

**Modeling of Image Formation
in
Cryo-Electron Microscopy**

Miloš Vulović

Modeling of Image Formation
in
Cryo-Electron Microscopy

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 13 juni 2013 om 12.30 uur
door

Miloš VULOVIĆ

Diplomirani inženjer elektrotehnike, Univerzitet u Beogradu
geboren te Belgrado, Servië

Dit proefschrift is goedgekeurd door de promotoren:

Prof.dr.ir. L.J. van Vliet

Prof.dr.ir. A.J. Koster

Copromotoren:

Dr. B. Rieger

Dr. R.B.G. Ravelli

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter

Prof.dr.ir. L.J. van Vliet, Technische Universiteit Delft, promotor

Prof.dr.ir. A.J. Koster, Leids Universitair Medisch Centrum, promotor

Dr. B. Rieger, Technische Universiteit Delft, copromotor

Dr. R.B.G. Ravelli, Leids Universitair Medisch Centrum, copromotor

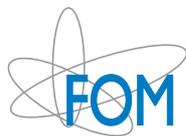
Prof.dr. D. van Dyck, University of Antwerp

Prof.dr.ir. P. Kruit, Technische Universiteit Delft

Dr. R.H.M. Schoenmakers, FEI Company

Dr. H. Friedrich, Technische Universiteit Eindhoven

Prof.dr. I.T. Young, Technische Universiteit Delft, reservelid



This work is part of the research programme *Microscopy and modification of nano-structures with focused electron and ion beams* (MMN) of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). The MMN programme is co-financed by FEI Company.

Cover design by Jelena Popadić

An electronic version of this dissertation is available at <http://repository.tudelft.nl/>.

ISBN 978-90-6464-679-9

Copyright © 2013 by Miloš Vulović

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage or retrieval system without permission from the author.

to H.U.B.I.

Contents

Contents	i
1 Introduction	1
1.1 Structural biology and electron microscopy (EM)	1
1.2 Cryo-EM	3
1.3 Phase contrast	5
1.4 Limiting factors in cryo-EM	6
1.5 Thesis challenges	8
1.5.1 Forward model	8
1.5.2 Detector	9
1.5.3 CTF	10
1.5.4 Thesis objectives	11
1.6 Thesis outline	12
2 Detector characterization	15
2.1 Introduction	16
2.2 Detector characterization	17
2.3 Measurement methods	21
2.3.1 Removal of outliers	21
2.3.2 Bias, dark current and readout noise	21
2.3.3 Detector effective gain measurements	22
2.3.4 Bias correction, gain normalization and pixel response	22
2.3.5 Modulation transfer function (MTF)	23
2.3.6 Conversion factor and detective quantum efficiency (DQE)	23
2.4 Results	25
2.4.1 Outliers	25
2.4.2 Bias and dark current	25
2.4.3 Readout noise	26
2.4.4 Lookup tables	27
2.4.5 Linearity of the response	27
2.4.6 MTF	28
2.4.7 Conversion factor and DQE	29
2.5 Discussion	29

2.6	Conclusion	34
3	Defocus and astigmatism estimation	35
3.1	Introduction	36
3.2	Theory	38
3.2.1	Phase contrast	38
3.2.2	Partial coherence and amplitude contrast	39
3.2.3	Weak-phase weak-amplitude object	40
3.2.4	Detector response	40
3.2.5	Power spectrum density and ellipticity of Thon rings due to the astigmatism	41
3.3	The algorithm	42
3.3.1	Power spectrum density processing	44
3.3.2	Polar representation	45
3.3.3	Template generation and template matching	45
3.3.4	Outlier rejection	48
3.3.5	Defocus and astigmatism estimation	48
3.3.6	Influence of spherical aberration C_s on the shape and frequency of Thon rings	49
3.4	Results	52
3.4.1	Validation by simulations	52
3.4.2	Results from measurements	56
3.4.3	Thon ring assessment	57
3.5	Discussion and conclusions	60
	Appendices	67
	A Weak-phase weak-amplitude approximation	67
	B Templates	69
B.1	Derivation of the template ellipticity R_t	69
B.2	The difference between detected q_{found} and central frequency q_c	70
	C Thon ring outlier rejection	71
	D Spherical aberration influence	72
D.1	C_s influence on the ellipticity	72
D.2	C_s influence on CTF minima position $q_{c,i}$	73
D.3	Correction for the C_s influence on the ring ellipticities	74
	E Thon ring averaging	75

4	Forward model	77
4.1	Introduction	77
4.2	Theory	79
4.2.1	Interaction potential (IP)	79
4.2.2	Electron wave propagation	81
4.2.3	Detector response	82
4.3	Computational methods	82
4.3.1	Interaction potential (IP)	82
4.3.2	Electron wave propagation and intensity detection	86
4.3.3	Parameter estimations	86
4.4	Experimental methods	87
4.4.1	Sample preparation	87
4.4.2	Image acquisitions/data collection	87
4.5	Results	88
4.5.1	“Bond” contributions	88
4.5.2	Defocus series	90
4.5.3	Integrated flux series and motion factor	90
4.5.4	Inelastic contributions	91
4.5.5	Camera’s DQE	92
4.5.6	Acceleration voltage influence	92
4.5.7	Amorphousness of the solvent	93
4.6	Discussion	94
4.6.1	Forward model	94
4.6.2	Poisson-Boltzmann (PB) approach	96
4.6.3	Beam-induced movements	97
4.6.4	Validation	98
4.6.5	Outlook	99
4.7	Conclusions	100
	Appendices	101
F	Theory of image formation in cryo-EM	101
F.1	Interaction potential (IP)	101
F.1.1	One-body, stationary Schrödinger equation	101
F.1.2	Isolated atom superposition approximation (IASA)	108
F.1.3	The Poisson-Boltzmann (PB) approach for modeling the electrostatic interactions between the atoms in a dielectric	110
F.1.4	Combining IASA- and PB -based potentials	114
F.2	High-energy electron-specimen interaction	116
F.2.1	Small angle approximation	116
F.2.2	Multislice method, projection assumption, weak-phase object approximation	117
F.2.3	Inelastic interactions	118

F.3	Optical system	122
F.4	Detector response	123
F.5	Fresnel diffraction from a phase step	126
F.6	Evaluation of the forward model on images of carbon edges and carbon nanotubes	128
F.6.1	Modeling carbon edges and carbon nanotubes	128
F.6.2	Experimental methods	129
F.6.3	Results	130
G	Projection assumption and weak-phase object approximation in cryo-EM	133
G.1	Introduction	133
G.2	High-energy electron and specimen interaction	134
G.2.1	Small angle approximation	135
G.3	Bounds to projection assumption and weak-phase object approximation	135
G.3.1	Projection assumption	136
G.3.2	Projection assumption and weak-phase object approximation	136
G.3.3	Weak-phase object approximation	137
G.3.4	Thick-phase grating approximation	137
G.4	Results	138
G.4.1	Hemoglobin	138
G.4.2	Exit waves of a tubulin tetramer	139
G.4.3	Synthetic amorphous test specimen	140
G.5	Discussion	142
5	Radiation damage	145
5.1	Introduction	146
5.2	Methods	149
5.2.1	Experimental methods	149
5.2.2	Computational methods	150
5.3	Results	153
5.3.1	Dose	153
5.3.2	Averaged intensity <i>versus</i> dose	154
5.3.3	Radiation damage series of Hemoglobin followed up to 5500 MGy	154
5.3.4	Defocus variation	156
5.3.5	Beam-induced movements of gold particles	156
5.3.6	Fourier ring correlation, Fourier ring phase residual and FOM plots	157
5.3.7	Dose-rate effects	158
5.3.8	Heating effects	159
5.3.9	The role of solvent constituents	160
5.4	Discussion	160
5.4.1	Dose	160
5.4.2	Dose-rate effect	162
5.5	Conclusions	165

Appendices	169
H Calculation of dose	169
6 Conclusions and perspectives	171
6.1 Recommendations	176
6.2 Outlook	178
References	181
Summary	197
Samenvatting	201
List of publications	205
Acknowledgments	207
Biography	211

Chapter 1

Introduction

1.1 Structural biology and electron microscopy (EM)

In order to understand the function of a living organism from a macroscopic scale (meter) down to atomic resolution (Ångström), roughly ten orders of magnitude must be mastered. Only after the invention of the microscope has it become possible to visualize and investigate the *microcosm* of the cell. Antonie van Leeuwenhoek, a Dutchman from Delft, became famous for his microscopes and discoveries with them in the 17th century and was one of the first to observe small unicellular organisms, which he named *animalcules*. In the early 19th century, cells were recognized as *building blocks of life*, establishing the field of cell biology.

The minimal separation between two resolvable objects in a light microscope is around 0.2-0.4 μm . This limit is directly related to the wavelength of light. Due to this resolution restriction, light microscopy cannot resolve various subcellular organelles, proteins, or viruses. In the 1930s, Ernst Ruska constructed the first transmission electron microscope (TEM), allowing researchers to investigate, among others, the cell to a resolution of a few nanometers. Electron microscopy (EM) provides superior resolving power due to the much shorter wavelength of high-energy electrons (2-5 pm), as compared to visible light (400-700 nm). However, largely due to aberrations and a small numerical aperture, even state-of-the-art electron microscopes have a resolving power in the range of 50 pm. This is more than sufficient for atomic resolution imaging of biological objects since a typical atomic radius is around 1 Å (0.1 nm). Although in materials science research, the atomic resolution is attained on certain specimens, a number of factors limit the resolution in EM of biological specimens to typically 4 – 6 Å.

In order to better understand life processes, *how* the various components within living organisms interact, and *what* is their function, knowledge of the structure of biological objects at all scales is essential [1]. In many cases, structural information complements biochemical studies and it allows validation of existing and formulation of new hypotheses on how structures interact. After the successful mapping of entire genomes of multiple species over the last decade, the challenge remains to understand how these genetic sequences relate to the wide variety of structures and how these structures undergo conformational changes when interacting with other structures. A systematic structural analysis of proteins, protein-ligand interactions as well as protein complexes (“Structural Proteomics” [1]) will become increasingly important [2].

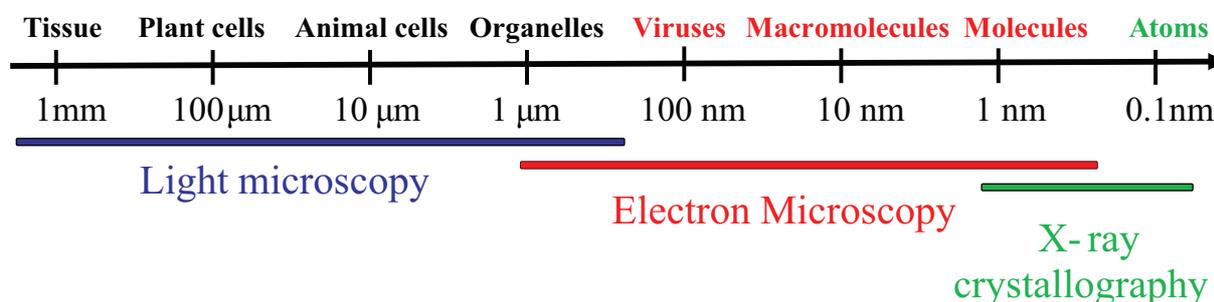


Fig. 1.1. The approximate range of biological structures covered by various techniques.

Macromolecules can adopt multiple configurations which are often crucial for their function. Therefore, 3D structure can provide insight into the complex biological processes at the cellular level and it allows the design of drugs that interfere with the action of a protein implicated in a disease [3]. Resolving macromolecular structures such as proteins, membranes and DNA at the highest possible resolution is both experimentally and computationally demanding and many Nobel prizes have resulted from the impact the developed tools and resolved structures have made on our understanding of biology (e.g. [4–7]).

The majority of the 3D atomic structures deposited in the RCSB Protein Data Bank (PDB) archive have been obtained by X-ray crystallography (over 80 %), nuclear magnetic resonance (NMR) spectroscopy (around 16%), and *via* theoretical modeling such as homology modeling (around 2%). X-ray crystallography has been able to analyze both small proteins and large complexes. The technique requires samples to be so pure and stable, that they can form a crystalline state. Traditionally, NMR structure determination has been limited to smaller macromolecules. A large number of structures cannot be resolved using X-ray and NMR. EM does not require crystals and it has become an increasingly powerful tool for structure determination. EM is particularly useful in studies where the aim is to distinguish conformations of molecular assemblies [8] as well as structural studies of large macromolecular complexes and their context within a whole cell [9]. However, with the exception of a few specific systems, EM typically does not provide atomic-resolution data for biological specimens. It is often necessary to incorporate information from X-ray crystallography or NMR to interpret detailed interactions. The complete structure of a macromolecular complex can be obtained at lower resolution using EM and subsequently used for model building based on docking of high-resolution components (obtained from X-ray crystallography or NMR). In this way, the atomic model of large structures can be obtained. As electrons interact with a specimen through Coulomb forces, EM effectively images a potential map. X-ray crystallography, however, obtains a map of the electron density. The incident electrons interact a few hundred times stronger with matter than X-rays or neutrons [10]. This high sensitivity combined with the availability of electron optics, high-coherence sources, and detectors has made EM an essential tool for studying properties of matter. Unfortunately, the high interaction efficiency in EM can also result in significant radiation damage of the specimen.

The resolution obtained by EM lies between those of light microscopy and X-ray crystal-

lography (see Fig. 1.1). One of the benefits of light microscopy is the possibility to image the dynamics of cellular assemblies *in vivo*. The achievable resolution, however, is not sufficient to distinguish individual macromolecular complexes that define many cellular functions. Super-resolution microscopy techniques such as STED [11], PALM [12] and STORM [13, 14] as well as correlative light and electron microscopy (CLEM) approaches [15] are increasingly applied to tackle these challenges. However, in case of STED, PALM and STORM this only effects the localization accuracy and will never result in a 3D structural model of the entire region of interest. In CLEM, fluorescence microscopy is used to navigate within cells or tissue, after which EM provides the ultrastructural information highlighted by the fluorescence. EM is therefore an essential tool in structural biology with the unique possibility of bridging the gap between cellular and molecular biology.

1.2 Cryo-EM

Since electrons also scatter from air molecules, the interior of an electron microscope has to be kept under high vacuum. Most biological specimens are incompatible with vacuum and therefore need to be immobilized (fixed) or dried prior to imaging. Cryo-electron microscopy (Cryo-EM) is a technique in which the specimen is vitrified and kept frozen-hydrated at liquid nitrogen (or helium) temperatures both during specimen preparation and imaging. At those temperatures the sublimation of the cryoimmobilized sample is negligible. The cryoimmobilization (fixation) is achieved by freezing the specimen at an extremely fast cooling rate by *plunge freezing* in liquid ethane, or, alternatively, through *high-pressure freezing*. The ultrarapid cooling prevents the formation of crystalline ice, resulting in an amorphous, vitreous ice which is, at the right thickness, transparent to electrons. During this phase transition, the vitreous ice volume expands which could cause stress to the biomolecules. Before cryo-EM was introduced [16, 17], the most common sample preparation protocol included staining with heavy metal salts. The high atomic numbers of the atoms in the salts compared to the light atoms of the organic material provide high contrast (see Fig. 1.2). This negative (i.e. contrast is produced by the stain, not by the structure of interest) staining approach has both advantages and disadvantages. Some disadvantages are that the staining step can lead to a deformation (flattening) of the specimen and that only surfaces accessible to the stain can be visualized. Negative staining is a good approach for early stages of molecular characterization and specimen preparation optimization as well as for imaging very small structures (e.g. smaller than 150 kDa) for which the contrast produced by cryo-EM is not sufficient. Advantages of cryo-EM include the excellent preservation of the molecular structure and the fact that at low temperatures, biological specimens are less vulnerable to radiation damage [18]. The motion of beam-induced radiolysis products are slowed down at low temperatures, reducing secondary damage. The main disadvantage of cryo-EM is the low contrast that originates from: 1) the small difference between scattering properties of a protein and the vitreous ice and 2) the high fraction of inelastic scattering events in materials with low atomic number such as vitreous ice.

Although TEM only generates 2D projections of the electrostatic potential of the specimen, the 3D volume can be reconstructed if many projections are attainable, each displaying the

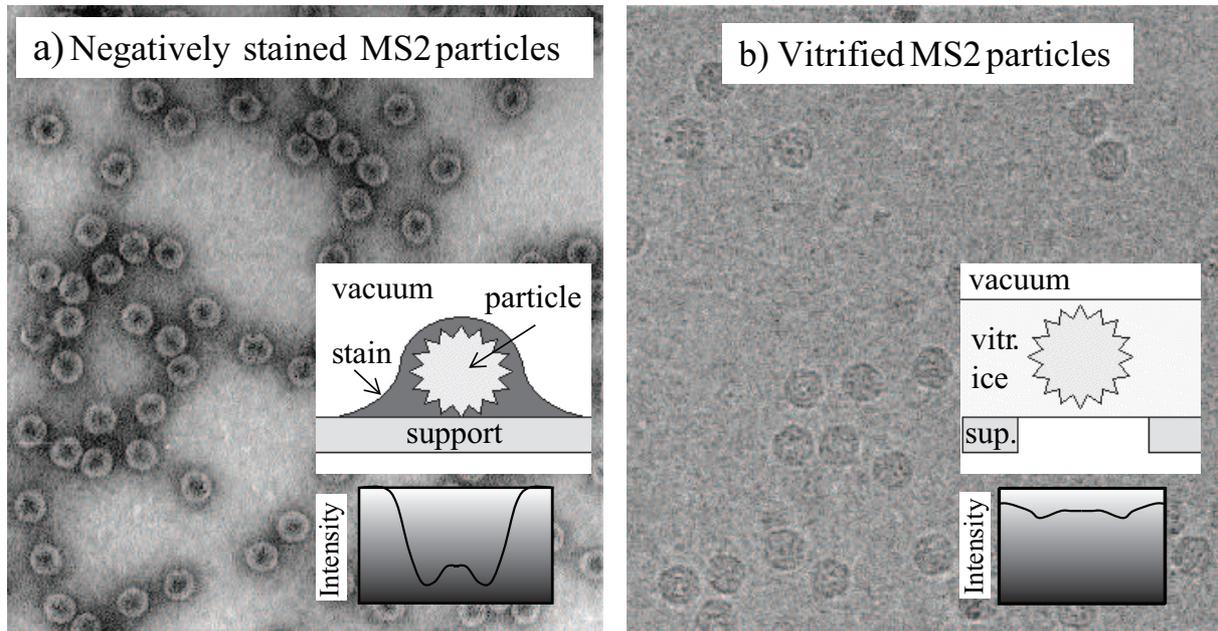


Fig. 1.2. a) Negatively stained and b) vitrified bacteriophage MS2 particles (*Courtesy of Roman Koning (LUMC)*). Note that scattering (amplitude) contrast is dominant for negatively stained sample, and phase contrast for vitrified sample.

object from a different angle. Two popular methods for obtaining 3D information in cryo-EM are single particle analysis (SPA) and electron tomography (ET).

Single Particle Analysis (SPA)

The averaging of different particles that have identical structure by functional demand can be used to reduce noise and electron dose, and consequently, radiation damage. In crystallography, redundancy of structural information is used to obtain a higher signal-to-noise ratio (SNR) by translational and rotational repetition of a unit cell consisting of one or more identical molecules.

In the EM field, the highest resolution so far (better than 3 Å [19]) was achieved by electron crystallography of monolayer (2D) crystals such as those formed by membrane proteins. The applicability of electron crystallography in determination of a 3D structure is, however, often compromised by insufficient image quality of the tilted crystalline specimen.

EM has the great advantage that the images of many identical objects can be treated in a similar way as the unit cells of a crystal. Isolated macromolecules that exist in structurally identical conformations will have identical projections in the electron microscope when viewed from the same orientation. This assumption is the basis of single particle analysis (SPA). In practice, the alignment of projection images for the purpose of averaging is much more challenging in the case of SPA than in the case of the electron crystallography. Furthermore, variations in the particle environment and shape restrict the attainable resolution. Even at resolutions in the range of 1 nm, SPA provides insights on quaternary structure of large macromolecular assemblies and

it might complement missing information caused by the phase problem in X-ray crystallography [2]. The high symmetry of helical and icosahedral structures allows additional averaging by well-known relations between symmetric building blocks.

Electron tomography (ET)

Samples that possess a unique, irreproducible 3D structure such as a whole cell or subcellular organelles such as mitochondrion require a tomographic approach for visualization. In electron tomography (ET), the sample is tilted and imaged from various directions. The specimen tilt is commonly limited to angles around ± 70 degrees due to restrictions of the specimen holder, obscurity of the specimen support, or the slab-geometry of the specimen which leads to increased apparent specimen thickness at high tilt angles. The problem associated with these missing projections is often referred to as the *missing wedge* and leads to artifacts and an anisotropic resolution of the reconstructed volume. Additionally, the presence of noise results in a relatively low-resolution of the reconstructed structures (> 2 nm) compared to X-ray crystallography and SPA. The main advantage of ET is the imaging of macro-molecular complexes within the cell, their various spatial configurations and indirectly, dynamics, as well as the capability to visualize whole prokaryotic cells or thin eukaryotic cells. With high enough resolution (in the range of 1 nm), the docking of high-resolution sub-units (obtained by X-ray crystallography) would be possible (as in the case with SPA) and ET would be able to deliver a pseudo-atomic atlas of a cell. If the particles have the same conformation and need to be imaged in a cellular context rather than isolated in a solution, sub-tomogram averaging can be applied. This approach is similar to SPA, except that alignment and averaging are performed on 3D datasets (with anisotropic resolution).

1.3 Phase contrast

Biological specimens consist mainly of light elements with similar atomic mass. The mass density of macromolecules deviates only little from that of vitreous water. Since scattering of the incident electrons is dependent on atomic number and density, the scattering contrast is inherently low. While passing through the specimen, however, a high-energy electron changes its wavelength and speed, maintaining its energy (elastic scattering). The specimen produces local phase shifts of incident electron wave. Since the detector can only capture the intensity of the electron wave, the information about the object encoded in the phase cannot be detected. Phase contrast can be produced by creating an additional phase shift between the scattered and unscattered part of the electron wave. In EM, phase plates are still uncommon (due to contamination problems) and therefore, the additional phase shifts needed for phase contrast are typically introduced by defocusing or by aberrant lenses. The signal transfer from the specimen to the imaging plane is described by the contrast transfer function (CTF). The CTF is an oscillating function in the spatial frequency domain and depends on the defocus and aberrations. In materials science, in particular for imaging crystalline structures, it is not uncommon to set the defocus to a value (the so-called Scherzer defocus) where the CTF is maximal over the entire frequency range of interest. In cryo-EM, Scherzer defocus is rarely used as the structures are

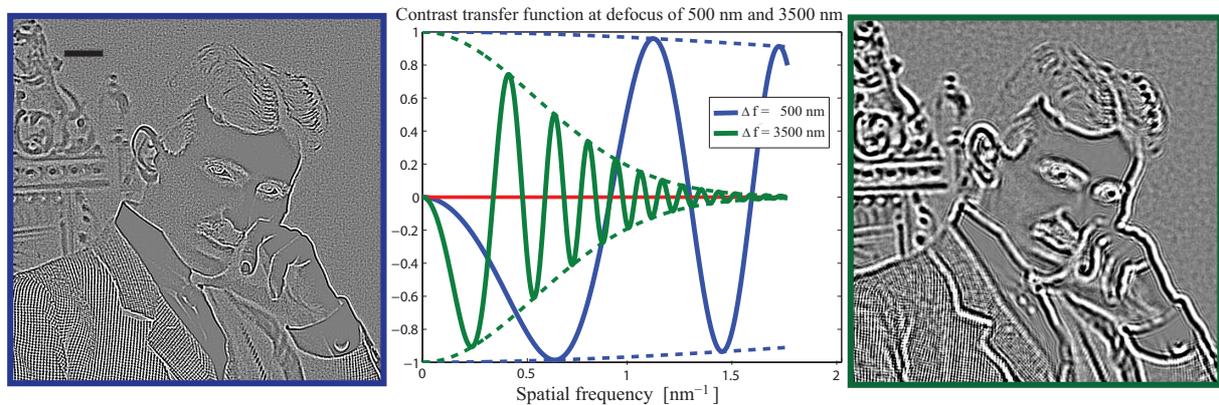


Fig. 1.3. The effects of the contrast transfer function (CTF) on an image of Nikola Tesla. Stronger defocusing of the objective lens generates low-frequency contrast at the expense of decreased contrast at higher frequencies. Scale bar corresponds to 20 nm.

complex and the frequency range of interest is quite wide. High-frequency contrast is needed for high resolution, while low-frequency contrast is important for visual and/or automatical localization of the macromolecules. Such low-frequency contrast is usually achieved at a defocus of a few micrometers.

1.4 Limiting factors in cryo-EM

The level of structural detail that can be obtained with cryo-EM is largely limited by 1) the noise and the blurring of the detector, 2) the CTF, and 3) radiation damage which limits the integrated electron flux that can be used, resulting in images with a poor SNR.

Detector

Ideally, the quality of the cryo-EM images would only depend on the shot (Poisson) noise resulting from the limited electron dose the sample can withstand before beam-induced deformations start to be apparent. Unfortunately, the images are degraded by the modulation transfer function (MTF) of the detector and several inherent noise components. The MTF describes how the signal is transferred for different spatial frequencies. In a TEM detector, the signal and the noise, however, are not transferred in the same way [20]. The detective quantum efficiency (DQE) describes the noise added by the detector and defines the final quality of the images. In recent years, efforts have been made to improve the DQE by utilizing direct electron detection.

Contrast transfer function (CTF)

The CTF exhibits an oscillatory character when defocus and aberrations are introduced. Furthermore, the spatial and temporal incoherencies of the electron source damp the contrast for

increasing spatial frequencies. Efforts have been made to improve instrumentation *via* high-coherency and high-brightness electron sources as well as by introducing phase plates for in-focus phase imaging. Although recently the usage of phase plates showed significant gain of contrast in cryo-EM images [21, 22], phase plates are still commercially unavailable. Phase contrast produced by an increasing value of defocus causes narrowing of the different CTF bands and consequently displaces its zero-crossings. Stronger defocusing of the objective lens generates low-frequency contrast at the expense of decreased contrast at higher frequencies (see Fig. 1.3). For spatial frequencies beyond the first zero-crossing of the CTF, phase contrast images cannot be quantitatively interpreted without an accurate image formation model. In order to be able to perform reliable image reconstructions, knowledge of the CTF parameters is crucial so that one can correct for its detrimental effects. Additional challenges include local axial variations of the defocus due to the specimen thickness as well as lateral defocus variations in tomography due to the tilt geometry.

Radiation damage

Radiation damage, unfortunately, will always limit the achievable resolution in cryo-EM [23, 24]. The damage arises from the deposition of energy into the specimen due to inelastic interactions between the incident electrons and matter. Furthermore, it has been suggested that radiation damage during the exposure causes beam-induced movements that attenuate the contrast further [23, 25–27]. Studies that describe radiation damage are as old as cryo-EM itself [16, 23]. Understanding of the radiation chemistry, however, is still lacking [28]. The integrated electron flux used to acquire cryo-EM data is a compromise between SNR and radiation damage. Additionally to radiation damage, inelastically scattered electrons that reach the detector lost their coherency which causes images to appear more blurry. These inelastic components can be partially suppressed by zero-loss energy filtering. On the bright side, inelastic scattered electrons can be used to map elements by means of electron energy-loss spectroscopy (EELS).

Other limitations

Some of the other limiting factors in obtaining higher resolution in cryo-EM include: 1) the specimen preparation methods (e.g. artifacts in thin sections); 2) thicker ice layers contributing to larger fractions of inelastically scattered electrons, thus decreasing the image quality; 3) structural variations among single particles or sub-tomograms (heterogeneity) hampering the full potential of redundancy and blurring the calculated average; 4) noise obstructing precise alignment of the particles in SPA or sub-tomograms as well as alignment of the images in a tilt series for ET; 5) non-parallel illumination introducing unwanted higher order aberrations of the objective lens; 6) geometrical distortions of the projection lens system for large fields of view; 7) missing wedge artifacts in tomography which can hamper the final structure determination; 8) reduction of contrast when macromolecules are imaged in aqueous solutions that contain additives which stabilize the protein of interest (e.g. detergents or lipids); 9) solvent boundary and surface tension effects that induce preferred orientations in the molecule; and 10) inhomogeneous sample distribution due to e.g. local hydrophobic patches in the carbon support.

1.5 Thesis challenges

The research performed in this thesis represents one of two subprojects of the FOM industrial partnership program with FEI Company (www.fei.com). The common aim is to obtain higher resolution in cryo-ET of biological specimens. The research is performed in collaboration between TU Delft and the Leiden University Medical Centrum (LUMC). Optimizing the data collection strategy is crucial for reliable image interpretability and achieving the highest attainable resolution. Simulation of image formation (forward modeling) provides possibility to easily and cost-effectively investigate the influence of a certain physical parameter on the final image. The data collection strategy and framework of our project involve a combined procedure that ranges from specimen preparation, through the actual data collection to forward modeling and reconstruction. Fig.1.4 depicts the project framework that can be divided in three parts: forward modeling (simulations), experimental data acquisition, and reconstruction. The focus of this thesis will be on accurate modeling of the image formation process (blue box in Fig.1.4), accurate characterization of the detector and CTF parameters (orange box in Fig.1.4), and experiments (green box in Fig.1.4). The simulated images are to be compared with experimental images for validation and to be utilized in reconstruction of the 3D electrostatic potential distribution by solving a complex inverse problem. Tomographic reconstruction with focus on the problem of spatially varying CTF [29, 30] (yellow box in Fig.1.4) is the main subject of the second subproject. Integration of two subprojects should lead to better design of experiments, forward modeling and 3D reconstruction.

1.5.1 Forward model

An accurate forward model is essential for optimization of data collection strategy, assisting the regularization (introduction of prior information) of the 3D reconstruction, improving image interpretation and achieving a resolution beyond the first zero-crossing of the CTF. Such a model has to account for the specimen's elastic and inelastic scattering properties, the effects of the CTF, and the influence of the detector on the image formation in cryo-EM. Simulations of TEM images of biological specimens are implemented in a number of software packages for SPA and ET [31–38]. Often, a virtual model of a specimen is created using simple 3D geometrical phantoms [32]. In some cases, the specimen volume is constructed based on information from the PDB. TEM images are then computed by projecting the 3D electron densities. These simulations are rather simplistic because the constructed specimen does not represent the actual physical electron-specimen scattering properties (interaction potential). The noise is often simplified as being additive Gaussian noise and the relevant detector properties have been neglected. Those simulations have been mostly used for determination of particle orientation in SPA and for evaluation of reconstruction algorithms in SPA and ET. For optimization of data collection, those models are insufficient.

TEM-simulator [39] aims to provide simulations based on physical principles. It was the first bimolecular simulator whose results were compared to experiments, albeit not in depth. The thickness of the specimen was neglected in that model, low-pass filtering to a certain resolution exceedingly damped the interaction potential (IP), and the solvent was assumed to be

water instead of slightly less dense vitreous ice. Although most simulation parameters described in [39] are based on physical principles, a calibration protocol needs to be employed for some nuisance parameters such as granularity of the solvent, absorption potential as well as detector parameters such as the MTF, DQE, and conversion factor. Furthermore, none of the aforementioned approaches have considered chemical bonding and/or interaction of the sample with solvent and ions.

The aim of this thesis is to understand and model the relevant physical processes governing the image formation and to address the aforementioned issues as well as to develop a platform for simulations, herein referred to as InSilicoTEM. The main ingredients of a forward model include the parameters of the specimen, microscope optics and the detector (see segments in blue box in Fig. 1.4).

Parameters

The experimental and imaging parameters are not only the settings for the data acquisition process, but are also the input parameters for the forward model (ellipses in Fig. 1.4). These include buffer conditions as well as TEM and camera settings. The parameters of the specimen (buffer conditions block) include pH , dielectric constant ϵ , temperature T , ion concentration n_0 , and the specimen thickness d . The specimen's interaction potential must be constructed and the influence of the solvent and ions addressed. The electron-specimen interaction describes how the electron wave at acceleration voltage E is propagated through the specimen. Microscope parameters include acceleration voltage E and its spread ΔE , opening angle α_i , defocus Δf , astigmatism A_1 , spherical C_s and chromatic C_c aberrations, objective aperture A_p , magnification M , and incident electron flux Φ_e . Relevant camera parameters are exposure time t_{exp} , binning, conversion factor CF , MTF, DQE, readout I_{rn} and dark current I_{dc} noise.

Some imaging parameters vary between acquisitions, while others are stable for a long period of time. To accurately model image formation, we need to know the numerical values of all parameters. When necessary, they must be estimated from the experiment, using independent measurements. The parameters that must be estimated (orange box Fig.1.4) include imaging parameters such as M , Φ_e , Δf , A_1 , d , as well as detector parameters such as the CF , MTF, DQE, I_{rn} and I_{dc} .

1.5.2 Detector

Although the detector characteristics significantly influence the image formation, in previous image simulation work they have been either neglected or phenomenologically introduced. The reason for this is that the quality of an image detector as used in TEM is not easily accessible. Different detector manufacturers provide different types of figures of merit when advertising their detector. Therefore, a comprehensive characterization of the detector including all relevant noise contributions is essential to an accurate image formation model and eventually to a 3D reconstruction. A careful characterization of TEM detectors will yield, among others, statistics for hot and bad pixels, the MTF, the conversion factor, the effective gain and the DQE. Furthermore, a correction of the fixed pattern noise based on insufficient statistics would spoil

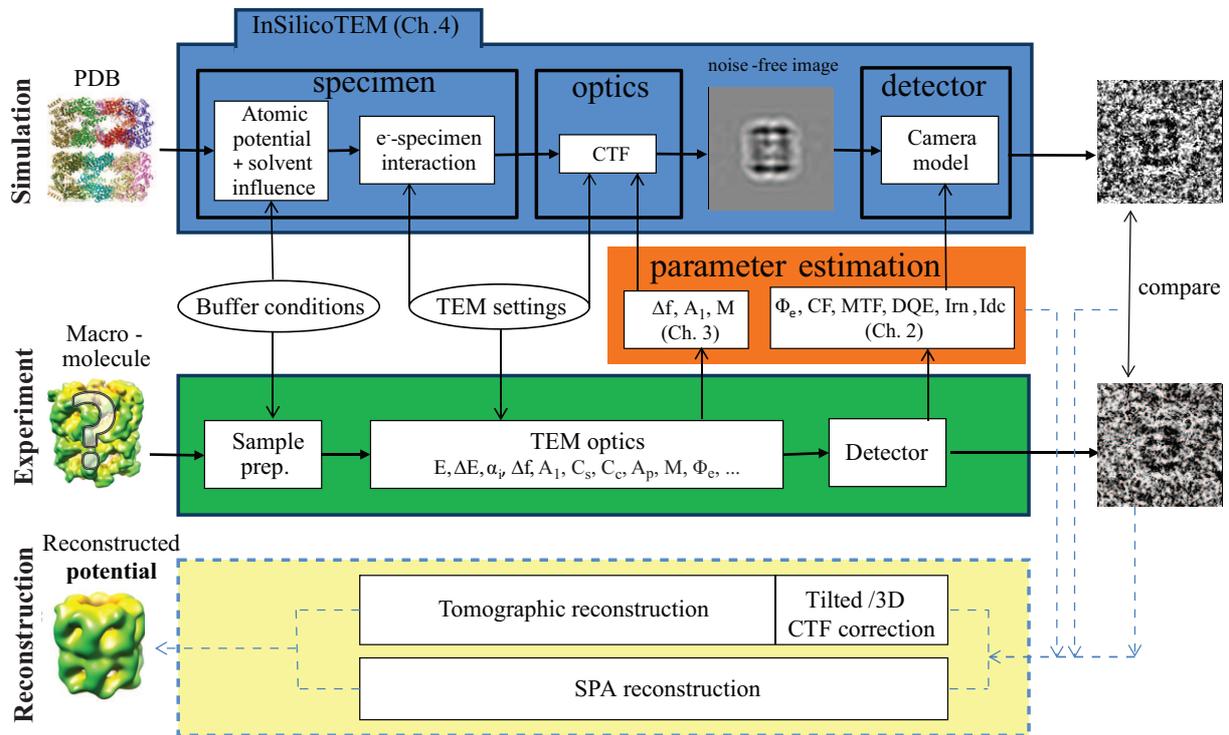


Fig. 1.4. Project workflow which involves a combined procedure that ranges from sample preparation and data collection (experiment) to forward modeling (simulation) and reconstruction. A forward model (blue box) includes the parameters of the specimen, microscope optics and the detector. The experimental and imaging parameters represent not only the acquisition settings but also the input parameters for the forward model (ellipses). To accurately model image formation and facilitate 3D reconstructions, the numerical values of all parameters need to be known and, when necessary, they must be estimated from the independent measurements (orange box). The simulated images are to be compared with experimental images for validation and to be utilized in reconstruction of the 3D electrostatic potential distribution by solving a complex inverse problem.

image interpretation, and automated procedures.

1.5.3 CTF

Determination of the CTF parameters, especially defocus and twofold astigmatism, is crucial in designing post-processing strategies to account for the effect of the CTF and for the interpretation of images at spatial frequencies beyond the first zero-crossing of the CTF. The defocus estimation is usually based on the detection of Thon rings in the power spectrum density (PSD) of the image. There are various software packages that provide defocus determination (e.g. [40–46]). Their accuracy can be limited by the fitting of the background in the PSD. Furthermore, the influence of the spherical aberration on the shape of the Thon rings has been ignored. A robust estimation of small astigmatism values were lacking and the uncertainty of the estimations was hard to assess. The CTF determination at the specimen area is very chal-

lensing due to the low SNR. Therefore, the defocus is often estimated from an adjacent carbon support area. Furthermore, for tilted specimens in tomography, at least three neighboring areas should be imaged to capture the full geometry of the tilt.

Radiation damage

Quantitative modeling of radiation damage is largely hampered by the limited knowledge of the processes involved. As a metric of radiation damage influence one can use *dose*, expressed as energy deposited per mass unit, which is derived from parameters including the electron energy, incident flux and measured sample thickness. Knowledge of the dose allows obtaining an upper estimate of radical concentrations (which build up in the vitreous sample), and performing heat transfer simulations. The analysis of these effects facilitates the optimization of data collection. Furthermore, beam-induced motions influence the image contrast, and must therefore be included in the forward model.

1.5.4 Thesis objectives

- **Accurate modeling of the image formation process in cryo-EM based on physical principles.**

In order to construct such a forward model and furthermore assist the CTF correction and/or regularization of the reconstructions it is necessary to:

- Construct the interaction potential based on electron scattering properties and investigate the embedding solvent contribution to that potential.
- Properly describe electron propagation through the specimen (expressed *via* weak-phase object approximation, projection assumption, their combination or multislice approach).
- Include the influence of the inelastic scattering.
- Characterize TEM detectors including all relevant statistics.
- Develop a method for accurate estimation of the CTF parameters, in particular defocus and astigmatism and their uncertainties.

Since radiation damage determines the allowable flux used for imaging and influences the image contrast, efforts will be made to

- Better understand certain aspects of radiation damage such as specimen heating, dose-rate effects, and beam-induced movements.

More detailed objectives can be found in introduction section of each chapter.

1.6 Thesis outline

This thesis is compiled from a collection of five journal papers and one conference paper. These papers are organized and presented in four chapters. The papers [47–50] are published, while the papers [51, 52] have been submitted at the time of publication of this thesis. To enhance the readability, the publications [48] and [49] have been combined in Chapter 3. Paper [52] is presented as Appendix G belonging to Chapter 4. All topics in this thesis are not limited to ET applications only, but are relevant for cryo-EM in general.

Chapter 2 - TEM camera characterization

In this chapter, a set of algorithms is provided to characterize on-axis slow-scan CCD-based TEM detectors. A careful characterization, yields, among others, statistics for hot and bad pixels, the modulation transfer function, the conversion factor, the effective gain and the detective quantum efficiency. Gain and bias corrections of raw images are presented along with the need for the use of lookup tables of defect pixels. The relative performance of the characterized detectors is discussed and a comparison is made with similar detectors that are used in the field of X-ray crystallography. These tools are not limited only to CCD-based detectors but can be extended for the characterization of a new generation direct-electron detectors. The chapter was published as a paper [47].

Chapter 3 - Accurate defocus and astigmatism estimation

This chapter presents an algorithm to accurately estimate defocus and astigmatism. The associated uncertainties are derived from a single image. The algorithm suppresses the background in the power spectrum density (PSD) using an adaptive filtering strategy, after which robust template matching is applied to estimate the shape of the Thon rings. The frequencies of the detected rings, together with outlier rejection and assignment of an order to the CTF zeros, are used to estimate the defocus and its uncertainty (k -trajectory method). From defocus and ellipticity, we derive astigmatism and its uncertainty. The accuracy of the algorithm is evaluated on simulated data and the reproducibility is investigated on experimental data. We introduce a Thon ring averaging method for contrast transfer assessment which takes into account the influence of spherical aberration on Thon rings shape. The chapter was published as a paper [49] and additionally, for coherency, a figure from the conference paper [48] is included.

Chapter 4 - Forward modeling in cryo-EM

The aim of this chapter is to construct an image formation model that accounts for the specimen's scattering properties, microscope optics, and detector response. The interaction potential is calculated *via* electron scattering factors of isolated atoms and extended with the influences of the solvent's dielectric and ionic properties as well as the molecular electrostatic distribution. Inelastic scattering is addressed. Subsequently, the electron wave is propagated through the specimen and the influence of the optics is included *via* the CTF. We incorporate the DQE in

the camera model, instead of using only the MTF. The full model was validated against experimental images of 20S proteasome, hemoglobin, and GroEL. We investigate the effects of defocusing, changes due to the integrated electron flux, inelastic scattering and acceleration voltage. The influences of the beam-induced specimen movements and the solvent amorphousness are considered. At higher SNRs, experimental and simulated intensity profiles across carbon edges and nanotubes are compared (Appendix F). All parameters in the analysis are based on physical principles and, when necessary, experimentally determined *via* tools described in Chapters 2 and 3. The main part of the chapter has been accepted for a publication as a paper [51]. As an addition to this chapter, Appendix F explains theoretical approximations and methods in more detail and has been submitted as supplementary material associated with the paper [51]. Appendix G represents a paper on applicability of the projection assumption and weak-phase object approximation in phase-contrast cryo-EM [52].

Chapter 5 - Radiation damage: effects of dose and dose rate

In this chapter, inspired by numerous radiation damage studies done by X-ray crystallographers, we investigate parameters such as dose, dose-rate and beam-heating in EM. We show how the incident electron flux, expressed in $e^{-}\text{\AA}^{-2}\text{s}^{-1}$, electron energy, and measured sample thickness and composition, can be related to the absorbed dose, expressed in grays ($1\text{Gy} = 1\text{J/kg}$). Stroboscopic exposure series were collected for different incident fluxes and integration times from a hemoglobin sample. The quantitative comparisons between different doses are presented along with the discussion about the benefit of stroboscopic data collection. The chapter was published as a paper [50]. As coauthor of the paper, I contributed mostly to the dose and heat-transfer calculations.

Chapter 6 - Conclusions and recommendations

In the last chapter, theoretical and practical work presented in the thesis is revisited and evaluated. Conclusions are drawn with respect to the validation of the developed models, their limitation and applicability. The chapter lists the main contributions of the thesis and gives recommendations for future work.

The software packages for camera characterization (Chapter 2), defocus and astigmatism estimation (Chapter 3), and simulations of image formation - InSilicoTEM (Chapter 4) have been implemented in DIPimage, a MATLAB toolbox, and are freely available for non-commercial use (<http://www.diplib.org/add-ons>).

The defocus and astigmatism estimation algorithm was developed in collaboration with FEI Company and besides my MATLAB implementation it also resulted in prototype software implemented by Dr E. Franken and used within FEI Company. The InSilicoTEM source code was transferred directly to FEI Company where it is currently used for modeling.

Chapter 2

Detector characterization

Published as [47]: M. Vulovic, B. Rieger, L. J. van Vliet, A. J. Koster, R. B. G. Ravelli, “A toolkit for the characterization of CCD cameras for transmission electron microscopy”, *Acta Crystallographica D* 66 (1) (2010) 97-109.

Abstract

Charged coupled devices (CCD) are nowadays commonly utilized in transmission electron microscopy (TEM) for applications in life sciences. The direct access to digitized images has revolutionized the use of electron microscopy, sparking developments of automated collection of e.g. tomographic data, focal series, random conical tilt pairs, and ultra-large single particle data sets. Nevertheless, for ultra-high resolution work, photographic plates are often still preferred. In the ideal case, the quality of the recorded image of a vitrified biological sample would solely be determined by the counting statistics of the limited integrated electron flux the sample can withstand before beam-induced alterations dominate. Unfortunately, the image is degraded by the non-ideal point-spread function of the detector - as a result of a scintillator coupled by fibre optics to a CCD - and the addition of several inherent noise components. Different detector manufacturers provide different types of figure-of-merits while advertising the quality of their detector. It is hard for most laboratories to verify if all the anticipated specifications are met. In this report, a set of algorithms is presented to characterize on-axis slow-scan large-area CCD-based TEM detectors. These tools have been added to a publicly available image processing toolbox for MATLAB. Three inhouse CCD cameras were carefully characterized, yielding, among others, the statistics of hot and bad pixels, the modulation transfer function, the conversion factor, the effective gain, and the detective quantum efficiency. These statistics will aid data collection strategy programs and provide prior information for quantitative imaging. The relative performance of the characterized detectors is discussed, and a comparison is made with similar detectors that are used in the field of X-ray crystallography.

2.1 Introduction

Charge-coupled devices (CCDs) are used in nearly every scientific domain of life science imaging, e.g. for transmission and fluorescence microscopy, optical and UV spectroscopy, digital photography, X-ray diffraction and imaging, and electron microscopy. Large area CCD-based systems are the most common detectors on modern synchrotron beamlines [53], complemented by multiwire gas-filled chambers and novel photon-counting pixel arrays. The instant image access in electronic form, high sensitivity, low noise, versatile coverage from submicrometric to millimetric spatial resolution, as well as high reliability of commercial CCD cameras, make them ideal for a wide range of applications. In transmission electron microscopy (TEM), however, there has been a considerable delay in the adaptation to CCD technologies.

In 1982, the use of an array of 100×100 photosensitive elements to detect 20-100 keV electrons directly was reported [54]. The system demonstrated an excellent linearity between input and output signal, and a high intrinsic gain, but had a limited spatial resolution compared to photographic film, and suffered from radiation damage. They suggested to first convert the electron image to its photon counterpart and to detect the latter by a CCD. In [55] was reported the use of such an indirect detection scheme, involving an electron scintillator, an optical coupler and a 576×382 pixel sensor. Many more experimental and commercial systems have been reported since then (see references in [56]). The direct access to digital data has enabled developments such as autotuning of the microscope [57], automated electron tomography [58], protein electron crystallography [59], and automated cryo-electron single particle micrograph collection [60].

Despite the many advantages of CCDs, some areas remain where applications of CCDs have been limited by certain characteristics inherent to CCD based detectors [61]. E.g., for high-resolution single-particle work, film is still significantly better [62] than fibre-optic coupled CCD detectors: without binning of the CCD camera and at a magnification of 70000x, film is better beyond 21 Å resolution. For 4-fold binning of the CCD camera and at very high magnification ($> 300000\times$), film is reported to be superior beyond 7 Å resolution. This might have contributed to the slow transition from film recording to digital imaging in the field of TEM. Until recently, large-area CCD cameras could only been offered as third party add-ons to new TEMs. The relatively slow pace of adaption partially reflects the satisfactory performance of film recordings in terms of resolution and number of pixels after digitization, although both gaps are being closed. Commercial digital cameras are now available that have a larger image area than film (<http://www.tvips.com/Prod.TF816.php>). Detector systems based on newly developed CMOS hybrid-pixel technology which operate in noiseless single-photon-counting mode, are already commercially available for X-ray imaging and diffraction applications (<http://www.dectris.com/>). Hybrid pixel detectors are being developed for TEM applications [63], [64] and offer considerable scope for better characteristics compared to phosphor/fibre optics-coupled CCDs [65].

The incremental improvements in CCD technology, number of pixels, quality of phosphors/scintillators, fibre-optic coupling and electronics, as well as emerging novel pixel array detector technology, will not make it easier for the user to select from this heterogeneous landscape the right detector for an experiment. Whereas well funded large user facilities might be

able to keep up to date with the latest detector technologies, most academic laboratories will have to select a particular detector and use it for at least a decade. Even among a given category of detectors such as CCD cameras, the wide range of inconsistent, sometimes incomprehensible, and often incomplete commercial specifications hamper the selection process. In this paper, we present a set of algorithms to characterize CCD detectors, which have been implemented in DIPLib, a publicly available software toolbox (www.diplib.org) for MATLAB (The MathWorks, Inc.). This should facilitate users to commission new detectors and help them to design better data collection strategies with existing ones. A number of detector characteristics are recapitulated, such as readout noise, conversion factor, effective gain, point-spread function, modulation transfer function, and detective quantum efficiency. Three of our own $4k \times 4k$ TEM imaging CCD detectors have been characterized. Only the user can judge if a detector meets the needs of an experiment, and the outcome depends on many other elements as well, including electron source, optics, and, above all, the sample. Therefore, the differences found for the three detectors are not judged upon and no reference is made to their manufacturers.

2.2 Detector characterization

To characterize a CCD detector and subsequently identify and correct artifacts, one needs to determine the contributions of all noise components, effective gain, conversion factor, linearity of response, modulation transfer function (MTF), and detective quantum efficiency (DQE). Temporal noise randomly changes from frame to frame. It includes stochastic contributions such as dark current noise, readout noise, photon noise, beam flicker, burst noise, and shutter noise. There is also a source of fixed pattern noise especially in fibre-optic coupled digital cameras. This spatial noise does not vary from frame to frame and is caused by spatial variation in the thickness of the scintillator, fibre-optic coupling (*chicken wire* or broken fibres), dust, CCD bias pattern (in particular if multiple readout ports or composite CCDs are used), and other artifacts that produce variations in the pixel-to-pixel sensitivity and/or distortions in the optical path to the CCD or in the CCD chip itself. Flat-field correction is used to suppress fixed pattern noise.

A corrected image $I_{\text{corr}}(x, y)$ can be obtained *via* [66]

$$I_{\text{corr}}(x, y) = \frac{I_{\text{raw}}(x, y) - \overline{I_{\text{bg}}}(x, y)}{I_{\text{gain}}(x, y)}, \quad (2.1)$$

where $I_{\text{raw}}(x, y)$ is the original, uncorrected image, $\overline{I_{\text{bg}}}(x, y)$ is the average background image (see below), and $I_{\text{gain}}(x, y)$ is the image with normalized gain values for each pixel. In X-ray crystallography, a fibre optic taper or lens system makes the conversion from raw images to corrected images more cumbersome, since the distortion of the demagnifying system needs to be accounted for. Furthermore, it is non-trivial to obtain a stable large uniform X-ray beam that is needed for the collection of the data from which $I_{\text{gain}}(x, y)$ is obtained. Therefore, most X-ray detector manufacturers deliver their camera with tables for distortion and flat-field correction, and the user only has to collect background images for the desired exposure time. The manufacturers' gain and distortion calibration would normally remain adequate for a number

of years. This also holds for X-ray detectors where fibre optic plates (1:1 magnification) rather than tapers are used.

Electron microscopy detectors typically employ fibre optic plates in combination with large sensor chips. Two popular large area CCD sensors are the Fairchild CCD 485 and 486 (Eagle 4k, Gatan 4k, Tvips 4k). These sensors are also used for a number of X-ray detectors (Bruker APEXI and II, platinum 135/200/200C, Rayonix 135 and 165). No distortion corrections are required when these sensors are bonded to a fibre optic plate. The electron microscopist can perform the background and gain calibration in a straightforward manner, as large uniform flat-field electron beam illumination conditions are readily obtained with modern electron microscopes. Academic and commercial electron microscope data collection packages, such as Tia (<http://www.fei.com/products/types/fei-software.aspx>), SerialEM [67], UCSF software [68], Leginon [69] and Digital Micrograph (<http://www.gatan.com/products/software/>), provide functionality for this camera calibration step. The rate of recurrence at which background and gain calibration is required is significantly higher compared to X-ray detectors, and can vary from once a month to a few times a day, depending on the camera manufacturer.

Correction of raw images does not require the same illumination conditions during acquisition of raw images and the white reference images. The optical density of a semi-thin scattering-contrast dominated TEM sample can be modeled *via* the Beer-Lambert law:

$$\log\left(\frac{I_{\text{sample}}}{I_0}\right) = -\alpha l, \quad (2.2)$$

where I_0 is the incoming intensity, I_{sample} is the outgoing intensity, α is the absorption coefficient, and l is the path length. In this equation, I_0 does not have to be a uniform beam. A near-uniform beam can be referred to as flood field [70]. The explicit measurement of I_{sample} and I_0 in electron microscopy is, for example, carried out in the Leginon package for the automatic characterization of the thickness of vitreous ice specimens [60]. In principle, the separate measurements of the gain normalized image I_{gain} in Eq. (2.1) and a flood field image I_0 of Eq. (2.2) could be combined in one measurement. However, such a characterization would only remain valid as long as I_0 does not change. The flood field image I_0 will change for different electron beam settings, whereas the gain normalized image I_{gain} is independent of the electron optics, and only alters with factors such as temperature.

To estimate the properties of a fibre-coupled CCD correctly, it is important to suppress statistical outliers (*zingers*, named after Zinger [71]) in the reference images. They can be detected by measuring a large number of images under identical conditions. Cosmic rays and muons in particular can produce a burst of photons in the scintillator leading to white spots or streaks in the image. Radioactive elements (essentially thorium) present in the fibre-optic tapers can also lead to zingers [72]. Other possible sources of zingers are X-rays and burst noise (*popcorn noise*), the latter referring to a variety of electronic effects that could yield both increased and decreased pixel values.

The average background image $\overline{I_{\text{bg}}}(x, y)$ will be different for different integration times. It has a time-independent offset, the average bias $I_{\text{bias}}(x, y)$, plus a time-dependent contribution from the spontaneous thermally-induced generation of electron-hole pairs within the CCD, which is referred to as dark current. For typical exposure times in bright field TEM imaging of

biological samples (0.1s to a few seconds) a linear relation may be assumed:

$$\overline{I_{\text{bg}}}(x, y) = I_{\text{bias}}(x, y) + t_{\text{exp}} I_{\text{dc}}(x, y), \quad (2.3)$$

where (x, y) denotes pixel position, t_{exp} the exposure time of the CCD camera (or integration for the dark images), and $I_{\text{dc}}(x, y)$ the average dark current in counts per second. The readout noise $I_{\text{rn}}(x, y)$ is the standard deviation of a large series of background images I_{bg} measured at an exposure time at or near zero seconds,

$$I_{\text{rn}}(x, y) = \left(\frac{1}{N} \sum_{i=1}^N [I_{\text{bg},i} - I_{\text{bias}}]^2 \right)^{1/2}. \quad (2.4)$$

A flat-field (uniform) illumination of the camera will not result in a uniform response of the CCD, as each of the conversion steps from high-energy electrons to photo-induced electrons read from the CCD will introduce local amplification or attenuation of the signal. The scintillator will have variations in thickness; some parts could be blocked by artefacts such as dust, the coupling of the scintillator to the fibre optic plate will have imperfections, the fibre optic plate itself will leave a very strong pattern of individual fibres and fibre bundles, the coupling of the fibre optic plate to the CCD will lead to location-dependent signal loss and the CCD itself has a non-uniform response. The combined effects are corrected for by means of a flat-fielding, which relies on the measurement of white reference (uniformly illuminated) images I_{white} at one or multiple exposure times:

$$I_{\text{gain}}(x, y) = \frac{\overline{I_{\text{white}}}(x, y) - \overline{I_{\text{bg}}}(x, y)}{\langle \overline{I_{\text{white}}} - \overline{I_{\text{bg}}} \rangle_{x,y}} \quad (2.5)$$

where $\overline{I_{\text{bg}}}$ is an average background image as calculated with Eq. (2.3) and $\overline{I_{\text{white}}}$ is an average white reference image calculated in a similar way. The notation $\langle \rangle_{x,y}$ is used to denote spatial averaging over the entire image.

The modulation transfer function (MTF) is a measure of how the signal amplitude is transferred for different spatial frequencies. It is calculated from the modulus of the Fourier transform of the point-spread function (PSF) of the detector. There are two common methods for experimental determination of the MTF, referred to as the noise and the edge method. The noise method is a stochastic method in which the camera is exposed to uniform illumination. The incoming signal may be considered as white noise that has a constant power spectrum over all spatial frequencies. The assumption is that this constant spectrum will be attenuated by the MTF of the camera as any other signal. We expect the detector PSF to be dominated by the fibre-optic plate scintillator and therefore isotropic. The absolute value of the Fourier transform of a uniformly illuminated image, angularly averaged, yields the MTF of the system [56, 73]. Angular averaging of the Fourier transform can be performed by creating rings in an image with a Gaussian profile $G(r, \sigma)$. The Gaussian-weighted sum of the modulus of the Fourier transform of the white noise image $|F(q)|$,

$$|F(r)| = \frac{\sum_q G(r, \sigma) |F(q)|}{\sum_q G(r, \sigma)}, \quad (2.6)$$

will yield the MTF after normalization. The edge method is a deterministic method and uses a uniformly illuminated straight sharp metal knife-edge which blocks the incident electrons on one side [74]. The knife-edge profile can be represented by a step function. An image of the knife-edge is taken with uniform illumination and is subjected to flat-field correction. The mean intensities on the dark and bright sides are calculated and used to normalize the image. An average edge profile from the slanted edge is extracted from the image. Differentiation of the 1D edge spread function (ESF) gives the point-spread function (PSF) and, after Fourier transform and taking the modulus, a 1D cross-section of the detector's 2D modulation transfer function. Assuming an isotropic MTF, an edge measurement in a single direction suffices.

Attenuation from the MTF alone would not spoil the image quality. If the signal is transferred up to Nyquist frequency and the MTF is known, one can, in theory, restore the image by deconvolution. In practice, deconvolution will be hampered by noise. The detective quantum efficiency (DQE) describes the noise added by the detector.

The DQE is defined as the squared ratio of the SNR between output and input signal

$$\text{DQE} = \left(\frac{\text{SNR}_{\text{out}}}{\text{SNR}_{\text{in}}} \right)^2. \quad (2.7)$$

The noise of a stochastic scattering process is not transferred in the same manner as the signal [75]. An electron is scattered in the scintillator and produces photons along its trajectory. These photons are scattered again. The process in the scintillator is therefore a complicated combination of scattering and amplification: the noise in the detected (output) signal is not simply the noise in the input signal attenuated by the MTF [76]. The signal and noise transfer differently as a function of spatial frequency q , thus the DQE becomes

$$\text{DQE}(q) = \frac{S_{\text{out}}(q)^2 / \text{NPS}_{\text{out}}(q)}{S_{\text{in}}(q)^2 / \text{NPS}_{\text{in}}(q)}, \quad (2.8)$$

where NPS refers to the noise power spectrum. In order to measure DQE the frequency dependence of the signal for a white image is approximated by

$$S_{\text{out}}(q) = \overline{S_{\text{out}}} \text{MTF}(q), \quad (2.9)$$

where $\overline{S_{\text{out}}}$ is the mean of the signal S_{out} . Since the input signal is a Poisson process with constant expected value across the image, the expected variance and the expected mean of the signal are the same and frequency independent i.e. $\text{NPS}_{\text{in}}(q) = S_{\text{in}}(q) = \overline{S_{\text{in}}}$. The mean of the incoming signal equals the integrated flux $\overline{S_{\text{in}}} = N$. The conversion factor is given as $CF = \overline{S_{\text{out}}} / \overline{S_{\text{in}}}$. The DQE can now be rewritten as

$$\text{DQE}(q) = CF^2 N \frac{\text{MTF}^2(q)}{\text{NPS}_{\text{out}}(q)}, \quad (2.10)$$

with

$$\text{NPS}_{\text{out}}(q) = \mathcal{F}[\sigma_{\text{shot}}(x, y)]^2 + \mathcal{F}[I_{\text{rn}}(x, y)]^2, \quad (2.11)$$

where $\mathcal{F}[\circ]$ denotes a Fourier Transform, $\sigma_{\text{shot}}(x, y)$ is the standard deviation per pixel due to Poisson noise and I_{rn} is the readout noise. The noise from dark current is usually dominated by

the readout noise and will be neglected in this analysis. The relative contribution of the readout noise to $NPS_{\text{out}}(q)$ is larger for lower integrated flux and higher frequencies. Eq. (2.10) can also be expressed as

$$DQE(q) = \frac{MTF^2(q)}{NNPS(q)}, \quad (2.12)$$

where NNPS is the normalized noise power spectrum,

$$NNPS = \frac{NPS_{\text{out}}}{CF^2N}. \quad (2.13)$$

2.3 Measurement methods

Three of our inhouse on-axis bottom-mounted cameras were characterized. These detectors, named X, Y, and Z, are mounted on Tecnai microscopes (FEI Company, The Netherlands) which were operated at 120 kV voltage. Two of the microscopes have a lanthanum hexaboride (LaB_6) tip as cathode, the third one a field emission gun (FEG). Each of the three CCD sensors has an active surface of $61.2 \times 61.2 \text{ mm}^2$, 4096×4096 pixels, a pixel pitch of $15 \mu\text{m}$, and a 100 % fill factor (<http://www.fairchildimaging.com/>). The cameras differ in the phosphor scintillator and fibre-optic plate that is coupled to the CCD sensor. The unbinned images were read out at 1 MHz by 4 parallel readout ports employing 16 bit AD converters; at the maximum speed one can obtain 7.5 unbinned images per minute. The square images are framed by respectively five (detector X and Y) and ten (detector Z) reference pixels in each direction: this frame should be excluded from the final image. All cameras are Peltier cooled to a set temperature of 248 K in order to decrease dark current. Image processing was done using MATLAB (Mathworks) and the *DIPimage* toolbox (TU Delft, The Netherlands, www.diplib.org). Data were collected using MATLAB scripts, inspired by the TOM toolbox [38] and employing the TEMScripting activeX server from Tecnai version 3.1.2 (<http://www.fei.com/products/types/fei-software.aspx>). All functions for camera characterizations can be found online at www.diplib.org/add-ons.

2.3.1 Removal of outliers

Dark reference images were acquired with the column valves closed, i.e. there was no beam. A series of at least ten images were acquired under identical conditions (with the same integration time). Pixels with intensity fluctuations larger than ten times the standard deviation of the intensity of a pixel within the series were marked as outliers. Occasionally, the iterative procedure identified two outliers within a series of ten. Outliers from the white reference images were removed in a similar way.

2.3.2 Bias, dark current and readout noise

After outlier removal, the bias and dark current was determined for every pixel by analyzing a total of 100 dark reference images measured at ten different exposure times. The range of exposure times was 0.05-10 s. The dark current was determined for each pixel from the slope

of a linear least-squares fit of the dark images *versus* exposure time. The offset of this fit gave the bias.

The ten dark reference images with the smallest exposure time were used to calculate the readout noise by computing the standard deviation per pixel within the series.

2.3.3 Detector effective gain measurements

Two different approaches were used to determine the effective gain of each detector, one based on white reference images and the other one based on gradient images.

White reference (I_{white}) images were acquired with different exposure times using a constant uniform illumination of the detector. The beam was spread to be wider than the diameter of the fluorescent screen (165 mm) and it was slightly (15 mm) shifted from the optical axis in random directions between the acquisitions of the successive images in order to average out any potential non-uniformities in the illumination. From these images the average background $\overline{I_{\text{bg}}}$ was subtracted. Outliers were removed as described above. Series of at least ten repeated exposures were made for seven different exposure times (range 0.05-2.5 s). Apart from the beam shift and exposure time, the illumination conditions were kept constant during the acquisition of all images. The spatial median of the intensities of all pixels within each quadrant was determined for each exposure time. Pixels with an average intensity that differed more than 1 % from this median were excluded in subsequent calculations. The variance of the pixel intensity within the series was determined for each selected pixel. The mean of the variance $\langle \text{var}(I) \rangle_{x,y}$ and the mean of intensities $\langle I \rangle_{x,y}$ over the selected pixels in each quadrant were computed. A plot of the mean variance *versus* the mean intensity was made with dots representing the pairs $[\langle I \rangle_{x,y}, \langle \text{var}(I) \rangle_{x,y}]$ for each exposure time. The slope of a linear least-squares fit of this plot gave the effective gain of the camera.

The effective gain was also determined using gradient images [77]. A series of at least ten repeated measurements were made of a highly non-uniform beam. For the LaB₆ microscopes, an intensity gradient was achieved by imaging the blurred beam edge at very high magnification. Since blurring of the beam edge is difficult to achieve for a FEG source, astigmatism of the condenser lens was used. Like in the aforementioned method, outliers were removed, the average background was subtracted, and the mask determined above was applied. The intensities in the gradient images were distributed into 100 bins. The variance and the mean of the intensity were calculated for each bin. The effective gain of the camera was again determined as described above.

2.3.4 Bias correction, gain normalization and pixel response

The white reference images described above were also used to check the linearity of the pixel response. Similar to the calculation of $\overline{I_{\text{bg}}}$ (Eq. (2.3)), a linear least-squares fit of intensity *versus* exposure time was computed for each pixel to yield the average white image $\overline{I_{\text{white}}}(x, y) = I_{\text{bg}}(x, y) + t_{\text{exp}} I_{\text{slope}}(x, y)$. The linearity of the pixel response as a function of the exposure time was checked by computing R^2 , the square of the sample correlation coefficient between the

measured and predicted values. The gain normalization image I_{gain} was calculated using $\overline{I_{\text{bg}}}$ and $\overline{I_{\text{white}}}$ (Eq. (2.5)).

Prior to the correction of a raw image using Eq. (2.1), border pixels had to be excluded from analysis. This border is 5 pixels wide for detector X and Y, and 10 pixels wide for detector Z.

2.3.5 Modulation transfer function (MTF)

Ten flat-field corrected, uniformly illuminated images taken with 1 s exposure time were used to calculate the MTF *via* the noise method. In order to avoid problems due to averaging too few data points at low spatial frequencies, we used a variable standard deviation for the Gaussian in Eq. (2.6), namely $\sigma = 2.5$ at low frequencies and $\sigma = 0.9$ at higher frequencies. Note that the angular averaging takes place in the reciprocal domain and that σ is expressed in the number of bins. Individual MTF curves were calculated for each of the ten images; the final noise-method MTF was an average of these.

Both the beam-stop and the diffraction aperture were used to determine the MTF *via* the edge method. The beam-stop was placed directly above the fluorescent screen under a slightly inclined angle with respect to the pixel array [78]. Ten images of the edge were taken with uniform illumination and subjected to outlier rejection and flat-field correction. The mean intensities on the bright and dark sides of the beam-stop were calculated and used to normalize the image. The average edge profiles from the slanted beam-stop edge were extracted from the image. The edge profiles were oversampled with a factor of eight and processed according to [79]. Averaging of 128 lines along the edge suppressed the noise and yielded a 1D edge-spread function (ESF). The point-spread function (PSF) of the detector was obtained by computing the derivative of the ESF *via* finite difference. Owing to Poisson statistics, it proved to be necessary to reduce the noise of the bright side of the edge by setting the tails of the PSF to zero. Individual MTFs were obtained after down-sampling the PSF to the original pixel pitch and computing the magnitude of the Fourier transform. We repeated this procedure for ten images and averaged the ten MTFs to obtain a more robust estimation.

Images of the diffraction aperture were taken at low magnification of the projection lens system ($1000\times$) and these images were normalized to yield an average value of one inside and zero outside the aperture hole. Edge profiles perpendicular to the edge were extracted, averaged, and further processed as described above. The curved edge of the aperture was found using the PLUS operator [80] with subpixel precision.

2.3.6 Conversion factor and detective quantum efficiency (DQE)

The conversion factor CF was measured by relating the beam current I_{beam} and exposure time t_{exp} to the integrated intensity O (in ADUs) in the corrected output image. The beam diameter was made to be smaller than the field of view of the camera to ensure that the detector captured all incident electrons. The incident beam current was obtained through the Tecnai TEMscripting activeX server interface, which reads the current from the fluorescence screen. For all microscopes, the incident beam current readings were postcalibrated using an independent current measurement from a Faraday cage of a double tilt analytical holder (Gatan, Inc., model 646).

A Keithley model 602 was used as a picoammeter. The conversion factor CF in ADUs per primary electron (ADU pe^{-1}) was calculated using the formula $CF = 1.6 \times 10^{-19} O / (I_{\text{beam}} t_{\text{exp}})$, where t_{exp} is the exposure time of the detector.

The MTF from the edge method was used for the DQE calculation. The NNPS was based on the subtraction of two raw uniformly illuminated, dark-subtracted images that were measured with the same exposure time $I_{\text{in}} = (I_1 - I_2) / \sqrt{2}$, with $I_1 = I_{\text{white1}} - \overline{I_{\text{bg}}}$ and $I_2 = I_{\text{white2}} - \overline{I_{\text{bg}}}$. The integrated flux $N = \langle I_1 + I_2 \rangle_{x,y} / (2CF)$ used for these images was 176, 149, and 124 primary electrons per pixel for detector X, Y, and Z, respectively. A sine-shaped windowing function (w) was applied to this image in order to avoid edge artifacts from the implementation of the discrete Fourier transform (\mathcal{F}). The square of the Fourier transform was multiplied by four to compensate for the power loss as a result of the windowing. Angular averaging of the spectrum was performed. The influence of the readout noise was represented through the term $\text{NNPS}_r = \mathcal{F}^2[I_{\text{rn}}(x, y) / CF]$. The normalized noise power spectrum NNPS was obtained from Eq. (2.11) and Eq. (2.13) after dividing the contributions from the Poisson noise and readout noise by the integrated flux N ,

$$\text{NNPS} = \frac{\langle 4\mathcal{F}[wI_{\text{in}}]^2 \rangle_{\phi} + \mathcal{F}[\frac{I_{\text{rn}}(x,y)}{C}]^2}{N}. \quad (2.14)$$

After determining the NNPS, the DQE was computed using Eq. (2.12).

Table 2.1. Characteristics of three in-house 4k TEM detectors at 120 kV.

The single chip sensors are read out from four different ports: upper left (UL), lower left (LL), upper right (UR) and lower right (LR).

Quadrant	Detector X				Detector Y				Detector Z			
	UL	UR	LL	LR	UL	UR	LL	LR	UL	UR	LL	LR
Bias (ADU)	496	498	487	485	505	504	508	505	1003	1002	1002	1002
Readout noise (ADU)	7.6	8.8	7.9	9.6	7.0	7.3	7.0	7.2	3.3	3.4	3.5	3.5
Readout noise (CCDe^{-}) [†]	11.4	13.2	11.8	14.4	10.5	10.9	10.5	10.8	9.8	10.2	10.5	10.5
Dark current($\text{ADU pixel}^{-1}\text{s}^{-1}$ / $\text{CCDe}^{-} \text{ pixel}^{-1}\text{s}^{-1}$)												
Mean	0.31/0.47				2.90/4.35				0.05/0.15			
Standard deviation	0.37/0.56				1.14/1.71				0.11/0.33			
No. of pixels with $\overline{I_{\text{dc}}} > 50 \text{ ADU pixel}^{-1}\text{s}^{-1}$	144				86				9			
Effective gain (ADU/pe^{-1})	5.0	5.6	5.3	5.6	7.7	7.7	7.7	7.7	4.6	4.6	4.8	5.0
Conversion factor (ADU/pe^{-1})	76				100				34			
MTF at 0.5 Nq (120 keV)	0.12				0.13				0.19			
DQE(0)	0.6				0.6				0.6			
DQE at 0.5 Nq	0.15				0.16				0.14			

[†] The nominal gain was estimated to be $1.5 \text{ CCDe}^{-} / \text{ADU}$ (binning 1) for detector X and Y and $3 \text{ CCDe}^{-} / \text{ADU}$ for detector Z.

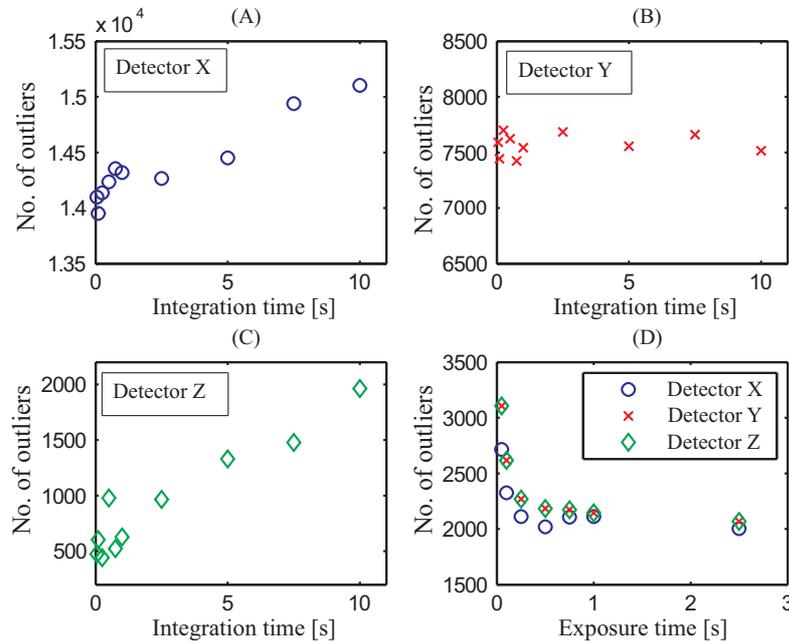


Fig. 2.1. Number of outliers *versus* time. Figures (A)-(C) show the number of outliers in the dark images and (D) shows those in the white reference images. Detector X and Z show an increase in the number of outliers with integration time in the dark references. The number of outliers in the white references was comparable for the three detectors.

2.4 Results

2.4.1 Outliers

Fig. 2.1 presents the number of outliers *versus* integration time (for the dark reference images) or exposure time (for the white reference images) for each detector. Fig. 2.1A and Fig. 2.1C show a comparable increase of almost 1500 outliers in the dark reference images when increasing the integration time from milliseconds to 10 s. This increase is not observed for detector Y (Fig. 2.1B). The number of outliers in the white reference images is highly similar for all the three detectors (Fig. 2.1D).

2.4.2 Bias and dark current

The bias in the images can differ for each of the four quadrants, as each readout port has its own AD converter. Fig. 2.2 shows a histogram of the bias for each quadrant of the three detectors. The average bias values for the four quadrants are 496, 498, 487, and 485 ADUs for detector X, 505, 504, 508, and 505 ADUs for detector Y, and 1003, 1002, 1002, and 1002 ADUs for detector Z (Table 2.1).

Fig. 2.3 shows a histogram of the dark current for each of the three detectors. The average (standard deviation) of the dark current is 0.31 (0.37), 2.9 (1.14), and 0.05 (0.11) $\text{ADU pixel}^{-1} \text{s}^{-1}$

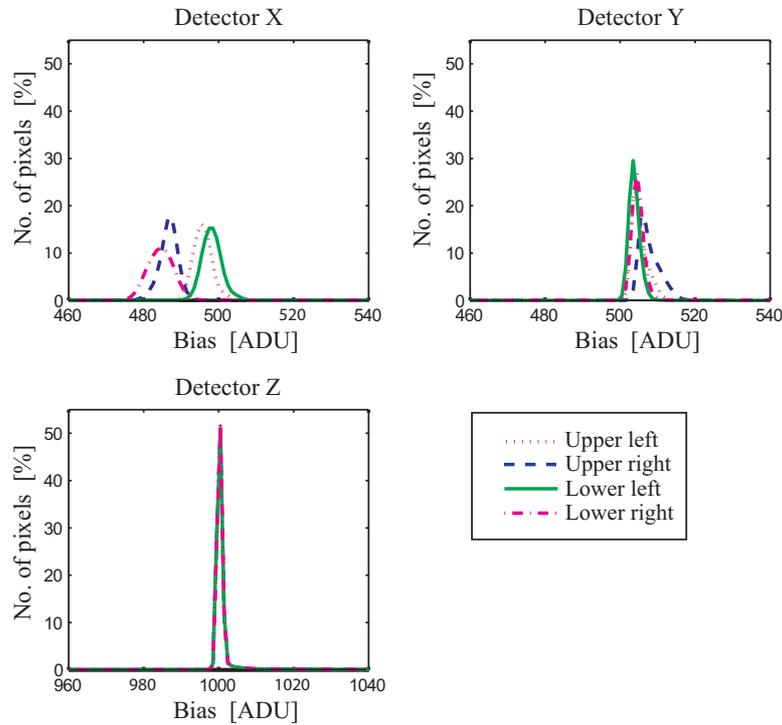


Fig. 2.2. Histogram of the bias for each quadrant of the three detectors. The bin width of the histogram is 1 ADU. Average bias is 492, 506 and 1002 ADUs for detectors X, Y, Z, respectively. Detector Z shows the smallest spread of the bias.

for detectors X, Y, and Z, respectively. Dark-current generation is a Poisson process. Therefore it is to be expected that pixels with a high dark current will also have a high standard deviation of the dark current.

Pixels with an excessive dark current are so-called *hot-pixels*. A complementary cumulative distribution of these is shown in Fig. 2.4. The numbers of pixels with a dark current larger than 100, 50, and 30 $\text{ADU pixel}^{-1} \text{s}^{-1}$ are 40, 144, 675 for detector X, 19, 86, 853 for detector Y, and 4, 9, 21 for detector Z, respectively.

2.4.3 Readout noise

Owing to the differences in readout circuitry, the readout noise is measured separately for each of the four quadrants of the image. The specification for the readout noise for a Fairchild CCD486 Image Sensor is 8 ADU (12 e^- with 1.5 CCDe^- per ADU nominal gain). The mean of the readout noise is 8.5 ADU for detector X, 7.1 ADU for detector Y, and 3.4 ADU for detector Z (Table 2.1). The nominal gain for each detector was determined from the comparison between full well capacity and saturation intensity in the image. It was estimated to be 1.5 CCDe^- per ADU for detector X and Y, and 3 CCDe^- per ADU for detector Z.

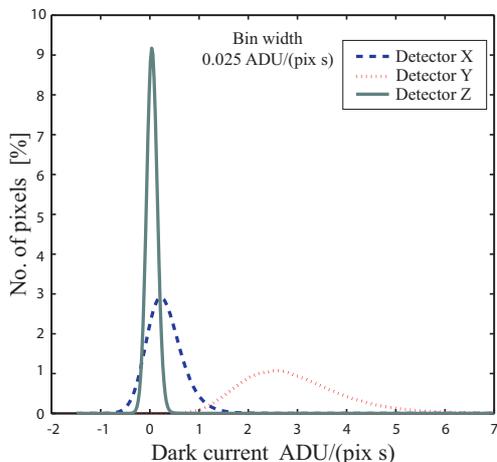


Fig. 2.3. Histogram of the dark current for three detectors. The bin width of the histogram is 0.025 ADU. Detector Z has the smallest spread of the dark current.

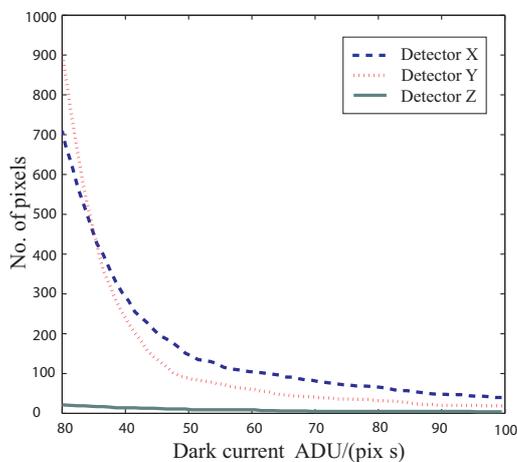


Fig. 2.4. Complementary cumulative distribution function of the hot pixels, showing the number of pixels that have a dark current higher than a certain value. The number of pixels with a dark current larger than 100 ADU pixel⁻¹s⁻¹ is 40, 19 and 4 for detectors X, Y, and Z, respectively.

2.4.4 Lookup tables

Fig. 2.5A depicts the average outlier-corrected and background-corrected I_{gain} image of detector X with normalized gain values. In close-up (Fig. 2.5B), the image fibre bundles and even the individual fibres can be clearly seen. Fig. 2.5C gives a close up of I_{gain} of detector Z, displayed at the same magnification as in Fig. 2.5B.

A mask is made for those pixels, where a very low signal was observed ($I_{\text{gain}} < 0.2$, e.g. due to dust or broken fibres), or where the signal was excessively large ($I_{\text{gain}} > 2$, e.g. due to thicker parts of the scintillator). The low and high threshold values (0.2 and 2.0, respectively) were selected empirically. Pixels within this mask could be either replaced by a value based on the mean and variance of the closest *normal* neighboring pixels, or remain marked as *unobserved* during subsequent processing. This mask forms a lookup table together with the list of pixel defects identified during the analysis of the dark reference images.

2.4.5 Linearity of the response

The linearity of response was assessed by making a linear least-squares fit to the intensity of the white reference images *versus* exposure time. R^2 was calculated for every pixel. For all three detectors, the linear response was good within the range of intensity values measured: R^2 was higher than 0.999 for almost all pixels. It proved to be unnecessary to extend the mask of bad

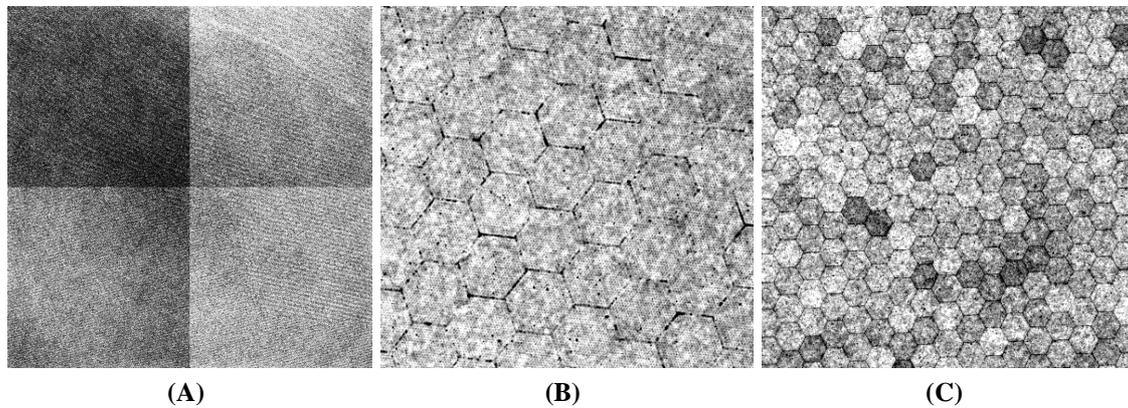


Fig. 2.5. The dark-corrected and scaled I_{gain} image. (A) Overview for detector X, showing the difference between the four quadrants. (B) Detail of (A), showing the fiber-optic coupling and individual fibers. (C) Detail of I_{gain} of detector Z, shown at the same magnification as in (B). The size of the fiber bundles is about 1.1 mm for detector X and Y and 450 μm for detector Z.

pixels (lookup table) with pixels that had a particular low R^2 value (e.g. < 0.9).

The average effective detector gain was calculated for each quadrant separately using both white reference and gradient images. Fig. 2.6 shows the effective gain for the upper right quadrant of the CCD using white reference images. Table 2.1 shows the effective gain in ADUs per primary electron for all quadrants. Detector X and Z have a comparable effective gain of on average 5 ADU/ pe^- . Detector Y gives more ADUs per primary electron (7.7), and its effective gain is very homogenous over each of the four quadrants. All three cameras showed excellent linearity of the variance in pixel response as a function of the pixel intensity. The method with gradient images was used for comparison and the effective gain values of the upper right quadrant are 5.6, 6.7 and 4.8 ADU/ pe^- for detectors X, Y, and Z respectively.

2.4.6 MTF

The modulation transfer function (MTF) was calculated with the noise (Fig. 2.7) and the edge method (Fig. 2.8). Both the beam-stop (Fig. 2.8A) and the diffraction aperture (Fig. 2.8B) were used to generate an edge. The MTF at half Nyquist was similar when determined with either of the two edge methods: 0.19 for detector Z, and 0.12 (beam-stop measurement) or 0.13 (aperture measurement) for detector X and Y. The MTF reached a higher minimum at higher frequencies for the noise method compared with the edge method.

The MTF of detector Z was also determined at 200 kV (Fig. 2.8A). It shows a more rapid decrease at lower frequencies. The MTF at half Nyquist was measured to be 0.19 at 120 kV and 0.13 at 200 kV.

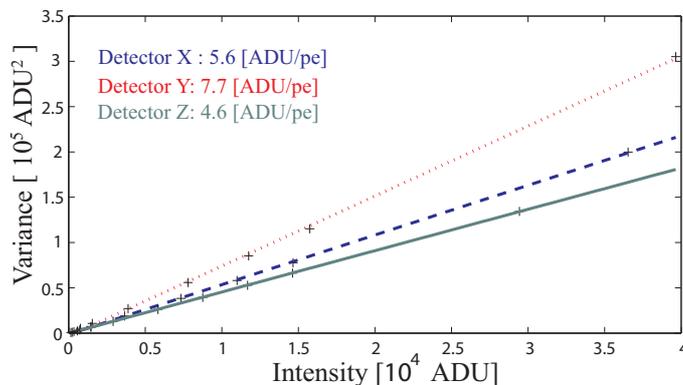


Fig. 2.6. Effective gain measurements from white reference images of the upper right quadrant for the detector X (dashed line), Y (dotted line) and Z (solid line). The values for the other quadrants are given in Table 2.1.

2.4.7 Conversion factor and DQE

The conversion factors at 120 kV as measured using the screen current method, are 76, 100, and 34 ADUs per primary electron for detector X, Y, and Z respectively. The conversion factor for detector Z at 200 kV was 23 ADU per primary electron. Fig. 2.9 shows the DQE for all three detectors. DQE at frequencies close to zero is about 0.6 for all three detectors (Table 2.1).

2.5 Discussion

Raw images provide useful system information. Quantification of noise based on raw (unprocessed) CCD images will give different numbers compared to quantification based on corrected (calibrated) images owing to image rescaling by flat-fielding. In this study, the characterization of the cameras was based on raw images, which could fortunately be obtained through scripting for all of our inhouse detectors. Data acquisition software, such as Digital Micrograph (<http://www.gatan.com/products/software/>) and Serial EM [67], typically collect one new dark reference image prior to the collection of each new series of images, thus ensuring that the dark reference image noise was representative for the imaging conditions used. The disadvantage of this approach is that the acquisition of new dark reference images for every new series of images takes time. Multiple dark reference images would be needed in order to reject zingers. The FEI Tecnai software (v.3.1.2; www.fei.com) also allows online dark subtraction, but relies on previously collected dark reference images. These images are collected for one exposure time only, making dark subtraction less accurate if deviating exposure times were used. The possible advantage of the use of a series of previously collected dark reference images (apart from gain in data-collection speed) is that more elaborate outlier-rejection schemes could be applied. Fig. 2.1 shows that the number of outliers, including decreased pixel values, can be quite substantial, up to 1 per 1000 (or 0.1 %) for detector X. The number of outliers in the dark reference images increased both for detector X and Z as a function of exposure time, with a rate of approximately

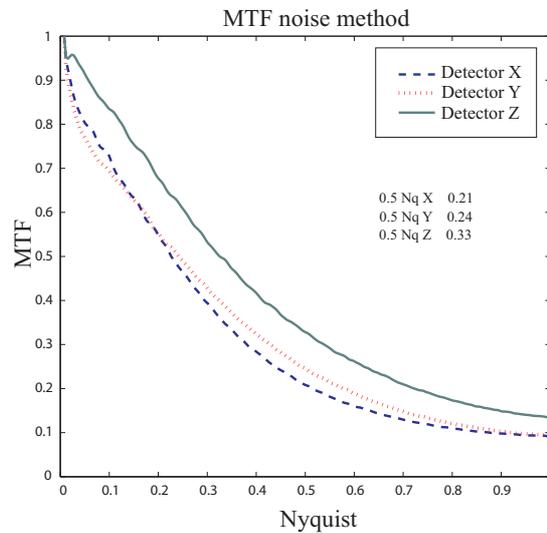


Fig. 2.7. MTF obtained with noise method. Radial averaging was performed with Gaussian rings in order to diminish discretization error.

150 pixels s^{-1} . This high rate can probably be attributed to the larger influence of cosmic rays and radioactive decay with increased integration time. In contrast, detector Y did not show such an increase. This detector has a much higher dark current (Fig. 2.3; Table 2.1), compared to detector X and Z. The increased noise level of detector Y at longer exposure times probably masks the detection of the increased occurrences of outliers as observed for the other two detectors. For all three detectors, the dark current is much higher than the tabulated nominal number of 0.005 $CCDe^{-}$ $pixel^{-1}s^{-1}$ for the 486 Fairchild sensor, cooled to 213 K. A doubling of the dark current for every 7 K of temperature increase (http://www.fairchildimaging.com/main/documents/Condor486-90_RevE.pdf) would suggest that these sensors, despite their identical set temperature of 248 K, are actually used at a temperature of 259 K (detector X), 281 K (detector Y), and 248 K (detector Z). An increased dark current could also be a consequence of radiation damage to the CCD itself [81]; however, detector Y was basically new at the time of characterization. We interpret these numbers as a strong indication that detectors X and Y are not cooled as well as detector Z.

Macromolecular crystallography (MX) CCD sensors are generally cooled to far lower temperatures compared with TEM CCD sensors. For instance, the Bruker APEXII detector (based on Fairchild 486 sensor) is cooled to 213 K and the Rayonix 165 detector to 203 K. The lower temperature is partly required because of the longer exposure times that are used at older X-ray sources and the lower conversion factors for X-ray photons compared to high-energy electrons. These X-ray detectors are thermally isolated units that are placed separately from the goniometer holding the specimen. This allows these detector manufacturers to accurately control the vacuum and temperature of the CCD, overcoming the need for routine recalibration. In contrast, TEM detectors are directly mounted on the electron microscope in a vacuum that is controlled by the electron microscope manufacturer rather than the detector manufacturer. This vacuum also contains the specimen, films etc., and is therefore not guaranteed to be of constant quality.

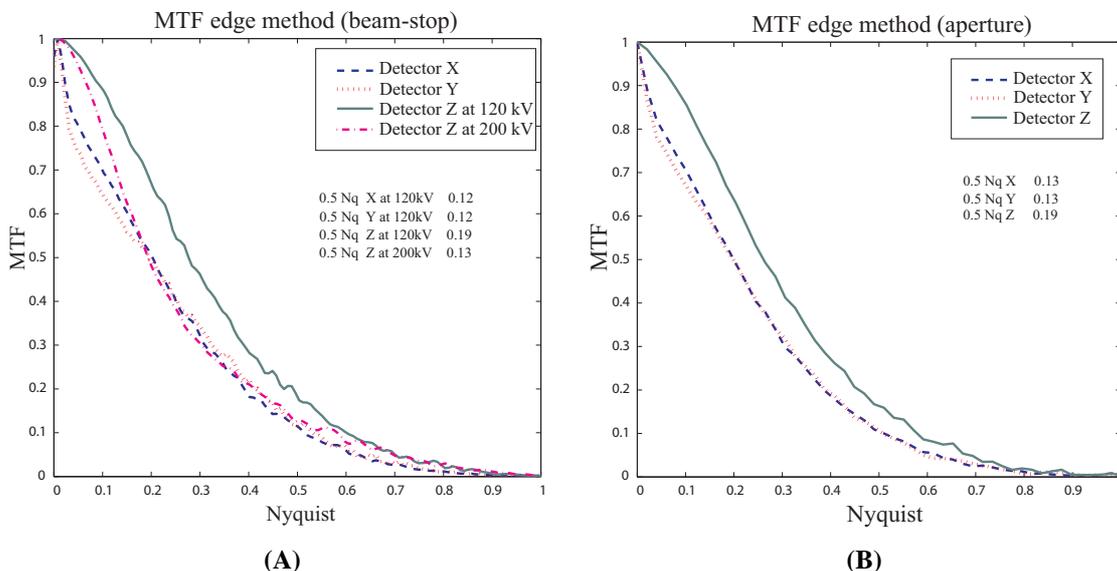


Fig. 2.8. MTF obtained with the edge method employing (A) a beam-stop (detector Z is characterized at 120 kV and 200 kV) and (B) an aperture.

Deeper cooling of the CCD sensor and the coupled fibre-optic plate could result in condensation on the detector surface. In contrast to X-ray CCD detectors, TEM detectors do seem to require repetitive recalibrations. It is our impression that the frequency of calibration could be lessened if the vacuum and cooling conditions of the camera could be better controlled; i.e. to a standard comparable to those of MX detectors. A more constant and deeper cooling of the TEM detector would allow the use of more accurate bias-correction and gain-correction schemes, faster data collection (no need to recollect dark image every time), and a decoupling of the correction for CCD fixed-pattern noise from the correction for beam inhomogeneities.

Fig. 2.4 shows the number of pixels with a dark current higher than a certain threshold. Various criteria can define a hot pixel, for example a dark current higher than ten times the average dark current, or dark signals higher than one per 1000 of the maximum encoding range at the nominal exposure time [53]. Fig. 2.4 seems to strongly favor detector Z above detector X and Y, but this difference would be less striking if the first criterion would have been used. The *hottest* pixels, particularly for detector X, will saturate, with the column valves closed, if exposure times between 10 and 60 s are used; leakage will result in pixel column defects. Not all detector manufacturers give image-blemish grades (point, cluster and column defects) as this is a delicate balance between system cost, industrial state-of-the-art and actual experimental needs. As long as no recalibration of the detector is needed, *hot* pixels can be reliably identified and taken into account during subsequent data processing by either replacing them by a value based on the statistics of neighboring pixels or marking them as *unknown*. This lookup table will also contain extreme values from the gain-normalized image as obtained using Eq. (2.5).

The impact of the use of lookup tables for image correction becomes particularly apparent during the calculation of cross-correlation functions with the purpose of measuring image

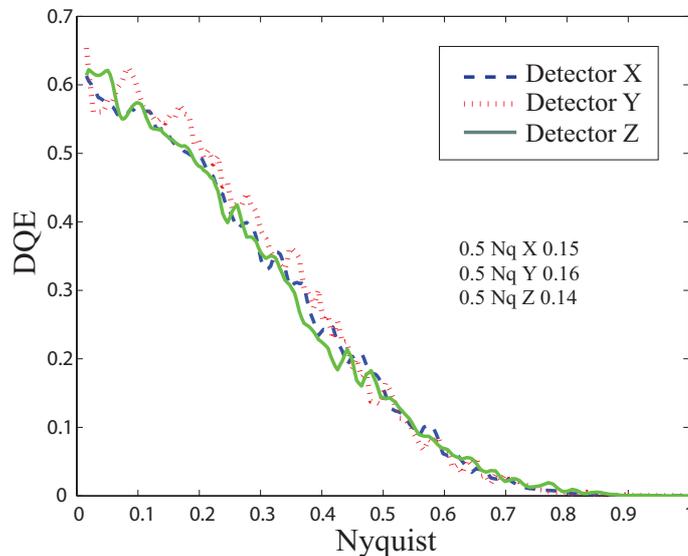


Fig. 2.9. DQE for three detectors, measured with an integrated flux of 182, 146, and 80 primary electrons per pixel for detectors X, Y and Z respectively. The similarity in the graphs indicates that the lower MTFs of detectors X and Y are compensated for by large conversion factors.

shifts. Image shifts are often measured in automated TEM procedures, e.g. during automated tomographic data collection [82]. These automated procedures fail if image shifts are not measured correctly. Therefore, if the fixed-pattern noise is not fully accounted for then images are not measured correctly because of the appearance of an additional, undesired peak at the origin of the cross-correlation function. This origin peak corresponds to the unshifted fixed-pattern between the two images. The height (intensity) of the origin-peak can dominate the true cross-correlation peak when low contrast specimens such as vitrified biological materials are imaged. Under these conditions, the true correlation peak will be relatively low and the appearance of an origin peak due to imperfect calibration may well pose limits to reliable automation. The use of lookup tables could mitigate part of the problem of fixed-pattern noise, but unfortunately not all software packages can employ these at present. Correction of the raw images with our own average dark and white images virtually eliminates the cross-correlation origin peak.

Uncorrected systematic outliers could result in undesirable artifacts if the data is used for 3D reconstructions. State-of-the-art tomographic reconstruction packages, such as IMOD [83] and Inspect3D (www.fei.com), can use statistical criteria to identify and correct cosmic rays and detector flaws prior to reconstruction. However, more subtle systematic errors would still propagate unless adequate lookup tables are used.

The use of four readout ports of data from a CCD chip can result in both bias (Fig. 2.2) and gain (Fig. 2.5A) inhomogeneities. Gain inhomogeneities in corrected diffraction images of $\pm 1\%$ or less with respect to the average values are deemed to be acceptable [53]. The quadrant gain inhomogeneities in the raw images are less than 1% for detector Z, whereas they are around 6% for detector X and Y. An improper correction of poorly balanced offsets could lead to quadrant-edge effects, especially in Fourier domains [68]. Correction will be more precise if

the spread of the bias is smaller.

The conversion factor is rather large for detector X (76 ADU/pe⁻) and Y (100 ADU/pe⁻), whereas detector Z has a conversion factor (34 ADU/pe⁻ at 120 kV, 23 ADU/pe⁻ at 200 kV) that is close to values given in literature for that type of detector. The readout noise, in CCDe⁻, is slightly higher for detector X than detector Y and Z (Table 2.1). The effective gain is rather similar for detector X (5.3 ADU/pe⁻) and for detector Z (4.8 ADU/pe⁻), whereas detector Y is the most sensitive with an effective gain of 7.7 ADU/pe⁻. For simulated data, the effective gain will converge to the conversion factor for increased pixel binning as the dampening effect of point-spread function will decrease for higher binning. For the data presented here, 16 × 16 rebinning reduces the difference between effective gain and conversion factor to less than 10%. However, for higher binning, the effective gain does not converge to the conversion factor due to detector response inhomogeneities [53].

Both the edge and the noise method give a comparable relative ranking of the three detectors. Detector Z shows a better propagation at low frequency compared to detectors X and Y. Even for 200 kV electrons, detector Z looks better between 0 and 0.2 Nyquist rate compared to detectors X and Y for 120 kV electrons, whereas it is comparable at higher frequencies. For higher voltages of the electron source, the percentage of electrons that are backscattered from the support layer of the CCD camera will be higher [84]. They re-enter the scintillator and give rise to intensity at a large lateral distance from the place they initially hit the scintillator and cause the more rapid decrease of signal for low frequencies. By changing the thickness of the phosphor layer one can alter the balance between sensitivity and resolution as a thicker layer gives a better sensitivity but also larger point spread. This might explain why detector X and Y have better sensitivity but lower resolution compared to detector Z, although the differences in size of the fibre optic bundles (1.1 mm for detector X, Fig. 2.5B; 450 μm for detector Z, Fig. 2.5C) are also likely to have an effect on the MTF at low resolution. The noise method gives too optimistic values for the MTF at higher frequencies where the noise contributions of the camera start to dominate. Both the beam-stop (Fig. 2.8B) and the aperture (Fig. 2.8A) MTF graphs approximate zero towards Nyquist frequency, which is reported to be an over-pessimistic estimate of the true MTF [56].

For all three detectors, the DQE at frequencies close to zero-frequency is about 0.6. Measurement errors in conversion factor would give proportional errors in the DQE measurement (Eq. (2.10)). TEM detectors with larger pixel sizes can show even better DQE(0) values of 0.8 [85] or 0.76 [84]. The normalized noise power spectrum will be integrated flux-dependent; Fig. 2.9 shows the DQE for our three detectors measured with a relatively high integrated flux of 182, 146, and 80 primary electrons per pixel. Overall, the DQEs of the more sensitive detectors X and Y are remarkably comparable with the DQE of the sharper and less noisy detector Z (Fig. 2.9).

A number of programs exists to aid macromolecular crystallographers in planning their data-collection strategy [86–88]. From one or a few images, these programs will characterize the specimen, simulate data-set statistics for different combinations of data-collection parameters, and suggest the most optimal ones. The program Best [88] honors its name by being able to suggest an optimal data-collection strategy based on the most complete set of parameters. These include anisotropic diffraction, background scattering, detector statistics, geometric pa-

rameters, and even radiation damage [89, 90]. Given a number of test images, the program will suggest exposure time, rotation range, number of images and starting angle, and predict data-set statistics such as signal-to-noise *versus* resolution for each suggestion. One could imagine a similar scheme for (cryo-) electron microscopy, in particular for tomographic data collection, where the effect of parameters such as defocus, rotation steps, number of angles, single *versus* double tilt, integrated electron flux and electron flux, magnification and detector binning could be simulated after an initial characterization of the specimen with a small number of test images. A detailed knowledge of all the parameters involved, including the characteristics of the camera as determined here, will aid the development of an *expert system* [91] that will aid the electron microscopist to make objective and reproducible decisions for their (tomographic) data collection.

2.6 Conclusion

A general methodology for characterizing TEM CCD detectors has been presented. The set of algorithms have been added to the publicly available image processing toolbox for MATLAB (www.diplib.org/add-ons) to allow non-expert electron microscopy users to characterize, based on uncorrected images, the properties of their CCD detector. Furthermore, it can facilitate information exchange between detector users and producers. Three 4k inhouse CCDs have been characterized, showing different strengths in terms of sensitivity, resolution, DQE and noise. The need for the use of lookup tables is demonstrated. Fixed pattern noise could be fully accounted for by using large sets of dark and white reference images. Unfortunately, the noise patterns seem to drift in time, possible because of unstable cooling of the CCD sensors, thereby limiting the useful lifetime of these reference sets.

Acknowledgements

The authors thank Frank Faas and Montserrat Barcena for useful discussions, Remco Schoenmakers for critical reading of the manuscript and Paul Mooney for crucial help in the determination of the conversion factors by calibration of the incident beam currents.

Chapter 3

Defocus and astigmatism estimation

Published as [49]: M. Vulović, E. Franken, R. B. G. Ravelli, L. J. van Vliet, B. Rieger, “Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy”, *Ultramicroscopy* 116 (1) (2012) 115-134.

with additional content from [48] M. Vulovic, P. Brandt, R. B. G. Ravelli, A. J. Koster, L. J. van Vliet, B. Rieger, “Estimation of defocus and astigmatism in transmission electron microscopy”, *In Proc. of the IEEE International Symposium on Biomedical Imaging* (2010) 1121-1124.

Abstract

Defocus and twofold astigmatism are the key parameters governing the contrast transfer function (CTF) in transmission electron microscopy (TEM) of weak-phase objects. We present a new algorithm to estimate these aberrations and the associated uncertainties. Tests show very good agreement between simulated and estimated defocus and astigmatism. We evaluate the reproducibility of the algorithm on experimental data by repeating measurements of an amorphous sample under identical imaging conditions and by analyzing the linearity of the stigmator response. By using a new Thon ring averaging method, the modulation depth of the rings in a 1D averaged power spectrum density (PSD) can be enhanced compared to elliptical averaging. This facilitates a better contrast transfer assessment in the presence of spherical aberration. Our algorithm for defocus and astigmatism estimation inverts the contrast of the Thon rings and suppresses the background in the PSD using an adaptive filtering strategy. Template matching with kernels of various ellipticities is applied to the filtered PSD after transformation into polar coordinates. Maxima in the resulting 3D parameter space provide multiple estimates of the long axis orientation, frequencies and apparent ellipticities of the rings. The frequencies of the detected rings, together with outlier rejection and assignment of an order to the CTF zeros, are used to estimate the defocus and its uncertainty. From estimations of defocus and ellipticity, we derive astigmatism and its uncertainty. A two-pass approach refines the astigmatism and defocus estimate by taking into account the influence of the known spherical aberration on the shape and frequencies of the rings. The implementation of the presented algorithm is freely available for non-commercial use.

3.1 Introduction

In order to improve resolution and allow reliable quantitative image analysis in transmission electron microscopy (TEM), it is essential to account for the effects of the oscillating contrast transfer function (CTF) on the image formation, the elastic and inelastic scattering properties of the sample, and the effects of the TEM detector. Determination of the CTF parameters, especially defocus and twofold astigmatism, is crucial in designing post-processing strategies to account for the effect of the CTF and for interpretation of the images at spatial frequencies beyond the first zero of the CTF. Additionally, in high resolution electron microscopy (HREM), the unbiased and precise estimation of defocus and astigmatism forms the basis for the assessment of the maximal contrast transfer of the microscope, the optimal adjustment of aberration correctors, exit wave reconstruction, and the modeling of image formation.

Early descriptions of the influence of these aberrations on the CTF can be found in [92, 93]. One of the most commonly used autofocus routines in TEM (especially for life-sciences) is based on a beam-tilt induced image displacement [94]. In order to obtain accurate estimates of defocus and astigmatism it is desirable to measure them from diffractograms of an amorphous sample, and avoid changes of the imaging conditions and possible introduction of higher order aberrations due to tilting of the beam. Many methods [31, 33, 35, 36, 40–46, 95–100] base the CTF parameters estimation on the patterns in a diffractogram known as Thon rings [93] (see also Fig. 3.1B). The CTF parameters are usually estimated by minimizing the discrepancy between the background-subtracted power spectrum densities (PSD) of simulated and measured projections [41, 42, 44–46, 95–100].

Some methods use 1D radial profiles obtained from circular averaging of 2D experimental PSD [41, 95, 97] or by elliptical averaging [40]. An inadequacy of circular averaging is that it neglects astigmatism. Astigmatism distorts the circular shape of the Thon rings and thus decreases their modulation depth in the obtained 1D profile. A few algorithms that consider astigmatism involve concepts such as dividing the PSD into sectors where Thon rings are approximated by circular arcs [44, 101], applying Canny edge detection to find the rings [40] prior to elliptical averaging, determining the relationship between the 1D circular averages with and without astigmatism [102], or using a brute-force scan of a database containing precalculated patterns as in ATLAS [103]. Some other approaches for estimating CTF parameters do a fully 2D PSD optimization [46, 98–100] but they usually regulate and fit numerous parameters by an extensive search that does not guarantee convergence. Furthermore, only a few schemes that were developed for defocus estimation provide an error analysis [103, 104].

The background in the PSD hampers the Thon ring detection and therefore should be suppressed prior to estimation of defocus and astigmatism. The background dominates at low frequencies and originates from various contributions such as inelastic scattering, camera noise, and object structure. At high frequencies the oscillations are damped by the envelopes originating from the energy spread, finite source size, and the detector's modulation transfer function (MTF); as a result they submerge in the noise. Most state-of-the-art algorithms for defocus determination mentioned above [40–46] base their estimation on procedures that calculate a 1D averaged PSD, fit a non-linear background model through the PSD minima, and finally subtract it in order to extract the CTF oscillations. Background fitting, however, is a difficult step and

often introduces systematic errors as no true model for background can be generated and the fitting is sensitive to the shape and the frequency range of the fitted model function. In [48] we analyzed the robustness of an approach based on background subtraction by characterizing the defocus estimation from each CTF zero individually. The minima at low frequencies were less reliable since they depend strongly on background subtraction. Hence, it is desirable to avoid fitting of a background function through the local PSD minima.

The precision of quantitative HREM image analysis is often limited by the precision of the related aberration estimations. The latest instrumentation improvements of aberration correctors require high precision and low bias of aberration estimates. For determination of higher-order aberrations, the Zemlin-tableau method [105] is commonly used which relies on accurate measurements of lower-order aberrations and requires acquisition of a number of images. In HREM, some of the alternative methods to Thon ring pattern recognition include estimation of defocus and astigmatism from crystalline regions [106] or using defocus series [107]. A number of algorithms developed for materials science applications report small absolute errors in defocus and astigmatism [103, 106–110]. However, none of these algorithms consider estimation of small astigmatism (few nm) at high defocus values (order of a few microns) which implies very small ellipticity of Thon rings. Such settings are common for life-sciences applications where phase contrast imaging is used mostly at significant defocus.

Most state-of-the-art algorithms mentioned above are sensitive to background estimation and subtraction, thresholding of the PSD, and involve numerous intermediate steps that must be optimized. Peaks in diffractograms from crystalline material, incomplete appearance of the rings in a certain direction as a result of astigmatism, temporal envelope and/or sample drift represent an additional challenge [103]. Furthermore, the presence of spherical aberration (C_s) changes the frequency and shape of individual Thon rings, such that they can be only in approximation considered as ellipses. Although elliptical averaging (e.g. [40]) of the PSD is an improvement over the commonly used circular averaging, none of the approaches so far have included the influence of C_s on the shape of the rings in the averaging procedure to get one-dimensional Thon ring profiles; this becomes more important for a relatively small ratio between defocus and spherical aberration terms in the aberration function.

This paper presents and validates an unbiased and precise algorithm to automatically estimate defocus and twofold astigmatism from diffractogram(s) of an amorphous sample together with the corresponding uncertainties. We assume that astigmatism is smaller than defocus, i.e. Thon rings are approximately elliptical. This requirement is typically met in life sciences applications where defocus is in the micrometers range. The algorithm, however, can also be applied to a range of parameter settings typical for materials science as long as the defocus is larger than astigmatism. The algorithm has been implemented in DIPimage, a MATLAB toolbox for scientific image processing and analysis, and will be freely available for noncommercial use *via* email upon request (<http://www.diplib.org/add-ons>).

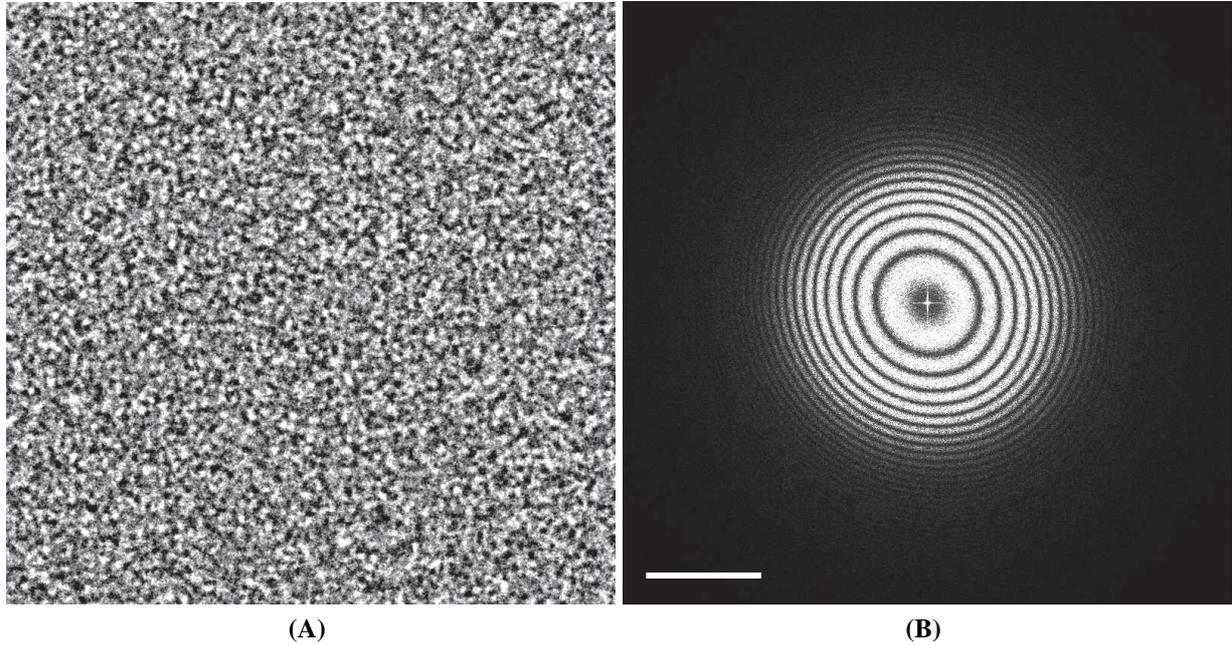


Fig. 3.1. (A) PtIr sample imaged at a requested microscope underfocus of 1000 nm and magnification of 62 kx; (B) Power spectrum density (PSD) of the same image showing Thon rings that are not perfectly circular due to astigmatism. The scale bar corresponds to 0.5 nm^{-1} .

3.2 Theory

3.2.1 Phase contrast

In approximation, image formation of weakly scattering objects in TEM can be considered as a linear process. For non-tilted and thin specimens, the defocus is constant across the field of view and therefore, the CTF is space-invariant. Phase contrast occurs as a result of interference between the unscattered part of the electron exit wave function and the elastically scattered part from the specimen. The electron wave is further subject to a frequency dependent phase shift introduced by the microscope aberrations. If we consider spherical aberration, defocus and twofold astigmatism, the total aberration function is

$$\chi(q, \alpha) = \frac{2\pi}{\lambda} \left(\frac{1}{4} C_s \lambda^4 q^4 - \frac{1}{2} \Delta f(\alpha) \lambda^2 q^2 \right) \quad (3.1)$$

where q is the magnitude of the spatial frequency (q_x, q_y). The relativistic electron wavelength λ depends on the energy of the incident electrons. It is assumed that the spherical aberration C_s is known. The defocus at eucentric height is Δf . We use the convention that underfocus implies $\Delta f > 0$, as in [111]. Twofold astigmatism (A_1, α_1) describes the azimuthal variation of (de)focus

$$\Delta f(\alpha) = \Delta f - A_1 \cos(2(\alpha - \alpha_1)). \quad (3.2)$$

The same sign convention is applied to A_1 as to defocus ($A_1 > 0$ corresponds to underfocus, and $\text{sgn}(A_1) = \text{sgn}(\Delta f)$). Fig. 3.2 illustrates the change of sign of A_1 while altering between underfocus and overfocus due to the fact that the focal distances of the tangential and the meridian rays interchange. The transfer function of the lens system is [111]

$$T(q, \alpha) = e^{-i\chi(q, \alpha)}. \quad (3.3)$$

The Fourier transform ($\mathcal{F}[\circ]$) of the electron wave at the back focal plain is given by

$$\tilde{\Psi}(q, \alpha) = \mathcal{F}\left[e^{i\sigma v_z(x, y)}\right] T(q, \alpha) \quad (3.4)$$

where $v_z(x, y) = \int V(x, y, z) dz$ describes the projected scattering potential of the sample in z -direction of the incident electrons, $\sigma = \lambda me / (2\pi\hbar^2)$ is the interaction constant, and the tilde refers to the Fourier domain. Finally, the intensity in the image plane is defined as

$$I(x, y) = |\Psi(x, y)|^2. \quad (3.5)$$

3.2.2 Partial coherence and amplitude contrast

The energy spread and the finite source size introduce temporal and spatial incoherence respectively. These can be modeled as damping envelopes in the spatial frequency domain. The temporal incoherency of the source can be modeled as a chromatic envelope function [111]:

$$K_c(q) = \exp\left[-\left(\frac{\pi\lambda q^2 C_c \Delta E}{4E \sqrt{\ln 2}}\right)^2\right]. \quad (3.6)$$

Here C_c is the chromatic aberration coefficient, which is usually of the same order of magnitude as C_s (a few mm). The energy of the incident electrons is E and the energy spread ΔE is around 1 – 2 eV for thermionic guns (LaB₆) and 0.3 – 0.5 eV for field-emission guns (FEG). See Table 3.1 for specifications used here. In the case of non-tilted illumination, K_c does not exhibit azimuthal dependency [112]. Furthermore, the finite source size introduces spatial incoherency which results in the spatial envelope:

$$K_s(q, \alpha) = \exp\left[-\frac{(\pi C_s \lambda^2 q^3 - \pi \Delta f(\alpha) q)^2 \alpha_i^2}{\ln 2}\right] \quad (3.7)$$

where α_i is the illumination aperture that is usually in the order of tenths or hundredths of mrad. The total incoherency of the source can be summarized as

$$K(q, \alpha) = K_s(q, \alpha) K_c(q). \quad (3.8)$$

Furthermore, the thickness of the sample (d) induces another damping envelope [29]

$$K_d(q) = \text{sinc}\left(\frac{1}{2}\lambda q^2 d\right).$$

In our analysis, however, we assume that the influence of $K_d(q)$ is negligible compared to $K(q, \alpha)$. The influence of the objective aperture is described as

$$A_p(q) = \begin{cases} 1 & |q| \leq q_{\text{cut}}, \\ 0 & |q| > q_{\text{cut}} \end{cases} \quad (3.9)$$

where $q_{\text{cut}} = 2\pi d_{\text{ap}}/(f\lambda)$ is the cut-off frequency, d_{ap} is the physical diameter of the aperture and f is the focal length of the objective lens. The amplitude contrast attenuation can be modeled by an imaginary term in the projected potential

$$v_z(x, y) = V_z(x, y) + i\Lambda_z(x, y). \quad (3.10)$$

The amount of amplitude contrast is given by the ratio of the attenuation term to the magnitude of the projected potential

$$W(q) = \frac{\tilde{\Lambda}_z(q)}{\sqrt{\tilde{\Lambda}_z(q)^2 + \tilde{V}_z(q)^2}}. \quad (3.11)$$

3.2.3 Weak-phase weak-amplitude object

In order to estimate the CTF parameters, the sample properties must be known. For that purpose the most convenient specimens are amorphous films. It is assumed that the overlap of atomic positions in a projection is significant and that the projected amorphous sample is essentially noise with a flat frequency spectrum. This is surely an approximation as every real specimen has limited scattering power. The mean inner potential of the sample introduces a constant phase change of the electron wave which can be neglected in this analysis as it is frequency independent. With these assumptions, the projected potential $v_z(x, y)$ is known and allows us to extract the CTF from the recorded image intensity. The total intensity for a weak-phase, weak-amplitude object is similarly as in [41, 113] given by

$$I_0(x, y) = \mathcal{F}^{-1} \left[\delta(q) + \sigma \tilde{V}_z(q) \text{CTF}(q, \alpha) \right] \quad (3.12)$$

and the CTF is

$$\text{CTF}(q, \alpha) = 2A_p(q)K(q, \alpha) \sin(\chi(q, \alpha) - \Phi_a(q)) \quad (3.13)$$

where $\Phi_a(q) = \arcsin(W(q))$. We refer to Appendix A for detailed derivation of Eqs. (3.12) and (3.13).

3.2.4 Detector response

The measurement process yields Poisson noise, adds readout noise I_{rn} and dark current I_{dc} to the final image, and blurs the image with a detector point-spread function PSF(x, y)

$$I(x, y) = \left[CF \cdot N_{\text{pois}}(\Phi_e \cdot I_0(x, y)) \right] * \text{PSF}(x, y) + I_{\text{rn}} + I_{\text{dc}} \quad (3.14)$$

where $N_{\text{pois}}(A)$ denotes Poisson noise yield, CF is the conversion factor of the camera in [ADU/ e^-], $\Phi_e \cdot I_0(x, y)$ is the incident integrated electron flux in [e^-/area], and $*$ represents the 2D convolution operator.

3.2.5 Power spectrum density and ellipticity of Thon rings due to the astigmatism

The PSD of a mean-subtracted image is given by

$$P(q, \alpha) = |\mathcal{F}[I(x, y) - \langle I(x, y) \rangle_{x,y}]|^2 \quad (3.15)$$

where $\langle I \rangle_{x,y}$ denotes the mean intensity of the image. The minima in the PSD correspond to the zeros of Eq. (3.1). Fig. 3.1B displays the PSD of a recorded image of PtIr (platinum-iridium) showing a pattern referred to as Thon rings [93]. The observed contrast is minimal (Thon rings frequencies) when the CTF is zero. That occurs for zeros of the sine term in Eq. (3.13)

$$\chi(q, \alpha) - \Phi_a(q) = k\pi, \quad k \in \mathbb{Z}. \quad (3.16)$$

The location of a CTF zero depends on the defocus, the accelerating voltage, and the spherical aberration. By including the amplitude contrast into a so called effective k_{eff} we get

$$k_{\text{eff}} = k + \frac{\Phi_a}{\pi}. \quad (3.17)$$

For thin objects $k_{\text{eff}} \approx k$ usually holds, but we will keep k_{eff} for generality.

The shape of the Thon rings in the PSD is circular if no astigmatism is present. With increasing astigmatism (and $C_s \approx 0$) the shape gradually transits from elliptical to parabolic and hyperbolic. In the following, it is assumed that the astigmatism is not excessive such that the PSD contains near-elliptical equi-phase contours. The q^2 term in Eq. (3.1) has an azimuthal dependency ($\Delta f(\alpha)$), whereas the q^4 term with C_s is isotropic. This results in a shape of Thon ring which is not perfectly elliptical, especially for high frequencies. Let us for a moment consider the case without spherical aberration. The influence of C_s on the rings will be addressed later (see 3.3.6). In the case $C_s = 0$, the rings are ellipses and the position of the CTF zeros can be found from

$$\pi q^2 \lambda (-\Delta f + A_1 \cos(2(\alpha - \alpha_1))) = k_{\text{eff}} \pi. \quad (3.18)$$

From this expression we can find that the defocus in the direction of the long axis ($\alpha = \alpha_1$) of the Thon rings is given by

$$-\Delta f_l = \frac{k_{\text{eff}}}{\lambda q_l^2} \quad (3.19a)$$

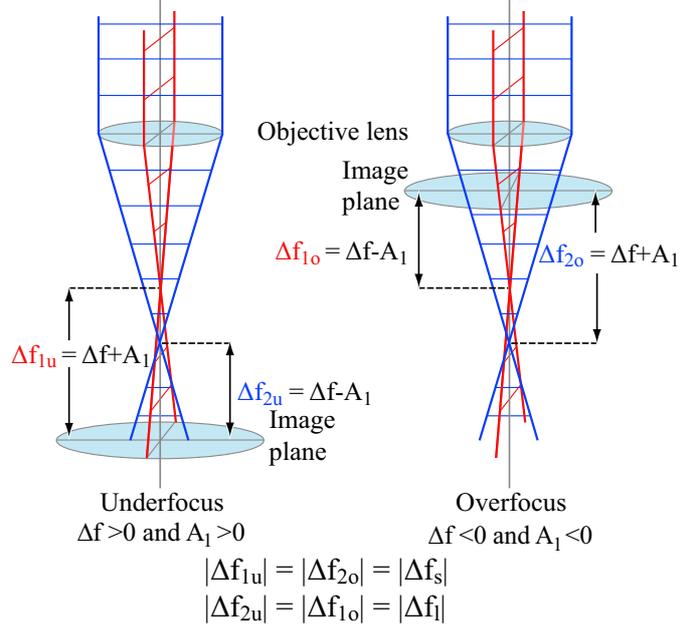
$$\text{with } \Delta f_l = \Delta f - A_1. \quad (3.19b)$$

Similarly, for the short axis ($\alpha = \alpha_1 \pm \pi/2$) we find

$$-\Delta f_s = \frac{k_{\text{eff}}}{\lambda q_s^2} \quad (3.20a)$$

$$\text{with } \Delta f_s = \Delta f + A_1. \quad (3.20b)$$

Fig. 3.2. Defocus and astigmatism follow the same sign convention $\Delta f > 0$, $A_1 > 0$ for underfocus and $\Delta f < 0$, $A_1 < 0$ for overfocus. Focal distances of tangential and meridian rays interchange while altering between underfocus and overfocus ($|\Delta f_{1u}| = |\Delta f_{2o}|$ and $|\Delta f_{2u}| = |\Delta f_{1o}|$). These defoci correspond respectively to the short q_s (Δf_s) and long q_l (Δf_l) axis of the Thon rings.



The frequencies q_l and q_s represent the PSD minima in the long and short axis direction respectively; Δf_l and Δf_s are the corresponding defoci. It holds that $q_s < q_l$ and $|\Delta f_s| > |\Delta f_l|$. The ellipticity of a Thon ring is given by

$$R_0 = \sqrt{\frac{\Delta f_s}{\Delta f_l}} = \sqrt{\frac{\Delta f + A_1}{\Delta f - A_1}}, \quad R_0^2 \geq 1. \quad (3.21)$$

In the case $C_s = 0$, the ellipticity represents the ratio between the long and short axes of the ellipse

$$R_0 = \frac{q_l}{q_s}. \quad (3.22)$$

The twofold astigmatism is then derived from the defocus Δf and the ellipticity R_0 as

$$A_1 = \Delta f \frac{R_0^2 - 1}{R_0^2 + 1}. \quad (3.23)$$

3.3 The algorithm

An overview of the algorithm is shown in Fig. 3.3. In the first step, the PSD is obtained using Eq. (3.15). Then, the PSD contrast is inverted, the background suppressed, and the pattern denoised by an adaptive filtering procedure. Subsequently, in step 3 the PSD is resampled to polar coordinates. In this polar power spectrum image, Thon rings manifest themselves as straight lines when there is no astigmatism, or 'sine-like' curves when there is astigmatism present. The Thon rings can be found by probing the polar power spectrum image with templates (step 4)

that resemble this expected Thon ring shape. This leads to a three-dimensional parameter space of frequency, orientation, and Thon ring ellipticity (step 5). In this space, the most dominant orientation and ellipticity of the Thon rings as well as their frequency are found by analyzing the local maxima. A model curve is fitted through the detected maxima peaks. The fit results in an estimate for the equivalent ellipticity R_0 , as defined in Eq. (3.21), which corresponds to the apparent ellipticity at the frequency of generated templates (step 6). Using the frequency of the found rings and by incorporating mechanisms (step 7) to remove outliers (false positives) and being able to deal with missing Thon rings (false negatives), the defocus value can be estimated. From the defocus value and ellipticity, the astigmatism can finally be calculated (step 8) using Eq. (3.23). If the ratio between the defocus and spherical aberration terms in Eq. (3.1) is low, we use a two-step approach and refine the initial astigmatism and defocus estimates (steps 6, 7, 8).

The next subsections explain all steps in more detail.

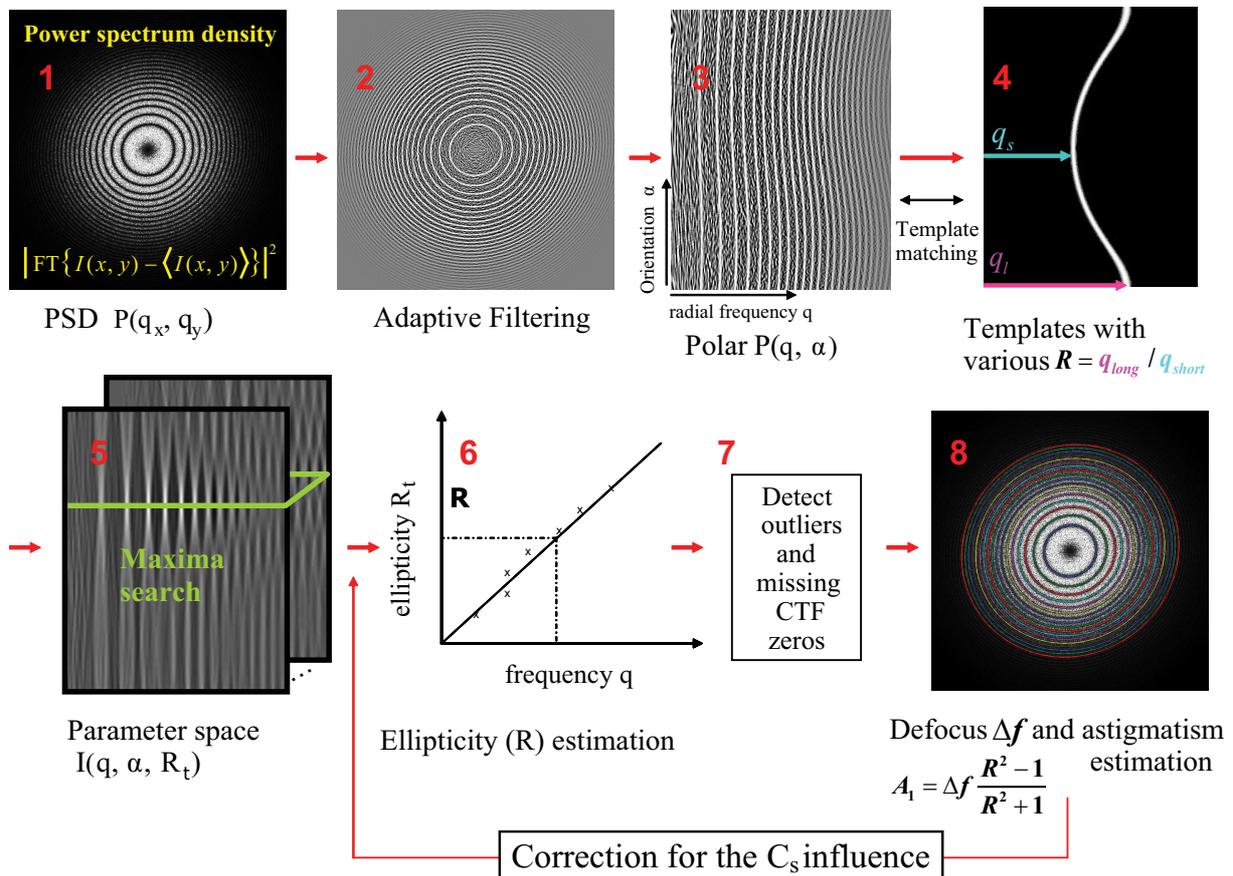


Fig. 3.3. Flow diagram of the algorithm. Note that we display the result after each step. Step 1: Compute the PSD from an image; Step 2: Suppress the background and invert the contrast of the rings by adaptive filtering; Step 3: Transform from Cartesian into polar coordinates; Step 4: Generate template and apply template matching; Step 5: Find local maxima in parameter space; Step 6: Find the ellipticity of the Thon rings; Step 7: Detect outliers, identify missing CTF zeros, assign ordinal number to each CTF zero; Step 8: Estimate defocus and astigmatism. Possible second pass for correction of the C_s influence.

3.3.1 Power spectrum density processing

The PSD in Eq. (3.15) is calculated using a fast Fourier transform (FFT). In order to avoid possible edge effects, a Hann window can be applied to the image prior to PSD calculation. Spatial or frequency rebinning could be used to speed up subsequent calculations.

Periodogram averaging

There are different ways to improve the signal-to-noise ratio (SNR) of the PSD. These include periodogram averaging [41, 96, 98], averaging the PSDs of images of individual particles [95, 97], additional angular averaging of the periodogram [41, 95–97], classification and averaging of the PSDs of different micrographs [33, 45], PSD enhancement [99, 114] and parametric PSD estimation technique using autoregressive modeling [96] or 2D-autoregressive moving average modeling [46]. For images that have such a low SNR that the rings are barely visible, we chose to perform periodogram averaging. Patches with a fraction of the size of the original image ($N_{\text{patch}} = N/j$) ($j \in \{2, 4, 8\}$) of an untilted sample are selected, and multiplied by a Hann (cosine) window in order to avoid edge effects, i.e.

$$I_i(x, y) = I(x + a_{x,i}, y + a_{y,i})w(x, y) \quad (3.24)$$

where $w(x, y)$ is the Hann window, $x, y \in [1, N_{\text{patch}}]$, and $a_{x,i}, a_{y,i} \in [0, N - N_{\text{patch}}]$. Note that $a_{x,i}, a_{y,i}$ are the offsets for the entire patch i . The periodogram averaged PSD is defined as:

$$P(q, \alpha) = \frac{1}{n} \sum_{i=1}^n P_i(q, \alpha) \quad (3.25)$$

where n is the number of patches and P_i is PSD of image I_i .

Background suppression

The background is suppressed and the contrast of the Thon rings is inverted using an adaptive filtering strategy. First, the logarithm of the PSD image is calculated which decreases the influence of the background slope. It also reduces the modulation depth variation of different rings. In this way, the widths of the Thon rings become more similar, and consequently, it is easier to detect them with a constant-width template.

An orientation-adaptive, second order Gaussian derivative filter [115] is applied to suppress the background and invert the contrast. Within the local footprint of the second order Gaussian derivative filter, the background is approximately linear and therefore suppressed. This adaptive filter assumes that the image is locally translation invariant along exactly one orientation (valid for line-like structures). As this is approximately true for all of the curved Thon rings which are straight within the filter's footprint, no disturbing artifacts are produced. As expected, we only perceive a slight compression of the contrast for the inner Thon rings. The method is in particular valuable for the dim outer Thon rings that obey the translation invariance to a very large extent. The filter kernel is anisotropic and smooths more along line-like structures such as the Thon rings than perpendicular to it. Furthermore, the spatial blurring of the adaptive filter

could be modified to make the the rings more prominent. The structure tensor [116, 117] is used to estimate the local orientation which steers the adaptive filter [118, 119]. The structure tensor was computed using a gradient scale of 1 and tensor scale of 20 pixels. These values proved to be robust against varying imaging conditions. Only in case of very small astigmatism, it is sensible, however, to avoid orientation estimation at all and assume a perfectly circular pattern. Any shifts between locations of the original Thon rings and the filter responses are corrected using the PLUS filter [80] as second derivative filter. Step 2 in Fig. 3.3 displays the PSD after applying this adaptive filtering.

3.3.2 Polar representation

The filtered PSD image is transformed into polar coordinates using cubic interpolation (step 3 in Fig. 3.3). This results in an image with one dimension (vertical in our display convention) representing angles (from 0 to π) and the other dimension representing frequency (horizontally from 0 to $N/2$, where N is the image size). Representing the angle α over an interval of π instead of 2π is possible since the PSD has Friedel's symmetry. The canonical implicit form of an ellipse whose long axis coincides with the q_x axis in Cartesian coordinates is given by

$$\frac{q_x^2}{q_l^2} + \frac{q_y^2}{q_s^2} = 1.$$

By substituting $q_x = q \cos \alpha$ and $q_y = q \sin \alpha$ and solving for q , an elliptical Thon ring in polar coordinates can be represented by

$$C(\alpha) = \frac{q_l q_s}{\sqrt{(q_s \cos(\alpha - \alpha_1))^2 + (q_l \sin(\alpha - \alpha_1))^2}}, \quad \alpha \in [0, \pi) \quad (3.26)$$

where α_1 is the angle between the long axis of the ellipse and the q_x axis. Step 3 in figure Fig. 3.3 suggests that the apparent curvature of the transformed rings (i.e. peak-to-peak amplitude) increases with frequency; however, all curves, when C_s is ignored, still have the same ellipticity q_l/q_s . It might be beneficial, although not necessary, to exclude the first few percent of the frequency range from the analysis where the original PSD was affected the most by the strong inelastic background.

3.3.3 Template generation and template matching

Template matching is performed by convolving templates of the shape of Eq. (3.26) with the polar image. The general approach would be to use the Radon transform. However, since in our case the shape of the template parameters are kept fixed, and only the position parameter is varied, the Radon transform can be implemented as a convolution [120, 121].

Template generation

Generated templates consist of ellipses in polar representation which all have a zero angle orientation of the long axis ($\alpha_1 = 0$) and a ‘‘central frequency’’ (q_c) in the middle of the frequency

range (at half Nyquist $N/4$, where N is the image size). We need to know this central frequency q_c of the Thon ring when aiming at estimating defocus. This is the frequency of the equivalent Thon ring without astigmatism, but with the same defocus. For the case that $C_s = 0$, we define, similarly to Eq. (3.19a) and Eq. (3.20a):

$$q_c^2 = \frac{k_{\text{eff}}}{\lambda \Delta f}. \quad (3.27)$$

Using Eq. (3.19b) and Eq. (3.20b) we observe the following relations for the short and long axis of a Thon ring:

$$\Delta f = \frac{1}{2}(\Delta f_l + \Delta f_s), \quad (3.28)$$

$$\frac{k_{\text{eff}}}{\lambda q_c^2} = \frac{1}{2} \left(\frac{k_{\text{eff}}}{\lambda q_l^2} + \frac{k_{\text{eff}}}{\lambda q_s^2} \right). \quad (3.29)$$

Solving the latter equations for q_c yields

$$q_c = \frac{\sqrt{2} q_l q_s}{\sqrt{q_l^2 + q_s^2}}. \quad (3.30)$$

The only parameter for the generated templates that is varied is the template ellipticity R_t , which ranges from 1 to R_{max} with increments of dR . There is a need for a good compromise between template matching computation speed and precision. However, it is not crucial to know the exact value of R_{max} for template generation. The user could specify either the value for R_{max} directly, or the uncertainty margins of the detected astigmatism. Given a specific uncertainty of the astigmatism estimation (e.g. 10 %), we can combine the expected maximal astigmatism and given defocus value from the microscope to derive a rough estimate for R_{max} . A realistic approach is to predict the maximal number of detected CTF zeros ($N_{0\text{max}}$) from the pixel size and requested defocus value. Then we have $dR = (R_{\text{max}} - 1)/(2N_{0\text{max}})$. It is always possible to perform an estimation of R_{max} with one additional iteration. Initially, templates are generated with a large R_{max} and coarse dR to get a rough estimate of the astigmatism, and then use R_{max} estimated by equation Eq. (B.3) in Appendix B.1 for the second iteration. We used a fixed number of 100 templates (as default) ranging from 1 to R_{max} . Making dR smaller did not further improve the accuracy.

Search for maxima in the parameter space

After convolution of the templates with the polar image, the resulting parameter space image has three dimensions (frequency q , azimuthal angle α , and template ellipticity R_t). Maxima in the parameter space are found by watershed-based segmentation on the inverted parameter space image. The lowest values in the watershed segmented regions are the local minima and the minimal height difference between peak and valley is 20 %. Sub-pixel localization is achieved by quadratic fitting through three points in each dimension at the same time. Each

maximum provides the orientation of the long axis α_1 , frequency q_i and apparent ellipticity $R_{t,i}$ for Thon ring i . We construct a histogram of the total weight of the found maxima with respect to azimuthal angle. The global mode in this histogram renders the angle of the long axis, since the angle of the long axis is common to all rings. Now the α coordinate is fixed, and a search for the maxima is performed again in the (q, R_t) -plane. In this way, the robustness of the algorithm is increased by imposing the constraint that all the rings must have an identical orientation of the long axis.

Zero astigmatism

If no astigmatism is present, the maxima in parameter space will be randomly placed along the long-axis orientation. Whatever value of the long-axis is selected has no influence on the estimated defocus value. Furthermore, the highest responses will be in the first plane ($R_{t,i} = 1$ for all rings i) of the three dimensional parameter space. In order to identify these responses as maxima, the watershed algorithm requires intensity comparison with neighboring pixels. For the responses that are at the edge of the parameter space we always expand the volume in the direction of $R < 1$ ellipticity. This is done by mirroring the first few slices in R direction at the plane $R = 1$, and then shifting them in α orientation direction by $\pi/2$ (now q_l becomes q_s and *vice versa*). Search for the maxima is performed only within $R \geq 1$. An additional control is performed by analyzing the slope of the responses in the (R, q) -plane. If the slope is smaller than 10^{-6} (which corresponds roughly to astigmatism less than 0.1 \AA per 1000 nm defocus), we assume that the responses are distributed at $R = 1$.

If no maxima are detected, the astigmatism will be ignored. All responses are projected in the direction of the angle and in the direction of the apparent ellipticity resulting in a reduced (one dimensional) parameter space where frequency q is the only remaining dimension. Maxima in this space represent frequency positions of the rings which are used to estimate only defocus, *via* the k -trajectory method (see 3.3.5). A similar approach (by reducing the parameter space from three to one dimensions) can be used for small astigmatism values to find defocus independently from the ellipticities.

If one is only interested in defocus estimation, the background-suppressed 2D PSD (3.3.1) is initially angularly averaged and the frequency positions of the rings are found by searching the maxima in the 1D spectrum in a similar manner as described in 3.3.3. The angular averaging could be performed either in a non-weighted or a weighted manner. Weighted angular averaging is performed by computing the weighted average inside rings with a Gaussian profile to avoid problems arising from averaging too few data points at low spatial frequencies (see [47] for details). Weighted averaging, however requires longer computational time. Note that by ignoring evident astigmatism, defocus estimation could be compromised as the SNR of the 1D angularly averaged spectrum decreases.

Correction for the difference between detected and template frequency positions

The radial frequency of a detected maximum does not reflect the true q_c of the Thon ring due to the difference between the mean values of the polar transformed PSD elliptical curve and

that of the template generated elliptical curve Eq. (3.26). The mean value is the solution of an incomplete elliptical integral of the first kind (see Appendix B.2 and Eq. (B.6)) which depends on R_t . Each detected q has its corresponding R_t which is used to solve Eq. (B.6) numerically. In B.2 we derive the relative error between the detected q values of the maxima and the expected central frequencies q_c cf. Eq. (3.30). This relative error depends only on the ellipticities R_t that are used to convert the detected q positions to the corresponding central frequencies q_c (Eq. (B.11)) which are further to be used for defocus and astigmatism estimation.

Derivation of Thon ring ellipticity from template ellipticity

Given a certain amount of astigmatism, templates with low ellipticities will match to the low frequency rings, and templates with a higher ellipticity to the higher frequency rings. We derived an analytical relation which predicts the behavior of the template matching ellipticities as a function of frequency (see Appendix B.1). This model is fitted through the detected maxima pairs $(q_i, R_{t,i})$. The ellipticity R_0 (common to all rings assuming $C_s \approx 0$) is the apparent ellipticity at the location of the generated templates (i.e. the middle of the frequency range, $N/4$). Additionally, if the number of detected maxima is larger than five (by default) we use robust fitting as implemented in the statistics toolbox of MATLAB. We define the uncertainty of the ellipticity value σ_{R_0} as a confidence interval of one standard deviation in the nonlinear regression.

3.3.4 Outlier rejection

If the number of detected maxima is larger than four (by default) we can perform outlier rejection and analyze the central frequencies in the squared frequency (q^2) domain. The minima of the CTF are equidistant in q^2 space (for $C_s = 0$). Using this knowledge we exclude the points that do not follow this pattern (i.e. outliers) and identify gaps in the sequence of detected rings. Next, an order is assigned to the CTF zeros which are the input for the k -trajectory method used for defocus estimation. We refer to Appendix C for detailed information about the outlier rejection.

3.3.5 Defocus and astigmatism estimation

After outlier rejection, identification of the missing or false CTF zeros, and assigning k -values to the detected Thon rings using k -trajectory method [48], the defocus is estimated. Fig. 3.4A shows the square of the frequency dependent sine term in Eq. (3.13) for various amounts of normalized defocus with the positions of the minima (red) and maxima (green) superimposed. The location of the CTF zeros from Eq. (3.16) can be used to solve for the defocus from each (ordered) individual zero i as:

$$\Delta f_i = \frac{C_s \lambda^3 q_{c,i}^4 - 2k_{\text{eff},i}}{2\lambda q_{c,i}^2}, \quad (3.31)$$

where $i \in \mathbb{N}$ is the assigned ordinal number of CTF zero and $q_{c,i}$ is the central frequency of ring i . For simplicity and without loss of generality let's assume a pure weak-phase object; i.e.

$k_{\text{eff}} = k$. Amplitude contrast is taken into account in the final implementation by keeping k_{eff} . The problem we now face is: which k_i corresponds to the frequency $q_{c,i}$? For convenience of the analysis we use normalized dimensionless frequency $q^* \equiv qC_s^{1/4}\lambda^{3/4}$ and defocus $\Delta f^* \equiv \Delta f(C_s\lambda)^{-1/2}$. In case of overfocus ($\Delta f^* < 0$) in Fig. 3.4A, the i -th zero-crossing corresponds to $k = i$. However, in case of underfocus ($\Delta f^* > 0$), in the first region $q_{c,i}^* = 1$ corresponds again to $k = 0$, but $q_{c,i}^*$ ($i > 1$) corresponds to $k = i - 1$. For a normalized underfocus larger than $2^{1/2}$, positive k values are encountered. We visually explain k -trajectories in Fig. 3.4B. For each k -sequence, the values of Δf_i can be calculated using Eq. (3.31). The k -sequence for which Δf_i has the smallest relative variance is assumed to be the correct one. The mean value of all Δf_i is the estimate of the actual defocus. $\Delta f_{\text{est}} = \Delta f \pm \sigma_{\Delta f}$ where $\sigma_{\Delta f}$ is the standard deviation of the best sequence. There exist situations, for a relatively small ratio between defocus and spherical aberration phase contribution, when minima in the squared CTF do not correspond to a zero crossing in the CTF. They might be falsely detected as zero crossings, and could hamper the k -trajectory method. Therefore, we allow one of the local minima not to be a CTF zero (see Fig. 3.5).

From defocus, ellipticity and their spreads we derive the astigmatism using Eq. (3.23). The standard deviation of the astigmatism is then

$$\sigma_{A_1} = \sqrt{\left(\frac{\partial A_1}{\partial \Delta f} \sigma_{\Delta f}\right)^2 + \left(\frac{\partial A_1}{\partial R_0} \sigma_{R_0}\right)^2} = \sqrt{\left(\frac{R_0^2 - 1}{R_0^2 + 1} \sigma_{\Delta f}\right)^2 + \left(\frac{4\Delta f R_0}{R_0^2 + 1} \sigma_{R_0}\right)^2}, \quad (3.32)$$

where σ_{R_0} is the standard deviation of the found ellipticity defined as one confidence interval of the fit (see 3.3.3).

3.3.6 Influence of spherical aberration C_s on the shape and frequency of Thon rings

The ratio between the spherical aberration and defocus terms in Eq. (3.1) is

$$\beta(q) = \frac{C_s \lambda^2 q^2}{2\Delta f}. \quad (3.33)$$

The presence of spherical aberration changes the positions of the high frequency Thon rings and in combination with astigmatism it might also change the ellipticity. This occurs for a relatively large value of $\beta(q)$ (e.g. > 0.2).

C_s influence on ellipticity

For non-zero C_s , the Thon rings do not have the same ellipticity. Therefore, we have to make a clear distinction in ellipticity of an individual Thon ring ellipse, which we will call Q_i for Thon ring i , given by

$$Q_i = \frac{q_{l,i}}{q_{s,i}} \quad (3.34)$$

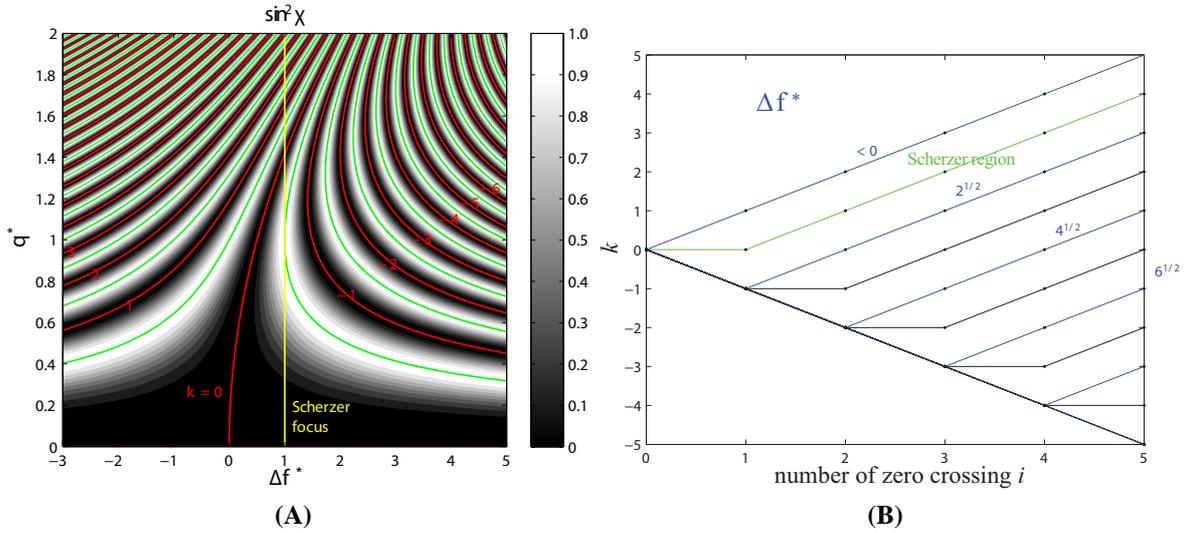


Fig. 3.4. (A) The square of the oscillating part of CTF in Eq. (3.13). The red and green lines indicate minima ($\sin^2(\chi(q)) = 0$) and maxima ($\sin^2(\chi(q)) = 1$), respectively. For simplicity and without loss of generality let's assume $k_{\text{eff}} = k$ (amplitude contrast is neglected). For convenience we use normalized dimensionless frequency $q^* \equiv qC_s^{1/4}\lambda^{3/4}$ and defocus $\Delta f^* \equiv \Delta f(C_s\lambda)^{-1/2}$. The Scherzer focus is represented by the yellow line. Following the q -axis direction, first a wide region of low contrast is encountered. In overfocus ($\Delta f^* < 0$) contrast improves, but the pass band is small and minima are quickly encountered. In underfocus ($\Delta f^* > 0$) there are regions where the maxima curves (green lines) are vertical. In those regions the contrast transfer is high for a wide frequency band. (B) The possible sequences of k -values for a certain zero crossing. In blue, the corresponding normalized defoci are indicated. In the vicinity of the Scherzer focus the k -sequence is equal to the green line.

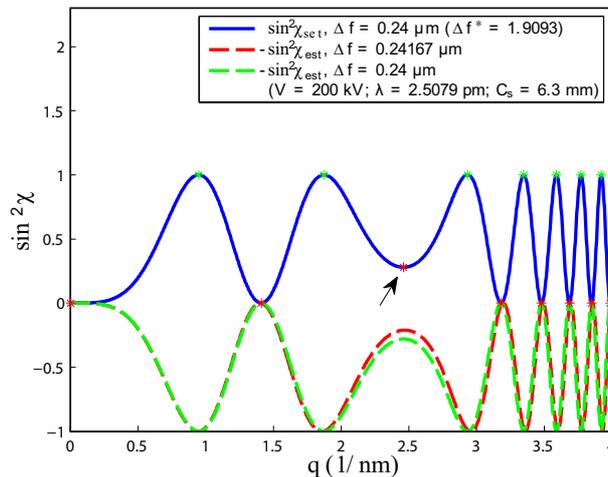


Fig. 3.5. An illustration of the flaw due to a local minimum which is not a zero crossing of the CTF (envelopes neglected). The blue line shows the simulated $\sin^2\chi$ at $\Delta f^* = 1.9$. The red and the green lines show $\sin^2\chi$ estimates with and without additional minimum, respectively. The additional minimum is indicated by the black arrow. The estimated $\sin^2\chi$ curves are flipped for the better visualization.

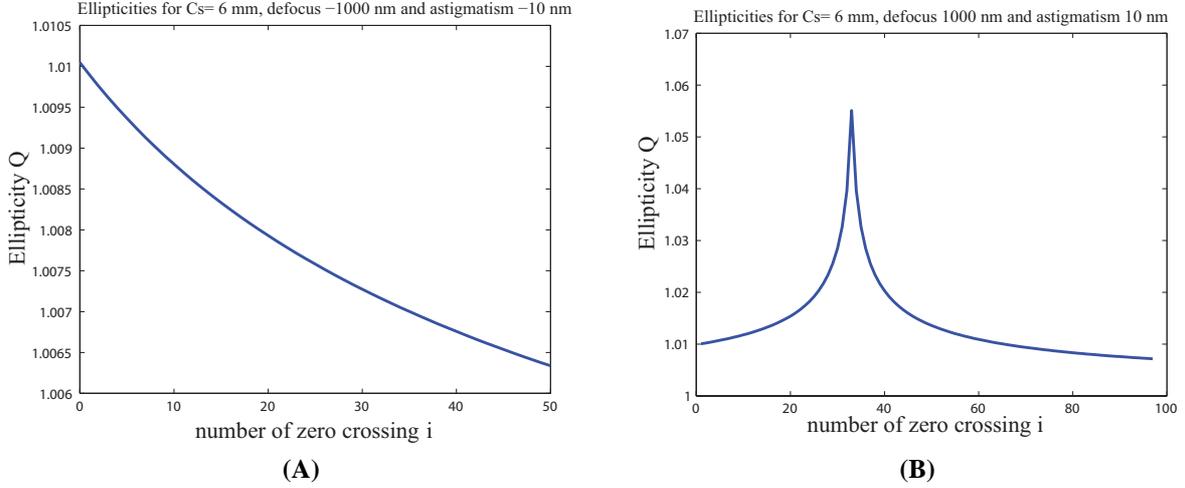


Fig. 3.6. The influence of the spherical aberration C_s on the Thon ring ellipticities. (A) In overfocus, ellipticity decreases monotonically with frequency. (B) In underfocus the ellipticity initially increases after which it decreases.

and the earlier introduced dimensionless measure R_0 given by Eq. (3.21). Note that $Q_i|_{C_s=0} = R_0$ for all Thon rings.

The ellipticity with C_s for different Thon rings (k_i values) is given by (see Appendix D.1 for derivation)

$$Q_i(k) = \sqrt{\frac{|\Delta f_s| + \sqrt{\Delta f_s^2 + 2C_s k_i}}{|\Delta f_l| + \sqrt{\Delta f_l^2 + 2C_s k_i}}}. \quad (3.35)$$

Note that for underfocus negative k_i -values exist cf. Fig. 3.4. As shown in Fig. 3.6, ellipticity monotonically decreases with frequency in overfocus, while in underfocus ellipticity initially increases after which it decreases.

C_s influence on the frequency of the rings in q^2 -space

For outlier rejection, we use the property that the minima are equidistant in q^2 -space. However, the presence of C_s alters the frequencies of the Thon rings (see Appendix D.2 for details). Similar to the ellipticities, in overfocus the distances between neighboring minima become smaller while in underfocus the distances first increase and then decrease. Therefore, we derive a criterion for applying an additional iteration resulting in a two-step approach. In case that the relative error in equidistance between neighboring minima in q^2 -space (Eq. (D.9)) is larger than 25 % (equally $\beta(q) > 10$ %), we decide to perform one additional iteration to correct for the C_s influence.

Correction for spherical aberration influence

From the parameter space of our template matching procedure as described in Appendix D.3, we can extract a value for Q_i for each Thon ring. However, for estimating the astigmatism, it is of interest to find the “equivalent ellipticity” $R_{\text{eq},i}$ when C_s would have been zero.

In D.3 we derive the “equivalent ellipticity” of a Thon ring as

$$R_{\text{eq}} = \sqrt{\frac{q_{l,i}^2 (2\Delta f_l - C_s \lambda^2 q_{l,i}^2)}{q_{s,i}^2 (2\Delta f_s - C_s \lambda^2 q_{s,i}^2)}}. \quad (3.36)$$

Note, that the expression contains values for $\Delta f_s = \Delta f + A_1$ and $\Delta f_l = \Delta f - A_1$. This means that in order to calculate the equivalent ellipticity, one first needs to have an initial estimate of defocus and astigmatism. Furthermore, in order to use outlier rejection it is desirable to know the C_s influence in Eq. (3.1) (i.e. β). Therefore, initially, we estimate the defocus from the first half of the PSD frequency range. The template matching function (Eq. (B.3)) is fitted to the frequencies for which $\beta < 0.1$. Now, using the estimated values, we estimate $R_{\text{eq},i}$ using Eq. (3.36) and from that, the defocus and astigmatism.

3.4 Results

3.4.1 Validation by simulations

PSD simulations of an amorphous sample

Simulated images are obtained by taking into account effects of the specimen scattering properties, microscope aberrations, and camera characteristics (cf. Eq. (3.14)). The Fourier transform of the projected potential of a weak-phase amorphous object is represented as:

$$\tilde{V}_z(q) = e^{i\varphi(q)} \quad (3.37)$$

where the amplitude of each frequency has the same constant value (equal to one) but the phase $\varphi(q)$ is random. Note that the phase distribution must be antisymmetric $\varphi(-q) = -\varphi(q)$ since the image is real. The Fourier transform of such a signal ($\tilde{V}_z(q)$) represents a white-noise object and its histogram is normally distributed with zero mean and standard deviation of one. The standard deviation of the generated $V_z(x, y)$ is normalized to 0.1 prior to applying the CTF and modulation transfer function of the camera (MTF). This normalization to 0.1 is necessary since Poisson noise can only be added to positive values; without the normalization, the inverse Fourier transform of the second term in Eq. (3.12) might become smaller than -1 , leading to negative intensity values. Furthermore, the normalization to 0.1 could be interpreted as phase-contrast initially set to 10 % of the image intensity but further modulated by CTF and MTF. The MTF *via* edge method, conversion factors, readout noise, dark current noise of the cameras used for simulations were determined experimentally for different types of TEM cameras [47], and can be measured, including detective quantum efficiency (DQE) for any camera using online toolbox [47]. Table 3.1 gives the values for aberration coefficients and

Table 3.1. Some parameters and aberration constants of evaluated TEM microscopes

source	LaB ₆	FEG	X-FEG
V [kV]	120	200	300
ΔE [eV]	1.0	0.7	0.7
λ [pm]	3.35	2.51	1.97
C_s [mm]	6.3	2.0	2.7
C_c [mm]	5.0	2.0	2.7
α_i [mrad]	0.3	0.1	0.03

electron source incoherency used to simulate images for different types of microscopes. The PSD background is considered to originate mainly from inelastically scattered electrons and has been modeled as a Lorentzian radial distribution [122]. Although amplitude contrast $W(q)$ is usually treated as a constant ($\sim 6 - 10$ %) [113], we allow a frequency dependency in the form of a Gaussian, as amplitude contrast is expected to give a larger contribution to the lower frequencies.

We simulated images with various values of defocus, various amounts and orientation of astigmatism, integrated electron flux, and magnification for three different types of electron guns (LaB₆, FEG, and X-FEG), energies and TEM cameras. In order to check the reproducibility of the estimation, for each parameter combination, we simulated 60 different noise realizations. Since the astigmatism is known in the simulations, the R_{\max} for template generation was predicted from Eq. (B.3) using the Nyquist frequency as q_c ; the number of generated templates was 100. Whenever necessary, in order to enhance SNR, rebinning in spatial or frequency domain is used.

Results from simulations

Precision and bias of defocus and astigmatism estimations are evaluated by simulations. Precision of the estimations as a function of astigmatism is shown in Fig. 3.7. Characterization of bias (absolute and/or relative error) of defocus and astigmatism estimations is presented in Table 3.2, Figs. 3.8-10. We observe a very good agreement between simulated and estimated defocus and astigmatism values. Given a particular magnification and camera size, defocus can be estimated with errors less than 4 % for LaB₆ and 1 % for X-FEG gun microscopes and with a small spread. Some examples from Table 3.2 include astigmatism values that range from 10 nm (LaB₆) down to 0.2 nm (X-FEG) with ~ 10 % spread (for defoci of 1 and 2 μm). An example of a correction for the C_s influence on the ellipticity of the rings (see 3.3.6) is presented in Fig. 3.12.

Fig. 3.7 shows the uncertainty of the astigmatism, and statistical uncertainty (precision) of defocus, ellipticity, and astigmatism angle estimation for the X-FEG gun type microscope at a magnification of 200 kx. The graphs show the precision represented by the standard deviation of the parameters estimation (+) as a function of astigmatism. For each defocus and astigmatism value, the estimation is characterized by its mean value and standard deviation. Each data

Table 3.2. Results from simulations for three different types of the electron guns (LaB₆, FEG, and X-FEG) and TEM cameras [47]. For each parameter combination, 60 noise realizations were processed and the number of outliers (failures) is provided. An estimation of defocus and ellipticity was considered to be an outlier (failure) if it differed more than 3 standard deviations from the median value of the set. Mean absolute and relative errors of defocus and astigmatism are presented for two different integrated electron fluxes: 25 and 1000 $e^-/\text{\AA}^2$.

Electron source	LaB ₆				FEG				X-FEG				X-FEG		
Camera size N	2k x 2k				2k x 2k				2k x 2k				4k x 4k		
Magnification	50 kx				50 kx				200 kx				200 kx		
# of CTF zeros	2-3				4-6				35-50				72-80		
Defocus [nm]	1000	2000			1000	2000			1000	2000			2000		
Astigmatism [nm]	10	10			5	5			0.5	0.5			0.2	0.2	
Int. flux [$e^-/\text{\AA}^2$]	25	1000	25	1000	25	1000	25	1000	25	1000	25	1000	25	1000	
Defocus error	%	1.7	1.5	3.5	2.6	1.2	0.8	0.4	0.3	0.06	0.005	1.0	0.01	0.6	0.002
	nm	17	15	70	52	12	7.6	8.2	5.0	0.60	0.05	20	0.20	12	0.04
Astigmatism error	%	13	8.6	67	39	14	9.4	10	7.8	4.6	3.1	14	14	37	9.6
	nm	1.3	0.9	6.7	3.9	0.7	0.5	0.4	0.02	0.02	0.02	0.07	0.1	0.07	0.02
Relative error of ellipticity [%]	0.14	0.08	0.31	0.20	0.07	0.05	0.03	0.02	0.002	0.002	0.01	0.004	0.004	0.001	
# of outliers (out of 60 repeats)	0	0	2	2	1	0	1	4	1	0	4	2	2	1	

point represents a series of 60 repeated measurements from which outliers were rejected. An estimation of defocus and ellipticity was considered to be an outlier (failure) if it differed more than 3 standard deviations from the median value of the set. The mean and standard deviation were re-calculated without the outliers and concurrently the number of outliers is provided. The mean of the predicted astigmatism uncertainty values (\circ) in Fig. 3.7A were derived from the measured defocus and ellipticity uncertainties but also from their estimated values (Eq. (3.32)). The number of outliers is only 1 – 2 out of 60 for a high SNR. Figs. 3.7A,B,C show astigmatism, defocus, and ellipticity uncertainties that are small compared to the absolute value. Furthermore, the spread (precision) of defocus and astigmatism estimations from repeated acquisitions (+) is often similar to the predicted uncertainty from one individual image (\circ). For astigmatism larger than 1 nm, Figs. 3.7A,B,C suggest that the estimated errors are smaller than the predicted errors. Estimations for higher integrated fluxes (better SNR) generally perform better. Although the ellipticity for a fixed astigmatism is smaller for 2000 nm defocus than for 1000 nm, the results indicate that data for larger defocus give slightly better results than for lower defocus. This probably relates to the larger number of rings for higher defocus. Determination of the astigmatism angle is shown in Fig. 3.7D and indicates that the uncertainty rises with smaller astigmatism strength. This is expected as the peak detection in parameter space is compromised for very small ellipticity values.

Fig. 3.8 shows the mean of the absolute and relative errors of astigmatism estimation within a series of repeats. Depending on the values of defocus, astigmatism, and integrated flux, the relative error varies from a few percent to a few tens of a percent. In general, the absolute value

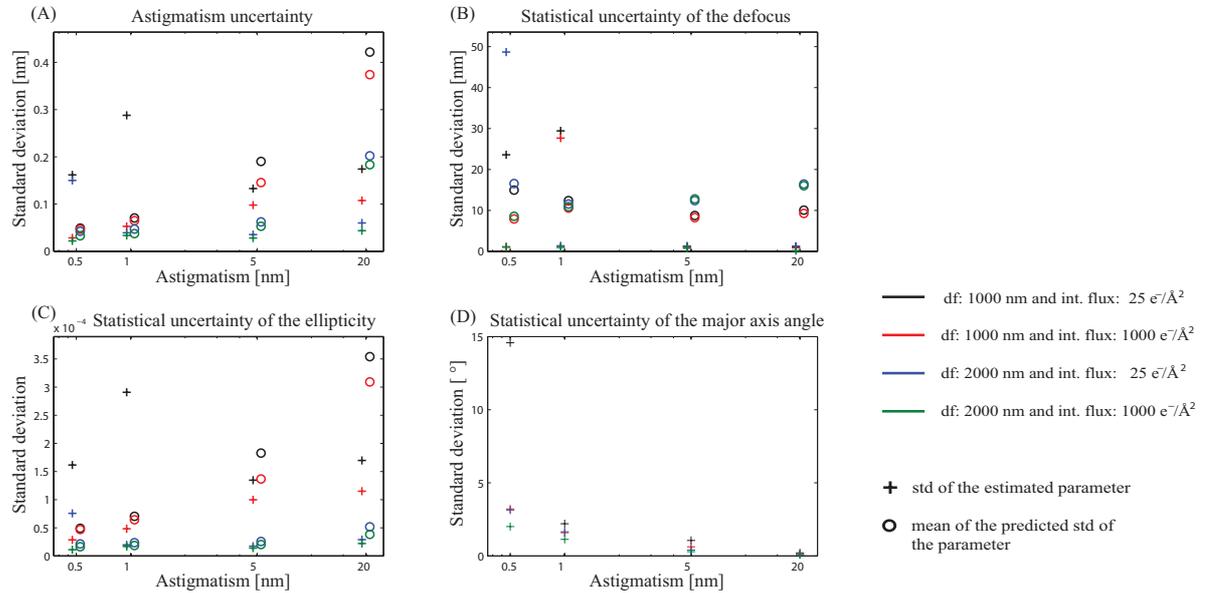


Fig. 3.7. Uncertainties of the estimated parameters for X-FEG gun type microscope at a magnification of 200 kx. Each data point represents a series of 60 repeated simulations from which outliers were rejected. The pluses (+) characterize the standard deviation (std) within the series of mean estimated values. The circles (o) characterize the mean of the predicted standard deviation of the estimation within the series. For better visibility pluses and circles are separated and shifted slightly to the left and to the right respectively from their real astigmatism values presented on the horizontal axis.

increases with astigmatism strength while the relative error decreases.

The mean absolute and relative errors of defocus estimation are shown in Fig. 3.9. The horizontal axis now represents three different defoci, the different colors denote different integrated fluxes and magnifications, while the mean errors of defocus are additionally averaged over four different values of astigmatism (the values on the horizontal axis in Fig. 3.8) since it is expected that defocus is independent of astigmatism. The estimation error is better than 1 %. In a similar manner we characterized the errors of the ellipticity estimates (Fig. 3.10), that were used for the calculation of astigmatism *via* Eq. (3.23). The sensitivity of the estimator is high, being able to detect ellipticity down to 1.0004 with a relative error of only 10^{-3} % (see Table 3.2). Fig. 3.11 demonstrates that errors in the estimated long axis orientation angle increase with smaller astigmatism which is in agreement with Fig. 3.7D. Along with the uncertainties of defocus and astigmatism estimation, Table 3.2 also indicates the mean number of outliers and the number of detected zeros (rings) for different integrated fluxes, defoci, and astigmatism values.

The images with isotropic CTF (no astigmatism) were furthermore simulated for a X-FEG type microscope and 2k x 2k camera size. The mean absolute errors of astigmatism were 0.04 nm and 0.08 nm for defoci of 1000 nm and 2000 nm respectively and for an integrated electron flux of $25 e^{-}/\text{\AA}^2$.

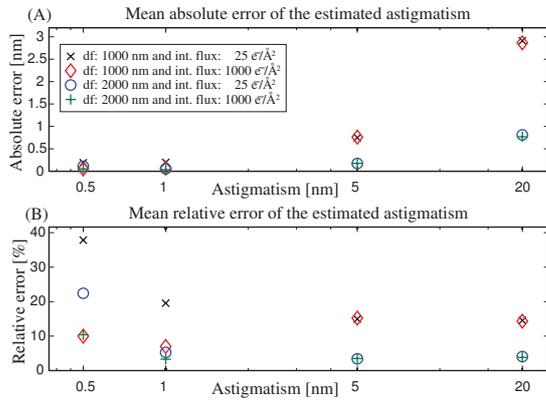


Fig. 3.8. Mean absolute (A) and mean relative (B) errors of estimated astigmatism as a function of the simulated astigmatism for X-FEG gun type microscope at a magnification of 200 kx (for two different defoci and integrated fluxes). Each data point represents a series of 60 repeated simulations from which outliers were rejected.

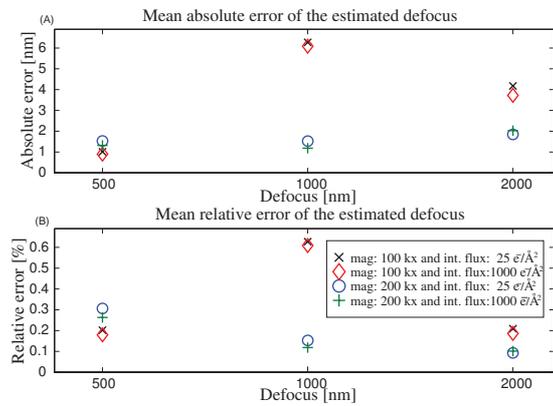


Fig. 3.9. Mean absolute (A) and mean relative (B) errors of estimated defocus as a function of simulated defocus for X-FEG gun type microscope. The error values were averaged over all astigmatism values presented in Fig. 3.8. For the comparison two different integrated fluxes and magnifications were presented. Each data point represents a series of 60 repeated simulations from which outliers were rejected.

3.4.2 Results from measurements

The reproducibility of the algorithm was evaluated using ten sequentially repeated measurements of a platinum-iridium (PtIr) sample under identical conditions for different combinations of magnification, defocus and astigmatism. Unbinned images (4k x 4k) were collected on a Tecnai F20 (FEI Company, The Netherlands), using MATLAB scripts inspired by the TOM toolbox [38] and employing the TEMScripting ActiveX server. Series of images with four different stigmator settings were collected for three defocus values (500 nm, 1000 nm and 2000 nm). Three different magnifications (62 kx, 100 kx, 150 kx) were used. The incident beam was parallel and the incident integrated electron flux was constant ($\approx 167 e^-/\text{\AA}^2$). Each series consists of ten repeated measurements under identical conditions. Whenever necessary, in order to enhance SNR, the rebinning or periodogram averaging was applied by using 20 patches of relatively large size $N_{\text{patch}} = N/2$ in order to maintain good sampling of high frequencies in the Fourier domain. Table 3.3 summarizes the results. The standard deviation of measured defocus and astigmatism within a series (+) is small and comparable to the mean value of the predicted standard deviations calculated from individual estimations (\circ).

The linearity of the stigmator response was evaluated on data acquired using the same sample on a Titan microscope. The microscope was equipped with a Falcon CMOS direct electron detector and operated at 300 kV voltage. A series of images with increasing strength of the stigmators (x and y) in both directions (positive and negative) were collected. The results of

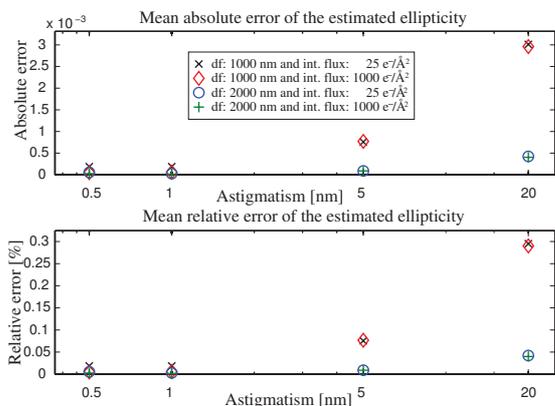


Fig. 3.10. Mean absolute (A) and mean relative (B) errors of estimated ellipticity as a function of simulated astigmatism for X-FEG gun type microscope at a magnification of 200 kx (for two different defoci and integrated fluxes). Each data point represents a series of 60 repeated simulations from which outliers were rejected.

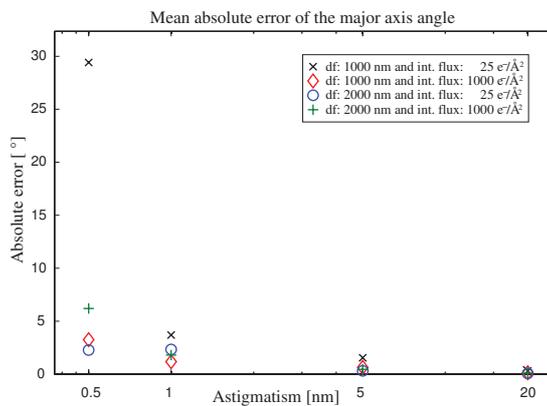


Fig. 3.11. Mean absolute errors of the long axis orientation as a function of simulated astigmatism for X-FEG gun type microscope at a magnification of 200 kx (for two different defoci and integrated fluxes). Each data point represents a series of 60 repeated simulations from which outliers were rejected.

the astigmatism estimation for 450 nm overfocus are shown in Fig. 3.13. The projections of astigmatism on the x - ($A_{1,x} = A_1 \cos \alpha_1$) and y -axes ($A_{1,y} = A_1 \sin \alpha_1$) were calculated. The linearity was assessed by making a linear least-squares fit to the estimated projected astigmatism *versus* stigmator strength (see Fig. 3.13A). The square of the sample correlation coefficient between the measured and predicted values, within the range of measured astigmatism values, was nearly one: 0.9998 and 0.9997 for negative and positive y stigmator strengths, respectively. Fig. 3.13B shows the relation between x and y projected astigmatism. Linear least-squares fits for all four data sets (increase and decrease of x and y stigmator strengths) were calculated. The angles between the introduced astigmatism were nearly 90° . This corresponds well to the expected orthogonality while altering between the positive and negative values of a stigmator. The introduced astigmatism changes with twice this angle (Eq. (3.2)). For the same reason, the angles between lines of the x and y stigmator close to 45° correspond well to the orthogonality between x and y stigmators. Equidistant data points within a series indicate linearity, already presented in Fig. 3.13A.

3.4.3 Thon ring assessment

In this section we will evaluate our CTF estimation algorithm as a tool for assessing Thon rings. In particular the modulation depth of the rings as a measure for useful contrast transfer as a function of spatial frequency. For this purpose, we first analyze the performance of our Thon ring averaging method, as this is an important prerequisite to objectively assess the Thon rings

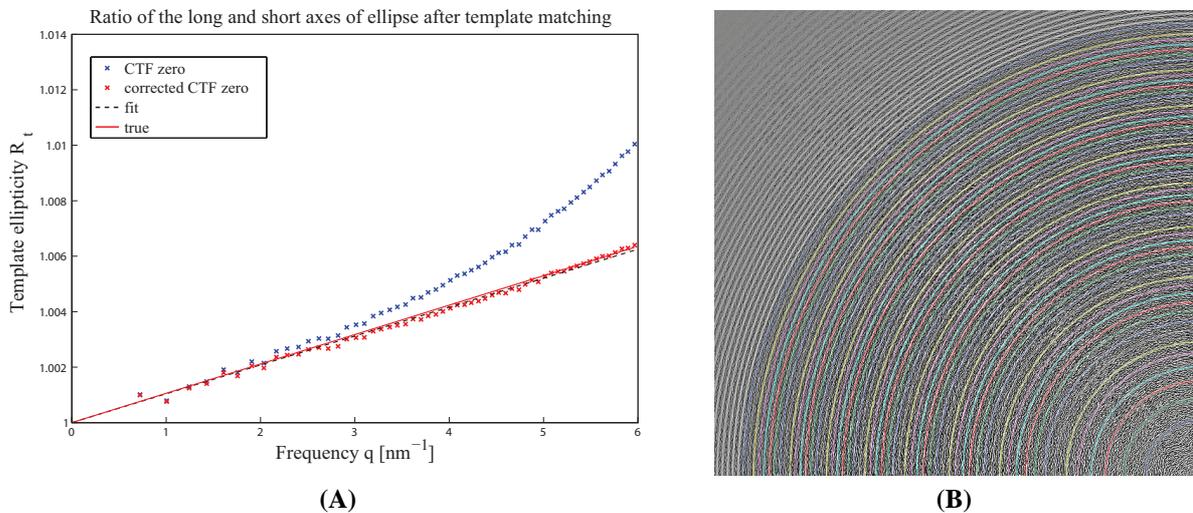


Fig. 3.12. (A) The apparent ellipticities of the rings after the template matching, with and without subsequent correction for the C_s influence (defocus 1000 nm, astigmatism 5 nm, $C_s = 2.7$ mm, magnification 200 kx, X-FEG source). (B) Overlay of positions and shapes of the found Thon rings with background suppressed PSD.

from 1D CTF profiles. Subsequently, we will introduce a quantitative measure for Thon ring visibility and show some results on real images.

Thon ring averaging

The algorithm for Thon ring averaging (TRA) is described in Appendix E. Our new TRA method extends the elliptical averaging method by taking into account C_s influence on the ellipticity of the rings. Fig. 3.14 illustrates the difference between circular, elliptical, and Thon ring averaging. For a certain combination of imaging parameters such as a large ratio β between the spherical aberration and defocus terms in Eq. (3.1), Thon ring averaging is advantageous to get a higher SNR of 1D PSD profiles.

A Thon ring visibility criterion

Defocus and astigmatism estimation is useful for assessing Thon rings and information transfer. That is, we want to quantify the contrast transfer of a TEM by Thon rings with regard to some criterion. For this purpose, we first accurately estimate the defocus and astigmatism, including the correction for the C_s effect (see 3.3.6). Subsequently, we calculate the Thon ring average as described in Appendix E and the theoretical positions of the maxima m_i and minima t_i (i.e. the Thon ring frequencies) in the angular average. The modulation of the amplitude of the Thon ring i is then given by

$$M_i = \frac{\text{PSD}_{1D}(m_{i-1}) + \text{PSD}_{1D}(m_i)}{2} - \text{PSD}_{1D}(t_i) \quad (3.38)$$

Table 3.3. Robustness of the estimation evaluated on images of a PtIr sample acquired on a microscope with a FEG electron gun and 4k x 4k camera. Series of images with four different stigmator settings were collected for three defocus values (500 nm, 1000 nm and 2000 nm). Three different magnifications, of 62 kx, 100 kx, and 150 kx were used. The incident beam was parallel and the incident integrated electron flux was constant ($\approx 167 e^-/\text{\AA}^2$). Each series consists of ten subsequently repeated measurements under identical conditions. The standard deviation of measured defocus and astigmatism within a series (+) is small and comparable to the mean value of predicted standard deviations calculated from individual estimations (\circ).

Requested defocus [nm]	Error [nm]	Δf_{est}	ast ₁ (62 kx)	ast ₂ (62 kx)	ast ₃ (62 kx)	ast ₄ (62 kx)
500	measured + predicted \circ	561.8 ± 5.5 3.0	16.2 ± 6.6 1.9	12.9 ± 1.8 2.0	14.5 ± 2.7 2.4	11.5 ± 4.6 1.6
1000	measured + predicted \circ	1051 ± 5.7 5.6	22.5 ± 1.0 0.8	18.1 ± 1.3 0.8	15.4 ± 1.3 1.0	7.9 ± 1.5 1.2
2000	measured + predicted \circ	2050 ± 6.6 44	32.7 ± 1.3 1.0	28.1 ± 0.7 0.9	25.4 ± 1.1 0.8	6.3 ± 1.0 1.0
Requested defocus [nm]	Error [nm]	Δf_{est}	ast ₁ (100 kx)	ast ₂ (100 kx)	ast ₃ (100 kx)	ast ₄ (100 kx)
500	measured + predicted \circ	300.9 ± 6.6 4.1	19.0 ± 1.9 1.5	14.8 ± 4.3 1.6	12.9 ± 1.9 4.2	22.4 ± 5.8 2.5
1000	measured + predicted \circ	732.6 ± 5.0 4.7	18.8 ± 1.8 2.6	14.9 ± 3.9 2.5	13.7 ± 1.1 2.2	11.6 ± 5.0 2.2
2000	measured + predicted \circ	1724 ± 8.0 24	25.6 ± 1.0 1.0	20.2 ± 1.2 1.0	18.4 ± 1.5 0.9	12.8 ± 2.6 4.3
Requested defocus [nm]	Error [nm]	Δf_{est}	ast ₁ (150 kx)	ast ₂ (150 kx)	ast ₃ (150 kx)	ast ₄ (150 kx)
500	measured + predicted \circ	551.6 ± 5.9 3.3	18.8 ± 1.8 1.5	14.7 ± 1.8 1.4	12.0 ± 1.2 1.7	10.6 ± 3.6 1.6
1000	measured + predicted \circ	1030 ± 4.1 2.6	21.2 ± 1.3 0.7	16.5 ± 0.7 0.7	14.5 ± 0.5 0.7	6.5 ± 1.1 1.1
2000	measured + predicted \circ	1982 ± 5.6 35	30.6 ± 0.7 1.1	25.6 ± 0.8 1.0	24.0 ± 1.0 1.0	4.7 ± 1.3 0.7

where m_{i-1} and m_i are the two closest maxima with $m_{i-1} < t_i < m_i$. The modulation depth of a Thon ring is defined as M_i/nf , where nf is the noise floor, found by calculating the average of the power spectrum that is outside of the Nyquist bound

$$nf = \frac{\sum_{|q| > \frac{N}{2}} \text{PSD}(q)}{\sum_{|q| > \frac{N}{2}} 1}. \quad (3.39)$$

A Thon ring is considered to be detected if its modulation depth is larger than two. Fig. 3.14 shows an example of the Thon ring assessment procedure.

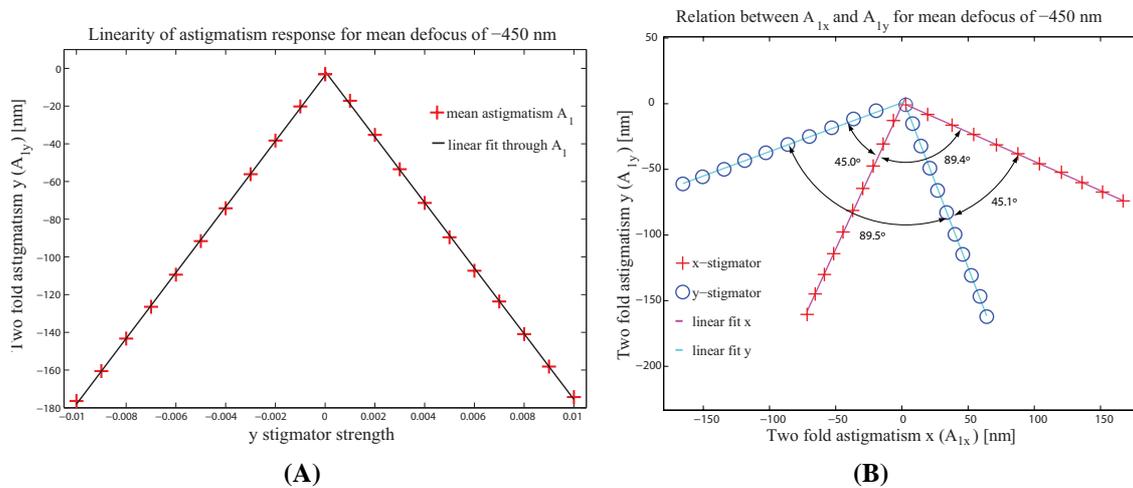


Fig. 3.13. The response of microscope’s stigmators evaluated using a PtIr sample on a Titan microscope (at 300 kV and 250 kx magnification) equipped with a Falcon CMOS direct electron detector. Series of images with increasing strength of the stigmators (x and y) in both directions (positive and negative) were collected. The projections of astigmatism on the x ($A_{1x} = A_1 \cos \alpha_1$) and y axes ($A_{1y} = A_1 \sin \alpha_1$) were calculated. (A) Linearity of estimated y -projected astigmatism A_{1y} versus y stigmator strength for 450 nm overfocus. The linearity was validated by high coefficient of determination: 0.9998 and 0.9997 for negative and positive y stigmator strengths, respectively. Additionally, the slopes of the lines show good agreement (-17.44 and 17.39). (B) Relation between x - and y -projected astigmatism values. The angles between linear least-squares fits cyan-cyan (magenta-magenta) lines were nearly 90° . The angles between cyan-magenta lines were close to 45° and correspond well to the final orthogonality between x and y stigmators. Equidistant data points within a series indicate linearity, already presented in Fig. 3.13A.

3.5 Discussion and conclusions

Unbiased and precise defocus and astigmatism determination is necessary for CTF estimation and correction, assessment of microscope contrast, image modeling, optimal adjustment of aberration correctors, and exit wave reconstruction. It is also beneficial for the calculation of resolution metrics such as Fourier ring correlation [123]. We have presented an algorithm for the unbiased and precise estimation of defocus and astigmatism from the PSD of TEM images of amorphous specimens. The algorithm provides an error estimate and automatically rejects outliers. Tests show very good agreement between simulated and estimated values of defocus and astigmatism (Table 3.2). Given a particular magnification and camera size, defocus can be estimated with a small spread and errors less than 4 % for LaB₆ and 1 % for X-FEG gun microscopes. Some examples include astigmatism values that range from 10 nm (LaB₆) down to 0.2 nm (X-FEG) with a ~ 10 % spread (for defoci of 1 and 2 μm). We chose relatively large defocus values, typical for life sciences, to demonstrate the ability to detect small astigmatism (very small ellipticity). We evaluated the reproducibility of the algorithm on experimental data by repeating measurements under identical TEM imaging conditions for a few

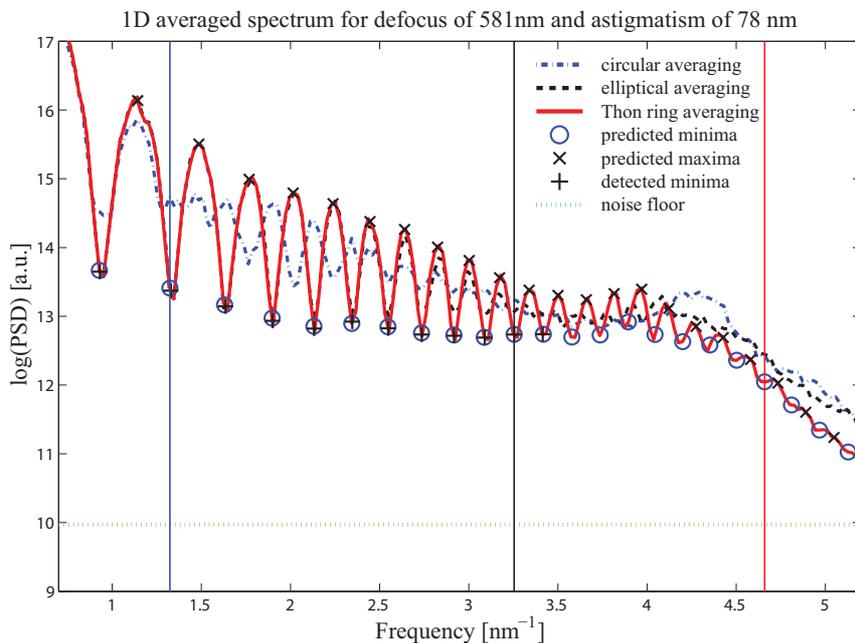


Fig. 3.14. Thon ring averaging and Thon ring assessment. Thon ring averaging (TRA), elliptical and circular averaging methods are compared. The horizontal axis represents the central frequency q_c given by Eq. (E.4). TRA is advantageous when C_s influence on the ellipticity of the rings is not negligible. The image of PtIr sample was acquired with a Titan microscope (at 300 kV and 380 kx magnification) equipped with a Falcon CMOS direct electron detector and FEG electron gun. Estimated defocus 581.4 ± 0.5 nm; estimated astigmatism 78.2 ± 0.4 nm; spherical aberration 2.7 mm. Note that up to ~ 3 nm^{-1} elliptical averaging and TRA are perfectly in phase, but they appear uncorrelated. Thon ring assessment: the green dotted horizontal line shows the estimated noise floor and the vertical lines show the result of the Thon ring assessment, i.e. modulation amplitude of the Thon ring is twice higher than the noise floor for all frequencies left of the vertical line.

defocus and astigmatism values (see Table 3.3). The autofocus routine (which works by measuring the beam-tilt induced image displacement) of the microscope was executed before each magnification series and then moved to the requested defocus. The reason for the mismatch between requested and estimated defocus at the magnification of 100 kx might be an inaccurate defocus calibration (i.e., the calibration that relates beam-tilt induced image shift to defocus values) for this particular magnification. Our approach requires that the sample is amorphous or near-amorphous. Both amorphous carbon and PtIr satisfy this requirement. Actually, for the PtIr sample, the grains of PtIr are evaporated on carbon film. The advantages of PtIr is that this specimen may be used to test the resolution of the electron microscope by the point separation test, gives an intrinsic magnification calibration by the PtIr reflexion at ~ 2.35 Å and might scatter to higher frequencies than carbon. However, we do not use calibration properties in our evaluations (only amorphousness). The algorithm was used to analyze the response of the stigmators which was validated to be linear (Fig. 3.13). The uncertainty of the defocus estimation from one image depends on the number of detected zeros. As shown in Fig. 3.7 and Table 3.3,

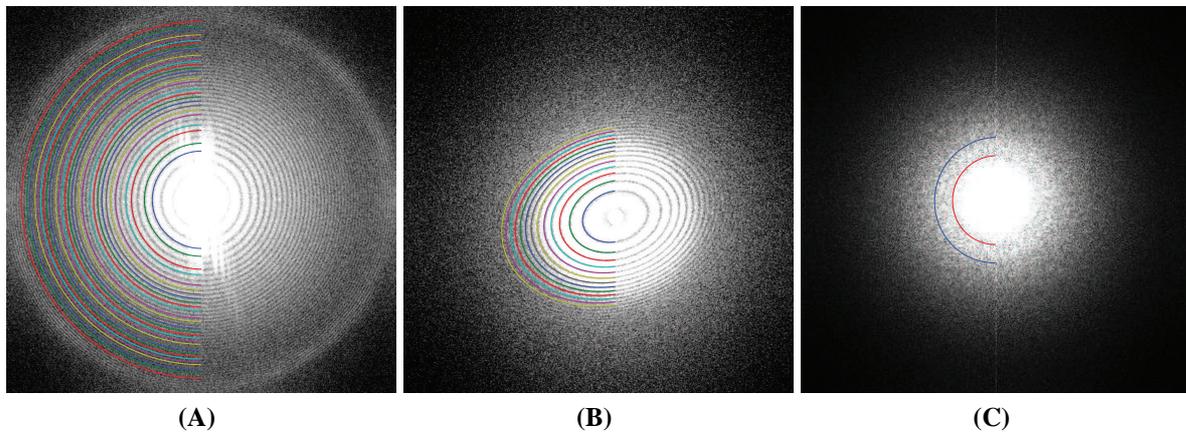


Fig. 3.15. (A) PSD of an image (size $4k \times 4k$ pixels) acquired on a Titan equipped with an X-FEG electron gun. The disturbance probably comes from specimen drift. Estimated defocus 1120 ± 2 nm; estimated astigmatism 1.6 ± 0.1 nm; Magnification 155 kx; (B) PSD of an image (size $4k \times 4k$ pixels) acquired on a Tecnai F20 (with a FEG electron gun). Estimated defocus 969 ± 13 nm; estimated astigmatism 269 ± 3 nm; magnification 62 kx. (C) PSD of an image (size 1536×1536 pixels) of hemoglobin embedded in vitreous ice acquired on a Tecnai T12 equipped with a LaB_6 electron gun. The image was taken with an integrated electron flux of $\sim 5e^- \text{\AA}^{-2}$ at the edge of a hole of a Cflat support film and includes $\sim 30\%$ of the support. Estimated defocus 4521 ± 444 nm; estimated astigmatism 166 ± 50 nm; Magnification 50 kx. The calculated Thon rings are mapped only over the angular range of 180° for better comparison. For the display a percentile stretch was used (the lower and upper 1 % of the gray values were clipped before stretching).

the spread of defocus and astigmatism estimations from repeated acquisitions is often similar to the predicted uncertainty from an individual image, although they inherently represent different statistical measures. Additionally, we show that accounting for the influence of astigmatism and C_s enhances the modulation depth of the 1D averaged PSD and helps assessing the quality of the contrast transfer.

The algorithm suppresses the background in the PSD using an adaptive filtering strategy that avoids the need for conventional estimation of the frequency range of the 1D background and fitting of a model through the PSD minima. Furthermore, an anisotropic background as mentioned in [99] can be addressed in this way. The method itself relies on template matching using kernels of various ellipticities. Maxima in the 3D parameter template space provide the long axis orientation, frequencies and apparent ellipticities of the rings. From these parameters we derive an equivalent ellipticity (R_0), common to all rings, which corresponds to the apparent ellipticity at the position of the generated template.

The frequencies of detected Thon rings are used to estimate the amount of defocus *via* the k -trajectory method as described in [48]. This method assigns an integer number k to each detected Thon ring (CTF zero). Several defoci can be computed from the CTF zeros, but the value with minimal normalized standard deviation is taken as the final defocus estimate. Accuracy is hard to assess in the actual experiments since the true values are unknown. However, theory governs that the estimated defocus values for the different Thon rings should be consistent. Each

defocus estimation based on more than one Thon ring is accompanied by the uncertainty $\sigma_{\Delta f}$ (see 3.3.5). If the provided C_s value, electron energy, measured magnification and the amount of amplitude contrast are correct, it is very unlikely that there exists a systematic disturbance which shifts the CTF zeros in such a way that $\sigma_{\Delta f}$ stays the same or decreases. This would be only possible if we falsely detect spurious CTF zeros at regular positions between every true CTF zero (including one before the first zero). Under all these assumptions, $\sigma_{\Delta f}$ can be used as a measure of accuracy which incorporates both bias and precision [124]. Additionally, it can be used as a sorting criterion, without having to evaluate repeated measurements.

Spurious or missed rings in the PSD are automatically identified and accounted for. This means that estimations can be done from any subset of rings, not relying exclusively on the first few minima in the PSD as is usually done. The outlier rejection and CTF zeros ordering use the fact that zeros of the CTF are equidistant in squared frequency space (for $C_s = 0$). An additional control is performed in the k -trajectory method where one possible false CTF zero that occurs for a small ratio between defocus and C_s phase terms is discarded. Furthermore, the k -trajectory method is capable of distinguishing between underfocus and overfocus (for $C_s \neq 0$ or amplitude contrast $W(q) \neq 0$).

Ignoring the influence of spherical aberration on the CTF results in a deviation of the apparent ellipticities from anticipated ones (blue crosses in Fig. 3.12) at high spatial frequencies and/or relatively low defocus values. We predict and correct for this C_s influence in a two-pass refining process (red crosses Fig. 3.12A) and accurately map the Thon rings (Fig. 3.12B). In addition, we introduce a new angular averaging method, Thon ring averaging (TRA), which takes into account the influence of C_s on the ellipticity of the rings; TRA averages over true Thon rings to get the 1D PSD, rather than averaging over circles or ellipses. TRA proves to be superior (Fig. 3.14) especially in cases when the ratio between the spherical aberration and defocus terms in Eq. (3.1) is relatively large (e.g. $\beta(q) > 0.5$ where $\beta(q)$ is defined in Eq. (3.33)). The Thon ring assessment as described in section 3.4.3 uses TRA and is a useful tool for microscope contrast transfer assessment.

The typical processing time depends on the input image size and the accuracy required. Spatial or frequency rebinning could be used to speed up subsequent calculations. The default settings in the software are currently such that images are binned to 512×512 pixels after which the estimation takes a few seconds if the templates were pre-computed and stored on disk or half a minute if 50 templates have to be generated (on a computer running at 2.7 GHz with 4GB RAM). However, a high accuracy and detection of very small astigmatism requires computation time. Another advantage of rebinning is that it can enhance the SNR. Nevertheless, one should use rebinning with caution. For the PSD that has wider rings which also extend to high fraction of Nyquist frequency (e.g. relatively lower magnification and lower defocus), binning in the Fourier domain might be beneficial. If the PSD has rings that are narrow and close to each other, but they do not extend to a high fraction of Nyquist frequency (e.g. relatively high magnification and high defocus), spatial binning is beneficial.

In order to avoid possible edge effects, a Hann window can be applied to the image prior to PSD calculation. Here, the Hann window is only used for periodogram averaging. It is very wide (one period over the whole image) and is therefore very narrow in Fourier domain (effectively a kernel of only $\sim 3 \times 3$ pixels in the Fourier spectrum). The convolution/blurring

of the logarithm of the PSD is therefore small and a shift of the CTF zeros is expected to be sub-pixel and only measurable if there is a steep slope in the PSD to begin with.

Although the accuracy of the defocus value provided by the microscope software is seldom sufficient, this defocus value can still be used to assist in a rough initial estimation, in a prediction of the C_s influence, and in a prediction of the (equi)distance of minima in the squared frequency (q^2) space. Astigmatism distorts the circular shape of the PSD rings and decreases the SNR of the 1D PSD angular averages. We have assumed that astigmatism is not excessive (astigmatism is not larger than defocus) and the Thon rings are still approximately elliptical. These requirements are typical for life sciences applications where phase contrast imaging is used mostly at relatively high defocus. The algorithm, however, can also be applied to a range of parameter settings typical for materials science as long as the defocus is larger than astigmatism. Provided that the astigmatism is relatively large but not excessive it is possible to extract the astigmatism even from the circularly averaged PSD [102]. Our algorithm, however, is able to detect very small astigmatism as well.

Although there are numerous aberrations in TEM, we focus here on robust and unbiased determination of defocus and astigmatism as they are crucial for the measurements based on diffractogram tableaux of all higher-order aberrations such as coma and threefold astigmatism. Ideally, the illumination of the sample should be parallel. Tilted illumination introduces higher ellipticity of Thon rings due to the higher-order aberrations [105]. In this work, we assume that CTF modulation is space-invariant over the entire micrograph. This is valid for most HREM and single particle EM studies in which the grid plane is perpendicular to the parallel incident electron beam. The astigmatism is usually constant for a sequential data collection, whereas the defocus is likely to show larger variation, in particular for tomography. Therefore, it is advisable to accurately measure astigmatism on a zero-tilt diffractogram, correct for the astigmatism if required, and then continue with image acquisition. Defoci in tomograms can be measured using procedures described in [125, 126].

Whereas algorithms that base their defocus estimation on 1D averaged PSD are sensitive to sample drift and missing rings in the PSD, our algorithm based on template matching proved to be robust (see e.g. Fig. 3.15A,B). The rings are successfully mapped even when their completeness is compromised by external disturbance. Estimation from images with larger astigmatism values is still possible, although the rings can be incomplete (see Fig. 3.15B), due to the fact that the spatial envelope Eq. (F.69) dampens the contrast of the rings in one direction more than in another. Although such bad images could be discarded, we can still use them for defocus and astigmatism estimation illustrating the robustness of our technique.

The method takes the C_s influence into account and thus can be used on all microscopes (with or without C_s corrector). The algorithm's accuracy increases with the number of rings (see Table 3.2). Consequently, it might be beneficial to first estimate and correct astigmatism using higher magnification and then go back to the desired magnification. If only one or a few Thon rings are visible, it might also be advantageous to use an alternative pre-processing strategy that relies on bilateral filtering [127, 128] and provides a better segmentation of low-frequency rings. This option is included in the provided software implementation of our algorithm. Furthermore, the spatial (and/or tonal) blurring of the adaptive and/or bilateral filter could be modified to make the rings more prominent. An example of defocus and astigmatism estimation from the PSD

with barely visible Thon ring is shown in Fig. 3.15C.

Most of the algorithms developed so far (including ours) base their defocus estimation on the frequency of one or more minima in the PSD. This becomes quite a challenging task when the specimen is embedded in vitreous ice due to extremely low SNR. Alternatively to the PSD, some other measures can be used as the input for our algorithm, such as differential phase residual [129,130] or figure of merit [50,131,132]. These measures, however, rely on more than one acquisition. A remaining challenge is to accurately estimate the defocus at each location of the (non-)tilted specimens embedded in vitreous ice, especially if no amorphous carbon is present in the image.

The set of presented algorithms have been implemented in MATLAB and are available as a part of the image-processing toolbox DIPimage (<http://www.diplib.org>). Some of the possible applications of the algorithm are described in [133].

Acknowledgements

The authors would like to thank Bram Koster for useful discussions, Pieter Brandt for help on the k -trajectory method, and Remco Schoenmakers for critical reading of the manuscript.

Appendix A

Weak-phase weak-amplitude approximation

Assuming an incident plane wave $\Psi_{\text{in}}(x, y) = 1$, a weak potential (both the real and imaginary part of $v_z(x, y)$ in Eq. (3.10)), and applying the first order Taylor expansion, the exit wave from the specimen can be written as:

$$\Psi_{\text{ex}}(x, y) = e^{i\sigma v_z(x, y)} \approx 1 + i\sigma V_z(x, y) - \sigma \Lambda_z(x, y). \quad (\text{A.1})$$

The Fourier transform of the exit wave is then

$$\tilde{\Psi}_{\text{ex}}(q) = \delta(q) + i\sigma \tilde{V}_z(q) - \sigma \tilde{\Lambda}_z(q). \quad (\text{A.2})$$

Without loss of generality let us assume no astigmatism is present (i.e. $T(q, \alpha) = T(q)$). Substituting Eq. (3.3) in Eq. (3.4) we obtain

$$\tilde{\Psi}_{\text{im}}(q) = \tilde{\Psi}_{\text{ex}}(q) K_a(q) e^{-i\chi(q)} = \tilde{\Psi}_{\text{ex}}(q) K_a(q) [\cos(\chi(q)) - i \sin(\chi(q))] \quad (\text{A.3})$$

where $K_a(q) = K(q)A_p(q)$. $K(q)$ and $A_p(q)$ are defined in Eq. (3.8) and Eq. (3.9) respectively. The Fourier transform of the image intensity can be written as

$$\tilde{I}(q) = \mathcal{F}[|\Psi_{\text{im}}(x, y)|^2] = \tilde{\Psi}_{\text{im}}^*(-q) * \tilde{\Psi}_{\text{im}}(q). \quad (\text{A.4})$$

For point-symmetric aberrations such as defocus, astigmatism and spherical aberration it holds that $\chi(-q) = \chi(q)$. Considering that $V_z(x, y)$ and $\Lambda_z(x, y)$ are real we have $\tilde{V}_z^*(-q) = \tilde{V}_z(q)$ and $\tilde{\Lambda}_z^*(-q) = \tilde{\Lambda}_z(q)$ and therefore

$$\begin{aligned} \tilde{I}(q) = & \int_{-\infty}^{\infty} \left(\delta(q') - i\sigma \tilde{V}_z(q') - \sigma \tilde{\Lambda}_z(q') \right) \times K_a(q') [\cos(\chi(q')) + i \sin(\chi(q'))] \\ & \times \left(\delta(q - q') + i\sigma \tilde{V}_z(q - q') - \sigma \tilde{\Lambda}_z(q - q') \right) \times K_a(q - q') [\cos(\chi(q - q')) - i \sin(\chi(q - q'))] dq'. \end{aligned}$$

Since $\sigma \tilde{V}_z(q') \ll \delta(q)$ and $\sigma \tilde{\Lambda}_z(q') \ll \delta(q)$ we can neglect the second order terms in q and the convolution reduces to

$$\tilde{I}(q) = \delta(q) + 2K_a(q)\sigma \left[\tilde{V}_z(q) \sin(\chi(q)) - \tilde{\Lambda}_z(q) \cos(\chi(q)) \right]. \quad (\text{A.5})$$

Since

$$a \sin(x) - b \cos(x) = \operatorname{sgn}(a) \sqrt{a^2 + b^2} \sin\left(x - \arcsin\left(\frac{b}{\sqrt{a^2 + b^2}}\right)\right), \quad (\text{A.6})$$

equation Eq. (A.5) can be rewritten as

$$\tilde{I}(q) = \delta(q) + \operatorname{sgn}(\tilde{V}_z(q)) 2\sigma \sqrt{\tilde{V}_z(q)^2 + \tilde{\Lambda}_z(q)^2} K_a(q) \sin(\chi(q) - \arcsin(W(q))), \quad (\text{A.7})$$

where $W(q)$ is the amount of the amplitude contrast as defined in Eq. (3.11). Since $W(q)$ for thin amorphous carbon (PtIr) is typically $\sim 6 - 10\%$ [113], $\sqrt{\tilde{V}_z(q)^2 + \tilde{\Lambda}_z(q)^2} \approx |\tilde{V}_z(q)|$ and the final intensity can be expressed as

$$\tilde{I}(q) = \delta(q) + \sigma \tilde{V}_z(q) 2K_a(q) \sin(\chi(q) - \arcsin(W(q))). \quad (\text{A.8})$$

Appendix B

Templates

B.1 Derivation of the template ellipticity R_t

The central frequency of each generated template is in the middle of the frequency range (i.e. $q_{ct} = \frac{N}{4}$). For simplicity and without loss of generality, let the generated templates have $\alpha_1 = \frac{\pi}{4}$. From Eq. (3.22) and Eq. (3.30) the frequencies of the long and short axis can be expressed as:

$$q_l = \sqrt{\frac{q_c^2 (R_0^2 + 1)}{2}} \quad (\text{B.1a})$$

$$q_s = \sqrt{\frac{q_c^2 (R_0^2 + 1)}{2R_0^2}}. \quad (\text{B.1b})$$

From step 4 in Fig. 3.3, the peak-to-peak amplitude of the curve can be expressed as

$$A = q_l - q_s = q_s(R_0 - 1) = q_c(R_0 - 1) \sqrt{\frac{R_0^2 + 1}{2R_0^2}}. \quad (\text{B.2})$$

Templates match when the peak-to-peak amplitudes of the template and the pattern in the polar image are the same, i.e. $A_t = A_p$

$$(R_t - 1) \sqrt{\frac{R_t^2 + 1}{2R_t^2}} q_{ct} = (R_0 - 1) \sqrt{\frac{R_0^2 + 1}{2R_0^2}} q_c$$

$$(R_t - 1)^2 (R_t^2 + 1) = 2cR_t^2 q_c^2$$

where $c = \frac{R_0 - 1}{q_{ct}} \sqrt{\frac{R_0^2 + 1}{2R_0^2}}$ is constant. The solution that has physical meaning ($R_t \in \mathbb{R}^+$) gives the relation between the template ellipticity and the central frequency:

$$R_t(q_c) = \frac{1}{2} + \frac{1}{2} \sqrt{2c^2 q_c^2 + 1} + \frac{1}{\sqrt{2}} [\sqrt{2c^2 q_c^2 + 1} + c^2 q_c^2 - 1]^{\frac{1}{2}} \quad (\text{B.3})$$

B.2 The difference between detected q_{found} and central frequency q_c

Combining Eq. (3.22) and Eq. (3.26) yields the polar representation of an ellipse

$$C(\alpha) = \frac{q_l}{\sqrt{\cos^2 \alpha + (R_0 \sin \alpha)^2}}. \quad (\text{B.4})$$

Its mean value is

$$q_m(R_t) = \langle C(\alpha) \rangle_\alpha = \frac{q_l}{\pi} I_{\text{el}}(R_t), \quad (\text{B.5})$$

where

$$I_{\text{el}}(R_t) = \int_0^\pi \frac{d\theta}{\sqrt{1 + (R_t^2 - 1) \sin^2 \theta}} \quad (\text{B.6})$$

is the incomplete elliptic integral of the first kind. Since the maxima in the parameter space provide also R_t we can use it to numerically solve the integral $I_{\text{el}}(R_t)$. Using Eq. (B.1a) the relative error between the mean (Eq. (B.5)) and central frequency of Eq. (B.4) is:

$$\varepsilon_{R_t} = \frac{q_m - q_c}{q_m} = \frac{\sqrt{1 + R_t^2} I_{\text{el}}(R_t) - \sqrt{2}\pi}{\sqrt{1 + R_t^2} I_{\text{el}}(R_t)}. \quad (\text{B.7})$$

The response of the template matching depends on the difference between the mean value of the polar transformed Thon ring q_m and the mean value of the generated template $q_{m,t}$. Since central frequencies of the templates are fixed to $\frac{N}{4}$, the mean values $q_{m,t}$ are slightly shifted and that indicates that

$$q_{\text{found}}(R_t) = q_m - q_{m,t} + \frac{N}{4}. \quad (\text{B.8})$$

The central frequency that is needed for defocus and astigmatism estimation is

$$q_c = q_m(1 - \varepsilon_{R_t}). \quad (\text{B.9})$$

From Eq. (B.8) and Eq. (B.9) we can write:

$$q_{\text{found}}(R_t) = \frac{q_c - \frac{N}{4}}{1 - \varepsilon_{R_t}} + \frac{N}{4}. \quad (\text{B.10})$$

Thus the central frequency as a function of the found response in parameter space is

$$q_c = (q_{\text{found}} - \frac{N}{4})(1 - \varepsilon_{R_t}) + \frac{N}{4}. \quad (\text{B.11})$$

Appendix C

Thon ring outlier rejection

From the collection of possible Thon ring candidates \mathcal{C} ordered by frequency q we calculate a list of selected Thon rings \mathcal{S} given by

$$\mathcal{S} = \{(q_1, s_1), (q_2, s_2), \dots, (q_N, s_N)\} \quad (\text{C.1})$$

where \mathcal{S} is a subset of \mathcal{C} with an extra element s_i added to the tuple, which specifies how many Thon rings are skipped between the selected Thon ring i and $i - 1$.

Outlier rejection restricts the number of possible subsets \mathcal{S} by the following restrictions:

$$\forall_i : \left| \frac{q_i^2 - q_{i-1}^2}{s_i \cdot d_i} \right| \leq \text{maxRelativeError} \quad (\text{C.2})$$

and

$$\sum_{i=1}^N s_i \leq \text{maxThonRingsSkip} \quad (\text{C.3})$$

where d_i represents the expected q^2 -distance between Thon rings $i - 1$ and i , which is recursively defined as

$$d_{i+1} = \frac{1}{2} \left(\frac{q_i^2 - q_{i-1}^2}{s_i} + d_i \right), \quad (\text{C.4})$$

$$d_1 = \text{median}(q_i^2 - q_{i-1}^2). \quad (\text{C.5})$$

By default our implementation allows an error of equidistance of 20 % ($\text{maxRelativeError} = 0.2$) and the maximal number of skipped Thon rings is set to $\text{maxThonRingsSkip} = 6$. The reason for the recursive definition in Eq. (C.4) is that we get an IIR-filter-like refinement of the q^2 -distance between Thon rings as we increase q , which is desirable as in fact the distance is not truly constant for $C_s \neq 0$. Furthermore, the distance also changes in the presence of amplitude contrast.

Appendix D

Spherical aberration influence

D.1 C_s influence on the ellipticity

Thon ring frequencies in the PSD correspond to the zeros of the CTF ($\chi(q, \alpha) = k_{\text{eff}}\pi$). The frequencies of the Thon rings in long/short axis orientation can be found from

$$\frac{C_s \lambda^3}{2} q_{l,s}^4 - \Delta f_{l,s} \lambda q_{l,s}^2 - k_{\text{eff}} = 0. \quad (\text{D.1})$$

It follows that

$$q_{l,s}^2 = \frac{\Delta f_{l,s} \pm \sqrt{\Delta f_{l,s}^2 + 2C_s k_{\text{eff}}}}{C_s \lambda^2}. \quad (\text{D.2})$$

For weak-amplitude samples $k_{\text{eff}} \approx k$ holds. The apparent ellipticity of the ring i is then

$$Q_i^2(k) = \frac{\Delta f_s \pm \sqrt{\Delta f_s^2 + 2C_s k}}{\Delta f_i \pm \sqrt{\Delta f_i^2 + 2C_s k}}, \quad (\text{D.3})$$

where i is the order of CTF zero for corresponding k -value. Since $Q_i^2(k) > 1$, we keep $-\sqrt{\Delta f_{l,s}^2 + 2C_s k_{\text{eff}}}$ for overfocus ($\Delta f < 0$, $A_1 < 0$ and $k \in \mathbb{N}$), and $+\sqrt{\Delta f_{l,s}^2 + 2C_s k_{\text{eff}}}$ for underfocus ($k \in \mathbb{Z}$ and $|k| \leq N_{0\text{max}}$). Eq. (D.3) can be written as

$$Q_i(k) = \sqrt{\frac{|\Delta f_s| + \sqrt{\Delta f_s^2 + 2C_s k}}{|\Delta f_i| + \sqrt{\Delta f_i^2 + 2C_s k}}} \quad (\text{35})$$

and its solutions are real and Thon rings are elliptic-like as long as

$$k \geq -\frac{\Delta f_i^2}{2C_s \lambda}. \quad (\text{D.4})$$

From Eq. (3.1) it is expected that the ellipticity of the rings decreases with frequency due to the influence of C_s which is angularly symmetric. Similarly, by increasing C_s , the apparent

ellipticity at a certain frequency should decrease (note, however, that changes in C_s are less influential than changes in q in Eq. (3.1)). This is directly visible in overfocus where ellipticity decreases monotonically with frequency and/or C_s . In underfocus, however, initially it increases after which it decreases. If the initial increase (in underfocus) is large, the condition Eq. (D.4) might not be satisfied, implying the formation of the rings that are no more elliptic-like but rather hyperbolic-like.

D.2 C_s influence on CTF minima position $q_{c,i}$

For the case $C_s = 0$, the neighboring CTF minima in squared frequency space are equidistant:

$$\Delta q_{c,i}^2|_{C_s=0} = \frac{\Delta k_i}{-\lambda \Delta f}, \text{ with } |\Delta k_i| \equiv |k_{i+1} - k_i| = 1. \quad (\text{D.5})$$

When C_s cannot be neglected, the position of the CTF minima can be found from

$$\begin{aligned} \frac{C_s \lambda^3}{2} q_c^4 - \lambda \Delta f q_c^2 &= k, \\ \frac{C_s \lambda^3}{2} (q_{c,i+1}^4 - q_{c,i}^4) - \lambda \Delta f (q_{c,i+1}^2 - q_{c,i}^2) &= \Delta k_i. \end{aligned}$$

The distance between neighboring minima in squared frequency space is now:

$$\Delta q_{c,i}^2|_{C_s \neq 0} = \frac{\Delta k_i}{-\lambda \Delta f + \frac{C_s \lambda^3}{2} (q_{c,i+1}^2 + q_{c,i}^2)}. \quad (\text{D.6})$$

If β is the fraction of the C_s influence defined in Eq. (3.33) then we have

$$q^2 = \frac{2\beta|\Delta f|}{C_s \lambda^2}. \quad (\text{D.7})$$

Substituting Eq. (D.7) in Eq. (D.6) we obtain

$$\Delta q_{c,i}^2|_{C_s \neq 0} = \frac{\Delta k_i}{-\lambda \Delta f + \lambda \Delta f (\beta_{i+1} + \beta_i)}. \quad (\text{D.8})$$

The relative error between equidistant CTF zeros ($C_s = 0$) and distances when $C_s \neq 0$ can be presented as

$$\varepsilon_{C_s} = \frac{\Delta q_{c,i}^2|_{C_s \neq 0} - \Delta q_{c,i}^2|_{C_s=0}}{\Delta q_{c,i}^2|_{C_s=0}} = \frac{1}{1 - (\beta_{i+1} + \beta_i)} - 1. \quad (\text{D.9})$$

For example if $\beta_i \approx 10\%$ then $\varepsilon_{C_s} = 25\%$.

D.3 Correction for the C_s influence on the ring ellipticities

When $C_s > 0$, the Thon ring ellipses (that is, approximate ellipses), do not all have the same ellipticity. Therefore, we have to make a clear distinction in ellipticity of an individual Thon ring ellipse, which we will call Q_i for Thon ring i , given by

$$Q_i = \frac{q_{l,i}}{q_{s,i}} \quad (3.34)$$

where the long axis in the PSD is given by frequency $q_{l,i}$ and short axis by $q_{s,i}$. We will keep on using the symbol R_0 as the dimensionless measure of astigmatism given by

$$R_0 = \sqrt{\frac{\Delta f + A_1}{\Delta f - A_1}}. \quad (3.21)$$

Note that $Q_i|_{C_s=0} = R_0$ for all Thon rings. For $C_s > 0$, however, we detect Q_i for each Thon ring, but how to find the equivalent ellipticity R_{eq} for all rings? To obtain this relation, we define the frequency q_v that is equivalent to the frequency q if C_s would be zero. That is, their phases and k -values in Eq. (3.1) are equal.

$$\frac{1}{2}C_s\lambda^3q^4 - \lambda q^2\Delta f = -\lambda q_v^2\Delta f. \quad (D.10)$$

Solving for q_v^2 yields

$$q_v^2 = \frac{q^2\Delta f - \frac{1}{2}C_s\lambda^2q^4}{\Delta f}. \quad (D.11)$$

The frequency q_v is always real in overfocus. However, in underfocus the additional relation $\frac{2|\Delta f|}{C_s\lambda^2q^2} > 1$ must be fulfilled. If we use Eq. (D.11) to get values q_{vl} and q_{vs} for long and short axes, we recalculate R_{eq} by using

$$R_{eq}^2 = \frac{q_{vl}^2}{q_{vs}^2} = \frac{(2\Delta f_l q_l^2 - C_s\lambda^2 q_l^4) \Delta f_s}{(2\Delta f_s q_s^2 - C_s\lambda^2 q_s^4) \Delta f_l}. \quad (D.12)$$

The numerator and denominator of the first fraction in the right-hand-side term are equal to k and the whole first fraction is equal to one. Thus, $R_{eq} = R_0$ equivalent ellipticity is equal to the ellipticity when $C_s = 0$. The problem is that we do not know this R_0 . R_{eq} can be further rewritten as:

$$R_{eq} = R_0 \sqrt{\frac{q_{l,i}^2}{q_{s,i}^2}} \sqrt{\frac{(2\Delta f_l - C_s\lambda^2 q_{l,i}^2)}{(2\Delta f_s - C_s\lambda^2 q_{s,i}^2)}}. \quad (D.13)$$

From the first estimate (up to $\beta = 10\%$) we get initial values for R_0 , Δf_l and Δf_s . Furthermore, we refine the estimate by finding Q_i from the whole spectrum. These values are scaled with the second fraction in Eq. (D.13) and in this way the final R_{eq} is obtained.

Appendix E

Thon ring averaging

This section describes our new method for obtaining 1D profiles from the PSD of a micrograph. The most basic method used to obtain such a 1D profile is circular averaging, calculated using the discretized form (i.e. integration becomes summation) of the following equation

$$p(q) = \frac{1}{\pi} \int_0^{\pi} d\alpha \int_{-3\sigma(q)}^{+3\sigma(q)} P(q + q', \alpha) G_{\sigma(q)}(q') dq'. \quad (\text{E.1})$$

where G_{σ} is a Gaussian kernel of scale σ , which can be a function of the radial frequency q . Some blurring with the Gaussian is applied to ensure smooth results on the discretized power spectrum. The sum over q' is bound to an interval of e.g. $-3\sigma, +3\sigma$ to make the implementation efficient but also approximate the Gaussian accurately. Circular averaging only exactly follows the Thon rings when there is no astigmatism. With astigmatism, one should use elliptical averaging, defined as

$$p_{R,\alpha_1}(q) = \frac{1}{\pi} \int_0^{\pi} d\alpha \int_{-3\sigma(q)}^{+3\sigma(q)} P(q', \alpha) G_{\sigma(q)}(q') dq', \quad (\text{E.2})$$

$$P(q', \alpha) = P\left(\frac{q + q'}{\sqrt{1 + (R^2 - 1) \sin^2(\alpha - \alpha_1)}}, \alpha\right),$$

where ellipticity R and α_1 represent the astigmatism influence.

When $C_s \neq 0$, Thon rings start to deviate from ellipses. With Thon ring averaging, we aim at getting averages over Thon rings as function of their central frequencies q_c . To correctly average over Thon rings we consider Eq. (3.1). Using this equation, we can find the ‘‘nominal radius’’ q_c of any position in the PSD (so not only frequencies of the Thon rings) by equating the latter formula to the same formula without the astigmatism term:

$$\frac{1}{2} C_s \lambda^3 q^4 - \lambda q^2 (\Delta f - A_1 \cos(2(\alpha - \alpha_1))) = \frac{1}{2} C_s \lambda^3 q_c^4 - \lambda q_c^2 \Delta f. \quad (\text{E.3})$$

Solving for q_c^2 we find

$$q_c^2 = \frac{\Delta f \pm \sqrt{\Delta f^2 + 2C_s \lambda k}}{C_s \lambda^2}. \quad (\text{E.4})$$

where $k = \frac{1}{2}C_s\lambda^3q^4 + \lambda q^2(\Delta f - A_1 \cos(2(\alpha - \alpha_1)))$. The “ \pm ” sign in Eq. (E.4) is plus for overfocus and for monotonic increase of k values in underfocus, and minus when k values in underfocus monotonically decrease. The implementation of Thon ring averaging works as follows:

1. Create two empty 1D arrays `result` and `sum` of size $N/2$ and initialize with zeroes.
2. For each power spectrum position (q_x, q_y) :
 - (a) Convert coordinates (q_x, q_y) to polar coordinates (q, α) and calculate the corresponding q_c using Eq. (E.4)
 - (b) Add the Gaussian weighted response $G(q' - q_c)P(q, \alpha)$ to `result` by adding its value to the bins in the interval $[q_c - 3\sigma, q_c + 3\sigma]$.
 - (c) Add the responses of the Gaussian weight in the corresponding bin of the array `sum`.
3. Divide all elements of `result` componentwise by the elements of `sum`. Return `result`.

Chapter 4

Forward model

Published as [51]: M. Vulović, R. B. G. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lüken, H. Rullgård, O. Öktem, B. Rieger, “Image formation modeling in cryo-electron microscopy”, *Journal of Structural Biology* (2013), accepted

Abstract

Accurate modeling of image formation in cryo-electron microscopy is an important requirement for quantitative image interpretation and optimization of the data acquisition strategy. Here we present a forward model that accounts for the specimen’s scattering properties, microscope optics, and detector response. The interaction potential is calculated with the isolated atom superposition approximation (IASA) and extended with the influences of solvent’s dielectric and ionic properties as well as the molecular electrostatic distribution. We account for an effective charge redistribution *via* the Poisson-Boltzmann approach and find that the IASA-based potential forms the dominant part of the interaction potential, as the contribution of the redistribution is less than 10 %. The electron wave is propagated through the specimen by a multislice approach and the influence of the optics is included *via* the contrast transfer function. We incorporate the detective quantum efficiency of the camera due to the difference between signal and noise transfer characteristics, instead of using only the modulation transfer function. The full model was validated against experimental images of 20S proteasome, hemoglobin, and GroEL. The simulations adequately predict the effects of phase contrast, changes due to the integrated electron flux, thickness, inelastic scattering, detective quantum efficiency and acceleration voltage. We suggest that beam-induced specimen movements are relevant in the experiments whereas the influence of the solvent amorphousness can be neglected. All simulation parameters are based on physical principles and, when necessary, experimentally determined.

4.1 Introduction

The structures of macromolecules, macromolecular complexes and subcellular assemblies provide insight into their functions. Knowledge of the 3D structure of a macromolecule is also the cornerstone for rational drug design [3].

Cryo-electron microscopy (cryo-EM) of biological specimens in an unstained, frozen-hydrated state has become an indispensable tool for structural biology [1]. Advances in cryo-EM single particle analysis (SPA) [2] and cryo-electron tomography (cryo-ET) [134–136] provide opportunities to characterize the structures of macromolecular complexes that are either too flexible, heterogeneous or transient to be explored by crystallographic methods [137, 138]. The level of structural detail that can be obtained by cryo-EM is largely limited by specimen heterogeneity, the effective contrast transfer function (CTF), the detector's detective quantum efficiency (DQE), and radiation damage which limits the integrated electron flux that can be used, resulting in a poor signal-to-noise ratio (SNR) in images.

In addition to hardware developments, computational methods will continue to improve, enabling more information to be extracted from inherently noisy cryo-EM images. Simulations of electron images will be increasingly important in order to optimize the data acquisition strategy, to improve image interpretation and resolution, and to provide insight on ways to improve instrumentation. An accurate forward model of image formation in cryo-EM should rely on all relevant physical properties such as the specimen's elastic and inelastic scattering properties and the effects of the CTF and the detector.

Simulation of transmission electron microscope (TEM) images of biological specimens is implemented in a number of software packages for SPA and ET such as Xmipp [31, 32], IMAGIC [33], SPIDER [34, 35], EMAN2 [36], Bsoft [37], and TOMToolbox [38]. In most cases, these simulations are used to facilitate Euler angles determination in SPA and to evaluate reconstruction methods for SPA [139, 140] and ET [141]. Usually a virtual model of a biological specimen is created using 3D primitives (phantoms) such as spheres, ellipsoids, cubes, and cylinders [32]. In some cases, the specimen volume is constructed based on information from the RCSB Protein Data Bank (PDB) and TEM images are computed by projecting the 3D specimen; the effects of the solvent and detector are rarely accounted for. In general, projecting the 3D electron density distribution into a 2D image is not correct, since it does not represent the actual physical electron-specimen scattering properties (interaction potential). In addition, the noise is often simplified as being additive Gaussian noise. Below, we discuss two related work that aim to provide more realistic simulations.

In [142], image simulations were performed to assess the attainable benefits of phase plates. The solvent (water) was treated explicitly *via* molecular dynamics (MD) simulations generating a box of amorphous water and a multislice approach was used to account for the specimen thickness and multiple scattering. The generated noise was Poisson distributed, but the detector response was not included. Unfortunately, the methods were not validated experimentally.

TEM-simulator [39] aims to provide accurate simulations based on physical principles. It was the first simulator whose results were compared to experimental data, albeit not in depth. There, the specimen thickness has been neglected, and low-pass filtering to a certain resolution exceedingly damps the interaction potential (IP). Although most simulation parameters described there are based on physical principles, a calibration protocol needs to be employed for some parameters that are phenomenologically introduced, leading to a situation where nuisance parameter tuning is required. Examples of such phenomenological parameters are amorphousness (granularity), absorption potential, as well as camera parameters such as the modulation-transfer function (MTF), detective quantum efficiency (DQE), and conversion factor. Further-

more, none of the aforementioned approaches have considered chemical bonding and/or interaction of the sample with solvent and ions. For completeness, it should be mentioned that recently [143] parameterized a function that describes the distribution of water molecules around a protein. In previous work the solvent was assumed to be water, instead of less dense vitreous ice, leading to possible artificial damping of the contrast between the protein (which has a higher density than water) and solvent.

For material science applications, numerous TEM simulators have been developed (reviewed by [144]). Many assume that the atoms of a specimen are periodically ordered which is not fulfilled for non-crystalline biological specimens. Some of the simulators, such as YAMS [145, 146] and SimulaTEM [147], have been used for image simulations of biological specimens. They do not assume that the specimen is periodic and although YAMS propagates the mutual coherence function through the specimen, a method more appropriate for treating the partial incoherence, only elastic scattering was assumed for biological specimens [140]. In both simulators the specimen thickness and multiple scattering events were treated *via* a multislice approach [148], but inelastic scattering, the detector response, and solvent were ignored. In high resolution electron microscopy (HREM) the contrast in experimental images has been frequently reported to be much less, typically about a factor of three, than predicted by image simulation [149, 150]. It was suggested in [151] that this discrepancy, often called the Stobbs-factor, originates from neglecting the detector's MTF in image simulations.

Here we present, analyze and validate an image formation model in TEM based on physical principles. In addition to computing the 3D potential distribution where atoms are treated in isolation, the interaction redistribution potential due to the solvent, ions and molecular interactions is computed. Beam-induced motion and amorphousness of the vitreous ice are also addressed. For validation, comparisons between experiments and simulations were performed on cryo-embedded specimens. Some of the parameters such as defocus, astigmatism and camera properties are accurately estimated from experiments *via* available toolboxes [47, 49], without introducing nuisance parameters. The simulator presented here, InSilicoTEM, has been implemented in DIPimage (www.DIPlib.org), a MATLAB toolbox for scientific image processing and analysis, and is freely available for non-commercial use upon request.

4.2 Theory

Forward modeling approaches in cryo-EM describe the complex image formation process. Below, we will shortly outline our image formation model whose main ingredients are: the interaction potential, electron wave propagation, and intensity detection by the camera. Appendix F provides a detailed description of all steps and approximations.

4.2.1 Interaction potential (IP)

The interaction between the incident electron wave and a macromolecule embedded in the surrounding medium is modeled as a sum of two interaction potential components: (1) “atom” contributions, i.e. the superposition of atomic potentials as if each atom was in isolation; and

(2) “bond” contributions, i.e. the influence of the charge redistribution due to the solvent, ions and molecular interactions

$$\mathcal{V}^{\text{int}}(\mathbf{r}) = \mathcal{V}_{\text{atom}}(\mathbf{r}) + \mathcal{V}_{\text{bond}}(\mathbf{r}), \quad (4.1)$$

where $\mathbf{r} = (x, y, z)$ is the position of the electron wave. Since $\mathcal{V}_{\text{atom}}$ considers the specimen as a set of isolated atoms, we get $\mathcal{V}_{\text{atom}}(\mathbf{r}) = \sum_{j=1}^m \mathcal{V}_{Z_j}(\mathbf{r} - \mathbf{R}_j)$, where \mathcal{V}_{Z_j} is the electrostatic potential of an isolated neutral atom with atomic number Z_j centered at \mathbf{R}_j . With the first Born approximation, such a potential can be written as the inverse Fourier transform of the electron scattering factor of the atom [10, 39] (see F.1.2 in Appendix F).

The isolated atom superposition approximation (IASA) ignores the potential due to the charge redistributions, $\mathcal{V}_{\text{bond}}$, which accounts for the interaction with neighboring atoms, solvent and ions. As $\mathcal{V}_{\text{atom}}$ provides the most significant contribution to the scattering of the incident electron, this computationally convenient approximation provides a good starting point for initial interpretation of high-energy electron diffraction and microscopy experiments [144, 152]. Biological specimens are embedded in an amorphous solvent and the potential distribution depends also on the dielectric and ionic properties of the solvent. It seems appropriate to include the contribution of the solvent and ions modeled by $\mathcal{V}_{\text{bond}}$. This potential due to the charge redistribution can be accounted for *via* a continuum electrostatics approach (see F.1.3 in Appendix F), described by the solution of the linearized Poisson-Boltzmann equation:

$$-\epsilon_0 \nabla(\epsilon_r(\mathbf{r}) \nabla \mathcal{V}_{\text{bond}}(\mathbf{r})) = \rho_{\text{mol}}^{\text{bond}}(\mathbf{r}) + \rho_{\text{sol}}^{\text{bond}}(\mathbf{r}) - \alpha(\mathbf{r}) \sum_i \frac{q_i^2 n_i^0 \mathcal{V}_{\text{bond}}(\mathbf{r})}{k_B T}, \quad (4.2)$$

where ϵ_0 is the permittivity of the vacuum, ϵ_r the relative permittivity (ϵ_{sol} for the sub-volume occupied by solvent and ϵ_{mol} for the molecule), $\rho_{\text{mol}}^{\text{bond}}(\mathbf{r})$ and $\rho_{\text{sol}}^{\text{bond}}(\mathbf{r})$ are the partial (net) charges of the molecule and solvent respectively; q_i and n_i^0 are respectively the charge and the concentration of an ion of type i ; k_B the Boltzmann constant, T the temperature, and $\alpha(\mathbf{r}) = 1$ for sub-volume occupied by solvent, and $\alpha(\mathbf{r}) = 0$ otherwise. In order to meaningfully add the two potential contributions (equation (4.1)), the assumptions specified in F.1.4 must be fulfilled.

Inelastic contributions

The effects of inelastic scattering are modeled as the imaginary part of the interaction potential. The total complex potential is $\mathcal{V}_{\text{tot}}^{\text{int}} = V_{\text{ph}} + iV_{\text{ab}}$ (see F.2.3), where V_{ph} is the interaction potential (real value) as described in the previous section. V_{ph} contributes to the phase contrast while V_{ab} influences the amplitude (absorption) contrast. Contributions to the amplitude contrast can be roughly separated into “plasmons”, electrons scattered outside the aperture, and atom core losses. “Plasmons” are not strictly oscillations of free electrons like in metals, but they are producing a similar amount of energy loss (~ 20 eV), hence this commonly used terminology [153]. In a typical electron energy-loss spectroscopy (EELS) spectrum, the intensities due to atom core losses ($\Delta E > 100$ eV) are a couple of order of magnitudes smaller than those of the plasmons ($\Delta E \sim 20$ eV). The influence of the aperture will be taken into account *via* the optical system. Therefore, the plasmons are considered the most dominant contribution to the inelastic interactions. Since a large part of the specimen consists of embedding medium, the plasmons

of vitreous ice damp the useful phase signal. For an amorphous solvent such as vitreous ice and a certain incident electron energy the plasmons can be described *via* the inelastic mean free path Λ_{in} . We performed Monte-Carlo simulations [154, 155] to validate the assumption that delocalized processes (represented by a constant absorption potential) are dominant (see F.2.3). For our purpose the imaginary part of the potential is modeled as

$$V_{\text{ab}}(x, y, z) = 1/(2\sigma\Lambda_{\text{in}}), \quad (4.3)$$

where $\sigma = \lambda me/(2\pi\hbar^2)$ is the interaction constant, λ , e , and m the relativistic wavelength, charge and mass of the incident electron, and h Planck's constant.

4.2.2 Electron wave propagation

The electron wave propagation through the specimen is based on a multislice method [148] that accounts for the thickness of the specimen and multiple scattering [144]. An incident electron is described by its wave function and at the top of the $(n + 1)$ th slice of the specimen, the wave function is given by

$$\Psi_{n+1}(x, y) = \mathcal{F}^{-1} \left[P_n(q_x, q_y, \Delta z_n) \mathcal{F}[\exp(i\sigma V_z(x, y, z)) \Psi_n(x, y)] \right], \quad (4.4)$$

where $P(q, \Delta z) = \exp(-i\pi\lambda\Delta z q^2)$ is the Fresnel propagator over a slice of thickness Δz , q is the magnitude of the spatial frequency (q_x, q_y), $\mathcal{F}[\circ]$ denotes the Fourier transformation, and

$V_z(x, y, z) = \int_z^{z+\Delta z} \mathcal{V}^{\text{int}}(x, y, z') dz'$ is the projected potential within the slice. Parallel illumination is modeled as an incident plane wave ($\Psi_0(x, y) = 1$). The propagation of the electron wave through the specimen can be interpreted as recursive transmission and propagation of the wave function through each slice until the wave leaves the specimen ($\Psi_{\text{exit}}(x, y)$).

In cryo-EM the images are mostly generated by phase contrast, as a result of interference between the unscattered and scattered part of the electron exit wave function. The electron wave exiting the specimen $\Psi_{\text{exit}}(x, y)$ is further subject to a frequency dependent phase shift introduced by the defocus Δf and microscope aberrations such as spherical aberration C_s , and twofold astigmatism (A_1, α_1). The contrast transfer function (CTF) of the lens system in polar coordinates is [111]

$$T(q, \alpha) = KA_p \exp\left(-i\frac{2\pi}{\lambda} \left(\frac{1}{4}C_s\lambda^4 q^4 - \frac{1}{2}(\Delta f - A_1 \cos(2(\alpha - \alpha_1))) \lambda^2 q^2\right)\right), \quad (4.5)$$

where A_p is the objective aperture function and K describes spatial and chromatic envelopes. Note that underfocus implies $\Delta f > 0$, as in [111]. The intensity in the image plane is the probability density function given by

$$I_0(x, y) = |\Psi(x, y)|^2 = |\mathcal{F}^{-1}[\mathcal{F}[\Psi_{\text{exit}}(x, y)] T(q, \alpha)]|^2. \quad (4.6)$$

For details see section F.3 in Appendix F.

4.2.3 Detector response

Capturing the final image involves the conversion of the electron wave intensity distribution into a digital signal *via* a detector. Electron detectors are characterized by parameters such as conversion factor CF in [ADU/e⁻], modulation transfer function (MTF) and detective quantum efficiency (DQE). The measurement process obeys Poisson statistics giving rise to shot noise; the detector adds readout noise I_m and dark current I_{dc} to the final image, and blurs the image with a detector point-spread function PSF (x, y) which Fourier transform is the MTF.

The MTF describes transfer of the signal amplitude for different spatial frequencies. However, the signal and the noise in a TEM detector are not transferred in the same way [20]. The DQE is defined as $DQE(q) = MTF^2(q) / NTF^2(q)$, where the NTF is the noise transfer function ($NTF^2(q) = NPS_{out} / (CF^2 \Phi_e)$) with NPS being the noise power spectrum, and Φ_e the incident electron flux in [e⁻ / area]. We model the signal and noise propagation as follows: 1) the Fourier spectrum of the noise-free signal ($\tilde{I}_0(q)$) is damped (multiplied) by the ratio between signal (MTF) and noise (NTF) transfer, 2) this signal is multiplied by the integrated electron flux and noise contributions are added, 3) the Fourier spectrum of that (noisy) signal is damped by the NTF, and 4) the number of electrons are scaled with CF to the image gray values in [ADU]. Hence, we can write the detected image as

$$I(x, y) = CF \cdot \mathcal{F}^{-1} \left[\mathcal{F} \left[P_{oiss} \left(\Phi_e \cdot \mathcal{F}^{-1} [\tilde{I}_0(q) \sqrt{DQE(q)}] \right) \right] \cdot NTF(q) \right] + I_m + I_{dc}, \quad (4.7)$$

where $P_{oiss}(A)$ returns a random number from a Poisson distribution with expected value A . Section F.4 in Appendix F explains the steps in more detail.

4.3 Computational methods

The main steps of image formation simulations are i) construction of the interaction potential (IP) and ii) electron wave propagation and recording intensity.

Physical parameters of the specimen include pH, dielectric constant, temperature, ion concentration, motion factor and thickness. Microscope parameters involve acceleration voltage and its spread, opening angle, defocus, astigmatism, spherical and chromatic aberrations, objective aperture, magnification, and incident electron flux. Relevant camera parameters are exposure time, binning, conversion factor, MTF, DQE, readout and dark current noise. All parameters influencing the image formation are based on physical principles and when necessary, they were estimated from the experiment, using independent measurements (except beam-induced movements), without introducing nuisance parameters.

In this section, we outline the computational methods for image simulation of biological specimens and parameters estimation.

4.3.1 Interaction potential (IP)

A forward simulation requires a known model of the specimen. In case of biological specimens, we construct the IP using a hybrid approach combining the isolated atom superposition

Table 4.1. Some APBS parameters

Symbol	Value	Meaning
pdie	2.00	dielectric constant of the solute
sdie	78.54	dielectric constant of the solvent
temp	298.15	temperature of the system [K]
srad	1.40	radius of the solvent molecules [\AA]
ion	+1 0 2	ion species, concentration [M] and radius [\AA]

approximation (IASA) and a Poisson-Boltzmann (PB) description of the interaction between the macromolecule and its solvent and ions (see 4.2.1). The input for computing the IP is a high-resolution X-ray structure as deposited within the PDB which contains a detailed specification of type and position of most atoms in the molecule. Here we used PDB files 1RYP, 1GR5 and 2GTL to model 20S proteasome, GroEL, and earthworm hemoglobin, respectively. The plasmons are accounted for *via* the inelastic mean free path. The amorphousness of the solvent was generated by an explicit atomic model *via* MD simulations. An empirical post-blurring can be applied, which results in a similar effect that beam-induced movements could have. The next subsections explain these procedures in more detail.

Isolated atom superposition approximation (IASA)

The dominant part of the interaction potential is the sum of the individual isolated atomic potentials calculated as the Fourier transforms of tabulated electron scattering factors. There are several empirical closed-form approximations of electron scattering factors available [144]. We use scattering factors that are parameterized as a weighted sum of five Gaussians as given in Table 1 in [152] and implemented in TEM-simulator [39]. The real potential map calculations are based on a slight modification of their map in such a way that low-pass filtering to a certain resolution does not exceedingly damp the interaction potential (IP), and the solvent is assumed to be vitreous ice instead of water (see equation (F.20)). The input PDB file is converted into the electrostatic potential map $\mathcal{V}_{\text{atom}}$. The voxel size of the map in this analysis was set to 1 \AA .

The influence of the embedding environment *via* the Poisson-Boltzmann (PB) approach

We use a continuous electrostatics method to model the influence of the solvent and ions as well as the coarse electrostatic potential redistributions within the macromolecule. There are different implementations for solving the Poisson-Boltzmann (PB) equation. In this study we used APBS (adaptive Poisson-Boltzmann solver) [156] that numerically solves the PB equation for solvation energy and potential.

The input for APBS is a modified PDB file (PQR) where the occupancy and temperature fields are substituted with partial charges and the radii fields using PDB2PQR [157]. Since protein structures deposited in the PDB format usually lack hydrogen atoms, PDB2PQR offers

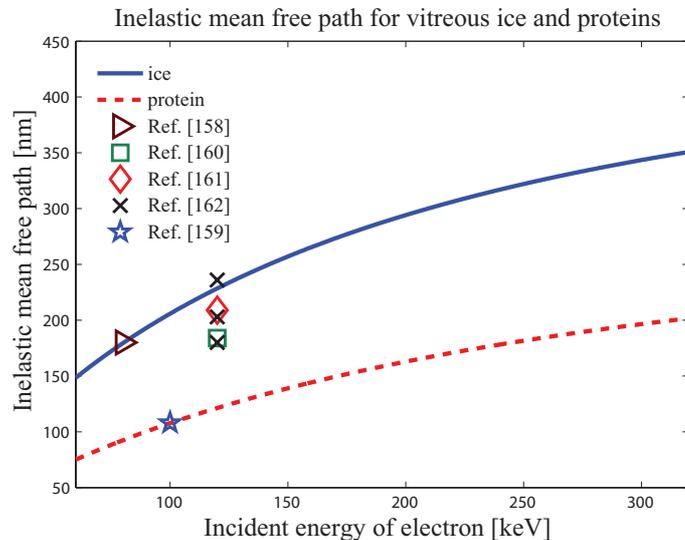


Fig. 4.1. Inelastic mean free path as a function of the incident electron energy for vitreous ice (blue solid) and protein (red dashed line) using equation (4.8) and data derived from [158] and [159]. The data points from references [158, 160–162] and [159] are included.

the functionality of adding missing hydrogens atoms and removing steric clashes caused by the newly added hydrogens. Partial (net) charges were calculated with AMBER, one of the forcefields available in PDB2PQR. The pH value was set to 7. The APBS input file contains both numerical aspects of the computation and physical parameters (c.f. Table 4.1).

For large molecules such as earthworm hemoglobin we adapted the procedure for calculating PQR files and APBS potentials. For large (constructed) PDB files we adopted a variable column width. The parsers also allowed a more flexible spacing between all fields and larger (unrestricted) field size. In order to calculate the potential map of hemoglobin, the molecule was split into eight parts with an overlap of 10 %. A single potential map was assembled from all parts. 20S Proteasome was simulated without ions in the solvent while the ion concentrations for the earthworm hemoglobin sample were 0.05, 1, and 3 M (mol/l), respectively.

In contrast to typical PB solvers that include two-step solvation energy calculations, here we used a one-step approach. For chemistry and biophysics applications, the reaction fields due to the polarization of the solvent and ions around the molecule are of interest and a two-step approach is needed. In that case, homogeneous dielectric calculations (dielectric constants of the molecule and solvent are equal) are subtracted from heterogeneous calculations (dielectric constants differ). Since the knowledge about the electrostatic potential redistribution within the molecule is beneficial for us, we did not need to perform homogeneous dielectric calculations (see F.1.4), resulting in reduced computation times.

Inelastic scattering

For our purpose the imaginary part of the potential was modeled *via* the inelastic mean free path (see equation (4.3)). As described in [158] and [163] the inelastic scattering cross section σ_{in}

can be represented *via* equation (F.60). The inelastic mean free path is related to σ_{in} as

$$\Lambda_{\text{in}} = \frac{M_{\text{W}}}{\rho N_{\text{A}} \sigma_{\text{in}}} = \frac{M_{\text{W}} \beta^2 \cdot 10^{10}}{9.03 \rho Z^{\frac{1}{2}} \ln \frac{\beta^2 (U_0 + mc^2)}{10}} [\text{nm}], \quad (4.8)$$

where Z is the atomic number, β the ratio between the velocity of electron and light ($\beta^2 = 1 - [mc^2 / (U_0 + mc^2)]^2$), U_0 the incident electron energy, mc^2 the rest energy of electron, M_{W} the molar mass, ρ the mass density, and N_{A} Avogadro's number. The dependency of the mean free path on the incident electron energy is given by (4.8) and plotted in Fig. 4.1. Experimentally determined values of the inelastic mean free path reported in the literature vary noticeably [164]. The reasons for these apparent discrepancies are not always clear. Some of the reported values for a couple of energies are included in Fig. 4.1. The fractional composition of a protein was taken to be 0.492, 0.313, 0.094, and 0.101 for elements H, C, N, and O, respectively [159, 165]. We used the values of Λ_{in} for vitreous ice at 80 kV and protein at 100 kV provided by [158] and [159], respectively. The values for any other incident energy of electron U_0 were calculated *via* equation (4.8).

Amorphousness of the solvent - specimen

As described in sections 4.3.1 and 4.3.1, the solvent has been modeled as a continuum. Although its influence on the potential redistribution of the macromolecule is accounted for (section 4.3.1), the solvent potential is calculated from the known density of water molecules using an averaging procedure (see equation (F.20)). However, amorphousness of the solvent can influence the appearance of the noise in the image. In order to assess the influence of the amorphousness in cryo-EM under low-flux conditions or to allow one to model it for high fluxes in HREM, we propose two methods for modeling this amorphousness: (i) adding a fixed noise pattern to the specimen's projected potential, and (ii) performing molecular dynamics (MD) simulations *via* GROMACS [166].

(i) Adding a fixed noise pattern to the projections:

This simple method assumes that the overlap of atomic positions in a projection of an amorphous sample is significant and that it is essentially noise with a flat frequency spectrum. This is surely an approximation as every real specimen has limited scattering power. Therefore, we multiply the frequency spectrum by $\exp(-2\pi(qr_d/\Delta_{x,y})^2)$ where q is the spatial frequency, r_d is the average minimum distance between atoms in the amorphous specimen and $\Delta_{x,y}$ is the pixel size in the object plane. The covalent sp^3 radius in carbon is 0.77 Å [167], and a model of amorphous carbon should thus have a minimum distance of $r_d = 1.54$ Å. For vitreous ice, the distance between oxygen atoms would be 2.88 Å [168].

(ii) MD simulations:

In order to produce an explicit description of the solvent (water), we used GROMACS [166], a MD simulation package which solves Newton's equations of motion for a system of N interacting atoms. The equations are solved simultaneously in small time steps reaching an equilibrium state of the system. The input was PDB file 1GR5 (GroEL). The missing hydrogens atoms were added and a topology file was generated containing the physical information about all

interactions between the atoms of the protein (bonds, angles, torsion angles, Lennard-Jones interactions and electrostatic interactions). Furthermore, the protein was solvated in a 20 x 20 x 50 nm water box with a simple point charge (SPC) 216 model. The specimen box consists of a multitude of small boxes, each containing 216 water molecules. In order to circumvent a crystalline arrangement of small water boxes, energy minimization was performed followed by a short MD simulation (20 ps), effectively randomizing the solvent molecules positions and solvating the protein.

Beam-induced movements

Beam-induced movements can significantly influence the contrast in cryo-EM [26, 169]. The whole layer of ice encapsulating the macromolecule seems to deform upon exposure in a complicated manner. Here, we model these effects empirically *via* an isotropic motion factor σ_M , which blurs the IP as follows:

$$\tilde{V}(q) = \tilde{V}^{int}(q) \exp(-2\pi^2 \sigma_M^2 q^2), \quad (4.9)$$

where $\tilde{V}(q)$ and $\tilde{V}^{int}(q)$ are the Fourier transforms of the potential $V(\mathbf{r})$ and $V^{int}(\mathbf{r})$, respectively. This is equivalent to damping of the electron scattering factors in the Fourier domain.

4.3.2 Electron wave propagation and intensity detection

The incident electron plane wave is propagated through the specimen by a multislice approach inspired by [144]. The slice thickness was kept constant at ~ 2 nm. The effective projected potential within this slice thickness in all our simulations $\sigma V_z(\mathbf{r})$ proved to be smaller than 0.36 suggesting that, within a slice, the probability of multiple scattering events is less than 5 % and that the weak-phase object approximation and projection assumption are valid [52]. As described in section 4.2, the CTF accounts for all relevant microscope aberrations, apertures and partial coherence of the electron source. Finally, the image intensity is captured by the detector modeled by the MTF, DQE and various noise sources.

4.3.3 Parameter estimations

Some imaging parameters vary between acquisitions, while others are stable for a long period of time. To accurately model image formation and validate it with experimental data, we need to know the numerical values of all parameters that influence image formation (see section 4.2). The detector parameters are characterized independently *via* methods described in [47]. The parameters that must be determined during the data acquisition are magnification, integrated flux, defocus, astigmatism and local ice thickness. The magnification of the microscope was calibrated prior to the acquisitions with a cross grating containing 2160 lines per mm. The integrated electron flux in [$e^-/\text{\AA}$] was estimated from the measured intensities in areas without specimen using the conversion factor of the detector. For each low-flux cryo image, an image of an adjacent carbon support was acquired to accurately measure defocus and astigmatism as well as their uncertainties using the publicly available toolbox described in [49]. Measurements

of the local ice thickness d are based on the Beer-Lambert law and were calculated from the ratio of the integrated intensity of an EELS zero-loss peak I_{z1} relative to the integral of the whole spectrum I . Similar to equation (F.64) in Appendix F we have

$$d = \Lambda_{\text{in}} \ln \frac{I}{I_{z1}}. \quad (4.10)$$

4.4 Experimental methods

In order to validate our image formation model, cryo-EM experiments were carried out using various test samples and experimental conditions. Numerous defocus and flux series of unfiltered and zero-loss energy filtered images of 20S proteasome and hemoglobin were acquired at 80 kV and 300 kV.

4.4.1 Sample preparation

Our modeling approach was evaluated with 20S proteasome from *S. cerevisiae*, *Lumbricus terrestris erythrocrucorin* (earthworm hemoglobin) and GroEL. Proteasome (Sigma Aldrich, 10 mg/ml) was diluted tenfold in 50 mM HEPES pH 7.4, 150 mM NaCl, and 1 mM DTT. The hemoglobin sample (Hb) was prepared by diluting the hemoglobin stock solution 25-fold in 50 mM NH₄Ac pH 6.6 (a protocol adapted from [50, 170]). The GroEL chaperonin (Sigma Aldrich, 5 mg/ml) was diluted fivefold in 200 mM MOPS pH 7.4, 100 mM KCl, and 4 mM MgCl₂. Diluted (1:10) protein A (a bacterial surface protein commonly used because of its ability to bind immunoglobins) conjugated with 5 nm colloidal gold particles (CMC-UMC, Utrecht, the Netherlands) was added ($\sim 3 \mu\text{l}$) as fiducial markers to the samples just before EM grid preparation. Aliquots of $3 \mu\text{l}$ samples at $\sim 1 \text{ mg/ml}$ protein concentration were applied to 200 mesh thick C-flat grids (Protochips Inc., NC, USA) (1.2 mm hole size). All grids were freshly glow discharged for 30 s with a current of 20 mA. Excessive liquid was blotted at room temperature from one side inside a Leica EM GP freezing plunger using 3 s blotting time and 2 s postblotting time with 95 % relative humidity. Subsequently, the blotted grid was plunged into liquid ethane for vitrification. The grids were stored in liquid nitrogen pending examination in the electron microscope. In addition to the low-salt hemoglobin sample described above, two more ion concentrations were tested, 1 M and 3 M NH₄Ac, respectively.

4.4.2 Image acquisitions/data collection

Frozen-hydrated specimens were examined with a Titan Krios electron microscope (FEI Company, The Netherlands), equipped with a field emission gun (FEG) operated at acceleration voltages of 80 and 300 kV. A post-column GIF energy filter (Gatan, USA) equipped with 2k x 2k Gatan US1000 camera was used. The energy slit was adjusted to select only electrons with an energy loss less than 5 eV. Other microscope settings were: condenser aperture number 3 (size of 100 μm), objective aperture 4 (100 μm), spot size index 5, and beam diameter of 2 μm . The spherical (C_s) and chromatic (C_c) aberrations for this Titan microscope are both 2.7 mm, while

the energy spread (ΔE) and illumination aperture (α_i) are 0.7 eV and 0.03 mrad, respectively. The grids were mounted using the Krios autoloader. A cross-grating was used for magnification calibration. Images of proteasome, hemoglobin and GroEL at 80 kV and hemoglobin at 300 kV were recorded on a 2k x 2k Gatan CCD (US1000) camera with a magnification at the detector plane of 44.5 kx. The pixel size of the detector is 14 μm and the final sampling density in the object plane was 3.15 $\text{\AA}/\text{pixel}$. The requested underfocus ranged from 500 nm to 4000 nm in five steps. The incident flux was derived from the detector analog-to-digital units (ADUs) by taking 1 s exposures without sample and using a conversion factor (in ADU/e⁻) as calibrated by [47]. Each defocus series was collected from a previously unexposed sample suspended across one of the holes in the C-flat grid. Electron fluxes of $\sim 2.5 e^- \text{\AA}^{-2} \text{s}^{-1}$ and $\sim 5.5 e^- \text{\AA}^{-2} \text{s}^{-1}$ at respectively 80 keV and 300 keV, were used to record each single frame, while the exposure times used were 0.5, 1, and 2 s. Images in a defocus series of the same view were taken with and without energy filtering. After each defocus series an image of the adjacent carbon support was acquired using image shift in order to accurately measure defocus and astigmatism on that area [49]. These values are then also used for the region of interest.

4.5 Results

The validation of our image formation model is based on a systematic comparison between simulated and experimental images under various experimental conditions. We present the influence of the solvent including ion concentration, defocus, integrated electron flux, motion factor, amorphousness of the specimen, ice thickness, MTF and DQE of the camera, and incident electron energy on the image formation of samples embedded in vitreous ice (proteasome and hemoglobin). For an unbiased comparison between experimental and simulated images, the display for each image was stretched between mean value plus/minus 2.2 standard deviations of the corresponding experimental image. Estimated ice thickness d and defocus Δf are specified accordingly.

4.5.1 “Bond” contributions

As described in section 4.2, $\mathcal{V}_{\text{atom}}$ is modeled using the isolated atom superposition approximation (IASA), while the redistribution potential as a result of the bond contributions $\mathcal{V}_{\text{bond}}$ is modeled by a Poisson-Boltzmann (PB) method. The ratio between the mean squared “bond” and “atom” potential contributions $R_{\text{bond}} = \langle \mathcal{V}_{\text{bond},0}^2 \rangle / \langle \mathcal{V}_{\text{atom},0}^2 \rangle$ was calculated for each of the simulated interaction potentials \mathcal{V}^{int} (equation (4.1)). $\mathcal{V}_{\text{bond},0}$ and $\mathcal{V}_{\text{atom},0}$ represent mean-subtracted $\mathcal{V}_{\text{bond}}$ and $\mathcal{V}_{\text{atom}}$ potentials, respectively. The values of R_{bond} for proteasome, hemoglobin in 50 mM, 1 M and 3 M NH_4Ac are 5.3 %, 9.5 %, 7.9 %, and 7.7 %, respectively. Fig. 4.2 permits comparisons between (1) experimental images, (2) simulated images which potential is calculated using only $\mathcal{V}_{\text{atom}}$, and (3) using combined potential $\mathcal{V}_{\text{atom}} + \mathcal{V}_{\text{bond}}$. For the experimental conditions used here, the $\mathcal{V}_{\text{bond}}$ contribution to \mathcal{V}^{int} is not significant. In general, the combined potential produces weaker ringing effects on the surface of the molecule and lower peaks inside the proteins (Fig. 4.2). The SNR in the experimental images was not high enough to notice ap-

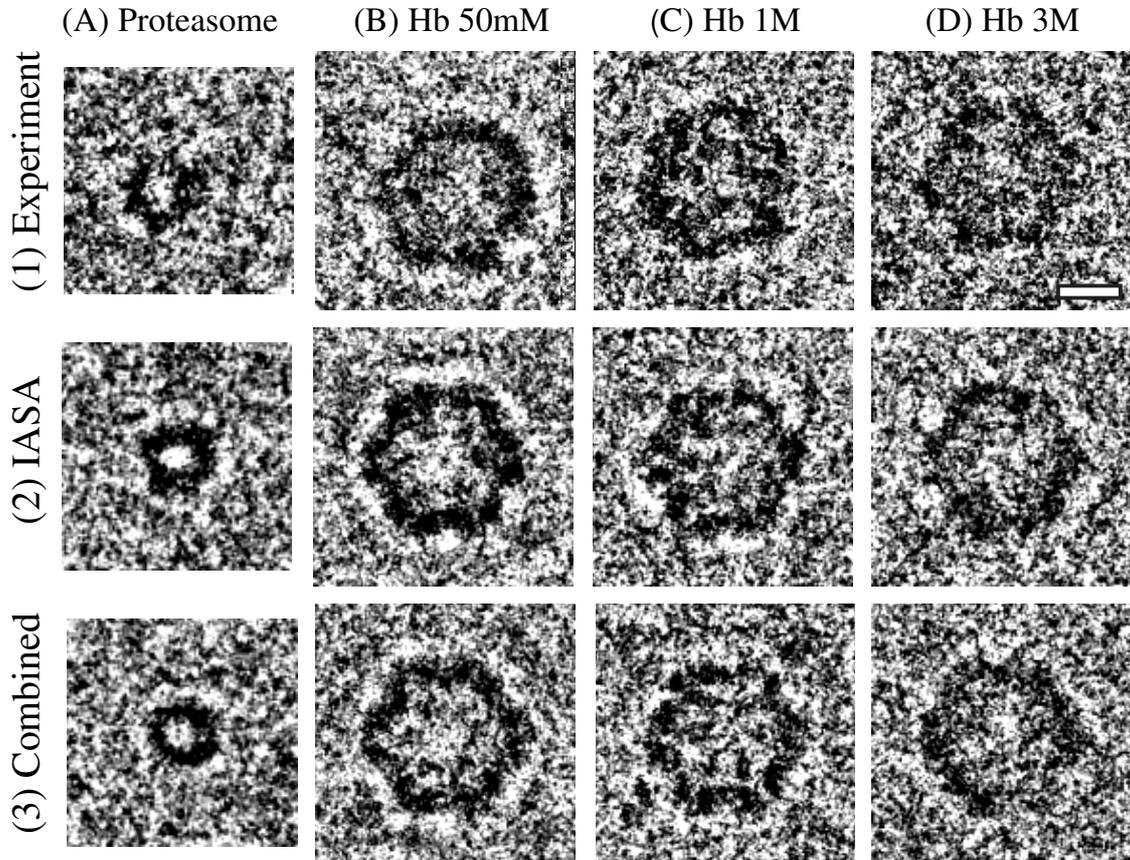


Fig. 4.2. Examples of (1) experimental images, (2) simulated images where the interaction potential (IP) was constructed from only $\mathcal{V}_{\text{atom}}$, and (3) simulated images with the IP calculated as combined potential $\mathcal{V}_{\text{atom}} + \mathcal{V}_{\text{bond}}$. The flux was $\sim 2.5 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ at 80 kV. From left to right are examples of (A) proteasome ($t_{\text{exp}}=2$ s, $\Delta f = 2509$ nm, $d = 69$ nm), (B) hemoglobin (Hb) in 50 mM ($t_{\text{exp}}=2$ s, $\Delta f = 4621$ nm, $d = 82$ nm), (C) Hb in 1M ($t_{\text{exp}}=2$ s, $\Delta f = 4505$ nm, $d = 196$ nm), and (D) Hb in 3M NH_4Ac ($t_{\text{exp}}=1$ s, $\Delta f = 2754$ nm, $d = 169$ nm). The scale bar corresponds to 10 nm.

parent differences due to the redistribution potential $\mathcal{V}_{\text{bond}}$ within the molecule. We performed simulations with various integrated fluxes, magnifications, defoci and acceleration voltages to assess when it is needed to include $\mathcal{V}_{\text{bond}}$ in the modeling. Fig. 4.3 compares images from $\mathcal{V}_{\text{atom}}$ and combined potential for some of the parameters. Figs. 4.3B and 4.3D suggest that the dark hexagon produced by large defocusing ($\Delta f = 6 \mu\text{m}$) is weaker when using the combined potential. The differences inside the protein are more pronounced at higher magnification (Figs. 4.3A and 4.3C), and at 300 kV (Figs. 4.3C and 4.3D), producing stronger signal for $\mathcal{V}_{\text{atom}}$ than for the combined potential. In general, assuming no beam-induced motion, a higher integrated flux better reveals minute differences inside the molecule. In Figs. 4.3B and 4.3D we used an integrated flux of $10 e^{-/\text{\AA}^2}$, which is four times higher than in the actual experiments (at the same magnification). A corresponding SNR (assuming perfect alignment and no beam-induced motion) would be achieved experimentally by averaging 16 equivalent particles.

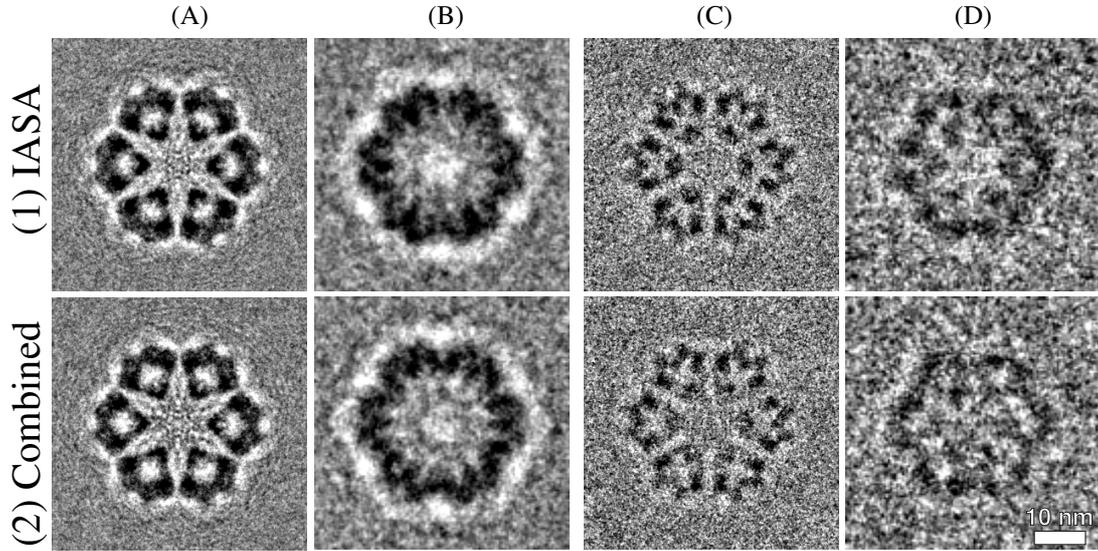


Fig. 4.3. Examples of (1) simulated images where interaction potential (IP) was constructed from only $\mathcal{V}_{\text{atom}}$, and (2) with the IP calculated as combined potential $\mathcal{V}_{\text{atom}} + \mathcal{V}_{\text{bond}}$. **(A)** Voltage 80 kV, Magnification 100 kx, $\Delta f = 2000$ nm, and integrated flux $100 e^-/\text{\AA}^2$; **(B)** Voltage 80 kV, Magnification 42 kx, $\Delta f = 6000$ nm, and integrated flux $10 e^-/\text{\AA}^2$; **(C)** and **(D)** are similar to **(A)** and **(B)**, respectively but at a voltage of 300 kV.

4.5.2 Defocus series

Various defocus series were acquired with a requested defocus ranging from 500 nm to 4000 nm. From the adjacent carbon area next to each region of interest the defocus values Δf are estimated and provided in the figures captions. For readability we omit to display the astigmatism values as well as uncertainties of the defocus estimations as provided by tools described in [49]. The astigmatism was always smaller than 6 % of the defocus value. The uncertainties of defocus estimation were on average 1.6 %.

Fig. 4.4 shows experimental and simulated defocus series of 20S proteasome, top and side view at 80 kV for 0.5 and 1 s exposure time, respectively. The simulations correctly predict the changes in the image when the defocus value is altered. For small defocus values the contrast at low frequencies is too small to be distinguished from the noise. However, at larger amounts of underfocus the white fringes and the central channel in the top view (second and third column) are readily recognized and they appear comparable in both experimental and simulated images. The experimental images at higher defocus values provide less details as is predicted by the simulations (the forth and the fifth column (side view) in Fig. 4.4).

4.5.3 Integrated flux series and motion factor

Subsequently, we tested whether the simulations can predict the effect of different integrated fluxes. After each defocus series, another region of interest was selected and imaged with a different integrated electron flux. The flux was kept constant ($\sim 2.5 e^- \text{\AA}^{-1} \text{s}^{-1}$ at 80 kV), while exposure times were set to 0.5, 1, and 2 s, producing an integrated flux per single frame of \sim

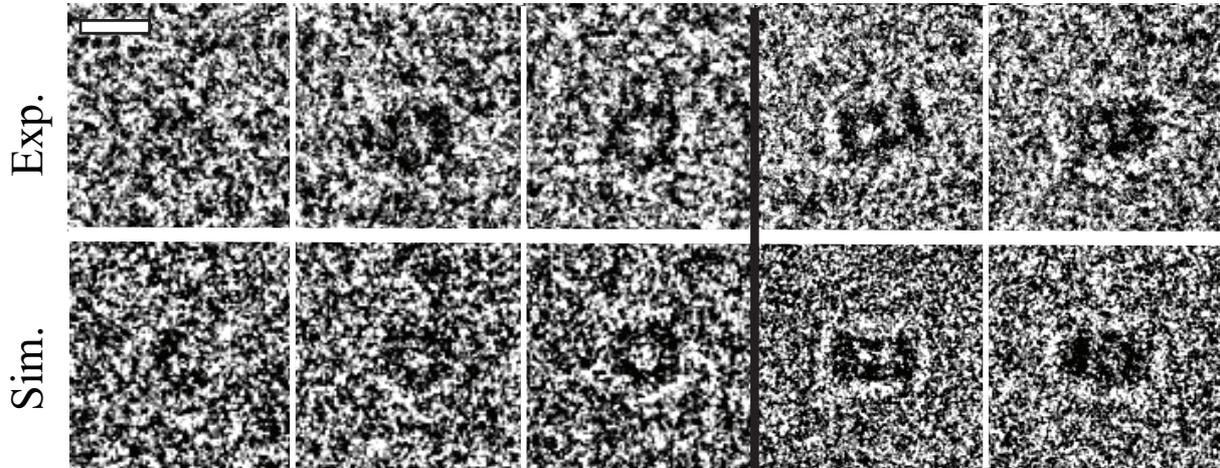


Fig. 4.4. Experimental and simulated defocus series of proteasome 20S at 80 kV and at a flux of $\sim 2.5 e^{-}\text{\AA}^{-2}\text{s}^{-1}$. First 3 columns (top view): $t_{\text{exp}} = 0.5$ s and defoci from left to right 0.75, 1.3, 1.9 μm , respectively. Last 2 columns (side view): $t_{\text{exp}} = 1$ s and defoci 4.4, and 6.7 μm , respectively. The scale bar corresponds to 10 nm.

1.25 $e^{-}/\text{\AA}$, $\sim 2.5 e^{-}/\text{\AA}$, and $\sim 5 e^{-}/\text{\AA}$, respectively. Fig. 4.5 shows experimental and simulated integrated flux series of 20S proteasome top view (three parts subdivided in quadrants). We expect, based on the experiments shown in [26], that the beam-induced motion depends on the integrated flux. The effective motion factors ranged from 4 \AA to 10 \AA . Modeling smaller motion factors is not needed given our sampling density of 3.15 $\text{\AA}/\text{pixel}$. It can be seen that in the absence of motion factor modeling ($\sigma_{\text{mot}} = 0$ \AA) the simulated images at higher integrated fluxes display a higher contrast and appear sharper than the experimental data. Incorporating a motion factor of $\sigma_{\text{mot}} \sim 4 - 8$ \AA and $\sigma_{\text{mot}} \sim 6 - 10$ \AA at $\sim 2.5 e^{-}/\text{\AA}$ and $\sim 5 e^{-}/\text{\AA}$, respectively let the simulations be in good agreement with the experiments. Note that the particles were selected from different areas of the specimen, so they differ slightly in defocus and specimen thickness.

4.5.4 Inelastic contributions

Fig. 4.6A shows simulations where only pure phase contrast is considered for the image formation and electron-specimen interaction. When inelastic events are considered (Fig. 4.6B), the vitreous ice will damp the amplitude of the propagating wave exponentially with increasing ice thickness. However, the difference between inelastic scattering properties of the protein and that of the vitreous ice (see Fig. 4.1) produces amplitude contrast. Since the inelastics are modeled as the imaginary part of the interaction potential, they are assumed to be removed from the image. Therefore, the simulated images must be compared with zero-loss energy filtered experimental images (Fig. 4.6C). The latter excludes most of the electrons with plasmon energy-losses. The objective aperture was large (100 μm) allowing us to assume that all elastically scattered electrons reached the detector. Fig. 4.6D shows unfiltered experimental images where both elastics and inelastics were detected, contributing to a stronger signal. However, the

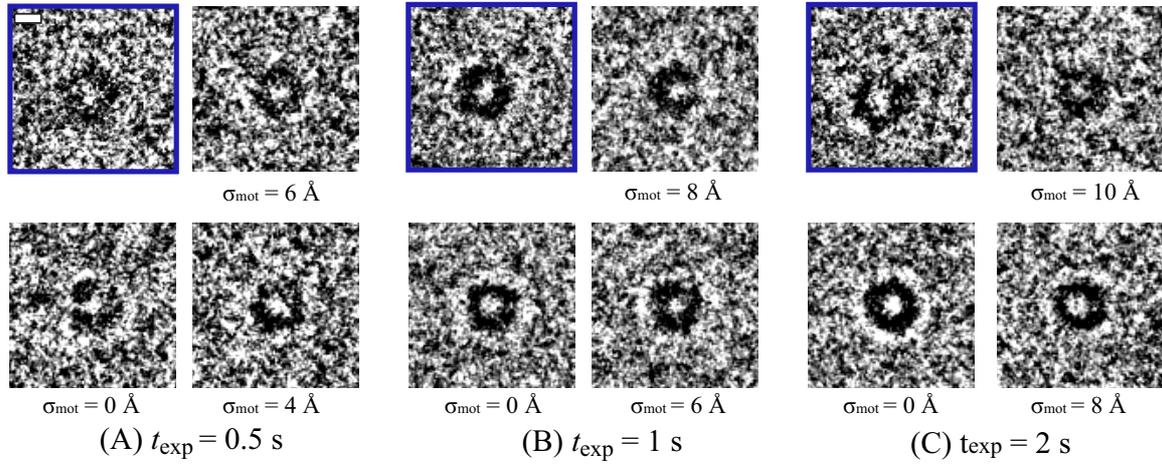


Fig. 4.5. Integrated flux series and varying motion factors σ_{mot} at 80 kV. The flux was $\sim 2.5 e^{-\text{\AA}^{-2}\text{s}^{-1}}$. The experimental images (upper left quadrants) are framed. The simulated images with increasing motion factor are presented in anticlockwise direction. A higher integrated flux requires a larger motion factor. (A) $t_{\text{exp}}=0.5$ s, $\Delta f = 2492$ nm, $d = 85$ nm (B) $t_{\text{exp}}=1$ s, $\Delta f = 4392$ nm, $d = 92$ nm, and (C) $t_{\text{exp}}=2$ s, $\Delta f = 2509$ nm, $d = 69$ nm. The scale bar corresponds to 5 nm.

images appear more blurry because the inelastics that reached the detector lost their coherency.

4.5.5 Camera's DQE

Fig. 4.7 illustrates the necessity of modeling the detector's DQE instead of the commonly used MTF-only approach. The left image (Fig. 4.7A) was simulated using equation (4.7), assuming that the signal and noise are transferred with the same MTF (DQE = 1). Fig. 4.7B shows a simulation which takes into account the DQE and the influence of the conversion factor on the image quality. The experimental image (Fig. 4.7C) is comparable to Fig. 4.7B, showing the importance of modeling the DQE.

4.5.6 Acceleration voltage influence

The low-frequency contrast in experimental and simulated images at 300 kV acceleration voltage is smaller than at 80 kV whereas the incident integrated flux was higher (see Figs. 4.8 and 4.2). This is in agreement with the energy dependent scattering properties of the incident electrons, interaction constant (see equation (4.4)), and the CTF. Additionally, the MTF and DQE of the CCD camera decrease with increasing acceleration voltage contributing to a reduced low-frequency contrast [20]. However, these combined effects provide an apparent higher level of details in the images (see Fig. 4.3). At 300 kV the motion factor appears to be smaller (Fig. 4.8), (data not shown for $\sigma_{\text{mot}} > \text{\AA}$). In Fig. 4.8, it appears that the simulated images at 300 kV using only $\mathcal{V}_{\text{atom}}$ (2) provide a stronger signal compared to the experimental images (1) and to the images that use the combined potential $\mathcal{V}_{\text{atom}} + \mathcal{V}_{\text{bond}}$ (3) (see also Fig. 4.3).

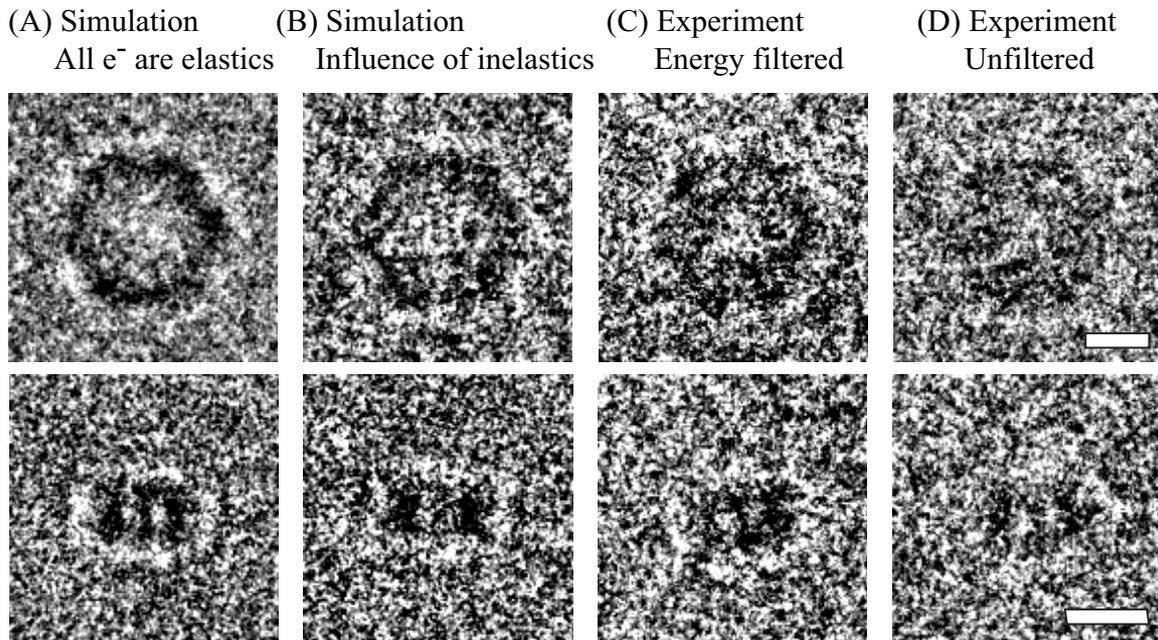


Fig. 4.6. Influence of inelastic scattering. **(A)** simulations of pure phase contrast, **(B)** simulations with inelastic scattering, **(C)** experimental zero-loss filtered images, and **(D)** experimental unfiltered images. From top to bottom are presented hemoglobin in 3M NH_4Ac ($t_{\text{exp}}=1$ s, $\Delta f = 4918$ nm, $d = 142$ nm, $\sigma_{\text{mot}} = 8\text{\AA}$), and side view of proteasome 20S ($t_{\text{exp}}=1$ s, $\Delta f = 6713$ nm, $d = 80$ nm, $\sigma_{\text{mot}} = 0\text{\AA}$). In order to use the same display stretching as in the other examples, the overall higher intensity in **(A)** was scaled with a thickness dependent constant $\exp(-d/\Lambda_{\text{in}})$, while in **(D)** we used the ratio between the median value of the filtered and unfiltered images. The scale bars correspond to 10 nm.

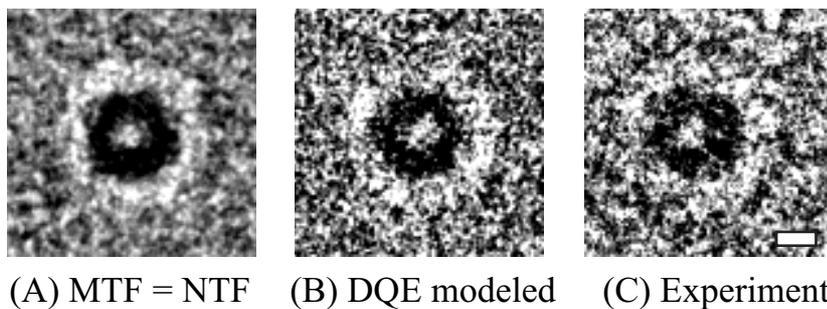
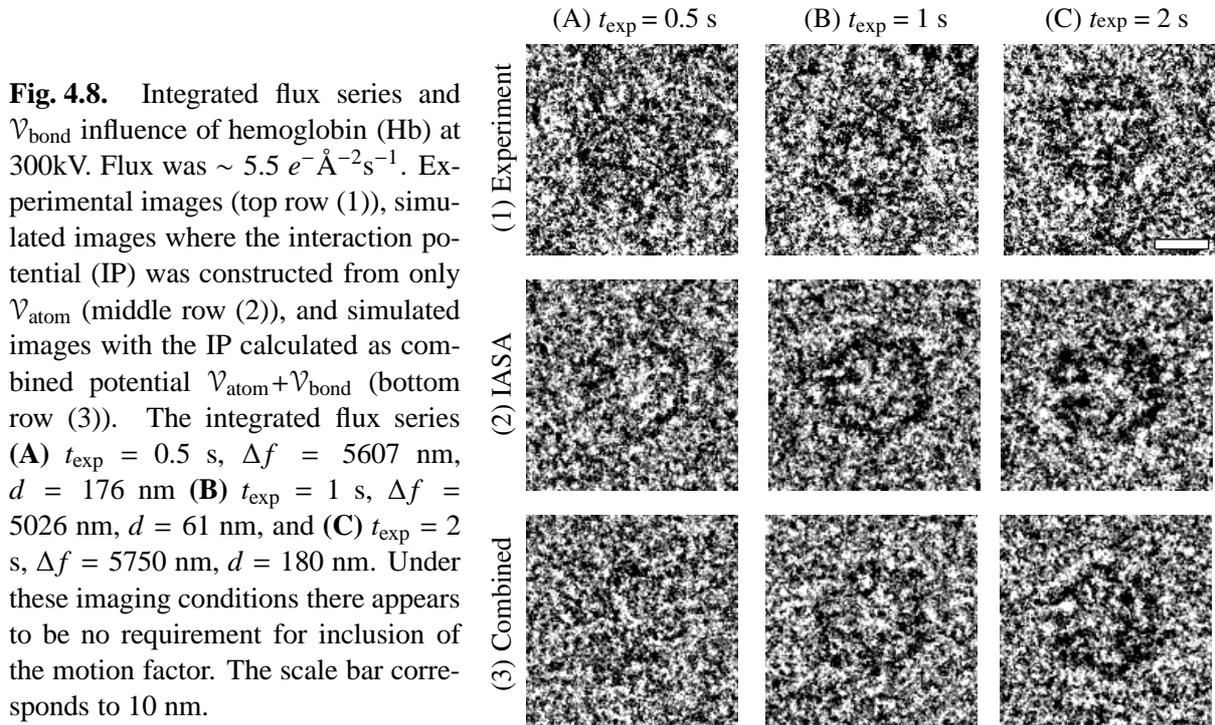


Fig. 4.7. Influence of the camera's DQE. **(A)** Simulated image assuming the same MTF for the signal and the noise. **(B)** Simulated image by taking into account the measured DQE. **(C)** Experimental image ($t_{\text{exp}}=1$ s, $d = 92$ nm, $\Delta f = 6713$ nm, and $\sigma_{\text{mot}} = 6\text{\AA}$). The scale bar corresponds to 5 nm.

4.5.7 Amorphousness of the solvent

Fig. 4.9 shows the influence of the amorphousness of the solvent on the image. The positions of the water molecules were simulated via MD (see section 4.3.1) in a $20 \times 20 \times 50$ nm box and the interaction potential was generated via IASA. A region 1 modeled with amorphousness is compared to a region 2 where the solvent is modeled as a constant potential and the noise



is only due to Poisson statistics. At the integrated flux used in experiments (Fig. 4.9A), the difference between those two regions is not noticeable. Simulated integrated flux series (Figs. 4.9C-E) suggests that only at high integrated fluxes ($> 100 e^{-/\text{\AA}^2}$), high magnification, and without beam-induced motion, the difference between Poisson noise and solvent amorphousness becomes apparent (Fig. 4.9E). At 300 kV, the differences are less pronounced, even at a high integrated flux (Fig. 4.9F).

4.6 Discussion

Here we highlight and discuss the unique aspects of our simulation model.

4.6.1 Forward model

A structure deposited in the PDB contains type and position of atoms in the molecule, although hydrogen atoms are usually lacking. Some of the programs that offer the functionality of adding hydrogen and other missing atoms are described in [157, 166, 171]. In our case, to calculate $\mathcal{V}_{\text{atom}}$, scattering factors for frequencies up to $q = 4 \text{\AA}^{-1}$ are parameterized as a weighted sum of five Gaussians and provided in Table 1 in [172]. Parameterizations of the scattering factors up to $q = 12 \text{\AA}^{-1}$ (provided by Table 3 in [172]) or using a combination of Gaussians and Lorentzians [144] would only be beneficial for very high scattering angles and/or heavy atomic elements. Biological specimens mainly consist of lighter elements such as H, C, O, and N, and the deviations of the parameterized curves in [172] for these elements, from the parameteriza-

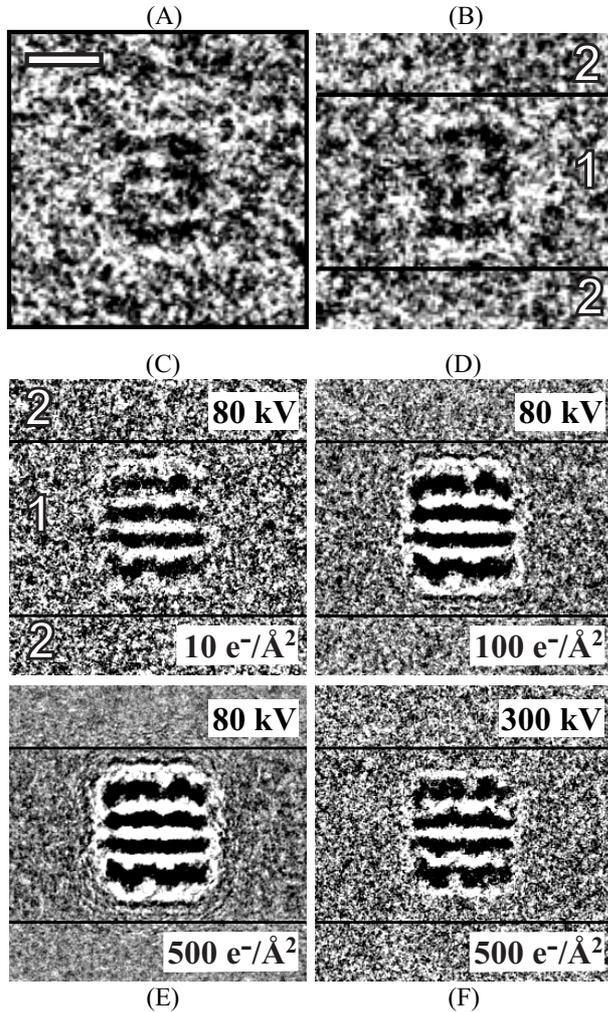


Fig. 4.9. Influence of the amorphousness of the solvent (region 1 between two horizontal lines) compared to Poisson noise only (region 2 below and above lines). **(A)** experiment and **(B)** simulation (voltage 80 kV, magnification 42 kx, integrated flux $\sim 5 e^-/\text{\AA}^2$, $\Delta f = 2718$ nm, $d = 120$ nm, $\sigma_{\text{mot}} = 6\text{\AA}$); **(C)-(F)** Amorphousness dependence in simulations with varying integrated flux and voltage (magnification 100 kx, $\Delta f = 1 \mu\text{m}$, $d = 20$ nm); for display purposes a percentile stretch was used (the lower and upper 1% of the gray values were clipped before stretching) **(C)** integrated flux $10 e^-/\text{\AA}^2$ at 80 kV; **(D)** integrated flux $100 e^-/\text{\AA}^2$ at 80 kV; **(E)** integrated flux $500 e^-/\text{\AA}^2$ at 80 kV; **(F)** integrated flux $500 e^-/\text{\AA}^2$ at 300 kV. The scale bar corresponds to 10 nm.

tions in [144], are less than 0.1 %. An advantage of using the parametrization as implemented here is that it avoids singularities at zero distances from the atomic nucleus. Here, calculations of $\mathcal{V}_{\text{atom}}$ are based on a slight modification of [39], in such way that low-pass filtering to a certain resolution does not exceedingly damp the IP, and the solvent is assumed to be vitreous ice instead of water (see equation (F.20) in Appendix F). Note that the difference between the inelastic mean free paths of vitreous ice and protein (Fig. 4.1) contributes to the amplitude contrast, but the plasmons of the vitreous ice attenuate the useful phase signal.

To describe electron wave propagation through a specimen with a finite thickness and to account for multiple scattering events, a multislice approach, inspired by [144] is used. Criteria for applicability of the weak-phase object approximation, projection assumption and multislice approach are presented in [52]. The criteria indicate that the projected potential map of hemoglobin sampled with a 3\AA pixel size does not, strictly speaking, satisfy the projection assumption, while the weak-phase object approximation holds. This implies that the thickness of the specimen cannot be neglected. Here, we simulated images of a single protein in a field of view smaller than 400×400 pixels for which the multislice approach took only a couple of seconds to compute. However, if one simulates a (tilt) series of e.g. $4k \times 4k$ images, several

hours of computational time would be required. If we assume not more than one (weak) elastic scattering event per incoming electron (first Born approximation), the free-space (Fresnel) propagation through a thick and/or tilted sample can be incorporated in the CTF [173–176]. We provide the possibility of including such a geometry in the CTF, which speeds up the forward computation [29] as well as the 3DCTF correction [30]. This approach assumes the weak-phase object approximation which is in our case satisfied but might not be fulfilled for thicker and/or tilted specimens and for higher resolution.

Performing MD simulations on a system consisting of both protein (GroEL) and solvent instead of doing it separately [142] should provide a more realistic modeling of the hydration shell of the protein [143]. We expect it should reduce the contrast between protein and environment, thereby further bridging the gap between simulations and experiments. Incorporating such a model might be subject for further studies. It has been reported recently [143] that such MD simulations can be used to derive a continuum model which describes the density of the water molecules surrounding a protein surface. Our current multislice algorithm does not require an explicit atomistic model of the solvent as in [142].

As TEM image formation usually involves only small angle scattering events, it is possible to ignore off-axis and higher order aberrations and only consider axial aberrations [10]. As our implementation is modular, there is a possibility of including higher order axial and non-axial aberrations in the future.

The insufficient SNR due to the low-flux imaging conditions and/or due to the beam-induced movements caused that we could not provide experimental evidence of amorphousness due to the solvent in our samples (compare Fig. 4.9). Consequently, the solvent can be modeled as a continuous medium, simplifying the simulations. Furthermore, by modeling the solvent as vitreous ice, 7 % less dense than water at room temperature [177], our predicted contrast would be slightly increased compared to [39, 142].

Noise in the images is mainly Poisson distributed, and strictly speaking signal-dependent. Here, we introduce a new way of modeling DQE which separates the signal and the noise transfer. An accurate description of the signal/noise transfer may facilitate the regularization in the reconstruction methods.

4.6.2 Poisson-Boltzmann (PB) approach

We characterize the influence of the solvent dielectric properties, ionic strength and electrostatic distribution within a molecule for TEM simulations and compare these with the isolated atom superposition approximation (IASA) where atoms are treated in isolation. The redistribution of the potential due to the interactions is modeled *via* a PB approach. The ratio R_{bond} between the mean squared “bond” (PB) and “atom” (IASA) potential contributions ranged from 6 to 10 %, suggesting that the $\mathcal{V}_{\text{atom}}$ contribution is the dominant part of the interaction potential. The mean value was subtracted from these potentials prior to calculating R_{bond} since the mean value does not influence the phase contrast [52]. Comparing simulated images where the IP was constructed only *via* $\mathcal{V}_{\text{atom}}$ with the ones where the IP was calculated by combining $\mathcal{V}_{\text{atom}}$ and $\mathcal{V}_{\text{bond}}$ did not show significant differences. In general, the images with the combined IP show weaker ringing effects around the protein edges (Fig. 4.2), better matching the corresponding

experimental images. The simulations suggested that the differences would be more pronounced for higher SNR (Fig. 4.3). Increasing the ion concentration in the hemoglobin solution resulted in a slight decrease of the mean $\mathcal{V}_{\text{bond}}$. A possible explanation is that the electrostatic shielding of the protein with a negative net charge produces smaller absolute values of $\mathcal{V}_{\text{bond}}$ for higher ion concentrations.

In material science, it has been reported [144] that errors up to 10% in calculation of electron scattering factors can occur due to the modeling of atoms as isotropic. The PB approach does not aim to accurately characterize the bonding between the individual atoms but focuses on the macroscopic influence of the solvent and ions on the potential distribution. The accuracy of PB approaches decreases in the region very close to the nucleus since the partial (net) charges are placed at the position of the nucleus (see F.1.4). The main benefit of the PB method lays therefore in the description of the potential redistribution due to the interaction with the solvent and its ions. Here, we have chosen APBS [156], a software package for numerically solving the PB equation based on finite elements. There are, however, other approaches, such as the boundary element solution [178], which may provide a faster and more accurate description of the potential at the protein boundaries. In this analysis we studied oligomeric macromolecules. The influence of the PB approach might be different for non-oligomeric macromolecules. Furthermore, the PB approach might facilitate the interpretation of transient states.

4.6.3 Beam-induced movements

Beam-induced specimen movements have long been recognized as one of the main factors attenuating the signal in cryo-EM [169, 179]. It has been suggested that the main causes of this local motion are specimen deformation and radiation damage during the exposure [23, 25–27], and/or charging [180, 181]. The inclusion of the motion factor blurs the simulated images to better match the experimental images. This effect is analogous to the damping envelope due to misalignment in SPA [182]. Our approach to include this damping effect is inspired by recent experiments of [26] who aimed to quantify the flux-dependent beam-induced movements.

In our analysis, the derived motion factors are similar to the displacement values reported in [26, 183]. Our observations are consistent with their suggestions that i) the motion is larger for higher fluxes, and ii) the motion rate decreases with exposure time showing that the motion is worst at the beginning of the exposure. In our case, the total dose that a specimen received could be up to 10x larger than the dose used to acquire individual images since we acquired numerous exposure series, e.g. at different defoci or with/without energy filter.

Henderson & Glaeser (1985) [169] suggested that some type of beam-induced movements of the specimens (around 5 Å or more) must occur in approximately equal amounts in all directions. Li *et al.* (2013) [183] found that this motion is not unidirectional, whereas Brilot *et al.* (2012) [26] reported directional preference of movements. The simulator accounts for isotropic motion and can be extended to model any particle trajectory during acquisition. However, if such a trajectory is known, it is preferred to correct for it by aligning and averaging the frames captured by a direct electron detector.

Our effective motion factor is smaller at 300 kV than at 80 kV. This could be related to numerous effects including differences in inelastic cross-sections, beam quality, or ice thickness.

Note that the integrated electron flux used for 300 kV was about two times larger than at 80 kV in order to have similar deposited energy per mass (dose). The ability to recognize amorphousness of the specimen/solvent decreases due to the beam-induced motion. Due to the large variation in the magnitude of the movements, some particles can have better contrast than others within the same field of view. The challenge remains how to avoid or automatically correct for beam-induced motion [26].

4.6.4 Validation

In previous work on accurate forward modeling, only the Tobacco Mosaic Virus (TMV) was characterized [142] and compared with experiments [39]. The advantage of TMV as a test sample is that the averaged 1D profile across the virus yields a high SNR and can be used for quantitative comparisons. Here, we analyzed three different biological specimens (20S proteasome, hemoglobin, and GroEL) in different embedding materials. Each simulation parameter relates directly to a physical quantity, but a strict quantitative comparison to cryo-EM images is difficult due to the high level of noise and challenging alignment, even for 1D-averaged profiles. For unbiased signal comparison, the display of each image was stretched to match the corresponding experimental image. For visual comparisons we simulated ten different noise realizations (data not shown), confirming that the noise did not change the appearance of the features. A comprehensive quantitative comparison in cryo-EM is mainly compromised by the low SNR. For completeness, it should be mentioned that also in material science, although dealing with much higher SNR, validation of simulations is usually done only visually [144, 184], even though there are attempts of using more quantitative approaches [185–188]. In section F.6, we present simulated and experimental images of carbon edges and carbon nanotubes and their 1D-averaged profiles for a more quantitative comparison. Advantages of using carbon edges and nanotubes for validation include the simplicity of their model and radiation hardness compared to cryo-EM.

Most simulation parameters described in [39] are based on physical principles. They need to employ, however, a calibration protocol for some parameters that are phenomenologically introduced, requiring their tuning. Examples of such parameters are amorphousness, absorption potential, as well as camera parameters such as the MTF, DQE, and conversion factor. The ice thickness in [39] was estimated from one spot although the thickness can vary significantly throughout the field of view. An advantage of the ice thickness measurements *via* an energy filter as described in this paper is that it is relatively fast experimentally and provides information about the local thickness. We assume that the energy filter was stable during acquisition as characterized in [189], without significantly compromising the accuracy of the thickness measurements. However, the experimentally determined values of mean free inelastic path used to estimate the thickness can vary noticeably [164]. Defocus values normally deviate from the values requested from the microscope. We estimate defocus and astigmatism on the adjacent carbon area which could, in principle, differ from the values at the region of interest due to the non-perpendicular pose of the sample relative to the beam. Although a model for the absorption potential was introduced in [39], simulated data were compared against unfiltered experimental images which also contain inelastically scattered electrons. However, any modeling based on

the imaginary part of the IP (analogously to Beer-Lambert law) assumes that those inelastic electrons are removed from the image, requiring a comparison with zero-loss energy filtered images.

Implemented in the C programming language, the user-friendly TEM-simulator [39] represents a good starting point for image simulations in cryo-EM. The novel aspects described and analyzed in this paper are included in InSilicoTEM, a simulator implemented in MATLAB .

4.6.5 Outlook

Accurate image simulations help to understand how the recorded image is formed, indicate ways to optimize data acquisition and microscope settings, and provide insight on ways to improve instrumentation. As an integral part of an accurate forward model, the estimation of parameters such as defocus and astigmatism [49] and camera's MTF and DQE [47] is essential and necessary for the CTF correction and/or regularization of the reconstruction approaches.

In addition to improving computational methods, the experimental developments should allow better transfer of the scattered electron wave onto the recorded image intensity. These experimental improvements are being achieved mostly by better sample preparation, by minimizing noise using direct electron detectors and electron counters, by improving the effective CTF *via* phase plates and more coherent electron sources, and by minimizing the effective beam-induced movements of the specimen.

The magnitude of the beam-induced movements must be reduced in order to increase the contrast in the images. Their effect can be somewhat decreased by lowering the flux, using a smaller carbon hole size, or by pre-irradiation. Postprocessing alignment and averaging the frames captured by a direct electron detector can further reduce blurring in the final images [26, 27, 183, 190, 191]. It is expected that dose fractionation and superresolution (beyond-Nyquist) EM using electron counting devices can reduce the influence of beam-induced movements, improving the achievable contrast in cryo-EM images. The modularity of InSilicoTEM allows integration of new physical parameters as well as modeling the influence of new hardware components such as the new generation of direct electron detectors.

The simulator could help to assess whether it is possible to resolve a specific macromolecule using a certain set of instrumental and processing parameters. It will be possible to easily and cost-effectively investigate the influence of new data acquisition techniques and advanced instrumentation, and to facilitate the development and evaluation of reconstruction and image processing techniques. In addition to the known PDB file, the input for InSilicoTEM simulator can also be a previously reconstructed 3D potential map of a sample. The simulator could furthermore facilitate the identification of molecular assemblies within the cell, a docking process where atomic models are fitted into cryo-EM obtained maps, and testing whether a proposed 3D model of a macromolecule is in agreement with the features observed in the micrographs. In electron tomography, iterative reconstruction schemes such as simultaneous iterative reconstruction technique attempt to minimize the difference between projections and simulated reprojections of the 3D map. From the differences between observed and simulated images one can often derive information to refine the model. The model parameters are iterated until simulated images best describe those observed. We expect that an accurate forward model

based on physical principles will facilitate such iterative scheme and reconstruction resulting in the 3D potential map.

4.7 Conclusions

We described an approach to simulate image formation in cryo-EM based on physical principles and taking into account the influence of the specimen and its surroundings, the optics of the microscope and the detector. Simulated and experimental images were generated under various settings and visually compared. Generated images adequately predict the effects of the phase contrast introduced by defocusing (Fig. 4.4), the changes due to the electron flux (Fig. 4.5), the influence of inelastic scattering (Fig. 4.6), camera DQE (Fig. 4.7), and the acceleration voltage (Fig. 4.8).

Various buffer compositions have been used to evaluate the influence of charge redistributions for the hemoglobin sample. The contribution of this redistribution to the interaction potential appears to be less than 10 % for all these cases and is mostly visible by the slightly less contrast at protein-solvent interfaces compared to the images calculated using only the IASA-based potential.

Increasing the integrated electron flux showed the benefit of introducing a motion factor which could be related to the beam-induced motions. For the 20S proteasome images taken at exposure times of 0.5 s, 1 s, and 2 s, the motion factors were in the range of $\sim 4 \text{ \AA}$, $\sim 6 \text{ \AA}$, $\sim 8 \text{ \AA}$, respectively (see Fig. 4.5). At 300 kV the motion factor appears to be smaller (Fig. 4.8). The varying contrast of the particles within a field of view can be explained by the apparently space-variant beam-induced movements.

For typical electron fluxes in cryo-EM ($< 100 e^-/\text{\AA}^2$), the influence of the amorphousness of the solvent can be neglected since Poisson noise is the dominant noise source in the image (see Fig. 4.9) and the solvent can be modeled as a continuum.

The theory and methods provided here represent the basis of an expert system that could optimize the data collection strategy and inexpensively and efficiently investigate the influence of the new hardware.

Acknowledgements

The authors would like to thank Paul Mooney (Gatan) for his help related to the MTF and DQE of the GIF camera, Cristina Avramut (LUMC) for assistance in preparation of hemoglobin, Chandra Bajaj and Alex Rand (ICES) for discussions related to the PB methods, Ronald Ligtnering (TU Delft) for help with pdb2pqr parsers, Lennard Voortman (TU Delft) for useful discussions, Willem van Dorp (RUG) for providing graphite flakes, Eric Bosch (FEI) for assistance in using Geant4, Erik Franken (FEI) for critical reading of the manuscript, and to acknowledge the use of cryo-EM facility at the Netherlands Center for Electron Nanoscopy (NeCEN) and help of Gert Oostergetel (RUG) and Roman Koning (LUMC).

Appendix F

Theory of image formation in cryo-EM

F.1 Interaction potential (IP)

Many imaging modalities are based on probing the object under investigation by a particle/wave. A central part in simulating such imaging modalities is to understand the interaction between the object and the particle/wave used for probing the object. In transmission electron microscopy (TEM), the particles are electrons and the object is the electrostatic potential of the specimen. Modeling the interaction potential of a specimen in TEM is a method for generating a model of the specimen interaction with high-energy electrons. We consider only incident electrons in the range of a few keV to a few MeV. If the acceleration voltage is lower than about 1 kV, the incident electrons are not distinguishable from the electrons of the specimen. The treatment of scattering then requires taking into account exchange effects [192] and virtual inelastic scattering effects [193]. Furthermore, if the acceleration voltage is very high, i.e. greater than 10 MV, Bremsstrahlung energy losses, relativistic effects and knock-on damage become significant [111].

F.1.1 One-body, stationary Schrödinger equation

In order to describe a closed system consisting of the incident electron and the specimen we use the one-body stationary Schrödinger equation for the electron wave function

$$\left[-\frac{\hbar^2}{2m} \nabla_r^2 + e\mathcal{V}^{\text{int}}(\mathbf{r}) \right] (\psi_e)(\mathbf{r}) = E_e \psi_e(\mathbf{r}), \quad (\text{F.1})$$

where the operator $-\frac{\hbar^2}{2m} \nabla_r^2$ is the Hamiltonian of the incident high-energy electron, which in this case represents its *kinetic energy*, $\mathcal{V}^{\text{int}}(\mathbf{r})$ the interaction potential, \hbar the reduced Planck constant, m the relativistic mass of the electron, e the electron charge, $\mathbf{r} = (x, y, z)$ the position, ψ_e the wave function, and E_e the energy of the incident electron. Equation (F.1) is valid if the following assumptions are fulfilled:

- (a) magnetic field is approximately constant on the scale of the specimen thickness

- (b) successive incident electrons can be treated as independent events
- (c) the relativistic effects can be approximated by the non-relativistic Schrödinger equation (neglecting electron spin) with the relativistic wavelength and mass of the electron.
- (d) the phenomenon is time independent, i.e. the specimen does not change during transition of the electron.
- (e) assuming high-energy incident electrons, it is possible to reduce the multi-body to the one-body Schrödinger equation and to model the scattering properties of the specimen by the interaction potential.

The validity of these assumptions is discussed in more detail in *Intermezzo 1*.

Additionally, using the Born-Oppenheimer approximation which allows the wave function of the specimen to be separated into its electronic and nuclear components, and assuming that nuclei are represented as point-masses, we can express the interaction potential as

$$\mathcal{V}^{\text{int}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left[\int_{\mathbb{R}^3} \frac{\rho_e(\mathbf{y})}{|\mathbf{r} - \mathbf{y}|} d\mathbf{y} - \sum_{j=1}^m \frac{eZ_j}{|\mathbf{r} - \mathbf{R}_j|} \right], \quad (\text{F.2})$$

where ρ_e is the electron density, \mathbf{R}_j and Z_j are the position of the nuclei and the atomic number of the j -th atom, respectively, and ϵ_0 denotes the permittivity of the vacuum. The Fourier spectrum of an atom in equation (F.2) can be measured by means of electron diffraction experiments [10]. The expression for the interaction potential of the whole specimen is still computationally unfeasible due to the appearance of the electron density function ρ_e which cannot be computed *ab initio* for all electrons of a macromolecular complex.

Alternatively, the interaction potential can be expressed as

$$\mathcal{V}^{\text{int}}(\mathbf{r}) = \mathcal{V}_{\text{atom}}(\mathbf{r}) + \mathcal{V}_{\text{bond}}(\mathbf{r}) \quad (\text{F.3})$$

where the first term $\mathcal{V}_{\text{atom}}$ represents the potential one would get by considering the specimen as an ensemble of non-interacting atoms (the atomic contribution) and the second term $\mathcal{V}_{\text{bond}}$ represents changes in the charge density due to electrostatic interaction between the atoms in the specimen (interatomic charge distribution in the specimen) and due to the influence of the surrounding solvent and ions.

----- *Intermezzo 1* -----

Validity of the one-body, stationary Schrödinger equation

The validity of the above assumptions are discussed here in more detail.

(a) The electrons spiral in a strong magnetic field. They penetrate through the specimen which is usually embedded in the magnetic field of the objective lenses with a strength of ~ 1 T. The change of the electron path due to the magnetic field is on the scale of the focal length of the lens ≈ 1 mm, while the changes due to the electrostatic interactions occur on the scale of the specimen thickness ($< 1 \mu\text{m}$). Therefore, we can separate the influence of the magnetic and electric field while an electron passes through the specimen [144].

(b) Phase contrast in TEM is formed as a result of interaction between the unscattered part of the electron exit wave from the specimen and its (coherently) scattered part. Since electrons in TEM possess high kinetic energy, their speed is a large fraction of the speed of light c . For an incident electron with rest mass m_0 and an acceleration voltage of $U = 200$ kV the speed is

$$v = c \sqrt{1 - \left(\frac{m_0 c^2}{m_0 c^2 + eU} \right)^2} = 0.53c = 1.6 \cdot 10^8 \frac{\text{m}}{\text{s}}.$$

A typical current in the specimen is 10^{-8} A $\approx 10^{11} \frac{e^-}{\text{s}}$ [194] and successive electrons' centers are separated by a mean distance of $s = vt \approx 10^8 \frac{\text{m}}{\text{s}} \cdot 10^{11} \text{ s} = 10^{-3}$ m. However, electrons are not points and we need to consider electron wavepackets to calculate the distance between them. If the energy spread of electrons is 1 eV for 200 keV incident energy, an electron has a coherence length of $l_c = 2 \cdot 10^5 \lambda$ where λ is the wavelength of the electron wave which is $\sim 2.5 \cdot 10^{-12}$ m for $U = 200$ eV. The length of the electron packet is therefore $l_c \approx 5 \cdot 10^{-7}$ m. If d is the thickness of the typical TEM specimen $d < 1 \cdot 10^{-6}$ m then it follows $d \ll s - l_c$. It can be concluded that the interaction between the electron beam and the specimen is occurring one electron at the time. The mean separation between two successive electrons is much larger than the specimen thickness and the length of the electron packet. Therefore, the interference concepts refer to the interference between wave parts of an individual electron.

(c) The relativistic Dirac equation including spin would be the correct wave equation for relativistic electrons at high energies. A simplification using the Schrödinger equation where the wavelength λ and mass m are relativistically corrected, proved to be accurate enough for typical energy ranges in electron microscope [195, 196] with

$$m = m_0 \left(1 + \frac{eV}{m_0 c^2} \right), \lambda = \frac{hc}{\sqrt{eV(2m_0 c^2 + eV)}}, \quad (\text{F.4})$$

where h is the Planck's constant.

(d) A strict theory of electron diffraction deals with time-dependent processes [197]. The system in question consists of a specimen and an incident electron that scatters against this specimen. The specimen is assumed to be fully described by a specification of the position and type of its constituting atoms. More precisely, here, a specimen \mathcal{S} consisting of N atoms is specified by the atom positions \mathbf{R}_i^0 and associated charges Q_i ($\mathcal{S} := (\mathbf{R}_0, \mathbf{Q})$, $\mathbf{R}_0 := (\mathbf{R}_1^0, \dots, \mathbf{R}_N^0)$, $\mathbf{Q} := (Q_1, \dots, Q_N)$). The system (the incident high-energy electron and specimen) is described quantum mechanically by its wave function Ψ_{tot} . The time evolution of the system is given by:

$$i\hbar \frac{\partial \Psi_{\text{tot}}}{\partial t}(\mathcal{S}, \mathbf{r}, t) = \mathcal{H}^{\text{tot}}(\Psi_{\text{tot}})(\mathcal{S}, \mathbf{r}, t), \quad (\text{F.5})$$

where \mathbf{r} is the position of the incident electron, and \mathcal{H}^{tot} is the Hamiltonian of the system.

For purely elastic scattering process, where the specimen does not change its state, the Hamiltonian is time-independent [10] and the wave function Ψ_{tot} is time harmonic:

$$\Psi_{\text{tot}}(\mathbf{S}, \mathbf{r}, t) = \exp\left(-i\frac{E}{\hbar}t\right)\psi_{\text{tot}}(\mathbf{S}, \mathbf{r}), \quad (\text{F.6})$$

where E denotes the energy of the system. Inserting (F.6) into (F.5) gives us the stationary model of the system:

$$\mathcal{H}^{\text{tot}}(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) = E \psi_{\text{tot}}(\mathbf{S}, \mathbf{r}). \quad (\text{F.7})$$

Assuming that an incoming wavepacket has a well-defined energy (and hence momentum), and it is many wavelengths long, the problem is well approximated by solving the time-independent Schrödinger equation with an incoming plane wave, which simplifies the analysis. This is not strictly fulfilled for inelastic scattering, as it will be discussed in section F.2.3.

(e) The Hamiltonian of the system \mathcal{H}^{tot} is defined as

$$\mathcal{H}^{\text{tot}}(\Psi) := -\frac{\hbar^2}{2m}\nabla_{\mathbf{r}}^2\Psi + \mathcal{H}^{\text{int}}(\Psi) + \mathcal{H}^{\text{sp}}(\Psi) \quad (\text{F.8})$$

where the operator \mathcal{H}^{int} is the interaction Hamiltonian defined as

$$\mathcal{H}^{\text{int}}(\Psi)(\mathbf{S}, \mathbf{r}) := \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \frac{eQ_i}{|\mathbf{r} - \mathbf{R}_i^0|} \Psi(\mathbf{S}, \mathbf{r}); \quad (\text{F.9})$$

\mathcal{H}^{sp} is the Hamiltonian of the specimen given as the sum of the kinetic energy operators of all nuclei and electrons in the specimen and their interaction Hamiltonians. It is possible to separate contributions of the nuclei and electrons in the interaction Hamiltonian. So, assuming there are m atoms in the specimen and n electrons in total, we get

$$\mathcal{H}^{\text{int}}(\Psi)(\mathbf{S}, \mathbf{r}) := \frac{1}{4\pi\epsilon_0} \left(\sum_{k=1}^n \frac{e^2}{|\mathbf{r} - \mathbf{r}_k^0|} - \sum_{j=1}^m \frac{Z_j e^2}{|\mathbf{r} - \mathbf{R}_j|} \right) \Psi(\mathbf{S}, \mathbf{r}) \quad (\text{F.10})$$

where \mathbf{r}_k^0 denotes the position of the k -th electron, \mathbf{R}_j and Z_j refer to the position of the nuclei and the atomic number of the j -th atom. Henceforth, whenever convenient, we will switch between both formulas (F.9) and (F.10).

The energy of the system is equal to the sum of energy of the incident high-energy electron and the energy of the specimen. The equation (F.7) becomes

$$\mathcal{H}^{\text{tot}}(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) = (E_e + E_{\text{sp}}) \psi_{\text{tot}}(\mathbf{S}, \mathbf{r}) \quad (\text{F.11})$$

and describes a complex multi-body problem involving the incident electron and the electrons and nuclei of the specimen. It now turns out that this multi-body problem can be simplified and transformed into an one-body equation.

In TEM, the incident electron energy is very high compared with the interaction energies of the electrons in the specimen. The spatial component of the wave function for the system can be separated into a part related only to the incident electron and a part related only to the specimen, such that $\psi_{\text{tot}}(\mathbf{S}, \mathbf{r}) = \psi_{\text{sp}}(\mathbf{S}) \psi_e(\mathbf{r})$ where ψ_{tot} is given as in (F.6). Let us now define

$$\tilde{\mathcal{H}} := -\frac{\hbar^2}{2m} \nabla_{\mathbf{r}}^2 + \mathcal{H}^{\text{int}}, \quad \text{so } \mathcal{H}^{\text{tot}} = \tilde{\mathcal{H}} + \mathcal{H}^{\text{sp}} \quad (\text{F.12})$$

and

$$\begin{aligned} \text{LHS in (F.11)} &= (\tilde{\mathcal{H}} + \mathcal{H}^{\text{sp}})(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) = \tilde{\mathcal{H}}(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) + \mathcal{H}^{\text{sp}}(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) \\ &= \tilde{\mathcal{H}}(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) + \psi_e(\mathbf{r}) E_{\text{sp}} \psi_{\text{sp}}(\mathbf{S}). \end{aligned}$$

In the last equality we used

$$\mathcal{H}^{\text{sp}}(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) = \mathcal{H}^{\text{sp}}(\psi_e \psi_{\text{sp}})(\mathbf{S}, \mathbf{r}) = \psi_e(\mathbf{r}) E_{\text{sp}} \psi_{\text{sp}}(\mathbf{S}).$$

Furthermore, we have

$$\begin{aligned} \text{RHS in (F.11)} &= E_e \psi_{\text{tot}}(\mathbf{S}, \mathbf{r}) + E_{\text{sp}} \psi_{\text{tot}}(\mathbf{S}, \mathbf{r}) \\ &= E_e \psi_{\text{tot}}(\mathbf{S}, \mathbf{r}) + \psi_e(\mathbf{r}) E_{\text{sp}} \psi_{\text{sp}}(\mathbf{S}). \end{aligned}$$

Thus, equating the left and right hand sides of (F.11) results in

$$\tilde{\mathcal{H}}(\psi_{\text{tot}})(\mathbf{S}, \mathbf{r}) + \psi_e(\mathbf{r}) E_{\text{sp}} \psi_{\text{sp}}(\mathbf{S}) = E_e \psi_{\text{tot}}(\mathbf{S}, \mathbf{r}) + \psi_e(\mathbf{r}) E_{\text{sp}} \psi_{\text{sp}}(\mathbf{S})$$

which in turn simplifies to

$$\tilde{\mathcal{H}}(\psi_{\text{tot}}) = E_e \psi_{\text{tot}}. \quad (\text{F.13})$$

An important observation is that the Hamiltonian for the specimen does not appear in (F.13). The interaction between the incident high-energy electron and the specimen can be expressed without introducing the Hamiltonian of the specimen. This is a huge simplification as the latter is very difficult, if not impossible, to calculate numerically in the context of TEM image simulation.

Now, to get a one-body equation involving only the wave function of the electron, we multiply (F.13) by ψ_{sp}^* from the left and integrate over the $(\mathbf{R}, \mathbf{r}^0)$ -space. Then, for $\mathbf{S} = (\mathbf{R}, \mathbf{r}^0, \mathbf{Q})$ we get

$$-\frac{\hbar^2}{2m} \nabla_{\mathbf{r}}^2 \psi_e(\mathbf{r}) + \int_{\mathbb{R}^{3m} \times \mathbb{R}^{3n}} \psi_{\text{sp}}(\mathbf{S})^* \mathcal{H}^{\text{int}}(\psi_{\text{sp}} \psi_e)(\mathbf{S}, \mathbf{r}) \, d\mathbf{R} d\mathbf{r}^0 = E_e \psi_e(\mathbf{r}),$$

where we used the definition of $\tilde{\mathcal{H}}$ given in (F.12) and the fact that

$$\int_{\mathbb{R}^{3m} \times \mathbb{R}^{3n}} \psi_{\text{sp}}(\mathbf{S})^* \psi_{\text{sp}}(\mathbf{S}) \, d\mathbf{R} d\mathbf{r}^0 = 1 \quad \text{for } \mathbf{S} = (\mathbf{R}, \mathbf{r}^0, \mathbf{Q}).$$

Hence, if we define the *interaction potential* as

$$\mathcal{V}^{\text{int}}(\mathbf{x}) := \frac{1}{e} \int_{\mathbb{R}^{3m} \times \mathbb{R}^{3n}} \psi_{\text{sp}}(\mathbf{S})^* \mathcal{H}^{\text{int}}(\psi_{\text{sp}} \psi_e)(\mathbf{S}, \mathbf{r}) \, d\mathbf{R} d\mathbf{r}^0. \quad (\text{F.14})$$

Using this definition, we end up with the one-body stationary Schrödinger equation for the electron wave function (equation (F.1)):

$$\left[-\frac{\hbar^2}{2m} \nabla_{\mathbf{r}}^2 + e\mathcal{V}^{\text{int}}(\mathbf{r}) \right] (\psi_e)(\mathbf{r}) = E_e \psi_e(\mathbf{r}).$$

To summarize, given that the system consists of high-energy electron and specimen, modeling the electron-specimen interaction in a TEM by the multi-body Schrödinger equation in (F.7) simplifies to a one-body Schrödinger equation (F.1) for the incident high-energy electron. In this equation, all the properties of the specimen are encoded in the interaction potential \mathcal{V}^{int} .

The interaction potential in (F.14) is still computationally unfeasible due to the appearance of the specimen wave function ψ_{sp} , so the next task is to investigate that potential.

The interaction potential is computed given a specification of the specimen. By definition, we have

$$\mathcal{H}^{\text{int}}(\psi_{\text{sp}} \psi_e)(\mathbf{S}, \mathbf{r}) = \psi_e(\mathbf{r}) \frac{1}{4\pi\epsilon_0} \left[\sum_{j=1}^m \frac{-Z_j e^2}{|\mathbf{r} - \mathbf{R}_j|} + \sum_{k=1}^n \frac{e^2}{|\mathbf{r} - \mathbf{r}_k^0|} \right] \psi_{\text{sp}}(\mathbf{S})$$

so the the interaction potential can be written as

$$\begin{aligned} \mathcal{V}^{\text{int}}(\mathbf{r})(\psi_e) = & \frac{1}{4\pi\epsilon_0} \left[\int_{\mathbb{R}^{3m} \times \mathbb{R}^{3n}} \psi_{\text{sp}}(\mathbf{S})^* \sum_{k=1}^n \frac{e}{|\mathbf{r} - \mathbf{r}_k^0|} \psi_{\text{sp}}(\mathbf{S}) \, d\mathbf{R} d\mathbf{r}^0 \right. \\ & \left. - \int_{\mathbb{R}^{3m} \times \mathbb{R}^{3n}} \psi_{\text{sp}}(\mathbf{S})^* \sum_{j=1}^m \frac{Z_j e}{|\mathbf{r} - \mathbf{R}_j|} \psi_{\text{sp}}(\mathbf{S}) \, d\mathbf{R} d\mathbf{r}^0 \right] \psi_e(\mathbf{r}). \quad (\text{F.15}) \end{aligned}$$

We now make use of the Born-Oppenheimer approximation which allows the wave function of the specimen to be separated into its electronic and nuclear (vibrational, rotational) components

$$\psi_{\text{sp}}(\mathbf{S}) = \phi_{\text{sp}}^e(\mathbf{r}^0) \phi_{\text{sp}}^{\text{nucl}}(\mathbf{R}, \mathbf{Z}).$$

The interaction potential can be now expressed as

$$\begin{aligned} \mathcal{V}^{\text{int}}(\mathbf{r})(\psi_e) = & \frac{1}{4\pi\epsilon_0} \left[\int_{\mathbb{R}^{3n}} \phi_{\text{sp}}^e(\mathbf{r}^0)^* \sum_{k=1}^n \frac{e}{|\mathbf{r} - \mathbf{r}_k^0|} \phi_{\text{sp}}^e(\mathbf{r}^0) \, d\mathbf{r}^0 \right. \\ & \left. - \int_{\mathbb{R}^{3m}} \phi_{\text{sp}}^{\text{nucl}}(\mathbf{R}, \mathbf{Z})^* \sum_{j=1}^m \frac{Z_j e}{|\mathbf{r} - \mathbf{R}_j|} \phi_{\text{sp}}^{\text{nucl}}(\mathbf{R}, \mathbf{Z}) \, d\mathbf{R} \right] \psi_e(\mathbf{r}). \end{aligned}$$

Hence, if we define

$$\mathcal{V}_{\text{nucl}}^{\text{int}}(\mathbf{r}) := -\frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^{3m}} \phi_{\text{sp}}^{\text{nucl}}(\mathbf{R}, \mathbf{Z})^* \sum_{j=1}^m \frac{Z_j e}{|\mathbf{r} - \mathbf{R}_j|} \phi_{\text{sp}}^{\text{nucl}}(\mathbf{R}, \mathbf{Z}) \, d\mathbf{R}$$

$$\mathcal{V}_e^{\text{int}}(\mathbf{r}) := \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^{3n}} \phi_{\text{sp}}^e(\mathbf{r})^* \sum_{k=1}^n \frac{e}{|\mathbf{r} - \mathbf{r}_k|} \phi_{\text{sp}}^e(\mathbf{r}) \, d\mathbf{r},$$

then

$$\mathcal{V}^{\text{int}}(\mathbf{x})(\psi_e) = [\mathcal{V}_e^{\text{int}}(\mathbf{r}) + \mathcal{V}_{\text{nucl}}^{\text{int}}(\mathbf{r})] \psi_e(\mathbf{r}).$$

Now, consider the nuclei as point-masses located at points $\mathbf{R}_1, \dots, \mathbf{R}_m$, i.e. as Dirac distributions centered at those points, then

$$\mathcal{V}_{\text{nucl}}^{\text{int}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \delta_{\mathbf{R}}\left(-\frac{Z_1 e}{|\mathbf{r} - \mathbf{R}_1|}, \dots, -\frac{Z_m e}{|\mathbf{r} - \mathbf{R}_m|}\right) = -\frac{1}{4\pi\epsilon_0} \sum_{j=1}^m \frac{Z_j e}{|\mathbf{r} - \mathbf{R}_j|}. \quad (\text{F.16})$$

Next,

$$\begin{aligned} \mathcal{V}_e^{\text{int}}(\mathbf{r}) &= \frac{n}{4\pi\epsilon_0} \int_{\mathbb{R}^{3n}} |\phi_{\text{sp}}^e(\mathbf{y}, \mathbf{r}_2^0, \dots, \mathbf{r}_n^0)|^2 \frac{e}{|\mathbf{r} - \mathbf{y}|} \, d\mathbf{y} d\mathbf{r}_2^0 \dots d\mathbf{r}_n^0 \\ &= \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^3} \left[n \int_{\mathbb{R}^{3(n-1)}} |\phi_{\text{sp}}^e(\mathbf{y}, \mathbf{r}_2^0, \dots, \mathbf{r}_n^0)|^2 \frac{e}{|\mathbf{r} - \mathbf{y}|} \, d\mathbf{r}_2^0 \dots d\mathbf{r}_n^0 \right] d\mathbf{y} \\ &= \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^3} \frac{\rho_e(\mathbf{y})}{|\mathbf{r} - \mathbf{y}|} \, d\mathbf{y} \end{aligned} \quad (\text{F.17})$$

where ρ_e is the electron density function of the specimen

$$\rho_e(\mathbf{y}) := n \int_{\mathbb{R}^{3(n-1)}} e |\phi_{\text{sp}}^e(\mathbf{y}, \mathbf{r}_2^0, \dots, \mathbf{r}_n^0)|^2 \, d\mathbf{r}_2^0 \dots d\mathbf{r}_n^0.$$

and

$$\frac{1}{ne} \int_{\Omega} \rho_e(\mathbf{y}) \, d\mathbf{y} = \text{Probability of finding an electron in } \Omega \subset \mathbb{R}^3.$$

The study of ρ_e is a central topic in quantum chemistry and solid state physics since it is directly related to the electronic structure of a solid. A variety of approaches, such as Hartree–Fock and density functional theory (DFT), have been developed for efficiently calculating ρ_e for a given solid. State-of-the-art approaches in this field with a description of associated mathematical challenges are reviewed in [198]. These methods, however, are not applicable for non-periodic structures with many atoms.

To summarize, using the Born-Oppenheimer approximation and the assumption that nuclei are point-masses, we can express interaction potential as

$$\mathcal{V}^{\text{int}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left[\int_{\mathbb{R}^3} \frac{\rho_e(\mathbf{y})}{|\mathbf{r} - \mathbf{y}|} \, d\mathbf{y} - \sum_{j=1}^m \frac{eZ_j}{|\mathbf{r} - \mathbf{R}_j|} \right]. \quad (\text{F.18})$$

F.1.2 Isolated atom superposition approximation (IASA)

As a first approximation one can disregard the redistribution of charges due to solvent, ions and electrostatic forces within the molecule ($\mathcal{V}_{\text{bond}}$ in equation (F.3)). This brings us to the isolated atom superposition approximation (IASA). $\mathcal{V}_{\text{bond}}$ contribution for crystals and material science specimens is in comparison to the atomic contribution $\mathcal{V}_{\text{atom}}$ reported to be $\sim 5\%$ [10, p. 8 and 428]. For atoms with low Z -number the influence is 5–10% [144].

Since $\mathcal{V}_{\text{atom}}$ considers the specimen as a set of isolated atoms, we get

$$\mathcal{V}_{\text{atom}}(\mathbf{r}) = \sum_{j=1}^m \mathcal{V}_{Z_j}(\mathbf{r} - \mathbf{R}_j)$$

where \mathcal{V}_Z is the potential of an isolated atom centered at \mathbf{R}_j with atomic number Z , so

$$\mathcal{V}_Z(\mathbf{r} - \mathbf{R}_j) = \frac{1}{4\pi\epsilon_0} \left[\int_{\mathbb{R}^3} \frac{\rho_{e,Z}(\mathbf{y})}{|\mathbf{r} - \mathbf{y}|} d\mathbf{y} - \frac{eZ}{|\mathbf{r} - \mathbf{R}_j|} \right]$$

with $\rho_{e,Z}$ denoting the electron density function associated to the shell electrons of the isolated atom.

With the first Born approximation (see *Intermezzo 2*), such a potential can be written as the inverse Fourier transform of the electron scattering factor $f_Z^{(e)}$ of the atom [10, 39]:

$$\mathcal{V}^{\text{int}}(\mathbf{r}) = \frac{16\pi\hbar^2}{me} \int f_Z^{(e)}(\boldsymbol{\xi}) e^{4\pi i \boldsymbol{\xi} \cdot \mathbf{r}} d^3 \boldsymbol{\xi} \quad (\text{F.19})$$

where 2ξ relates to the spatial frequency, as often used in electron crystallography [10].

The electron scattering factors are normally calculated from experimental X-ray scattering factors using the Mott-Bethe formula. For an overview of various methods and parametrization see [144]. To calculate the IASA-based potential, we used the scattering factors for frequencies up to $q = 4 \text{ \AA}^{-1}$ parameterized as a weighted sum of 5 Gaussians and provided in Table 1 in [152].

The mean inner potential of an amorphous specimen with n molecules per unit volume, molar mass M_W and mass density ρ is estimated *via* zero-frequency electron scattering factors $f_Z^{(e)}(0)$ as

$$V_0 = \frac{2\pi\hbar^2}{me} n f_Z^{(e)}(0) = \frac{2\pi\hbar^2}{me} \frac{\rho N_A}{M_W} f_Z^{(e)}(0). \quad (\text{F.20})$$

For low-density amorphous ice ($\rho = 0.93 \text{ g/cm}^3$), this corresponds to $V_0 = 4.5301 \text{ V}$.

As atomic potentials calculated *via* IASA provide the most significant contribution to the scattering of incident electrons, this computationally convenient approximation provides a good starting point for initial interpretation of high-energy electron diffraction and microscopy experiments. Biological specimens are embedded in an amorphous solution and the potential distribution depends also on the dielectric and ionic properties of the solution. It seems appropriate to include the contribution of the solvent and ions. This is in turn modeled by $\mathcal{V}_{\text{bond}}$.

In the next section we will investigate the relation of $\mathcal{V}_{\text{bond}}$ to the total interaction potential \mathcal{V}^{int} .

----- *Intermezzo 2* -----

Born approximation for scattering against a single atom

Here we will consider a single elastic scattering event (kinematic theory). Although strictly speaking the single scattering approximation of a high-energy electron by an individual light atom in vacuum $V_0 = 0$ is not satisfied [10], the approximation is valid for a distance of a fraction of an Ångström away from the center of the atom. If the specimen is characterized by a bulk mean potential $V_0 \neq 0$, an electron can be scattered by this constant potential many times. However, as diffraction effects result from the spatial distribution of the potential, the process of multiple scattering by a constant potential V_0 is identical to rescaling the wavelength i.e. adding a constant phase factor to the incident electron wave. If the incident stationary electron wave is $\Psi_{\text{inc}} = e^{ik_0 r}$ then the scattered wave is a spherical wave modified with a $f_Z^{(e)}$ (scattering factor): $\Psi_{\text{sc}} = f \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}$. If we consider (F.1) and rewrite some of the terms ($k_0^2 := 2mE_e/\hbar^2$ and $U(\mathbf{r}') := 2me\mathcal{V}^{\text{int}}/\hbar^2$) then we have

$$(\nabla^2 + k_0^2)\Psi(\mathbf{r}') = U(\mathbf{r}')\Psi(\mathbf{r}'). \quad (\text{F.21})$$

The general solution involves the impulse response (Green's function) $G(\mathbf{r}, \mathbf{r}')$ to \mathbf{r} for a point scatterer at \mathbf{r}' , which is

$$G(\mathbf{r}, \mathbf{r}') \simeq -\frac{1}{4\pi} \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}. \quad (\text{F.22})$$

The scattered wave solution is

$$\Psi_{\text{sc}}(\mathbf{r}) = \int U(\mathbf{r}')\Psi(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')d^3\mathbf{r}'. \quad (\text{F.23})$$

The resulting integral equation for the wave function is known as Lippmann-Schwinger equation:

$$\begin{aligned} \Psi(\mathbf{r}) &= \Psi_{\text{inc}} + \Psi_{\text{sc}} \\ \Psi(\mathbf{r}) &= e^{ik_0 r} + \frac{2me}{\hbar^2} \int \mathcal{V}^{\text{int}}(\mathbf{r}')\Psi(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')d^3\mathbf{r}'. \end{aligned} \quad (\text{F.24})$$

The zero order solution of (F.24) is a plane wave $\Psi(\mathbf{r}) = e^{ik_0 r}$. This is a solution in vacuum ($\mathcal{V}^{\text{int}} = 0$). If the potential $V(r)$ is weak enough, it will just slightly distort the incident plane wave. The first order solution is the “first Born approximation” and is obtained by inserting $\Psi(\mathbf{r}') \simeq e^{ik_0 r'}$ into the integral of equation (F.24). It approximates the outgoing wave as a plane wave. The first Born approximation is valid for large incident energies and weak scattering potentials. In a scattering experiment, as the detector is located far away from the scatterer, it holds that $\mathbf{r} - \mathbf{r}' \approx \mathbf{r}$, where \mathbf{r} represents the distance to the detector and \mathbf{r}' the size of the feature and

$$k|\mathbf{r} - \mathbf{r}'| = k \sqrt{\mathbf{r}^2 - 2\mathbf{r}\mathbf{r}' + \mathbf{r}'^2} \approx kr \sqrt{1 - 2\frac{\mathbf{r}'}{r}} \approx kr - k\frac{\mathbf{r}}{r}\mathbf{r}' = kr - \mathbf{k}\mathbf{r}' \quad (\text{F.25})$$

$$\begin{aligned}
G(\mathbf{r}, \mathbf{r}') &\simeq -\frac{1}{4\pi} \frac{e^{ikr} e^{-ikr'}}{|\mathbf{r}|} \\
\Psi(\mathbf{r}) &= e^{ik_0 r} - \frac{me}{2\pi\hbar^2} \int \mathcal{V}^{\text{int}}(\mathbf{r}') e^{ik_0 r'} \frac{e^{ikr} e^{-ikr'}}{|\mathbf{r}|} d^3 \mathbf{r}' \\
\Psi(\mathbf{r}) &= e^{ik_0 r} - \frac{me}{2\pi\hbar^2} \frac{e^{ikr}}{|\mathbf{r}|} \int \mathcal{V}^{\text{int}}(\mathbf{r}') e^{i(k_0 - k)r'} d^3 \mathbf{r}' \\
\Delta \mathbf{k} &= \mathbf{k} - \mathbf{k}_0 \\
\Psi(\mathbf{r}) &= e^{ik_0 r} - \frac{me}{2\pi\hbar^2} \frac{e^{ikr}}{|\mathbf{r}|} \int \mathcal{V}^{\text{int}}(\mathbf{r}') e^{-i\Delta k r'} d^3 \mathbf{r}'
\end{aligned} \tag{F.26}$$

In the asymptotic case ($\mathbf{r} \rightarrow \infty$) it can be concluded that

$$\begin{aligned}
\Psi_{\text{sc}}(\Delta \mathbf{k}, \mathbf{r}) &= \frac{e^{ikr}}{|\mathbf{r}|} f_Z^{(e)}(\Delta \mathbf{k}) \\
f_Z^{(e)}(\Delta \mathbf{k}) &= -\frac{me}{2\pi\hbar^2} \int \mathcal{V}^{\text{int}}(\mathbf{r}') e^{-i\Delta k r'} d^3 \mathbf{r}'
\end{aligned} \tag{F.27}$$

The electric charges give rise to an electrostatic potential at a point in space which Fourier transform is the (electron) scattering factor $f_Z^{(e)}$ of the atom.

F.1.3 The Poisson-Boltzmann (PB) approach for modeling the electrostatic interactions between the atoms in a dielectric

Electrostatic interactions in macromolecular systems originate from: (i) the local charges (ions), (ii) the polarization from the non-spherical distribution of electron density around atoms, (iii) the redistribution of electrons caused by local electrical fields (electronic polarization), and (iv) the reorientation of polar groups in the solute (macromolecule) and solvent (buffer) molecules in response to the electric field (orientation polarization) [199]. All these electrostatic contributions can be adequately addressed in theory. Molecular dynamic simulations can provide sufficient knowledge of a set of favorable positions of both the macromolecules and solvent atoms. Distributions of charges can be determined by multipole expansions, and electronic polarization can be treated by polarization tensors. The practical difficulty remains in the accurate parameterizations of these properties and in large computational resources. Therefore, approximations are necessary.

Since biomolecules are always studied in solvents such as ionic water solution, the explicit approaches such as molecular dynamics must incorporate the electrostatic effects of an extremely large number of small solvent molecules. The treatment of atomic electrostatic interactions can be reduced to a problem of continuum electrostatics. The simplest model of electrostatic interactions in macromolecules assumes that charges on a molecule interact through a homogenous medium, and that all interactions can be characterized by Coulomb's law. However, this model is quite inaccurate since the solute (macromolecule) and the solvent have different dielectric properties. A more realistic model assumes that the solute and solvent have different

dielectric constants. Now, the Poisson equation of the system of charges and dielectrics has to be solved. This equation represents the starting point for modeling electrostatic interactions in macromolecules. Furthermore, in practise the solution in which the protein is embedded has counterions. This can be described by the Debye-Huckel theory of electrolytes.

In more detailed models, partial charges on all atoms can be included. The partial charge calculations are based on fitting quantum mechanical electrostatic potentials. The most common concept is to place an atomic partial charge at each atomic center (nucleus). These charges then interact by Gauss' law. In reality many electrons and nuclei come together to form a molecule – partial charges give a crude representation of what a neighboring atom will on average see due to this collection.

Let ρ be the total charge density in a sample, then

$$\rho = \rho_{\text{mol}} + \rho_{\text{sol}} + \rho_{\text{ion}} \quad (\text{F.28})$$

where ρ_{mol} is the charge density of the macromolecules, ρ_{sol} is the charge density of the solvent (typically water) and ρ_{ion} is the charge density of the ions in the specimen. Furthermore, let us assume that we can distinguish between the charge density that would originate from a collection of isolated atoms, ρ^{atom} , and the charge density ρ^{bond} that accounts for redistribution of charge due to electrostatic interactions between atoms, i.e. partial charge densities. Both ρ^{atom} and ρ^{bond} can be written as a sum of contributions from macromolecules, solvent, and ions:

$$\rho^{\text{atom}} = \rho_{\text{mol}}^{\text{atom}} + \rho_{\text{sol}}^{\text{atom}} + \rho_{\text{ion}}^{\text{atom}} \quad (\text{F.29})$$

$$\rho^{\text{bond}} = \rho_{\text{mol}}^{\text{bond}} + \rho_{\text{sol}}^{\text{bond}} + \rho_{\text{ion}}^{\text{bond}}. \quad (\text{F.30})$$

The linearized Poisson-Boltzmann equation (see *Intermezzo 3*) is

$$-\epsilon_0 \nabla(\epsilon_r(\mathbf{r}) \nabla \mathcal{V}^{\text{int}}(\mathbf{r})) = \rho^{\text{atom}} + \rho_{\text{mol}}^{\text{bond}} + \rho_{\text{sol}}^{\text{bond}} - \alpha(\mathbf{r}) \sum_i \frac{q_i^2 n_i^0 \mathcal{V}^{\text{int}}(\mathbf{r})}{k_B T} \quad (\text{F.31})$$

where q_i and n_i^0 are respectively the charge and concentration far away from the molecules of the ion of type i , k_B is Boltzmann constant, and T the temperature of the specimen immediately before vitrification (after vitrification the ions cannot move freely anymore). The relative permittivity is

$$\epsilon_r(\mathbf{r}) = \begin{cases} \epsilon_{\text{sol}} & \mathbf{r} \in \Omega_{\text{sol}} (\alpha(\mathbf{r}) = 1), \\ \epsilon_{\text{mol}} & \text{otherwise } (\alpha(\mathbf{r}) = 0), \end{cases} \quad (\text{F.32})$$

where Ω_{sol} is the region occupied by solvent; ϵ_{mol} takes values around 2, and ϵ_{sol} is typically around 80.

----- *Intermezzo 3* -----

The Poisson-Boltzmann equation (PBE)

Using Coulomb's law that was at the basis of our interaction Hamiltonian (F.9), the interaction potential (F.15) can be rewritten into

$$\mathcal{V}^{\text{int}}(\mathbf{r})(\psi_e) = \frac{1}{4\pi\epsilon_0} \left[\int_{\mathbb{R}^{3m} \times \mathbb{R}^{3n}} \psi_{\text{sp}}(\mathbf{S})^* \sum_{i=1}^{n+m} \frac{eQ_i}{|\mathbf{r} - \mathbf{r}_i^0|} \psi_{\text{sp}}(\mathbf{S}) d\mathbf{R}_0 \right] \psi_e(\mathbf{r}). \quad (\text{F.33})$$

Similarly to (F.17), we can write (F.16) as

$$\mathcal{V}_{\text{nucl}}^{\text{int}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^3} \frac{\rho_{\text{nucl}}(\mathbf{y})}{|\mathbf{r} - \mathbf{y}|} d\mathbf{y}, \quad (\text{F.34})$$

where ρ_{nucl} represents the charge density of the nuclei and is defined as

$$\rho_{\text{nucl}}(\mathbf{y}) := - \int_{\mathbb{R}^{3(n-1)}} \sum_{i=1}^m Z_i e |\phi_{\text{sp}}^e(\mathbf{y}, \mathbf{r}_2^0, \dots, \mathbf{r}_n^0)|^2 d\mathbf{r}_2^0 \dots d\mathbf{r}_n^0.$$

Adding the contribution of charge densities of nuclei and electrons $\rho = \rho_e + \rho_{\text{nucl}}$ we get

$$\mathcal{V}^{\text{int}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^3} \frac{\rho(\mathbf{y})}{|\mathbf{r} - \mathbf{y}|} d\mathbf{y}$$

where ρ now represents the total charge density. The electric field is then

$$\mathbf{E}(\mathbf{r}) = -\nabla\mathcal{V}^{\text{int}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \rho(\mathbf{y}) \frac{\mathbf{r} - \mathbf{y}}{|\mathbf{r} - \mathbf{y}|^3} d\mathbf{y}. \quad (\text{F.35})$$

Taking the divergence with respect to \mathbf{r} of both sides of (F.35) and use the identity

$$\nabla \cdot \left(\frac{\mathbf{r}}{|\mathbf{r}|^3} \right) = 4\pi\delta(\mathbf{r}),$$

we get

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{1}{\epsilon_0} \int \rho(\mathbf{y}) \delta(\mathbf{r} - \mathbf{y}) d\mathbf{y}. \quad (\text{F.36})$$

This represents a form of Gauss' law in vacuum:

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0}.$$

The Poisson equation is obtained by expressing the above in terms of the potential:

$$\nabla^2 \mathcal{V}^{\text{int}}(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0}. \quad (\text{F.37})$$

Gauss' law for a dielectric is given by

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho(\mathbf{r})$$

where $\mathbf{D}(\mathbf{r}) = \epsilon(\mathbf{r})\mathbf{E}(\mathbf{r})$ holds for linear dielectrics and $\epsilon(\mathbf{r}) = \epsilon_r(\mathbf{r})\epsilon_0$ with $\epsilon_r(\mathbf{r})$ being dielectric constant or relative permittivity. The Poisson equation for linear dielectrics is then

$$-\nabla \cdot (\epsilon(\mathbf{r}) \nabla \mathcal{V}^{\text{int}}(\mathbf{r})) = \rho(\mathbf{r}), \quad (\text{F.38})$$

which, using (F.29)–(F.30), can be written as

$$-\nabla(\epsilon(\mathbf{r})\nabla\mathcal{V}^{\text{int}}(\mathbf{r})) = \rho^{\text{atom}}(\mathbf{r}) + \rho_{\text{mol}}^{\text{bond}}(\mathbf{r}) + \rho_{\text{sol}}^{\text{bond}}(\mathbf{r}) + \rho_{\text{ion}}^{\text{bond}}(\mathbf{r}). \quad (\text{F.39})$$

Furthermore, it is reasonable to assume that the solution has counterions. Electroneutrality of the solution implies

$$\sum_{i=1}^p N_i q_i = 0. \quad (\text{F.40})$$

where N_i and q_i are respectively the number and charge of the ion of type i . The work to bring an ion of type i from infinity to the position \mathbf{r} is equal to $q_i\mathcal{V}^{\text{int}}(\mathbf{r})$ where $\mathcal{V}^{\text{int}}(\mathbf{r})$ is the interaction potential at the point \mathbf{r} . Ions are in thermal equilibrium and relatively free to move, so their concentrations follow a Boltzmann distribution:

$$n_i(\mathbf{r}) = n_i^0 \exp\left(-\frac{q_i\mathcal{V}^{\text{int}}(\mathbf{r})}{k_B T}\right),$$

where n_i^0 is the concentration of the ion of type i far away from the molecules. The charge density of the interacting ions is therefore given by

$$\rho_{\text{ion}}^{\text{bond}}(\mathbf{r}) = \sum_i q_i n_i(\mathbf{r}) = \sum_i q_i n_i^0 \exp\left(-\frac{q_i\mathcal{V}^{\text{int}}(\mathbf{r})}{k_B T}\right). \quad (\text{F.41})$$

Inserting (F.41) into (F.39) gives us the Poisson-Boltzmann equation (PBE) for the total charge density:

$$-\nabla(\epsilon(\mathbf{r})\nabla\mathcal{V}^{\text{int}}(\mathbf{r})) = \rho^{\text{atom}}(\mathbf{r}) + \rho_{\text{mol}}^{\text{bond}}(\mathbf{r}) + \rho_{\text{sol}}^{\text{bond}}(\mathbf{r}) + \sum_i q_i n_i^0 \exp\left(-\frac{q_i\mathcal{V}^{\text{int}}(\mathbf{r})}{k_B T}\right). \quad (\text{F.42})$$

This is a transcendental equation in \mathcal{V}^{int} , so we cannot easily solve for the potential. For low concentration of ions (i.e. the electrostatic potential energy is small compared to $k_B T \approx 25$ meV at 300 K), a first order Taylor series approximation can be used to linearize the ionic part of PBE

$$\sum_i q_i n_i^0 \exp\left(-\frac{q_i\mathcal{V}^{\text{int}}(\mathbf{r})}{k_B T}\right) \approx \sum_i \left(q_i n_i^0 - \frac{q_i^2 n_i^0 \mathcal{V}^{\text{int}}(\mathbf{r})}{k_B T} \right).$$

The first term on the right hand side is zero due to electroneutrality of the solution (see equation (F.40)). Hence, the linearized PBE is

$$-\epsilon_0 \nabla(\epsilon_r(\mathbf{r})\nabla\mathcal{V}^{\text{int}}(\mathbf{r})) = \rho^{\text{atom}} + \rho_{\text{mol}}^{\text{bond}} + \rho_{\text{sol}}^{\text{bond}} - \alpha(\mathbf{r}) \sum_i \frac{q_i^2 n_i^0 \mathcal{V}^{\text{int}}(\mathbf{r})}{k_B T} \quad (\text{F.43})$$

with

$$\epsilon_r(\mathbf{r}) = \begin{cases} \epsilon_{\text{sol}} & \mathbf{r} \in \Omega_{\text{sol}} (\alpha(\mathbf{r}) = 1), \\ \epsilon_{\text{mol}} & \text{otherwise } (\alpha(\mathbf{r}) = 0), \end{cases} \quad (\text{F.44})$$

where ϵ_{mol} takes values between 1 and 4, and ϵ_{sol} is typically around 80.

F.1.4 Combining IASA- and PB -based potentials

Considering equations (F.31) and (F.32) and taking into account that the Laplace operator is linear, the electrostatic potential can be decomposed in two ways: (1) ‘‘atom’’ contributions which consider the superposition of each atom potential as if each atom was in isolation and ‘‘bond’’ contributions which provide a correction due to the molecular interactions; (2) atoms are divided into three classes: those belonging to the molecule to be studied, the solvent and the ions in the solution. So the potential is a sum of six terms

$$\mathcal{V}^{\text{int}} = \mathcal{V}_{\text{mol}}^{\text{atom}} + \mathcal{V}_{\text{sol}}^{\text{atom}} + \mathcal{V}_{\text{ion}}^{\text{atom}} + \mathcal{V}_{\text{mol}}^{\text{bond}} + \mathcal{V}_{\text{sol}}^{\text{bond}} + \mathcal{V}_{\text{ion}}^{\text{bond}},$$

where each term satisfies an equation similar to (F.31) with the corresponding charge density.

Discriminating between $\mathcal{V}_{\text{atom}}$ and $\mathcal{V}_{\text{bond}}$, as we assumed in section F.1.1, is important because $\mathcal{V}_{\text{atom}}$ is confined within the electron shells of each atom of the specimen. We consider it as independent of the other atoms in the molecule, ions or solvent. By contrast, $\mathcal{V}_{\text{bond}}$ has a range that extends over distances of several atoms. In order to calculate the total electrostatic potential we make the following assumptions:

- (a) The positions of all atoms in the macromolecule are known, *e.g.* specified by means of a PDB file.
- (b) Space not occupied by the macromolecule, *i.e.* Ω_{sol} , is filled with solvent (can be treated either explicitly *via* molecular dynamics (MD) approaches or implicitly by continuum electrostatics).
- (c) Concentration and type of ions in the solution far away from the macromolecules n_i^0 is known.
- (d) Atomic partial charges are calculated using a force field and are placed at each atomic center (nucleus).
- (e) Close to the nucleus, the contribution of the potential from the partial charge $\mathcal{V}_{\text{bond}}$ is smaller than $\mathcal{V}_{\text{atom}}$, so it is sensible to place the partial charge density as a point charge located at the nucleus.
- (f) The charge density of the nuclei is significantly larger than partial charge densities and its potential is the major cause of scattering of high-energy incident electrons.

Calculation of $\mathcal{V}_{\text{atom}}$ is described in section F.1.2. This potential is positive and decreases rapidly with the distance from the nucleus. Its value should correspond to the potential calculated *via* (F.31) at various distances close to the nucleus (at points $|\mathbf{r}| \leq r_i^0$ where r_i^0 is the radius of the i -th isolated atom). In general, the positions of ions and atoms in the solvent are not known and therefore $\rho_{\text{sol}}^{\text{atom}}$ and $\rho_{\text{ion}}^{\text{atom}}$ cannot be computed exactly. However, there are two approaches to address that issue: (i) Find the optimal positions of solvent and ion atoms by performing MD simulations and then compute the potential by using tabulated potentials as for the case of the macromolecules (section F.1.2). (ii) Compute the average value of $\mathcal{V}_{\text{sol}}^{\text{atom}}(\mathbf{r})$ given the density of the solvent. Note that the latter averaging approach is significantly less computationally expensive.

Let us focus on equation (F.30) and $\mathcal{V}_{\text{bond}}(\mathbf{r})$. The interactions between the atoms in the sample could be represented by a force field. The basic form of the force field includes terms corresponding to the potential energy describing covalent bonds, long-range electrostatic and Van der Waals forces. However, since we showed that our one-body Schrödinger equation (F.1) does not depend on the Hamiltonian of the specimen and that we can ignore all the interactions between atoms except the electrostatic ones, $\mathcal{V}_{\text{bond}}$ can be calculated from

$$-\epsilon_0 \nabla(\epsilon_r(\mathbf{r}) \nabla \mathcal{V}_{\text{bond}}(\mathbf{r})) = \rho_{\text{mol}}^{\text{bond}}(\mathbf{r}) + \rho_{\text{sol}}^{\text{bond}}(\mathbf{r}) - \alpha(\mathbf{r}) \sum_i \frac{q_i^2 n_i^0 \mathcal{V}_{\text{bond}}(\mathbf{r})}{k_B T}. \quad (\text{F.45})$$

Both $\rho_{\text{mol}}^{\text{bond}}$ and $\rho_{\text{sol}}^{\text{bond}}$ can be considered as “fixed” partial charge densities and modeled as delta functions $\delta(\mathbf{r} - \mathbf{r}_i)$ located at the atom centers \mathbf{r}_i with magnitudes (partial charges) Q_i :

$$\rho_{\text{fixed}}(\mathbf{r}) = \rho_{\text{mol}}^{\text{bond}} + \rho_{\text{sol}}^{\text{bond}} = \sum_i Q_i \delta(\mathbf{r} - \mathbf{r}_i).$$

Further simplification can include the averaging method for solvent partial charge densities. Changing of the solvent properties such as pH, ions, and dielectric constant will redistribute the partial charges (especially at the surface of the macromolecule) and consequently alter $\mathcal{V}_{\text{bond}}$. Here we consider the continuous model and the connection between each of the partial charge densities ρ^{bond} and the PBE model.

- $\rho_{\text{ion}}^{\text{bond}}$ corresponds to the Boltzmann term in the PBE representing the partial charge density of ions in solution.
- $\rho_{\text{sol}}^{\text{bond}}$ is modeling the impact of the solvent implicitly (primarily through molecular polarization) *via* the external dielectric constant.
- $\rho_{\text{mol}}^{\text{bond}}$ corresponds to the partial charges used for a coarse electrostatics description inside the molecule.

PB solvers are mostly interested in the calculation of solvation potential and energy (on the surface of a molecule) in order to understand active binding sites [178, 200–203]. Hence, a typical solution of a PB equation is $\mathcal{V}_{\text{sol}}^{\text{bond}} + \mathcal{V}_{\text{ion}}^{\text{bond}}$. This result is obtained from the difference between $\mathcal{V}_{\text{bond}}$ and $\mathcal{V}_{\text{mol}}^{\text{bond}}$. The first step is the calculation of the total $\mathcal{V}_{\text{bond}}$ *via* (F.45), and the

second step is the calculation of $\mathcal{V}_{\text{mol}}^{\text{bond}}$ only. $\mathcal{V}_{\text{mol}}^{\text{bond}}$ represents the potential of the molecule in an environment with the same dielectric constant and it is obtained *via* the solution of the Poisson equation for a uniform dielectric (Gauss law)

$$-\epsilon_0 \epsilon_{\text{mol}}(\mathbf{r}) \nabla^2 \mathcal{V}_{\text{mol}}^{\text{bond}}(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r}). \quad (\text{F.46})$$

The advantage of this approach is that $\mathcal{V}_{\text{bond}}$ and $\mathcal{V}_{\text{mol}}^{\text{bond}}$ result from the same set of fixed partial charges and thus have identical singularities, the difference $\mathcal{V}_{\text{bond}} - \mathcal{V}_{\text{mol}}^{\text{bond}} = \mathcal{V}_{\text{sol}}^{\text{bond}} + \mathcal{V}_{\text{ion}}^{\text{bond}}$ is a smooth function. The knowledge about the electrostatic distribution within the molecule is beneficial for us. The final potential due to redistribution is obtained in one step by calculating $\mathcal{V}_{\text{bond}}$ from (F.45) ($\mathcal{V}_{\text{bond}} = \mathcal{V}_{\text{mol}}^{\text{bond}} + \mathcal{V}_{\text{sol}}^{\text{bond}} + \mathcal{V}_{\text{ion}}^{\text{bond}}$). After considering assumption (e) above, the total potential in the specimen can be determined by $\mathcal{V}^{\text{int}} = \mathcal{V}_{\text{atom}} + \mathcal{V}_{\text{bond}}$, i.e. combining the contributions from potentials of isolated atoms calculated *via* scattering factors and contributions from the redistribution of charges calculated *via* PBE (F.45). The final sum corresponds to the total interaction potential \mathcal{V}^{int} . We argue that $\mathcal{V}_{\text{bond}}$ calculated *via* PBE gives an adequate correction to the electrostatic potential calculated using the isolated atom superposition approximation as described in section F.1.2.

$\mathcal{V}_{\text{bond}}$ calculated *via* a Poisson-Boltzmann approach could be also used to identify active and ligand-binding sites, to predict protein-protein and protein-membrane interfaces and to categorize biomolecules on the basis of the potential surrounding their surfaces. It is anticipated that such approaches will become increasingly important as the number of experimentally resolved structures increases.

F.2 High-energy electron-specimen interaction

The incident electron interacts with the electrostatic (interaction) potential of the specimen. The common description of the interaction between the incident electron and the specimen involves approximations such as projection assumption (PA) and/or weak-phase object approximation (WPOA), both based on the small angle approximation. A multislice approach extends the limits of the PA and WPOA. Appendix G provides criteria for applicability of these methods.

F.2.1 Small angle approximation

The incident electron possesses a high-energy and it travels predominately along the optical axis in the z direction. The specimen is a relatively small perturbation on this motion. Let us consider again the differential form of the one-body stationary Schrödinger equation (F.1). Similarly to the high-energy assumption in section F.1.1, it is useful to represent the total electron wave function $\psi_e(\mathbf{r})$ as a product of a plane wave traveling in the z direction and the wave function which varies slowly with z , i.e. $\psi_e(\mathbf{r}) = \Psi(x, y, z)e^{ikz}$. Now we have

$$\nabla_{\mathbf{r}}^2(\psi_e)(\mathbf{r}) = (\nabla_{x,y}^2 + \nabla_z^2)(\psi_e)(\mathbf{r}) = e^{ikz} \nabla_{x,y}^2 \Psi(x, y, z) + \nabla_z^2[\Psi(x, y, z)e^{ikz}]. \quad (\text{F.47})$$

Using the short notation $\Psi(x, y, z) = \Psi$, rearranging the terms in the Schrödinger equation and dividing by the common factor e^{ikz} we get:

$$\left[\nabla_r^2 - \frac{2me}{\hbar^2} \mathcal{V}^{\text{int}}(\mathbf{r}) \right] \Psi = \frac{2mE_e}{\hbar^2} \Psi. \quad (\text{F.48})$$

Since $k = \sqrt{2mE_e}/\hbar$, this simplifies to

$$\left[\nabla_{x,y}^2 + \frac{\partial^2 \Psi}{\partial z^2} + 2ik \frac{\partial \Psi}{\partial z} - \frac{2me}{\hbar^2} \mathcal{V}^{\text{int}}(\mathbf{r}) \right] \Psi = 0. \quad (\text{F.49})$$

Given our assumptions that the energy of the incident electron is high and that Ψ varies slowly with z we adhere to the small angle approximation $\left| \frac{\partial^2 \Psi}{\partial z^2} \right| \ll \left| k \frac{\partial \Psi}{\partial z} \right|$ ($k^2 \gg k_x^2 + k_y^2$) and therefore equation (F.49) can be represented, similarly as in [144], as

$$\frac{\partial \Psi}{\partial z} = \left[\frac{i\lambda}{4\pi} \nabla_{x,y}^2 + i\sigma \mathcal{V}^{\text{int}}(\mathbf{r}) \right] \Psi, \quad (\text{F.50})$$

where $\lambda = 2\pi/k$ is the wavelength of the incident electron and $\sigma = \lambda me/(2\pi\hbar^2)$ is the interaction constant.

F.2.2 Multislice method, projection assumption, weak-phase object approximation

Two popular methods for modeling electron transmission through the specimen (including the effects of multiple scattering and 'geometrical' thickness) are the Bloch wave method and the multislice method [144]. Here we will shortly describe the multislice method presented in [144]

as it is advantageous for non-crystalline and large specimens. Remembering that $\int_z^{z+\Delta z} \frac{\partial \Psi}{\partial z} dz = \ln(\Psi)|_z^{z+\Delta z}$ the solution of equation (F.50) can be written as

$$\Psi(x, y, z + \Delta z) = \exp \left[\frac{i\lambda}{4\pi} \Delta z \nabla_{x,y}^2 + i\sigma \int_z^{z+\Delta z} \mathcal{V}^{\text{int}}(x, y, z') dz' \right] \Psi(x, y, z). \quad (\text{F.51})$$

Assuming Δz is the thickness of a thin slice through the specimen, $V_z(x, y, z) = \int_z^{z+\Delta z} \mathcal{V}^{\text{int}}(x, y, z') dz'$ is the projected potential within the slice, and $t(x, y, z) = \exp(i\sigma V_z(x, y, z))$ is the transmission function (phase grating), we have

$$\Psi(x, y, z + \Delta z) \approx \exp \left(\frac{i\lambda}{4\pi} \Delta z \nabla_{x,y}^2 \right) t(x, y, z) \Psi(x, y, z). \quad (\text{F.52})$$

If we define $q_x = k_x/(2\pi)$ and $q_y = k_y/(2\pi)$, it can be shown [144,204] that the following relation holds

$$\mathcal{F} \left[\exp \left(\frac{i\lambda}{4\pi} \Delta z \nabla_{x,y}^2 \right) (t(x, y, z) \Psi(x, y, z)) \right] = \exp[-i\pi\lambda\Delta z(q_x^2 + q_y^2)] \mathcal{F} [t(x, y, z) \Psi(x, y, z)] \quad (\text{F.53})$$

$$= P(q, \Delta z) \mathcal{F} [t(x, y, z) \Psi(x, y, z)], \quad (\text{F.54})$$

where $P(q, \Delta z)$ is the Fresnel propagator. The wave function at the top of the specimen's $(n+1)$ th slice is

$$\Psi_{n+1}(x, y) = \mathcal{F}^{-1} [P_n(q, \Delta z_n) \mathcal{F}[t_n(x, y) \Psi_n(x, y)]] . \quad (\text{F.55})$$

Therefore, the propagation of the electron wave through the specimen can be interpreted as a recursive transmission and propagation of the wave function through each slice until the wave leaves the specimen ($\Psi_{\text{exit}}(x, y)$). The accuracy of the multislice approach increases with smaller slice thickness at the expense of a longer computational time. The slice thickness should not be smaller than the range of the atomic potential or the electron wavelength.

The projection assumption (PA) is commonly used for a thin object, where, given that the incident wave is a plane wave, the transmitted wave function is equal to the transmission function:

$$\Psi_{\text{exit}}(x, y) = t(x, y, z) \Psi_{\text{inc}}(x, y) = \exp(i\sigma V_z(x, y, z)). \quad (\text{F.56})$$

If the scattering is weak, especially in the case of light atoms, the weak-phase object approximation (WPAO) ($\sigma V_z \ll 1$) can be used. The wavefunction at distance z then becomes

$$\Psi(x, y, z) = \exp(i\sigma V_z(x, y, z)) \approx 1 + i\sigma V_z(x, y, z). \quad (\text{F.57})$$

The similarity between WPOA and the first Born approximation (see section F.1.2) is discussed in [205]. A closed form expression for the intensity using the first Born approximation can be found e.g. in [206].

Criteria for applicability of weak-phase object approximation, projection assumption and multislice have been suggested in Appendix G [52].

After passing through the specimen, the electron wave is convolved with the point-spread function of the microscope whose Fourier transform is called the contrast transfer function (CTF)

$$T(q) \sim \exp[-i\pi\lambda q^2(0.5C_s\lambda^2 q^2 - \Delta f),] \quad (\text{F.58})$$

where Δf is the defocus of the objective lens, and C_s the coefficient of spherical aberration. If one assumes not more than one weak elastic scattering event on the interaction potential (WPOA), the effects of geometrical thickness and/or tilt of the sample can be incorporated in the CTF [29, 173–176]. After being scattered once, the electron wave propagates in free space. That is equivalent to Fresnel propagation defined in equation (F.53). The Fresnel propagation over a distance Δz is equivalent to a defocus of $\Delta f = -\Delta z$ allowing us to model the specimen thickness as a part of the CTF, speeding up the forward computation [29] as well as 3DCTF correction [30].

F.2.3 Inelastic interactions

The quantum states of both the incident electron and the specimen change in an inelastic interaction. This change is accompanied by a transfer of energy between the scattered electron and the object. A proper modeling of inelastic scattering requires to treat both electrons and nuclei in the specimen using quantum mechanics. One can, however, make a simplified treatment. Inelastic scattering can be taken into account using a first-order perturbation analysis giving

Table F.1. Comparison between inelastic and elastic treatment

Scattering	Inelastic	Elastic
Incident	$\Gamma_{\text{in}} = \langle \Psi_{\text{in}}^*(\mathbf{r}, t) \Psi_{\text{in}}(\mathbf{r}', t - \tau) \rangle_T$	$\Psi_{\text{in}}(\mathbf{r})$
Object	$\gamma = \langle e^{i\chi(\rho, t)} e^{-i\chi(\rho', t')} \rangle$	$e^{i\chi(\rho)}$
Exit	$\Gamma_{\text{ex}} = \Gamma_{\text{in}} \cdot \gamma$	$\Psi_{\text{ex}}(\mathbf{r}) = \Psi_{\text{in}}(\mathbf{r}) \cdot e^{i\chi(\rho)}$

rise to an imaginary potential. The final potential is a complex quantity and it is usually called “optical potential”, by analogy with the complex refractive index in optics [207, 208]. The difference is that in optics the imaginary part of the refractive index causes the absorption of light. In TEM, however, this is usually not an actual absorption of electrons, but a loss of the electrons from the incident elastic energy channel. The probability that an inelastically scattered electron reappears in the elastic channel is negligible [209], and therefore inelastic scattering is responsible for the appearance of an imaginary part of the optical potential V_{ab} . Inelastically scattered electrons are superimposed on the TEM image formed by elastically scattered electrons forming a background. In principle they should be filtered with an energy filter. Inelastic processes can be generally separated into two components: (1) processes that generate secondary electrons or light (slow and fast secondary electrons, Auger electrons, X-rays (characteristic and Bremsstrahlung X-rays), cathodoluminescence); (2) processes that result from collective interactions with many atoms (plasmons and phonons). Even for energy filtered high resolution imaging, inelastic scattering effects are present in the image because electrons which have suffered a very small energy loss cannot be separated from the unscattered or elastically scattered electrons by a conventional energy filter with a slit of 1 – 5 eV [210]. Phonon scattering, for example, produces very small energy losses in fractions of 1 eV and contributes to the intensity for very high scattering angles [211]. For imaging of biological specimens this is expected not to be relevant.

A comprehensive theory of image formation includes the inelastic scattering effects which are roughly equivalent to that of a rearrangement of the object scatterers during the exposure time [212]. The specimen is free to move around its initial bound state or transfer to an excited state. The strict treatment of the problem of incorporating inelastic scattering into image simulation must include mutual coherence [212, 213] or a density matrix approach [214]. The intensity (probability density function) is related to a time-independent wave function only in the case of purely elastic scattering and coherent illumination. The fluctuations within the object, the source and the optics during the exposure time T exert an influence on the final detected signal which is time averaged. The mutual coherence function is defined as $MCF = \Gamma_{\text{in}} = \langle \Psi_{\text{in}}^*(\mathbf{r}, t) \Psi_{\text{in}}(\mathbf{r}', t - \tau) \rangle_T$ [212]. Inelastic scattering results from excitations of the internal degrees of freedom of the object and its effect can be incorporated in time-dependent potential. Further, the concept of mutual dynamic object transparency ($MDOT = \gamma = \langle e^{i\chi(\rho, t)} e^{-i\chi(\rho', t')} \rangle$), $\rho = (x, y)$ is introduced by [212]. This measure represents the relative change of MCF caused by the transmission through the specimen. It is a complex degree of coherence introduced by the specimen. If we consider a thin object, we can compare the model for propagation of elastically and inelastically scattered electrons (see Table F.1). From [212], the $MDOT$ can be expressed

as

$$\gamma \approx \exp[i\mu_1(\rho) - i\mu_1(\rho') - \frac{1}{2}\mu_2(\rho) - \frac{1}{2}\mu_2(\rho') + \mu_{11}(\rho, \rho', \tau)], \quad (\text{F.59})$$

where the function μ_1 represents the static projected potential and elastic scattering; absorption in the specimen is described by μ_2 and inelastic scattering by μ_{11} . Note that μ_2 is the energy averaged μ_{11} . The challenges of calculating μ_{11} are described in [215]. However, it was shown recently that the contribution of the inelastic μ_{11} to the elastic part μ_1 for silicon sample Si(110) at 20 kV acceleration voltage is 0.15 % [216]. Let us scale this influence as a function of atomic number and acceleration voltage. As described in [158] and [163] the elastic σ_{el} and inelastic σ_{in} scattering cross sections can be represented as:

$$\begin{aligned} \sigma_{\text{el}} &= \frac{1.4 \cdot 10^{-6} Z^{\frac{3}{2}}}{\beta^2} \left[1 - \frac{0.26Z}{137\beta}\right] [\text{nm}^2], \\ \sigma_{\text{in}} &= \frac{1.5 \cdot 10^{-6} Z^{\frac{1}{2}}}{\beta^2} \ln \frac{\beta^2(U_0 + mc^2)}{10} [\text{nm}^2], \end{aligned} \quad (\text{F.60})$$

where Z is atomic number, β the ratio between the velocity of the electron and light ($\beta^2 = 1 - [mc^2/(U_0 + mc^2)]^2$), U_0 the incident electron energy, and mc^2 the rest energy of electron. The ratio of inelastic and elastic scattering cross sections is therefore

$$R(Z, U_0) = \frac{\sigma_{\text{in}}}{\sigma_{\text{el}}} = \frac{1.5 \ln \frac{\beta^2(U_0 + mc^2)}{10}}{1.4Z \left[1 - \frac{0.26Z}{137\beta}\right]}. \quad (\text{F.61})$$

As a rough estimation we can scale the 0.15 % contribution (of μ_{11} to μ_1) for a silicon ($Z = 14$) at 20 keV to e.g. carbon ($Z = 6$) at 80 keV as $0.15 \% \cdot \frac{R(6,80000)}{R(14,20000)} = 0.37 \%$. This contribution decreases with increasing atomic numbers and/or acceleration voltage. The contribution of time dependent μ_{11} for a single atom is therefore small. However, the percentage of inelastically scattered electrons from thick vitreous ice is large leading to damping of the useful contrast. The time-dependent part of the interaction potential could be therefore important in modeling the formation of unfiltered images. Given the complexity and computational challenges of the mutual coherence function approach, we will consider only the time averaged part of the inelasticities μ_2 (absorption). This contributes to the imaginary part of the total potential V_{ab} and requires zero-loss energy filtered images for comparison. The total complex potential is therefore

$$\mathcal{V}_{\text{tot}}^{\text{int}} = V_{\text{ph}} + iV_{\text{ab}}. \quad (\text{F.62})$$

The most dominant inelastic interactions in cryo-EM are plasmons of the solvent (vitreous ice) [217]. For such an amorphous specimen, one can assume that the overlap of atomic positions in a projection is significant and that the projected amorphous sample is essentially noise with a flat frequency spectrum. This is surely an approximation as every real specimen has limited scattering power. The mean real part of the potential V_{ph} of the sample introduces a constant phase change of the electron wave while the imaginary part of the potential V_{ab} causes exponential damping of the wave amplitude and intensity:

$$I_{\text{exit}}(x, y) = |\Psi_{\text{exit}}|^2 = |\exp(i\sigma V_z(x, y) - \sigma V_{z,\text{ab}}(x, y))|^2 = \exp(-2\sigma V_{z,\text{ab}}). \quad (\text{F.63})$$

Since this damping increases with the thickness d of the specimen we have

$$I_{\text{exit}}(x, y) = I_{\text{inc}}(x, y) \exp\left(-\frac{d}{\Lambda_{\text{in}}}\right) \Rightarrow \frac{d}{\Lambda_{\text{in}}} = \ln \frac{I_{\text{inc}}}{I_{\text{exit}}}, \quad (\text{F.64})$$

where Λ_{in} is the inelastic mean free path at a certain electron energy. From the comparison between equations F.63 and F.64, the following relation holds

$$\frac{d}{\Lambda_{\text{in}}} = 2\sigma V_{z,\text{ab}}. \quad (\text{F.65})$$

The inelastic mean free path depends on the incident electron energy and is related to the scattering cross section (equation (F.60)) as

$$\Lambda_{\text{in}} = \frac{M_{\text{W}}}{\rho N_{\text{A}} \sigma_{\text{in}}} = \frac{M_{\text{W}} \beta^2 \cdot 10^{10}}{9.03 \rho Z^{\frac{1}{2}} \ln \frac{\beta^2 (U_0 + mc^2)}{10}} [\text{nm}]. \quad (\text{F.66})$$

Experimentally determined values for inelastic mean free paths reported in the literature vary noticeably [164]. The reasons for these apparent discrepancies are not always clear. We used the values of $\Lambda_{\text{in}} = 180$ nm for vitreous ice at 80 keV and $\Lambda_{\text{in}} = 108$ nm for protein at 100 keV provided by [158] and [159], respectively. The values for any other incident energy of electron U_0 are calculated *via* equation (H.4). The fractional composition of a protein was taken to be 0.492, 0.313, 0.094, and 0.101 for elements H, C, N, and O, respectively [159, 165].

Monte-Carlo simulations for inelastics

We generated electron energy-loss spectra for amorphous carbon and vitreous water utilizing the Monte-Carlo simulation package Geant4 [154, 155]. Using Geant4, it is possible to follow the trajectory and the energy of a given particle (as well as all secondary generated particles) within a material. The densities were taken to be 1.8 g/cm³ and 0.93 g/cm³ for amorphous carbon and vitreous water, respectively.

In our case, blocks of amorphous water and carbon of different thicknesses d (5 to 150 nm) were generated. A number of primary electrons $N \sim 10^6$, of various incident energies E_e (80 keV, 100 keV, 120 keV, 200 keV, and 300 keV), were introduced perpendicular to the specimen surface acting as a parallel illumination. From Fig F.1A it is clear that higher energy losses are increasing considerably with the thickness of the sample forming clear plasmon peaks. The thickness dependent ratio between the number of electrons that have lost less than 3 eV (I_{z1}) and the total number of electrons $N = I$ is shown in Fig. F.1B.

The values in Fig. F.1B obtained for different sample thicknesses perfectly match an exponential decay (solid curves). This validates equation (4.3) and the assumption that delocalized processes (represented by a constant absorption potential) are dominant. From Fig. F.1B we also see that for thicknesses used in cryo imaging (~ 100 nm) around 43 % and 26 % of the electrons at 80 keV and 300 keV, respectively, experience inelastic scattering. Thus, the absorption potential in that case plays an important role.

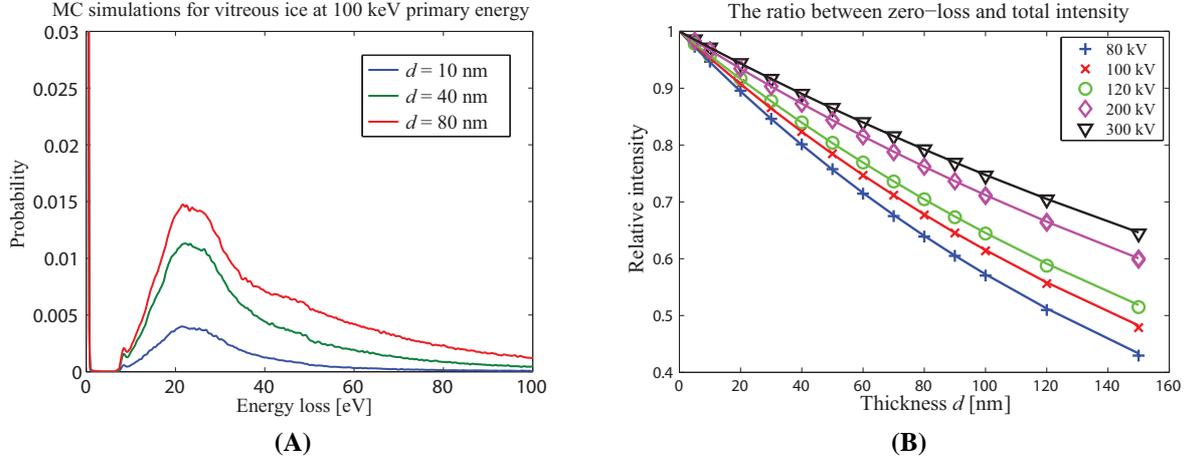


Fig. F.1. Monte-Carlo simulations (Geant4) of amorphous vitreous water. (A) EELS spectrum showing the increase of the plasmon peak with thickness. (B) The thickness dependence of the ratio between number of electrons that have lost less than 3 eV and the total number of electrons for inelastic mean free paths at various energies specified in section 4.3.1.

F.3 Optical system

In cryo-EM we record mostly images generated by phase contrast, as a result of interference between the unscattered and scattered part of the electron exit wave function. The electron wave exiting the specimen $\Psi_{\text{exit}}(x, y)$ is further subject to a frequency dependent phase shift introduced by the microscope aberrations. If we consider spherical aberration C_s , defocus Δf , and twofold astigmatism (A_1, α_1) , the total aberration function in polar coordinates is

$$\chi(q, \alpha) = \frac{2\pi}{\lambda} \left(\frac{1}{4} C_s \lambda^4 q^4 - \frac{1}{2} (\Delta f - A_1 \cos(2(\alpha - \alpha_1))) \lambda^2 q^2 \right), \quad (\text{F.67})$$

where q is the magnitude of the spatial frequency (q_x, q_y) . Note that underfocus implies $\Delta f > 0$, as in [111].

The energy spread of the incident electrons and the finite virtual source size give rise to temporal and spatial incoherence, respectively. These can be modeled as damping envelopes in the spatial frequency domain. The temporal incoherence of the source can be modeled as a chromatic envelope function [111]:

$$K_c(q) = \exp \left[- \left(\frac{\pi \lambda q^2 C_c \Delta E}{4E \sqrt{\ln 2}} \right)^2 \right], \quad (\text{F.68})$$

where C_c is the chromatic aberration coefficient; E and ΔE are the energy and the energy spread of the incident electrons, respectively. Furthermore, the finite source size introduces spatial incoherence which results in the spatial envelope function [111]:

$$K_s(q, \alpha) = \exp \left[- \frac{(\pi C_s \lambda^2 q^3 - \pi \Delta f(\alpha) q)^2 \alpha_i^2}{\ln 2} \right] \quad (\text{F.69})$$

where α_i is the illumination aperture.

The objective aperture function, optionally including a phase plate $A_p(q)$ is described as

$$A_p(q) = \begin{cases} 1 & q \leq q_{\text{cuton}}, \\ e^{i\frac{\pi}{2}} & q_{\text{cuton}} < q \leq q_{\text{cutoff}}, \\ 0 & q > q_{\text{cutoff}}, \end{cases} \quad (\text{F.70})$$

where $q_{\text{cutoff}} = 2\pi d_{\text{ap}}/(f\lambda)$ is the cut-off frequency, d_{ap} is the physical diameter of the aperture, f is the focal length of the objective lens, and q_{cuton} is the cut-on frequency of the phase plate. If only the objective aperture is considered then $q_{\text{cuton}} = q_{\text{cutoff}}$.

The contrast transfer function (CTF) of the lens system is [111]

$$T(q, \alpha) = K_s(q, \alpha) K_c(q) A_p(q) e^{-i\chi(q, \alpha)}. \quad (\text{F.71})$$

The Fourier transform of the electron wave at the back focal plane is given by

$$\tilde{\Psi}(q, \alpha) = \mathcal{F}[\Psi_{\text{exit}}(x, y)] T(q, \alpha). \quad (\text{F.72})$$

Finally, the intensity in the image plane is the probability density function given by

$$I_0(x, y) = |\Psi(x, y)|^2. \quad (\text{F.73})$$

F.4 Detector response

The realization of the final image involves the conversion of electron wave function into a digital signal *via* detection by a camera. Such a camera is characterized by several properties such as the conversion factor of the camera CF in $[\text{ADU}/e^-]$, various noise sources, as well as frequency response measures such as the modulation transfer function (MTF) and detective quantum efficiency (DQE). The measurement process obeys Poisson statistics which gives rise to shot noise, adds readout noise I_{m} and dark current I_{dc} to the final image, and blurs the image with a detector point-spread function PSF(x, y)

$$I(x, y) = [CF \cdot P_{\text{ois}}(N_{\text{tot}} \cdot I_0(x, y))] * \text{PSF}(x, y) + I_{\text{m}} + I_{\text{dc}} \quad (\text{F.74})$$

where $P_{\text{ois}}(A)$ denotes Poisson noise yield, $N_{\text{tot}} \cdot I_0(x, y)$ is the incident integrated electron flux in $\left[\frac{e^-}{\text{area}}\right]$, and $*$ represents the 2D convolution operator.

Poisson (shot) noise is related to the uncertainty associated with the measurement of electron current (or light), inherent to the discrete nature of electrons and the independence of electron detections. Its variance is signal-dependent and constitutes the dominant source of image noise. For higher integrated electron fluxes the Poisson distribution approaches a normal (Gaussian) distribution. However, for cryo microscopy the integrated electron flux is kept very low in order to avoid radiation damage and to preserve the structural details of the specimen. Therefore, it is important to model the noise *via* a Poisson distribution. The signals are usually captured with pixelized detectors (CCD or CMOS cameras) with pixel sizes Δx and Δy . The probability that

a pixel i will detect an electron is therefore $P_i = P(x_i, y_i) = |\Psi(x, y)|^2 \Delta x_i \Delta y_i$ (see *Intermezzo 4* for more details).

Without loss of generality let us consider only Poisson noise, which is the most dominant noise source in the detection process. Then the final realization of the intensity would be

$$I(x, y) = \mathcal{F}^{-1}\{\mathcal{F}[CF \cdot P_{ois} (N_{tot} \cdot I_0(x, y))] \cdot \text{MTF}(q)\}. \quad (\text{F.75})$$

The MTF describes how the signal amplitude is transferred for different spatial frequencies. It is the modulus of the Fourier transform of the detector's PSF. Attenuation from the MTF alone would not spoil the image quality. If the signal is transferred up to Nyquist frequency and the MTF is known, one can, in theory, restore the image by deconvolution. In practice, deconvolution will be hampered by the presence of noise. The DQE describes the noise added by the detector. The noise of a stochastic scattering process (like the one in TEM detectors) is not transferred in the same way as the signal [75]. The noise in the detected signal is not simply the noise in the input signal attenuated by the MTF [20]. The detective quantum efficiency is defined as the squared ratio of the signal-to-noise ratio (SNR) between output and input signal

$$\text{DQE}(q) = \left(\frac{\text{SNR}_{\text{out}}}{\text{SNR}_{\text{in}}} \right)^2. \quad (\text{F.76})$$

In [47] it was shown that this is equivalent to

$$\text{DQE}(q) = CF^2 N_{\text{tot}} \frac{\text{MTF}^2(q)}{\text{NPS}_{\text{out}}(q)}, \quad (\text{F.77})$$

where NPS refers to the noise power spectrum. If we further define the noise transfer function as $\text{NTF}^2(q) = \text{NPS}_{\text{out}} / (CF^2 N_{\text{tot}})$ then DQE is simply

$$\text{DQE}(q) = \frac{\text{MTF}^2(q)}{\text{NTF}^2(q)}. \quad (\text{F.78})$$

The problem we are facing is how to treat the signal and noise transfer separately since the Poisson (shot) noise depends on the signal. We decouple them in the following way: 1) the Fourier spectrum of the noise-free signal ($\tilde{I}_0(q) = \mathcal{F}[I_0(x, y)]$) is damped (multiplied) by the ratio between signal (MTF) and noise (NTF) transfer, 2) this signal is multiplied with the integrated electron flux and all noise contributions are added, 3) the Fourier spectrum of that (noisy) signal is damped by the NTF, and 4) the number of electrons are scaled with CF to the image gray values in [ADU]. Hence, we can write the detected image as

$$\tilde{I}_0(q) = \mathcal{F}[I_0(x, y)]$$

$$\begin{aligned} I_1(x, y) &= \mathcal{F}^{-1}\left[\tilde{I}_0(q) \frac{\text{MTF}(q)}{\text{NTF}(q)}\right] \\ I(x, y) &= \mathcal{F}^{-1}\{\mathcal{F}[CF \cdot P_{ois} (N_{tot} \cdot I_1(x, y))] \cdot \text{NTF}(q)\}, \end{aligned} \quad (\text{F.79})$$

or everything combined

$$I(x, y) = \mathcal{F}^{-1}\{\mathcal{F}[CF \cdot P_{oiss}(N_{\text{tot}} \cdot \mathcal{F}^{-1}[\mathcal{F}[I_0(x, y)] \sqrt{\text{DQE}(q)}]) \cdot \text{NTF}(q)] + I_{\text{rn}} + I_{\text{dc}}\}. \quad (\text{F.80})$$

----- *Intermezzo 4* -----

Poisson distribution of noise

Assuming that the total number of detected electrons N_{tot} is Poisson distributed, we will show that the intensity within pixel i follows the Poisson distribution, as well.

Let us assume for the moment that there is no uncertainty in the number of detected electrons (no Poisson noise). The probability that a subset of k electrons (from exactly N_{tot}) will be detected by the pixel i , while the remaining $N_{\text{tot}} - k$ will be detected by other pixels is $P_i^k (1 - P_i)^{N_{\text{tot}} - k}$. There are $\binom{N_{\text{tot}}}{k}$ combinations of k subsets of N_{tot} . The total probability of any subset of k from N_{tot} electrons detected by the pixel i is the well-known binomial distribution

$$P_i^{N_{\text{tot}}(k)} = \binom{N_{\text{tot}}}{k} P_i^k (1 - P_i)^{N_{\text{tot}} - k}. \quad (\text{F.81})$$

The expected value is $\langle k \rangle = N_{\text{tot}} P_i$ and variance $\text{Var}(k) = N_{\text{tot}} P_i (1 - P_i)$. Therefore, if exactly N_{tot} electrons would be detected, we would need to calculate the probability density function N_{tot} times and on average $N_{\text{tot}} P_i$ electrons will be detected by pixel i with a binomial distribution of the values around it. This allows us to calculate the probability density function of N_{tot} electrons in one run by scaling the pixels values as $N_{\text{tot}} P_i$ instead of calculating the probability density function N_{tot} times, i.e. $I^{N_{\text{tot}}}(x_i, y_i) = \text{Binom}(N_{\text{tot}}, |\Psi(x, y)|^2)$.

In reality, we do not know the exact number of electrons n falling on the detector, only that they obey a Poisson distribution with expected value of $\langle n \rangle = N_{\text{tot}}$, i.e. $P(n, N_{\text{tot}}) = \frac{N_{\text{tot}}^n}{n!} e^{-N_{\text{tot}}}$. Similar to [218], the probability of detecting exactly k electrons by the pixel i is now the sum of all conditional probabilities that $n > k$ electrons will fall on the detector and k electrons will be detected by the pixel i

$$\begin{aligned} P_i(k|\langle n \rangle = N_{\text{tot}}) &= \sum_{n=k}^{\infty} P(n, N_{\text{tot}}) \binom{n}{k} P_i^k (1 - P_i)^{n-k} \\ &= \sum_{n=k}^{\infty} \frac{N_{\text{tot}}^n e^{-N_{\text{tot}}}}{n!} \frac{n!}{k!(n-k)!} P_i^k (1 - P_i)^{n-k} \\ &= \frac{P_i^k}{k!} e^{-N_{\text{tot}}} \sum_{n=k}^{\infty} \frac{N_{\text{tot}}^n}{(n-k)!} (1 - P_i)^{n-k} \end{aligned}$$

Introducing a variable $n' = n - k$ we have

$$P_i(k|\langle n \rangle = N_{\text{tot}}) = \frac{P_i^k}{k!} e^{-N_{\text{tot}}} \sum_{n'=0}^{\infty} \frac{N_{\text{tot}}^{n'} N_{\text{tot}}^k}{(n')!} (1 - P_i)^{n'} \quad (\text{F.82})$$

The sum represents now the Taylor expansion of an exponential function and therefore we

get

$$P_i(k|n) = N_{\text{tot}} = \frac{P_i^k N_{\text{tot}}^k}{k!} e^{-N_{\text{tot}}} e^{N_{\text{tot}}(1-P_i)} = \frac{(P_i N_{\text{tot}})^k}{k!} e^{-P_i N_{\text{tot}}}. \quad (\text{F.83})$$

Therefore, instead of calculating n times our probability density function $|\Psi(x, y)|^2$, we can realize the intensity within the pixel i as Poisson distributed with the expected value of $P_i N_{\text{tot}}$, i.e. $I^{(n)=N_{\text{tot}}}(x_i, y_i) = P_{\text{ois}}(N_{\text{tot}} * |\Psi(x, y)|^2)$. The Poisson noise is present in every detection system and cannot be avoided.

F.5 Fresnel diffraction from a phase step

In this section the analytical solution for Fresnel diffraction from a semi-infinite phase step is derived and compared to a numerical solution. In the first approximation, a carbon edge can be considered as a semi-infinite phase step in the object plane:

$$\Psi(x, y, \Delta z = 0) = \begin{cases} e^{i\Delta\phi} & x < 0, \\ 1 & x \geq 0, \end{cases} \quad (\text{F.84})$$

where $\Delta\phi$ represents the phase step determined *via* equation (F.95). In order to express the analytical solution for this case, let us consider the Helmholtz equation (which can be derived by setting $\nabla^{\text{int}}(\mathbf{r}) = 0$ in equation (F.50)):

$$\left[\nabla_r^2 - k^2 \right] \Psi = 0. \quad (\text{F.85})$$

The solution of (F.85) is given by the first Rayleigh-Sommerfeld diffraction integral [219], a mathematical formulation of the Huygens-Fresnel principle:

$$\Psi(x, y, \Delta z) = -\frac{i}{\lambda} \int_{x', y'} \Psi(x', y', 0) \frac{e^{ikr}}{r} \cos \theta dx' dy', \quad (\text{F.86})$$

where θ is the angle between the Δz axis and the vector $\mathbf{r} = (x - x', y - y', \Delta z)$, i.e. $\cos \theta = \Delta z / r$. From the small angle approximation we have $\Delta z \gg \sqrt{(x - x')^2 + (y - y')^2}$ and

$$r = \sqrt{(x - x')^2 + (y - y')^2 + \Delta z^2} = \Delta z \sqrt{1 + \frac{(x - x')^2 + (y - y')^2}{\Delta z^2}} \approx \Delta z + \frac{(x - x')^2 + (y - y')^2}{2\Delta z}. \quad (\text{F.87})$$

Substituting r in equation (F.86), we get the Fresnel propagator integral formulation which models the spherical waves from equation (F.86) as parabolic waves:

$$\Psi(x, y, \Delta z) = -\frac{i}{\lambda} \frac{e^{ik\Delta z}}{\Delta z} \int_{x', y'} \Psi(x', y', 0) e^{\frac{ik[(x-x')^2 + (y-y')^2]}{2\Delta z}} dx' dy'. \quad (\text{F.88})$$

In case of a semi-infinite opaque screen, the analytical solution can be obtained from the following Fresnel diffraction integral:

$$\Psi(x, y, \Delta z) = -\frac{i}{\lambda} \frac{e^{ik\Delta z}}{\Delta z} \int_{-\infty}^{\infty} dy' \int_0^{\infty} e^{\frac{ik[(x-x')^2 + (y-y')^2]}{2\Delta z}} dx', \quad (\text{F.89})$$

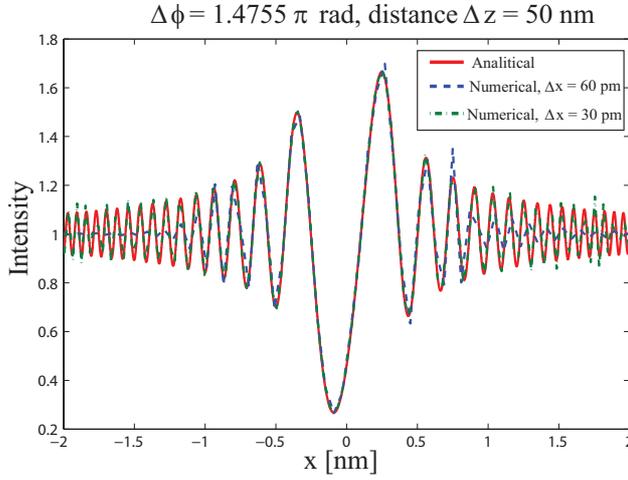


Fig. F.2. Intensity profiles of the analytical and numerical solutions of Fresnel diffraction from a phase step. Red solid curve shows the analytical solution implemented in Maple. The blue dashed and green dash-dotted line represent the discrete fast-Fourier transform solutions implemented in MATLAB for sampling distances of 60 and 30 pm, respectively. The phase step of 1.48π rad represents the phase change introduced by an edge of 50 nm thick amorphous carbon ($\rho = 2.3 \text{ g/cm}^3$) imaged at 200 keV (see equation (F.95))

By using the Fresnel sine \mathcal{S} and cosine \mathcal{C} integrals [220]

$$\mathcal{S}(w) = \int_0^w \sin\left(\frac{\pi}{2}w'^2\right) dw' \quad \mathcal{C}(w) = \int_0^w \cos\left(\frac{\pi}{2}w'^2\right) dw', \quad (\text{F.90})$$

defining $u \equiv (x-x')\sqrt{k/(\pi\Delta z)} = (x-x')\sqrt{2/(\lambda\Delta z)}$ and $v \equiv (y-y')\sqrt{k/(\pi\Delta z)} = (y-y')\sqrt{2/(\lambda\Delta z)}$ and using the properties $\mathcal{S}(-w) = -\mathcal{S}(w)$, $\mathcal{C}(-w) = -\mathcal{C}(w)$, and $\mathcal{S}(\infty) = \mathcal{C}(\infty) = 1/2$ we get

$$\begin{aligned} \Psi_{\text{semiop}}(x, y, \Delta z) &= -\frac{i}{\lambda} \frac{e^{ik\Delta z}}{\Delta z} \frac{\lambda\Delta z}{2} \int_{-x\sqrt{\frac{2}{\lambda\Delta z}}}^{\infty} e^{i\frac{\pi}{2}u^2} du' \int_{-\infty}^{\infty} e^{i\frac{\pi}{2}v'^2} dv', \\ &= -\frac{ie^{ik\Delta z}}{2} \left[\mathcal{C}(\infty) + i\mathcal{S}(\infty) - \mathcal{C}\left(-x\sqrt{\frac{2}{\lambda\Delta z}}\right) - i\mathcal{S}\left(-x\sqrt{\frac{2}{\lambda\Delta z}}\right) \right] [\mathcal{C}(\infty) + i\mathcal{S}(\infty) - \mathcal{C}(-\infty) - i\mathcal{S}(-\infty)], \\ &= -\frac{i}{2} e^{ik\Delta z} \left[\frac{1}{2} + i\frac{1}{2} + \mathcal{C}\left(x\sqrt{\frac{2}{\lambda\Delta z}}\right) + i\mathcal{S}\left(x\sqrt{\frac{2}{\lambda\Delta z}}\right) \right] \left[\frac{1}{2} + i\frac{1}{2} + \frac{1}{2} + i\frac{1}{2} \right], \\ &= -\frac{i(1+i)}{2} e^{ik\Delta z} \left[\frac{1+i}{2} + \mathcal{C}\left(x\sqrt{\frac{2}{\lambda\Delta z}}\right) + i\mathcal{S}\left(x\sqrt{\frac{2}{\lambda\Delta z}}\right) \right]. \end{aligned} \quad (\text{F.91})$$

The semi-infinite phase step can be represented as a combination of two semi-infinite opaque screens placed in the opposite x direction and shifted with a phase $e^{i\Delta\phi}$:

$$\Psi_{\text{phstep}}(x, y, \Delta z) = \Psi_{\text{semiop}}(x, y, \Delta z) + \Psi_{\text{semiop}}(-x, y, \Delta z)e^{i\Delta\phi}. \quad (\text{F.92})$$

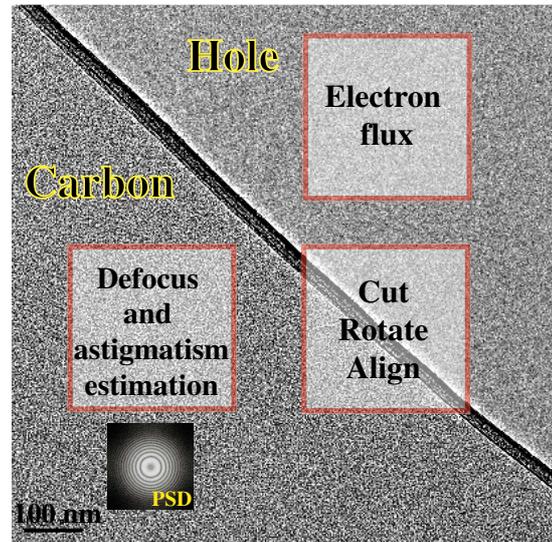
The intensity is defined as:

$$I_{\text{phstep}}(x, y, \Delta z) = |\Psi_{\text{phstep}}(x, y, \Delta z)|^2. \quad (\text{F.93})$$

This is the analytic solution for Fresnel diffraction from semi-infinite phase step. The discrete solution is based on the Fourier transform:

$$\Psi_{\text{phstep}}(x, y, \Delta z) = \mathcal{F}^{-1}[P(q, \Delta z)\mathcal{F}(\Psi(x, y, \Delta z = 0))], \quad (\text{F.94})$$

Fig. F.3. A carbon hole in a Quantifoil grid imaged at 2000 nm underfocus. The integrated flux can be estimated from intensity in the hole and given the magnification, while defocus and astigmatism can be estimated from the amorphous carbon area by calculating the power spectrum density (PSD) and using the toolbox provided in [49]. For defocus series reconstruction, edge images were cut and aligned.



where $P(q, \Delta z) = \exp(-i\pi\lambda\Delta zq^2)$ is the Fresnel propagator. Fig. F.2 allows the intensity comparison of analytical solution with numerical solutions for two sampling densities. In the discrete case, the sampling Δx of the edge should be much finer than the argument of the Fresnel sine and cosine integrals i.e. $\Delta x \ll \sqrt{2/(\lambda\Delta z)}$ for this idealized non-band limited object.

F.6 Evaluation of the forward model on images of carbon edges and carbon nanotubes

F.6.1 Modeling carbon edges and carbon nanotubes

Although the main purpose of InSilicoTEM is to simulate the images of biological samples embedded in vitreous ice, we also model two non-biological specimens: carbon edges and carbon nanotubes with buckyballs. The images of these specimens can be recorded at a high integrated flux providing a high SNR. The averaged intensity profiles across the carbon edges and nanotubes further increase the SNR and provide a better comparison for validating the simulations. The profile of carbon edges was modeled as an error function or was reconstructed from a defocus series; nanotubes with buckyballs were modeled *via* IASA in vacuum.

Carbon edges

Specimen grids in TEM usually contain a carbon film support because of their relatively low background signal and good electrical conductivity. An advantage of imaging support films for the purpose of validating simulations is that no additional sample preparation is required. While imaging carbon edges, the defocus and astigmatism can be estimated from the amorphous carbon area and the integrated electron flux can be estimated from the hole at the same time (see Fig. F.3). The mean inner potential of an amorphous specimen such as carbon film introduces a constant phase change of the electron wave which can usually be neglected as it is frequency

independent. In this case it is important to include the phase shift since we model the carbon-vacuum phase transition *via* an error function $\text{erf}(x)$ like $\frac{\Delta\phi}{2}[\text{erf}(\frac{x}{\sigma_{\text{err}}}) + 1]$. The analytical and discrete solutions when the transition is modeled via a step edge are compared in section F.5. This is an example of the forward model where no atomic model is required. The thickness-dependent phase shift through the carbon $\Delta\phi$ can be determined *via* [221]

$$\Delta\phi = \sigma V_z = \sigma V_0 d, \quad (\text{F.95})$$

where σ is the interaction constant depending on the voltage, V_0 the mean inner potential, and d the thickness. Following density-functional theory calculations [222] and holographic measurements [221], the mean inner potentials V_0 of amorphous carbon ($\rho = 1.8 \text{ g/cm}^3$) and graphite ($\rho = 2.3 \text{ g/cm}^3$) are 10.1 V and 12.7 V, respectively.

The profile of carbon edges in practise is usually much more complex due to manufacturing issues, cracking and folding of the carbon film, as well as wrapping under the electron beam. The model of an edge in this case is not a simple function. Another approach is to estimate the projected potential profile of such a complex edge *via* Wiener filtering applied to a defocus series [2]. The profile F is estimated from N images at different defoci as

$$\tilde{F}(q) = \sum_{i=1}^N \frac{\text{CTF}_i^*(q)\tilde{I}_i(q)}{\sum_{i=1}^N |\text{CTF}_i(q)|^2 + \lambda_w(q)} \quad (\text{F.96})$$

where CTF_i is the CTF for image I_i corresponding to equation (F.71) and $\lambda_w(q)$ the regularization factor which was assumed to be constant ($\lambda_w = 0.01$). The reconstructed edge was further used as the input for the simulations.

Carbon nanotubes with buckyballs

Carbon nanotubes containing spherical C60 fullerenes were modeled as another example that allowed high-flux imaging and comparison between simulations and experiment at a higher SNR. An atomic layer of carbon was rolled up in zigzag configuration with (16,0) chirality and 0° chiral angle. Furthermore, the Buckminsterfullerenes (buckyballs) with formula C60 are truncated icosahedrons, with a carbon atom at each vertex of each polygon. The buckyballs were placed along the nanotube with a distance of 1 nm from each other. The coordinates of the system composed of nanotubes and C60 buckyballs were used to construct an artificial PDB file consisting of only carbon atoms. Finally, the interaction potential was generated by IASA, without any solvent.

F.6.2 Experimental methods

Images of carbon edges were recorded using a Tecnai F20 electron microscope (FEI, The Netherlands) equipped with a GIF energy filter and FEG operated at 200 kV. Other microscope settings were: condenser and objective aperture size of $100 \mu\text{m}$, spherical (C_s) and chromatic (C_c) aberrations for this microscope are both 2.0 mm, while energy spread (ΔE) and illumination aperture (α_i) are 0.7 eV and 0.1 rad, respectively. Defocus series of a thin graphite edge

and a holey Quantifoil (R 2/2) carbon edge (requested defocus range 1-8 μm in nine steps) were acquired. The final sampling densities in the object plane were 1.85 $\text{\AA}/\text{pixel}$ and 2.29 $\text{\AA}/\text{pixel}$, while the estimated incident integrated flux was $\sim 590 \text{ e}^-/\text{\AA}^2$ and $\sim 43 \text{ e}^-/\text{\AA}^2$ for the thin graphite and Quantifoil edge, respectively.

Carbon nanotubes were obtained by a method described in [223]. The samples were solubilized with sonication in ultra clean acetone and brought onto glow discharged C-flats. Defocus pair images ($\Delta f = 35 \text{ nm}$ and $\Delta f = 70 \text{ nm}$) of nanotubes with buckyballs were collected on a 4k x 4k Falcon direct electron detector (FEI, The Netherlands) using a Titan electron microscope (FEI, The Netherlands) equipped with a C_s corrector and a GIF energy filter and FEG operated at 80 kV. Some microscope parameters such as aperture sizes, C_c , α_i , ΔE are described in section 4.4.2 while $C_s \sim 5 \mu\text{m}$. A magnification at the detector plane of 253 kx produced a sampling density in the object plane of 0.553 $\text{\AA}/\text{pixel}$. The estimated flux was $\sim 112 \text{ e}^-/\text{\AA}^2$.

F.6.3 Results

Images of carbon edges and carbon nanotubes were acquired with a high integrated flux and their averaged intensity profiles were used to further quantify the comparisons between experiments and simulations.

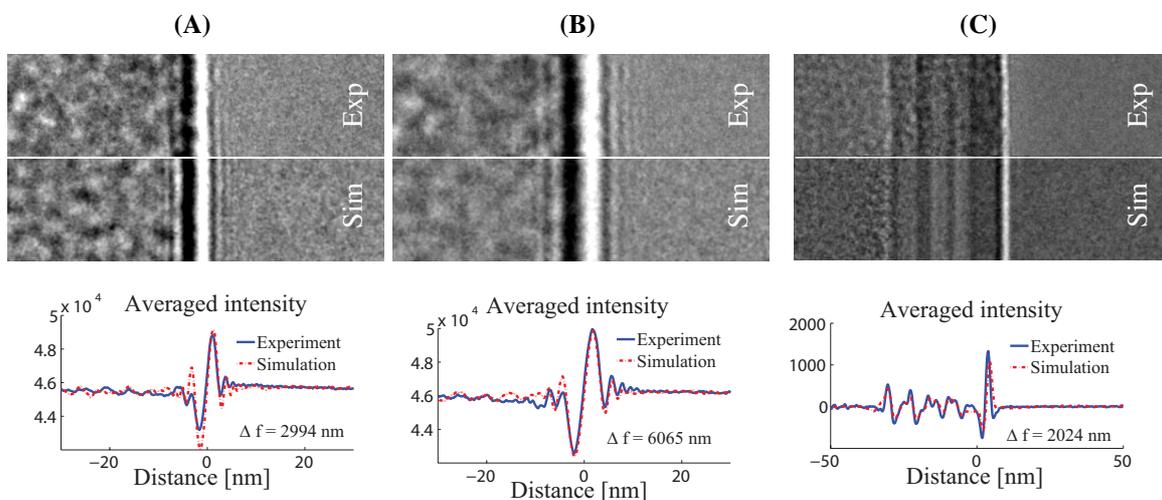


Fig. F.4. Experimental and simulated images of carbon edges. (A) and (B) image of a graphite flake edge at defoci of 2994 and 6065 nm, respectively. The edge is modeled as an error function with $\sigma_{\text{err}} = 1.85 \text{ \AA}$. The sampling density is 1.85 $\text{\AA}/\text{pixel}$, the incident integrated flux $\sim 590 \text{ e}^-/\text{\AA}^2$ and the thickness $d = 2.5 \text{ nm}$. (C) Experimental and simulated image of a Quantifoil carbon edge after defocus series reconstruction of the projected edge profile. Defocus is $\Delta f = 2024 \text{ nm}$, the sampling density 2.29 $\text{\AA}/\text{pixel}$, and the incident integrated flux $\sim 43 \text{ e}^-/\text{\AA}^2$. Note that very low frequencies of the projected profile could not be reconstructed *via* the Wiener approach and therefore, for the comparisons, all frequencies lower than 0.04 nm^{-1} were removed from the image. The averaged intensities along the edges over a distance of 18 nm are shown in the bottom row.

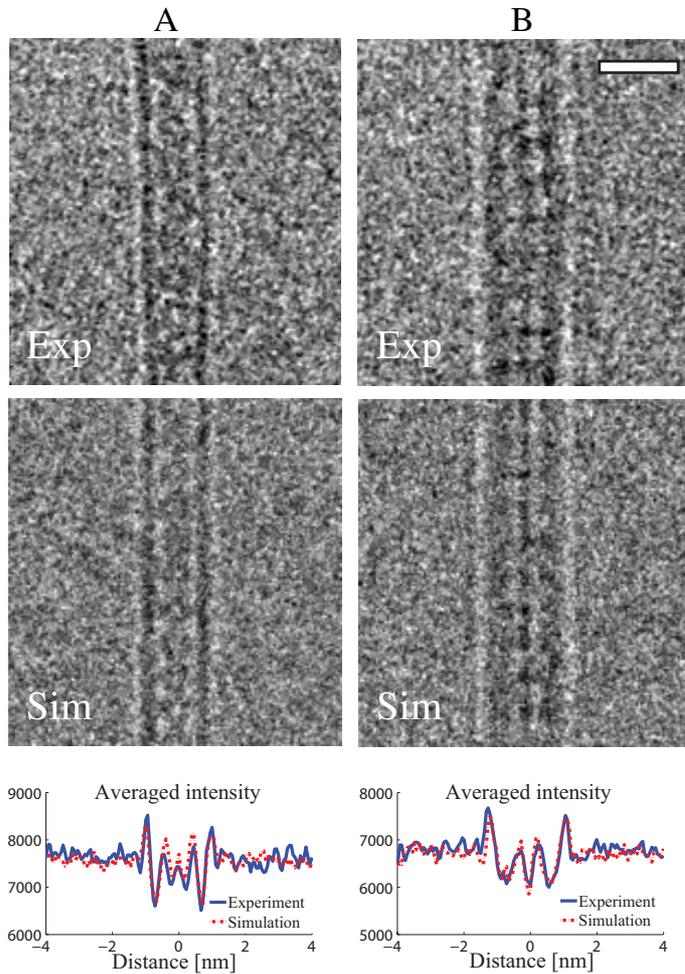


Fig. F.5. Experimental and simulated images of a carbon nanotube with fullerenes at 80 kV. **(A)** and **(B)** are the images and averaged profiles at defoci of 35 and 70 nm, respectively. The carbon nanotube was tilted by 30° . The estimated flux was $\sim 112 \text{ e}^-/\text{\AA}^2$. The profiles were averaged along the bottom half of the image making sure that both experimental and simulated profiles have the same number of C60 fullerenes included. The scale bar corresponds to 2 nm.

Carbon edges

Figs. F.4.A and F.4.B show the simulated and experimental images of a graphite flake edge whose profile was modeled as an error function (see section F.6.1). Additionally, the averaged profiles along the edge over a distance of 18 nm are presented. The amorphousness of the carbon was modeled as a fixed noise pattern added to the projections (see section 4.3.1). The estimated defoci were 2994 nm and 6065 nm, respectively. Furthermore, a defocus series of a Quantifoil carbon edge was acquired at an integrated flux of $44 \text{ e}^-/\text{\AA}^2$ estimated from a hole (Fig. F.4.C). The estimated defoci were 1005 nm, 2024 nm, 3068 nm, 4097 nm, 6195 nm, respectively. The profile of Quantifoil edge proved to be difficult to represent by a simple model such as an error function. After reconstructing the projected edge profile *via* Wiener filtering applied to a defocus series (see equation (F.96)), a simulated image of the profile at defocus of 2024 nm is compared to the experimental image (see Fig. F.4.C). The intensity variations across the carbon edges in the simulated and experimental images are very similar.

Carbon nanotubes

Defocus pair images ($\Delta f = 35$ nm and $\Delta f = 70$ nm) of carbon nanotubes with spherical C60 fullerenes were collected and a single nanotube was extracted for this analysis (see Fig. F.5). The diameter of the nanotube is 1.5 nm and the expected distance between C60 fullerenes (from untilted view and sample preparation) is 1 nm. The projected distance between C60 fullerenes in the extracted nanotube was 0.86 nm, suggesting that the nanotube was tilted by 30° with respect to the focal plane, which was also considered in this model. Defocus was determined by simulating a defocus series with a step size of 5 nm close to the requested defocus value and comparing it with the experimental image pair. The averaged profiles show that the experimental and the simulated images have similar intensity variations across the nanotubes (Fig. F.5).

Appendix G

Projection assumption and weak-phase object approximation in cryo-EM

Submitted as [52]: M. Vulović, L. M. Voortman, L. J. van Vliet, B. Rieger, “When to use the projection assumption and weak-phase object approximation in phase contrast cryo-EM”

Abstract

The projection assumption (PA) and the weak-phase object approximation (WPOA) are commonly used to model image formation in cryo-electron microscopy. For making the next step in resolution improvement we show that it is important to revisit these two approximations as well as their limitations. Here we start off inspecting both approximations separately to derive their respective conditions of applicability. The thick-phase grating (TPG) model imposes less strict conditions on the interaction potential than PA or WPOA and gives comparable exit waves as a multislice calculation. We suggest the ranges of applicability for four models (PA, PA+WPOA, WPOA and TPG) given different interaction potentials using exit wave simulations. The conditions for applicability are based on two measures delivering a worst-case (safest) and an average criterion. This allows us to present a practical guideline for when to use each image formation model depending on the spatial frequency, thickness and strength of the interaction potential of a macromolecular complex.

G.1 Introduction

Quantitative forward modeling of image formation and the simulation of images is becoming increasingly important in order to optimize the data acquisition strategy, facilitate reconstruction schemes, improve image interpretation and resolution, and provide insight into ways to improve instrumentation. An accurate description of the interaction between incident electrons and the specimen is one of the important steps in forward modeling, contrast transfer function (CTF) correction and 3D reconstruction in cryo-electron microscopy (cryo-EM).

In cryo-EM, incident electrons with typical energies of 80-300 keV interact with the electrostatic interaction potential (IP) of the specimen, e.g. macromolecules that are similar in density

to the surrounding vitreous ice. In order to describe the electron-specimen interaction (analytically) two approximations are often made: the weak-phase object approximation (WPOA) and the projection assumption (PA). The WPOA holds for weakly scattering objects [2] and the PA assumes that the exit wave from the specimen can be computed *via* the projected IP of the whole specimen [144]. Both approximations rely on the small angle approximation [205] and are frequently used at the same time. Applying both approximations greatly simplifies the computational complexity of forward modeling and 3D reconstruction and therefore they have been implemented in most software packages for single particle analysis (SPA) and electron tomography (ET) [31, 33, 35, 36, 38, 39].

These approximations have, of course, limitations as they cannot account for e.g. the curvature of the Ewald sphere or multiple scattering events [111]; effects which become more critical for high resolution imaging. In materials science high resolution electron microscopy (HREM), where atomic resolution is attained on certain specimens, a multislice calculation [144] is commonly used to overcome the limitations of the aforementioned approximations in modeling the transmission of the electron wave through the specimen. There, the specimen is divided into slices and propagation of the electron wave can be interpreted as a successive transmission and propagation through each slice until the wave leaves the specimen. The PA must hold within each slice and therefore, it is also important to formulate a quantitative criterion to determine the appropriated slice thickness. The multislice approach has been rarely used in cryo-EM [51, 142], mainly because of the lower resolution of cryo-EM compared to HREM. Due to the need for higher resolution in cryo-EM, it is important to revisit the WPOA and PA and investigate their applicability.

The thick-phase grating (TPG) approximation was introduced in HREM of perfect crystals in 1962 [224], but to the best of our knowledge, it has not received much attention since [225]. Furthermore, a rough indication was provided for the validity of various approximations depending on the thickness and atomic number of the crystals. Here, we introduce TPG to the field of cryo-EM and discuss its potential benefits. We provide practical boundaries to various approximations based on the thickness, strength and frequency of the interaction potential map.

G.2 High-energy electron and specimen interaction

To discuss the validity of the PA and WPOA it is convenient to start from the stationary one-body Schrödinger equation with a correction for the relativistic mass and wavelength of the electron. This is permitted for elastic scattering processes as *inter alia* i) the Hamiltonians of the electron and the specimen can be separated because the incident electrons have a much higher energy than the interaction energy of the particles within the specimen, ii) spin-spin interactions may be neglected, and iii) the electron current in cryo-EM is so low that effectively only one electron interacts with the specimen at the time, which guarantees independence of all incident electrons. Below we will shortly recapitulate the formulae commonly used in HREM [144, 225].

G.2.1 Small angle approximation

The stationary one-body Schrödinger equation for the electron wave function in a closed system is given by

$$\left(-\frac{\hbar^2}{2m}\nabla_{\mathbf{r}}^2 + e\mathcal{V}(\mathbf{r})\right)\psi_e(\mathbf{r}) = E_e\psi_e(\mathbf{r}), \quad (\text{G.1})$$

where $-\hbar^2/(2m)\nabla_{\mathbf{r}}^2$ is the Hamiltonian of the incident high-energy electron, which in this case represents its kinetic energy, $\mathcal{V}(\mathbf{r})$ the interaction potential, \hbar the reduced Planck constant, m the relativistic mass of the electron, e the electron charge, $\mathbf{r} = (x, y, z) = (\boldsymbol{\rho}, z)$ the position, ψ_e the electron wave function, and E_e the energy of the incident electron.

The incident electron travels (spirals) predominately along the optical axis, i.e. the z -direction. The specimen constitutes a relatively small perturbation to this motion. Therefore the total electron wave function $\psi_e(\mathbf{r})$ can be written as a product of a plane wave traveling in the z -direction and a wave function Ψ which varies slowly with z , i.e. $\psi_e(\mathbf{r}) = \Psi(\mathbf{r})e^{ikz}$, with the wave vector $k = 2\pi/\lambda = \sqrt{2mE_e}/\hbar$, and λ the wavelength. Now it follows from Eq. (G.1)

$$\left(\nabla_{\boldsymbol{\rho}}^2 + \partial_z^2 + 2ik\partial_z - \frac{2me}{\hbar^2}\mathcal{V}(\mathbf{r})\right)\Psi(\mathbf{r}) = 0. \quad (\text{G.2})$$

Given the assumptions that the energy of the incident electron is high and that Ψ varies slowly with z , it holds that $|\partial_z^2\Psi| \ll |k\partial_z\Psi|$ and $k^2 \gg k_x^2 + k_y^2$, which is known as the small angle approximation. With the definition of the interaction constant $\sigma = \lambda me/(2\pi\hbar^2)$, this leads to

$$\partial_z\Psi(\mathbf{r}) = \left(\frac{i\lambda}{4\pi}\nabla_{\boldsymbol{\rho}}^2 + i\sigma\mathcal{V}(\mathbf{r})\right)\Psi(\mathbf{r}). \quad (\text{G.3})$$

Taking the 2D Fourier-transform in $\boldsymbol{\rho} = (x, y)$ we get our common starting point for all further approximations

$$\partial_z\mathcal{F}_{\boldsymbol{\rho}}[\Psi] = -i\lambda\pi q^2\mathcal{F}_{\boldsymbol{\rho}}[\Psi] + i\sigma\mathcal{F}_{\boldsymbol{\rho}}[\mathcal{V}\Psi], \quad (\text{G.4})$$

in which the Fourier-transform is defined as $\mathcal{F}_{\boldsymbol{\rho}}[f(\boldsymbol{\rho})](\mathbf{q}) = \int f(\boldsymbol{\rho})e^{-2\pi i\boldsymbol{\rho}\mathbf{q}}d\boldsymbol{\rho}$.

G.3 Bounds to projection assumption and weak-phase object approximation

To solve Eq. (G.4) analytically, further simplifications are needed. Two common approximations in cryo-EM are the projection assumption (PA) and the weak-phase object approximation (WPOA), where the latter is also known as kinematic approximation [10]. These two approximations lead to four different models describing the electron-specimen interaction. Below we will provide rules-of-thumb when to use each of these models.

Without loss of generality it is assumed that before the wave function Ψ is scattered by the potential \mathcal{V} it has a constant magnitude and zero phase. The magnitude of the incoming wave is conveniently set to 1. The scattered part of the wave function Ψ_{sc} is then given by $\Psi = 1 + \Psi_{sc}$.

Contrast in cryo-EM is formed predominately by phase contrast [111]. Because scattering by a constant \mathcal{V}_0 is identical to rescaling the wavelength, i.e. adding a constant phase factor to the incident electron wave, elastic scattering from the mean bulk potential does not contribute to contrast generation. Since we are interested in that part of the scattering process that produces contrast, we subtract the mean bulk potential. This is known as the quasi-kinematic approximation [10].

G.3.1 Projection assumption

When the specimen is sufficiently thin, the projection assumption (PA) is commonly used [144]. Then the propagation term of Eq. (G.3) is small compared to the interaction term, i.e. $|\frac{i\lambda}{4\pi}\nabla_\rho^2\Psi| \ll |i\sigma\mathcal{V}\Psi|$. From Eq. (G.3) it follows

$$\partial_z\Psi(\mathbf{r}) = i\sigma\mathcal{V}(\mathbf{r})\Psi(\mathbf{r}) \Rightarrow \Psi = \exp\left\{i\sigma\int_{-\infty}^z\mathcal{V}dz'\right\},$$

which leads to the exit wave

$$\Psi_{\text{exit}} = \exp\{i\sigma\mathcal{V}_z\}, \quad (\text{G.5})$$

with the projected potential $\mathcal{V}_z = \int_{-\infty}^{\infty}\mathcal{V}dz$. The validity of the PA was addressed by [204]. They argue that the potential should not vary significantly over a distance $d_r \geq \sqrt{\lambda\Delta z}/(2\pi)$, where Δz is the thickness of the specimen. Here, we will define a quantitative criterion for the validity of the assumption based on the Fresnel number. We define it in analogy to optics as $F = \Delta r^2/(\lambda\Delta z)$ [219], where Δr is the voxel size of the discretized potential map. Note that the regime $F \gg 1$ corresponds to ray optics and $F \geq 1$ to the small angle approximation. If we assume Nyquist sampling of the potential map, we have $q < 1/(2\Delta r)$ and the spatial frequencies up to which the projection assumption holds, is given by

$$q \ll \sqrt{1/(4\lambda\Delta z)}. \quad (\text{G.6})$$

In the above considerations there is no requirement for weak scattering. In this case, the absolute value of the potential is not relevant and the PA can also be valid for a strong-phase object. Note that the PA is also known as *phase-object approximation* [205, 225].

G.3.2 Projection assumption and weak-phase object approximation

If the scattering is weak, which is the case for most atoms in biological samples, the weak-phase object approximation (WPOA) $\sigma\mathcal{V}_z < 1$ can be used. When both PA and WPOA hold, Eq. (G.5) can be approximated by

$$\Psi_{\text{exit}} = 1 + i\sigma\mathcal{V}_z + \mathcal{O}(\sigma^2\mathcal{V}_z^2). \quad (\text{G.7})$$

Since $\sigma\mathcal{V}_z < 1$ leads to a scattered wave $\Psi_{sc} < 1$, the above result can also be obtained by substituting $\Psi = 1$ into the rhs. of Eq. (G.5) giving $\partial_z\Psi = i\sigma\mathcal{V}$. We will refer to Eq. (G.7) as PA+WPOA.

G.3.3 Weak-phase object approximation

The applicability of the WPOA depends on how well $\exp\{i\sigma\mathcal{V}_z\}$ can be approximated by a first order Taylor series expansion with $\mu = \sigma\mathcal{V}_z$,

$$e^{i\mu} = 1 + i\mu - \frac{1}{2}\mu^2 + \mathcal{O}(\mu^3). \quad (\text{G.8})$$

The relative residual in orders m or higher is given by

$$p(m, \mu) = e^{-\mu} \sum_{n=m}^{\infty} \frac{\mu^n}{n!}, \quad (\text{G.9})$$

where $e^{-\mu}$ normalizes the total sum $p(0, \mu)$ to 1. If we allow for a maximum of e.g. 5 % in second and higher order terms, we solve $p(2, \sigma\mathcal{V}) = 0.05$ to find

$$\sigma\mathcal{V}_z < 0.36. \quad (\text{G.10})$$

We will use this condition for applying the WPOA.

Note that Eq. (G.9) is identical to the probability of multiple scattering events described by a Poisson distribution with scattering probability $\mu = d/\Lambda$, in which d is the path length and Λ the mean free path [226]. This allows the interpretation of the different orders $\mathcal{O}(\sigma\mathcal{V}_z)$ as scattering events.

In a typical cryo-EM experiment, the macromolecular complex is embedded in vitreous ice whose thickness is larger than the thickness of the macromolecular complex. If we assume that vitreous ice is characterized by a bulk mean potential $\mathcal{V}_{\text{ice}} > 0$, the process of multiple scattering by a constant \mathcal{V}_{ice} can be neglected in the quasi-kinematic approach. Therefore, the condition given by Eq. (G.10) can only be applied to the mean-subtracted projected potential.

When the resolution of the potential map is too high to allow satisfying the PA condition, we can still use the WPOA. Furthermore, using only the WPOA results in an easy to implement algorithm for forward modeling. With the assumption $\sigma\mathcal{V}_z < 1$ or equally $\Psi_{sc} < 1$, Eq. (G.4) can be solved as follows

$$\begin{aligned} \partial_z \mathcal{F}_\rho[\Psi] &= -i\lambda\pi q^2 \mathcal{F}_\rho[\Psi] + i\sigma \mathcal{F}_\rho[\mathcal{V}] \Rightarrow \mathcal{F}_\rho[\Psi] = 1 + i\sigma \int_{-\infty}^z e^{-i\lambda\pi q^2 z'} \mathcal{F}_\rho[\mathcal{V}] dz' \\ \mathcal{F}_\rho[\Psi_{\text{exit}}] &= 1 + i\sigma \int \mathcal{V}(\boldsymbol{\rho}, z) e^{-2\pi i(\boldsymbol{\rho}\mathbf{q} + \frac{1}{2}\lambda q^2 z)} d\mathbf{r} \\ \Psi_{\text{exit}} &= 1 + i\sigma \mathcal{F}_\rho^{-1} \left[\mathcal{F}[\mathcal{V}] \left(\mathbf{q}, \lambda q^2 / 2 \right) \right]. \end{aligned} \quad (\text{G.11})$$

Here $\mathcal{F}[\mathcal{V}]$ is the 3D Fourier transform of the potential evaluated at coordinate $(\mathbf{q}, \lambda q^2 / 2)$, with $\mathbf{q} = (q_x, q_y)$. Sampling the 3D Fourier-transform on the parabola $(\mathbf{q}, \lambda q^2 / 2)$ can be done accurately and fast, as in [29].

G.3.4 Thick-phase grating approximation

The limitations of the PA and WPOA can be overcome by thick-phase grating approximation [224]. Initially developed for perfect crystals both with respect to diffraction and imaging, the

thick-phase grating approximation in cryo-EM gives the following forward model

$$\Psi_{\text{exit}} = \exp \left\{ i\sigma \mathcal{F}_\rho^{-1} \left[\mathcal{F}[\mathcal{V}] \left(\mathbf{q}, \lambda q^2/2 \right) \right] \right\}. \quad (\text{G.12})$$

The advantage of this combination of the conditions is that in the limit of $F \gg 1$, Eq. (G.12) converges to Eq. (G.5) and in the limit of $\sigma\mathcal{V} \ll 1$, Eq. (G.12) converges to Eq. (G.11). This means we get the corresponding image models of PA or WPOA directly from the above equation.

The approximations of Eqs. (G.5), (G.7) and (G.11) were derived in a similar way in [205]. Quantitative useful conditions for the validity of the approximations of Eqs. (G.10) and (G.6) are presented here. The advantages will be demonstrated below.

G.4 Results

G.4.1 Hemoglobin

Here we investigate the validity of the PA and WPOA for *Lumbricus terrestris erythrocrucorin* (earth worm hemoglobin - PDBid 2GTL) interacting with 80 keV electrons. This is a representative sample in terms of scattering power and size in cryo-EM. The interaction potential (IP) is computed as the sum of isolated atomic potentials. The atomic potential is calculated as the Fourier transform of the electron scattering factor which is parameterized as a weighted sum of five Gaussians [152]. All samples in this analysis are embedded in vitreous ice ($\rho = 0.93 \text{ g/cm}^3$). Detailed description of how the IP is constructed can be found elsewhere [51].

Fig. G.1 shows the validity of both approximations for this sample as a function of spatial frequency for various slice thicknesses. The graph shows the maximum value of the projected IP for a given slice thickness that we computationally extracted from the full IP. By doing so we can simulate the influence of the sample thickness and hereby indirectly the influence of the potential strength on the validity of the assumptions. The thickness of the slices was varied from 2.0 to 32.5 nm, eventually containing the entire specimen.

The values on the $\sigma\mathcal{V}_z$ -axis are calculated using the maximum projected potential of a slice extracted from the middle of the full map. We show one line for a potential map sampled at 1 Å (green) and one at 3 Å (blue) which are given by Eq. (G.6), i.e. the Fresnel number is equal to one. The uncertainty of the plotted values due to specimen orientation is depicted by the shaded area around the lines. Left/below of the respective lines the PA starts becoming suitable, whereas right/above it is violated. As given by Eq. (G.10), below the horizontal line $\sigma\mathcal{V}_z = 0.36$ the WPOA holds. For the full potential map sampled at 1 Å (green circle), neither PA nor WPOA hold, whereas for the potential map sampled at 3 Å (blue circle) the WPOA is satisfied and the PA is found to be right at the border. We see from Fig. G.1 that the criteria for WPOA and PA are easier fulfilled for low-frequency potential maps (e.g. the potential is blurred by beam-induced movements, CTF and/or the camera transfer). For comparison we show in Fig. G.1 the quasi-kinematic (QK) and the kinematic (K) potentials as circles and triangles, respectively. The kinematic potential represents the absolute strength of the potential, while

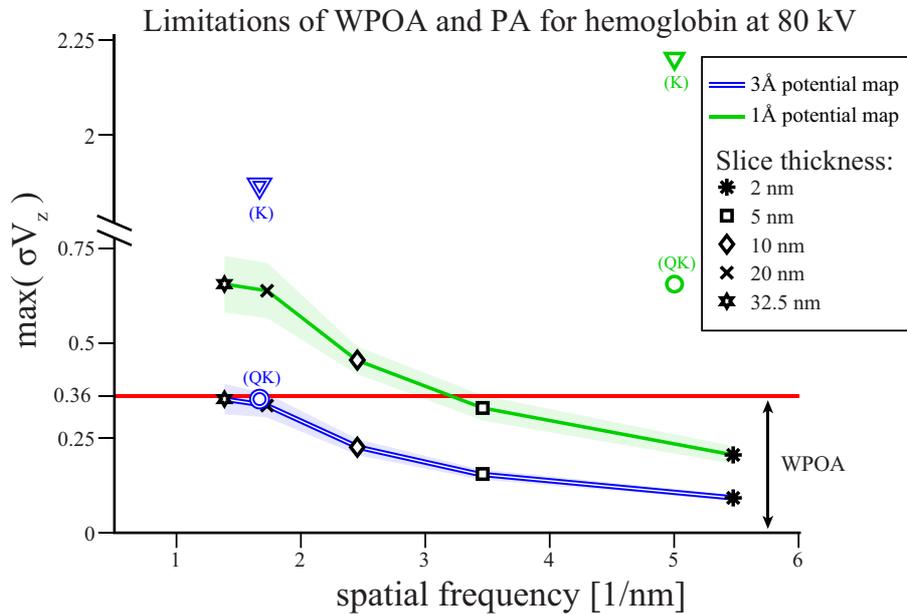


Fig. G.1. Validity of the PA and WPOA for hemoglobin interacting with 80 keV electrons for various slice thicknesses. The green and blue lines depict the boundary given by the Fresnel number $F = 1$ (compare Eq. (G.6)) for a potential map sampled at 1 Å and 3 Å respectively as a function of slice thickness. The shaded area around the lines denotes the variation due to possible slice orientations. The WPOA is valid below the red line, $\sigma\mathcal{V}_z < 0.36$, while the PA starts to hold for regions left/below to the blue or green line depending on the sampling of the map. The circles indicate the full map of hemoglobin at the respective sampling in the quasi-kinematic (QK) approach, whereas the triangles show the kinematic approach (K).

the quasi-kinematic potential refers to the mean-subtracted potential relevant for the generated phase contrast. Here we used the $\max(\sigma\mathcal{V}_z)$ as condition for the ranges of application for the different approximations, which gives a so-called worst-case (safest) condition.

G.4.2 Exit waves of a tubulin tetramer

For a tubulin tetramer (TT) constructed from PDBid 1SA0 ($\Delta z = 27$ nm) we show in Fig. G.2A the computed phase of the exit wave after interaction with 80 keV electrons using the four approximations discussed above, i.e. PA, PA+WPOA, WPOA and TPG. The potential map was sampled at 1 Å. In order to better visualize the effect of the approximations, we show in Fig. G.2B the differences of the four exit waves with a reference. This reference is computed by a multislice (MS) approach inspired by [144]. Since we use the MS method here only for computing the reference, the slice thickness is set equal to the resolution of the potential map. In the difference images of Fig. G.2B we observe that the TPG is nearly identical to the MS reference, whereas the WPOA shows deviations mostly in the stronger phase parts. For the PA we see deviations especially at the periphery of TT and, of course, for the combined PA+WPOA the deviations are the largest.

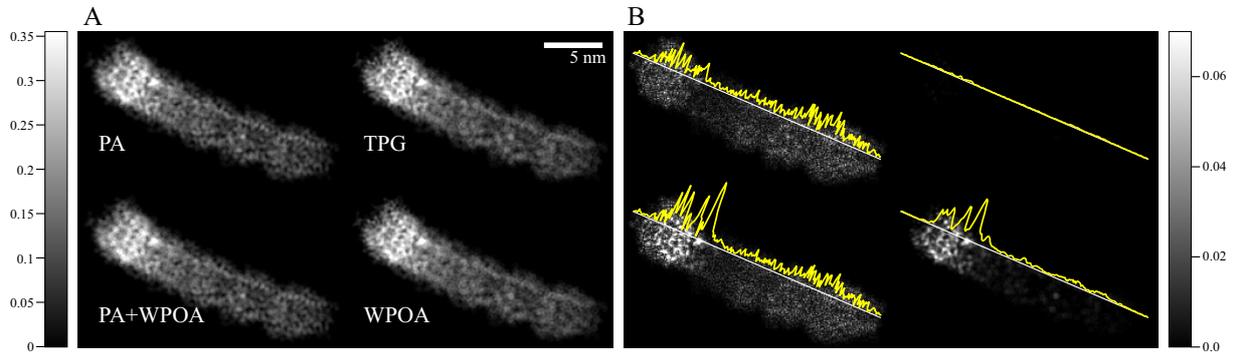


Fig. G.2. A) Simulated phases of exit waves of a tubulin tetramer ($HT = 80\text{kV}$) using the PA *via* Eq. (G.5), WPOA *via* Eq. (G.11), PA+WPOA *via* Eq. (G.7) and TPG *via* Eq. (G.12). B) Difference image of the exit waves in A) and the exit wave computed with a MS approach. The intensity scale bar indicate the phase of the exit wave subtracted by the mean.

G.4.3 Synthetic amorphous test specimen

We simulate exit waves of a synthetic test specimen using Eqs. (G.5), (G.11), (G.7) and (G.12) to study the validity of the predicted limits for the cases PA and WPOA. For the cases PA+WPOA and TPG we want to investigate where the limits of the validity of these combined approximations lie. Our derived conditions of Eqs. (G.6) and (G.10) are functions of the maximum spatial frequency, thickness and strength of the interaction potential. Therefore, a synthetic test potential must have these properties as well. The simplest potential that fulfills these criteria is a low-pass filtered Gaussian white-noise specimen of a specified thickness. This synthetic specimen resembles an amorphous material such as a carbon film.

The criterion for the WPOA Eq. (G.10) depends on the strength of the interaction potential. But since we are only interested in the scattering that produces phase contrast, the mean bulk potential can be ignored (quasi-kinematic). As a consequence $\sigma\mathcal{V}_z$ is not well defined as $\langle\sigma\mathcal{V}_z\rangle = 0$. An alternative is to consider $\max(|\sigma\mathcal{V}_z|)$ as we did in section G.4.1. This measure, however, depends for the synthetic test specimen on its spatial extent in (x, y) . Therefore, we will examine the standard deviation $\text{std}(\sigma\mathcal{V}_z)$ for our synthetic test specimens. For potential maps of a macromolecule, the $\text{std}(\sigma\mathcal{V}_z)$ depends on the size of the (vacuum) bounding box, in contrast to $\max(|\sigma\mathcal{V}_z|)$, which does not.

To test the applicability of the different approximations we again compare the four simulated exit waves against the MS reference. To quantify the difference between two of exit waves we use the normalized mean squared error (MSE), where the standard deviation of the reference exit wave is used for normalization. This normalization is necessary to ensure a proper comparison of MSEs originating from exit waves with varying $\text{std}(\sigma\mathcal{V}_z)$. Fig. G.3A shows the result of thresholding the MSE at 10%. We find a horizontal boundary for the WPOA and a vertical boundary for the PA, as expected from Eq. (G.6) and Eq. (G.10). The combined models have boundaries which asymptotically approach the individual (WPOA and PA) approximations. In Fig. G.3B a sketched version depicts the qualitative results in terms of regions where the different approximations hold.

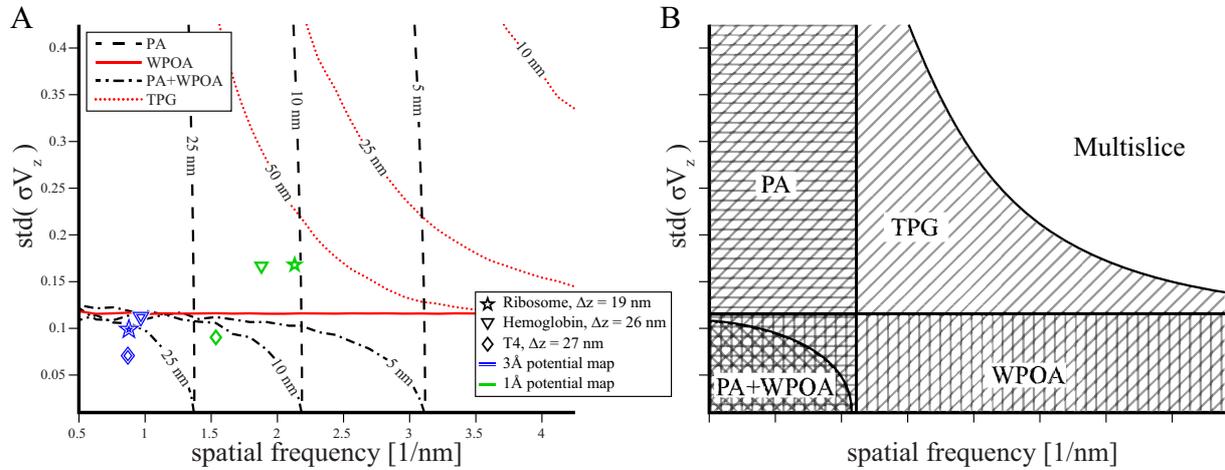


Fig. G.3. The applicability (at $HT = 80\text{kV}$) of the PA, WPOA, PA+WPOA and TPG. A) Boundaries for each approximation where different lines represent different specimen thickness. Lines indicate 10% MSE error of the respective approximations with a MS reference. Left/below the boundary the approximation holds for a particular thickness. Three protein-complexes potentials map (ribosome, hemoglobin, TT) sampled at 1 \AA and 3 \AA are included (see main text for details). B) A sketched diagram showing the qualitative results of A). The various striped regions depict region where each approximation holds.

In addition to the conditions that quantify the applicability for our synthetic specimen (Fig. G.3), we want to make a reproducible classification of the approximations for actual three-dimensional potential maps of macromolecules based on their potential properties. Therefore, we need to estimate the potential properties such that a synthetic specimen with that specification behaves similar to the actual potential under the different approximations (i.e. results expressed in similar MSEs against a MS reference). In Fig. G.3A we show the characteristics of three macromolecules (ribosomal subunit from *haloarcularia marismortui* - PDBid 1FFK, earth worm hemoglobin and TT) sampled at a 1 \AA and 3 \AA voxel size.

For the characteristic properties of each potential map we must calculate i) the maximum spatial frequency, ii) the thickness, and iii) the strength of the interaction potential. These properties can be ambiguous for a macromolecular potential as e.g. the size of the bounding box of the complex influences $\text{std}(\sigma V_z)$. As a solution we propose i) to retrieve the maximum spatial frequency by finding the 65th percentile of the 2D power spectrum of V_z , ii) to obtain the thickness by first computing $\text{std}(V(\rho, z))$ as a function of z , then finding the 2.5 and 97.5th percentile (i.e. the top and bottom of the protein respectively), and iii) to estimate the strength of the IP by masking any background from the map, then finding the 80th percentile of the histogram of $|V_z - \langle V_z \rangle|$. The corresponding values for the three macromolecules are depicted in Fig. G.3A (star, triangle and diamond). The specific values for each percentile were chosen such that a synthetic specimen with the estimated properties yields similar MSEs as the actual potential. The aim of the above procedure is to transfer the general conclusions from synthetic test specimens to actual macromolecular potentials. This procedure allows other macromolecules to be classified into regions based on the boundaries of applicability as depicted in Fig. G.3A.

Now we see in Fig. G.3 that the three proteins sampled at 3 \AA satisfy both the PA and

WPOA and are close to the PA+WPOA boundary. When sampled at 1 \AA the PA is not satisfied and only TT satisfies the WPOA. The hemoglobin results agree with those shown in Fig. G.1. Judging from Fig. G.2, which shows TT, we could conclude that the WPOA is violated for some parts of the molecule. In Fig. G.3, however, we see that on average TT satisfies the WPOA. This apparent contradiction is due to the fact that Fig. G.3 is computed from the average measure $\text{std}(\sigma\mathcal{V}_z)$.

G.5 Discussion

In this article we proposed quantitative criteria for the applicability of the PA (*via* the Fresnel number) and WPOA (*via* the probability of multiple interactions) in phase contrast cryo-EM. Rough indications for validity of various forward approximations in HREM were provided in [224]. In addition to the MS approach, proposed criteria motivate the existence of four models describing the electron wave propagation through the specimen (WPOA, PA, PA+WPOA and TPG). The choice of the model depends on the strength, frequency content and thickness of the interaction potential map. Here, we introduced the TPG approximation, known in HREM [224], to the cryo-EM field.

The MS method is the most accurate of the aforementioned methods and was utilized as a reference. The reasons for the little usage of MS in cryo-EM [51, 142] can be related to the lower resolution of the structures determined by cryo-EM, and to a more complicated inverse problem in 3D reconstruction. Potential difficulties of the 3D reconstruction based on MS can be partially avoided by using a directly invertible approximation (e.g. WPOA or WPOA+PA) in the first iteration of a typical tomographic scheme. As shown in Fig. G.2 the forward simulations indicate that the direct TPG approach gives nearly identical exit waves as a recursive MS calculation. We expect, however, that TPG can be advantageous for 3D reconstructions due to its invertibility and possibility to utilize non-uniform fast Fourier transform sampling of the Ewald sphere [29, 30].

The presented simulations of an amorphous test specimen serve as a practical reference to facilitate the model choice for electron wave propagation through an actual macromolecule such as hemoglobin, ribosome, or tubulin. The accuracy of each approximation depends on the properties of the potential under investigation. In order to describe the relevant potential properties we introduced two measures: $\max(|\sigma\mathcal{V}_z|)$ and $\text{std}(\sigma\mathcal{V}_z)$. The former represents the worst-case (safest) boundary and the latter an average boundary for which the approximations hold.

We deliberately present all our results for $\text{HT} = 80\text{kV}$ because for higher HT (shorter wavelength), the approximations given by Eqs. (G.10) and (G.6) are relaxed as $\sigma \propto \lambda$. The criteria for WPOA and PA are also easier to satisfy for potential maps of lower resolution (compare Figs. G.1 and G.3). Note that we do not make claims about the resolution in the final recorded images as it depends for a large part on the electron count, beam-induced movements, CTF and camera characteristics.

Under typical circumstances inelastic scattering influences the total contrast and we do not record pure phase contrast. Nevertheless, the findings in this article are important since phase

contrast is the primary contrast mechanism in cryo-EM [2]. In our analysis the mean value of the IP was subtracted (quasi-kinematic approach) since it does not contribute to the phase contrast. For inelastic scattering, modeled as the imaginary part of the IP [10], the mean potential cannot be neglected since it damps the useful phase signal.

As practical conclusions we find that, when simulating images at resolutions of $\sim 5 \text{ \AA}$, the applicability of the PA and WPOA need to be re-considered. Here, the TPG offers an excellent solution, as an alternative to the multislice approach. For tomograms with typical resolutions $> 30 \text{ \AA}$, the PA and WPOA are generally applicable. In single particle analysis, structures can be obtained up to 3.3 \AA resolution [227] at which the PA and WPOA may be violated depending on the size of the macromolecule, while the TPG again offers a solid solution. The implementation of the exit wave simulations is freely available for non-commercial use upon request.

Chapter 5

Radiation damage

Published as [50]: M. Karuppasamy, F. Karimi Nejadasl, M. Vulovic, A. J. Koster, R. B. G. Ravelli, “Radiation damage in single particle cryo-electron microscopy: effects of dose and dose rate”, *Journal of Synchrotron Radiation* 18 (3) (2011) 398-412.

Abstract

Radiation damage is an important resolution limiting factor both in macromolecular X-ray crystallography and cryo-electron microscopy. Systematic studies in macromolecular X-ray crystallography greatly benefited from the use of dose, expressed as energy deposited per mass unit, which is derived from parameters such as incident flux, beam energy, beam size, sample composition and sample size. In here, the use of dose is reintroduced for electron microscopy, accounting for the electron energy, incident flux, and measured sample thickness and composition. Knowledge of the amount of energy deposited allowed us to compare doses with experimental limits in macromolecular X-ray crystallography, to get an upper estimate of radical concentrations that are built up in the vitreous sample, and to translate heat-transfer simulations done for macromolecular X-ray crystallography to cryo-electron microscopy. Stroboscopic exposure series of 50-250 images were collected for different incident fluxes and integration times from *Lumbricus terrestris* extracellular hemoglobin as a test sample. The images within each series were computationally aligned and analyzed with similarity metrics such as Fourier ring correlation, Fourier ring phase residual, and figure of merit. Prior to gas-bubble formation, the images become linearly brighter with dose, at a rate of approximately 0.1 % per 10 MGy. The gradual decomposition of a vitrified hemoglobin sample could be visualized at a series of doses up to 5500 MGy, by which dose the sample was sublimated. Comparison of equal-dose series collected with different incident fluxes showed a dose-rate effect favoring lower fluxes. Heat simulations predict that sample heating will only become an issue for very large dose rates ($50 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$ or higher) combined with poor thermal contact between the grid and cryo-holder. Secondary radiolytic effects are likely to play a role in dose-rate effects. Stroboscopic data collection combined with an improved understanding of the effects of dose and dose-rate, will aid the single-particle cryo-electron microscopists to have a better control of the outcome of their experiments.

5.1 Introduction

Single particle cryo-electron microscopy (SP cryo-EM) is a unique technique widely used to elucidate the three-dimensional (3D) structures of macromolecules of molecular mass greater than few hundred kDa [228–231]. It provides complementary structural information as compared to macromolecular X-ray crystallography (MX) and nuclear magnetic resonance (NMR) techniques which require single crystals and labeled proteins respectively as a prerequisite to be studied by such methods. In SP cryo-EM studies, numerous projection images are collected from randomly (or sometimes preferentially) oriented macromolecules in a thin layer of vitreous sample (vitreous being an amorphous state). By computational reconstruction methods, a 3D electron density map of molecules to a resolution of about 10 Å (1 nm) can be obtained from these projection images [229, 232]. Further, it is becoming common to achieve pseudo-atomic models of macromolecular complexes to 6–4 Å resolution by fitting the atomic models of some of the components coming from X-ray diffraction studies into the reconstructed EM map of the entire complex (for example [233–237]). A full-atom model of a non-enveloped aquareovirus at 3.3 Å was recently obtained by SP reconstruction in which side-chain densities for non-Gly amino acids were clearly visible [227]. Technological improvements in electron optics, sample preparation, data collection and processing have enabled these recent advances.

Radiation damage, unfortunately, will always limit the achievable resolution in single particle cryo-EM [23, 24]. The damage results from the deposition of energy into the macromolecules owing to the inelastic interactions between the ionizing electron radiation and matter. Traditionally, radiation damage is treated as a binary nuisance. The dose used to collect SP cryo-EM data is a compromise between signal-to-noise ratio and radiation damage. Very high-quality images can be obtained, although, at the same time it is usual to discard an unpredictable amount of particles for reasons such as beam-induced movements [23]. At the typical energies used in transmission electron microscopy (TEM), 100 – 300 keV, inelastic scattering is about 3 times more likely than elastic scattering [158, 238]. Inelastic scattering events include, in order of importance, plasmon scattering, K- and L-shell ionization, Bremsstrahlung, fast and slow secondary electron emission. The deposited energy invariably destroys the biological specimen. Studies that describe these effects are as old as cryo-electron microscopy itself [16, 23].

Radiation damage studies done in cryo-EM received full attention from macromolecular X-ray crystallographers, in particular since radiation-damage became a daily nuisance on highly intensive third generation (3G) wiggler and undulator beamlines (reviewed by [90, 239]). *Vice versa*, systematic radiation damage studies in MX might be of interest to the SP cryo-EM community. Below, a concise background of these studies in MX is given.

The X-ray beam will introduce structural changes in the sample during the experiment, resulting in non-isomorphism, which is thought to be a major cause of unsuccessful multiple anomalous dispersion structure determinations [240, 241]. However, by collecting multiple complete data sets within the usable lifetime of a crystal, it has been possible to study radiation damage at an unprecedented detail. These studies have been complemented by experimental methods such as UV/VIS microscopy [242], fluorescence lifetime microscopy [243], X-ray spectroscopy [244], Raman spectroscopy [245, 246], and SAXS (Small Angle X-ray Scattering) [247], as well as theoretical simulations [248–250].

Radiation damage, in general, can be classified as primary and secondary in nature. The most dominant primary inelastic interaction between X-rays and matter at the energies typically used in MX (8 – 14 keV) is photoelectric absorption. The atom undergoing photoelectric absorption, typically in the order of 10 per unit cell per synchrotron data set, is a site of primary damage. The energy of the ejected electron depends on the energy of the incoming photon. An emitted photoelectron with ~ 12 keV for a 12 keV photon, will have a mean free path length of a few micrometers [251] and will cause secondary damage due to the excitation and formation of another ~ 500 ionization events. The resulting electron-loss and electron-gain centers might cause direct damage to the protein or indirect by diffusion through the vitrified cryo-buffer. Diffusible radicals may or may not recombine and might be intercepted by radical scavengers [251–254].

Early synchrotron studies of radiation damage in macromolecular crystals at cryogenic temperatures showed that site-specific damage will occur in a well defined order. Disulphide bonds are in particular susceptible, followed by decarboxylation of aspartate and glutamate residues [255–257]. The fact that there is a large range in susceptibility among different disulphide bonds and carboxyl groups illustrates the importance of secondary processes. The radical species that are formed upon irradiation of water include hydrogen (H^+) and hydroxyl (OH^-) radicals, electrons (e^-) and hydrated electrons (e_{aq}^-). Protons are only known to become mobile in amorphous ice at ~ 115 K. OH^- radicals become mobile above 130 K in crystalline ice [258]. Positive holes are rapidly trapped at 77 K (boiling point liquid nitrogen) forming amido radicals on the protein backbone chain, whereas electrons are able to move efficiently at 77 K until they encounter disulphide bonds where they are trapped [254, 256, 259]. The role of secondary processes will be temperature dependent; all radicals will gain mobility at higher temperatures but not all radicals can be frozen out at 77 K. Hydrated electrons will still be mobile under helium cooling. At room temperature and neutral pH the yields of hydrated electrons and hydroxyl radicals are approximately equal, while the yield of H atoms is much smaller [260]. At acidic pH , hydrated electrons rapidly recombine with protons to form hydrogen atoms. Both reducing radicals, the hydrated electron and the H atom, react rapidly with oxygen, if present, to yield oxygen-centered radicals that can attack components of the protein. The oxidizing OH^- radical is highly reactive and will abstract hydrogen atoms from C-H and N-H bonds to form carbon- and/or nitrogen-centered radicals. At room temperature, with many radicals being mobile, an inversed dose-rate effect has been observed and attributed to the increased importance of radical recombination at higher dose rate [260]. It was shown that OH^- radicals can be effectively scavenged in MX at room temperature [254]. Investigations into dose-rate effects in MX at cryogenic temperatures has indicated that such effects are in general small for vitrified samples [261–265].

The dose in grays ($1 \text{ Gy} = 1 \text{ J/kg}$) can be calculated with the aid of programs such as RADDOSE [266–268] from the incident-beam parameters (X-ray flux, photon energy and beam shape) and the crystal size, together with the absorption and attenuation coefficients obtained from the knowledge of the total number of different atom types in the unit cell. The widespread use of dose rather than incident flux and integration times, greatly facilitated objective comparisons between experiments performed at a large variety of X-ray sources, ranging from sealed tubes to microfocus synchrotron beamlines. The tolerable dose limit for a macromolecu-

lar crystal before it loses half of its diffraction intensity, $D_{1/2}$ was predicted to be 20 MGy [269] based on lifetime measurements on crystalline biological samples in the electron microscope. Dose limit in MX was experimentally measured $D_{1/2} = 43$ MGy [265] and a maximum dose of 30 MGy is recommended. Others [270–273] related the fading of the average intensity with dose through a resolution-dependent formula

$$\langle I \rangle = \langle I \rangle_{\text{ND}} \exp[-\ln(2)D/(Hd)], \quad (5.1)$$

where D is the absorbed dose, $\langle I \rangle$ is the average spot intensity after absorbing a dose D , $\langle I \rangle_{\text{ND}}$ is the average spot intensity in the absence of radiation damage, $\ln(2)$ is the natural log of two, d is the resolution in \AA , and H is a constant [271] of $10 \text{ MGy}\text{\AA}^{-1}$.

Radical recombination has been postulated as a plausible cause for dose-rate effects [260]. Excessive heating of the sample would also result in a dose-rate effect [262]. A thorough study of the thermal interactions of a cryo-cooled biological sample exposed to a strong X-ray beam based on classical heat-transfer theory is presented in [248]. The sample is internally heated as the energy of the X-ray beam is absorbed and externally cooled at its surface by convection to a cold N_2 gas stream. Two theoretical models were presented, a spatially uniform heating of a thin sample for the so-called *lumped model*. Here the temperature in the sample is a simple function of time. For thicker samples the temperature will be both a function of time and space; for this a *distributed model* was derived. They showed that heat transfer is limited by the rate of external convection; internal temperature gradients within the crystal are small. Some of the parameters used in the models described above are refined in [249] and it was concluded that crystal heating by X-ray absorption on present high-flux beamlines should be small (< 20 K), although there are new beamlines with fluxes larger than those used in their calculations. Using an IR camera, [274,275] gave an experimental verification of the calculations of [249] and [248]. Glass bead samples were used as a surrogate for the biological samples, and the spatial and temporal distribution of a cryo-cooled glass bead heated by a smaller X-ray beam could be carefully measured and visualized. They confirmed that the heating is not sufficient to raise the sample temperature to the amorphous/crystalline ice transition region of $\sim 130 - 140$ K [276,277].

In this work studies on the effects of dose and dose-rate for SP cryo-EM are presented and related to systematic radiation damage studies in MX. The deposited energy per mass unit (dose) used in our SP cryo-EM experiments were estimated from parameters such as flux, integration time, beam size and energy, protein concentration, sample thickness and the main contribution to inelastic scattering, namely plasmon interaction. The sample thickness was measured using electron tomography. Dose-rate effects were investigated by collecting several series of single-particle data with identical cumulative doses, but with variable incident fluxes and integration times. Analogous to MX, a figure-of-merit (FOM) term is defined to describe the average cosine of phase-errors within an aligned image series. It is shown that FOM can be used as a metric for radiation-damage studies. Unlike MX, a clear dose-rate effect could be observed, favoring the use of lower dose-rates. Dose-rate effects could originate, as mentioned above, from radical recombination and (or) sample heating. The process of sample heating by the electron beam was studied by simulated systems based on classical heat transfer models. The potential influence of radical recombination was studied by altering the solvent constituents of the SP sample.

High salt and glycerol concentrations, typically used as cryo-protectants in MX, are examined at cryo-temperatures within the TEM to see if they alter the radiation robustness of the sample. Similar, a low concentration of fixative was used. Results are discussed and compared with recent findings in literature [23, 24, 278–281].

5.2 Methods

5.2.1 Experimental methods

Sample preparation

We used *Lumbricus terrestris* erythrocyruorin (Hb) as a test sample. This 3.6 MDa extracellular respiratory protein complex, termed either erythrocyruorins or hemoglobins [282, 283] consists of 144 hemoglobin and 36 linker subunits. The hemoglobin subunits are organized into 12 dodecamers, each of which binds to a heterotrimer of linker proteins. Each dodecamer is a trimer of heterotetramers. The 12 dodecamers form a core complex with D_6 symmetry. The sample was prepared using a protocol adapted from [170]. The harvested concentrated Hb solution was stored at 277 K in 50 mM ammonium acetate (measured pH of 6.5) until use. Protein A (a bacterial surface protein commonly used because of its ability to bind immunoglobins) conjugated with 5 nm colloidal gold particles (CMC-UMC, Utrecht, The Netherlands) was added as fiducial markers to the protein sample just before preparation of the EM grids. Aliquots of 3 μ l sample at 0.5 – 1 mg/ml protein concentration were applied to 200 mesh glow discharged C – flat™ (Protochips Inc., NC, USA) grids (1.2 μ m hole size) and blotted from both sides inside an FEI Vitrobot using 3 s blotting time with 100 % relative humidity. Subsequently, the blotted grid was rapidly plunged into liquid ethane for vitrification. The grid was stored in liquid nitrogen pending examination in the electron microscope.

In addition to the low-salt control sample described above, three more solvent constituents were tested. The required amount of stock was dissolved to 0.5 – 1 mg/ml final protein concentration in i) 2 M ammonium acetate, ii) 50 % (v/v) glycerol, and iii) 0.2 % (v/v) glutaraldehyde. The sample prepared in 2 M NH_4Ac (as high salt) and 50 % (v/v) glycerol served as a model system for cryo-protectants commonly used in MX. Glutaraldehyde was chosen as it has been used as stabilizing organic molecule for protein complexes studied in SP cryo-EM [284, 285]. For the glutaraldehyde sample, the protein was incubated in a solution containing 0.2 % (v/v) glutaraldehyde in 50 mM ammonium acetate for about 10 min prior to use. Grids were prepared as above.

Image acquisition/data collection

Images were recorded on a 4k \times 4k Eagle on-axis CCD camera using a FEI (www.fei.com) TECNAI Biotwin electron microscope with a LaB₆ filament operating at 120 kV without using an energy-filter. Other microscope settings used were: condenser aperture number 3 (size of 100 μ m), objective aperture 3 (70 μ m), and spot size index 6. The grid was kept in a Gatan 626 (Gatan Inc., USA) cryo-holder at a temperature of 103 K, as monitored by the temperature

control unit. The magnification at the detector plane was $\sim 68000\times$, the requested underfocus $3\ \mu\text{m}$ and the exposure time 1 s. Images were hardware binned and consist of 2048×2048 pixels. The field of view was $0.9\ \mu\text{m} \times 0.9\ \mu\text{m}$, the pixel size $4.5\ \text{\AA}$ square. The incident flux was derived from the detector analog-to-digital units (ADUs) by taking 1 s exposures without sample and using conversion factors (in ADU/e^-) as calibrated by [47] for these systems. Each exposure series was collected from a previously unexposed sample suspended across one of the holes in the C-flat grid. A series of 50 successive images was recorded with an incident flux of $5\ e^- \text{\AA}^{-2}\text{s}^{-1}$ (medium flux), corresponding to an integrated flux for the final images of $250\ e^- \text{\AA}^{-2}$. Similarly, a series of 50 images was acquired with an incident flux of $50\ e^- \text{\AA}^{-2}\text{s}^{-1}$ (high-flux) and another series of 250 images with an incident flux of $1\ e^- \text{\AA}^{-2}\text{s}^{-1}$ (low-flux). In addition, 50 high-flux images ($50\ e^- \text{\AA}^{-2}\text{s}^{-1}$) were collected with an exposure time of 0.1 s (high-flux short-exposure), resulting in an integrated flux for the final images of $250\ e^- \text{\AA}^{-2}$. The pre-specimen shutter was used for all the experiments: the specimen was only exposed during the data recording. The pre-specimen shutter response of the microscope was checked by comparing the median intensity of the sum of ten images with exposure time of 0.1 s to the median intensity of one image with 1 s exposure time. The difference was less than 0.09 %. All images were collected as fast as possible after each other, resulting, on average, 13 images per minute.

Sample thickness measurements

In order to calculate the approximate sample thickness, tilt series were acquired and thickness was calculated from the reconstructed tomograms. Single-axis tilt series were recorded using FEI Inspect3D software for tilt angles from -52° to $+52^\circ$ in steps of 1° at a detector magnification of $\sim 68000\times$, and an incident flux of $1.3\ e^- \text{\AA}^{-2}\text{s}^{-1}$. The defocus was set to $5\ \mu\text{m}$ at 0° tilt angle. The IMOD software package [286] was used for data processing and 3D tomographic reconstruction. The approximate sample thickness was derived from the number of sample-containing tomogram slices in the beam direction.

5.2.2 Computational methods

Image alignment

Where relevant, images were corrected for statistical outliers [47]. Account was taken of sample drift by aligning the images to the first image of each series using a normalized cross-correlation function. The translation vectors were calculated with sub-pixel accuracy. The real-space images were translated by applying a corresponding phase shift in Fourier space.

Dose and heat calculations

The dose, in grays (Gy), was calculated based on the incident flux, exposure time, electron beam size and energy, and the molecular weight and number of Hb particles, in a manner similar to the program RADDOS [266–268]. As the product of the dominant form of inelastic scattering,

only plasmons were taken into account, depositing on average 20 eV per inelastic event into the sample [158].

The temperature rise of the vitreous ice was estimated based on lumped model calculations [248]. The total deposited energy as determined by the dose calculations (see Appendix H) was assumed to contribute to heating of the sample. In the *lumped system*, the internal temperature spatial variations in the sample are neglected and the temperature changes only with time. The energy balance is given by [248]

$$\rho c_p V \frac{dT}{dt} = P_{\text{dep}} - hA_s(T - T_0), \quad (5.2)$$

where ρ is the density of vitreous ice (0.93 g/cm³), V the volume of the illuminated sample, P_{dep} is the deposited power (energy per time) to the specimen, A_s is the area through which heat is conducted, T_0 is the initial temperature of the sample (103 K) and h is the heat-transfer coefficient. The heat capacity of the sample (c_p) was taken to be 900 Jkg⁻¹K⁻¹ [249]. In a lumped system with isolated walls (adiabatic model), this model predicts a rate of temperature increase of $P_{\text{dep}}/(\rho c_p V) \approx 61121 \text{ Ks}^{-1}$ (Fig. 5.6A). This is unrealistic and shows the importance of incorporating the cooling from the ambient and grid into the model. The evolution of the temperature could be written as [248]:

$$T(t) = T_0 + \frac{P_{\text{dep}}}{hA_s} [1 - \exp(-t/t_{\text{sys}})], \quad (5.3)$$

where

$$t_{\text{sys}} = \rho c_p V / (hA_s) \quad (5.4)$$

is the system time constant which characterizes the cooling rate. For a short time after the onset of the exposure, the system acts like an adiabatic system and the temperature increases linearly with time [248]. After a time corresponding to three system time constants ($3t_{\text{sys}}$), the sample reaches 95 % of the final temperature. If the exposure is shorter than this, the final maximum temperature will not be reached.

In the *distributed system*, the temperature is non-uniform both in time and position. The spatial and temporal thermal behavior of the system was simulated as heat diffusion in one dimension from the illuminated spot area to the cryo-cooled copper grid. The temperature distribution is derived from the diffusion equation,

$$c_p \rho \frac{\partial T(x, t)}{\partial t} = \rho_{\text{HS}} + k \frac{\partial^2 T(x, t)}{\partial x^2}, \quad (5.5)$$

where k is the thermal conductivity of vitrified ice. For simplicity, k is assumed to be constant. The parameter $\alpha = k/(c_p \rho)$ is called the thermal diffusion coefficient and determines the rate of the diffusion process. ρ_{HS} is the power density of the heat source derived from Eq. (5.2):

$$\rho_{\text{HS}} = \frac{P_{\text{dep}} - hA_s [T(x, t) - T_0]}{V} \Big|_{|x| < d_b}. \quad (5.6)$$

In order to solve Eq. (5.6) numerically, time and space were discretized. Potential stability problems were overcome by using the Crank-Nicolson method [287]. Since the thin cryo-EM samples are relative transparent to the electron beam, heat diffusion in the direction of the

beam (axial) can be considered instantaneous. As boundary conditions, it was assumed that the supporting copper mesh was in perfect thermal contact with the liquid nitrogen cooled sample holder rod, and kept at a constant temperature of 103 K. The illuminated specimen area $|x| < d_b$ (d_b being beam diameter) was approximated as a lumped system. Simulations were performed for vitreous ice of 50 μm diameter and 0.15 μm thickness, a uniform beam (a top-hat function) with a diameter of 10 μm , an incident flux of both $5 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ and $50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ at 120 kV accelerating voltage, and a heat transfer coefficient $k = 1.1 \text{ Wm}^{-1}\text{K}^{-1}$ [249]. Since the grid mesh is larger than the electron beam diameter, heat is transported from the illuminated region to the grid *via* the sample. Energy loss into the vacuum through black body radiation has been neglected. The temperature difference between the grid and the edge of illuminated specimen is given by $\Delta T = P_{\text{dep}}l/(kA_s)$, where l is the distance from illuminated area to the grid bars. If this is compared with the stationary case of the lumped system $\Delta T(t \rightarrow \infty) = P_{\text{dep}}/(hA_s)$, the heat-transfer coefficient h can approximately be expressed by k/l .

Mass loss

For each series, the common subarea was defined and its mean intensity was calculated for each image. The slope of $\Delta I/I_0$ ($\Delta I = I - I_0$) *versus* dose was tabulated together with the intensity of the first image of each series, the estimated sample thickness, and the number of hemoglobin molecules per unit area.

Beam-induced movement

Fiducial gold particles in the aligned images were used to track beam-induced movements that might have occurred during data collection. Distance matrices were calculated from the gold marker positions for the first and the last image of each series. The movement of the gold particles was measured by a change in these distance matrices within a series. The mean of the distance differences provides a metric for beam-induced movements [280].

The gold marker detection was challenging because of several difficulties. The gold markers are on average 5 nm in diameter, but can vary significantly in shape and size. The different series showed differences in signal-to-noise ratio. Inspired by [288] and [289], the above problems were overcome by using the Laplacian of Gaussian-filtered images. The Gaussian filtering was performed for a range of sigma values, varying around the gold size in pixels. The Laplacian of each of these Gaussian filtered images were summed, which is defined here as the sum of the Laplacian of Gaussian functions (sLOG). Gold particles were detected as the brightest regions in the sLOG images. The centers of the gold particle positions were found from the center of mass of the brightest regions. For each gold particle in the reference image, the vicinity area in the aligned image was used to locate the corresponding gold particle in that image.

Figure-of-merit as a measure of phase error

After alignment, a common subarea was defined for each exposure series. The Fourier transforms ($\mathcal{F}[\]$) of these subimages were averaged to yield averaged complex structure factors. A

figure-of-merit was defined as

$$\text{FOM} = \langle \cos(\varphi_j - \langle \varphi \rangle) \rangle, \quad (5.7)$$

where φ_j is the phase of the \mathcal{F} of individual subimage j , $\langle \varphi \rangle$ is the phase of the averaged complex structure factor described above, and the averaging is carried out for each pixel over N number of images within a series. N varied between 10 and 250 in our calculations. The FOM can vary between zero for random data and one for ideal noise-free data.

Defocus estimation

Periodogram-averaged power spectra were calculated as described previously [96]. The power spectra of the individual (medium- and low-dose) images were too noisy for defocus estimation through contrast transfer function (CTF) fitting.

The defocus could be derived from the radial averaging of the mean cosine of the difference phase, FOM [131]. These FOMs were calculated after splitting each data series into five parts, with each part corresponding to an integrated flux of 50, 100, 150, 200, and 250 $e^- \text{\AA}^{-2}$ respectively.

Fourier ring correlation and Fourier ring phase residual

The radiation damage was scrutinized closely by different similarity metrics. Two metrics were computed, the Fourier ring correlation (FRC) and the Fourier ring phase residual (FRPR) [130, 290]. They are obtained from

$$\text{FRC} = \frac{\sum F_1 F_2^*}{(\sum |F_1|^2 \sum |F_2|^2)^{1/2}}, \quad (5.8)$$

$$\text{FRPR} = \frac{\sum |F_1| |F_2| |\varphi_2 - \varphi_1|}{\sum |F_1| |F_2|}, \quad (5.9)$$

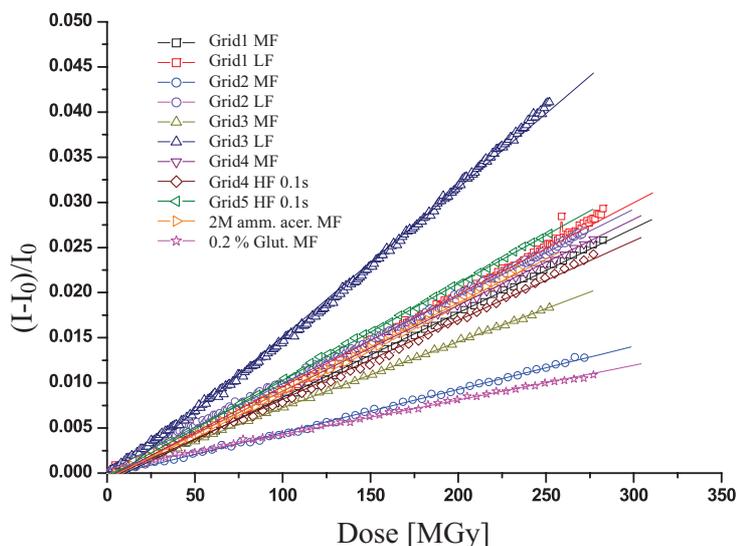
where F_j , $|F_j|$, and φ_j are, respectively, the Fourier transform of the j -th image for $j = 1, 2$ and its magnitude and phase. The metrics were computed up to the first crossing of the CTF, namely corresponding to 3.5 nm. Images were first aligned and then summed up to the specified integrated flux.

5.3 Results

5.3.1 Dose

Table 5.1 shows the relation between incident flux and dose for all the data. The dose was calculated (see Appendix H) based on the following parameters. The electron beam had a diameter of 10 μm as measured at lower magnification, using the same condenser and objective lens settings as for the experiments. Tomographic reconstructions (see section 5.2.1) showed that the typical vitreous sample layer thickness was ~ 150 nm. A volume of 11.8 fl was irradiated with, for the medium-flux exposure series, 5 $e^- \text{\AA}^{-2} \text{s}^{-1}$ during 1 s per image. The counted number of Hb

Fig. 5.1. Normalized intensity change as a function of dose for different exposure series. LF refers to the low-flux series ($1 e^{-\text{\AA}^{-2}\text{s}^{-1}}$), MF to medium-flux ($5 e^{-\text{\AA}^{-2}\text{s}^{-1}}$), and HF to high-flux ($50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$). The normalized intensity change was found to be linear with dose as shown by least-square fit to the data.



molecules per unit area (for example 1000 molecules in $1 \mu\text{m}^2$) is given in Table 5.1. A density for low density amorphous ice of 0.93 g/cm^3 [177] was used, resulting in a total of 3.6×10^{11} water molecules in the path of the beam. Based on all these parameters, an approximation for the total atomic content of the irradiated volume could be calculated. The total mass of the irradiated volume, based on these atom counts, was 10.9 pg. Using the atomic scattering factors of [158] and an incoming beam energy of 120 kV, we calculate that a fraction of 48 % of the incoming electrons was scattered inelastically, each depositing 20 eV, resulting in a total amount of energy deposited of 60.1 nJ. The dose for each individual medium-flux image corresponds to $\sim 5.5 \text{ MGy}$.

5.3.2 Averaged intensity versus dose

Table 5.1 gives the slopes of the normalized intensity change $\Delta I/I_0$ versus dose for the common subareas of each exposure series. The different incident fluxes and integration times can be found in the same table, together with the dose (in grays) per exposure. The $\Delta I/I_0$ graphs are shown in Fig. 5.1: the metric is highly linear with dose for all the low-, medium- and high-flux short-exposure (0.1 s) series that were collected. However, the high-flux (1 s exposure) series had to be excluded due to non-uniform events such as gas bubble formation, image blurring, or crystalline ice formation. The images became approximately 1 % brighter per 100 MGy dose.

5.3.3 Radiation damage series of Hemoglobin followed up to 5500 MGy

Movie S1 (see supplementary material in online version at doi:10.1107/S090904951100820X) shows a high-flux series of 100 images. Each image was taken with $50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ and 1 s integration time, corresponding to a dose of $\sim 55 \text{ MGy}$ per image. This series was taken at the edge of a hole of a C-flat support film, showing the support film on the right-hand side of the image. Comparing the first with the second image in this series, one can already observe a

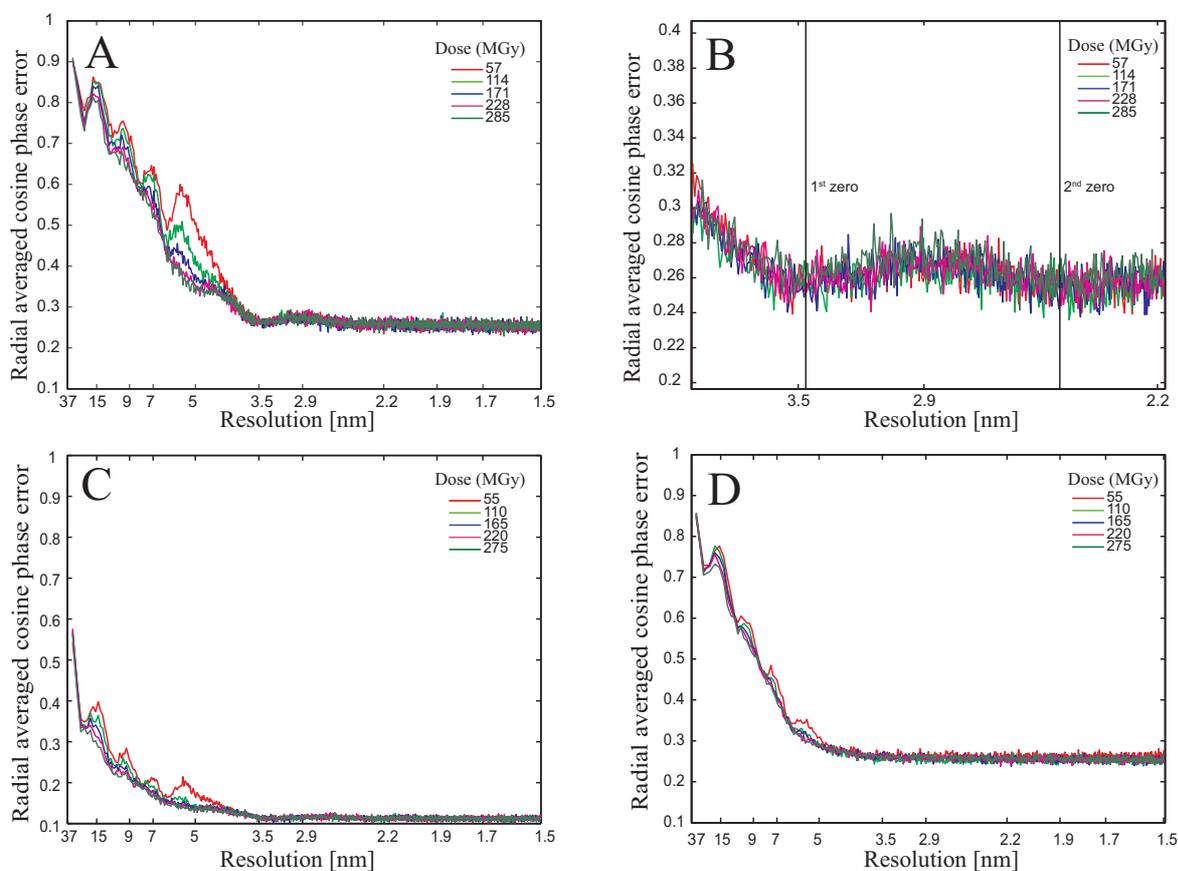


Fig. 5.2. Plots of radial averaged cosine phase error *versus* resolution for different dose rates. (A) Radial averaged FOMs are given for a medium-flux series on Hb in a low-salt sample for integrated fluxes of 50, 100, 150, 200, and 250 $e^- \text{Å}^{-2}$. (A) Close-up of (B) showing the first and second zero crossing of the CTF for a defocus of 3.37 μm . Radial averaged FOMs for the (C) low-flux and (D) high-flux short-exposure series.

blurring of the particles. This loss of resolution proceeds monotonically throughout the first 10–20 images. Cryo-electron tomography regularly shows the presence of loose ice particles on top of the vitreous sample layer. In our movie, such ice particle can be seen in the lower-left part of the image. This ice crystal seems to dissolve into the sample layer within the first seven images. Starting from image number seven, macroscopic bubbles appear at the protein sites. This is most obvious for the vitreous sample layer in the hole. One to four nanobubbles appear per hemoglobin complex, and a maximum number of bubbles is seen around image number 14. Hereafter, bubbles fuse and, eventually, disappear. Most bubbles in the hole area have disappeared at image number 40. The structure of the individual protein molecules disintegrates together with the bubble formation. At image 10, a remnant of the sixfold symmetry can still be seen for some particles, whereas towards image 40, all resemblance with the original particles is gone. Strikingly, the relative positions of the fiducial gold markers do not seem to alter significantly. Later in the series, from frame 60 onwards, the images start to show more detail. Sharp

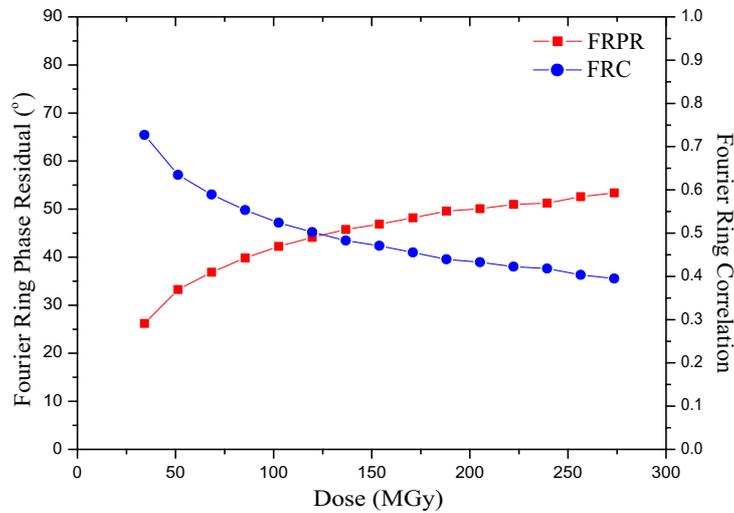


Fig. 5.3. Fourier ring phase residual (FRPR) and Fourier ring correlation (FRC) as a function of dose. Medium-flux data were combined in groups of three images, corresponding to an integrated flux of $15 e^{-\text{\AA}^{-2}}$ per combined image. The first combined image was used as a reference.

black worm-like features start to form, residuals of the protein molecules. The whole series had to be corrected for a linear change in intensity, as the sample was becoming more transparent with dose (Fig. 5.1). After image number 97, a hole formed from the top part of the image. In total, an excessive dose of 5500 MGy was used for this series, which was collected in a time span of 7.5 minutes.

5.3.4 Defocus variation

Changes in image contrast and particle resolution could, in principle, be a consequence of a drift of the defocus during the exposure series acquisition. The general stability of the microscope, therefore, was investigated by imaging a thin layer of carbon at room temperature 30 times. A series of measurements at three consecutive levels of defocus was recorded: 1, 1.25 and 1.5 μm . The standard deviation of the series was in the range of a few nanometers [48].

Radial averaged FOM figures were calculated (see Eq. (5.7)). Fig. 5.2 shows these graphs for five different successive cumulative doses for a medium-flux series on Hb in a low-salt sample. The first and second zero of the CTF would correspond to 3.46 nm and 2.45 nm respectively for an estimated defocus of 3.57 μm (the requested defocus was 3 μm). Both positions are found in these data and do not drift significantly as a function of cumulative dose. Fig. 5.2 is representative in this respect for all of the exposure series used in this study. It is found that the defocus ranged between 2.83 and 3.57 μm for the different medium-flux series.

5.3.5 Beam-induced movements of gold particles

The mean value (and its standard deviation) of change in distance between all possible pairs of gold particles is shown in Table 5.1. The average values for all low, medium, and high incident

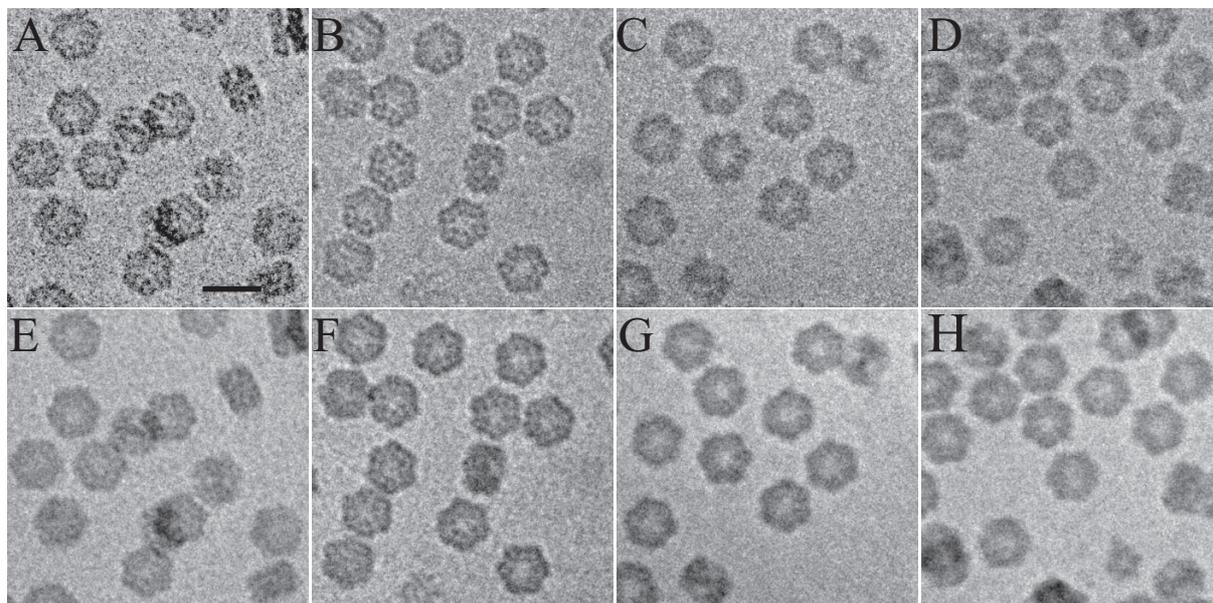


Fig. 5.4. Qualitative investigation of the dose-rate effect. The aligned and summed images of (A) and (E) low-flux, (B) and (F) medium-flux, (C) and (G) high-flux, and (D) and (H) high-flux short-exposure series are shown at two different integrated fluxes of (A)-(D) $50 e^{-}\text{\AA}^{-2}$ and (e)-(h) $250 e^{-}\text{\AA}^{-2}$, respectively. The scale bar shown in (a) corresponds to 30 nm.

flux series data are found to be $\sim 2 - 8$ pixels regardless of the solvent constituent used except for the glycerol sample. The value is about the same for the high-flux short-exposure series. A large distance of 54 pixels is seen between the first (integrated flux $5 e^{-}\text{\AA}^{-2}$) and fifth image ($25 e^{-}\text{\AA}^{-2}$) from the 50 % glycerol medium-flux series. The value becomes 89 pixels when the first image is compared with the tenth image ($50 e^{-}\text{\AA}^{-2}$), indicating an excessive amount of beam-induced movements within the glycerol sample.

5.3.6 Fourier ring correlation, Fourier ring phase residual and FOM plots

The aligned medium-flux images were grouped over a variable number N . Fig. 5.3 shows FRC (see Eq. (5.8)) and FRPR (see Eq. (5.9)) for combined images that contain the sum of three original images. Each combined image corresponds to an integrated flux of $15 e^{-}\text{\AA}^{-2}$. The first summed image was taken as a reference and compared with the successive summed images within a series. The metrics were calculated for different resolution ranges: Fig. 5.3 shows only the low-resolution data. The FRC decreases as a function of cumulative dose whereas the phase residual FRPR increases. Similarly, the FOM values decrease (corresponding to an increase in phase errors) as a function of accumulated dose for all three dose-rate series, low, medium and high flux (Figs. 5.2A, 5.2C-5.2D).

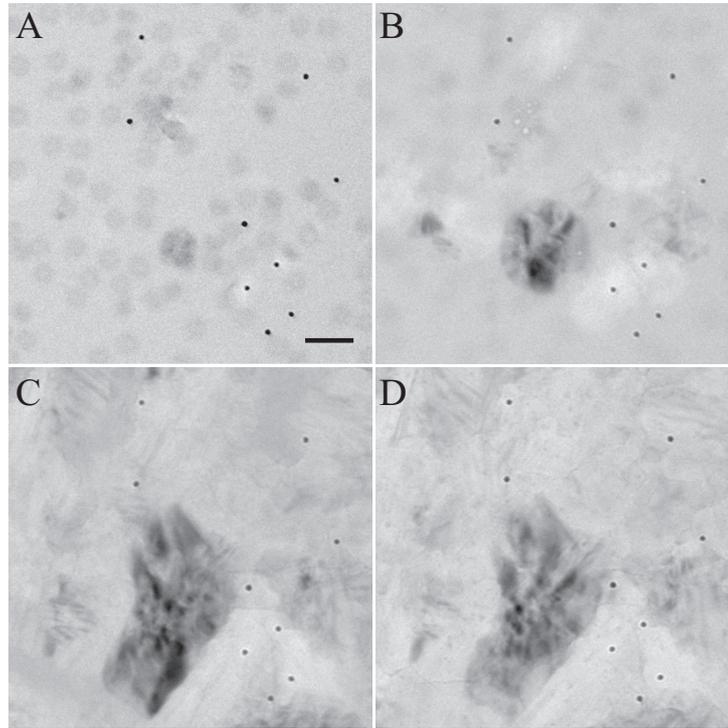


Fig. 5.5. Beam-induced ice crystallization at high-flux rate ($50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$). Images (A-D) are shown after a dose of 57, 565, 1695, and 2825 MGy, respectively. The beam-induced movements calculated for the fiducial gold markers in these images is surprisingly small (see Table 5.1, grid 1, high-flux series). The scale bar shown in (A) corresponds to 60 nm.

5.3.7 Dose-rate effects

Fig. 5.4 shows the summed image of an aligned low-flux (Fig. 5.4A), medium-flux (Fig. 5.4B), and high-flux short-exposure (Fig. 5.4D) series of a control set of Hb for an equal integrated flux of $50 e^{-\text{\AA}^{-2}}$. As a comparison, the first image of a high-flux series ($50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$, 1 s exposure) collected from the same grid is also shown (Fig. 5.4C). Figs. 5.2E-H show respective images for an integrated flux of $250 e^{-\text{\AA}^{-2}}$. It can be seen that for an equal accumulated incident flux, the images of high-flux series are invariably blurred. Furthermore, for an equal integrated flux, the appearance of gas bubbles (data not shown) occurred earlier in the high-flux short-exposure series compared with the medium- and low-flux series.

Figs. 5.2C and 5.2C show FOM plots for the low-flux and high-flux short-exposure series, respectively. The identical dose was fractionated over the same number of images as plotted in Figs. 5.2A and 5.2B. Both graphs start with comparable FOMs at low resolution, but fewer high-resolution details can be seen for the high-flux short-exposure series. The low-flux series (5.2C) shows high resolution details, although in absolute terms, all FOMs are smaller compared with the medium-flux series, probably due to an accumulation of alignment errors for the fivefold larger number of images. We measured camera statistics such as readout noise and dark current [47] and note that these sources of error are relatively small even for the low-flux series.

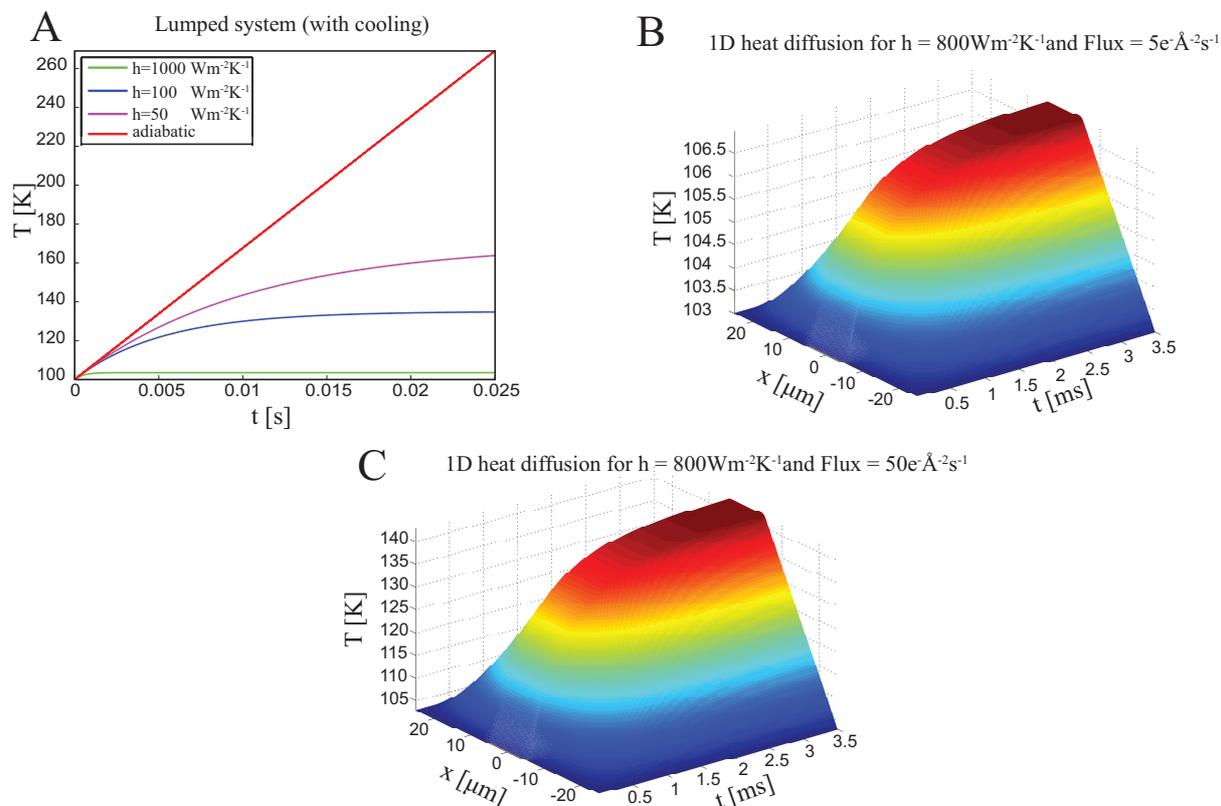


Fig. 5.6. Simulations of the heating of a vitreous sample by the electron beam. (A) Temperature *versus* time plot for a lumped system model for three different heat transfer coefficients h . The adiabatic model temperature rise is shown in comparison. One-dimensional heatdiffusion plots for a distributed model are shown for (B) medium- and (C) high-flux incident beam for a low heat transfer coefficient of $800 \text{ Wm}^{-2}\text{K}^{-1}$.

5.3.8 Heating effects

Electron irradiation could induce crystallization in the sample, as observed for the high-flux exposure series on the 50 mM NH_4Ac sample (see Fig. 5.5). In another high-flux series during which 100 images were collected, ice crystallization was not observed: instead, dark flake-like particles appeared prior to a complete sublimation of the illuminated area at a cumulative dose of 5500 MGy (Movie S1). Could this crystallization be due to sample heating?

The calculated dose for the parameters given in section 5.3.1 is 5.5 MGy. Heating simulations for a sample treated as a lumped system are shown in Fig. 5.6A for different values of h and compared with those for an adiabatic process. Figs. 5.6B and 5.6C show the temperature distribution calculated from the distributed model (Eq. (5.6)) for the medium (5.5 MGys^{-1}) and high-flux series (55 MGys^{-1}). The simulated temperature rise is strongly dependent on the incident flux and on the cooling rate given by the heat transfer coefficient h . For $h = 800 \text{ Wm}^{-2}\text{K}^{-1}$, the temperature is predicted to rise within milliseconds from 103 to 140 K when using the high-flux of $50 e^{-19} \text{ A}^{-2}\text{s}^{-1}$.

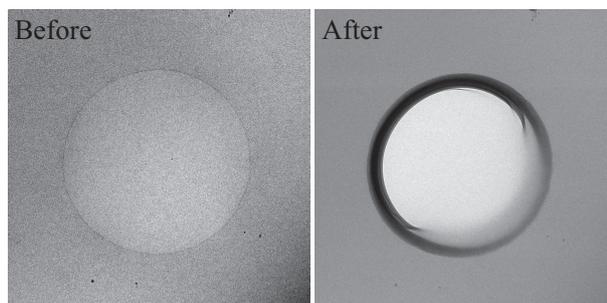


Fig. 5.7. Low magnification image of Hb sample in 50 % (v/v) glycerol before and after the collection of a medium-flux series data collection. The hole (sized $1.2 \mu\text{m}$) in the carbon support film contains a thin layer of vitreous sample in the before image, which is completely destroyed after the collection of 50 images, corresponding to a dose of 270 MGy.

5.3.9 The role of solvent constituents

Four different solvent constituents were used: 50 mM NH_4Ac , 2 M NH_4Ac , 50 % (v/v) glycerol, and 0.2 % (v/v) glutaraldehyde. Among these, the images of the higher-density glycerol sample (Fig. 5.7) showed less contrast as compared with other samples, although the specified defocus was the same for all the exposure series collected. The beam-induced movements were excessive for the medium-flux series of the glycerol sample: these movements occurred concurrently with the formation of gas bubbles. At high-flux, gas bubbles formed on all the samples. Among the solvents studied, gas bubble formation within the high-flux series was most clearly localized at the protein sites for the 0.2 % glutaraldehyde sample (Fig. 5.8).

5.4 Discussion

5.4.1 Dose

The incident flux is a poor metric to use for radiation damage studies, as the probability of sample-electron interaction does, apart from the incident flux, depend on the integration time, the sample, and the electron energy. Whereas an older paper on specimen damage [291] calculates the absorbed dose in energy per mass unit (erg/g , $1 \text{ erg} = 10^{-7} \text{ J}$), most recent electron microscopy papers use $e^{-\text{\AA}^{-2}}$ as the unit for dose. Analogous to dose calculations done for MX [266], we estimated the absorbed dose in grays based on the electron beam energy and size, the protein concentration, sample thickness, incident flux, exposure time, and tabulated inelastic scattering coefficients.

The typical integrated fluxes used in single particle cryo-EM range between 15 and $25 e^{-\text{\AA}^{-2}}$ [229]. For example, [227] recorded micrographs at approximately $25 e^{-\text{\AA}^{-2}}$ for the 3.3\AA reconstruction of a primed aquareovirus. In [292] single frame images were taken at $15 e^{-\text{\AA}^{-2}}$ for the study of kinesin-microtubule complexes, whereas [280] used $25\text{-}36 e^{-\text{\AA}^{-2}}$ for bacteriorhodopsin and $\epsilon 15$ bacteriophage. The typical integration time is 1 s, although the latter authors used 1.4 and 2 s. For helical reconstruction or cryo-electron tomography studies, a larger integrated flux

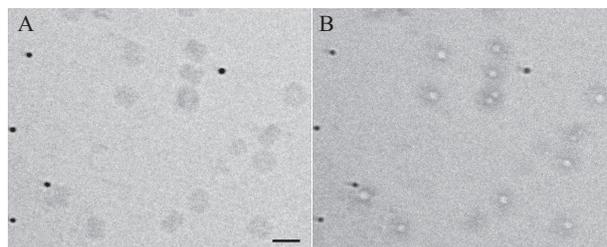


Fig. 5.8. Representative high-flux images from (A) the 1st, and (B) the 12th exposure from a 0.2 % glutaraldehyde sample. Clear bubbling can be observed on every protein particle. The scale bar shown in (A) corresponds to 30 nm.

is used, corresponding to 40-100 $e^- \text{Å}^{-2}$ [292, 293] or even 24-150 $e^- \text{Å}^{-2}$ [294]. In tomography, the dose is divided over a large number of images [295].

Table 5.1 shows the relation between incident flux, integration time, and dose, for the data presented here. The dose used to record the individual images of the medium-flux series with an incident flux of $5 e^- \text{Å}^{-2} \text{s}^{-1}$ varies between 5 and 6 MGy. For our sample and the electron energy used, the *Henderson* dose limit (20 MGy: [269]) and the *Garman* dose limit (30 MGy: [265]) would correspond to an integrated flux of ~ 20 and $\sim 30 e^- \text{Å}^{-2}$ respectively. Unlike MX, cryo-EM offers a unique way to study the decay of macromolecules at doses that exceed these limits by at least one order of magnitude (Movie S1).

Fig. 5.3 illustrates the gradual alteration of the low-resolution information in our data. In here, the medium-flux data are combined in groups of three images, corresponding to an integrated flux of $15 e^- \text{Å}^{-2}$ per combined image. According to the criteria of the FRPR function being less than 45° and the FRC value being larger than 0.5 [130, 290], one could combine these low-resolution data up to 125 MGy. However, these criteria would indicate that one could also combine data from e.g. 50-150 MGy yielding good statistics on radiation-damage compromised particles. The main cause of loss of correlation is the spread in radiation-damage-induced particle conformations.

Eq. (5.1) gives an empirical relation between radiation damage, dose and resolution for MX studies. According to this formula, the same fractional loss of diffracted intensities is obtained for constant ratios of dose over resolution length. Thus radiation damage observations obtained with high dose at low resolution would also be of relevance for lower dose at high resolution. Cryo-electron tomography is, compared with SP cryo-EM, a lower-resolution technique that is performed with a higher dose. We hypothesize that a relation similar to Eq. (5.1) exists for SP cryo-EM: the rate of loss of signal at high dose at low resolution is likely to be related to the rate of loss of signal with low dose at high resolution.

We would advocate the collection of data series (*stroboscopic data collection*; [296]) rather than individual images, with a dose ranging from e.g. 5 to 100 MGy. For particle picking, radiation damage is less of a problem. In fact, the gas bubble formation observed at a higher dose could even be helpful in locating the particles (Fig. 5.8). Radiation-damage compromised images might still be useful for alignment, as a minimum dose is required to align particles of a certain size to a certain resolution [229, 297]. Constant-dose interpolation schemes could be explored for stroboscopic data, similar to that which has been implemented in MX [298].

Alternatively, only the very first few images from each dose series could be used in the final reconstruction steps, providing a minimal and controlled amount of damage.

Radiation damage should not be treated as a binary nuisance, neither in MX nor in SP cryo-EM. Right from the first exposure of the sample to ionizing radiation, structural changes will occur [255–257]. The amount of dose is the main determinant of the amount of radiation damage.

5.4.2 Dose-rate effect

We observed a dose-rate effect, in accordance with [280] who introduced a LINDA imaging protocol: Low Intensity aNd low Dose Acquisition. For the same integrated flux of $15 e^{-\text{\AA}^{-2}}$, [280] compared data that were collected with an incident flux of $15 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ for 1 s (HiFlux) with data collected at $1.5 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ for 10 s (LINDA). Reconstructed models from successive data sets showed fewer signs of radiation damage for the data that were collected with the LINDA protocol compared with the HiFlux data. The 10 s data collection poses strict requirements on the cryo-stage of the electron microscope, as the sample should move less than a fraction of a pixel (e.g. $< 1 \text{\AA}$) within that time. In [280] FEI Polara microscope was used, whereas our analyzes were based on data that were collected with a more common electron microscope, a FEI Tecnai T12. The mechanical drifting of the stage was overcome by dose-fractionation. Fig. 5.4B shows a summed image for the medium-flux data, where 10 images of $5 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ with 1 s integration time were aligned and added. The low flux series (Fig. 5.4A) where 50 images of $1 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ with 1 s integration time were added, showed less detail, possibly because of the accumulation of alignment errors due to the low signal-to-noise ratios in the individual images. Fig. 5.4C comes from a single image, taken at $50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ with a 1 s integration time. This figure is representative of all high-flux series which never showed great detail. Adding ten aligned high flux images ($50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$) recorded with short exposure times (0.1 s) did not show clear improvements. Figs. 5.4E-H show corresponding images for an integrated flux of $250 e^{-\text{\AA}^{-2}}$.

A more quantitative analysis of these images is shown in Fig. 5.2. We introduced a new metric, analogous to MX, for ascertaining phase qualities, namely the average cosine of phase errors (FOM). The FOM plots enabled us to estimate the defocus values from the images taken from the vitreous sample area that excludes any carbon support (Fig. 5.2B).

The medium-flux series (Fig. 5.2A) shows the most detail at higher resolution compared with the high-flux (Fig. 5.2D) and low-flux (Fig. 5.2C) series. Unlike the high-flux series, there is still signal between the first and second zero of the CTF (Fig. 5.2B) in the low-flux series (Fig. 5.2C). We believe that this signal has been dampened due to an accumulation of alignment errors for the larger number of images used in the low-flux series, a problem that would be overcome by the LINDA protocol. Alternatively, use of larger fiducial markers combined with more sophisticated alignment schemes could help when the data are fractioned over a larger number of images.

Fig. 5.2A shows the medium-flux series, grouped in subsets of 10 images corresponding to an integrated flux of $50 e^{-\text{\AA}^{-2}}$. The peaks observed in this graph relate to the radial averaged Fourier transform of the hemoglobin particles. They are most pronounced for the medium-flux

series (Fig. 5.2A), demonstrating that this series not only provides higher resolution data, but also better signal at lower resolution. The peaks quickly reduce in height as a function of dose: at higher resolution the loss of signal is faster than at lower resolution.

Larger beam-induced movements of the fiducial markers for the high flux series were observed in [280]. This effect is not obvious from our data Table 5.1 as the beam-induced movements, measured between the first and last image of each series and given in pixels, are scattered. A typical value is five pixels, although the range is between one and nine pixels. In [280] 10 nm gold particles were used and frames recorded with $15 e^{-\text{\AA}^{-2}}$, whereas we used 5 nm gold particles and frames recorded with 1, 5, or $50 e^{-\text{\AA}^{-2}}$. The uncertainty in fiducial marker localization is larger for most of our data. The fiducial marker distances between successive images varied between two pixels for the low-dose series and one pixel for the high-flux series. For some of the data series, part of the beam-induced movements could be modeled with global parameters such as scaling or rotation. It is remarkable how little the fiducial markers move in the extreme case of ice crystallization (Fig. 5.5), challenging the credence of using differences in gold positions distances as a metric for beam-induced movements.

Analogous to MX studies [260,262], we discuss two possible causes for the observed dose-rate effect: excessive heating and radical recombination.

Sample heating

Sample heating could cause dose-rate effects, since the balance between heating by the electron beam and cooling by conduction will depend on the rate the energy is deposited in the sample. Analogous to [248] and [249], we simulated the heating of the vitrified sample using a lumped and a distributed model for different values of the heat transfer coefficient h . In the adiabatic case, a thermally isolated sample of the same size as the beam would melt quickly (Fig. 5.6). Both the lumped and the distributed models indicate that the temperature will rise most rapidly within the first milliseconds after exposure of the sample to the electron beam. Compared to MX, the system time constant (Eq. (5.4)) is much smaller in SP cryo-EM due to the lower volume-surface ratio and the larger heat transfer coefficient. Figs 5.4D and 5.4H seem to indicate that fast (sub-100 ms) processes are indeed responsible for the observed dose-rate effects. The images from the series of Fig. 5.4B and Fig. 5.4D were recorded with the same integrated flux per image, namely $5 e^{-\text{\AA}^{-2}}$, however, the images from Fig. 5.4B were integrated over 1 s at $5 e^{-\text{\AA}^{-2}}$ whereas the images from Fig. 5.4D were integrated over 0.1 s at $50 e^{-\text{\AA}^{-2}}$. The latter images are clearly worse, indicating that the additional damage induced by the high flux occurs in less than 100 ms.

Only for very high dose rates and low values of h , representing e.g. poor thermal contact between the grid and the cryoholder, is sample heating predicted to become an issue for SP cryo-EM, as the temperature of the sample is calculated to rise (Fig. 5.6C) above the glass transition [277,299], triggering an exothermic ice crystallization process. In fact, for one high-flux series, radiation-induced ice crystallization was observed (Fig. 5.5). However, this result was exceptional, suggesting poor thermal contact for that particular grid.

The heat model presented here complements existing specimen heating models used in TEM (see e.g. [111]) and could form the basis for an elaborate refinement that studies the influence

of supporting mesh size, size and spacing of holes within the support film, distance of the beam with respect to the grid bars, etc. Some experimental verification of h for different combinations of grids and holders would be required [111]. Such studies are beyond the scope of this manuscript; however, we can postulate that the effect of beam heating is felt within milliseconds after exposure, and beam heating is not expected to be a problem for cryo-EM samples with good thermal contact at medium- or low-fluxes.

Radical recombination

In MX, it is believed that the photo-electric absorption of a ~ 12 keV X-ray photon will produce ~ 500 radicals, assuming 25 eV per ionization event [251]. For our medium-flux data series taken at $5 e^- \text{Å}^{-2}$, we estimated 1.9×10^{10} inelastic scattering events per frame within a volume of 11.8 fl. If each inelastic event acts on a different target and produces one radical and ignoring radical recombination processes, then the radical concentration at the end of the first exposure would be 2.6 M. We extrapolate that for typical SP cryo-EM data collections, the biological molecules would be exposed to molar concentrations of radicals. Some of these radicals, in particular electrons, must be mobile [256] as the damage seems to accumulate at the interface of protein sites (Movie S1; Fig. 5.8; [138,236]).

Ignoring radiation recombination processes, one would calculate 52 M as the radical concentration for the high-flux series after 2 s of exposure, which is comparable to the concentration of water within the sample. Such radical concentrations are unlikely to be present, thus radical recombination must play a role for our data.

Dose-rate effects could be caused by concentration-dependent radical chemistry and diffusion of gas molecules within the sample. Supplementary Movie S1 illustrates the formation, diffusion, fusion, and rupture of these bubbles. For high-intensity beams, the pressure can become so high that it generates mechanical fractures within the specimen [280], and since this would negatively effect the conductive cooling of the sample, it might lead to local beam heating. The absolute temperature of the sample could play a role for dose and dose-rate effects: recently, a temperature of 50 K instead of 100 K was shown to reduce specific damage in MX by a factor three to four [247], whereas for cryo-EM diffraction studies, 100 K was found to be the optimal temperature [281]. Higher dose-rates could also lead to an inverse dose-rate effect, as radical recombination could become more important, in particular at elevated temperatures [260]. The dose-rates used in this SP cryo-EM study varied between 1 and 56 MGy/s, which is very high compared with the dose-rate studies carried out in MX (e.g., in [260], 6–10 MGy/s was used; in [264] 0.2 MGy/s was used). The data recorded with 56 MGy/s were inferior to the lower dose-rate series. This raises the question as to whether the typical dose-rates used in SP cryo-EM (~ 25 MGy/s), is optimal. It would be worth investigating whether further improvements could be obtained by lowering the dose-rate in SP cryo-EM studies by another order of magnitude. Simulations suggest that it should be possible to align extremely low-dose images for essentially noise- and point-spread function-free detectors [296]. Actual developments in detector technology yield promise for dose-fractioning in SP cryo-EM.

The high dose rates used in SP cryo-EM make it likely that radiation chemistry will play an even larger role compared to MX. There are indications that scavengers could prolong the

lifetime of cryo-cooled crystals in the X-ray beam [252, 300, 301] by neutralizing immobile ionized groups or quenching radical species. Unfortunately, the addition of a high concentration of scavengers can be harmful for fragile protein crystals. This difficulty does not exist in SP cryo-EM, although other problems, such as reduced sample contrast, might arise.

Hydrogen trapping was proposed [247] to be the cause of unit-cell volume expansion observed in MX [262]. In SP cryo-EM, the sample shrinks with dose, as radiolytic products, in particular hydrogen gas, diffuse out of the sample into the high-vacuum column of the electron microscope, resulting in mass loss. This process is linear with dose and seems to be highly reproducible among different samples tested (Fig. 5.1). The observed linear relationship between the relative intensity change and the dose could be a useful metric for studying the effects of scavengers.

Other metrics presented in this manuscript include FOM (Fig. 5.2), Fourier ring correlation and Fourier ring phase residual (Fig. 5.3), and beam-induced movements (Table 5.1). Here, radioprotectants were not tested, but rather one fixative and two cryoprotectants, among which was glycerol, the most widely used cryoprotectant in MX. The 50 % glycerol sample showed very little contrast between the protein and the solvent as its density (1.181 g/cm^3 at 72 K, [177]) is comparable with the average density of protein molecules (1.35 g/cm^3). Bubbling was observed throughout the glycerol sample, not only at the protein sites, consistent with the discussion by [247] that hydrogen gas [28] is formed upon radiolysis of organic molecules. Within 50 medium-flux images, the vitrified layer of the sample within the hole was completely sublimated (Fig. 5.7), unlike the other samples at medium-flux (Figs. 5.4E-5.4H). The gold fiducial markers showed large beam-induced movements (Table 5.1). The observed increased sensitivity to radiation damage upon addition of glycerol calls for further studies, in particular for MX.

The 2 M NH_4Ac sample did not show clear differences in radiation damage susceptibility: the relative intensity change (Fig. 5.1) and beam-induced movements (Table 5.1) were comparable to the low-salt samples. The distribution of the Hb particles within the sample was slightly different, as some Hb particles packed regularly. Similar to all the other samples, the 2 M NH_4Ac sample was vitrified in liquid ethane. High concentrations of salt are routinely used as cryoprotectants in MX: we could have vitrified this sample with liquid nitrogen thus overcoming some of the disadvantages of using liquid ethane.

The localized appearance of gas bubbles was most obvious for the 0.2 % glutaraldehyde sample (Fig. 5.8). The research in [284] advocated the use of 0.2 % glutaraldehyde for improving the sample quality for structure determination by SP cryo-EM. The described benefits of using a chemical fixation reagent in stabilizing individual macromolecules during sample preparation might also help in keeping the macromolecules together upon radiolysis.

5.5 Conclusions

Radiation damage should not be treated as a binary nuisance. It gradually changes the quality of SP cryo-EM data: the amount of alteration that is acceptable depends on what one aims for, for example, for particle picking or defocus estimation, a larger dose could be used compared

to the calculation of a 3D reconstruction. We advocate the use of stroboscopic data collection, with which variable amounts of dose can be used for the different steps of SP reconstruction.

Throughout this paper, the gray is used as the unit of dose. It is estimated from the incident flux, beam size, sample composition and thickness, and beam energy. The use of this unit provides direct access to the power deposited in the sample, which has been used for beam heating simulations. Furthermore, it allowed us to make direct comparisons with systematic radiation damage studies in MX, yielding, among other parameters, an upper estimate of the radical concentrations formed during cryo-EM experiments.

The usual dose applied in SP cryo-EM to collect single images is similar to the experimental dose limit for MX (30 MGy [265]) that is typically used to collect an entire data set of hundreds of diffraction images. These high doses in SP cryo-EM are necessary to counteract the low signal-to-noise ratios, but will inevitably cause radiation damage issues. The use of dose (in grays) is expected to be of help in characterizing the exact extent of these issues now that higher resolution SP cryo-EM studies are more frequently being performed. Unlike MX, SP cryo-EM could offer a unique insight into the later stages of radiation damage to macromolecules, as one could continue to record SP cryo-EM data at doses that exceed 30 MGy by at least one order of magnitude.

A clear dose-rate effect could be observed, favoring lower flux-rates. Data that were collected with an incident flux of $50 e^{-\text{\AA}^{-2}\text{s}^{-1}}$ were inferior to those that were collected at $5 e^{-\text{\AA}^{-2}\text{s}^{-1}}$. Beam heating simulations indicate that:

- the effect of beam heating is felt within milliseconds after exposure, and
- beam heating is not expected to be a problem for cryo-EM samples with good thermal contact at medium- or low-fluxes.

The electron beam deposits enough energy to form molar concentrations of radicals and radical recombination is likely to play a role in the observed dose-rate effects. This gives hope for future scavenger studies. A number of metrics have been presented, such as relative intensity change *versus* dose, FOM, FRC, FRPR, and beam-induced movements, which could aid such studies.

Acknowledgements

Authors kindly acknowledge Cristina Avramut (LUMC) for assistance in preparation of hemoglobin, Frank Faas and Remco Schoenmakers for high-throughput data collection tools, and Montse Barcena for discussions.

Table 5.1. Mass loss upon electron-beam irradiation.

Solvent, estimated number of Hb molecules (Hb/ μm^2), sample thickness (nm)	Incident flux ($e^- \text{Å}^{-2} \text{s}^{-1}$), integration time (s), dose per exposure (MGy)	I_0 (ADU)	$\Delta I/I_0$ versus dose ($10^{-10} \text{ Gy}^{-1} \dagger$)	Beam-induced movements [pixel (std)] \ddagger
50 mM NH_4Ac , 500, 140	5, 1, 5.7	1420	0.915	4.2 (3.0)
		1457	0.937	2.1 (1.6)
		1240	0.921	4.9 (3.7)
	1, 1, 1.1	306	1.02	3.6 (2.7)
		307	0.949	2.8 (2.1)
		298	1.03	5.3 (4.4)
50 mM NH_4Ac , 520, 160	50, 1, 56.5	18543		0.9 (0.8)
	5, 1, 5.4	1056	0.515	2.3 (1.8)
		1028	0.479	5.5 (3.0)
		1027	0.453	5.2 (3.5)
	1, 1, 1.1	998	0.456	5.5 (3.3)
		281	0.968	2.8 (2.0)
289		0.966	3.1 (3.2)	
50 mM NH_4Ac , 700, 200	5, 1, 5.0	281	0.969	2.6 (1.9)
		1234	0.719	2.8 (2.1)
		1134	0.735	2.9 (2.1)
	1, 1, 1.0	1184	0.700	5.0 (3.4)
		406	1.97	8.8 (5.4)
		429	1.65	8.8 (5.4)
50 mM NH_4Ac , 570, 150	5, 1, 5.5	428	1.63	7.3 (4.4)
		2083	0.898	2.0 (1.8)
		2025	0.887	2.9 (2.1)
		2075	1.07	3.7 (2.8)
		2132	0.953	2.4 (2.0)
		2115	0.978	2.7 (1.8)
	50, 0.1, 5.5	1996	0.862	6.9 (4.9)
		2009	0.707	5.6 (4.5)
		1886	0.905	2.5 (2.2)
		1910	0.860	3.4 (2.3)
		1987	0.878	2.7 (2.0)
		19359		0.9 (0.7)
50 mM NH_4Ac , 691, 200	50, 1, 55.4	1422	0.956	3.3 (2.7)
	50, 0.1, 5.0	1403	0.874	3.0 (3.0)
		1369	1.11	1.9 (1.6)
		1304	1.10	3.1 (1.4)
		1396	1.07	4.2 (4.0)
		1494	1.03	5.0 (4.0)
2 M NH_4Ac , 1120, 240	5, 1, 4.7	1517	0.973	
		1452	0.928	
		1469	0.976	
		1458	0.942	
		1547	0.217	8.2 (6.1)
		1417	0.384	
0.2 % glutaraldehyde, 120, 150	5, 1, 5.5	1593	0.291	
		1547	0.217	
		1417	0.384	
50 % glycerol §, 280, 150	5, 1, 5.4	1566		53.9 (89.5)¶

\dagger The correlation coefficients for all linear fits are around 0.99. \ddagger First and last images within each series were compared. \S The glycerol sample showed large movements of the fiducial gold markers.

$\¶$ The two values correspond to the first to the fifth, and the first to the tenth image of the series, respectively.

Appendix H

Calculation of dose

Dose is expressed as energy deposited per unit mass

$$D = \frac{\Delta N_{\text{in,e}} \Delta E}{m_{\text{tot}}}, \quad (\text{H.1})$$

where $\Delta N_{\text{in,e}}$ is the number of inelastically scattered electrons each depositing on average ΔE energy into the specimen with mass m_{tot} . As the dominant form of inelastic scattering, only plasmons are considered (assuming $\Delta E = 20$ eV [158]).

If the incident electron beam is described by diameter d_b , electron flux Φ_0 ($e^- \text{Å}^{-2} \text{s}^{-1}$) and exposure time t_{exp} , then the number of incident electrons equals

$$N_{\text{e,0}} = \Phi_0 t_{\text{exp}} \pi \left(\frac{d_b}{2} \right)^2. \quad (\text{H.2})$$

Analogously to Beer-Lambert law, the number of inelastic scattering events from a layer with thickness l can be expressed as

$$\Delta N_{\text{in,e}} = N_{\text{e,0}} \left[1 - \exp\left(-\frac{l}{\Lambda_{\text{in}}}\right) \right], \quad (\text{H.3})$$

where Λ_{in} is the total inelastic mean free path defined as

$$\frac{1}{\Lambda_{\text{in}}} = \frac{1}{V} \sum_Z N_Z \sigma_{\text{in}}(Z), \quad (\text{H.4})$$

with N_Z corresponding to the total number of atoms of type Z (atomic number) and $\sigma_{\text{in}}(Z)$ representing inelastic scattering cross section of that atom type. V is the volume of the specimen irradiated by the electron beam which can be calculated as

$$V = l \pi \left(\frac{d_b}{2} \right)^2. \quad (\text{H.5})$$

The volume occupied by molecules is

$$V_{\text{mol}} = N_{\text{mol}} V_{1\text{mol}} = N_{\text{mol}} \frac{M_{\text{mol}} \rho_{\text{mol}}}{N_A}, \quad (\text{H.6})$$

Table H.1. Number of H, C, O and N atoms in a hemoglobin protein (Hb). The number of residues in one Hb is 31824. The contribution of other atoms such as S and Fe have been ignored.

Z atom	average # of Z atoms in one residue	total # of Z atoms in a Hb
H	8	254592
C	4.869	154951
O	1.492	47481
N	1.351	42994

where N_{mol} is the total number of protein particles (e.g. 1000 in $1 \mu\text{m}^2$ area, Table 5.1), V_{mol} is the volume of one protein molecule, M_{mol} is molecular mass of one molecule (e.g for hemoglobin (Hb) $3.5 \cdot 10^6$ Daltons), N_A Avogadro's number, and ρ_{mol} average density of a protein molecule (1.35 g/cm^3 [159]).

From N_{mol} one can estimate the total number of hydrogen, carbon, oxygen, nitrogen and other atoms, by multiplying the total number of amino acids in one protein with average number of corresponding atoms (see Table H.1). If macromolecule is not post-translationally modified and its primary sequences are known, one can calculate actual total number of each atom type.

The solvent is considered to be vitreous ice with density of $\rho_{\text{sol}} = 0.93 \text{ g/cm}^3$. The volume occupied by solvent is $V_{\text{sol}} = V - V_{\text{mol}}$ and the total number of solvent molecules is

$$N_{\text{H}_2\text{O}} = \frac{V_{\text{sol}} \rho_{\text{sol}} N_A}{M_{\text{H}_2\text{O}}}. \quad (\text{H.7})$$

The solvent atomic contents are therefore: $N_{\text{sol,H}} = 2 \times N_{\text{H}_2\text{O}}$ and $N_{\text{sol,O}} = 1 \times N_{\text{H}_2\text{O}}$.

Total number of atoms of type Z in the specimen is then

$$N_Z = \begin{cases} N_{\text{mol,Z}} + N_{\text{sol,Z}} & Z \in \{H, O\}, \\ N_{\text{mol,Z}} & \text{otherwise.} \end{cases} \quad (\text{H.8})$$

Total mass of the specimen being irradiated can be expressed as

$$m_{\text{tot}} = \frac{\sum_Z N_Z M_Z}{N_A}, \quad (\text{H.9})$$

where M_Z is the molar mass of the atom with atomic number Z.

As described in [158] and [163] inelastic scattering cross sections can be calculated as

$$\sigma_{\text{in}}(Z) = \frac{1.5 \cdot 10^{-6} Z^{\frac{1}{2}}}{\beta^2} \ln \frac{2\beta^2(U_0 + mc^2)}{\Delta E} [\text{nm}^2], \quad (\text{H.10})$$

where β the ratio between the velocity of the electron and light ($\beta^2 = 1 - [mc^2/(U_0 + mc^2)]^2$), U_0 the incident electron energy, mc^2 the rest energy of electron.

For each atom type, the individual inelastic cross-section is calculated from Eq. (H.10). Combining Eqs. (H.5), (H.8) and (H.10) into Eq. (H.4) results in the total mean free inelastic path. Eq. (H.3) provides the total number of inelastic scattered electrons after which the dose is calculated *via* Eq. (H.1).

Chapter 6

Conclusions and perspectives

Quantitative forward modeling of image formation in transmission electron microscopy (TEM) of frozen-hydrated biological specimens is becoming increasingly important in order to optimize the data acquisition strategy, facilitate 3D reconstruction schemes, provide insight into ways to ameliorate instrumentation, improve image interpretation and ultimately, provide higher resolution structures. In order to create such a forward model and furthermore assist the contrast transfer function (CTF) correction and/or regularization of the 3D reconstructions it is necessary to: i) construct the electron-specimen interaction potential based on elastic and inelastic electron scattering properties and adequately describe electron propagation through the specimen; ii) characterize TEM detectors including all relevant statistics; iii) accurately estimate the CTF parameters, in particular defocus and astigmatism and their uncertainties. Since radiation damage limits the allowable electron flux used for imaging, better understanding of certain aspects of radiation damage such as specimen heating, dose-rate effects, and beam-induced movements is needed.

The research described in this thesis addresses the aforementioned points and includes: i) forward modeling of image formation in cryo-electron microscopy (cryo-EM); ii) methods for characterization of TEM detectors; iii) methods to accurately estimate defocus and astigmatism, and iv) an investigation of dose and dose-rate effects. The investigated methods and developed tools form a part of a larger project that aims to improve the resolution at which biological structures can be studied with cryo-EM. Below, I recapitulate the main conclusions of chapters 2-5 and provide an outlook.

Forward model

Simulations of TEM images of biological specimens, as implemented in a number of software packages [31–38], are oversimplified in one or more of the following points: the generated specimen volume does not represent actual physical electron-specimen scattering properties (potential), the thickness of the specimen is often ignored, the influence of the solvent is not accounted for, and accurate detector properties are not accounted for. Existing simulations have been mainly used for the determination of particle orientation in single particle analysis (SPA), contrast transfer function (CTF) correction, and evaluation of reconstruction algorithms in SPA

and electron tomography (ET). For optimization of data collection strategies, those models are insufficient. Furthermore, a (proper) validation of the simulations against experimental data was lacking.

The main goal of this project has been to improve our understanding of the relevant physical processes that govern image formation and to develop a quantitative forward model. I present such model in Chapter 4. It is based on physical principles and takes into account the influence of the specimen and its surroundings, the optics of the microscope as well as the noise and signal transfer by the camera. The model has been validated by comparing simulated and experimental images of 20S proteasome, hemoglobin, and GroEL for various microscope settings and different experimental conditions studied. The main conclusions are:

- Simulated images adequately predict the observed phase contrast introduced by defocusing, changes due to the integrated electron flux, influence of inelastic scattering, camera's detective quantum efficiency (DQE), and acceleration voltage.
- The dominant part of the interaction potential can be calculated *via* electron scattering factors using the isolated atom superposition approximation ($\mathcal{V}_{\text{atom}}$). This potential has been extended for the influences of the solvent, ions and molecular interactions that cause an effective charge redistribution. The potential from these redistributions ($\mathcal{V}_{\text{bond}}$) is computed via a Poisson-Boltzmann (PB) continuum electrostatics approach. The conditions under which it is possible to combine the potential directly constructed from the atomic model ($\mathcal{V}_{\text{atom}}$) with portions obtained from a continuous approach ($\mathcal{V}_{\text{bond}}$) are provided in Appendix F. The contribution of $\mathcal{V}_{\text{bond}}$ is shown to be less than 10 % compared to $\mathcal{V}_{\text{atom}}$. Its influence on the final image was not significant due to the low signal-to-noise ratio (SNR) at which we recorded our experimental data. In simulations, a higher acceleration voltage and/or a higher integrated flux reveals a higher level of detail and therefore slightly larger differences between $\mathcal{V}_{\text{atom}}$ and the combined interaction potential ($\mathcal{V}_{\text{atom}} + \mathcal{V}_{\text{bond}}$).
- We suggest that beam-induced specimen movements are relevant (4-10 Å) and appear to be stronger for higher integrated electron fluxes. The apparent motion factor magnitudes were smaller at 300 kV than at 80 kV which could be related to numerous effects including differences in inelastic cross-sections, beam quality, or ice thickness. The derived motion factors are similar to the displacement values reported in [26, 183]. The varying contrast of the particles within a field of view can be explained by the apparently space-variant beam-induced movements.
- For typical electron fluxes in cryo-EM, the influence of the amorphousness of the solvent (modeled *via* molecular dynamics simulations) can be neglected since shot noise is the dominant source of noise in the image and consequently, the solvent can be modeled as a continuum.
- The detective quantum efficiency (DQE) of the detector must be included in the model in order to properly describe the noise and signal transfer. Ignoring the DQE results in simulated images with an erroneously higher SNR.

-
- Intensity profiles across carbon edges and carbon nanotubes provide another source of data to quantitatively validate our model. The simulated and experimental images of such specimens exhibit similar modulations.
 - Practical criteria for applicability of the weak-phase object approximation (WPOA) and projection assumption (PA) in phase contrast cryo-EM are introduced and they motivate the existence of four methods (in addition to the multislice approach) for description of the electron wave propagation through the specimen.
 - At an acceleration voltage of 80 kV and a potential map with a voxel size of 3 Å, the earth worm hemoglobin sample can be considered as a weak-phase object, while the projection assumption is not satisfied. This implies that the thickness cannot be neglected in modeling such specimen.
 - Simulating the interaction potentials of an amorphous sample facilitates the choice of the appropriate model used for the electron wave propagation through a biological sample (macromolecule). The strength, frequency content and thickness of the interaction potential map determine if the interaction should be described *via* WPOA, PA, both WPOA and PA, thick-phase grating (TPG) or *via* a multislice approach.
 - For validation of the forward model, some parameters that influence the image formation such as defocus, astigmatism and camera parameters must be estimated independently from experimental data. For most of them we used the tools developed and described in Chapters 2 and 3.

TEM detectors

Although the choice of the detector significantly influences the image quality, in previous image simulation work it has been either neglected or empirically introduced. The quality of a detector is not easily accessible. Different manufacturers provide different types of figures of merit when advertising their detector. A comprehensive characterization of the detector including all relevant noise contributions, modulation transfer function (MTF), and DQE is essential for an accurate forward model and eventually for 3D reconstruction. Furthermore, correction of the fixed pattern noise based on poor statistics would spoil image interpretation. A general methodology for characterizing TEM CCD detectors is presented in Chapter 2. The methods can be extended to the new generation of direct electron detectors. The main conclusions are:

- A set of algorithms have been developed to allow TEM users to characterize, based on raw images, the aforementioned properties of their CCD detector.
- Three in-house CCD cameras were characterized, yielding, statistics for hot and bad pixels, modulation transfer function, conversion factor, effective gain, detective quantum efficiency, as well as readout and dark-current noise.
- Two detectors with different MTFs and conversion factors can have very similar DQE curves. The effects of a slightly inferior MTF of the detector with a thicker scintillator layer are, at the same time, balanced by a higher conversion factor.

- Fixed-pattern noise could be fully corrected by using large sets of dark and white reference images. Unfortunately, the noise patterns appear to drift in time, possibly due to unstable cooling of the detectors, thereby limiting the useful lifetime of these reference sets.
- The need for using a lookup table of bad pixels is demonstrated. Systematic outliers (*hot* and defect pixels) hamper the alignments of images and could cause artifacts in the 3D reconstruction.

Defocus and astigmatism

Defocus and twofold astigmatism are key parameters governing the CTF in TEM. A precise and unbiased estimation of these aberrations is crucial in modeling image formation, interpretation of high resolution images, (tilted) CTF correction, assessment of microscope information transfer, optimal adjustment of aberration correctors, determination of higher-order aberrations and exit wave reconstruction. The accuracy of most defocus estimators developed so far (e.g. [40–46]) can be limited by the fitting of the background in the power spectrum density (PSD) of an image. The influence of the spherical aberration (C_s) on the shape of the Thon rings has been ignored. Furthermore, a robust estimation of small astigmatism has been lacking and the uncertainty of the estimation has been hard to assess.

An algorithm to estimate defocus and astigmatism and the associated uncertainties is presented in Chapter 3. The algorithm estimates these aberrations from the Thon rings in the PSD of an amorphous sample image. The method suppresses the background using an adaptive filtering strategy and uses template matching to estimate the shape of the rings. The frequencies of the detected rings are used to estimate the defocus and its uncertainty, while the ellipticity provides the information for estimation of astigmatism. In our applications astigmatism is usually smaller than defocus, i.e. Thon rings are approximately elliptical. To assess the performance of the algorithm, relatively large defocus values were chosen, typical for cryo-EM in life sciences.

- By utilizing information from various rings within a PSD, an outlier rejection routine is applied contributing to the high robustness of the algorithm.
- Uncertainty of the estimation is derived from a single image if more than one Thon ring is present in the PSD.
- A two-pass approach refines the astigmatism and defocus estimate by taking into account the influence of the known spherical aberration on the shape and frequencies of the rings.
- The reproducibility of the algorithm is validated on experimental data by repeating measurements of an amorphous sample under identical imaging conditions. The standard deviation of measured defocus and astigmatism within a series is small ($< 1\%$) and comparable to the mean value of the predicted standard deviations calculated from individual estimations. The linearity of the stigmator response was assessed *via* coefficient of determination (R^2) which was better than 0.9997.

- Tests show very good agreement between simulated and estimated defocus and astigmatism. Given a particular magnification and camera size, defocus can be estimated with errors less than 4 % for LaB₆ and 1 % for X-FEG gun microscopes and with a small spread.
- Robust template matching in a polar representation of the PSD permits detection of very small astigmatism. Some examples include astigmatism values that range from 10 nm (LaB₆) down to 0.2 nm (X-FEG) with ~ 10 % spread (for defoci of 1 and 2 μm).
- By using a new Thon ring averaging method, the modulation depth of the rings in a 1D averaged PSD can be enhanced compared to elliptical averaging. Given particular values for defocus, C_s and astigmatism it has been shown that contrast transfer assessment can be improved from 3.26 nm⁻¹ (elliptical) to 4.65 nm⁻¹ (Thon ring averaging).

Radiation damage

Radiation damage will always be a resolution limiting factor in cryo-EM. In Chapter 5, we investigate, inspired by numerous radiation damage studies done in X-ray crystallography, the influence of parameters such as dose, dose-rate and beam-heating. Knowledge about the amount of energy deposited allowed us to get an upper estimate of the radical concentrations that are built up in the vitreous sample, and to perform heat-transfer simulations. Stroboscopic exposure series were collected for different incident fluxes and integration times from earth worm hemoglobin as a test sample.

- Dose (in units of Gray) is estimated from the incident flux, beam size, average sample composition and thickness, and beam energy. This measure can be used as an estimate of the power deposited in the sample.
- Prior to gas-bubble formation, the images become linearly brighter (more electrons are detected) with dose, at a rate of approximately 0.1 % per 10 MGy.
- The complete disintegration of a vitrified hemoglobin sample could be visualized up to a dose of 5500 MGy.
- We advocate the use of stroboscopic data collection, with which variable amounts of dose can be used for the different steps of acquisition and reconstruction in SPA providing a minimal and controlled amount of damage.
- The estimated dose used in a typical SPA cryo-EM image is similar to the experimental dose limit for an entire data set of hundreds of images in X-ray crystallography (30 MGy).
- Comparison of equal-dose series collected with different incident fluxes showed a dose-rate effect favoring lower fluxes. Data that were collected with an incident electron flux of 50 e⁻A⁻²s⁻¹ were inferior to those that were collected at 5 e⁻A⁻²s⁻¹.

- Beam-induced heating simulations (for vitreous ice of $50\ \mu\text{m}$ diameter and $0.15\ \mu\text{m}$ thickness, beam diameter of $10\ \mu\text{m}$, fluxes of 5 and $50\ \text{e}^{-}\text{\AA}^{-2}\text{s}^{-1}$ at 120 kV) indicate that 1) the effect of beam heating is felt within milliseconds after the exposure, and 2) the beam heating is not expected to be a problem for cryo-EM samples with good thermal contact at low ($1\ \text{e}^{-}\text{\AA}^{-2}\text{s}^{-1}$) to medium ($5\ \text{e}^{-}\text{\AA}^{-2}\text{s}^{-1}$) electron fluxes.
- The electron beam deposits enough energy to form, in the absence of radical recombination, molar concentrations of radicals. We postulate that radical recombination must play a role in the observed dose and dose-rate effects.

Software

The software packages for camera characterization (Chapter 2), defocus and astigmatism estimation (Chapter 3), and simulations of image formation - InSilicoTEM (Chapter 4) are implemented in DIPimage, a MATLAB toolbox for scientific image processing and analysis, and are freely available for non-commercial use (<http://www.diplib.org/add-ons>). As being a part of a FOM industrial partnership program, the research in this thesis was performed in collaboration with FEI company. This resulted, among others, in a prototype software for defocus and astigmatism estimation implemented by Dr E. Franken and used within FEI Company. The InSilicoTEM source code was transferred directly to FEI Company where it is currently used for modeling.

6.1 Recommendations

There are numerous aspects that could further be investigated in order to ameliorate and extend the image formation model as well as estimations of the CTF, camera and radiation damage parameters. Some of the recommendations are listed below. The modularity of the simulator developed in this thesis (InSilicoTEM) allows integration of new physical phenomena without modifying the entire model of the image formation.

- The effects of the beam-induced movements must be reduced in order to increase the contrast in the images. Their influence can be somewhat diminished experimentally by lowering the flux, using a smaller carbon hole size, or by pre-irradiation. Computationally, one could reduce blurring in the final images by retrospective alignment [26, 27, 183, 190, 191] and averaging the raw frames captured by a direct electron detector [26, 183, 190, 191]. It is expected that dose fractionation using electron counting devices can significantly reduce the influence of beam-induced movements.
- Each simulation parameter relates directly to a physical quantity, but a strict quantitative comparison to cryo-EM images is difficult due to the high level of noise and challenging alignment. To facilitate an unbiased comparison, the display of each image was stretched in the same way as the corresponding experimental image. For visual comparisons we

simulated ten different noise realizations, confirming that the noise did not change the appearance of the features. However, statistical image-based measures could be introduced to allow a quantitative comparison of such data.

- The current model recognizes the difference between inelastic scattering properties of an average protein composition and vitreous ice. It could be improved by treating the scattering events on the atomic scale. Furthermore, one of the most challenging aspects remains how to model inelastics that actually fall on the detector and contribute to the unfiltered image.
- MD simulations were performed on a system consisting of both protein and solvent. These studies could be followed up to provide a more realistic modeling of the hydration shell of the protein. We expect that these should lead to an improved description of the contrast between the protein and the solvent. These effects could be important if beam-induced motions can be reduced. Furthermore, the result of the solvent influence predicted by atomistic MD simulations could be compared with continuous approaches such as Poisson-Boltzmann.
- As TEM image formation usually involves small angle scattering events, only axial aberrations such as defocus, astigmatism and spherical aberration have been considered in this analysis. As shown in [302, 303], beam-tilt and coma must be taken into consideration for resolutions around 4 Å (at 300 keV) since they introduce large phase errors. For lower acceleration voltages, the errors become even more significant. Extending the forward model to include the influence of higher order and off-axis aberrations should facilitate better understanding of the effects of non-parallel illumination and/or C_s corrector on the final image.
- If beam-induced motions can be minimized, validation of the model and comparisons with experimental data taken at a higher magnification (resolution) than presented in Chapter 4 could give more insights in the influence of the solvent ($\mathcal{V}_{\text{bond}}$) and amorphousness.
- The thickness of the vitreous ice determines the degree of dampening of the useful signal. In Chapter 4, the thickness is measured exploiting Lambert-Beer law and zero-loss energy filtering. The accuracy of the measurement depends on the stability of the energy slit of the filter and on the values of inelastic mean free paths. The latter varies noticeably in the literature [164]. Additional coarse tomograms could be acquired in order to confirm the measured thickness.
- In this analysis we studied oligomeric macromolecules. The contribution of the potential from charge redistribution $\mathcal{V}_{\text{bond}}$ might be different for non-oligomeric macromolecules.
- The arrangement of ions in the solvent is described by the Boltzmann distribution at room temperature and was considered not to change by rapid-freezing. The density of ice is 7 % less than water. Simulations could be performed to investigate the motion of ions as well as the influence of the strain introduced to biomolecules by ultra-rapid freezing.

- The camera characterization described in Chapter 2 has become even more important in the last years because of the emergence of highly expensive direct electron detectors by several competing manufacturers. The tools for characterizations were originally developed for CCD-based cameras and should be adapted to the new generation of CMOS direct electron detector where one image represents a sum of individual frames. The possibility for processing image stacks when such a camera works in electron counting mode should be also considered.
- For certain molecules, PDB files at different stages of radiation damage caused by X-ray crystallography are available. Forward modeling using such PDB files could be used to investigate radiation damage stages in cryo-EM.
- As an estimate of the power deposited in the sample, *dose* in units of grays could be used for comparison of experiments done at different acceleration voltages or different sample composition. This will facilitate data acquisition strategy and the choice of minimal possible radiation damage.
- Although experimental results on phase plates are not presented in this thesis, their basic model is incorporated in InSilicoTEM and can be further extended. In addition to phase plates, hybrid double-sideband/single-sideband objective apertures [304] can be included in InSilicoTEM in order to facilitate their design and study their impact on the reconstructed images.
- Using the reciprocity principle for TEM and bright-field scanning TEM (BF-STEM) (described in [144]), basic simulations of STEM images should be possible by appropriate modification of InSilicoTEM. Similarity between two imaging modes and some necessary modifications can be found in [144] (e.g. the incoherency of the source in TEM is equivalent to the blurring of the detector in BF-STEM). A practical limitation of applying the reciprocity principle also to annular dark field (ADF) STEM might lay in large collecting angles of the ADF detector.

6.2 Outlook

Cryo-EM for life sciences is going through an exciting epoch. Recent advances in instrumentation and experimental techniques include efforts to improve sample preparation (e.g. micromachining [305] or GraFix [285]) and transfer of information from the specimen to the image by: i) minimizing noise using direct electron detectors and counters, ii) improving CTF band-pass *via* phase plates, better lenses and more coherent electron sources, and iii) minimizing beam-induced movements of the specimen. In addition to the experimental developments, computational methods are continuously improving enabling more information to be extracted from inherently noisy cryo-EM images.

Automation of acquisition and reconstruction procedures is crucial for performing experiments that were previously thought to be impractical or even impossible. Direct access to digital

data has enabled developments such as autotuning of the microscope [57], automated electron tomography [58], and automated cryo-electron single particle micrograph collection [60].

Setting acquisition parameters such as acceleration voltage, defocus and electron flux is done by the microscopist and cannot be addressed in a completely automatic way. The relation between the data quality and acquisition parameters is complex. A skilled microscopist with substantial experience in data acquisition of a specific biological system at a specific instrument would generally be able to predict near-optimal acquisition conditions. A wrong choice of data acquisition parameters, however, may result in an unsuccessful experiment. Different applications often require different parameter settings.

Improving the data acquisition strategy becomes important especially when one strives for the highest attainable resolution. Note that a typical experiment may require many days of data collection and subsequent processing time using expensive equipment.

Simulations of image formation (forward modeling) provide the possibility to easily and cost-effectively investigate the influence of new data acquisition techniques and advanced instrumentation, and facilitates the development and evaluation of 3D reconstruction and image processing techniques. An accurate forward model leads to better understanding of the influence of a certain physical process on the final image and provides insight on ways to improve instrumentation and its use. The research described in this thesis and the resulting simulator (InSilicoTEM) contribute to that goal. An accurate forward model is an essential tool for optimizing the acquisition strategy, assisting the regularization of the 3D reconstruction, improving image interpretation, and achieving resolution beyond the first zero-crossing of the CTF.

The optimization of the data acquisition strategy involves an integrated approach that tackles the entire workflow from specimen preparation, through the actual data collection to forward modeling and reconstruction.

The theory and methods provided in this thesis represent a basis for modeling the outcome of data acquisition for any combination of the requested parameters. Image simulations with various combinations of acquisition parameters can be used for automatic optimization of the experiments.

A criterion for the optimal set of parameters could be related to the SNR at a certain resolution using the metrics such as Fourier ring correlation (FRC) [306] in 2D, or Fourier Shell Correlation (FSC) [307] in 3D case.

The numerical values of all parameters need to be known in order to accurately model image formation and facilitate 3D reconstructions. When necessary, the parameters such as defocus, astigmatism, vitreous ice thickness, electron flux, and camera characteristics must be estimated from the experiment using independent measurements (test images).

For quantitative modeling and analysis it is crucial to use adequate physical measures. As electrons interact with a specimen through Coulomb forces, EM effectively images a potential map. However, in cryo-EM, erroneously electron densities (obtained by X-ray techniques) were commonly used. Furthermore, as cryo-EM is approaching near-atomic resolution the differences in structural information between potential map and electron density maps should become apparent. Parameters based on physical principles will assist the electron microscopist to make objective and reproducible decisions for their data acquisition and bring discussions and individual expertise on a quantitative level.

The determination of the specimen thickness from the test images will predict the dampening of the signal due to the inelastic scattering events in the ice. Furthermore, the information about the average specimen composition will allow estimation of the dose deposited in the specimen at a certain voltage, indicative for the possible amount of radiation damage.

In addition to the opportunity to correct for beam-induced movements, the use of stroboscopic data collection provides possibilities of having variable amounts of dose (and possible radiation damage) used for the different steps of acquisition and reconstruction in SPA. For example, the first images from series could be used in the final reconstruction step (if corrected for the beam-induced movements [183]), while the last ones, with a higher dose, could be used for better localization of the particles [27].

With the simulator, one can optimize tomographic data collection, where the effect of parameters such as defocus, angular range and increment, single *versus* double tilt, integrated flux and flux-rate, magnification and detector binning can be simulated after an initial characterization of test images. Furthermore, the results of Saxton's acquisition scheme [308] can be investigated where tilt intervals are decreasing with the tilt angle.

It would be possible to obtain a simulated 3D reconstruction subset before or concurrently with automatic recording of a (tilt) series, enabling the (re)adjustment of the certain parameters such as defocus and flux. This is particularly interesting in low-flux conditions where the objects cannot be distinguished in the projections.

Near-atomic resolution in cryo-EM requires averaging over millions of asymmetric units [309]. This is nearly two orders of magnitude larger number than recently reported [183, 191] and a couple orders of magnitude larger than what would be achievable [238]. The forward model could facilitate estimations on size and number of particles needed to produce a 3D reconstruction at a certain resolution.

In addition to finding the optimal parameters for the available equipment, the influence of the new hardware components such as a direct electron detector or a phase plate can be inexpensively and efficiently investigated. For example, the outcome of *in silico* experiments can explore the potential benefits of single electron counters and optionally their *superresolution* (beyond-Nyquist) working mode.

The simulator could furthermore aid the identification of molecular assemblies within the cell, selection of the best particles for SPA reconstruction, template matching in SPA and subtomogram averaging, a docking process where atomic models are fitted into cryo-EM maps, and testing whether a proposed 3D model of a macromolecule is in agreement with the features observed in the micrographs.

In electron tomography, iterative reconstruction schemes attempt to minimize the difference between projections and simulated reprojections of the 3D map. From the differences between observed and simulated images one can often derive information to refine the model. The model parameters are iterated until simulated images best describe those observed. For such methods, the image formation model used for simulations is a relevant factor that affects the quality of the reconstructions as currently the volume is just projected. We expect that a more accurate forward model will improve such an iterative reconstruction schemes resulting in better 3D potential maps.

Bibliography

- [1] A. Sali, R. M. Glaeser, T. Earnest, W. Baumeister, From words to literature in structural proteomics, *Nature* 422 (2003) 216–225.
- [2] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*, 2nd Edition, Oxford University Press, 2006.
- [3] Y. Wang, J. Chiu, Q. He, Proteomics in computer-aided drug design, *Current Computer-Aided Drug Design* 1 (2005) 43–52.
- [4] D. C. Hodgkin, for determinations by X-ray techniques of the structures of important biochemical substances, Nobel Prize (1964) –.
- [5] A. Klug, for development of crystallographic electron microscopy and his structural elucidation of biologically important nucleic acid-protein complexes, Nobel Prize (1982) –.
- [6] R. Kornberg, for studies of the molecular basis of eukaryotic transcription, Nobel Prize (2006) –.
- [7] V. Ramakrishnan, T. Steitz, A. E. Yonath, for studies of the structure and function of the ribosome, Nobel Prize (2009) –.
- [8] N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, H. Stark, Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy, *Nature* 466 (2010) 329–333.
- [9] S. Nickell, C. Kofler, A. P. Leis, W. Baumeister, A visual approach to proteomics, *Nature Reviews Molecular Cell Biology* 7 (3) (2006) 225–230.
- [10] L.-M. Peng, S. L. Dudarev, M. J. Whelan, *High-Energy Electron Diffraction and Microscopy*, Vol. 61 of *Monographs on the Physics and Chemistry of Materials*, Oxford University Press, 2004.
- [11] S. W. Hell, J. Wichmann, Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy, *Optics Letters* 19 (11) (1994) 780–782.
- [12] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, H. F. Hess, et al., Imaging intracellular fluorescent proteins at nanometer resolution., *Science* 313 (2006) 1643–1645.
- [13] M. J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), *Nature Methods* 3 (2006) 793–796.
- [14] M. Heilemann, S. van de Linde, M. Schttpelz, R. Kasper, M. Sauer, et al., Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes, *Angewandte Chemie International Edition* 47 (33) (2008) 6172–6176.
- [15] C. van Rijnsoever, V. Oorschot, J. Klumperman, Correlative light-electron microscopy (CLEM) combining live-cell imaging and immunolabeling of ultrathin cryosections, *Nature methods* 5 (2008) 973–980.
- [16] K. A. Taylor, R. M. Glaeser, Electron microscopy of frozen hydrated biological specimens, *Journal of Ultrastructure Research* 55 (3) (1976) 448–456.
- [17] J. Dubochet, J. Lepault, R. Freeman, J. A. Berriman, J.-C. Homo, Electron microscopy of frozen water and aqueous solutions, *Journal of Microscopy* 128 (3) (1982) 219–237.

- [18] W. Chiu, K. H. Downing, J. Dubochet, R. M. Glaeser, H. Heide, F. Zemlin, et al., Cryoprotection in electron microscopy, *Journal of Microscopy* 141 (3) (1986) 385–391.
- [19] T. Gonen, Y. Cheng, P. Sliz, Y. Hiroaki, Y. Fujiyoshi, S. C. Harrison, T. Walz, Lipid-protein interactions in double-layered two-dimensional AQP0 crystals, *Nature* 438 (2005) 633–638.
- [20] R. R. Meyer, A. 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) (1998) 23–33.
- [21] K. Murata, X. Liu, R. Danev, J. Jakana, M. F. Schmid, J. King, K. Nagayama, W. Chiu, Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions, *Structure* 18 (8) (2010) 903–912.
- [22] R. Danev, S. Kanamaru, M. Marko, K. Nagayama, Zernike phase contrast cryo-electron tomography, *Journal of Structural Biology* 171 (2) (2010) 174–181.
- [23] R. M. Glaeser, Retrospective: Radiation damage and its associated “information limitations”, *Journal of Structural Biology* 163 (2008) 271–276.
- [24] W. H. Massover, Biological specimen preparation and preservation for high resolution microscopies, *Micron* 42 (2) (2011) 141–151.
- [25] R. M. Glaeser, K. A. Taylor, Radiation damage relative to transmission electron microscopy of biological specimens at low temperature: A review, *Journal of Microscopy* 112 (1) (1978) 127–138.
- [26] A. F. Brilot, J. Z. Chen, A. Cheng, J. Pan, S. C. Harrison, C. Potter, B. Carragher, R. Henderson, N. Grigorieff, Beam-induced motion of vitrified specimen on holey carbon film, *Journal of Structural Biology* 177 (3) (2012) 630–637.
- [27] F. Karimi Nejadasl, M. Karuppasamy, E. R. Newman, J. E. McGeehan, R. B. G. Ravelli, Non-rigid image registration to reduce beam-induced blurring of cryo-electron microscopy images, *Journal of Synchrotron Radiation* 20 (1) (2013) 58–66.
- [28] R. D. Leapman, S. Sun, Cryo-electron energy loss spectroscopy: observations on vitrified hydrated specimens and radiation damage, *Ultramicroscopy* 59 (1) (1995) 71–79.
- [29] L. M. Voortman, S. Stallinga, R. H. M. Schoenmakers, L. J. van Vliet, B. Rieger, A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM, *Ultramicroscopy* 111 (8) (2011) 1029–1036.
- [30] L. M. Voortman, E. M. Franken, L. J. van Vliet, B. Rieger, Fast, spatially varying CTF correction in TEM, *Ultramicroscopy* 118 (2012) 26–34.
- [31] C. O. S. Sorzano, R. Marabini, J. Velázquez-Muriel, J. Bilbao-Castro, S. H. W. Scheres, J. M. Carazo, A. Pascual-Montano, XMIPP: a new generation of an open-source image processing package for electron microscopy, *Journal of Structural Biology* 148 (2) (2004) 194–204.
- [32] J. R. Bilbao-Castro, C. O. S. Sorzano, I. García, J. J. Fernández, Phan3D: Design of biological phantoms in 3D electron microscopy, *Bioinformatics* 20 (17) (2004) 3286–3288.
- [33] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, M. Schatz, A new generation of the IMAGIC image processing system, *Journal of Structural Biology* 116 (1) (1996) 17–24.
- [34] J. Frank, B. Shimkin, A new image processing software for structural analysis and contrast enhancement, in: J. M. Sturgess (Ed.), *Electron Microscopy 1978; Papers Presented at the Ninth International Congress on Electron Microscopy*, Toronto, Microscopical Society of Canada, Toronto, 1978, Vol. 1, 1978, pp. 210–211.
- [35] T. R. Shaikh, H. Gao, W. T. Baxter, F. J. Asturias, N. Boisset, A. Leith, F. Frank, SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs, *Nature Protocols* 3 (2008) 1941–1974.

- [36] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, EMAN2: An extensible image processing suite for electron microscopy, *Journal of Structural Biology* 157 (1) (2007) 38–46.
- [37] J. B. Heymann, D. M. Belnap, Bsoft: Image processing and molecular modeling for electron microscopy, *Journal of Structural Biology* 157 (1) (2007) 3–18.
- [38] S. Nickell, F. Förster, A. Linaroudis, W. Del Net, F. Beck, R. Hegerl, W. Baumeister, J. M. Plitzko, TOM software toolbox: acquisition and analysis for electron tomography, *Journal of Structural Biology* 149 (3) (2005) 227–234.
- [39] H. Rullgård, L.-G. Öfverstedt, S. Masich, B. Daneholt, O. Öktem, Simulation of transmission electron microscope images of biological specimens, *Journal of Microscopy* 242 (2) (2011) 234–256.
- [40] S. P. Mallick, B. Carragher, C. S. Potter, D. J. Kriegman, ACE: Automated CTF estimation, *Ultramicroscopy* 104 (1) (2005) 8–29.
- [41] J. Zhu, P. A. Penczek, R. Schröder, J. Frank, Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70S *Escherichia coli* ribosome, *Journal of Structural Biology* 118 (3) (1997) 197–219.
- [42] J. F. Conway, A. C. Steven, Methods for reconstructing density maps of “single” particles from cryoelectron micrographs to subnanometer resolution, *Journal of Structural Biology* 128 (1) (1999) 106–118.
- [43] S. J. Ludtke, P. R. Baldwin, W. Chiu, EMAN: semiautomated software for high-resolution single-particle reconstructions, *Journal of Structural Biology* 128 (1) (1999) 82–97.
- [44] Z. Huang, P. R. Baldwin, S. Mullapudi, P. A. Penczek, Automated determination of parameters describing power spectra of micrograph images in electron microscopy, *Journal of Structural Biology* 144 (1-2) (2003) 79–94.
- [45] B. Sander, M. M. Golas, H. Stark, Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra, *Journal of Structural Biology* 142 (3) (2003) 392–401.
- [46] J. A. Velázquez-Muriel, C. O. S. Sorzano, J. J. Fernández, J. M. Carazo, A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment, *Ultramicroscopy* 96 (1) (2003) 17–35.
- [47] M. Vulovic, B. Rieger, L. J. van Vliet, A. J. Koster, R. B. G. Ravelli, A toolkit for the characterization of CCD cameras for transmission electron microscopy, *Acta Crystallographica D* 66 (1) (2010) 97–109.
- [48] M. Vulovic, P. Brandt, R. B. G. Ravelli, A. J. Koster, L. J. van Vliet, B. Rieger, Estimation of defocus and astigmatism in transmission electron microscopy, In *Proc. of the IEEE International Symposium on Biomedical Imaging* (2010) 1121–1124.
- [49] M. Vulović, E. Franken, R. B. G. Ravelli, L. J. van Vliet, B. Rieger, Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy, *Ultramicroscopy* 116 (1) (2012) 115–134.
- [50] M. Karuppasamy, F. Karimi Nejadasl, M. Vulovic, A. J. Koster, R. B. G. Ravelli, Radiation damage in single particle cryo-electron microscopy: effects of dose and dose rate, *Journal of Synchrotron Radiation* 18 (3) (2011) 398–412.
- [51] M. Vulović, R. B. G. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, H. Rullgård, O. Öktem, B. Rieger, Image formation modeling in cryo-electron microscopy, *Journal of Structural Biology*, accepted (2013) –.
- [52] M. Vulović, L. M. Voortman, L. J. van Vliet, B. Rieger, When to use the projection assumption and weak-phase object approximation in phase contrast cryo-EM, submitted (2013) –.
- [53] C. Ponchut, Characterization of X-ray area detectors for synchrotron beamlines, *Journal of Synchrotron Radiation* 13 (2006) 195–203.
- [54] P. T. E. Roberts, J. N. Chapman, A. M. MacLeod, A CCD-based image recording system for the CTEM, *Ultramicroscopy* 8 (1982) 385–396.

- [55] J. C. H. Spence, J. M. Zuo, Large dynamic-range, parallel detection system for electron-diffraction and imaging, *Review of Scientific Instruments* 59 (1988) 2102–2105.
- [56] G. Y. Fan, M. H. Ellisman, Digital imaging in transmission electron microscopy, *Journal of Microscopy-Oxford* 200 (2000) 1–13.
- [57] O. L. Krivanek, P. Mooney, Applications of slow-scan CCD cameras in transmission electron microscopy, *Ultramicroscopy* 49 (1) (1993) 95–108.
- [58] A. Koster, H. Chen, J. W. Sedat, D. A. Agard, Automated microscopy for electron tomography, *Ultramicroscopy* 46 (1) (1992) 95–108.
- [59] J. Brink, W. Chiu, Applications of a slow-scan CCD camera in protein electron crystallography, *Journal of Structural Biology* 113 (1994) 23–34.
- [60] B. Carragher, N. Kisseberth, D. Kriegman, R. Milligan, C. S. Potter, J. Pulokas, A. Reilein, Legion: An automated system for acquisition of images from vitreous ice specimens, *Journal of Structural Biology* 132 (1) (2000) 33–45.
- [61] K. H. Downing, F. M. Hendrickson, Performance of a 2k CCD camera designed for electron crystallography at 400 kV, *Ultramicroscopy* 75 (1999) 215–233.
- [62] B. Sander, M. M. Golas, H. Stark, Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy, *Journal of Structural Biology* 151 (2005) 92–105.
- [63] G. McMullan, D. M. Cattermole, S. Chen, R. Henderson, X. Llopart, C. Summerfield, L. Tlustos, A. R. Faruqi, Electron imaging with medipix2 hybrid pixel detector, *Ultramicroscopy* 107 (2007) 401–413.
- [64] A. R. Faruqi, D. M. Cattermole, R. Henderson, B. Mikulec, C. Raeburn, Evaluation of a hybrid pixel detector for electron microscopy, *Ultramicroscopy* 94 (2003) 263–276.
- [65] A. R. Faruqi, R. Henderson, Electronic detectors for electron microscopy, *Current Opinion in Structural Biology* 17 (2007) 549–555.
- [66] R. S. Aikens, D. A. Agard, J. W. Sedat, Solid-state imagers for microscopy, *Methods Cell Biology* 29 (1989) 291–313.
- [67] D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements, *Journal of Structural Biology* 152 (2005) 36–51.
- [68] S. Q. Zheng, B. Keszthelyi, E. Branlund, J. M. Lyle, M. B. Braunfeld, J. W. Sedat, D. A. Agard, Automated acquisition of electron microscopic random conical tilt sets, *Journal of Structural Biology* 157 (2007) 138–147.
- [69] C. Suloway, J. Pulokas, D. Fellmann, A. Cheng, F. Guerra, J. Quispe, S. Stagg, C. S. Potter, B. Carragher, Automated molecular microscopy: the new legion system, *Journal of Structural Biology* 151 (2005) 41–60.
- [70] J. P. Moy, A. P. Hammersley, S. O. Svensson, A. Thompson, K. Brown, L. Claustre, A. Gonzalez, S. McSweeney, A novel technique for accurate intensity calibration of area X-ray detectors at almost arbitrary energy, *Journal of Synchrotron Radiation* 3 (1996) 1–5.
- [71] A. Zinger, Detection of best and outlying normal populations with known variances, *Biometrika* 48 (1988) 457–461.
- [72] D. Bourgeois, J. P. Moy, S. O. Svensson, A. Kvik, The point-spread function of X-ray image-intensifiers/CCD-camera and imaging-plate systems in crystallography: assessment and consequences for the dynamic range, *Journal of Applied Crystallography* 27 (1994) 868–877.
- [73] W. J. De Ruijter, Imaging properties and applications of slow-scan charge-coupled device cameras suitable for electron microscopy, *Micron* 26 (1995) 247–275.

- [74] J. C. Dainty, R. Shaw, *Image science: principles, analysis and evaluation of photographic-type imaging processes*, 1st Edition, Academic Press, 1974.
- [75] M. Rabbani, R. Shaw, R. Van Metter, Detective quantum efficiency of imaging systems with amplifying and scattering mechanisms, *Journal of the Optical Society of America A* 4 (1987) 895–901.
- [76] R. R. Meyer, A. I. Kirkland, Characterisation of the signal and noise transfer of CCD cameras for electron detection, *Microscopy Research and Technique* 49 (2000) 269–280.
- [77] L. J. van Vliet, D. Sudar, I. T. Young, Digital fluorescence imaging using cooled charge-coupled device array cameras, in: G. E. Celis (Ed.), *Cell biology*, Academic Press, 1998, pp. 109–120.
- [78] E. Samei, M. J. Flynn, D. A. Reimann, A method for measuring the presampled MTF of digital radiographic systems using an edge test device, *Medical Physics* 25 (1998) 102–113.
- [79] J. C. Mullikin, L. J. van Vliet, H. Netten, F. R. Boddeke, G. W. van der Feltz, I. T. Young, Methods for CCD camera characterization, *Image Acquisition and Scientific Imaging Systems*, Proc. SPIE 2173 (1996) 73–84.
- [80] P. W. Verbeek, L. J. van Vliet, On the location error of curved edges in low-pass filtered 2-D and 3-D images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (7) (1994) 726–733.
- [81] N. M. Allinson, Development of non-intensified charge-coupled device area X-ray detectors, *Journal of Synchrotron Radiation* 1 (1994) 54–62.
- [82] A. J. Koster, R. Grimm, D. Typke, R. Hegerl, A. Stoschek, J. Walz, W. Baumeister, Perspectives of molecular and cellular electron tomography, *Journal of Structural Biology* 120 (3) (1997) 276–308.
- [83] D. N. Mastrorade, Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package, *Journal of Structural Biology* 230 (2008) 212–217.
- [84] R. R. Meyer, A. I. Kirkland, R. E. Dunin-Borkowski, J. L. Hutchison, Experimental characterisation of CCD cameras for HREM at 300 kV, *Ultramicroscopy* 85 (2000) 9–13.
- [85] Y. M. Kim, J. Y. Lee, D. Moonen, K. I. Jang, Y. J. Kim, Quantitative evaluations of a high-voltage multiscan CCD camera, *Journal of Electron Microscopy (Tokyo)* 56 (2007) 217–224.
- [86] A. G. W. Leslie, Automatic fold and structural motif elucidation from 3D EM maps of macromolecules, in: *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* 26, 1992, pp. 264–275.
- [87] R. B. G. Ravelli, R. M. Sweet, J. M. Skinner, A. J. M. Duisenberg, J. Kroon, STRATEGY: a program to optimize the starting spindle angle and scan range for X-ray data collection, *Journal of Applied Crystallography* 30 (1997) 551–554.
- [88] A. N. Popov, G. P. Bourenkov, Choice of data-collection parameters based on statistic modelling, *Acta Crystallographica D* 59 (2003) 1145–1153.
- [89] G. P. Bourenkov, A. N. Popov, A quantitative approach to data-collection strategies, *Acta Crystallographica D* 62 (2006) 58–64.
- [90] R. B. G. Ravelli, E. F. Garman, Radiation damage in macromolecular cryocrystallography, *Current Opinion in Structural Biology* 16 (2006) 624–629.
- [91] A. G. W. Leslie, H. R. Powell, G. Winter, O. Svensson, C. Nave, et al., Automation of the collection and processing of x-ray diffraction data a generic approach, *Acta Crystallographica D* 58 (2002) 1924–1928.
- [92] H. P. Erickson, A. Klug, Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs, *Phil. Trans. R. Soc. Lond. Ser. B* 261 (105) (1971) 105–118.
- [93] F. Thon, Phase contrast electron microscopy, *Electron Microscopy in Materials Science* (1971) 570–625.
- [94] A. J. Koster, A. van den Bos, K. van der Mast, An autofocus method for a TEM, *Ultramicroscopy* 21 (3) (1987) 209 – 222.

- [95] Z. H. Zhou, S. Hardt, B. Wang, M. B. Sherman, J. Jakana, W. Chiu, CTF determination of images of ice-embedded single particles using a graphics interface, *Journal of Structural Biology* 116 (1) (1996) 216–222.
- [96] J. J. Fernández, J. R. Sanjurjo, J. M. Carazo, A spectral estimation approach to contrast transfer function detection in electron microscopy, *Ultramicroscopy* 68 (4) (1997) 267–295.
- [97] A. Saad, S. J. Ludtke, J. Jakana, F. J. Rixon, H. Tsuruta, W. Chiu, Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination, *Journal of Structural Biology* 133 (1) (2001) 32–42.
- [98] J. A. Mindell, N. Grigorieff, Accurate determination of local defocus and specimen tilt in electron microscopy, *Journal of Structural Biology* 142 (3) (2003) 334–347.
- [99] C. O. S. Sorzano, S. Jonic, R. Nez-Ramrez, N. Boisset, J. M. Carazo, Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function, *Journal of Structural Biology* 160 (2) (2007) 249–262.
- [100] C. Yang, W. Jiang, D. H. Chen, U. Adiga, E. G. Ng, W. Chiu, Estimating contrast transfer function and associated parameters by constrained non-linear optimization, *Journal of Microscopy* 233 (3) (2009) 391–403.
- [101] K. Tani, H. Sasabe, C. Toyoshima, A set of computer programs for determining defocus and astigmatism in electron images, *Ultramicroscopy* 65 (1-2) (1996) 31–44.
- [102] K. V. Fernando, Radial averages of astigmatic TEM images, *Journal of Structural Biology* 164 (1) (2008) 49–59.
- [103] J. Barthel, A. Thust, Aberration measurement in HRTEM: Implementation and diagnostic use of numerical procedures for the highly precise recognition of diffractogram patterns, *Ultramicroscopy* 111 (1) (2010) 27–46.
- [104] C. O. S. Sorzano, A. Otero, E. M. Olmos, J. M. Carazo, Error analysis in the determination of the electron microscopical contrast transfer function parameters from experimental power spectra, *BMC Structural Biology* 9 (2009) 18–34.
- [105] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, K.-H. Herrmann, Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms, *Ultramicroscopy* 3 (1978) 49–60.
- [106] A. Thust, C. Jia, K. Urban, Direct determination of imaging parameters from wave functions in HRTEM, *Microscopy and Microanalysis* 9 (2003) 140–141.
- [107] R. Meyer, A. Kirkland, W. Saxton, A new method for the determination of the wave aberration function for high resolution TEM 1. Measurement of the symmetric aberrations, *Ultramicroscopy* 92 (2002) 89–109.
- [108] A. Steinecker, W. Mader, Measurement of lens aberrations by means of image displacements in beam-tilt series, *Ultramicroscopy* 81 (2000) 149–161.
- [109] S. Uhlemann, M. Haider, Residual wave aberrations in the first spherical aberration corrected transmission electron microscope, *Ultramicroscopy* 72 (1998) 109–119.
- [110] A. Kirkland, W. Saxton, R. Meyer, Super resolved microscopy and aberration determination in the TEM, in: *Electron Microscopy and Analysis Group Conference EMAG97*, Cambridge, UK, 1997, pp. 105–108.
- [111] L. Reimer, H. Kohl, *Transmission Electron Microscopy*, 5th Edition, Vol. 36 of Springer series in optical sciences, Springer Verlag, 2008.
- [112] J. Barthel, A. Thust, Quantification of the information limit of transmission electron microscopes, *Physical Review Letters* 101 (20) (2008) 2008011–2008014.
- [113] K. Yonekura, M. Braunfeld, S. Mari-Yonekura, D. Agard, Electron energy filtering significantly improves amplitude contrast of frozen-hydrated protein at 300 kV, *Journal of Structural Biology* 156 (3) (2006) 524–536.

- [114] S. Jonic, C. O. S. Sorzano, M. Cottevaille, E. Larquet, N. Boisset, A novel method for improvement of visualization of power spectra for sorting cryo-electron micrographs and their local areas, *Journal of Structural Biology* 157 (1) (2007) 156–167.
- [115] P. Bakker, L. J. van Vliet, P. W. Verbeek, Edge preserving orientation adaptive filtering, in: *Conference on Computer Vision and Pattern Recognition*, Vol. 2, IEEE Computer Society Press, Los Alamitos, CA, USA, 1999, pp. 535–540.
- [116] J. van de Weijer, L. J. van Vliet, P. W. Verbeek, R. van Ginkel, Curvature estimation in oriented patterns using curvilinear models applied to gradient vector fields, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 23 (9) (2001) 1035–1042.
- [117] T. P. Harrigan, R. W. Mann, Characterization of microstructural anisotropy in orthotropic materials using a second rank tensor, *Journal of Materials Science* 19 (1984) 761–767.
- [118] M. Kass, A. Witkin, Analyzing oriented patterns, *Computer Vision, Graphics and Image Processing* 37 (1987) 362–385.
- [119] W. Förstner, E. Gülch, A fast operator for detection and precise location of distinct point, corners and centres of circular features, in: *Proceedings of the ISPRS Conference on Fast Processing of Photogrammetric Data*, 1987, pp. 281–305.
- [120] C. L. Luengo Hendriks, M. van Ginkel, P. W. Verbeek, L. J. van Vliet, The generalized Radon transform: Sampling, accuracy and memory considerations, *Pattern Recognition* 38 (12) (2005) 2494–2505.
- [121] G. C. Stockman, A. K. Agrawala, Equivalence of Hough curve detection to template matching, *Communications of the ACM* 20 (11) (1977) 820–822.
- [122] R. F. Egerton, K. Wong, Some practical consequences of the Lorentzian angular distribution of inelastic scattering, *Ultramicroscopy* 59 (1-4) (1995) 169 – 180.
- [123] W. Saxton, W. Baumeister, The correlation averaging of a regularly arranged bacterial cell envelope protein, *Journal of microscopy* 127 (2) (1982) 127 – 138.
- [124] B. A. Walther, J. L. Moore, The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance, *Ecography* 28 (6) (2005) 815–829.
- [125] J. J. Fernández, S. Li, R. A. Crowther, CTF determination and correction in electron cryotomography, *Ultramicroscopy* 106 (7) (2006) 587–596.
- [126] Q. Xiong, M. K. Morphew, C. L. Schwartz, A. H. Hoenger, D. N. Mastronarde, CTF determination and correction for low dose tomographic tilt series, *Journal of Structural Biology* 168 (3) (2009) 378 – 387.
- [127] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: *Sixth International Conference on Computer Vision*, 1998, pp. 839 –846.
- [128] T. Q. Pham, L. J. van Vliet, Separable bilateral filtering for fast video preprocessing, in: *ICME 2005. IEEE International Conference on Multimedia and Expo*, 2005, p. 4.
- [129] J. Frank, A. Verschoor, M. Boublik, Computer averaging of electron micrographs of 40S ribosomal subunits, *Science* 214 (4527) (1981) 1353–1355.
- [130] M. van Heel, Similarity measures between images, *Ultramicroscopy* 21 (1) (1987) 95 – 100.
- [131] F. Karimi Nejadasl, M. Karuppasamy, A. J. Koster, R. B. G. Ravelli, Defocus estimation from stroboscopic cryo-electron microscopy data, *Ultramicroscopy* 111 (11) (2011) 1592–1598.
- [132] D. M. Blow, F. H. C. Crick, The treatment of errors in the isomorphous replacement method, *Acta Crystallographica* 12 (10) (1959) 794–802.

- [133] E. Franken, R. Wagner, R. Schoenmakers, M. Vulović, Applications of accurate defocus and astigmatism estimation, in: EMBL Conference: Sixth International Congress on Electron Tomography, Heidelberg, Germany, 2011, p. 80.
- [134] V. Lucić, F. Förster, W. Baumeister, Structural studies by electron tomography: from cells to molecules, *Annual Review of Biochemistry* 74 (2005) 833–865.
- [135] R. McIntosh, D. Nicastro, D. Mastronarde, New views of cells in 3D: an introduction to electron tomography, *Trends Cell Biology* 15 (2005) 43–51.
- [136] A. Leis, B. Rockel, L. Andrees, W. Baumeister, Visualizing cells at the nanoscale, *Trends in biochemical sciences* 34 (2) (2009) 60–70.
- [137] R. Henderson, Realizing the potential of electron cryo-microscopy, *Quarterly Review of Biophysics* 37 (2004) 3–13.
- [138] R. M. Glaeser, K. H. Downing, D. DeRosier, W. Chu, J. Frank, *Electron Crystallography of Biological Macromolecules*, Oxford University Press, 2006.
- [139] R. Marabini, G. T. Herman, J.-M. Carazo, 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs), *Ultramicroscopy* 72 (1998) 53–65.
- [140] C. O. S. Sorzano, R. Marabini, N. Boisset, E. Rietzel, R. Schröder, G. T. Herman, J.-M. Carazo, The effect of overabundant projection directions on 3D reconstruction algorithms, *Journal of Structural Biology* 133 (2–3) (2001) 108–118.
- [141] R. Marabini, E. Rietzel, R. Schröder, G. T. Herman, J.-M. Carazo, Three-dimensional reconstruction from reduced sets of very noisy images acquired following a single-axis tilt scheme: Application of a new three-dimensional reconstruction algorithm and objective comparison with weighted backprojection, *Journal of Structural Biology* 120 (1997) 363–371.
- [142] R. J. Hall, E. Nogales, R. M. Glaeser, Accurate modeling of single-particle cryo-EM images quantitates the benefits expected from using zernike phase contrast, *Journal of Structural Biology* 174 (3) (2011) 468–475.
- [143] Z. Shang, F. J. Sigworth, Hydration-layer models for cryo-em image simulation, *Journal of Structural Biology* 180 (1) (2012) 10–16.
- [144] E. J. Kirkland, *Advanced Computing in Electron Microscopy*, 2nd Edition, Springer Verlag, 2010.
- [145] C. Dinges, H. Rose, Simulation of filtered and unfiltered TEM images and diffraction patterns, *Physica Status Solidi A. Applications and Material Science* 150 (1) (1995) 23–29.
- [146] H. Müller, H. Rose, P. Schorsch, A coherence function approach to image simulation, *Journal of Microscopy* 190 (1–2) (1998) 73–88.
- [147] A. Gómez-Rodríguez, L. M. Beltrán-del Río, R. Herrera-Becerra, SimulaTEM: Multislice simulations for general objects, *Ultramicroscopy* 110 (2) (2009) 95–104.
- [148] J. M. Cowley, A. F. Moodie, The scattering of electrons by atoms and crystals. i. a new theoretical approach, *Acta Crystallographica* 10 (1957) 609–613.
- [149] M. J. Hytch, W. M. Stobbs, Quantitative comparison of high resolution TEM images with image simulations, *Ultramicroscopy* 53 (1994) 191–203.
- [150] C. B. Boothroyd, R. E. Dunin-Borkowski, W. M. Stobbs, C. J. Humphreys, Quantifying the effects of amorphous layers on image contrast using energy filtered transmission electron microscopy, in: *MRS Proceedings*, Vol. 354, 1995, pp. 495–500.
- [151] A. Thust, High-resolution transmission electron microscopy on an absolute contrast scale, *Physical Review Letters* 102 (2009) 2208011–2208014.

- [152] L.-M. Peng, G. Ren, S. L. Dudarev, M. J. Whelan, Robust parameterization of elastic and absorptive electron atomic scattering factors, *Acta Crystallographica A* 52 (2) (1996) 257–276.
- [153] R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd Edition, Springer, 1996.
- [154] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, D. Zschiesche, et al., Geant4 - a simulation toolkit, *Nuclear Instruments and Methods in Physics Research* 506 (3) (2003) 250–303.
- [155] E. Kieft, E. Bosch, Refinement of monte carlo simulations of electronspecimen interaction in low-voltage SEM, *Journal of Physics D: Applied Physics* 41 (21) (2008) 215310.
- [156] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, *Proceedings of the National Academy of Sciences of the United States of America* 98 (2001) 10037–10041.
- [157] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, PDB2PQR: an automated pipeline for the setup of poissonboltzmann electrostatics calculations, *Nucleic Acids Research* 32 (2004) 665 – 667.
- [158] J. P. Langmore, M. F. Smith, Quantitative energy-filtered electron microscopy of biological molecules in ice, *Ultramicroscopy* 46 (1992) 349–373.
- [159] C. Colliex, C. Jeanguillaume, C. Mory, Unconventional modes for STEM imaging of biological structures, *J Ultrastructural Research* 88 (1984) 177 – 206.
- [160] R. Egerton, A data base for energy-loss cross sections and mean free paths, in: *Proc. 50th Ann. Meet. Elec. Microsc.*, Vol. 136, 1992, pp. 1248 – 1249.
- [161] S. Sun, S. Shi, R. Leapman, Water distributions of hydrated biological specimens by valence electron energy loss spectroscopy, *Ultramicroscopy* 50 (2) (1993) 127 – 139.
- [162] R. Grimm, D. Typke, M. Barmann, W. Baumeister, Determination of the inelastic mean free path in ice by examination of tilted vesicles and automated most probable loss imaging, *Ultramicroscopy* 63 (3) (1996) 169 – 179.
- [163] J. Wall, M. Isaacson, J. P. Langmore, Collection of scattered electrons in dark field electron microscopy - 2. inelastic scattering, *Optik* 39 (1974) 359–374.
- [164] B. Feja, U. Aebi, Determination of inelastic mean free path of electrons in vitrified ice layers for on-line thickness measurements by zero-loss imaging, *Journal of Microscopy* 193 (1) (1999) 15 – 19.
- [165] R. Reichelt, A. Engel, Monte carlo calculations of elastic and inelastic electron scattering in biological and plastic materials, *Ultramicroscopy* 13 (3) (1984) 279 – 293.
- [166] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *Journal of Chemical Theory and Computation* 4 (3) (2008) 435 – 447.
- [167] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, R. Taylor, Tables of bond lengths determined by X-ray and neutron diffraction. part 1. bond lengths in organic compounds, *Journal of the Chemical Society, Perkin Trans. 2* 12 (1) (1987) 1–19.
- [168] M. M. Teeter, Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin, *Proc. Natl. Acad. Sci* 81 (19) (1984) 6014–6018.
- [169] R. Henderson, R. M. Glaeser, Quantitative analysis of image contrast in electron micrographs of beam-sensitive crystals, *Ultramicroscopy* 16 (2) (1985) 139–150.
- [170] S. N. Vinogradov, P. K. Sharma, Preparation and characterization of invertebrate globin complexes, in: J. Everse, K. D. Vandegriff, W. R. M. (Eds.), *Hemoglobins Part B: Biochemical and Analytical Methods*, Academic Press, 1994, Ch. 8, pp. 112–124.

- [171] G. Vriend, WHAT IF: A molecular modeling and drug design program, *Journal of Molecular Graphics* 8 (1) (1990) 52–56.
- [172] L.-M. Peng, S. L. Dudarev, M. J. Whelan, Electron scattering factors of ions and dynamical RHEED from surfaces of ionic crystals, *Physical Review B* 57 (12) (1998) 7259–7265.
- [173] A. Philippsen, H. Engel, A. Engel, The contrast-imaging function for tilted specimens, *Ultramicroscopy* 107 (2-3) (2007) 202 – 212.
- [174] H. Winkler, K. A. Taylor, Focus gradient correction applied to tilt series image data used in electron tomography, *Journal of Structural Biology* 143 (1) (2003) 24 – 32.
- [175] D. J. DeRosier, Correction of high-resolution data for curvature of the Ewald sphere, *Ultramicroscopy* 81 (2) (2000) 83–98.
- [176] Y. Wan, W. Chiu, Z. H. Zhou, Full contrast transfer function correction in 3D cryo-EM reconstruction, in: *IEEE Proceedings of ICCAS 2004*, Vol. 2, 2004, pp. 960–964.
- [177] T. Alcorn, D. H. Juers, Progress in rational methods of cryoprotection in macromolecular crystallography, *Acta Crystallographica D* 66 (4) (2010) 366–373.
- [178] C. Bajaj, S. Chen, A. Rand, An efficient higher-order fast multipole boundary element solution for poissonboltzmann-based molecular electrostatics, *SIAM Journal on Scientific Computing* 33 (2) (2011) 826–848.
- [179] P. Bullough, R. Henderson, Use of spot-scan procedure for recording low-dose micrographs of beam-sensitive specimens, *Ultramicroscopy* 21 (3) (1987) 223–230.
- [180] R. Henderson, Image contrast in high-resolution electron microscopy of biological macromolecules: TMV in ice, *Ultramicroscopy* 46 (1992) 1–18.
- [181] R. M. Glaeser, K. H. Downing, Specimen charging on thin films with one conducting layer: Discussion of physical principles, *Microscopy and Microanalysis* 10 (6) (2004) 790–796.
- [182] G. J. Jensen, Alignment error envelopes for single particle analysis, *Journal of Structural Biology* 133 (3) (2001) 143 – 155.
- [183] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, Y. Cheng, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM, *Nature Methods* doi:10.1038/nmeth.2472.
- [184] A. I. Kirkland, L.-Y. S. Chang, J. L. Hutchison, Atomic resolution transmission electron microscopy, in: P. W. Hawkes, J. C. H. Spence (Eds.), *Science of Microscopy*, Springer Verlag, 2008, Ch. 1, pp. 3–64.
- [185] G. Mobus, R. Schweinfest, T. Gemming, T. Wagner, M. Rühle, Iterative structure retrieval techniques in HREM: a comparative study and a modular program package, *Journal of Microscopy* 190 (1) (1998) 109–130.
- [186] D. Tang, A. Kirkland, D. Jefferson, Optimisation of high-resolution image simulations: I. Image selection in real space, *Ultramicroscopy* 48 (3) (1993) 321–331.
- [187] D. Tang, A. Kirkland, D. Jefferson, Optimisation of high-resolution image simulations: II. Image selection in reciprocal space, *Ultramicroscopy* 53 (2) (1994) 137–146.
- [188] O. Saxton, Quantitative comparison of images and transforms, *Journal of Microscopy* 190 (1997) 52–60.
- [189] U. Lücken, P. Tiemeijer, M. Barfels, P. Mooney, B. Bailey, D. Agard, Image information transfer through a post-column energy filter with detection by a lens-coupled transmission-scintillator ccd camera, *Microscopy and Microanalysis* 14 (2008) 1322–1323.

- [190] M. G. Campbell, A. Cheng, A. Brilot, A. Moeller, D. Lyumkis, D. Veessler, J. Pan, S. Harrison, C. Potter, B. Carragher, N. Grigorieff, Movies of ice-embedded particles enhance resolution in electron cryo-microscopy, *Structure* 20 (11) (2012) 1823 – 1828.
- [191] X. Bai, I. S. Fernandez, G. McMullan, S. H. W. Scheres, Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles, *eLife Sciences* 2 (2013) –.
- [192] V. Ochkur, On the Born-Oppenheimer method in the theory of atomic collisions, *Soviet Physics - JETP* 18 (1964) 503–508.
- [193] J. B. Pendry, *Low energy electron diffraction*, Academic Press, New York, 1974.
- [194] B. Fultz, J. M. Howe, *Transmission Electron Microscopy and Diffractometry of Materials*, 2nd Edition, *Advanced Texts in Physics*, Springer Verlag, 2002.
- [195] K. Fujiwara, Relativistic dynamical theory of electron diffraction, *Journal of the Physical Society of Japan* 16 (11) (1961) 2226–2238.
- [196] A. Rother, K. Scheerschmidt, Relativistic effects in elastic scattering of electrons in TEM, *Ultramicroscopy* 109 (2) (2009) 154–160.
- [197] G. Newton, *Scattering Theory of Waves and Particles*, 2nd Edition, Springer Verlag, 1966.
- [198] C. Le Bris, P.-L. Lions, From atoms to crystals: A mathematical journey, *Bulletin of the American Mathematical Society* 42 (3) (2005) 291–363.
- [199] J. Moulton, Electrostatics, *Current Opinion in Structural Biology* 2 (2) (1992) 223–229.
- [200] N. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci* (1998) 10037–10041.
- [201] M. K. Gilson, A. Rashin, R. Fine, B. Honig, On the calculation of electrostatic interactions in proteins, *Journal of Molecular Biology* 184 (3) (1985) 503–516.
- [202] Y. C. Zhou, M. Feig, G. W. Wei, Highly accurate biomolecular electrostatics in continuum dielectric environments, *Journal of Computational Chemistry* 29 (1) (2008) 87–97.
- [203] W. Im, D. Beglov, B. Roux, Continuum solvation model: computation of electrostatic forces from numerical solutions to the poisson-boltzmann equation, *Computer Physics Communications* 111 (1998) 59–75.
- [204] K. Ishizuka, N. Uyeda, A new theoretical and practical approach to the multislice method, *Acta Crystallographica A* 33 (1977) 740–749.
- [205] M. M. J. Treacy, D. van Dyck, A surprise in the first born approximation for electron scattering, *Ultramicroscopy* 119 (2012) 57–62.
- [206] D. Fanelli, O. Öktem, Electron tomography: A short overview with an emphasis on the absorption potential model for the forward problem, *Inverse Problems* 24 (1) (2008) 013001.
- [207] H. Yoshioka, Effects of inelastic waves on electron diffraction, *Journal of the Physical Society of Japan* 12 (1957) 618–628.
- [208] P. Hodgson, *The Optical Model of Elastic Scattering*, Oxford University Press, 1963.
- [209] P. Rez, *The theory of inelastic scattering in the electron microscopy of crystals*, Ph.D. thesis, St. Catherine's College Oxford (1976).
- [210] L. Reimer, *Energy filtering Transmission Electron Microscopy*, 5th Edition, Springer Verlag, 1995.
- [211] P. Hartel, H. Rose, C. Dinges, Conditions and reasons for incoherent imaging in STEM, *Ultramicroscopy* 63 (1) (1996) 93–114.
- [212] H. Rose, Information transfer in transmission electron microscopy, *Ultramicroscopy* 15 (1984) 173–192.

- [213] P. Hartel, H. Rose, C. Dinges, Image formation by inelastically scattered electrons in electron microscopy, *Optik* 45 (1976) 139–158.
- [214] P. Schattschneider, M. Nelhiebel, B. Jouffrey, Density matrix of inelastically scattered fast electrons, *Phys. Rev. B* 59 (1999) 10959–10969.
- [215] H. Müller, A coherence function approach to image simulation, Ph.D. thesis, Technical University Darmstadt (2000).
- [216] Z. Lee, H. Rose, R. Hambach, U. Kaiser, EFTEM image calculation based on mutual coherence approach, in: *European Microscopy Congress 2012*, 2012.
- [217] E. M. Slayter, H. S. Slayter, *Light and Electron Microscopy*, 5th Edition, Cambridge University Press, 1992.
- [218] D. L. Fried, Noise in photoemission current, *Applied Optics* 4 (1) (1965) 79–80.
- [219] J. W. Goodman, *Introduction to Fourier optics*, 3rd Edition, Roberts and Company, 2005.
- [220] E. Hecht, *Optics*, 4th Edition, Addison-Wesley, 2002.
- [221] M. Wanner, D. Bach, D. Gerthsen, R. Werner, B. Tesche, Electron holography of thin amorphous carbon films: measurement of the mean inner potential and a thickness-independent phase shift, *Ultramicroscopy* 106 (2006) 341 – 345.
- [222] M. Schowalter, J. T. Titantah, D. Lamoen, P. Kruse, Ab initio computation of the mean inner coulomb potential of amorphous carbon structures, *Applied Physics Letters* 86 (11) (2005) 341 – 345.
- [223] D. S. Bethune, R. D. Johnson, J. R. Salem, M. S. de Vries, C. S. Yannoni, Atoms in carbon cages: the structure and properties of endohedral fullerenes, *Nature* 366 (1993) 123–128.
- [224] J. M. Cowley, A. F. Moodie, The scattering of electrons by thin crystals, *Journal of the Physical Society of Japan, Supplement B* 17 (2) (1962) 86–91.
- [225] J. C. H. Spence, *High-Resolution Electron Microscopy*, 3rd Edition, Oxford University Press, 2008.
- [226] S. Goudsmith, J. Saunderson, Multiple scattering of electrons, *Physical review* 57 (1940) 24–29.
- [227] X. Zhang, L. Jin, W. H. Fang, Q. Hui, Z. H. Zhou, 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry, *Cell* 141 (3) (2010) 472–482.
- [228] H. R. Saibil, Macromolecular structure determination by cryo-electron microscopy, *Acta Crystallographica D* 56 (2000) 1215–1222.
- [229] J. Frank, Single-particle reconstruction of biological macromolecules in electron microscopy – 30 years, *Quarterly Review of Biophysics* 42 (2009) 139–158.
- [230] S. Jonic, C. Véenien-Bryan, Protein structure determination by electron cryo microscopy, *Current Opinions in Pharmacology* 9 (2009) 636–642.
- [231] E. Orlova, H. Saibil, Methods for three-dimensional reconstruction of heterogeneous assemblies, *Methods Enzymology* 482 (2010) 321–341.
- [232] P. Wendler, H. R. Saibil, Cryo-electron microscopy structures of hsp100 proteins - crowbars in or out?, *Biochemistry and Cell Biology* 88 (2010) 89–96.
- [233] Z. H. Zhou, Towards atomic resolution structural determination by single-particle cryo-electron microscopy, *Current Opinion in Structural Biology* 18 (2) (2008) 218–228.
- [234] S. Bhushan, M. Gartmann, M. Halic, J. P. Armache, A. Jarasch, T. Mielke, O. Berninghausen, D. N. Wilson, R. Beckmann, Alpha-helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel, *Nature Structural and Molecular Biology* 17 (2010) 313–317.
- [235] C. V. Sindelar, K. H. Downing, An atomic-level mechanism for activation of the kinesin molecular motors, *Proceedings of the National Academy of Sciences* 107 (2010) 4111–4116.

- [236] M. L. Baker, J. Zhang, S. Ludtke, W. Chiu, Cryo-EM of macromolecular assemblies at near-atomic resolution, *Nature Protocols* 5 (2010) 1697–1708.
- [237] T. Fujii, A. H. Iwane, T. Yanagida, K. Namba, Direct visualization of secondary structures of F-actin by electron cryomicroscopy., *Nature* 59 (2010) 724–728.
- [238] R. Henderson, The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules, *Quarterly Reviews of Biophysics* 28 (1995) 171–193.
- [239] E. F. Garman, Radiation damage in macromolecular crystallography: what is it and why should we care?, *Acta Crystallographica D* 66 (2010) 339–351.
- [240] L. M. Rice, T. N. Earnest, A. T. Brunger, Single wavelength anomalous diffraction phasing revisited, *Acta Crystallographica D* 56 (2000) 1413–1420.
- [241] R. B. G. Ravelli, M. H. Nanao, A. Lovering, S. White, S. McSweeney, Phasing in the presence of radiation damage, *Journal of Synchrotron Radiation* 12 (2005) 276–284.
- [242] J. McGeehan, R. B. G. Ravelli, J. W. Murray, R. L. Owen, F. Cipriani, S. McSweeney, M. Weik, E. F. Garman, Colouring cryo-cooled crystals: online microspectrophotometry, *Journal of Synchrotron Radiation* 16 (2009) 163–172.
- [243] A. Royant, P. Carpentier, J. Ohana, J. McGeehan, B. Paetzold, M. Noirclerc-Savoye, X. Vernede, V. Adam, D. Bourgeois, Advances in spectroscopic methods for biological crystals. 1. fluorescence lifetime measurements, *Journal of Applied Crystallography* 40 (2007) 1105–1112.
- [244] J. Yano, J. Kern, K. D. Irrgang, V. K. Latimer, M. J. and Yachandra, et al., X-ray damage to the mn4ca complex in single crystals of photosystem ii: A case study for metalloprotein crystallography, *Proc. Natl Acad. Sci.* 102 (2005) 12047–12052.
- [245] J. E. McGeehan, P. Carpentier, A. Royant, D. Bourgeois, R. B. G. Ravelli, X-ray radiation-induced damage in DNA monitored by online raman, *Journal of Synchrotron Radiation* 14 (2007) 99–108.
- [246] P. Carpentier, A. Royant, J. Ohana, D. Bourgeois, Advances in spectroscopic methods for biological crystals. 2. raman spectroscopy, *Journal of Applied Crystallography* 40 (2007) 1113–1122.
- [247] A. Meents, S. Gutmann, A. Wagner, C. Schulze-Briese, Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures., *Proc. Natl Acad. Sci.* 107 (2010) 1094–1099.
- [248] T. M. Kuzay, M. Kazmierczak, B. J. Hsieh, X-ray beam/biomaterial thermal interactions in third-generation synchrotron sources, *Acta Crystallographica D* 57 (2001) 69–81.
- [249] S. Kriminski, M. Kazmierczak, R. E. Thorne, Heat transfer from protein crystals: implications for flash-cooling and X-ray beam heating, *Acta Crystallographica D* 59 (2003) 697–708.
- [250] C. Nave, M. A. Hill, Will reduced radiation damage occur with very small crystals?, *Journal of Synchrotron Radiation* 12 (2005) 299–303.
- [251] P. O'Neill, D. L. Stevens, E. F. Garman, Physical and chemical considerations of damage induced in protein crystals by synchrotron radiation: a radiation chemical perspective., *Journal of Synchrotron Radiation* 9 (2002) 329–332.
- [252] R. J. Southworth-Davies, E. F. Garman, Radioprotectant screening for cryocrystallography, *Journal of Synchrotron Radiation* 14 (2007) 73–83.
- [253] E. Nowak, A. Brzuszkiewicz, M. Dauter, Z. Dauter, G. Rosenbaum, To scavenge or not to scavenge: that is the question., *Acta Crystallographica D* 65 (2009) 1004–1006.
- [254] A. I. Barker, R. J. Southworth-Davies, K. S. Paithankar, I. Carmichael, E. F. Garman, Room-temperature scavengers for macromolecular crystallography: increased lifetimes and modified dose dependence of the intensity decay, *Journal of Synchrotron Radiation* 16 (2009) 205–216.

- [255] W. Burmeister, Structural changes in a cryo-cooled protein crystal owing to radiation damage, *Acta Crystallographica D* 56 (2000) 328–341.
- [256] R. B. G. Ravelli, S. M. McSweeney, The fingerprint that X-rays can leave on structures, *Structure* 8 (2000) 315–328.
- [257] M. Weik, R. B. G. Ravelli, G. Kryger, McSweeney, J. L. Sussman, et al., Specific chemical and structural damage to proteins produced by synchrotron radiation, *Proc. Natl Acad. Sci.* 97 (2000) 623–628.
- [258] M. C. R. Symons, Mechanism of radiation damage to proteins and DNA – an EPR perspective, *Progress in Reaction Kinetics and Mechanism* 24 (1999) 139–164.
- [259] G. Jones, J. Lea, F. Symons, M.C. Taiwo, Structure and mobility of electron gain and loss centres in proteins, *Nature* 330 (1987) 772–773.
- [260] R. J. Southworth-Davies, M. A. Medina, I. Carmichael, E. F. Garman, Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography, *Structure* 15 (2007) 1531–1541.
- [261] H. K. Leiros, S. M. McSweeney, A. O. Smalas, Atomic resolution structures of trypsin provide insight into structural radiation damage, *Acta Crystallographica D* 57 (2001) 488–497.
- [262] R. B. G. Ravelli, P. Theveneau, S. McSweeney, M. Caffrey, Unit-cell volume change as a metric of radiation damage in crystals of macromolecules, *Journal of Synchrotron Radiation* 9 (2002) 355–360.
- [263] P. Sliz, S. C. Harrison, G. Rosenbaum, How does radiation damage in protein crystals depend on X-ray dose?, *Structure* 11 (2003) 13–19.
- [264] H. K. Leiros, J. Timmins, R. B. G. Ravelli, S. M. McSweeney, Is radiation damage dependent on the dose rate used during macromolecular crystallography data collection?, *Acta Crystallographica D* 62 (2006) 125–132.
- [265] R. L. Owen, E. Rudino-Pinera, E. F. Garman, Experimental determination of the radiation dose limit for cryocooled protein crystals, *Proc. Natl Acad. Sci.* 103 (2006) 4912–4917.
- [266] J. W. Murray, E. F. Garman, R. B. G. Ravelli, X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose, *Journal of Applied Crystallography* 37 (2004) 513–522.
- [267] J. W. Murray, E. Rudino-Pinera, R. L. Owen, M. Grininger, R. B. G. Ravelli, E. F. Garman, Parameters affecting the X-ray dose absorbed by macromolecular crystals, *Journal of Synchrotron Radiation* 12 (2005) 268–275.
- [268] R. L. Owen, E. Rudino-Pinera, E. F. Garman, Know your dose: RADDPOSE, *Acta Crystallographica D* 66 (2010) 381–388.
- [269] R. Henderson, in: *Proc. R. Soc. London B*241, 1990, pp. 6–8.
- [270] J. Kmetko, N. S. Husseini, M. Naidas, Y. Kalinin, R. E. Thorne, Quantifying X-ray radiation damage in protein crystals at cryogenic temperatures, *Acta Crystallographica D* 62 (2006) 1030–1038.
- [271] M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, D. Starodub, et al., An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy, *Journal of Electron Spectroscopy and Related Phenomena* 170 (2009) 4–12.
- [272] J. M. Holton, A beginner’s guide to radiation damage, *Journal of Synchrotron Radiation* 16 (2009) 133–142.
- [273] J. M. Holton, K. A. Frankel, The minimum crystal size needed for a complete diffraction data set, *Acta Crystallographica D* 66 (2010) 393–408.
- [274] E. H. Snell, M. J. van der Woerd, M. D. Miller, A. M. Deacon, Finding a cold needle in a warm haystack: Infrared imaging applied to locating cryocooled crystals, *Journal of Applied Crystallography* 38 (2005) 69–77.

- [275] E. H. Snell, H. D. Bellamy, G. Rosenbaum, M. J. van der Woerd, Non-invasive measurement of X-ray beam heating on a surrogate crystal sample, *Journal of Synchrotron Radiation* 14 (2007) 109–115.
- [276] J. A. McMillan, S. C. Los, Vitreous ice: Irreversible transformations during warm-up, *Nature* 206 (1965) 806–807.
- [277] M. Weik, R. B. G. Ravelli, I. Silman, J. L. Sussman, P. Gros, J. Kroon, Specific protein dynamics near the solvent glass transition assayed by radiation-induced structural changes, *Protein Science* 10 (2001) 1953–1961.
- [278] C. V. Iancu, E. R. Wright, J. B. Heymann, G. J. Jensen, A comparison of liquid nitrogen and liquid helium as potential cryogens for electron cryotomography, *Journal of Structural Biology* 153 (2006) 231–240.
- [279] W. H. Massover, Radiation damage to protein specimens from electron beam imaging and diffraction: a mini-review of anti-damage approaches, with special reference to synchrotron X-ray crystallography, *Journal of Synchrotron Radiation* 14 (2007) 116–127.
- [280] J. Z. Chen, C. Sachse, C. Xu, T. Mielke, C. M. T. Spahn, N. Grigorieff, A dose-rate effect in single-particle electron microscopy, *Journal of Structural Biology* 161 (2008) 92–100.
- [281] B. E. Bammes, J. Jakana, M. F. Schmid, W. Chiu, Radiation damage effects at four specimen temperatures from 4 to 100 K, *Journal of Structural Biology* 169 (2010) 331–341.
- [282] W. J. Royer, K. Strand, M. van Heel, W. Hendrickson, Structural hierarchy in erythrocrucorin, the giant respiratory assemblage of annelids, *Proc Natl Acad Sci* 97 (13) (2000) 7107–7111.
- [283] W. E. Royer Jr, H. Sharma, K. Strand, J. E. Knapp, B. Bhyravbhatla, Lumbricus erythrocrucorin at 3.5 Å resolution: architecture of a megadalton respiratory complex, *Structure* 14 (2006) 1167–1177.
- [284] B. Kastner, N. Fischer, M. M. Golas, B. Sander, H. Stark, GraFix: sample preparation for single-particle electron cryomicroscopy, *Nature Methods* 5 (2008) 53–55.
- [285] H. Stark, Chapter five - graFix: Stabilization of fragile macromolecular complexes for single particle cryo-EM, in: G. J. Jensen (Ed.), *Cryo-EM Part A Sample Preparation and Data Collection*, Vol. 481 of *Methods in Enzymology*, Academic Press, 2010, pp. 109–126.
- [286] J. R. Kremer, D. N. Mastrorade, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD, *Journal of Structural Biology* 116 (1996) 71–76.
- [287] J. Crank, P. Nicolson, A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type, *Advances in Computational Mathematics* 6 (1996) 207–226.
- [288] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2004) 91–110.
- [289] K. Mikolajczyk, T. Tuytelaars, C. Schmid, L. V. Gool, et al., A comparison of affine region detectors, *International Journal of Computer Vision* 65 (2006) 43–72.
- [290] H. Y. Liao, J. Frank, Definition and estimation of resolution in single-particle reconstructions, *Structure* 18 (2010) 768–775.
- [291] K. Stenn, G. F. Bahr, Specimen damage caused by the beam of the transmission electron microscope, a correlative reconsideration, *Journal of Ultrastructure Research* 31 (1970) 526–550.
- [292] J. Cope, S. Gilbert, I. Rayment, D. Mastrorade, A. Hoenger, Cryo-electron tomography of microtubule-kinesin motor complexes, *Journal of Structural Biology* 170 (2010) 257–265.
- [293] L. A. Baker, J. L. Rubinstein, Radiation damage in electron cryomicroscopy, in: G. J. Jensen (Ed.), *Cryo-EM Part A Sample Preparation and Data Collection*, Academic Press, 2010, Ch. 15, pp. 371–388.
- [294] M. Barcena, A. J. Koster, Electron tomography in life science, *Seminars in Cell and Developmental Biology* 20 (2009) 920–930.

- [295] B. F. McEwen, K. H. Downing, R. M. Glaeser, The relevance of dose-fractionation in tomography of radiation-sensitive specimens, *Ultramicroscopy* 60 (3) (1995) 357–373.
- [296] D. Typke, C. J. Gilpin, K. H. Downing, R. M. Glaeser, Stroboscopic image capture: Reducing the dose per frame by a factor of 30 does not prevent beam-induced specimen movement in paraffin, *Ultramicroscopy* 107 (2-3) (2007) 106–115.
- [297] W. O. Saxton, J. Frank, Motif detection in quantum noise-limited electron micrographs by cross-correlation, *Ultramicroscopy* 2 (1977) 219–227.
- [298] K. Diederichs, S. McSweeney, R. B. G. Ravelli, Zero-dose extrapolation as part of macromolecular synchrotron data reduction, *Acta Crystallographica D* 59 (2003) 903–909.
- [299] M. Weik, J.-P. Colletier, Temperature-dependent macromolecular X-ray crystallography, *Acta Crystallographica D* 66 (2010) 437–446.
- [300] B. Kauffmann, M. Weiss, V. Lamzin, A. Schmidt, How to avoid premature decay of your macromolecular crystal - a quick soak for long life, *Structure* 14 (2006) 1099–1105.
- [301] E. De la Mora, I. Carmichael, E. Garman, Effective scavenging at cryotemperatures: further increasing the dose tolerance of protein crystals, *Journal of Synchrotron Radiation* 18 (2000) 313–317.
- [302] R. M. Glaeser, D. Typke, P. C. Tiemeijer, J. Pulokas, A. Cheng, Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM, *Journal of Structural Biology* 174 (1) (2011) 1–10.
- [303] X. Zhang, Z. H. Zhou, Limiting factors in atomic resolution cryo-electron microscopy: No simple tricks, *Journal of Structural Biology* 175 (3) (2011) 253–263.
- [304] B. Buijsse, F. van Laarhoven, A. K. Schmid, R. Cambie, S. Cabrini, J. Jin, R. M. Glaeser, Design of a hybrid double-sideband/single-sideband (schlieren) objective aperture suitable for electron microscopy, *Ultramicroscopy* 111 (12) (2011) 1688–1695.
- [305] A. Rigort, F. Bäuerlein, A. Leis, M. Gruska, C. Hoffmann, T. Laugks, U. Böhm, M. Eibauer, H. Gnaegi, W. Baumeister, J. Plitzko, Micromachining tools and correlative approaches for cellular cryo-electron tomography, *Journal of Structural Biology* 172 (2) (2010) 169–179.
- [306] M. van Heel, W. Keegstra, E. F. J. Schutter, W. G. van Bruggen, Arthropod hemocyanin studied by image analysis, *Life Chemistry Reports Supplement 1* (1982) 69–73.
- [307] G. Harauz, M. van Heel, Exact filters for general geometry 3-dimensional reconstruction, *Optik* 73 (1986) 146–156.
- [308] W. Saxton, W. Baumeister, M. Hahn, Three-dimensional reconstruction of imperfect two-dimensional crystals, *Ultramicroscopy* 13 (1984) 57–70.
- [309] N. Grigorieff, S. C. Harrison, Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy, *Current Opinion in Structural Biology* 21 (2011) 265–273.

Summary

Modeling of Image Formation in Cryo-Electron Microscopy

PhD thesis
by Miloš Vulović

Knowledge of the structure of biological specimens is crucial for understanding life. Cryo-electron microscopy (cryo-EM) permits structural studies of biological specimen at their near-native state. The resolution is limited to typically 4 – 6 Å largely owing to: i) the noise and blurring of the detector; ii) the oscillatory and dampening character of the contrast transfer function (CTF) originating from defocusing which is employed to produce contrast; and iii) the radiation damage which limits the integrated electron flux that can be used, resulting in images with poor signal-to-noise ratio.

In order to obtain the best cryo-EM data, it is important to optimize the data acquisition strategy. Simulations of image formation (forward modeling) provide the possibility to easily and cost-effectively investigate the performance as a function of specimen and microscope settings.

The main goal of this thesis is to improve our understanding of the relevant physical processes that govern image formation and to develop a quantitative forward model. An accurate forward model is an essential tool for optimizing the acquisition strategy, assisting the regularization (introduction of prior information) in the 3D reconstruction, improving image interpretation, and achieving a resolution beyond the limits imposed by the oscillatory CTF.

This thesis addresses the following challenges: i) construction of the electron-specimen interaction potential based on elastic and inelastic electron scattering properties and adequate description of the electron propagation through the specimen; ii) accurate estimation of the CTF parameters, in particular defocus and astigmatism and their uncertainties, iii) characterization of the detector including all relevant statistics; iv) better understanding of certain aspects of radiation damage such as specimen heating, dose-rate effects, and beam-induced movements.

The specimen's interaction potential is constructed from available atomic structures. It is calculated initially *via* electron scattering factors treating the atoms as if they are in isolation. This potential is extended for the influence of the embedding medium that causes an effective charge

redistribution. The potential from these redistributions is computed *via* a Poisson-Boltzmann continuum electrostatics approach and its contribution is shown to be less than 10 %. Depending on the strength, frequency content and thickness of the interaction potential map, the electron wave is propagated by one (or a combination) of the following approaches: weak-phase object approximation, projection assumption, thick-phase object approximation and multislice approach. For most specimens imaged in this thesis, the projection assumption is strictly speaking violated, implying that the thickness cannot be neglected.

The full forward model was validated against experimental images of hemoglobin, 20S proteasome and GroEL. Simulated images adequately predict the observed phase contrast introduced by defocusing, changes due to the integrated electron flux, inelastic scattering, camera detective quantum efficiency, and acceleration voltage. We discuss that beam-induced specimen movements are relevant in the experiments, while the influence of the solvent amorphousness, modeled *via* a molecular dynamics approach, can be neglected. All parameters are based on physical principles and, when necessary, estimated from independent measurements.

The key parameters governing the CTF are defocus and twofold astigmatism. Estimation of these aberrations and the associated uncertainties is based on patterns in the power spectrum density known as Thon rings. Our presented method suppresses the background using an adaptive filtering strategy and uses template matching to estimate the shape of the rings. The frequencies of the detected rings are used to estimate the defocus and its uncertainty, while the ellipticity provides the estimation of astigmatism. By utilizing information from various rings, an outlier rejection routine is applied contributing to the high robustness of the algorithm. Tests show high accuracy and very good agreement between simulated and estimated defocus and astigmatism. The reproducibility of the algorithm is evaluated on experimental data by repeating measurements under identical imaging conditions and by analyzing the linearity of the stigmator response. By using a new Thon ring averaging method, the modulation depth of the rings in a 1D averaged power spectrum density can be enhanced compared to elliptical averaging. This provides a better information-transfer assessment of the microscope in the presence of spherical aberration.

A set of algorithms is presented for a comprehensive characterization of the detector. The methods provide fixed-pattern noise correction and statistics for hot and bad pixels, modulation transfer function, conversion factor, readout noise, dark-current and shot noise as well as detective quantum efficiency.

Radiation damage is related to the energy deposited in the specimen by inelastic scattering. The deposited energy per mass, *dose*, is estimated from the incident flux, beam size, average sample composition and thickness, and acceleration voltage. The dose allows us to get an upper estimate of the concentrations of radicals that are built up in the specimen, and to perform heat-transfer calculations. Comparison of equal-dose stroboscopic exposure series collected with different incident fluxes and integration times showed a dose-rate effect favoring lower fluxes. For typical experimental settings the effect of beam-induced heating is felt within milliseconds and is not expected to be a problem for samples with good thermal contact at low to medium electron fluxes. The potential radical concentrations are very high, suggesting that the radical recombination must play a role in the observed dose-rate effects. We advocate the use of strobo-

scopic data collection, with which variable dose can be used for the various images, providing a controlled amount of damage. With a new generation of direct electron detectors, exposure series will become common practice. Such data could facilitate reduction of the effect of beam-induced movements by retrospective alignment and averaging of the raw frames.

The theory and methods provided in this thesis represent a basis for modeling the outcome of data acquisition for any combination of the requested parameters, and form the essence of an expert system that would optimize the data collection strategy. Furthermore, the influence of new hardware components could be inexpensively and efficiently investigated.

Samenvatting

Modellering van Beeldvorming in Cryo Elektronenmicroscopie

Proefschrift
door Miloš Vulović

Kennis van de structuur van biologische monsters is essentieel om het leven te begrijpen. Cryo- elektronenmicroscopie (cryo-EM) maakt studie van biologische specimen in hun nabij-natuurlijke staat. De resolutie is beperkt tot typisch 4-6 Å wat grotendeels bepaald wordt door: i) de ruis van en het uitsmeren door de detector; ii) het oscillerende en dempende karakter van de contrast overdrachts functie (COF), veroorzaakt door het onscherp stellen dat wordt gebruikt om contrast te produceren, en iii) stralingsschade die de geïntegreerde flux van elektronen beperkt, wat resulteert in beelden met een slechte signaal-ruis verhouding.

Om de beste cryo-EM data te verkrijgen, is het belangrijk om de data-acquisitie strategie te optimaliseren. Simulaties van de beeldformatie (voorwaartse modellering) bieden de mogelijkheid om eenvoudig en kosteneffectief de prestaties als een functie van het monster en microscoop instellingen te onderzoeken.

Het belangrijkste doel van dit proefschrift is om ons begrip van de relevante fysische processen te vergroten die de beeldformatie beheersen en om een kwantitatief voorwaarts model te ontwikkelen. Een nauwkeurig voorwaarts model is een essentieel hulpmiddel voor het optimaliseren van de acquisitiestrategie, het bijstaan van de regularisatie (introductie van *a priori* informatie) in de 3D-reconstructie, het verbeteren van beeld interpretatie, en het bereiken van een resolutie die de grenzen, opgelegd door de oscillerende COF, overschrijd.

Dit proefschrift richt zich op de volgende uitdagingen: i) de opbouw van de elektron-monster interactiepotentiaal op basis van elastische en niet-elastische elektron verstrooiings eigenschappen en het adequaat beschrijven van elektron propagatie door het specimen; ii) nauwkeurige schatting van de COF parameters, met name onscherpte en astigmatisme en de daarbij horende onzekerheden, iii) karakterisering van de detector inclusief alle relevante statistieken; iv) beter begrip van bepaalde aspecten van stralingsschade zoals specimen opwarming, dosiseffecten, en bundel geïnduceerde bewegingen.

De interactiepotentiaal van het monster wordt geconstrueerd uit beschikbare atomaire struc-

turen. Deze wordt aanvankelijk berekend *via* elektron verstrooiingsfactoren die de atomen geïsoleerd behandelen. Deze potentiaal wordt verfijnd met de invloed van het inbeddingsmedium, wat effectief een ladingsherverdeling veroorzaakt. De potentiaal van deze herverdelingen wordt berekend *via* een Poisson- Boltzmann continuüm elektrostatische benadering; haar bijdrage blijkt minder dan 10 % te zijn. Afhankelijk van de sterkte, frequentieinhoud en de dikte van de interactiepotentiaal, wordt de voortplanting van de elektrongolf beschreven door één (of een combinatie) van de volgende methoden: zwakke faseobject benadering, projectie aanname, dikkefaserasteren benadering, en meerplaksaanpak. Voor de meeste specimina afgebeeld in dit proefschrift, wordt de projectie aanname strikt genomen, geschonden, hetgeen impliceert dat de dikte niet kan worden verwaarloosd.

Het volledige voorwaartse model werd gevalideerd met experimentele beelden van hemoglobine, 20S proteasoom en GroEL. Gesimuleerde afbeeldingen voorspellen adequaat het waargenomen fasecontrast geïntroduceerd door het onscherp stellen, veranderingen als gevolg van de geïntegreerde flux van elektronen, niet-elastische verstrooiingen, detectieve kwantumefficiëntie van de camera en versnellingsspanning. Wij stellen dat bundel-geïnduceerde monsterbewegingen relevant zijn in de experimenten, terwijl de invloed van het oplosmiddel amorfheid, gemodelleerd *via* een moleculaire dynamica benadering, kan worden verwaarloosd. Alle parameters zijn gebaseerd op fysische principes en, indien nodig, worden geschat op basis van onafhankelijke metingen.

De belangrijkste parameters die de COF bepalen zijn onscherpte en tweeledig astigmatisme. Schatting van deze aberraties en hun bijbehorende onzekerheden wordt gedaan aan de hand van de patronen in het vermogensdichtheidsspectrum, bekend als Thon ringen. Onze gepresenteerde methode onderdrukt de achtergrond door gebruik van een adaptieve filtering strategie en gebruikt een sjabloon vergelijkingsmethode om de vorm van de ringen te schatten. De frequenties van de gedetecteerde ringen bepalen de onscherpte en de bijbehorende onzekerheid, terwijl de ellipticiteit een schatting geeft van het astigmatisme. Door gebruik te maken van informatie uit verschillende ringen, wordt een uitschieter verwerping routine toegepast die bijdraagt aan de hoge robuustheid van het algoritme. Tests tonen een hoge nauwkeurigheid en een zeer goede overeenkomst tussen gesimuleerde en de geschatte onscherpte en astigmatisme. De reproduceerbaarheid van het algoritme wordt beoordeeld op experimentele data met herhaalde metingen onder identieke beeldomstandigheden en analyse van de lineariteit van de stigmator. Door gebruik te maken van een nieuwe Thon ring middelingsmethode, kan de modulatie diepte van de ringen in een 1D gemiddeld vermogensspectrum worden vergroot in vergelijking tot elliptische middeling. Dit geeft een betere evaluatie van de informatieoverdracht van een microscoop in de aanwezigheid van sferische aberraties.

Een set van algoritmes is ontwikkeld voor een uitvoerige karakterisering van de detector. De methodes leveren vast-patroon ruis correctie en statistieken voor hete en defecte pixels, modulatieoverdrachtsfunctie, omrekeningsfactor, uitlezingruis, donkere stroom en hagelruis evenals detectieve kwantumefficiëntie.

Stralingschade is gerelateerd aan de energie die wordt overgedragen aan het monster door niet-elastische verstrooiing. De overgedragen energie per massa, dosis, wordt geschat op basis van de invallende elektronen flux, bundelgrootte, gemiddeld monster samenstelling en dikte,

en versnellingsspanning. De dosis kan worden gebruikt om een bovengrens te schatten van de concentraties van opgebouwde radicalen in het monster, en om de hitte overdracht te kunnen afschatten. Vergelijking van gelijke dosis stroboscopische opnameseries verzameld met verschillende invallende flux en integratietijden toonden een dosistempo effect die de lagere flux bevoordeelden. Voor een typische experimentele situatie wordt verondersteld dat het effect van de bundel-geïnduceerde verwarming wordt waargenomen binnen enkele milliseconden en geen problemen veroorzaakt voor monsters met een goed thermisch contact op lage tot middelmatige elektronen flux. De potentiële concentratie van radicalen is zeer hoog, wat suggereert dat radicalen recombinitie een rol moet spelen bij de waargenomen dosistempo effecten. Wij pleiten voor het gebruik van stroboscopische dataverzameling, waarmee variabele doseringen kunnen worden gebruikt voor de verschillende beelden, wat resulteert in een gecontroleerde hoeveelheid schade. Met een nieuwe generatie van directe-elektronendetectors worden opnameseries gemeengoed. Dergelijke data kan bijdragen aan het verminderen van bewegingsonscherpte in beelden veroorzaakt door bundel-geïnduceerde bewegingen door retrospectieve uitlijning en middeling van de ruwe frames.

De theorie en methoden uit dit proefschrift vormen een basis voor het modelleren van de uitkomsten van data acquisitie voor elke combinatie van de gewenste parameters and zijn de essentie van een systeem dat de data acquisitie strategie automatisch zou kunnen optimaliseren. Bovendien kan de invloed van nieuwe hardware onderdelen voordelig en doeltreffend onderzocht worden.

List of publications

Journals

1. **M. Vulović**, R. B. G. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lübben, H. Rullgård, O. Öktem, B. Rieger: “Image formation modeling in cryo-electron microscopy”, *Journal of Structural Biology*, 2013, *Accepted*
2. **M. Vulović**, E. Franken, R. B. G. Ravelli, L. J. van Vliet, B. Rieger: “Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy”, *Ultramicroscopy*, 2012, Vol. 116, No. 1, pp. 115-134.
3. M. Karuppasamy, F. Karimi-Nejadasl, **M. Vulovic**, A. J. Koster, R. B. G. Ravelli: “Radiation damage in single particle cryo-electron microscopy: effects of dose and dose rate”, *Journal of Synchrotron Radiation*, 2011, Vol. 18, No. 3, pp. 398-412.
4. **M. Vulovic**, B. Rieger, L.J. van Vliet, A.J. Koster, R.B.G. Ravelli: “A toolkit for the characterization of CCD cameras for transmission electron microscopy”, *Acta Crystallographica D*, 2010, Vol. 66, No. 1, pp. 97-109.

Submitted

5. **M. Vulović**, L. M. Voortman, L. J. van Vliet, B. Rieger: “When to use the projection assumption and weak-phase object approximation in phase contrast cryo-EM”, 2013.

Conference papers

1. **M. Vulovic**, P. Brandt, R. B. G. Ravelli, A. J. Koster, L. J. van Vliet, B. Rieger: “Estimation of defocus and astigmatism in transmission electron microscopy”, *IEEE International Symposium on Biomedical Imaging*, 2010, pp. 1121-1124.

Conference proceedings

1. **M. Vulovic**, R.B.G. Ravelli, L.J van Vliet, A.J. Koster, and B. Rieger: “Experimental evaluation of a forward model in cryo-electron microscopy”, *Microscopy Conference (MC2013)*, Regensburg, 2013
2. L.M. Voortman, **M. Vulovic**, L.J. van Vliet and B. Rieger: “When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM”, *Microscopy Conference (MC2013)*, Regensburg, 2013
3. **M. Vulovic**, R.B.G. Ravelli, L.J van Vliet, A.J. Koster, I. Lazić, and B. Rieger: “Modelling of image formation in cryo electron microscopy”, *Dutch meeting on Molecular and Cellular Biophysics*, Veldhoven, 2012
4. **M. Vulovic**, R.B.G. Ravelli, L.J van Vliet, A.J. Koster, O. Öktem, H. Rullgård, I. Lazić, C. Bajaj, A. Rand, and B. Rieger: “Modelling of image formation in cryo electron microscopy”, *The 15th European Microscopy Congress (EMC2012)*, Manchester, 2012
5. **M. Vulovic**, E. Franken, L.J. van Vliet, A.J. Koster, R.B.G. Ravelli, and B. Rieger: “Un-biased and Precise Determination of Defocus and Astigmatism in Transmission Electron Microscopy”, *Microscopy Conference (MC 2011)*, Kiel, 2011
6. **M. Vulovic**, E. Franken, L. J. van Vliet, and B. Rieger: “Accurate astigmatism and defocus estimation in TEM”, *Sixth International Congress on Electron Tomography*, EMBL Heidelberg, 2011
7. E. Franken, **M. Vulovic**, B. Rieger, R. Wagner, and R. Schoenmakers: “Applications of Accurate Defocus and Astigmatism Estimation”, *Sixth International Congress on Electron Tomography*, EMBL Heidelberg, 2011
8. F. Karimi-Nejadasl, M. Karuppasamy, **M. Vulovic**, B. Koster, and R. Ravelli: “Radiation Damage Studies in Single Particle Cryo-Electron Microscopy”, *Nederlandse Vereniging voor Microscopie, Annual Meeting*, Amsterdam, 2010
9. **M. Vulovic**, R.B.G. Ravelli, A.J. Koster, L.J. van Vliet, and B. Rieger: “Forward modeling in cryo-electron tomography of macromolecular complexes”, *17th International Microscopy Congress*, Rio de Janeiro, (IMC17), 2010
10. **M. Vulovic**, P.L. Brandt, L. Voortman, A.A. Joseph, R.G.B. Ravelli, A.J. Koster, L.J. van Vliet, and B. Rieger: “Contrast Transfer Function estimation for tilted, non-thin, weakphase objects”, *Microscopy Conference*, Graz, (MC 2009), 2009

Acknowledgments

In 2008 I moved to The Netherlands and started my PhD on a research collaboration between the Quantitative Imaging (QI) group at TU Delft and Electron Microscopy (MCB-EM) group at the LUMC. By living and working in The Netherlands, I have gained invaluable personal and professional experience. I feel I owe my gratitude to many great people who were all, in their own way, important to the realization of this thesis. My apologies to all those I might have forgotten to include.

Firstly, I would like to express my gratitude to my co-promoter and daily supervisor at TU Delft, Bernd Rieger, whose supervision was crucial for me and with whom I shared most of the challenges of my PhD work. He has been dedicated, reliable, and always available to me. Many thanks especially for asking the right questions, showing me how to best present the results, and for the patience in editing (shortening) my over-enthusiastic manuscripts.

To my daily supervisor at the LUMC, Raimond Ravelli, I owe my gratitude for teaching me numerous things, from how to fly a kite and my first phrases in Dutch to practical electron microscopy, as well as for showing me what it means to be truly devoted to science. Special thanks for believing in me when I needed it most. I would also like to acknowledge his input to the manuscripts in Chapters 2 and 5 related to X-ray crystallography and for his editing.

I am sincerely grateful to my promoter Lucas van Vliet (TU Delft) for helping me manage my PhD work. Our meetings and his feedback and support were essential in finding my way through this research. Once again, I would like to thank him and Bernd for giving me the opportunity to do the PhD and allowing me to continue in a post-doctoral position.

I would like to thank my promoter Bram Koster (LUMC) for useful discussions and for openness, enthusiasm and support. He always kept a positive attitude no matter what the situation was.

My PhD project was a part of a larger industrial partnership program with FEI Company. Close and active collaboration with FEI introduced me to the company environment. Numerous people were involved, and I especially want to thank the ones who contributed the most to my work. First, I would like to thank Erik Franken who has been in charge of this project within FEI. He gave valuable input and made a contribution to the development of defocus and astigmatism software (Chapter 3). In addition to my MATLAB implementation, a prototype software has been implemented by Erik and used within FEI company. Furthermore, his approachability and expertise made our scientific discussions friendly and serious at the same time. I owe special thanks to Ivan Lazić, my fellow Serbian, in addition to friendly relations, for very enthusiastic discussions about physics. Forward modeling software InSilicoTEM (Chapter

4) was transferred to FEI and further used for modeling. Ivan's feedback has helped in further improvement of the software. I would also like to thank Remco Schoenmakers for his initial contribution to the project, friendly approach and prudent remarks.

During October and November 2011, I was a visiting scientist at CIAM (Center for Industrial and Applied Mathematics) at KTH in Stockholm. I would like to thank Ozan Öktem for great hospitality and interesting discussions. I would like to acknowledge his input related to Intermezzo 1d and 1e in Appendix F. My gratitude also goes to Hans Rullgård for contributing to useful discussion on construction of the interaction potential (Chapter 4). The internship in Sweden gave me an important additional international experience. And I fell in love with Stockholm.

I would like to thank people from the Netherlands Center for Electron Nanoscopy (NeCEN) where I acquired my experimental data presented in Chapter 4. Gert Oostergetel, for his patience and kind approach and for providing earthworms as a sample. Although the results of characterization of Falcon direct electron detector are not included in this thesis, I would like to thank Sacha De Carlo for help during those experiments.

The help of Paul Mooney (Gatan) was crucial for determining the conversion factors of the detectors characterized in Chapter 2. I am most grateful to him for help related to the MTF and DQE of the GIF cameras and for all interesting discussions at the Gordon conferences.

The names of several other people contributing to the work are already mentioned in the acknowledgments of the particular chapters and in the author list of the publications.

I would like to thank Q. Xu and S. Malladi for a day behind a C_s -corrected microscope.

Special thanks go to my present and former officemates both at TU Delft and the LUMC: Lennard for countless inspiring discussions, for help with many small and big daily issues, for motivating me in *het Nederlands te spreken* and for all the good times spent at conferences; Fatemeh for being so supportive, for various programming tips and interesting discussions; Mani for always being ready to help; Pieter Brandt for fun and inspirational time during his Master thesis project; and Sanneke for having many useful advices for a PhD newcomer and for the help with DipImage. I would like to acknowledge contribution of Lennard to the Appendix G, Fatemeh and Mani to Chapter 5, and Pieter to Chapter 3.

I would like to thank all members of QI for great coffee breaks, lab uitjes, borrels, lunches and various other gatherings. You all make this group quite interesting (QI). I am most grateful to the QI secretary, Mandy, for being so caring and helping me with any kind of problem. Ronald L. is actually making a lot of sense by walking around wearing the T-shirt with the phrase "Stop making sense". His extraterrestrial positive spirit makes our daily work much more fun and stubborn practical computer problems easy. I am thankful to Wim for being helpful and for all the coffees, *kruidnoten* and chocolate Easter eggs. I especially want to thank Heidi for helping me cope with challenges of the PhD road, being such a lovely and understanding person and for the translation of Goethe quote in my propositions from German to Dutch (thanks to Frankie as well). I cannot express enough gratitude to Frans Fos for all the involvement and help when visiting AMC. I thank Ted, among others for organizing borrels before Bernd and for the punch. Sjoerd for being approachable. Alex for being great drinks- and Dutch course- buddy.

Thanks also to all QI PhD students for the fun times in and outside the office: Nadya, Mojtaba, Jelena, Robiël, Zhang, Martijn B., Thijs, Jeroen H., Qiaole, Rahil, Robert, Jianfei, Jeroen v S., TT, Ganesh, Kedir, Christiaan and former employees Arun, Martijn vd G., Matthan, Melanie, Vincent, Arno, Rosalie, Piet, Fred.

I haven't enjoyed myself any less at the LUMC. Many thanks to dear Cristina for such a friendly attitude, and for help with extracting the hemoglobin sample from the worms we dug up together from the ground behind the LUMC. Roman and Montse, being the real experts in (cryo-)EM tomography, I feel I could never learn enough from them. I am most grateful to them for all the discussions and help with experiments (Roman for plunge freezing as well) and for the great times at Gordon conferences. I would like to thank Frank F. for being willing to help me whenever he could with computer and network related issues (acknowledgements to Willem as well). Many thanks go to Erik B. and Ronald L. for always being ready to help me with microscope related issues. Although I have not visited the LUMC very often in the last 6 months, I still feel like a part of the *family*: also thankfully to Marjon, Barbara, Christoph and Kasia C. Many thanks as well to all former employees Sandra, Karen, Kasia M., Kevin, Jack, Linda, Karine, Jos, and Mieke for making my working days enjoyable.

Thanks also go to all other members and master/bachelor students of QI/MCB-EM groups for the good work atmosphere.

Returning to my thesis, I am grateful to FOM for the financing of the project and Ms Annette Bor for all the help. Veronika Pišorn for proofreading some parts of the text and propositions, Jelena Popadić for design of the cover, Jacob Gesink and Lennard for helping me with translation of the *sammenvating*, Bojan for the help with translation of *stellingen*. I am most grateful to committee members for agreeing to participate in my PhD defense.

While living in Delft I met a lot of wonderful people from all around the world. Surrounded with such a large number of cultural influences enriched me as a person. I thank them all for that. I am also very thankful to my friends in The Netherlands from ex-Yugoslavia and for their love and support throughout all these years. I owe special thanks to: Marko Mihailović for being so enthusiastic, for organizing all massive gatherings and being so contagiously positive; Darko i Jasmina for being so interesting and inspiring (Darko also for the help with Intermezzo 4 in Appendix F); Miloš i Jelena Popadić for always being ready to offer and provide help; Andjela for always reminding me how great is it to be different; Ljubiša and Aileen for good company and help in dealing with the administration; Yannick and Sandra for helping me get around at the very beginning of my arrival to Delft.

Finally, I would like to express my deepest gratitude to the following people:

To Maja Rudinac for helping me come to Delft and for always being so supportive. She and Stevan took care of me like family when I moved to The Netherlands.

To Tanja Matić for keeping my cultural life in the Netherlands alive.

To Aca and Veronika for their love, for help, for inspirational discussions, for making me happy and preparing healthy food.

To Dubravka Arandjelović for being there through better and worse; for making me laugh

out loud every day. Together with Bojan, for always making me feel like home and for helping me with Dutch language.

To Jale, for everything.

To Vukica, Vule and Dušan, my wonderful family, for their love and dedication. Mama, tata, Dule, hvala vam mnogo.

To my close friends and family in Serbia, who I carry in my heart wherever I might be.

Miloš Vulović
Delft, May 2013

Biography

Miloš Vulović obtained a graduate engineering degree from the Faculty of Electrical Engineering, University of Belgrade in 2008, graduating from the Department of Physical Electronics (Technical Physics) Group for Optoelectronics and Laser Engineering. His thesis included the design and simulation of specific features of anti-resonant reflecting optical waveguides (ARROW) and was conducted within the Royal Society International joint project of ATI (UK), ETF (Serbia), Intel Corporation (USA), ITC-IRST (Italia) and Politecnico di Bari (Italia).

He started to work towards a PhD degree in 2008 on a research collaboration between Delft University of Technology and Leiden University Medical Center. The project was a part of a larger industrial partnership program with FEI Company. His research has been conducted in the field of cryo-electron microscopy and focused on the physical modeling of image formation. This work has resulted in methods (related to forward model and parameters estimation) developed to aid, among others, the optimization of the data collection strategy and resolution improvements in cryo-electron microscopy. In October and November 2011 he was a visiting scientist at CIAM at the Royal Institute of Technology (KTH), Stockholm, Sweden.

Since March 2013, he has been working as a post-doctoral researcher at TU Delft/LUMC.

