the images.

The formation of high resolution transmission electron microscope (HRTEM) images is a complex process involving electron wave interference and the final image contrast is subject to influences from the microscope configurations, in particular the lens aberrations. Aside from the direct imaging approach, exit wave reconstruction is an existing technique that restores the electron wave at the exit sample plane, where the wave is free of objective lens aberrations and more closely related to the electron potential of the sample. Because all reconstruction methods require serial acquisition with varied objective lens configurations, it has traditionally been considered not favourable for the characterisation of radiation-sensitive materials, despite its potential benefits.

This thesis describes research that aims to improve the quantification and control of the imaging dose while at the same time attempting to extend the application of exit wave reconstruction technique to low-dose HRTEM imaging of soft materials.

1.2 Overview of thesis

Chapter 2. This chapter reviews the origin, the mechanisms and the reduction methods of radiation damage in electron microscopy, with an emphasis in the context of the TEM. Some basic concepts, such as electron dose and dose rate, in particular their units used in electron microscopy, are discussed and clarified for later use.

Chapter 3. The theory of HRTEM image formation and the key instrumental components of aberration-corrected TEM are reviewed respectively. The theoretical background of electron exit wave reconstruction is also reviewed. Different reconstruction methods are surveyed, together with the software packages that are available within the community.

Chapter 4. The experimental procedures of low-dose HRTEM imaging are described in detail. In order to achieve a quantitative measurement of imaging dose, the calibration of the electron microscope has been emphasised and rigorously conducted. This chapter uses the time series imaging of hydroxyapatite from real human tooth tissue as an example, demonstrating how low-dose HRTEM imaging is carried out under controlled conditions at very high spatial resolution. The image registration problem has been identified as a crucial issue when the traditional cross-correlation function (XCF) based image registration fails to align the experimental image series due to the extreme low imaging dose. A quantitative measure of image quality, IQ factor, is introduced with both a descriptive concept and a clear mathematical form. This criterion for image quality measurement is adapted in the later parts of this thesis to evaluate the characterisation results of different materials.

Chapter 5. This part of the work investigates the application of exit wave reconstruction to low-dose focal series data. The high noise level in low-dose images and the registration effects on the reconstruction are investigated and quantitatively compared with low-dose focal series data. New registration routines and a new denoising method are developed to assist the application.

Chapter 6. This chapter applies the low-dose imaging techniques to the very radiation-sensitive material, ZSM-5, an important zeolite. Low-dose time series are recorded and registered with the improved rigid XCF registration method and a new non-rigid registration method based on multiscale gradient flow algorithm. Low-dose focal series data of ZSM-5 were also acquired and used to reconstruct the exit wave. The result demonstrates, for the first time, that a near-atomic resolution exit electron wave can be successfully restored even though the imaging dose is greatly curtailed, as required by the extreme beam sensitivity of the sample. This shows a promising future for applying exit wave reconstruction to investigations of many other radiation-

sensitive materials with complex structures at very high resolution.

Chapter 7. The final chapter summarises the achievements in this research as a whole, together with a list of potential directions for future exploration.

Chapter 2

Literature Review: Radiation Damage

2.1 Radiation damage mechanisms

Radiation damage has long been one of the main obstacles in high resolution electron microscopy (HREM), limiting the achievable resolution in the characterisation of radiation-sensitive materials (Glaeser, 1971; Egerton et al., 2004).

In a 200 kV medium-voltage electron microscope, electrons are accelerated to a speed of $2.65 \times 10^8\,\mathrm{m\,s^{-1}}$; the energy carried by a high speed electron is more than 50000 times that of the C-C bond energy (3.60-3.69eV). Interactions between the electron beam and the sample are highly complicated and are influenced by both the microscope configuration and the nature of the material itself. Although materials such as metals have proved to be radiation-resistant and can be imaged using an intense electron beam, many other materials, including most organic samples and some inorganic samples, e.g. zeolites (Treacy and Newsam, 1987; Rice et al., 1990; Terasaki et al., 1991), are much more prone to electron radiation damage.

When radiation damage occurs during EM imaging, there may be multiple damage

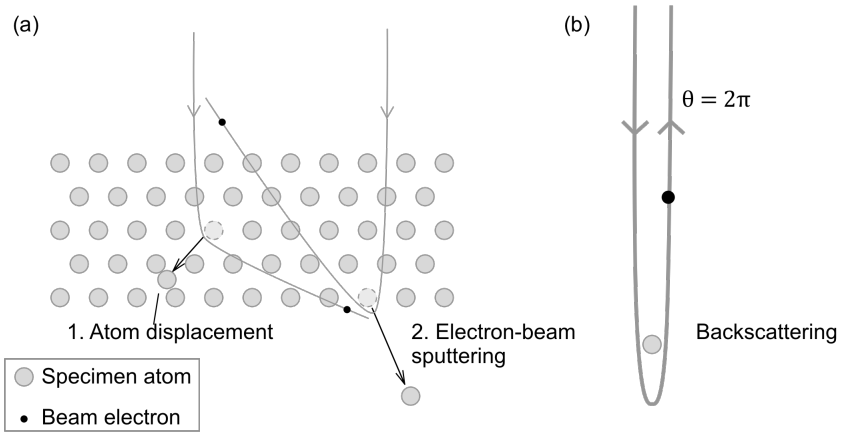


Figure 2.1: Schematic diagrams showing (a) Knock-on damage: 1. atom displacement, 2. electron-beam sputtering (b) Electron backscattering when the scattering angle, $\theta = 2\pi$.

mechanisms in effect simultaneously. Understanding the differences between these mechanisms and identifying their roles in the damage process are the prerequisite steps towards effective radiation damage control. In general, there are two main categories of damage mechanisms, namely *knock-on damage* and *radiolysis*. In addition, in some cases, as a result of energy deposition from these two mechanisms *sample heating* is also an important aspect of radiation damage.

2.1.1 Knock-on damage

Atom displacement: Elastic scattering is an electron-sample interaction process, in which the incident electrons are scattered by the Coulomb force between them and the nuclei of the sample atoms without any loss in total kinetic energy.

The energy transfer, E, is described by Eqs. 2.1 - 2.2;

$$E = E_{max} \sin^2 \frac{\theta}{2} = E_{max} \frac{1 - \cos\theta}{2} \tag{2.1}$$

$$E_{max} = \frac{E_0(E_0 + 2m_e c^2)}{E_0 + (1 + m_e/M)^2} \approx \frac{2E_0(E_0 + 2m_e c^2)}{Mc^2} \tag{2.2}$$

where, θ is the scattering angle, and c is the speed of light. The incident electron energy is given by E_0 (Eq. 2.1), with the rest electron mass, m_e, and the nucleus mass, M, the latter varying the atomic number of the element (Reimer and Kohl, 2008; Egerton et al., 2010). E_{max} is the maximum energy transfer, which takes place in the special case of electron backscattering ($\theta = 2\pi$) during an elastic collision.

For high-angle elastic scattering events ($\theta > 90°$), the energy transfer can be significant and lead to atom displacement, resulting in a Frenkel pair (Reimer and Kohl, 2008), in which a sample atom that absorbs sufficient kinetic energy to overcome the barrier of its potential well moves to a nearby interstitial position and leaving a vacancy at the original position. The energy required for such a displacement event is quantified by the displacement energy, E_d, determined by the bond strength, the atomic number and the crystal orientation of the sample. According to Eq. 2.2, when $E_{max} = \min E_d$, E_0 becomes the threshold energy E_{th}, which is the smallest incoming electron energy needed for atom displacement events to take place.

If the incident electron energy, E_0, is below E_{th} for all the elements in the material, then knock-on damage does not take place. The mean displacement energy values of some common materials and their corresponding threshold energy are listed in Table 2.1.

The concentration of displacement events c_d is the ratio of displacements to the number of atoms per unit volume, having the following relation;

$$c_d = \frac{\sigma_d j t}{e} \tag{2.3}$$

in which σ_d is the cross section, j is the current density and t is the irradiation time. The cross section is a property of the material (Table 2.1), while the beam current density is determined by the number of incoming electrons within a unit amount of time. For two accelerating voltages, both higher than E_{th} of the material,

Element	C	Si	Cu	Mo	Au
$E_{d,mean}$(eV)	5 (10)	13	19	33	33
E_{th}(keV)	27.2 (54.4)	145	400	810	1300
$\sigma_d(10^{-24}\text{cm}^2)$	89 (43.6)	69	59	7.2	–
E_{max}(eV)	366	155	68	45	22

Table 2.1: Mean displacement energy $E_{d,mean}$, threshold energy E_{th} of some common elements, along with their displacement cross sections, σ_d, and maximum energy transfer E_{max} for a backscattering event ($\theta = 180°$) at $E = 1\,\text{MeV}$ (Reimer and Kohl, 2008)

.

the higher voltage will have a higher probability to cause knock-on damage than the lower one for the same exposure time.

Electron-beam sputtering: When knock-on damage happens at or close to the surface of the specimen, atoms can be sputtered into the column vacuum instead of moving to interstitial sites. This type of knock-on damage is known as *electron-beam sputtering* (Cherns et al., 1976; Thomas, 1985; Crozier et al., 1990). Most sputtering events occur on the back surface of the specimen, where atoms are more likely to gain enough momentum to leave the sample.

Being a sub-type of knock-on damage, the only way to eliminate electron sputtering is to lower the accelerating voltage of the microscope so that the beam energy is below the threshold energy for the investigated material. The displacement energy is influenced not only by atomic number, but also by the position of an atom in the specimen. For many materials, the displacement energy of surface atoms are different from (usually smaller than) that of the atoms in the bulk. Hence, coating the beam-exit surface with a layer of different material that has a high threshold energy is a way to control electron-beam sputtering (Egerton et al., 2010).

2.1.2 Radiolysis

Radiolysis is radiation damage caused by ionisation, which involves interaction between the incident electrons and the electrons in the sample. Unlike knock-on damage, which happens during elastic collisions and has a threshold energy specific to the material atoms, radiolytic damage arises from inelastic scattering and can occur at any voltage in samples with covalent or ionic bonds. Because the energy transfer in this process occurs between electrons, which share the same rest mass m_e, the transferred momentum is substantially higher than that in electron-nucleus interactions according to classical mechanics.

Radiolysis is often the dominant mechanism in organic materials, polymers, and ionic crystals, and is also often the hardest type of damage to avoid (Glaeser, 2007). The damage process can be viewed in two steps: electron excitation and bond recovery. The excitation reaction, independent of accelerating voltage and sample temperature, happens immediately after the sample is exposed to the electron beam, causing breakage of chemical bonds. This produces free radicals in organic materials or electron-hole pairs (also known as excitons) in ionic crystals. These broken chemical bonds are unstable and have a tendency to cross-link and form new bonds in the subsequent de-excitation. This arbitrary rebonding process can gradually alter the original structure of the specimen. The degree of crystallinity decreases as the cumulative electron dose increases until the sample is uniformly amorphous. When the degree of radiolysis is high, molecule fragments can migrate away from their original positions and settle at new sites on the sample, or even leave the specimen altogether into the column vacuum. Electron beam damage research on various phthalocyanine films shows that coating both sides of the specimen with a continuous layer of carbon film can reduce radiation damage (Fryer and Holland, 1983). This protective effect is explained by a *sandwiching* mechanism which encapsulates the fragmented parts after the bond scission process and increases the probability of recovering bond fusion

processes at the damage site (Fryer et al., 1984; Fryer, 1984). However, coating the sample with two extra layers of materials increases the thickness of the TEM specimen which inevitably has a negative impact on the directly interpretable resolution.

2.1.3 Sample heating

Fast energy deposition in the specimen can accumulate heat and result in a rise in sample temperature. This heating effect and the local temperature change are linked by the following equation;

$$IE_{ie}(\frac{z}{\lambda_{mfp}}) = \frac{4\pi\kappa z(T-T_0)}{0.58+2\ln\frac{R_0}{R}} + 2\pi R^2\varepsilon\sigma(T^4-T_0^4) \tag{2.4}$$

Here, I is the beam current; E_{ie} is the average energy transfer in each inelastic collision; z is the thickness of the heated area and λ_{mfp} is the mean free path for inelastic scattering. The left side of Eq. 2.4 therefore estimates the energy deposition in the sample. For the right side of the equation, R is the incident beam radius; ε is the emissivity from both surfaces of the specimen; the thermal conductivity over a distance R_0 is κ; σ is the Stefan-Boltzmann constant and the temperature rise $\Delta T = T - T_0$ (Egerton et al., 2004). The thermal conductivity of the sample is the crucial factor in sample heating. Metals are in general much better thermal conductors ($> 100\,\mathrm{W\,m^{-1}\,K^{-1}}$), and are harder to heat up by the electron beam than most polymers ($0.2 \sim 2\,\mathrm{W\,m^{-1}\,K^{-1}}$) and biological samples (Reimer and Kohl, 2008; Egerton et al., 2004).

The role that sample heating plays in radiation damage is worth attention because it is part of the argument as to whether TEM or scanning transmission electron microscope (STEM) is better for studying radiation-sensitive materials. For STEM, specimen heating is different from that in TEM in two ways. First, the current density of a STEM probe is much higher than that of spread illumination in TEM.

Some research shows that STEM may have some advantages over TEM in terms of reducing radiation damage related to sample heating due to the fact that the temperature rise is not strongly affected by the current density (Egerton and Takeuchi, 1999). Secondly, the scanning mechanism in STEM means that the highly focused probe only stops at each position for a very short time ($\sim 1 \times 10^{-5}$ s), compared to a constant exposure throughout the TEM exposure process (~ 1 s). As a result, the scanning mechanism leaves the irradiated area more time to diffuse the heat away and recover and hence makes STEM the more favourable technique as far as sample heating is concerned (Egerton et al., 2004).

The generally accepted rule of reducing radiation damage by controlling the specimen temperature is to keep it low during experiments. It was reported in some early literature that radiolysis in alkali halides is strongly affected by temperature and that by cooling the specimen to liquid helium temperature, the defect mobility in the specimen is reduced (Hobbs et al., 1973; Hobbs, 1973, 1978). More recently, the increasing use of cryo-electron microscopy in biological EM imaging has proven very effective in retarding radiolysis (Section 2.4.3).

2.2 Dose and dose rate

2.2.1 Dose

Radiation damage is not an unique problem that only exists in electron microscopy. A notable amount of research in radiation damage can be found in other fields such as X-ray crystallography (Sliz et al., 2003), nuclear materials (Jenkins and Kirk, 2000), particle physics (Van Lint, 1987), etc. Historically, there are some variations between the definitions and notations of the concept of "*dose*".

In ionising radiation and radiobiology, electron dose is usually discussed in the form of energy absorption per unit mass, given the unit Gray (Gy) in the International

System of Units (SI);

$$1\,\mathrm{Gy} = 1\,\mathrm{J\,kg^{-1}}. \tag{2.5}$$

This definition emphasises the energy deposition by the radiation source. However, in the case of EM imaging, most electrons pass through the sample losing either none or only a very small fraction of their kinetic energy. It is therefore hard to accurately estimate the energy deposited but also less important to know the exact amount. In contrast, the electron beam current is relatively easy to measure. Therefore, it is more practical to use units involving the electric charge (or the number of electrons) instead of energy when describing dose in EM. In some radiation damage studies of radiation chemistry, Coulombs per unit volume, $\mathrm{C/m^3}$, are commonly used. This unit can be easily converted to the number of electrons per unit volume by dividing by the charge of a single electron, $e = -1.602 \times 10^{-19}\,\mathrm{C}$.

HREM samples have to be thin in the beam direction in order to achieve high spatial resolution. There is often no direct and quantitative way to experimentally measure the specimen thickness during imaging. The commonly accepted thickness estimation usually requires HRTEM image simulation. Therefore, it is more convenient to consider the electron dose in the context of two dimensions instead of three. For the sake of consistency, the electron dose is defined as the number of electrons per unit area, for example $e/\mathrm{nm^2}$ or $e/\mathrm{\AA^2}$, in this thesis. The other commonly used units of electron dose in EM literature include $\mathrm{C\,m^{-2}}$ and $\mathrm{C\,cm^{-2}}$, which can be converted to $e/\mathrm{\AA^2}$ by the following relationships;

$$1\,\mathrm{C\,m^{-2}} = 0.0625\,e\,\mathrm{\AA^{-2}} \tag{2.6}$$

$$1\,\mathrm{C\,cm^{-2}} = 6.25\,e\,\mathrm{\AA^{-2}} \tag{2.7}$$

Electron dose can also be understood from an instrumental perspective. Using j to denote the beam current density and t the exposure time, the dose, D is calculated

as;

$$D = \frac{jt}{e} \tag{2.8}$$

This relation applies to both TEM and STEM, with the difference that, t is the acquisition time for TEM, but, instead for STEM, t refers to the probe dwell time at each individual pixel.

2.2.2 Dose rate

Regardless of damage mechanisms, once radiation damage starts to occur, the amount of damage is expected to grow as the cumulative electron dose increases. It is often useful to know the speed at which the dose accumulates in order to estimate how fast radiation damage progresses. As can be seen in Eq. 2.8, moving the exposure time t from the right hand side to the left gives

$$D_{rate} = D/t = \frac{j}{e} \tag{2.9}$$

Thus, for a unit amount of time, the electron dose is solely determined by the current density in the electron beam. Hence, the control of the dose rate D_{rate} (Eq. 2.9) can be controlled by changing the value of j.

Intuitively, keeping the total dose as a constant, the change in dose rate may have an impact on radiation damage. However, assuming such a dose-rate effect exists, demonstrating and quantifying this variance in radiation damage faces a number of experimental difficulties that need to be addressed. Foremost, to get the same total dose with a dose rate several orders of magnitude lower than the original dose rate, one needs to increase the exposure time by the same amount. For example, a total electron dose of 1 s exposure at $1 \times 10^5\, e/(\mathrm{nm}^2\,\mathrm{s})$ a is equivalent to ca.17 min at $1 \times 10^2\, e/(\mathrm{nm}^2\,\mathrm{s})$. Although both of these dose rates are feasible by adjustment of the electron beam, in practice images are rarely taken with exposure times more than

a few seconds, on account of practical issues, in particular sample drift and external sources of vibration. The comparison between a short-exposure high flux image and a long-exposure low flux image will inevitably need to address these additional variables. Adding a stack of short-exposure low-flux images may be a workaround for the sample drift problem, but it will require reliable low-dose image registration, which by itself is a formidable task as will be discussed in later chapters.

Apart from occasional references in the early literature (Fryer, 1987; Varlot et al., 1997), the role of the dose rate in radiation damage has gathered more interest in some recent studies (Karuppasamy et al., 2011; Barton et al., 2011; Jiang and Spence, 2012). These works have provided some evidence of a proposed "dose rate effect", but a comprehensive agreement on what systematic comparison methodology should be used and how other complications can be eliminated has not emerged.

2.3 Measurement of radiation damage

In order to correlate the amount of electron irradiation to the amount of damage, it is necessary to have a quantitative measure of radiation damage. There are three established methods for measuring the level of radiation damage: *loss of mass*, *electron diffraction patterns*, and *Electron Energy Loss Spectroscopy (EELS)* (Reimer and Kohl, 2008).

2.3.1 Loss of mass

During radiation damage, molecules can leave the surface of the sample and enter the column vacuum after the bond breaking process. Therefore beam irradiation will cause a thinning effect or a hole-drilling effect. Assuming all damaged parts of the sample enter the vacuum, then the loss in specimen mass should be proportional to the radiation damage (Ramamurti et al., 1975; Freeman and Leonard, 1981; Lin,

1974).

Problems with using mass loss as the indicator of radiation damage include, a) the assumption that all the damaged sample leaves the specimen grid is often not true, b) accurate mass measurement is a difficult task in the microscope, c) sample contamination can compromise the effectiveness of the mass measurement, and d) damage in a small local area on the sample is not measurable by this approach.

2.3.2 Decay of electron diffraction pattern

Electron diffraction (ED) patterns are an indication of the crystallinity in samples. Radiation-sensitive crystals deteriorating under electron radiation leads to a loss in crystallinity and hence a loss in the intensity of reflections in the ED patterns. By recording the diffraction pattern with films or digital detectors, the change in ED intensity can be used as an effective proxy to quantitatively measure the radiation damage (Fryer, 1984).

2.3.3 Electron energy-loss spectroscopy (EELS)

Electron energy-loss spectroscopy (EELS) is a technique for analysing compositional and bonding changes in the sample (Reimer, 2013). It also provides an alternative approach to quantitatively measuring radiation damage (Egerton et al., 1987; Egerton, 2011b). Elements in the sample can be monitored by core-loss spectroscopy and structural order changes can be detected by the fine structure in the low-loss region (Egerton, 2011a). It has been shown that a relationship between the resonance energy of π-electrons and the radiation sensitivity can be determined by analysing the low-loss spectra of polyethylene terephthalate (PET) (Isaacson, 1975; Varlot et al., 1997). A study of changes in the aromatic structure of polystyrene under electron beam exposure shows a dependence of critical dose on probe size rather than dose rate, an effect attributed to fast secondary electron generation when using smaller

electron probe (Siangchaew and Libera, 2009).

2.3.4 Critical dose

The *critical dose* (or *characteristic dose*), D_c, is often mentioned in the literature as a maximum dose limit for characterising a specific radiation-sensitive material. However, the definition of this "maximum dose" appears to be somewhat arbitrary. Some researchers define critical dose as the dose when the sample "begins" to be destroyed (Spence, 2013), whereas others define it as the dose of "destruction" (Saxton and Frank, 1977) or "complete destruction" (Reimer, 1984; Egerton et al., 2004) of the ED pattern. More ambiguous definitions such as the "tolerable dose" without losing sample structure are also sometimes used (Egerton, 2012; Kourkoutis et al., 2012). However, a clearer and more quantitative definition has been used in various studies, which consider the critical dose D_c to be the electron dose that attenuates a chosen feature in the ED pattern (or power spectrum) to $1/e$ ($\approx 37\%$) of its original intensity (Pan and Crozier, 1993; Schäfer et al., 2009).

2.4 Reducing radiation damage

Reducing the influence of radiation damage can be approached from various aspects of low-dose HRTEM imaging. Strategies include targeting the dominant damage mechanism, improving operational techniques to maximise the efficiency of electron use and preserving the radiation-sensitive sample with minimum side effects on image quality.

2.4.1 Accelerating voltage choice

The accelerating voltage of the electron source has a direct impact on radiation damage effect. Knock-on damage is reduced by using lower accelerating voltage as elec-

tions with less energy have a smaller cross-section with the sample atoms and less likelihood of causing atom displacements. Knock-on damage can even be completely avoided if the electron energy is below the threshold energy of all atoms in the sample. However, the voltage-damage relationship is opposite in the case of radiolysis, which will be worsened by a lower-energy beam because the interaction time between incident electrons and the sample electrons is increased allowing more ionisation events to occur.

Between knock-on damage and radiolysis, one tends to overshadow the other for a specific type of sample. Therefore, choosing the accelerating voltage according to the dominant damage mechanism of the sample can help reduce radiation damage. Controlling the electron energy below the threshold can prevent knock-on damage in metals. However, if radiolysis is the main mechanism, studies have shown that for organic and biological materials radiation damage is reduced by a factor of 4 at 3 MV compared to 100 kV (Dupouy, 1976).

2.4.2 Minimum Dose System (MDS)

An automated image acquisition routine can reduce unnecessary electron exposure during the microscope adjustment and improve the efficiency of the use of a limited electron dose budget. On some JEOL microscope control systems, a minimum dose system (MDS) (Fig. 2.2) is pre-installed for low-dose imaging (Evans et al., 2008). The main function of the MDS is to automate the adjustments of lens values, beam shift and image shift. Three operational modes - **Search**, **Focus**, and **Photo** - are respectively configured in advance to suit the needs of each experimental step. The **Search** mode operates at a low magnification with a spread beam, keeping the intensity of the beam much lower than that actually used for imaging. This mode is used when searching for a suitable sample area (red circle in Fig. 2.2) for later imaging. The **Focus** mode and the **Photo** mode have essentially the same settings,

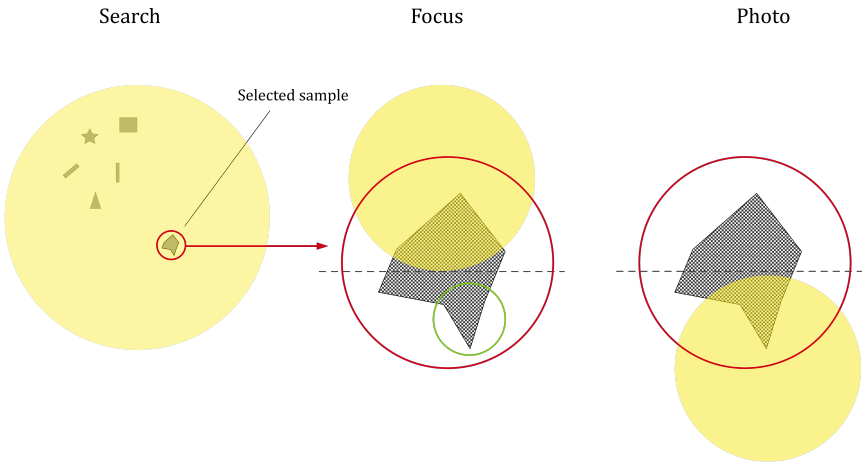


Figure 2.2: Three modes of the minimum dose system. Red and green circles mark the chosen sample and the region of interest respectively.

except that the beam position of the **Focus** mode is deliberately offset by a small amount so that the electron beam does not damage the region of interest (ROI) (green circle in Fig. 2.2), but is still close enough to it so that the sample has a practically identical crystal orientation as the ROI. Delicate adjustment of sample tilt, defocus and other microscope parameters can now be carried out in this similar area without damaging the final imaging region. After reaching the optimal imaging conditions, one can switch from the **Focus** mode to the **Photo** mode and record the image immediately to ensure that the electron exposure in the ROI is kept to a minimum.

2.4.3 Negative staining and cryo-electron microscopy

Negative staining Negative staining is an established EM sample preparation technique for biological materials, and has proved to be an effective procedure in protecting the sensitive organic structures from electron radiation (Watson, 1958). The fundamental idea behind negative staining is to replace the water component in biological samples with a radiation-resistant substance which retards the process of degradation. Heavy metal salts, such as uranyl acetate are most commonly used as the staining substance. Alternative staining substances include neutralised phospho-

tungstic acid and ammonium molybdate. Reviews of the detailed sample preparation procedures can be found in Harris and Horne (1994); Glaeser (2007).

In addition to increasing robustness against radiation damage, the negative staining treatment also enables the sample to generate better image contrast due to the stronger scattering effect of the heavy elements than the low atomic number elements in unstained biological samples.

Negative stained samples are suitable for TEM imaging with resolutions up to ca.2 nm to 3 nm. For resolution higher than this, the interpretation of the image contrast becomes difficult (Glaeser, 2007), as most biological TEM imaging uses large defocus values and the contrast transfer function oscillation may become severe even in the lower frequency range.

Cryo-electron microscopy (Cryo-EM) It was discovered that by plunge-freezing the specimen and embedding it in vitreous ice, biological structures can be preserved in a near-natural hydrated state (Meryman, 1956; Fernández-Morán, 1960; Dubochet et al., 1982, 1988). Such treatment creates a layer of cryo-protection for the TEM samples making them more resistant to electron radiation during observation and imaging. It was reported that the tolerable dose for radiation-sensitive samples are increased by a factor of 5 to 7 after cryo-protection (Glaeser, 2008). Cryo-electron microscopy (Cryo-EM) imaging has been largely enabled by the advances in dedicated hardware. Over the years, cold TEM stages with cooling systems using liquid-nitrogen or liquid-helium have been developed and integrated into microscope systems to help keep the sample in its frozen state (Dietrich et al., 1976; Carlson and Evans, 2011; Rigort et al., 2010; Duke et al., 2014).

Early work using Cryo-EM took advantage of the prolonged sample lifetime and studied radiation damage processes in hydrated biological samples (Taylor and Glaeser, 1974, 1976; Stewart and Vigers, 1986). With the fast growth in computing power

and the establishment of the *single-particle analysis* technique (Frank, 2009), applications of Cryo-EM have quickly gained popularity within the structural biology community. Single-particle analysis exploits the fact that structures of the same biological molecules are identical and records many images containing a great number of particles in random orientations. As a result, assuming all particles can be correctly classified and aligned, which is not an easily achievable task due to the poor signal-to-noise ratio (SNR) in low-dose images, high quality projections from different directions can be obtained by averaging hundreds of thousands of identical particles with the same orientation. Because each projection is the average of many images, the SNR in the original data is increased. Many structures, which could not be directly imaged by other techniques, have been reconstructed at near-atomic resolution, from symmetric biological samples (in particular, the structure of GroEL has been extensively studied (Harris et al., 1994; Ranson et al., 2001; Ludtke et al., 2004; Kanno et al., 2009)) as well as more complicated asymmetric large protein molecules (Zhang et al., 2010; Liu et al., 2010; Liao et al., 2013). In fact, as the resolution of single-particle cryo-EM continuous to improve, it has now become more and more accepted as a method for *de novo* model building in protein structure determination, which traditionally relied heavily on X-ray crystallography (Cheng, 2015; Bai et al., 2015).

Chapter 3

Imaging Theory and Instrumentation

3.1 Electron-specimen interaction

An uniform electron wave function $\psi_e(\mathbf{x},z) = 1$ can be used to describe a coherent incident electron beam moving in a direction z, parallel to the optical axis. The elastic interaction between the incident beam and the atoms of the specimen can be viewed as a potential well function $\phi(\mathbf{x}, z)$ (Cowley and Pogany, 1968) which modifies the phase and the amplitude of the incident electron wave. The electron wave function at the exit plane of the sample, or simply the *exit wave*, $\psi_o(\mathbf{x})$ is

$$\psi_o(\mathbf{x}) = \exp(a(\mathbf{x}) + i\eta(\mathbf{x})) \tag{3.1}$$

With $a(\mathbf{x})$ the change in the wave amplitude, and $\eta(\mathbf{x})$ the phase shift. Both $a(\mathbf{x})$ and $\eta(\mathbf{x})$ are conjugate symmetric, with the phase given by;

$$\eta(\mathbf{x}) = \sigma \int \phi(\mathbf{x}, z)dz \tag{3.2}$$

$\mathbf{x}$ is a position vector in the sample plane, z denotes the thickness which the electron travels through the sample. The integral $\int \phi(\mathbf{x}, z)dz$ represents the projected

electron potential in the z direction.

σ is an interaction constant given by;

$$\sigma = \frac{2\pi m_e e \lambda}{h^2} \tag{3.3}$$

where, m_e and e are the rest mass and the unit charge of the electron, h is Plank's constant, λ is the wave length related to the accelerating voltage V, which is calculated with relativistic correction as;

$$\lambda = \frac{h}{\sqrt{2m_e eV(1 + \frac{eV}{2m_e c^2})}} \tag{3.4}$$

If the interaction is weak, the amplitude term $a(\mathbf{x})$ will be insignificant. In this case, the electron-specimen interaction only affects the phase and the sample can then be considered as a *phase object*.

$$\psi_o(\mathbf{x}) \simeq \exp(i\eta(\mathbf{x})) \tag{3.5}$$

When the sample is sufficiently thin or mainly contains light elements such that the resulting change in the electron wave phase $|\eta| \ll 1$, and the exit wave function can be further expanded as

$$\psi_o(\mathbf{x}) \simeq 1 + i\eta(\mathbf{x}) \tag{3.6}$$

A specimen that satisfies the above conditions is referred to as a *weak phase object* (WPO). A different assumption of a negligible $\eta(\mathbf{x})$ and a small $a(\mathbf{x})$ in theory leads to a *weak amplitude object* (WAO). However, this is rarely the case for most TEM samples and is therefore not considered further.

For high resolution TEM imaging, the weak phase object approximation (WPOA) is convenient but is only valid for samples that are extremely thin. Imaging of thick

samples is accompanied by multiple and inelastic scattering and electron absorption, both of which complicate the scattering model and image interpretation.

3.2 Imaging Theory

Imaging theory is the mathematical representation that describes the process after the electron-specimen interaction to the image plane before detection. At the detection plane the complex electron wave is recorded as a real image intensity with a loss of phase information. It is at the exit plane of the sample that the electron wave $\psi_o(\mathbf{x})$ carries the maximum amount of information about the sample and contains few microscope-induced effects.

While enabling the TEM to produce high resolution images, various parts of the microscope also have potentially negative influences on the electron wave. Therefore improvements in electron microscopy have focussed on reducing these effects. When it is impossible to correct these imperfections by instrumental design or engineering, efforts are made to accurately measure and model these microscope effects so that they may be removed by offline data processing.

TEM imaging can be interpreted as a convolution between the exit wave and an all-encompassing *point response function* $t(\mathbf{x})$, with $*$ denoting the convolution operation. The convolution product is the electron wave $\psi_i(\mathbf{x})$ at the image plane

$$\psi_i(\mathbf{x}) = \psi_o(\mathbf{x}) * t(\mathbf{x}) \tag{3.7}$$

In wave optics, real space and reciprocal space (also known as frequency space) are linked by a Fourier transform (FT) operation. Taking the Fourier transforms of both sides of Eq. 3.7, the convolution in real space is simplified to a multiplication as shown in Eq. 3.8, which represents the wave function at the back focal plane of the objective lens.

$$\Psi_i(\mathbf{k}) = \Psi_o(\mathbf{k})T(\mathbf{k}) \tag{3.8}$$

Capital letters are used to denote functions in reciprocal space and lower-case letters for the corresponding real space functions, with $\mathbf{k}$ being the spatial frequency. Ψ_i, Ψ_o, and T are the Fourier transforms of the real space functions ψ_i, ψ_o, and t. The function arguments, $\mathbf{x}$ and $\mathbf{k}$, also indicate the function space.

$T(\mathbf{k})$ is the Fourier transform of the point response function. It is referred to by different names in the literature, but more often as the *transfer function*, or the *contrast transfer function*. The structure of transfer functions varies based on the specific assumptions used for describing the imaging process. Here, I first start without any approximations about the sample or the exit wave.

After the sample plane, the electron beam travels as a complex wave, initially ψ_o, until it reaches the detector, where it is recorded as an image intensity distribution $g(\mathbf{x})$. The image intensity is by definition the probability of the electron landing position on the recording device, which equals to the squared modulus of the image wave function in real space (Eq. 3.9).

$$\begin{aligned} g(\mathbf{x}) &= |\psi_i(\mathbf{x})|^2 \\ &= \psi_i^*(\mathbf{x})\psi_i(\mathbf{x}) \\ &= [\psi_o^*(\mathbf{x}) * t^*(\mathbf{x})][\psi_o(\mathbf{x}) * t(\mathbf{x})] \end{aligned} \tag{3.9}$$

Taking Fourier transform on both sides of Eq. 3.9, it becomes;

$$\begin{aligned} G(\mathbf{k}) &= \Psi_i^*(-\mathbf{k}) * \Psi_i(\mathbf{k}) \\ &= [\Psi_o^*(-\mathbf{k})T^*(-\mathbf{k})] * [\Psi_o(\mathbf{k})T(\mathbf{k})] \\ &= \int [\Psi_o^*(\mathbf{k}'+\mathbf{k})T^*(\mathbf{k}'+\mathbf{k})][\Psi_o(\mathbf{k}')T(\mathbf{k}')]d\mathbf{k}' \\ &= \int T(k')T^*(k'+\mathbf{k})\Psi_o(\mathbf{k}')\Psi_o^*(\mathbf{k}'+\mathbf{k})d\mathbf{k}' \end{aligned} \tag{3.10}$$

$$T_{cc}(\mathbf{k},\mathbf{k}') \equiv T(\mathbf{k})T^*(\mathbf{k}') \tag{3.11}$$

The function $T_{cc}(\mathbf{k},\mathbf{k}')$, known as the *transmission cross-coefficient*, describes the interference between two spatial frequencies, $\mathbf{k}$ and $\mathbf{k}'$, at the back focal plane of the objective lens. Eq. 3.10 is general and its transfer of the exit wave to the image wave $\Psi_i(\mathbf{k})$ gives rise to both linear and the non-linear components (Van Aert et al., 2010) in the image spectrum.

Under perfectly coherent illumination conditions, the transmission cross-coefficient function is simply

$$T_{cc}(\mathbf{k},\mathbf{k}') = \exp[-i\chi(\mathbf{k})]\exp[-i\chi(\mathbf{k}')] \tag{3.12}$$

The Fourier transform of the exit wave function (Eq. 3.6) under the WPOA is given by;

$$\Psi_o(\mathbf{k}) = \delta(\mathbf{k}) + iH(\mathbf{k}) \tag{3.13}$$

where $\delta(\mathbf{k})$ denotes the unscattered beam.

Hence,

$$
\begin{aligned}
G(\mathbf{k}) &= \int T_{cc}(\mathbf{k}', \mathbf{k}' + \mathbf{k})[\delta(\mathbf{k}') + iH(\mathbf{k}')][\delta(\mathbf{k}' + \mathbf{k}) + iH^*(\mathbf{k}' + \mathbf{k})]d\mathbf{k}' \\
&= [T_{cc}(0, \mathbf{k})\delta(\mathbf{k}) + T_{cc}(-\mathbf{k}, 0)\delta(-\mathbf{k})] + i[T_{cc}(0, \mathbf{k})H(\mathbf{k}) + T_{cc}(-\mathbf{k}, 0)H^*(-\mathbf{k})] \\
&\quad - \int [T_{cc}(\mathbf{k}', \mathbf{k}' + \mathbf{k})H(\mathbf{k}')H^*(\mathbf{k}' + \mathbf{k})]d\mathbf{k}' \qquad (3.14)
\end{aligned}
$$

Therefore the expression for the image intensity $G(\mathbf{k})$ becomes the sum of three parts,

$$
G(\mathbf{k}) = G_b(\mathbf{k}) + G_l(\mathbf{k}) + G_{nl}(\mathbf{k}) \qquad (3.15)
$$

the background $G_b(\mathbf{k})$, the interference $G_l(\mathbf{k})$ of the central unscatterred beam and the scattered beams, and the interference $G_{nl}(\mathbf{k})$ between different scattered beams. $H(\mathbf{k})$ is also conjugate symmetric;

$$
H^*(-\mathbf{k}) = H(\mathbf{k}) \qquad (3.16)
$$

Hence,

$$
\begin{aligned}
G_b(\mathbf{k}) &\equiv T_{cc}(0, \mathbf{k})\delta(\mathbf{k}) + T_{cc}(-\mathbf{k}, 0)\delta(-\mathbf{k}) \qquad (3.17) \\
G_l(\mathbf{k}) &\equiv iH(\mathbf{k})[T_{cc}(0, \mathbf{k}) + T_{cc}(-\mathbf{k}, 0)] \qquad (3.18) \\
G_{nl}(\mathbf{k}) &\equiv -\int [T_{cc}(\mathbf{k}', \mathbf{k}' + \mathbf{k})H(\mathbf{k}')H^*(\mathbf{k}' + \mathbf{k})]d\mathbf{k}' \qquad (3.19)
\end{aligned}
$$

$G_l(\mathbf{k})$ and $G_{nl}(\mathbf{k})$ are known as the *linear component* and the *non-linear component* in the image intensity respectively. Given that $g_b(\mathbf{x}) = \mathcal{F}[G_b(\mathbf{k})]$ is merely an invariable background intensity in real space and that the features in an image are due to the variations in the image intensity (image contrast), it is a common practice

to define the image contrast $C(\mathbf{k})$ (and $c(\mathbf{x})$ in real space) by ignoring the background component $G_b(\mathbf{k})$, to give;

$$C(\mathbf{k}) \equiv G_l(\mathbf{k}) + G_{nl}(\mathbf{k}) \tag{3.20}$$

When the sample is thin and the electron scattering is weak, only the interference of the scattered beams with the unscatterred beam is considered as represented by $G_l(\mathbf{k})$, making $G_{nl}(\mathbf{k})$ negligible. In this case the image contrast is further reduced to

$$C(\mathbf{k}) = G_l(\mathbf{k}) = iH(\mathbf{k})[T_{cc}(0,\mathbf{k}) + T_{cc}(-\mathbf{k},0)] \tag{3.21}$$

However, it should be noted that the assumption of perfect coherence in the above is never realised in practice. Hence this approach needs to be modified for practical application to include partial coherence as described later.

3.2.1 The wave aberration function

For a perfect objective lens with no aberration, the wavefront in the diffraction plane should be spherical centred at the image point P_g (Fig. 3.1). Due to the imperfect formation of the magnetic field or to the unavoidable aberrations of round electromagnetic lenses (Scherzer, 1936), various kinds of aberrations exist and the actual aberrated wavefront is shifted away from the ideal wavefront by an amount of $W(\omega)$, with $\omega = \lambda\mathbf{k} = \lambda k e^{i\varphi}$ being a complex variable ($\mathbf{k}$ is the reciprocal space vector at the diffraction plane, and $k = |\mathbf{k}|$). $W(\omega)$ can be expanded as a Taylor series (Eq. 3.22), with the coefficients in each term of the Taylor series representing one type of aberration (for example $A_1 = |A_1|e^{i\alpha_1}$). Table 3.1 gives a complete list of aberrations contained in Eq. 3.22.[1]

[1]For high angle annular dark field STEM imaging, the effects of the aberrations are to alter the phase function.

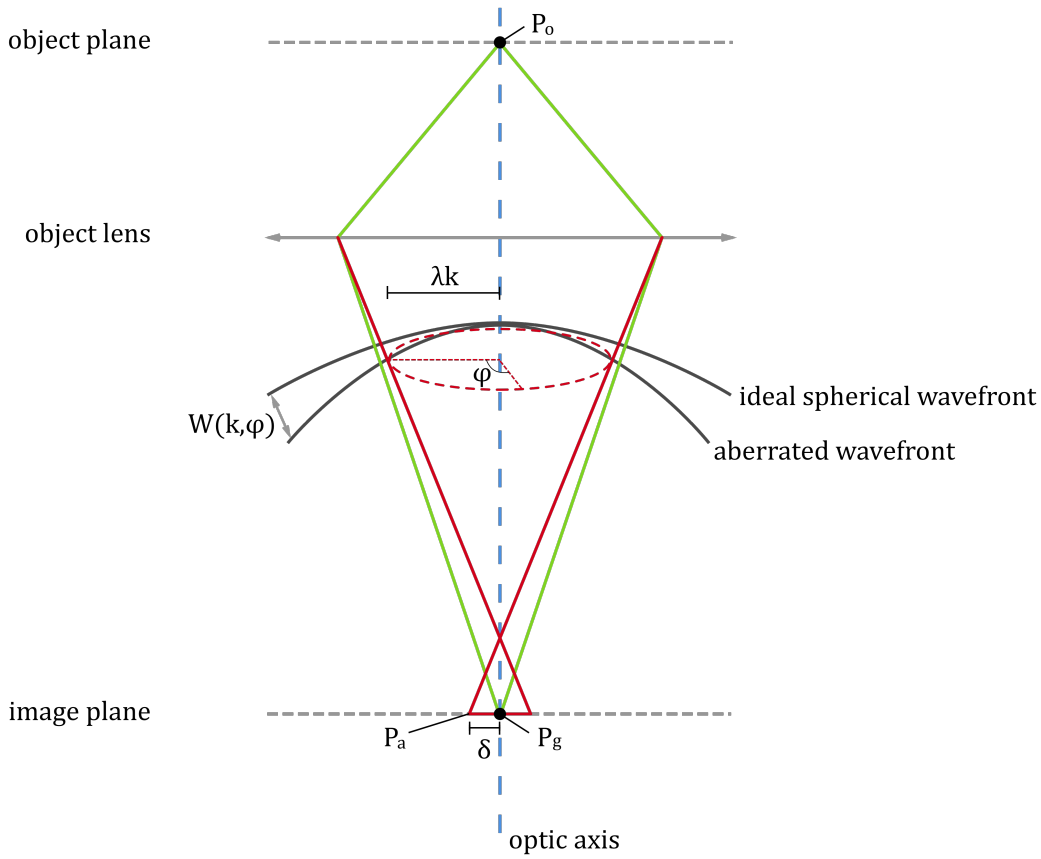


Figure 3.1: Lens aberrations cause the actual wavefront to shift away from the ideal spherical wavefront at the diffraction plane. For an object point P_o, the image at the ideal image plane, where the ideal wavefront of P_o converges into an image point P_g, the actual aberrated image "point" P_i may be of irregular shape. (In the special example here, P_o is on the optic axis and the aberrated wavefront maintains a rotationally symmetric shape, the aberrated image point becomes a blurred disk with radius δ).

Mathematically, applying the effects of aberrations for phase contrast imaging can be achieved through multiplying the wave function with a phase shift factor of $\exp[-i\chi(\mathbf{k})]$, in which $\chi(\mathbf{k}) = \frac{2\pi}{\lambda}W(\lambda\mathbf{k})$ (Saxton, 1995, 2015).

Aberration	Name	Order in ω	Azimuthal symmetry
A_0	Image shift	1	1
A_1	Two fold astigmatism	2	2
C_1	Defocus (over focus as positive)	2	∞
A_2	Three fold astigmatism	3	3
B_2	Axial coma	3	1
A_3	Four-fold astigmatism	4	4
S_3	Axial star aberration	4	2
C_3	Spherical aberration	4	∞
A_4	Five fold astigmatism	5	5
D_4	Three lobe aberration	5	3
B_4	Fourth order axial coma	5	1
A_5	Six fold astigmatism	6	6
S_5	Fifth order axial star aberration	6	4
C_5	Fifth order spherical aberration	6	∞
R_5	Fifth order rosette aberration	6	2

Table 3.1: Aberration coefficients up to sixth order in ω.

$$
\begin{aligned}
W(\omega) \;=\; & \mathcal{R}\{A_0\omega^* \\
& +\frac{1}{2}A_1\omega^{*2} + \frac{1}{2}C_1\omega^*\omega \\
& +\frac{1}{3}A_2\omega^{*3} + \frac{1}{3}B_2\omega^{*2}\omega \\
& +\frac{1}{4}A_3\omega^{*2}\omega^2 + \frac{1}{4}S_3\omega^{*3}\omega + \frac{1}{4}C_3\omega^{*2}\omega^2 \\
& +\frac{1}{5}A_4\omega^{*5} + \frac{1}{5}B_4\omega^{*3}\omega^2 + \frac{1}{5}D_4\omega^{*4}\omega \\
& +\frac{1}{6}A_5\omega^{*6} + \frac{1}{6}R_5\omega^{*5}\omega + \frac{1}{6}S_5\omega^{*4}\omega^2 + \frac{1}{6}C_5\omega^{*3}\omega^3\}
\end{aligned}
\tag{3.22}
$$

The symmetry of each aberration can be seen from the cosine functions in Eq. 3.23, where the wave aberration function is rewritten as a function of $k = |\mathbf{k}|$ and φ, or from the graphic illustration of phase plates generated by each aberration as shown in Fig. 3.2.

$$
\begin{aligned}
W(k,\varphi) &= |A_0|\,\lambda k\cos(\varphi-\alpha_0) \\
&\quad +\frac{1}{2}|A_1|\,\lambda^2k^2\cos 2(\varphi-\alpha_1)+\frac{1}{2}|C_1|\,\lambda^2k^2 \\
&\quad +\frac{1}{3}|A_2|\,\lambda^3k^3\cos 3(\varphi-\alpha_2)+\frac{1}{3}|B_2|\,\lambda^3k^3\cos(\varphi-\beta_2) \\
&\quad +\frac{1}{4}|A_3|\,\lambda^4k^4\cos 4(\varphi-\alpha_3)+\frac{1}{4}|S_3|\,\lambda^4k^4\cos 2(\varphi-\sigma_3)+\frac{1}{4}|C_3|\,\lambda^4k^4 \\
&\quad +\frac{1}{5}|A_4|\,\lambda^5k^5\cos 5(\varphi-\alpha_4)+\frac{1}{5}|D_4|\,\lambda^5k^5\cos 3(\varphi-\delta_4)+\frac{1}{5}|B_4|\,\lambda^5k^5\cos(\varphi-\beta_4) \\
&\quad +\frac{1}{6}|A_5|\,\lambda^6k^6\cos 6(\varphi-\alpha_5).+\frac{1}{6}|S_5|\,\lambda^6k^6\cos 4(\varphi-\sigma_5)+\frac{1}{6}|C_5|\,\lambda^6k^6 \\
&\quad +\frac{1}{6}|R_5|\,\lambda^6k^6\cos 2(\varphi-\rho_5)+\cdots
\end{aligned}
\tag{3.23}
$$

For the magnetic field of a perfect rotationally symmetrical electromagnetic lens, only round aberrations such as defocus (C_1) and spherical aberrations (C_3, C_5) are present, and the variation in their values is a function of spatial frequency $|\mathbf{k}|$. Non-round aberrations arise from the imperfection of the magnetic field, usually as a result of lens manufacturing and lens misalignment.

Compensation between aberrations Aberrations are different terms in the aberration function. When adjusted correctly, the collective wavefront deviation caused by multiple aberrations can be much less severe than the effect from individual aberrations alone. An example of using one aberration to compensate another is between defocus (C_1) and spherical aberration (C_3). Both of these are rotationally symmetric but have different dependencies on $\mathbf{k}$, if C_1 and C_3 are the only included aberrations, the two dimensional phase plate is equivalent to the one-dimensional radial averaged profile. As shown in Fig. 3.3, the first zero crossing in the transfer function with $C_1 = 10\,nm$, $C_3 = -30\,\mu m$ (Fig. 3.3e) is at a larger spatial frequency than that of either aberration alone (Fig. 3.3a and Fig. 3.3c) . An alternative combination (Fig. 3.3f) with both smaller $|C_1|$ and $|C_3|$ gives an even bigger plateau transfer range

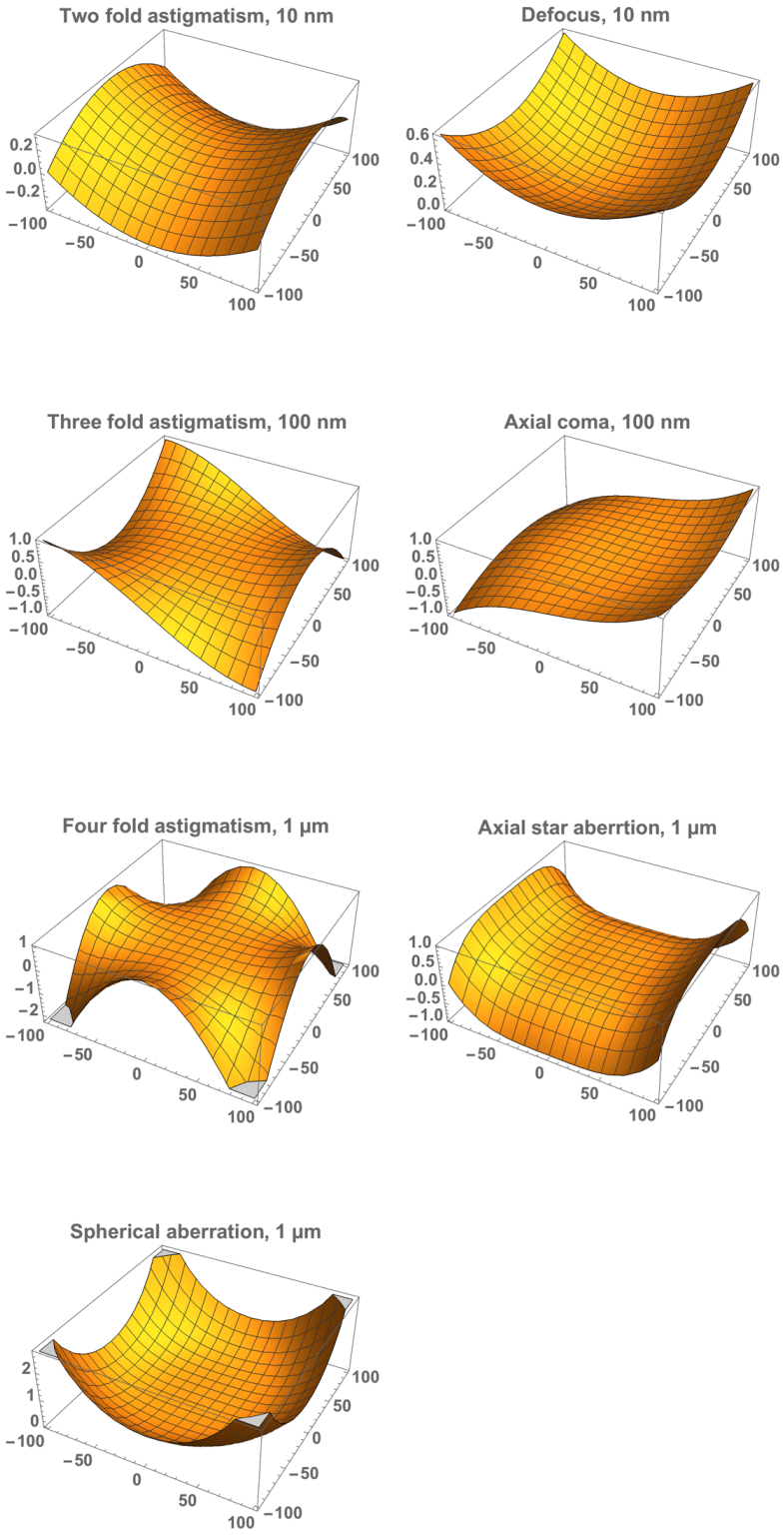


Figure 3.2: Illustration of phase plates generated by individual aberrations up to sixth order in **k** (the units of all axes are scaled to nm).

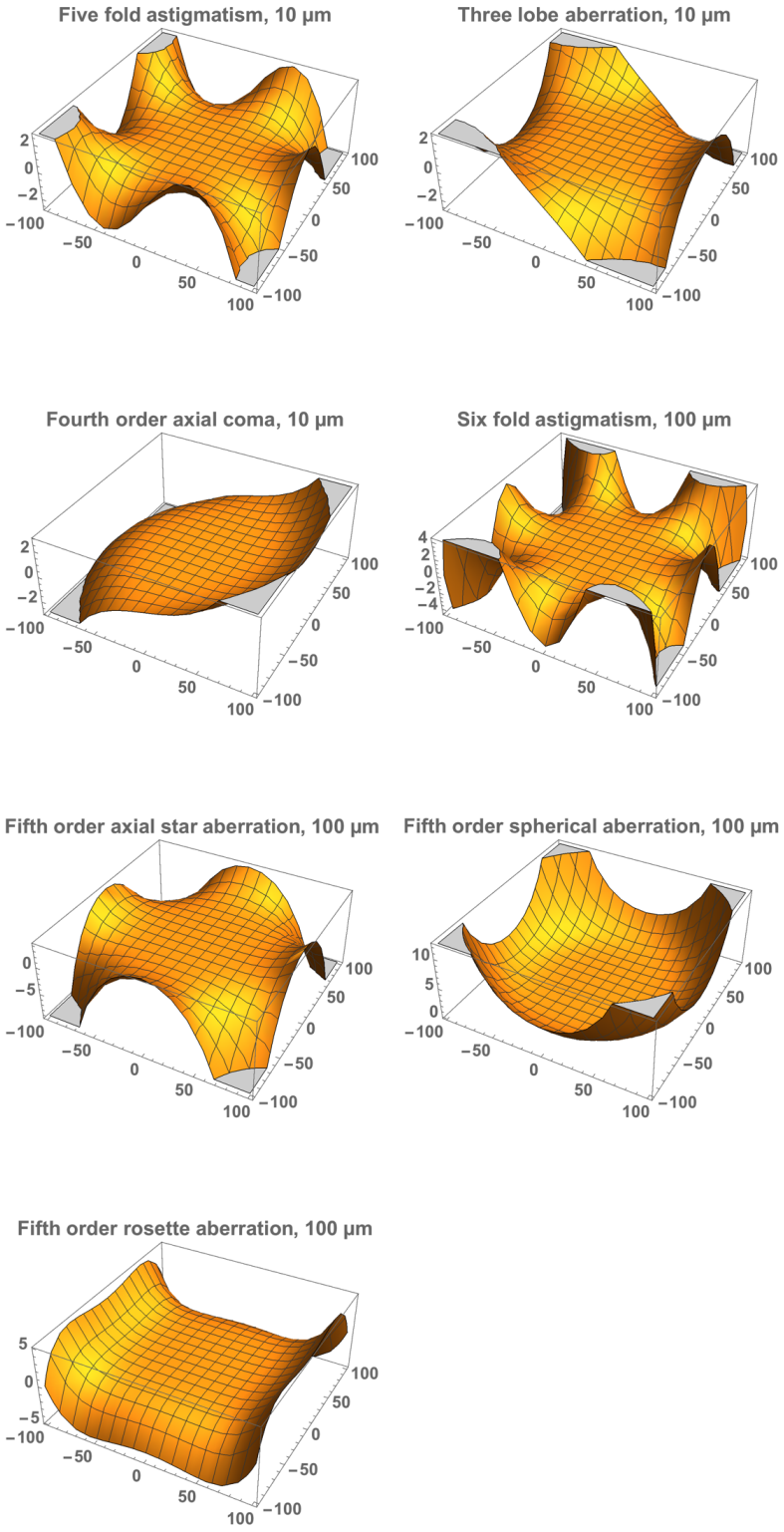
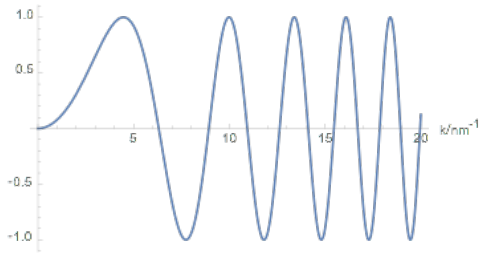
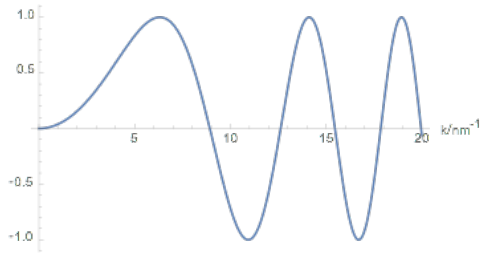


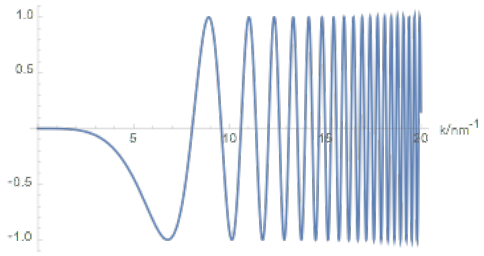
Figure 3.2: (*continued*) Illustration of phase plates generated by individual aberrations up to sixth order in $\mathbf{k}$ (the units of all axes are scaled to nm).



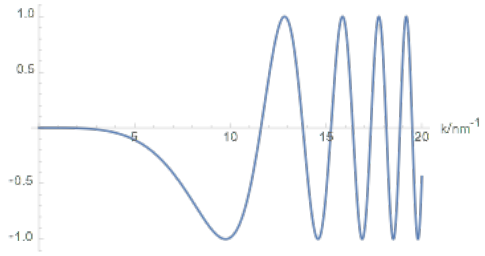
(a) $C_1 = 10\,nm$, $C_3 = 0\,\mu m$



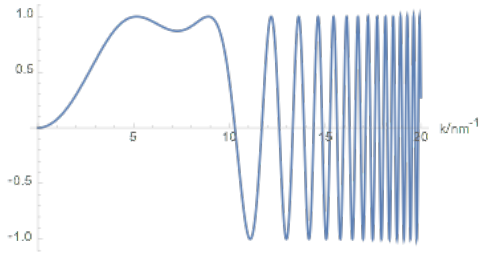
(b) $C_1 = 5\,nm$, $C_3 = 0\,\mu m$



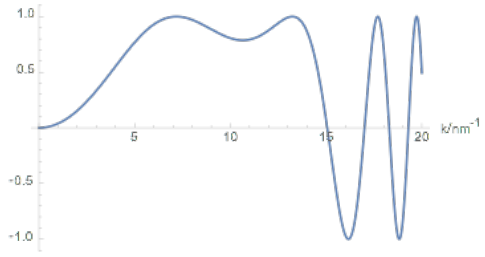
(c) $C_1 = 0\,nm$, $C_3 = -30\,\mu m$



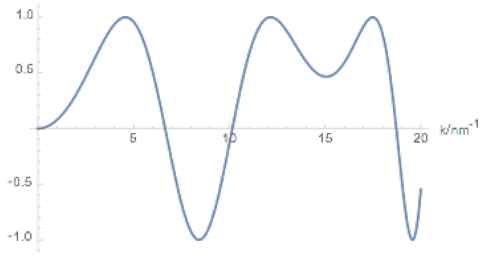
(d) $C_1 = 0\,nm$, $C_3 = -7\,\mu m$



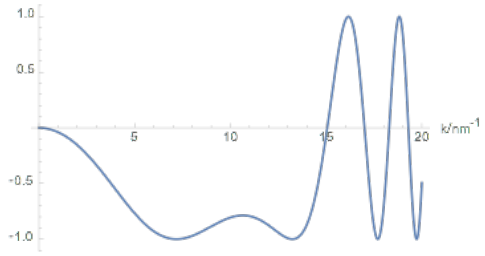
(e) $C_1 = 10\,nm$, $C_3 = -30\,\mu m$



(f) $C_1 = 5\,nm$, $C_3 = -7\,\mu m$



(g) $C_1 = 10\,nm$, $C_3 = -7\,\mu m$



(h) $C_1 = -5\,nm$, $C_3 = 7\,\mu m$

Figure 3.3: Coherent contrast transfer functions with different defocus and spherical aberration values at 200 kV.

in the contrast transfer function. However it should be noted that a larger first zero-crossing often means a smaller proportion of low-frequency information transfer and therefore this may not always be the best setting, particularly for imaging biological structures which requires strong transfer at low frequencies.

Due to the power dependence on λ and $\mathbf{k}$, the effects of high order aberrations are more visible when the lower order aberrations are restrained to small values. However the overall phase plate is more sensitive to changes in lower order aberrations, an effect that can also be seen in Fig. 3.3. The low order aberrations are unfortunately also less stable and need to be frequently adjusted during experiments, while the high order aberrations stay relatively constant over a longer period of time.

For practical reasons, a variety of guidelines for choosing the values of C_1 and C_5 have been proposed. The modified Scherzer defocus (Scherzer, 1949) is defined as;

$$C_{1,Scherzer} = -1.2\sqrt{\lambda C_3} \tag{3.24}$$

which gives the highest first zero-crossing value.

When the fifth order spherical aberration C_5 is included, the "optimised" choice (Chang et al., 2006; Kirkland et al., 2007) for C_1 and C_3 becomes;

$$C_1 = 1.56(C_5\lambda^2)^{1/3},\ C_3 = -2.88(C_5^2\lambda)^{1/3} \tag{3.25}$$

Urban et al. (2009) have also suggested the use of a negative C_3 and a corresponding positive C_1 for HRTEM. This is now easily achievable on a Cs-corrected instrument.

Measurement of aberrations To correct the aberrations of the objective lens one has to first accurately measure the aberrations. Two of the aberration measurement methods require recording a series of images with the beam illumination tilted to dif-

ferent incident angles against the original incident beam (Zemlin et al., 1978) (Zemlin, 1979) (Uhlemann and Haider, 1998). For a beam at an angle of τ from the axial direction, the new effective wave aberration function has the same form as Eq. 3.22, with the prime symbol $\prime$ indicating the new set of aberrations.

$$\begin{aligned} W'(\omega) &= W(\omega+\tau) - W(\tau) \\ &= \mathcal{R}\{A'_0\omega^* \\ &\quad +\frac{1}{2}A'_1\omega^{*2} + \frac{1}{2}C'_1\omega^*\omega \\ &\quad +\frac{1}{3}A'_2\omega^{*3} + \frac{1}{3}B'_2\omega^{*2}\omega \\ &\quad +\frac{1}{4}A'_3\omega^{*2}\omega^2 + \frac{1}{4}S'_3\omega^{*3}\omega + \frac{1}{4}C'_3\omega^{*2}\omega^2 \\ &\quad +\frac{1}{5}A'_4\omega^{*5} + \frac{1}{5}B'_4\omega^{*3}\omega^2 + \frac{1}{5}D'_4\omega^{*4}\omega \\ &\quad +\frac{1}{6}A'_5\omega^{*6} + \frac{1}{6}R'_5\omega^{*5}\omega + \frac{1}{6}S'_5\omega^{*4}\omega^2 + \frac{1}{6}C'_5\omega^{*3}\omega^3\} \end{aligned} \tag{3.26}$$

Hence, the increment in aberrations results in a expression of the new aberrations in terms of the original ones as shown in Eq. 3.27 (Haigh, 2008).

$$
\begin{aligned}
A_0' &= A_0 + A_1\tau^* + C_1\tau + A_2\tau^{*2} + \frac{1}{3}B_2^*\tau^2 + \frac{2}{3}B_2\tau^*\tau + C_3\tau^*\tau^2 \\
&\quad + A_3\tau^{*3} + \frac{3}{4}S_3\tau^{*2}\tau + \frac{1}{4}S_3^*\tau^3 + A_4\tau^{*4} + \frac{3}{5}B_4\tau^{*2}\tau^2 + \frac{2}{5}B_4^*\tau^*\tau^3 \\
&\quad + \frac{4}{5}D_4\tau^{*3}\tau + \frac{1}{5}D_4^*\tau^4 + A_5\tau^{*5} + \frac{1}{2}C_5\tau^{*2}\tau^3 + \frac{5}{6}R_5\tau^{*4}\tau + \frac{1}{6}R_5^*\tau^5 \\
&\quad + \frac{1}{3}S_5\tau^{*3}\tau^2 + \frac{2}{3}S_5\tau^*\tau^4 \qquad (3.27) \\
A_1' &= A_1 + 2A_2\tau^* + \frac{2}{3}B_2\tau + C_3\tau^2 + 3A_3\tau^{*2} + \frac{2}{3}S_3\tau^*\tau + 4A_4\tau^{*3} \\
&\quad + \frac{2}{5}B_4^*\tau^3 + \frac{9}{5}B_4\tau^*\tau^2 + \frac{12}{5}D_4\tau^{*2}\tau + 5A_5\tau^{*4} + 2C_5\tau^*\tau^3 \\
C_1' &= C_1 + \mathcal{R}\{\frac{4}{3}B_2\tau^* + \frac{3}{2}S_3\tau^{*2} + 2C_3\tau^*\tau + \frac{12}{5}B_4\tau^{*2}\tau + \frac{8}{5}D_4\tau^{*3}\} + 3C_5\tau^{*2}\tau^2 \\
A_2' &= A_2 + 4A_3\tau^* + \frac{3}{4}S_3\tau + 6A_4\tau^{*2} + \frac{3}{5}B_4\tau^2 + 10A_5\tau^{*3} + C_5\tau^3 \\
B_2' &= B_2 + \frac{9}{4}S_3\tau^* + 3C_3\tau + \frac{18}{5}B_4\tau^*\tau + \frac{9}{5}B_4^*\tau^2 + 9C_5\tau^*\tau^2 \\
A_3' &= A_3 + 4A_5\tau^* + \frac{4}{5}D_4\tau + 10A_5\tau^{*2} + \frac{10}{3}R_5\tau^*\tau + \frac{2}{3}S_5\tau^2 \\
S_3' &= S_3 + \frac{8}{5}B_4\tau + \frac{16}{5}D_4\tau^* + \frac{20}{3}R_5\tau^{*2} + 4C_5\tau^2 + \frac{16}{3}S_5\tau^*\tau \\
C_3' &= C_3 + \mathcal{R}\{\frac{12}{5}B_4\tau^* + 6C_5\tau^*\tau\} \\
A_4' &= A_4 + 5A_5\tau^* + \frac{5}{6}R_5\tau \\
B_4' &= B_4 + 5C_5\tau + \frac{10}{3}S_5\tau^* \\
D_4' &= D_4 + \frac{5}{3}S_5\tau + 5R_5\tau^* \\
A_5' &= A_5 \\
S_5' &= S_5 \\
C_5' &= C_5 \\
R_5' &= R_5
\end{aligned}
$$

It is apparent from the expression of A_0', the tilt induced image shift, that all aberrations are involved in the expression of A_0' (Eq. 3.27). This leads to a conclusion that if enough tilted images were taken and the image shifts could be accurately

evaluated, it should be theoretically possible to calculate all other lens aberrations. Practically there are two main difficulties in applying this method. The first issue is that the sample drift is indistinguishable from the image shift and will introduce large errors in the measurement of A_0'. The other issue with the method is that at high resolution, the image contrast will be more sensitive to aberration changes and almost all registration methods fail to align images whose contrast varies from one image to another, as is discussed subsequently in Section 5.4. Therefore, using this real space image shift approach to determine the lens aberrations is limited to applications at low resolution.

The widely adapted diffractogram tableau method was initially developed by Zemlin et al. (1978), after whom the method is often referred to as the *Zemlin tableau* method (Zemlin, 1979). It collects a series of images at differently tilted beam angles and calculates the power spectrum of each image, from which the defocus C_1 and the two-fold astigmatism A_1 are fitted from the rings in the power spectra of amorphous images. The method makes use of amorphous structures and it is not affected by sample drift. Another advantage of the Zemlin tableau method is that for each image taken, three observables (one from the defocus and two from the complex two-fold astigmatism) can be obtained, which makes the number of tilt angles required to fit the same number of aberrations fewer than the image shift method.

3.2.2 Partially coherent imaging

The discussion on imaging theory hitherto has assumed that the electron source of the microscope is perfectly coherent in that all electrons arise from a singular point with the same intrinsic energy (temporally coherent) and in the same direction (spatially coherent). This is not true for all existing electron sources.

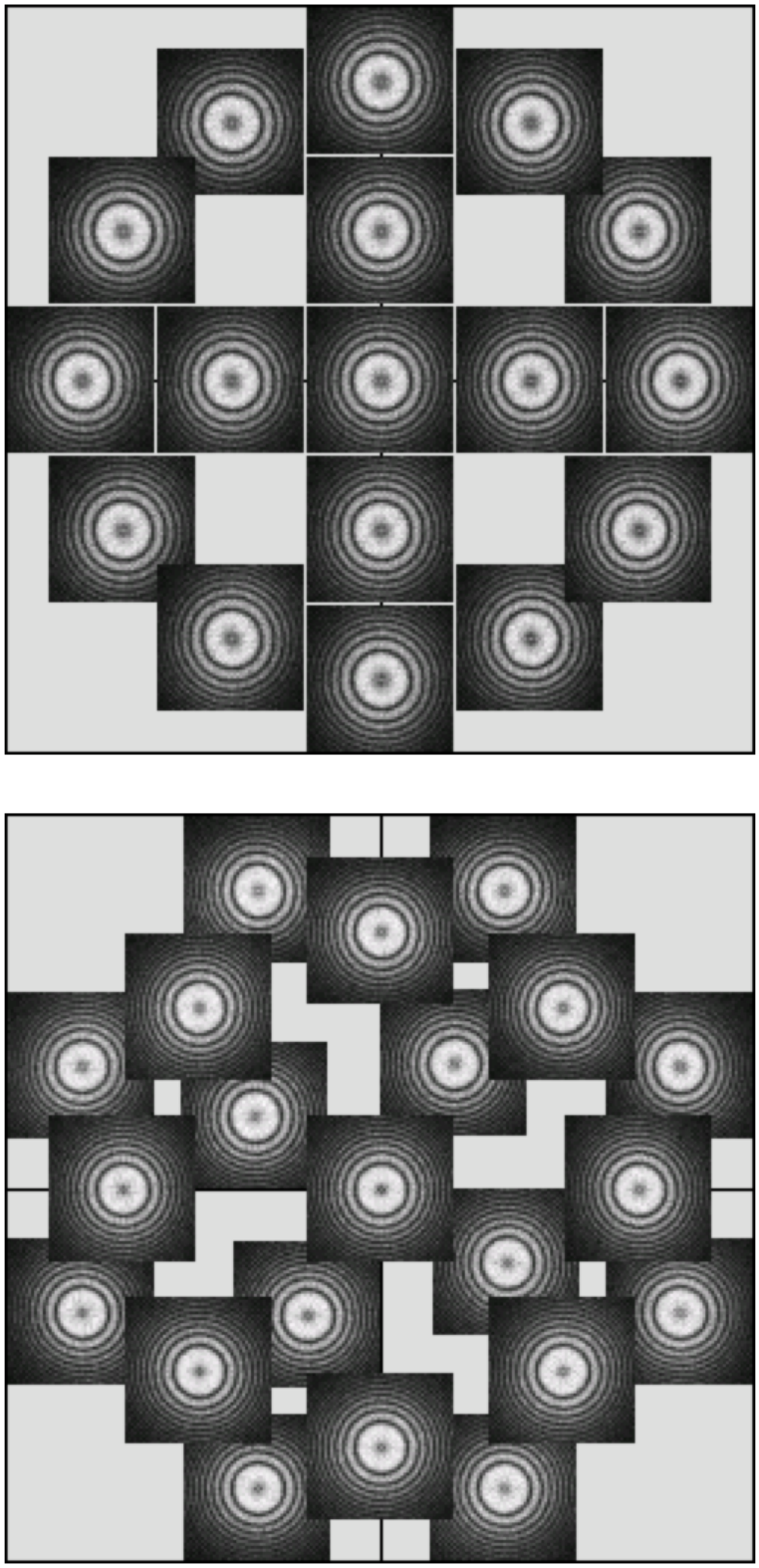


Figure 3.4: Zemlin tableau taken by the CEOS aberration corrector under "standard" mode (measuring aberrations up to A_4 from 17 image power spectra), and "extended" mode (measuring aberrations up to A_5 from 21 image power spectra)

Partial temporal coherence Electrons emitted from the source are not monochromatic. Even for the more advanced cold field-emission guns, the energy range of the electrons, ΔE, is about $0.2 - 0.3\,\mathrm{eV}$. This value is the main contributor to the energy spread in the incident beam, but can be significantly reduced by the implementation of a monochromator. In addition to the initial energy spread, the accelerating voltage too has a certain level of variation, ΔU, and the magnetic field intensity in the objective lens and spectrometer is also subject to fluctuations in the electrical current, ΔI, in the lens coils. All three factors contribute to an effective spread in the defocus.

Although the initial electron energy dispersion of thermionic emitters follows closely to an asymmetric Maxwell-Boltzman distribution (De Graef, 2003), more commonly a symmetric Gaussian distribution is adapted as the approximated form for the collective effects of all three factors above on the focus variation.

$$f(\Delta f) = \frac{2\sqrt{\ln 2}}{\sqrt{\pi} f_s} \exp[-\ln 2 (\frac{\Delta f}{f_s/2})^2] \tag{3.28}$$

Notice that $f(\Delta f)$ is normalised so that

$$\int_{-\infty}^{+\infty} f(\Delta f)\, d\Delta f = 1 \tag{3.29}$$

The full width at half maximum (FWHM) of this distribution is defined as the *focal spread* f_s,

$$f_s = C_c \frac{1 + E/E_0}{1 + E/2E_0} [(\frac{\Delta E}{E})^2 + (\frac{\Delta U}{U})^2 + 4(\frac{\Delta I}{I})^2]^{1/2} \tag{3.30}$$

where C_c is the chromatic aberration coefficient, $E_0 = m_e c^2 = 511\,\mathrm{keV}$ is the rest energy of a single electron, and E, U, and I correspond to the initial energy of electrons after extraction, the accelerating voltage and the objective lens current respectively. The reason that the lens current term has a factor of 4 is due to the focus value being proportional to the square of the field $B(z)$ (Eq. A.1), while the

other two terms are linearly related to the electron energy.

Eq. 3.30 indicates several means by which the focal spread could be minimised. Reducing the $(\frac{\Delta E}{E})^2$ term is achievable by installing a monochromator after the electron source; the reduction of fluctuations in the accelerating voltage and objective lens current is by and large an engineering problem imposed on the power supply system, which if improved will also benefit the suppression of ΔE, ΔU, and the instability of all other electron optics. Lastly, the correction of the electron optical C_c, known as the chromatic aberration, is a topic that has been under active development (Haider et al., 2009).

Partial spatial coherence A similar Gaussian distribution treatment can be applied to the spatial incoherence, only in a two-dimensional form as the source is an object with a physical shape and the possible wave vectors lie on a half sphere around the effective point source. With a divergence angle θ_d, the *beam divergence*, s_d, is defined as;

$$s(\mathbf{\Delta s}) = \frac{\ln 2}{\pi s_d^2} \exp[-\ln 2(\frac{\mathbf{\Delta s}}{s_d})^2] \tag{3.31}$$

where;

$$s_d = \frac{\theta_d}{\lambda} \tag{3.32}$$

The Gaussian distribution $s(\mathbf{\Delta s})$ is again normalised to 1.

$$2\pi \int_0^{+\infty} s(\Delta s)\Delta s \, d\Delta s = 1 \tag{3.33}$$

The partially coherent transmission cross-coefficient function The complete form of $T_{cc}(\mathbf{k}, \mathbf{k}')$ that includes the effects of partial coherence can be calculated by including the two partial coherence distributions (Eq. 3.28 and Eq. 3.31) into Eq. 3.12, substituting C_1 with $C_1 + \Delta f$, and $\chi(\mathbf{k})$ with $\chi(\mathbf{k} + \Delta s)$. The new transmission cross-

coefficient function has the following form;

$$\begin{aligned} T_{cc}(\mathbf{k},\mathbf{k}') &= \exp[-i\chi(\mathbf{k})]\exp[-i\chi(\mathbf{k}')] \\ &\int_{-\infty}^{+\infty} \exp[-i\pi\lambda(\mathbf{k}^2-\mathbf{k}'^2)]\, f(\Delta f)\, d\Delta f \\ &\int_{0}^{+\infty} \exp[-i(\nabla\chi(\mathbf{k})-\nabla\chi(\mathbf{k}')]\, s(\Delta s)\, \Delta s \end{aligned} \tag{3.34}$$

For linear imaging, the temporal and spatial coherence can be reduced to two envelope functions given as;

$$E_t(\mathbf{k}) = \exp[-(\frac{\pi\lambda f_s}{4\sqrt{\ln 2}})^2\mathbf{k}^4] \tag{3.35}$$

$$E_s(\mathbf{k}) = \exp[-\frac{(\nabla\chi(\mathbf{k}))^2 s_d^2}{4\ln 2}] \tag{3.36}$$

and the new partial coherent TCC is then;

$$T_{cc}(\mathbf{k},0) = \exp[-i\chi(\mathbf{k})]E_t(\mathbf{k})E_s(\mathbf{k}) \tag{3.37}$$

In contrast to the temporal coherence envelope function $E_t(\mathbf{k})$ which is only related to the focal spread, $E_s(\mathbf{k})$ has a dependence on both the beam divergence and the aberration function $W(\mathbf{k})$ and its actual value will therefore vary with the aberrations at different spatial frequencies.

If the non-linear components are strong, partial coherence damping effects will also become relevant for the non-linear terms in Eq. 3.14, but there is no simple analytical form available for this more complex situation. This is therefore calculated through integration of the more general TCC (Eq. 3.34).

In calculating the integration, numerical approaches are almost always needed. The image intensity of the incoherent illumination, g_{inc}, is modelled as a summation

of numerous images of coherent illumination, g_{coh} (Eq. 3.38), weighted by their proportions in the samplings. The sampling method can be either a classic equal-spaced sampling, which gives the effective TCC the form of Eq. 3.39, or a *nonuniform importance sampling* adapted from a Monte Carlo integration approach of image intensity calculation (Eq. 3.40) (Chang et al., 2005).

$$g_{coh}(\mathbf{x}, \Delta f, \mathbf{\Delta s}) = |\mathcal{F}^{-1}(\Psi_o(\mathbf{k}) \exp[-i\chi(\mathbf{k} + \mathbf{\Delta s}, C_1 + \Delta f)])|^2 \tag{3.38}$$

$$g_{inc}(\mathbf{x}) = \frac{1}{h_{\Delta f}} \frac{1}{h^2_{\mathbf{\Delta s}}} \sum_{\Delta f} \sum_{\mathbf{\Delta s}} g_{coh}(\mathbf{x}, \Delta f, \mathbf{\Delta s})\ f(\Delta f)\ s(\mathbf{\Delta s}) \tag{3.39}$$

$$g_{inc}(\mathbf{x}) = \frac{1}{h_{\Delta f}} \frac{1}{h^2_{\mathbf{\Delta s}}} \sum_{\Delta f} \sum_{\mathbf{\Delta s}} g_{coh}(\mathbf{x}, \Delta f, \mathbf{\Delta s})\ f(\Delta f)\ s(\mathbf{\Delta s}) \tag{3.40}$$

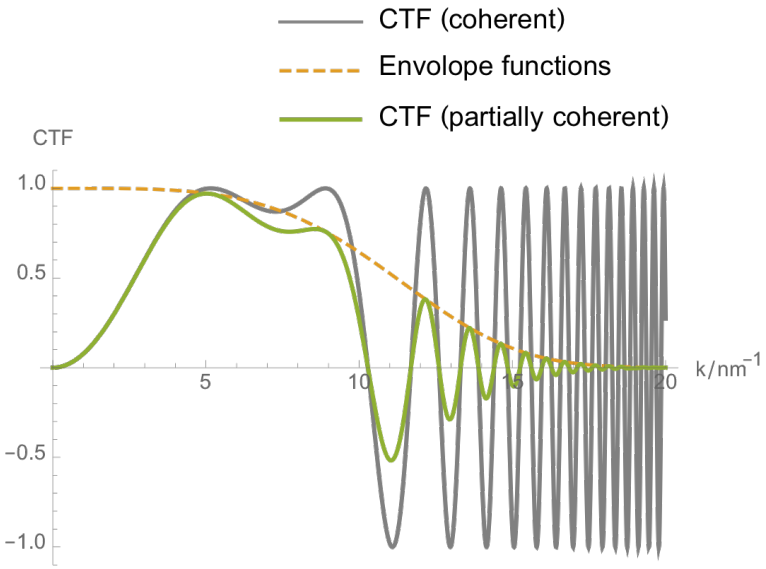
3.2.3 Detector effects

The modulation transfer function (MTF) is another envelope function to be added into the TCC. This additional factor describes how the detector responds to incoming signals as a function of spatial frequency. It varies from one detector to another and can be measured experimentally (Meyer et al., 2000), as;

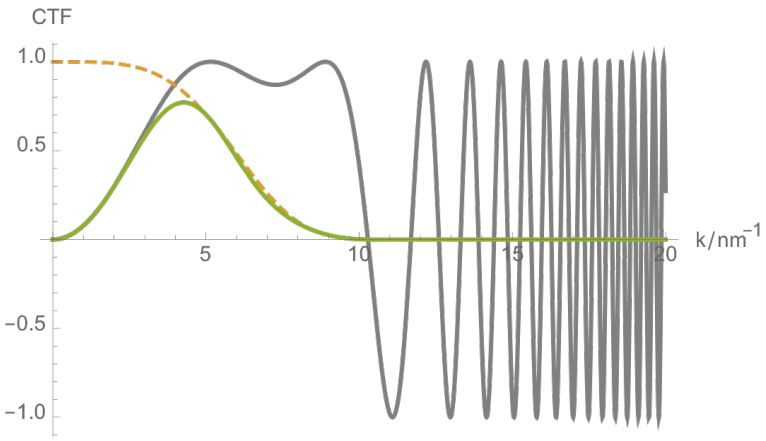
$$MTF(\mathbf{k}) = \frac{\mathcal{F}[I_{out}(\mathbf{u})]}{\mathcal{F}[I_{in}(\mathbf{u})]} \tag{3.41}$$

I_{in} and I_{out} in Eq. 3.41 denote the input and output signals respectively. For TEM images, $\mathcal{F}[I_{in}]$ and $\mathcal{F}[I_{out}]$ correspond to $\Psi_i^* \Psi_i$ and $G(\mathbf{k})$ respectively.

Nyquist frequency For a discrete signal, such as a TEM image collected by a pixelated CCD camera, the Nyquist frequency is half the sampling rate and gives the highest frequency that can be reconstructed from a recorded signal. If each pixel in



(a) $C_1 = 10\,\mathrm{nm}$, $C_3 = -30\,\mu\mathrm{m}$, $f_s = 2.8\,\mathrm{nm}$.



(b) $C_1 = 10\,\mathrm{nm}$, $C_3 = -30\,\mu\mathrm{m}$, $f_s = 10\,\mathrm{nm}$.

Figure 3.5: Influence of the temporal coherence envelope function on the transfer function.

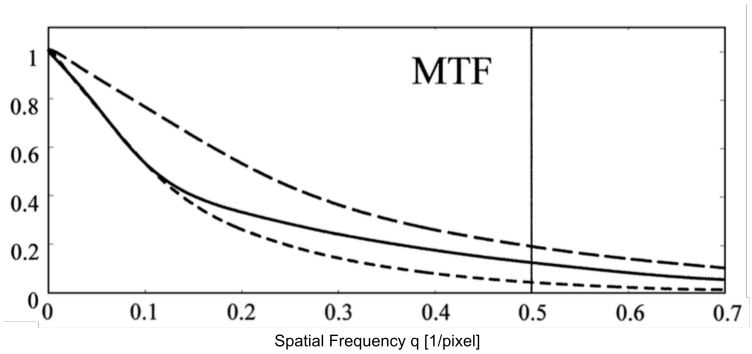


Figure 3.6: (Reproduced from (Meyer et al., 2000)) MTF measured for a 1k×1k (solid line) and a 2k×2k camera (short dashed line: no binning, long dashed line: 2×2 binning).

the image represents a distance, a, in real space given by a certain magnification of the microscope, the Nyquist frequency is then given by;

$$k_{Nyquist} = \frac{1}{2a} \tag{3.42}$$

From MTF to detector envelope function The MTF with a real spatial frequency as the argument then becomes a detector envelope function;

$$E_d(\mathbf{k}) = MTF(\frac{\mathbf{k}}{f_{Nyquist}}) \tag{3.43}$$

$$\mathbf{k} = f_{Nyquist}\mathbf{u}, \quad 0 \leq |\mathbf{u}| \leq 1 \tag{3.44}$$

To measure the MTF, both $\mathcal{F}[I_{in}]$ and $\mathcal{F}[I_{out}]$ (Eq. 3.41) need to be known. Several methods for MTF determination are available (De Ruijter and Weiss, 1992). The easiest and most common one is by recording an empty image of pure white noise with uniform illumination. The electron distribution in such an image is a stochastic process that follows Poisson statistics and therefore the attenuation at high frequencies (Fig. 3.6) in the power spectrum is the sole result of the MTF. Because the detector property should have no directional effect, the power spectrum should be rotationally symmetric and can be fully represented by a radial profile. The MTF is then the

squared root of this radial profile. Alternatively, the sinusoidal fringes generated by electron holography can be used as a deterministic input signal for the MTF measurement. However, for a given configuration, there is only one particular set of spatial frequencies in the hologram, and hence multiple measurements are needed to provide enough data points before an estimate can be made. Another method of MTF measurement involves using a sharp edge feature in an image (Meyer et al., 2000; Zuo, 1996; Daberkow et al., 1991). Ideally, the edge should be made of Al or platinum to minimise the hard X-ray generation, and placed close to the detector in an inclined position relative to the pixel arrays, so that aliasing is avoided. If the edge image is sufficiently sharp, the highest resolution information at the edge will surpass the Nyquist limit, and hence the MTF across the whole range of spatial frequency can be determined from the power spectrum.

It has been argued that the MTF is the origin of the so-called *Stobbs factor*, a discrepancy between the experimental image intensity and the simulated image intensity (Thust, 2009).

Apart from the modulation transfer function, another key factor in detector performance evaluation is the detective quantum efficiency (DQE). This describes the detector's impact on the signal-to-noise ratio before and after the image acquisition. Given an input image, SNR_{in}, if the output image has a signal-to-noise ratio, SNR_{out}, then the DQE is defined as

$$DQE = \frac{(SNR_{out})^2}{(SNR_{in})^2} = (\frac{S_{out}N_{in}}{S_{in}N_{out}})^2 \tag{3.45}$$

3.2.4 Resolution and information limit

Although resolution is one of the most fundamental concepts in microscopy, there is no universal agreement on which definition should be exclusively used (den Dekker and van den Bos, 1997). The classical two-point resolution defined by John William

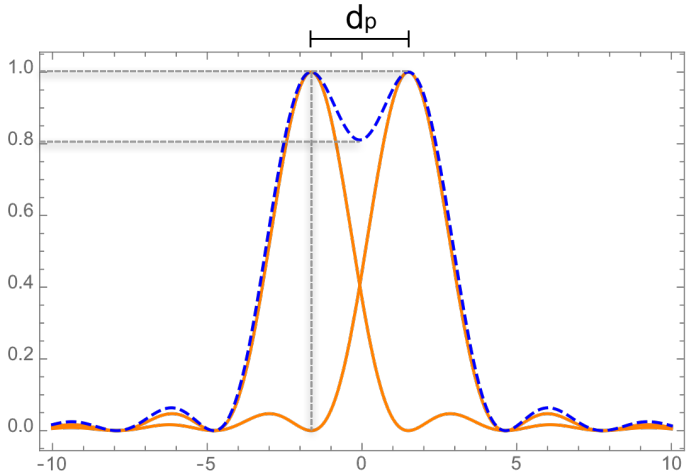


Figure 3.7: Rayleigh criterion for image resolution. The blue dashed line is the composite intensity of the two orange intensity profiles

Strutt, or Lord Rayleigh, for whom the definition is better known as the *Rayleigh criterion*, describes two points modelled by Delta functions (Fig. 3.7). When the one of the point's maximum meets the first minimum of the other point, it is then considered the two points are *resolved*. When Rayleigh criterion is satisfied, the lowest point between the two peaks is of 81% of the intensity of the peak maximum. This definition of resolution is more suitable for incoherent imaging systems, such as STEM, while for coherent imaging system, an ideal TEM for example, whether the two point sources are in phase or not will have a significant impact on the combined intensity (Van Aert et al., 2006).

The context of Rayleigh criterion was originally the diffraction points of two light sources, therefore when adopted in analysis of electron microscope images, the two Delta functions are usually replaced by 2D Gaussians. In that case when the composite intensity at the central "dip" is 81% of the maxima on either side, the distance d_p between the two maxima is recognised as the resolution.

There has also been discussion of resolution in the frequency domain. The use of a *diffraction limit* for resolution measurement looks at the transfer cutoff frequency

f_c in the transfer function. In the case of TEM imaging, such a cutoff frequency is also called the *information limit*. It has been demonstrated that the diffraction limit definition of resolution is closely related to the Rayleigh criterion (Van Aert et al., 2006) and the two have the following relation;

$$f_c = \frac{1}{d_p} \tag{3.46}$$

The cutoff frequency in the transfer function can be limited by the wave aberrations of the lens or by the partial coherence envelope functions (Fig. 3.5). When detector effects are considered, the sampling choice and the detector functions can also contribute.

It is helpful to have a concise equation describing the linear imaging process (Eq. 3.47) alongside a simplified illustration that relates Eq. 3.47 to the actual microscope components (Fig. 3.8).

$$G_i(\mathbf{k}) = \Psi_o(\mathbf{k})[\exp[-\frac{2\pi i}{\lambda}W(\mathbf{k})]E_t(\mathbf{k})E_s(\mathbf{k})E_d(\mathbf{k})] \tag{3.47}$$

When attempting to reach the ultimate spatial resolution on a specific microscope, it is important to identify the primary resolution limiting factor. The effective transfer range of both temporal coherence and spatial coherence envelope functions can be improved by choosing a field-emission source over the traditional thermionic emitter. However in an uncorrected TEM with a large intrinsic spherical aberration causes the transfer function to oscillate between positive and negative values within the information limit even if the aberration setting is optimised (Fig. 3.5a). Similarly, a properly adjusted aberration corrector moves the first zero crossing of the wave aberration function to a higher spacial frequency. However, if the first crossing point of CTF shifts beyond the information limit imposed by the partial coherence of the electron source, the overall transfer function will drop to zero at spatial frequencies

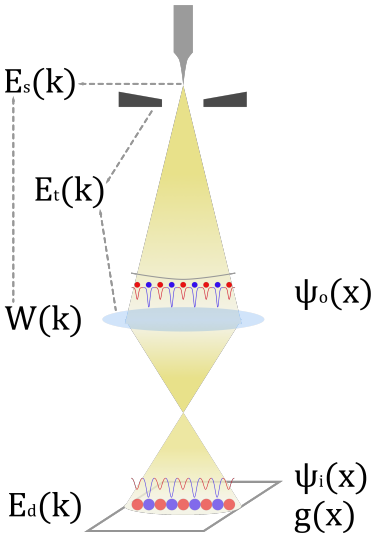


Figure 3.8: Relationships between the imaging equation and the microscope components. $W(\mathbf{k})$ is the wave aberration function, $E_t(\mathbf{k})$ the temporal incoherence envelope function, $E_s(\mathbf{k})$ the spatial incoherence envelope function, and $E_d(\mathbf{k})$ the detector envelope function

higher than the information limit(Fig. 3.5b).

The influence of the detector is adjustable, in that the MTF has a much stronger damping effect at frequencies close to the $f_{Nyquist}$. When working at low magnifications, this may cause a loss in high spatial frequency information that could have been better included at higher magnifications, albeit at the expense of a smaller field of view.

3.3 Exit Wave Reconstruction

Since the microscope induces wavefront deviation and damping effects on the electron wave, it is natural to think of procedures that may potentially compensate the imperfect transfer function with post-experimental processing to give direct access to the electron wave at the exit plane of the sample. This gives rise to the development of the "indirect" method of exit wave reconstruction.

3.3.1 Linear restoration filters

Various versions of exit wave reconstruction theory have been developed, but the basic idea behind them can all be, to a certain extend, traced back to the early wave retrieval methods proposed by Schiske (1968).

Again, we assume that the linear imaging condition is satisfied and that the exit wave function, $\psi_o(\mathbf{x})$, is purely imaginary (taking away the "1" in Eq. 3.6; i.e. $\delta(\mathbf{k})$ in Eq. 3.13) and therefore its Fourier transform is conjugate anti-symmetric;

$$\Psi_o^*(-\mathbf{k}) = -\Psi_o(\mathbf{k}) \tag{3.48}$$

Linear imaging contrast in Fourier space is equal to the sum of $\Psi_i(\mathbf{k})$ and $\Psi_i^*(-\mathbf{k})$. Considering the transfer functions as known and the noise in the image added as $n(\mathbf{k})$, two images are needed to determine the exit wave $\Psi_o(\mathbf{k})$ and its conjugate $\Psi_o^*(-\mathbf{k})$. The image contrast becomes;

$$\begin{aligned} G(\mathbf{k}) &= \Psi_i(\mathbf{k}) + \Psi_i^*(-\mathbf{k}) + n(\mathbf{k}) \\ &= \Psi_o(\mathbf{k})T(\mathbf{k}) + \Psi_o^*(-\mathbf{k})T^*(-\mathbf{k}) + n(\mathbf{k}) \\ &= \Psi_o(\mathbf{k})[T(\mathbf{k}) - T^*(-\mathbf{k})] + n(\mathbf{k}) \end{aligned} \tag{3.49}$$

Given;

$$\exp(-i\chi(\mathbf{k})) = \cos(\chi(\mathbf{k})) - i\sin(\chi(\mathbf{k})) \tag{3.50}$$

When only defocus and spherical aberrations are considered;

$$\chi(-\mathbf{k}) = \chi(\mathbf{k}) \tag{3.51}$$

Then;

$$\begin{aligned} T(\mathbf{k}) - T^*(-\mathbf{k}) &= (\exp(-i\chi(\mathbf{k})) - \exp(i\chi(-\mathbf{k})))E_t(\mathbf{k}) \\ &= \{[\cos(\chi(\mathbf{k})) - i\sin(\chi(\mathbf{k}))] - [\cos(\chi(-\mathbf{k})) + i\sin(\chi(-\mathbf{k}))]\}E_t(\mathbf{k}) \\ &= -2i\sin(\chi(\mathbf{k}))E_t(\mathbf{k}) \end{aligned} \tag{3.52}$$

Defining an effective transfer function T_{eff} as Eq. 3.53;

$$T_{eff} = -2i\sin(\chi(\mathbf{k}))E_t(\mathbf{k}) = -2i\sin(\frac{2\pi}{\lambda}W(\mathbf{k}))E_t(\mathbf{k}) \tag{3.53}$$

The obvious way to retrieve an approximated exit wave $\Psi_0(\mathbf{k})$ then appears to be to simply apply an linear filter, r, to the image contrast with r defined as;

$$r = \frac{1}{T_{eff}(\mathbf{k})} = \frac{i}{2\sin(\chi(\mathbf{k}))E_t(\mathbf{k})} \tag{3.54}$$

The obvious shortcoming of this simple treatment is that the effective transfer function $T_{eff}(\mathbf{k}) = T(\mathbf{k}) - T^*(-\mathbf{k})$ may be zero at a number of frequencies, causing an amplification of image noise. Fig. 3.9a visualises this noise amplification problem with a simplified version of T_{eff},

Fig. 3.9 shows that the filter r generates strong transfer of the restored signal Ψ_r when $T_{eff}(\mathbf{k})$ reaches zero, where there is little information in the original image. Therefore, only the restored noise n_r (red line in Fig. 3.9) around these singularity points will experience this amplification.

$$\Psi_r = \frac{|T_{eff}|}{|T_{eff}| + n} \tag{3.55}$$

$$n_r = \frac{n}{|T_{eff}| + n} \tag{3.56}$$

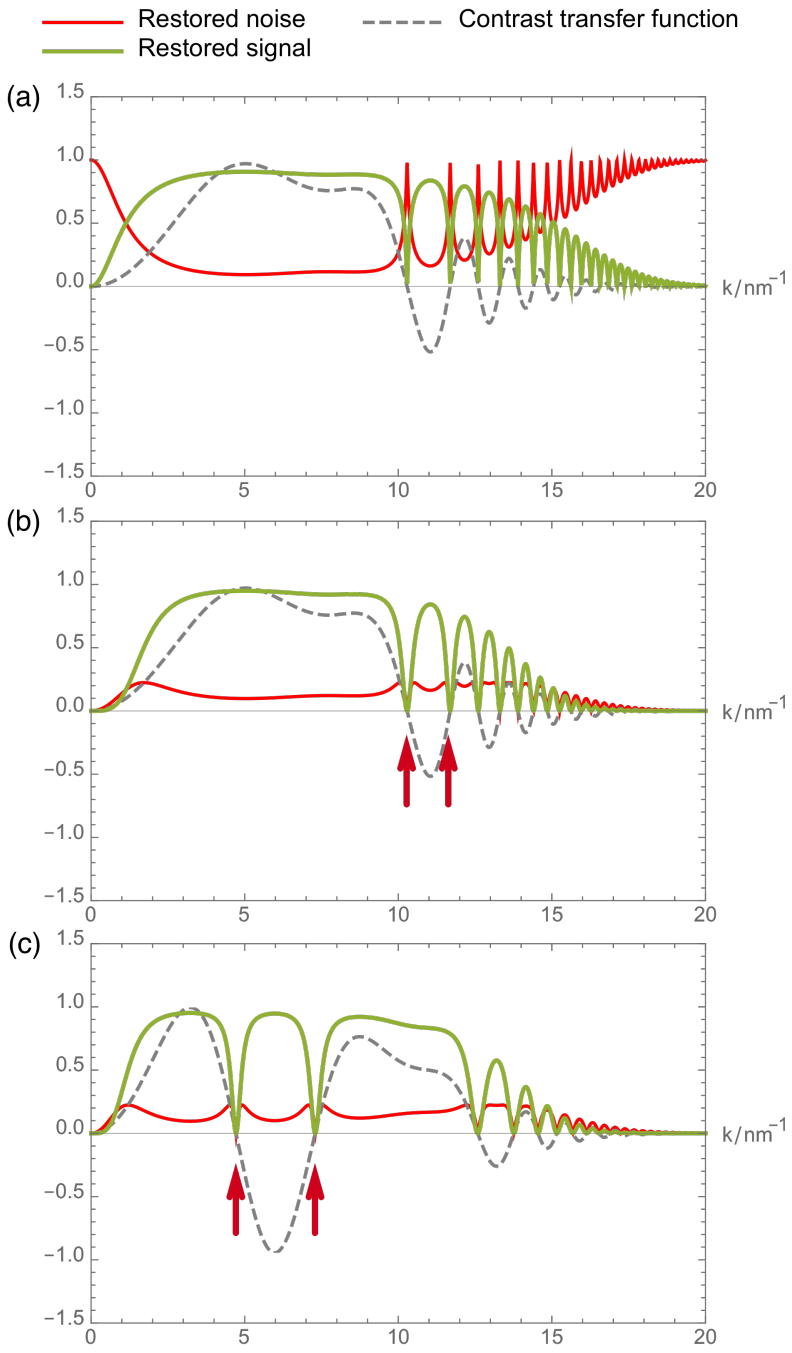


Figure 3.9: (a) $C_1 = 10\,nm$, $C_3 = -30\,\mu m$. The simple division approach for wave restoration and the consequent noise amplification issue. (b) Restored signal using a Wiener filter, $C_1 = 10\,nm$, $C_3 = -30\,\mu m$ (c) Restored signal using a Wiener filter, $C_1 = 20\,nm$, $C_3 = -30\,\mu m$. The restored noise component is suppressed when contrast transfer function is close to 1 or very small. The red cursor in (b) and (c) indicates that zero-crossing positions are different in images at different defocus values, compensating each other in the final restored transfer function after summation. Noise component $n(\mathbf{k})$ is assumed to be 0.2 across all frequencies.

Alternatively, if a Wiener filter (Wiener, 1949) is adopted (Eq. 3.57), the transfer of noise at these zero transfer points will be also suppressed, as shown in Fig. 3.9b.

$$r = \frac{T^*_{eff}}{T^2_{eff} + n} \tag{3.57}$$

The restored signal Ψ_r and restored noise n_r are then;

$$\Psi_r = \frac{T^2_{eff}}{T^2_{eff} + n} \tag{3.58}$$

$$n_r = \frac{T^*_{eff} n}{T^2_{eff} + n} \tag{3.59}$$

It is apparent that the restored wave from only one image will not be able to include information at those zero transfer frequencies. This can be solved by performing the same restoration on another image at a different defocus whose zero transfer frequencies are also located differently (Fig. 3.9c).

An improved linear restoring filter based on a Wiener filter method was proposed by Saxton (Saxton, 1988; Saxton and Smith, 1985) to cope with the low transfer problem described above.

$$\begin{aligned} G(\mathbf{k}) &= G_b(\mathbf{k}) + G_l(\mathbf{k}) + n(\mathbf{k}) \\ &= \Psi_i(\mathbf{k}) + \Psi^*_i(-\mathbf{k}) + n(\mathbf{k}) \\ &= \Psi_o(\mathbf{k})T(\mathbf{k}) + \Psi^*_o(-\mathbf{k})T^*(-\mathbf{k}) + n(\mathbf{k}) \end{aligned} \tag{3.60}$$

With N images taken under different imaging conditions, generally a focal series (Schiske, 1968) or a tilt azimuth series (Kirkland et al., 1995), the exit wave function is overdetermined and can be estimated from the image intensities by applying a linear filter to each image in the series and calculating the sum. The construction

of the filter aims to suppress the conjugate wave and the noise $n(\mathbf{k})$. Searching for such a filter can be achieved by least-squares fitting to minimise the difference between the estimated exit wave and the calculated exit wave using the linear filter. Dropping the argument $\mathbf{k}$ in the functions ($-\mathbf{k}$ is represented by a subscript "_") for conciseness, an evaluation cost function M can be defined as;

$$M = \left\langle \left| \sum_{i=1}^{N} r_i G_i - \Psi \right|^2 \right\rangle \tag{3.61}$$

in which $\langle \cdots \rangle$ denotes an ensemble average. Solving the following derivative of M (Eq. 3.62) should give the solution to the linear filter for the restoration.

$$\frac{dM}{dr_j^*} = \left\langle [\sum_{i=1}^{N} r_i G_i - \Psi] G_j^* \right\rangle = 0 \tag{3.62}$$

Bringing Eq. 3.49 into Eq. 3.62, the expanded expression of Eq. 3.62 would contain several different terms below:

$\langle \Psi \Psi^* \rangle$: the expected power spectrum of the object; the ideal image with uniform transfer;

$\langle \Psi \Psi_- \rangle$: when no assumption is made about the wave, this term should be zero;

$\langle \Psi n_i \rangle$: the correlation between the object and the noise, which should average to zero for random noise;

$\langle n_i n_j \rangle$: the correlation between noise in images i and j, which should also be approximately zero when $i \neq j$.

Eq. 3.62 will then be simplified as;

$$T_j^* \Psi \sum r_i T_i - T_j^* \Psi + T_{j-} \Psi \sum r_i T_i + N r_j = 0 \tag{3.63}$$

The resultant filter r_i has the following form:

$$r_i = \frac{T_i^*(\sum |T_{i-}|^2 + \nu) - T_{i-} \sum T_i T_{i-}}{(\sum |T_{i-}|^2 + \nu))(\sum |T_i|^2 + \nu)) - (\sum T_i T_{i-})^2} \tag{3.64}$$

with the noise power given as a fraction of the object power $\nu = \frac{\langle n_i n_i^* \rangle}{\langle \Psi_i \Psi_i^* \rangle} = \frac{N}{\Psi}$.

To restore the exit wave with this improved filter, one can simply sum up all the images multiplied respectively by their corresponding filters.

$$\Psi_r = \sum_{i=1}^{N} r_i G_i \tag{3.65}$$

The initial restored exit wave function can be iterated using Eq. 3.62 and the refinement of the reconstruction essentially further suppresses the non-linear terms and the noise, by including more images in the image series. For N images, the non-linear terms and noise term decrease at a rate of $\sqrt{N}$ for Poisson noise.

3.3.2 The parabola method (PAM)

A *parabola* method based on a three-dimensional Fourier transform interpretation of the electron wave was also proposed as an advancement of the traditional linear method (Op de Beeck and Van Dyck, 1996; de Beeck et al., 1996; Thust et al., 1996b,a). This approach introduces a quasi-coherent approximation to resemble the complete form of the transmission cross-coefficient (TCC) function when calculating the image intensity in order to more accurately simulate the non-linear components in the images. The main improvement is in calculation speed over other methods which has gradually become less attractive as the number of input images is often small (Saxton, 1994a) and computing power has significantly increased in the last few decades and is no longer a limiting factor for most linear reconstruction methods.

3.3.3 The multiple input maximum *a posteriori* method

For thick samples, the validity of the linear image assumption can be compromised, and more general reconstruction strategies are needed. Based on the improved method of maximum *a posteriori* (MAP) restoration originally developed in optical image reconstruction (Hunt, 1977; Trussell and Hunt, 1979), an exit wave reconstruction method suitable for reconstructing both linear and non-linear components was proposed (Kirkland, 1984; Kirkland et al., 1985). By generalising conventional MAP theory, which only allows for a single input image, to solve the reconstruction problems for an image series, this new method (known as multiple input MAP, or simply MIMAP) provides a rigorous treatment for non-linear image reconstruction. The application of this scheme on real experimental data reportedly obtained a $30 - 50\%$ resolution improvement (Kirkland et al., 1985). Some other researchers who had applied this MIMAP method reported it being computing-power demanding, as the number of iterations to reach the convergence can be prohibitively large (Huang and Ximen, 1991; Gribelyuk and Hutchison, 1992).

3.3.4 The maximum-likelihood method

The maximum-likelihood (MAL) method proposed in (Coene et al., 1992, 1996) follows the pioneering work of Kirkland on the original MIMAP method (Kirkland, 1984; Kirkland et al., 1985). This method concentrates on improving the computational efficiency of the original method and on decoupling the exit wave function and its complex conjugate. In theory, the MAL method should perform best when the initial guess provided is close to the true exit wave. This can be partially achieved by calculating an initial exit wave using one of the linear methods and using this wave to seed the MAL algorithm. This approach can also help the method avoid traps of local minima. A workflow digram of MAL is given in Fig. 3.10.

In addition to the reconstruction methods described above, attention has also

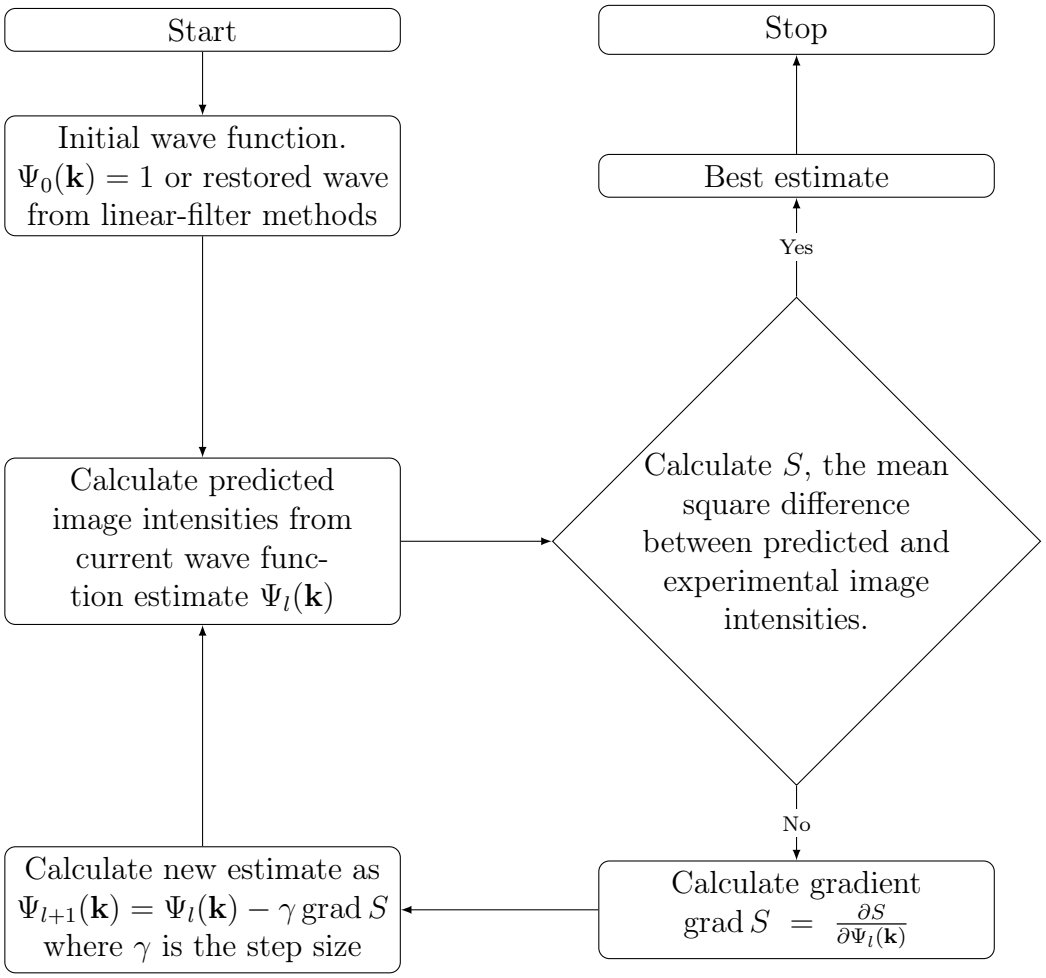


Figure 3.10: Workflow diagram of the MAL method.

been paid to quantitative interpretation of reconstructed exit waves and extending the technique to three dimensional investigation of materials. While the theoretical work has produced attractive outcomes, experimental applications on real data are still limited to samples of simple structures such as graphene (Van Dyck et al., 2012).

3.3.5 Software for exit wave reconstruction

There are a number of commercially available software packages that can be used for performing exit wave reconstruction. The subsequent chapters have used the first two packages as the tools for reconstruction tasks in this thesis. The other packages are seen in the literature from various sources, and are included here for completeness.

FTSR for DigitalMicrograph™ has been developed for both focal series reconstruction and tilt series reconstruction using an improved linear Wiener filter algorithm and a novel phase-compensated PCF/PCI method for focus determination (Meyer et al., 2002) (Meyer et al., 2004).

SEMPER is a digital image processing environment developed by Saxton et al. (1979). The environment uses its own programming language, which bears a relationship to Fortran. Several exit wave reconstruction methods were programmed in SEMPER, including the original Wiener filter method later transferred into FTSR and the MAL method with a Monte-Carlo integration for image intensity calculation (Chang et al., 2005).

Iterative Wave Function Reconstruction (IWFR) is a software package based on the reconstruction method of its namesake. It is essentially a linear method using an iterative approach to update the estimated wave and propagate the newly obtained wave to different defocus level where simulated images are generated for intensity comparison (Gerchberg and Saxton, 1972; Fienup, 1982; Allen et al., 2004; Ishizuka, 2013). Although claiming to be robust against noise, it is unclear how the software addresses the noise issue when the transfer function is close to zero.

TrueImage is a reconstruction package developed by FEI (Kübel and Thust). Its uses the parabola method as a first step of reconstruction, whose result is then used as the input wave for a further refinement stage using the maximum-likelihood method (MAL), which attempts to recover the non-linear components in the wave (Coene et al., 1996).

3.3.6 Focal/tilt series acquisition

Two acquisition geometries are commonly used in image series collection for exit wave reconstruction - *focal series acquisition* and *tilt series acquisition*.

A focal series only alters one aberration - the defocus of the images. All other microscope settings should be determined beforehand and kept constant over the series acquisition. The restriction on the choice of starting focus, focal step, and the range of the focus values mainly comes from the spherical aberration. The combinations of the spherical aberration and the defoci should aim to be sufficiently different that all spatial frequencies are strongly transferred by one or more images in the series. For an uncorrected TEM, the spherical aberration can be large and therefore the change in defocus, both in step and range, would need to be larger than those for an aberration-corrected TEM. Most researchers choose to use a through-focus (focal series) acquisition strategy (Kirkland and Meyer, 2004; Coene et al., 1992; Zandbergen and Van Dyck, 2000). Unequal step focal series, which can potentially balance the proportions of high spatial frequency and low spatial frequency information with fewer images compared to a equal-step focal series needs, has also been proposed for exit wave reconstruction (Haigh and Kirkland, 2010; Miedema et al., 1994).

The tilt series acquisition geometry records a series of images with differently tilted illumination settings (Kirkland et al., 1995, 1997). By combining images from different tilt angles, this method effectively increases the size of the virtual Fourier aperture and extends the effective information limit to a higher spacial frequency range than the axial limit.

A hybrid tilt-focal series acquisition were explored as an attempt to take advantages of both acquisition geometries, as illustrated in Fig. 3.11 (Kirkland and Meyer, 2004).

The number of images in an image series is typically no less than three (Ishizuka and Allman, 2005). Although there is no theoretical upper limit, the benefit of using

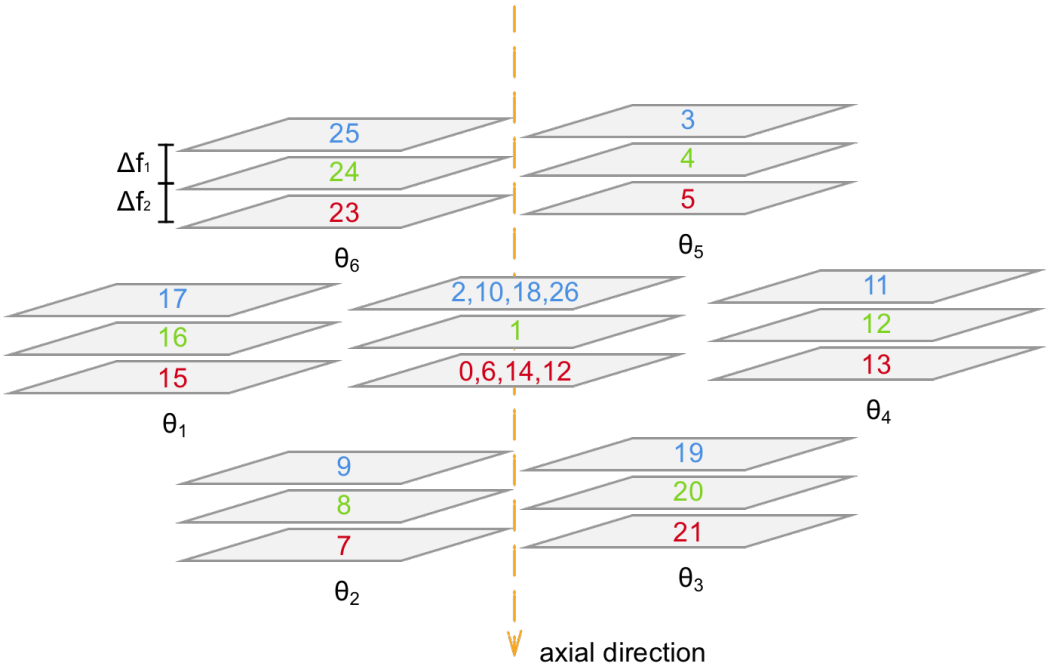


Figure 3.11: Schematic illustration of the tilt-focal series acquisition. Images $(0-26)$ are taken in the numbered sequence. Each set of images are recorded with the beam tilt angle displayed at beneath them. Images, whose numbers share the same colour, are at taken at the same defocus.

more images to improve the reconstruction quality diminishes quickly, given that the reduction of non-linear image components and Poisson noise follows a square root dependence on the number of images (Saxton, 1994a). Therefore, practical reconstructions are rarely performed with more than 64 images, which brings down the non-linear and noise components by a factor of 8. It should also be noted that, if a dose budget is imposed due to radiation damage, dividing the budget into N low-dose images will lower the signal-to-noise ratio to $1/N$ in comparison with a single acquisition of the same total dose as the series, although the dose rate is also reduced by a factor of N, which could be beneficial if the radiation damage of the investigated material is dose-rate dependent.

3.4 Instrumentation

The transmission electron microscope (TEM) has become a powerful instrument with a large number of complex components (Fig. 3.12). This section will cover the essential parts of a TEM, with emphasis on their relationship to radiation damage control and aberration correction.

3.4.1 Source

There are two general types of electron sources, the more conventional and economical thermionic-emission emitters and the more advanced but expensive field-emission sources.

Thermionic source By applying a voltage V_{heat}, the filament in a thermionic electron gun is heated to a high temperature, T, at which the electrons inside the filament materials accumulate enough energy to overcome the work function barrier, Φ and escape from the filament surface. The Wehnelt cylinder and the gun bias together

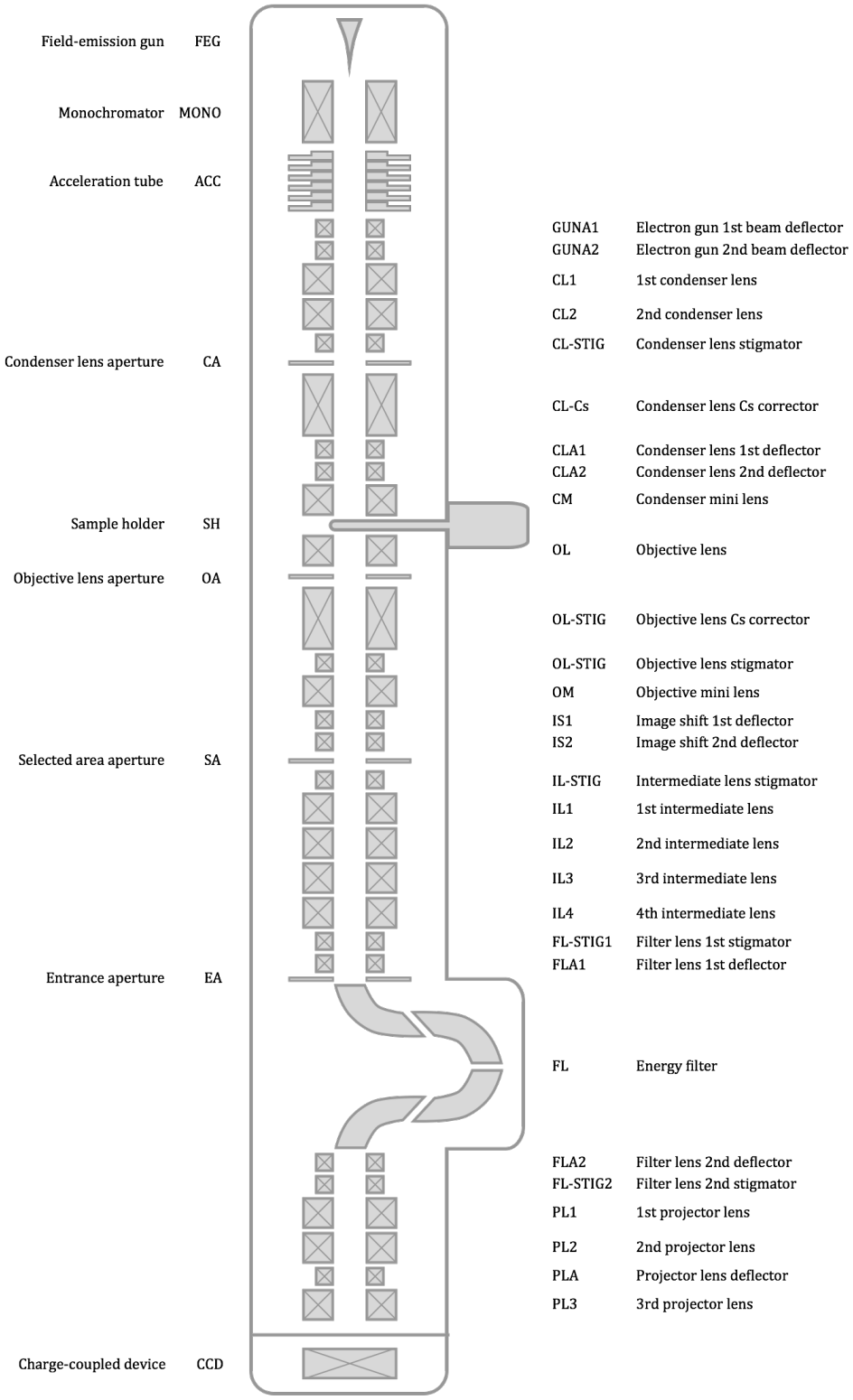


Figure 3.12: Schematic diagram of the JEOL2200MCO optical column

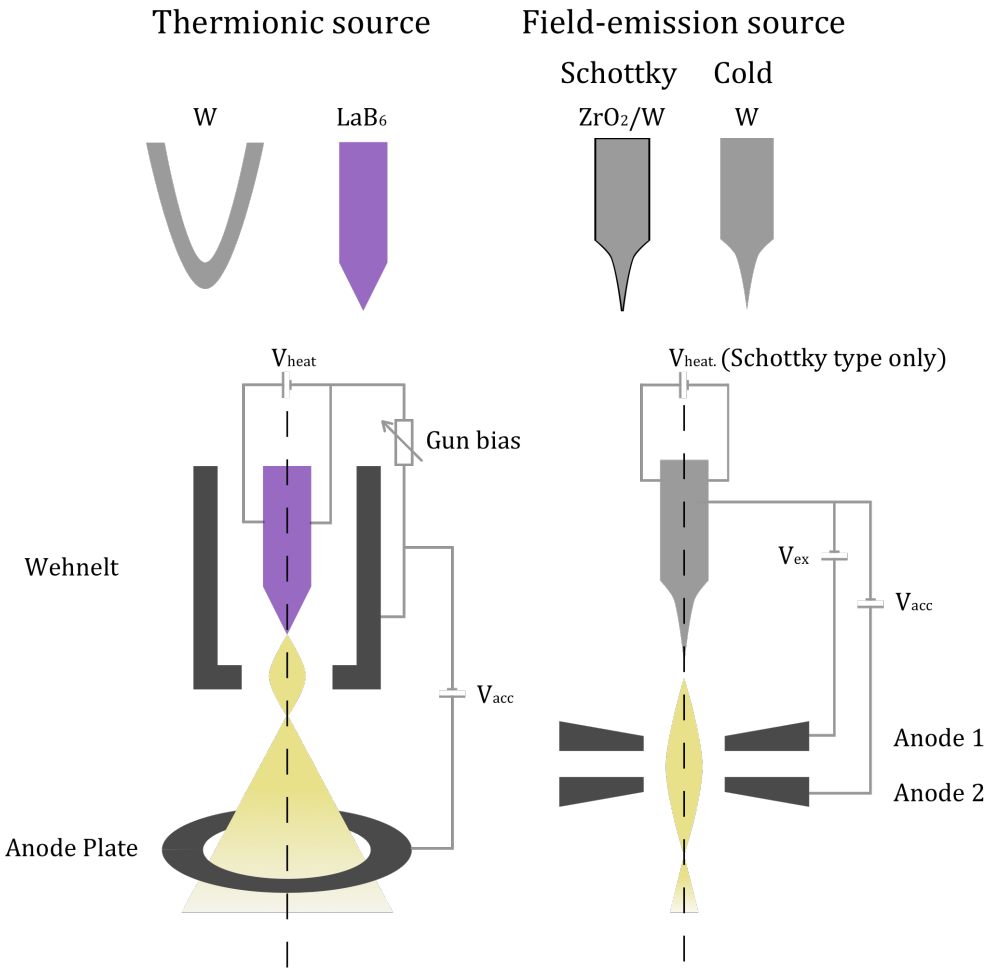


Figure 3.13: Types of electron source.

generate an inward focusing field that focuses the electrons leaving the filament. The material of which the filaments are made needs to have a high melting point, for example Tungsten at 3660 K, to survive the high heating temperature and a low work function, such as that of the lanthanum hexaboride (2.4 eV), so that the electrons have a high probability of leaving the cathode. The emission current density J can be calculated as;

$$J = \frac{4\pi m_e e}{h^3}(kT)^2 e^{-\frac{\Phi}{kT}} \tag{3.66}$$

As can be seen from Eq. 3.66, with $k = 8.62 \times 10^{-5}\,\mathrm{eV\,K^{-1}}$ being the Boltzmann constant, $h = 6.63 \times 10^{-34}\,\mathrm{m^2\,kg\,s^{-1}}$ being the Planck constant. For a specific thermionic source and Wehneld configuration, J is only determined by the working temperature of the source.

Field-emission source In the presence of a strong electrostatic field E around the filament tip, electron can be extracted from the tip without needing to heat the filament. The strength of the electric field is governed by the extraction voltage V_{ex}, and the effective radius of the tip r, as;

$$E = \frac{V_{ex}}{r} \tag{3.67}$$

The extraction voltage is typically only a few kilovolts in practice. The tip of the filament in a field-emission gun is specially manufactured into a sharp single crystal to reduce the effective radius. The crystal is oriented such that electrons are emitted from a low work function surface, such as W$\{3\,1\,0\}$. In addition, the small tip radius is also the reason that field-emission guns generally have a superior spatial coherence compared to thermionic sources.

Whether the filaments in the field-emission guns are heated or not further divides them into two subtypes: *Schottky FEGs* and *cold FEGs*. The Schottky FEGs work

Type of electron gun	Tungsten	LaB_6	Schottky FEG	Cold FEG
Work function (kV)	4.5	2.4	3.0	4.5
Richardson's constant ($A\,m^{-2}\,K^{-2}$)	6×10^9	6×10^9	n/a	n/a
Operating temperature (K)	2700	1700	1700	300
Current density ($A\,m^{-2}$)	5	10^2	10^5	10^6
Crossover size (nm)	$> 10^5$	10^4	15	3
Brightness ($A\,m^{-2}\,sr^{-1}$)	10^{10}	5×10^{11}	5×10^{12}	5×10^{13}
Energy spread (eV)	3	1.5	0.7	0.3
Emission current stability (%/h)	< 1	< 1	< 1	5
Vacuum (Pa)	10^{-2}	10^{-4}	10^{-6}	10^{-9}
Lifetime (h)	100	1000	> 5000	> 5000

Table 3.2: Comparison of different electron sources operating at 100 kV. Reproduced from (Williams and Carter, 2009).

under heated conditions and the field emission is therefore thermally assisted. The working temperature of Schottky FEGs can be as high as some of the thermionic guns. In contrast, the emission process in cold FEGs takes place under room temperature and is solely initiated by the strong electric field applied to the tip.

Cold FEGs are brighter and have smaller energy spread than Schottky FEGs. They are the better choice for experiments that require good energy resolution. However, cold FEGs are less stable compared to Schottky FEGs in that the tip of the W crystal in a cold FEG needs to have a pristine surface while the actual environment around the tip is never completely free of contamination even under ultra high vacuum conditions. As contamination accumulates, the brightness of a cold FEG will decrease, hence regular "flashing" is needed to clean the tip by heating it to a high temperature for a short amount of time. This will force the deposited contamination to evaporate. Such tip cleaning is not necessary for Schottky FEGs as they are constantly heated.

Table 3.2 lists some common parameters for different electron sources.

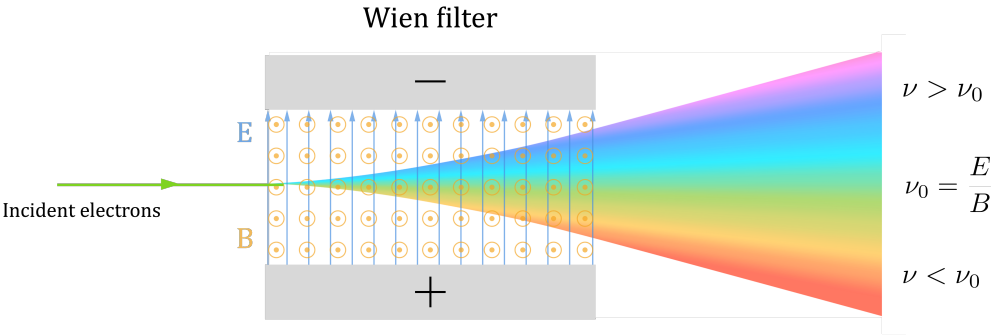


Figure 3.14: Trajectory of electrons in a single Wien filter

3.4.2 Monochromators

In practice, three types of monochromators have been designed and manufactured: the *single Wien-filter type* (Martínez and Tsuno, 2008; Tiemeijer et al., 2012b,a), the *double Wien-filter type* (Kirkland et al., 2013), and the *Omega-filter type* (Probst et al., 1993).

Single Wien-filter monochromator A Wien filter is a velocity selector for charged particles (Fig. 3.14). When travelling at a speed of ν through a combined electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$, a particle with an electrical charge, q, will experience a Coulomb force $\mathbf{F_E}$

$$\mathbf{F_E} = q\mathbf{E} \tag{3.68}$$

and a magnetic force $\mathbf{F_B}$.

$$\mathbf{F_B} = q\nu \times \mathbf{B} \tag{3.69}$$

If the spatial relations between $\mathbf{E}$, $\mathbf{B}$, and ν are as illustrated in Fig. 3.14, the Coulomb force and the magnetic force on the charged particles act in opposite directions. For a known geometry of the electric and magnetic fields, $\mathbf{F_E}$ is solely related

to the particle charge, e^- in the case of electrons, and $\mathbf{F_B}$ is proportional to both the electrical charge and the speed of the particle. This implies the existence of a critical velocity, ν_0, defined by;

$$F_E = F_B \tag{3.70}$$

$$\nu_0 = \frac{E}{B} \tag{3.71}$$

At ν_0 the two forces are balanced and the particle should travel through the Wien filter in a straight line. Particles of any other velocity will be dispersed by the combined field.

As shown in Fig. 3.15, the single Wien-filter monochromator produces an elongated crossover at its exit plane and uses a physical slit to select electrons with a desired energy E_e. The width of the slit determines the allowed energy spread ΔE_e.

Double Wien-filter monochromator The JEOL2200MCO microscope at Oxford used in this work is installed with a double Wien-filter monochromator (Kirkland et al., 2013). The advantage of having a second Wien-filter after the energy slit, in comparison to the single Wien-filter monochromator, is to cancel out any change in beam shape induced by the first filter.

Omega-filter monochromator In contrast to the first two designs, the electrostatic Omega filter chooses a different route for dispersing the achromatic electron beam (Essers et al., 2010). The Omega filter takes in the achromatic incoming beam image at its entrance plane, and forms an energy-filtered image at the exit plane of the filter, exactly the same as how normal energy-filtered TEM forms filtered images of the sample, except for that the energy of the electrons is much smaller. The output beam will have the identical beam shape as before the monochromator.

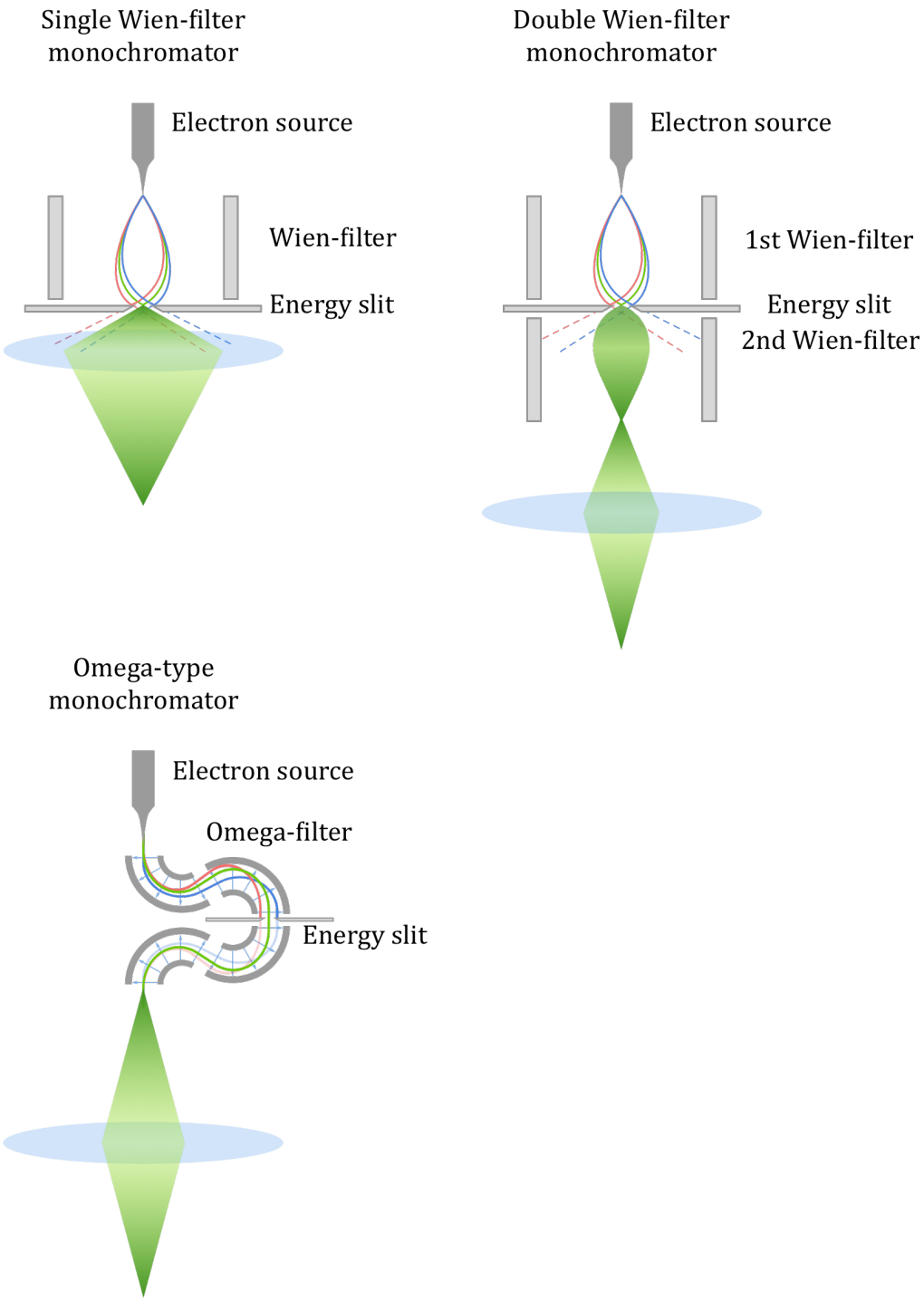


Figure 3.15: Three types of monochromators.

3.4.3 Multipole electromagnetic elements

Electromagnetic lenses are the essential components in electron microscopes. Among them, the objective lens, a round lens, is the pivotal part of the electron optical system. The theory of round electromagnetic lens is well studied in particle optics (Orloff, 2008). A brief description of round lenses and the origin of lens aberrations is summarised in Appendix A. The section intends to explain the working mechanism of multipole electromagnetic elements, which is the building block of many microscope components, particularly aberration correctors (Section 3.4.4).

In a more general sense, most parts listed on the right hand side of Fig. 3.12 are actually either an individual electromagnetic element or a certain type of lens combination. Dipoles (Fig. 3.16a) can be used for deflecting the electron beam, and when used in pairs, dipoles can shift the beam without introducing changes in beam tilt (Fig. 3.17). Quadrupoles (Fig. 3.16b) elongate the beam in one direction and compress it in the perpendicular direction. This makes them suitable for reducing two-fold astigmatism and other image distortions with two-fold symmetry. For a $2N$ multipole, with a polar coordinate system defined as shown in Fig. 3.17, the magnetic scalar potential (Bleloch and Ramasse, 2011) is;

$$\Phi(r,\theta) = r^N[p_N \cos(N\theta) + q_N \sin(N\theta)] \tag{3.72}$$

The magnetic field equals the gradient of $\Phi(r,\theta)$, and therefore is proportional to r^{N-1}, with N-fold symmetry. In fact, a $2N$ multipole in theory is capable of perfectly compensating the N-fold astigmatism if the aberration angle is exactly aligned with the multipole element. However, more often a pair of the same multipoles are needed to cope with the rotational misalignment.

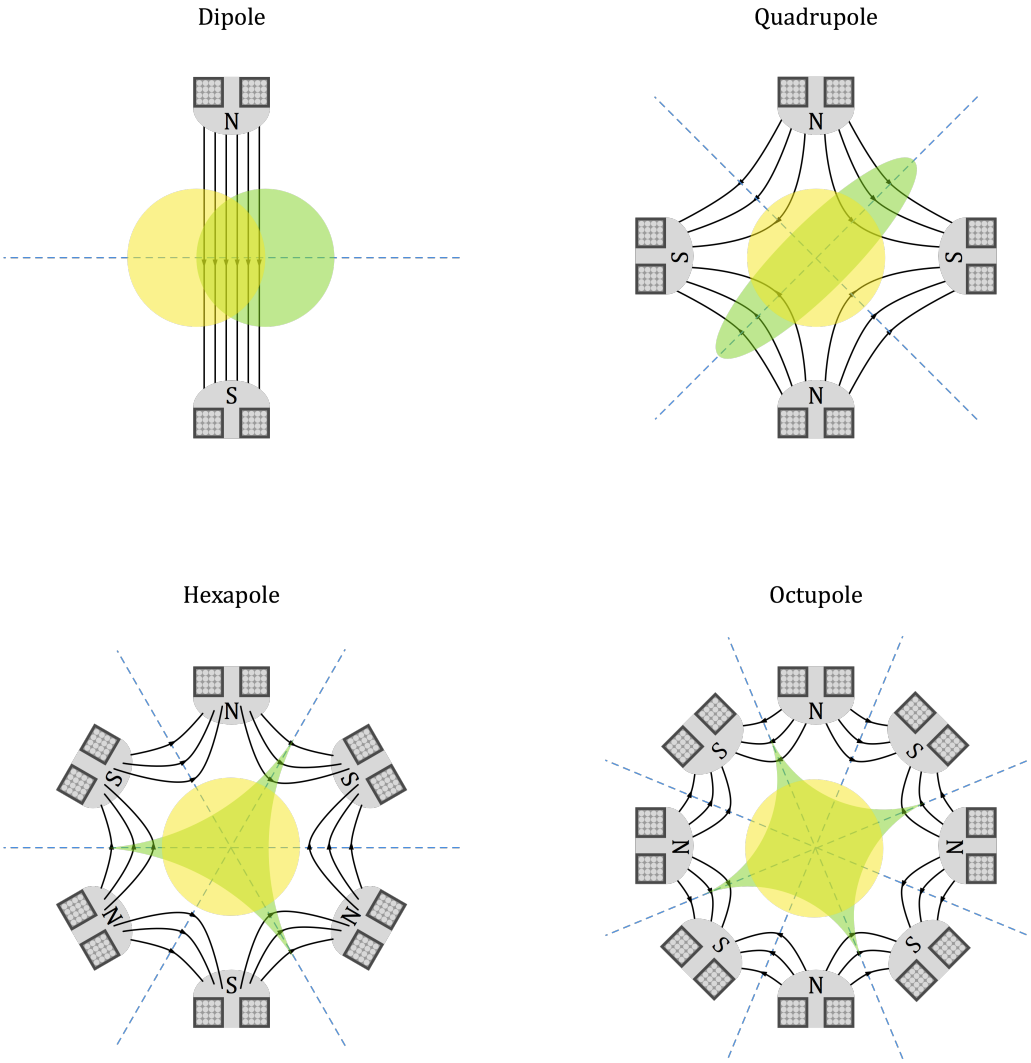


Figure 3.16: Schematic llustration of (a) dipole, (b) quadrupole, (c) hexapole, and (d) octupole element and their effects on the electron beam (the round yellow beam is the incoming electron beam and the green beam is the outgoing beam after the optical elements).

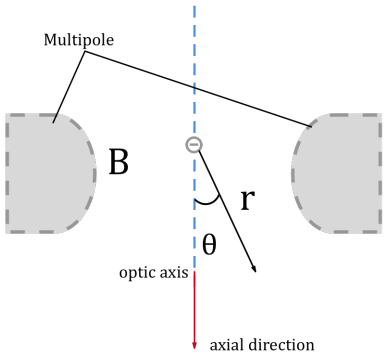


Figure 3.17: An electron in a multipole of a magnetic field, B, with polar coordinates (r, θ).

3.4.4 Aberration correctors

In the last two decades, the advancement of atomic resolution imaging in materials science is in large part thanks to the successful development of electron optical aberration correctors.

Otto Scherzer in his famous Scherzer theorem proved that the spherical aberration of a rotationally symmetric static electromagnetic objective lens[2] is unavoidable and always takes a positive value (Scherzer, 1936). It also has long been known that a simple negative (diverging) lens solution, as in light optics (Urban, 2008), is not available for electromagnetic lenses as discussed in Appendix A. Over the history of electron microscopy, feasible designs of spherical aberration correctors consisting of multipole lenses have been proposed, but the realisation of usable aberration correctors has only happened fairly recently (see Hawkes (2015) for a review).

Quadrupole-octupole corrector Spherical aberration is a round aberration with infinite azimuthal symmetry and therefore cannot be completely corrected by single multipoles. Just as the quadrupole stigmator for correcting the astigmatism, the quadrupole-octupole corrector designed for spherical aberration is based on the fact

[2]An *objective* lens forms a real image of a real object.

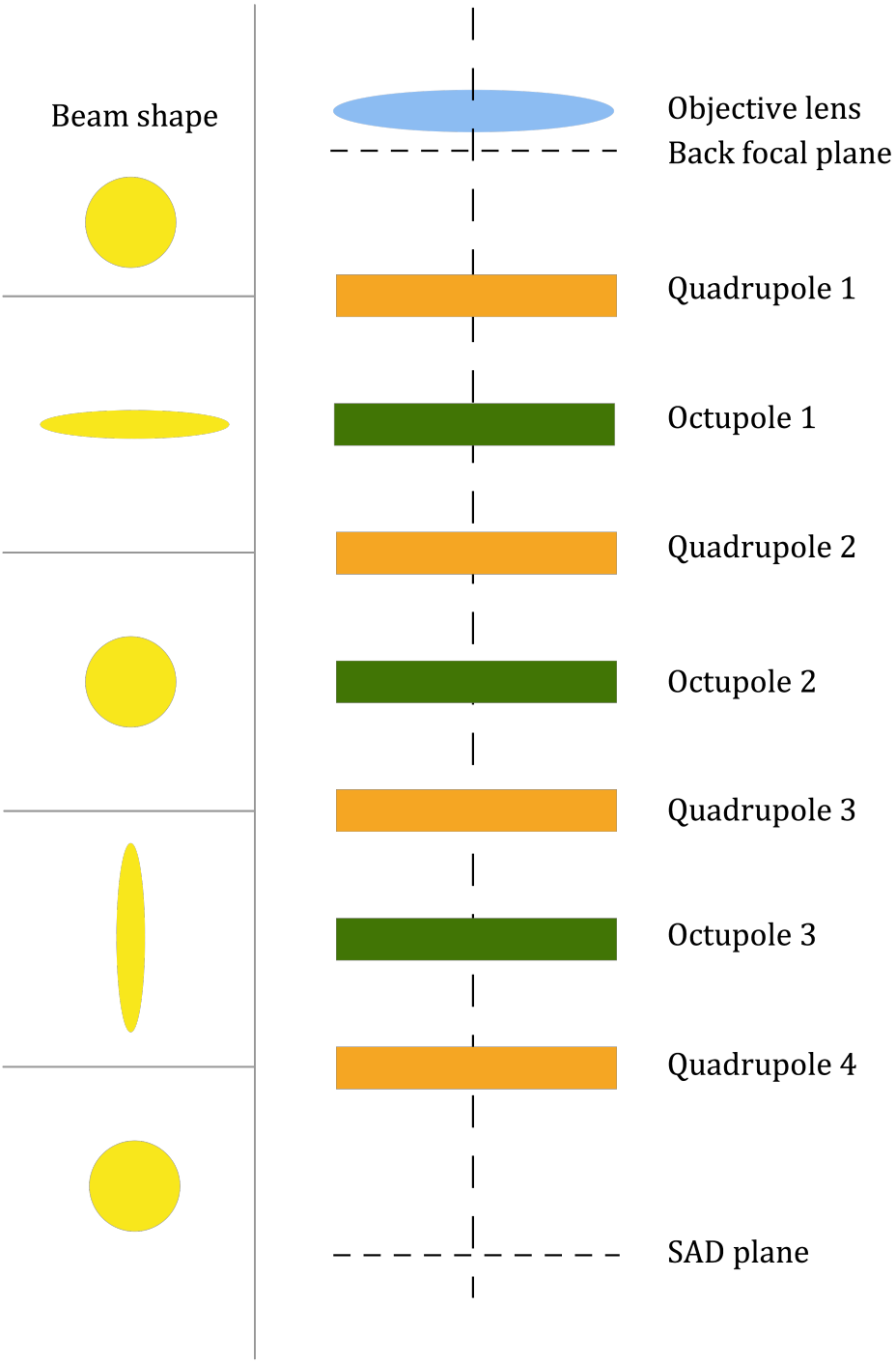


Figure 3.18: Optical elements in a quadrupole-octupole aberration corrector and the corresponding beam shapes (adapted from Hawkes (2015)).

that the magnetic field strength of an octupole ($N = 4$) varies with r^3 which is of the same order as the third order aberrations. The quadrupole ($N = 2$) pairs were added into the corrector to deliberately distort the beam into a long and thin elliptical shape so that when the beam passes through one of octupoles, the positive C_3 direction has little influence while the negative direction compensates C_3. Repeating the process in the perpendicular direction, C_3 is then fully compensated. The reason for having an additional third octupole in the corrector optical path is to cancel out the four-fold astigmatism introduced by the other two octupoles.

The quadrupole-octupole design is more commonly used in scanning transmission electron microscopes (STEM).

Hexapole spherical aberration corrector Almost all Cs-correctors for TEM are hexapole correctors. Originally proposed in 1965 by Hawkes (1965), whose idea of modifying the primary aberrations of the round lenses with the secondary aberrations of the hexapole elements was later further developed and confirmed by Beck (1979) and Rose (1981), the first working hexapole corrector was reported more than 30 years later (Haider et al., 1998), after several important design improvements such as the introduction of the transfer lenses and an "adaptor" lens (Rose, 2008).

The hexapole corrector has two hexapoles which are aligned by transfer lenses and rotationally offset by 60° relative to each other which cancel the primary three-fold astigmatism. The reason that hexapoles ($N = 3$) can be used for correcting third order aberrations is that, when the multipole lens is not thin but has a depth along the axial direction, i.e. the hexapoles are "long", the electron trajectory involves a movement in the direction away from the optical axis, where the magnetic field is stronger than it would be had the electron not travelled towards the radial direction. This increases the rate of the lens effect by one extra order and therefore the hexapoles are able to correct third order aberrations. This second-order effect is homogeneous in

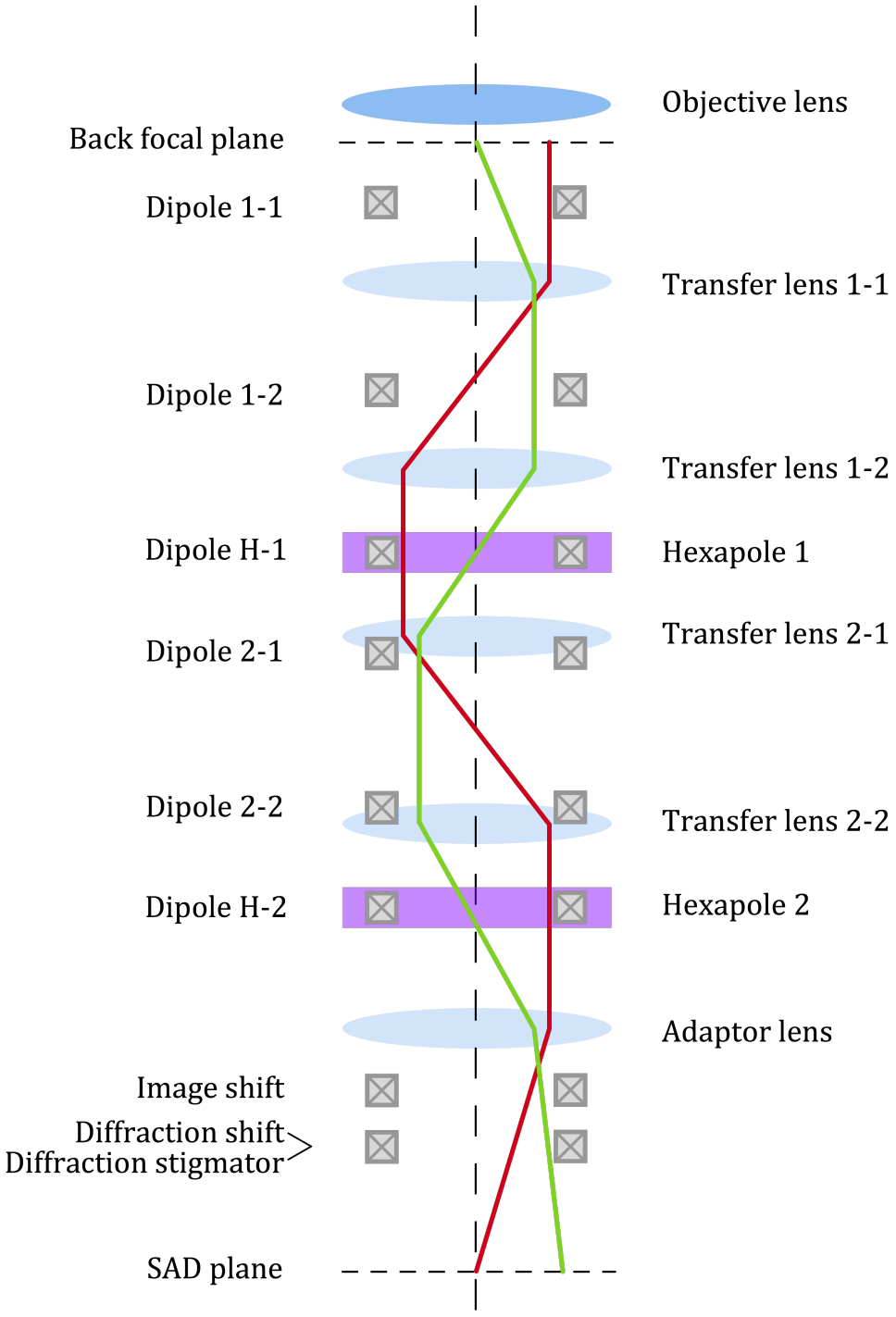


Figure 3.19: Optical elements in a hexapole aberration corrector (adapted from Hawkes (2015)).

all azimuthal directions and matches the infinite azimuthal symmetry of the spherical aberration. The complete design also includes two pairs of round transfer lenses to ensure that the two hexapoles share equivalent optical planes, a prerequisite condition for the hexapole corrector to function.

The future of aberration correctors aim at correcting higher order aberrations up to C_5 as well as C_c correction (Rose, 2005; Rose and Wan, 2005).

3.4.5 Sample stages

Sample stages are the only part of a microscope that have direct contact with the specimen. The functions of sample stages have evolved based on more and more specialised experimental needs. The traditional single-tilt holder is still widely used for nanoparticles and samples that do not require anything but a stable support for the specimen grid. For crystalline materials, a double-tilt holder can be used to adjust the sample orientation to the required zone axis before taking HRTEM images. Cold TEM stages are widely used in Cryo-EM, especially in experiments involving imaging of biological structures and other radiation-sensitive materials because keeping specimens at low temperatures can reduce the speed of radiation damage and hence prolong the observation time (Section 2.4.3). The operation of cold holders needs additional steps in the preparation stage prior to the TEM observation, including pumping the dewar to high vacuum and filling the liquid nitrogen in the holder chamber. A common problem with using a cold holder is the vibration caused by the boiling of liquid nitrogen (LN). Filtering the liquid nitrogen can reduce the residual water in the LN and lower the nucleation initiated by ice impurities.

At high resolution it is still often noticed that the sample drift is more severe when using a cooling holder compared to the more traditional holders. This could be caused by the sharp temperature gradient near the cooling area.

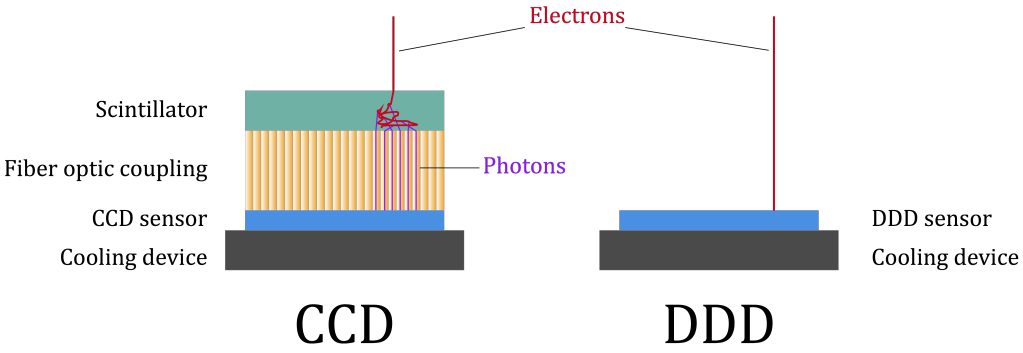


Figure 3.20: Illustration of a CCD camera and a DDD camera.

3.4.6 Detectors

Although the detector on a TEM has no direct control over electron dose or radiation damage, a well calibrated high-performance detector is still vital for low-dose imaging and data processing. The signal-to-noise ratio is often very poor in low-dose images making the detection of sample features difficult. The high noise level limits not only the interpretation of individual images, but also the application of exit wave reconstruction as all restoration techniques require well aligned image series, which is hard when the signal is not strong enough for the registration algorithms (Saxton, 1994b).

Charge coupled devices (CCDs) An illustration of a simple CCD camera is shown in Fig. 3.20. The main feature that distinguishes a CCD detector in the TEM from a CCD chip in a normal digital camera is the extra scintillation layer. The phosphor scintillator, when hit by the incoming electrons generates photons. Between the scintillation layer and the CCD element, a fibre optics or an optical lens is used to direct the photons onto the CCD sensor, where the phonons create electron-hole pairs that are trapped in the potential well of separated pixels. Hence, for TEM CCD cameras, the detection process is not direct, but instead through an electron-photon-electron conversion process. The incoming primary electron can ac-

tivate photon generation while it travels inside the scintillation layer and the photons coming out of the enlarged interaction area cause a point spread effect which may limit the attainable resolution (Meyer et al., 2000). Had the electron incoming position been detected directly instead of through the secondarily generated photons, the efficiency and accuracy of the image acquisition could be improved. This leads to the growing interest in direct electron detection technology.

When using a CCD camera for quantitative electron microscopy, it is important to keep the electron exposure of an individual image within the limit of the dynamic range, which is the full well capacity of the camera divided by the readout noise. A large dynamic range allows the image intensity to be linearly proportional to the electron dose over a large dose range.

In the "post-film" age of electron microscopy, the charge coupled device (CCD) camera has become the most common image recording device for TEM. Its advantages over traditional film are clear; linearity of response and real-time access to the recorded images are particularly important for quantitative HRTEM research. There are however some drawbacks, for example the limited dynamic range is often not suitable for recording intense signals such as long-exposure diffraction patterns, and the number of pixels on a single CCD chip imposes restrictions on the highest spatial frequency at low magnifications, as subsequently discussed in Section 4.1.2.

Direct Detection Devices (DDDs) Direct detection devices (DDDs) have started to appear in the recent years, but have already gained attention from various microscopy fields. Unlike CCDs, DDDs do not require a scintillation layer nor the optic fibre/lens coupling (Fig. 3.20)). The signal is generated by letting the primary electron directly exposing a CMOS chip where each primary electron will generate a large number of excited electrons to form the signal in the pixel wells. It has been reported (Milazzo et al., 2010) that the DDD systems are superior to CCDs in that the

MTF of a DDD demonstrates a higher transfer at half Nyquist frequency. DDDs also have fast readout speed compared to CCDs, allowing quick multiple image acquisition, which can be used to correct fixed-pattern reset noise arisen from the accumulated charges.

Although the advantages of using DDDs over CCDs have not become significant enough to overthrow the dominant market share of the latter, the superior acquisition speed and smaller pixel sizes of DDDs will likely enable novel imaging techniques, i.e. high-speed HRTEM at low magnifications in the future.

Chapter 4

Low-Dose HRTEM Imaging

4.1 Instrument calibration and parameter measurement

4.1.1 Electron dose calibration

Most electron microscopes provide an estimated value of the beam current via the fluorescent screen, which gives a qualitative indication of the dose level for routine TEM imaging. More quantitative measurement of electron dose and its relation to radiation damage depends upon a more accurate electron dose calibration.

With a typical Schottky FEG, the incident electron beam current at the sample plane is of the order of $10^{-9} \sim 10^{-10}$A. Measurement of such a small beam current inside the narrow column space requires a dedicated Faraday cup. The Faraday cup should be installed along the electron path above the CCD camera. During normal TEM imaging with a sample presented in the field of view, some of the electrons in the beam are absorbed, and the beam current before and after the sample is therefore different. To eliminate this effect, the calibration should be carried out either with an empty sample holder with no sample grid or with a very large hole on the specimen

grid where the beam can pass through without any obstruction.

A step-by-step illustration of the calibration procedure is shown in Fig. 4.1, and the detailed execution steps are described as follows:

Step 1: Insert the Faraday cup when the fluorescent screen is well illuminated. Identify the approximate centre of the Faraday cup which provides an estimate of where to place the focused beam in a subsequent step.

Step 2: Extract the Faraday cup. Focus the beam until it reaches its brightest and smallest. Move the beam to the imaginary centre of the Faraday cup.[1]

Step 3: Insert the Faraday cup again, and monitor the beam current value measured by the picoammeter. Finely adjust the beam shift until the measured current reaches its maximum, at which, when perfectly centred the value should stay unchanged against small adjustments of beam shift in any direction. Record the current value i.

Step 4: Extract the Faraday cup once more, leaving the beam at the centre of the screen.

Step 5: Insert the CCD camera and spread the beam to the extend that the beam fills most of the area of image with the edges of the beam at the boundaries of the CCD field of view.

Step 6: Acquire images of the beam with a preset exposure time, t, and calculate the total pixel counts, s_c, in the image of the beam.

$$g = \frac{s_c}{it/e} \tag{4.1}$$

[1]Magnification of this step is not particularly important, while it is still helpful to choose a relatively low magnification when performing the calibration, as this makes it easier to focus the beam to a spot that is small enough to fit into the Faraday cup entrance.

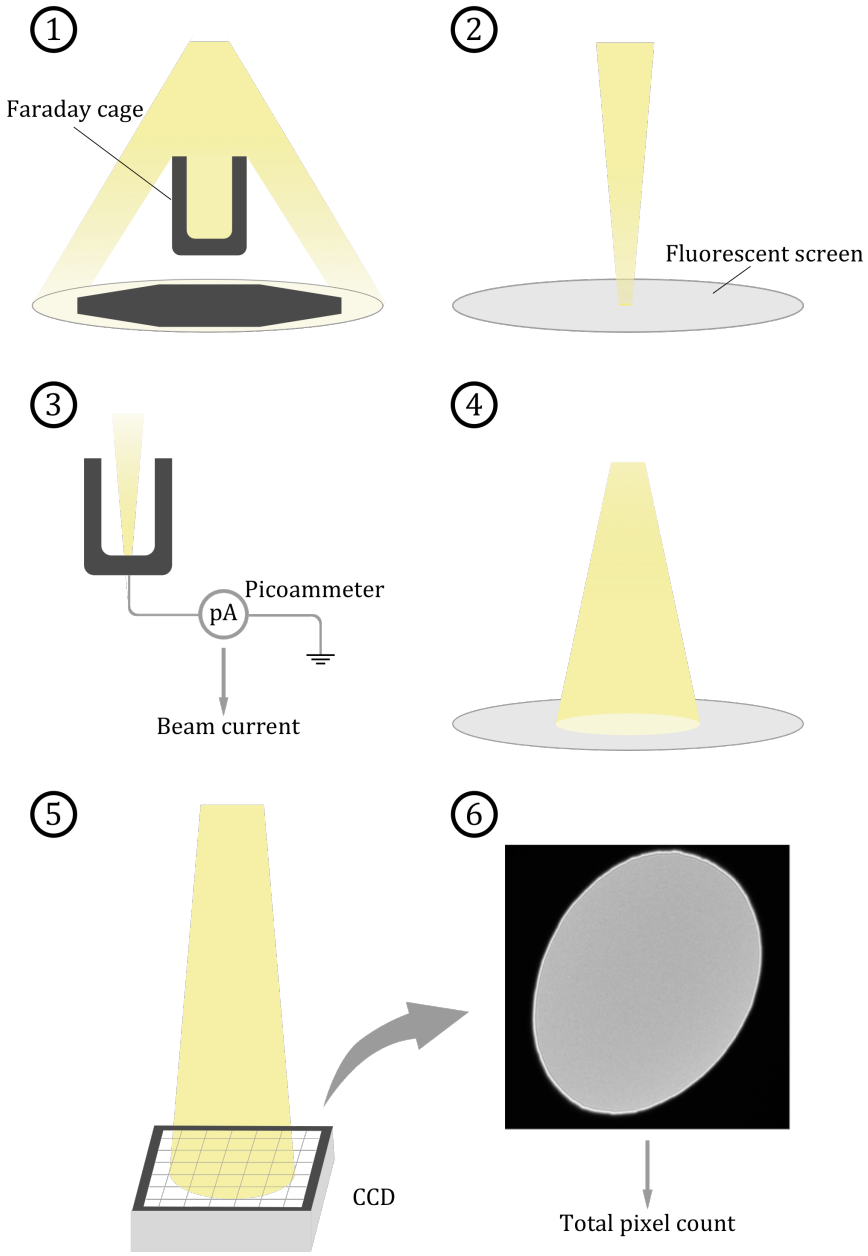


Figure 4.1: Schematic illustration of the dose calibration procedures. (1) Faraday cup alignment; (2) Beam centring; (3) Beam current measurement; (4 - 5) Beam image acquisition; (6) Total pixel count of a beam image.

Accelerating voltage	200 kV	80 kV
Counts/primary electron	6.83	12.07

Table 4.1: Dose calibration results, units: counts/electron

Within the full well capacity of the CCD camera, there is a linear response between the number of electrons detected by the camera and the digital counts recorded (De Ruijter and Weiss, 1992). The ratio between these two values is the gain of the CCD camera. From the definition of gain in Eq. 4.1, it is straightforward to plot the CCD counts as a function of the product of the beam current and the exposure time divided by the electron charge, e, as;

$$s_c(i,t) = g(\frac{it}{e}) \tag{4.2}$$

From Eq. 4.2, it can be seen that the gain of the CCD camera is the slope of a straight line and can be obtained by linear fitting if enough data points of CCD counts s_c against number of electrons are given. The data point collection can be done by repeating the above calibration procedure with different condenser lens apertures which modify the beam current i, and with different exposure times t.

Two sets of calibration results of dose calibration data collected on the JEOL 2200MCO at Oxford at 200 kV and 80 kV, are plotted in Fig. 4.2. Table 4.1 shows that one primary electron on average generates 6.63 counts at 200 kV, and 12.07 counts at 80 kV. The fact that lower-energy electrons on average generate higher counts than higher-energy electrons is attributed to the fact that low-energy electrons are expected to have a shorter mean free path (Ziegler et al., 2014) in the scintillator material and a greater chance to be deflected away from their incoming direction and travels a relatively longer distance inside of the scintillator. Hence, more photons are generated by a single electron at 80 kV than at 200 kV.

One final step prior to fitting the electron-counts curve is to eliminate the abnormal

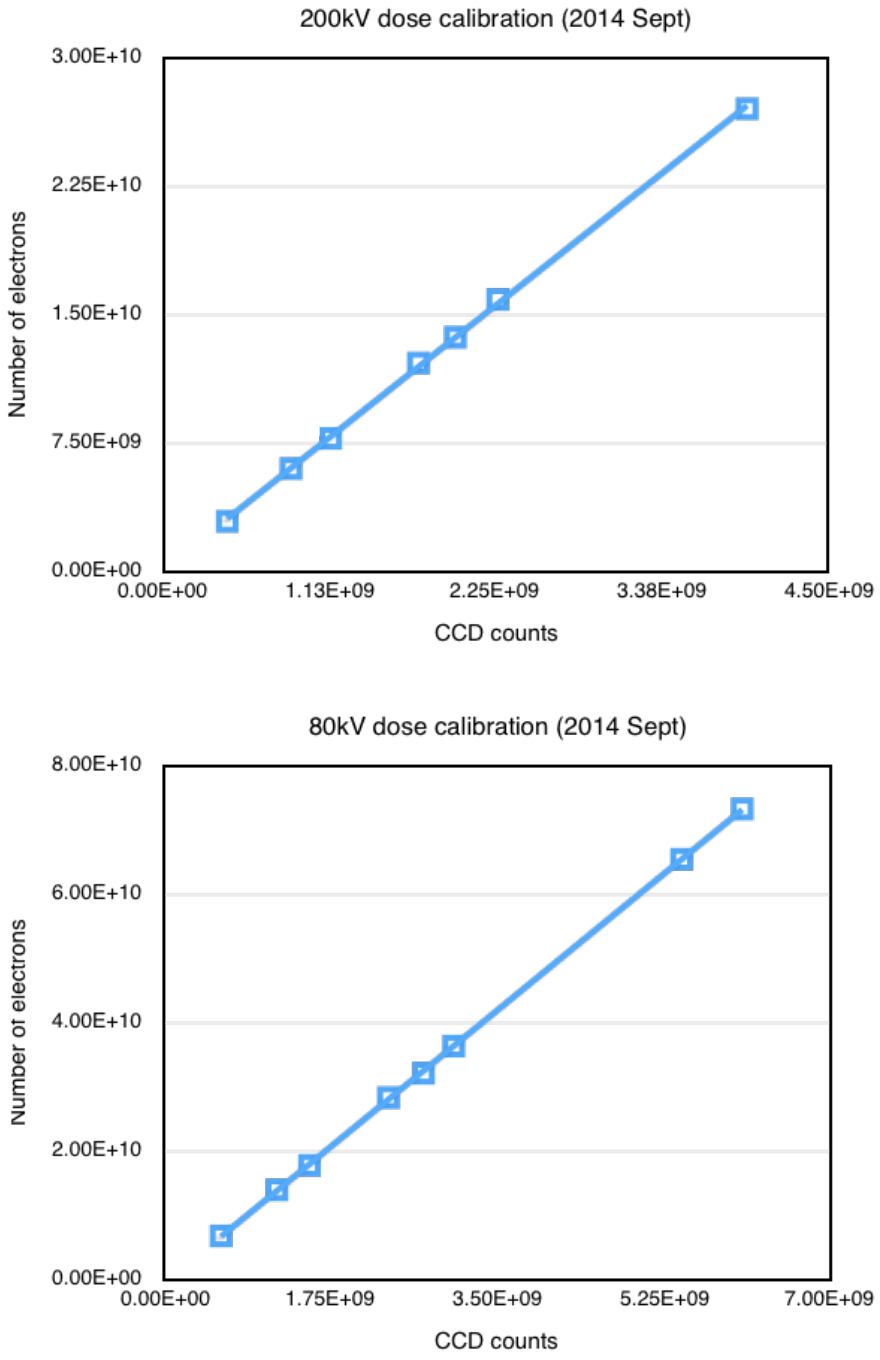


Figure 4.2: Dose calibration for the JEOL2200MCO at 200 kV and 80 kV.

data points that arise when the CCD is over-saturated. This can be caused by a prolonged exposure time or by an intense electron beam when the condenser lens aperture is large. When the outliers are excluded, the error in the linear fitting to a straight line function should be small. In the case presented in Fig. 4.2, the coefficient of determination, r^2, is 0.9995 and 0.9999 for 200 kV and 80 kV respectively.

An issue that may cause an increased calibration error is *abnormal pixels*. Abnormal pixels will give either extremely large or very small pixel values, which should not appear for near uniform illumination. Some simple procedures that can be used to qualitatively check the status of the calibration include:

- Take an empty image with the beam off; check the mean value and the histogram of the image. When the detector is carefully tuned and gain-normalised[2], the mean of an empty image should be zero and the full width at half maximum (FWHM) of the distribution in the histogram should be narrow, typically no larger than 5 to 10 counts.

- Turn the beam on at a moderate beam intensity, approximately half of the CCD dynamic range and record an empty image of an area with no samples or any obstructions. The image should be uniformly flat and free of detector features.

- Starting from a high dose, gradually reduce the dose and take multiple images at different stages. All images should look similarly feature-free, except for those at very low dose which may have a higher noise level.

4.1.2 Magnification and image sampling

Digital EM images are two dimensional discrete signals. The choice of sampling interval of the signal will directly affect the signal representation. According to the

[2]The correct procedures of detector tuning and gain normalisation should be found in the corresponding manual of the instrument.

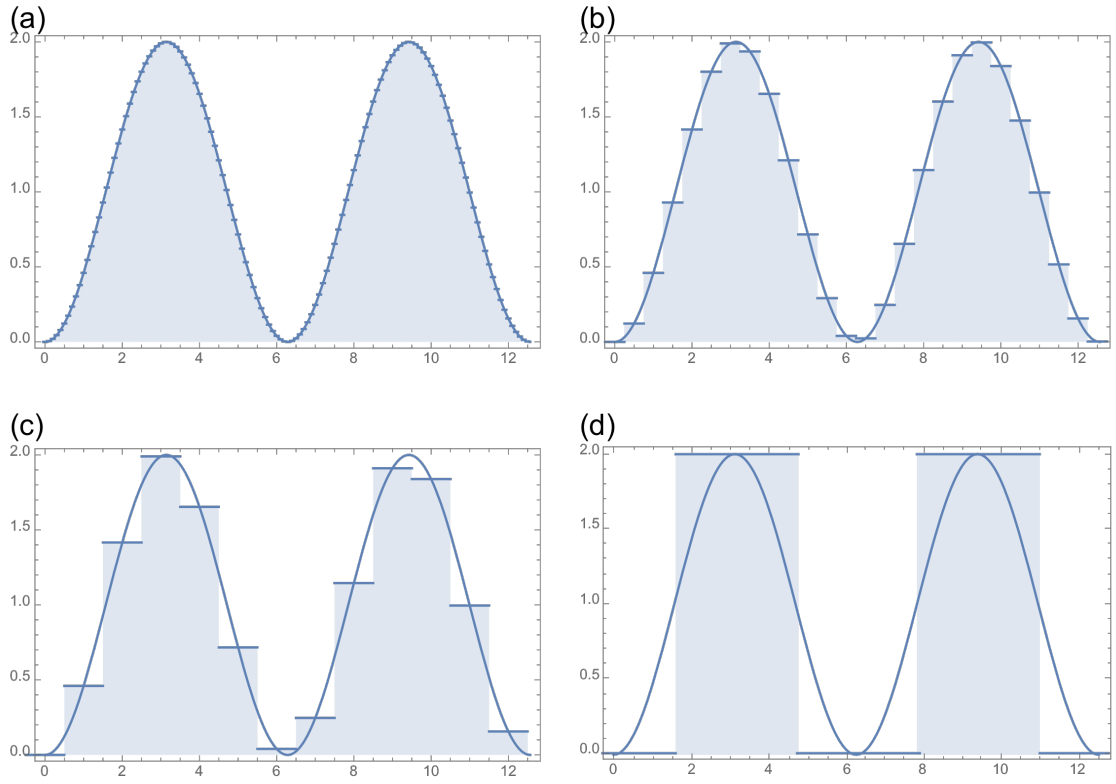


Figure 4.3: A discrete representation of the positive sine function $\sin x - \pi/2 + 1$ with different sampling intervals (a) Sampling interval: 0.1; (b) Sampling interval: 0.2; (c) Sampling interval: 1; (d) Sampling interval: π.

Nyquist-Shannon sampling theorem, the attainable resolution is restricted to twice the size of the smallest sampling interval (Nyquist, 1928; Shannon, 1949). Fig. 4.3 shows the same periodic sine signal represented discretely by four different sampling intervals, and the value of each "stair", or "pixel" in the images, is the signal value at the mid-point. Fig. 4.3a uses the smallest sampling interval among the four and gives the finest description of the original signal, while Fig. 4.3b and Fig. 4.3c are coarser than their predecessors, but still allow the two sine peaks to be well pronounced and separated. The sampling interval in Fig. 4.3(d) is half the period of the sine signal, and it is the coarsest sampling that can separate the peaks at all places, with a dark pixel (the trough) in the middle of two bright pixels (the crest). Any smaller sampling interval than that will cause two consecutive bright pixels, in which case the two peak points are no longer resolved.

For a given CCD camera, the higher the magnification, the smaller the image sampling interval in the image. While undersampling the data can result in loss of high spatial frequency information, oversampling the data may cause problems from a limited field of view.

The general rule for choosing the image sampling interval is to include all the interpretable spatial frequencies in the image while keeping the magnification at a value where there are enough electrons within the field of view to generate sufficient signal. Low contrast images are often severely corrupted by noise and more prone to errors from the calibration of the CCD. Uncorrected profile features and defects in a CCD are commonly seen in low-dose images. Placing the highest spatial frequency at between half to two thirds of the Nyquist frequency has been considered to be optimum for imaging with the CCD camera used. A smaller sampling interval is sometimes adopted for the purpose of avoiding the severe attenuation from the modulation transfer function at high frequency range, as previously described in Section 3.2.3. If a direct detector device (DDD) is available, the improved MTF allows a larger usable

frequency range, and therefore imaging can take place at a lower magnification.

4.1.3 Sampling interval calibration

Precise magnification calibration makes use of a known material structure that contains periodic features. Calibration methods are available in both real space and reciprocal space. For the real space approach, a known periodic feature is chosen, such as n_{uc} unit cells with a period length of L_c along a certain crystallographic direction and the number of corresponding pixels n_{pixel} in the experimental images is measured, knowing the actual distance from a crystal model. The sampling interval is then given by;

$$d_{si} = \frac{n_{uc} L_c}{n_{pixel}} \tag{4.3}$$

Using an image of zeolite ZSM-5 as an example (Fig. 4.4). The number of pixels across 10 unit cells in the [1 0 0] direction is 411. The unit cell parameter in the same direction $a = 2.0064\,\text{nm}$. The sampling interval is then calculated as;

$$d_{si} = \frac{10 \times 2.0064\,\text{nm}}{411} = 0.048\,82\,\text{nm} \tag{4.4}$$

The reciprocal space approach measures a specific Fourier reflection and relates it to the corresponding real space crystallographic spacing. The required relation is given in Eq. 4.5,

$$d_{si} = \frac{n_{fpixel} L_c}{N} \tag{4.5}$$

where n_{fpixel} is the number of pixels in Fourier space, L_c is the corresponding unit cell parameter in real space, N is the number of pixels in one of the dimensions of the CCD (2048 for a 2K×2K CCD).

$$d_{si} = \frac{(996 \div 20) \times 2.0064\,\text{nm}}{2048} = 0.048\,79\,\text{nm} \tag{4.6}$$

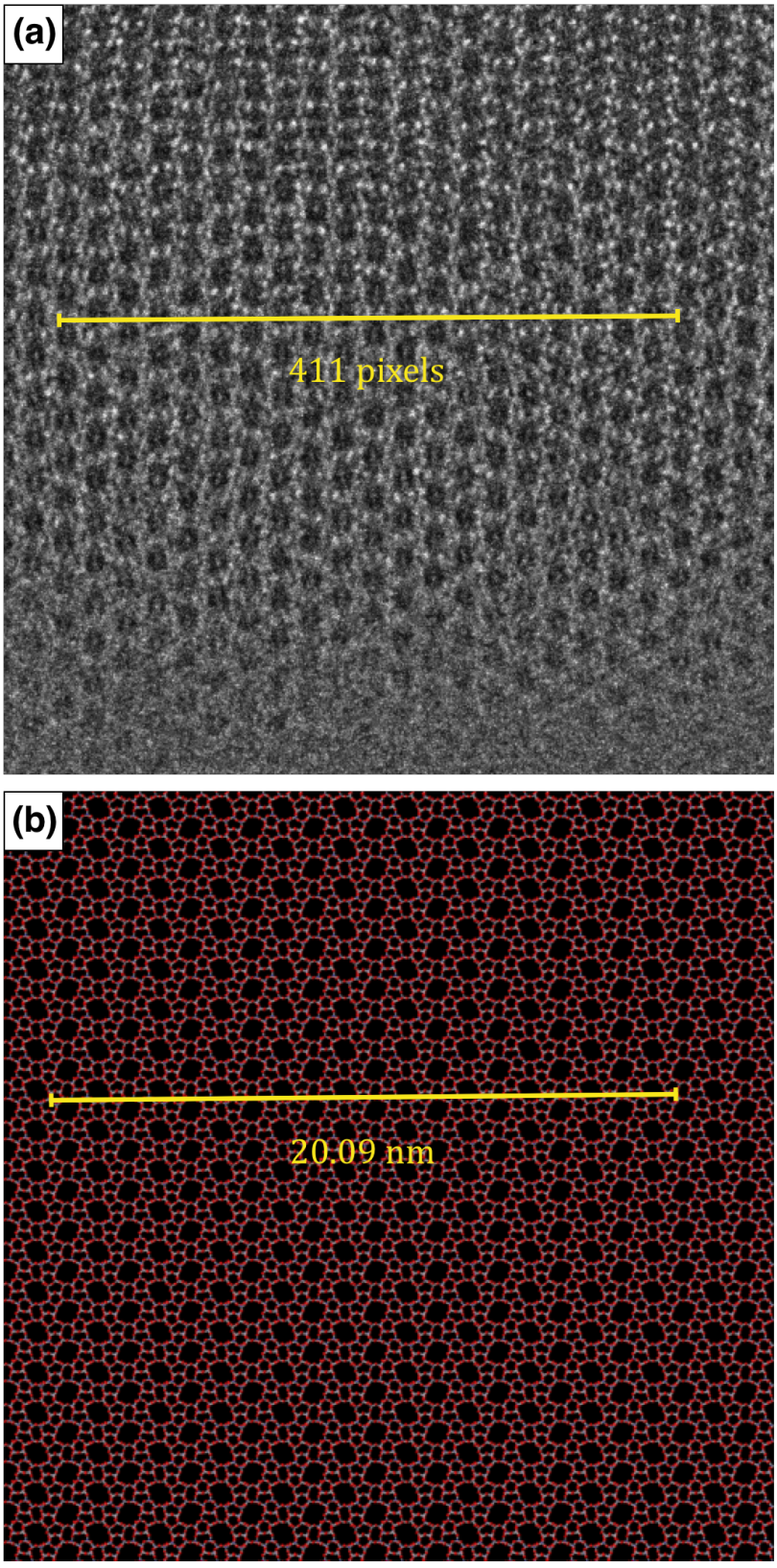


Figure 4.4: (a) Image of ZSM-5 with the yellow line indicating 10 repeating unit cells. (b) Corresponding atomic model.

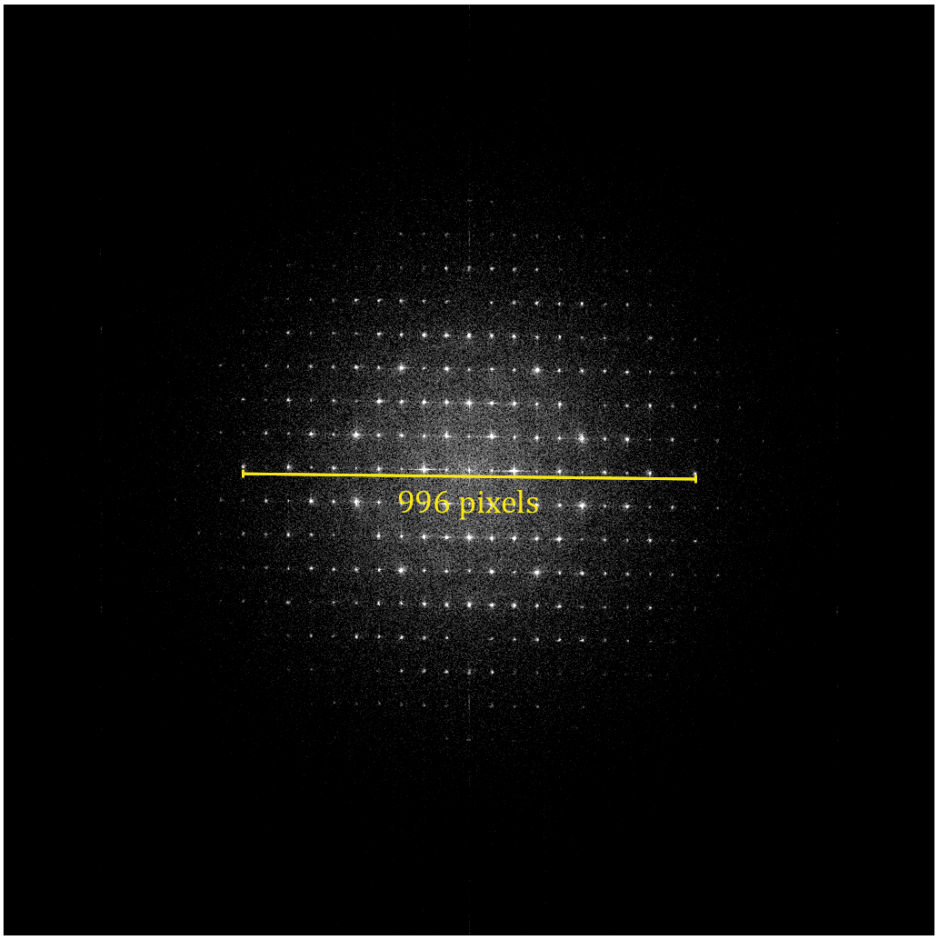


Figure 4.5: Corresponding power spectrum from the original image of Fig. 4.4a with the yellow line indicating the distance between two equivalent reflections in pixels.

The difference between the results given by the two methods is less than 0.6%.

The area that a pixel represents in real space is subsequently calculated as;

$$a_{si} = d_{si}^2 = 0.002\,387\,\mathrm{nm}^2. \tag{4.7}$$

4.1.4 Post-experiment dose determination

When both the gain of the CCD camera and the sampling internal are known, only the mean value of the image is required to calculate the dose used for the imaging exposure. For example, the average counts in the ZSM-5 image in Fig. 4.4 is 98. Using the 2014 September dose calibration results in Table 4.1 and Eq. 4.7, the imaging dose of the experimental image is;

$$D = \frac{\frac{1}{N\times N}\sum_{i,j=N} p_{i,j}}{g \times a_{si}} = \frac{98}{6.83 \times 0.0023865} = 6012\,e/\mathrm{nm}^2 \tag{4.8}$$

In addition to using the calibration results for post-experiment data processing, it is helpful in experiments to have the exchange relationships between mean pixel value and the electron dose (in e/nm^2 or $e/\mathrm{\AA}^2$) at each magnification for quick reference.

4.2 TEM sample preparation

Most sample materials studied in this thesis come in either powder or small crystal forms. TEM specimen preparation for such materials is typically carried out via a crush-and-disperse approach. The prepared TEM sample grids should contain enough usable sample for efficient imaging experiments. To examine the materials in TEM, crushed samples containing thin areas need to be dispersed onto TEM specimen grids via suspension of the material fragments.

The unprocessed sample is very often too big in size or too thick for high resolution TEM experiments. Reducing the size requires finely crushing the powder sample

with an agate mortar and pestle. Areas that are thin enough for atomic resolution imaging tend to only exist on the edges of the sample fragments. In order to embrittle the sample, which increases the occurrence of sharp regions on sample edges, liquid nitrogen (LN) is used to chill the mortar and pestle before adding the sample. When the mortar is thoroughly chilled to liquid nitrogen temperature, $-195.79\,^{\circ}\mathrm{C}$ (77 K), the mortar should be able to hold a small amount of liquid nitrogen, without it quickly evaporating or boiling.

After preparatory cooling, the powder sample is placed into the mortar and filled with more liquid nitrogen before immediately grinding with a LN-cooled pestle. The grinding generates heat and therefore further LN refills to keep the sample at a low temperature until the powder crushing is finished. The resultant fine powder sample is then transferred into a vial containing ethanol before being placed into an ultrasonic bath to go through the final dispersion process. Dependent on the suspension concentration, five to ten drops of suspension containing the ground sample are dripped onto the holey carbon film TEM grid. The optimum suspension concentration and the number of suspension drops on one grid require trial and error. The prepared TEM specimens need to be thoroughly dried for the residual ethanol to evaporate before being stored in the sample box or being put into the TEM. Prior to a TEM session, in order to eliminate residual organic contamination which may have deposited onto the sample grid during sample preparation and storage, further cleaning of the specimen is required. This can be done by baking the sample grid inside a plasma cleaner overnight before the experiments. The baking temperature is dependent on the specific sample, as well as on restrictions of the sample holder used for baking.

4.3 Low-dose imaging experiments

4.3.1 Dose control

There is more than one way of controlling the dose and the dose rate during experiments. This section summarises the three approaches used over the course of this project.

Condenser lenses The primary mechanism for controlling the beam current density by using the condenser lenses after the gun deflectors. By focusing or spreading the beam illuminating the sample, the condenser lenses can change the beam intensity almost continuously.

There are some complications about using condenser lenses to reduce radiation damage. When the illumination is spread, the beam extends beyond the boundary of the field of view of the CCD (red dashed circle in Fig. 4.6). Since the total beam current, i, is constant, for a given time, the lower the dose rate d/t, the bigger the illuminated area on the sample. Assuming that the beam is round with a radius, R, the following relation between R and d (Eq. 4.9) shows how the region of exposure changes with dose.

$$\frac{d}{t} = \frac{i}{4\pi R^2} \tag{4.9}$$

This makes it difficult to determine where the sample is free from electron exposure before imaging and where it is not, and this may introduce errors in the measurement of the total dose on a given area throughout the experiment. Using a smaller condenser lens aperture can reduce the incident beam current and reduce the illumination area. Another solution to partly avoid this problem is to only use sample regions that are distant from each other on the TEM grid, so that the sample is less likely to be pre-exposed. If the illuminated area of previous imaging does reach the new area of

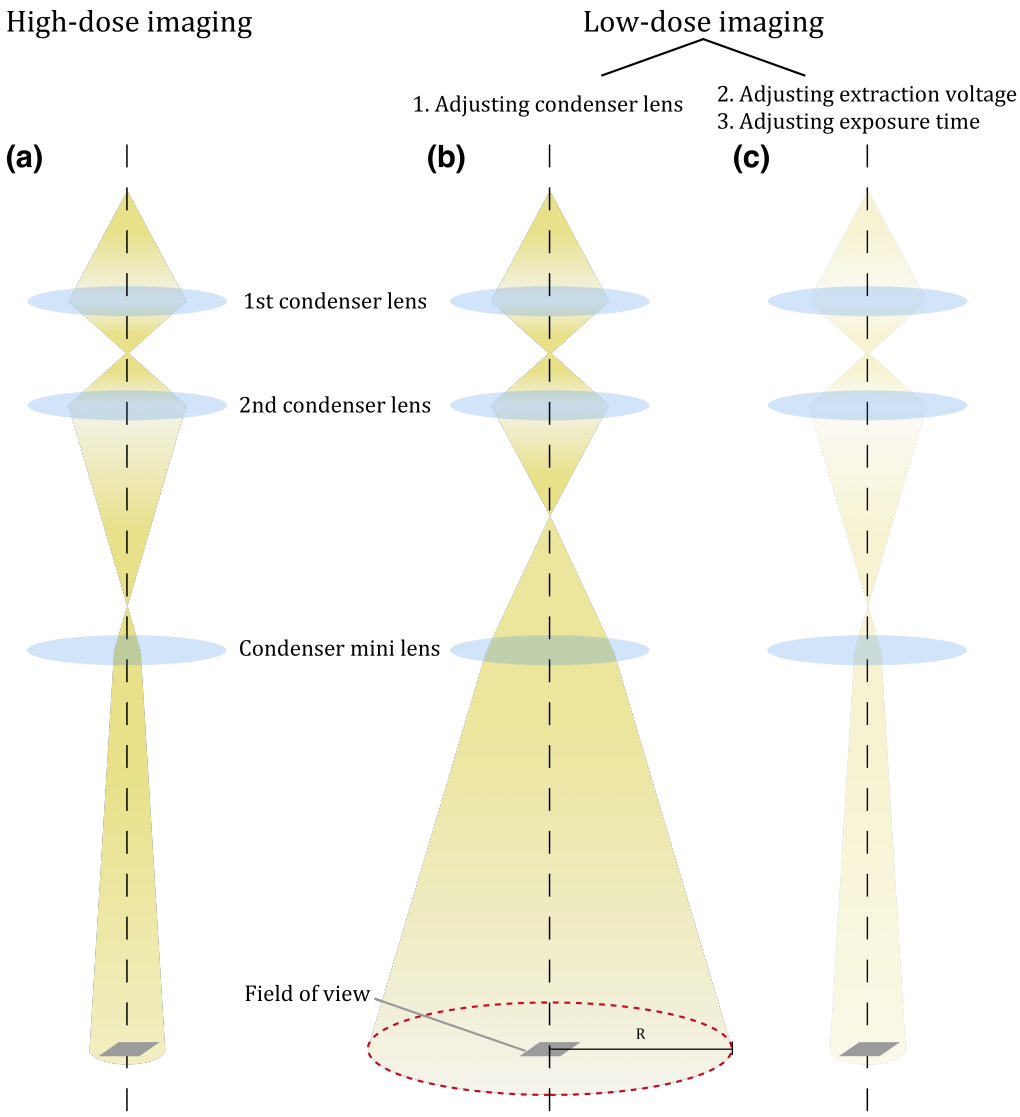


Figure 4.6: A comparison of the microscope configuration for (a) high-dose imaging. (b) Low-dose configuration through defocusing the beam with the condenser lens. (c) Low-dose configuration through reducing the total beam exposure of an image with a lower extraction voltage or a shorter exposure time.

interest, the magnitude of the dose rate will be negligibly small negligible as R will be large in such situations.

Extraction voltage Eq. 3.67 in Section 3.4.1 implies that, for a Schottky FEG, another way of dose control is to change the extraction voltage V_{ex}, which is linearly related to the electron beam current. With the illumination area unchanged, dose control by changing the extraction voltage can be used for experiments aiming to compare imaging at different doses. By altering the beam current, i in Eq. 4.9, all imaging conditions remain the same before and after the extraction voltage change.

Exposure time As discussed in Section 2.2, changing the dose rate can significantly affect the total dose for the same exposure time. Similarly, changing the exposure time can also be used for dose control at a constant dose rate. This method for dose control is best suited for stable samples that are not sensitive to charging effects so that the TEM images are free of blurring which is usually caused by sample movement.

4.3.2 Tilting the sample

For crystalline specimens, aligning the zone axis of the crystal with the beam direction is a crucial step towards obtaining atomic resolution. For radiation-sensitive samples, the diffraction mode needs less dose to generate good diffraction data and causes less radiation damage over the same period of time compared to the imaging mode. In spite of this, one should still choose sample pieces that are close to the zone axis to limit the amount of required tilting and hence to further limit the beam exposure. Increasing the exposure time when recording diffraction patterns will help to give a high quality diffraction pattern even at low dose rates, and since the dose rate is low in this mode, there is more time available for adjusting the sample orientation.

The radiation-sensitive nature of many materials leads to a large deformation caused by electron illumination, often enough to alter the zone axis locally. Therefore

it is important to record the data promptly after the tilting of zone axis is finished.

4.4 Low-dose imaging of hydroxyapatite

In this section, hydroxyapatite is used as an example material to demonstrate a typical low-dose HRTEM imaging experiment.

Mineralised hydroxyapatite (HAP) is the building block of tissues in human teeth and bones. It occupies 96% of the volume in the enamel within human teeth, with the remaining 4% taken up by organic components. Understanding the microscopic structure of such materials is of great importance in studies of the mechanical properties of teeth and bones. Recent research has suggested that localised structural variation inside the crystallites is related to the growth mechanisms of these biocrystals (Yanagisawa, 2003). Although the hydroxyapatite crystals are much more resistant to radiation damage compared to other biological tissues, one still needs to conduct high resolution TEM imaging experiments under low-dose conditions as not only the crystalline structures can be damaged by an intense electron beam, but in addition the gluing cementum is prone to beam damage. Its fast degradation during image acquisition can lead to misalignment between the electron beam and the crystal orientation, compromising the obtainable spatial resolution.

Hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$) is a hexagonal crystal with space group symmetry P63/m and lattice parameters, $a = b = 0.943\,\mathrm{nm}$, $c = 0.688\,\mathrm{nm}$ and $\gamma = 120°$. There are two mirror planes located at $z = \frac{1}{4}$ and $z = \frac{3}{4}$. An array of PO_4 tetrahedra are linked by Ca^{2+} ions, and the OH^- ions occupy columns on the screw axes (Reyes-Gasga et al., 2012).

The imaging of hydroxyapatite under different doses used the condenser lens dose control mechanism previously described. Each series of images were taken with an exposure time of 2 s. For two adjacent images within the same time series, there is a

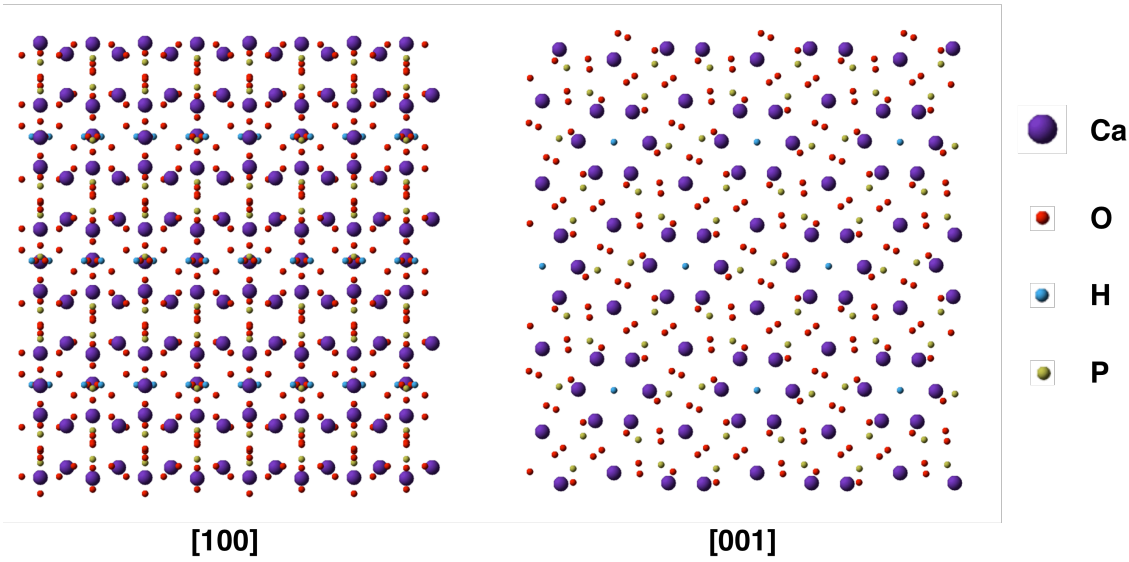


Figure 4.7: Atomic model of hydroxyapatite.

1 s interval in between acquisition to allow time for the beam to stabilise.

4.4.1 Time series acquisition

Image series recorded with long exposure times often suffer from image blurring due to sample drift, therefore one should generally avoid taking images with prolonged acquisition time. The maximum acceptable length of exposure depends on how stable the sample and the microscope is on the day and on the operating magnification. One way to boost the signal-to-noise ratio in low-dose images is to take multiple short-exposure images (a time series) of the same area under the same imaging condition and use image registration to compensate the sample drift between images. The aligned images are averaged into one image with an enhanced signal-to-noise ratio. For Poisson noise, this increase in SNR is $\sqrt{N}$ for a time series containing N images. However, the key part of this procedure is a successful image registration scheme that can compensate for sample movement over the acquisition of the low-dose time series.

To explore the limits of using few electrons to acquire TEM image series at atomic resolution, nine different time series images at various doses were recorded, with each

Series name	Dose ($e/Å^2$)	Average count/pixel
HAP_ts1	253	156
HAP_ts2	82	53
HAP_ts3	44	27
HAP_ts4	23	14
HAP_ts5	11	6.8
HAP_ts6	5.6	3.4
HAP_ts7	3.2	2.0
HAP_ts8	2.2	1.4
HAP_ts9	1.0	0.62

Table 4.2: Dose and average counts of different hydroxylapatite time series.

.

subsequent series using less imaging electrons (Table 4.2). With each series using fewer electrons than the previous series, the images in the lower dose series are much more noisy than the ones in the starting series (HAP_ts1).

4.5 Image Registration

It was noticed over the course of this project that image registration sets a major restriction on various post-processing techniques for high resolution image series. There are a number of different registration strategies explored.

4.5.1 The cross-correlation function (XCF)

The cross-correlation function (XCF) is the most commonly used registration method in image processing. For an image $g(\mathbf{x})$ and a reference image $f(\mathbf{x})$, with both being 2D discrete signals, the cross-correlation function is defined as;

$$XCF(f(\mathbf{x}), g(\mathbf{x}), \mathbf{r}) = \sum_{\mathbf{x}} f(\mathbf{x}) g(\mathbf{x} - \mathbf{r}) \tag{4.10}$$

where image g is shifted with respect to image f by a vector $\mathbf{r}$. A normalised form

of the cross-correlation function can also be defined;

$$NXCF(f(\mathbf{x}), g(\mathbf{x}), \mathbf{r}) = \frac{\sum_{\mathbf{x}}(f(\mathbf{x}) - \overline{f})(g(\mathbf{x}-\mathbf{r}) - \overline{g})}{\{\sum_{\mathbf{x}}[f(\mathbf{x}) - \overline{f}]^2 \sum_{\mathbf{x}}[g(\mathbf{x}-\mathbf{r}) - \overline{g}]^2\}^{0.5}} \quad (4.11)$$

$$= \frac{1}{n}\sum_{\mathbf{x}} \frac{(f(\mathbf{x}) - \overline{f})(g(\mathbf{x}-\mathbf{r}) - \overline{g})}{\sigma_f \sigma_g} \quad (4.12)$$

which is widely used in image processing. It is also sometimes referred as the cross-correlation coefficient, where n is the number of pixels in the images, $\overline{f}$,$\overline{g}$ and σ_f, σ_g denote the average and the standard deviation of the two images respectively.

The XCF can be calculated between any images, regardless of their shapes. But for most real applications, images are often are usually square and their dimensions are the power of 2. Using the convolution theorem, this gives a convenient approach to quickly calculating the cross-correlation functions by taking advantage of the discrete fast Fourier transform (FFT). Using $\mathcal{F}[\cdot]$ and $\mathcal{F}^{-1}[\cdot]$ to denote the Fourier transform and the inverse Fourier transform, the XCF can be written as;

$$XCF(f, g) = \mathcal{F}^{-1}[\mathcal{F}^*[f(\mathbf{x})]\mathcal{F}[g(\mathbf{x})]] \quad (4.13)$$

where $\mathcal{F}^*[f(\mathbf{x})]$ denotes the complex conjugate of $\mathcal{F}[f(\mathbf{x})]$.

The auto-correlation function (ACF) is a special case of the XCF when the two input images are identical to each other, and is defined as;

$$ACF = \mathcal{F}^{-1}[\mathcal{F}^*(f(\mathbf{k}))\mathcal{F}(f(\mathbf{k}))] \quad (4.14)$$

4.5.2 Reduction of aliasing

Although the XCF of two images can be calculated using the definition given by Eq. 4.10, it is more commonly calculated by a more efficient element-wise multiplication in Fourier space. The downside of this computational efficiency are *aliasing*

artefacts arising from the correlation calculation.The aliasing artefact arises due to the periodicity assumption in the discrete Fourier transform. Although a real image has finite limits in both dimensions, the Fourier transform considers the image as an infinitely repeating periodic object. For this reason, this artefact is also often known as the "wrap-around" artefact. Simple solutions using image filtering are described here to reduce the influence of aliasing.

By apply an anti-aliasing filter to the real image before performing the Fourier transform, the edges of the image are tapered to zero. The most common filter for this purpose is a two-dimensional Hanning window (Harris, 1978). The one-dimensional form of a Hanning window $w(x)$ has the following form;

$$w(x) = 0.5(1 + cos(\frac{2\pi x}{n})), x \in [-\frac{n}{2}, \frac{n}{2}] \tag{4.15}$$

where n is the width of the image. A two-dimensional Hanning window can be created by rotating the one-dimensional Hanning window around its centre (Fig. 4.8). The new window function is isotropic and simply creates a damping effect on the image. As can be seen in Fig. 4.9, the filtered image will lose some image intensity, especially in regions near the image boundaries and hence it is advisable to have the region of interest close to the image centre. If due to experimental reasons the intended region is not placed in the middle but the edge of the image, it is necessary to either create a smaller selection that has the ROI in the centre, or to use the mean-frame method for aliasing artefact correction.

It should be noted that there is no aliasing from the linear correlation operation. Hence if calculating the cross-correlation function using the classic linear definition (Eq. 4.10), anti-aliasing processing is unnecessary. However, due to the performance boost of the fast Fourier transform in the correlation function calculation, it is rare that one would choose the linear correlation over the circular correlation (Eq. 4.13).

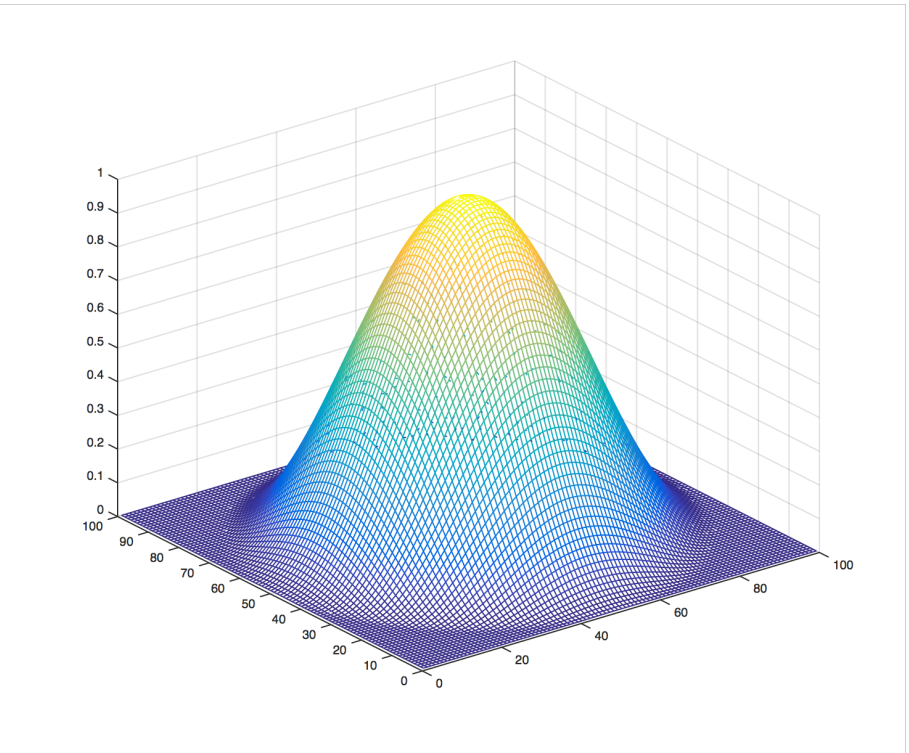


Figure 4.8: A two-dimensional Hanning window.

More recent development in signal processing have made efforts towards further minimising the influence of aliasing by using zero-aliasing correlation while retaining the calculation in the Fourier domain. However, to date, this is not yet a mature technique and is still computationally challenging for many applications (Fernandez et al., 2014).

4.5.3 Image registration of hydroxyapatite time series

Artefacts in the XCF of low-dose HRTEM images The simple XCF is a fairly effective and robust approach for high dose TEM image registration. However, registering the testing data set of a low-dose HAP time series, it appears to fail after the imaging dose drops below a certain threshold. The reason for this failure could be simply as a low signal-to-noise ratio, but a more thorough inspection shows that it is artefacts within the XCF that cause the registration to fail.

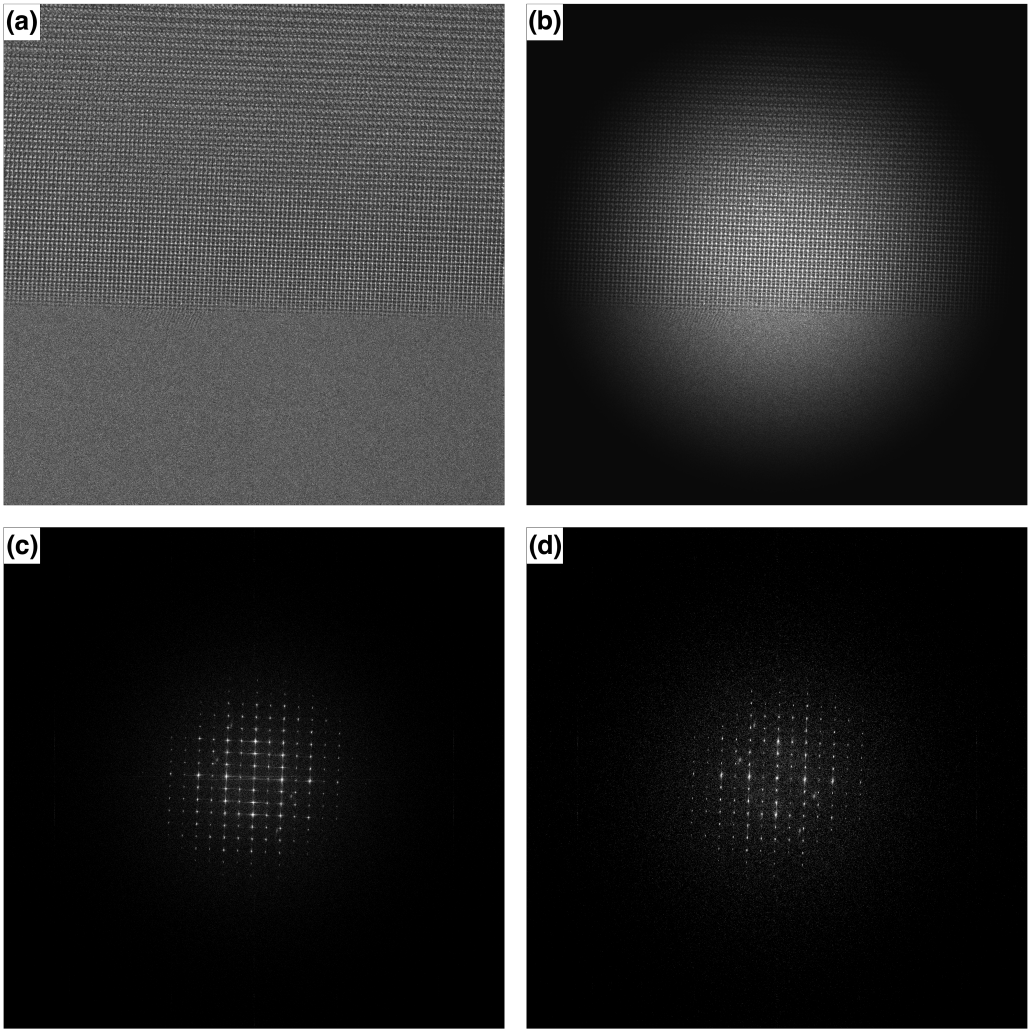


Figure 4.9: Anti-aliasing with a Hanning window. (a) Raw image (b) Image after applying the Hanning window. (c) Power spectrum of (a). (d) Power spectrum of (b).

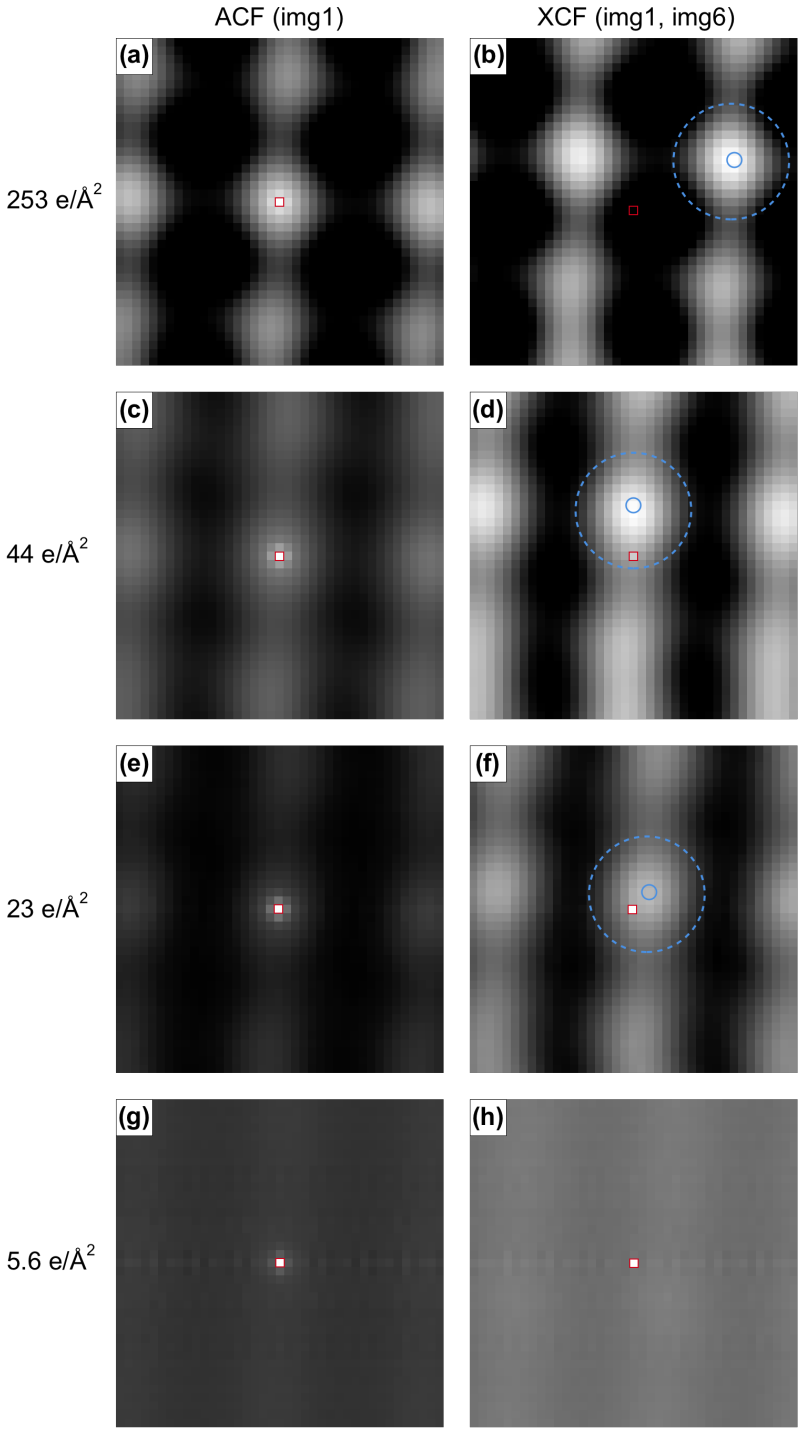


Figure 4.10: (a,c,e,g) Auto-correlation functions of the first image in the HAP time series at four different doses ($253\,e/\text{Å}^2$, $44\,e/\text{Å}^2$, $23\,e/\text{Å}^2$, and $5.6\,e/\text{Å}^2$); (b,d,f,h) Cross-correlation functions of the first image and and the sixth image. The red square marks the centre pixel of the ACF/XCF; the small blue circle marks the brightest pixel after eliminate the centre pixel of the XCF (the large dashed blue circle is manually placed to indicate the XCF peak when the two images are close to their best match position).

Registration methods that are based on the XCF calculate a two-dimensional XCF map of the reference image and the subject image. Searching for the highest intensity pixel gives the position relative to the central pixel, the shift between the initial image position and the position where the two images are most correlated. Correlation functions are sensitive to periodic signals and image shifts that are integer multiples of the signal period will generate correlation peaks of similar strength to the perfect alignment position. For atomic resolution TEM images of periodic structures, a distinctive feature of the XCF maps is that the periodicity is also displayed, making the map image look similar to the input images (Fig. 4.10). Having multiple correlation peaks is often considered to have a negative impact on image registration, and might cause the algorithm to confuse periodic side peaks with the correct peak when noise is present. Non-periodic features, such as sample edges and thickness-variations in contrast, can strengthen the correct correlation peak and weaken its periodic repetitions.

The abnormal central pixel in XCF In a times series, when calculating the auto-correlation function (XCF with the image itself), the brightest pixel will always appear at the centre of the XCF map. This global maximum pixel corresponds to zero relative displacement, at which the image perfectly overlaps with itself, in terms of both signal and noise components, as shown in the four images in the left part of Fig. 4.10. For images in a time series taken under exactly the same imaging conditions, the only differences between the reference image and the rest of the series are random noise and sample drift which the registration aims to remove. When the images are aligned, the XCF map of any two images from the same series should have an appearance much like the ACF of each individual image, with the maximum intensity appearing at the centre of a broadened XCF peak (top right in Fig. 4.10). This is true for series recorded at higher imaging dose, for example the two images

from a hydroxyapatite time series taken at $253\,e/\text{Å}^2$ (top line in Fig. 4.10). However, as the beam intensity decreases, to doses that are equal or lower than $23\,e/\text{Å}^2$, an abnormally high intensity pixel, whose location is the geometric centre of the XCF map, grows increasingly sharp in the XCF map and finally exceeds all other pixels and becomes the strongest spot. When the imaging dose is below a certain value, in this case $23\,e/\text{Å}^2$, the traditional XCF registration fails and always returns zero image shifts.

The origin of this unwanted bright pixel is most likely to be an underlying fixed detector profile that has not been fully compensated by the camera calibration. Unlike random noise, the residual detector profile is constant across different images. If we ignore the random noise component and consider that each image f and g consists of image signals, f_s and g_s, and the residual detector profiles, f_d and g_d, then;

$$f = f_s + f_d,\ g = g_{signal} + g_d \tag{4.16}$$

Given that the correlation operation is distributive, though not commutative, the cross correlation of the two images above can be further expanded as;

$$f \otimes g = f_s \otimes g_s + f_s \otimes g_d + f_d \otimes g_s + f_d \otimes g_d \tag{4.17}$$

with $\otimes$ denoting the cross-correlation operation.

The signal shows no correlation with the detector profile in either of images, f or g, and the correlation terms of the image signal in one image with the detector profile in another image should reduce to zero, i.e.;

$$f_s \otimes g_d = f_d \otimes g_s = 0 \tag{4.18}$$

This equation can be simplified as;

$$f \otimes g = f_s \otimes g_s + f_d \otimes g_d \tag{4.19}$$

For typical HRTEM images, due to the large incoming dose, the detector component is negligible and the correlation term between the signals of the two images is significantly larger than the correlation of the detector profile, therefore;

$$f_s \otimes g_s \gg n_d \tag{4.20}$$

$$f \otimes g \approx f_s \otimes g_s \tag{4.21}$$

and the XCF peak should indicate the image shift vector as predicted.

In low electron dose situations, where the correlation between signals diminishes as the signal weakens, the detector correlation term n_d will become dominant in the registration result. The detector profile is essentially static across all images and the XCF of detector profiles in different images is the ACF of each profile component ($ACF(f_d)$ and $ACF(g_d)$). The registration result is then compromised by this zero-shifting background, described by;

$$f_d = g_d = const n_d = f_d \otimes g_d = const \tag{4.22}$$

and as;

$$f_s \otimes g_s \to 0 \tag{4.23}$$

implies that;

$$f \otimes g \approx n_d \tag{4.24}$$

It is worth noticing that this registration failure is not due to poor detector calibration but the finite elimination of gain-normalisation errors that the calibration is able to achieve. The corresponding CCD counts for a dose of $23\,e/\text{Å}^2$ is 14 per pixel, which is extremely low.

Given that further improvement in the accuracy of camera calibration is hard, a simple workaround to this problem is to eliminate defective pixels. This treatment has proven to be able to extend the effectiveness of XCF registration to the lower dose series, as shown in the later part of this section (Fig. 4.11).

Periodicity in the detector profile If the exposure intensity is further reduced, a more extreme version of the above artefact will start to emerge. For a CCD camera with rectangular pixels, the residual detector profile is not a completely random background. Instead, it may have its own periodicity in certain directions that corresponds to fringes or bright spots in its Fourier transform and its ACF. In the particular case of a residual camera profile, with the centre pixel already been removed, the correlation peak search may locate one of the strong peaks related to the periodicity in the camera profile. If the correlation peak of the image signal is too small to outweigh these periodic peaks, the result of XCF registration will point to a shift based on the detector profile. The dose threshold for this artefact is lower than the ACF central peak of the detector profile discussed previously, which does not require any periodic feature in the detector background.

The second stage of misalignment caused by a residual detector profile is significant at 6.8 counts per pixel on average, equivalent to $11\,e/\text{Å}^2$ at the sample. Importantly, it is nearly impossible to improve the registration by individually eliminating the misleading local maxima, as there are too many and their location is random.

Until now, we have not considered any human intervention on the registration procedure. The methods and additional improvements all work under a situation

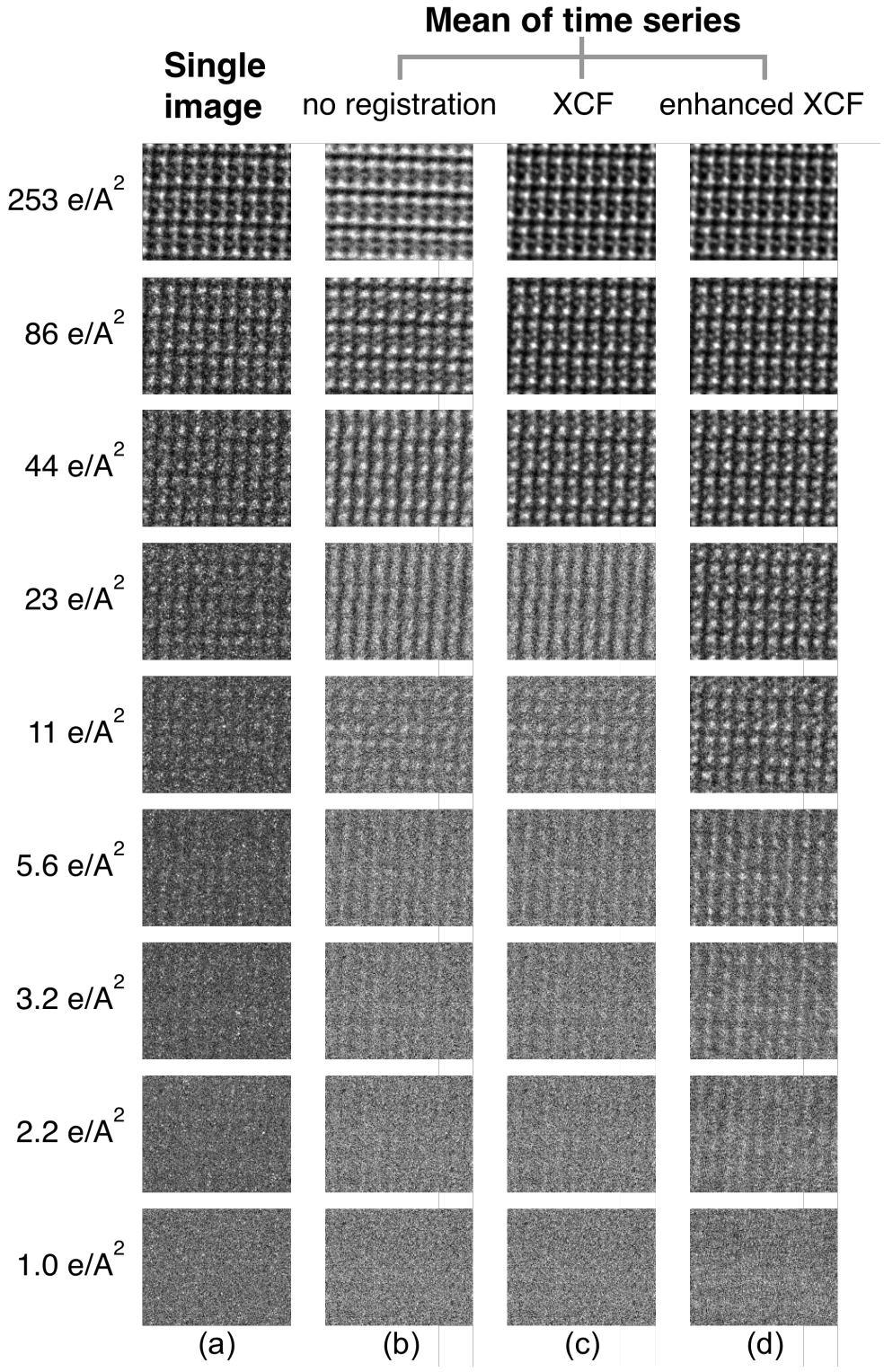


Figure 4.11: Comparison between (a) a single low-dose image of each dose series and the mean images of a time series (b) without registration, (c) with simple cross-correlation registration and (d) with the enhanced registration procedure, where the abnormal central pixel is excluded.

where no prior knowledge about the sample or the experimental settings is needed for image registration. However, reasonable use of some prior knowledge can be helpful for registration. When examining the registration results quantitatively, there are some obvious errors which can be corrected by introducing certain restrictions based on an understanding of the sample and the image drift during time series imaging. It has been observed during experiments that the sample drift tends to have a general trend direction over the course of a time series. The main source of sample drift, the instability of the sample holder should result in a directional drift. In addition, charging effects may result in an additional random sample movement. If the sample experiences a sudden and substantial movement during acquisition, the image taken at that moment should show a severe blurring compared to the rest of the series. If the microscope is stable, it is acceptable to assume a maximum value of possible drift between the acquisitions of neighbouring images and only search for the correlation peak within a limited area when registering the two images. A circular search area is ideal for the model but not convenient for programming, and hence in the enhanced registration code a square search area is adopted. The size of the search area was estimated from high dose series be less than five pixels between two neighbouring images and was set to be eight pixels in any direction, i.e. a 16×16 pixel area for peak searching.

Fig. 4.11 demonstrates the improvements that can be achieved by the enhanced XCF registration compared to the conventional XCF as well as to a simple averaging of the time series images without any registration. The comparison done in this section was done by direct visual observation and is therefore only qualitative. Quantitative comparision will be presented subsequently in this chapter after a comparison metric is defined. The conventional XCF works well for the highest three doses ($253\,e/\text{Å}^2$, $86\,e/\text{Å}^2$, $44\,e/\text{Å}^2$), as there is an evident improvement in image contrast and the Calcium atom columns are clearly resolved. As the imaging dose is

reduced, the conventional XCF method starts to fail, primarily due to the abnormal pixel in the geometric centre of the correlation map. The method returns zero image shift for all series that were taken with lower doses than $44\,e/\text{Å}^2$, meaning the output series is the same as the raw data without any registration. This implies that under this dose threshold, there will be no practical effect of using the conventional XCF method as opposed to no registration before averaging the time series into one mean image. The mean image of the $23\,e/\text{Å}^2$ series without registration, although having a better contrast in the horizontal direction, shows a noticeable blurring issue in the vertical direction, with supposedly separated atom columns merging into a line in the averaged image. The enhanced XCF registration is able to successfully compensate this misalignment and returns a mean image in which the atoms are resolved. This robustness when working with ultra low-dose image series data is still in effect as the imaging dose is reduced to $2.2\,e/\text{Å}^2$. For the time series taken at $1.0\,e/\text{Å}^2$, it is difficult to tell just from the image whether the registration is still functioning properly, as the mean image is significantly degraded by noise and only a small variance in image contrast can be observed.

Importantly, between two fringes, the spatial separation is only $3.46\,\text{Å}^2$, which can be clearly resolved with only 10 images at an ultra low dose of $2.2\,e/\text{Å}^2$ each image. This total dose of $22\,e/\text{Å}^2$ has dropped into the realm of biological HRTEM imaging, but the $3.46\,\text{Å}^2$ is previously not achievable in biological imaging without the help of single particle analysis, which generally requires hundreds of images.

4.6 Metrics for success of times series registration

4.6.1 Noise-normalised reflection peak profile

The improvement in image quality can be observed qualitatively both in the mean images (Fig. 4.11) and in their power spectra, where the reflections become clearer

and more distinct. To quantitatively evaluate and demonstrate this improvement needs a consistently measurable quantity with physical meaning. Image contrast in real space images is intuitively easy to understand but is not a good quantitative measure as the image contrast is not the same everywhere.

In the power spectra, sharp peaks are directly related to the sample's periodicity while the noise present in the background follows a noise power spectrum depending on the detector (Meyer and Kirkland, 1998). Moreover, a two-dimensional Gaussian fitting can provide a robust approach to estimating both the peak height and the background noise.

The [0 9 6] reflection peak was chosen as the subject of the analysis. This peak is visible in the power spectra of most image series taken, and therefore comparisons can be made between the high dose series and the ultra low ones.

In comparison with the real space analysis that has been discussed above (Fig. 4.11), the reflection peak comparison has also included the peak analysis result from a single image and the mean image of each series together with two additional series mean images after registrations using the traditional XCF method and the enhanced version. For high dose series, with an imaging dose of $253\,e/\text{Å}^2$ per image (Fig. 4.14), the traditional XCF registration performs equally well compared to the enhanced method, giving the same registration results. The Gaussian fitting can accurately locate the [0 9 6] reflection peak in all four power spectra. However, the absolute peak height h_{abs} might not be the most useful measure of the image quality. For example, the mean image of the unregistered series has a substantially lower h_{abs} than the other three, although a higher level of noise in the background of the single image can be seen in the Gaussian fitted peak profile, compared to that of the mean images.

Two supplementary quantities are defined for comparison purpose: the relative peak height h_r being the peak height above the background noise level h_n (Eq. 4.25) and the noise-normalised peak height h_{norm} which is the relative peak height divided

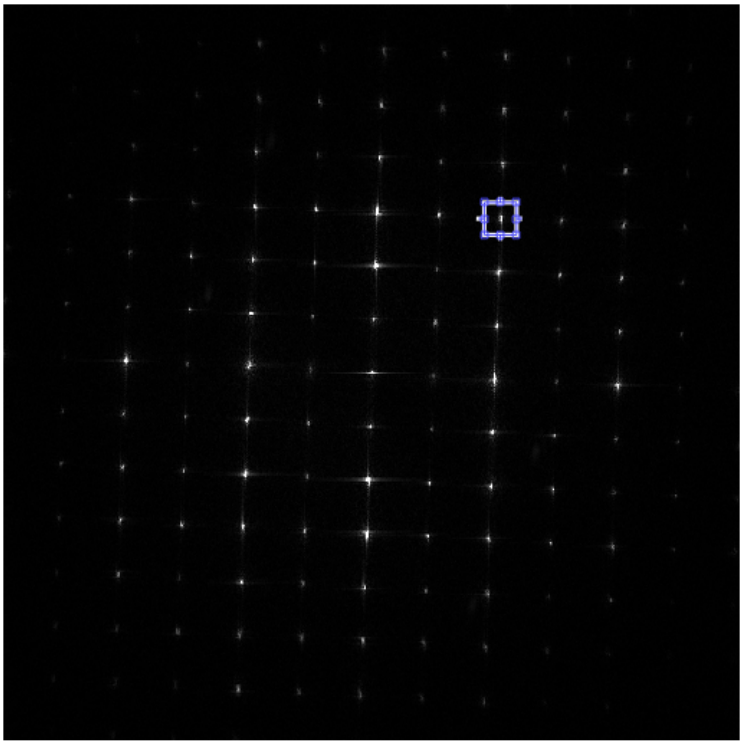


Figure 4.12: Central part ($< \frac{1}{2}$ Nyquist frequency) of the power spectrum region of a hydroxyapatite time series image. The [0 9 6] reflection is marked by the blue square.

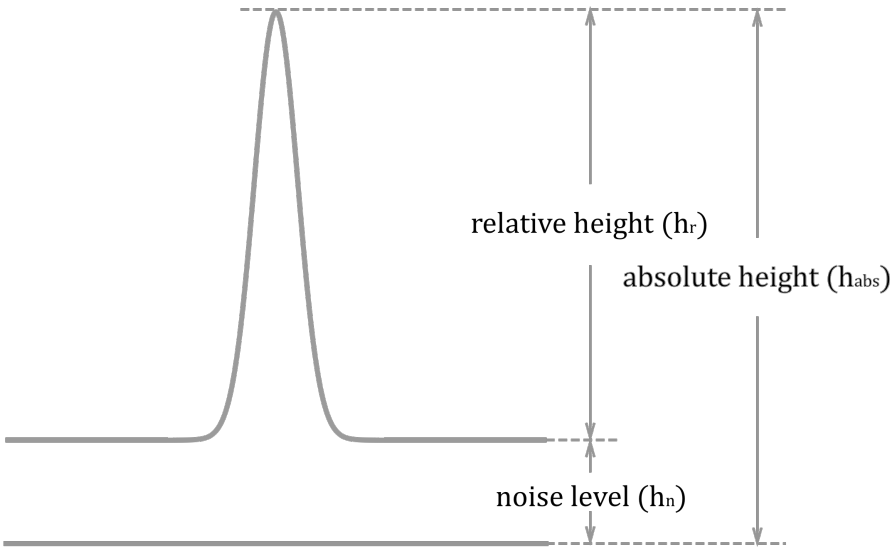


Figure 4.13: Illustration of a Gaussian fitted reflection peak, and the relation between relative peak height h_r and the absolute peak height h_{abs}.

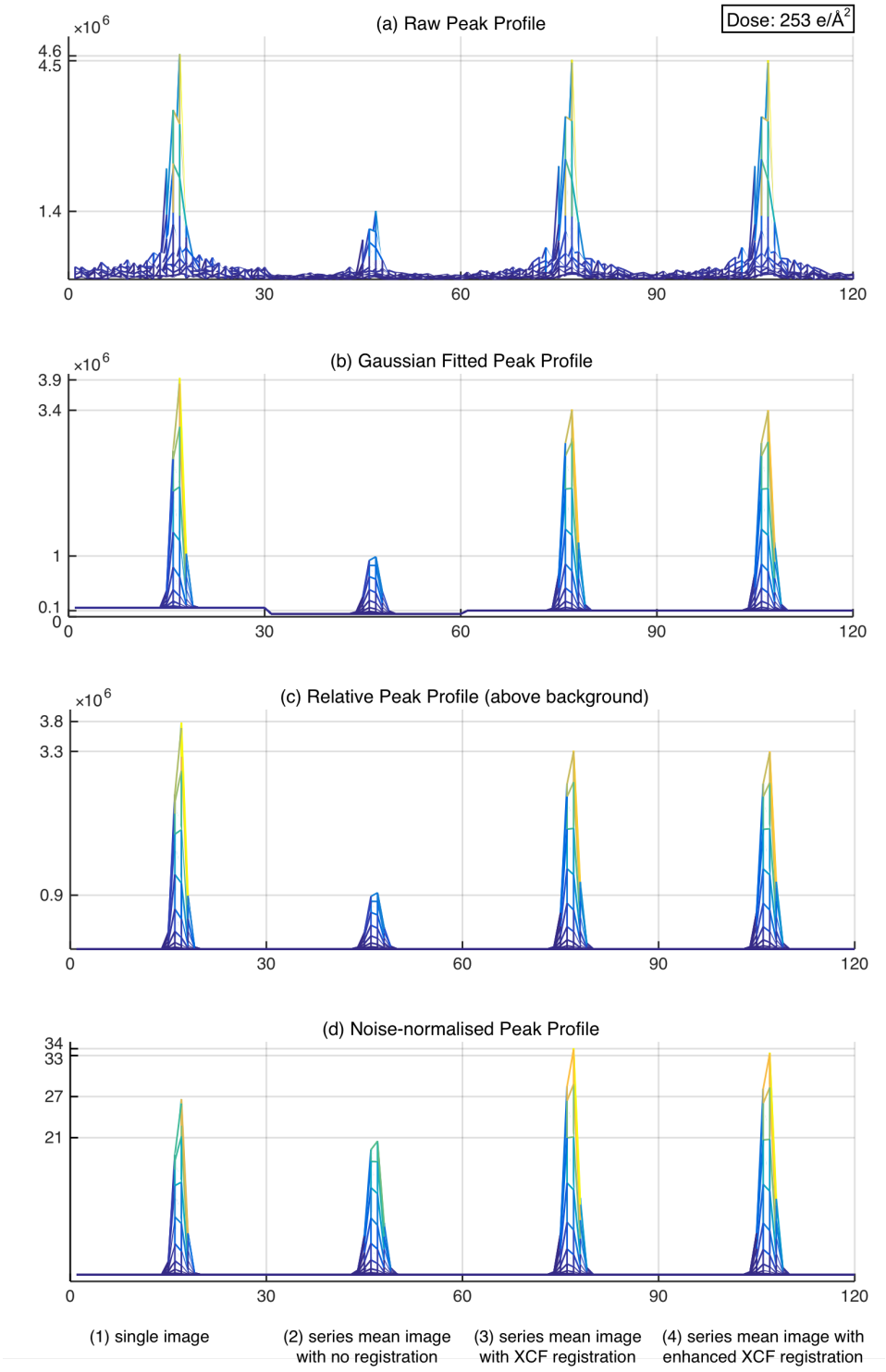


Figure 4.14: The [0 9 6] reflection peak analysis with the HAP time series of 253 $e/\text{Å}^2$. (a) Raw peak profile; (b) Gaussian fitted peak profile; (c) relative peak height (above background); (d) noise-normalised peak height. In the plot, the four peaks are [0 9 6] reflections from the power spectrum of (1) a single image; (2) the series mean image with no registration; (3) the series mean image with XCF registration; (4) the series mean image with enhanced XCF registration.

by the noise background(Eq. 4.26).

$$h_r = h_{abs} - h_n \tag{4.25}$$

$$h_{norm} = \frac{h_r}{h_n} \tag{4.26}$$

The series mean images after shift correction of the two registration methods preserves the relative height of the reflection peak similar to the one in a single image power spectrum.

When looking at the noise-normalised peak height, the improvement brought by the image registration is evident, especially for the enhanced XCF method. With better image alignment, the mean images have a higher noise-normalised peak than the single low-dose image. This is an indication of more pronounced reflection peak and a higher signal-to-noise ratio at the specific spatial frequency. It can be seen from Fig. 4.14, that the growth in signal-to-noise ratio does not come from an amplification of the signal component but from suppression of the noise in the images.

For the lowest-dose series (below $44\,e/\text{Å}^2$), the simple XCF stops functioning and output mean image reduces to a simple mean image without registration. At this dose, the Gaussian fitting is no long capable of detecting the [0 9 6] reflection as it has been merged with the noise spectrum. The enhanced XCF is still effective on series with doses as low as $2.2\,e/\text{Å}^2$. The noise-normalised peak height also demonstrates an appreciable increase in the mean image against each individual images.

4.6.2 IQ factor

The noise-normalised 3D peak profile is a pragmatic and physically sensible measure of the amount of information at a specific spatial frequency. The less ideal aspect of this measure is its limitation in using a very small part of the information presented

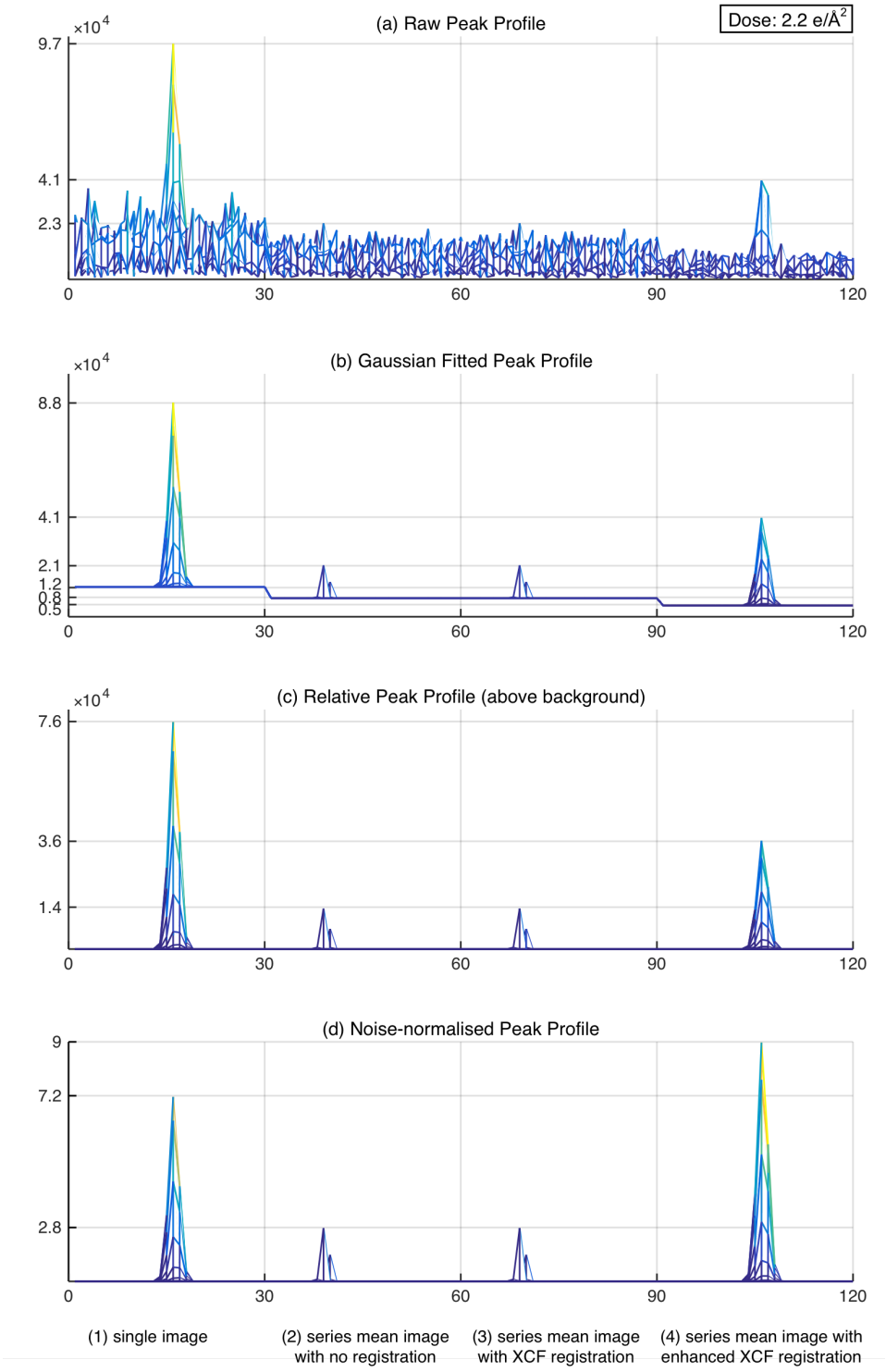


Figure 4.15: The [0 9 6] reflection peak analysis with the HAP time series of 2.2 $e/\text{Å}^2$. (a) Raw peak profile; (b) Gaussian fitted peak profile; (c) relative peak height (above background); (d) noise-normalised peak height. In the plot, the four peaks are [0 9 6] reflections from the power spectrum of (1) a single image; (2) the series mean image with no registration; (3) the series mean image with XCF registration; (4) the series mean image with enhanced XCF registration.

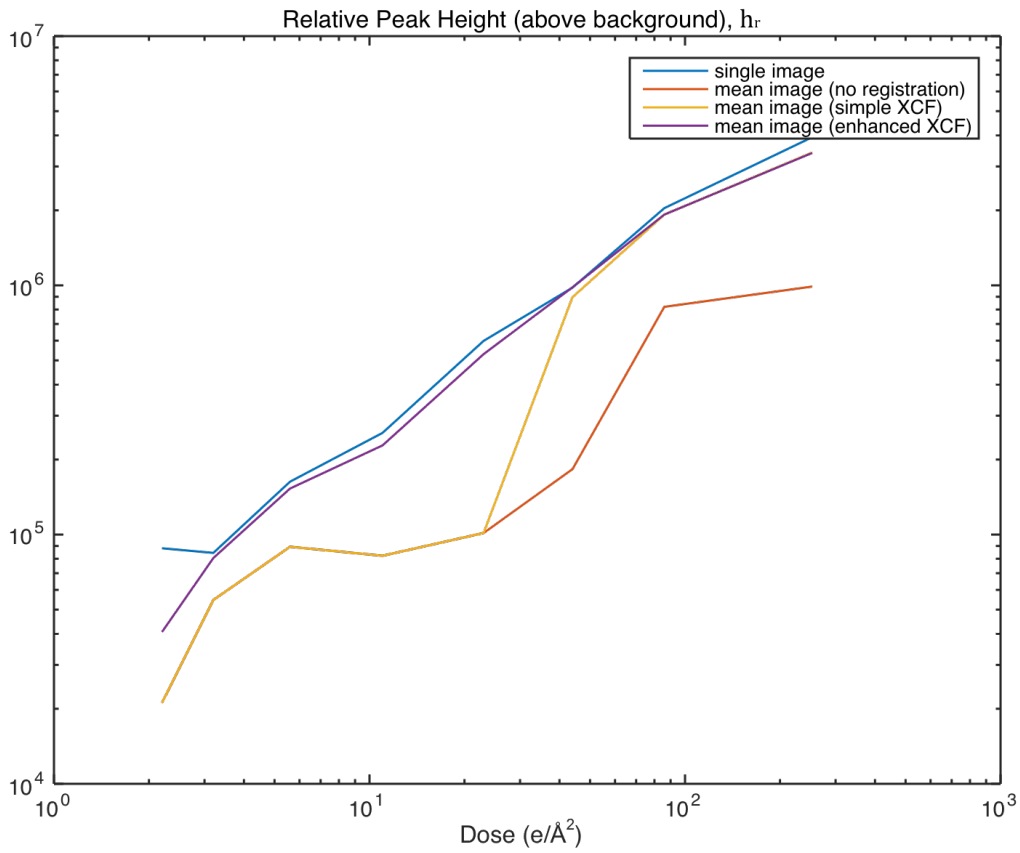


Figure 4.16: The [0 9 6] reflection relative peak height plotted against dose.

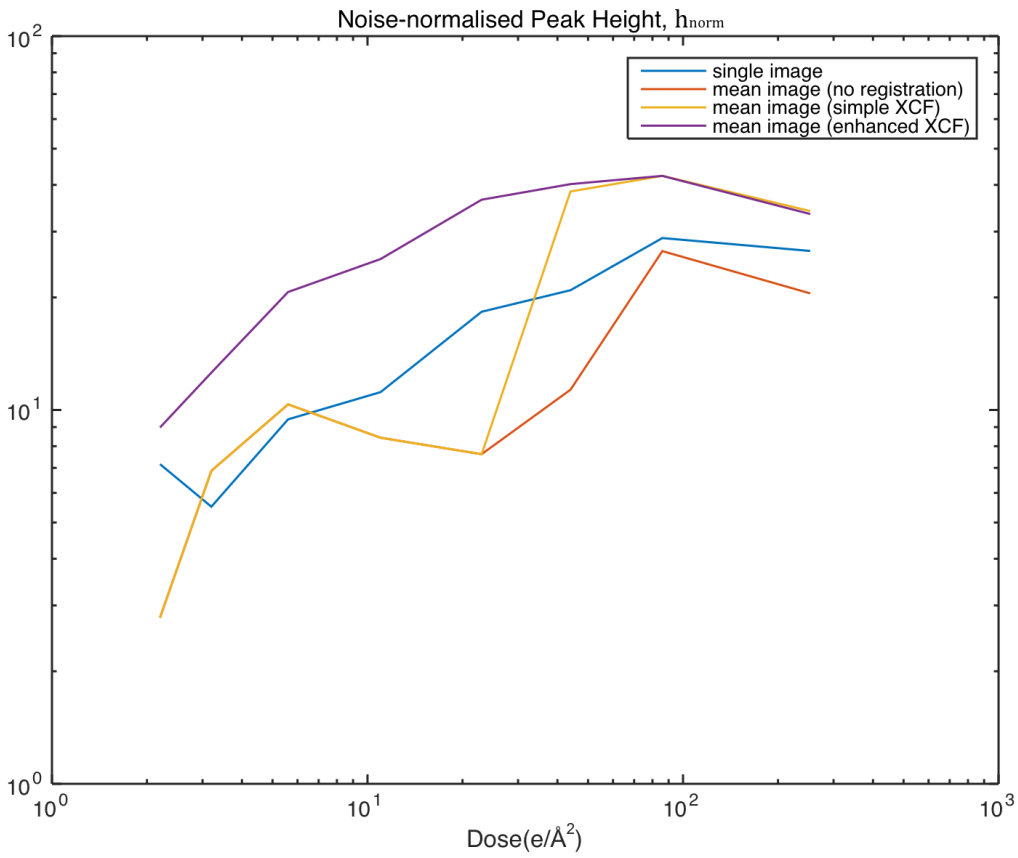


Figure 4.17: The [0 9 6] reflection noise-normalised peak height plotted against dose.

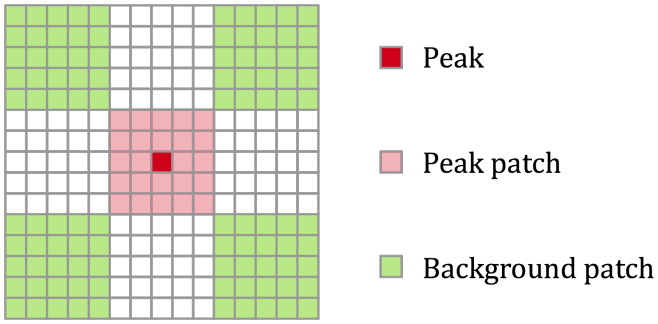


Figure 4.18: A schematic illustration of the IQ factor.

in the power spectrum. This casts doubts on its effectiveness as a comprehensive metric for image quality evaluation.

In search of a quantitative measure that can include all the reflection peaks instead of a specific few while still making use of the similar noise-normalised reflection peak height, a new quantity called the *IQ factor* is proposed and implemented, with "IQ" short for *image quality*.

The implementation of the IQ factor calculation can been divided into two main steps - peak finding and IQ factor calculation.

The peak finding routine is performed entirely automatically with a set of initial parameters and some preset restrictions. The sampling interval and a frequency threshold are set to eliminate aliasing peaks that may appear in the high frequency range where there should not be any interpretable information. A minimal amount of separation is an important parameter which forbids peaks to be too close to each other so that a badly shaped peak would not be recognised as multiple peaks, and abnormal pixels in the vicinity of the reflections are also not selected.

Learning from the experience of the using this peak finding routine, it is often helpful to apply a 2D Hanning window in advance, to reduce aliasing artefacts that affect the search process.

Because the IQ factor is designed to be utilised as a comparison measure, peak searching needs to have a reference image, which is typically a good quality image with reflections well resolved in the PS. The reference image will first go through the peak finding process and the positions of the located peaks will be saved as a guide line for the next image, in which the search of peaks will start from the positions in the reference with restrained adjustment.

After the initial peak finding, individual IQ factors are then calculated for each selected reflection peak as below.

Given an image of $N \times N$ pixels. For a pixel $\mathbf{i} = (i_1,i_2) \in \{1+p, \ldots, N-p\}^2$ let $P_p^{\mathbf{i}}$ denoting the patch of size $(2p+1) \times (2p+1)$ centred at the pixel *veci*, i.e.;

$$P_p^{\mathbf{i}} = \{\mathbf{k} \in \{1, \ldots, N\}^2 : (|i_1 - k_1| \leq p) \wedge (|i_2 - k_2| \leq p)\} \tag{4.27}$$

where $p \in \mathcal{N}$ determines the patch size. A suitable choice of p depends on the magnification of the micrograph and the size of the features in it (for the micrographs of the zeolites the appropriate value is $p = 4$). Furthermore, denote the modulus of the Fourier transform of the image f by $M[f]$ and the modulus at a pixel $\mathbf{i}$ by $M[f](\mathbf{i})$.

If $\mathbf{i}$ is a peak for which one can fit a window $P_{3p+1}^{\mathbf{i}}$ inside the DFT image, the corresponding IQ factor is computed as follows. We calculate the average local background around $\mathbf{i}$ as the average value of M at the union $W(\mathbf{i}) := W_1(\mathbf{i}) \cup W_2(\mathbf{i}) \cup W_3(\mathbf{i}) \cup W_4(\mathbf{i})$ of the windows;

$$\begin{aligned} W_1(\mathbf{i}) &:= P_p^{(i_1+(2p+1),i_2+(2p+1))} \\ W_2(\mathbf{i}) &:= P_p^{(i_1-(2p+1),i_2+(2p+1))} \\ W_3(\mathbf{i}) &:= P_p^{(i_1-(2p+1),i_2-(2p+1))} \\ W_4(\mathbf{i}) &:= P_p^{(i_1+(2p+1),i_2-(2p+1))} \end{aligned}$$

This gives that the IQ factor at a spot $\mathbf{i}$ is;

$$IQ[f](\mathbf{i}) := \frac{M(\mathbf{i})}{\frac{1}{4(2p+1)^2}\sum M[f](\mathbf{k})} - 1 \tag{4.28}$$

The calculation is also illustrated in Fig. 4.18, which shows an example of an identified peak, with a patch size of 5×5 ($p = 2$). The mean value of the four green patches is used as the background value, which when divided by the deep red peak pixel becomes the IQ factor of this peak selection. This makes it very similar to the noise-normalised peak profile, except for that the peak is not 2D Gaussian fitted. However, the value calculated here should very close to that of the latter, as the Gaussian fitting uses the maximum pixel as the initial position of the centre and the averaged background should not be much different from a fitted background.

Algorithm 1 Peak initialization.

Parameter θ_{ratio}, the minimal peak intensity ratio
Parameter θ, the minimal peak background ratio
Parameter q, half of minimal peak distance
$g = M[f_{\text{ref}} - \overline{f_{\text{ref}}}]$ ▷ initialize g with the modulus of f, removing frequency for the mean value of f
$g = \frac{g - \min g}{\max g - \min g}$ ▷ normalize the values of g to $[0, 1]$
$\mathcal{I} = \emptyset$ ▷ build a list with potential peaks
repeat
    Select the pixel position $\mathbf{i}$ in g with the highest intensity.
    if $g(\mathbf{i}) > \theta_{\text{ratio}}$ **then**
        $\mathcal{I} \leftarrow \mathcal{I} \cup \{\mathbf{i}\}$ ▷ Add $\mathbf{i}$ to the list of potential peaks.
        Replace the values of g in the patch $P_q^{\mathbf{i}}$ with zero
    else
        break ▷ no potential peak candidates left
    end if
until true
$\mathcal{I} \leftarrow \{\mathbf{i} \in \mathcal{I} : \mathbf{i} = (i_1, i_2) \in \{1 + p, \ldots, N - p\}^2\}$ ▷ Discard peaks too close to the boundary
$\mathcal{I} \leftarrow \left\{\mathbf{i} \in \mathcal{I} : \frac{M[f_{\text{ref}}](\mathbf{i})}{\text{average}(M[f_{\text{ref}}](P_p^{\mathbf{i}}))} \geq \theta\right\}$ ▷ Discard peaks that don't exceed the background by a given ratio

The list of peaks from a reference image f_{ref} is determined as follows: First, a list

of potential peaks is generated with Algorithm 1. Then, the final list of peaks is determined by further peak exclusion with Algorithm 2. Here, the frequency corresponding to $\mathbf{i}$ is

$$\nu(\mathbf{i}) = ||(\tfrac{1}{N-1}(N/2+1-i_1), \tfrac{1}{N-1}(N/2+1-i_2))||$$

The last criterion in Algorithm 2 is essentially a value threshold, which tries to exclude peaks with very small IQ factors. These peaks are often pseudo maxima that are nothing more than a brighter pixel in the background noise. The resulting list of peaks $\mathcal{I}_{\text{ref}} \subset \{1,\dots,N\} \times \{1,\dots,N\}$ is saved and used as basis for computing the IQ factor sum of an image, which is a global measure of the image quality.

Algorithm 2 Peak refinement.

1: **Parameter** θ_{freq}, the maximal frequency threshold
2: **Parameter** θ_{IQ}, the minimal IQ value threshold
3: $\mathcal{I} \leftarrow \{\mathbf{i} \in \mathcal{I} : \mathbf{i} = (i_1,i_2) \in \{1+(3p+1),\dots,N-(3p+1)\}^2\}$ ▷ Discard peaks for which one cannot fit a window $P^{\mathbf{i}}_{3p+1}$ inside the DFT image
4: $\mathcal{I} \leftarrow \{\mathbf{i} \in \mathcal{I} : i_1 \geq N/2\}$ ▷ Discard peaks in the left quadrants
5: $\mathcal{I} \leftarrow \{\mathbf{i} \in \mathcal{I} : \nu(\mathbf{i}) \leq \theta_{\text{freq}}\}$ ▷ Discard peaks with very high frequency
6: $\mathcal{I} \leftarrow \{\mathbf{i} \in \mathcal{I} : IQ[f_{\text{ref}}](\mathbf{i}) > \theta_{\text{IQ}}\}$ ▷ Discard peaks with very small IQ factors

The IQ factor sum of an image f is determined as follows. The list of potential peaks is initialised with $\mathcal{I}_{\text{ref}}$. Note that this initialisation only depends on the reference image, not the the current image. Then these peaks are locally aligned to the maxima of the modulus of f, i.e. each $\mathbf{i}$ is replaced by the position of the maximum of $M[f]$ in the local windows $P^{\mathbf{i}}_{q/2}$ (for a suitable width q). The resulting list is cleaned with Algorithm 2, which gives the list of peaks used in the IQ factor sum:

$$IQ_{sum} := \sum IQ[f](\mathbf{i}) \tag{4.29}$$

The reason for adding the negative offset of 1 to the equation (4.28) is to get rid of the increase in the IQ_{sum} when the number peaks increases. With this bias term, the IQ factor of area of a uniform intensity would be 0 instead of 1 and the value of

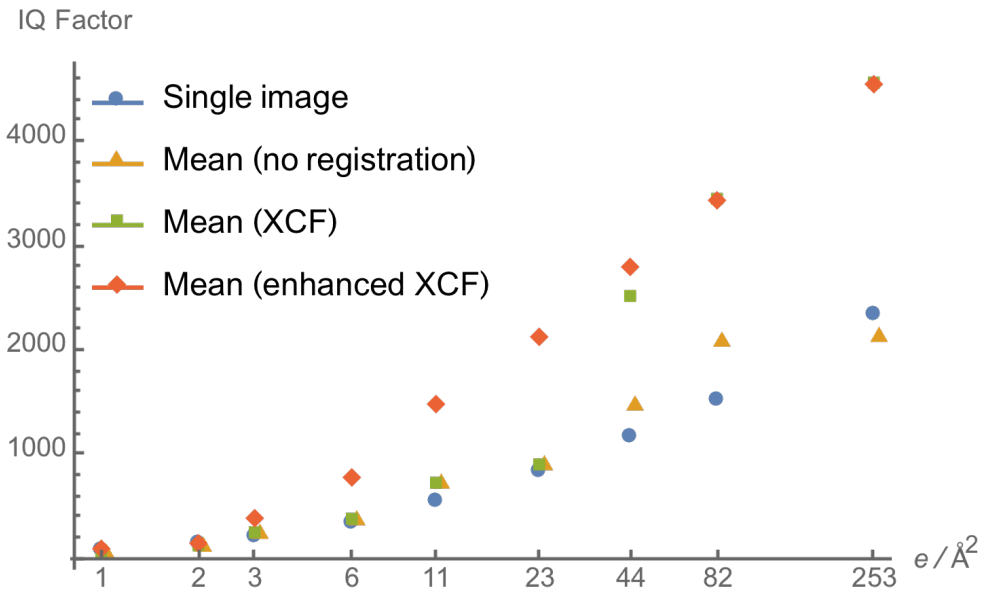


Figure 4.19: IQ factor comparison of single image, mean image with no registration, mean image after conventional XCF, and mean image after enhanced XCF.

an area of pure noise but with small variations would be little higher than 0. This is an important modification to ensure that the increase in the IQ factor is driven by the improving SNR against their surrounding background rather than the total number of peaks used for the calculation, especially when there are a large number of small weak high frequency reflections.

4.6.3 IQ factor comparison, HAP time series

The same set of mean images analysed by the noise-normalised 3D peak profile in Section 4.6.1 were processed by the new IQ factor evaluation for cross-check comparison. The calculated IQ factors are summarised into a graph as shown in Fig. 4.19. The mean images calculated after the enhanced XCF have demonstrated superiority over the conventional XCF in the low dose data below $82\,e/\text{Å}^2$. The higher IQ factors indicate better transfer for crystallographic information collectively across the frequency domain. It can be seen that mean images after the conventional XCF give identical

IQ factors as mean images without any registration due to the artefact pixels in the XCF map. The IQ factor results are therefore consistent with the improvement that was observed in real space mean images Fig. 4.11.

A novel evaluation method of the global image quality in the frequency domain has been established and tested. In addition to evaluating typical HRTEM images and mean images of a time series acquisition, it will be seen in the next chapter that the IQ factor can also be used as a measure of the exit wave reconstruction quality, for which there has not previously been an agreed measure. It is believed that this measurement has the potential for an even wider application to image quality measurement of any images that aim at recording periodic structures, as well as to image processing techniques that aim at retrieving information from such images.

Chapter 5

Exit Wave Reconstruction

5.1 Materials

5.1.1 Cerium dioxide

Cerium dioxide (CeO_2) is a rare earth metal oxide with a cubic unit cell, belonging to the space group $Fm\overline{3}m$ (Fig. 5.1). Cerium dioxide nanoparticles are being actively studied for catalysis (Kašpar et al., 1999; Trovarelli, 1996). Its other applications range from solid-oxide fuel cells (Stambouli and Traversa, 2002) and solar cells (Zitova and Flusser, 2003) to potential pharmacological agents (Celardo et al., 2011) and treatments for neurodegenerative diseases (Xu and Qu, 2014; Das et al., 2013).

The CeO_2 sample used in this work comes in the form of nanoparticles and TEM specimens were prepared by the ultrasonic dispersion method that is standard for nanoparticle TEM sample preparation.

5.1.2 Silicon nitride

Silicon nitride is a well-known ceramic material. The β-Si_3N_4 studied here belongs to space group $P6_3$ and has a hexagonal structure. The atomic structure of Si_3N_4 attracts particular interests from the electron microscopy community because of the close

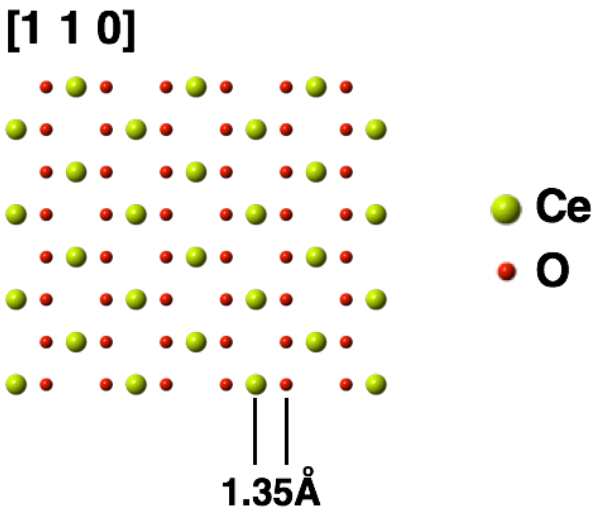


Figure 5.1: Atomic model of Cerium dioxide, [1 1 0] direction.

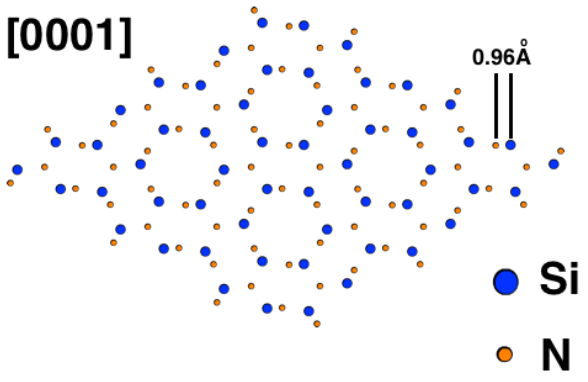


Figure 5.2: Atomic model of Silicon nitride, [0 0 0 1] direction.

spacing (approximately 0.96 Å) between the silicon and nitrogen columns projected along the [0 0 0 1] zone axis, as shown in Fig. 5.2.

The sample preparation of the Si_3N_4 TEM specimen was carried out by low-energy ion milling, which produces thin clean edges that are ideal for HRTEM imaging. Compared to the other materials studied in this thesis, silicon nitride is relatively radiation-resistant. Under the chosen imaging settings (Table 5.3), the thin edge appear to be very robust and was unchanged over the course of the TEM experiment.

5.2 Low-dose focal series imaging

Focal series imaging is a common acquisition geometry for exit wave reconstruction. The defocus of each image in the series needs to be known. In order to determine the defocus values, the focal step between each image and the absolute defocus of one of the images is in principle sufficient to calculate the absolute defocus values of all other images.

The true focal step is closely related to the digital-to-analogue converter (DAC) of the objective lens, which links the digital control and the analog objective lens current. For a calibrated instrument, the change in defocus of the objective lens corresponding to 1 DAC is known and the chosen focal step for a focal series acquisition needs to be an integer multiple of the smallest defocus corresponding to 1 DAC.

Other residual aberrations, such as the spherical aberration, can be determined during the aberration corrector adjustment. Subsequently they can be input into the exit wave reconstruction computation.

Focal series data acquisition was carried out on both of the materials described in Section 5.1. Between the two studied materials, cerium dioxide is a good example of a nanoparticle sample, and silicon nitride is representative of a bulk material.

Parameter	Value
Voltage	300 kV
Defocus range	−20 nm to 20 nm
Focal step	2 nm
Number of images	21
Spherical aberration (C_3)	−1.6 µm
Sampling interval	0.009 87 nm

Table 5.1: Imaging conditions for the CeO_2 focal series data.

.

Series name	Dose per image ($e/Å^2$)	Average count/pixel
CeO_fs1	223	6.339
CeO_fs2	525	15.717
CeO_fs3	1010	30.096
CeO_fs4	1940	58.561
CeO_fs5	3650	108.957

Table 5.2: Dose and average count for different CeO_2 focal series.

.

5.2.1 CeO_2 focal series

The focal series acquisition of a CeO_2 nanoparticle used five different doses (Table 5.2). The dose adjustment in the experiment was carried out by altering the extraction voltage of the electron gun. Apart from the extraction voltage, all other microscope configurations were kept constant between the series. Experimental parameters relevant to the exit wave reconstruction are listed in Table 5.1. Each series contained 21 images with a defocus range from −20 nm to 20 nm. At a magnification of $\times 10^6$, the sampling interval was measured to be 0.009 87 nm at the detector.

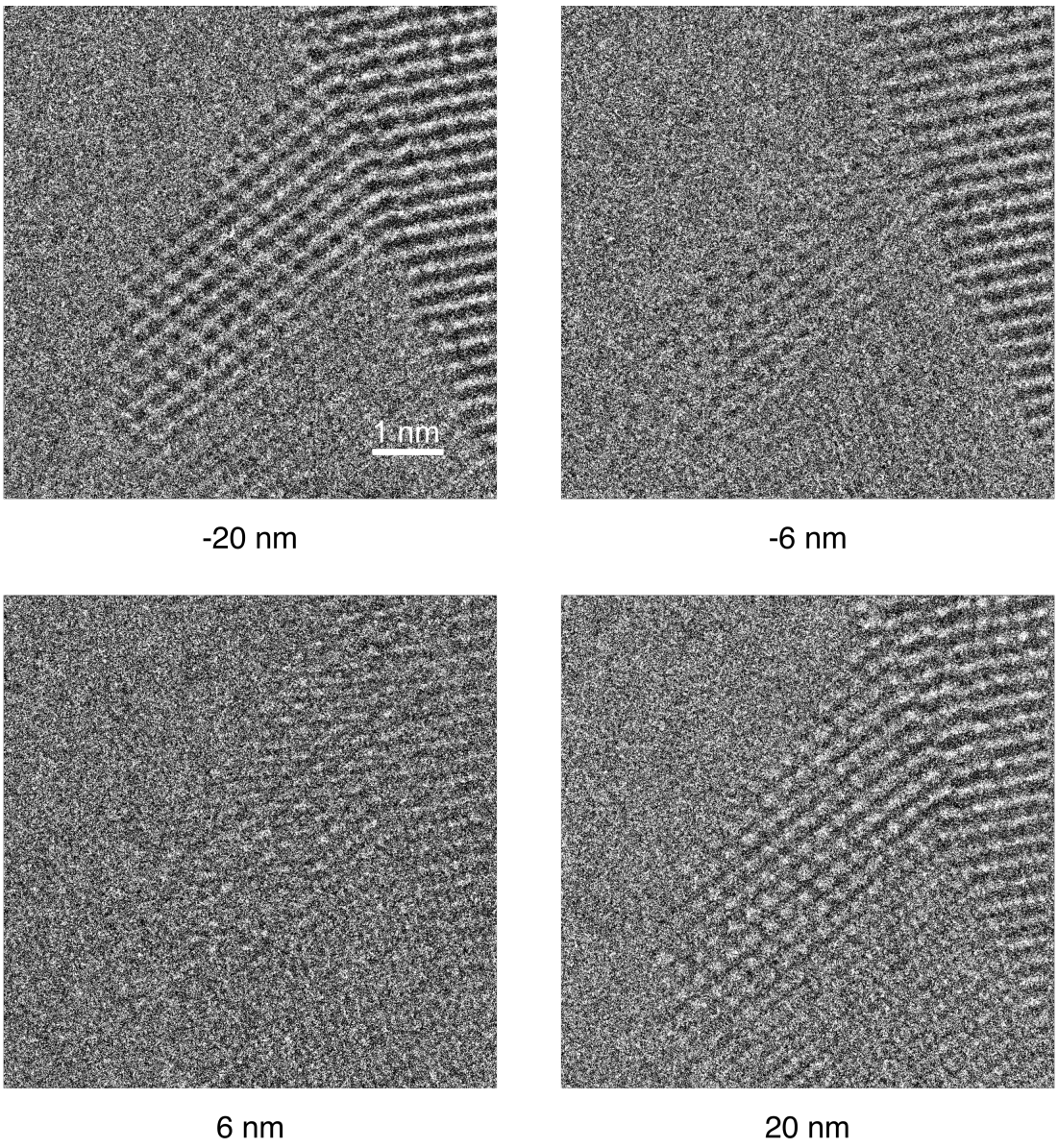


Figure 5.3: Focal series images of a cerium dioxide nanoparticle along [1 1 0] taken at $223\,e/\text{Å}^2$.

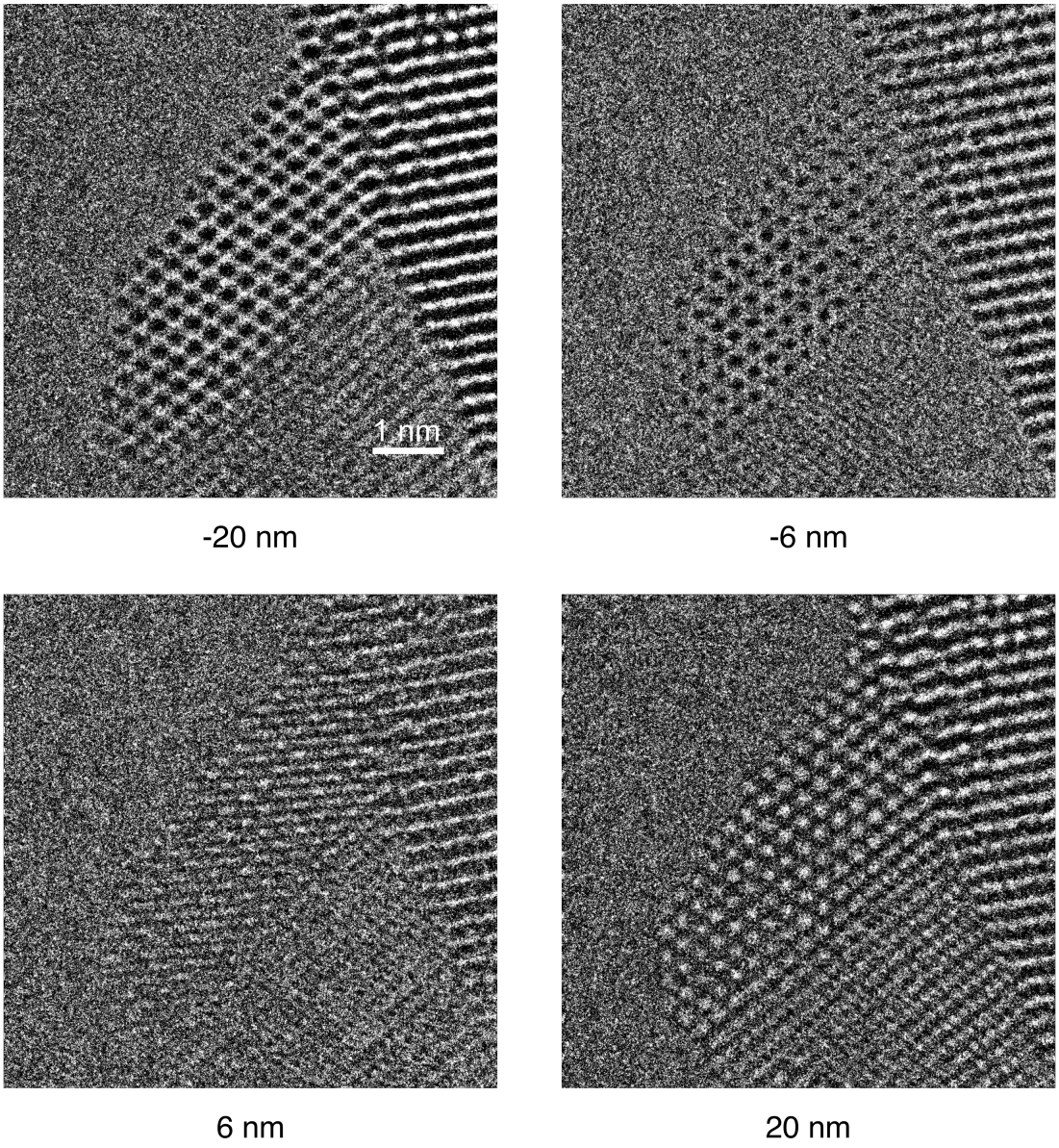


Figure 5.4: Focal series images of a cerium dioxide nanoparticle along [1 1 0] taken at $525\,e/\text{Å}^2$.

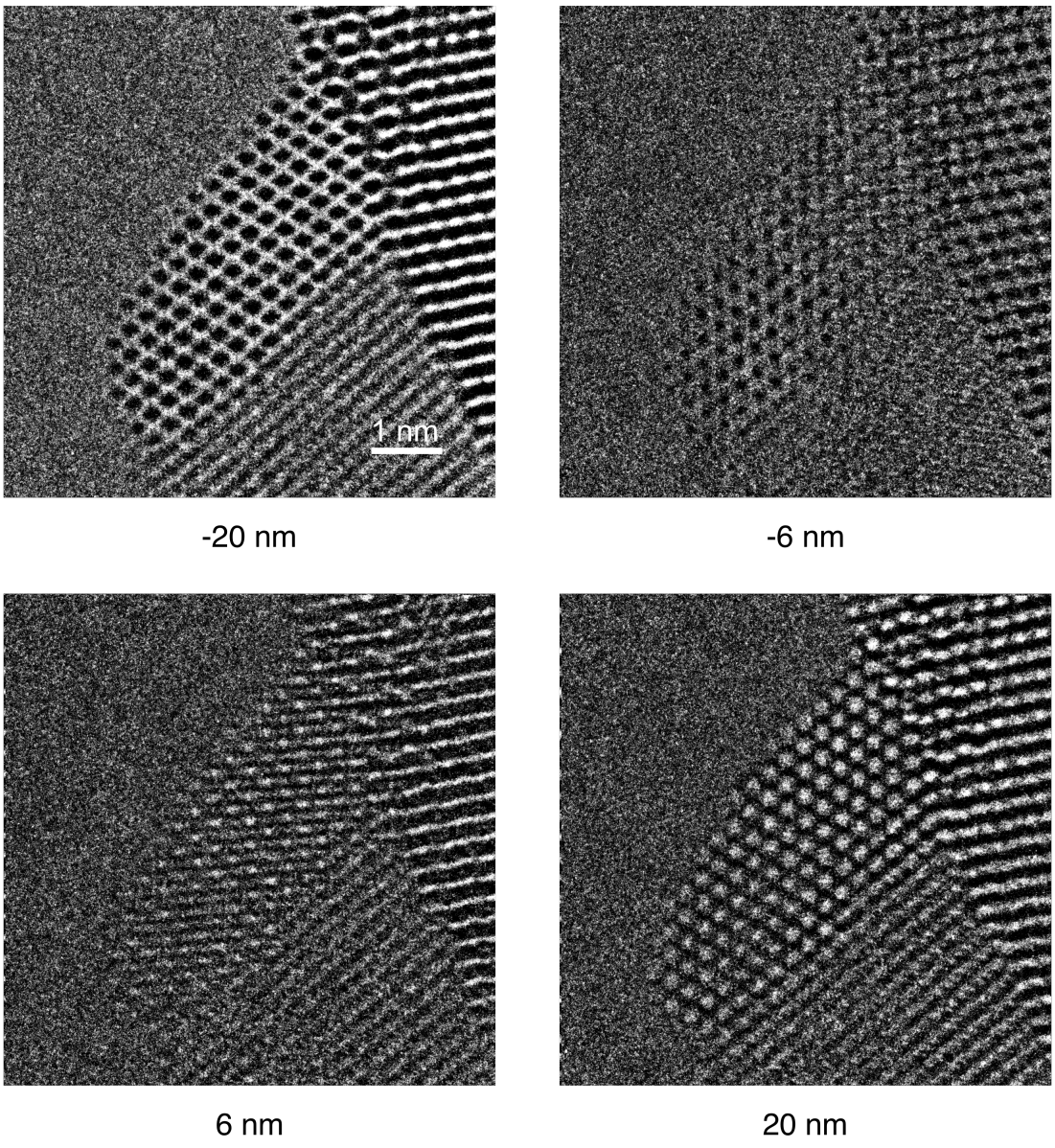


Figure 5.5: Focal series images of a cerium dioxide nanoparticle along [1 1 0] taken at $1010\,e/\text{Å}^2$.

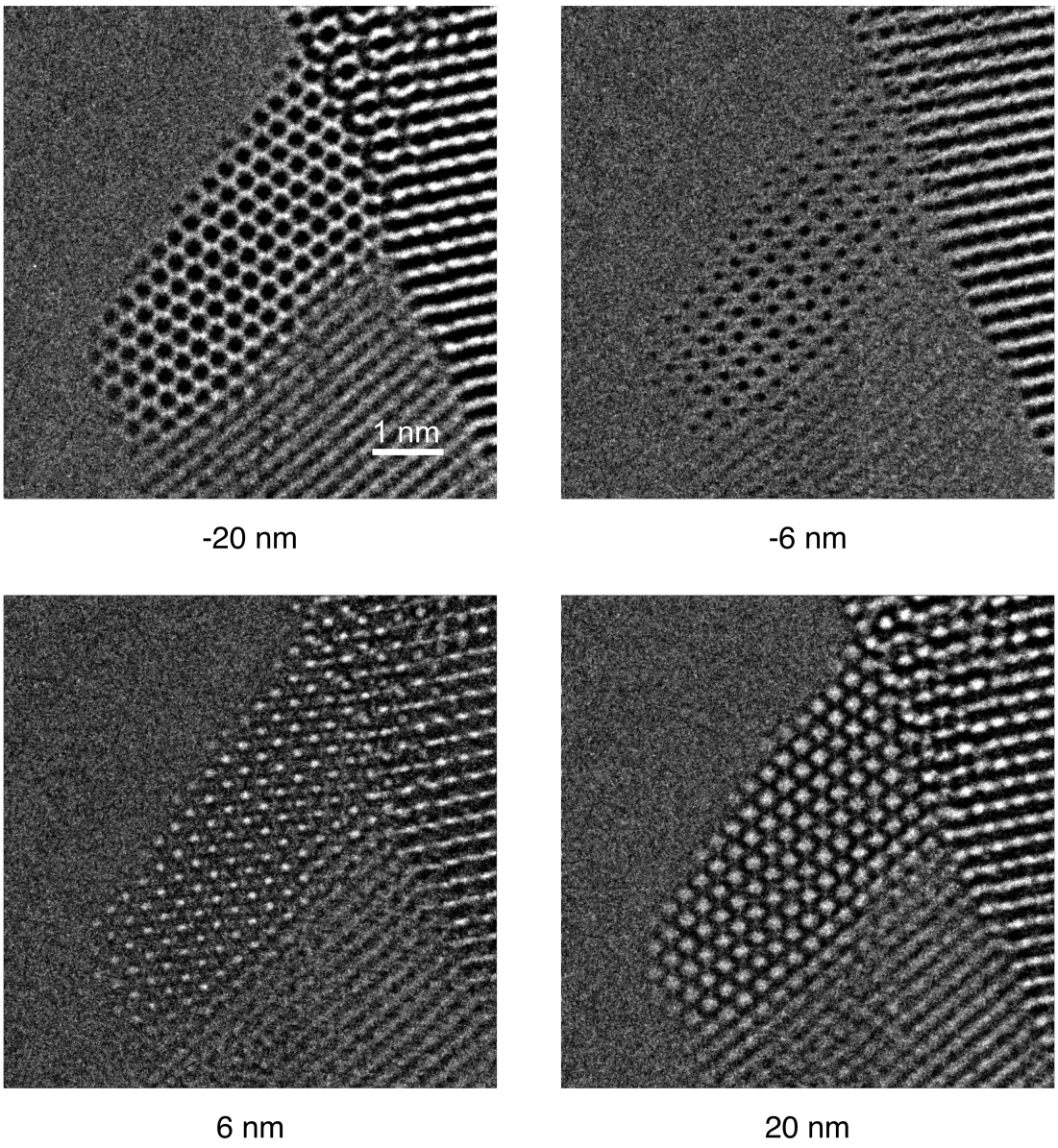


Figure 5.6: Focal series images of a cerium dioxide nanoparticle along [1 1 0] taken at $1940\,e/\text{Å}^2$.

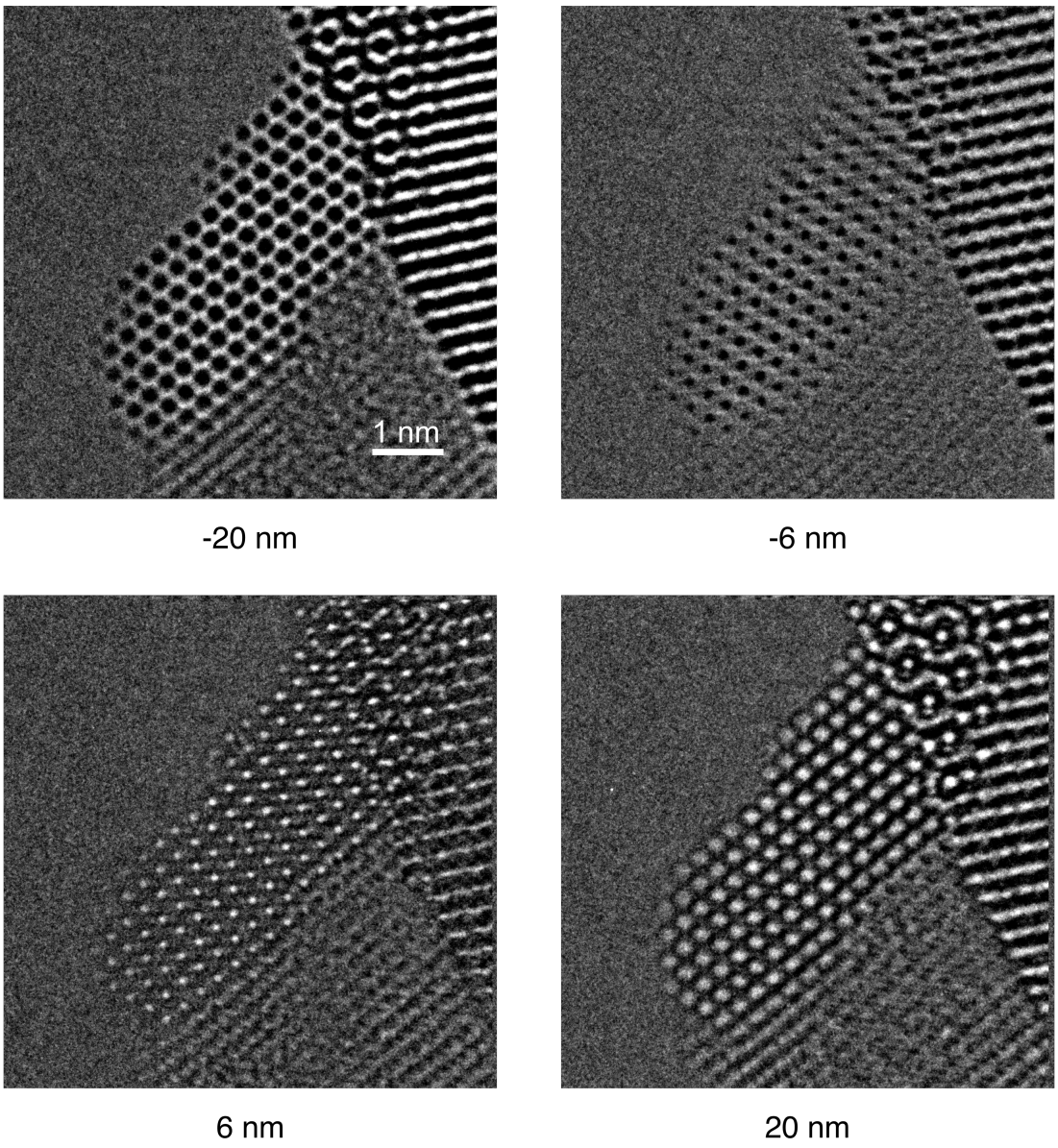


Figure 5.7: Focal series images of a cerium dioxide nanoparticle along [1 1 0] taken at $3650\,e/\text{Å}^2$.

5.2.2 Si_3N_4 focal series

Four focal series of the Si_3N_4 sample were recorded (Table 5.4). The dose control was carried out by changing the exposure time. This method for controlling the dose is easier to operate than changing the extraction voltage and can achieve consistent

Parameter	Value
Voltage	300 kV
Defocus range	−40 nm to 38 nm
Focal step	2 nm
Number of images	40
Spherical aberration (C_3)	−5.66 µm
Sampling interval	0.008 24 nm

Table 5.3: Imaging conditions for the Si_3N_4 focal series data.

.

Series name	Dose per image ($e/Å^2$)	Average count/pixel
SiN_fs1	16	0.756
SiN_fs2	31	1.522
SiN_fs3	177	8.486
SiN_fs4	1889	90.707

Table 5.4: Dose and average count for different Si_3N_4 focal series.

.

imaging results (Fig. 5.8 and Fig. 5.9) with an identical microscope configuration. The imaging parameters of this focal series are listed in Table 5.3.

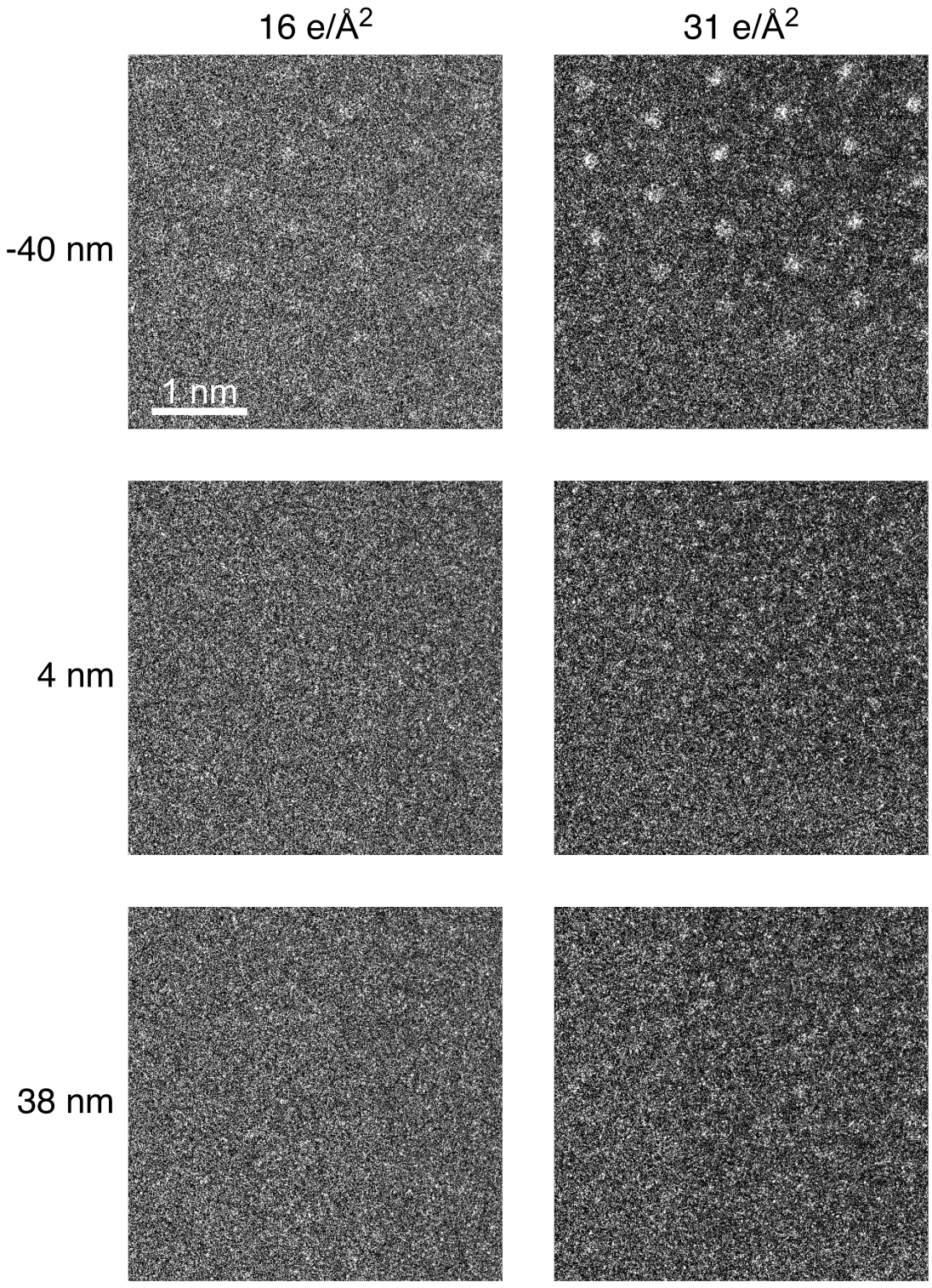


Figure 5.8: Focal series image data of silicon nitride along [0 0 0 1] taken at $16\,e/\text{Å}^2$ and $31\,e/\text{Å}^2$.

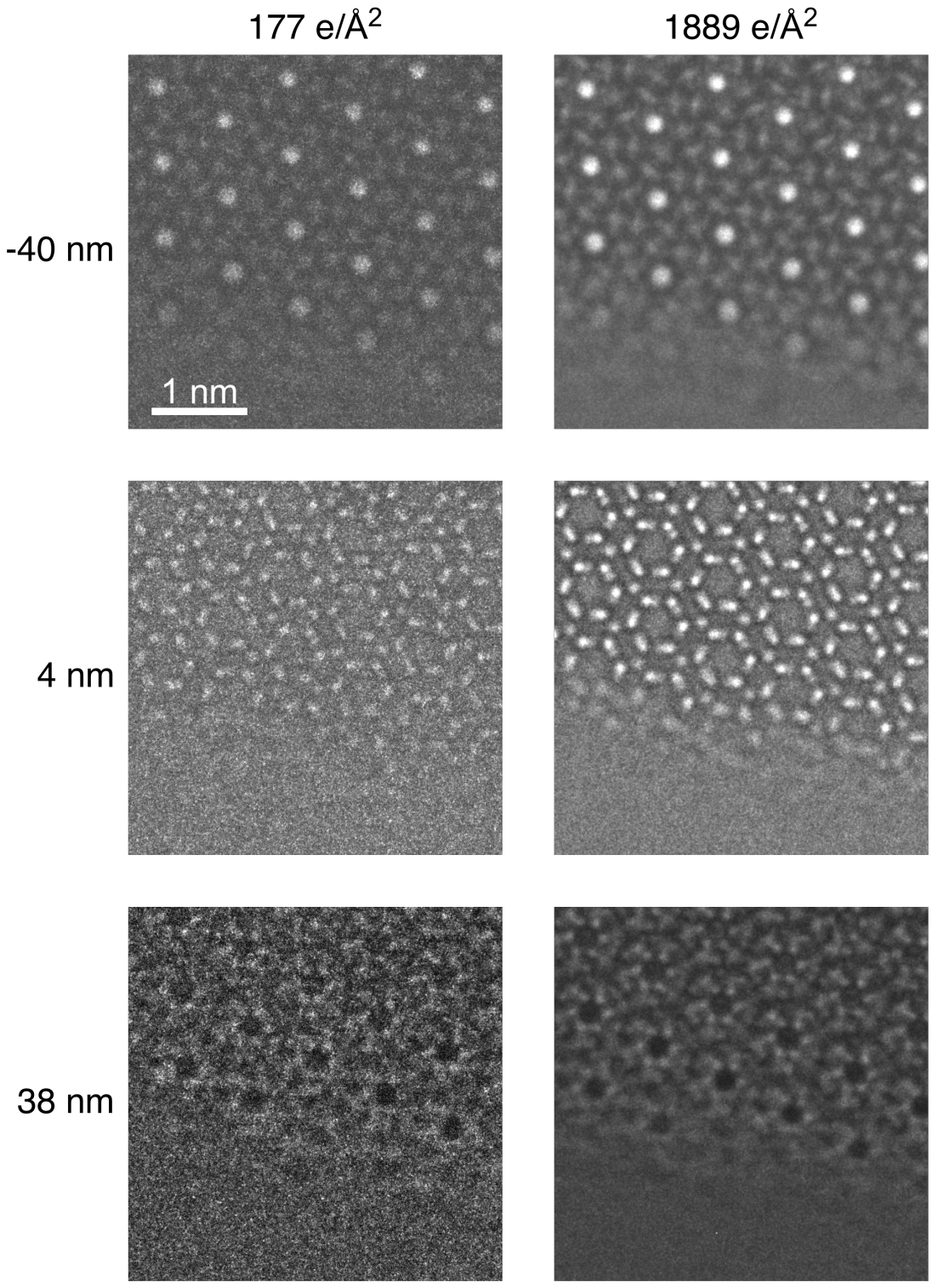


Figure 5.9: Focal series image data of silicon nitride along [0 0 0 1] taken at 177 $e/\text{Å}^2$ and 1889 $e/\text{Å}^2$.

5.3 Striping artefact removal

It has been noticed that when imaging at very low dose levels, striping artefacts (or striping noise) can appear in the image data. Images corrupted by striping noise show horizontal (or vertical, depending on how the detector is placed) stripes that are parallel to each other. These stripes can appear to be either darker or brighter than their adjacent lines of pixels. For example, one of the images in series SiN_fs2 (Fig. 5.10a) was taken with a very short exposure time and the average CCD count in a single image is 1.522 per pixel. It can be clearly seen that some bright/dark stripes are corrupting the raw experimental images. Similar striping artefacts are also evident in the low-dose cerium dioxide data (series CeO_fs1 and CeO_fs2)[1]. Fundamentally, this striping noise problem is detector-related. The seriousness of this type of artefact is associated with the averaged digital count but less with the absolute dose.

Striping noise does not only exist in low-dose electron microscopy, but can also be found in other research fields that use CCD detectors to acquire images at low signal, such as astronomy (Amraei and Mobasheri, 2014). Solutions have already been proposed and tested in astronomical image processing, but have not yet been introduced to electron microscopy (Amraei and Mobasheri, 2014; Grogin et al., 2010).

In order to remove the striping artefact in the low-dose images of the CeO_2 and Si_3N_4 data, a histogram correction method has been developed to remove the striping noise as illustrated in Fig. 5.11. Because the 4K×4K UltraScan CCD detector is an assembly of four smaller 2K×2K CCDs, the image correction needs to be done on each row of the four component CCDs individually.

A complete description of the stripe reduction procedure is as below;

Step 1: Eliminate outliers by forcing the brightest 0.1% pixels to the value of the

[1]The selected images shown in Fig. 5.3 - Fig. 5.9 have all been processed by the striping noise reduction routine and hence do not contain the artefact described here.

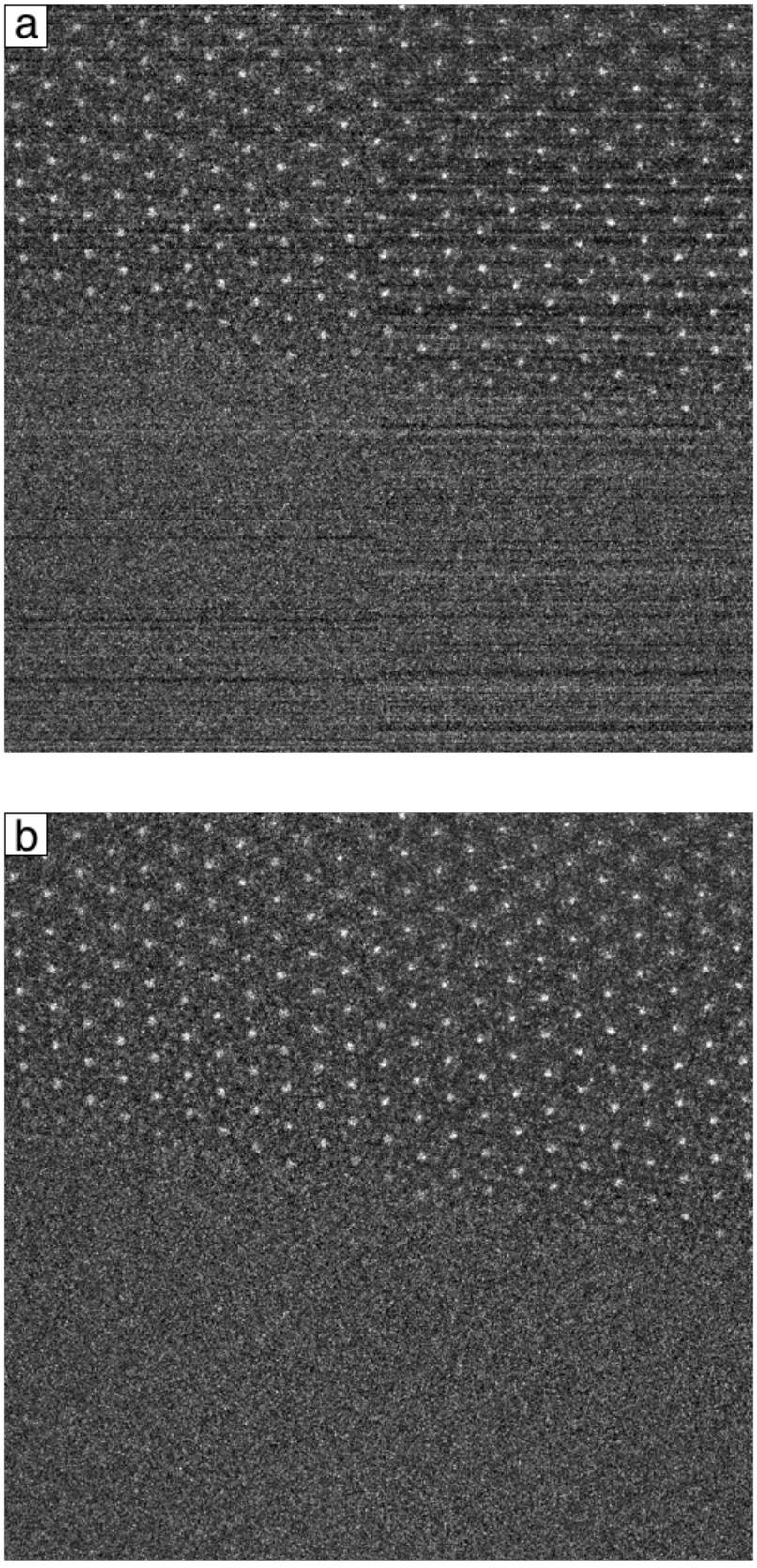


Figure 5.10: A Si_3N_4 image (a) before and (b) after striping noise reduction.

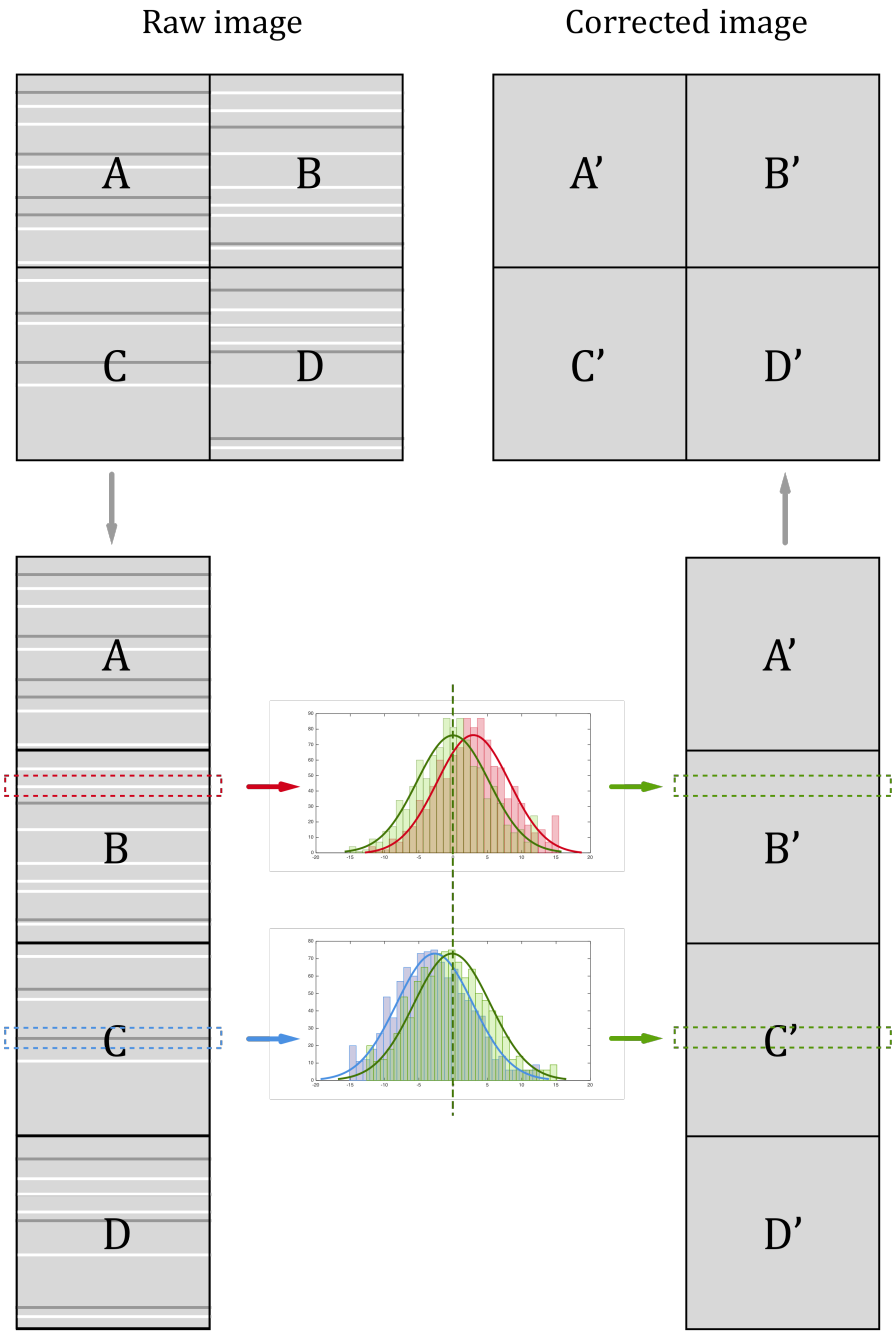


Figure 5.11: Illustration of the striping noise removal procedure. The raw image that contains horizontal striping noise undergoes a line-by-line histogram adjustment to neutralise the erroneous detector bias and force every line to have the same mean value. The camera used here is composed of four smaller CCDs, which need to be adjusted individually.

99.9 percentile and darkest 0.1% pixels to the 0.1 percentile.

Step 2: Normalise the image to a chosen range (for example, [0, 1]); calculate the mean of the image for later use.

Step 3: Separate the four 2K×2K sections and rearrange these into a 2K×8K image with the short side being the parallel direction to the stripes.

Step 4: Plot and fit the histogram of each row of the data and shift the fitted histogram centre to zero.

Step 5: Perform the reverse of Step 3 to reshape the image back to its original state.

Examining Fig. 5.10b, the applied striping noise reduction proved to be very effective in cleaning the defective stripes in ultra low dose images. The images used in this work have therefore all been processed by this striping noise reduction routine for consistency, although the improvement is more important for lower dose images than for the higher dose ones.

5.4 Registration of focal series

Registration of images in a focal series aims to align the same objects in images taken at variable focus values, which is important for the application of exit wave reconstruction. As discussed in Section 3.2.1, the HRTEM imaging process is related to the contrast transfer function which is an oscillating function that may contain either positive or negative phase information transfer. The contrast transfer function is determined by the objective lens aberrations, among which the low order aberrations have the most influential effects on the overall shapes of the transfer function (Fig.3.3). This makes defocus one of the most important factors in determining image phase contrast, especially for high resolution images where defocus is typically

chosen to be close to the Gaussian defocus ($C_1 = 0$). The image contrast is therefore extremely sensitive to any focus variation. Moreover, phase contrast in HRTEM images can reverse with changes of only few nanometres in the focus value. In the case of the CeO_2 focal series where the crystalline structure is relatively simple and the atom columns are well separated, when the series crosses the zero focus point, dark atoms in the image become bright (Fig. 5.3 to Fig. 5.7). If the unit cell structure is more complicated and the atoms are more compact in the direction of imaging, the contrast change can cause significant difficulty in correlating different images with each other. All of the changes complicate the image registration, as the correlation functions between images are calculated based on image contrast and the reversed contrast gives a low correlation. No existing registration method can overcome the image contrast change problem arising from different defocus values. In the following sections, attempts to minimise the negative impact of this contrast reversal problem have been made, giving rise to a more sophisticated and improved registration routine as discussed subsequently.

5.4.1 The neighbour-reference XCF registration

The contrast change in a focal series is a gradual process. The similarity between two consecutive images is the highest among all image pair combinations. This property can be made use of in image registration. It is natural to consider implementing a registration procedure that is based on neighbour-reference XCF, as illustrated in Fig. 5.12 for N images and propagating the image shift through the focal series. Ideally, if the registration is perfect for all adjacent image pairs, the output image series should yield the best alignment regardless of any contrast reversal. In reality, however, this neighbour-reference scheme still faces restrictions in several aspects.

Firstly, the discrete nature of the cross-correlation function causes it to only return integer pixel numbers for the image shift. Registration between two digital images

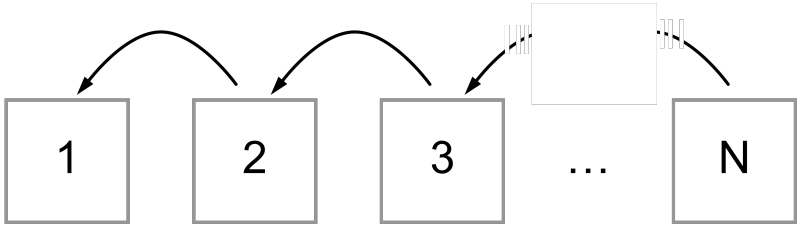


Figure 5.12: Schematic illustration of the neighbour-reference XCF registration procedure. Each image uses its previous image as the reference.

is limited to pixel level accuracy, implying that the best alignment attempt can be no more accurate than the size of a pixel in the images. Hence, the error of one registration operation can be as large as half the side length of a pixel. This imposes constraints on the length of series suitable for the neighbour-reference method. If the focal series is long, propagating all previous image shifts to the registration can result in large misalignment between two images that are far apart in the series, as all alignment errors from the earlier registration steps are passed on to the later steps.

Secondly, the neighbour-reference scheme is superior to the single-reference scheme in that for each registration the reference and the subject images are less likely to be affected by the contrast change than when using a specific reference image to register the whole series. However, the minimal image contrast at zero defocus can remain an major obstacle. The contrast transfer function changing from positive transfer to negative transfer may happen within a very narrow focus range near zero defocus and even a single focal step may prove to be too large. If one or more of the images have a defocus value very close to zero, the contrast transfer can be virtually zero across the frequency domain (green lines in Fig. 5.13), i.e. minimum contrast in the image. This results in an image containing only noise. In most failed registrations, the registration of near-focus images is indeed the "weak link in the chain" for any focal series image alignment. Eliminating the low contrast images does not resolve this problem as the contrast reversal will then be even more abrupt at the transition

stage.

Moreover, as shown in Chapter 4, the fewer electrons that are used for imaging, the more likely it is to have registration failures. Because of the propagation of shift vectors, the neighbour-reference registration method is unable to identify and correct these alignment errors in future registration steps, resulting in an even bigger misalignment in addition to the inherent registration error for long focal series.

5.4.2 Simulation-assisted XCF registration

To overcome the shortcomings of the existing registration methods when dealing with low-dose focal series, a simulation-assisted registration method has been developed with dedicated efforts made to improve the input image contrast and to integrate a simulation reference series into the registration workflow.

Noise reduction through motif-averaging In the lowest dose image series, the signal-to-noise ratio can be so low that features in the raw images are obscured and swamped by random noise. More importantly, the intensity of the maxima in the XCF map of such series are often inadequate to identify the correct registration vector. Denoising these images with the method outlined in Section 5.3 before registration becomes even more crucial. However, little can be done to the random background noise in the images directly. Therefore a compromise of using prior knowledge by considering that the samples has a distinctive periodicity, has to be made in order to suppress the noise in the image series before registration. By selecting and averaging repetitive motifs in a single image, the crystalline structure is amplified and the random background noise is reduced. The key step for this noise reduction to work is to correctly select the repeating motif. This is achieved by utilising the auto-correlation function (ACF) of the image, which is the XCF between the image and itself (Eq. 5.1).

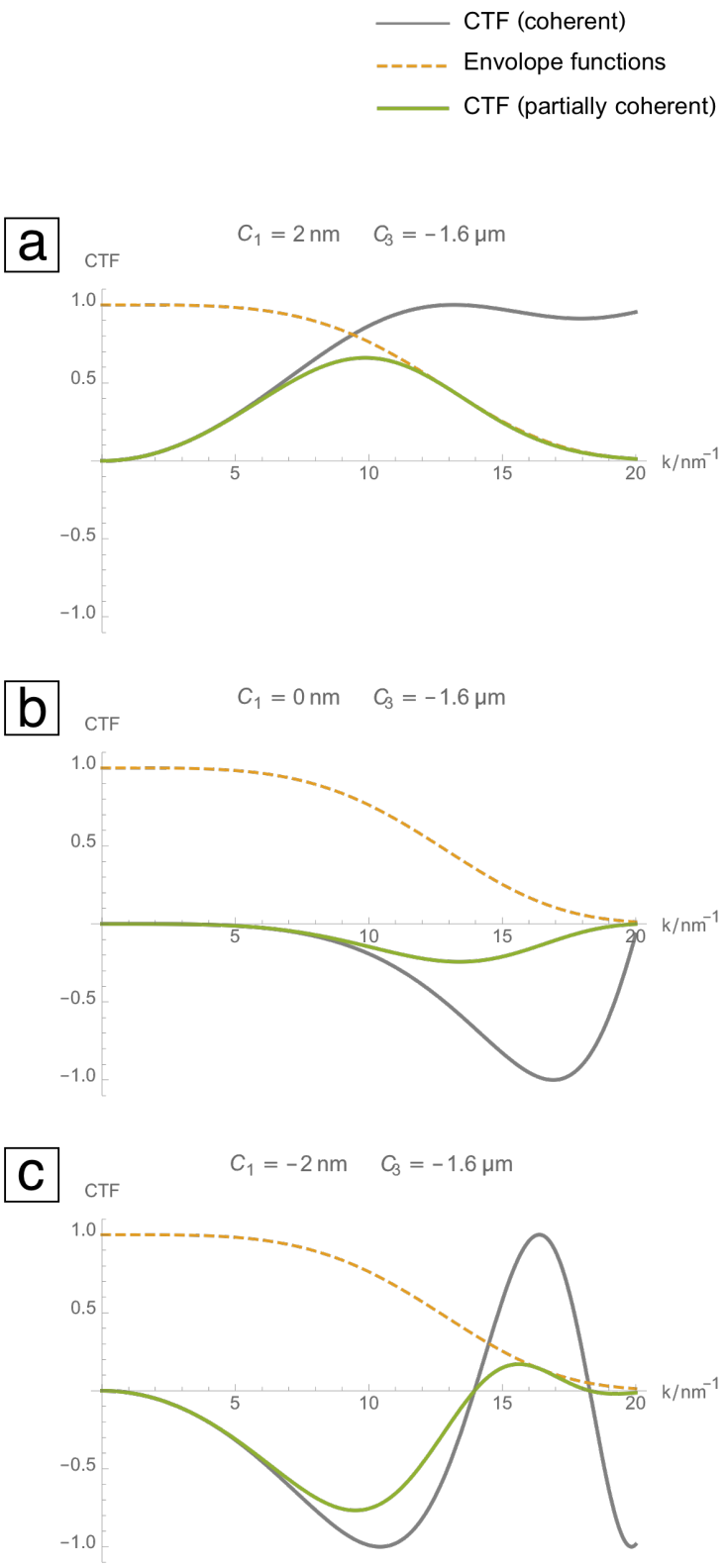


Figure 5.13: Changes in the contrast transfer function close to zero defocus with accelerating voltage $V_{acc} = 200\,\text{kV}$, focal spread $f_s = 2.8\,\text{nm}$, spherical aberration $C_3 = -1.6\,\mu\text{m}$. The defocus values are (a) 2 nm, (b) 0 nm, (c) −2 nm respectively.

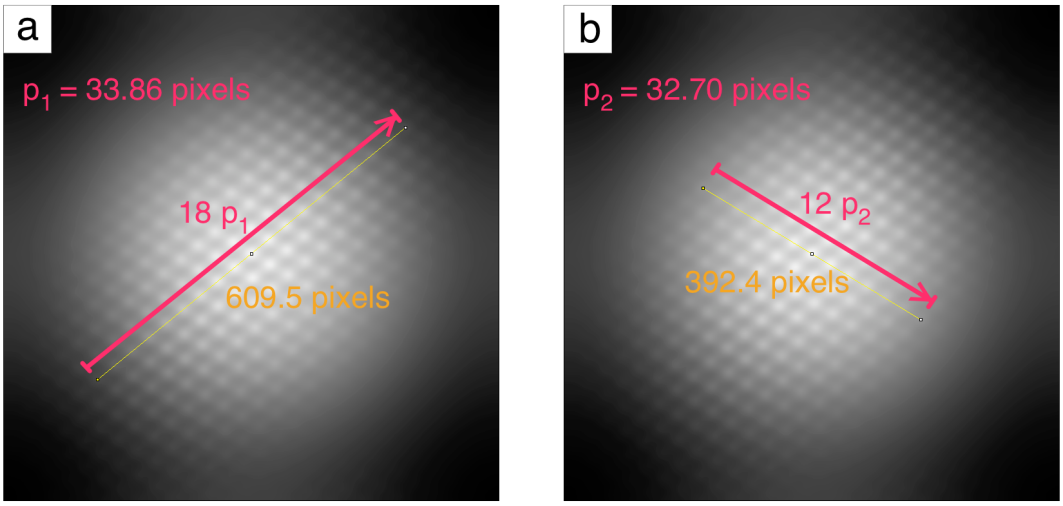


Figure 5.14: Motif determination using the auto-correlation function. (a) and (b) measure two independent vectors (in pixels), $\mathbf{p_1}$ and $\mathbf{p_2}$, representing the periodicity of the motif.

$$ACF[f(\mathbf{x})] = XCF[f(\mathbf{x}), f(\mathbf{x})] \tag{5.1}$$

The motif detection is based on the ACF property that repetitive features generate correlation peaks in the ACF map, in which the periods of repetitive features are equal to the distances between adjacent peaks along the corresponding directions. Two non-parallel vectors, $\mathbf{p_1}$ and $\mathbf{p_2}$, are needed to describe the periodicity of the motif. Similar to the treatment for the sampling interval measurement, the measurement accuracy of the vectors $\mathbf{p_1}$ and $\mathbf{p_2}$ can be improved by measuring multiple periods and dividing the length by the number of periods (Fig. 5.14).

Using an image from the CeO_2 data as an example, the two motif vectors, $\mathbf{p_1}$ and $\mathbf{p_2}$, measured from the ACF shown in Fig. 5.14, are respectively;

$$|\mathbf{p_1}| = \frac{609.5}{18} = 33.86\,(pixels) \tag{5.2}$$

$$|\mathbf{p_2}| = \frac{392.4}{12} = 32.70\,(pixels) \tag{5.3}$$

In the CeO_2 focal series, the nanoparticle studied only occupies the central part of the field of view, with the rest of the image containing either vacuum, or other particles which, if included, will diminish the improvement in SNR of averaging multiple motif images. These unwanted areas are excluded from motif selection by applying additional rules to the motif selection process. As illustrated in Fig. 5.15, the shift vectors $\mathbf{a_1}$ and $\mathbf{a_2}$ are chosen such that each of them corresponds to an edge of the particle. $\mathbf{a_1}$ and $\mathbf{a_2}$ are linear combinations of the previously determined $\mathbf{p_1}$ and $\mathbf{p_2}$ as;

$$\mathbf{a_1} = \mathbf{p_2} - \mathbf{p_1} \tag{5.4}$$

$$\mathbf{a_2} = \mathbf{p_1} \tag{5.5}$$

The algorithm leads the selection box to move along $\mathbf{a_1}$ until reaching the particle edge at the other end before moving one step of $\mathbf{a_2}$ and repeating the previous motion. At each box position, the image clip within the box is added to the average motif. By setting the number of steps in each direction, natural boundaries are set for the search region. Additionally, extra boundaries are introduced by adding conditions to the coordinates of the selections. For example, the red dashed line in Fig. 5.15 is chosen to avoid motif selection that includes any part of the other particle in the upper right corner of the field of view. The white rectangle box marks the motif that is currently selected. A magnified view of the upper right corner in each sub images is shown in Fig. 5.15, and the other small image below it indicates the averaged motif of all the motifs collected so far (marked by green rectangles). A significant increase in the image contrast can be seen in the averaged motif while more individual motifs are added into the averaged motif. This enhancement in contrast and SNR is beneficial for the later registration step.

The averaged motif contains all the structural features of the sample that are

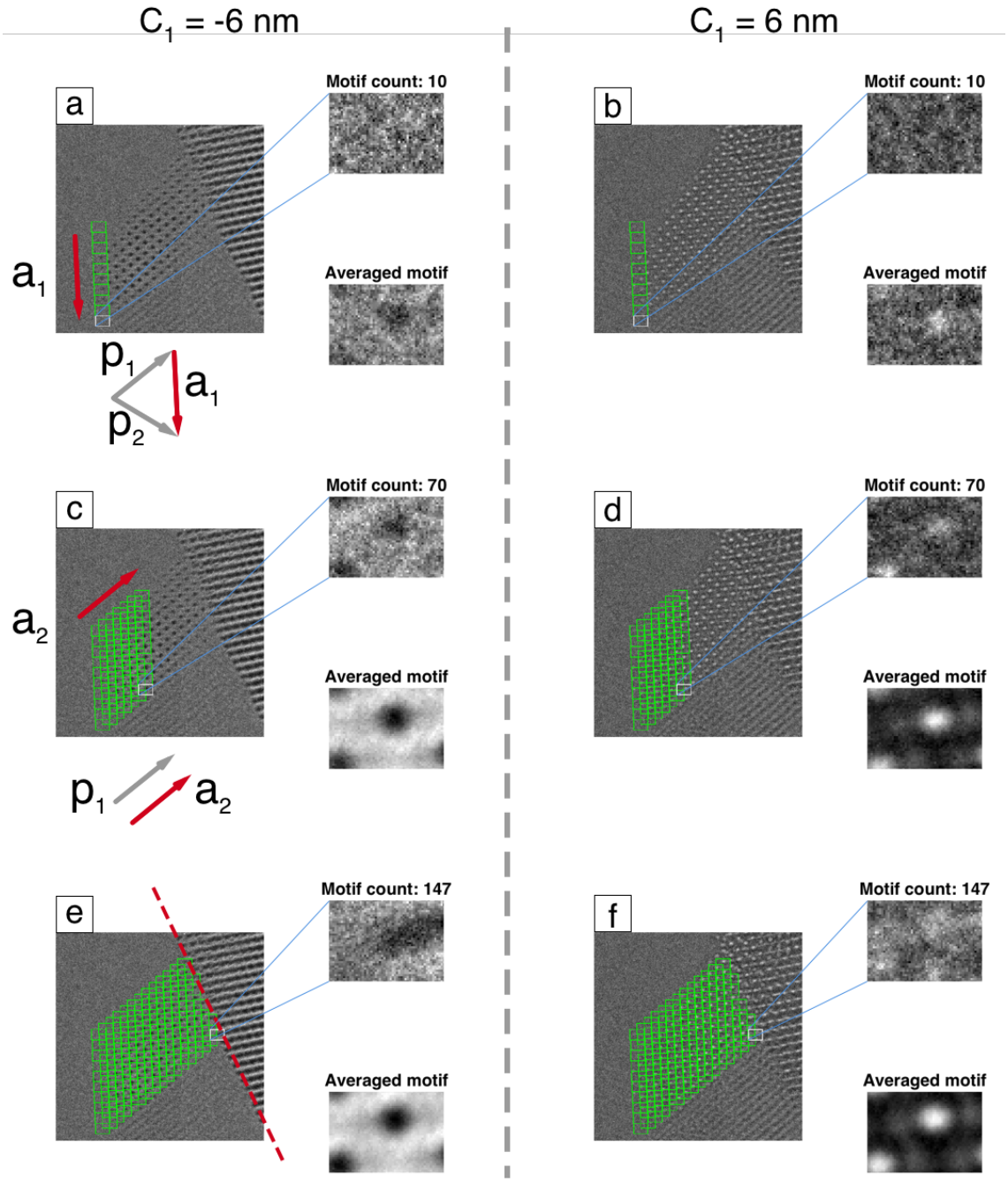


Figure 5.15: Motif selection and averaging in individual images of a focal series. The two enlarged clips from each sub-image are the selected motif and the averaged motif. Progressive improvement in the signal-to-noise ratio of the averaged motifs as the number of motif increases can be seen in both (a)(c)(e) the −6 nm image and (b)(d)(f) the 6 nm image. The number of the motif count indicates how many motifs have been used to obtain the average motif.

needed for the simulation-assisted registration. In practice, the averaged motif is extended by repeating along $\mathbf{a_1}$ and $\mathbf{a_2}$ to form larger images for the registration function to work (Fig. 5.16). This indirectly compensates for the damping effect of the Hanning window applied to the input images during the registration and ensures that enough structural information is included in the calculation of the cross-correlation function.

The image simulation was carried out with the software package MULTEM (Lobato and Van Dyck, 2015). Because the simulated images are perfectly aligned, registering individual experimental images to their simulated counterpart will also effectively align them with each other.

To demonstrate the improvement of using this model-based registration scheme over the conventional XCF method using the best contrast image as the reference, a pair of over/under focus images are registered with each other by the two schemes respectively (Fig. 5.17). Atoms appear dark in the under focus image but bright in the over focus image. The comparative result in Fig. 5.17(b,c) shows that the simulation-assisted method is able to correctly match the atoms in the under focus image to the corresponding atoms in the over focus image, while the direct XCF registration fails to do the same, due to the image contrast reversal and mistakenly registers the dark area in the over focus image to the dark atoms in the under focus image.

It is worth noting that the image motif is only the size of a single unit cell, hence the simulation-assisted registration is only suitable for correcting misalignment of less than half a unit cell length. Image shifts that are larger than half a unit cell but smaller than a whole unit cell are equivalent to a smaller shift in the opposite direction. An image shift that is larger than one unit cell is equivalent to it being subtracted by a integer number of period length. In fact, it is helpful to think of this simulation-assisted registration as a fine tuning step after a coarse registration by, for example, the neighbour-reference XCF.

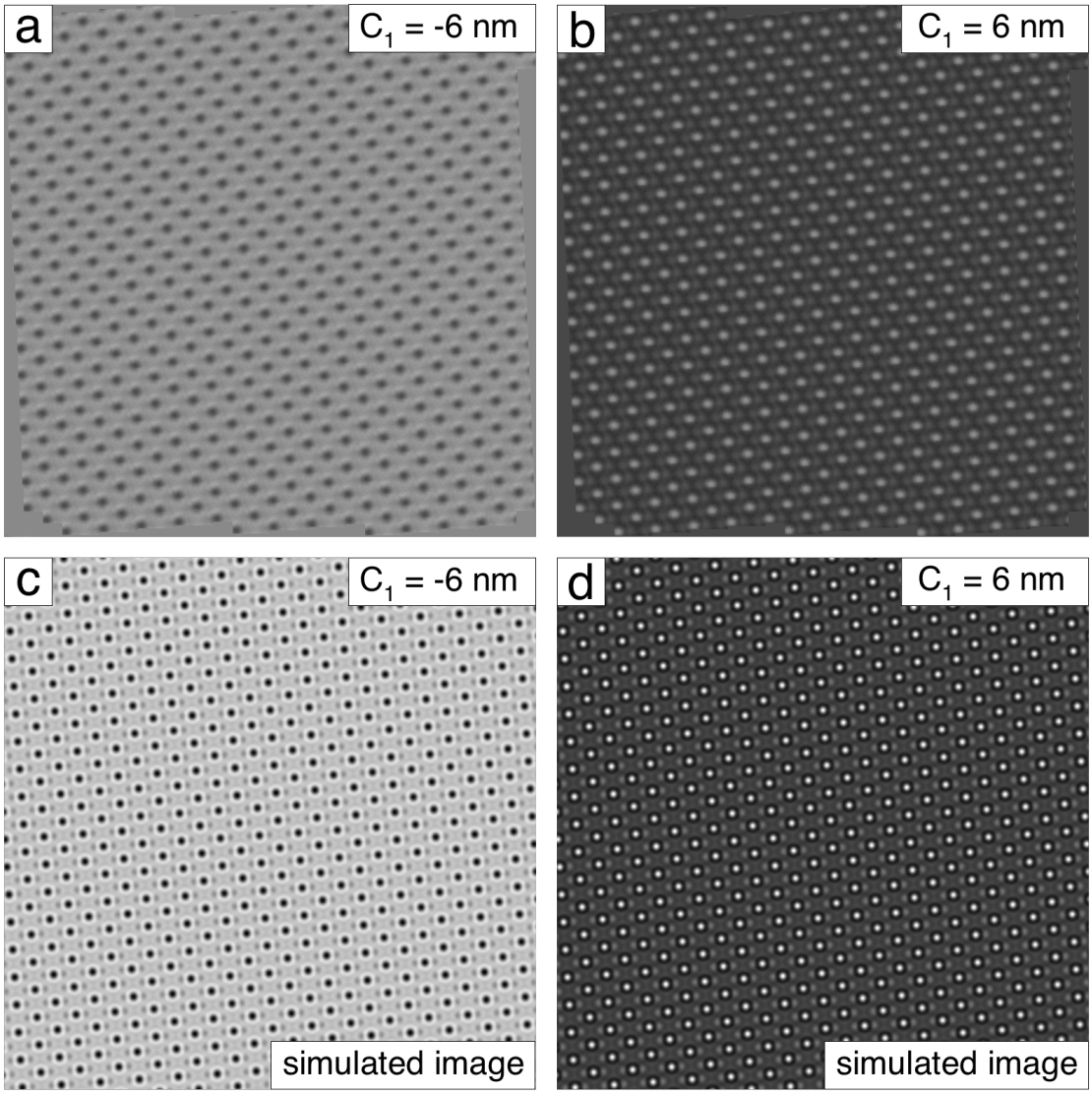


Figure 5.16: Extended images ((a) −6 nm defocus, (b) 6 nm defocus) obtained by repeating the averaged motifs from the motif selection process in Fig. 5.15 and (c,d) the corresponding simulated images.

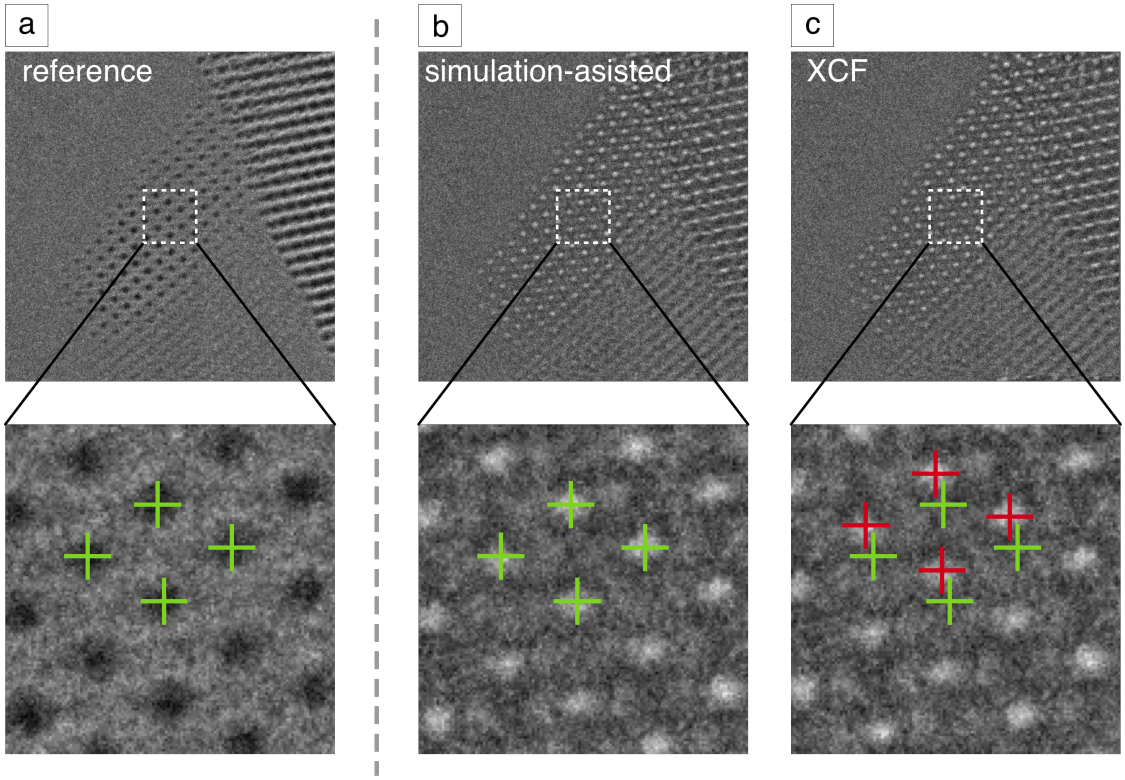


Figure 5.17: An example comparison between the registration results using the simulation-assisted registration and the simple XCF registration performed on two images from focal series at $1940\,e/\text{Å}^2$. Green crosses are at the same positions in different images. (a) and (b) show that the simulation-assisted registration can overcome the contrast reversal problem, while in (c) the atom position marked by red crosses are shifted away from their correct positions.

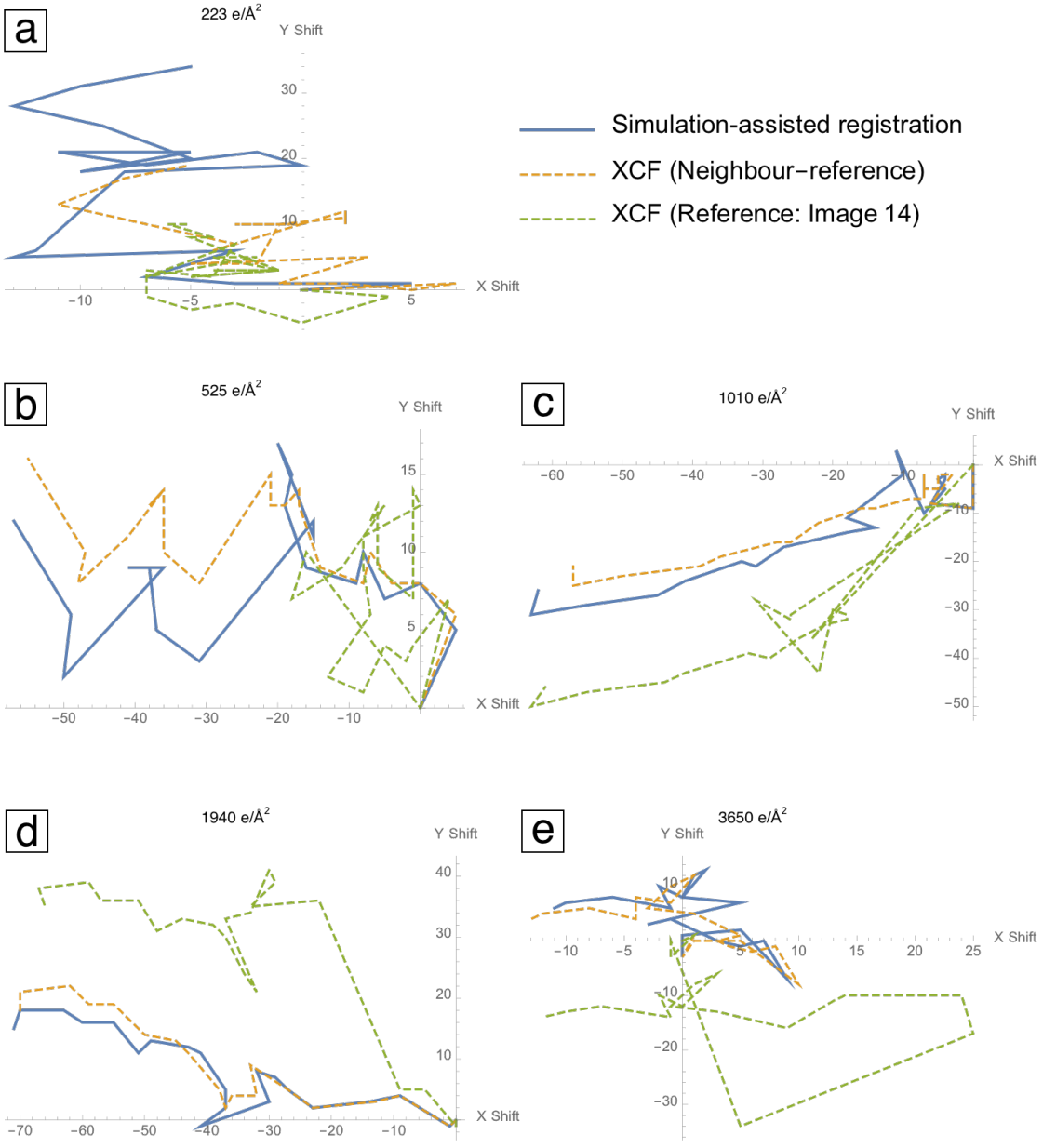


Figure 5.18: Comparison of registration results for CeO_2 by simulation-assisted registration (blue line), XCF with neighbour-reference (orange dashed line), and XCF using Image No.14 as reference (green dashed line). The title of each plot indicates the electron dose of a single image in the series.

The simulation-assisted registration returns very satisfactory registration results. The misalignment is imperceptible by human vision, which is capable of detecting any misalignments larger than 1 to 2 pixels with the help of markers[2]. Therefore, it is reasonable to use the simulation-assisted method result as the benchmark and consider the difference between the it and another registration as the error in the latter. As can be seen in Fig. 5.18, for all five focal series, the traditional XCF registration with a fixed reference image (green dashed lines) left a large amount of misalignment uncorrected, regardless of the dose level. In particular, big displacement jumps in the simple XCF results occur when the reference image and the subject image are on different sides of zero defocus, causing failures in determining the correct image shifts. The neighbour-reference XCF registration (orange dashed lines) show similar registration vectors to the ones determined from simulation-assisted registration in the high dose series, with the image shift difference limited to approximately 10 pixels. This agrees with our previous discussion that the simulation-assisted registration is best used as a fine tuning process following the neighbour-reference XCF registration. However, for the ultimate low-dose series (Fig. 5.18(a,b)), the difference in the determined image shifts increased significantly, which is largely due to the high noise level in the images.

5.5 Exit wave reconstruction

Established exit wave reconstruction methods require pre-aligned image series as the input. Therefore, it is only now, after overcoming the registration problem, that the reconstruction of low-dose focal series becomes possible. This section will first demonstrate how to apply existing reconstruction methods, previously reviewed in Section 3.3, to retrieve the exit waves from the low-dose focal series. In addition,

[2]For some extreme noisy and low-contrast images, the accuracy of visual inspection may reduce to 3 to 4 pixels

the effect of different dose on the resultant exit wave reconstruction will be explored and discussed. Lastly, the registration problem will be revisited, demonstrating the importance of good image registration for successful exit wave reconstruction.

5.5.1 CeO_2

The variable dose CeO_2 focal series data offers a great opportunity to study the dose effect in exit wave reconstruction. Mathematically, the wave optical description of the reconstruction process does not include electron dose. However, it is unrealistic to expect retrieval of the exit wave with a very small number of electrons, as the wave property of electrons would only be observable when there are enough of them, in very similar manner of the famous double-slit experiment, which demonstrates wave-particle duality. A more practical difficulty of exit wave retrieval from low-dose focal series is noise. How the noise influences the quality of the reconstructed exit wave has not been studied.

To answer these questions, the registered CeO_2 focal series are used as input for both the Wiener filter reconstruction approach used by the FTSR code and the maximum-likelihood approach implemented in SEMPER. Both reconstructions were provided with the same imaging conditions to ensure comparability.

The reconstructed exit waves of the CeO_2 nanoparticle are shown in Fig. 5.19 (Wiener-filter method) and Fig. 5.20 (MAL method), in the form of amplitude and phase. As expected for an approximate weak phase object, the phase of the reconstructed complex waves contain more information than the amplitude. Overall, the two methods demonstrate good general agreement in reconstruction, although subtle differences do exist, which will be discussed in the subsequent sections.

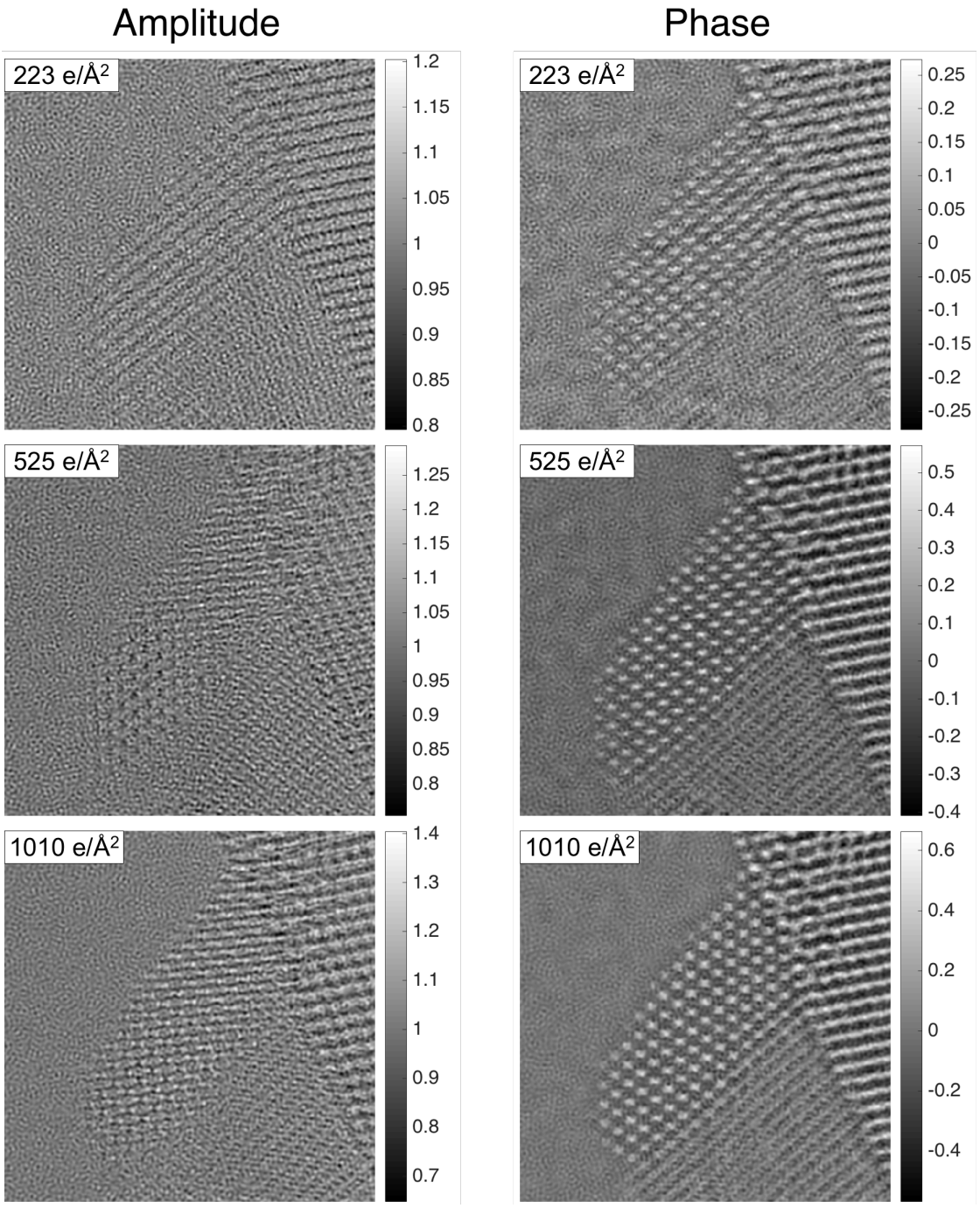


Figure 5.19: Reconstructed exit waves (phase unit: radian) from cerium dioxide nanoparticle focal series using the Wiener filter method in FTSR at various doses (the imaging dose of a single image of each series is indicated).

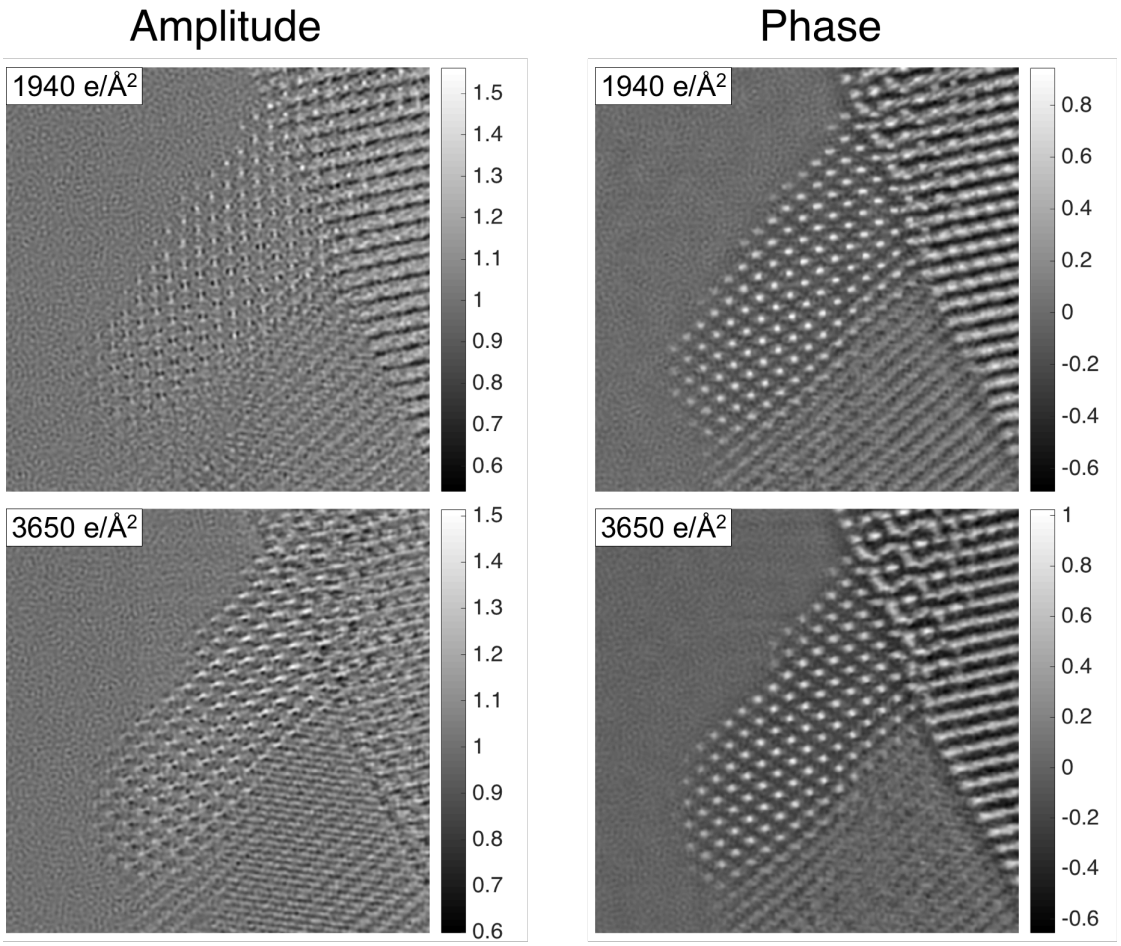


Figure 5.19: (*continued*) Reconstructed exit waves (phase unit: radian) from cerium dioxide nanoparticle focal series using the Wiener filter method in FTSR at various doses (the imaging dose of a single image of each series is indicated).

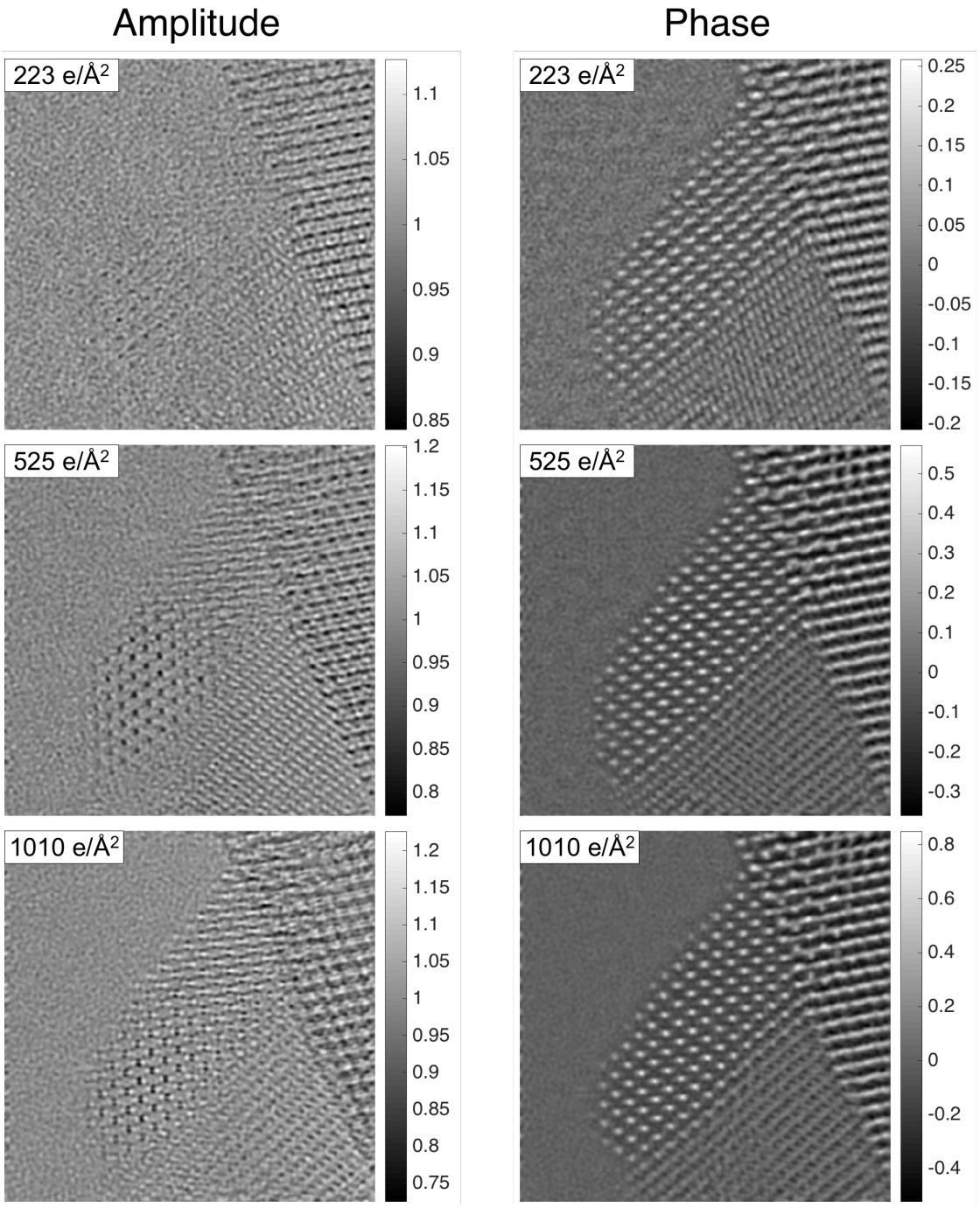


Figure 5.20: Reconstructed exit waves (phase unit: radian) from cerium dioxide nanoparticle focal series using the maximum-likelihood method in SEMPER at various doses (the imaging dose of a single image of each series is indicated).

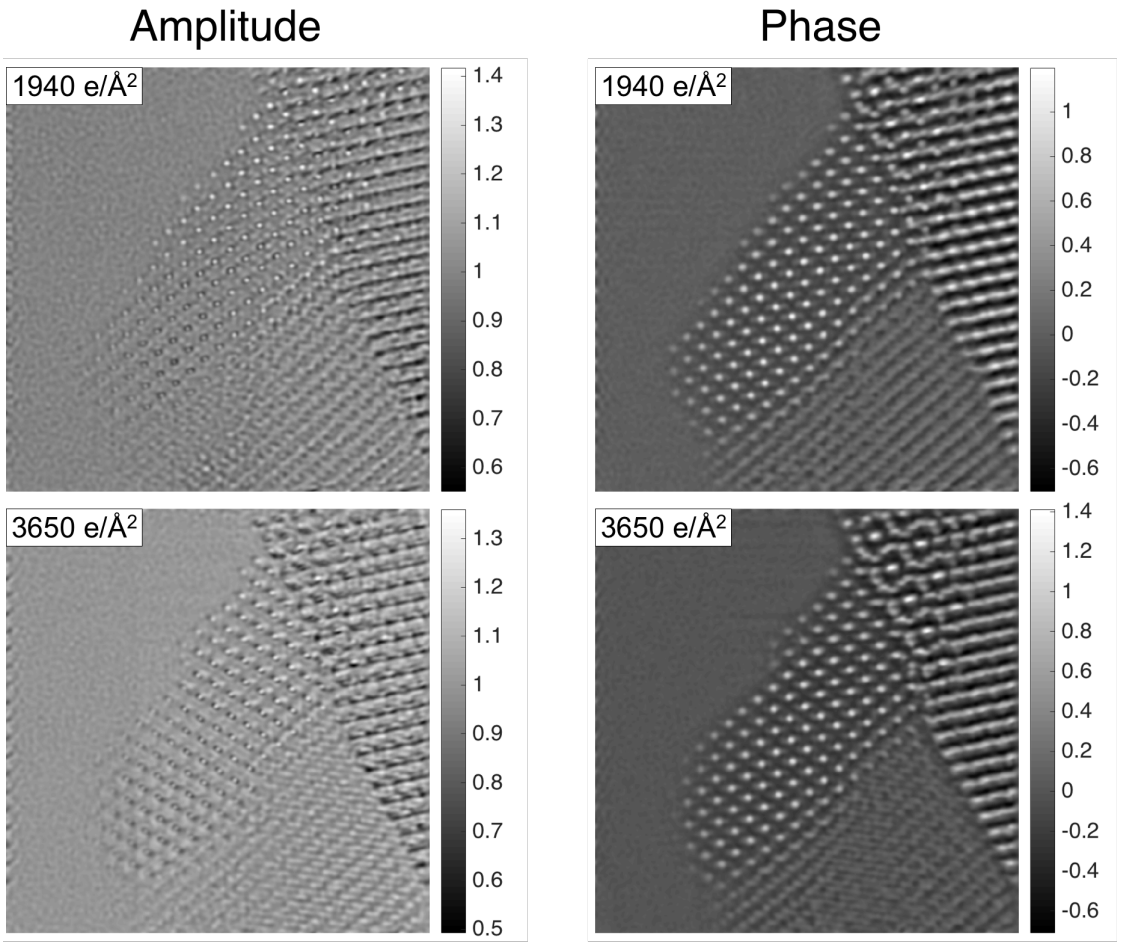


Figure 5.20: (*continued*) Reconstructed exit waves (phase unit: radian) from cerium dioxide nanoparticle focal series using the maximum-likelihood method in SEMPER at various doses (the imaging dose of a single image of each series is indicated).

5.5.2 Comparison of the reconstruction results

5.5.2.1 Wiener-filter method vs. MAL

For the highest two doses ($3650\,e/\text{Å}^2$ and $1940\,e/\text{Å}^2$), the restored wave phases are similar and agree with each other, judged by direct visual inspection (Fig. 5.19 and Fig. 5.20). The phases of the lower dose series are visibly more noisy and the sample structural features also become blurred as the dose decreases. Furthermore, the range of the phase shifts are smaller at lower doses than the two high dose series, as indicated by the colour bar of each phase image. The phase shift of an electron wave should be a function only determined by the electron-sample interaction while being independent of the dose. Therefore this observed *diminished phase phenomenon* can only be explained from the point of view that the wave phase recovery process is affected by the limited amount of signal information conveyed by the electrons, and by the large proportion of image noise that corrupts the phase information. Because the electron wave function describes the probability of an electron landing position on the detector, to restore the wave one needs enough electrons to be statistically representative. The main source of noise that affects low-dose images is in the detecting process. This noise is unrelated to the electron signal and generates a randomised phase shift that partly cancels out the actual phase shift.

While quantitative interpretation of restored exit waves is still an ongoing topic in electron microscopy, it is important to ensure consistent reconstruction results and to avoid any uncertainty induced by imaging dose. The presented comparison result of the CeO_2 exit wave reconstruction suggests that there is a certain minimum dose, only above which the phase shift of the restored wave is reliable. For waves that are reconstructed from lower dose data, the "wave" is useful only in a qualitative sense and provides an enhanced signal-to-noise ratio sample representation compared to the individual images in the series.

The IQ factor analysis introduced in Chapter 4 has been used to quantitatively compare the two reconstruction methods at different doses. Since the IQ factor analysis can only be applied to images of single crystalline structures, all phase data were masked to block unwanted extra regions of the sample and vacuum, leaving only the investigated particle included in the analysis. The edges of the applied mask were softened by a mean filter to suppress aliasing effects during the Fourier transform process (Fig. 5.21 and Fig. 5.22).

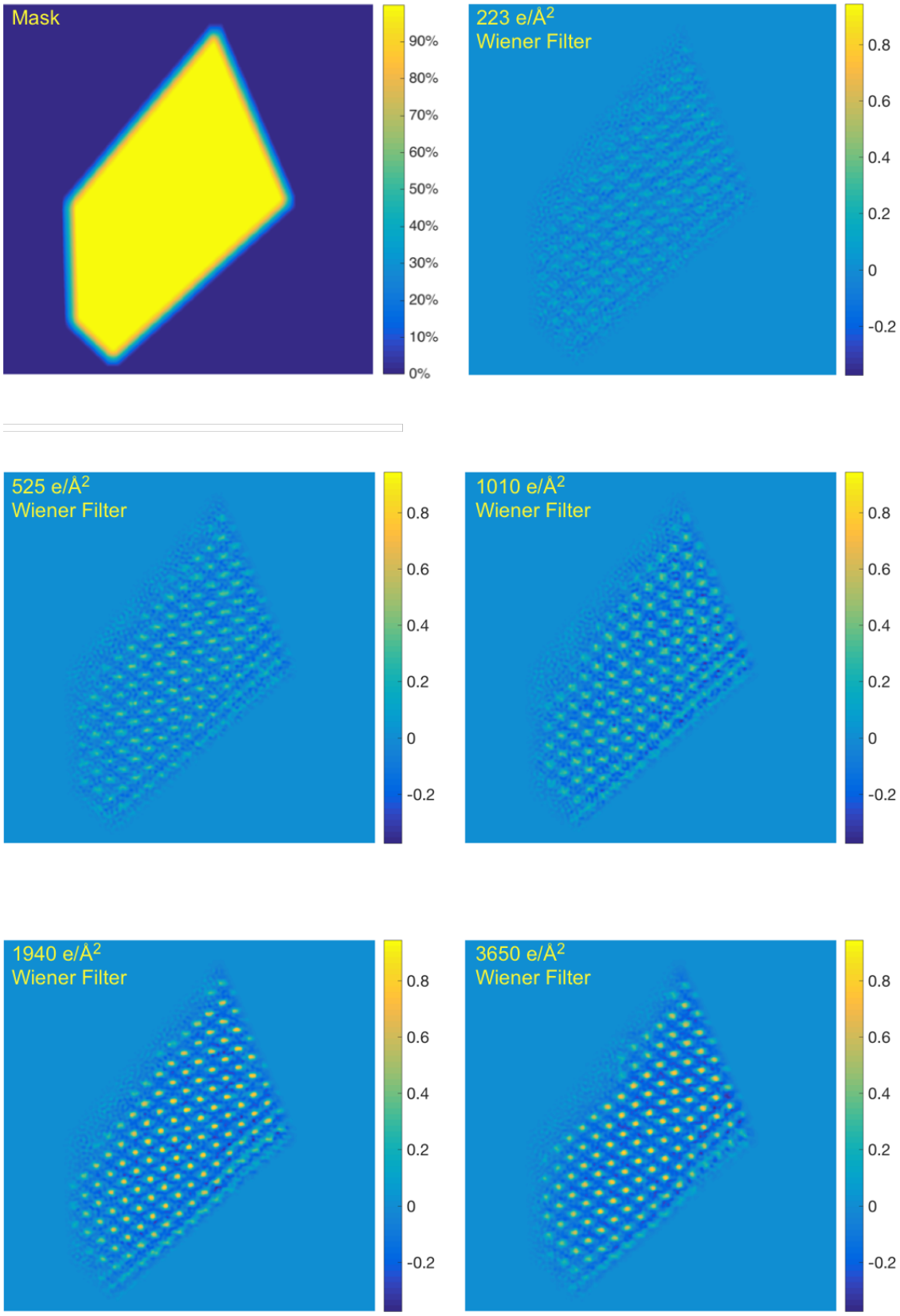


Figure 5.21: The mask applied to the exit wave phases and the masked restored phases (unit: radian) of CeO_2 exit waves reconstructed using the linear Wiener filter method at various doses (the imaging dose of a single image of each series is indicated).

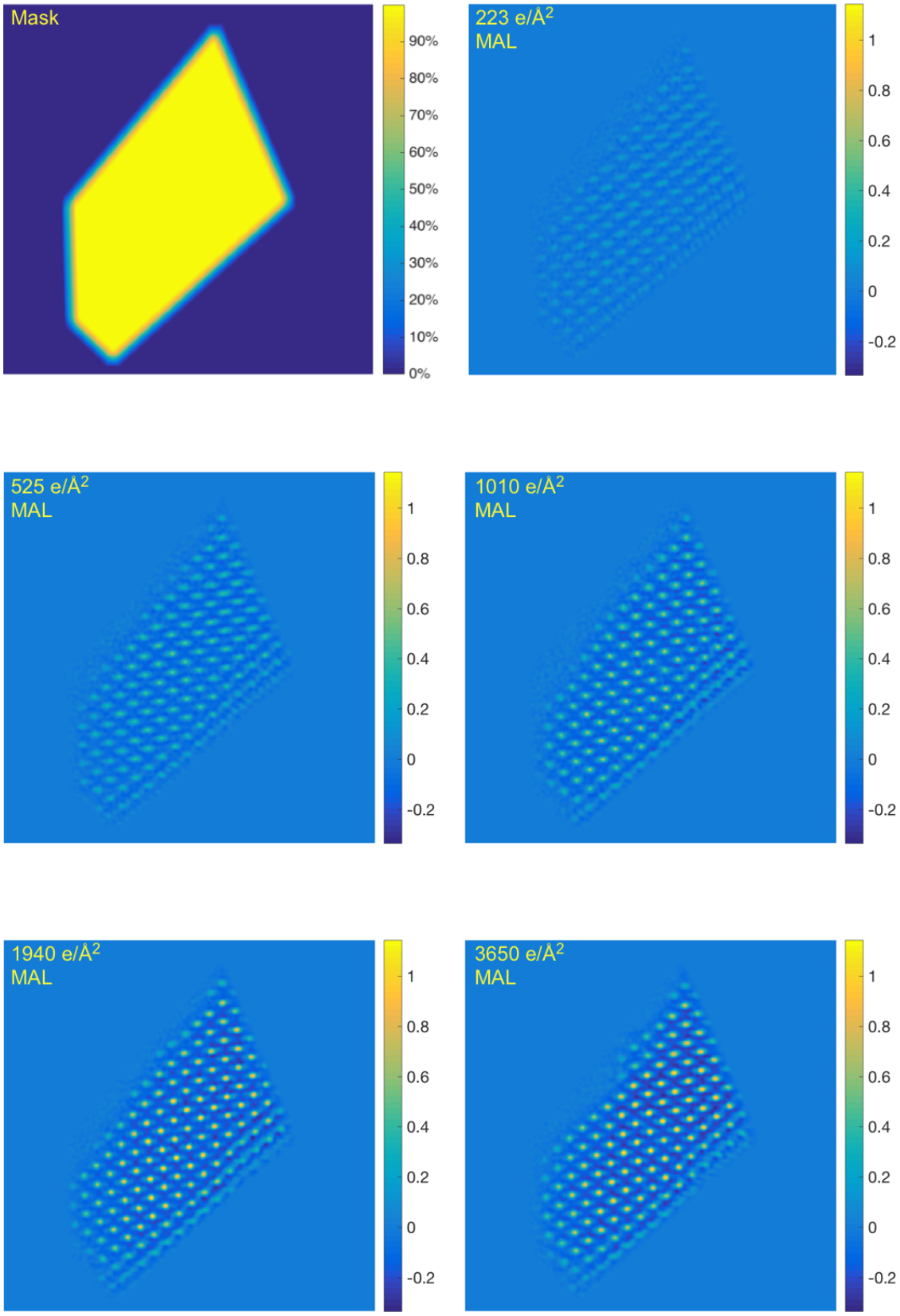


Figure 5.22: The mask applied to the exit wave phases and the masked restored phases (unit: radian) of CeO_2 exit waves reconstructed using the maximum-likelihood method at various doses (the imaging dose of a single image of each series is indicated).

The IQ factor was calculated using the selected reflections (marked by red dots in Fig. 5.23). The results confirm the proposed explanation of the effects of dose on exit wave reconstruction. Restored phases from both the linear Wiener filter method and the maximum likelihood method increase in IQ factor as the imaging dose is increased, up to a dose of $1940\,e/\text{Å}^2$. The highest dose series, at $3650\,e/\text{Å}^2$, gives an almost identical IQ factor to the $1940\,e/\text{Å}^2$ series.

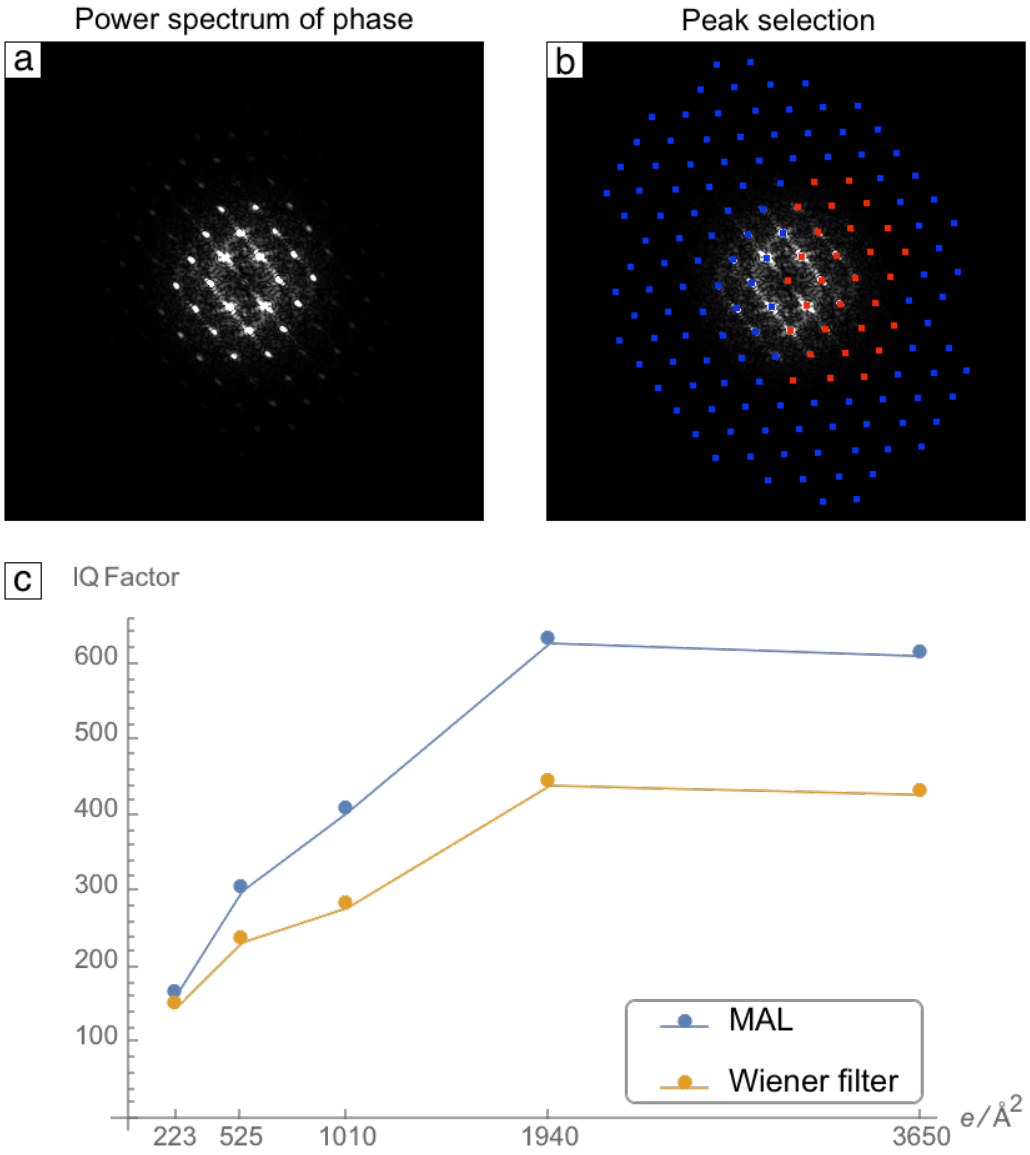


Figure 5.23: IQ factor comparison of the exit wave phase reconstructed using FTSR and SEMPER. (a) The power spectrum of the exit wave phase; (b) Automated peak selection performed on (a), with peaks marked by red dots being the ones used in the IQ factor calculation.

5.5.3 Exit wave reconstruction comparison with different registration methods

In Section 5.4.2, the differences between registration results from the newly developed simulation-assisted registration and the two XCF-based registration methods, one

using a fixed reference image and the other using a neighbour-reference scheme, were described (Fig. 5.18). Again by using the IQ factor analysis, the registration effects on exit wave reconstruction can now be investigated quantitatively.

The maximum-likelihood method was used exclusively for reconstruction to eliminate the influence of different reconstruction methods. The restored phases of the three registration methods at all five doses are shown side-by-side in Fig. 5.24, and the corresponding calculated IQ factors are plotted in Fig. 5.25.

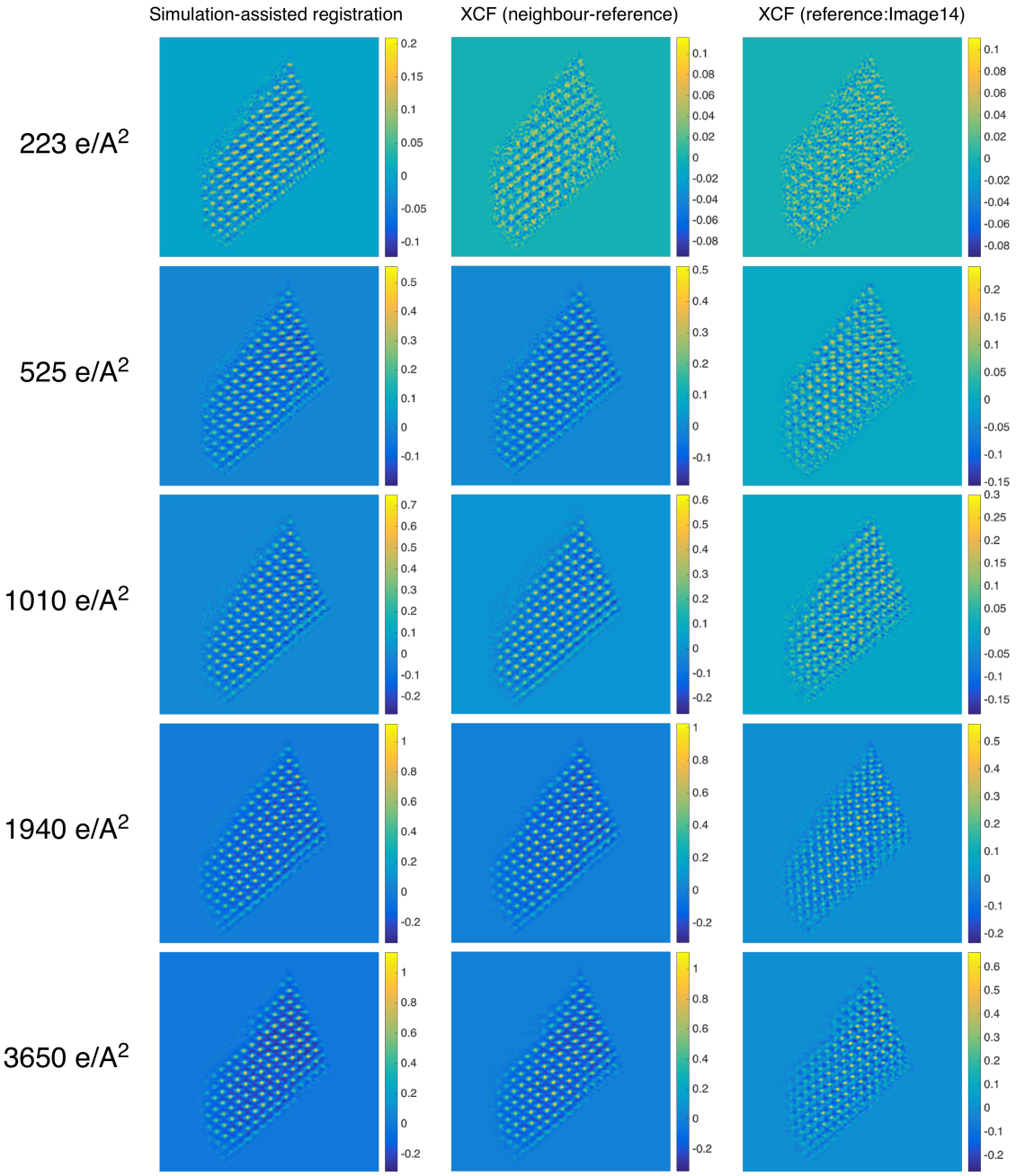


Figure 5.24: Phases (unit: radian) reconstructed by three registration methods at various doses (the imaging dose of a single image of each series is indicated).

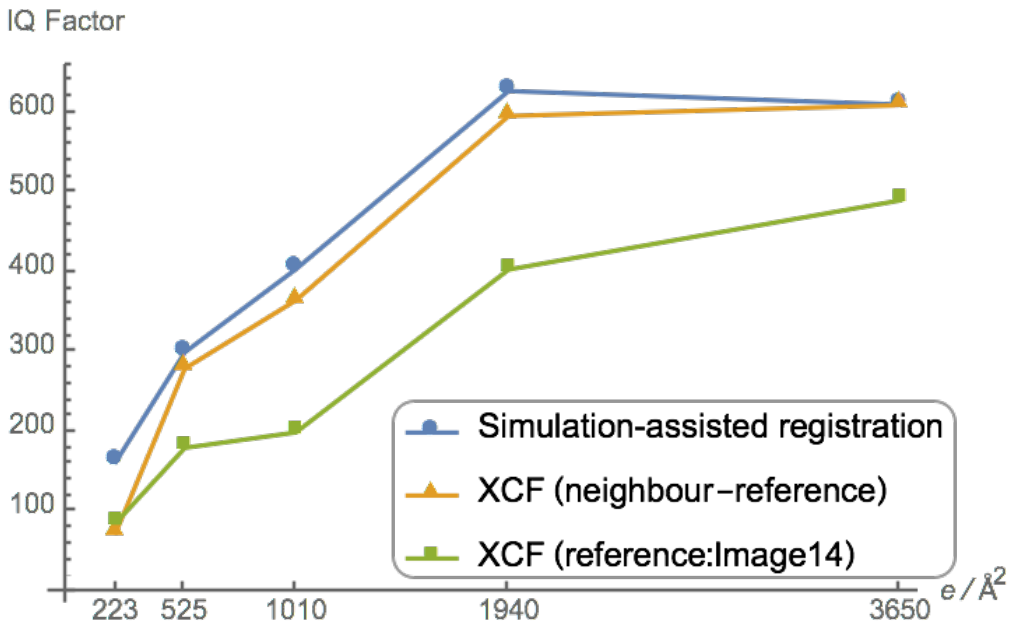


Figure 5.25: IQ factor comparison between restored phases with registration results from different methods at various doses (the imaging dose of a single image of each series is indicated).

It is shown that the IQ factor calculated for data registered by the simulation-assisted registration method is consistently higher compared to the other two methods, although this difference is less at high dose.

It should also be noticed in Fig. 5.24 shows that better registration also gives a larger reconstructed phase shift range, which implies that, if there is a dose threshold above which the phase shift range is consistent, good image registration would allow the reconstruction to be performed with focal series collected at a lower dose.

5.5.4 Si_3N_4

The exit wave reconstruction of three focal series of silicon nitride were also carried out using the maximum-likelihood method (Fig. 5.26). The restored wave phases confirm the dose effect on exit wave reconstruction, as was demonstrated by the

CeO_2 reconstruction. The two higher dose series, taken at $1889\,e/\text{Å}^2$ and $117\,e/\text{Å}^2$, show close agreement, with the phase shift ranges being approximately within -0.6 and 1.4. The Si N atom pair, although not fully resolved everywhere, can be seen at some local positions. For the lower dose focal series ($31\,e/\text{Å}^2$ and $16\,e/\text{Å}^2$), the reconstruction suffers from the high level of noise and cannot reproduce the phases within the same phase shift range.

5.5.4.1 An application limitation of the IQ factor

It was expected that the influence of dose and registration on the reconstruction of the Si_3N_4 data would be similar to the analysis results of the CeO_2 data. Unfortunately, attempts to perform the IQ factor analysis were unsuccessful for the Si_3N_4 exit waves. It was realised that the reflection peaks are too close to each other in the power spectra due to the use of very high magnification during imaging such that it is unable to obtain the peak background patches without them including other peaks, as shown in Fig. 5.27. Therefore, no IQ factor analysis was done with the restored Si_3N_4 phases.

5.5.5 Conclusions

There are at least three elements that are clearly linked to the quality of the exit wave reconstruction from low-dose focal series: dose, image registration, and noise. These elements are not isolated factors, instead they are closely intertwined as low-dose images are noisy and hence are harder for image registration. However, each one of them must also have their own effect. It would be unrealistic to expect a perfectly aligned focal series without any noise but sparsely formed from only a few electrons to fully recover a good exit wave. A small misalignment causes the phase shift of a sharp peak to spread into a reduced and blurred one and a big misalignment may even overlap a positive phase shift to a negative phase shift which cancel each other. Image noise can be seen as a random signal that generate random phase shifts, which

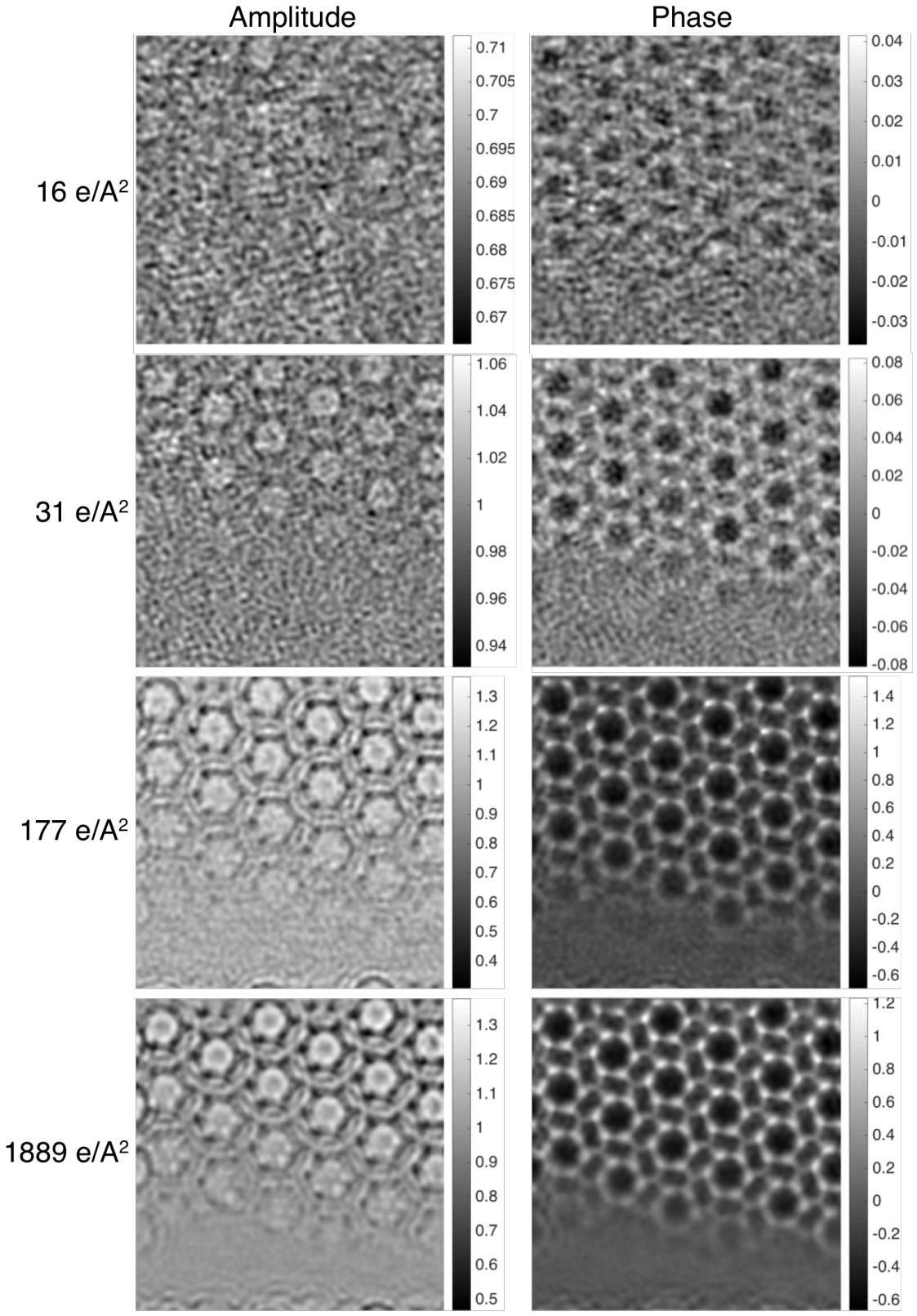


Figure 5.26: Reconstructed exit waves (phase unit: radian) from silicon nitride focal series data using the maximum-likelihood method in SEMPER at various doses (the imaging dose of a single image of each series is indicated).

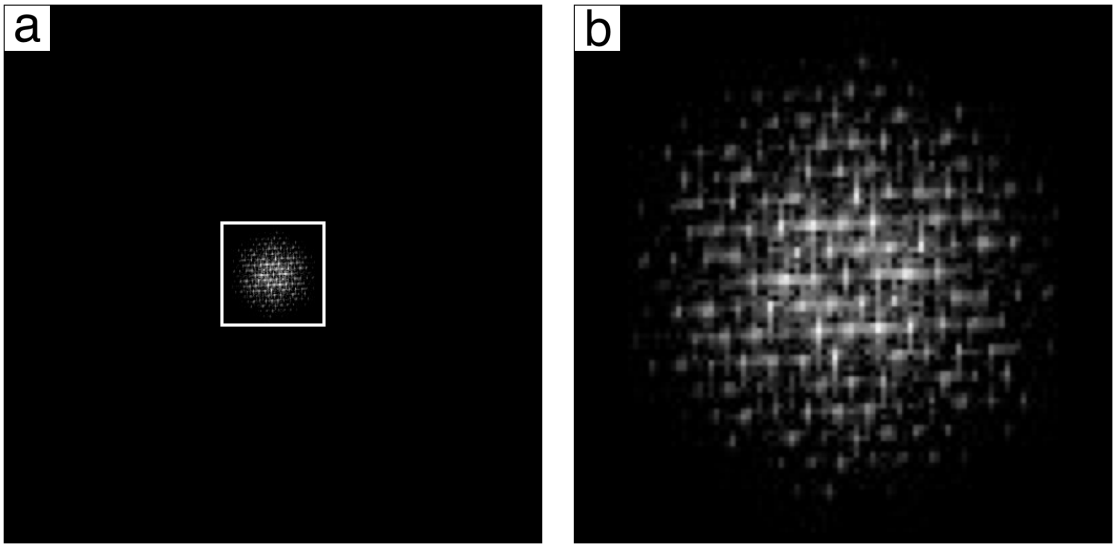


Figure 5.27: (a) Power spectrum of the restored phase (dose: 1889 $e/\text{Å}^2$) in Fig. 5.26. (b) is the enlarged central part of (a), marked by the white box, showing the close separation of individual reflections.

statistically reduce the true phase shift in the reconstruction. The reduction in phase shift range in the experiments in this chapter is a direct reflection of this phenomenon.

Chapter 6

Applications to Radiation-sensitive Materials

6.1 Introduction

As the ultimate aim of this thesis is to apply the techniques described in the previous chapters - low-dose imaging, improved image registration and exit wave reconstruction to radiation-sensitive materials, this chapter gives examples of how they can be used to characterise some very radiation-sensitive material samples, zeolites.

6.1.1 Zeolites

Zeolites are a class of microporous materials, which are both found naturally and synthesised industrially (Niwa et al., 2010). They are used in a wide range of applications, from solid acid catalysts to adsorbers of toxic components (Marakatti and Halgeri, 2015; Marakatti et al., 2014; Wang and Peng, 2010; Palomino et al., 2011). Zeolites are the most common heterogeneous catalysts for refinement of petrochemicals in the oil industry, due to their excellent catalytic performance and cost-effectiveness, whilst also being environmentally friendly (Choudary and Newalkar, 2011). There has been

an increasing use of zeolites in NO_x removal in the car industry (Beale et al., 2015).

Because of their radiation-sensitivity, high resolution TEM imaging of zeolite materials was prohibitively difficult until advances in dose control and detector technology enabled a sufficient electron dose to be used to record a single image under low-dose conditions (Pan and Crozier, 1993). However, the noise level in these early images was generally very high such that the unit cell of the zeolite framework was only visible after averaging periodic motifs in the image. Some later reports were able to resolve the zeolite frameworks to various extents without the need for averaging (Terasaki et al., 1997; Thomas et al., 2001; Ohsuna et al., 2004), but the true determination of these complex structures often requires X-ray diffraction. It has been reported that some complex zeolite structures can now be determined from electron diffraction with the help of computers undertaking a large amount of computation (Nicolopoulos et al., 1995; Dorset and McCourt, 1997). If a Cs-corrected TEM is available, negative spherical aberration imaging (with a corresponding positive defocus) can be carried out for zeolite characterisation, which provides a more intuitive representation of the zeolite structure, where atomic columns are bright (Yoshida et al., 2013). The optimum accelerating voltage for zeolite imaging has also been explored by Yoshida and Sasaki (2013), who found that the electron radiation damage to zeolites was the least severe at 200 kV.

With all the works mentioned above, the existing literature has largely been focusing on single frame imaging of the zeolite materials, while focal/time series imaging of zeolite materials, although not non-existent (Pan, 1996), has never been achieved at the highest possible resolution compared to the single-exposure results. Furthermore, the exit wave reconstruction techinique has not been applied to the characterisation of zeolites.

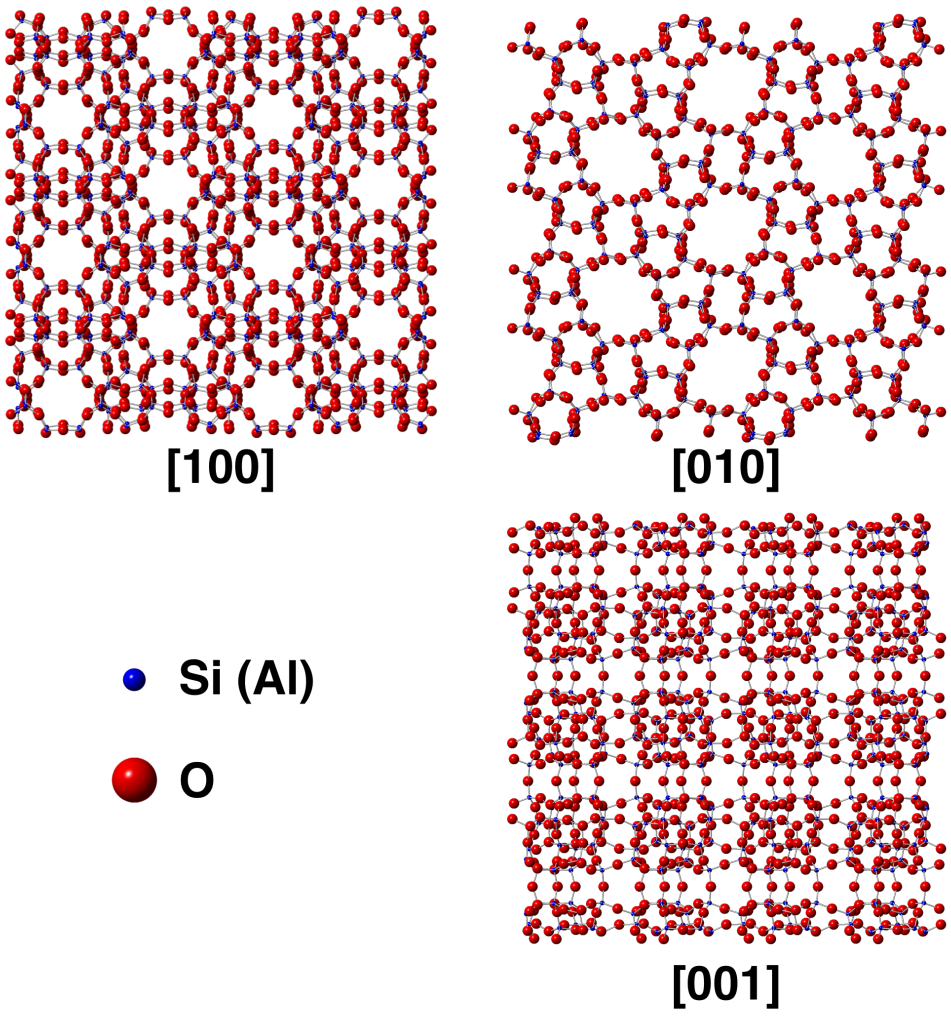


Figure 6.1: Atomic models of ZSM-5 projected along the zone axes [1 0 0], [0 1 0], and [0 0 1] as shown.

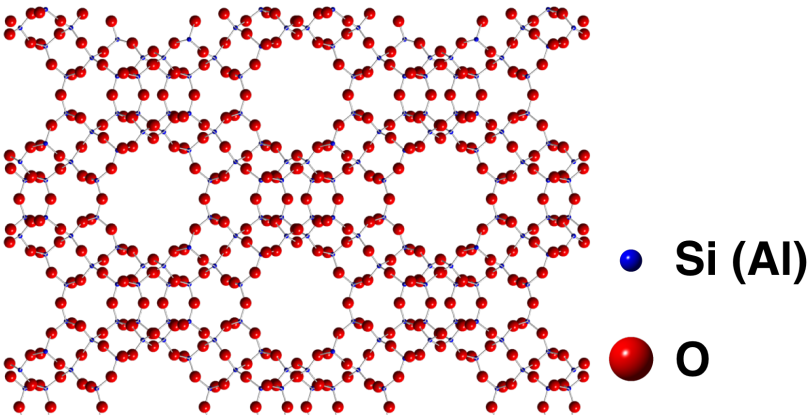


Figure 6.2: Atomic models of zeolite Y projected along the zone axes [1 1 0].

ZSM-5 ZSM-5 is an MFI-framework aluminium silicate zeolite, belonging to the space group Pnma, with cell parameters $a = 20.064$ Å, $b = 19.938$ Å, $c = 13.409$ Å (Grosse-Kunstleve et al., 1999). As illustrated in Fig. 6.1, ZSM-5 has a complex framework structure composed of SiO_4 tetrahedra, typically with high silicon-to-aluminium (Si/Al) ratios. The ZSM-5 sample studied in this work had a Si/Al ratio of 300, indicating that for every 300 silicate tetrahedra, there will be one tetrahedra with an aluminium atom substituting at the central site which is normally occupied by a silicon atom. Each Al^{3+} replacement of the Si^{4+} will leave one unpaired electron in the framework, resulting in a negative charge. The negative charge of the zeolite framework is compensated by cations attached to the framework; the type of the cations also differentiate the properties of the zeolite from other zeolite materials. Looking along [0 1 0] direction, the mesoporous structure contains large holes (10-membered rings), each surrounded by ten smaller rings (two 6-membered rings and eight 5-membered rings).

Zeolite Y Another type of zeolite, siliceous zeolite Y, was used for the analysis of STEM time series images. The sample is chemically similar to the ZSM-5 sample used in TEM imaging, except that the zeolite Y is siliceous and contains only silicon

tetrahedra. Zeolite Y belongs to space group $Fd\bar{3}m$ and has a cubic structure with the cell parameter $a = 24.345\,\text{Å}$. When looking along [1 1 0] direction, the framework structure of zeolite Y are composed of large 12-membered rings surrounded by smaller 6-membered and 4-membered rings (Fig. 6.2).

6.1.2 Sample preparation

The raw ZSM-5 sample came in the form of powder, with an average particle size of 1 µm estimated from low magnification images. These unprocessed particles are too thick for HRTEM imaging and therefore need to be crushed before deposition onto a TEM sample grid. During the crushing process, the mortar and pestle were cooled to liquid nitrogen temperature and the powders were submerged in liquid nitrogen, using the methods described in Section 4.2. Because the zeolite particles are more brittle at low temperature, thin areas are more easily found on the particle edges where fractures tend to happen.

The importance of having the specimen dry should never be neglected before TEM imaging of zeolitic materials (Ortalan et al., 2010; Klingstedt, 2011). It is believed that polar solvents such as water and ethanol can increase the radiation sensitivity, worsening the radiation damage, especially when radiolysis (ionisation damage) is the main damage mechanism. Specimens with residual solvent readily generate $H^{\cdot}$, $HO^{\cdot}$, or $R^{\cdot}$ radicals ($R^{\cdot}$ denotes residual organic molecules), which weaken the bonds and lead to secondary damage in the sample. Experimentally, drying may involve leaving the specimen in a pumped plasma cleaner or in the high vacuum microscope column, typically overnight.

6.2 Low-dose time series imaging and non-rigid image registration

6.2.1 Time series imaging

When characterising zeolites using the TEM, all non-imaging electron exposure must be kept to the minimum possible to avoid any unnecessary damage. The search for a suitable area on the specimen was carried out at low magnifications (between 10000× to 30000×), with a spread beam illumination and a dose rate less than $2\,e/(\text{Å}^2\,\text{s})$. An example low magnification image is shown in Fig. 6.3. Sample tilting and defocus adjustment were also carried out at a minimal dose rate of approximately $0.5\,e/(\text{Å}^2\,\text{s})$ as electron diffraction only needs a low flux parallel illumination. Dose rates at each chosen magnification and the corresponding condenser lens configurations were calibrated prior to the experiments using the calibration procedure previously described in Section 4.1. The time that the sample was exposed to the electron beam at each step was recorded; only calibrated settings were used to ensure that the electron exposure was always known and under control. The whole process results in approximately $150\,e/\text{Å}^2$ pre-acquisition exposure over a typical 5 min adjustment stage before the high resolution imaging took place. Because any small increase in the dose rate would have significantly raised the total dose over a long adjustment time, special attention was paid to carefully limiting the exposure dose rate.

After all microscope adjustments were completed, a low-dose time series of a ZSM-5 crystal was acquired with a constant dose rate of $25.3\,e/(\text{Å}^2\,\text{s})$. The series was composed of 60 consecutively recorded low-dose images, each image recorded with an electron dose of $76\,e/\text{Å}^2$ over a 3 s exposure period. The total electron dose is then $4710\,e/\text{Å}^2$, including $150\,e/\text{Å}^2$ from the adjustment stage before the final imaging. By comparison, the $150\,e/\text{Å}^2$ pre-acquisition electron exposure equals the dose of only two images, which is a small fraction of the total electron exposure. Table 6.1 gives

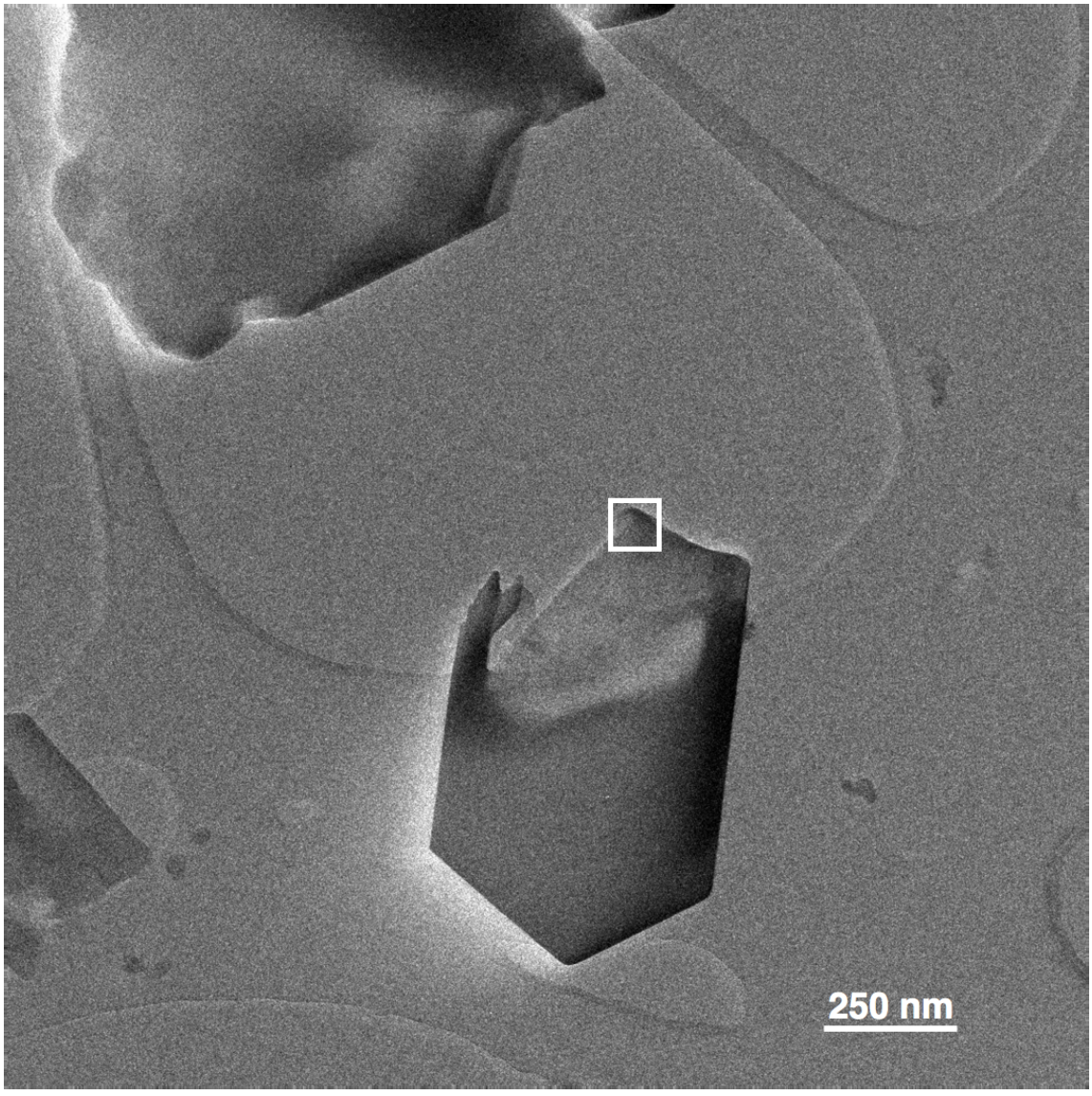


Figure 6.3: A low magnification (20000×) image of a ZSM-5 crystal recorded during the sample search stage. The white box indicates the region subsequently shown in Fig. 6.4.

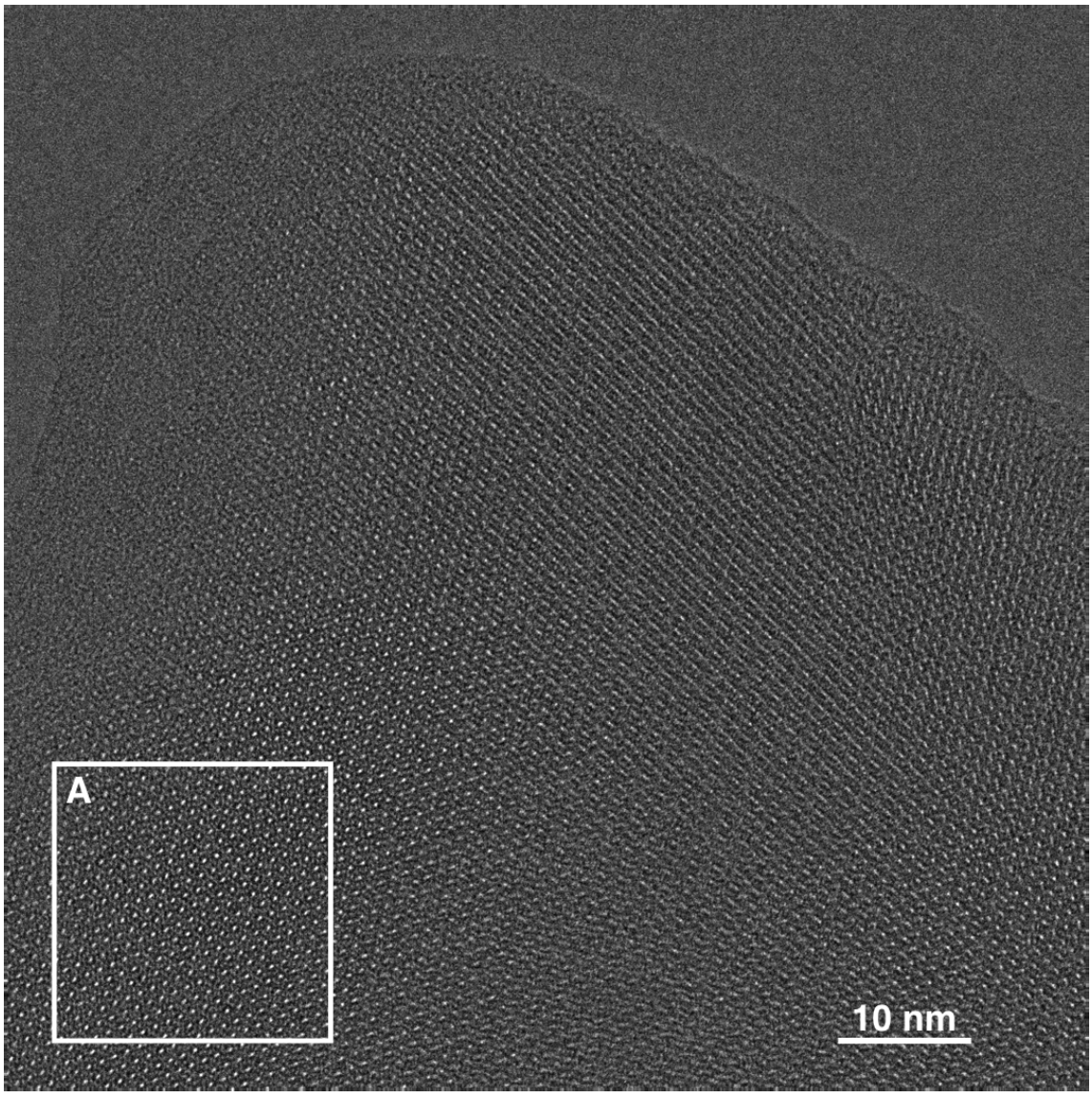


Figure 6.4: The first image of the ZSM-5 time series dataset, taken along the [0 1 0] zone axis. Box A marks regions used subsequently in Fig. 6.5.

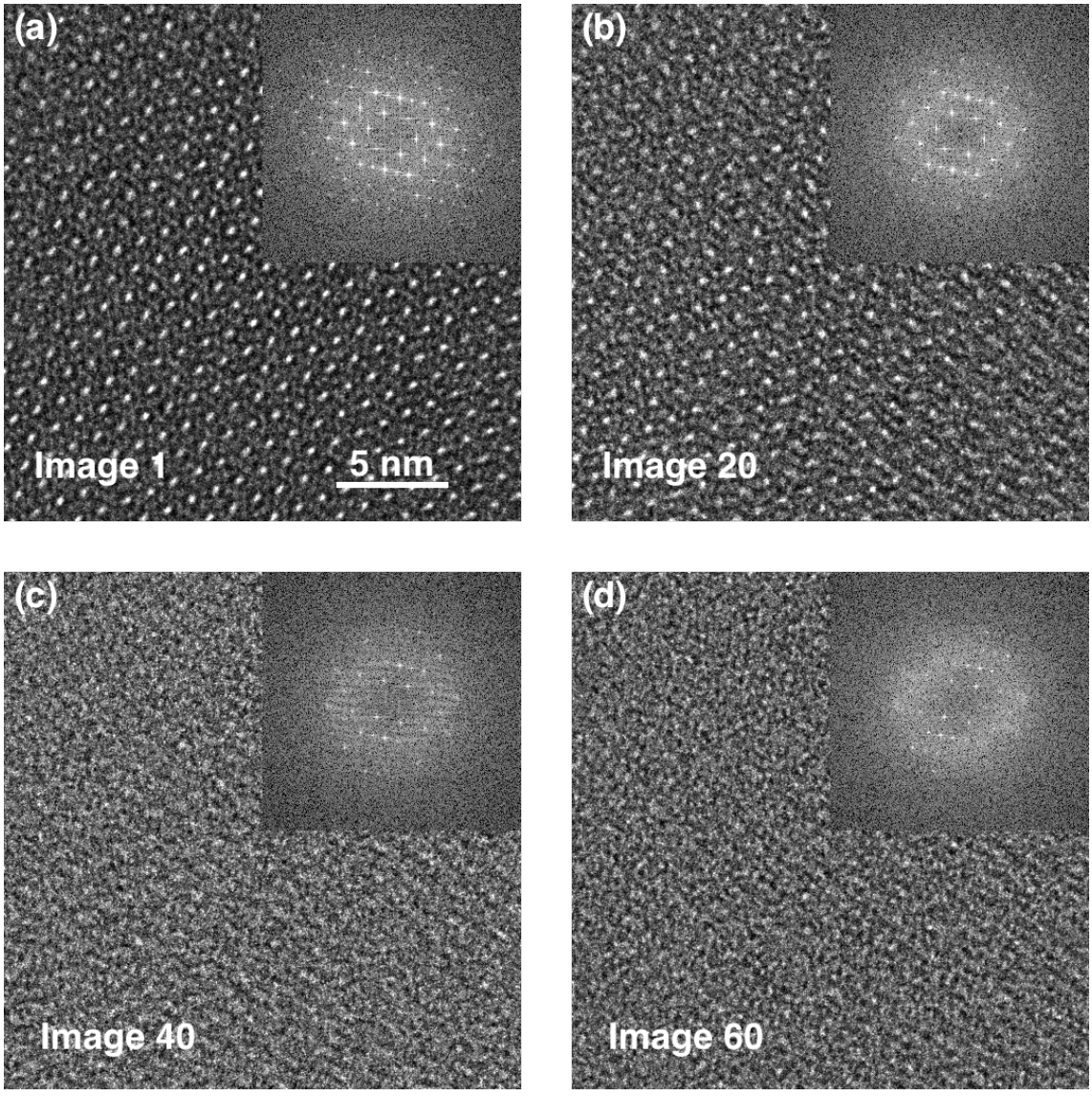


Figure 6.5: (a) 1st, (b) 20th, (c) 40th, and (d) 60th images in the time series of the marked region shown in Fig. 6.4A. The insets are the power spectrum calculated from each image.

Aberrations	Magnitude	Degree
C_1	$-10\,\mathrm{nm}$	n/a
A_1	$1.369\,\mathrm{nm}$	$-66.3°$
B_2	$6.995\,\mathrm{nm}$	$59.5°$
C_3	$144.2\,\mathrm{nm}$	n/a
A_3	$274.5\,\mathrm{nm}$	$-133.8°$
S_3	$64.84\,\mathrm{nm}$	$22.6°$

Table 6.1: Aberration coefficients used in the ZSM-5 time series data.

the lens aberrations after correction.

Fig. 6.5 shows the enlarged version of region A in Fig. 6.4. It shows four frames from different stages of the time series and demonstrates how the zeolite particle gradually lost its crystalline structure as radiation damage progressed. In the real space images, a growing amount of amorphous substance appeared in the region. The insets in each images show that the reflections in the power spectra gradually decreased in intensity, and many of the high spatial frequency reflections completely disappeared into the background in the late stage of the acquisition. To further quantify these observations of radiation damage, three groups of reflections in different directions were chosen for analysis (Fig. 6.6).

A clear reflection peak can be easily fitted by a two-dimensional Gaussian fitting, while a peak submerged in the background noise would not be correctly picked up by the Gaussian fitting. This is used as a criterion for whether a reflection has disappeared from the power spectrum. Because the reflection positions do not change significantly between images, once the Gaussian peak fitting returns a peak position that deviates from its previous position by more than five pixels, it is safe to say the particular reflection has disappeared from the power spectrum and should be excluded from the analysis of future images.

In Fig. 6.6(a1) and Fig. 6.6(a2), the $[\bar{4}\,0\,0]$, $[\bar{8}\,0\,0]$, $[\overline{12}\,0\,0]$, and $[\overline{16}\,0\,0]$ reflections correspond to spatial frequencies of $0.10\,\mathrm{Å}^{-1}$, $0.20\,\mathrm{Å}^{-1}$, $0.30\,\mathrm{Å}^{-1}$, and $0.40\,\mathrm{Å}^{-1}$ respectively. High index reflections were weak at the beginning of the time series, and

quickly decreased in intensity into the background as the acquisition progressed.

After the first nine images in the time series, equivalent to $684\,e/\text{Å}^2$ of imaging dose, the $[\overline{16}\,0\,0]$ peak could no longer be identified from the background noise by Gaussian fitting. For the other reflections in the group, the $[\overline{12}\,0\,0]$ peak disappeared after $1367\,e/\text{Å}^2$ of electron exposure, and the $[\overline{8}\,0\,0]$ peak after $4028\,e/\text{Å}^2$. The vertical axes in Fig. 6.6 are on a logarithmic scale, and hence as seen from the linear trend line fitting, the peak intensities of the $[\overline{4}\,0\,0]$ and the $[\overline{8}\,0\,0]$ reflections exhibit an exponential decay in intensities with the increasing total exposure dose. The same analysis performed on two other groups of peaks confirmed this pattern of reflection decay (Fig. 6.6(b1,c1)).

It should be kept in mind that changes in the aberrations have a direct impact on the contrast transfer function, and hence on the peak intensities in the power spectrum. Therefore, when using the power spectrum to analyse radiation induced peak fading, the microscope settings need to be kept constant. Since image focus is a low order aberration, a deliberately altered focus value, such as those used in a focal series, will strongly influence this intensity measurement. As an example, using the same peak analysis method as used for the time series, the reflection intensities oscillate and no clear trend is evident (Fig. 6.7).

One drawback of using any specific reflection for the tracking of radiation damage is that the reflection intensities decrease at different rates and may not represent the overall degree of damage. A viable way to bypass this problem is to use the IQ factor analysis as an alternative damage evaluation method. Fig. 6.8 plots the IQ factor of the whole time series and shows that the IQ factor decreases in an exponential fashion as the accumulative dose increases, reflected by the fitted linear line in the semi-log plot. The red dots in Fig. 6.9 mark the reflections used for the IQ factor calculation, with the peak selection process carried out automatically by the IQ factor software. Because the IQ factor is a global indicator of the reflection intensities relative to the

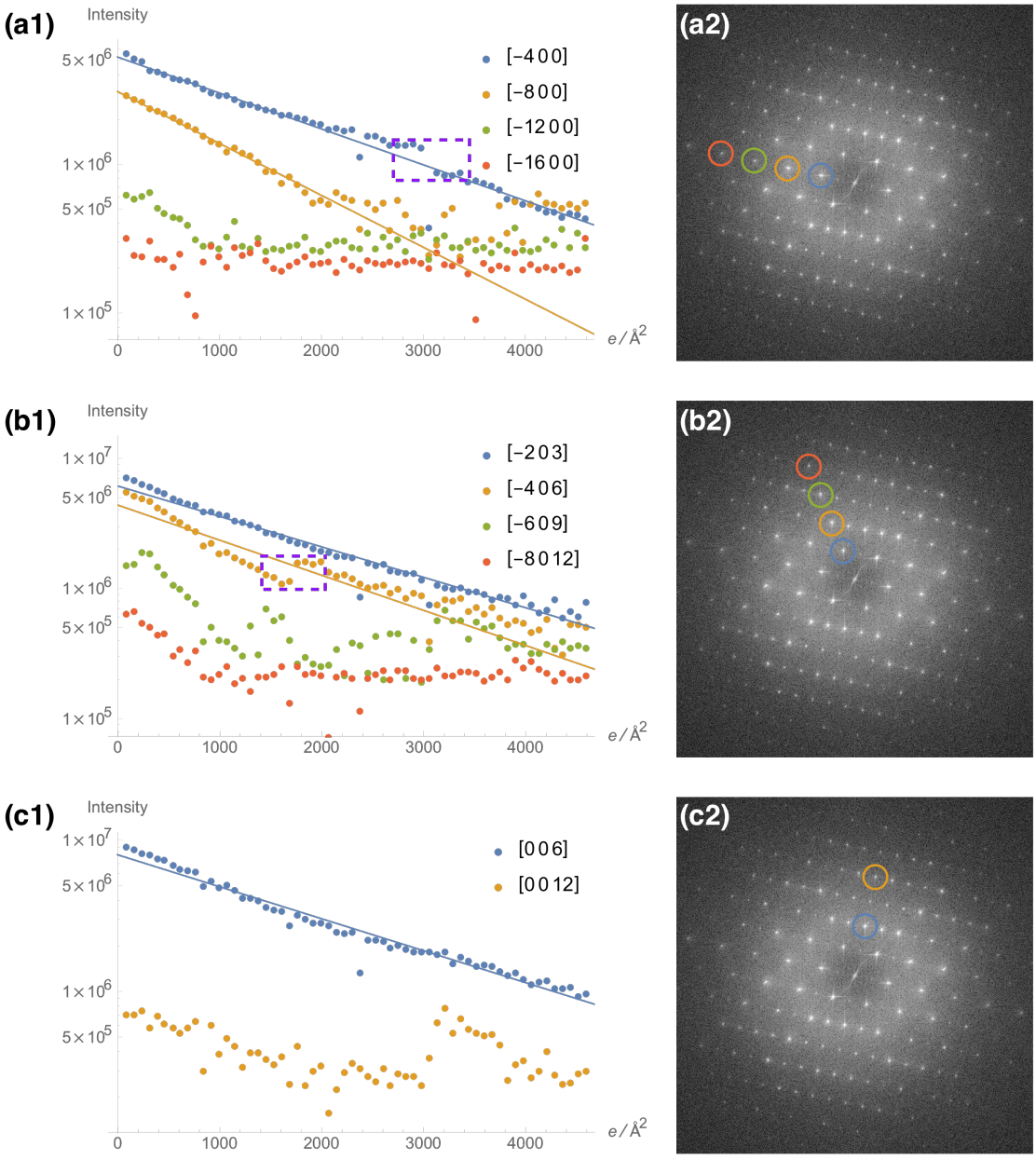


Figure 6.6: Reflection intensity decay curves of three groups of reflections as shown for a time series of 60 images. Reflections in power spectra (a2, b2, c2) are marked by circles with the same colours as the data points in the plots (a1, b1, c1).

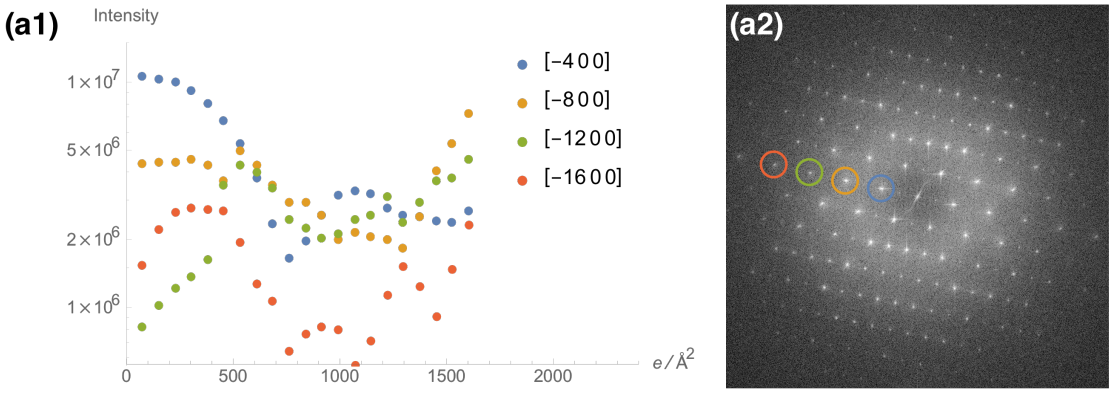


Figure 6.7: Reflection intensity decay curves of the reflections shown in Fig. 6.6a for a focal series of 20 images. Reflections in power spectra (a2) are marked by circles with the same colours as the data points in the plot (a1).

background noise in the power spectrum, the determination of the critical dose, which by definition is the dose needed for the reflection to decrease to $1/e$ of its starting value, can now be unambiguously determined. For this specific case, the critical dose for the ZSM-5 sample is calculated to be $4467\,e/\text{Å}^2$.

Historically, the determination of the critical dose was carried out using diffraction patterns, as described in Chapter 2. No electron microscope images can be recorded during this process. The advantage of using the IQ factor for critical dose determination is that it allows simultaneous HRTEM imaging which provides a clear demonstration of how a radiation-sensitive sample changes at each stage of the damage process. Alternatively, the corresponding image at any data point in Fig. 6.8 is available for visual inspection.

6.2.2 Multi-level XCF registration of ZSM-5 times series

The image registration methods that have previously been discussed in this thesis are all rigid and designed to only deal with translational displacement correction. More often than not, sample change induced by radiation damage is a non-uniform process, with different regions on the same piece of sample shifting towards different

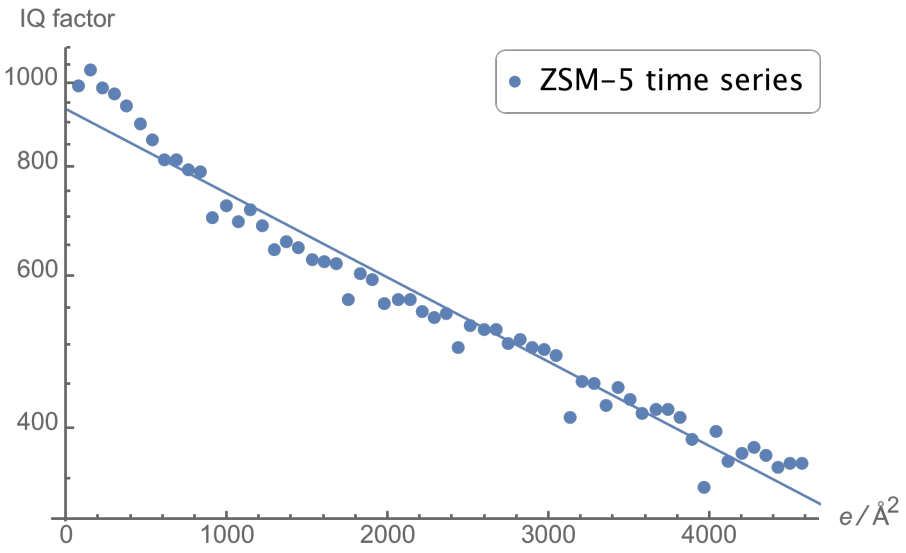


Figure 6.8: Semi-log plot showing the decreasing IQ factor of the time series images.

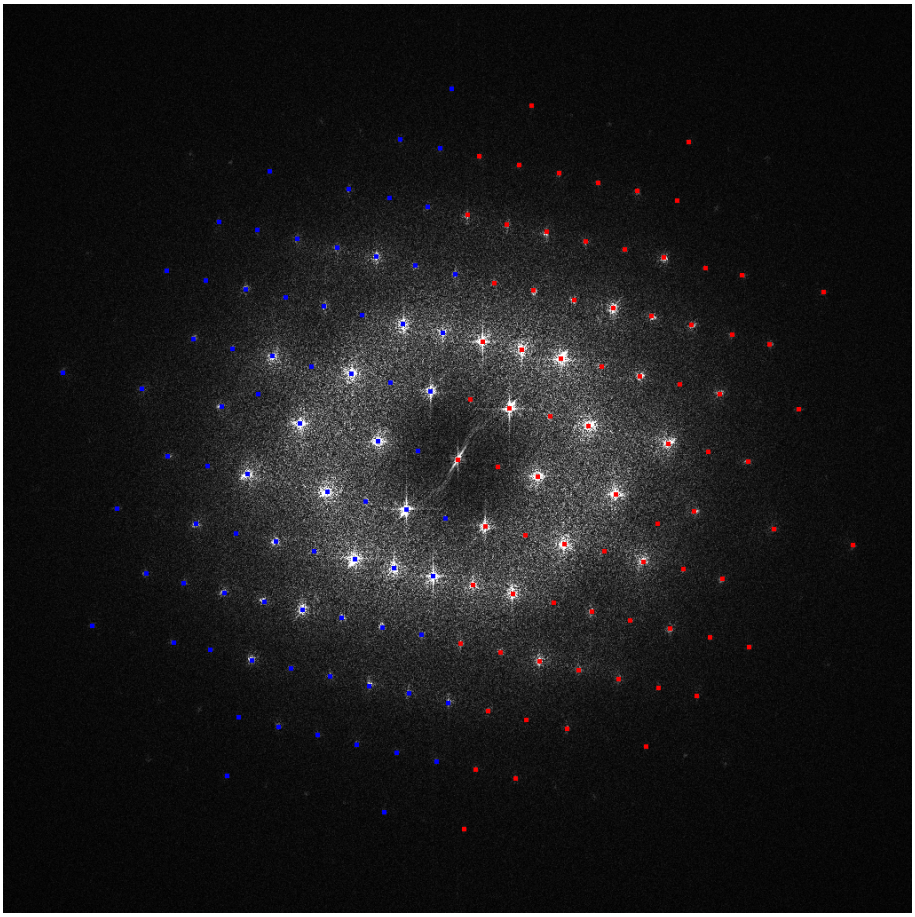


Figure 6.9: All detected reflection peaks in the power spectra of the ZSM-5 time series images (both red and blue dots). Because the power spectrum is centrosymmetric, only half of the reflections (red dots) are needed for the IQ factor calculation.

directions. In such cases, a non-rigid registration scheme can be potentially helpful for aligning image series of radiation-sensitive samples. To first test this hypothesis in a semi-quantitative way, a simple multi-level registration scheme was developed. The core algorithm of this multi-level method is still based on the cross-correlation function, but instead of performing the XCF registration only once between the two images, additional rounds of registration will follow. Each round of registration is carried out between the subregions of the two images to locally optimise the image shift.

Here we used the first image of the ZSM-5 time series as the reference and the thirtieth image as the subject image to demonstrate the multi-level XCF registration. The choice of using the 30th image as the subject image is based on the fact that it is distant enough from the reference image to exhibit some degree of sample deformation, but with the sample still retaining much of its crystallinity. Between the time points when these two images were taken, the cumulative electron exposure is estimated to be $2280\,e/\text{Å}^2$.

The multi-level registration process starts with the initial (level 0) registration, a rigid XCF registration which eliminates the overall sample drift in the subject image, outputting a coarsely aligned image stack. For the subsequent levels, both the reference and the subject image are divided into sub-images. Each sub-image is half the width and height of its parent image, hence the number of reference-subject image pairs multiplies by a factor of four. For example, at level 1 registration, a 2048×2048 image is divided into four 1024×1024 sub-images, each of which will be registered to its counterpart from the reference image, before forming a new subject image passed onto the next level. This resultant image will then be further divided into sixteen 512×512 sub-images at level 2 (Fig. 6.10) which are then registered. This refining process will go on for a number of levels specified in advance.

Two-dimensional "quiver" plots at each level illustrate how the subject image

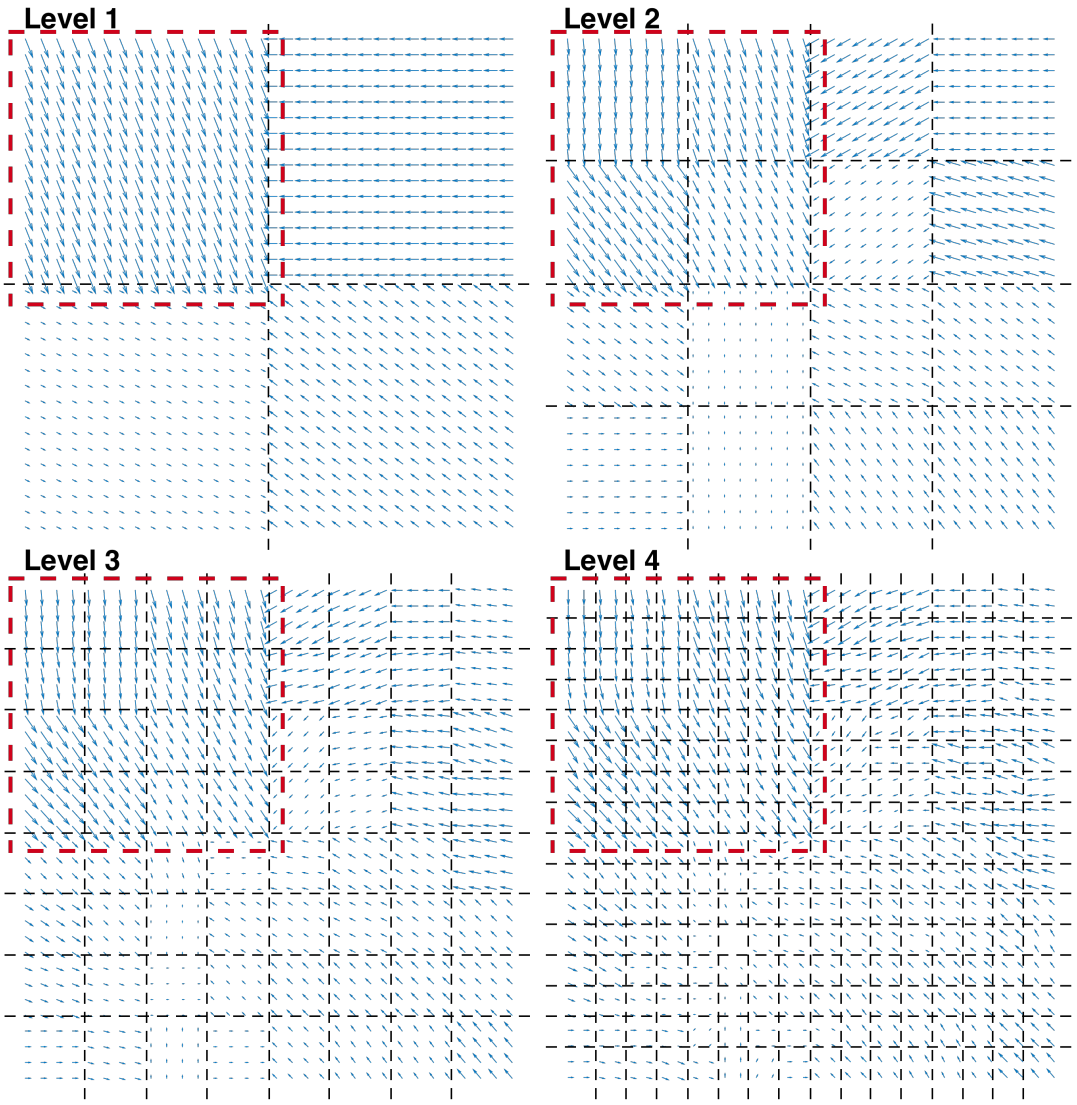


Figure 6.10: Quiver map of the registration vectors for various parts of the subject image over the course of multi-level XCF registration (Level 1 to Level 4). In order to accommodate the chronologically intuitive perception of sample contraction, the arrows in the quiver plots point from the positions in the reference image towards the positions in the subject image. The red dashed box marks the upper left quarter that is abnormally shifted relatively to the other three quarters.

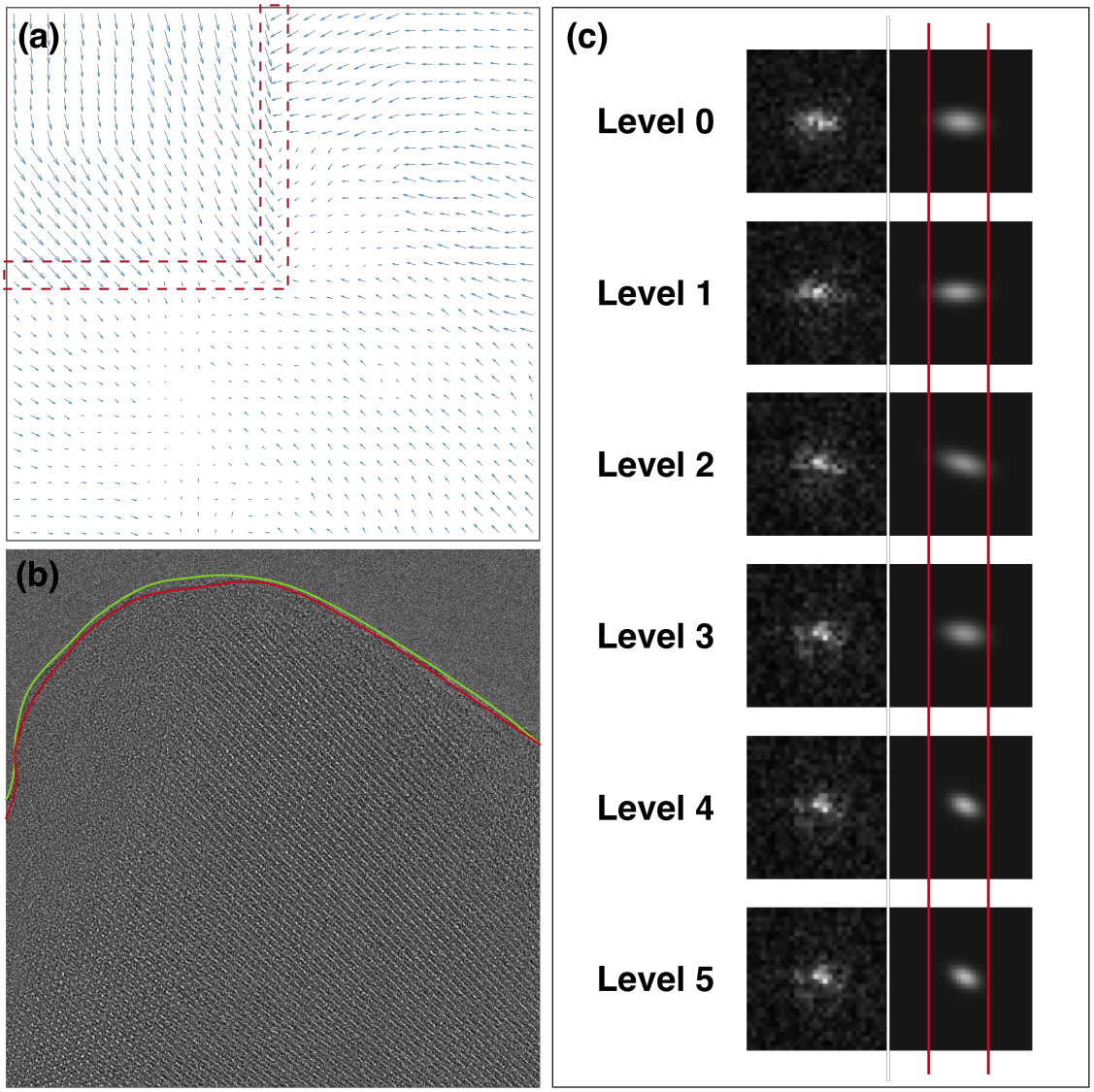


Figure 6.11: (a) Final deformation map of the subject image, with the red box marking the region where abrupt vector changes occur. (b) The subject image prior to the registration, in which the red and green curves are the particle profile before and after registration. (c) An observed sharpening of the $[\overline{6}\,0\,\overline{3}]$ reflection as the multi-level registration proceeds.

is modified towards the best alignment with the reference image (Fig. 6.10 and Fig.6.11a). An evident sample contraction phenomenon is observed from the quiver plots, which can be effectively considered as a deformation map of the sample[1]. The upper left quarter of the deformation map (Fig. 6.11a) shows a collective sharp change in the registration vectors at the edges between one quarter area and the other three quarters, marked by the red dashed box, suggesting an alignment defect in this quarter subregion. The other three quarters of the image, have gradually come to alignment as the registration moves up into higher levels, forming a deformation map free of abrupt changes. Fig. 6.11b shows the particle profile before (red curve) and after multi-level registration (green curve), which overlaps with the particle outline in the first image in the series. It can be seen that due to radiation damage the sample has shrunk towards its mass centre after the acquisition of these thirty images.

More interestingly, at each level of the registration, individual reflections in the power spectrum appear to go through a "sharpening " process as the registration level increases (an example of this is given in Fig. 6.11c for the $[\overline{6}\,0\,\overline{3}]$ reflection). The physical meaning of this phenomenon could be attributed to the two steps of typical radiolysis damage; bond breaking (or fragmentation) and bond reforming. The sample was first damaged in various places where chemical bonds were broken, and then at the damage sites the sample underwent a relaxation process when hanging bonds were reconnected, altering the location of the atoms around them. The cumulative effect of this subtle local disturbance of the periodic lattice is that the unit cells far apart may no longer be in phase with each other and this is reflected by broadened reflection peaks in the power spectrum. The multi-level registration was able to independently adjust different parts of the image according to the undamaged reference image, and as a result, the reflection peak became sharper than its original form.

[1]In order to accommodate the chronologically intuitive perception of sample contraction, the arrows in the quiver plots point from the positions in the reference image towards the positions in the subject image.

6.3 Non-rigid registration of the ZSM-5 times series

The multi-level registration method in the previous section is a simple form of non-rigid registration; although it is sufficient to demonstrate the potential usefulness of applying non-rigid registration to serial imaging in high resolution TEM, it has the shortcoming of being unable to avoid abrupt shifting discrepancies in some local areas, such as that shown in Fig. 6.11a. A more sophisticated non-rigid registration method with proper deformation regularisation will now be introduced and comparatively tested alongside the rigid registration registration in this section. This method is largely based on the previously published work by Berkels et al. (2014), with the important modification of adding an iteration step based on the IQ factor.

6.3.1 Implementation

For convenience, the concept of a deformation map is formally defined as a two-dimensional vector matrix, ϕ, assigning a shift vector to each individual pixel in the subject image. In Eq. 6.1, using the symbol $\circ$ to denote the process of applying the deformation map of non-rigid registration to the subject image, the matrix obtained by image registration will accordingly be altered, f to align with g, the chosen reference[2]. If the deformation map $\phi_{f,g}$ is written with two subscripts, as in Eq. 6.1, the subscripts specify the subject image and the reference image respectively.

$$f \circ \phi_{f,g} \approx g \tag{6.1}$$

6.3.1.1 Multi-level vs. multi-scale

The realisation of obtaining the best deformation map depends on a gradual optimisation process that rolls out through sequential levels. It is necessary to clarify one key

[2]The shift vector in rigid registration can be perceived as a deformation map with identical vectors for every pixel.

difference in the definition of "level" to this advanced non-rigid registration approach and the previously discussed multi-level XCF (ML-XCF) approach. For the ML-XCF method, when the registration moves from a lower level to a higher level, the image is further divided into smaller sub-images, while the pixels within each sub-image remain rigidly positioned with each other. The advanced non-rigid method adopts a different operation; at each level of the registration, the image is *binned* into a down sampled version of the original image. The down sampling is more aggressive at coarser levels (Fig. 6.12(d)(e)(f)) than at finer levels (Fig. 6.12(a)(b)(c)). This reduction in sampling interval neutralises random noise from nearby pixels. After each step of the registration, the image can be then "unbinned" for future rounds of registration. For the purpose of distinguishing the new method from the previously discussed ML-XCF method, the term multi-scale non-rigid approach (MS-NR) is used to refer to the new method, and the concept of *level* in the MS-NR is referred to as *scale*.

Examining Fig. 6.12 it is apparent that starting from too coarse a scale is not helpful for registration. The structural details left in the image may no long be representative after the down sampling, e.g. in the situations shown in Fig. 6.12(e)(f). Therefore, the non-rigid registration performed on the ZSM-5 time series was initiated with images that were binned by 4 (Fig. 6.12(c)), considering that the unit cells of the sample can still be discerned at this level and that the periodicity is also preserved.

The MS-NR registration approach is advanced in three aspects: *regularisation*, *gradient flow*, and *iteration*.

6.3.1.2 Regularisation

The main limitation of the previously discussed ML-XCF method is the lack of ability to avoid discontinuities inherited from the lower levels of registration in the deformation map.

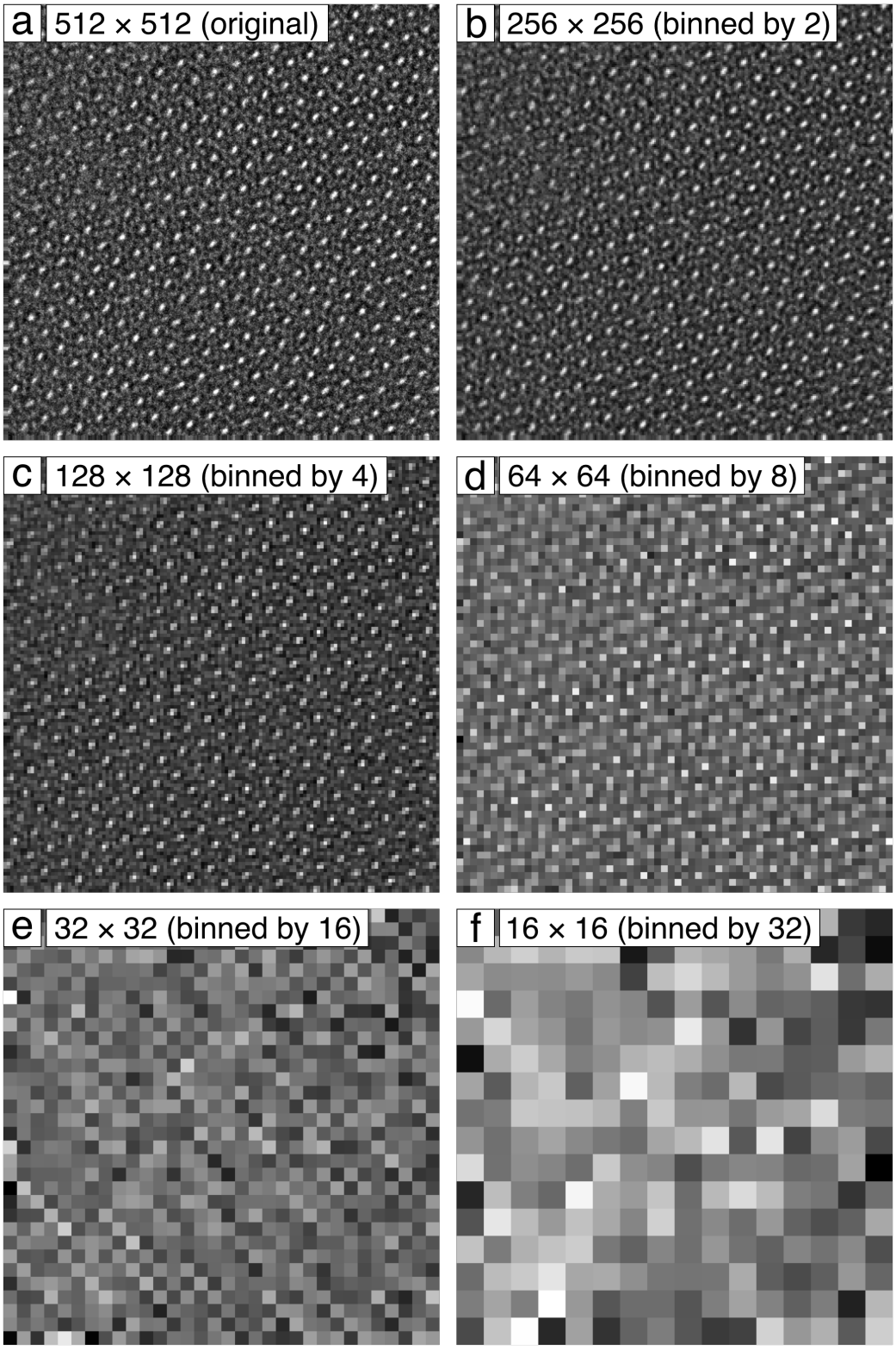


Figure 6.12: The original HRTEM image of ZSM-5, down sampled (binned) by various rates. The actual non-rigid registration starts from images binned by 4, as shown by (c).

The solution to this problem is to introduce a regularisation mechanism to the algorithm, which will punish the occurrence of any discontinuous distortion in the deformation map while minimising its influence on the search for correct image shifts. A generally accepted way of implementing this regularisation, as reviewed in Modersitzki (2003), is through an objective function which has the following form

$$E_{f,g}[\phi] = S_{f,g}[\phi] + \lambda R[\phi] \tag{6.2}$$

in which $S_{f,g}$ is the similarity measure between the reference image, g and the non-rigidly deformed subject image, $f \circ \phi$, is obtained by calculating the negative normalised cross-correlation function (Eq. 6.3).

$$S_{f,g}[\phi] = -NXCF[f \circ \phi, g] \tag{6.3}$$

The weighting factor λ in Eq. 6.2 is adjustable in order to balance the rigidity of the registration and the smoothness of the resultant image. The smaller the factor λ, the more rigid the registration. Unlike the rigid XCF registration methods, where the correlation function is used both as a similarity measure and as a means to determine the shift vector, the correlation function here only evaluates the similarity between the reference image and the deformed subject image. The search for the optimised deformation is carried out through the gradient flow calculation, as discussed later in Section 6.3.1.3.

The role of the regulariser term, $R[\phi]$ is to ensure smoothness of the deformation. It has the form given in Eq. 6.4 (Berkels et al., 2014).

$$\begin{aligned} R[\phi] &= \frac{1}{2}\int_\Omega \|D(\phi(x) - x)\|^2 dx \\ &= \frac{1}{2}\int_\Omega |\frac{\partial \phi_1}{\partial x_1} - 1|^2 + \int_\Omega |\frac{\partial \phi_1}{\partial x_2}|^2 + \int_\Omega |\frac{\partial \phi_2}{\partial x_1}|^2 + \int_\Omega |\frac{\partial \phi_2}{\partial x_2} - 1|^2 dx \end{aligned} \tag{6.4}$$

Symbol	Function	↑	↓
$S_{f,g}[\cdot]$	− Similarity	Worse match	Better match
$R[\cdot]$	Regularisation	Lumpier	Smoother
λ	Weighting coefficient	More rigid	Less rigid

Table 6.2: Terms in Eq. 6.2 and the effects on the objective function.

Table 6.2 specifies the physical meanings of larger (↑) or smaller (↓) values of the terms in the objective function Eq. 6.2. When Eq. 6.2 is minimised, the similarity term[3] will be the smallest under the condition that the regularisation term is minimised at the same time for a given value of λ.

6.3.1.3 Gradient flow

The search of optimum deformation which minimises Eq. 6.2 is realised by a gradient flow algorithm (Berkels et al., 2014). Because the gradient of a function is the first order term in the best linear approximation of the function, its form is not restricted to the Euclidean inner product, as shown by Eq. 6.5.

$$E[\phi] = E[\phi_0] + \operatorname{grad} E[\phi_0] \cdot (\phi - \phi_0) + o(\|\phi - \phi_0\|) \tag{6.5}$$

The gradient flow approach to the minimiser search task adapts a different inner product $g(\cdot)$ (Eq. 6.6), where $g(\cdot)$ is defined by Eq. 6.7.

$$E[\phi] = E[\phi_0] + g(\operatorname{grad} E[\phi_0], (\phi - \phi_0)) + o(\|\phi - \phi_0\|) \tag{6.6}$$

$$g(v, w) = \int vw + \frac{\sigma^2}{2} \int \nabla v \cdot \nabla w \tag{6.7}$$

Hence, we denote this gradient of $E[\phi^l]$ as $\operatorname{grad}_G E[\phi^l]$, where ϕ^l is the deformation map at the registration scale l.

[3]The negative sign clarify that the *similarity* is the highest when the *similarity term* is the smallest.

$$\phi^{l+1} = \phi^{l} - \tau_g \operatorname{grad}_G E[\phi^l] \tag{6.8}$$

with τ_g being the step size. With the objective function and the gradient flow defined, **Algorithm 3** summarises the procedure for the MS-NR registration between two images, with f being the subject image and g the reference image. The registration algorithm also needs specified inputs of the starting level m_0 and the ending level m_1. The initial deformation can be a simple identity matrix.

Algorithm 3 Pair-wise registration.

procedure MULTI-SCALE GRADIENT FLOW
Require: starting level m_0 and ending level m_1
Require: image $f = f^{m_1}$ and $g = g^{m_1}$
Require: initial deformation $\phi = \phi^{m_1}$
    for $m = m_1 - 1, \cdots, m_0$ **do**
        Initialise $[f^m, g^m, \phi^m]$ by restricting $[f^{m+1}, g^{m+1}, \phi^{m+1}]$
    end for
    for $m = m_0, \cdots, m_1$ **do**
        Register f^m and g^m using the gradient flow (Eq. 6.8) on level m, with ϕ^m as initial deformation
        Store the resulting deformation in ϕ^m
        if $m < m_1$ **then**
            Set ϕ^{m+1} to the prolongation of ϕ^m
        end if
    end for
    return ϕ_1^m
end procedure

6.3.1.4 Iteration

Aligned images in a time series stack can be averaged to give a mean image which has a better image quality and a higher SNR where Poisson noise is reduced by $\sqrt{N}$ (N being the number of aligned images). An incremental mean image (IMI), $f_N^{(im)}$, is defined as the averaged image from the first to the Nth image after registration (Fig. 6.13), with f_i and ϕ_i being individual images from the series and their deforma-

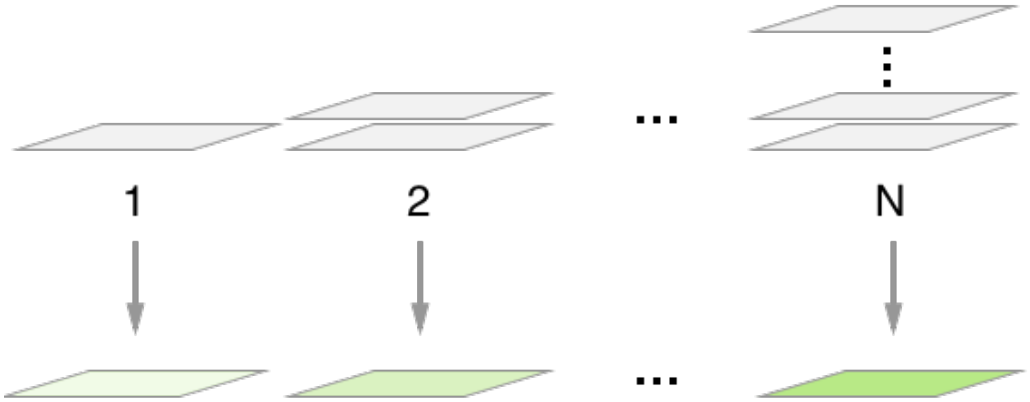


Figure 6.13: The incremental mean image defined.

tion, i.e.;

$$f_N^{(im)} = \frac{1}{N} \sum_{i=1}^{N} f_i \circ \phi_i. \tag{6.9}$$

The IMI of the already aligned part of the series can be used as the new reference in the next iteration, after which more images are registered and added into the updated mean image. The situation becomes more complicated if the sample is prone to radiation damage. Damaged substance will likely be aperiodic and will appear in the image as effectively random noise in the background. There will be a point when adding new images to the mean image will result in a worse reference than the existing mean image. Therefore an IQ factor of the mean image is integrated into **Algorithm 4** to determine the point at which it starts to decline.

Algorithm 4 Time series registration.

procedure SERIES AVERAGING PROCEDURE
  for $i = 2, \cdots, n$ **do**
    Compute $\phi_{i,i-1}$ with MS-NR (initial guess identity)
  end for
  Initialise 0-th average with $f^0 := f^1$
  for $k = 1, \cdots, K$ **do**
    Compute $\phi^k_{1,0}$ with MS-NR (initial guess identity)
    for $i = 2, \cdots, N$ **do**
      Compute $\phi^k_{i,0}$ with MS-NR (initial guess $\phi_{i,i-1} \circ \phi^k_{i-1,0}$)
    end for
    if $k == 1$ **then** ▷ Determine how many frames to use
      for $i = 2, \cdots, N$ **do**
        if $IQ_{sum}(f_i^{(im)}) < IQ_{sum}(f_{i-1}^{(im)})$ **then**
          break
        end if
      end for
      $N_{\text{opt}} = i - 1$
    else
    end if
    $f^k = f^{(im)}_{N_{\text{opt}}}$
  end for
end procedure

6.3.2 Comparison of rigid and non-rigid registration methods using the IQ factor

6.3.2.1 ZSM-5 TEM time series

The ZSM-5 TEM time series was aligned using both the rigid XCF registration and the multi-scale non-rigid registration. To compare the effectiveness of the two methods, the registration results were analysed with the assistance of the IQ factors of the incremental mean images (IMIs) obtained from the aligned series. If the sample was not damaged during TEM imaging and the image series are aligned, the IQ factor of IMIs would increase as the number of images, N, increases (Fig. 6.13). For the same time series, a higher IQ factor of the IMI would indicate a better image registration as the averaging operation reduces the image noise and amplifies the signal.

Fig. 6.15 shows the IQ sum values calculated from the incremental mean images of the ZSM-5 time series after the rigid XCF registration and the non-rigid registration. It appears that the non-rigid algorithm is more effective than the rigid algorithm for the first four images of the series, due to the iterative refinement implemented within the non-rigid method. However, as the number of images increases, the rigid algorithm catches up and surpasses the non-rigid method at near the peak IQ sum value by a margin of 37 (2.8 %), when the first nine images were included. The registered low-dose image stack shows a noticeable improvement in signal-to-noise ratio and resolution by averaging several images together (Fig. 6.14(c,d)). On the other hand, for images taken after a large amount of beam damage has occurred to the fragile sample, further adding images into the IMIs would not improve, but instead would deteriorate, the final image quality. The rigid method is slightly superior until about 40 images into the time series when the damage became very severe and the incremental mean images obtained from the rigid method starts to drop quickly in the IQ sum while the results from non-rigid registration only decreased by less than

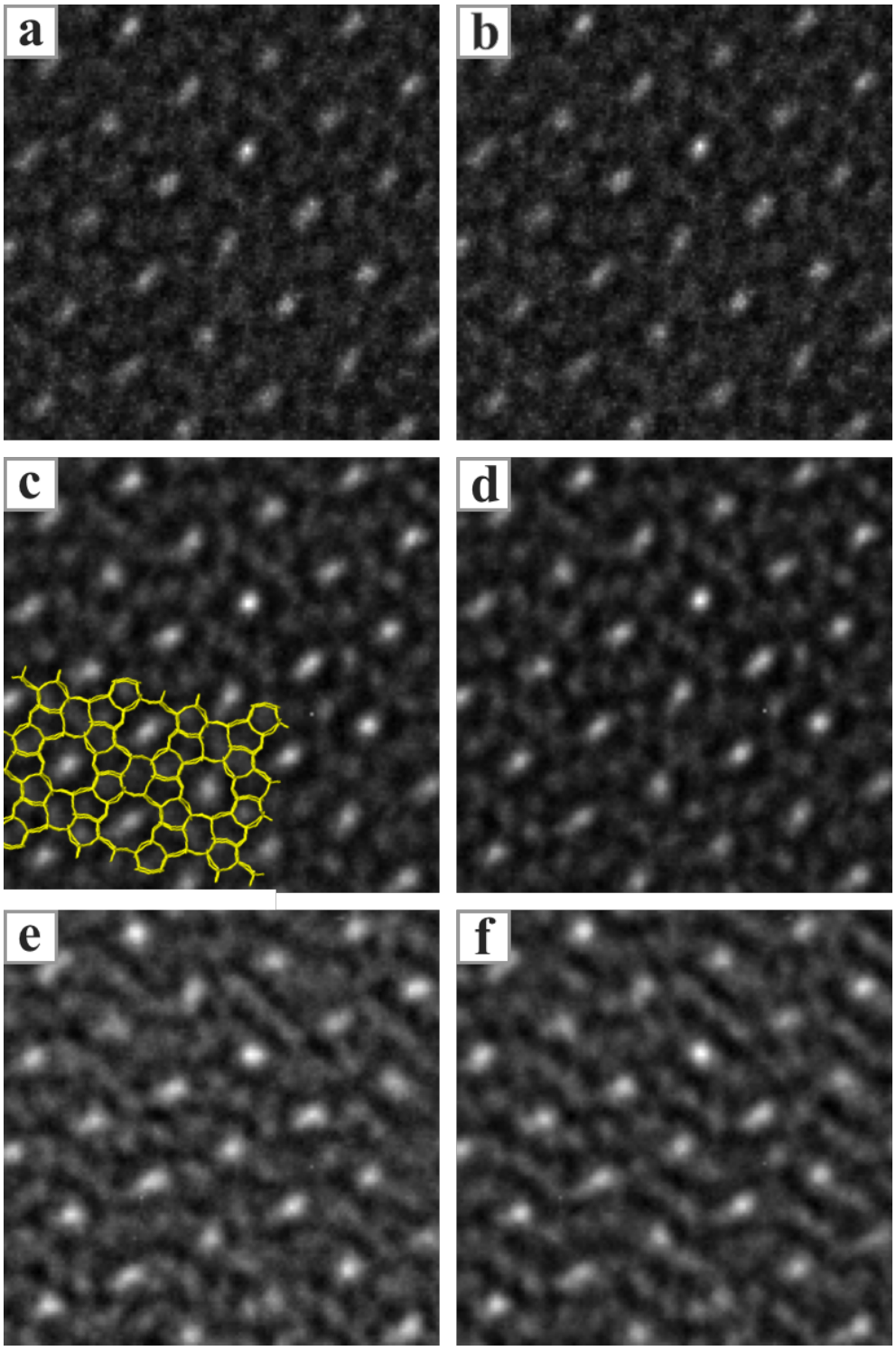


Figure 6.14: Mean images of (a,b) the first, (c,d) the first nine, and (e,f) all sixty images of the HRTEM time series of ZSM-5. (a,c,e) are results from the rigid XCF registration, (b,d,f) are results from the nonrigid registration. The ZSM-5 structure is visible in the averaged images (c) and (d), though the contrast is reversed due to setting of defocus and spherical aberration, leaving the supposedly empty round-shaped channels brighter than the atoms forming the framework.

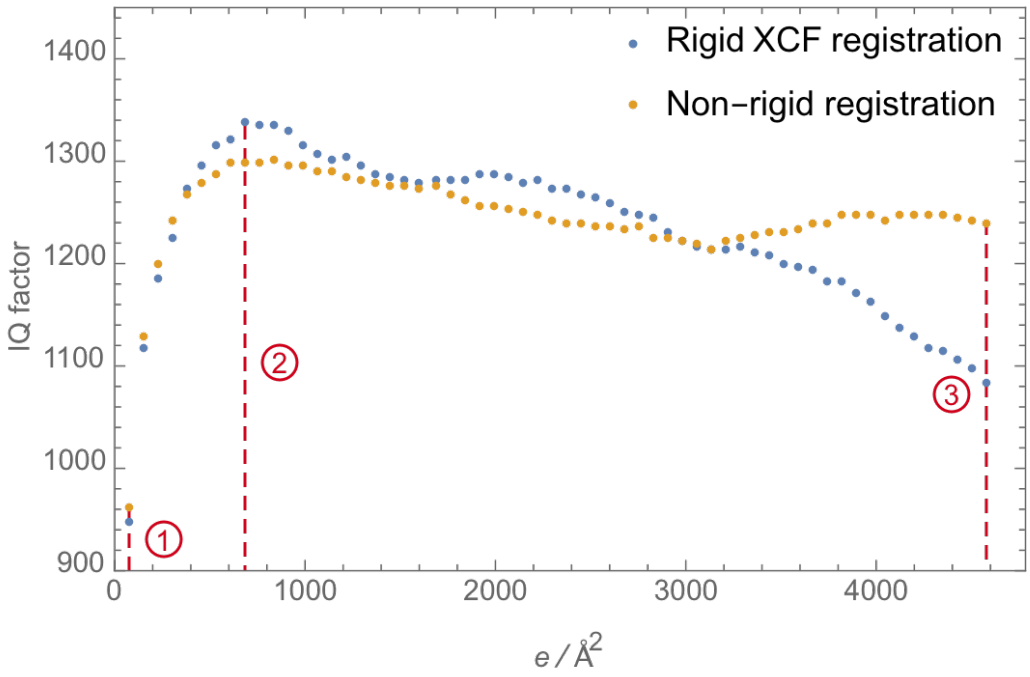


Figure 6.15: Comparison of IQ factors of the ZSM-5 TEM time series incremental mean images after two registration procedures. Number 1, 2, and 3 correspond to the first, the ninth, and the sixtieth incremental mean images.

100 in the IQ sum and reached a stable value.

Discussion: The reason why the non-rigid registration method cannot produce the same or a higher IQ factor is unclear. A possible cause could be that the ZSM-5 sample, albeit radiation-sensitive, has a damage rate that is not too fast for the rigid registration to adapt. In particular, the rigid registration uses the neighbouring-reference scheme and the changes between adjacent images, as briefly mentioned in Section 6.2.2, are not substantial. Moreover, the TEM image, given its simultaneous recording, does not suffer much from the image distortions that are typically worse in microscopy techniques that relies on a scanning probe, such as scanning transmission electron microscopy (STEM) and scanning electron microscopy (SEM).

There has also been suspicion that the resolution of the images might have an impact on the quality of the MS-NR results. The TEM time series of the ZSM-5

particle was taken at a moderate magnification (×500000) due to restriction of the electron dose. This choice of magnification is beneficial in boosting the SNR of the low-dose image. However, as can be seen in Fig. 6.12, only the first three scales of the down sampling treatment preserve enough structure information for sensible registration. The other three scales (Fig. 6.12(d)(e)(f)) clearly misrepresented the sample structure and therefore should not be included in the actual registration workflow. Because of a lack of comparable data on this issue, this relation between the data resolution and the attainable registration quality is yet to be experimentally confirmed.

6.3.2.2 STEM time series: zeolite Y

A time series of another zeolite sample, zeolite Y, was recorded by Douglas A. Blom from University of South Carolina, using a 200 kV aberration-corrected JEOL JEM 2100F TEM/STEM in the high-angle annular dark-field (HAADF) imaging mode (Fig. 6.16). The illumination semi-angle was 15.5 mrad, while the collection angle of the detector ranged from 50 mrad to 284 mrad. The electron dose in each frame of the series was estimated to be 1400 $e/\text{Å}^2$. Radiation damage happens faster than that in the TEM ZSM-5 time series and the image drift is considerably worse. Therefore, only the first fourteen frames of the series were used for comparative analysis.

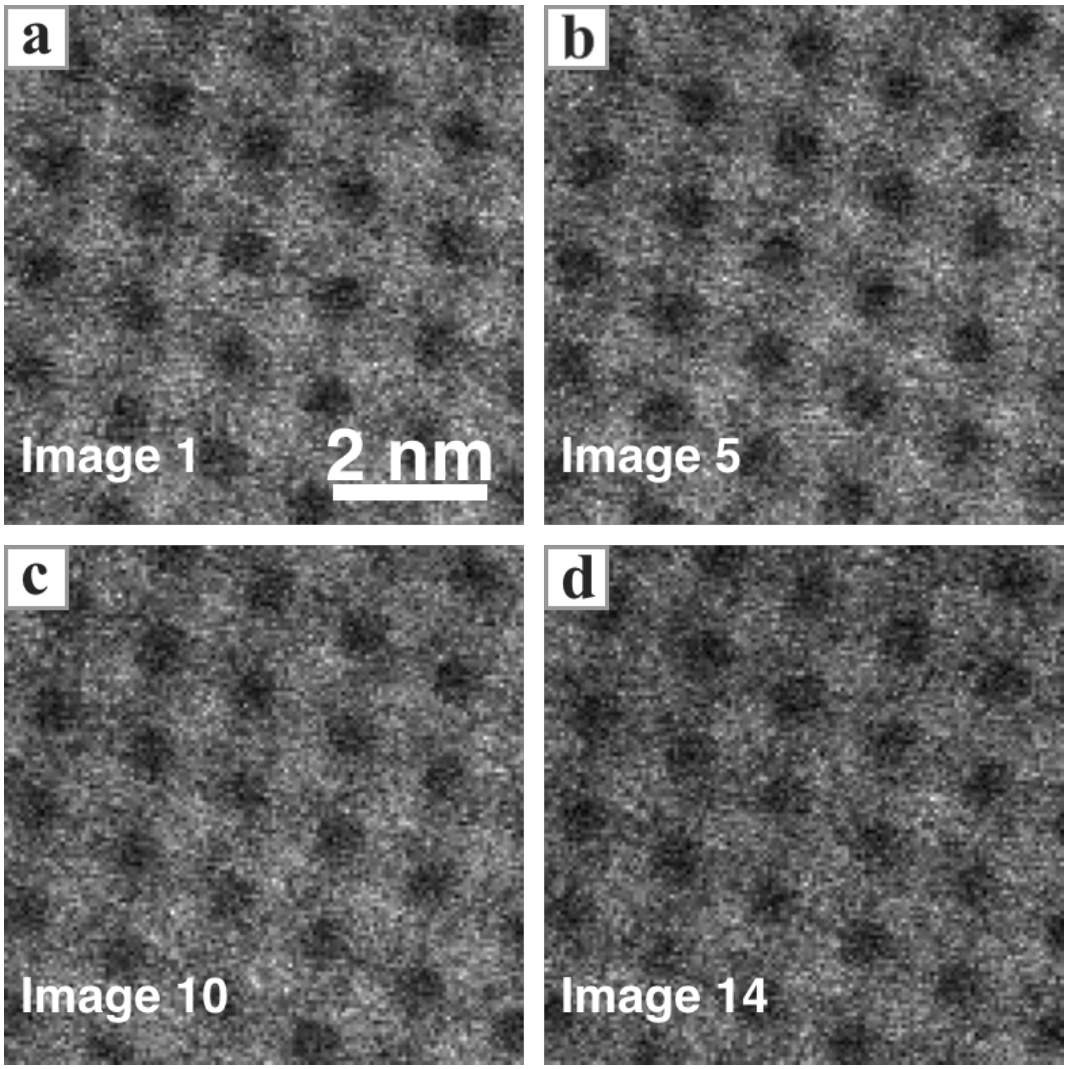


Figure 6.16: The 1st, 5th, 10th, and 14th images of the zeolite Y STEM time series.

The incremental mean images of the STEM series were obtained in the same way as in the analysis of the TEM series. Fig. 6.17 demonstrates the improvement in the resolution of the IMIs using both the rigid and non-rigid registration. There is noticeably more fine structure visible after averaging fourteen non-rigidly aligned images, which are not clearly resolved in the IMI after the rigid registration. The IQ sums of the IMIs calculated from the aligned zeolite Y STEM time series are plotted in Fig. 6.18. The increasing trends in IQ factors from both methods confirm the effectiveness of the registration and averaging approach in enhancing the image quality. In comparison, the MS-NR registration of the STEM series demonstrates a

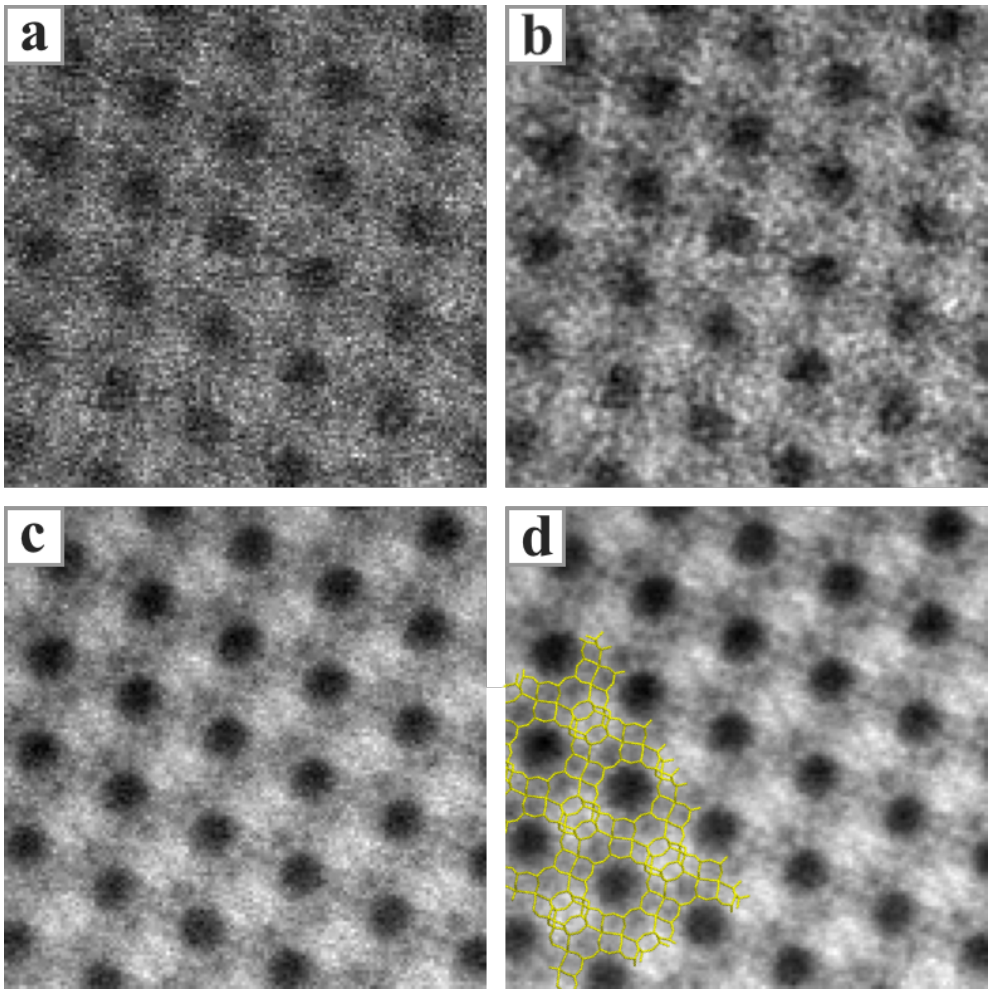


Figure 6.17: Mean images of (a,b) the first, (c,d) all fourteen images of the STEM time series of zeolite Y. (a,c) are results from the rigid XCF registration, (b,d) are results from the non-rigid registration. More structural details are resolved in (d) than in the others.

superior performance over the rigid XCF registration, performing consistently better than the latter, particularly for the second half of the series when more damage has accumulated and the benefits of adapting the non-rigid method become more obvious. If using all fourteen frames of the STEM series are used to form the mean images, the MS-NR registration gives a 11.3 % better IQ sum than the rigid method.

Conclusion: Both examples showed that the conventional rigid XCF methods remains a robust and computationally efficient registration scheme, although the simple

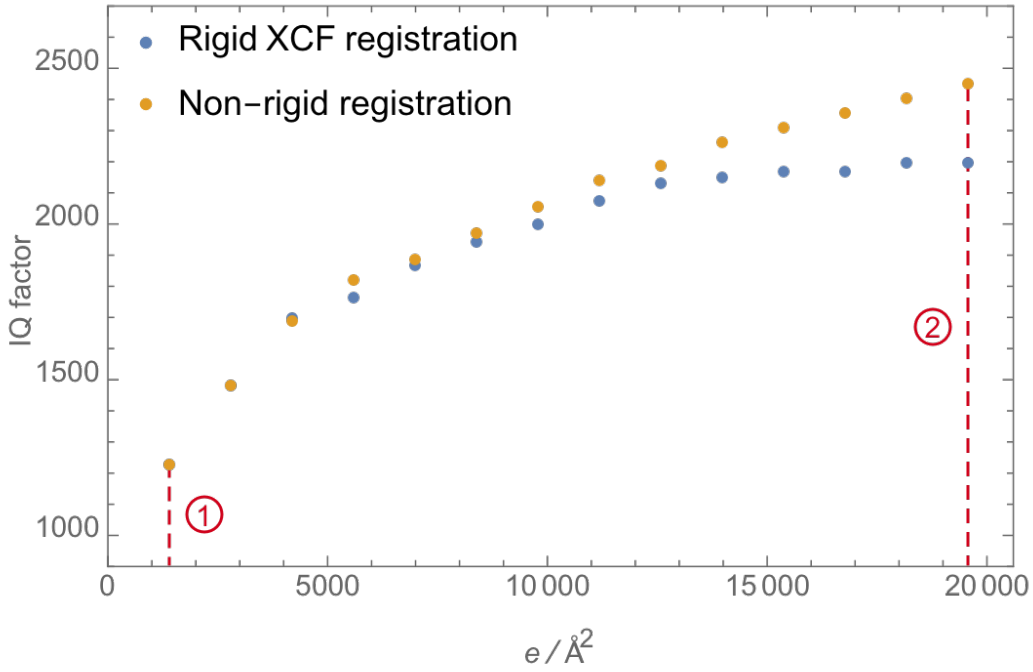


Figure 6.18: Comparison of IQ factors for the zeolite Y STEM time series incremental mean images after two registration procedures. Number 1 and 2 correspond to the first and the fourteenth incre- mental mean images.

treatment of image misalignment as a rigid translational movement would only hold true when radiation damage is negligible. TEM images do not suffer much from the image distortions that are typically worse in microscopy techniques that rely on a scanning probe, such as scanning transmission electron microscopy (STEM) and scanning electron microscopy (SEM). Nevertheless, The ZSM-5 TEM image series did not continue to produce better IMIs after a certain number of images when the radiation damage had become serious, as shown by the rigid registration result of the long TEM time series of ZSM-5. Because the locations on a sample where damage happens are arbitrary and asynchronous, the deformation will be complicated and hence requires the more sophisticated MS-NR registration approach. In contrast, the distortion caused by acquisition mechanism in the STEM time series is much more relevant to image registration. While the damage was apparent after the time series, it was not the primary limiting factor. Because the scanning distortion in each image

Parameter	Value
Voltage	200 kV
Defocus range	−9 nm to 11 nm
Focal step	4 nm
Number of images	6
Spherical aberration (C_3)	0.23 µm
Sampling interval	0.048 85 nm
Dose per frame	63 e/Å^2

Table 6.3: Imaging conditions for the ZSM-5 focal series data.

is not inherently related to that of the others, the advantage of using the MS-NR registration is more evident; 11.3 % higher in IQ factor sum for the fourteen images in this example in comparison to registration by the rigid method.

6.4 Focal series low-dose imaging and exit wave reconstruction of ZSM-5

6.4.1 Focal series imaging

Focal series imaging adopts essentially the same low-dose experimental procedure as for the time series, except that the final acquisition step is controlled by a computer script which automatically changes the defocus value before recording each individual frame. For the experiments described here, the aberration-corrected TEM was operated at 200 kV and the spherical aberration was corrected to 0.23 µm. The focal series contained six images, with a focus range between −9 nm to 11 nm and the focal step was a constant 4 nm throughout (Table 6.3).

For radiation-sensitive materials, the number of images is restricted by the critical dose and by the dose used for each single acquisition. A larger number of images means a lower dose budget for individual images. The electron dose for the focal series studied here was controlled to be 63 e/Å^2 for each image in the series, giving

a total imaging electron dose of $378\,e/\text{Å}^2$ over 6 images (Fig. 6.19). This was well below the critical dose of ZSM-5 ($4467\,e/\text{Å}^2$) measured by the time series experiment in Section 6.2.1. Consequently, the sample showed limited radiation damage over the process of data acquisition.

6.4.2 Exit wave reconstruction

Because the focal series is short, comprising only 6 images, the neighbour-reference XCF registration can be used to align the images, as the propagation of registration errors is negligible for short image series.

After omitting the -1 nm defocus image which is very close to the Gaussian focus and has very low image contrast, the exit wave reconstruction was carried out using the linear Wiener filter method on the remaining five images, with the input experimental parameters listed in Table 6.3. Examining the exit wave of the ZSM-5 focal series (Fig. 6.20), it is found that the signature structure of ZSM-5, in which ten smaller rings surround a large hole, is clearly resolved across the field of view in the wave phase, while the wave amplitude shows low contrast, implying that the phase object approximation assumed by the linear imaging model was satisfactory and that most structure information about the material was transferred in the phase shift of the electron wave. The direct correlation between the exit wave phase and the schematic atomic model of ZSM-5 (embedded in Fig. 6.20b, in which bright contrast shows the SiO_4 tetrahedra) proves the effectiveness of exit wave reconstruction for complex structural characterisation. In contrast, the individual low-dose HRTEM images in the series at different defoci (Fig. 6.19) exhibit substantial differences in image contrast, leaving much ambiguity over direct image interpretation. For example, the tiny voids in the middle of 5-membered and 6-membered rings are bright in under focus images but dark in over focus images.

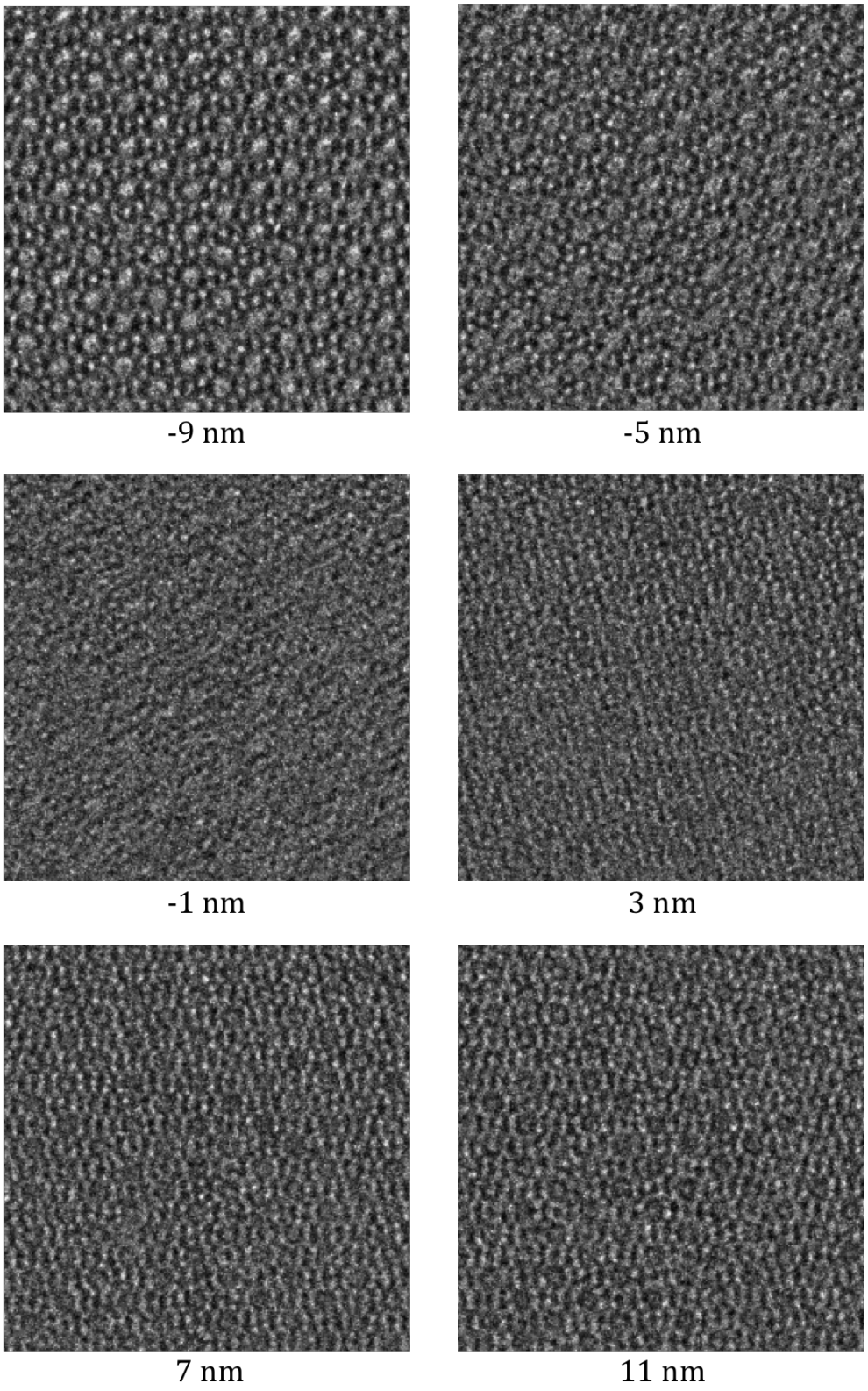


Figure 6.19: Low-dose focal series of ZSM-5 in [1 0 0] direction, with a focal step of 4 nm between adjacent images. (The −1 nm image was not included in the exit wave reconstruction due to low image contrast.)

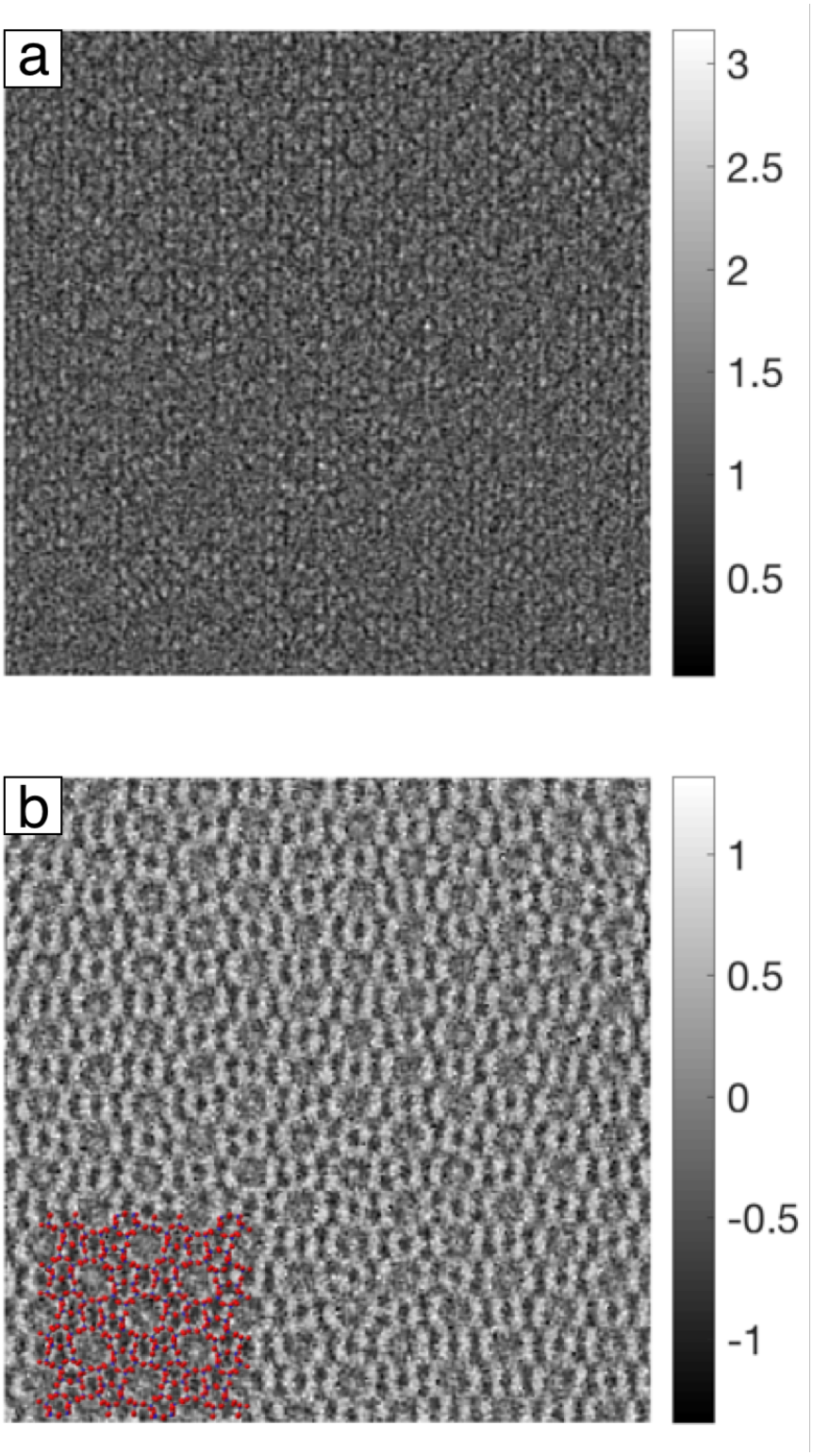


Figure 6.20: (a) Amplitude and (b) phase of the exit wave (unit: radian) reconstructed from 5 images of the ZSM-5 focal series. An atomic model of ZSM-5 overlaps the lower left corner of the phase, indicating the positions of framework and the voids.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

7.1.1 Low-dose HRTEM imaging under controlled dose conditions

- A quantitative dose measurement scheme dedicated for high resolution TEM imaging has been developed, composed of an instrumental dose calibration prosedure (Section 4.1.4) and an image sampling interval calibration procedure (Section 4.1.3).
- Ultra high resolution TEM characterisation has been carried out on hydroxyapatite (Section 4.4.1), cerium dioxide nano particles (Section 5.2.1), silicon nitride (Section 5.2.2), and ZSM-5 (Section 6.2.1 and Section 6.4.1).
- Critical dose of the very radiation-sensitive ZSM-5 ($4467\,e/\text{Å}^2$) has been quantitatively determined (Section 6.2.1).
- A noise reduction method suppressing the striping noise in low-dose images has been developed and proved to be effective at removing striping noise caused by inconsistent detector bias (Section 5.3).

7.1.2 Image registration of low-dose HRTEM images

- The "bright central pixel" artefact found in the cross-correlation function (XCF) when registering low-dose HRTEM images has been overcome with an enhanced XCF registration method that eliminates this abnormal pixel during the shift vector determination process. This significantly extends the lowest dose, from the $44\,e/\text{Å}^2$ to the $2.2\,e/\text{Å}^2$, at which the XCF registration method can still effectively aligning the hydroxyapatite time series data (Section 4.5.3).

- A new evaluation metric, the IQ factor, has been introduced and implemented for quantitative assessment of the image quality. The IQ factor also becomes a important way of determining which registration method gives better result than the others, by judge the quality of the mean image after registration (Section 4.6.2).

- A simulation-assisted XCF registration has been developed and applied to focal series data that experiences contrast reversal at close to zero defocus (Section 5.4.2). This method shows a significant accuracy improvement in registering atomic resolution focal series images, when the influence of contrast reversal is the most severe.

7.1.3 Exit wave reconstruction of low-dose focal series

- Focal series of CeO_2 at different doses have been registered with three different registration strategies, fixed-reference XCF, neighbour-reference XCF, and simulated-assisted XCF. The effects of electron dose and registration have been studied using the IQ factor as a measurement. The results has suggested that a dose threshold must be reached for the reconstruction methods to retrieve quantitatively meaningful exit waves.

- It has been shown that the existing exit wave reconstruction methods can be

used on low-dose HRTEM focal series, provided that the series is correctly registered (Section 5.5). Exit waves of a CeO_2 nanoparticle and part of a Si_3N_4 crystal were reconstructed.

- Below a certain threshold, the amount of phase shift will decline as the imaging dose of the focal series decreases (Section 5.5.2.1).

- The quality of image registration has an impact on the exit wave reconstruction results. The best registration, which was given by the simulation-assisted XCF, leads to better restored exit wave phases, which have higher IQ factor for the CeO_2 focal series data at all doses (Section 5.5.3).

- Exit wave reconstruction was successfully carried out on a ZSM-5 focal series containing 6 images.

7.2 Suggestions for future work

7.2.1 Noise reduction and direct electron detection

Noise is the biggest issue for analysing low-dose HRTEM images. Avoiding introducing noise during experiments and eliminating noise after the data is acquired are paramount when trying to characterise radiation-sensitive materials that have a very limited dose budget. New generation detectors may help to improve the efficiency of electron detection with lower level of noise. Compressed sensing has recently generated interest among the STEM community, while more feasible experimental designs of its potential application to TEM are still needed (one recent progress in this area can be seen in (Stevens et al., 2015)). For off-line noise reduction methods, it would be worth to explore the existing methods in the field of signal processing.

The low signal-to-noise ratio problem of low-dose TEM imaging can be ultimately overcome by using a direct (electron counting) detector, which would not induce any

detector noise.

Appendix A

Round Electronmagnetic Lenses

The structure of the simplest round electromagnetic lens in the TEM is illustrated in Fig. A.1. A set of copper coils (or "solenoid"), when carrying an electric current, generates a magnetic field according to the Ampere's circuital law. The polepieces, which are made of soft iron or similar material, confine the magnetic field in a very small gap between the upper polepiece and the lower polepiece. The trajectory of the electrons are altered because of the interaction between the negatively charged electron and the magnetic field, showing a similar result to light being converged by a convex lens. However, the underlying physical process is different.

An intuitive way to understand the focusing process is to perform a simple analysis of the magnetic forces that an electron experience inside the lens. As shown in Fig. A.2, when entering the magnetic field with a speed of v_z along the optical axis, an electron is accelerated in the direction normal to the radical component of the magnetic field B_r inside the horizontal plane. After the electron accumulates a speed in the same direction as the force $F_{B(r)}$, a subsequent force $F_{B(z)}$ pointing towards the field centre directs the electron closer to the axis. If the magnetic field is sufficiently deep along the axis and its radical component does not flip direction, the trajectory of the electron will spiral and the electron beam will be focused to the axial centre.

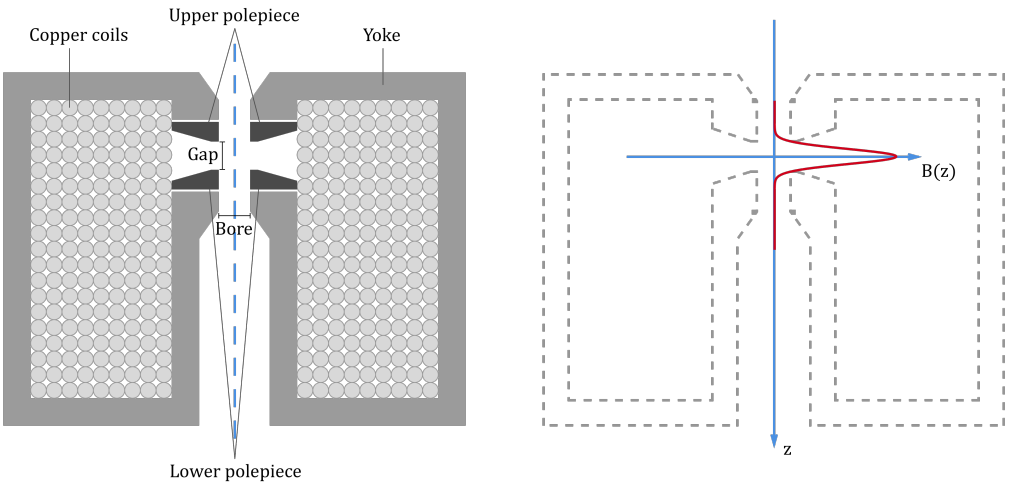


Figure A.1: Illustration of the structure of a simple round lens, and the change of magnetic field strength $B(z)$ along the axial direction z.

With only some qualitative analysis, it is clear that this model is inaccurate in at least two aspects. Firstly, electromagnetic lenses are thin and the vertical length of the magnetic field is limited, especially considering that the high energy electrons are travelling at high speed. In reality, the electron will only complete a fraction of the spiral track as opposed to the illustration in Fig. A.2. Secondly, the radical component of the magnetic field will flip the electron trajectory direction and the corresponding magnetic force $F_{B(r)}$ will also reverse direction. As stated by the Busch's theorem (Dalglish and Kelly, 1975), the angular velocity of any electron at any point depends solely on the difference between the total magnetic fluxes linked by the charge rings at the two electron positions and is independent of the detailed trajectory. This suggests that, for the symmetric magnetic field of the lens, the canonical angular momentum of the electrons passing through it should retain its value before and after the lens. For an electron with no azimuthal velocity prior to entering the lens field, when it exits the lens, its azimuthal velocity returns to zero, meaning that the lower part of the lens field will exactly cancel out the azimuthal velocity given by the upper part of the lens. However, a small but important radical velocity is preserved and it

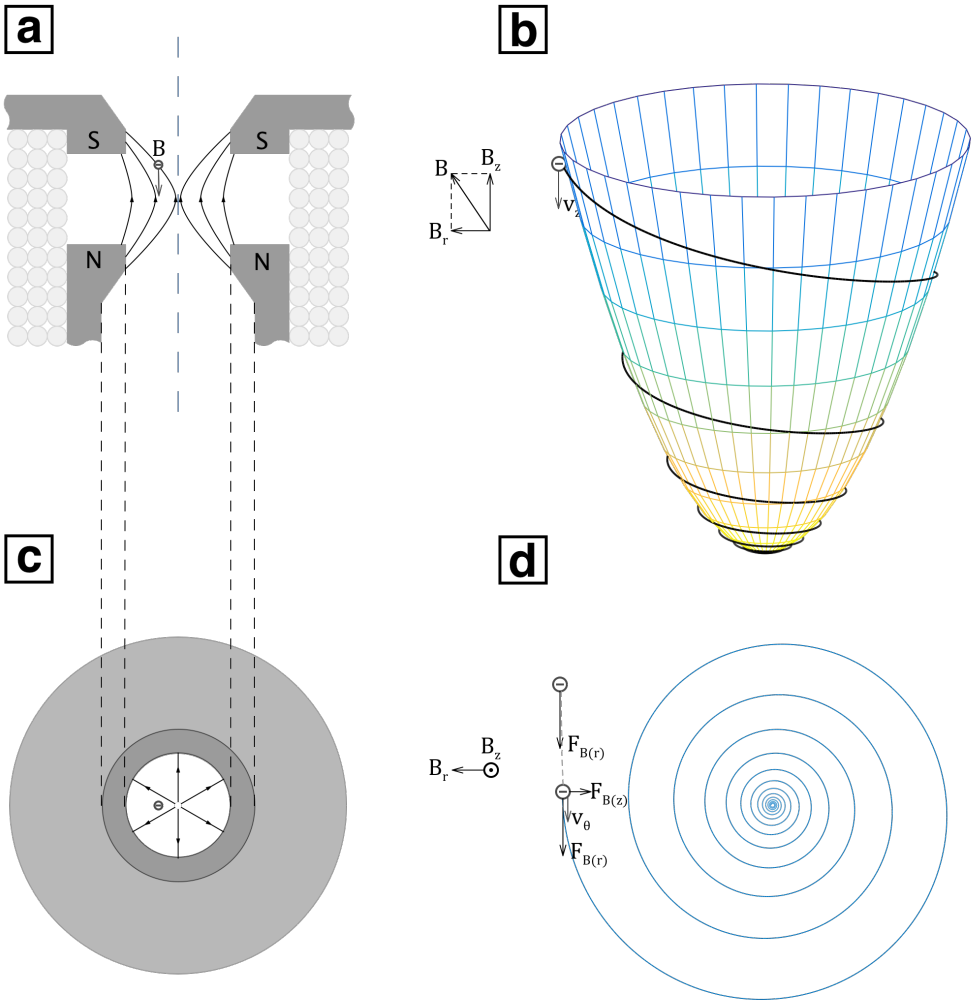


Figure A.2: A round magnetic lens viewing from (a) the side and (b) the top. (b,d) The electron trajectories of (a) and (b).

is this velocity that focuses the electron beam as it passes through the lens, instead of the spiral movement mentioned above. The trace of the azimuthal velocity can be identified as an image rotation before and after the lens.

The focal length f of a thin electromagnetic lens is derived in various textbooks, and is given below;

$$\frac{1}{f} = \int \frac{e^2}{4\gamma^2 m^2 v_z^2} B^2 dz \tag{A.1}$$

in which, f is the focal length, B is the magnetic field, e and m are the electric charge and the mass of an electron, and $\gamma = \sqrt{1 - v_z^2/c^2}$ is the Lorentz factor. The approximation involved for a "thin" lens is that its length $L \ll v_z/\omega_L$, where ω_L is the azimuthal velocity of the electron which stays approximately constant over the path inside the lens. Eq. A.1 has the important implication that the focal length of a magnetic lens is always positive, meaning only "convex" electromagnetic lenses can be manufactured. The absence of "diverging" or negative lenses in electron optics is one reason why spherical aberration cannot be corrected in the same way as in light optics by simply applying a diverging lens after the convex lens.

Bibliography

L. J. Allen, W. McBride, N. L. O'Leary, and M. P. Oxley. Exit wave reconstruction at atomic resolution. *Ultramicroscopy*, 100(1-2):91–104, 2004.

E. Amraei and M. R. Mobasheri. Striping noise removal of images acquired by CBERS 2 CCD camera sensor. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-2/W3:53–58, 2014.

X.-c. Bai, G. McMullan, and S. H. W. Scheres. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1):49–57, Jan. 2015.

B. Barton, C. Song, and C. Kisielowski. Atomic resolution at 50 - 300 kV obtained using low dose rate HRTEM. *Microscopy and Microanalysis*, 17(S2):1266–1267, Oct. 2011.

A. M. Beale, F. Gao, I. Lezcano-Gonzalez, C. H. F. Peden, and J. Szanyi. Recent advances in automotive catalysis for NOx emission control by small-pore microporous materials. *Chem. Soc. Rev.*, 44(20):7371–7405, Oct. 2015.

V. D. Beck. Hexapole spherical-aberration corrector. *Optik*, 53(4):241–255, 1979.

B. Berkels, P. Binev, D. A. Blom, W. Dahmen, and R. C. Sharpley. Optimized imaging using non-rigid registration. *Ultramicroscopy*, 138:46–56, 2014.

A. Bleloch and Q. Ramasse. Lens aberrations: diagnosis and correction. In

Aberration-Corrected Analytical Transmission Electron Microscopy, pages 55–87. John Wiley & Sons, Ltd, Chichester, UK, 2011.

D. B. Carlson and J. E. Evans. Low-cost cryo-light microscopy stage fabrication for correlated light/electron microscopy. *Journal of Visualized Experiments*, (52): e2909–e2909, June 2011.

I. Celardo, J. Z. Pedersen, E. Traversa, and L. Ghibelli. Pharmacological potential of cerium oxide nanoparticles. *Nanoscale*, 3(4):1411–1420, Apr. 2011.

L.-Y. Chang, R. R. Meyer, and A. I. Kirkland. Calculations of HREM image intensity using Monte Carlo integration. *Ultramicroscopy*, 104(3-4):271–280, Oct. 2005.

L. Y. Chang, A. I. Kirkland, and J. M. Titchmarsh. On the importance of fifth-order spherical aberration for a fully corrected electron microscope. *Ultramicroscopy*, 106 (4-5):301–306, 2006.

Y. Cheng. Single-particle cryo-EM at crystallographic resolution. *Cell*, 161(3):450–457, Apr. 2015.

D. Cherns, F. J. Minter, and R. S. Nelson. Sputtering in the high voltage electron microscope. *Nuclear Instruments and Methods*, 132:369–376, Jan. 1976.

N. V. Choudary and B. L. Newalkar. Use of zeolites in petroleum refining and petrochemical processes: recent advances. *Journal of Porous Materials*, 18(6):685–692, 2011.

W. Coene, G. Janssen, M. Op de Beeck, and D. Van Dyck. Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy. *Physical review letters*, 69(26):3743–3746, Dec. 1992.

W. M. J. Coene, A. Thust, M. Op De Beeck, and D. Van Dyck. Maximum-likelihood

method for focus-variation image reconstruction in high resolution transmission electron microscopy. *Ultramicroscopy*, 64(1-4):109–135, Aug. 1996.

J. M. Cowley and A. P. Pogany. Diffuse scattering in electron diffraction patterns. I. General theory and computational methods. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 24(1):109–116, Jan. 1968.

P. A. Crozier, M. R. McCartney, and D. J. Smith. Observation of exit surface sputtering in TiO2 using biased secondary electron imaging. *Surface Science*, 237(1-3): 232–240, 1990.

I. Daberkow, K. H. Herrmann, L. Liu, and W. D. Rau. Performance of electron image converters with YAG single-crystal screen and CCD sensor. *Ultramicroscopy*, 38 (3-4):215–223, Dec. 1991.

R. L. Dalglish and J. C. Kelly. The application of Busch's theorem to magnetic lenses for beam focussing. *Physics Letters A*, 53(3):249–250, June 1975.

S. Das, J. M. Dowding, K. E. Klump, and J. F. McGinnis. Cerium oxide nanoparticles: applications and prospects in nanomedicine. *International Journal of Computer Vision*, 8(9):1483–1508, 2013.

M. O. de Beeck, D. Van Dyck, and W. Coene. Wave function reconstruction in HRTEM: the parabola method. *Ultramicroscopy*, 64(1-4):167–183, 1996.

M. De Graef. *Introduction to Conventional Transmission Electron Microscopy*. Cambridge University Press, Mar. 2003.

W. J. De Ruijter and J. K. Weiss. Methods to measure properties of slow-scan CCD cameras for electron detection. *Review of Scientific Instruments*, 63(10):4314–9, 1992.

A. J. den Dekker and A. van den Bos. Resolution: a survey. *Journal of the Optical Society of America A*, 14(3):547, 1997.

I. Dietrich, F. Fox, E. Knapek, G. Lefranc, K. Nachtrieb, R. Weyl, and H. Zerbst. Improvements in electron microscopy by application of superconductivity. *Ultramicroscopy*, 2:241–249, Jan. 1976.

D. L. Dorset and M. P. McCourt. Prospects for the direct electron crystallographic determination of zeolite structures. *Microscopy Research and Technique*, 36(3): 212–223, 1997.

J. Dubochet, J. J. Chang, R. Freeman, J. Lepault, and A. W. McDowall. Frozen aqueous suspensions. *Ultramicroscopy*, 10(1-2):55–61, Jan. 1982.

J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz. Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21(02):129–228, 1988.

E. M. H. Duke, M. Razi, A. Weston, P. Guttmann, S. Werner, K. Henzler, G. Schneider, S. A. Tooze, and L. M. Collinson. Imaging endosomes and autophagosomes in whole mammalian cells using correlative cryo-fluorescence and cryo-soft X-ray microscopy (cryo-CLXM). *Ultramicroscopy*, 143:77–87, Aug. 2014.

G. Dupouy. Advantages of megavolt electron microscopy in biological research. *Ultramicroscopy*, 2:199–203, Jan. 1976.

R. F. Egerton. *Electron Energy-Loss Spectroscopy in the Electron Microscope*. Springer Science & Business Media, Boston, MA, July 2011a.

R. F. Egerton. Measurement of radiation damage by electron energy-loss spectroscopy. *Journal of Microscopy*, 118(4):389–399, Aug. 2011b.

R. F. Egerton. Mechanisms of radiation damage in beam-sensitive specimens, for TEM accelerating voltages between 10 and 300 kV. *Microscopy Research and Technique*, 2012.

R. F. Egerton and M. Takeuchi. Radiation damage to fullerite (C60) in the transmission electron microscope. *Applied Physics Letters*, 75(13):1884–1886, 1999.

R. F. Egerton, P. A. Crozier, and P. Rice. Electron energy-loss spectroscopy and chemical change. *Ultramicroscopy*, 23(3):305–312, 1987.

R. F. Egerton, P. Li, and M. Malac. Radiation damage in the TEM and SEM. *Micron*, 35(6):399–409, 2004.

R. F. Egerton, R. McLeod, F. Wang, and M. Malac. Basic questions related to electron-induced sputtering in the TEM. *Ultramicroscopy*, 110(8):991–997, 2010.

E. Essers, G. Benner, T. Mandler, S. Meyer, D. Mittmann, M. Schnell, and R. Höschen. Energy resolution of an Omega-type monochromator and imaging properties of the MANDOLINE filter. *Ultramicroscopy*, 110:971–980, 2010.

J. E. Evans, C. Hetherington, A. I. Kirkland, L.-Y. Chang, H. Stahlberg, and N. Browning. Low-dose aberration corrected cryo-electron microscopy of organic specimens. *Ultramicroscopy*, 108(12):1636–1644, 2008.

J. A. Fernandez, V. N. Boddeti, and A. Rodriguez. Zero-aliasing correlation filters for object recognition. *CoRR*, abs/1411.2316, Dec. 2014.

H. Fernández-Morán. Low-temperature preparation techniques for electron microscopy of biological specimens based on rapid freezing with liquid helium II. *Annals of the New York Academy of Sciences*, 85(2):689–713, Apr. 1960.

J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15): 2758–2769, 1982.

J. Frank. Single-particle reconstruction of biological macromolecules in electron microscopy–30 years. *Quarterly Reviews of Biophysics*, 42(03):139, 2009.

R. Freeman and K. R. Leonard. Comparative mass measurement of biological macromolecules by scanning transmission electron microscopy. *Journal of Microscopy*, 122(3):275–286, June 1981.

J. R. Fryer. Radiation damage in organic crystalline films. *Ultramicroscopy*, 14(3): 227–236, 1984.

J. R. Fryer. The effect of dose rate on imaging aromatic organic crystals. *Ultramicroscopy*, 23(3):321–327, 1987.

J. R. Fryer and F. Holland. The reduction of radiation damage in the electron microscope. *Ultramicroscopy*, 11(1):67–70, 1983.

J. R. Fryer, C. McNee, and F. M. Holland. The further reduction of radiation damage in the electron microscope. *Ultramicroscopy*, 14(4):357–358, Jan. 1984.

R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237, 1972.

R. M. Glaeser. Limitations to significant information in biological electron microscopy as a result of radiation damage. *Ultramicroscopy*, 36(3-4):466–482, Aug. 1971.

R. M. Glaeser. *Electron crystallography of biological macromolecules*. Oxford University Press, USA, 2007.

R. M. Glaeser. Retrospective: radiation damage and its associated "information limitations". *Journal of Structural Biology*, 163(3):271–276, Sept. 2008.

M. A. Gribelyuk and J. L. Hutchison. On the iterative restoration of the exit plane wave function from defocus series in HREM. *Ultramicroscopy*, 45(1):127–143, 1992.

N. A. Grogin, P. L. Lim, A. Maybhate, R. N. Hook, and M. Loose. Post-SM4 ACS/WFC bias striping: characterization and mitigation. In S. Destua and C. Oliveira, editors, *2010 HST Calibration Workshop*. Space Telescope Science Institute, 2010.

R. W. Grosse-Kunstleve, L. B. McCusker, and C. Baerlocher. Zeolite structure determination from powder diffraction data: applications of the FOCUS method. *Journal of Applied Crystallography*, 32(3):536–542, June 1999.

M. Haider, S. Uhlemann, E. Schwan, H. Rose, and B. Kabius. Electron microscopy image enhanced. *Nature*, 392(6678):768–769, 1998.

M. Haider, P. Hartel, H. Müller, S. Uhlemann, and J. Zach. Current and future aberration correctors for the improvement of resolution in electron microscopy. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1903):3665–3682, Sept. 2009.

S. Haigh. *Super resolution tilt series exit wave restoration from aberration corrected images*. PhD thesis, University of Oxford, Oct. 2008.

S. J. Haigh and A. I. Kirkland. Finding phase information in the darkness. *Journal of Physics: Conference Series*, 241:012013, 2010.

F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

J. R. Harris and R. W. Horne. Negative staining: a brief assessment of current technical benefits, limitations and future possibilities. *Micron*, 25(1):5–13, Jan. 1994.

J. R. Harris, A. Plückthun, and R. Zahn. Transmission electron microscopy of GroEL,

GroES, and the symmetrical GroEL/ES complex. *Journal of Structural Biology*, 112(3):216–230, 1994.

P. W. Hawkes. The geometrical aberrations of general electron optical systems. I. The conditions imposed by symmetry. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 257(1086): 479–522, June 1965.

P. W. Hawkes. The correction of electron lens aberrations. *Ultramicroscopy*, 156: A1–A64, Sept. 2015.

L. W. Hobbs. Transmission electron microscopy of defects in alkali halides. *Journal de Physique Colloques*, 34:C9–227, 1973.

L. W. Hobbs. Radiation damage in electron microscopy of inorganic solids. *Ultramicroscopy*, 3(4):381–386, 1978.

L. W. Hobbs, A. E. Hughes, and D. Pooley. A study of interstitial clusters in irradiated alkali halides using direct electron microscopy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 332(1589):167–185, Mar. 1973.

X. Huang and J. Ximen. Discussions on the multiple-input maximum a-posteriori wave-function restoration method in high-resolution electron microscopy. *Journal of Electron Microscopy Technique*, 17(3):344–350, Mar. 1991.

B. R. Hunt. Bayesian methods in nonlinear digital image restoration. *IEEE Transactions on Computers*, C-26(3):219–229, Mar. 1977.

M. S. Isaacson. Inelastic scattering and beam damage in biological molecules . Wiely, New York, 1975.

K. Ishizuka. Phase retrieval from image intensities: Why does exit wave restoration using IWFR work so well? *Microscopy*, page dft005, Mar. 2013.

K. Ishizuka and B. Allman. Phase measurement in electron microscopy using the transport of intensity equation. *Microscopy Today*, pages 22–24, 2005.

M. L. Jenkins and M. A. Kirk. *Characterisation of Radiation Damage by Transmission Electron Microscopy*. CRC Press, Nov. 2000.

N. Jiang and J. Spence. On the dose-rate threshold of beam damage in TEM. *Ultramicroscopy*, 113:77–82, 2012.

R. Kanno, A. Koike-Takeshita, K. Yokoyama, H. Taguchi, and K. Mitsuoka. Cryo-EM structure of the native GroEL-GroES complex from thermus thermophilus encapsulating substrate inside the cavity. *Structure*, 17(2):287–293, Feb. 2009.

M. Karuppasamy, F. Karimi Nejadasl, M. Vulovic, A. J. Koster, and R. B. G. Ravelli. Radiation damage in single-particle cryo-electron microscopy: effects of dose and dose rate. *Journal of Synchrotron Radiation*, 18(3):398–412, Apr. 2011.

J. Kašpar, P. Fornasiero, and M. Graziani. Use of CeO_2-based oxides in the three-way catalysis. *Catalysis Today*, 50(2):285–298, Apr. 1999.

A. Kirkland, J. Warner, J. S. Kim, and P. Nellist. The design and performance of a double Wien filter monochromator for application in TEM. *Microscopy and Microanalysis*, 19:310–311, 2013.

A. I. Kirkland and R. R. Meyer. "indirect" high-resolution transmission electron microscopy: aberration measurement and wavefunction reconstruction. *Microscopy and Microanalysis*, 10(04):401–413, 2004.

A. I. Kirkland, W. O. Saxton, K. Chau, K. Tsuno, and M. Kawasaki. Super-resolution by aperture synthesis: tilt series reconstruction in CTEM. *Ultramicroscopy*, 57(4): 355–374, 1995.

A. I. Kirkland, W. O. Saxton, and G. Chand. Multiple beam tilt microscopy for super resolved imaging. *Journal of Electron Microscopy*, 46(1):11–22, 1997.

A. I. Kirkland, S. L. Y. Chang, and J. L. Hutchison. Atomic resolution transmission electron microscopy. In P. W. Hawkes and J. C. H. Spence, editors, *Science of Microscopy*, pages 3–64. Springer New York, New York, NY, 2007.

E. J. Kirkland. Improved high resolution image processing of bright field electron micrographs: I. Theory. *Ultramicroscopy*, 15(3):151–172, 1984.

E. J. Kirkland, B. M. Siegel, N. Uyeda, and Y. Fujiyoshi. Improved high resolution image processing of bright field electron micrographs: II. Experiment. *Ultramicroscopy*, 17(2):87–103, Jan. 1985.

M. Klingstedt. *Characterizing cavity containing materials using electron microscopy : A study of metal oxides, mesoporous crystals and porous material containing nanosized metal-particles*. PhD thesis, Department of Materials and Environmental Chemistry (MMK), Stockholm University, Stockholm, 2011.

L. F. Kourkoutis, J. M. Plitzko, and W. Baumeister. Electron microscopy of biological materials at the nanometer scale. *Annual Review of Materials Research*, 42(1):33–58, Aug. 2012.

C. Kübel and A. Thust. TrueImage: a software package for focal-series reconstruction in HRTEM . URL `http://www.totalresolution.com/download/ReconstructionManual.pdf`.

M. Liao, E. Cao, D. Julius, and Y. Cheng. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107–112, Dec. 2013.

S. D. Lin. Electron radiation damage of thin films of glycine, diglycine, and aromatic amino acids. *Radiation Research*, 59(3):521, Sept. 1974.

H. Liu, L. Jin, S. B. S. Koh, I. Atanasov, S. Schein, L. Wu, and Z. H. Zhou. Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science*, 329(5995):1038–1043, Aug. 2010.

I. Lobato and D. Van Dyck. MULTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA. *Ultramicroscopy*, 156:9–17, Sept. 2015.

S. J. Ludtke, D.-H. Chen, J.-L. Song, D. T. Chuang, and W. Chiu. Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. *Structure*, 12(7):1129–1136, July 2004.

V. S. Marakatti and A. B. Halgeri. Metal ion-exchanged zeolites as highly active solid acid catalysts for the green synthesis of glycerol carbonate from glycerol. *RSC Advances*, 5(19):14286–14293, Jan. 2015.

V. S. Marakatti, A. B. Halgeri, and G. V. Shanbhag. Metal ion-exchanged zeolites as solid acid catalysts for the green synthesis of nopol from Prins reaction. *Catalysis Science & Technology*, 4(11):4065–4074, Oct. 2014.

G. Martínez and K. Tsuno. The use of multipole fields for aberration correction in $\pi/2$ Wien filters. *Physics Procedia*, 1(1):193–198, Aug. 2008.

H. T. Meryman. Mechanics of freezing in living cells and tissues. *Science*, 124(3221): 515–521, Sept. 1956.

R. R. Meyer and A. I. Kirkland. The effects of electron and photon scattering on signal and noise transfer properties of scintillators in CCD cameras used for electron detection. *Ultramicroscopy*, 75(1):23–33, Oct. 1998.

R. R. Meyer, A. I. Kirkland, R. E. Dunin-Borkowski, and J. L. Hutchison. Experi-

mental characterisation of CCD cameras for HREM at 300kV. *Ultramicroscopy*, 85 (1):9–13, Sept. 2000.

R. R. Meyer, A. I. Kirkland, and W. O. Saxton. A new method for the determination of the wave aberration function for high resolution TEM: 1. Measurement of the symmetric aberrations. *Ultramicroscopy*, 92(2):89–109, 2002.

R. R. Meyer, A. I. Kirkland, and W. O. Saxton. A new method for the determination of the wave aberration function for high-resolution TEM.: 2. Measurement of the antisymmetric aberrations. *Ultramicroscopy*, 99(2-3):115–123, 2004.

M. A. O. Miedema, A. van den Bos, and A. H. Buist. Experimental design of exit wave reconstruction from a transmission electron microscope defocus series. *Instrumentation and Measurement, IEEE Transactions on*, 43(2):181–186, Apr. 1994.

A.-C. Milazzo, G. Moldovan, J. Lanman, L. Jin, J. C. Bouwer, S. Klienfelder, S. T. Peltier, M. H. Ellisman, A. I. Kirkland, and N.-H. Xuong. Characterization of a direct detection device imaging camera for transmission electron microscopy. *Ultramicroscopy*, 110(7):741–744, June 2010.

J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, Oxford, Dec. 2003.

S. Nicolopoulos, J. M. Gonzalez-Calbet, M. Vallet-Regi, A. Corma, C. Corell, J. M. Guil, and J. Perez-Pariente. Direct Phasing in Electron Crystallography: Ab Initio Determination of a New MCM-22 Zeolite Structure. *Journal of the American Chemical Society*, 117(35):8947–8956, Sept. 1995.

M. Niwa, N. Katada, and K. Okumura. *Characterization and Design of Zeolite Catalysts*. Solid Acidity, Shape Selectivity and Loading Properties. Springer Science & Business Media, Aug. 2010.

H. Nyquist. Certain topics in telegraph transmission theory. In *Winter Convention of the A. I. E. E.*, pages 617–644, New York, NY, Feb. 1928.

T. Ohsuna, B. Slater, F. Gao, J. Yu, Y. Sakamoto, G. Zhu, O. Terasaki, D. E. W. Vaughan, S. Qiu, and C. R. A. Catlow. Fine structures of zeolite-linde-L (LTL): surface structures, growth unit and defects. *Chemistry - A European Journal*, 10 (20):5031–5040, Oct. 2004.

M. Op de Beeck and D. Van Dyck. Direct structure reconstruction in HRTEM. *Ultramicroscopy*, 64(1-4):153–165, Aug. 1996.

J. Orloff. *Handbook of Charged Particle Optics, Second Edition*. CRC Press, Oct. 2008.

V. Ortalan, A. Uzun, B. C. Gates, and N. D. Browning. Direct imaging of single metal atoms and clusters in the pores of dealuminated HY zeolite. *Nature Nanotechnology*, 5(7):506–510, May 2010.

M. Palomino, A. Corma, J. L. Jordá, F. Rey, and S. Valencia. Zeolite Rho: a highly selective adsorbent for CO2/CH4 separation induced by a structural phase modification. *Chem. Commun.*, 48(2):215–217, Nov. 2011.

M. Pan. High resolution electron microscopy of zeolites. *Micron*, 27(3-4):219–238, June 1996.

M. Pan and P. Crozier. Low-dose high-resolution electron microscopy of zeolite materials with a slow-scan CCD camera. *Ultramicroscopy*, 48(3):332–340, 1993.

W. Probst, G. Benner, J. Bihr, and E. Weimer. An "omega" energy filtering TEM — principles and applications. *Advanced Materials*, 5(4):297–300, Apr. 1993.

K. Ramamurti, A. V. Crewe, and M. S. Isaacson. Low temperature mass loss of thin

films of l-phenylalanine and l-tryptophan upon electron irradiation — A preliminary report. *Ultramicroscopy*, 1(2):156–158, Jan. 1975.

N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, and H. R. Saibil. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, 107(7):869–879, Dec. 2001.

L. Reimer. Methods of detection of radiation damage in electron microscopy. *Ultramicroscopy*, 14(3):291–303, Jan. 1984.

L. Reimer. *Energy-Filtering Transmission Electron Microscopy*, volume 71 of *Springer Series in Optical Sciences*. Springer, Berlin, Heidelberg, June 2013.

L. Reimer and H. Kohl. *Transmission Electron Microscopy*. Physics of Image Formation. Springer Science & Business Media, Aug. 2008.

J. Reyes-Gasga, E. L. Martínez-Piñeiro, and É. F. Brès. Crystallographic structure of human tooth enamel by electron microscopy and x-ray diffraction: hexagonal or monoclinic? *Journal of Microscopy*, 248(1):102–109, 2012.

S. B. Rice, J. Y. Koo, M. M. Disko, and M. M. J. Treacy. On the imaging of Pt atoms in zeolite frameworks. *Ultramicroscopy*, 34(1-2):108–118, Nov. 1990.

A. Rigort, F. J. B. Bäuerlein, A. Leis, M. Gruska, C. Hoffmann, T. Laugks, U. Böhm, M. Eibauer, H. Gnaegi, W. Baumeister, and J. M. Plitzko. Micromachining tools and correlative approaches for cellular cryo-electron tomography. *Journal of Structural Biology*, 172(2):169–179, Nov. 2010.

H. Rose. Correction of aperture aberrations in magnetic systems with threefold symmetry. *Nuclear Instruments and Methods in Physics Research*, 187(1):187–199, Aug. 1981.

H. Rose. Prospects for aberration-free electron microscopy. *Ultramicroscopy*, 103(1): 1–6, Apr. 2005.

H. Rose and W. Wan. Aberration correction in electron microscopy. In *2005 Particle Accelerator Conference*, pages 44–48, Knoxville, Tennessee, 2005. IEEE.

H. H. Rose. Optics of high-performance electron microscopes. *Science and Technology of Advanced Materials*, 9(1):014107, Apr. 2008.

W. Saxton and J. Frank. Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy*, 2(2-3):219, 1977.

W. O. Saxton. Accurate atom positions from focal and tilted beam series of high resolution electron micrographs. In *Scanning Microscopy International*, pages 213–224, Chicago, IL, 1988. Scanning Microsc.

W. O. Saxton. What is the focus variation method? Is it new? Is it direct? *Ultramicroscopy*, 55(2):171–181, Aug. 1994a.

W. O. Saxton. Accurate alignment of sets of images. *Journal of Microscopy*, 174(2): 61–68, 1994b.

W. O. Saxton. Observation of lens aberrations for very high-resolution electron microscopy. I. Theory. *Journal of Microscopy*, 179(2):201–213, Aug. 1995.

W. O. Saxton. Observation of lens aberrations for high resolution electron microscopy II: Simple expressions for optimal estimates. *Ultramicroscopy*, 2015.

W. O. Saxton and D. J. Smith. The determination of atomic positions in high-resolution electron micrographs. *Ultramicroscopy*, 18(1-4):39–47, Jan. 1985.

W. O. Saxton, T. J. Pitt, and M. Horner. Digital image processing: The semper system. *Ultramicroscopy*, 4(3):343–353, Jan. 1979.

T. Schäfer, P. Michel, F. Claret, T. Beetz, S. Wirick, and C. Jacobsen. Radiation sensitivity of natural organic matter: Clay mineral association effects in the Callovo-Oxfordian argillite. *Journal of Electron Spectroscopy and Related Phenomena*, 170 (1-3):49–56, Mar. 2009.

O. Scherzer. Über einige Fehler von Elektronenlinsen. *Zeitschrift für Physik*, 101 (9-10):593–603, 1936.

O. Scherzer. The theoretical resolution limit of the electron microscope. *Journal of Applied Physics*, 20(1):20, 1949.

P. Schiske. Zur frage der bildrekonstruktion durch fokusreihen. In *Proc. Eur. Reg. Conf. Electron. Microsc., 4th*, page 145. Proc. Eur. Reg. Conf. Electron. Microsc., 1968.

C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37 (1):10–21, 1949.

K. Siangchaew and M. Libera. The influence of fast secondary electrons on the aromatic structure of polystyrene. *Philosophical Magazine A*, 80(4):1001–1016, Aug. 2009.

P. Sliz, S. C. Harrison, and G. Rosenbaum. How does radiation damage in protein crystals depend on X-Ray dose? *Structure*, 11(1):13–19, Jan. 2003.

J. C. H. Spence. *High-Resolution Electron Microscopy*. Oxford University Press, Oxford, Sept. 2013.

A. B. Stambouli and E. Traversa. Solid oxide fuel cells (SOFCs): a review of an environmentally clean and efficient source of energy. *Renewable and Sustainable Energy Reviews*, 6(5):433–455, Oct. 2002.

A. Stevens, L. Kovarik, P. Abellan, X. Yuan, L. Carin, and N. D. Browning. Applying compressive sensing to TEM video: a substantial frame rate increase on any camera. *Advanced Structural and Chemical Imaging*, 1(1):10, Aug. 2015.

M. Stewart and G. Vigers. Electron microscopy of frozen-hydrated biological material. *Nature*, 319:631–636, Feb. 1986.

K. A. Taylor and R. Glaeser. Electron diffraction of frozen, hydrated protein crystals. *Science*, 186(4168):1036–1037, Dec. 1974.

K. A. Taylor and R. M. Glaeser. Electron microscopy of frozen hydrated biological specimens. *Journal of Ultrastructure Research*, 55:448–456, 1976.

O. Terasaki, T. Ohsuna, V. Alfredson, and J. O. Bovin. The study of zeolites by HVHREM. *Ultramicroscopy*, 1991.

O. Terasaki, T. Ohsuna, N. Ohnishi, and K. Hiraga. Zeolites and related materials studied by electron microscopy. *Current Opinion In Solid State & Materials Science*, 2(1):94–100, 1997.

J. M. Thomas, O. Terasaki, P. L. Gai, W. Zhou, and J. Gonzalez-Calbet. Structural elucidation of microporous and mesoporous catalysts and molecular sieves by high-resolution electron microscopy. *Accounts of Chemical Research*, 34(7):583–594, July 2001.

L. E. Thomas. Light-element analysis with electrons and X-rays in a high-resolution STEM. *Ultramicroscopy*, 18(1-4):173–184, 1985.

A. Thust. High-resolution transmission electron microscopy on an absolute contrast scale. *Physical review letters*, 102(22):220801, 2009.

A. Thust, W. Coene, M. Op De Beeck, and D. Van Dyck. Focal-series reconstruction

in HRTEM: simulation studies on non-periodic objects. *Ultramicroscopy*, 64(1): 211–230, 1996a.

A. Thust, M. Overwijk, W. Coene, and M. Lentzen. Numerical correction of lens aberrations in phase-retrieval HRTEM. *Ultramicroscopy*, 64(1):249–264, 1996b.

P. C. Tiemeijer, M. Bischoff, B. Freitag, and C. Kisielowski. Using a monochromator to improve the resolution in TEM to below 0.5Å. Part II: Application to focal series reconstruction. *Ultramicroscopy*, 118:35–43, July 2012a.

P. C. Tiemeijer, M. Bischoff, B. Freitag, and C. Kisielowski. Using a monochromator to improve the resolution in TEM to below 0.5Å. Part I: Creating highly coherent monochromated illumination. *Ultramicroscopy*, 114:72–81, Mar. 2012b.

M. M. J. Treacy and J. M. Newsam. Electron beam sensitivity of zeolite L. *Ultramicroscopy*, 23(3-4):411–419, Jan. 1987.

A. Trovarelli. Catalytic properties of ceria and CeO_2-containing materials. *Catalysis Reviews-Science and Engineering*, 38(4):439–520, 1996.

H. J. Trussell and B. R. Hunt. Improved methods of maximum a posteriori restoration. *IEEE Transactions on Computers*, C-28(1):57–62, 1979.

S. Uhlemann and M. Haider. Residual wave aberrations in the first spherical aberration corrected transmission electron microscope. *Ultramicroscopy*, 72(3-4):109–119, 1998.

K. W. Urban. Studying atomic structures by aberration-corrected transmission electron microscopy. *Science*, 321(5888):506–510, 2008.

K. W. Urban, C. L. Jia, L. Houben, M. Lentzen, S. B. Mi, and K. Tillmann. Negative spherical aberration ultrahigh-resolution imaging in corrected transmission electron

microscopy. *Philosophical Transactions of The Royal Society A-Mathematical Physical And Engineering Sciences*, 367(1903):3735–3753, Aug. 2009.

S. Van Aert, D. Van Dyck, and A. J. den Dekker. Resolution of coherent and incoherent imaging systems reconsidered - Classical criteria and a statistical alternative. *Optics Express*, 14(9):3830–3839, May 2006.

S. Van Aert, J. H. Chen, and D. Van Dyck. Linear versus non-linear structural information limit in high-resolution transmission electron microscopy. *Ultramicroscopy*, 110(11):1404–1410, 2010.

D. Van Dyck, J. R. Jinschek, and F. R. Chen. 'Big Bang' tomography as a new route to atomic-resolution electron tomography. *Nature*, 486(7402):243–246, 2012.

V. A. J. Van Lint. The physics of radiation damage in particle detectors. *Nuclear Instruments and Methods in Physics Research*, A253:453–459, 1987.

K. Varlot, J. M. Martin, C. Quet, and Y. Kihn. Towards sub-nanometer scale EELS analysis of polymers in the TEM. *Ultramicroscopy*, 68(2):123–133, June 1997.

S. Wang and Y. Peng. Natural zeolites as effective adsorbents in water and wastewater treatment. *Chemical Engineering Journal*, 156(1):11–24, Jan. 2010.

M. L. Watson. Staining of tissue sections for electron microscopy with heavy metals. *The Journal of Cell Biology*, 4(4):475–478, July 1958.

N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*. The MIT Press, 1949.

D. B. Williams and C. B. Carter. *Transmission Electron Microscopy*. A Textbook for Materials Science. Springer Science & Business Media, Boston, MA, Aug. 2009.

C. Xu and X. Qu. Cerium oxide nanoparticle: a remarkably versatile rare earth nanomaterial for biological applications. *NPG Asia Materials*, 6(3):e90, 2014.

T. Yanagisawa. High-resolution electron microscopy of enamel-crystal demineralization and remineralization in carious lesions. *Journal of Electron Microscopy*, 52(6): 605–613, Dec. 2003.

K. Yoshida and Y. Sasaki. Optimal accelerating voltage for HRTEM imaging of zeolite. *Microscopy*, 62(3):369–375, 2013.

K. Yoshida, Y. Sasaki, and H. Kurata. High-resolution imaging of zeolite with aberration-corrected transmission electron microscopy. *AIP Advances*, 3(4):042113, 2013.

H. W. Zandbergen and D. Van Dyck. Exit wave reconstructions using through focus series of HREM images. *Microscopy Research and Technique*, 49(3):301–323, 2000.

F. Zemlin. A practical procedure for alignment of a high resolution electron microscope. *Ultramicroscopy*, 4(2):241–245, Jan. 1979.

F. Zemlin, K. Weiss, P. Schiske, W. Kunath, and K. H. Herrmann. Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy*, 3:49–60, Jan. 1978.

X. Zhang, L. Jin, Q. Fang, W. H. Hui, and Z. H. Zhou. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell*, 141(3): 472–482, Apr. 2010.

A. Ziegler, H. Graafsma, X. F. Zhang, and J. W. M. Frenken. *In-situ Materials Characterization*, volume 193 of *Across Spatial and Temporal Scales*. Springer Science & Business Media, Berlin, Heidelberg, Apr. 2014.

B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

J. M. Zuo. Electron detection characteristics of slow-scan CCD camera. *Ultramicroscopy*, 66(1-2):21–33, Nov. 1996.