CTF Correction *in* Cryo-Electron Tomography

Simulation, Experiments and Reconstruction

CTF Correction *in* Cryo-Electron Tomography

SIMULATION, EXPERIMENTS AND RECONSTRUCTION

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 maandag 12 mei 2014 om 15:00 uur

door

Lenard Maarten VOORTMAN

natuurkundig ingenieur
geboren te Geldrop.

Dit proefschrift is goedgekeurd door de promotor:

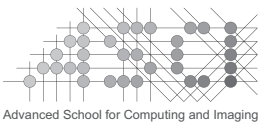
Prof. dr. ir. L. J. van Vliet

Copromotor:

Dr. B. Rieger

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. L. J. van Vliet,	Technische Universiteit Delft, promotor
Dr. B. Rieger,	Technische Universiteit Delft, copromotor
Prof. dr. A. Engel,	Technische Universiteit Delft
Prof. dr. J. M. Carazo,	Spanish National Center for Biotechnology
Prof. dr. J. M. Plitzko,	Universiteit Utrecht
	Max Planck Institute of Biochemistry
Dr. E. M. Franken,	FEI Company
Dr. J. A. G. Briggs,	European Molecular Biology Laboratory Heidelberg
Prof. dr. ir. P. Kruit,	Technische Universiteit Delft, reservelid



This work was carried out in the ASCI graduate school.
ASCI dissertation series number 298.

Printed by: Ridderprint BV, Ridderkerk, The Netherlands

Front & Back: Illustration of 70S Ribosome with superimposed defocus gradient.

ISBN 978-90-5335-844-3

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

Reis ver, drink wijn, denk na
Lach hard, duik diep
Kom terug
ERIK DE JONG — Spinvis

This work is part of the research programme *Microscopy and modification of nanostructures 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.

Contents

1

Introduction

Ontrent halff Septembr. 1675. ontdeckten
ick in regenwater kleijne diertgens.

ANTONIE VAN LEEUWENHOEK

I believe that we do not know anything for certain,
but everything probably.

CHRISTIAAN HUYGENS

1

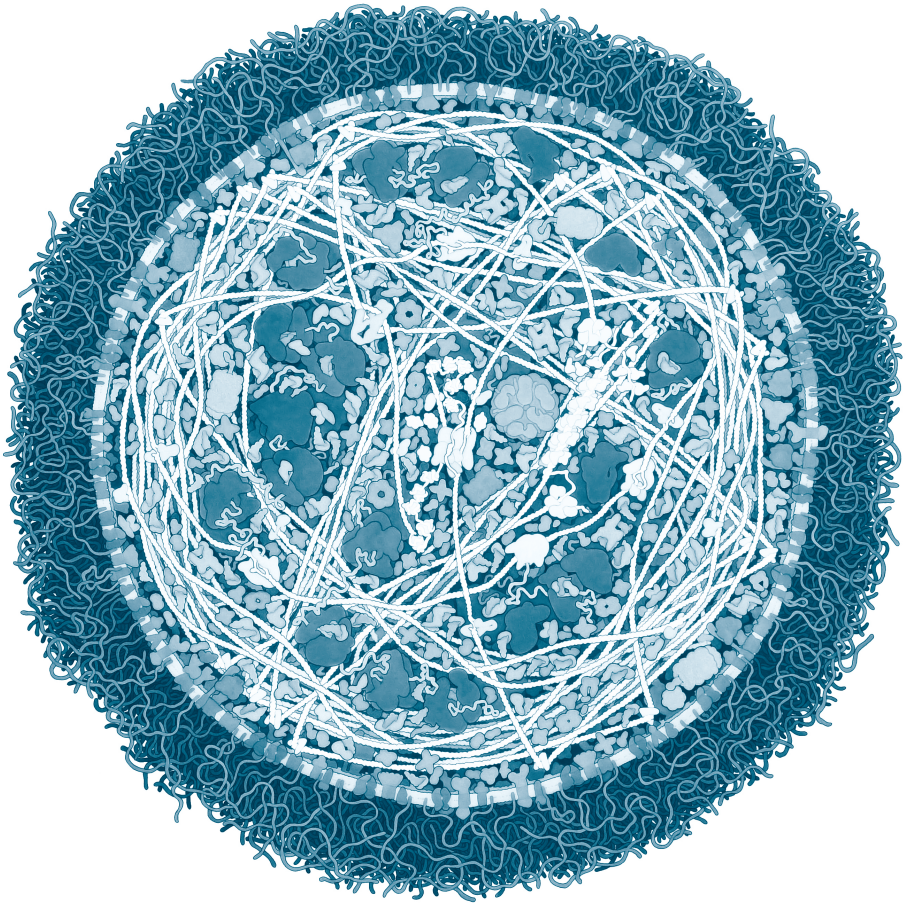


Illustration of Mycoplasma mycoides bacterium, by David S. Goodsell, the Scripps Research Institute.

1.1. *Imaging the building blocks of life*

"Every cell in nature is a thing of wonder. Even the simplest are far beyond the limits of human ingenuity. To build the most basic yeast cell, for example, you would have to miniaturize about the same number of components as are found in a Boeing 777 jetliner and fit them into a sphere just 5 microns across; then somehow you would have to persuade that sphere to reproduce."

BILL BRYSON — A Short History *of* Nearly Everything

Life inside a cell is complicated, crowded and highly dynamic. All this whirling around of proteins goes on continuously without us taking much notice. Nevertheless, what happens on this microscopic scale has a dramatic influence on our daily live. Many human diseases are caused by viruses, bacteria, fungi, prions or parasites. What almost all these pathogens have in common, is that they are too small to see with the naked human eye.

The study of this elusive microscopic world made a huge leap in the 17^{th} century when Antonie van Leeuwenhoek, a Dutch tradesman from Delft, handcrafted the world's first high-magnification microscope. Before Van Leeuwenhoek, microscopes could magnify only up to approximately 30 times. Hence, the scientific community was surprised when a self-taught tradesman achieved magnifications of around 300 times. This remarkable magnification led Van Leeuwenhoek to the discovery of single-celled organisms, the vacuole of cells, spermatozoa and numerous other first observations. His discoveries revolutionized the contemporary understanding of life. This was the start of microbiology.

The smallest structures that can be resolved using conventional light microscopy is approximately 200 nm. This limit is due to the diffraction limit (Abbe's famous formula $d = \lambda/(2n\sin\alpha)$), a direct consequence of the finite wavelength of visible light and the finite opening angle of a lens. It was the invention of the electron microscope in the 1930's that allowed scientist to carefully study the internal structure of the cell at a much finer scale. Electron microscopes are able to achieve a much higher resolution because the wavelength of an accelerated electron is orders of magnitude smaller than that of light.

With the help of high resolution electron microscopes — together with complementary biochemical studies — we now know what the fundamental building blocks of life are and how they look like. The cell's main machinery consists of proteins and protein-complexes. Every cell contains a huge number of proteins, some estimate at least 200,000. Proteins are built from chains of amino acids and different proteins have different sequences of these amino acids. The number of

unique proteins that exist in nature is unknown; currently 10^6 protein sequences are known. What is remarkable, however, is that this vast number of proteins is built from only 20 different amino acids.

Made from only 20 different amino acids, proteins display an enormous diversity in function. This is possible because the long chains of amino acids are folded into specific three-dimensional structures. This structure dictates for a large part the functionality of the protein. Some protein-complexes can change their conformation by using some of the cell's available energy. These different conformational states are essential for the protein's functionality in a highly dynamic environment. Therefore, to study proteins we need a high-resolution 3D imaging technique in combination with a method of arresting proteins in their momentary state of function [1]. Ideally, we would like to arrest the structures in a close-to-native environment, i.e. inside a cell.

1.2. *Cryo-electron tomography*

Cryo-electron tomography (CET) is an imaging modality with the unique potential to visualize proteins, protein-complexes and other molecular assemblies in a close-to-native environment at a high resolution in three dimensions. These properties allow CET to bridge-the-gap of knowledge between the molecular structures determined using X-ray crystallography, nuclear magnetic resonance or single-particle cryo-electron microscopy and the cellular context provided by light microscopy or conventional electron microscopy. Fig. 1.1 schematically depicts the workflow of the *experiments* to record images and the subsequent *reconstruction* leading to a 3D structure (also shown is the *simulation* workflow which is mentioned below). In essence, CET is a combination of three separate techniques: cryo-fixation, transmission electron microscopy and tomographic reconstruction.

1.2.1. Cryo-fixation

Electrons not only scatter against our specimen, but also against air molecules. To make imaging with electrons possible, the path of the electron beam must be kept in a high vacuum. Hence, the specimen under investigation must be able to withstand this vacuum. The traditional way to prepare specimens that are suitable for this imaging environment involves chemical fixation and heavy metal staining. For high-resolution studies this protocol is inadequate since one effectively images the stain instead of the original structure.

Cryo-fixation, on the other hand, can potentially preserve the native environment of the macromolecules with minimal changes. The basis of cryo-fixation is

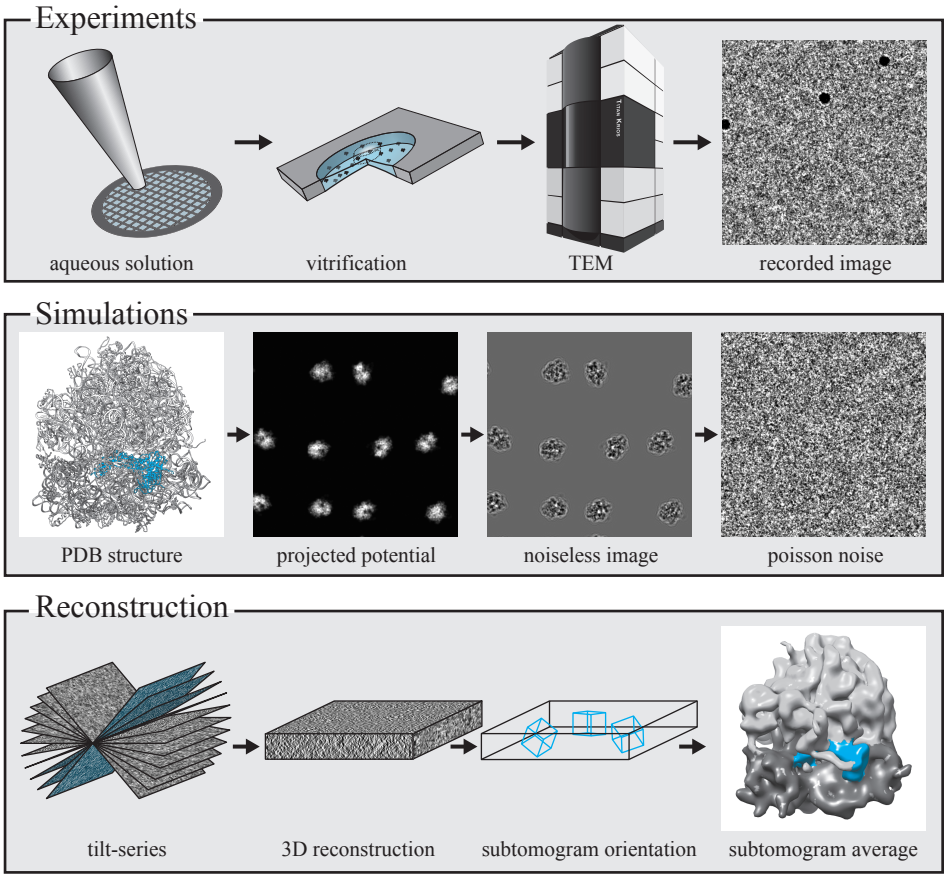


Figure 1.1. Overview of the project workflow ranging from experiments to simulations and reconstruction. *(Experiments)* Specimen preparation for CET can be performed by applying an aqueous solution containing a purified sample such as proteins or cells to an EM grid which is covered by a carbon film which in turn contains a vast number of small holes. Vitrification of the specimen results in a thin layer of amorphous ice (which contains the sample) suspended in the holes of the carbon film. The specimen is then transferred under cryogenic conditions to a TEM where a tilt-series can be acquired using automated data collection. *(Simulations)* Starting from the atomic structure of a known structure deposited in the protein RCSB Protein Data Bank (PDB), a potential map is computed. Based on the electron beam characteristics and this potential map an exit-wave is simulated; possibly using the projection assumption, weak-phase object or thick-phase grating approximation. This wave function is then modified by the CTF followed by an accurate camera model resulting in a quantitative simulation of a CET image. *(Reconstruction)* First the defocus of each image in a tilt-series is estimated after which CTF correction can be performed. From the CTF corrected tilt-series a 3D volume is computed using tomographic reconstruction (typically filtered back projection for CET). Additionally, subtomogram averaging can be performed by first detecting identical copies of a specific structure. Prior to averaging, the relative orientation and location of each subtomogram can be estimated and iteratively refined.

that when water is frozen very rapidly, the formation of ice crystals is prevented, resulting in amorphous ice. Formation of crystalline ice must be prevented since ice crystal growth not only destroys the structural integrity of the specimen, it will also diffract the electron beam. Rapid freezing can be achieved by plunging a specimen into liquid ethane[1] or by high-pressure freezing where the specimen is simultaneously cooled and pressurized in a metal container. Conceptually, cryo-fixation is much simpler than chemical fixation; practically, cryo-fixation is very challenging. After cryo-fixation, all specimen handling should be done at liquid nitrogen temperatures. Furthermore, imaging these kind of specimens demands a more delicate imaging strategy compared to chemical fixation because the specimens are much more sensitive to damage caused by the electron beam. The advantages of cryo-fixation are superior preservation of the ultrastructure and almost instantaneous fixation, i.e. within milliseconds. The specimen can then be imaged without any additional staining.

1.2.2. Transmission electron microscopy

Transmission electron microscopy (TEM) is in some ways similar to regular bright field light microscopy. For both techniques the specimen is illuminated from one side and the light or electron-wave that is transmitted through the specimen is magnified to form an image. Modern bright field light microscopes are limited by both the wavelength of visible light (400 to 700 nm) and the numerical aperture. In TEM, electrons are typically accelerated to an energy between 80 and 300 keV resulting in a wavelength between 4 and 2 pm. However, due to the small numerical aperture of (state-of-the-art) electron microscopes the resolving power is 'only' 50 pm. This resolution is, nevertheless, more than sufficient to resolve individual atoms. As we shall see below, for imaging biological specimens the resolution is not limited by the diffraction limit but rather by the damage that these accelerated electrons inflict to the specimen.

1.2.3. Tomographic reconstruction

Non-invasive, three-dimensional imaging techniques have revolutionized healthcare in the past century. The most familiar modality, computed tomography (CT) uses a series of X-ray projections taken at different projection angles to make a three-dimensional reconstruction of the body. Already in 1917, the mathematical principles required to make this reconstruction were developed by Radon [2]. It was only after the digital revolution, however, that it became practically possible

[1] Since nitrogen gas is a good insulator, liquid nitrogen is unsuitable due to the Leidenfrost effect.

since tomographic reconstruction in essence is a computational method. For CET another technical advance was essential. Since unstained biological specimens are sensitive to radiation damage, projections should be acquired using a minimal electron dose. This was made possible by the development of automated data collection using digital cameras [3]. In CET a tilt-series is acquired by tilting the specimen from approximately $-70°$ up to $+70°$. Higher tilt-angles are typically impossible since the effective specimen thickness increases dramatically beyond the mean free path of the electron at high tilt-angles, and also the specimen holder can block the electron beam. This limited tilt range results in so called missing wedge artifacts after tomographic reconstruction.

1.3. *Phase contrast*

When electrons scatter against the specimen, two types of contrast are generated: amplitude contrast and phase contrast. For negatively stained specimens, amplitude contrast is the primary contrast mechanism due to the large difference in atomic number between the heavy metal stains and the aqueous biological material. Without any staining, the low atomic number of the biological material does not produce strong enough inelastic scattering to yield useful amplitude contrast. The elastically scattered electrons, although they have the same intensity as the unscattered electrons, do have a different phase. In order to measure this phase-shift, phase information needs to be transformed into intensity contrast. The most used technique in CET to convert this phase information into measurable intensity contrast, is by intentionally recording images with underfocus. This defocussing allows for contrast generating interference of the undiffracted beam with the beam that is phase-shifted by the specimen. Other methods to measure phase contrast include phase-plates (which is still in a research and development phase) and electron holography (requiring a high dose and therefore primarily useful for material science).

1.3.1. CTF correction

Defocussing of the image can be modeled as an extra propagation of the exit-wave formed by the electrons exiting the specimen. Due to the small scattering angles and very short wavelengths, this propagation can be modeled by Fresnel diffraction. This accurately models the phase of a spherical wave emerging from the specimen, where the spherical wave is approximated by a quadratic function.

The amount of phase contrast that is generated by a biological specimen is relatively weak. The phase difference between the diffracted and undiffracted

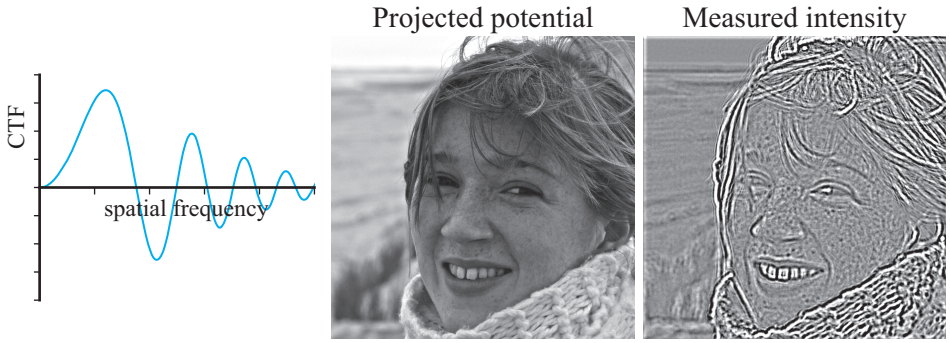


Figure 1.2. Example of a typical contrast transfer function (CTF) and how this affects the recorded image intensity. The projected potential of the specimen is proportional to the accumulated phase-shift of the electron wave when it exits the specimen. The phase of the electron wave, however, cannot be measured directly. Intensity contrast appears when a defocus is applied, modeled by the CTF. The loss of low frequency contrast and the ringing pattern around sharp edges is clearly visible.

beam rarely exceeds $\pi/2$. In this regime, the weak-phase object approximation (WPOA) can be applied. By combining the WPOA and Fresnel diffraction we can describe the recorded intensity in the image plane using the contrast transfer function (CTF). The CTF can not only model the defocussing, but also other aberrations such as astigmatism and spherical aberration. However, due to the relatively low resolution of CET (in comparison with material science applications) only the lower order aberrations need to be considered.

Fig. 1.2 shows a typical CTF for CET and a test image to show how the specimen's projected potential is translated to measurable intensity contrast on the detector. The CTF is an oscillating function of spatial frequency. Therefore, information at certain spatial frequencies is imaged with inverted contrast. As a consequence, unambiguous interpretation of image contrast is hampered. These effects, however, only occur at spatial frequencies beyond the first zero-crossing of the CTF. The location of the first zero-crossing can be set to any value by changing the applied defocus. Thus, it is possible to choose the defocus such that CTF correction is not required. However, this is only a suitable approach for medium to low resolution phase contrast imaging. This is caused by the strong attenuation of low spatial frequencies by the CTF when the defocus is small. As a consequence, the applied defocus is usually larger than 2 µm. For smaller defoci it becomes difficult to localize the structure of interest due to the lack of contrast at low spatial frequencies. The first zero-crossing for a defocus of 2 µm is at 1/2.0 nm for

300 keV electrons. Therefore, to achieve nanometer to sub-nanometer resolution it is necessary to do CTF correction.

The simplest form of CTF correction, so called phase flipping, inverts the contrast of those spatial frequencies which were imaged with inverted contrast. More elaborate correction methods also correct for the damping of certain spatial frequencies, so called amplitude correction. The difficulty with amplitude correction is the trade-off between amplifying attenuated signal and amplifying noise. A Wiener filter should produce an optimal trade-off but requires accurate knowledge of the signal-to-noise ratio (SNR). Since the SNR for CET is typically very low and difficult to estimate, phase flipping is the most used CTF correction type.

Since the CTF depends on the amount of defocus, the defocus needs to be known to allow proper CTF correction. Even though the defocus is set by the microscope, this requested value is typically not accurate enough for CTF correction. Therefore, the defocus needs to be estimated from the acquired images. The accuracy with which the defocus can be estimated determines, for a large part, how effective CTF correction is.

1.4. *Radiation damage*

Electrons used for imaging have such a high energy that they can inflict severe damage to the specimen. This damage may occur when electrons scatter inelastically, i.e. when energy is deposited. As a consequence, continuous imaging the specimen eventually destroys the structures of interest. Therefore, we must limit the total amount of electrons that are used for finding the target structures and subsequent imaging. Typically, the total electron dose used for tomography ranges from 25 to 100 $e^-/Å^2$. After exposing the specimen to such an electron dose the structure is degraded to such an extent that details are lost. The SNR, however, increases with increasing electron dose. This leads to conflicting requirements: the dose must be as high as possible to increase the SNR and as low as possible to minimize radiation damage.

For other non-biological materials, such as graphene and carbon nanotubes (CNT), radiation damage poses less of a restriction. If the acceleration voltage of the microscope is lowered to energies below the bond energy between the atoms, knock-on damage no longer occurs. Imaging at lower acceleration voltages does, however, require sophisticated aberration corrected microscopes. Nevertheless, tomography offers the exciting possibility to study the complex 3D structures formed by graphene and CNTs.

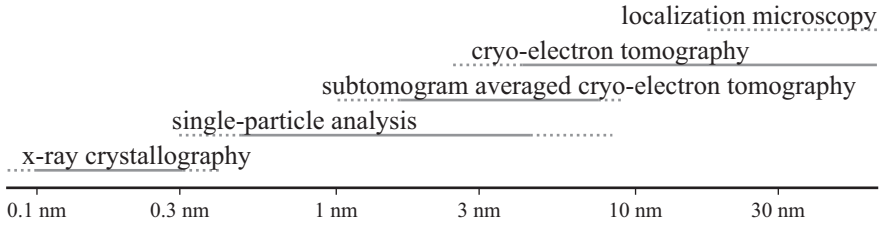


Figure 1.3. Approximate resolution ranges covered by different microscopy modalities.

1.5. *Subtomogram averaging*

One method to increase the SNR while keeping the electron dose as low as possible, is subtomogram averaging. If a specimen contains many copies of an identical structure, the reconstructed subvolumes of these structures can be aligned and averaged to increase the SNR of the final reconstruction. Prerequisites for subtomogram averaging are that the different copies can be identified in the reconstructed volume and aligned. Subtomogram averaging not only increases the SNR of the final reconstruction, it can also removes missing wedge artifacts since this missing wedge can be oriented differently for different particles.

The challenging aspect of subtomogram averaging is that the method is prone to template bias and over alignment. Subtomogram averaging typically uses a cross-correlation based search using an initial template or the average of a previous alignment iteration. It is therefore essential to low-pass filter search templates such that no artificial structures are introduced.

1.6. *Complementary techniques*

CET is unique in its possibility to bridge-the-gap between acquiring high resolution molecular structures on one hand and their conformational state and function in cellular context on the other hand. The full potential of CET, therefore, can only be unlocked when combined with information achieved with other techniques. Fig. 1.3 presents an overview of the approximate resolution ranges of CET and complementary methods.

1.6.1. Single particle analysis

Even though subtomogram averaging is a powerful technique, more than three quarters of the electron microscopy density maps deposited in the EMDataBank [4] have been achieved using single particle analysis (SPA). When it is possible to prepare the biological specimen *in vitro* and in such a way that it consists of iso-

lated, non-overlapping macromolecules, SPA typically outperforms CET in terms of resolution. Just as CET, SPA is a cryo-electron microscopy technique, i.e. cryo-fixation is used to immobilize the specimen. The term single particle analysis is, however, rather deceptive since it relies on averaging a great number of identical particles. Similar to subtomogram averaging, identical copies of the same structure are imaged and subsequently reconstructed into a single 3D structure. The difference with subtomogram averaging is, however, that the exact location and orientation of the particles is estimated already from their individual 2D projections. This elevates the need for acquiring a tilt-series and allows the individual projections to be imaged with a much higher dose (and thus SNR) since the dose does not need to be distributed over an entire tilt-series. The difficulty now is to estimate the orientation (and handedness) of the particles in the projections. This involves an iterative reconstruction and the generation of class-averages (an average of 2D projections of particles with similar orientation). Using SPA it is possible to achieve near-atomic resolution (0.3 – 0.5 nm) of biomolecules, see Fig. 1.3. The SPA approach is, however, not applicable when projections of particles are superimposed with other structures present in the specimen. Superimposing structures prevent the accurate determination of location and orientation of the desired structure. This limitation renders SPA impractical for usage *in situ*, i.e. within cells, viruses, vesicles and even complex *in vitro* systems [5].

1.6.2. X-ray crystallography

For determining structure at truly atomic resolution, X-ray crystallography has been the most successful method; more than 80% of the biological macromolecules deposited in the RCSB Protein Data Bank (PDB) were determined using X-ray crystallography. For this method to work, it is necessary to purify the protein after which they need to crystallize. The crystalline specimen is irradiated by an intense monochromatic X-ray beam after which the diffraction pattern is recorded. Using several intricate procedures the phases of all the spots comprising the diffraction pattern are determined after which a 3D structure can be build. Even though the resolution obtained with X-ray crystallography is often superior to EM methods, due to the restriction that proteins need to crystallize *in vitro*, the cellular context can only be studied using complementary methods.

1.6.3. Light microscopy & conventional electron tomography

Resolutions ranging from 1 mm down to ~ 10 nm can effectively be studied using a combination of light and conventional transmission electron microscopy. Light microscopy has the advantages of fluorescent labeling and the possibility to image

in vivo. This makes light microscopy the ideal method for functional studies. Conventional electron tomography preparation techniques use chemical fixation and negative staining. This limits the resolution to ~ 10 nm but results is supreme contrast at lower resolutions. Advances in superresolution microscopy extend the resolution of fluorescence microscopy of single molecules down to tens of nanometers, thereby closing the gap to conventional electron microscopy. See Fig. 1.3 for an overview of the different resolution ranges.

1.7. *Project outline*

The research in this thesis is part of the FOM industrial partnership program with FEI Company. This project was a collaboration between Delft University of Technology (TU Delft), Leiden University Medical Center (LUMC) and FEI Company. The goal of this project was to substantially improve the 3D resolution in electron tomography of frozen hydrated biological specimen after tomographic reconstruction. To achieve this, we made an integral analysis of the entire imaging pipeline. A study of image formation, experiments and reconstruction led to a TEM simulator for biomolecules and new image processing and image analysis methods for experimental data. Fig. 1.1 presents an overview of the entire project comprising *experiments*, *simulations* and *reconstruction*. Within this project there were two sub-projects of which this thesis represents one. This thesis is primarily concerned with CTF correction in the 3D reconstruction from tilt-series (experiments or simulations). The second sub-project revolved around the development and experimental validation of an accurate forward model of the image formation process and estimation of the required microscope parameters from experiments.

1.7.1. Thesis challenges

Today, the resolution in phase-contrast CET is not only limited by the poor SNR, but also by the CTF of the microscope. Even though CTF correction for SPA is well established, for CET it has not been adopted in daily practice. Reasons for this are twofold: first, due to the low SNR of CET the resolution is typically much lower than SPA, thereby reducing the necessity of CTF correction, and second, the computational complexity of CTF correction techniques for tomography makes them less practical. With the advent of subtomogram averaging the achievable resolution in CET has been substantially improved, alleviating the first obstacle. This thesis tackles the second obstacle by introducing 100 times faster algorithms for forward modeling the CTF and for CTF correction. Furthermore, we present a thorough theoretical analysis of different CTF correction methods and study the effects using simulations. We present an extended acquisition scheme which

makes precise and accurate defocus estimation — which is required for proper CTF correction — for low-dose CET possible. Finally, we quantify the influence of CTF correction and defocus estimation on experimental data and simulations.

1.7.2. Thesis outline

In **Chapter 2** we summarize the theoretical background of regular (untilted) CTF, tilted CTF and full 3D CTF forward models. We introduce a new algorithm, based on the nonuniform fast Fourier transform (NUFFT), that speeds up the simulation of projections with a spatially varying CTF. This speed-up allows spatially varying CTF models to be used in daily practice. Using simulations we study what the influence of these CTF models is on the projections and how we can predict this difference theoretically. Furthermore, we make a first attempt in global (i.e. to process the entire image at once) CTF correction methods using a matrix multiplication. These new tools also allow us to study the influence of defocus error and whether defocus-series might be a useful addition to tomography.

Chapter 3 introduces new, fast, spatially varying CTF correction methods. These CTF correction methods rely on the NUFFT algorithm used to simulate the influence of the spatially varying CTF on the recorded image in chapter 2. Where the simulations, however, use the NUFFT to sample the Fourier domain at nonuniform locations, the CTF correction computes an inverse Fourier transform from nonuniformly sampled Fourier domain data. We show that this reduces the required computation time for a global CTF correction method on a tilt-series (image size 4096^2) from months to hours.

In **Chapter 4** we take a closer look at two approximations commonly used to model image formation: the projection assumption (PA) and the weak-phase object approximation (WPOA). We investigate the limitations of both approximations separately and derive their respective conditions of applicability. Then we combine these two approximations in different ways to find four models: PA, PA+WPOA, WPOA and the thick-phase grating approximations (TPGA). We suggest ranges of applicability for these four models using exit wave simulations. This leads to a practical guideline for when to use which image formation model depending on the desired resolution (spatial frequencies), specimen thickness, and strength of the interaction potential of the macromolecular complex.

Chapter 5 culminates the entire project in quantifying the influence of the developed CTF correction and defocus estimation methods using *experiments*, *simulations* and *reconstruction*. We combine the efforts of both subprojects, namely: detector characterization, defocus and astigmatism estimation, the developed TEM simulator (InSilicoTEM) making use of the TPGA, and tilted CTF

correction. We collected tilt-series of ribosomes on a Titan microscope (FEI Company) at NeCEN[2] using a novel extended acquisition scheme which allows precise determination and correction of the defocus for each macromolecule in the tomogram. Subsequently, we applied subtomogram averaging to increase the SNR. Using both simulations and experiments we quantify the influence of defocus estimation, CTF correction, tilt-series alignment and subtomogram alignment.

Finally, **Chapter 6** presents an outlook and recommendations for future work.

References

[1] O. Medalia, I. Weber, A. S. Frangakis, D. Nicastro, G. Gerisch, and W. Baumeister, *Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography,* Science **298**, 1209 (2002).

[2] J. Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkiten,* Verh. Sächs. Akad. Wiss (1917).

[3] A. J. Koster, R. Grimm, D. Typke, R. Hegerl, A. Stoschek, J. Walz, and W. Baumeister, *Perspectives of molecular and cellular electron tomography,* Journal of Structural Biology **120**, 276 (1997).

[4] C. L. Lawson, M. L. Baker, C. Best, C. Bi, M. Dougherty, P. Feng, G. van Ginkel, B. Devkota, I. Lagerstedt, S. J. Ludtke, R. H. Newman, T. J. Oldfield, I. Rees, G. Sahni, R. Sala, S. Velankar, J. Warren, J. D. Westbrook, K. Henrick, G. J. Kleywegt, H. M. Berman, and W. Chiu, *EMDataBank.org: unified data resource for CryoEM,* Nucleic Acids Research **39**, D456 (2011).

[5] F. K. M. Schur, W. Hagen, A. de Marco, and J. A. G. Briggs, *Determination of protein structure at 8.5Å resolution using cryo-electron tomography and sub-tomogram averaging,* Journal of Structural Biology **184**, 394 (2013).

[2] Netherlands Centre for Electron Nanoscopy

2

A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM

Equipped with our five senses — along with telescopes and microscopes and mass spectrometers and seismographs and magnetometers and particle accelerators and detectors sensitive to the entire electromagnetic spectrum — we explore the universe around us and call the adventure science.

EDWIN HUBBLE

Lenard M. Voortman
Sjoerd Stallinga
Remco H. M. Schoenmakers
Lucas J. van Vliet
Bernd Rieger

Ultramicroscopy 111 (2011), pp. 1029–1036

2

Abstract

Today, the resolution in phase contrast cryo-electron tomography is for a significant part limited by the contrast transfer function (CTF) of the microscope. The CTF is a function of defocus and thus varies spatially as a result of the tilting of the specimen and the finite specimen thickness. Models that include spatial dependencies have not been adopted in daily practice because of their high computational complexity. Here we present an algorithm which reduces the processing time for computing the 'tilted' CTF by more than a factor 100. Our implementation of the full three-dimensional CTF has a processing time on the order of a Radon transform of a full tilt-series. We derive and validate an expression for the damping envelope function describing the loss of resolution due to specimen thickness. Using simulations we quantify the effects of specimen thickness on the accuracy of various forward models. We study the influence of spatially varying CTF correction and subsequent tomographic reconstruction by simulation and present a new approach for space-variant phase-flipping. We show that our CTF correction strategies are successful in increasing the resolution after tomographic reconstruction.

2.1. *Introduction*

Cryo-electron tomography (CET) is an important tool for studying the three-dimensional (3D) structure of biological specimens in their (close to) native state [1]. In CET the 3D scattering potential is reconstructed from many projections at different tilt-angles. The reconstruction of a 3D image from the two-dimensional (2D) projections is typically done using filtered back-projection (FBP) or the simultaneous iterative reconstruction technique (SIRT). The specimens consist mostly of light atoms such as C, N and H, and therefore are very weak electron scatterers. As a result amplitude contrast is not very strong and images are recorded in a phase contrast mode. This phase contrast is, however, only generated if the specimen is sufficiently defocused to allow interference of the unscattered wave with the scattered wave. The phase-shifts due to elastic scattering are relatively small and are described using the well established weak-phase approximation [2]. Here the image formation is dictated by the contrast transfer function (CTF). The CTF is an oscillating function of the spatial frequency; therefore contrast for certain spatial frequency ranges is inverted. To maximize the transfer of certain spatial frequencies (corresponding to sizes of ~ 4 to 10 nm in the specimen) without contrast inversions, typically several μm underfocus is used in practice. Transmission of higher spatial frequencies is suppressed by the apparent finite source-size of the electron gun. This cut-off point can only be shifted to a higher spatial frequency by using smaller defoci. This comes at the expense of possible contrast inversions in the image. In the presence of these contrast inversions, image interpretation is no longer intuitive. To remove possible contrast inversion, various authors apply a low-pass filter to the projections with its cut-off frequency at the first zero-crossing of the CTF [3, 4].

CTF correction is already well established in the field of single-particle analysis [5]. It usually comprises phase-flipping of the spatial frequencies which were imaged with inverted contrast. Sometimes the magnitude of the CTF is also corrected. For each detected particle the defocus is assumed to be constant, which makes CTF correction spatially invariant. When imaging tilted specimens, as in tomography, the assumption that the defocus is constant over the field of view no longer holds. Recently there has been an increased interest to characterize and correct for the effects of this spatially varying CTF [6–11]. As the thickness of the specimen can easily be on the order of 100 nm or more there is also a relevant defocus gradient in the axial direction to be considered [12]. There have also been efforts to correct for these effects [13–16]. These methods for forward modelling and for correcting the spatially varying CTF for tilted and/or thick specimens have not been widely adopted for tomographic reconstruction. Partially because of

their complexity, but more importantly due to the burden they pose on processing times (up to several weeks). We present a hundredfold reduction of the processing time for the forward modelling of tilted specimens. This makes our spatially varying method for computing CTFs applicable in practice.

The structure of this article is as follows. In Section 2.2 we present an overview of the different approximations to the CTF: untilted thin, tilted thin and tilted thick. Next, in Section 2.3 we propose novel algorithms to reduce the computational complexity of modelling a spatially varying CTF. In Section 2.4 we use simulations to show the limitations and advantages of the different CTF models for forward modelling and finally in Section 2.5 we apply them to CTF correction.

2.2. *Contrast Transfer Function for 3D specimens*

In order to describe the CTF, we introduce the following terminology. The Fourier transform (FT) of the 3D scattering potential $V(\mathbf{x}, z)$ is defined as

$$\hat{V}(\mathbf{q}, q_z) = \int V(\mathbf{x}, z) e^{-2\pi i(\mathbf{x}\cdot\mathbf{q} + z q_z)} d\mathbf{x} dz,$$

where $\mathbf{x} = (x, y)$ and $\mathbf{q} = (q_x, q_y)$. Equivalently, we define the 2D FT as $\hat{f}(\mathbf{q}) = \int f(\mathbf{x}) e^{-2\pi i \mathbf{x}\cdot\mathbf{q}} d\mathbf{x}$.

We choose to fix the specimen coordinates $V(\tilde{\mathbf{x}}, \tilde{z})$ and rotate the microscope coordinate system $V_\alpha(\mathbf{x}, z)$. The rotated coordinate system is defined by two variables: the tilt-angle α and the orientation of the tilt-axis (azimuth) β. Since a rotation of the specimen in real space corresponds to the same rotation in Fourier space we define[1]

$$\begin{aligned}\hat{V}_\alpha(\mathbf{q}, q_z) = \hat{V}\big(&q_x \cos\beta\cos\alpha + q_y \sin\beta\cos\alpha + q_z \sin\alpha,\\ &- q_x \sin\beta + q_y \cos\beta,\\ &- q_x \cos\beta\sin\alpha - q_y \sin\beta\sin\alpha + q_z \cos\alpha\big).\end{aligned} \tag{2.1}$$

Hence, for $\beta = 0$ the tilt-axis is the y-axis, which we will assume in the remainder of this article.

Since scattering in biological specimens is usually very weak, it is common to use the weak-phase approximation [2]. We will use this approximation throughout this article and thus the FT of the recorded intensity is expressed as [17]

$$\hat{I}(\mathbf{q}) = 2\sin\left(\chi(q)\right)\hat{V}_\alpha(\mathbf{q}, 0), \tag{2.2}$$

[1] We implicitly define $\hat{V}(\tilde{q_x}, \tilde{q_y}, \tilde{q_z}) = \hat{V}(\tilde{\mathbf{q}}, \tilde{q_z})$.

with $\chi(q) = \frac{2\pi}{\lambda}\left(\frac{1}{4}C_s\lambda^4 q^4 + \frac{1}{2}\Delta f\lambda^2 q^2\right)$ the aberration function, C_s the spherical aberration, Δf the defocus at $z = 0$, λ the electron wavelength and $q = \|\mathbf{q}\|$. We use the convention that underfocus implies $\Delta f < 0$. The projected scattering potential $\hat{V}_\alpha(\mathbf{q}, 0)$ is equivalent (using the projection slice theorem [18]) to the FT of $\int V_\alpha(\mathbf{x}, z)dz$.

We ignore amplitude contrast in Eq. (2.2) to keep the following derivations short and to improve the readability. In practice, amplitude contrast can be incorporated simply as an additional phase-shift [2].

Note that Eq. (2.2) is an approximation to the full contrast transfer function as the defocus is assumed to be constant for the whole specimen. We will refer to Eq. (2.2) as CTF_0.

2.2.1. Tilted, thin specimens

To allow for tilted geometries Philippsen *et al.* [6] derived an extension to the CTF_0. We will refer to this function (which they called the tilted contrast imaging function) as the tilted CTF (TCTF). The FT of the recorded image intensity for tilted, thin specimens is

$$\begin{aligned}\hat{I}(\mathbf{q}) = {} & ie^{-i\chi(q)}\hat{V}_\alpha\left(\mathbf{q} - \frac{1}{2}\lambda q^2\boldsymbol{\beta}\tan\alpha, 0\right) \\ & - ie^{i\chi(q)}\hat{V}_\alpha\left(\mathbf{q} + \frac{1}{2}\lambda q^2\boldsymbol{\beta}\tan\alpha, 0\right),\end{aligned} \tag{2.3}$$

where $\boldsymbol{\beta}$ is a unit-vector perpendicular to the tilt-axis with $\boldsymbol{\beta} = (\cos\beta, \sin\beta)$. For small tilt-angles, i.e. $\alpha \to 0$, Eq. (2.3) gradually reduces to Eq. (2.2).

2.2.2. Thick specimens

The validity of Eq. (2.2) is not only limited to untilted specimens, it is also limited to thin specimens. In the weak-phase regime all scattering events can be treated as weak and independent. Therefore all points in the specimen contribute linearly to the recorded intensity. Incorporating the effects of the thickness of the specimen is simply a matter of integrating along the z direction of the scattering potential times the appropriate CTF

$$\hat{I}(\mathbf{q}) = \int 2\sin\left(\chi(q) - \pi\lambda z q^2\right)\mathscr{F}_{\mathbf{x}}\left[V_\alpha(\mathbf{x}, z)\right]dz. \tag{2.4}$$

Wan *et al.* [12] showed that the evaluation of Eq. (2.4) can be simplified by taking the FT in the z direction

$$\mathscr{F}_z\left[2\sin\left(\chi\left(q\right)-\pi\lambda z q^2\right)\right] = ie^{-i\chi(q)}\delta\left(q_z-\frac{1}{2}\lambda q^2\right) - ie^{i\chi(q)}\delta\left(q_z+\frac{1}{2}\lambda q^2\right). \tag{2.5}$$

Using this equation, the multiplication in Eq. (2.4) of the scattering potential with the CTF can be written as a convolution in q_z. The integration over z is equal to evaluating for $q_z = 0$. Combining these steps leads to [12]

$$\hat{I}\left(\mathbf{q}\right) = ie^{-i\chi(q)}\hat{V}_\alpha\left(\mathbf{q},-\frac{1}{2}\lambda q^2\right) - ie^{i\chi(q)}\hat{V}_\alpha\left(\mathbf{q},\frac{1}{2}\lambda q^2\right), \tag{2.6}$$

which we will refer to as the three-dimensional CTF (3DCTF). We show in the following section that Eq. (2.6) reduces to Eq. (2.3) for infinitely thin specimens. Here we also derive the approximation error of Eq. (2.3) with respect to Eq. (2.6).

2.2.3. Thickness induced damping envelope

One very common assumption in CET is that the 3D scattering potential can be approximated by a 2D projected scattering potential. The TCTF (Eq. (2.3)) is an approximation to the 3DCTF (Eq. (2.6)) for thin specimens. We will quantify the correctness of this approximation as a function of the specimen thickness. In order to show the relation between these two equations, we first ideally describe a specimen with limited thickness as

$$V(\tilde{\mathbf{x}},\tilde{z}) = V_\infty(\tilde{\mathbf{x}},\tilde{z})\frac{1}{t}\operatorname{rect}\left(\frac{\tilde{z}}{t}\right), \tag{2.7}$$

where $V_\infty(\tilde{\mathbf{x}},\tilde{z})$ is a theoretical specimen with infinite thickness, t the thickness of the real specimen $V(\tilde{\mathbf{x}},\tilde{z})$ and $\frac{1}{t}\operatorname{rect}\left(\frac{\tilde{z}}{t}\right)$ a normalized block function of unity area. A multiplication with the block function in Eq. (2.7) is equivalent to

$$\hat{V}(\tilde{\mathbf{q}},\tilde{q_z}) = \hat{V}_\infty(\tilde{\mathbf{q}},\tilde{q_z}) \star \operatorname{sinc}\left(t\tilde{q_z}\right), \tag{2.8}$$

where $\star$ is the convolution operator in $\tilde{q_z}$ and $\operatorname{sinc}(t\tilde{q_z})$ is the normalized sinc function. For infinitely thin specimens, $\hat{V}(\tilde{\mathbf{q}},\tilde{q_z})$ is independent of $\tilde{q_z}$.

Revisiting the TCTF (Eq. (2.3)) and the 3DCTF (Eq. (2.6)) we see that the scattering potential $\hat{V}_\alpha$ is sampled, in the microscope coordinates, at $(\mathbf{q}\mp\frac{1}{2}\lambda q^2\boldsymbol{\beta}\tan\alpha,0)$ and $(\mathbf{q},\mp\frac{1}{2}\lambda q^2)$ respectively. For a graphical representation see Fig. 2.1. From

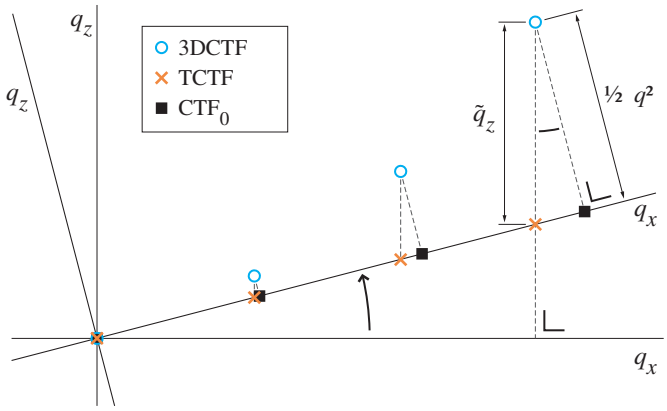


Figure 2.1. A comparison of the CTF_0, TCTF and 3DCTF based on the points sampled in Fourier space ($\beta = 0$). The rotated axes (q_x, q_z) correspond to the coordinate system of the microscope, perpendicular to the tilt-axis. As a result, $\hat{V}_\alpha(\mathbf{q}, 0)$ is the central section which represents the projection data under a certain tilt-angle. The fixed axes $(\tilde{q}_x, \tilde{q}_z)$ correspond to the Fourier space of the specimen: $\hat{V}(\tilde{\mathbf{q}}, \tilde{q}_z)$. The relation between $\hat{V}_\alpha$ and $\hat{V}$ is described by Eq. (2.1). For clarity, only $\hat{V}_\alpha(\mathbf{q}, 0)$ (■), $\hat{V}_\alpha(\mathbf{q} + \frac{1}{2}\lambda q^2 \beta \tan\alpha, 0)$ (×) and $\hat{V}_\alpha(\mathbf{q}, +\frac{1}{2}\lambda q^2)$ (○) for $\mathbf{q} \geq 0$ and $q_y = 0$ are plotted. Here $\alpha < 0$.

this figure (or from Eq. (2.1)) we see that the lateral (in the specimen coordinates) spatial frequency axes $(\tilde{q}_x, \tilde{q}_y)$ coincide. Along the axial spatial frequency however, there is a discrepancy of

$$\Delta\tilde{q}_z = \frac{1}{2}\lambda q^2 (\sin\alpha \tan\alpha + \cos\alpha) = \frac{1}{2}\lambda q^2 \frac{1}{\cos\alpha}.$$

In combination with Eq. (2.8) we conclude that the fraction of the transferred signal is given by

$$E_{t,\alpha}(\mathbf{q}) = \operatorname{sinc}\left(\frac{1}{2}\lambda q^2 \frac{t}{\cos\alpha}\right) \tag{2.9}$$

if instead of sampling $\hat{V}(\tilde{\mathbf{q}}, \tilde{q}_z)$ we sample $\hat{V}(\tilde{\mathbf{q}}, \tilde{q}_z + \Delta\tilde{q}_z)$. This function can be used to estimate the overall loss of resolution when a specimen with thickness t is approximated by an infinitely thin specimen. Neglecting the thickness produces the effect of a damping envelope, similar to the spatial and temporal damping envelopes [17, Section 6.4.2].

From Eq. (2.9) we also see that in the limit of an infinitely thin specimen ($t \to 0$), Eq. (2.3) can be used instead of Eq. (2.6) without loss of transfer.

2.3. *Algorithmic improvements*

The effects of tilting and/or specimen thickness on the CTF have not been widely adopted in reconstruction or forward modeling due to the large burden they pose on processing times. Even though Eq. (2.2) of the CTF_0 and Eq. (2.3) of the TCTF look very similar, their discrete counterparts highlight some challenging differences. The function $\hat{V}_\alpha(\mathbf{q},0)$ in Eq. (2.2) can be discretized by sampling on a regular grid. This results in a discrete Fourier transform (DFT) which can be computed by a standard fast Fourier transform (FFT). In contrast, the samples of $\hat{V}_\alpha$ in Eq. (2.3) do not form a regular grid ($\mathbf{q} \mp \frac{1}{2}\lambda q^2 \boldsymbol{\beta} \tan\alpha$). These sample points do not coincide (in general) with the discretized spatial frequencies of the DFT. As a result the FFT can not be used to compute $\hat{V}_\alpha$.

Let us first consider the general case of spatially varying filtering

$$I(\mathbf{x}) = \int V(\mathbf{s})\, h(\mathbf{x}-\mathbf{s},\mathbf{s})\, d\mathbf{s}, \tag{2.10}$$

where $h(\mathbf{x},\mathbf{s})$ is a spatially varying point-spread function corresponding to the input (object) V at location $\mathbf{s}$. We can relate the point-spread function to a transfer function $T(\mathbf{q},\mathbf{s})$ in the Fourier domain by

$$h(\mathbf{x},\mathbf{s}) = \int T(\mathbf{q},\mathbf{s})\, e^{2\pi i \mathbf{q}\cdot\mathbf{x}}\, d\mathbf{q}. \tag{2.11}$$

For example, $T_{CTF}(\mathbf{q},\mathbf{x}) = 2\sin\left(\chi\left(q\right)\right)$ is simply the transfer function corresponding to the CTF_0. The extension to a spatially varying transfer function T_{TCTF} is given by

$$T_{TCTF}(\mathbf{q},\mathbf{x}) = 2\sin\left(\chi\left(q\right) - \pi\lambda q^2 \mathbf{x}\cdot\boldsymbol{\beta}\tan\alpha\right). \tag{2.12}$$

To derive a general expression for a spatially varying transfer function, let us insert Eq. (2.11) into Eq. (2.10) and change to the Fourier domain $\hat{I}(\mathbf{p})$ ($\mathbf{q}$ is already in use by $T(\mathbf{q},\mathbf{s})$). Now we can first integrate over $\mathbf{x}$ which results in a $\delta(\mathbf{q}-\mathbf{p})$ term. Then, integrate over $\mathbf{q}$ and replace the remaining $\mathbf{p}$ by $\mathbf{q}$. This leads us to the following result

$$\hat{I}(\mathbf{q}) = \int V(\mathbf{x})\, T(\mathbf{q},\mathbf{x})\, e^{-2\pi i \mathbf{q}\cdot\mathbf{x}}\, d\mathbf{x}, \tag{2.13}$$

where $V(\mathbf{x})$ is the input scattering potential, $T(\mathbf{q},\mathbf{x})$ a spatially varying transfer function and $\hat{I}(\mathbf{q})$ the output with $\mathbf{x},\mathbf{q} \in \mathbb{R}^2$.

Evaluation of $\hat{I}(\mathbf{q})$ in the form of Eq. (2.13) can only be implemented for discretized data as a matrix multiplication. In contrast, when $T(\mathbf{q},\mathbf{x})$ does not depend on the location $\mathbf{x}$, Eq. (2.13) simplifies to $\hat{I}(\mathbf{q}) = \hat{V}(\mathbf{q})T(\mathbf{q},0)$. Such spatially invariant equations can clearly be computed much faster. They can be implemented using the FFT algorithm which has a computational complexity of $\mathcal{O}(N \log N)$ [19] compared to $\mathcal{O}(N^2)$ for a regular DFT or matrix multiplication needed to compute Eq. (2.13).

2.3.1. Nonuniform fast Fourier transform

One way of implementing the TCTF (Eq. (2.3)) is to revert to a set of linear equations similar to Eq. (2.13) with Eq. (2.12) as a transfer function (as done by Philippsen *et al.* [6]). We developed an algorithm which results in a substantially faster implementation.

The TCTF of Eq. (2.3) is presented as two separate terms each sampling the FT of the scattering potential at slightly different points. Computing the Fourier tranform of a nonuniform grid of points is, however, not straightforward. We propose to compute this Fourier transform by using the nonuniform fast Fourier transform (NUFFT) (also called gridding) [20, 21]. Here we have used a specific implementation which uses a Gaussian kernel [22].

The fact that the CTF for tilted (TCTF, Eq. (2.3)) and/or thick (3DCTF, Eq. (2.6)) specimens can be computed by nonuniform sampling in the Fourier domain is a result of the particular (oscillating) cos-like dependence of the CTF as a function of defocus and not a general property of spatially varying transfer functions. For example, this does not work for the finite source-size envelope because the FT in the axial direction does not result in a (finite) set of δ-functions.

2.3.2. Taylor expansion of transfer functions

In order to speed up the calculation of those space variant transfer functions for which the FT does not result in a finite set of δ-functions, we propose a Taylor series approximation. The Taylor series of $T(\mathbf{q},\mathbf{x})$ in $\mathbf{x}$ combined with Eq. (2.13) yields

$$\hat{I}(\mathbf{q}) = \sum_{n=0}^{n_{max}} \frac{T^{(n)}(\mathbf{q},0)}{n!} \mathscr{F}_{\mathbf{x}}\left[\mathbf{x}^n V(\mathbf{x})\right] + \mathcal{O}(n_{max}+1), \tag{2.14}$$

with $T^{(n)}(\mathbf{q},0) = \left.\frac{\partial^n T(\mathbf{q},\mathbf{x})}{\partial \mathbf{x}^n}\right|_{\mathbf{x}=0}$ and n_{max} the order at which the Taylor series is truncated. The derivatives of $T(\mathbf{q},\mathbf{x})$ can be derived analytically which makes their implementation straightforward. At the expense of a summation over $n_{max}+1$ terms, Eq. (2.14) can now be implemented using the FFT. Fortunately, when the

spatial dependence is weak, the Taylor series can be truncated after a relatively small number of terms.

In Appendix 2.A we present an implementation for the finite source-size envelope in the form of Eq. (2.14) as well as an illustrative example.

The processing time for Eq. (2.14) depends linearly on the number of terms. The approximation error of the expansion can be estimated from Taylor series theory. To achieve a certain accuracy requires a lower order of the expansion for low tilt-angles than for higher tilt-angles. Therefore we change the order n_{max} adaptively depending on the tilt-angle and some specific microscope parameters.

2.3.3. Speed improvement

In order to evaluate the improvements in processing times for each of the proposed methods, we implemented them to simulate a full tilt-series for an artificial specimen. The processing times of the different methods only depend on the image size N^2. All methods were implemented in Matlab (The MathWorks, USA) and the results are shown in Table 2.1. Without loss of generality the 3D (x, y, z) specimen was reduced to a set of 2D (x, z) slices perpendicular to the tilt-axis. For each 2D slice, a tilt-series of 141 one-dimensional projections with a maximum tilt-angle of $\pm 70°$ was simulated. Processing times were evaluated on a PC equipped with an Intel E8400 dual core 3 GHz processor and 8 GB RAM. The processing times of the one-dimensional projections were scaled to resemble the computation of a full tilt-series where the 2D projections have a square image size.

To calculate the TCTF (Eq. (2.3)) efficiently one takes the 1D NUFFT of each projection in the tilt-series at the appropriate spatial frequencies. In Table 2.1 we show that this approach is more than two orders of magnitude faster for an image size of 4096^2 than by computing Eq. (2.13) through a full matrix multiplication (along a single image dimension). We also noticed that simulating the TCTF using the 1D NUFFT is only four times slower than simulating CTF_0. This means that the TCTF is now within reach for iterative reconstruction algorithms.

Next to the speed advantage, implementing the TCTF using the 1D NUFFT is also intuitive and does not depend on any additional parameters such as strip width or tile size as required by other approaches [7, 9–11] where the image is split in (overlapping) strips or tiles of approximately equal defocus for processing.

To calculate the 3DCTF (Eq. (2.6)) the 2D NUFFT of the scattering potential must be calculated at the appropriate spatial frequencies[2]. The 2D NUFFT can

[2] Since we calculate each (x, z) slice of the 3D scattering potential independently, a 2D NUFFT is sufficient.

Table 2.1. Average run-times for two different methods to calculate the TCTF (Eq. (2.3)) are shown: (first row) the full linear model of Eq. (2.13) using Eq. (2.12), and (second row) the NUFFT method described in the text. Rows 3 and 4 show the average run-times of respectively the build-in Matlab function and the NUFFT for calculating the Radon transform. The last row shows the average run-time of a combination of the Radon transform and the 3DCTF (Eq. (2.6)) using the NUFFT. Notice that some values are in minutes (m) and some in hours (h).

Image size (pixel)	1024^2	2048^2	4096^2
Method	Runtime		
TCTF, Eq. (2.13)	83.1 (m)	12.3 (h)	95.9 (h)
TCTF, 1D NUFFT	2.9 (m)	11.3 (m)	44.4 (m)
Radon, Matlab	14.2 (m)	1.9 (h)	14.9 (h)
Radon, 2D NUFFT	14.0 (m)	1.8 (h)	13.3 (h)
3DCTF, 2D NUFFT	18.3 (m)	2.1 (h)	14.5 (h)

Values represent averages over 100 2D image slices, each projected for 141 tilt-angles. Standard deviations were all below 5%

also be used to compute forward projections [23]. Table 2.1 shows that this approach results in processing times of the same order as the (real-space) Radon transform implemented in Matlab. When calculating the 3DCTF one implicitly projects the (x, z) slice of the scattering potential to a one-dimensional signal. We further notice that simulating the 3DCTF indeed results in processing times of the same order as calculating the forward projection using the 2D NUFFT.

2.4. *Simulating the effects of the complete CTF*

In the previous sections we presented three different forward models and explored their relation.

1. Untilted, thin specimens: CTF_0, Eq. (2.2)
2. Tilted, thin specimens: TCTF, Eq. (2.3)
3. Tilted, thick specimens: 3DCTF, Eq. (2.6)

In this section we will investigate the frequency transfer characteristics of these models using simulations.

2.4.1. White noise test specimen

To compare the different forward models a specimen model is required. To avoid a bias towards a specific spatial frequency range, a synthetic specimen with a flat

spectrum is desirable. Therefore, we choose white noise as an input specimen. To overcome the inherent stochastic nature of such a specimen, the results of 150 realizations are averaged.

We also reduce the 3D (x, y, z) problem to 2D (x, z) slices perpendicular to the tilt-axis. Without loss of generality this reduces the processing time while maintaining the full complexity of the problem. We simulate specimens with a thickness of 50, 100 and 200 nm and a field of view (FOV) of 1 μm. We avoid local-tomography effects [24, 25]. We also do not consider possible loss of resolution and additional noise at detection by plate or CCD camera, given by the modulation transfer function [26].

The CTF_0 and TCTF forward models both act on projected data (which is 1D in the following simulations). The 3DCTF, however, is a function of the 2D (x, z) slice. In order to avoid aliasing effects sufficient zero-padding in the z-direction is required. The specimen is padded with zeros to make the (x, z) slice square. We verified that the amount of zero-padding is sufficient by comparing projections generated using the regular Radon transform with those generated by direct sampling using the NUFFT.

2.4.2. Forward modelling

If a forward model is used in an iterative reconstruction algorithm, its computational complexity is of great importance. A trade-off has to be made between the accuracy (or correctness) of a forward model and its computational complexity.

We simulated projections of the white noise test specimen with the three different forward models mentioned above: CTF_0, TCTF and 3DCTF. To test the accuracy of the two approximations (CTF_0 and TCTF), we compare projections generated by these models with projections generated with the 3DCTF forward model. The microscope parameters were set to: $\Delta f = -1$ μm, $\lambda = 2.5$ pm (HT = 200 kV), $C_s = 2$ mm and the convergence angle $\alpha_i = 0.2$ mrad (see Appendix 2.A). The 1D projections consisted of 4096 pixels with a pixel size of 0.24 nm.

In order to verify the quality of the different forward models, i.e. the resolution, a comparison of the projections is required. There are numerous ways to measure the resolution in a reconstructed volume. In the field of single-particle analysis, Fourier shell correlation (FSC) is used extensively to check the consistency of a reconstruction [27, 28]. There the initial dataset is usually split in two and each half is used to reconstruct the particle after which the two reconstructions are correlated with each other. For our purpose, we define the normalized Fourier correlation (NFC) of two functions f and g as the one-dimensional equivalent of

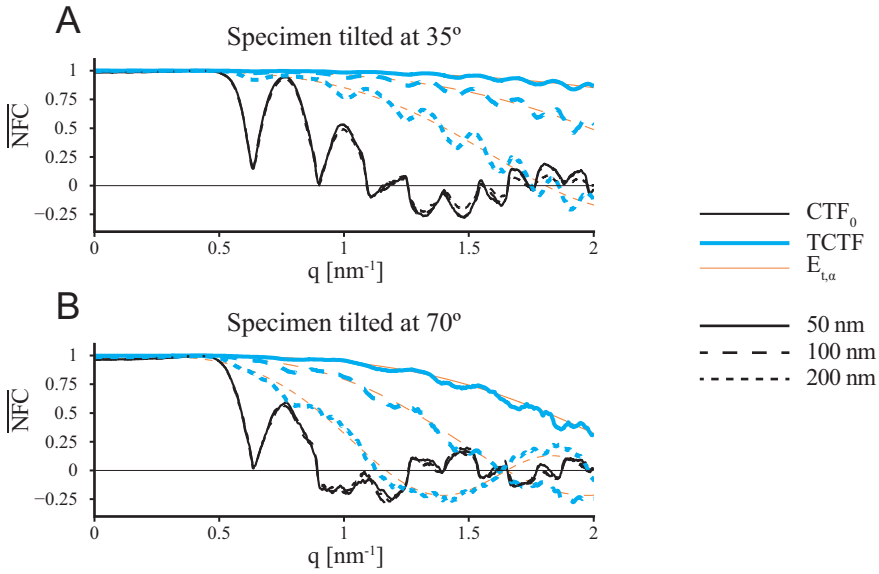


Figure 2.2. Average NFC over 150 realizations between the 3DCTF and the two approximations: CTF_0 in black, TCTF in red. The simulations were carried out for specimens with a thickness of 50, 100 and 200 nm. Overlaid in blue is Eq. (2.9), the analytic prediction of the resolution loss (damping envelope) as a result of neglecting the specimen thickness.

Fourier ring correlation (FRC) and Fourier shell correlation (FSC)

$$\mathrm{NFC}_{f,g}(\mathbf{q}) = \frac{\sum_{\mathbf{q}'\in\mathbf{q}_i} \hat{f}(\mathbf{q}')\hat{g}^*(\mathbf{q}')}{\sqrt{\sum_{\mathbf{q}'\in\mathbf{q}_i} \left\|\hat{f}(\mathbf{q}')\right\|^2 \sum_{\mathbf{q}'\in\mathbf{q}_i} \left\|\hat{g}(\mathbf{q}')\right\|^2}}, \tag{2.15}$$

where $\mathbf{q}_i$ is a small range of spatial frequencies around $\mathbf{q}$ and $-\mathbf{q}$.

In Fig. 2.2 we show the result of the NFC between the two approximate models (CTF_0 and TCTF) and the 3DCTF. From this figure we conclude that the loss of transfer when using the projection assumption can be estimated using Eq. (2.9). From Fig. 2.2 we also conclude that for specimens tilted up to ±70°, with a thickness of 100 nm or less the TCTF is a very good approximation to the 3DCTF for spatial frequencies up to (at least) $1.3\,\mathrm{nm}^{-1}$. The expected loss of resolution (Eq. (2.9)) when using TCTF instead of 3DCTF agrees very well with the simulations. This implies that the accuracy of the TCTF can be approximated accurately with the help of Eq. (2.9). Eq. (2.9) can also be useful to find the tolerable thickness or alternatively to estimate the achievable resolution using the TCTF.

Furthermore, we see in Fig. 2.2 that for CTF_0, the results only slightly vary when changing the thickness. However, it is also clear that the CTF_0 is only a

valid approximations up to approximately 1 nm^{-1} for specimens tilted at ±35° and only up to 0.7 nm^{-1} at ±70°. The clearly visible zero-crossings of the CTF for the CTF_0 result (indicated by the steep dips of the $\overline{NFC}$) are in some way also present in the TCTF model. But in the TCTF model the defocus varies spatially, and as a result the zero-crossings do as well. Hence, parts without contribution (the zero-crossings) are averaged with parts of non-zero contribution.

2.5. *CTF correction*

Next to forward modeling of the CTF, one ultimately wants to correct for these effects in order to reconstruct an unambiguous (without phase-reversals) estimate of the specimen under investigation. We are interested how much the resolution in a reconstructed tomogram can be improved as a result of proper correction for the CTF with spatially varying defocus.

To this end we simulated a tilt-series of 141 projections using the 3DCTF with an underfocus of 1 µm (tilt-angles were uniformly distributed between ±70°). White noise was again used as a specimen model. The thickness of the specimen was 100 nm. The 1D projections consisted of 4096 pixels with a pixel size of 0.24 nm. For more details we refer to Section 2.4 in which the same settings are used.

Poisson noise was added to the projections to simulate realistic low-dose noise conditions. We used a total dose of 150 $e^-/Å^2$ for the entire tilt-series, assuming that the signal of the 1D projections represent a line of square pixels (this corresponds to ~ 6.3 counts per pixel per tilt-angle). The modulation depth of the phase-contrast signal was set to 10%.

2.5.1. Tomographic reconstruction

There exists a variety of algorithms to reconstruct a specimen from its projections. Here we used filtered back projection (FBP) because compared to other algorithms this method depends on the smallest number of parameters. This allows us to do a comparison of reconstructions which depends primarily on the different CTF correction methods and not on how well we succeeded in the fine-tuning of the reconstruction algorithm.

In Fig. 2.3 we show the results for the three different CTF correction methods (discussed below) after tomographic reconstruction. To analyze the difference in the lateral direction ($\mathbf{x}$) we calculate the NFC in the z direction of the reconstruction.

In a simulation the specimen is known. We therefore choose to correlate the different reconstructions to the original specimen. This results in relatively low

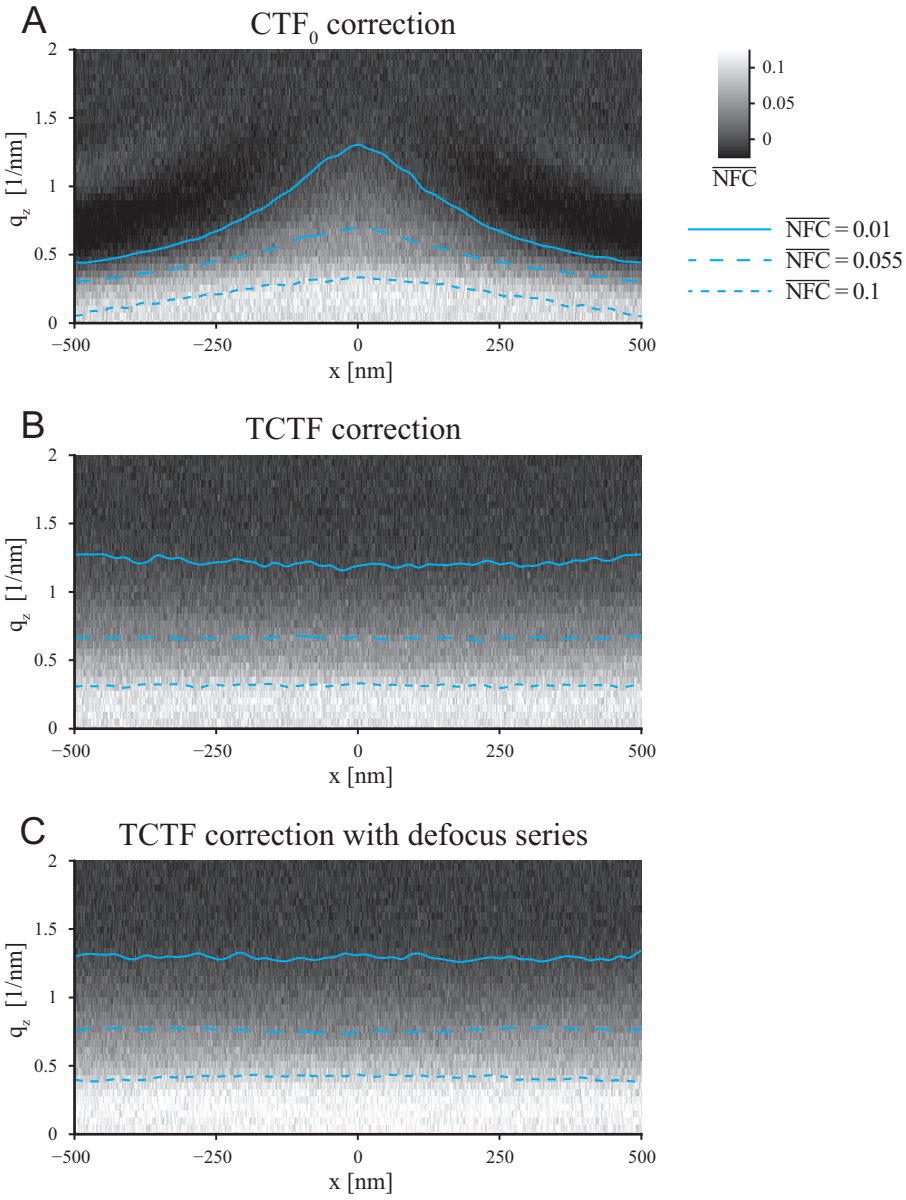


Figure 2.3. Average NFC between an (x, z) slice of the CTF corrected reconstruction and the original (known) specimen over 150 realizations. Simulated projections were CTF corrected without accounting for spatially varying defocus (A), with correction of spatially varying defocus (B), and finally by combining different defocus values and spatially varying defocus correction (C). All CTF correction methods use phase flipping. Isolines are plotted for easier comparison. The average difference of the NFC between (B) and (C), is 16% of the average value of (B).

correlation values. We also correlated the reconstruction to a reconstruction of the undistorted projections. Both suffer from the same 'missing wedge' artifact and therefore correlate much better. Nevertheless, we settled for correlating to the original specimen because this leads to the most objective measure of resolution. The choice of tomographic reconstruction method also affects the absolute correlation values. It is well accepted that iterative reconstruction methods, such as SIRT, perform much better under noisy conditions. Nevertheless, the results in Fig. 2.3 show a clear difference between the different CTF correction methods.

2.5.2. Phase flipping

The two most common methods for CTF correction in CET are phase flipping (which corrects the parts of the frequency domain data which have negative contrast) and amplitude correction [7, 10, 11]. We also implemented and tested amplitude correction using a Wiener filter approach, but we opted for phase-flipping because this gives the best results under the chosen noise conditions.

To apply phase flipping to tilted specimens we developed the following approach. The TCTF is written in the form of Eq. (2.13), leading to a transfer function T_{TCTF} given by Eq. (2.12). Subsequently, our CTF correction algorithm is described by

$$V'(\mathbf{x}) = \int \hat{I}(\mathbf{q}) \operatorname{sgn}\left(T_{TCTF}(\mathbf{q},\mathbf{x})\right) e^{-2\pi i \mathbf{q}\cdot\mathbf{x}} d\mathbf{q}, \tag{2.16}$$

where $V'(\mathbf{x})$ is an estimate of $V(\mathbf{x})$ and $\operatorname{sgn}(x)$ the sign function.

When we compare the result of CTF_0 phase-flipping in Fig. 2.3A with the result of TCTF phase-flipping of Eq. (2.16) in Fig. 2.3B we see a dramatic increase in resolution at the edges of the specimen ($x = \pm 500$ nm). Around the tilt-axis ($x = 0$ nm) the difference between the two methods disappears because tilting the specimen does not induce a significant difference in defocus.

2.5.3. Defocus-series

The zero-crossings of the CTF make it impossible to obtain a good estimate of $V(\mathbf{x})$ (from a single projection) because at these points in the Fourier domain all information about the specimen is lost. A solution to this problem is to acquire a defocus series after which the different projections can be combined to estimate $V(\mathbf{x})$ [29].

We propose to (also) apply this approach to tomography by acquiring a defocus-series at each tilt-angle of a tilt-series while not increasing the total dose. The dose available for a certain tilt-angle is distributed over the defocus-series.

Note that this does not influence the signal-to-noise (SNR) ratio of the estimate $V'(\mathbf{x})$ [30]. It will, however, result in lower SNR per recording which might hamper the alignment procedure in practice. Assuming that a perfect alignment exists, we investigated whether a defocus-series will result in an increased resolution.

We simulated a defocus-series of three different average defocus values: -0.5, -1 and −1.5 µm, at each tilt-angle. This choice was made to properly correct for the first zero-crossing ($q = 0.63\,\mathrm{nm}^{-1}$) at a defocus of −1 µm. CTF correction for defocus-series is usually done with a modified Wiener filter [29], our experience with this approach is that it is too sensitive to noise. We chose an alternative method which resembles phase flipping for a single projection. This requires no regularisation and gives better results under our noise conditions. The CTF correction for each defocus-series at a certain tilt-angle is given by

$$\hat{V}'(\mathbf{q}) = \frac{\sum_i \hat{V}'_i(\mathbf{q})}{\sqrt{\sum_i \int T_i(\mathbf{q},\mathbf{x})^2 d\mathbf{x}}}, \tag{2.17}$$

where the summation runs over the different defocus values and $\hat{V}'_i(\mathbf{q})$ is the FT of

$$V'_i(\mathbf{x}) = \int \hat{I}_i(\mathbf{q})\, T_i(\mathbf{q},\mathbf{x}) e^{-2\pi i \mathbf{q}\cdot\mathbf{x}} d\mathbf{q}.$$

This reduces to regular phase-flipping for a single, untilted projection.

In Fig. 2.3C we notice that distributing the dose at each tilt-angle over three projections at different defocus values has the potential to increase the achievable resolution compared to TCTF phase-flipping of a single projection as shown in Fig. 2.3B. The advantage of the defocus-series based TCTF correction of Eq. (2.17) is that it does not require regularisation and that it will not amplify noise because the magnitude per spatial frequency stays fixed.

2.5.4. Influence of defocus accuracy

So far we have assumed that the exact defocus is known as an input to the CTF correction step. In reality, the defocus needs to be estimated. It is therefore important to investigate the sensitivity of the discussed methods to an error in the defocus estimation. To this end we simulated a tilt-series with an underfocus of 1 µm and subsequently used a slightly different systematic defocus value for CTF correction of the tilt-series. The results of these simulations are shown in Fig. 2.4A.

Judging from Fig. 2.4A, a systematic error in defocus estimation on the order of 50 nm is acceptable when using TCTF correction (possibly with defocus-series).

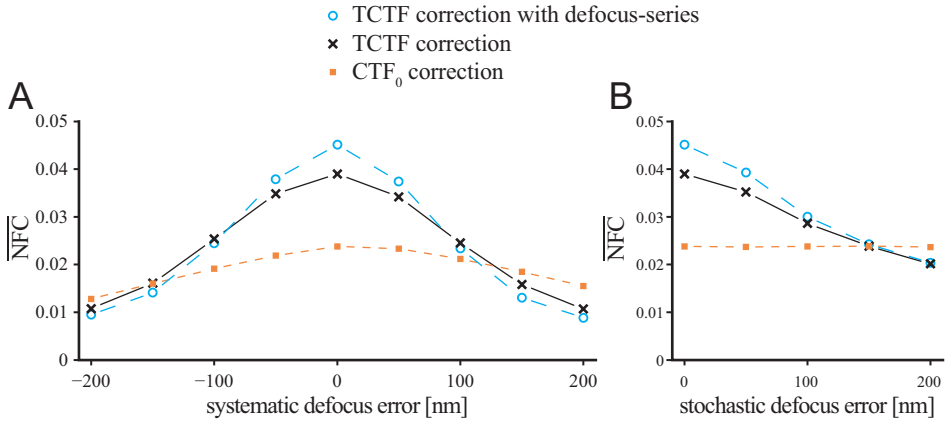


Figure 2.4. Averaged NFC values as seen in Fig. 2.3 using the discussed CTF correction methods but with an erroneous defocus for the CTF correction step. (A) Systematic defocus error for all projections. (B) Positive values are closer to focus. Normal distributed defocus error for each projection.

The thickness of the simulated specimen was 100 nm which also introduces an ambiguity of the defocus for the TCTF method. CTF_0 correction shows much less dependence on an error in the defocus estimation. This result is expected since for tilted specimens CTF_0 correction based on an erroneous defocus value is still valid for some parts of the projection.

Next to a systematic error in defocus estimation, the defocus for each projection in a tilt-series is not constant [10, 11]. We have also investigated the sensitivity of CTF correction to these fluctuations by imposing a normal distributed stochastic defocus error to each projection of the tilt-series. The result is shown in Fig. 2.4B. When compared to a systematic defocus error (shown in Fig. 2.4A) the effects are almost the same. Only for larger stochastic errors the influence is slightly smaller.

2.6. *Conclusions*

We derived an analytical expression for the loss of resolution by using the projection assumption, i.e. ignoring the actual thickness of the specimen Eq. (2.9). This estimated loss of resolution was verified by the simulation of projections at different tilt-angles and specimen thicknesses. Eq. (2.9) can now be used as a rule-of-thumb for estimating whether the projection assumption is sufficient to image at a certain resolution, or alternatively to estimate up to which resolution the TCTF is a good approximation to the 3DCTF. For actual acquisitions of a

specimen with a certain thickness the transfer beyond a certain spatial frequency is severely damped. With the help of Eq. (2.9) the signal-to-noise ratio of each projection can now be increased by insertion of the correct objective aperture that blocks the scattered waves above this frequency.

We successfully reduced the processing time required for computing the TCTF by more than a factor 100. The processing time of the 3DCTF is reduced to the time it takes to compute one Radon transform. We anticipate that this reduction of processing time will commence the development of new iterative reconstruction algorithms which incorporate these models to improve the resolution of reconstructed tomograms.

We used simulations to quantify the effects of the specimen thickness on the accuracy of the TCTF and CTF_0. This yields a first indication of the applicability of the TCTF for specimens of finite thickness. In contrast to Philippsen *et al.* [6], who used a direct inversion of projections simulated by the TCTF, we applied TCTF correction to projections simulated by the 3DCTF. This approach is much closer to reality and therefore provides a better view on the applicability of the TCTF for CTF correction.

The approach taken by Philippsen *et al.* [6] and Winkler and Taylor [8] for CTF correction is based on the direct inversion of the TCTF operator, without regularisation and with truncated singular value decomposition respectively. We proposed a new method for TCTF correction which resembles phase-flipping of the CTF_0. It does not require regularisation and cannot amplify noise. Furthermore, we also described a phase-flipping-like TCTF correction method for defocus series. Finally we showed that these CTF correction strategies are successful in increasing the resolution after tomographic reconstruction.

2.A. *Taylor expansion of transfer functions*

As an illustrative example we have worked out the first three terms of Eq. (2.14) using the T_{TCTF} (Eq. (2.12)) as a transfer function:

$$\begin{aligned}\hat{I}(\mathbf{q}) = {} & 2\sin\big(\chi(q)\big)\,\mathscr{F}_{\mathbf{x}}\left[V(\mathbf{x})\right] - \\ & 2\cos\big(\chi(q)\big)\left(\pi\lambda q^2\tan\alpha\right)\mathscr{F}_{\mathbf{x}}\left[(\mathbf{x}\cdot\boldsymbol{\beta})\,V(\mathbf{x})\right] - \\ & \sin\big(\chi(q)\big)\left(\pi\lambda q^2\tan\alpha\right)^2\mathscr{F}_{\mathbf{x}}\left[(\mathbf{x}\cdot\boldsymbol{\beta})^2\,V(\mathbf{x})\right] \\ & + \ldots . \end{aligned} \tag{2.18}$$

Please notice that Eq. (2.18) is merely meant as an example to illustrate how Eq. (2.14) should be interpreted. The TCTF can be implemented using the much faster NUFFT algorithm described earlier.

In order to accurately describe the imaging model for thick tilted specimens, the finite source-size envelope needs to be reconsidered. As derived by [17, Section 6.4.2] the envelope function belonging to the effects of the finite source size is

$$K_s(q) = \exp\left(-\frac{\left(\nabla\chi(q)\right)^2 H^2}{4\ln 2}\right) = \exp\left(-\frac{\left(\pi C_s\lambda^2 q^3 - \pi\Delta f\, q\right)^2 \alpha_i^2}{\ln 2}\right),$$

where $\nabla\chi(q)$ is the gradient of the aberration function, $H = \frac{\alpha_i}{\lambda}$, and α_i is the convergence angle. Since this envelope depends on Δf, it must be reconsidered for thick and/or tilted specimens. We define the spatially varying finite source-size envelope as

$$K_{s,z}(\mathbf{q}, z) = \exp\left(-\frac{\left(\nabla\chi(q) - \pi\lambda z q\right)^2 H^2}{4\ln 2}\right),$$

where z is the deviation from the average defocus. Applying Eq. (2.14) leads to

$$\hat{I}_{K_s}(\mathbf{q}) = \sum_{n=0}^{n_{max}} \frac{K_{s,z}^{(n)}(\mathbf{q}, 0)}{n!} \mathscr{F}_{\mathbf{x}}\left[(\mathbf{x}\cdot\boldsymbol{\beta}\tan\alpha)^n I(\mathbf{x})\right], \tag{2.19}$$

where

$$\frac{K_{s,z}^{(n)}(\mathbf{q}, 0)}{n!} = \frac{\partial^n K_{s,z}(\mathbf{q}, z)}{\partial z^n}\bigg|_{z=0} \frac{1}{n!} = K_s(q) \sum_{k=0}^{n} A_{n,k} \left(\nabla\chi(q)\right)^k \left(\frac{H}{2\sqrt{\ln 2}}\right)^{k+n},$$

and with

$$A_{n,k} = \begin{cases} (-1)^{\left(\frac{n-k}{2}\right)} \frac{2^k}{\left(\frac{n-k}{2}\right)!k!} & \text{if } n-k \text{ is even and } k < n, \\ 0 & \text{otherwise.} \end{cases}$$

For the spatially variant envelope Eq. (2.19) we found that $n_{max} = 4$ is in most circumstances sufficient.

References

[1] W. Baumeister, *Electron tomography: Towards visualizing the molecular organization of the cytoplasm,* Current Opinion in Structural Biology **12**, 679 (2002).

[2] J. Zhu, P. A. Penczek, R. Schröder, and 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**, 197 (1997).

[3] G. E. Murphy, J. R. Leadbetter, and G. J. Jensen, *In situ structure of the complete Treponema primitia flagellar motor,* Nature **442**, 1062 (2006).

[4] J. Liu, A. Bartesaghi, M. J. Borgnia, G. Sapiro, and S. Subramaniam, *Molecular architecture of native HIV-1 gp120 trimers,* Nature **455**, 109 (2008).

[5] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan, *Single-particle electron cryo-microscopy: Towards atomic resolution,* Quarterly Reviews of Biophysics **33**, 307 (2000).

[6] A. Philippsen, H.-A. Engel, and A. Engel, *The contrast-imaging function for tilted specimens,* Ultramicroscopy **107**, 202 (2007).

[7] J. J. Fernández, S. Li, and R. A. Crowther, *CTF determination and correction in electron cryotomography,* Ultramicroscopy **106**, 587 (2006).

[8] H. Winkler and K. A. Taylor, *Focus gradient correction applied to tilt series image data used in electron tomography,* Journal of Structural Biology **143**, 24 (2003).

[9] J. A. Mindell and N. Grigorieff, *Accurate determination of local defocus and specimen tilt in electron microscopy,* Journal of Structural Biology **142**, 334 (2003).

[10] Q. Xiong, M. K. Morphew, C. L. Schwartz, A. H. Hoenger, and D. N. Mastronarde, *CTF determination and correction for low dose tomographic tilt series,* Journal of Structural Biology **168**, 378 (2009).

[11] G. Zanetti, J. D. Riches, S. D. Fuller, and J. A. G. Briggs, *Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging,* Journal of Structural Biology **168**, 305 (2009).

[12] Y. Wan, W. Chiu, and Z. H. Zhou, *Full contrast transfer function correction in 3D cryo-EM reconstruction,* in *IEEE Proceedings of ICCCAS 2004*, Vol. 2 (2004) pp. 960–964.

[13] G. J. Jensen and R. D. Kornberg, *Defocus-gradient corrected back-projection,* Ultramicroscopy **84**, 57 (2000).

[14] I. G. Kazantsev, J. Klukowska, G. T. Herman, and L. Cernetic, *Fully three-dimensional defocus-gradient corrected backprojection in cryoelectron microscopy,* Ultramicroscopy **110**, 1128 (2010).

[15] J. N. Dubowy and G. T. Herman, *An approach to the correction of distance-dependent defocus in electron microscopic reconstruction,* in *IEEE International Conference on Image Processing, ICIP 2005*, Vol. 3 (2005) pp. 748–751.

[16] I. G. Kazantsev, G. T. Herman, and L. Cernetic, *Backprojection-based reconstruction and correction for distance-dependent defocus in cryoelectron microscopy,* in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2008* (2008) pp. 133–136.

[17] L. Reimer and H. Kohl, *Transmission Electron Microscopy*, 5th ed. (Springer-Verlag, Berlin, 1984).

[18] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging* (IEEE Service Center, 1988).

[19] M. Frigo and S. G. Johnson, *The design and implementation of FFTW 3,* Proceedings of the IEEE **93**, 216 (2005).

[20] A. Dutt and V. Rokhlin, *Fast Fourier transforms for nonequispaced data,* SIAM Journal on Scientific Computing **14**, 1368 (1993).

[21] A. Dutt and V. Rokhlin, *Fast Fourier transforms for nonequispaced data, ii,* Applied and Computational Harmonic Analysis **2**, 85 (1995).

[22] L. Greengard and J.-Y. Lee, *Accelerating the nonuniform fast Fourier transform,* SIAM Review **46**, 443 (2004).

[23] S. Matej, J. A. Fessler, and I. G. Kazantsev, *Iterative tomographic image reconstruction using Fourier-based forward and back-projectors,* IEEE Transactions on Medical Imaging **23**, 401 (2004).

[24] A. G. Ramm and A. I. Katsevich, *The Radon transform and local tomography* (CRC, 1996).

[25] E. T. Quinto and O. Öktem, *Local tomography in electron microscopy,* SIAM Journal on Applied Mathematics **68**, 1282 (2008).

[26] M. Vulović, B. Rieger, L. J. van Vliet, A. J. Koster, and R. B. G. Ravelli, *A toolkit for the characterization of CCD cameras for transmission electron microscopy,* Acta Crystallographica Section D **66**, 97 (2010).

[27] M. van Heel and M. Schatz, *Fourier shell correlation threshold criteria,* Journal of Structural Biology **151**, 250 (2005).

[28] M. van Heel, *Similarity measures between images,* Ultramicroscopy **21**, 95 (1987).

[29] P. A. Penczek, J. Zhu, R. Schröder, and J. Frank, *Three dimensional reconstruction with contrast transfer compensation from defocus series,* Scanning Microscopy **11**, 147 (1997).

[30] D. J. DeRosier, *Correction of high-resolution data for curvature of the Ewald sphere,* Ultramicroscopy **81**, 83 (2000).

3

Fast, spatially varying CTF correction in TEM

All that is gold does not glitter,
Not all those who wander are lost;
The old that is strong does not wither,
Deep roots are not reached by the frost.

JOHN RONALD REUEL TOLKIEN

Any intelligent fool can make things bigger, more complex, and more violent. It takes a touch of genius — and a lot of courage — to move in the opposite direction.

ERNST FRIEDRICH SCHUMACHER

Lenard M. Voortman
Erik M. Franken
Lucas J. van Vliet
Bernd Rieger

Ultramicroscopy 118 (2012), pp. 26–34

3

Abstract

We have developed new methods for contrast transfer function (CTF) correction of tilted and/or thick specimens. In order to achieve higher resolutions in cryo-electron tomography (CET), it is necessary to account for the defocus gradient on a tilted specimen and possibly the defocus gradient within a thick specimen. CTF correction methods which account for these defocus differences have recently gained interest. However, there is no global CTF correction method available to this date (to process the entire field-of-view at once) which can use different inverse filters, e.g. phase-flipping or Wiener filter, and which can do so within a reasonable time for realistic image sizes. We show that the CTF correction methods presented in this article correctly account for the spatially varying defocus, can employ different inverse filters and are significantly faster (more than 50 times) than existing methods. We provide proof-of-principle implementations of all the presented CTF correction methods online.

3.1. *Introduction*

Cryo-electron tomography (CET) is an essential tool to study the three-dimensional (3D) structure of biological specimens at molecular resolution. To push the achievable resolution beyond its current limitations, it is necessary to take more details of the image formation into account.

Biological specimens are weak electron scatterers; the main source of image contrast is therefore phase contrast. Deliberately imaging the specimen out of focus allows the weakly scattered wave to interfere with the unscattered wave, thereby generating contrast. In this weak-phase contrast regime, image formation is for an important part described by the contrast transfer function (CTF). The CTF is an oscillating function of the spatial frequency and depends, among other parameters of the imaging system, on the applied defocus.

In CET many two-dimensional (2D) projections of the 3D structure of a specimen are recorded at different tilt-angles. After acquiring such a tilt-series, one estimates the 3D structure of the specimen using tomographic reconstruction. Tilting of the specimen induces a defocus gradient perpendicular to its tilt-axis. This defocus gradient has, up to recently, been considered to be too small to have a significant influence on the resolution.

Recent advances in the technology of electron microscopy, together with the challenges of imaging larger specimens, have put more demanding requirements on the imaging models [1]. To achieve high resolution, the imaging models should be accurate up to higher spatial frequencies. They have to include effects such as the defocus gradient due to tilting of the specimen (resolution beyond 5 nm [2]) but also the defocus differences due to the finite thickness of the specimen (resolution beyond 1 nm [3]).

The influence of the defocus differences within a specimen on the reconstruction quality has been studied by various groups. Some studies focused on the forward model for tilted, thin specimens [2–4]. Others take a local (strip/tiling) approach to correct for tilted and thin specimens [4–7]. A few studies focused on the defocus gradient along the projection direction [8–10]. All of these CTF correction methods [2, 4–10], however, are either local approaches or very limited in the choice of inverse filter (e.g. phase-flipping is not possible).

A practical drawback of the global approaches currently available is that they are particularly slow for large image sizes (above 1024^2) with computation times in the order of days. The allure of a global approach is that it does not require fine-tuning of the size of a local neighborhood or stitching of local results. We prefer these global methods for their elegance and show in this article that it is possible to significantly decrease the required computation time.

We derive two global methods for CTF correction of tilted and/or thick specimens. Both methods can be used with different inverse filters (including phase flipping and Wiener filtering) and employ fast algorithms. For large images (4096^2), tilted CTF correction can be computed in a couple of hours on a modern workstation.

3.2. *Theory*

Image formation in CET is governed by the CTF. The CTF model assumes weak interaction (Born approximation) and small scattering angles (weak phase). The Born approximation allows us to treat each slice of the potential $V_\alpha(\mathbf{x}, z)$ independently. Each slice is convolved with the appropriate point-spread function (PSF), which is the inverse Fourier transform of the CTF: $\mathscr{F}_{\mathbf{q}}^{-1}\left[H(\mathbf{q}, z)\right]$, where $H(\mathbf{q}, z)$ is the CTF at an axial distance z from the center of the specimen, $\mathbf{q} = (q_x, q_y)$ and $\mathscr{F}_{\mathbf{q}}^{-1}$ is the inverse Fourier transform from $\mathbf{q} \in \mathbb{R}^2$ to $\mathbf{x} \in \mathbb{R}^2$. Finally, all slices are added together to give the image intensity $I_\alpha(\mathbf{x})$. This leads to a model for the 2D intensity map in the camera plane for a 3D specimen

$$I_\alpha(\mathbf{x}) = \int \mathscr{F}_{\mathbf{q}}^{-1}\left[H(\mathbf{q}, z)\right] * V_\alpha(\mathbf{x}, z)\, dz, \tag{3.1}$$

where $*$ is the convolution operator which only acts on $\mathbf{x}$ and $V_\alpha(\mathbf{x}, z)$ is the potential of the (rotated) specimen at tilt-angle α in the microscope coordinates $(\mathbf{x}, z)$.

The CTF is generally given by [11]

$$H(\mathbf{q}, z) = 2\sin\left(\chi(q) - \pi\lambda z q^2\right), \tag{3.2}$$

with $\chi(q) = \frac{2\pi}{\lambda}\left(\frac{1}{4}C_s\lambda^4 q^4 + \frac{1}{2}\Delta f \lambda^2 q^2\right)$ the aberration function, C_s the spherical aberration, Δf the defocus at the center of the specimen $z = 0$, z the axial coordinate relative to the microscope, λ the electron wavelength and $q = \|\mathbf{q}\|$. We use the convention that underfocus implies $\Delta f < 0$.

Many different methods exist for tomographic reconstruction, of which back-projection (BP) is the simplest. BP is known to overestimate the contribution of low spatial frequencies. It is therefore common to incorporate an additional filtering step which leads to filtered back-projection (FBP). Nevertheless, we will use the BP algorithm to derive our algorithms in order to keep the notation as simple as possible. Furthermore, the additional filtering step does not alter our conclusions. At the end of this section we will indicate the steps required for the additional filtering.

Since we are working with tomography, we define two coordinate systems: the specimen coordinates $(\tilde{\mathbf{x}}, \tilde{z})$ and the microscope coordinate system $(\mathbf{x}, z)$. The rotation between the two coordinate systems is defined by two variables: the tilt-angle α and the orientation of the tilt-axis (azimuth) β. We relate the microscope coordinates to the specimen coordinates using a rotation operator $\mathscr{R}_\alpha$:

$$V(\tilde{\mathbf{x}}, \tilde{z}) = \mathscr{R}_\alpha \left[V_\alpha(\mathbf{x}, z) \right], \tag{3.3}$$

where $V(\tilde{\mathbf{x}}, \tilde{z})$ is the specimen in specimen coordinates, and $V_\alpha(\mathbf{x}, z)$ is the specimen in microscope coordinates. The specific form of the rotation operator is not relevant for the rest of our derivations.

The BP method consists of three steps. The first step is the spreading-back along the projection direction z for all projection angles α

$$I_{\text{spread},\alpha}(\mathbf{x}, z) = I_\alpha(\mathbf{x}), \tag{3.4}$$

where $I_\alpha(\mathbf{x})$ is an intensity map of the 2-dimensional spatial coordinate $\mathbf{x}$. The result of this operation is that $I_{\text{spread},\alpha}$ is now a function of $\mathbf{x}$ and z, but invariant along z. These kinds of functions are commonly referred to as ridge functions, a term coined by Logan and Shepp [12]. The next step is to rotate all $I_{\text{spread},\alpha}(\mathbf{x}, z)$ to the specimen coordinates $(\tilde{\mathbf{x}}, \tilde{z})$ using the rotation operator $\mathscr{R}_\alpha$. The third step is to add all spread-back and rotated contributions into a single volume:

$$BP(\tilde{\mathbf{x}}, \tilde{z}) = \sum_\alpha \mathscr{R}_\alpha \left[I_{\text{spread},\alpha}(\mathbf{x}, z) \right] d\alpha. \tag{3.5}$$

In electron tomography the available tilt-angles are usually limited to $\pm 70°$, resulting in the well-known missing wedge artifacts in the reconstruction volumes.

3.2.1. Three-dimensional CTF correction

To account for the defocus dependence of the CTF, Jensen and Kornberg proposed the defocus-gradient correct back-projection (DGCBP) method [8], to which we will refer as three-dimensional CTF (3DCTF) correction. This method introduces the following extra step between Eqs. (3.4) and (3.5)

$$I_{\text{3DCTF},\alpha}(\mathbf{x}, z) = \mathscr{F}_{\mathbf{q}}^{-1} \left[H_{\text{inv}}(\mathbf{q}, z) \right] * I_{\text{spread},\alpha}(\mathbf{x}, z), \tag{3.6}$$

where $H_{\text{inv}}(\mathbf{q}, z)$ is the inverse filter (which approximates $1/H(\mathbf{q}, z)$) as a function of z. The convolution operator $*$ only acts on $\mathbf{x}$. Similar to Eq. (3.1), Eq. (3.6) treats each slice in the z-direction independently. For a rigorous mathematical

justification of 3DCTF correction, see Kazantsev *et al.* [10]. The computational complexity of 3DCTF as described in Jensen and Kornberg [8] and Kazantsev *et al.* [10] is relatively large because the 3D volume $I_{\mathrm{3DCTF},\alpha}(\mathbf{x}, z)$ needs to be computed for each tilt-angle and subsequently rotated and added to $BP(\tilde{\mathbf{x}}, \tilde{z})$.

We now take the first step towards an efficient algorithm via the 3D Fourier transform of Eq. (3.6)

$$\hat{I}_{\mathrm{3DCTF},\alpha}(\mathbf{q}, q_z) = \mathscr{F}_z\left[H_{\mathrm{inv}}(\mathbf{q}, z)\right] *_{q_z} \mathscr{F}_{\mathbf{x},z}\left[I_{\mathrm{spread},\alpha}(\mathbf{x}, z)\right], \tag{3.7}$$

where $*_{q_z}$ is a convolution in q_z only. Notice that the convolution in Eq. (3.6), which acts only on $\mathbf{x}$, is replaced by a multiplication. However, the Fourier transform is also carried out along the z-direction, resulting in a convolution operator in q_z. Since $I_{\mathrm{spread},\alpha}$ is constant along z, $\mathscr{F}_{\mathbf{x},z}\left[I_{\mathrm{spread},\alpha}(\mathbf{x}, z)\right]$ can be replaced by $\mathscr{F}_{\mathbf{x}}\left[I_\alpha(\mathbf{x})\right]\delta(q_z)$. The convolution in q_z results in

$$\hat{I}_{\mathrm{3DCTF},\alpha}(\mathbf{q}, q_z) = \mathscr{F}_z\left[H_{\mathrm{inv}}(\mathbf{q}, z)\right]\mathscr{F}_{\mathbf{x}}\left[I_\alpha(\mathbf{x})\right]. \tag{3.8}$$

Due to the periodic nature of the CTF $H(\mathbf{q}, z)$ (from Eq. (3.2)) as a function of z, we can rewrite Eq. (3.8) into a special form. The $\sin(az)$-like behavior of the CTF results in a Fourier transform in the form of two δ-functions at $\delta(q_z \pm a)$. In general, the Fourier transform for periodic functions consists of a set of equidistant δ-functions. We assume that the inverse filter $H_{\mathrm{inv}}(\mathbf{q}, z)$ can be expanded as a sum of $\sin(anz)$. Therefore, we can state that

$$\mathscr{F}_z\left[H_{\mathrm{inv}}(\mathbf{q}, z)\right] = \sum_{n=\pm1,\pm3,\ldots} i\,a(n)e^{-in\chi(q)}\delta(q_z - \frac{1}{2}n\lambda q^2), \tag{3.9}$$

where $a(n) \in \mathbb{C}$ is a set of coefficients depending on the type of inverse filter.

Inserting Eq. (3.9) into Eq. (3.8) and using the notation $\hat{I}_\alpha(\mathbf{q}) = \mathscr{F}_{\mathbf{x}}\left[I_\alpha(\mathbf{x})\right]$ leads to

$$\hat{I}_{\mathrm{3DCTF},\alpha}\left(\mathbf{q}, q_z\right) = \sum_{n=\pm1,\pm3,\ldots} i\,a(n)e^{-in\chi(q)}\hat{I}_\alpha\left(\mathbf{q}\right)\delta(q_z - \frac{1}{2}n\lambda q^2). \tag{3.10}$$

This is one of the central results in this article as it will lead us to a computationally efficient implementation of CTF correction.

We derived the coefficients $a(n)$ for three different inverse filters. The first, H_{inv}^{CTF}, is included as an illustrative example only,

$$H_{\mathrm{inv}}^{\mathrm{CTF}}(\mathbf{q}, z) = \frac{1}{4}H(\mathbf{q}, z), \tag{3.11}$$

$$H_{\text{inv}}^{\text{PhaseFlip}}(\mathbf{q},z) = \frac{1}{2}\operatorname{sgn}\big(H(\mathbf{q},z)\big), \tag{3.12}$$

$$H_{\text{inv}}^{\text{Wiener}}(\mathbf{q},z) = \frac{H(\mathbf{q},z)}{\big(H(\mathbf{q},z)\big)^2 + b^2} \text{ where } b \in \mathbb{R}^+, \tag{3.13}$$

$$a^{CTF}(n) = \begin{cases} \frac{1}{4} & \text{if } n = 1, \\ -\frac{1}{4} & \text{if } n = -1, \\ 0 & \text{otherwise,} \end{cases}$$

$$a^{\text{PhaseFlip}}(n) = \frac{1}{n\pi},$$

$$a^{\text{Wiener}}(n) = \operatorname{sgn}(n)\frac{\left(\sqrt{1+(b/2)^2} - (b/2)\right)^{|n|}}{2\sqrt{1+(b/2)^2}}.$$

3.2.2. Truncation of Fourier series

In order to make Eq. (3.10) useful for actual computation, the summation needs to be truncated. The summation in this equation runs over all odd numbers. For large n, however, the coefficients $a(n)$ will approach zero for all functions which we might consider for CTF correction. Furthermore, for all square-integrable functions, the sum of $|a(n)|^2$ will always be finite. Using Parseval's identity, the relative error due to truncation of the Fourier series is given by

$$E(N) = \frac{1}{M}\sum_{|n|>N}^{\infty} |a(n)|^2, \tag{3.14}$$

where $\{n \in \mathbb{Z} : n \text{ is odd and } |n| > N\}$ and $M = \sum_m^{\infty} |a(m)|^2$ is a summation over all odd m.

For phase-flipping Eq. (3.12) and Wiener Eq. (3.13) filter, we have worked out Eq. (3.14):

$$E(N)_{\text{PhaseFlip}} = \frac{2}{\pi^2}\psi_1\left(N/2+1\right),$$

$$E(N)_{\text{Wiener}} = \left(b/2 - \sqrt{1+(b/2)^2}\right)^{2N+2},$$

where ψ_1 is the trigamma function.

Truncation of the Fourier series is therefore justifiable when the relative truncation error $E(N)$ passes below a certain threshold ϵ:

$$\hat{I}_{\text{3DCTF}}\left(\mathbf{q}, q_z\right) \approx \sum_{n=\pm 1,\pm 3,\ldots}^{\pm N} i\, a(n) e^{-in\chi(q)} \hat{I}_{\alpha}\left(\mathbf{q}\right) \delta(q_z - \frac{1}{2} n\lambda q^2), \tag{3.15}$$

where N is the smallest odd integer for which $E(N) < \epsilon$.

The advantage in computational complexity of Eq. (3.15) over Eq. (3.6) is the sparsity of $\hat{I}_{\text{3DCTF}}\left(\mathbf{q}, q_z\right)$. This complexity can be reduced even further by applying the rotation and summation of $\hat{I}_{\text{3DCTF}}\left(\mathbf{q}, q_z\right)$ into $\widehat{BP}(\tilde{\mathbf{q}}, \tilde{q}_z)$ in the Fourier domain as outlined below.

3.2.3. Direct Fourier reconstruction

The most common technique for tomographic reconstruction is real-space back-projection. In essence, Eq. (3.5) and Eq. (3.4) are computed simultaneously by interpolation in the spatial domain. Some of the advantages of this technique are that the implementation is straightforward and parallelizable and that real-space interpolation does not produce additional artifacts.

An alternative technique, called direct Fourier reconstruction, is not so commonly used. This technique relies on accurate interpolation in the Fourier domain. Since this interpolation is not straightforward and prone to errors, Fourier domain reconstruction is primarily used to explain the theory behind back-projection. However, when implemented correctly, the results of direct Fourier reconstruction and those obtained by real-space back-projection are of similar quality.

How direct Fourier reconstruction works is explained below. The effect of the spreading-back operation in Eq. (3.4) is to put the Fourier transformed projection on a plane in the 3D Fourier domain of the reconstruction volume. The rotation and summation in Eq. (3.5), when transformed to the Fourier domain, results in an identical rotation and summation of the individual contributions. Therefore, the task of direct Fourier reconstruction is to put the Fourier transformed projections on tilted planes into a single 3D (Fourier) space. In reality, however, all our measured data is discretized and we only have a limited number of projections. As a result, most of the data points that must be added to the reconstruction volume do not end up on a (rectangular) grid point, making interpolation necessary. Direct Fourier reconstruction relies on a suitable interpolation method otherwise very strong artifacts appear. We use the non-uniform fast Fourier transform (NUFFT) [13] as interpolation method in the Fourier domain. Fig. 3.1A gives a simple illustration of direct Fourier reconstruction.

3.2.4. Implementation of 3DCTF correction

Algorithm 1 Fourier domain implementation of 3DCTF correction

```
Allocate Fourier domain of reconstruction volume V̂
for all projections I_α in tilt-series do
    Î_α ← Fourier transform I_α
    create rotation matrix R
    for all q do
        compute χ(q)
        for all n ∈ ℤ such that n is odd and |n| ≤ N do
            compute a(n)
            put i a(n) Î_α(q) e^{-i n χ(q)} at R(q, -½ n λ q²)^T in V̂
        end for
    end for
end for
V ← inverse Fourier transform V̂
```

The reason to introduce direct Fourier reconstruction is that Eq. (3.10) suggests an efficient Fourier domain implementation of 3DCTF correction. Instead of putting the Fourier transformed projections onto tilted planes, we suggest to put the same data on tilted parabola with different phase and amplitude factors. The phase factors correspond to the phase factors of the CTF ($ie^{-in\chi(q)}$), the amplitude factors are given by $a(n)$. The different parabola are constituted by $\delta(q_z - \frac{1}{2}n\lambda q^2)$ in Eq. (3.10). We use the NUFFT [13] to put the data on these tilted parabola. We use a type 1 NUFFT, which allows the reconstruction of a signal from Fourier domain data which has been sampled at non-uniform locations. The reconstruction from non-uniformly sampled Fourier domain data is conceptually the same as putting data on tilted parabola in the Fourier domain. Figs. 3.1B&C give a visual representation of 3DCTF correction via Eq. (3.10). To fully clarify our implementation, Alg. 1 provides the corresponding pseudo-code.

3.2.5. CTF correction for tilted and thin specimens

In the previous subsection, we introduced a new algorithm for full three-dimensional CTF correction. Here we will introduce an approximation to Alg. 1 which is specifically designed for thin specimens.

In CET, the specimen under investigation is typically thin compared to the field-of-view ($\sim 100\,\text{nm}$ compared to $> 1\,\mu\text{m}$). For these thin specimens, the defocus difference due to the finite thickness can be ignored (up to a certain resolution [3]). It is therefore quite common to separate the (back-)projection of

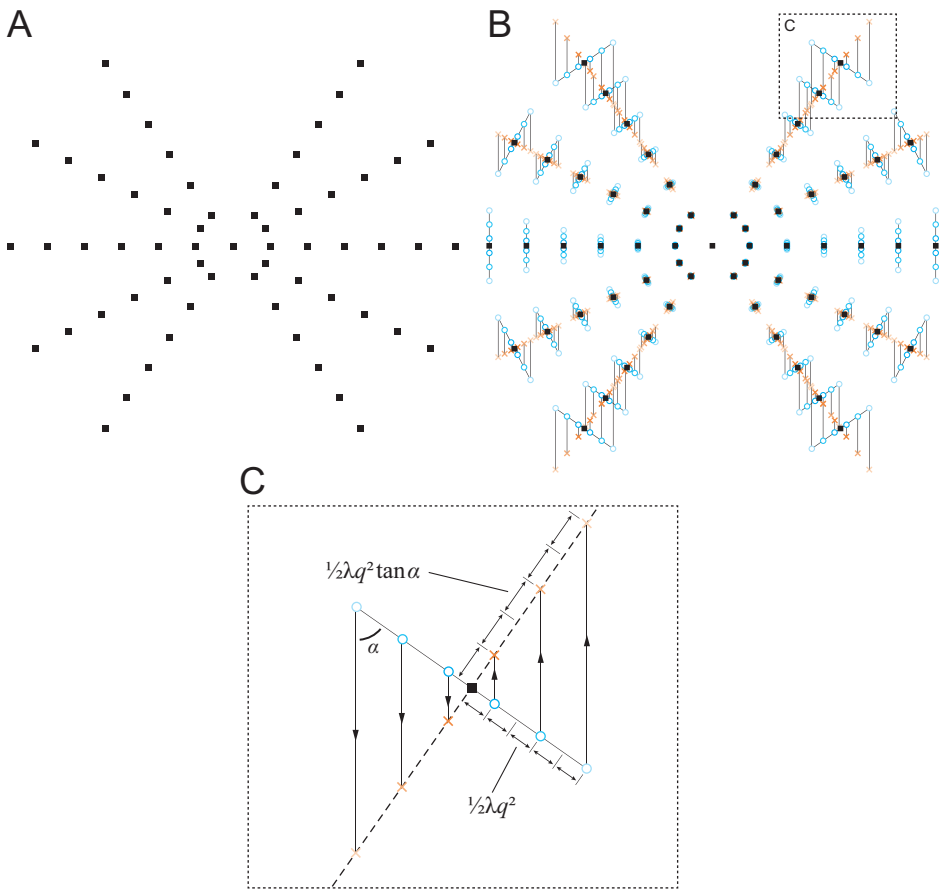


Figure 3.1. Fourier domain representation of the CTF correction methods. (A) Direct Fourier reconstruction puts the Fourier transformed projections on planes in the Fourier domain. It is clear that for discrete data, the data points (■) in general do not form a rectangular grid. (B) & (C) A visual representation of 3DCTF correction Eq. (3.10) (○) and TCTF correction Eq. (3.19) (×) when combined with direct Fourier reconstruction. For comparison, the data points used by regular CTF correction Eq. (3.21) are presented in (■). The detail in (C) shows the series expansion of Eq. (3.10) and Eq. (3.19), respectively perpendicular to and along the central section. Data points (○ and ×) located further away from the in-plane data point ■ receive smaller weights. For clarity, only those points relative to a single data point of the projection are shown. Notice that relative distances have been greatly exaggerated to illustrate the method.

the specimen from the influence of the CTF. For untilted specimens this means that the defocus is assumed to be constant throughout the specimen. However, even an infinitely thin specimen can have a large defocus gradient over the field-of-view due to tilting. In order to account for this defocus gradient, Philippsen *et al.* [2] derived the tilted contrast imaging function:

$$\begin{aligned}\hat{I}_\alpha(\mathbf{q}) = {} & ie^{-i\chi(q)}\hat{V}_{\text{proj},\alpha}\left(\mathbf{q}-\frac{1}{2}\lambda q^2\boldsymbol{\beta}\tan\alpha\right) \\ & - ie^{i\chi(q)}\hat{V}_{\text{proj},\alpha}\left(\mathbf{q}+\frac{1}{2}\lambda q^2\boldsymbol{\beta}\tan\alpha\right),\end{aligned} \tag{3.16}$$

where $V_{\text{proj},\alpha}(\mathbf{x})$ is the projection of the specimen potential along the z-direction of the microscope and $\hat{V}_{\text{proj},\alpha}(\mathbf{q})$ its Fourier transform. $\boldsymbol{\beta}$ is a unit-vector perpendicular to the tilt-axis $\boldsymbol{\beta} = (\cos\beta, \sin\beta)$. β denotes the orientation of the tilt-axis (azimuth). We will refer to Eq. (3.16) as tilted CTF (TCTF). One of the advantages of the TCTF is that it separates the projection of the specimen from the CTF. As a result, the TCTF is a function of the (geometrically) projected specimen which allows a more efficient implementation than Eqs. (3.6) and (3.10). For thick specimens, on the other hand, we showed in the previous section that the (back-)projection procedure cannot be separated from the CTF correction.

The difference between the TCTF Eq. (3.16) and a CTF for untilted specimens is the shift of $\mp\frac{1}{2}\lambda q^2\tan\alpha$ in the Fourier domain. To correct for the TCTF, we should correct not only for the phases $\pm ie^{\mp i\chi(q)}$ but also for the shifts in the Fourier domain. An approximation to the solution is to reapply the TCTF to the projections $\hat{I}_\alpha$. For the very simple inverse filter of Eq. (3.11) and in case of untilted and thin specimens it is easy to show that the forward model is essentially the same as the CTF correction. Hence, in order to correct for the CTF, the CTF is applied again. For spatially varying CTF models, however, this approach is only partially correct.

For the TCTF, the forward model differs from the corresponding CTF because of the combination of spatial dependence due to the tilt and the delocalization effects of the CTF. In the formalism of Philippsen *et al.* [2], this difference can be seen from the shifts in the Fourier domain which are proportional to $\mp\frac{1}{2}\lambda q^2\tan\alpha$. When the TCTF is reapplied, the shifts in the Fourier domain do not cancel each other completely due to the dependence on q [2, 14]. This difference in shifts is visualized in Fig. 3.2.

If we use Fig. 3.2 as a reference, it is straightforward to modify the forward model of the CTF (Eq. (3.16)) into a simple CTF correction method with an inverse

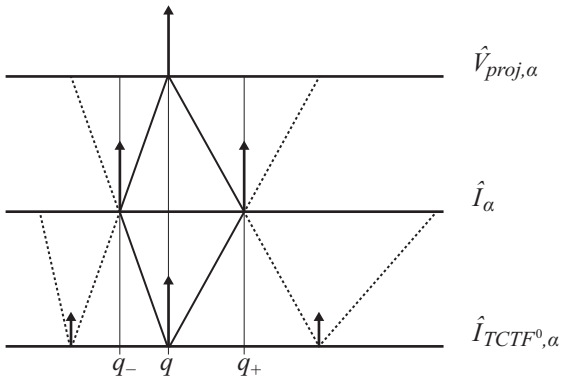


Figure 3.2. An overview of the shifts in the Fourier domain of the TCTF Eq. (3.16) and TCTF^0 correction Eq. (3.17). Each point in $\hat{I}_{\alpha}$ is related to two equidistant points in $\hat{V}_{\text{proj},\alpha}$. Similarly, each point in $\hat{I}_{\alpha}$ is related to two equidistant points in $\hat{I}_{\text{TCTF}^0,\alpha}$. In contrast, $\hat{I}_{\text{TCTF}^0,\alpha}$ is related to two points in $\hat{I}_{\alpha}$ which are not equidistant.

filter similar to Eq. (3.11):

$$\hat{I}_{\text{TCTF}^0,\alpha}(\mathbf{q}) = \frac{1}{4} i e^{-i\chi(q_-)} \hat{I}_{\alpha}(\mathbf{q}_-) - \frac{1}{4} i e^{i\chi(q_+)} \hat{I}_{\alpha}(\mathbf{q}_+), \tag{3.17}$$

where $\mathbf{q}_-$ and $\mathbf{q}_+$ are implicitly defined as

$$\begin{aligned} \mathbf{q}_- + \frac{1}{2}\lambda q_-^2 \boldsymbol{\beta} \tan\alpha &= \mathbf{q}, \\ \mathbf{q}_+ - \frac{1}{2}\lambda q_+^2 \boldsymbol{\beta} \tan\alpha &= \mathbf{q}. \end{aligned} \tag{3.18}$$

The quadratic equations of Eq. (3.18) formally have two solutions. However, only one solution is within the range of $\mathbf{q}$ that are captured in TEM since $\|\lambda\mathbf{q}\| \ll 1$. Fig. 3.2 illustrates the relation between $\mathbf{q}_{\mp}$ and $\mathbf{q}$.

As an alternative to analyzing the shifts in the TCTF forward model, TCTF correction can be derived from the model of 3DCTF correction. We use the concept that a thin specimen has relatively slow variations in the $\tilde{q}_z$ direction of its Fourier domain, where $\tilde{q}_z$ is along the $\tilde{z}$ direction in specimen coordinates. For infinitely thin specimens it is trivial to show that its Fourier domain is constant in the $\tilde{q}_z$ direction.

From the theory of direct Fourier reconstruction, we learned that the Fourier transform of a projection yields a tilted plane in the Fourier domain (the dashed line in Fig. 3.1C). To approximate 3DCTF correction (the circles in Fig. 3.1C), we shift $\delta(q_z \mp \frac{1}{2} n\lambda q^2)$ of Eq. (3.10) onto the projection plane. This shift is carried out

along the $\tilde{q}_z$ direction of the specimen (see Fig. 3.1C). The assumption that the specimen is thin, and therefore slowly varying in the Fourier domain along $\tilde{q}_z$, permits shifting along this direction. The delta-peaks move to $\delta(\mathbf{q} \mp \frac{1}{2} n\lambda q^2 \boldsymbol{\beta} \tan\alpha)$.

The general form of the TCTF correction method is

$$\hat{I}_{\mathrm{TCTF},\alpha}(\mathbf{q}) = \sum_{n=\pm 1,\pm 3,\ldots} i\, a(n) e^{-in\chi(q_n)} \hat{I}_\alpha(\mathbf{q_n}), \tag{3.19}$$

where $a(n) \in \mathbb{C}$ is a set of filter coefficients and $\mathbf{q_n}$ is given by

$$\mathbf{q_n} + \frac{1}{2} n\lambda q_n^2 \boldsymbol{\beta} \tan\alpha = \mathbf{q}. \tag{3.20}$$

Whereas the 3DCTF correction method of Eq. (3.10) is a 3D function in $(\mathbf{q}, q_z)$, TCTF correction is applied to the projections $I_\alpha(\mathbf{x})$. Hence, we have successfully separated the (back-)projection from the CTF correction.

3.2.6. Implementation of the TCTF

The implementation of Eq. (3.19) for discrete datasets is not straightforward. The difficulty lies in the shifted sample locations $\mathbf{q_n}$ in the Fourier domain. As these sample locations do not form a regular grid, the fast Fourier transform (FFT) can no longer be used to compute $\hat{I}_\alpha(\mathbf{q_n})$. This problem is similar to those encountered when implementing Eq. (3.16). For Eq. (3.16), however, there already exists an efficient implementation which uses the non-uniform fast Fourier transform (NUFFT) [3]. Although a similar approach can be used to implement the TCTF, a more elegant solution is presented here.

The NUFFT used to implement the forward TCTF in Voortman *et al.* [3] allows fast computation of sample points in the Fourier domain at arbitrary locations (called type 2 NUFFT by Greengard and Lee [13]). For TCTF correction we propose to use a different kind of NUFFT, which allows the reconstruction of a signal from Fourier domain data which has been sampled at non-uniform locations (type 1 NUFFT). The procedure for Eq. (3.19) is to use the data $\hat{I}_\alpha(\mathbf{q_n})$ (which is on a regular grid), multiply each point with its appropriate phase factor $i\, a(n) e^{-in\chi(q_n)}$ and finally place it at $\hat{I}_{\mathrm{TCTF},\alpha}(\mathbf{q_n} + \frac{1}{2} n\lambda q_n^2 \boldsymbol{\beta} \tan\alpha)$. Whereas in the implementation of Eq. (3.16) we sample at non-uniform locations and transform to a regular grid $\mathbf{q}$, for Eq. (3.19) we take $\mathbf{q_n}$ on a regular grid and combine these points at non-uniform locations $\mathbf{q}$ in $\hat{I}_{\mathrm{TCTF},\alpha}$.

The pseudo-code for TCTF correction is given in Alg. 2. A visual representation is given in Fig. 3.1B&C.

Algorithm 2 Fourier domain implementation of TCTF correction

Allocate Fourier domain of CTF corrected projection $\hat{I}_{\mathrm{TCTF},\alpha}$
$\hat{I}_\alpha \leftarrow$ Fourier transform I_α
for all q **do**
  compute $\chi(q)$
  for all $n \in \mathbb{Z}$ such that n is odd and $|n| \leq N$ **do**
    compute $a(n)$
    put $i\,a(n)\hat{I}_\alpha(q)e^{-in\chi(q)}$ at $q - \frac{1}{2}n\lambda q^2$ in $\hat{I}_{\mathrm{TCTF},\alpha}$
  end for
end for
$I_{\mathrm{TCTF},\alpha} \leftarrow$ inverse Fourier transform $\hat{I}_{\mathrm{TCTF},\alpha}$

3.2.7. Regular CTF correction

In this article we compare TCTF correction (Eq. (3.19)) and 3DCTF correction (Eq. (3.10)) to spatially invariant CTF correction. The latter we will refer to as regular CTF correction. For completeness, we write regular CTF correction in the same form as Eqs. (3.19) and (3.10), i.e. expanded in a Fourier series:

$$\hat{I}_{CTF,\alpha}(\mathbf{q}) = \sum_{n=\pm1,\pm3,\ldots} i\,a(n)e^{-in\chi(q)}\,\hat{I}_\alpha(\mathbf{q}), \tag{3.21}$$

where $a(n) \in \mathbb{C}$ is a set of filter coefficients. Notice that even though Eqs. (3.10) and (3.19) look very similar to Eq. (3.21), there are some important differences. Eq. (3.10) is a 3D function and explicitly depends on q_z. Eq. (3.19) is a function of $\mathbf{q_n}$ which implicitly depends on $\mathbf{q}$ and the tilt-angle α.

3.2.8. Weighting

The difference between filtered back-projection (FBP) and back-projection (BP) is the additional filtering step. This filter is required because BP overestimates the contribution of the low spatial frequencies. In our implementation of direct Fourier reconstruction, we weight the data with a ramp filter before putting it on a tilted plane in the Fourier domain of the reconstruction volume.

We also added a filtering step to Alg. 1 by weighting $\hat{I}_\alpha(\mathbf{q})$ with $\|\mathbf{q}\|$, thereby ignoring the minute difference between $\|\mathbf{q}\|$ and $\|(q, -\frac{1}{2}n\lambda q^2)^T\|$, respectively q and $q\sqrt{(1+1/4\,n^2\lambda^2q^2)}$. More sophisticated filtering methods can be implemented in a similar manner. For our simulations, however, a simple weighting with $\|\mathbf{q}\|$ was sufficient.

3.3. *Results*

In this article two new CTF correction methods are introduced: 3DCTF correction (Eq. (3.10)) and TCTF correction (Eq. (3.19)). In this section we will use simulated images to test whether the presented methods produce the desired results and assess the improvement in speed.

To simulate projections, a specimen model is required. Since our goal is to evaluate the different correction methods, the specimen model was kept as simple as possible. We used spherical, phase-contrast objects with a diameter of 2.6 nm. The spheres were simulated using an error-function (normal cumulative distribution) of the radial coordinate with a standard deviation $\sigma = 0.18$ nm. This sharp edge was used to show the differences in step-response of the different correction methods.

The projections were simulated using the 3DCTF since this method accurately models the spatially varying defocus of a 3D specimen [3]. We simulated a tilt-series of 141 projections (tilt-angles were uniformly distributed between $\pm 70°$). The simulated microscope parameters were $\Delta f = -1\,\mu$m, $\lambda = 2.5$ pm (HT = 200 kV), $C_s = 2$ mm. No envelopes were simulated and no noise was added. The 3D (x, y, z) model was reduced to 2D (x, z) slices perpendicular to the tilt-axis. Without loss of generality, this reduces the processing time while maintaining the full complexity of the problem.

Simulations were performed using Matlab (The MathWorks, USA) and the DIPimage toolbox (TU Delft, The Netherlands; `http://www.diplib.org`), a publicly available image-processing toolbox for Matlab. The algorithms introduced in the previous section and all the scripts that were used to generate the following results can be found online (`http://www.diplib.org/add-ons`).

3.3.1. Comparison of CTF correction methods

To compare the two newly presented CTF correction methods with the traditional regular CTF correction, we simulated projections and processed these images using the three different correction methods. In order to differentiate between the spatially varying behavior of the correction methods, we placed phase-contrast spheres at three different locations. One sphere was placed at the center of the specimen ($\tilde{x} = \tilde{z} = 0$); this coincides with the tilt-axis. At this location the defocus does not change and we expect very similar results for the different correction methods. One sphere was displaced horizontally from the center ($\tilde{x} = 500$ nm, $\tilde{z} = 0$). At this location the defocus is correctly modeled by the TCTF and the 3DCTF. Finally, one sphere was displaced vertically from the center ($\tilde{x} = 0$, $\tilde{z} = 250$ nm), this defocus is only modeled correctly by the 3DCTF.

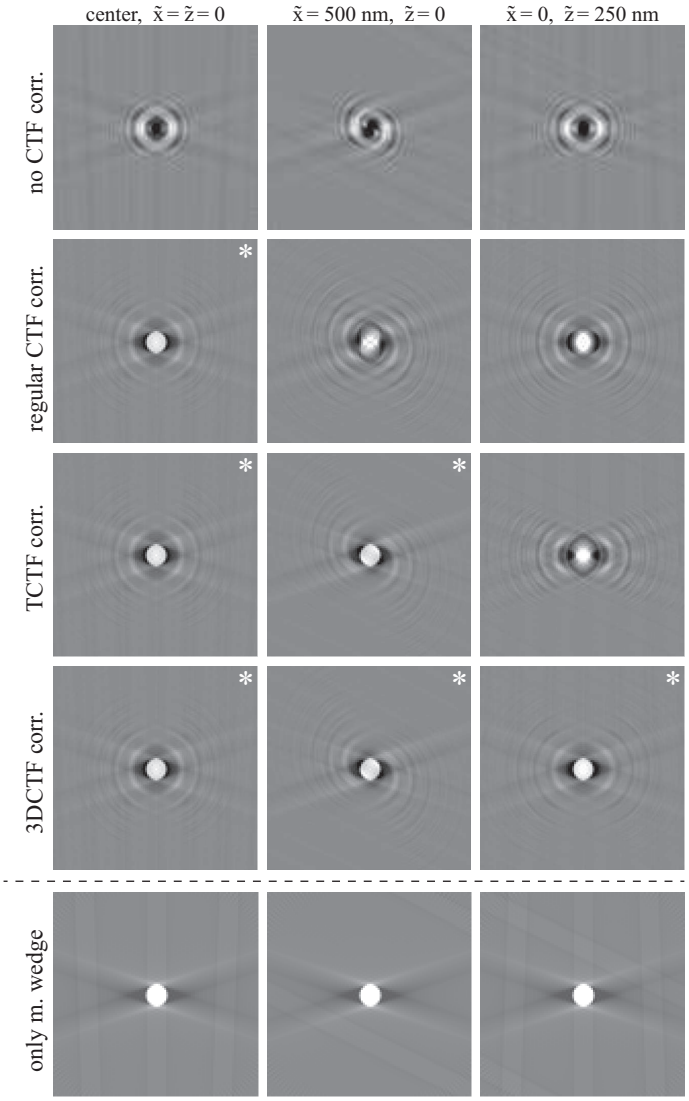


Figure 3.3. Phase contrast test sphere (diameter = 2.6 nm) after tomographic reconstruction without CTF correction and for three different CTF correction methods, at three different locations. A Wiener filter was used for all three correction methods (regularization parameter $b = 0.2$). The artifacts present in the images marked with an $*$ are primarily due to the missing wedge and the regularization of the Wiener filter. $\Delta f = -1$ µm, $\lambda = 2.5$ pm, $C_s = 2$ mm. The final row shows reconstructions with only missing wedge artifacts, i.e. no CTF was used to simulate the projections and no CTF correction was applied.

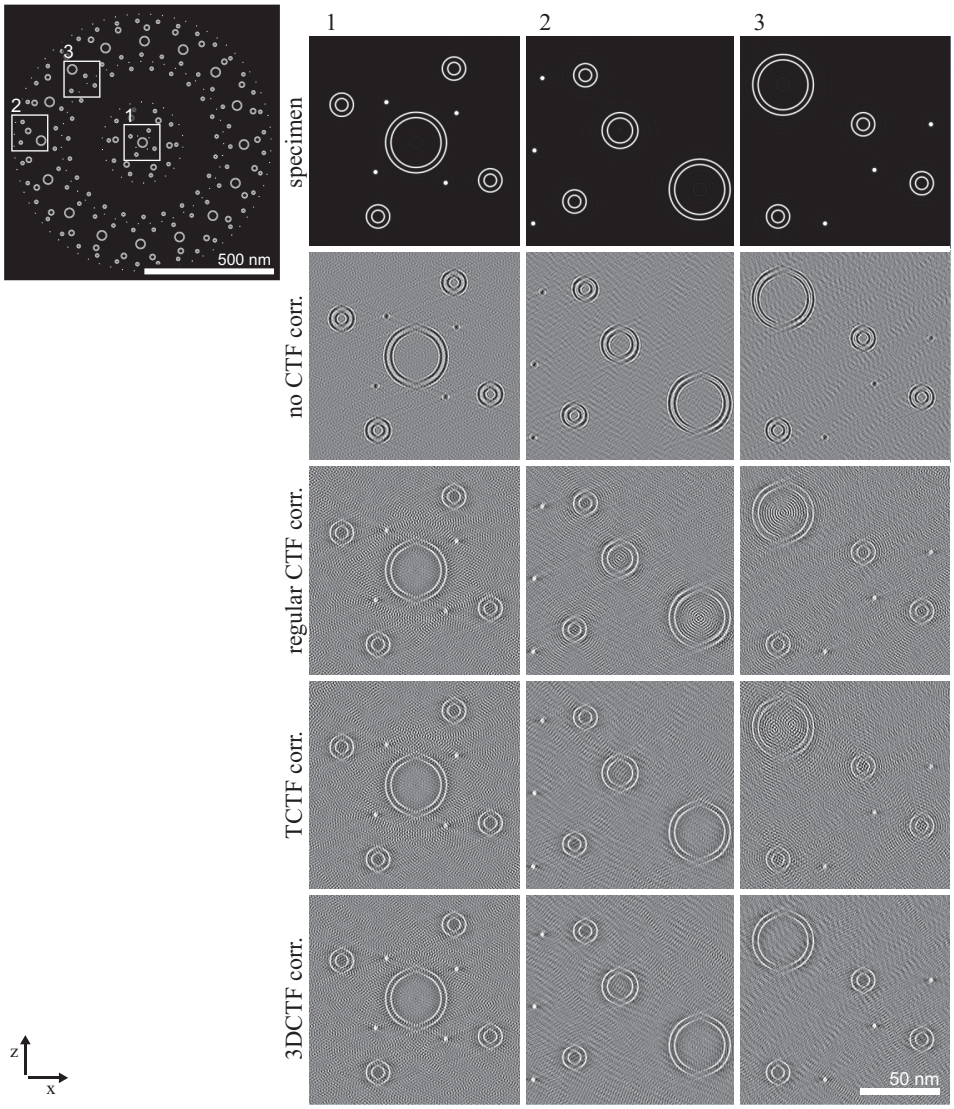


Figure 3.4. The influence of different CTF correction methods on a complex test object, composed of small solid spheres and spherical shells of varying diameter, which are placed around the tilt-axis. All images presented are slices in the (x, z) plane. For three different locations zoomed-in results are shown. The specimen (far left) is used to simulate projections using the 3DCTF method. The second column shows the tomographic reconstruction without CTF correction. In the last three columns the effect of the three different CTF correction methods after reconstruction is shown. Around the tilt-axis (location 1) the three correction methods give similar results. Away from the center on the $z = 0$ plane (location 2) only TCTF correction and 3DCTF correction give satisfactory results. Only 3DCTF correction gives a good reconstruction above the tilt-axis (location 3; when $|z| > 0$). $\Delta f = -1$ µm, $\lambda = 2.5$ pm, $C_s = 2$ mm

We used a Wiener filter for all three correction methods. The regularization coefficient in Eq. (3.13) was set to $b = 0.2$. The Fourier series of the filter was truncated at $N = 13$ (in Eqs. (3.10), (3.19) and (3.21) respectively).

Fig. 3.3 shows a comparison of the three different CTF correction methods: regular CTF correction (Eq. (3.21)), TCTF correction (Eq. (3.19)) and 3DCTF correction (Eq. (3.10)). The first two methods process the projections, which are then used for tomographic reconstruction. We used Direct Fourier reconstruction as tomographic reconstruction method. In contrast, 3DCTF correction is an integral part of the tomographic reconstruction.

The results in Fig. 3.3 clearly show that the three correction methods behave as predicted. The object in the center (left column in the figure) is reconstructed equally well for all three methods. The artifacts that are still visible can be attributed to the missing wedge and the regularization of the Wiener filter.

As expected, regular CTF correction does not give satisfactory results for objects displaced horizontally or vertically. The vertically displaced object is less affected, since the absolute distance to the center is smaller.

TCTF correction accurately accounts for the defocus differences of the horizontally displaced object but not for the vertically displaced object. The vertically displaced object suffers from more artifacts after TCTF correction than after regular CTF correction. This is caused by the fact that at high tilt-angles ($\pm 70^\circ$), the error in defocus made by the TCTF method ($\tilde{z}/\cos(\alpha)$) is almost nine times larger than the error made by regular CTF correction ($\tilde{z}\cos(\alpha)$). However, this effect only occurs close to the center of the specimen. Furthermore, the horizontal displacement is usually of much larger influence than the vertical displacement.

Finally, in the third row of Fig. 3.3 one can clearly see that the 3DCTF correction method correctly accounts for all the defocus differences throughout the specimen.

The results shown in Fig. 3.3 use a test object which is as simple as possible. This helps in identifying the differences between the CTF correction methods. Nevertheless, due to the simplicity of the test object the practical implications are difficult to discern. Fig. 3.4 shows results similar to Fig. 3.3 but for a more complex test object.

3.3.2. Different inverse filters

The CTF correction methods described in this article can employ different inverse filters by changing the weighting coefficients $a(n)$ in Eqs. (3.10), (3.19) and (3.21). Fig. 3.5 shows the results of three different inverse filters. The bright halo (visible in Fig. 3.5A) and the small dip around the center of the object (visible in Fig. 3.5B)

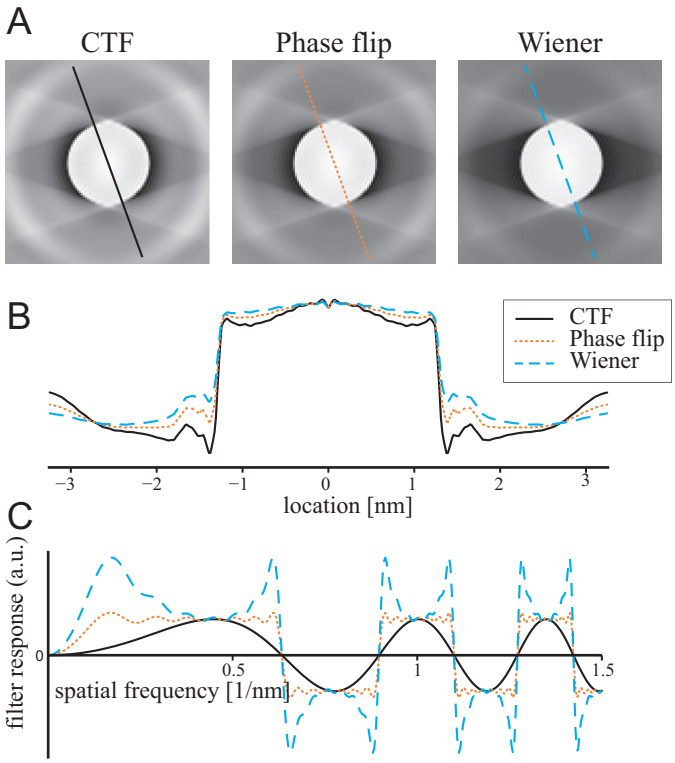


Figure 3.5. Differences between different inverse filters, for a test sphere in the middle of the specimen. The different inverse filters and corresponding coefficients are given in Eqs. (3.11), (3.12) and (3.13). (A) Reconstruction slice after tomographic reconstruction. (B) Intensity profiles on a tilted line (70°) through the middle of the test sphere. (C) Corresponding inverse filters used. The truncation of the Fourier series ($N = 13$) results in some oscillations on the filter response.

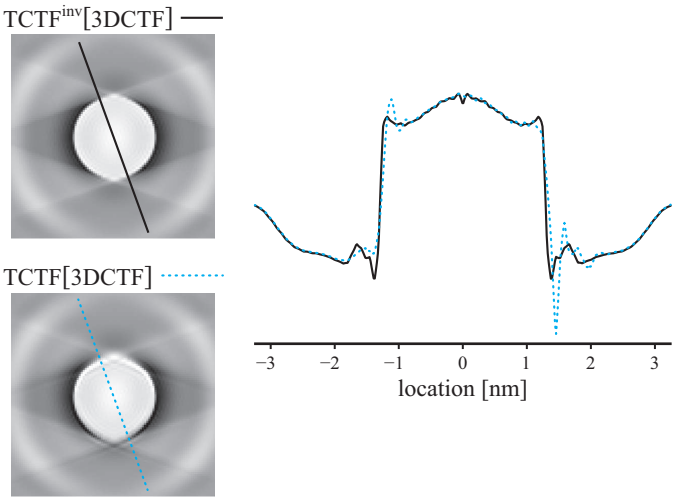


Figure 3.6. The TCTF forward model (Eq. (3.16)) can be used as an approximation to the corresponding TCTF correction method (Eq. (3.17)). At high resolutions, the TCTF forward model suffers from overshoot artifacts.

are noticeably reduced when using a phase-flipping filter instead of the CTF filter. Using a Wiener filter reduces these artifacts even further.

The Fourier series of the phase-flipping and Wiener filter was truncated at $N = 13$. This can be seen in Fig. 3.5C; some oscillations are visible in the filter responses. However, we did not see any visual difference in the reconstruction result when increasing N beyond 13. Nevertheless, we measured the difference between a truncated Fourier series of the filter and the original filter; these results were in agreement with Eq. (3.14).

3.3.3. Implementation details of TCTF correction

Earlier in this article we made an effort to point out the difference between the forward model of the TCTF (Eq. (3.16)) and the corresponding TCTF correction method (Eq. (3.19)). This difference has been noted in several studies [2, 14], but these studies concluded that this difference only occurs at resolutions too high to be of practical interest. We agree with this statement, but we think it is worth mentioning that our solution for TCTF correction is quite elegant and not more complicated or obscure than its forward model counterpart.

Fig. 3.6 shows that using the TCTF forward model as a correction method does not give a correct result at high resolutions ($> 2\ \mathrm{nm}^{-1}$). This can be seen from the overshoot around the edges and the asymmetry in the reconstruction. It is clear

Table 3.1. Average run-times for different CTF correction methods. The Fourier series of the TCTF correction method Eq. (3.19) (Alg. 2) and the 3DCTF correction method Eq. (3.10) (Alg. 1) were truncated at different values of N (as given by Eq. (3.15)). The two reference methods are a TCTF correction method using Tikhonov regularization and a 3DCTF correction method using a real space implementation. Regular CTF correction (Eq. (3.21)) and direct Fourier reconstruction are included for comparison. Notice that some values are in minutes (m), some in hours (h) and some in days (d).

Image size (pixel)	1024^2	2048^2	4096^2
Method	Runtime		
Regular CTF corr.	1.0 (m)	3.2 (m)	11.6 (m)
TCTF corr, N=7	10 (m)	37 (m)	2.4 (h)
TCTF corr, N=15	20 (m)	1.2 (h)	4.7 (h)
TCTF corr, Tikhonov	9.2 (h)	5.1 (d)	69 (d)
Direct Fourier rec.	22 (m)	2.7 (h)	22 (h)
3DCTF corr, N=7	55 (m)	4.9 (h)	31 (h)
3DCTF corr, N=15	93 (m)	7.5 (h)	42 (h)
3DCTF corr, real space	20 (h)	7.2 (d)	62 (d)

Values are extrapolated from run-time averages of 100 image slices. Each slice is projected at 141 tilt-angles. Standard deviations were all below 5% with the exception of the first row (20%, 13%, 3%) and second row (6%, 3%, 4%).

that TCTF correction as described by Eq. (3.19) and Alg. 2 reduces the overshoot significantly and gives a symmetric result.

3.3.4. Speed improvement

The methods presented in this article were designed to be faster than previously presented global methods. We implemented the two methods described here (Alg. 1 and Alg. 2) and also two reference methods. Comparing methods based on their computation time can be difficult due to implementation differences. Nevertheless, the results give an indication of the relative computation times of the different methods as the differences span several orders of magnitude.

All methods were implemented in Matlab (The MathWorks, USA). Without loss of generality, the 3D (x, y, z) specimen was reduced to a set of 2D (x, z) slices perpendicular to the tilt-axis. For each 2D slice, a tilt-series of 141 one-dimensional projections with a maximum tilt-angle of $\pm 70°$ was simulated. Processing times were evaluated on a PC equipped with an Intel X5550 quad-core 2.66 GHz processor and 24 GB RAM. The processing times of the one-dimensional projec-

tions were scaled to resemble the computation of a full tilt-series where the 2D projections have a square image size.

We could not find a suitable candidate for benchmarking the TCTF correction method presented here (Alg. 2). Alternative methods were either fundamentally different (i.e. strip/tile-based [5–7] instead of processing the entire image) or too slow [15] to provide a meaningful comparison (for more details see the discussion section). Therefore, we implemented Tikhonov regularization, a well-known linear regularized inversion method. See Appendix 3.A for a very brief description of the regularization. Table 3.1 shows that the method presented in this article clearly outperforms the reference method, especially for large image sizes.

We compared our 3DCTF correction as described in Alg. 1 to a method similar to the one presented in Jensen and Kornberg [8] and Kazantsev *et al.* [10]. This method evaluates Eq. (3.6) for each tilt-angle in real-space before summing up contributions from all angles. Table 3.1 shows that the reference method (real-space 3DCTF correction) is up to 48 times slower than Alg. 1 ($N = 7$, image size $= 4096^2$). The computation times scale almost linear in the number of iterations (N) used for the series expansion of the filter. For a very low number of iterations ($N = 1$), the computation times are close to those of direct Fourier reconstruction (not shown); this is due to the similarities between both methods.

3.3.5. Global vs. local

In the previous subsection we discarded local CTF correction methods as a suitable candidate for a speed comparison. Local approaches, however, are currently the standard in the field. For example the program `CTFPHASEFLIP` [6] as incorporated in the widely used software package IMOD [16] (`http://bio3d.colorado.edu/imod/`), uses a local approach. Here we provide an example where there is a distinct difference between a local and our global method.

Fig. 3.7 shows the difference between the global TCTF correction method as presented in Alg. 2 and the local `CTFPHASEFLIP` program. In this example the local method produces artifacts because the size of the strips is of the same order as the size of the point spread function corresponding to the CTF. Even though the tilt-angle is small ($\alpha = 10°$) and the influence of the defocus gradient is negligible, artifacts appear simply because the image is divided into strips. It is clear that similar artifacts do not occur for all setups and depend on the particular setting of the parameters. Nevertheless, for a certain width of the strips, these artifacts will appear eventually when the resolution is increased sufficiently. The solution would be to increase the width of the strips but this drastically increases the computation time and is currently not possible using `CTFPHASEFLIP`.

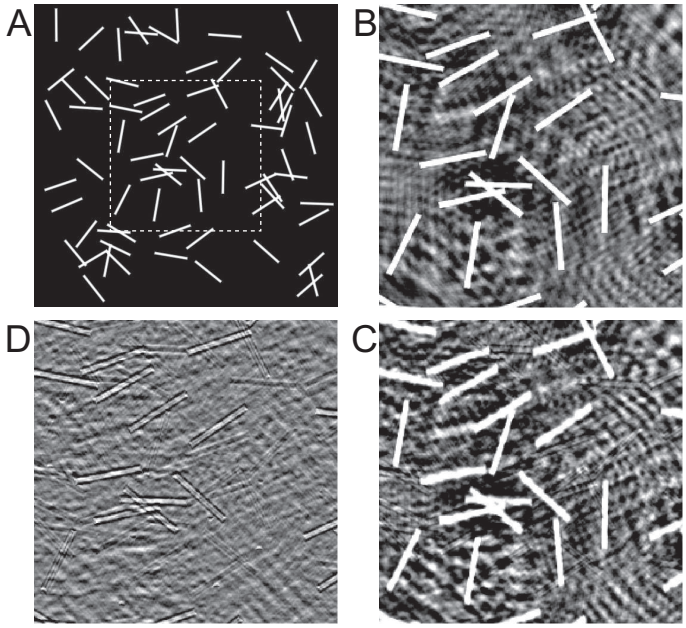


Figure 3.7. Difference between local and global TCTF correction methods. The test image consists of randomly oriented rods of 5.6 nm length. (A) (x, y) projection without CTF. (B) projection after global TCTF correction Alg. 2. (C) projection after (local) tilted phase flipping using `CTFPHASEFLIP`, which is part of IMOD. (D) difference between (B) and (C). The tilt-axis is horizontal. In contrast to horizontal rods, vertical rods are reconstructed the same in (B) and (C) and are therefore not visible in the difference image (D). $\Delta f = -2$ µm, $\lambda = 2.5$ pm, $C_s = 0$ mm, tilt-angle $\alpha = 10°$, pixel size = 1 Å. The parameters for `CTFPHASEFLIP` were set such that `stripPixelNum` = 128, `interPixelNum` = 32.

3.4. *Discussion*

The motivation for our research on spatially varying CTF correction was primarily a need for speed as today's global correction algorithms are too slow for practical application. Table 3.1 shows a significant improvement for both new correction methods compared to their reference methods. It is interesting to note, considering that TCTF correction is followed by direct Fourier reconstruction, that for low values of N the difference in computation time between TCTF correction and 3DCTF correction becomes relatively small.

How large N needs to be, can be estimated using Eq. (3.14), given a certain desired accuracy. Under typical noise conditions, the Wiener filter requires a fair amount of regularization, which results in a smooth filter response. Therefore, $N = 7$ is usually sufficient for a Wiener filter. For a phase-flipping filter, more orders are required due to the sharp transitions in the filter response. Especially at low spatial frequencies, the approximation of the phase-flipping filter can have a considerable effect.

The presented CTF correction methods assume that the defocus at the tilt-axis does not change throughout the tilt-series. Recent studies, however, showed that this assumption is not realistic and that the variation of the defocus needs to be accounted for [6, 7]. Nevertheless, the presented CTF correction methods treat each projection independently, accounting for a known varying defocus is therefore a trivial modification.

A popular alternative to filtered back-projection (FBP) is the simultaneous iterative reconstruction technique (SIRT). Even though we only considered FBP in this article, the choice of tomographic reconstruction method is not essential. Especially for the TCTF correction method, the CTF correction step is simply a pre-filtering step after which any reconstruction method can be used. We are currently investigating the possibilities for and influence of 3DCTF correction in an iterative reconstruction scheme.

The decision to use Tikhonov regularization as reference method in Table 3.1 was due to the lack of alternatives. In the past few years a number of strip-based TCTF correction methods have been presented [5–7]. These methods rely on the subdivision of the image into strips or tiles of approximately equal defocus. Benchmarking against these strip/tile-based methods is difficult due to the parameters that can and must be tweaked (i.e. strip width and number of strips). Depending on the implementation, we expect the computation time of strip-based methods to be of the same order of magnitude as our TCTF correction method. For example, the computation time of `CTFPHASEFLIP` varies between much shorter and slightly longer than the computation times of our global TCTF correction

approach. The speed of this local method strongly depends on the particular setting of its parameters, which also dictate the obtainable accuracy.

Furthermore, the example in Fig. 3.7 shows that the tweaking of parameters in strip-based CTF correction methods can have a significant effect on the accuracy of the result. This is one of the reasons why one should prefer global methods over local (strip-based) methods if possible. In addition, an alternative to strip-based methods is also of scientific interest in the developing field of spatially varying CTF correction.

We also considered the TCTF correction approaches as presented in Philippsen *et al.* [2] and Winkler and Taylor [15] to benchmark against. These two approaches are based on the direct inversion of the TCTF operator, without regularization and with truncated singular value decomposition respectively. We regard methods without regularization to be unsuitable for realistic scenarios in which noise necessitates the use of regularization. Our experience with truncated singular value decomposition is that for large image sizes (i.e. 4096^2) it is too slow to be practical.

By changing the coefficients $a(n)$, the presented CTF correction methods can employ different inverse filters. This makes it possible to use phase-flipping, a popular inverse filter. However, one can now also think of new inverse filters simply by choosing different weights $a(n)$. Furthermore, it is even possible to output reconstruction volumes for each order n and assign the weights dynamically when summing the volumes.

3.A. *Tikhonov regularization*

We implemented Tikhonov regularization to benchmark the computation times of our TCTF correction method described in Alg. 2. The TCTF forward model Eq. (3.16) is rewritten as a set of linear equations: $\mathbf{Ax} = \mathbf{b}$, where the vectors $\mathbf{x}$ and $\mathbf{b}$ are discretized representations of the projected potential $\hat{V}_{\text{proj},\alpha}(\mathbf{q})$ and image intensity $\hat{I}_\alpha(\mathbf{q})$ respectively. The objective of Tikhonov regularization is to find $\mathbf{x}$ which minimizes $\|\mathbf{Ax}-\mathbf{b}\|^2 + \|\lambda\mathbf{x}\|^2$, where λ is the regularization parameter and $\|\cdot\|$ is the Euclidean norm. This inverse problem has an explicit solution:

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A} + \lambda^2\mathbf{I}\right)^{-1}\mathbf{A}^T\mathbf{b}, \tag{3.22}$$

where $\mathbf{I}$ is the identity matrix.

References

[1] G. J. Jensen and A. Briegel, *How electron cryotomography is opening a new window onto prokaryotic ultrastructure,* Current Opinion in Structural Biology **17**, 260 (2007).

[2] A. Philippsen, H.-A. Engel, and A. Engel, *The contrast-imaging function for tilted specimens,* Ultramicroscopy **107**, 202 (2007).

[3] L. M. Voortman, S. Stallinga, R. H. M. Schoenmakers, L. J. van Vliet, and B. Rieger, *A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM,* Ultramicroscopy **111**, 1029 (2011).

[4] V. Mariani, A. D. Schenk, A. Philippsen, and A. Engel, *Simulation and correction of electron images of tilted planar weak-phase samples,* Journal of Structural Biology **174**, 259 (2011).

[5] J. J. Fernández, S. Li, and R. A. Crowther, *CTF determination and correction in electron cryotomography,* Ultramicroscopy **106**, 587 (2006).

[6] Q. Xiong, M. K. Morphew, C. L. Schwartz, A. H. Hoenger, and D. N. Mastronarde, *CTF determination and correction for low dose tomographic tilt series,* Journal of Structural Biology **168**, 378 (2009).

[7] G. Zanetti, J. D. Riches, S. D. Fuller, and J. A. G. Briggs, *Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging,* Journal of Structural Biology **168**, 305 (2009).

[8] G. J. Jensen and R. D. Kornberg, *Defocus-gradient corrected back-projection,* Ultramicroscopy **84**, 57 (2000).

[9] Y. Wan, W. Chiu, and Z. H. Zhou, *Full contrast transfer function correction in 3D cryo-EM reconstruction,* in *IEEE Proceedings of ICCCAS 2004*, Vol. 2 (2004) pp. 960–964.

[10] I. G. Kazantsev, J. Klukowska, G. T. Herman, and L. Cernetic, *Fully three-dimensional defocus-gradient corrected backprojection in cryoelectron microscopy,* Ultramicroscopy **110**, 1128 (2010).

[11] L. Reimer and H. Kohl, *Transmission Electron Microscopy*, 5th ed. (Springer-Verlag, Berlin, 1984).

[12] B. F. Logan and L. A. Shepp, *Optimal reconstruction of a function from its projections,* Duke Math. J **42**, 645 (1975).

[13] L. Greengard and J.-Y. Lee, *Accelerating the nonuniform fast Fourier transform,* SIAM Review **46**, 443 (2004).

[14] V. Mariani, *Transfer of tilted sample information in transmission electron microscopy*, Ph.D. thesis, University of Basel (2010).

[15] H. Winkler and K. A. Taylor, *Focus gradient correction applied to tilt series image data used in electron tomography,* Journal of Structural Biology **143**, 24 (2003).

[16] J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, *Computer visualization of three-dimensional image data using IMOD,* Journal of Structural Biology **116**, 71 (1996).

4

When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM

The surface of the earth
is the shore of the cosmic ocean.
CARL SAGAN

Mountains are not Stadiums where I satisfy my ambition to achieve,
they are the cathedrals where I practice my religion.
ANATOLI BOUKREEV

Miloš Vulović*
Lenard M. Voortman*
Lucas J. van Vliet
Bernd Rieger

Ultramicroscopy 136 (2014), pp. 61–66

*These authors contributed equally to this work

4

Abstract

The projection assumption (PA) and the weak-phase object approximation (WPOA) are commonly used to model image formation in cryo-electron microscopy. For simulating 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 by inspecting both approximations separately to derive their respective conditions of applicability. The thick-phase grating approximations (TPGA) 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 TPGA) given different interaction potentials using exit wave simulations. The conditions of applicability for the models are based on two measures, 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.

4.1. *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, i.e. 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 [1] and the PA assumes that the exit wave from the specimen can be computed via the projected IP of the whole specimen [2]. Both approximations rely on the small angle approximation [3] 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) [4–9].

These approximations have, of course, limitations as they cannot account for, e.g. the curvature of the Ewald sphere or multiple scattering events [10]; 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 [2] 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 appropriate slice thickness. The multislice approach has been rarely used in cryo-EM [11, 12], 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 applicabilities.

The thick-phase grating approximation (TPGA) was introduced in HREM of perfect crystals in 1962 [13, 14], but to the best of our knowledge, it has not

received much attention since. They provided rough indications for the validity of various approximations depending on the thickness and atomic number of the crystals. In Gibson [15], the multislice (MS) method was used to discuss the breakdown of the PA in HREM of amorphous samples and the consequence of the breakdown on the measurement of microscope parameters, especially spherical aberration. In Jap and Glaeser [16] the ranges of validity for the WPOA for organic crystals are discussed in terms of thickness, resolution and incident electron energy. Although they use a quantitative measure (dissimilarity factor [17]), only single scattering approximations were considered.

Here, we introduce TPGA 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.

4.2. *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 (see e.g. Vulović [18]). 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 same time, which guarantees independence of all incident electrons. Below we will shortly recapitulate the formulae commonly used in HREM [2, 14].

4.2.1. Small angle approximation

The stationary one-body Schrödinger equation for the electron wave function in a closed system is given by

$$\left(-\frac{\hbar^2}{2m}\nabla_{\boldsymbol{r}}^2 + e\mathcal{V}(\boldsymbol{r})\right)\psi_{\mathrm{e}}(\boldsymbol{r}) = E_{\mathrm{e}}\psi_{\mathrm{e}}(\boldsymbol{r}), \tag{4.1}$$

where $-\hbar^2/(2m)\nabla_{\boldsymbol{r}}^2$ is the Hamiltonian of the incident high-energy electron, which in this case represents its kinetic energy, $\mathcal{V}(\boldsymbol{r})$ the interaction potential, $\hbar$ the reduced Planck constant, m the relativistic mass of the electron, e the electron charge, $\boldsymbol{r} = (x, y, z) = (\boldsymbol{\rho}, z)$ 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 $\psi_e(\boldsymbol{r})$ 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. $\psi_e(\boldsymbol{r}) = \Psi(\boldsymbol{r})e^{ikz}$, with the wave vector $k = 2\pi/\lambda = \sqrt{2mE_e}/\hbar$, and λ the wavelength. Now it follows from Eq. (4.1)

$$\left(\nabla^2_{\boldsymbol{\rho}} + \partial^2_z + 2ik\partial_z - \frac{2me}{\hbar^2}V(\boldsymbol{r})\right)\Psi(\boldsymbol{r}) = 0\,. \tag{4.2}$$

Given the assumptions that the energy of the incident electron is high and that Ψ varies slowly with z, it holds that $|\partial^2_z\Psi| \ll |k\partial_z\Psi|$ 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

$$\partial_z\Psi(\boldsymbol{r}) = \left(\frac{i\lambda}{4\pi}\nabla^2_{\boldsymbol{\rho}} + i\sigma V(\boldsymbol{r})\right)\Psi(\boldsymbol{r})\,. \tag{4.3}$$

Taking the 2D Fourier-transform in $\boldsymbol{\rho} = (x, y)$ we get our common starting point for all further approximations

$$\partial_z\mathcal{F}_{\boldsymbol{\rho}}[\Psi] = -i\lambda\pi q^2\mathcal{F}_{\boldsymbol{\rho}}[\Psi] + i\sigma\mathcal{F}_{\boldsymbol{\rho}}[V\Psi]\,, \tag{4.4}$$

in which the Fourier-transform is defined as $\mathcal{F}_{\boldsymbol{\rho}}[f(\boldsymbol{\rho})](\mathbf{q}) = \int f(\boldsymbol{\rho})e^{-2\pi i\boldsymbol{\rho}\mathbf{q}}d\boldsymbol{\rho}$.

4.3. *Bounds to PA and WPOA*

To solve Eq. (4.4) 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 [19]. 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 V it has a constant magnitude and zero phase. The magnitude of the incident 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 [10]. Because scattering by a constant 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 [19].

4.3.1. Projection assumption

When the specimen is sufficiently thin, the projection assumption (PA) is commonly used [2]. Then the propagation term of Eq. (4.3) is small compared to the interaction term, i.e. $|\frac{i\lambda}{4\pi}\nabla^2_{\boldsymbol{\rho}}\Psi| \ll |i\sigma\mathcal{V}\Psi|$. From Eq. (4.3) it follows

$$\partial_z\Psi(\boldsymbol{r}) = i\sigma\mathcal{V}(\boldsymbol{r})\Psi(\boldsymbol{r}) \Rightarrow \Psi = \exp\left\{i\sigma\int_{-\infty}^{z}\mathcal{V}\,dz'\right\}, \tag{4.5}$$

which leads to the exit wave

$$\Psi_{\text{exit}} = \exp\{i\sigma\mathcal{V}_z\}\,, \tag{4.6}$$

with the projected potential $\mathcal{V}_z = \int_{-\infty}^{\infty}\mathcal{V}\,dz$. The validity of the PA was addressed by Ishizuka and Uyeda [20]. 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 $F = \Delta r^2/(\lambda\Delta z)$ [21], 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

$$q \ll \sqrt{1/(4\lambda\Delta z)}. \tag{4.7}$$

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* or *phase-grating approximation* [3, 14].

4.3.2. 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. (4.6) can be approximated by

$$\Psi_{\text{exit}} = 1 + i\sigma\mathcal{V}_z + \mathcal{O}(\sigma^2\mathcal{V}_z^2)\,. \tag{4.8}$$

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. (4.5) giving $\partial_z\Psi = i\sigma\mathcal{V}$. We will refer to Eq. (4.8) as PA+WPOA.

4.3.3. 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 $\mu = \sigma\mathcal{V}_z$,

$$e^{i\mu} = 1 + i\mu + \mathcal{O}(\mu^2)\,. \tag{4.9}$$

The relative residual in orders m or higher is given by

$$p(m,\mu) = e^{-\mu}\sum_{n=m}^{\infty}\frac{\mu^n}{n!}\,, \tag{4.10}$$

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

$$\sigma\mathcal{V}_z < 0.36\,. \tag{4.11}$$

We will use this condition for applying the WPOA.

Note that Eq. (4.10) 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 [22]. 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 $\mathcal{V}_{\text{ice}}$ can be neglected in the quasi-kinematic approach. Therefore, the condition given by Eq. (4.11) 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. (4.4) can be solved as follows

$$\begin{aligned}&\partial_z\mathcal{F}_{\boldsymbol{\rho}}[\Psi] = -i\lambda\pi q^2\mathcal{F}_{\boldsymbol{\rho}}[\Psi] + i\sigma\mathcal{F}_{\boldsymbol{\rho}}[\mathcal{V}]\\ &\Rightarrow \mathcal{F}_{\boldsymbol{\rho}}[\Psi] = 1 + i\sigma\int_{-\infty}^{z} e^{-i\lambda\pi q^2 z'}\mathcal{F}_{\boldsymbol{\rho}}[\mathcal{V}]\,dz'\end{aligned}$$

$$\mathscr{F}_{\boldsymbol{\rho}}\left[\Psi_{\text{exit}}\right] = 1 + i\sigma \int \mathscr{V}(\boldsymbol{\rho}, z) e^{-2\pi i\left(\boldsymbol{\rho}\mathbf{q} + \frac{1}{2}\lambda q^2 z\right)} d\boldsymbol{r}$$

$$\Psi_{\text{exit}} = 1 + i\sigma \mathscr{F}_{\boldsymbol{\rho}}^{-1}\left[\mathscr{F}[V]\left(\mathbf{q}, \lambda q^2/2\right)\right]. \tag{4.12}$$

Here $\mathscr{F}[V]$ is the 3D Fourier transform of the potential evaluated at coordinate $(\mathbf{q}, \lambda q^2/2)$, with $\mathbf{q} = (q_x, q_y)$. Computing the 3D Fourier-transform sampled on the parabola $(\mathbf{q}, \lambda q^2/2)$ can be done accurately and fast, as in Voortman *et al.* [23].

4.3.4. Thick-phase grating approximation

The limitations of the PA and WPOA can be overcome by the thick-phase grating approximation (TPGA) [13, 14]. Initially developed for perfect crystals with respect to both diffraction and imaging, the TPGA applied to cryo-EM represents the fourth combination resulting from the PA and WPOA and gives the following forward model

$$\Psi_{\text{exit}} = \exp\left\{i\sigma \mathscr{F}_{\boldsymbol{\rho}}^{-1}\left[\mathscr{F}[V]\left(\mathbf{q}, \lambda q^2/2\right)\right]\right\}. \tag{4.13}$$

The advantage of this combination is that in the limit of $F \gg 1$, Eq. (4.13) converges to Eq. (4.6) and in the limit of $\sigma\mathscr{V} \ll 1$, Eq. (4.13) converges to Eq. (4.12). This means we get the corresponding image models of PA or WPOA directly from the above equation in their respective limits.

The approximations of Eqs. (4.6), (4.8) and (4.12) were derived in a similar way in Treacy and Dyck [3]. Quantitative useful conditions for the validity of the approximations of Eqs. (4.11) and (4.7) are presented here. Their advantages will be demonstrated below.

4.4. *Results*

4.4.1. Hemoglobin

Here we investigate the validity of the PA and WPOA for *Lumbricus terrestris erythrocruorin* (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 [24]. All samples in this analysis are embedded in vitreous ice (ρ = 0.93 g/cm^3) which was modeled as a continuous medium. A detailed description of how the IP is constructed can be found elsewhere [12].

Fig. 4.1 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 middle of 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. (4.7), 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. (4.11), 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. 4.1 that the criteria for WPOA and PA are easier fulfilled for low-frequency potential maps (e.g. when the potential is blurred by beam-induced movements, CTF and/or the camera transfer). For comparison we show in Fig. 4.1 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 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.

4.4.2. Exit waves of a tubulin tetramer

For a tubulin tetramer (TT) constructed from PDBid 1SA0 ($\Delta z = 27$ nm) we show in Fig. 4.2A the computed phase of the exit wave after interaction with 80 keV electrons using the four approximations discussed above: PA, PA+WPOA, WPOA and TPGA. The potential map was sampled at 1 Å. In order to better visualize the effect of the approximations, we show in Fig. 4.2B the differences of the four exit waves with a reference. This reference is computed by a multislice (MS) approach inspired by Kirkland [2]. 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. 4.2B we observe that the TPGA is nearly identical to the MS reference, whereas the WPOA shows deviations mostly in the stronger

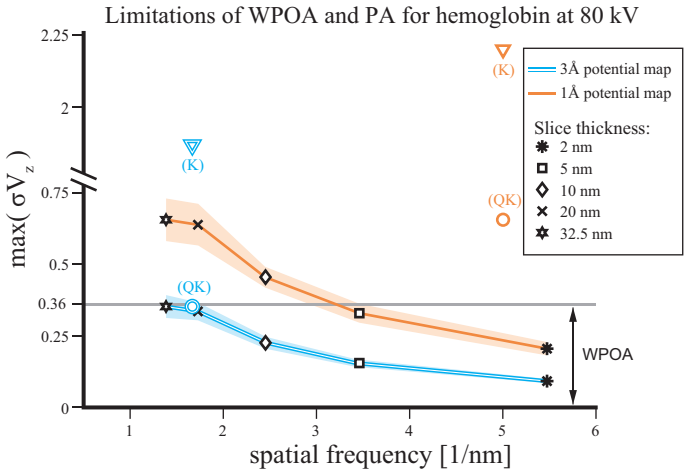


Figure 4.1. Validity of the PA and WPOA for hemoglobin interacting with 80 keV electrons for various slice thicknesses. The green and blue lines depict the boundary given by the Fresnel number $F = 1$ (compare Eq. (4.7)) for a potential map sampled at 1 Å and 3 Å respectively as a function of slice thickness. For each thickness, one slice is computationally extracted from the middle of the full IP (not to be confused with the multislice method). 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).

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.

4.4.3. Synthetic amorphous test specimen

We simulate exit waves of a synthetic test specimen using Eqs. (4.6), (4.12), (4.8) and (4.13) to study the validity of the predicted limits for the cases PA and WPOA. For the cases PA+WPOA and TPGA we want to investigate where the limits of the validity of these combined approximations lie. Our derived conditions of Eqs. (4.7) and (4.11) 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. (4.11) 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

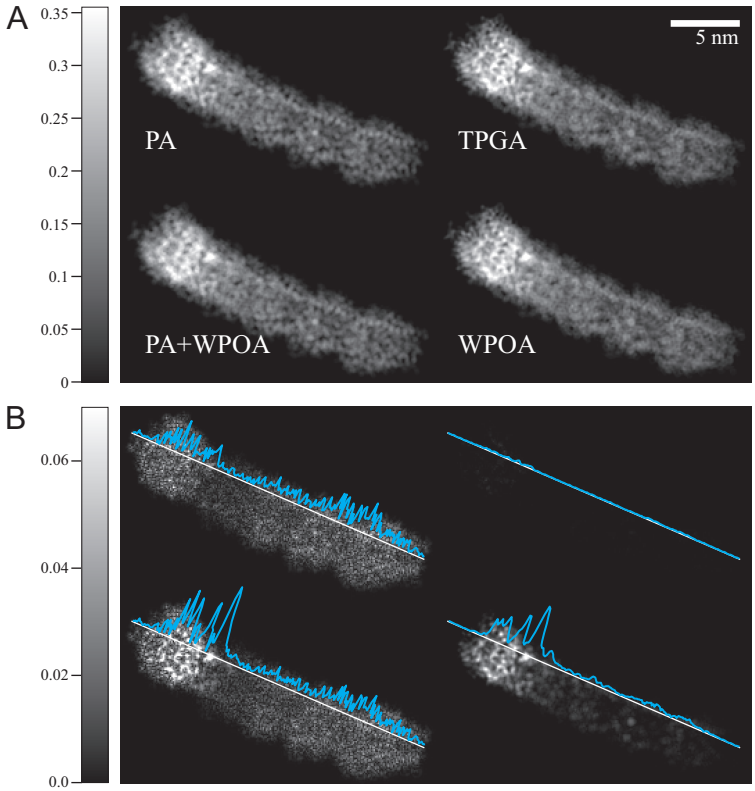


Figure 4.2. (A) Simulated phases of exit waves of a tubulin tetramer (HT = 80 kV) using the PA via Eq. (4.6), WPOA via Eq. (4.12), PA+WPOA via Eq. (4.8) and TPGA via Eq. (4.13). (B) Difference image of the exit waves in (A) and the exit wave computed with a MS approach. Graph inserts show the intensity along the line. The intensity scale bar indicate the phase of the exit wave subtracted by its mean.

max($|\sigma\mathcal{V}_z|$) as we did in section 4.4.1. This measure, however, depends for the synthetic test specimen on its spatial extent in (x, y). Therefore, we will examine the standard deviation std($\sigma\mathcal{V}_z$) for our synthetic test specimens. For potential maps of a macromolecule, the 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 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 std($\sigma\mathcal{V}_z$). Fig. 4.3A 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. (4.7) and Eq. (4.11). The combined models have boundaries which asymptotically approach the individual (WPOA and PA) approximations. In Fig. 4.3B a sketched version depicts the qualitative results in terms of regions where the different approximations hold.

In addition to the conditions that quantify the applicability for our synthetic specimen (Fig. 4.3), 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. 4.3A we show the characteristics of three macromolecules (ribosomal subunit from *haloarcula marismortui* - PDBid 1FFK, earth worm hemoglobin and TT) sampled at a 1 Å and 3 Å 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 std($\sigma\mathcal{V}_z$). As a solution we propose i) to retrieve the maximum spatial frequency by finding the 65$^{\text{th}}$ percentile of the 2D power spectrum of $\mathcal{V}_z$, ii) to obtain the thickness by first computing std($\mathcal{V}(\boldsymbol{\rho}, z)$) as a function of z, then finding the 2.5 and 97.5$^{\text{th}}$ 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 80$^{\text{th}}$ percentile of the histogram of $|\mathcal{V}_z - \langle\mathcal{V}_z\rangle|$. The corresponding values for the three macromolecules are depicted in Fig. 4.3A (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.

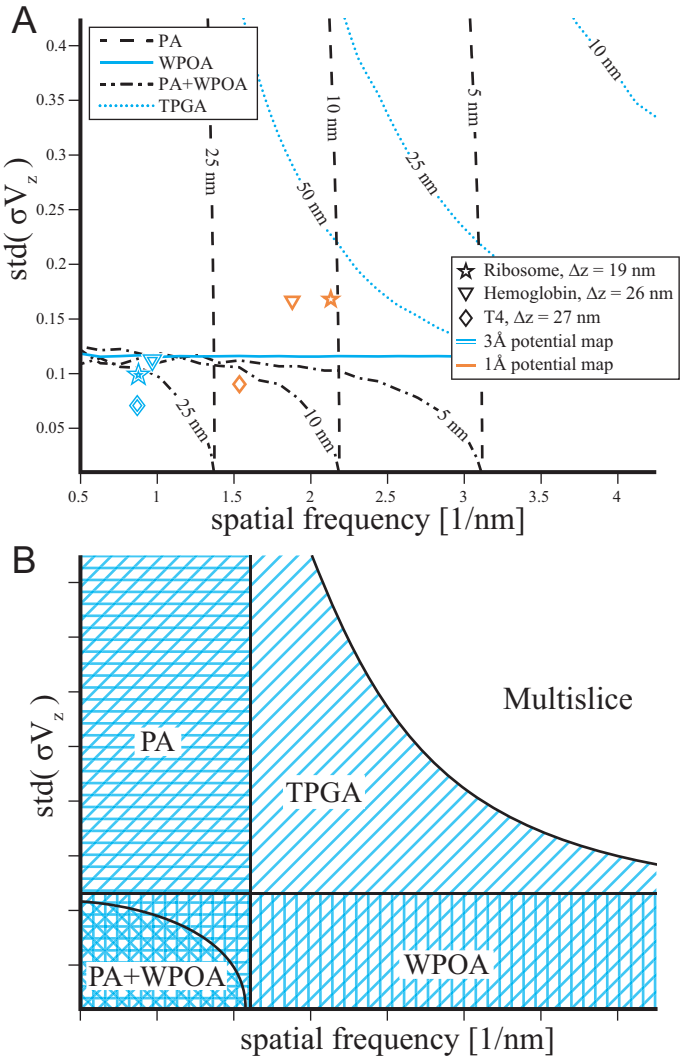


Figure 4.3. The applicability (at HT = 80 kV) of the PA, PA+WPOA, WPOA, and TPGA. (A) Boundaries for each approximation where different lines represent different specimen thickness. Lines indicate 10% MSE error of the respective approximations with a MS reference. Left/below the boundary the approximation holds for a particular thickness. Three protein-complexes potentials map (ribosome, hemoglobin, TT) sampled at 1 Å and 3 Å 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.

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. 4.3A.

Now we see in Fig. 4.3 that the three proteins sampled at 3 Å satisfy both the PA and WPOA and are close to the PA+WPOA boundary. When sampled at 1 Å the PA is not satisfied and only TT satisfies the WPOA. The hemoglobin results agree with those shown in Fig. 4.1. Judging from Fig. 4.2, which shows TT, we could conclude that the WPOA is violated for some parts of the molecule. In Fig. 4.3, however, we see that on average TT satisfies the WPOA. This apparent contradiction is due to the fact that Fig. 4.3 is computed from the average measure std($\sigma\mathcal{V}_z$), instead of max($|\sigma\mathcal{V}_z|$) in Fig. 4.1.

4.5. *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. In Cowley and Moodie [13] rough indications were provided for the validity of various forward approximations in HREM depending on the thickness and atomic number of the crystals. Depending on the magnitude of the error one considers acceptable, their boundaries (evaluated at one spatial frequency) are consistent with our criteria.

Here, in addition to the MS approach, the proposed criteria motivate the existence of four models describing the electron wave propagation through the specimen (PA, PA+WPOA, WPOA, and TPGA). The choice of the model depends on the strength, frequency content and thickness of the interaction potential map. Furthermore, the TPGA is applied here to realistic specimen models in 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 [11, 12] are related to the lower resolution of the structures studied by cryo-EM compared to HREM, and because it leads 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 PA+WPOA) in the first iteration of a typical iterative reconstruction scheme. As shown in Fig. 4.2, the simulations indicate that the direct TPGA approach gives nearly identical exit waves as a recursive MS calculation. We expect, however, that TPGA can be advantageous for 3D reconstructions due to its invertibility and the

possibility to utilize the non-uniform fast Fourier transform in the sampling of the Ewald sphere [23, 25].

For sake of completeness it should be mentioned that in materials science the projected charge density approximation (PCDA) [14, 26] is also used. The PCDA provides a linear expression between the intensity and projected charge density, but assumes that the CTF is parabolic. In cryo-EM, PCDA is usually violated owing to much higher defocus values compared to materials science. For example, at 300 kV and 1-2 μm defocus, the frequency up to which PCDA is satisfied (10 % error) would be around 3-4 nm^{-1}. This is not sufficiently accurate for SPA; for tomography, a higher defocus needs to be employed limiting the validity of the PCDA to frequencies lower than 6-12 nm^{-1}.

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 $\mathrm{std}(\sigma\mathcal{V}_z)$. 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 HT = 80 kV because for higher HT (shorter wavelength), the approximations given by Eqs. (4.11) and (4.7) 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. 4.1 and 4.3). 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. It is, however, clear that the electron-specimen interaction model needs to be accurate up to spatial frequencies as least as high as the resolution of the final image.

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 [1]. 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 [19], the mean potential cannot be neglected since it damps the magnitude of the exit wave. Here, only the relative difference between exit waves is analyzed. Therefore, damping due to the mean imaginary potential does not influence our findings. Nevertheless, the contrast in the final images does depend on inelastic scattering which requires a deeper investigation.

Although amorphousness of the vitreous ice would be apparent in the simulated images of the exit waves, we modeled the ice as a constant background. In Vulović *et al.* [12] it has been shown that for typical electron fluxes in cryo-EM ($< 100\ e^-/\text{Å}^2$), the influence of the solvent amorphousness in the final images can be neglected. Any averaging technique to enhance the SNR would blur out the structural noise.

As practical conclusions we find that, when simulating images at resolutions of ~ 5 Å, the applicability of the PA and WPOA needs to be re-considered. Here, the TPGA offers an excellent solution, as an alternative to the multislice approach. For tomograms with typical resolutions > 30 Å, the PA and WPOA are generally applicable. In single particle analysis, structures are being obtained up to 3.3 Å resolution [27] and are expected to improve further given advances in hardware developments such as direct electron detectors and phase plates. At those resolutions the PA and WPOA may be violated depending on the size of the macromolecule, whereas the TPGA again offers a good and fast alternative. The implementation of the exit wave simulations is freely available for non-commercial use upon request.

References

[1] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State,* 2nd ed. (Oxford University Press, 2006).

[2] E. J. Kirkland, *Advanced Computing in Electron Microscopy*, 2nd ed. (Springer Verlag, 2010).

[3] M. M. J. Treacy and D. v. Dyck, *A surprise in the first Born approximation for electron scattering,* Ultramicroscopy **119**, 57 (2012).

[4] C. O. S. Sorzano, R. Marabini, J. Velázquez-Muriel, J. R. Bilbao-Castro, S. H. W. Scheres, J. M. Carazo, and A. Pascual-Montano, *XMIPP: A new generation of an open-source image processing package for electron microscopy,* Journal of Structural Biology **148**, 194 (2004).

[5] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, and M. Schatz, *A new generation of the IMAGIC image processing system,* Journal of Structural Biology **116**, 17 (1996).

[6] T. R. Shaikh, H. Gao, W. T. Baxter, F. J. Asturias, N. Boisset, A. Leith, and F. Frank, *SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs,* Nature Protocols **3**, 1941 (2008).

[7] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, *EMAN2: An extensible image processing suite for electron microscopy,* Journal of Structural Biology **157**, 38 (2007).

[8] S. Nickell, F. Förster, A. Linaroudis, W. Del Net, F. Beck, R. Hegerl, W. Baumeister, and J. M. Plitzko, *TOM software toolbox: Acquisition and analysis for electron tomography,* Journal of Structural Biology **149**, 227 (2005).

[9] H. Rullgård, L.-G. Öfverstedt, S. Masich, B. Daneholt, and O. Öktem, *Simulation of transmission electron microscope images of biological specimens,* Journal of Microscopy **242**, 234 (2011).

[10] L. Reimer and H. Kohl, *Transmission Electron Microscopy*, 5th ed., Springer series in optical sciences, Vol. 36 (Springer Verlag, 2008).

[11] R. J. Hall, E. Nogales, and 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**, 468 (2011).

[12] M. Vulović, R. B. G. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, H. Rullgård, O. Öktem, and B. Rieger, *Image formation modeling in cryo-electron microscopy,* Journal of Structural Biology **183**, 19 (2013).

[13] J. M. Cowley and A. F. Moodie, *The scattering of electrons by thin crystals,* Journal of the Physical Society of Japan, Supplement B **17**, 86 (1962).

[14] J. C. H. Spence, *High-Resolution Electron Microscopy*, 3rd ed. (Oxford University Press, 2003).

[15] J. M. Gibson, *Breakdown of the weak-phase object approximation in amorphous objects and measurement of high-resolution electron optical parameters,* Ultramicroscopy **56**, 26 (1994).

[16] B. K. Jap and R. M. Glaeser, *The scattering of high-energy electrons. II. quantitative validity domains of the single-scattering approximations for organic crystals,* Acta Crystallographica Section A **36**, 57 (1980).

[17] E. H. Linfoot, *Transmission factors and optical design,* Journal of the Optical Society of America **46**, 740 (1956).

[18] M. Vulović, *Modeling of Image Formation in Cryo-Electron Microscopy*, Ph.D. thesis, Delft University of Technology (2013).

[19] L.-M. Peng, S. L. Dudarev, and M. J. Whelan, *High-Energy Electron Diffraction and Microscopy,* Monographs on the Physics

and Chemistry of Materials, Vol. 61 (Oxford University Press, 2004).

[20] K. Ishizuka and N. Uyeda, *A new theoretical and practical approach to the multislice method,* Acta Crystallographica Section A **33**, 740 (1977).

[21] J. W. Goodman, *Introduction to Fourier optics*, 3rd ed. (Roberts and Company, 2005).

[22] S. Goudsmith and J. L. Saunderson, *Multiple scattering of electrons,* Physical Review **57**, 24 (1940).

[23] L. M. Voortman, S. Stallinga, R. H. M. Schoenmakers, L. J. van Vliet, and B. Rieger, *A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM,* Ultramicroscopy **111**, 1029 (2011).

[24] L.-M. Peng, G. Ren, S. L. Dudarev, and M. J. Whelan, *Robust parameterization of elastic and absorptive electron atomic scattering factors,* Acta Crystallographica Section A **52**, 257 (1996).

[25] L. M. Voortman, E. M. Franken, L. J. van Vliet, and B. Rieger, *Fast, spatially varying CTF correction in TEM,* Ultramicroscopy **118**, 26 (2012).

[26] A. I. Kirkland, S. L.-Y. Chang, and J. L. Hutchison, *Atomic resolution transmission electron microscopy,* in *Science of Microscopy*, Vol. 1, edited by P. W. Hawkes and J. C. H. Spence (Springer, 2008) pp. 3–64.

[27] X. Zhang, L. Jin, W. H. Fang, Q. Hui, and Z. H. Zhou, *3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry,* Cell **141**, 472 (2010).

5

Quantifying resolution limiting factors in subtomogram averaged cryo-electron tomography using simulations

de waarheid is een raadsel
en dat gaat als volgt
het is een goeie vriend maar altijd te laat
ERIK DE JONG — Spinvis

Lenard M. Voortman
Miloš Vulović
Massimiliano Maletta
Andreas Voigt
Erik M. Franken
Angelita Simonetti
Peter J. Peters
Lucas J. van Vliet
Bernd Rieger

Manuscript submitted to Journal of Structural Biology

5

Abstract

Cryo-Electron Tomography (CET) is the only available technique capable of characterizing the structure of biological macromolecules in conditions close to the native state. With the advent of subtomogram averaging, as a post-processing step to CET, resolutions in the (sub-)nanometer range have become within reach. In addition to advances in instrumentation and experiments, the reconstruction scheme has improved by inclusion of more accurate contrast transfer function (CTF) correction methods, better defocus estimation, and better alignments of the tilt-series and subtomograms. To quantify the importance of each contribution, we have split the full process from data collection to reconstruction into different steps. For the purpose of evaluation we have acquired tilt-series of ribosomes in such a way that we could precisely determine the defocus of each macromolecule. Then, we simulated tilt-series using the InSilicoTEM package and applied tomogram reconstruction and subtomogram averaging. Through large scale simulations under different conditions and parameter settings we find that tilt-series alignment is the resolution limiting factor for our experimental data. Using simulations, we find that when this alignment inaccuracy is alleviated, tilted CTF correction improves the final resolution, or equivalently, the same resolution can be achieved using less particles. Furthermore, we predict from which resolution onwards better CTF correction and defocus estimation methods are required. We obtain a final average using 3198 *ribosomes with a resolution of* 2.2 nm *on the experimental data. Our simulations suggest that with the same number of particles a resolution of* 1.2 nm *could be achieved by improving the tilt-series alignment.*

5.1. *Introduction*

Cryo-electron tomography (CET) is an essential technique to study the structure of macromolecules *in situ*, i.e. embedded in their native environment. A typical CET acquisition consists of a thin specimen that is tilted in order to acquire projections of the specimen at different angles. These projections are then used to reconstruct a three-dimensional (3D) volume. A major problem of this technique is radiation damage imposed by the electrons onto the specimen. This limits the amount of electrons that can be used for imaging which results in very noisy images. Nevertheless, this restriction can be overcome using subtomogram averaging. If a specimen contains many copies of an identical structure, the reconstructed subvolumes of these structures can be aligned and averaged to increase the signal-to-noise ratio (SNR) of the final reconstruction. Prerequisites for subtomogram averaging are that these copies can be identified in the reconstructed volume and aligned.

Projections in CET are intentionally recorded with underfocus. This defocusing allows contrast-generating interference of the undiffracted beam with the beam that is phase-shifted by the specimen. The contrast transfer function (CTF) describes the contrast transfer due to aberrations such as defocusing, astigmatism and spherical aberration. In effect, the CTF is an oscillating function of spatial frequency and depends on the defocus. These oscillations result in contrast inversions at certain spatial frequencies. Therefore, it is necessary to correct for the oscillations in order to interpret structures at a resolution beyond the first zero-crossing.

Recently, different studies tackled a number of problems associated with CET. Due to the extremely low SNR per projection, defocus estimation requires either a special averaging technique [1, 2], using magnification correction [3] or a new acquisition procedure [4]. Due to the tilted geometry, required for tomography, CTF correction needs to account for the defocus gradient perpendicular to the tilt-axis [1–3, 5–8]. Furthermore, some researchers studied CTF correction methods that consider the defocus gradient within the specimen along the optical axis [7, 9, 10]. Using these methods together with subtomogram averaging, Eibauer *et al.* [4] reported a resolution of 1.68 nm on mycobacterial membrane protein MspA.

Similar to single-particle analysis (SPA) [11], the resolution after subtomogram averaging depends on the number of particles but also on the accuracy of defocus estimation, CTF correction method as well as tilt-series and subtomogram alignment. In order to get a better understanding of what is currently limiting the resolution, it is needed to quantify the influence of the different processing steps.

In this study we investigate the influence of defocus estimation, CTF correction, tilt-series alignment and subtomogram alignment using primarily simulations. We use an extended acquisition scheme to determine the defocus and defocus gradient for each projection in a tilt-series, providing us with orientation and planarity of the sample and thus the defocus at the position of each macromolecule. We acquired experimental data using this scheme and simulate tilt-series which match the experimental conditions. After subtomogram averaging, we show that the experimental data is in good agreement with the simulations. Using these simulations we quantify the influence of defocus estimation and CTF correction, but also tilt-series alignment and subtomogram alignment on the resolution.

5.2. *Experimental methods*

5.2.1. Protein purification and sample incubation

E.Coli MRE600 were cultured up to an OD600 of 1.0. Then the cell membranes were disrupted in a French Press and the ribosomes were purified following the protocol described in Fechter *et al.* [12].

The EF-G gene was inserted in E.Coli following the procedure introduced by Dümmler *et al.* [13]. BL21 E.Coli cells, transformed with the EF-G fuse gene, were cultured in LB medium at 37 °C up to an OD600 of 0.7. The expression of the EF-G was induced upon administration of isopropyl-β-D-1-thiogalactopyranoside (IPTG). Four hours after induction the cells were harvested and the pellet dissolved in a buffer containing 25 mM Tris-HCl pH 7.1, 5% glycerol, 700 mM NaCl, 6 mM β-mercapto-ethanol, 0.1 mM PMSF and 0.1 mM benzamidine. The cells were then lysed by sonication and the debris and cell-membranes were removed by centrifugation (13000 rpm, 45 min). The EF-G was separated by the contaminants present in the lysate loading the cell extract in a nickel-nitrilotriacetic acid (Ni-NTA) chromatographic column. The column was washed with 5 volumes of buffer containing 25 mM Tris-HCl pH 7.1, 300 mM NaCl, 5% glycerol, 20 mM Imidazole, 6 mM β-mercapto-ethanol, 0.1 mM PMSF and 0.1 mM benzamidine and then the EF-G was eluted with a buffer containing 25 mM Tris-HCl, pH 7.1, 300 mM NaCl, 5% glycerol, 300 mM imidazole, 6 mM β-mercapto-ethanol, 0.1 mM PMSF and 0.1 mM benzamidine.

The purified 50S, 30S and EF-G were dialysed in separate membrane against the same buffer: 2 mM MgCl2, 50 mM KCl, 20 mM Tris-HCl of pH 7.5, 0.1 mM EDTA, 1 mM DTT .

Equimolar amounts of 50S and 30S E.coli Ribosome were incubated for one hour with t-RNA fMet and m-RNA in a dialysis buffer. In another vial, with the same buffer composition, we incubated a 1.3 times excess of EF-G with a 10 times excess of fusidic acid. Then we mixed the solutions of the vials together in a 10mM MgCl2 buffer in order to obtain the 70S complex bound to t-RNA fMet and m-RNA and locked to EF-G by fusidic acid administration. 5 µL of 0.3 mg/mL of the complex were applied to glow-discharged quantifoil grids. The excess of liquid was blotted away in a vitrobot (FEI Company, 4 sec blot time, Force 0, 100% humidity, 25 °C) and then the grid was flash-frozen in liquid ethane cooled down by liquid nitrogen.

5.2.2. Data acquisition and reconstruction

Single axis tilt-series were collected at 300 kV, with a Titan microscope (FEI Company) equipped with a Falcon direct electron detector at NeCEN (www.necen.nl). The pixel size is 3.748 Å; the spherical and chromatic aberrations are both 2.7 mm, the objective aperture 100 µm, while the energy spread and illumination aperture are 0.7 eV and 0.03 mrad, respectively. To facilitate the estimation of defocus and orientation of the specimen, we made use of an extended acquisition protocol. The extended acquisition protocol is described in detail in Appendix 5.B. We acquired projections from −60° to +60°, with an angular increment of 2°, a total dose around 78 $e^-/Å^2$ and a requested underdefocus of 4.5 µm. CTF correction was performed using our previously developed publicly available toolbox [7, `http://www.diplib.org/add-ons`]. The tilt-series were reconstructed into tomograms using the standard IMOD procedure [14, 15] after gold-bead tracking and rigid alignments. The mean residual provided by IMOD from all gold markers was 1.83 pixels.

The tilt-series alignment included solving for all tilt-axis rotations, all magnifications and grouped tilt-angles. Distortions or local-alignments were not used. We selected the best four reconstructions based on optimal defocus estimation and lowest residual in tilt-series alignment.

For particle picking we followed the standard PyTom procedure [16]. We used a template of a ribosome (EMD-1920, Elmlund *et al.* [17]) filtered to 4 nm in order to localize similar structures in the tomogram using cross-correlation. The ribosomes were sorted from the false positives based on their shape. After this cleaning procedure we obtained 3198 ribosome particles.

After particle picking, the location and orientation of each particle were refined using PEET [18]. As an initial template we used a reference potential low-pass filtered to 5 nm. The reference potential was generated from a high resolution

X-ray structure as deposited within the RCSB Protein Data Bank (for details see section 5.3.1). We iteratively refined the orientation estimate using 6 steps with 4× binning, followed by two steps with 2× binning and one final orientation refinement step on unbinned data.

We filter the averaged subvolume to correct for the damping of high spatial frequencies due to the modulation transfer function (MTF) of the camera and processing artifacts. This is done by weighting the angular-averaged power-spectra of the reconstructed volumes to match the angular-averaged power-spectra of the simulated interaction potential of the ribosome. This effectively amplifies the high spatial frequencies. After weighting of the power-spectra, the volumes were low-pass filtered to the resolution found by thresholding the Fourier shell correlation (FSC) calculated on two half datasets to 0.143.

5.3. *Simulations*

5.3.1. Tilt-series

An advantage of using a simulator of TEM images is that it models the outcome of the data acquisition for many combinations of experimental parameters. We used the *InSilicoTEM* simulator [19] to generate tilt-series with similar imaging conditions (defocus, dose, magnification) as in our experimental data in order to investigate the influence of defocus estimation, tilt-series alignment, subtomogram orientation estimation, and number of particles on the final resolution.

We simulated six tilt-series each containing 600 ribosome particles. As in the experiments, the tilt angles ranged from −60° to 60° in steps of 2° resulting in 61 (4k × 4k) unbinned images with a pixel size of 3.748 Å. Fig. 5.1 shows an example of experimental and simulated projection, as well as slices from a tomogram.

The main steps in our image simulations are i) construction of the specimen's interaction potential, ii) modeling the electron-specimen interaction, iii) propagation of the electron wave through the optics, and iv) detection of the intensity.

The ribosome particles were embedded in a slab of vitreous ice modeled as a continuous medium with a potential of 4.5301 V. The thickness of the specimen was taken to be ~ 35 nm and used to model the signal damping due to the plasmons of the vitreous ice. The voxel size was chosen to be the same as the pixel size in the images (3.748 Å) resulting in a specimen volume with a size of 4k × 4k × 81. The volume was divided in x and y directions into 25 × 25 subvolumes within which particles were randomly positioned. The input for computing the interaction potential of an individual particle was a high resolution X-ray structure

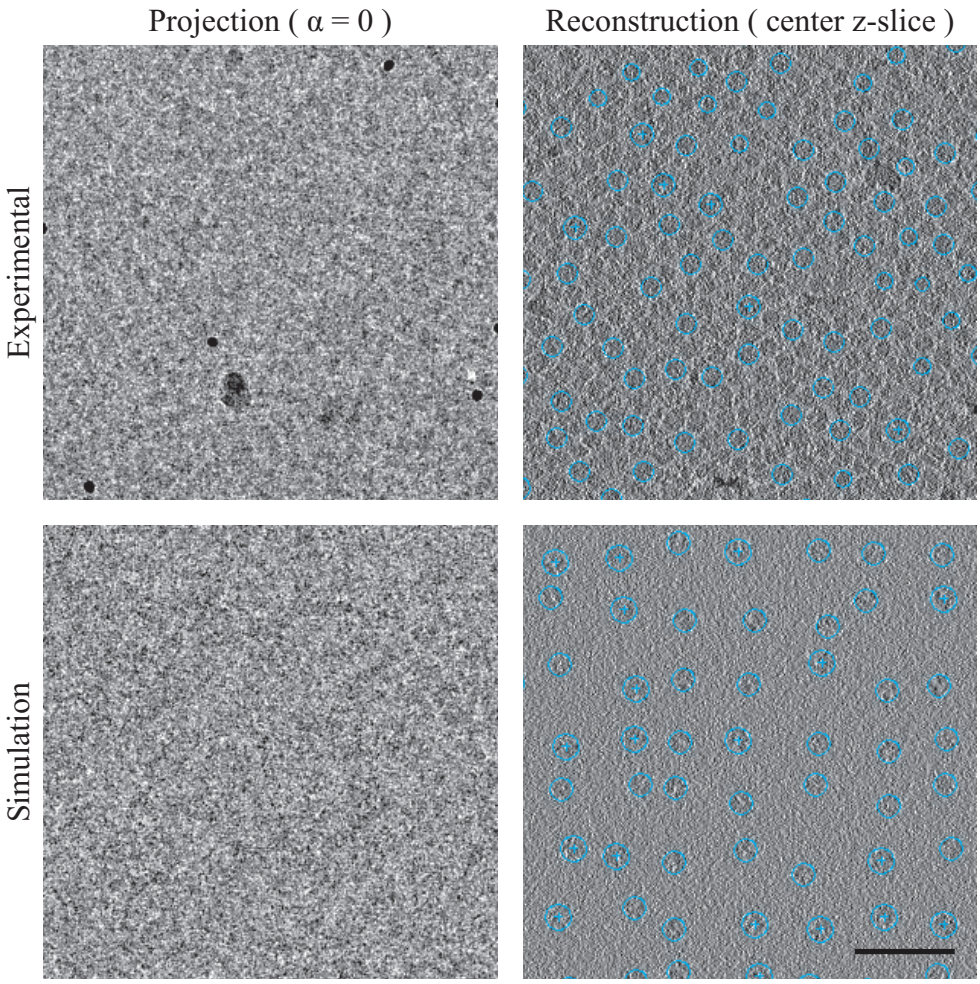


Figure 5.1. An example of experimental and simulated projections, as well as slices from a tomogram. The scale bar corresponds to 100 nm.

as deposited within the RCSB Protein Data Bank (PDB). To simulate the whole ribosome particle, we combined the PDB entries 2WRJ (large ribosome subunit) and 2WRI (small ribosome subunit with elongation factor G). The coordinates of each combined PDB (2WRI+2WRJ) were rotated with random Roll-Pitch-Yaw angles [20] producing uniform sampling of particle orientations. Rotation was performed on PDB coordinates before calculating the final potential map to avoid any possible interpolation errors. Finally, the interaction potential was calculated as the sum of the individual atomic potentials that were computed as the Fourier transforms of tabulated electron scattering factors. Prior to down-sampling to the pixel size, the resulting potential maps were low-pass filtered (Gaussian with standard deviation of 3.61 Å) in order to avoid aliasing. The influence of pH and the ions in the the solvent is neglected as well as possible beam-induced motion.

The electron-specimen interaction was modeled via the thick-phase grating approach (TPGA) [21] to account for the effects of the specimen thickness and tilt geometry. Our implementation of the TPGA utilizes non-uniform fast Fourier transform sampling of the parabolic Fresnel propagator in the frequency domain, to speed-up the calculations [6, 7].

The electron wave exiting the specimen is further subjected to the effective CTF which is simulated using the parameters of the Titan microscope described in 5.2.2. The defocus per projection used for the simulations matched the defocus values of the previously estimated defocus from the experimental datasets (estimated using an extended acquisition scheme, see Appendix 5.B) of on average 4.51 μm underfocus. Astigmatism was neglected.

The installed detector (Falcon I) was characterized using tools provided in [22]. Conversion factor, MTF and detective quantum efficiency (DQE) of the detector were included in the model as described in [19]. The integrated electron flux for each complete tilt-series was 78 $e^-/Å^2$.

Additionally, we wanted to compare the detector used in the experiments to a new generation of direct electron detectors in counting mode. Under ideal circumstances, such a detector in counting mode can be approximated as a *perfect* pixelated detector. Having square pixels of finite size, the detector's MTF is represented as a 2D sinc function (MTF = sinc(qΔx), where Δx is the pixel size) while the DQE of such detector is modeled as DQE = sinc^2(qΔx) [23].

5.3.2. Post-processing

The simulated data was processed in the same way as the experimental data, see section 5.2.2, except for a few differences due to the nature of the simulations.

Figure 5.2. Different post-processing steps were applied to the simulated data. For the defocus estimation, CTF correction, tilt-series alignment and subtomogram alignment steps two different variations were used resulting in 16 different subtomogram averaged results.

In the simulated tilt-series the defocus varied per tilt-angle (matching the experimental data). For the defocus estimation two different options were used. A defocus estimation procedure that accounts for the varying defocus in a tilt-series (such as the extended acquisition protocol) was emulated by using the simulated varying defocus values. To simulate defocus estimation procedures which only estimate an average defocus for the entire tilt-series, the defocus values of an entire tilt-series were averaged and this average was then used to correct all projections in that tilt-series.

Tilt-series alignment is not required since the projections were simulated with perfect alignment. In order to estimate the alignment accuracy of the experimental tilt-series together with any eventual beam-induced motion, gold-bead tracking was performed on all gold markers within the field-of-view. Using all gold markers gives a good indication of the global misalignment but the alignment itself was performed using a manually selected subset. The mean residual from all gold markers was 1.83 pixels (as provided by IMOD) and was used to introduce a certain misalignment into the simulations. The projections in the simulated tilt-series were randomly displaced using a Gaussian distribution with standard deviation equal to the mean residual of all gold markers after alignment of the experimental tilt-series. Here we assume that random shifting of the projections produces a similar response as the true effects related to the misalignment of the real particles. This assumption is justifiable since each particle is simulated with different orientations and in that way a bias is avoided.

We did not simulate any debris or contamination in the tilt-series. As a result, all particles present in the reconstructed volumes are identical and easy to identify. Therefore, the simulated datasets are not representative to verify particle picking and subsequent classification procedures. Instead, we used the locations that were used in the simulations. To compare the simulations with the experiments, however, we must analyze the resolution for different numbers of particles to see whether particle picking was successful in the experimental data.

Subtomogram alignment was carried out equivalently to the experimental data processing. Since the simulated orientations were known, however, a *perfect* subtomogram alignment was also created.

Simulated data was processed with all different combinations of the aforementioned post-processing steps, leading to 16 different subtomogram averaged results (see Fig. 5.2).

5.4. *Results*

5.4.1. Experimental data

After processing the experimental data as described in section 5.2.2, the final subtomogram average was calculated together with the corresponding FSC. By thresholding the FSC at 0.143 we measured a resolution of 2.2 nm. In order to correctly interpret the electron microscopy map we fitted the PDB structure of a ribosome bound to EF-G (2WRJ, 2WRI) to the map. The rigid body fitting was performed using the `Fit in Map` function of Chimera [24] and is shown in Fig. 5.3A. It is clear that all the densities are occupied by the features of the 70S ribosome and we can observe that the EF-G (in red in the picture) nicely fits in the protruding density left empty by the 70S. This confirms that the fucidic acid that we introduced in the solution stalled the ribosome in the late elongation state where the 70S is bound to EF-G.

The FSC in Fig. 5.3A for different number of particles, shows that the increase in correlation between 800 and 1600 is very small. Comparing the FSCs of Fig. 5.3A and Fig. 5.3B, we note that the FSCs match well for 100 to 800 particles. However, for the simulations (where all picked particles are identical) the correlation shows a larger increase for 1600 particles. This might indicate that the total set of picked particles in the experimental data does not have a uniform quality.

The relation between the number of averaged particles, and the final resolution is shown in Fig. 5.4. Using a threshold of 0.143 on the FSC we determine the resolution for different numbers of particles in each half-dataset. From Fig. 5.4 it is clear that the resolution only increases gradually. The sudden decrease in resolution for 100 and 200 particles is caused by the FSC dropping below 0.143 around the first zero-crossing.

There was only a minute difference between the resolution with TCTF or CTF0 correction. Similarly, the choice between a varying defocus estimate and the mean defocus estimate only has a small influence on the resolution.

For all FSC curves presented in Fig. 5.3 there are strong oscillations caused by the zero-crossings of the CTF. Given the amount of defocus variation in this

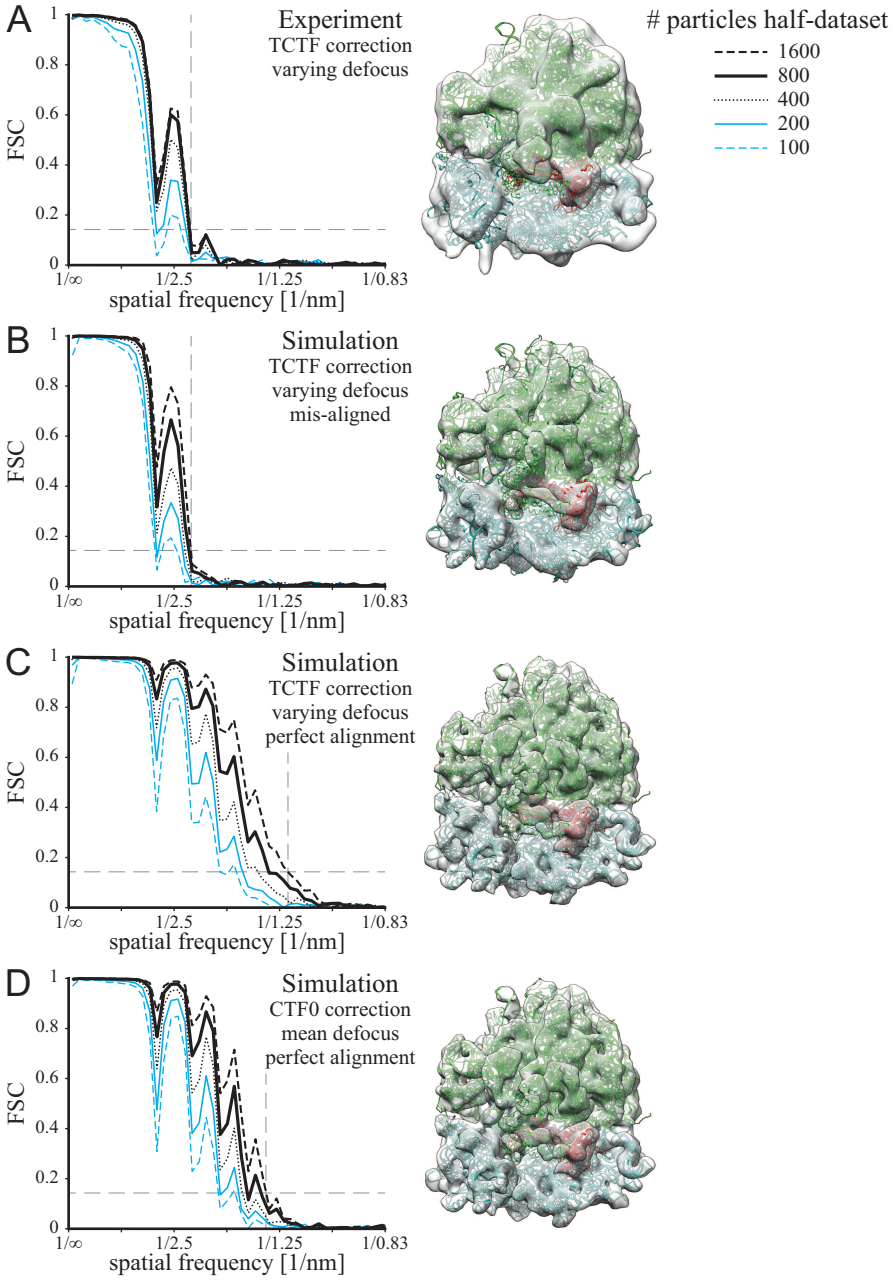


Figure 5.3. Fourier shell correlation (FSC) and isosurface representations for experimental and simulated data using different post-processing steps. (A) Experimental data, TCTF correction, varying defocus. (B) Simulated data, TCTF correction, varying defocus, misaligned. (C) Simulated data, CTF0 correction, mean defocus, *perfect* alignment. (D) Simulated data, TCTF correction, varying defocus, *perfect* alignment. Visualizations using Chimera [24], EF-G is highlighted in red.

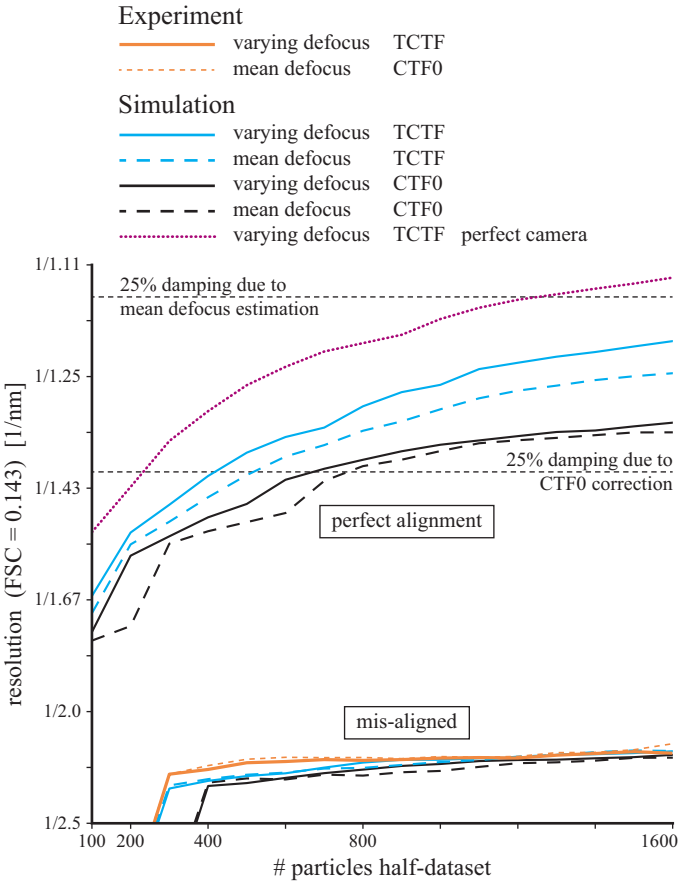


Figure 5.4. Resolution (FSC= 0.143) as a function of the number of particles, for experimental data and simulations. For the experiments and simulations with matching misalignment of the tilt-series, the resolution is slightly lower than 2 nm. The simulations with *perfect* tilt-series alignment show a clear influence of CTF correction and defocus estimation. The expected influence of a direct electron detector in counting mode (*perfect* camera) is also included. Uncertainty of the resolution is in the order of the line thickness. Two horizontal lines are included which are related to the predicted resolution at which the FSC is damped by more than 25% (see Appendix 5.A).

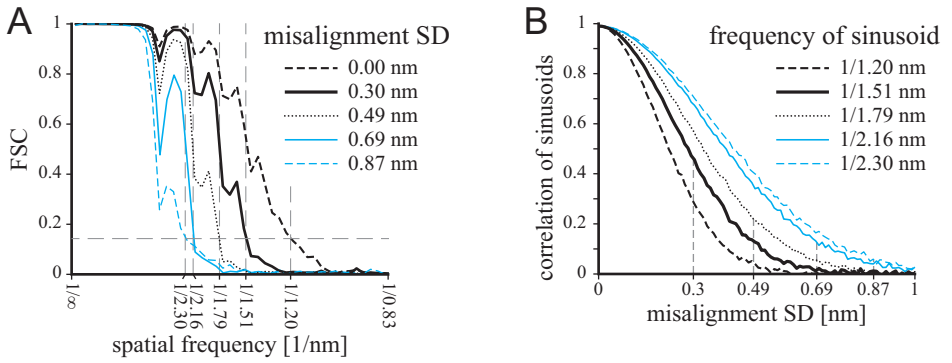


Figure 5.5. (A) Fourier shell correlation (FSC) of simulated data using 1600 particles in each half-dataset, TCTF correction and varying defocus for different amounts of misalignment. (B) Average normalized correlations of two sinusoids with frequencies chosen to correspond to the resolutions (FSC= 0.143) found in (A). One of the sinusoids was randomly displaced using a Gaussian distribution. The standard deviation (SD) of this random displacement is used as a measure for misalignment.

dataset, we expect these oscillations to appear. The appearance of these oscillations, however, is also influenced by how the FSC is computed (see Appendix 5.C).

5.4.2. Simulated data

We processed the simulations similarly to the experimental data. Fig. 5.3B shows the FSC and isosurface renderings where the post-processing steps were matched to that of the experimental data. The resolution from the simulated data is in good agreement with the experimental data in Fig. 5.3A. Furthermore, the FSC shows similar characteristics. We also see that the molecular structure fits accurately inside the isosurface rendering. This demonstrates self-consistency of the procedures since the simulations used this molecular structure to generate projections.

The simulations indicate that the tilt-series alignment has the largest influence on the final resolution. Fig. 5.3C and D show the FSC and isosurface renderings for simulations where the tilt-series alignment was *perfect*. From the FSC it is clear that TCTF correction and varying defocus estimation results in an improvement of the FSC and thus also in resolution.

In Fig. 5.4 the resolution is plotted for different numbers of particles. The simulations were performed using a different combination of post-processing steps (see Fig. 5.2). The effects of the defocus estimation, CTF correction and tilt-series alignments are shown in Fig. 5.4. The influence of subtomogram alignment was relatively small and therefore omitted from the figure.

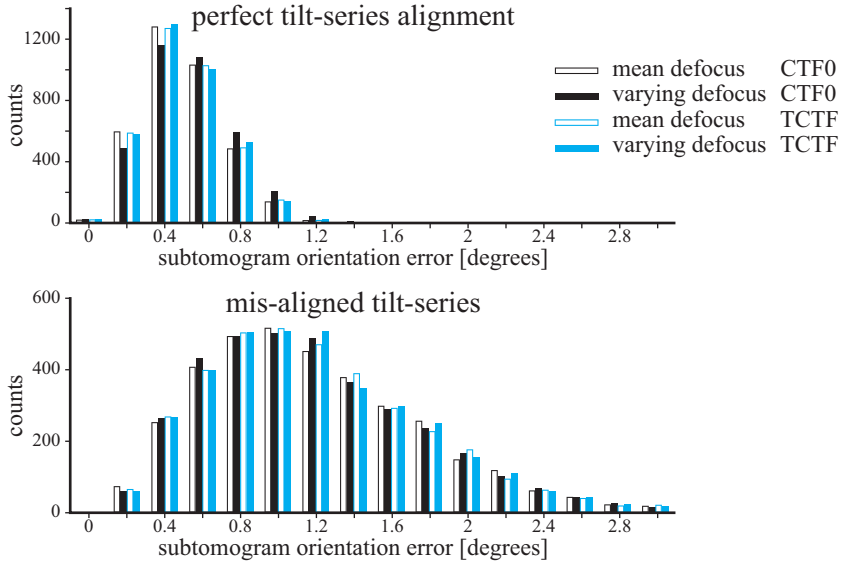


Figure 5.6. Accuracy of subtomogram alignment on simulated data using different post-processing procedures. Different CTF correction or defocus estimation methods do not seem to influence the orientation estimation. The tilt-series alignment does influence the subtomogram alignment. An angular mismatch of 2° corresponds to a 0.5 nm shift at the periphery of the ribosome.

From Fig. 5.4 it is clear that with an inadequate tilt-series alignment, there is almost no influence of the type of CTF correction or defocus estimation on the attained resolution. However, with *perfect* tilt-series alignment TCTF correction and varying defocus estimation are of clear influence. Also note that the influence of the varying defocus estimate has a larger influence when used in combination with TCTF correction.

The root-mean-square difference between mean defocus and varying defocus estimation as used in the simulations was 154 nm; this corresponds to a resolution of 1.15 nm using Eq. (5.7). From these limits, we expect that using the mean instead of varying defocus estimation has a smaller influence on the attainable resolution than using CTF0 instead of TCTF correction. This is confirmed in Fig. 5.4. However, we see in the simulations that already for lower resolutions there is a notable improvement using TCTF correction and varying defocus estimation. It is expected that using a *perfect* camera (direct electron detector in counting mode) a certain resolution can be achieved with approximately half the number of particles (dotted line in Fig. 5.4).

The amount of misalignment that was introduced in the simulations was 1.83 pixels (~ 0.69 nm). From Fig. 5.4 we see that this limits the resolution to 2.2 nm.

To better understand the influence of a certain misalignment on the FSC, we processed tilt-series using different amounts of misalignment. Fig. 5.5A shows the FSC of a reconstruction similar to those presented in Fig. 5.3B and C. Then, we investigated what is the influence of a certain misalignment on the obtained resolution. To this end, we calculated the correlation of two sinusoids of which one was randomly displaced. The frequency of the sinusoids was chosen to be the same as the obtained resolution in Fig. 5.5A. Fig. 5.5B shows that at a spatial frequency of $1/2.2\ \mathrm{nm}^{-1}$, a misalignment of ~ 0.69 nm results in a correlation of less than 20%.

To study the accuracy of the subtomogram orientation estimation, we compare the estimated (Euler) angles with the simulated angles. In order to quantify the difference between two sets of angles we first convert each set of angles to a rotation matrix: R_s the simulated rotation and R_e the estimated orientation. Now the orientation error can be found using $\mathrm{trace}\left(R_e R_s^{\top}\right) = 1 + 2\cos(\theta)$, where θ is the rotation angle required to rotate from R_s to R_e. Fig. 5.6 shows a histogram of the angular mismatch between the simulated orientations and the estimated orientations. It is clear that the orientation estimation is overall successful. The average orientation error for *perfect* tilt-series alignment or misaligned tilt-series result in a displacement of only ~ 0.1 nm and ~ 0.3 nm respectively at the periphery of the particle. This is also confirmed by the FSC of the simulations with *perfect* subtomogram alignment which is very similar to the FSC in which the orientation was estimated.

5.5. *Discussion*

The simulations are in good agreement with the experimental data which provides a good validation of the *InSilicoTEM* simulator [19]. These simulations allow us to conclude which factors need to be optimized in order to achieve a better resolution. Tilt-series alignment has the strongest influence on the final resolution, followed by the CTF correction method (CTF0 versus TCTF correction) and defocus estimation accuracy (varying defocus estimate versus mean defocus estimate). The subtomogram alignment had the least significant influence on the resolution. In other words, the subtomogram alignment is already sufficiently accurate.

We used an FSC threshold of 0.143 to determine the resolution since it is expected from theory that this is a good indicator for the true resolution of the map in the absence of overfitting [11, 25]. We are confident that the presented results do not suffer from overfitting since for the simulated data exchanging the

estimated orientations with the simulated orientations did not influence the FSC and resolution. A different threshold criterion, e.g. 0.5 or half-bit criterion, would lead to slightly different resolutions. Nevertheless, this does not alter the relative difference between the different processing methods discussed in this study.

Here, we used an extended acquisition scheme and developed a tool for assessing the specimen geometry. Although in the case of a reasonably thin and flat specimen, four positions (instead of only two on the tilt axis) do not notably improve defocus estimation, they are still useful for checking specimen orientation and planarity. Furthermore, the extra off-axis regions do not have to be acquired for all tilt angles. A couple of images of (un)tilted focus areas will be sufficient to estimate the orientation and planarity of the specimen.

In Fig. 5.8A from Appendix 5.A we predict a resolution from which CTF0 correction would damp the FSC by more than 25%. Furthermore, in Fig. 5.8B we predict the same threshold depending on the defocus estimation accuracy. For a comparison, we included these resolutions in Fig. 5.4 showing that the limits are in agreement with the results from simulations. A damping of 25% is relatively strong and we see in the simulations that already for lower resolutions there is a notable improvement using TCTF correction and varying defocus estimation. Similarly, Fig. 5.8 depends on the chosen threshold value of 0.75. Changing the threshold will shift the boundaries between the regions but does not change the relative position. From Fig. 5.4, however, we conclude that the choice for 0.75 does provide a good indication of the resolution at which the influence of CTF correction or defocus estimation becomes relevant. Furthermore, the relative importance of defocus estimation and CTF correction was predicted correctly using criteria from Appendix 5.A independent of the chosen threshold.

The resolution that we obtained for our 70S ribosome (2.2 nm) is comparable with the one obtained by Chen *et al.* [16] (2.1 nm) for a similar structure. Other studies reported resolutions of ~ 1.7 nm [4, 26] for particles with higher symmetry. For simulated particles with a *perfect* tilt-series alignment, however, a resolution up to 1.2 nm can be attained, see Fig. 5.4. This underlines the importance of tilt-series alignment which is apparently essential to improve the resolution. Here we used rigid alignments in IMOD. We did not use the possibility for local alignment since at this magnification there were not enough gold markers to produce a reliable local fit.

One approach to tackle the alignment problems involves a combination of SPA and subtomogram averaging techniques [27, 28]. They used subtomogram averaging to assign initial locations and angles to each subvolume and then refine the alignment of the original projections using a projection matching procedure.

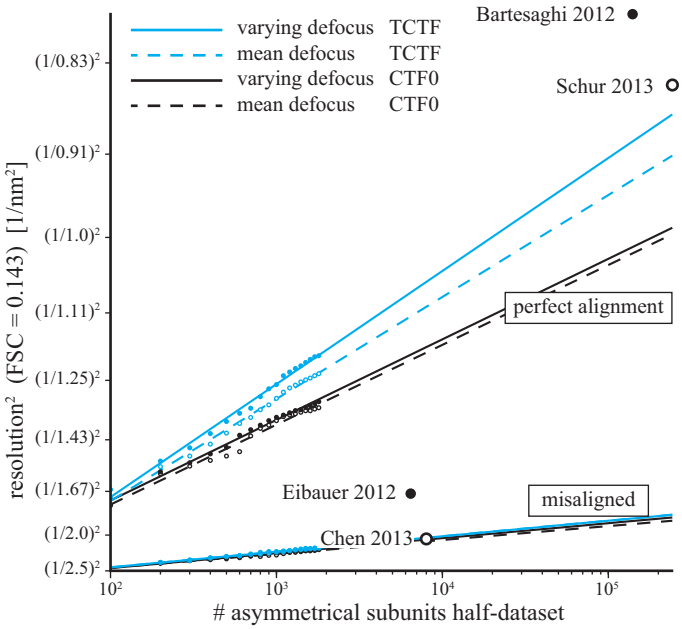


Figure 5.7. Extrapolation of resolution as a function of number of particles. Different datapoints of related publications [4, 16, 28, 29] of various structures are plotted for comparison. The horizontal axis represents the number of asymmetric units. The results from this study are extrapolated assuming a particular falloff of the particle's structure factor (see main text for details).

In this way the final alignment is effectively independent of the initial tilt-series misalignment or rotation of the particles during tilt-series collection. Another benefit is that this method reduces the interpolation steps required to produce the intermediate steps in classic subtomogram averaging. Using this technique, Bartesaghi *et al.* [28] reported a resolution of ~ 0.8 nm using 140.000 asymmetric units.

Recently, Schur *et al.* [29] reported a resolution of 0.85 nm using an optimized data collection protocol. To achieve this resolution they averaged 242,692 asymmetric units. Instead of measuring the defocus changes throughout the acquisition of a tilt-series, they used an optimized protocol that effectively stabilized the stage movement and defocus.

When we extrapolate our simulation results to the same number of asymmetric units, the predicted resolution is in the same range. Fig. 5.7 provides a graphical comparison of the resolution with respect to the number of particles reported in [4, 16, 28, 29]. The extrapolation is based on methods presented in Rosenthal and Henderson [11]. The data in Fig. 5.7 on which the extrapolation is

based is exactly the same as in Fig. 5.4. Please note that the achieved resolution inevitably depends on the scattering properties of the studied structure. Therefore, one should be careful when interpreting the results from Fig. 5.7. Nevertheless, it is interesting that the resolution reported by Bartesaghi *et al.* [28] is higher than the extrapolation of our optimal result. This is possibly due to a different reconstruction method, lower electron energy (80 kV) and a different detector.

We conclude that in our dataset suboptimal tilt-series alignment limits the resolution to around 2.2 nm. The origin of this misalignment is not easy to assess. By performing gold-bead tracking on all gold markers within the field-of-view, we measured an alignment residual of approximately 1.83 pixels (~ 0.69 nm). As we see from Fig. 5.5, this misalignment results in a severe damping of correlation between sinusoids with a period of 2.2 nm. This indicates that the residual of the gold-bead tracking is a good indication of the amount of misalignment. Furthermore, we see that a seemingly small alignment residual of ~ 0.69 nm already has a large influence at a resolution of 2.2 nm.

In this study, we could not further investigate the influence of radiation damage. The damage to the structure of the ribosome is difficult to quantify. In fact, there is no model to accurately simulate the degradation of the structure. Recent studies [28, 29] suggest, however, that using 25 to 40 $e^-/Å^2$ improves the resolution of the final average. Nevertheless, our simulations without influence of radiation damage accurately matched the experiments. If radiation damage is small and at random places, the majority of the particles will dictate the result at any location and *hide* the radiation damage.

Experimental data were acquired without an energy filter. In general, it is expected that energy filtering will improve the contrast. However, due to the thin ice the difference between energy filtered and unfiltered data might be inconspicuous. The reported simulation results did not include amplitude contrast. We did, however, simulate tilt-series with amplitude contrast but did not find a notable influence on the final resolution after subtomogram averaging.

The tomograms were acquired on very thin ice. This might make the ice more susceptible to beam-induced motion [30]. Therefore, the measured and simulated alignment error is a combined effect of the limited accuracy of the gold-bead tracking and any eventual beam-induced motion. Before a tomogram acquisition, the exposure area was visually inspected (in order to check the presence of ribosomes) with a dose of 1 $e^-/Å^2$. This pre-irradiation should have lowered the potential effects of beam-induced motion [30].

The latest generation of direct electron detectors [31–34] enables acquisition of exposure series and allows higher resolution reconstructions with fewer parti-

cles for SPA. Such acquisition facilitates reduction of the effect of beam induced movements by retrospective alignment and then averaging of the raw frames. This correction could be applied to tilt-series data collection as well. Nevertheless, each projection in a tilt series is acquired with a significantly smaller electron dose than in SPA. The extremely low SNR will severely hamper alignment of the sub-frames within a projection in a tilt-series. In order to process such data one would need a modified data acquisition, an independent set of gold markers dedicated to frame alignment and/or a mathematical model able to predict the radiation damage.

Using our simulations we give an estimate of the influence of a direct electron detector in counting mode (*perfect* detector) on the final resolution. The simulated *perfect* detector when compared to a Falcon I suggests that the same resolution would be achieved with approximately half the number of particles (dotted line in Fig. 5.4).

5.A. *Predicting the influence of CTF correction and defocus estimation*

5.A.1. CTF correction

The CTF inherently depends on the defocus, and the defocus varies over the specimen along the optical axis. Owing to the specific geometry associated with cryo-electron tomography (CET), special CTF correction methods have been developed. The tilting of a (thin) specimen results in a defocus gradient across the image, perpendicular to the tilt-axis. Furthermore, depending on the thickness, the image contributions from the top and bottom of the specimen can exhibit a significant difference in defocus.

CTF correction methods for tomography can be categorized into three groups.

First, the most basic form ignores the defocus gradient. We will refer to this method as regular CTF correction (CTF0):

$$\begin{aligned}\hat{I}_{\mathrm{CTF0}}\left(\mathbf{q}\right) = & \, i e^{-i\chi(q)} \hat{V}_{\alpha}\left(\mathbf{q},0\right) \\ & - i e^{i\chi(q)} \hat{V}_{\alpha}\left(\mathbf{q},0\right),\end{aligned} \tag{5.1}$$

where $\mathbf{q} = (q_x, q_y)$, $\hat{I}\left(\mathbf{q}\right)$ is the Fourier transform of the image intensity $I\left(\mathbf{x}\right)$ and $\hat{V}_{\alpha}\left(\mathbf{q}, q_z\right)$ is the Fourier transform of the 3D scattering potential V in a rotated coordinate system, i.e. $\hat{V}_{\alpha}\left(\mathbf{q},0\right)$ is a projection slice of $\hat{V}$ at tilt angle α. Furthermore, $\chi(q) = \frac{2\pi}{\lambda}\left(\frac{1}{4}C_s\lambda^4 q^4 + \frac{1}{2}\Delta f \lambda^2 q^2\right)$ is the aberration function, C_s the spherical aber-

ration, Δf the defocus at the center of the specimen, λ the electron wavelength and $q = \|\mathbf{q}\|$ the magnitude of the spatial frequency $\mathbf{q}$.

Second, CTF correction for tilted geometries has been studied in Fernández *et al.* [1], Xiong *et al.* [2], Zanetti *et al.* [3], Philippsen *et al.* [5], Voortman *et al.* [6, 7], Winkler and Taylor [8]. The analytical form of the tilted CTF (TCTF) is

$$\begin{aligned}\hat{I}_{\mathrm{TCTF}}(\mathbf{q}) = i e^{-i\chi(q)}\hat{V}_\alpha\left(\mathbf{q} - \frac{1}{2}\lambda q^2 \boldsymbol{\beta}\tan\alpha, 0\right) \\ - i e^{i\chi(q)}\hat{V}_\alpha\left(\mathbf{q} + \frac{1}{2}\lambda q^2 \boldsymbol{\beta}\tan\alpha, 0\right),\end{aligned} \tag{5.2}$$

where $\boldsymbol{\beta}$ is a unit-vector perpendicular to the tilt-axis $\boldsymbol{\beta} = (\cos\beta, \sin\beta)$ and β denotes the orientation of the tilt-axis (azimuth).

Third, the influence of the defocus gradient within the specimen (along the optical axis) has been studied by [6, 7, 9, 10]. We will refer to this method as 3D CTF correction (3DCTF):

$$\begin{aligned}\hat{I}_{\mathrm{3DCTF}}(\mathbf{q}) = i e^{-i\chi(q)}\hat{V}_\alpha\left(\mathbf{q}, -\frac{1}{2}\lambda q^2\right) \\ - i e^{i\chi(q)}\hat{V}_\alpha\left(\mathbf{q}, \frac{1}{2}\lambda q^2\right).\end{aligned} \tag{5.3}$$

From Eqs. (5.1), (5.2) and (5.3) it is clear that the difference between these CTF models can be fully characterized by shifts in the Fourier representation of the potential. The difference between Eqs. (5.1) and (5.2) is a shift of $\pm\frac{1}{2}\lambda q^2\tan\alpha$ along $\boldsymbol{\beta}$ in the Fourier domain. Similarly, the difference between Eqs. (5.2) and (5.3) is a shift of $\pm\frac{1}{2}\lambda q^2\frac{1}{\cos\alpha}$ along q_z [6].

Eq. (5.3) is the most accurate CTF model since it accounts for all defocus gradients. Eq. (5.1) is the least accurate but also the fastest to compute [6]. Therefore, we want to estimate the error that is introduced by approximating the 3DCTF. This error is closely related to a damping envelope we derived previously that estimates the loss of useful contrast when a specimen with finite thickness is approximated by an infinitely thin specimen [6]. This damping envelope is the result of a specific shift of sampling points in the Fourier domain.

In general, shifting in the Fourier domain results in a loss of correlation between the 'unshifted' and 'shifted' version. Similar to our previous work [6], we approximate this correlation loss as

$$E(\boldsymbol{\Delta}\mathbf{q}) = \mathrm{sinc}\left(d|\boldsymbol{\Delta}\mathbf{q}|\right), \tag{5.4}$$

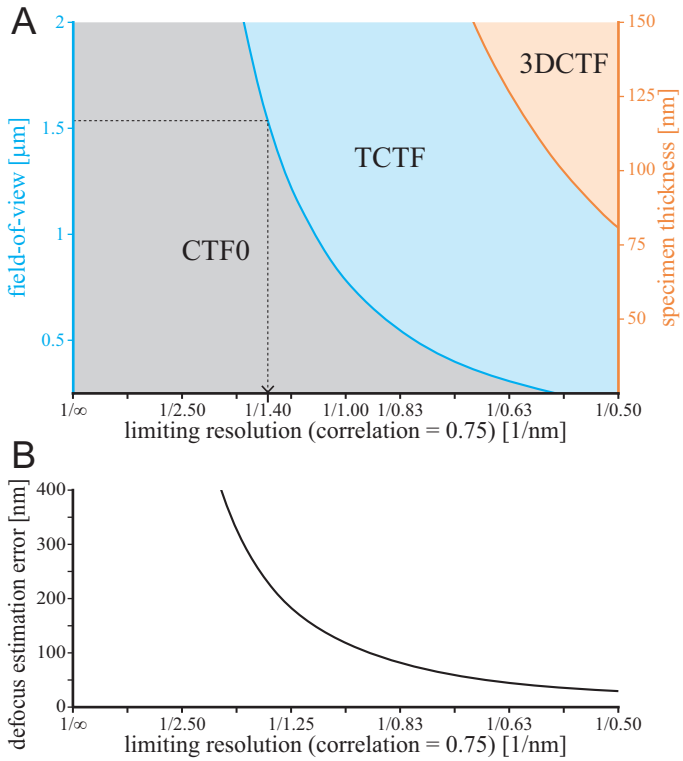


Figure 5.8. (A) Regions of applicability for the different CTF correction methods. The limiting resolution was determined by thresholding the correlation functions Eqs. (5.5) and (5.6) at 0.75. From this resolution onwards the Fourier shell correlation is expected to be damped by more than 25%. Limits are presented for CTF0 correction as a function of FOV and for TCTF correction as a function of the specimen thickness, for a tilt-series of ±60°. (B) Maximum allowable defocus estimation error as a function of resolution, determined by thresholding the correlation function Eq. (5.7) at 0.75.

where d is the spatial extent or field-of-view and sinc is the normalized sinc function. Eq. (5.4) can be used to quantify the difference between CTF0, TCTF and 3DCTF correction of a projection at a specific tilt-angle, field-of-view and thickness.

For tomography, however, different projections are combined to form a 3D volume. Therefore, the differences in the reconstructed volume are an average of the differences between the projections in the tilt-series. When the specimen is tilted, the part of the field-of-view which can be used effectively for the 3D reconstruction is reduced by $1/\cos\alpha$. Then the expected correlation between a reconstruction with CTF0 and TCTF correction is given by the average

$$E_{\mathrm{CTF0}}(q) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{sinc}\left(\frac{1}{2}\lambda q^2 d \sin\alpha_n\right), \tag{5.5}$$

where the summation is over all α_n in a tilt-series with a total of N projections.

The difference between 3DCTF correction and TCTF correction depends on the thickness of the specimen. The shift of magnitude $\frac{1}{2}\lambda q^2 \frac{1}{\cos\alpha}$ is in the axial direction of the specimen's Fourier transform. Therefore the effective d is the thickness t of the specimen and the expected correlation is given by

$$E_{\mathrm{TCTF}}(q) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{sinc}\left(\frac{1}{2}\lambda q^2 \frac{t}{\cos\alpha_n}\right). \tag{5.6}$$

We consider two CTF correction methods similar (no significant difference) when their correlation is larger than 0.75. Thresholding the expected correlation at 0.75 we find the maximum resolution up to which we considered CTF models to be similar. Fig. 5.8A shows this limiting resolution for a range of field-of-views and specimen thicknesses given a tilt-series of $\pm 60°$. The choice for 0.75 as a threshold is somewhat arbitrary. However, from Fig. 5.4 we conclude that the choice for 0.75 does provide a good indication of the resolution at which the influence of CTF correction or defocus estimation becomes relevant. Furthermore, the relative importance of defocus estimation and CTF correction was predicted correctly using Eqs. (5.5), (5.6) and (5.7).

5.A.2. Defocus estimation

Similar to the influence of CTF correction, we predict how the accuracy of defocus estimation influences the resolution. Whereas the difference between CTF correction methods can be characterized by shifts in the Fourier domain, a difference in defocus estimation results in an extra random phase contribution to the Fourier

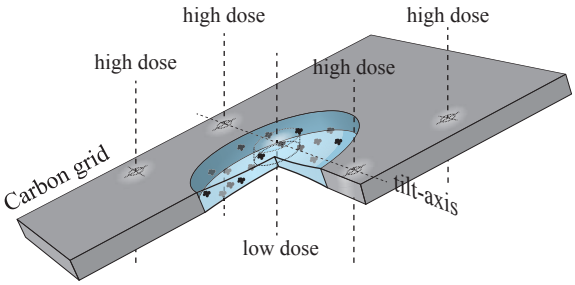


Figure 5.9. Schematic overview of the extended acquisition scheme. The low-dose *exposure image* is located in a hole of the carbon-grid. Two high-dose *focus images* are located on the carbon, on the tilt-axis. The other two high-dose areas are located off-axis and positioned such that even at high tilt-angles the electron beam does not overlap with the *exposure image*.

transform of the potential. This phase translates to a correlation function

$$E_{\sigma_{\Delta f}}(q) = \cos\left(\pi\lambda q^2 \sigma_{\Delta f}\right), \tag{5.7}$$

where $\sigma_{\Delta f}$ is the estimation error of the defocus Δf.

Fig. 5.8B shows the limiting resolution for a range of defocus estimation errors. Similar to Fig. 5.8A, a threshold of 0.75 was applied to Eq. (5.7) to find the limiting resolution.

5.B. *Extended acquisition scheme*

Improving the resolution beyond the first zero-crossing of the CTF requires a CTF correction step. In order to perform the CTF correction in such a way that it actually increases the resolution, the defocus (gradient) of each projection must be estimated with a certain accuracy [6]. During the acquisition of a tilt-series, the defocus on the tilt-axis can vary due to the non-perfect mechanics of the stage, the accuracy with which the eucentric height can be adjusted, the auto-focus area being offset from the exposure area or other fluctuations.

Estimating the defocus on the low-dose exposures of a tilt-series is very challenging and usually requires averaging over a certain tilt-range [1, 2]. When recorded images at different angles need to be averaged, however, it is no longer possible to assess defocus variations throughout the tilt-series. Eibauer *et al.* [4] introduced an extended acquisition scheme which can estimate the defocus per tilt-angle. They use two extra high-dose *focus images* per tilt-angle to estimate the defocus on the *exposure images*. We implemented and used an addition to this extended acquisition scheme. In addition to the two high-dose *focus images* which

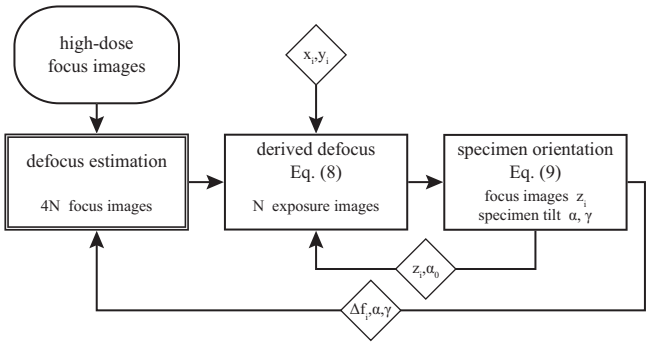


Figure 5.10. Work-flow to estimate the defocus on the *exposure images* from the high-dose *focus images*. The lateral location of the focus images is given by (x_i, y_i). The axial location z_i (w.r.t. unrotated coordinate system $\alpha = 0$) is estimated from the estimated defocus and subsequently used to improve the defocus estimate.

are positioned on the tilt-axis, we acquire two off-axis high-dose *focus images* (see Fig. 5.9). Having now four high-dose images has a number of additional benefits. Firstly, we can estimate the orientation of the specimen which can be used for TCTF correction. Secondly, the accuracy might be improved since we have more measurements to base the defocus estimate on. Finally, having four areas allows us to check whether the specimen is flat.

The acquisition procedure was implemented in a development version of FEI's Tomography acquisition package (version 4.0), making use of the batch mode and automatic tracking and focusing. The low dose acquisition template (which allows for graphical selection of areas for focusing and tracking) was extended such that an arbitrary number of extra acquisition areas could be added. These may have their own optics and camera acquisition settings. We used these extra acquisition areas to acquire extra high dose *focus images*.

5.B.1. Defocus estimation using extended acquisition scheme

After acquiring tilt-series using the extended acquisition scheme, the defocus at the *exposure images* can be estimated. The work-flow is presented in Fig. 5.10.

First, the defocus of the high-dose *focus images* is estimated using the methods described in Vulović *et al.* [35].

Next, the defocus at the location of the *exposure images* ($\Delta f(\alpha)$) is estimated from the defocus of each *focus image* ($\Delta f_i(\alpha)$) using

$$\Delta f(\alpha) \approx \Delta f_i(\alpha) - \sin(\alpha)\, x_i - \frac{\cos(\alpha_0)}{\cos(\alpha + \alpha_0)} z_i, \tag{5.8}$$

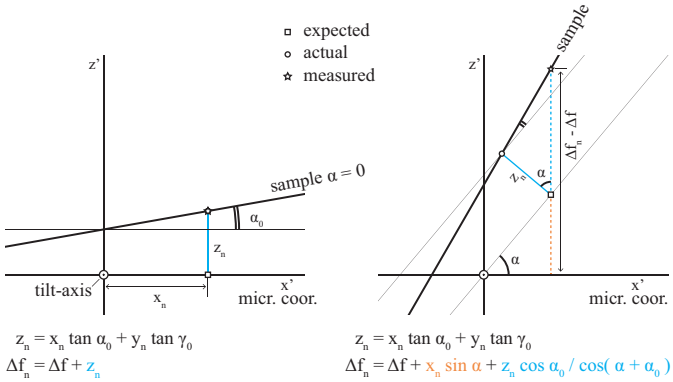


Figure 5.11. Graphical representation of Eq. (5.8). The acquisition software accounts for the movement of the focus area (*expected*). However, due to the specimen orientation α_0, γ_0 when $\alpha = 0$, the focus area has a different location (*actual*). As a result, the *measured* defocus depends on the axial location z_i of the focus area.

where α the tilt-angle and α_0 the specimen orientation at $\alpha = 0$. Furthermore, (x_i, y_i, z_i) is the location of the *focus image* with respect to the *exposure image* in specimen coordinates where y is aligned with the tilt-axis. For a graphical representation see Fig. 5.11.

However, initially Eq. (5.8) contains two unknowns: α_0 and z_i. Therefore, we use an iterative procedure described in Fig. 5.10. Using a rough estimate of $\Delta f(\alpha)$ we estimate z_i with

$$z_i = \text{median}_\alpha \left[\left(\Delta f_i(\alpha) - \Delta f(\alpha) - \sin(\alpha)\, x_i \right) \frac{\cos(\alpha_0 + \alpha)}{\cos(\alpha_0)} \right]. \tag{5.9}$$

We then refine the estimate of z_i by assuming that the specimen is (locally) a perfect plane. Using linear regression, we estimate the specimen orientation α_0, γ_0 assuming

$$z_i = \tan(\alpha_0) x_i + \tan(\gamma_0) y_i + \epsilon \quad \text{for } i = \{1,2,3,4\}. \tag{5.10}$$

Setting $z_i = \tan(\alpha_0) x_i + \tan(\gamma_0) y_i$ refines the estimate of z_i. This regression analysis also allows us to check whether the specimen is locally flat. The inner-loop in Fig. 5.10 is iterated until the specimen orientation α_0, γ_0 has converged, or to a maximum of 10 iterations.

Using the defocus $\Delta f_i(\alpha)$ and defocus gradient estimates, we can refine the defocus estimation on the focus images (outer loop in Fig. 5.10). Xiong *et al.* [2] introduced a method for shifting and averaging power spectra. This method

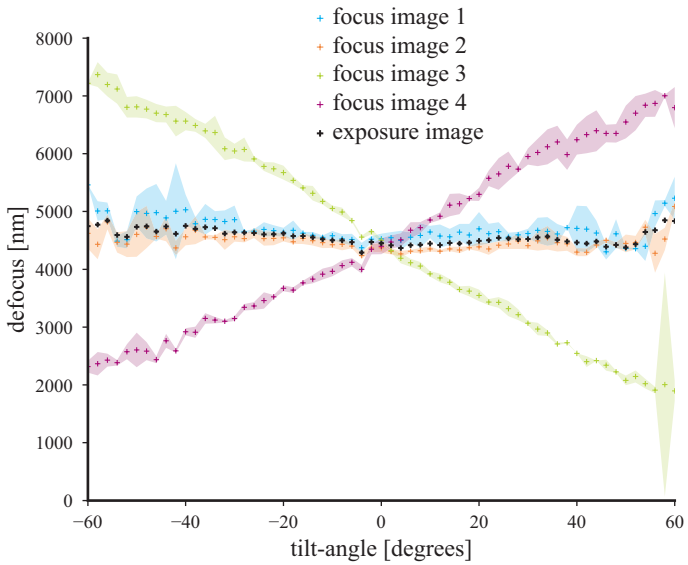


Figure 5.12. Example of defocus estimation using an extended acquisition scheme. Four *focus images* are used to compute the defocus at the position of the *exposure image*. The color-shaded areas correspond to the uncertainty of the defocus estimate.

is especially beneficial at the high tilt-angles in the tilt-series. At these high tilt-angles the strong defocus gradient blurs the Thon rings of the power spectrum. We implemented a periodogram averaging technique which aligns the zero-crossings, given an estimate of the defocus and defocus gradient. The defocus gradient is derived from the estimated specimen orientation, taking into account the tilt-axis orientation in the images. We noticed that, for the used experimental parameters, this extra iteration was essential. The outer loop is iterated only once.

Fig. 5.12 shows an example of defocus estimates on the four *focus images* and the resulting estimate of the defocus at the *exposure images*. The defocus estimation on the four *focus images* in this example was performed using the methods described previously in Vulović *et al.* [35]. The dose used for *focus images* was in this case only 4-5 times higher than for *exposure images* to minimaze any potential radiation damage of the exposure area.

5.C. *Frequency shells for FSC computation*

In Fig. 5.3 we observe oscillations in the FSC due to the zero-crossings of the CTF. Here we demonstrate that the visibility of these dips depends on the width of the

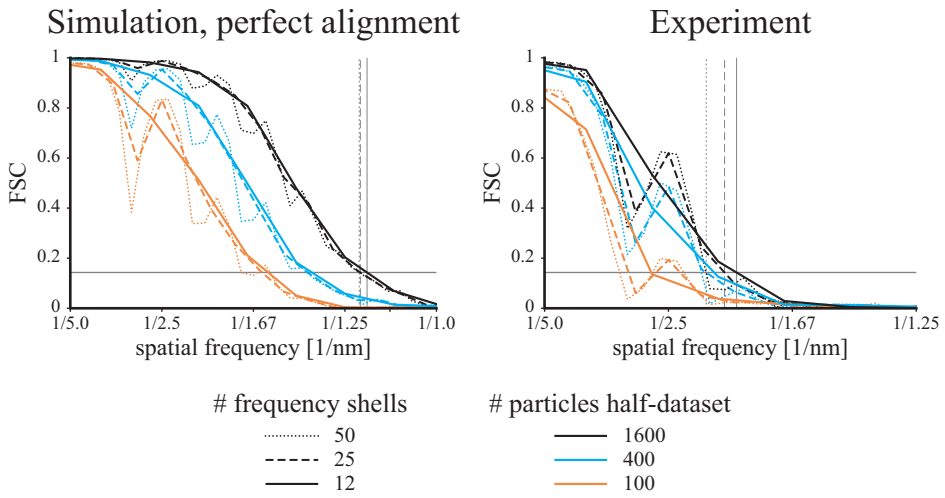


Figure 5.13. Fourier Shell Correlation (FSC) for different numbers of frequency shells and numbers of particles for one simulated dataset and one experimental dataset. Both datasets are processed using TCTF correction using a varying defocus estimate. The simulated data has a perfect tilt-series alignment whereas the experimental data has a non-perfect alignment.

frequency shells used for calculating the FSC. Consequently, this influences the found resolution by thresholding the FSC curve.

Dips in the FSC in subtomogram averaging occur if the particles are recorded with a similar defocus. Even though each tilt-series exhibits a certain variation in defocus and each tilted projection contains macromolecules imaged with a different defocus, this defocus spread does not guarantee that the SNR around the zero-crossings is similar to all other spatial frequencies. Only when a large number of particles is averaged or when tomograms are intentionally acquired using different defoci, can these dips be avoided.

Nevertheless, Fig. 5.13 shows that when the number of frequency shells used to calculate the FSC is too low, it is impossible to judge whether these dips disappeared due to undersampling of the FSC or that the SNR was indeed high enough. Furthermore, choosing too few frequency shells can lead to an erroneous increase in the reported resolution.

For the aforementioned reasons we calculate the FSC using 50 frequency shells, making the oscillations in the FSC more apparent. This number of frequency shells results in a frequency shell width that is close to the voxelsize of the Fourier transformed subtomograms in this study. (size of the subtomograms was 128^3 voxels). Fig. 5.13 shows that due to these oscillations it is possible that the FSC

crosses the threshold multiple times. We use the spatial frequency where the FSC crosses the threshold for the first time to determine the resolution.

Different choices in terms of the number of frequency shells, type of threshold and whether to use the first or last crossing of the threshold lead to different reported resolutions. The focus of this study is to quantify the influence of the different processing steps that lead to a reconstruction. Therefore, we primarily use FSC and the derived resolution to make a relative comparison between experiments and different simulations. In that sense, the choice of resolution criteria and the exact computation of the FSC does not influence our findings.

References

[1] J. J. Fernández, S. Li, and R. A. Crowther, *CTF determination and correction in electron cryotomography,* Ultramicroscopy **106**, 587 (2006).

[2] Q. Xiong, M. K. Morphew, C. L. Schwartz, A. H. Hoenger, and D. N. Mastronarde, *CTF determination and correction for low dose tomographic tilt series,* Journal of Structural Biology **168**, 378 (2009).

[3] G. Zanetti, J. D. Riches, S. D. Fuller, and J. A. G. Briggs, *Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging,* Journal of Structural Biology **168**, 305 (2009).

[4] M. Eibauer, C. Hoffmann, J. M. Plitzko, W. Baumeister, S. Nickell, and H. Engelhardt, *Unraveling the structure of membrane proteins in situ by transfer function corrected cryo-electron tomography,* Journal of Structural Biology **180**, 488 (2012).

[5] A. Philippsen, H.-A. Engel, and A. Engel, *The contrast-imaging function for tilted specimens,* Ultramicroscopy **107**, 202 (2007).

[6] L. M. Voortman, S. Stallinga, R. H. M. Schoenmakers, L. J. van Vliet, and B. Rieger, *A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM,* Ultramicroscopy **111**, 1029 (2011).

[7] L. M. Voortman, E. M. Franken, L. J. van Vliet, and B. Rieger, *Fast, spatially varying CTF correction in TEM,* Ultramicroscopy **118**, 26 (2012).

[8] H. Winkler and K. A. Taylor, *Focus gradient correction applied to tilt series image data used in electron tomography,* Journal of Structural Biology **143**, 24 (2003).

[9] I. G. Kazantsev, J. Klukowska, G. T. Herman, and L. Cernetic, *Fully three-dimensional defocus-gradient corrected backprojection in cryoelectron microscopy,* Ultramicroscopy **110**, 1128 (2010).

[10] G. J. Jensen and R. D. Kornberg, *Defocus-gradient corrected back-projection,* Ultramicroscopy **84**, 57 (2000).

[11] P. B. Rosenthal and R. Henderson, *Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy,* Journal of molecular biology **333**, 721 (2003).

[12] P. Fechter, C. Chevalier, G. Yusupova, M. Yusupov, P. Romby, and S. Marzi, *Ribosomal initiation complexes probed by toeprinting and effect of trans-acting translational regulators in bacteria,* in *Riboswitches,* Methods in Molecular Biology, Vol. 540, edited by A. Serganov (Humana Press, 2009) pp. 247–263.

[13] A. Dümmler, A.-M. Lawrence, and A. De Marco, *Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors,* Microbial cell factories **4**, 34 (2005).

[14] J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, *Computer visualization of three-dimensional image data using IMOD,* Journal of Structural Biology **116**, 71 (1996).

[15] D. N. Mastronarde, *Dual-axis tomography: An approach with alignment methods that preserve resolution,* Journal of Structural Biology **120**, 343 (1997).

[16] Y. Chen, S. Pfeffer, T. Hrabe, J. M. Schuller, and F. Förster, *Fast and accurate reference-free alignment of subtomograms,* Journal of Structural Biology **182**, 235 (2013).

[17] H. Elmlund, D. Elmlund, and S. Bengio, *PRIME: Probabilistic initial 3D model generation for single-particle cryo-electron microscopy,* Structure **21**, 1299 (2013).

[18] D. Nicastro, C. Schwartz, J. Pierson, R. Gaudette, M. E. Porter, and J. R. McIntosh, *The molecular architecture of axonemes revealed by cryoelectron tomography,* Science **313**, 944 (2006).

[19] M. Vulović, R. B. G. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, H. Rullgård, O. Öktem, and B. Rieger, *Image formation modeling in cryo-electron microscopy,* Journal of Structural Biology **183**, 19 (2013).

[20] J. J. Kuffner, *Effective sampling and distance metrics for 3D rigid body path planning,* in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2004)*, Vol. 4 (2004) pp. 3993–3998.

[21] M. Vulović, L. M. Voortman, L. J. van Vliet, and B. Rieger, *When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM,* Ultramicroscopy **136**, 61 (2014).

[22] M. Vulović, B. Rieger, L. J. van Vliet, A. J. Koster, and R. B. G. Ravelli, *A toolkit for the characterization of CCD cameras for transmission electron microscopy,* Acta Crystallographica Section D **66**, 97 (2010).

[23] G. McMullan, S. Chen, R. Henderson, and A. R. Faruqi, *Detective quantum efficiency of electron area detectors in electron microscopy,* Ultramicroscopy **109**, 1126 (2009).

[24] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, *UCSF chimera - a visualization system for exploratory research and analysis,* Journal of computational chemistry **25**, 1605 (2004).

[25] S. H. W. Scheres and S. Chen, *Prevention of overfitting in cryo-EM structure determination,* Nature Methods **9**, 853 (2012).

[26] J. A. G. Briggs, J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H.-G. Kräusslich, *Structure and assembly of immature HIV,* Proceedings of the National Academy of Sciences **106**, 11090 (2009).

[27] L. Zhang and G. Ren, *IPET and FETR: Experimental approach for studying molecular structure dynamics by cryo-electron tomography of a single-molecule structure,* PLoS ONE **7**, e30249 (2012).

[28] A. Bartesaghi, F. Lecumberry, G. Sapiro, and S. Subramaniam, *Protein secondary structure determination by constrained single-particle cryo-electron tomography,* Structure **20**, 2003 (2012).

[29] F. K. M. Schur, W. Hagen, A. de Marco, and J. A. G. Briggs, *Determination of protein structure at 8.5Å resolution using cryo-electron tomography and sub-tomogram averaging,* Journal of Structural Biology **184**, 394 (2013).

[30] A. F. Brilot, J. Z. Chen, A. Cheng, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher, R. Henderson, and N. Grigorieff, *Beam-induced motion of vitrified specimen on holey carbon film,* Journal of Structural Biology **177**, 630 (2012).

[31] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, and Y. Cheng, *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM,* Nature methods **10**, 584 (2013).

[32] X. Li, S. Q. Zheng, K. Egami, D. A. Agard, and Y. Cheng, *Influence of electron dose rate on electron counting images recorded with the K2 camera,* Journal of Structural Biology **184**, 251 (2013).

[33] M. G. Campbell, A. Cheng, A. F. Brilot, A. Moeller, D. Lyumkis, D. Veesler, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher, *et al.*, *Movies of ice-embedded particles enhance resolution in electron cryo-microscopy,* Structure **20**, 1823 (2012).

[34] X.-c. Bai, I. S. Fernandez, G. McMullan, and S. H. Scheres, *Ribosome structures to near-atomic resolution from thirty thousand cryo-em particles,* Elife **2** (2013).

[35] M. Vulović, E. Franken, R. B. G. Ravelli, L. J. van Vliet, and B. Rieger, *Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy,* Ultramicroscopy **116**, 115 (2012).

6

Outlook

What we call the beginning is often the end
And to make an end is to make a beginning.
The end is where we start from.

THOMAS STEARNS ELIOT

I think that only daring speculation can lead us further
and not accumulation of facts.

ALBERT EINSTEIN

This thesis presents fast algorithms for the computation and correction of spatially varying contrast transfer functions (CTF) in cryo-electron tomography (CET) [1, 2]. We introduced estimations of the application limits of important approximations in CET such as the projection assumption (PA), the weak-phase object approximation (WPOA) and combinations thereof, among which the thick phase grating approximation [3]. By applying our theoretical knowledge of the CTF and Fresnel propagation, we improved CET simulations using InSilicoTEM [4]. We determined the influence of inaccurate defocus estimation on the applicability of CTF correction and presented an extended acquisition scheme to achieve the desired defocus accuracy under realistic conditions. The extended acquisition scheme was used to acquire tilt-series of ribosomes on a Titan microscope (FEI Company) at NeCEN. Combined with developed defocus estimation procedures, we precisely determined the defocus of each macromolecule which was then used for tilted CTF correction. Subtomogram averaging was used to further improve the resolution. To conclude, we demonstrated how simulations can be used to study the influence of different processing steps and the number of subtomograms on the achievable resolution. We showed that a comparison between experiment and simulation allowed us to identify the factors that limit the resolution as well as the effect of tilted CTF correction [5].

In this final chapter, we draw conclusions of the presented research, highlight some important developments in the field of CET and give recommendations for future work.

6.1. *CTF correction in cryo-electron tomography*

The theoretical background of CTF correction with a spatially varying defocus is discussed in this thesis. Within the boundaries of the small-angle approximation and WPOA, we present a theoretical framework for CTF correction which covers all possible scenarios: untilted and thin, tilted and thin, and thick. We consider this part of the theory to be finished. Nevertheless, we envision possible extensions beyond the limits of applicability of the CTF, e.g. different contrast mechanisms. And even though we present algorithms that considerably speed-up computation times, further improvements are still desirable.

The differences between the discussed CTF correction methods is also studied in this thesis. We find that a measurable difference between regular and tilted CTF correction occurs at resolutions which are currently only achievable on biological specimens using subtomogram averaging (e.g. ~ 1 nm). Furthermore, we conclude that differences between tilted CTF correction and full 3D CTF correction

occur at resolutions which are currently reserved for single-particle averaging (SPA) (e.g. ~ 0.3 nm).

In this thesis we show that using a Fourier domain representation of the spatially varying CTF is an insightful and efficient approach for modeling and correcting the CTF. The speed-up we achieved using newly developed fast algorithms was essential for the conducted research. Previously, computation times were in the order of days, resulting in significant bottlenecks. By reducing the computation times to the order of hours allowed us to conduct more research, but this speed-up can also lead to more user friendly tools. Furthermore, the fast simulation and correction for the CTF opens up the possibility to include CTF correction within an iterative reconstruction algorithm.

In order to get a better understanding of the image formation process in CET we investigated the limits of the PA and the WPOA. We find that for resolutions typical in subtomogram averaging (up to ~ 1 nm), both approximations are applicable. For reconstructions from single-particle analysis, however, both the PA and WPOA may be violated depending on the achieved resolution and the size of the macromolecule. Using 3D CTF correction it is possible to reconstruct structures for which the PA was not applicable. Reconstructing structures which do not fulfill the WPOA, however, still requires more investigation. The WPOA can be easily modified to allow for a small, proportional amount of amplitude contrast. This approach fails, however, for example in phase-contrast X-ray imaging where the strong attenuation of the signal due to amplitude contrast complicates phase retrieval [6, 7].

We quantified the influence of different CTF correction and defocus estimation methods for subtomogram averaged CET. This quantification was, however, largely based on simulated data. A similar quantification on experimental data has not yet been conducted and is not available in the literature. The theory concerning global tilted CTF correction was introduced by Philippsen *et al.* [8], but no experimental verification was presented. Strip or patch based CTF correction methods have been presented by Fernández *et al.* [9], Zanetti *et al.* [10], Xiong *et al.* [11]. In these studies, however, the improvement from untilted CTF correction to tilted CTF correction was not quantified. Recently, Eibauer *et al.* [12] convincingly showed the influence of accurate defocus estimation on CTF correction. They could, however, only quantify this difference by comparing with a known molecular structure. Finally, Schur *et al.* [13] recently demonstrated that sub-nanometer resolution is achievable with accurate defocus estimation and CTF correction. An interesting addition would be to quantify the influence of the different processing

steps in a quantitative manner, similar to the analysis of simulated data presented here. The advantage of 3D CTF correction, has to our knowledge, not yet been quantified on experimental data.

Here, we aimed at quantifying the improvements of tilted CTF correction and accurate defocus estimation in simulation and experiment. This turned out to be challenging on experimental data due to other limiting factors such as tilt-series alignment. Nevertheless, we think that tilt-series alignment is not a fundamental problem, but rather depends on different experimental conditions. Therefore, we recommend future work to be focused in areas which include alignment, but also concerning technological developments, automation, image formation models and tomography of materials which are less sensitive to radiation damage.

6.2. *Improving alignment with subtomogram averaging*

One of the conclusions of this thesis is that tilt-series alignment can severely limit the resolution in subtomogram averaged CET. With respect to the alignment accuracy, very promising results have been obtained by combining the strengths of SPA and subtomogram averaging techniques [14, 15]. The rationale is that these two methods share the same goal, an averaged 3D reconstruction, but each has its strengths and weaknesses. In subtomogram averaging, particle picking and orientation estimation is done in 3D and, therefore, more robust. The drawback is that a suboptimal tilt-series alignment cannot be alleviated in later alignment steps. SPA has the advantage that alignment and orientation refinement is performed directly on the original projections. The difficulty lies in estimating ab initio the relative orientation of all particles and coping with superimposed structures in the projection.

Consolidating SPA and subtomogram averaging will benefit from both techniques. The accurate orientation estimation of tomography resolves the ab initio estimate required for SPA and the refinement procedures of SPA can resolve the need for two separate alignment steps in subtomogram averaged tomography. Radiation damage effects, typically associated with CET, can then easily be alleviated by using a subset of the original projections to compute the final average.

In this thesis we did not investigate the use of 3D CTF correction for SPA. The concept of 3D CTF correction (or Ewald sphere correction) is not new for SPA, but has remained largely theoretical. The fast 3D CTF correction method presented in this thesis, together with the recent improvements in resolution (primarily due to better microscopes and detectors), opens up an interesting research opportunity.

6.3. *Technological developments*

Shot noise (obeying Poisson statistics) is currently the major limitation of CET. The signal-to-noise ratio (SNR) can be improved by employing subtomogram averaging, but this is limited by practical constraints. Furthermore, the low SNR of the projections also influences defocus estimation, tilt-series alignment and subtomogram alignment. The current development of direct electron detectors (DEDs), capable of operating in counting mode, will therefore have a significant impact on the resolution improvement. Even though DEDs are impractical for anything but extreme-low-dose imaging, and it is still uncertain what the lifetime of such detectors is, for dedicated low-dose cryo applications the promise is a significant increase in SNR.

In this thesis we predict that an ideally functioning DED in counting mode can reach the same resolution with half the number of particles when compared to a Falcon I DED. There are, however, more advantages of this new generation of DED. Sub-frame alignment allows correction for beam-induced motion blur. Super-resolution counting mode effectively increases the pixel count of a detector [16]. Finally, an increased detective quantum efficiency (DQE) can greatly reduce the limiting influence of radiation damage.

Another promising development for CET is the phase plate for TEM. The difficulty of phase-contrast imaging is the loss of contrast at low spatial frequencies. High frequency information can be recovered by defocussing and subsequent CTF correction. Only when a very large defocus is used, low frequency contrast is generated but this inevitably damps the high frequencies. An ideal phase plate effectively solves this problem by introducing a $\pi/2$ phase-shift between the diffracted and undiffracted beam. This allows in-focus imaging with optimal contrast for low and high spatial frequencies. Currently, phase plates are not routinely available or used due to manufacturing difficulties, limitations of the microscope, as well as undesirable imaging artifacts. Nevertheless, we predict that phase plates will play an important role for cellular tomography and localization purposes in subtomogram averaging applications.

The CTF correction methods presented in this thesis can also be used, mutatis mutandis, for tomography applications with a phase plate. Due to the inherent defocus gradient that occurs during tilting, CTF correction is still necessary for the high spatial frequencies. Furthermore, the tools presented here to estimate the application limits of the projection assumption, weak-phase object approximation, regular CTF correction, tilted CTF correction, 3D CTF correction and defocus estimation are not influenced by imaging with a phase plate.

The results and methods from this thesis are useful when estimating which factors limit the resolution. When phase plates prove to be a reliable tool in achieving higher resolutions it is even more valuable to know which are the resolution limiting factors. In general, we find that for resolutions beyond ~ 1 nm, tilted CTF correction is a necessary technique, with or without phase plate.

6.4. *Automation and high-throughput*

At this time, CET is a slow, and only partially automated technique. Subtomogram averaging requires an even greater amount of manual intervention and clever fine-tuning of alignment iterations. Defocus estimation and CTF correction require specialized tools in the hands of experienced users. This makes subtomogram averaged, CTF corrected CET a highly specialized, painstakingly slow process. Computation time, however, is thanks to the fast algorithms described in this thesis, as well as increasingly better computer hardware, no longer the biggest bottleneck. The following steps should be improved to increase the total throughput.

Tilt-series alignment needs to be a fully automated procedure, preferably integrated with the acquisition software. It would then be possible to either alert the user that there are too few markers to align or automatically search for a more suitable area of interest. Defocus estimation on exposure images and/or defocus estimation using the extended acquisition protocol should also be integrated in the acquisition software. Estimating the defocus, and from the defocus the specimen orientation, during the acquisition can improve the stability and repeatability of the defocus on the exposure, as well as improve the reliability of image tracking. When the acquisition software is combined with tilt-series alignment and defocus estimation, the output can be an aligned tilt-series in which each recorded image is labeled with the correct defocus value, tilt-axis and tilt-angle.

The CTF correction methods presented in this thesis are currently only available as a Matlab (The MathWorks, USA) toolbox. This makes them not user friendly enough to be used for routine work. More effort is required to shape the algorithms presented here into an all-round, user friendly, CTF correction tool.

Primarily because subtomogram averaging is still an emerging method, a plethora of subtomogram averaging software packages exists. The difficulty of automating subtomogram averaging beyond what is currently available lies in the diversity of samples and their need for tailored solutions. Nevertheless, the solutions that are developed allow more and more structures to be studied.

6.5. *Image contrast*

Interpretation of image contrast in CET is currently limited to phase contrast (and possibly a small fixed ratio of amplitude contrast). In order to get the maximum of information out of the low number of electrons available for imaging, a better understanding of the interaction between the electron beam and the specimen is required. This can also aid in concluding whether aberration correctors and energy filters can improve the SNR.

Due to the low SNR of CET, iterative tomographic reconstruction schemes are typically avoided. When the image contrast is fully understood, however, it becomes possible to use a priori information in combination with iterative reconstruction methods.

6.6. *Other materials*

The difficulties associated with CET for a large part originate from the sensitivity of biological materials to radiation. The CTF correction methods developed in this thesis, however, are also applicable to radiation resistant materials. Carbon nanotubes, for example, form 3D structures which are largely resistant to exposure with 80 keV electrons. Much higher resolutions can be obtained on these structures, which increases the necessity of tilted CTF and 3D CTF correction.

6.7. *Recommendations*

Acquisition and Reconstruction — Software for tomographic acquisitions should integrate tilt-series alignment and defocus estimation to improve the stability and repeatability of tilt-series acquisition. When identical particles are averaged, SPA and subtomogram averaging techniques should be combined to provide the best results. Use of 3D CTF correction methods for high resolution SPA reconstructions (possibly iterative reconstructions) is worth an extra investigation.

Image Formation — More research of image formation is required since the current understanding is not sufficient to decide about hardware developments and the settings of all acquisition parameters. More research of image formation is required in order to decide about hardware developments and the settings of all acquisition parameters. The newest generation of direct electron detectors, however, show such promising results that they warrant use even without a complete understanding of contrast formation. When image formation is better understood, the use of model-based and constrained iterative reconstruction schemes should be re-evaluated.

References

[1] L. M. Voortman, S. Stallinga, R. H. M. Schoenmakers, L. J. van Vliet, and B. Rieger, *A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM,* Ultramicroscopy **111**, 1029 (2011).

[2] L. M. Voortman, E. M. Franken, L. J. van Vliet, and B. Rieger, *Fast, spatially varying CTF correction in TEM,* Ultramicroscopy **118**, 26 (2012).

[3] M. Vulović, L. M. Voortman, L. J. van Vliet, and B. Rieger, *When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM,* Ultramicroscopy **136**, 61 (2014).

[4] M. Vulović, R. B. G. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, H. Rullgård, O. Öktem, and B. Rieger, *Image formation modeling in cryo-electron microscopy,* Journal of Structural Biology **183**, 19 (2013).

[5] L. M. Voortman, M. Vulović, M. Maletta, A. Voigt, E. M. Franken, A. Simonetti, P. J. Peters, L. J. van Vliet, and B. Rieger, *Quantifying resolution improvements in subtomogram averaged cryo-electron tomography,* Journal of Structural Biology (submitted).

[6] J. Klukowska and G. T. Herman, *Reconstruction from microscopic projections with defocus-gradient and attenuation effects,* in *Computational Methods for Three-Dimensional Microscopy Reconstruction* (Springer, 2014) pp. 157–186.

[7] J. Oton, C. Sorzano, E. Pereiro, J. Cuenca-Alba, R. Navarro, J. M. Carazo, and R. Marabini, *Image formation in cellular x-ray microscopy,* Journal of structural biology **178**, 29 (2012).

[8] A. Philippsen, H.-A. Engel, and A. Engel, *The contrast-imaging function for tilted specimens,* Ultramicroscopy **107**, 202 (2007).

[9] J. J. Fernández, S. Li, and R. A. Crowther, *CTF determination and correction in electron cryotomography,* Ultramicroscopy **106**, 587 (2006).

[10] G. Zanetti, J. D. Riches, S. D. Fuller, and J. A. G. Briggs, *Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging,* Journal of Structural Biology **168**, 305 (2009).

[11] Q. Xiong, M. K. Morphew, C. L. Schwartz, A. H. Hoenger, and D. N. Mastronarde, *CTF determination and correction for low dose tomographic tilt series,* Journal of Structural Biology **168**, 378 (2009).

[12] M. Eibauer, C. Hoffmann, J. M. Plitzko, W. Baumeister, S. Nickell, and H. Engelhardt, *Unraveling the structure of membrane proteins in situ by transfer function corrected cryo-electron tomography,* Journal of Structural Biology **180**, 488 (2012).

[13] F. K. M. Schur, W. Hagen, A. de Marco, and J. A. G. Briggs, *Determination of protein structure at 8.5Å resolution using cryo-electron tomography and sub-tomogram averaging,* Journal of Structural Biology **184**, 394 (2013).

[14] L. Zhang and G. Ren, *IPET and FETR: Experimental approach for studying molecular structure dynamics by cryo-electron tomography of a single-molecule structure,* PLoS ONE **7**, e30249 (2012).

[15] A. Bartesaghi, F. Lecumberry, G. Sapiro, and S. Subramaniam, *Protein secondary structure determination by constrained single-particle cryo-electron tomography,* Structure **20**, 2003 (2012).

[16] X. Li, S. Q. Zheng, K. Egami, D. A. Agard, and Y. Cheng, *Influence of electron dose rate on electron counting images recorded with the K2 camera,* Journal of Structural Biology **184**, 251 (2013).

List of Publications

4. **L. M. Voortman**, M. Vulović, M. Maletta, A. Voigt, E. M. Franken, A. Simonetti, P. J. Peters, L. J. van Vliet, and B. Rieger, *Quantifying resolution improvements in subtomogram averaged cryo-electron tomography,* Journal of Structural Biology *submitted*

3. M. Vulović[†], **L. M. Voortman**[†], L. J. van Vliet, and B. Rieger, *When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM,* Ultramicroscopy **136**, 61 (2014).

2. **L. M. Voortman**, E. M. Franken, L. J. van Vliet, and B. Rieger, *Fast, spatially varying CTF correction in TEM,* Ultramicroscopy **118**, 26 (2012).

1. **L. M. Voortman**, S. Stallinga, R. H. M. Schoenmakers, L. J. van Vliet, and B. Rieger, *A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM,* Ultramicroscopy **111**, 1029 (2011).

[*]These authors contributed equally to this work

Summary

CTF Correction *in* Cryo-Electron Tomography

SIMULATION, EXPERIMENTS AND RECONSTRUCTION

Nanometer resolution inside the cell will allow us to study the fundamentals of life at the smallest scale. This thesis addresses what is needed to obtain this resolution using cryo-electron tomography (CET).

CET is a microscopy modality with the unique potential to visualize proteins, protein-complexes and other molecular assemblies in a close-to-native environment at a high resolution in three dimensions. In CET, a thin specimen embedded in vitreous ice is tilted in the electron beam to acquire projections under different angles. The primary contrast mechanism is phase contrast which is obtained by intentionally defocussing the specimen. The contrast transfer function (CTF) describes how aberrations, such as defocussing, generate detectable intensity contrast. The CTF is an oscillating function of spatial frequency, resulting in contrast inversions at certain spatial frequencies. To interpret structures at a resolution beyond the first zero-crossing, it is necessary to correct for the CTF. In this thesis we answer the questions: how can we model the CTF for tomographic geometries, what is the influence of CTF correction, which processing steps need to be improved to fully exploit CTF correction in combination with subtomogram averaging, and how big is the improvement in resolution?

This thesis presents fast and efficient algorithms for both forward modeling and correction of the CTF for tilted geometries of various thicknesses, as well as methods to model the specimen-beam interaction. To avoid a brute-force multislice procedure to model the specimen-beam interaction, we study the influence of the projection assumption, the weak-phase object approximation, and the thick-phase grating approximation, as well as their limits of applicability. Fast algorithms for computing and correcting the CTF in tilted geometries are mandatory for practical use. Our algorithm reduces the computation time for a tilt-series from ~ 100 hours down to ~ 45 minutes. Using simulations, we also

study how different CTF models influence the projections and what the influence of CTF correction is on the final reconstruction.

We quantify the influence of the developed CTF correction methods in subtomogram averaged CET. Subtomogram averaging is the solution to raise the signal-to-noise ratio for high spatial frequencies above the noise floor. To achieve the required defocus estimation accuracy under realistic experimental conditions, we present an extended acquisition scheme in combination with a previously developed defocus estimation procedure. Using simulations and experimental data of ribosomes, acquired on a Titan microscope (FEI Company) at the NeCEN, we study the influence on the achievable resolution of different processing steps, including CTF correction, as well as the number of subtomograms. A comparison of simulations and experiments allows us to identify the factors that limit the resolution as well as the effect of tilted CTF correction. We obtained a final average using 3198 ribosomes with a resolution of 2.2 nm on the experimental data. Our simulations suggest that with the same number of particles a resolution of 1.2 nm could be achieved by improving the tilt-series alignment.

Lenard M. Voortman

Samenvatting

CTF-Correctie *in* Cryo-Elektronentomografie

SIMULATIES, EXPERIMENTEN EN RECONSTRUCTIE

Nanometer resolutie in de cel zal ons toelaten om op de kleinste schaal de fundamenten van het leven te bestuderen. Dit proefschrift beschrijft wat nodig is om deze resolutie met cryo-elektronentomografie (CET) te verkrijgen.

CET is een microscopie modaliteit met de unieke mogelijkheid om eiwitten, eiwitcomplexen en andere moleculaire assemblages te visualiseren in een bijna natuurlijke omgeving met een hoge resolutie in drie dimensies. Voor CET wordt een dun, in amorf ijs ingebed, preparaat gekanteld in de elektronenbundel om projecties onder verschillende hoeken te verkrijgen. Het primaire contrast mechanisme is fasecontrast dat wordt verkregen door opzettelijk het preparaat te defocusseren. De contrast transfer functie (CTF) beschrijft hoe aberraties, waaronder defocussering, detecteerbaar intensiteitcontrast genereren. De CTF is een oscillerende functie van spatiële frequentie, wat resulteert in contrast inversies op bepaalde spatiële frequenties. Om structuren te interpreteren met een resolutie voorbij de eerste nuldoorgang, is het noodzakelijk om te corrigeren voor de CTF. In dit proefschrift beantwoorden we de vragen: hoe kunnen we de CTF voor tomografische geometrieën modelleren, wat is de invloed van CTF-correctie, welke processtappen moeten worden verbeterd om CTF-correctie ten volle te benutten in combinatie met subtomogram-middeling, en hoe groot is de verbetering van de resolutie?

Dit proefschrift introduceert snelle en efficiënte algoritmen voor zowel voorwaartse modellering als correctie van de CTF voor gekantelde geometrieën van verschillende diktes, alsook methodes om de interactie tussen preparaat en elektronenbundel te modelleren. Om te voorkomen dat een meerplaksaanpak met brute kracht gebruikt moet worden voor het modelleren van de preparaat-bundelinteractie, bestuderen we de invloed van de projectieveronderstelling, de zwakke-fase-objectbenadering, en de dikke-fase-rasterbenadering, evenals hun

grenzen van toepasbaarheid. Snelle algoritmen voor het berekenen en het corrigeren van de CTF in gekantelde geometrieën zijn een vereiste voor praktisch gebruik. Ons algoritme reduceert de rekentijd voor een kantelserie van ~ 100 uur tot slechts ~ 45 minuten. Met behulp van simulaties, bestuderen we ook hoe de verschillende CTF modellen de projecties beïnvloeden en wat de invloed van CTF-correctie is op de uiteindelijke reconstructie.

We kwantificeren de invloed van de ontwikkelde CTF-correctiemethoden in subtomogram-gemiddelde CET. Subtomogram-middeling is de oplossing om de signaal-ruisverhouding voor hoge spatiële frequenties boven het ruisniveau te brengen. Om de vereiste defocusschattingsnauwkeurigheid onder realistische experimentele omstandigheden te bereiken, presenteren we een uitgebreid acquisitieprotocol in combinatie met een eerder ontwikkelde defocusschattingsmethode. Met simulaties en experimentele data van ribosomen, verkregen met een Titan microscoop (FEI) bij het NeCEN, bestuderen we de invloed op de bereikbare resolutie van verschillende processtappen, waaronder CTF-correctie, alsmede het aantal subtomogrammen. Een vergelijking tussen simulaties en experimenten stelt ons in staat de factoren te identificeren die de resolutie beperken, alsook de invloed van gekantelde CTF-correctie. Gebruikmakend van 3198 ribosomen, verkrijgen we met de experimentele data uiteindelijk een gemiddelde met een resolutie van 2.2 nm. Onze simulaties suggereren dat met hetzelfde aantal deeltjes een resolutie van 1.2 nm bereikt zou kunnen worden door de kantelserie-uitlijning te verbeteren.

Lenard M. Voortman

Acknowledgments

This thesis would not have been possible without the help from others. Lucas, thank you for all the good advice you gave me during the past years, even for the times that I willingly went the other way. Bernd, I savored the moments when my results were 'not wrong' or 'not bad'. You are a great supervisor with a big heart and short comments. Milos, my office buddy, I thoroughly enjoyed working together with you. In the future, I will miss your company during conferences.

Then of course the people at FEI: Remco, Erik and Andreas. Thanks for a fruitful collaboration.

The measurements of Ribosomes would not have been possible without Max.

Bram and Raymond from the LUMC, our meetings always seemed to shift my perspective, thanks for these different viewing angles. But also Frank, who was the first to introduce me to image processing of electron microscopy data. And Roman, thanks for all the good times during conferences.

QI: I would like to thank everyone. Thanks for all the good times, the *dagjes uit*, the dinners and the coffee. Mandy, thanks for all the office related help, as well as the numerous WordFeud games. Ronald, thanks for being the best computer guru ever, but even more for all the help that was not IT related.

Finally, a big thanks to all friends and family. Thanks for all the support, but especially for all the distraction.

En Marjon, natuurlijk!

Curriculum Vitæ

Lenard Maarten Voortman was born in Geldrop, The Netherlands on December 11, 1984. After graduating in 2003 from the Rythovius College (secondary school) in Eersel, he went on to study Applied Physics at Delft University of Technology. He received his Bachelor's degree cum laude in 2006. The following year he spent as full-time board member of the Delftse studenten vereniging Sint Jansbrug. He then continued his studies and received his Master's degree cum laude in 2009 on the subject 'Tilted CTF in Electron Tomography: Forward Modelling & Tomographic Reconstruction'.

He then pursued a PhD on this topic in a joint collaboration between Leiden University Medical Center and Delft University of Technology. The project was part of an industrial partnership program with FEI Company.

After his PhD project he started working as application specialist at DELMIC, a start-up company that produces integrated correlative light and electron microscopes.