Progressive Stochastic Reconstruction Technique for Cryo Electron Tomography

Beata Turoňová

Computer Graphics Group
Saarland University
66123 Saarbrücken, Germany

Thesis for obtaining the title of
Doctor of Engineering Science (Dr.-Ing.)
of the Faculty of Natural Science and Technology I
of Saarland University

Saarbrücken, Germany, 2016

Date of the colloquim

September 15^{th}, 2016

Dean

Prof. Dr. Frank-Olaf Schreyer

Reporters

Prof. Dr. Philipp Slusallek
Prof. Dr. Dr. h.c. mult. Alfred K. Louis

Chairman of the examination board

Prof. Dr. Joachim Weickert

Scientific assistant

Dr. Richard Membarth

To my parents
Naděžda and Bronislav

ABSTRACT

Cryo Electron Tomography (cryoET) plays an essential role in Structural Biology, as it is the only technique that allows us to study the structure of large macromolecular complexes in their close to native environment in *situ*. The reconstruction process is a challenging task as the single-tilt acquisition scheme imposes severe limitation on the input projections. High-resolution protocols such as Subtomogram Averaging (SA) can alleviate some of these limitations. Results of these protocols are highly dependent on the quality of the reconstruction. State-of-the-art methods deliver low-contrast and noisy reconstructions, which complicates their processing during the SA.

In this thesis we focus on improvement of the quality of tomograms in cryoET in order to facilitate their subsequent processing. We propose a Progressive Stochastic Reconstruction Technique (PSRT) – a novel iterative approach to tomographic reconstruction in cryoET that is based on Monte Carlo random walks. We design a progressive scheme to suit conditions present in cryoET and integrate PSRT into the SA pipeline, where it delivers high-contrast reconstructions that significantly improve template-based localization without any loss of high-resolution structural information. Furthermore, we perform a systematic study of the geometry-related acquisition artifacts and draw recommendations regarding the mutual influence of these artifacts and implications to the interpretation of both cryoET and SA experiments.

KURZFASSUNG

Kryoelektronentomographie (cryoET) spielt eine wichtige Rolle in der Strukturbiologie, da sie als die einzige Technik ermöglicht, intakte Zellstrukturen auf der molekülaren Ebene zu untersuchen. Die 3D Rekonstruktion wird durch nicht vollständigen Informationen in Bildaufnahmen des Elektronenmikroskops deutlich kompliziert. Mehrere hochauslösende Verfahren, z.B. Subtomogram Averaging (SA), reduzieren diese Begrenzung. Ihre Ergebnisse hängen jedoch stark von der Rekonstruktionsqualität ab. Die gegenwärtigen Methoden erzeugen Rekonstruktionen, die meistens sehr verrauscht und kontrastarm sind.

Im Rahmen dieser Arbeit befassen wir uns mit der Verbesserung der Rekonstruktionsqualität in cryoET, um die Weiterverarbeitung der Daten zu vereinfachen. Wir stellen eine neue iterative stochastische Rekonstruktionsmethode für cryoET vor, die auf Monte-Carlo Random Walks basiert und so entworfen wurde, dass sie den spezifischen Umständen in cryoET entspricht. In der Kombination mit SA wird gezeigt, dass unsere Methode Rekonstruktionen mit höhem Kontrast erzeugt und damit die automatische Erkennung der Zellstrukturen verbessert ohne die hochauslösenden Informationen über den Zellstrukturen zu verlieren. Weiterhin untersuchen wir die geometrische Rekonstruktionsartefakten in cryoET und deren Einfluss auf die Informationsqualität der Rekonstruktionen.

ACKNOWLEDGEMENTS

The work presented in this thesis would not be possible without support from my family, friends and colleagues.

I would like to thank my supervisor Philipp Slusallek for letting me work on such an interesting topic: The reconstruction problem in electron tomography. His support, guidance and insightful suggestions and ideas greatly helped me in my research during the past years and ultimately improved the work I am presenting in this thesis. I would also like to thank the Max Planck Institute for Informatics and the Grad School at Saarland University for providing funding for my research.

Special thanks go to my main collaborator, adviser, and good friend Lukáš Maršálek for introducing me to the fascinating world of electron tomography, for helping me to overcome the difficulties I had to face, for patiently answering all my questions, and for making the work on this topic such an interesting and enjoyable experience. I would also like to thank all my colleagues and friends from the Computer Graphics chair: Tomáš Davidovič, Javor Kalojanov, Piotr Danilewski, Michala Maršálková, Iliyan Georgiev, Vincent Pegoraro, Richard Membarth, Arsène Pérard-Gayot, and Felix Klein for their help, support, ideas, and most importantly for making my time at the chair such a fun experience.

I would also like to thank my colleagues from the 3D Imaging Group at DFKI: Tim Dahmen, Stefan Nickels, Nico Marniok, and Patrick Trampert for the fruitful Monday discussions, for help with the seminar on tomography, and lessons on clean coding. Thanks also belong to Prof. Louis for his helpful comments and suggestions regarding tomographic reconstruction problems and to Max Maletta, Friedrich Förster, Florian Schur, and John Briggs for providing me with the experimental datasets to test my method on. The computationally demanding Monte Carlo reconstructions would not be possible without our admins Sebastian Wendland, Umar Chikobava, Igor Wozniak, and Davit Hovhannisyan and I would like to thank them not only for their excellent IT support but also for (almost) silently overlooking the fact that I was not installing updates for months. I would also like to thank the anonymous reviewers for their comments and questions that helped to improve my publications.

Special thanks belong to my mother and my sister for their endless moral support, their motivation, and for being here for me all the time. Thanks also go to all my friends for the distractions from my work that were always welcomed. Last but not least, I would like to thank my partner Martin Šimonovský for his insightful thoughts and challenging questions regarding my work that often led to deepening my knowledge, for patiently listening to me while I was talking (endlessly) about my work, for helping me whenever I needed it, for cheering me up on the bad days, and most importantly for all the fun we had.

CONTENTS

Abstract **I**

Acknowledgements **V**

List of Figures **X**

1 Introduction **1**
1.1 Original Contributions 3
1.2 Thesis Outline 4
1.2.1 Terminology 4

2 Cryo Electron Tomography **5**
2.1 Specimen Preparation 6
2.2 Image Acquisition in TEM 6
2.2.1 Electron Imaging 6
2.2.2 Image Acquisition 7
2.3 Tilt-Series Preprocessing 9
2.3.1 Alignment 9
2.3.2 Contrast Transfer Function Correction 9
2.4 3D Reconstruction 10
2.5 Subtomogram Averaging 10
2.5.1 Localization 11
2.5.2 Classification 11
2.5.3 Alignment and Averaging 11

3 Tomographic reconstruction in cryoET **13**
3.1 3D Tomography 14
3.2 Tomographic Reconstruction Algorithms 15
3.2.1 Mathematical Interpretation 15
3.2.2 Geometric Interpretation 17
3.3 Resolution and Quality Assessment 19

4 Monte Carlo Minimization **21**
4.1 Monte Carlo Methods 21
4.1.1 Uniform Sampling 22
4.1.2 Importance Sampling 22
4.1.3 Metropolis-Hasting Sampling 23
4.2 Monte Carlo Approach for Reconstructions 24
4.2.1 Previous Work 24
4.2.2 Geometric Interpretation 25

5 Progressive Stochastic Reconstruction Technique - PSRT 27
5.1 Algorithm Overview . . . 27
5.1.1 Samples . . . 28
5.1.2 Random Walk . . . 29
5.1.3 Energy of the Sample . . . 30
5.2 Progressive Reconstruction Scheme . . . 30
5.3 Memory Efficient Solution for Specimen-Level Interior Problem . . . 32
5.4 Locality of PSRT . . . 32
5.4.1 Region of Interest Reconstruction . . . 32

6 Implementation 35
6.1 Volume representation . . . 35
6.2 Evaluation of the Objective Function . . . 36
6.2.1 Sample Representation . . . 36
6.2.2 Projection Computation . . . 37
6.3 Energy Computation . . . 37
6.4 Random Walks Behavior on the Boundaries . . . 38
6.5 Memory Efficient Solution for Specimen-Level Interior Problem . . . 38
6.6 Region of Interest Reconstruction . . . 39

7 Properties of PSRT 41
7.1 Coarse-to-fine Reconstruction . . . 41
7.1.1 2D Synthetic Dataset . . . 42
7.1.2 3D Synthetic Dataset . . . 45
7.1.3 Noise . . . 46
7.1.4 Enhanced Contrast for Low-Defocus Data . . . 47
7.2 Memory Efficient Solution to Interior Problem . . . 49
7.3 Region of Interest Reconstruction . . . 49
7.3.1 ROI Using a Binary Mask . . . 52
7.4 Memory Requirements . . . 52
7.5 Computation Time . . . 54
7.6 Parameter Setup . . . 55
7.6.1 Automatization . . . 56

8 Integration to Subtomogram Averaging Pipeline 57
8.1 Subtomogram Averaging Pipeline . . . 57
8.2 Experimental Data . . . 58
8.3 Reconstruction . . . 58
8.4 Localization . . . 59
8.5 Final Alignment . . . 60
8.5.1 Stability . . . 61
8.6 Reconstruction on 4k Data . . . 61

9 Parallelization 65
9.1 Parallelization on the Reconstructed Area . . . 65
9.1.1 Validation on Experimental Data . . . 67
9.1.2 Computation Times . . . 68
9.2 Parallelization of Random Walks . . . 69

10 Geometric Artifacts in CryoET 71
10.1 Missing Wedge . . . 72

10.2 Missing Information 73
10.3 Interior Problem 74
10.4 Subtomogram Averaging 76
10.5 Experimental Evaluation 77
10.5.1 Missing Wedge 77
10.5.2 Missing Information 79
10.5.3 Interior Problem 80
10.5.4 Influence on Reconstructed Values 81
10.5.5 Experimental Data 90
10.6 Conclusions 90

11 Conclusions 95
11.1 Progressive Stochastic Reconstruction Technique 95
11.2 Study on Geometric Artifacts in CryoET 97
11.3 Perspectives on Tomographic Reconstructions in CryoET 98

A Supplemental Material for Geometric Artifacts in CryoET 101
A.1 Fourier Spectra for MW Reconstructions 101
A.2 Extensions for IP 101
A.3 Additional Interior Problem Results 101
A.4 Filtering 103

Bibliography 109

LIST OF FIGURES

2.1 Cryo Electron Tomography workflow 5
2.2 Scheme of TEM and image formation 7
2.3 Single-tilt scheme of a specimen tilting in a TEM. 8

3.1 Scheme of the coordinate system and notation used in this work 13
3.2 Radon transform and iterative methods 16
3.3 Geometric reconstruction scheme 17
3.4 Fourier based reconstruction scheme 18

5.1 PSRT scheme diagram 28
5.2 Seed distribution 31
5.3 Region of interest reconstruction 33

6.1 Sample subdivision scheme 36
6.2 Extension scheme 39
6.3 Creation of ROIs from a binary mask 39

7.1 2D reconstructions with constant sample size 43
7.2 2D reconstructions using coarse-to-fine scheme 43
7.3 2D reconstructions of Shepp-Logan phantom for complete projection set 44
7.4 2D reconstructions of Shepp-Logan phantom in the presence of missing wedge 44
7.5 Progress of PSRT reconstruction on 3D synthetic dataset 45
7.6 Comparison of reconstructed structure of synthetic ribosome 46
7.7 Graphs of achieved rRMSE and SNR for different noise levels 47
7.8 Ribosome structures at different noise levels 47
7.9 Results for low-defocus data 48
7.10 Suppression of interior problem artifacts 50
7.11 Region of interest reconstructions 51
7.12 ROI reconstruction using binary mask 53
7.13 Reconstruction times for different approaches within PSRT. 55

8.1 Subtomogram averaging pipeline 58
8.2 Localization histograms 59
8.3 Localization results 60
8.4 Contrast comparison 61
8.5 Progress of 80S ribosome resolution 62
8.6 Reconstruction results of PSRT-SA compared to WBP and SART 62
8.7 Reconstruction stability 63
8.8 Localization histograms for PSRT and INFR 64

9.1 Splitting of the reconstructed area along y-axis. 66
9.2 Artifacts on the splitting boundaries 67

9.3 Localization results for parallel scheme reconstruction 68
9.4 Paralellization speed-up 69

10.1 Scheme of the missing wedge problem 72
10.2 Scheme of the missing information problem 73
10.3 Scheme of the interior problem 75
10.4 Tomograms in presence of the missing wedge problem 78
10.5 Tomograms in presence of the missing information problem 79
10.6 Tomograms in presence of the interior problem 80
10.7 Graphs of reconstructed values in presence of the missing wedge problem 83
10.7 (Continued) 84
10.8 Graphs of reconstructed values in presence of various geometric artifacts 86
10.8 (Continued) 87
10.9 Graphs of reconstructed values in presence of the missing wedge problem 88
10.10 Graphs of reconstructed values in presence of various geometric artifacts 89
10.11 Tomograms of experimental dataset 91

A.1 Fourier spectra of tomograms with MW 102
A.2 Effects of extension in x dimension 103
A.3 Comparison between two approaches for suppressing IP artifacts 104
A.4 Results of filtering in presence of MI problem 105
A.5 Results of filtering in presence of IP problem 105
A.6 Results of filtering on experimental data 106
A.7 Effects of filtering on reconstructions obtained with SART 107

CHAPTER 1

INTRODUCTION

The existence of unseen microorganisms was hypothesized as early as in the 6th century BC but it could not be confirmed until the 17th century, when the desire to observe the things that are not visible with naked eye lead to invention of the first Visible Light Microscopes (VLM). While the inventor of the first VLM is disputable, the credit for popularizing microscopes goes to Antonie van Leeuwenhoek, who among the first ones observed bacteria and other microorganisms using his own improved design of single-lens VLM and thereby became on of the pioneers in the field of microbiology. Almost two hundreds years later, Ernst Abbe derived the formula according to which the resolving power of a VLM is limited by the wavelength of visible light. The connection between the wavelength and achievable resolution was later confirmed also by Lord Rayleigh, whose criterion for resolution limit is nowadays prevalent. It estimates the best achievable resolving power of VLM with conventional lenses to be roughly 200 nm, which albeit impressive, is not enough to resolve the structures at near atomic level, given the range of atoms' radii being between $3-0.3$Å. It was not until 1925 when Louis de Broglie laid the groundwork for overcoming this limitation by postulating the wave-like character of electrons with their wavelength being significantly shorter than that of visible light, allowing (at least theoretically) for higher resolution. In 1931, Ernst Ruska and Max Knoll developed the first Transmission Electron Microscope (TEM) and within a year the resolution limit of VLM was surpassed [Williams and Carter, 2009]. Further development followed and already in 1945 the milestone resolution of 1 nm was achieved. During the following decades, TEM become widely used in material science as well as in biology where it allowed for significant discoveries in the field of Structural Biology [Frank, 2006].

Structural Biology studies the structures of biological macromolecules in order to gain better understanding of their functionality as well as interactions. This is of great importance, as the macromolecules perform most of the functions of cells, whereas their ability to carry out specific actions depends on their current conformational state. Thus, understanding the influence of the current state of the macromolecular structure on its functionality can provide valuable information about the molecular mechanisms within a cell. There are numerous imaging techniques that are used in Structural Biology. The low-resolution methods, such as VLM or X-ray tomography, can provide cellular context however without sufficient amount of details due to the limited resolution. By contrast, the high-resolution techniques, such as X-ray crystallography, Nuclear Magnetic Resonance (NMR), electron crystallography, or Single Particle Analysis (SPA), can provide us with detailed structures of isolated complexes or crystals but cannot reveal how the structures function or interact within a cell [Lučić et al., 2005]. This is caused by the specimen preparation techniques that often require crystallization and/or purification, which leads to destruction of any context information. For a long time, there was a gap between the low-detail contextual techniques with resolution up to 30 nm and the high-resolution methods that could reveal structures at nearly atomic level but not in their natural environment. This gap was bridged by Electron Tomography (ET) that has the ability to recover both the context as well as relatively high-resolution structures of macromolecules. Although the principles of ET are not new, it was the technological advancement and automatic data collection developed in past few years that made this method an invaluable tool for structural analysis. Nevertheless, the intracellular distributions as well as interactions of macromolecular complexes might be violated if the specimen preparation is based on chemical fixation, dehydration, and/or heavy-metal staining. In recent years, the improvement of vitrification techniques together with development of thinning methods that could be completely carried out under cryogenic conditions gave rise to Cryo Electron Tomography (cryoET). Due to the special specimen preparation, cryoET is currently the only technique that allows us to study the structure and intracellular distribution of large macromolecular complexes in their close to native environment *in situ* and thus plays an essential role in Structural Biology. In combination with novel high-resolution protocols, such as Subtomogram Averaging (SA) [Winkler and Taylor, 1999; Förster and Hegerl, 2007], cryoET is able to reveal both structural and contextual information with resolution as high as 8 Å.

As the name suggest, the driving principle behind ET/CryoET is tomography - a technique that recovers a 3D map of the internal structures of an object that was imaged at multiple angles using some kind of penetrating wave. The mathematical basis for tomographic reconstruction was laid down by Johann Radon in 1917 when he described the relationship between a 2D object and a set of 1D images obtained by integrating the density of the object along parallel straight lines that are perpendicular to the images. The ability to describe a 2D object by infinitely many 1D projections taken under different angles is known as Radon transform. In order to recover the object from its projections one has to invert either the Radon transform [Radon, 1917] (in case of 2D object) or the X-ray transform [John, 1938] that adapts the Radon transform for 3D objects.

In cryoET, a vitrified specimen is imaged from multiple angles using TEM and the acquired images are then used to computationally reconstruct a model of the 3D scattering potential of the specimen, called tomogram. This reconstruction is especially challenging in cryoET, since the acquisition process imposes some significant limitations. First, the specimen can be tilted only from roughly ±60°, leaving us with an incomplete set of projections (called the *missing wedge problem*). Second, the specimen is larger than a detector and thus some projections contain information about the specimen outside of the reconstructed area. This issue is a specific case of the problem known as the *interior problem*. Third, the thinness of the biological specimen in combination with the low electron dose causes the projections to be very noisy with low contrast. These limitations complicate the reconstruction process and influence the achievable resolution. When the structure of an abundant macromolecular complex is sought, the resolution can be improved and missing wedge distortion removed by processing the tomograms using high-resolution protocols such as SA. The results of these protocols however strongly depend on the quality of provided tomograms. The reconstruction methods currently in use, such as Weighted Back Projection (WBP) [Radermacher, 1992] or Simultaneous Iterative Reconstruction Technique (SIRT) [Gilbert, 1972a], deliver noisy and low-contrast reconstructions. The low quality complicates further processing of the tomograms, such as segmentation or pattern matching, and prevents fully automatic application of SA.

The main goal of this work is to improve the quality of tomograms obtained from cryoET projections in order to facilitate their subsequent processing. We propose a Progressive Stochastic Reconstruction Technique (PSRT) – a novel iterative approach to tomographic reconstruction in cryoET that is based on a different mathematical framework than the state-of-the-art techniques as it uses Monte Carlo random walks to perform a coarse-to-fine reconstruction. PSRT delivers tomograms with reduced noise and enhanced contrast, which improves the performance of automatic tomogram postprocessing compared to the currently used methods. Moreover, it provides a memory efficient solution to the interior problem and allows for region of interest reconstructions, which are of special importance in SA.

In this work we also study the negative influence of the geometry-related acquisition artifacts on the quality of tomograms. Though individually described by different authors before, a systematic study of these artifacts in one place and with focus on their manifestation in tomograms reconstructed with different techniques has not been, to our knowledge, performed before. We illustratively explain how these artifacts arise and when they can be avoided. We draw recommendations on numerous points which, we believe, will help a reader to better interpret the artifacts and choose the optimal way how to lower their negative impact, thereby improving the quality of tomograms.

1.1 Original Contributions

Parts of this thesis are based on published works, where the author was both the primary investigator and author and wrote the majority of the text. In this work, we use passages from the text with the explicit permission of the co-authors. Below we highlight the major contributions of this thesis.

- **Progressive Stochastic Reconstruction Technique (PSRT).** We design a novel iterative approach to the tomographic reconstruction in cryoET based on Monte Carlo random walks guided by a sampling strategy similar to the Metropolis-Hastings sampling strategy. We design a progressive reconstruction scheme to suit the conditions present in cryoET and apply it successfully to reconstructions of macromolecular complexes from both synthetic and experimental datasets. We integrates PSRT into SA, where it provides an elegant solution to the region-of-interest problem and delivers high-contrast reconstructions that significantly improve template-based localization without any loss of high-resolution structural information. Furthermore, the locality of SA is exploited to design an importance sampling scheme that significantly speeds up the otherwise slow Monte Carlo approach. We design a new memory efficient solution for the specimen-level interior problem of cryoET, removing all associated artifacts. Finally, we implement a parallelization scheme that allows us to process large datasets more efficiently. These contributions are based on [Turoňová et al., 2015], where the author developed the proposed techniques, performed all the experiments, as well as wrote the majority of the text.

- **Study on Geometric Artifacts in cryoET.** We perform a systematic study of the acquisition geometry-related artifacts that negatively influence the achievable resolution as well as the processing of the tomograms. We first present the problems that cause the artifacts in a unified geometrical framework. We characterize not only the individual artifacts but also their interplay and combined effects. We attribute the relevant visual issues of cryoET reconstructions to one or more of those artifacts, disentangling through comprehensive tests the relative influences of each of them. We further study the severity and manifestation of the artifacts across a representative set of reconstruction methods and their influence on the recovered values. Finally, we draw recommendations on numerous points, especially regarding the mutual influence of the geometric artifacts, ability of different reconstruction methods to suppress them as well as implications to the interpretation of both cryoET and SA experiments. These contributions are based on [Turoňová et al., 2016], where the author performed all the experiments and wrote the majority of the text.

1.2 Thesis Outline

Further on, this work is structured as follows: In Chapter 2 we review the underlying principles of cryoET that are necessary to fully understand the character of the data we are dealing with as well as the required properties of the reconstructed tomograms. In Chapter 3 we provide a mathematical formulation of the reconstruction problem and discuss various approaches for its solution. Chapter 4 briefly introduces basic principles of Monte Carlo methods and explains how they can be used for tomographic reconstruction. In Chapter 5 we present our novel iterative approach to the reconstruction problem in cryoET together with its core properties, while Chapter 6 reveals the implementation details of our method. In Chapter 7 we validate our method on synthetic data, compare it to state-of-the-art methods, and discuss its performance on experimental datasets. The integration of our method within the standard subtomogram averaging pipeline is described in Chapter 8. We discuss possible parallelization schemes in Chapter 9. In Chapter 10 we perform a systematic study of geometric artifacts present in cryoET tomograms and evaluate their influence on quality of a tomogram. Finally, Chapter 11 summarizes our main contributions and outlines possible follow-ups of our work.

1.2.1 Terminology

Due to rather interdisciplinary character of this work, we introduce the terminology that is used in this thesis to avoid any potential misunderstandings. The term cryoET is often misused since it often comprises also preparation techniques that do not preserve the native state. In line with [Lucić et al., 2013], we denote as cryoET a workflow that uses only nonfixed, nonstained, and fully hydrated specimens that were vitrified and imaged while still in the fully hydrated, vitrified state. Note that specimens are also often called samples. However, in this thesis we strictly use the term specimen, while the term sample is reserved for Monte Carlo theory and PSRT. In Single Particle Analysis (SPA), the word particle denotes a biological structure of interest and this terminology was later adopted in Subtomogram Averaging (SA) where one denotes as a particle a single instance of macromolecular complex present in a tomogram. In this work, we follow this convention and in addition to the physical meaning of the term particle we use this term also to denote the biological structure in the context of SA.

CHAPTER 2

CRYO ELECTRON TOMOGRAPHY

Cryo Electron Tomography (cryoET) plays an essential role in Structural Biology, as it is the only technique that allows for study of the structure of large macromolecular complexes in their close to native environment *in situ*. Although the word tomography typically refers only to the reconstruction of 3D volume (called tomogram) out of 2D projections, the term cryoET is used to describe the whole process from the specimen preparation to the analysis of the reconstructed tomograms (see Fig. 2.1).

In this chapter we review the basic principles behind cryoET that are necessary to understand the character of the data we will be working with as well as the requirements that are posed on the reconstructed volumes. We first briefly describe the specimen preparation and its influence on quality of a tomogram. Then we focus on the image formation in a Transmission Electron Microscope (TEM) and the limitations it poses on the acquired data. We also review preprocessing steps for the acquired images that should be done prior to reconstruction. The reconstruction itself is mentioned only very briefly as it is the main subject of the next chapter. Finally, we describe the individual steps of the Subtomogram Averaging (SA) pipeline as this high-resolution protocol is important in the context of this thesis.

Note that this chapter provides only a very brief overview of cryoET. For a comprehensive description on electron tomography we refer the reader to the excellent work by Frank [2006].

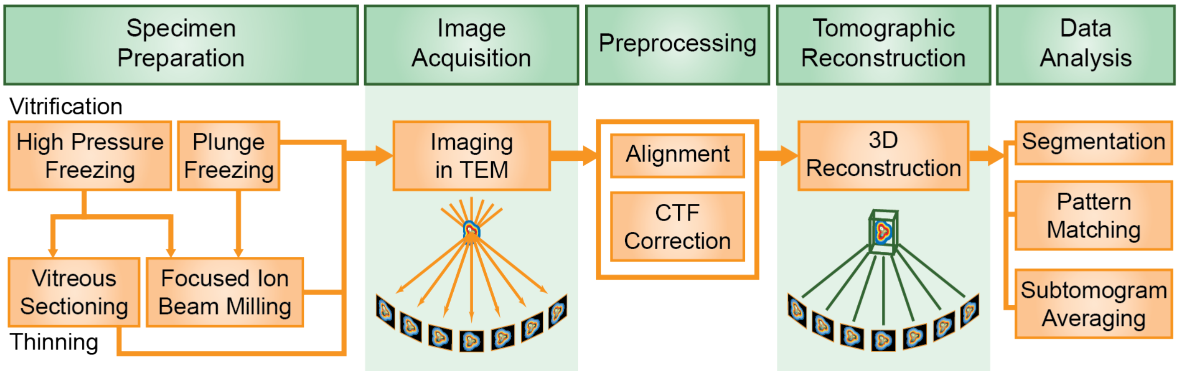

Specimen Preparation
Image Acquisition
Preprocessing
Tomographic Reconstruction
Data Analysis
Vitrification
High Pressure Freezing
Plunge Freezing
Vitreous Sectioning
Focused Ion Beam Milling
Thinning
Imaging in TEM
Alignment
CTF Correction
3D Reconstruction
Segmentation
Pattern Matching
Subtomogram Averaging

Figure 2.1: Cryo Electron Tomography workflow.

2.1 Specimen Preparation

Specimen preparation represents the main difference between the “classic” electron tomography and cryoET. As the name suggests, the preparation technique is based on freezing which, unlike other procedures that involve chemical fixation, dehydration, and heavy-metal staining, is able to preserve the native environment. The freezing method used in cryoET is vitrification, a special type of freezing that prevents formation of ice crystals within the specimen. Based on their thickness, there are two basic approaches for vitrification of biological specimens [Lucić et al., 2013]. Plunge freezing [Dubochet et al., 1988] for specimens thinner than 10 µm and high pressure freezing [Studer et al., 2008] for thickness up to 200 µm.

Unfortunately, frozen-hydrated unstained specimens are very sensitive to radiation damage [Glaeser, 2008]. Higher doses of electrons cause not only progressive deterioration of structures withing the specimen but also induce movement of them. As a consequence, the images in TEM have to be recorded using only very low number of electrons, which ultimately limits the achievable resolution. Moreover, due to the short mean free path of electrons with low energy, the thickness of the specimen should be less than 500 nm [Lucić et al., 2013]. Although there are complexes that could be scanned directly, most of the eukaryotic cells exceed this limit and thus have to be sectioned (thinned) in cryo-conditions (i.e. under temperature less than 140°). One approach for thinning is vitreous sectioning [Al-Amoudi et al., 2004] – an effective but very challenging procedure that unfortunately introduces artifacts in form of knife marks and compression [Al-Amoudi et al., 2005; Hsieh et al., 2006]. A completely different approach is focused ion beam milling, where the thinning is done by sputtering with gallium ions [Marko et al., 2007; Rigort et al., 2012]. So far, no artifacts caused by this method have been detected [Marko et al., 2006].

2.2 Image Acquisition in TEM

2.2.1 Electron Imaging

To understand the image formation in TEM, it is important to know how electrons interact with the specimen. As electrons have a dual character we can understand them either as negatively charged particles or as waves.

In the first case, the interaction with atoms in the specimen is characterized by scattering, which can be elastic or inelastic. During elastic scattering the electrons maintain their energy and are only slightly deviated from their path which results in an amplitude contrast in the scanned image. Inelastic scattering, on the other hand, causes a loss of the electrons’ energy, which is deposited within the specimen and leads to radiation damage. Moreover, direct electrons are scattered strongly from their original path, making the inelastic scattering useless for imaging. It is worth noting that electrons are scattered much stronger than X-Rays and that the number of scattering events increases with the thickness of the specimen. For example, for 500 Å thick unstained biological specimen and $300 kV$ incoming electron beam there is around 15% inelastic scattering, only 5% elastic scattering and the rest is unchanged. In other words, for every useful scattering event, there are three damaging ones.

Considering the wave-like character of electrons, the incident electron beam can be seen as an extended plane wave that is diffracted by atoms of the specimen (see Fig. 2.2b). The diffraction causes phase shifts that yields phase contrast in the scanned images. Since the unstained biological specimens consist essentially only of atoms with low atomic number and are very thin, the phase changes are small in comparison to the electron wavelength, which results in low phase contrast. Despite that,

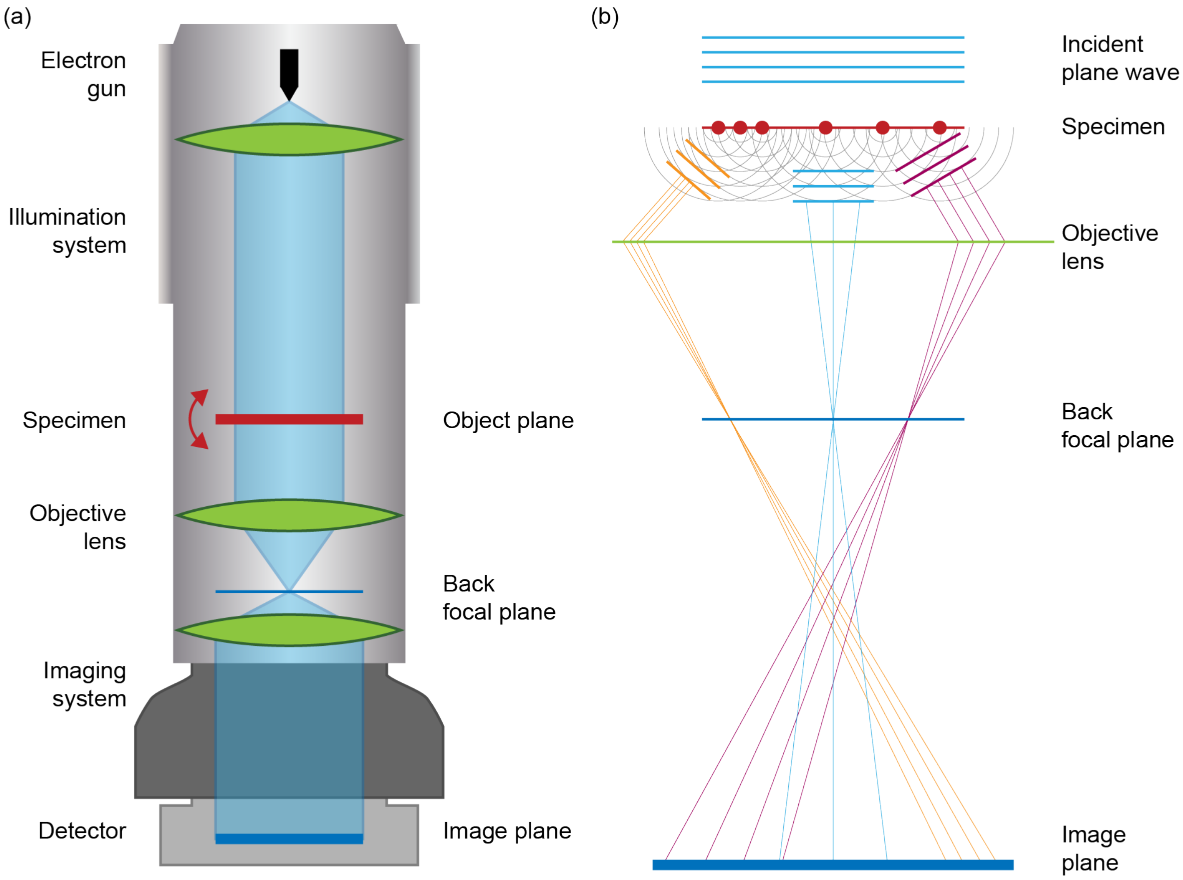

(a)
Electron gun
Illumination system
Specimen
Object plane
Objective lens
Back focal plane
Imaging system
Detector
Image plane
(b)
Incident plane wave
Specimen
Objective lens
Back focal plane
Image plane

Figure 2.2: (a) Simplified scheme of Transmission Electron Microscope (TEM) loosely redrawn from [Lucić et al., 2013]. (b) Image formation in TEM. Once the incident plane wave hits the specimen, the atoms within it become scattering centers for secondary wavelets. These wavelets interfere, resulting in many coherent beams diffracted at specific angles which are focused by the objective lens to first form a diffraction pattern at the back focal plane and then a final image of the scattering potential at the image plane.

for very thin specimens the phase contrast is significantly stronger than the amplitude contrast and thus dominates in the scanned images obtained with cryoET.

In order to reconstruct a 3D map of a specimen from 2D images, one has to know the relationship between the specimen and the scanned images. This relationship can be described using Weak Phase Object Approximation (WPOA). The theory is beyond the scope of this thesis and we refer the reader to [Williams and Carter, 2009] for details. The important result of WPOA is that for thin specimens consisting of atoms with low atomic numbers the contrast in the recorded image is linearly related to the projected specimen scattering potential. As we will see in the next chapter, this property is crucial for forming a proper forward model of image acquisition and thus for tomographic reconstruction.

2.2.2 Image Acquisition

The specimen is scanned in Transmission Electron Microscope (TEM) whose main parts present electron gun, a set of electromagnetic lenses, a stage (holder) for the specimen, and a detector (nowadays mostly direct electron detector). Its simplified scheme is shown in Fig. 2.2a, while an exhaustive description of the whole construction can be found in [Williams and Carter, 2009]. For our purposes, it

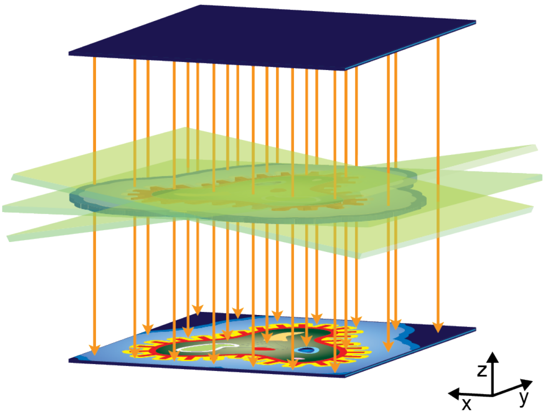


Figure 2.3: Single-tilt scheme of a specimen imaged in a TEM.

is sufficient to know that the whole imaging system operates in vacuum, that its optical principles are similar to traditional visible light microscopy, and finally that despite constant improvement it still suffers from many mechanical imperfections. There is one major difference to classical microscopy that should be mentioned though - the specimens are not imaged in focus. The underfocus (typically in range between $-2\,\mu\text{m}$ and $-5\,\mu\text{m}$) decreases the achievable resolution but increases the phase-contrast by modulating both high and low frequencies. This process is mathematically described by the Contrast Transfer Function (CTF), which will be discussed in more detail in the following section.

The imaging itself works in the following way: a specimen is introduced into the TEM on the stage which is perpendicular to the direction of an electron beam, the beam than passes through the specimen and is recorded in the form of 2D image. Since tomographic reconstruction requires the specimen to be imaged from many different angles, we have to tilt the stage with the specimen around y-axis (see Fig. 2.3) and repeat the scanning process. In this way, we obtain a set of 2D projections, a so-called tilt-series, that are later used for the reconstruction. The tilting significantly complicates both the image acquisition and subsequent processing of the images.

Firstly, the mechanical imperfections of TEM cause not only shifts and rotations but also various image distortions such as changes in magnification. Although the development of automated collection of tilt series [Dierksen et al., 1992; Koster et al., 1992, 1997] helped to minimize some of these unwanted effects, the acquired images have to undergo further processing before they can be used for reconstruction. It can also happen that the stage is not perpendicular to the beam but is slightly tilted around the x-axis, which complicates any subsequent processing of the tilt-series including the reconstruction. This problem is referred to as non-perpendicularity of the beam or simply as x-tilt.

Secondly, the specimen is scanned multiple times, once for each angle, which increases the danger of radiation damage. In other words, we have to split an already small electron dose among all the projections. For a tilt-series consisting of 60 projections and a dose of about $60\,e^-/\text{Å}^2$, we have only $1\,e^-/\text{Å}^2$ for a single image. This not only decreases the contrast but also poses a limitation on a number of projections we can acquire. Consequently, the tilt-series typically contain only up to 60 projections.

Thirdly, the tilting causes anisotropic resolution and information inconsistency in the reconstruction. As mentioned in Ch. 1, there are three main reasons for that. The first is that the tilt angle is limited to roughly $\pm 60°$ as at higher angles the effective thickness of the specimen increases and at some point the specimen is no longer electron transparent. This limitation is also known as the missing wedge problem due to its characteristic shape in Fourier space. The second problem is known as the interior problem and is caused by the fact that the specimen is larger than the detector and thus the projections at higher tilt angles contain redundant information w.r.t. the reconstructed area. Lastly, while scanning a “thick” specimen we might loose some information from its corners. All three problems introduce artifacts into the reconstructions and will be therefore discussed in more details in Ch. 10.

2.3 Tilt-Series Preprocessing

The acquired images have to be processed before we can start with the reconstruction. Although there are many possible image modifications (e.g. filtering or gold beads removal), we focus in this section only on image alignment and on the correction for the Contrast Transfer Function (CTF). While the image alignment is essential for obtaining correct reconstruction, the correction of CTF is necessary to increase the resolution achievable with subtomogram averaging.

2.3.1 Alignment

Due to the mechanical imperfections in TEM, the specimen shifts during the image acquisition process, which results in misaligned tilt-series. Prior to any reconstruction, projections from such tilt-series have to be aligned to a common coordinate system and thus their relative shifts and orientations have to be computed. There are two basic approaches that are common in electron tomography. The first one is marker-free alignment which tries to find the relative positions based on specific features or patches present the images [Liu et al., 1995; Brandt et al., 2001a; Brandt and Ziese, 2006; Castaño-Díez et al., 2007, 2010; Sorzano et al., 2009; Winkler and Taylor, 2006, 2013; Han et al., 2014]. These approaches require high contrast and are thus better suited for stained samples. The second approach uses so-called fiducial markers for alignment [Ress et al., 1999; Brandt et al., 2001b; Amat et al., 2008]. The markers are some electron-dense particles (typically gold beads) within the specimen that have high contrast in the scanned images and thus can be tracked more accurately. The high-contrast however also introduces strong artifacts near the markers in the reconstructed tomograms. Despite that, the marker-based alignment is the dominant approach scheme in cryoET as the low SNR conditions often prevent a precise marker-free alignment. Note that some of the above mentioned methods are not able to handle general tilt geometry and thus in case of x-tilt their alignment model has to be changed to accommodate for the non-perpendicular beam [Castaño-Díez et al., 2006; Mastronarde, 2008].

2.3.2 Contrast Transfer Function Correction

The Contrast Transfer Function (CTF) describes the influence of aberrations in TEM on frequencies in acquired images. CTF can be seen as a sinus function that oscillates between negative and positive values and gradually attenuates with higher spatial frequencies. This causes contrast flipping and attenuation of details in the projections. Moreover, at zero-crossings the structural information is lost completely. Due to the contrast reversal, the information with frequencies larger than at the first zero-crossing is inconsistent with the values at the frequencies lower than at the first-zero crossing. Thus, the information beyond the first zero-crossing is not reliable and the achievable resolution is limited by the frequency of the first zero-crossing. The CTF depends on the defocus at which the specimen is scanned: At lower defocus we can achieve higher resolution but the contrast is very low which makes the reliable processing of such images and subsequently tomograms more difficult. For defocus between -2 and $-5\mu m$ and $300kV$ microscope the first zero-crossing occurs in the range of $20-30$Å [Fernández, 2012]. Thus CTF does not pose limitation on classical tomography but constrains the achievable resolution in Subtomogram Averaging (SA).

In SA we typically want to recover structures with resolution that lies beyond the first zero-crossing and thus we have to apply the CTF correction to remove the contrast reversal. There are two major problems that methods for CTF correction have to face. First, due to the low SNR in tilt-series that are close to focus (e.g. for datasets between -2 and $-5\mu m$) it is difficult to reliably restore the CTF. The second problem is that the defocus is not constant but varies along the beam direction. There are numerous approaches how to tackle these problems [Winkler and Taylor, 2003; Fernández et al., 2006; Philippsen et al., 2007; Zanetti et al., 2009; Xiong et al., 2009; Voortman et al., 2011; Bharat et al., 2015] that are now being successfully used in recovering high resolution structures [Ortiz et al., 2010; Nicastro et al., 2011; Li et al., 2012].

Even with the CTF correction the information at zero-crossings remains unavailable and thus a different approach has to be used to recover the missing frequencies. The most common technique is to use focal pairs (two tilt-series taken at low and high defocus) and combine the information from both of them [Kudryashev et al., 2012].

The recent hardware development enables to use phase plates during the image acquisition that allow for imaging nearly at focus. In simple terms, the phase plates shift the CTF and transform it to a cosine function, thereby improving the contrast at low frequencies. Although there are some promising results using this approach [Danev et al., 2010; Kishchenko et al., 2015; Fukuda et al., 2015], its use on large scale is still problematic due to hardware imperfections and thus the low-contrast of the dataset has to be improved using software solutions.

2.4 3D Reconstruction

The preprocessed tilt-series are used to computationally reconstruct a model of the 3D scattering potential of the specimen, called tomogram. As mentioned above, the single-tilt geometry imposes some significant limitations on the acquired images. Moreover, the thinness of the specimen in combination with the low electron dose causes the projections to be very noisy and of low contrast. These limitations make the reconstruction in cryoET especially challenging task. The mathematical formulation of tomographic reconstruction as well as an overview of the state-of-the-art techniques is provided in the next chapter.

2.5 Subtomogram Averaging

Subtomogram Averaging (SA) is a computational processing protocol for obtaining high-resolution 3D structures of individual macromolecular complexes present in tomograms [Winkler and Taylor, 1999; Förster and Hegerl, 2007]. The essence of this approach is to average together structures at the same conformation state to enhance their features and improve resolution. The averaging increases the SNR and due to different orientations of particles in the tomogram, the final structure suffers less from distortions caused by the missing wedge problem. Although the currently highest resolution achieved by SA is 8.5Å [Schur et al., 2013], the more common resolution is up to $15-20$Å [Briggs et al., 2009; Eibauer et al., 2012; Chen et al., 2013]. The detailed overview of successful SA applications can be found in the review by Briggs [2013].

A typical SA pipeline consists of localization, classification, alignment, and averaging and many different approaches were designed to perform the individual steps. Recently, unified frameworks for the entire protocol were developed, with the two most outstanding being Dynamo [Castaño-Díez et al., 2012] and PyTom [Hrabe et al., 2012]. Recently, the Relion framework, which was primarily designed for Single Particle Analysis (SPA), has been extended also for the use in SA [Bharat et al., 2015].

2.5.1 Localization

The first step in the SA pipeline is to localize the structures of interest within the tomogram. Provided that there are only relatively few, well separated particles for averaging, the localization can be done manually. Otherwise, automatic localization is necessary. There are two approaches how to perform the automatic localization. The first one is based on the identification of features such as edges, variance in density or density peaks (e.g. in [Xu et al., 2011]). The second approach is reference-based – it uses cross correlation techniques to find particles that match a given template [Förster et al., 2010; Chen et al., 2012; Bharat et al., 2015], which is usually obtained by downsampling a model from crystallography or SPA. This template-based matching is currently the most common approach in SA as it seems to be more robust w.r.t. noise than feature-based approaches. However, this property also presents a danger of finding structures that are actually nothing but pure noise [Henderson, 2013]. Therefore, it is important that the localized particles are always visible. The second problem, which also concerns feature-based methods, is localization of false-positive candidates – often gold markers or some contamination present in a tomogram are falsely marked as particles one is searching for. These false positive results have to be manually removed, which prevents the SA pipeline to be fully automatic.

2.5.2 Classification

The localized particles can be in different conformational states and thus, in order to obtain the best possible alignment, the structurally different particles should be classified and subsequently clustered together. The classification is complicated by the fact that the number of classes is usually unknown and the number of particles in each class might significantly differ. Moreover, the distinctive features of particles might be distorted or completely lost due to limitations imposed on the reconstruction, such as the missing wedge problem. The most common approaches are based on constrained Principal Component Analysis (PCA) in conjunction with k-means clustering [Förster et al., 2008; Bartesaghi et al., 2008], probabilistic PCA using maximum expectation [Yu et al., 2010; Hrabe et al., 2012], or statistic approaches based on maximum likelihood optimization [Scheres et al., 2009; Stölken et al., 2011]. Recently, approaches deploying neural networks and unsupervised learning emerged [Yu and Frangakis, 2011; Chen et al., 2014]. Note that some of these algorithms iteratively refine the results by performing classification together with precise alignment.

2.5.3 Alignment and Averaging

The alignment and subsequent averaging of subtomograms are the last crucial steps in the SA pipeline. As mentioned above, the alignment is often combined with classification and thus similar principles are used for handling the missing wedge problem and finding the similarities between the subtomograms. Unlike in localization and classification where only rough orientations were sufficient, during alignment the space of possible rotations has to be sampled more densely in order to achieve the best result. Thus, in order to reduce the computational costs, the cross-correlation methods for alignment usually start with the rough orientations obtained during the localization and the dense sampling of angular space is performed only in the vicinity of the initial rotations [Förster and Hegerl, 2007; Amat et al., 2010; Hrabe et al., 2012]. This approach might however lead to reference bias [Subramaniam, 2006]. Computational less demanding and reference-free alternative presents Fast Rotational Matching (FRM) [Bartesaghi et al., 2008; Xu et al., 2012; Chen et al., 2013].

CHAPTER 3

TOMOGRAPHIC RECONSTRUCTION IN CRYOET

Tomographic reconstruction is the mathematical process that recovers an internal structure of an object that was imaged at multiple angles using some kind of penetrating wave. The mathematical basis for tomographic reconstruction was presented in the work of Radon [1917] in which he described the relationship between a 2D object and its 1D projections. This relationship was later extended to 3D objects by formulating the X-ray transform [John, 1938]. Many advances followed and nowadays the tomography is used in a wide range of sciences from medicine and biology, over material science to archeology.

In cryoET, we aim to reconstruct a model of the 3D scattering potential of the specimen and as indicated in the previous chapter, this task is very challenging under cryoET conditions. In this chapter, we present the basic mathematical principles behind tomography and review existing techniques both in terms of their mathematical as well as geometric meaning. We also discuss the problem of resolution and quality assessment that complicates evaluation of reconstruction methods.

In this work, we use the coordinate system depicted in Fig. 3.1. We denote the size of a reconstructed area in x as width w, in y as height h, and in z as depth d. Since the electron beam is orthogonal to the specimen along z-axis, we can obtain a 3D tomogram in two ways: by performing a full 3D reconstruction out of 2D projections or by splitting the volume along y-axis and doing set of 2D reconstructions out of 1D projections and stacking them together (with the splitting being done along y-axis). In this work, we consider the full 3D reconstruction and use the second approach only for illustration purposes (with the 2D reconstruction lying in xz plane).

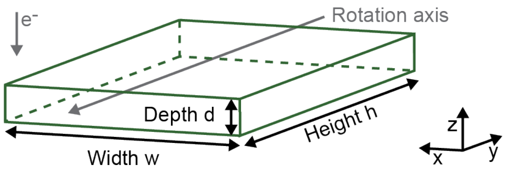


Figure 3.1: Scheme of the coordinate system and notation used in this work. We denote the size of a reconstructed area in x as width w, in y as height h, and in z as depth d.

3.1 3D Tomography

The work on tomography is extremely extensive and here we only very briefly highlight the key terms that are required for the understanding of our method. The reader is kindly referred to the numerous publications on the topic for additional details [Louis, 1989; Natterer and Wübbeling, 2001; Frank, 2006; Fanelli and Öktem, 2008; Herman, 2009].

In 3D tomography one seeks to reconstruct a 3D volume from a set of measured 2D projections by computationally reversing the process of projection acquisition. The acquisition of projections $\bar{p}$ of a given specimen can be mathematically expressed as an operator $\mathscr{R}$ (so-called forward operator) acting on a volume v: $\bar{p} = \mathscr{R}[v]$, where $\bar{p} = \{p_i, i \in I\}$.

A common approach to defining the forward operator $\mathscr{R}$ that we also use in this work is to simplify the acquisition process by seeing a measurement p_i as an integration along the ray Θ_i transmitting through the volume v

$$p_i = \int_{\Theta_i} v(x)\,dx. \tag{3.1}$$

This relationship in 2D is known as the Radon transform and was first described in [Radon, 1917]. Tomographic reconstruction is then based on the theory of inverting the Radon transform (in case of a sequence of 2D reconstructions) or the X-Ray transform (for a full 3D reconstruction). Thus, it can be formulated as an inverse problem, where one tries to find an inverse of the operator $\mathscr{R}$ to recover the unknown volume v:

$$v = \mathscr{R}^{-1}[\bar{p}].$$

However, to obtain the unknown volume in a unique and stable way, the theory requires that a) the sample is imaged from all possible directions on a unit sphere around the object, b) the sample wholly fits into the reconstructed area and into the projections even from oblique angles [Louis, 1989; Natterer and Wübbeling, 2001]. Unfortunately, the single-tilt acquisition scheme does not fulfill either of these conditions due to the limited tilt-angle and the specimen-level interior problem (see Sec. 2.2.2 or Ch.10 for more details). Moreover, due to the specimen thinness and low electron dosage used during the acquisition, the input data suffers from low contrast and is severely contaminated with noise. Consequently, the inverse we are seeking is not only severely ill-posed but typically even does not exist. Therefore, the reconstruction is often formulated as a least square problem

$$\bar{r} = \left\| \bar{p} - \mathscr{R}[\tilde{v}] \right\| \tag{3.2}$$

where one does not try to find an exact inverse, but rather an estimate $\tilde{v}$ to the unknown specimen's scattering potential v such that the difference $\bar{r}$ ($\bar{r} = \{r_i, i \in I\}$, so-called residuals) between the measured projections $\bar{p}$ and the forward projection of the estimate $\mathscr{R}[\tilde{v}]$ is minimal.

Visual Interpretation of Results

Unlike in Computed Tomography (CT), the relationship described in Eq. 3.1 does not contain any attenuation factor. This is in line with the fact that the measured values correspond to the scattering potential of the specimen and not to its density. Consequently, the features (e.g. macromolecules, membranes etc.) appear darker in the acquired images than their environment as they scatter electrons more strongly. However, many algorithms rely on the opposite gray-level interpretation that corresponds to the density interpretation, i.e. the features are bright while darker gray values indicate less dense areas. Moreover, the gray-levels in phantoms usually also correspond the density interpretation as it facilitates perception and presentation of the results. For that reason, all results presented in this work use the density gray-level coding as well. To achieve that also for experimental data we invert the values in tomograms by simple multiplication by -1. Unlike in CT, where the values in tomograms can be quantitatively described using Hounsfield units, the values in cryoET tomograms have neither a physical meaning nor a quantitative interpretation and thus this inversion has no effect on the tomogram's interpretability. Especially, since it is done only for presentation purposes and any other processing of tomograms is done using their original values.

3.2 Tomographic Reconstruction Algorithms

The most common way to classify different reconstruction methods is based on their mathematical approach for solving the tomographic problem. However, as we will see in Ch. 10, it is useful to divide the algorithms also w.r.t. the geometric aspects of the reconstruction problem. Therefore, we provide in this section the "standard" mathematical classification as well as the geometric one.

3.2.1 Mathematical Interpretation

We can divide the reconstruction algorithms in cryoET into three major groups. The first one is formed by direct inversion methods that try to reconstruct the unknown scattering potential by explicitly finding a mathematical inverse to the forward operator $\mathscr{R}$, which for 2D case is expressed as the Radon transform. For 2D we can reformulate Eq. 3.1 in the following way

$$\mathscr{R}f(\theta,l) = \int_{-\infty}^{\infty} f(x(t),y(t))\,dt, \tag{3.3}$$

where f is a function describing the 2D volume, $(x(t),y(t))$ denotes parametrization of the line, whose perpendicular distance to the origin is l and the angle formed by the distance vector with the x axis is θ (see Fig. 3.2a). The inverse Radon transform $\mathscr{R}^{-1}$ at points (r,ϕ) is then defined as

$$\mathscr{R}^{-1}\mathscr{R}f(r,\phi) = -\frac{1}{2\pi^2}\int_{0}^{\pi}\int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial l}\mathscr{R}f(l,\theta)}{r\cos(\theta-\phi)-l}\,dl\,d\theta. \tag{3.4}$$

To better understand the inverse Radon transform it is useful to express Eq. 3.4 as sequence of operations:

$$\mathscr{R}^{-1} = -\frac{1}{2\pi}\mathscr{B}\mathscr{H}_Y\frac{\partial}{\partial l}\mathscr{R}f(l,\theta),$$

where $\mathscr{H}_Y$ is the Hilbert transform with respect to the first variable and operator $\mathscr{B}$ denotes back projection. For a function t of two variables, the value of $\mathscr{B}t$ at any point (r,ϕ) is defined as:

$$\mathscr{B}t(r,\phi) = \int_{0}^{\pi} t(r\cos(\theta-\phi),\theta)\,d\theta.$$

Eq. 3.4 requires that the exact values of $f(l,\theta)$ are known at all θ and l. In cryoET settings (and not only there) is this requirement not satisfied and thus, as stated above, the inverse generally does not exist. Therefore, the direct inversion methods try to find its approximation either by requiring a complete projection set so that the inverse exists [Radermacher, 1992] or by using an approximate inverse [Kohr and Louis, 2011]. Despite those problems, the most common technique currently used in cryoET is Weighted Back Projection (WBP) [Bracewell and Riddle, 1967; Crowther et al., 1970; Gilbert, 1972b;

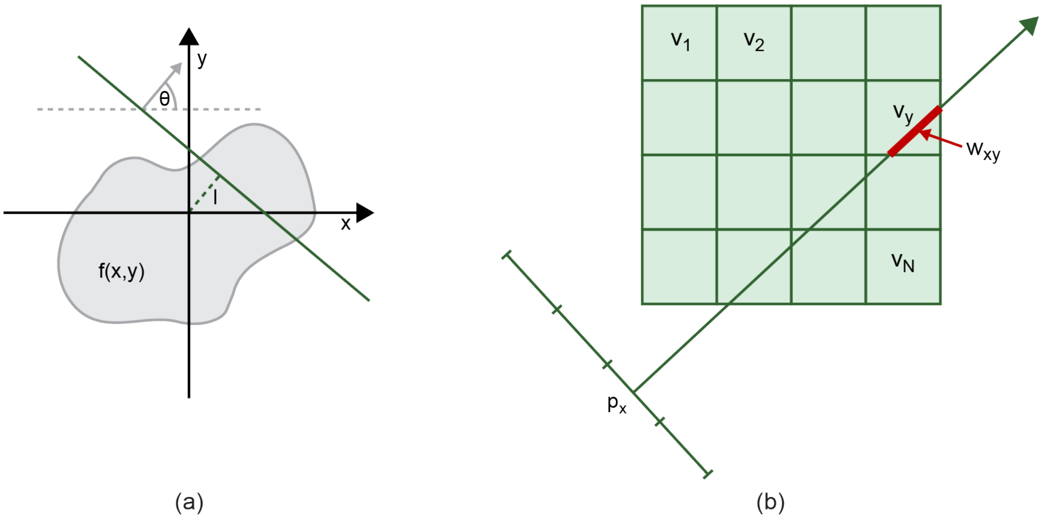


Figure 3.2: (a) Line parametrization used in Eq. 3.3. (b) In the iterative methods, the weighting factor w_{xy} can be defined as the intersection length of the ray corresponding to the pixel p_x with the voxel v_y.

Radermacher et al., 1987; Radermacher, 1992] which belongs to the first group. In cryoET settings, WBP is known to deliver reliable but very noisy reconstructions with low contrast.

The second approach formulates the reconstruction problem as a large system of linear equations and tries to solve it iteratively. As there are more unknowns than equations, this system is undetermined. There are various methods that can be used to solve the system, with the most common in cryoET being the Algebraic Reconstruction Technique (ART) [Gordon et al., 1970], the Simultaneous Iterative Reconstruction Technique (SIRT) [Gilbert, 1972a] based on the Landweber approach [Landweber, 1951], and the Simultaneous Algebraic Reconstruction Technique (SART) [Andersen and Kak, 1984] which is using Kaczmarz method [Kaczmarz, 1937]. All three approaches follow the same pattern – in each step they compute the error of the current volume estimate and then update the volume estimate by distributing a certain amount of this error within the volume. The error is computed as a difference between the input projections and the projections of the current volume estimate. Let N denote the number of voxels in the volume, then for the voxel v_j, where $j \in N$, the update step at $(k+1)^{th}$ iteration can be expressed as

$$v_j^{k+1} = v_j^k + \lambda \frac{\sum\limits_{p_i \in S} \left(\frac{p_i - \sum_{n=1}^{N} w_{in} v_n^k}{\sum_{n=1}^{N} w_{in}} \right) w_{ij}}{\sum\limits_{p_i \in S} w_{ij}}, \tag{3.5}$$

where p_i denotes a pixel value from the input projections with I being the total number of all pixels in all projections and $S \in I$, λ is so-called relaxation parameter that determines the amount of the error that should be distributed within the volume, and w_{xy} is the weighting factor for the y^{th} voxel hit by the ray that corresponds to the x^{th} pixel (see Fig. 3.2b). The difference between ART, SART, and SIRT lies in the definition of the subset S. For ART, the subset S consists of only one pixel. In case of SART, the subset S contains all pixels from one projection. Finally for SIRT, the subset contains all pixels from all projections. The size of the subset influences the convergence speed of the methods – ART converges relatively quickly but the reconstruction quality is relatively low, SIRT delivers better results but its convergence is significantly slower, and SART represents a compromise between these two by delivering results of reasonable quality relatively quickly [Kak and Slaney, 1988]. Generally, the iterative methods are more robust against the specific limitations found in cryoET [Lučić et al., 2005; Leis et al., 2006] and deliver reconstructions with higher contrast compared to WBP [Fernández, 2012]. Since all the approaches are computationally demanding, various parallelization schemes were proposed [Fernández, 2008; Xu et al., 2010; Agulleiro and Fernández, 2011; Dahmen et al., 2016] to decrease the processing time and make them a viable alternative to WBP.

The third approach is based on a principle of maximum entropy and incorporates a-priori information into the tomographic reconstruction. A method based on this principle was proposed by Skoglund et al. [1996] using advanced noise model as an a-priori knowledge.

3.2.2 Geometric Interpretation

Geometric-Based Methods

From the geometric point of view, using the simplification from Eq. 3.1 one tries to back project the measured values into the volume along the corresponding rays such that the differences between the measured values and values of the projections of the volume estimate are minimal. Most of the iterative methods belong to this group, such as SIRT and SART. In addition, also the variant of the Weighted Back Projection (WBP) that does the back projection in real space (e.g. [Gilbert, 1972b]) belongs here. The simplified scheme for the geometric reconstruction is depicted in Fig. 3.3 and looks like follows:

1. Take an initial volume estimate.
2. Optional: filter the input projections either in real space or in Fourier space.
3. Until a stopping criterion is met, repeat:
    - 3.1. Forward projection: Project the volume estimate $\tilde{v}$ to obtain so-called virtual projections $\bar{t}$.
    - 3.2. Error computation: Calculate the difference $\bar{r}$ between the input projections $\bar{p}$ and virtual projections $\bar{t}$.
    - 3.3. Back projection: Distribute the difference back into the volume to obtain a new volume estimate.

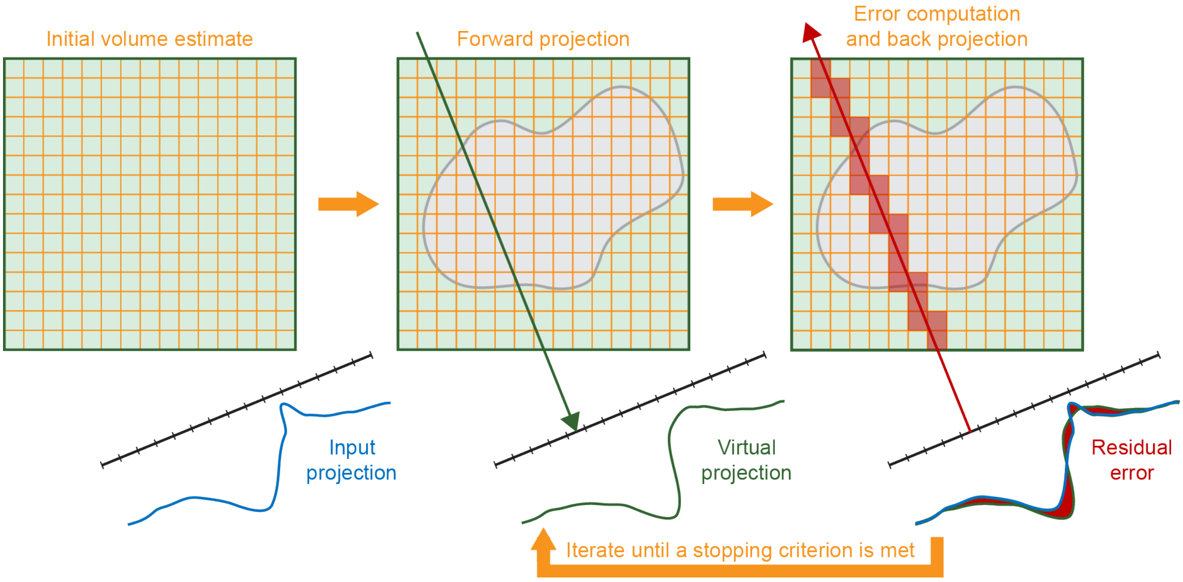


Figure 3.3: A simplified scheme for a geometric reconstruction.

Note that WBP as described in [Gilbert, 1972b] is a special case, in which the reconstruction process consists of only one iteration [Norton, 1985].

Fourier-Based Methods

Fourier-Based methods perform the reconstruction in Fourier space and are based on the Fourier Slice Theorem (FST) [Bracewell, 1956]. According to the FST for the 2D case, the Fourier transform of a 1D projection with a given tilt angle is equivalent to a slice under the same angle through the origin of 2D Fourier transform of a reconstructed object. Thus, the essence of this approach is to take the Fourier Transforms (FT) of the input projections, convert their polar coordinates to Cartesian and perform Inverse Fourier Transform (IFT) of the Cartesian arrangement to obtain the reconstruction. This approach was described e.g. in [Bracewell and Riddle, 1967; De Rosier and Klug, 1968; Crowther et al., 1970] and its simplified scheme for 2D case is depicted in Fig. 3.4 and can be summarized as follows:

1. Perform Fourier transform on the 1D input projections.
2. Filter each projection.
3. Convert projections' polar coordinates to Cartesian ones.
4. Perform 2D inverse Fourier transform of the Cartesian arrangement to obtain the reconstruction in real space.

All WBP techniques with back projection in Fourier space are in this group, such as implementations in IMOD [Sandberg et al., 2003], BSoft [Heymann et al., 2008], or PyTom [Hrabe et al., 2012]. Recently, a new iterative approach called Iterative Nonuniform fast Fourier transform based Reconstruction method (INFR) was introduced in [Chen and Förster, 2013] and also belongs to this group.

The conversion from polar to Cartesian coordinates differs for the various Fourier based approaches and requires either non-trivial interpolation in Fourier space [Sandberg et al., 2003; Grigorieff, 1998; Penczek et al., 2004; Zhang et al., 2008; Fernández, 2012] or the use of non-uniform fast Fourier transforms [Chen and Förster, 2013; Fessler and Sutton, 2003; Matej et al., 2004; Knopp et al., 2007]. Moreover, the 3D Fourier transform is sampled more densely near its origin which promotes lower frequencies and consequently leads to blurred reconstructions. To compensate for that, filtering of the input projections is necessary for all Fourier based methods [Fernández, 2012].

Note that these methods rely on certain geometric constraints and any changes in the acquisition setup (e.g. convergent beam, x-tilt) require a new mathematical derivation. Moreover, the dimensions of the reconstructed volume are not flexible, i.e. for 2D reconstruction out of 1D projections with size w, the reconstructed area is implicitly $w \times w$. To overcome this limitation, more complex algorithms have to be employed (e.g. [Sandberg et al., 2003]).

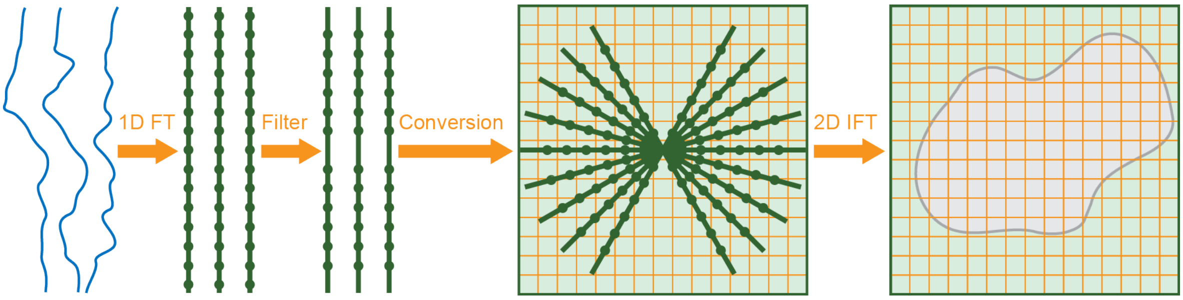


Figure 3.4: A simplified scheme for a Fourier based reconstruction.

3.3 Resolution and Quality Assessment

One of the major problem in cryoET (or ET as such) is the assessment of resolution of the reconstructed tomograms [Frank, 2006]. Although many approaches were suggested (e.g. [Cardone et al., 2005; Frank, 2006]), most of them did not find widespread use and resolution estimation thus remains an open problem [Leis et al., 2009; Henderson et al., 2012; Fernández, 2012].

This is fortunately not the case for Subtomogram Averaging (SA) technique where the resolution can be approximated by comparing two 3D volumes that were obtained by splitting the subtomograms into halves and averaging them. The resolution is then computed using Fourier Shell Correlation (FSC) [Harauz and van Heel, 1986], which represents a normalized cross-correlation between (complex) Fourier components:

$$FSC(r) = \frac{\sum_{r_i \in r} F(r_i)G(r_i)^*}{\sqrt{\sum_{r_i \in r} |F(r_i)|^2 \sum_{r_i \in r} |G(r_i)|^2}}, \tag{3.6}$$

where F is Fourier transform of the first volume, G^* is complex conjugate of the Fourier transform of the second volume, and r_i denotes the individual voxel element at radius r. The FSC yields a 1D curve of correlation coefficients as a function of r. The main problem while using FSC represents a threshold criterion at which one wants to estimate the resolution as it is a user specified parameter and its optimal value is a matter of discussion [van Heel and Schatz, 2005]. Typical values are 0.147, 0.333, or 0.5 (which is the one we use in our experiments).

The lack of quality metrics for entire tomograms complicates not only the interpretation of tomogram but also the quality assessment of a technique used for its reconstruction. It poses a major problem for assessing the quality and usefulness of a new method. The relative quality estimation can be made by comparison to tomograms obtained by well established methods, such as WBP, ideally on experimental datasets containing well-known structures.

In this work, we validate our method on experimental datasets by comparing its results to reconstructions obtained by WBP (performed in Fourier space) [Sandberg et al., 2003] and SART [Andersen and Kak, 1984]. For SA datasets, we use the FSC to calculate achieved resolution. For synthetic datasets we use the Relative RMSE (RRMSE):

$$RRMSE(x, y) = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{\sum_{i=1}^{N} y_i^2}}, \tag{3.7}$$

where x_i denotes value in the i^{th} voxel in a reconstructed tomogram, y_i denotes value in the corresponding voxel of a phantom and N is number of voxels. We use this metric to measure the error between the reconstructed tomogram and the ground-truth. It is also possible to use this for error estimation between the measured and the virtual projections in which case this metric would be applicable also for experimental data. Note however, that the error would not be a reliable indicator of the quality in this case as the input projections are very noisy and low error might indicate well reconstructed noise rather than useful information.

It is worth noting that the quality itself is a relative term. It should be always regarded w.r.t a specific property we want a tomogram to have. For instance, for automatic localization of structures one prefers high contrast over high frequencies, whereas for averaging one wants to have high (but still reliable) details present in the tomogram even at the cost of low contrast. Thus, in this work we evaluate the quality of tomograms also w.r.t. their intended use, whenever possible.

CHAPTER 4

MONTE CARLO MINIMIZATION

A different approach to the tomographic problem as described in the previous chapter is to use Monte Carlo (MC) Minimization. Unlike previously mentioned methods, MC techniques are based on randomly sampling the domain which contains the solution to the given problem. In this chapter, we first review the basic principles of Monte Carlo approaches and give a short overview of the sampling techniques that are relevant for this work. Then we briefly explain how the tomographic reconstruction can be formulated within a MC framework and discuss the related work in which different MC approaches were used in tomography. Similarly to the previous chapter, we show the geometric meaning of the MC approach to clarify the main differences to the previous methods.

The body of literature on Monte Carlo algorithms is large and we review here only the basic principles necessary to understand our work. For a comprehensive description of MC methods, we refer reader for example to work by Hammersley and Handscomb [1964] or Robert and Casella [2004].

4.1 Monte Carlo Methods

Informally, we can denote as a Monte Carlo (MC) method any approach that tries to find a solution to a given problem by randomly sampling a domain containing this solution. Monte Carlo methods are used for solving a wide range of problems, with the best known being numerical integration of functions over a complicated (often multidimensional) domain, simulation of stochastic processes, and optimization. In the context of this work, we will focus on the latter however most of the principles described here are general and apply also for different applications.

We distinguish two types of MC optimization: *Stochastic approximation* and *random search optimization* [Gilks et al., 1999]. In the stochastic approximation we seek to minimize a function $F(x)$ whose exact form is unknown and has to be obtained using Monte Carlo simulation. The function $F(x)$ is specified as an expectation value

$$F(x) = \mathbf{E}_Y[g(x, Y)] = \int_\Omega g(x, y) p_x(y) dy, \tag{4.1}$$

where $g(x, y)$ is a known function parametrized by noise $Y \in \Omega$ with probability distribution $p_x(y)$, possibly depending on specific x. As the expectation cannot be evaluated exactly, we use Monte Carlo evaluation of the integral, i.e. to evaluate $F(x)$ for some x we obtain samples Y_i, $i = 1..N$, from p_x and calculate

$$F_N(x) = \frac{1}{N}\sum_{i=1}^{N} g(x, Y_i).$$

Stochastic approximation methods include e.g. Robbins-Monro [Robbins and Monro, 1951] algorithm or Kiefer-Wolfowitz [Kiefer and Wolfowitz, 1952] algorithm.

In the random search optimization, the function $F(x)$ is known and can be typically evaluated directly. However, the deterministic approaches might fail to find the global minimum if the function $F(x)$ has many local minima and/or they might be impractical for high-dimensional spaces. The random search optimization overcomes these problems by evaluating the $F(x)$ at large number of points x_i, $i = 1..N$, that are generated randomly. The approximation $\hat{F}(x)$ of the global minimizer is then

$$\hat{F}(x) = \min_{i\in[1,N]} F(x_i).$$

Random search optimization techniques include e.g. *simulated annealing* [Kirkpatrick et al., 1983] or *cross-entropy* [Rubinstein and Kroese, 2004] approaches. The Progressive Stochastic Reconstruction Technique (PSRT) proposed in this thesis belongs to this category as well. It is worth noting that some optimization methods might combine the stochastic approximation with together with the random search optimization [Spall, 2003].

Intuitively, the quality of the solution is proportional to the number of samples – more samples yield more accurate solution. At worst, the rate of convergence is $N^{-\frac{1}{2}}$, i.e. to double the quality of a solution one has to quadruple the number of samples [Hammersley and Handscomb, 1964].

Due to processing of a large amount of samples, MC methods are generally slow and one often has to find a balance between the accuracy of the solution and computational time.

The efficiency of MC methods strongly depends on their sampling strategy. Its main purpose is to provide a probability distribution either for drawing of the samples or for deciding, whether the drawn sample should be accepted as a part of the final solution. Note that this distinction is not strict and that some strategies provide both. From the existing sampling techniques we focus here only on those that are important in the context of this thesis.

4.1.1 Uniform Sampling

The simplest probability distribution to sample from is the uniform distribution, where the whole domain is sampled with equal probability. The distribution is typically created by scaling random numbers taken from range $[0, 1)$ to the size of the domain. This approach can result in rather poor samples arrangement due to clustering of the samples. One approach to address this problem is to use the stratified sampling. It divides the domain into equally sized sub-domains and distributes the samples within these sub-domains. Alternatively, we can use deterministic pseudo-random sequences with good discrepancy properties, e.g. Halton sequence [Halton, 1964]. The main advantage of the Halton sequence is that we can stop sampling at any time and still obtain a reasonably uniform distribution (i.e. we do not have to know the exact number of samples in advance).

4.1.2 Importance Sampling

Even a good uniform distribution might not be very effective for sampling as there might be regions in the target function that should be sampled more densely. If we possess such information, i.e. if we can approximate the target function $F(x)$ by a function $f(x)$ with a similar shape, we can then draw samples from $f(x)$. This technique is called Importance Sampling (IS) because it concentrates samples in those regions where the function is strong. Consequently, for optimal $f(x)$ IS leads to a significant convergence speed-up. Note that importance sampling is orthogonal to the methods described in Sec. 4.1.1 and we can thus combine it with any of them arbitrarily.

4.1.3 Metropolis-Hasting Sampling

Metropolis-Hastings (MH) algorithm [Metropolis et al., 1953; Hastings, 1970] also samples more densely in areas with higher potential for improvement w.r.t. the target distribution. The advantage of this method is its ability to draw samples even from an unknown target Probability Density Function (PDF) $F(x)$ as long as we can locally evaluate a function $f(x)$ that is proportional to $F(x)$.

MH generates samples by performing a biased random walk that consists of randomly generating a new sample using a transition function t and acceptance/rejection of that sample based on an acceptance function A. Thus, for a given target PDF $F(x)$ the transition probability $F(x_{k-1} \rightarrow x_k)$ from the sample x_{k-1} to x_k can be expressed as

$$F(x_{k-1} \rightarrow x_k) = t(x_{k-1} \rightarrow x_k)A(x_{k-1} \rightarrow x_k).$$

On one side, the task of the transition function is to generate the next sample such that promising areas identified by the previously accepted sample are explored more effectively. On the other side, it must ensure that every point in the volume has a chance of being explored, a property called ergodicity. A typical choice for the transition function is a Gaussian blob centered on the previously accepted sample.

The acceptance function has to fulfill the condition of *detailed balance*, i.e. the probability of accepting the sample x_k after accepting the sample x_{k-1} has to be the same as the probability of accepting the sample x_{k-1} after x_k has been accepted:

$$F(x_{k-1})F(x_{k-1} \rightarrow x_k) = F(x_k)F(x_k \rightarrow x_{k-1}).$$

Moreover, the acceptance function has to make sure that a sample is accepted only if it improves on $F(x)$ and even then only if its contribution is on par with its neighbours. MH defines the acceptance function as:

$$A(x_{k-1} \rightarrow x_k) = min\left\{1, \frac{F(x_k)\,t(x_k \rightarrow x_{k-1})}{F(x_{k-1})\,t(x_{k-1} \rightarrow x_k)}\right\}.$$

The samples for which $A \geq 1$ are always accepted while the rest is accepted with probability corresponding to A.

Let $f(x)$ be a function proportional to $F(x)$:

$$\frac{F(x_k)}{F(x_{k-1})} = \frac{f(x_k)}{f(x_{k-1})}$$

then the acceptance probability can be computed as follows:

$$A(x_{k-1} \rightarrow x_k) = min\left\{1, \frac{f(x_k)\,t(x_k \rightarrow x_{k-1})}{f(x_{k-1})\,t(x_{k-1} \rightarrow x_k)}\right\}.$$

Note that if we use symmetric transition function t than the acceptance function reads:

$$A(x_{k-1} \rightarrow x_k) = min\left\{1, \frac{f(x_k)}{f(x_{k-1})}\right\} \tag{4.2}$$

and corresponds to Metropolis sampling.

Both transition and acceptance are based only on the last accepted sample and thus the random walk in MH creates a Markov chain. We shall note that the distribution of the samples in the chain created by MH converges to the target distribution in the limit $N \to \infty$, where N denotes the number of samples.

4.2 Monte Carlo Approach for Reconstructions

In the previous chapter we formulated the tomographic reconstruction as a least square problem and presented different mathematical approaches for its solution. Another way how to solve the least square problem is to use a Monte Carlo approach. If we regard the volume we want to reconstruct as the unknown target PDF $F(x)$ then we can try to estimate it by using a Metropolis-Hastings algorithm, with Eq. 3.2 providing the basis for function $f(x)$. The local evaluation of $f(x)$ guides the placement of the random walk samples. How exactly such an $f(x)$ should be defined is the topic of the next chapter.

4.2.1 Previous Work

The first attempt to solve tomographic reconstruction problem by using MC approach was introduced by Gordon and Herman [1971]. For each measurement p_i the authors identify a set of voxels that contribute to it and try to satisfy Eq. 3.1 by randomly choosing voxels from the set and changing their values until they obtain the measurement value. Of course, the correction on one set of voxels leads to the disruption of already satisfied measurements. The authors proposed three approaches how to deal with that and showed for small amount of projections that for a certain angular distribution of projections it is possible to satisfy all measurements. Due to the randomness of the algorithms, they could perform multiple reconstructions on the same dataset and obtain different results with the final tomogram being an average from all of them. The methods delivered good results but were computationally very demanding. Moreover, an extension to a larger number of projections was not straightforward and thus the approach has never found widespread use.

A similar approach was proposed by [Barbuzza et al., 2007; Barbuzza and Clausse, 2011] for tomographic reconstructions in the field of material science radiography. The authors perform one random walk over the voxels of the volume. The random walk is guided by MH and in each step of the random walk the transition function is used to pick a voxel whose value is then altered. This change is accepted only if it improves on the objective function. In their earlier work, they actually employ coarse-to-fine approach using multigrids to speed up the convergence and improve the quality of their results. The progressive multigrid approach was abandoned in the follow-up work and replaced by the incorporation of both a-priori information and regularization into the objective function.

A method based on the MH algorithm was also developed for tomographic high contrast fluid imaging by [Gregson et al., 2012]. This technique, called Stochastic Tomography (ST), performs many random walks to create spherical elements of constant size and energy called samples that gradually build up the final volume estimate. Their approach is gridless, i.e. the reconstructed volume is continuous for the purposes of the sample placement. To speed up the reconstruction process, the authors use a convex hull approach to quickly eliminate samples that are not visible from all input projections. This is possible since the reconstructed object is fully contained within the tomogram and does not reach its boundaries. For this reason, the authors also did not have to deal with boundary conditions for random walks. Similar to the previous approach, this method deals with high-contrast but sparse data (up to 16 projections) and thus requires additional regularization to obtain reconstructions with reasonable quality. We build on this approach and adapt it to the specific conditions present in cryoET in order to deliver quality reconstructions without the need for explicit regularization or a priori information about the target specimen.

To complete the list on tomographic methods based on MC approach we should mention the wide range of reconstruction methods that use simulated annealing technique. Simulated annealing is an extension of Metropolis algorithm that basically mimics the way a thermal system approaches its equilibrium state by slowly decreasing the temperature and is thus often use for finding a global minimum of nonlinear non-convex optimization problems as it can overcome the multiple-minima problem [Kirkpatrick et al., 1983; Li and Scheraga, 1987]. As such, it has been used in Computed Tomography (CT) [Haneishi et al., 1989; Qureshi et al., 2009], Single Photon Emission Computed Tomography (SPECT) [Webb, 1989; Kearfott and Hill, 1990], Positron Emission Tomography [Sundermann and Lemahieu, 1995], or recently in Electrical Impedance Tomography (EIT) [de Castro Martins et al., 2012]. All the approaches alter the values in the volume while progressively decreasing the temperature until they reach a global minimum of a given optimization problem. Typically, the 3D reconstruction is performed as a set of 2D reconstructions out of 1D projections.

4.2.2 Geometric Interpretation

In the methods introduced in the previous chapter each measurement value is evenly distributed among all voxels that lie on the corresponding ray and this distribution for a given set of voxels happens at once (see Fig. 3.3 and Fig. 3.4). Consequently, the values at low intensity regions are overshot while regions with higher intensities remain underestimated which leads to lower contrast. By contrast, the tomographic methods based on MC act more locally. First, a position (voxel) in the volume is randomly picked and then the value at this position is changed. If the new value improves the current volume estimate, the change is accepted, otherwise, it is rejected. It is thus possible to sample the high-intensity regions more densely, e.g. by using Metropolis-Hastings sampling strategy. It not only enhances contrast but also allows for full saturation of these regions. We show in the subsequent chapters how important this locality of MH is.

CHAPTER 5

PROGRESSIVE STOCHASTIC RECONSTRUCTION TECHNIQUE - PSRT

In the previous chapter we discussed a recent technique for tomographic reconstruction from radiographs proposed in a work of Barbuzza and Clausse [2011], where Monte Carlo Minimization (MCM) technique together with Metropolis-Hastings (MH) sampling strategy was used to perform the reconstruction. This idea was further extended in the work of [Gregson et al., 2012] for visible light 3D imaging of fluid dynamics. In both cases, the ability of MCM to easily incorporate a-priori information and/or regularization was the main reason for choosing this approach. However, we believe that the real potential of MCM lies in its flexibility to easily integrate techniques compensating for some of the limitations present in cryoET as well as its ability to focus more on specific parts of the volume during the reconstruction process. We adapt the basic algorithm of Gregson et al. [2012], called Stochastic Tomography (ST), to the reconstruction of cryoET tilt-series and extend it with a progressive coarse-to-fine scheme, a novel memory efficient solution to the specimen-level interior problem, and ability to easily perform region of interest reconstruction.

In this chapter we present the basic principles behind our novel approach called Progressive Stochastic Reconstruction Technique (PSRT). First, we provide an overview of the core parts of the proposed algorithm and then we introduce the coarse-to-fine scheme for improving both the quality of tomograms and convergence speed. We explain how we can easily incorporate the memory efficient solution to the interior problem into PSRT. Finally, we discuss the local character of PSRT and describe how it can be used to perform reconstruction for specific regions.

5.1 Algorithm Overview

Fig. 5.1 illustrates the high-level flow of PSRT algorithm. PSRT is an iterative method that reconstructs the volume by filling it up with spherical samples. Each sample has a defined small contribution to the volume, which we call energy. By adding the samples into the volume, we progressively decrease the difference between the measured projections and the projections of the current volume, which equals to minimizing the residuals $\bar{r}$ from Eq. 3.2. Generation of samples is inspired by Metropolis-Hastings (MH) strategy: We perform series of random walks during which we randomly generate new samples

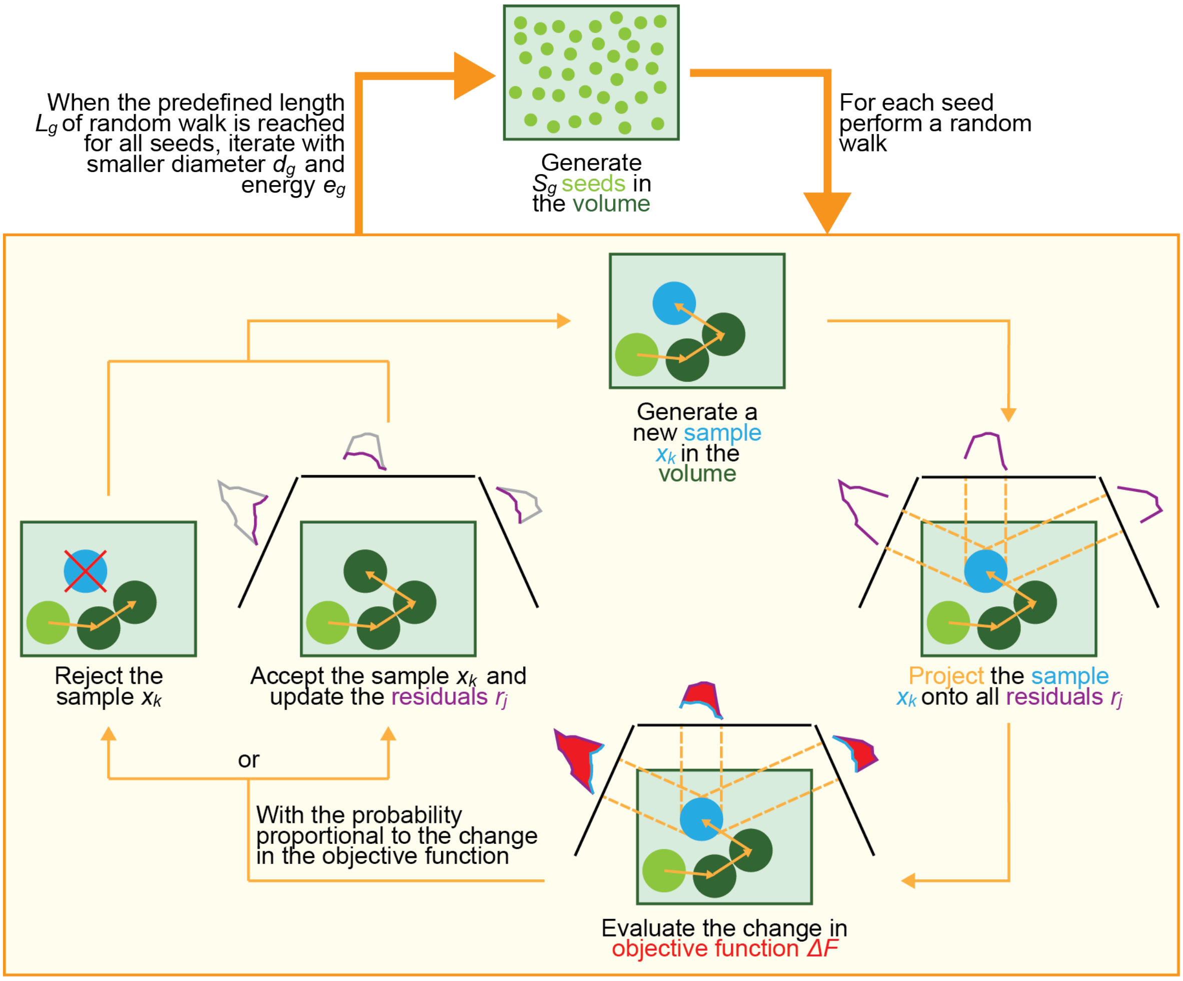

When the predefined length L_g of random walk is reached for all seeds, iterate with smaller diameter d_g and energy e_g
Generate S_g seeds in the volume
For each seed perform a random walk
Generate a new sample x_k in the volume
Reject the sample x_k
Accept the sample x_k and update the residuals r_j
Project the sample x_k onto all residuals r_j
or
With the probability proportional to the change in the objective function
Evaluate the change in objective function ΔF

Figure 5.1: A scheme diagram of one iteration of the proposed technique PSRT.

according to a given distribution and use a probabilistic acceptance function to decide whether the samples should be put into the volume or not. The starting samples of the random walks are called seeds.

The algorithm contains four free parameters that have to be specified by a user prior to the reconstruction: The starting diameter of a sample D_{start}, a total number of samples per voxel V, width of the transition function σ, and a scaling factor for the initial energy computation α. The meaning of each parameter is explained in the sections below.

5.1.1 Samples

Each sample is characterized by its coordinates, energy, and diameter. The coordinates are generated during the random walk. The starting energy of a sample is computed based on the input projections (see Sec. 5.1.3). The starting diameter is a user-specified parameter and it is decreased in discrete steps during the reconstruction, i.e. we perform a progressive, coarse-to-fine reconstruction. With this approach we first obtain a blurry but less noisy reconstruction produced by bigger samples and then progressively sharpen it using samples with a smaller diameter and spatially more coherent random walks.

The samples are not homogeneous but their energy is concentrated in the center and it is decreasing with the sample radius according to a Gaussian distribution. More details are provided in the next chapter in Sec. 6.2.1.

The coordinates of samples are, in line with Stochastic Tomography (ST), continuous which means that the samples can be placed anywhere within the reconstructed area (i.e. are not placed on any grid). Although we work with a continuous coordinate space, the reconstructed volume is discretized using voxels. Therefore we use voxel as a measurement unit in the following sections with its size corresponding to the size of a pixel in the input projections as they have the same side length.

5.1.2 Random Walk

The basic principle of minimization of Eq. 3.2 using a series of random walks was introduced by [Gregson et al., 2012] as the Stochastic Tomography (ST). This whole section reviews this principle, as it is essential for understanding the core of our algorithm.

Similar to other iterative reconstruction techniques, the algorithm seeks to minimize a suitably defined objective function F, which in the case of a least squares problem is defined as

$$F(\tilde{v}) = \sum_i (r_i^2),$$

where $\tilde{v}$ is a volume estimate and r_i are the corresponding residuals (see Eq. 3.2).

For a sample x_k the change of the objective function is evaluated as the difference between the volume without and with the sample:

$$\Delta F(x_k) = F(\tilde{v}_{k-1}) - F(\tilde{v}_k)$$

where $\tilde{v}_k = \{x_1, ..., x_k\}$ is simply a set of so far accepted samples. If $\Delta F(x_k) > 0$ the sample x_k improves on F. By accepting only samples that improve the objective function and rejecting all others, it is ensured that the algorithm will converge eventually.

During the random walk, the position of a new sample is generated by using a transition function t, which adds an offset to the position of the last accepted sample. The offset for each coordinate is computed randomly using a Gaussian probability function. Using a function with Gaussian probability ensures the ergodicity of the algorithm, since every point in the volume can be selected with non-zero probability. Note that the Gaussian transition function and the Gaussian distribution within the samples are two separate issues.

To evaluate the objective function $F(\tilde{v}_k)$, the new sample has to be projected onto the residuals. Since each sample affects only a small part of all the residuals $\bar{r}$, it is sufficient to evaluate only changes in these parts. Thus, the evaluation of ΔF is done only on a subset of residuals r_j, where $j \in I$, influenced by the new sample x_k:

$$\begin{aligned} \Delta^{\pm} F_j(x_k) &= r_j^2 - (r_j \pm e)^2 \qquad (5.1) \\ \Delta F(x_k) &= max\left\{ \sum_j \Delta^+ F_j(x_k), \sum_j \Delta^- F_j(x_k) \right\} \end{aligned}$$

where e denotes the sample energy which is defined in Sec. 5.2). In ST algorithm the energy is fixed during the whole reconstruction. Both positive and negative energy is used for the computation with the final sign to be chosen so the $\Delta F(x_k)$ is larger. This is necessary to allow for a correction in those areas where the energy is too high. If the sample is accepted, it is stored in the volume and the residuals are updated accordingly. Note that since the evaluation of ΔF uses values in residuals, the acceptance of a sample is influenced by all previously accepted samples, not only by the last one, thus the random walk does not create a Markov Chain [Gregson et al., 2012].

In line with Eq. 4.2, the acceptance function is computed as

$$A(x_k) = min\left\{1, \frac{\Delta F(x_k)}{\Delta F(x_{k-1})}\right\}.$$

The sample x_k is accepted if $a \leq A(x_k)$, where a is a random number from the range $(0, 1]$.

For larger volumes performing only one random walk is not practical. First, it is sensitive to a placement of a first sample in the sense that if this sample is placed in the region with lower probability acceptance it could significantly slow down convergence. Second, it would take a large number of samples to cover the whole reconstructed area. To overcome this, the authors perform many shorter random walks instead of only one. At the beginning of the reconstruction seeds are distributed over the whole reconstructed area and for each seed one random walk is initiated. Note however that the random walks are performed sequentially, since it is not straightforward to parallelize neither the evaluation of ΔF nor updating residuals in case of sample acceptance [Gregson et al., 2012].

5.1.3 Energy of the Sample

The energy of a sample for the first iteration is computed according to the following equation:

$$e_{start} = \alpha \frac{avg(P)}{VN}, \tag{5.2}$$

where V denotes the user parameter specifying the total number of samples per voxel, N the number of voxels in the volume, $avg(P)$ denotes average intensity of the input projections, and α is the scaling factor which compensates both for rejected samples and samples accepted with negative energy. Both the rejection ratio and the number of samples accepted with negative energy depend on the properties of the input projections. Intuitively, the rejection ratio as well as the number of correction samples will be higher in case of experimental datasets with low SNR than by the piecewise homogeneous tomograms reconstructed from noiseless projections. For an experimental dataset one should start the reconstruction with bigger samples to reduce the noise and set the width of the transition function relatively to the size of the samples. If the transition width is too large, the random walk can cover larger area which could increase the rejection rate as the experimental data are highly heterogeneous. On the other hand, if the width is too small, the random walk is very dense and thus the amount of samples accepted with a negative energy could be higher as there would be a greater need for correction.

5.2 Progressive Reconstruction Scheme

Our key contribution is the introduction of the progressive reconstruction scheme, which is one of the prerequisites for successful cryoET reconstruction. Instead of using samples with a fixed size, we introduce samples with different diameters. This allows PSRT to deliver smooth reconstructions with enhanced contrast without losing the high-frequency structural information. The main purpose of bigger samples is to fill the volume quickly and obtain a rough reconstruction with lower noise levels that is then further refined using samples with smaller diameter. This approach leads not only to a major speed-up in convergence, but more importantly, it makes the algorithm feasible without any need for further regularization. As it was shown in [Gregson et al., 2012], using small samples of a fixed

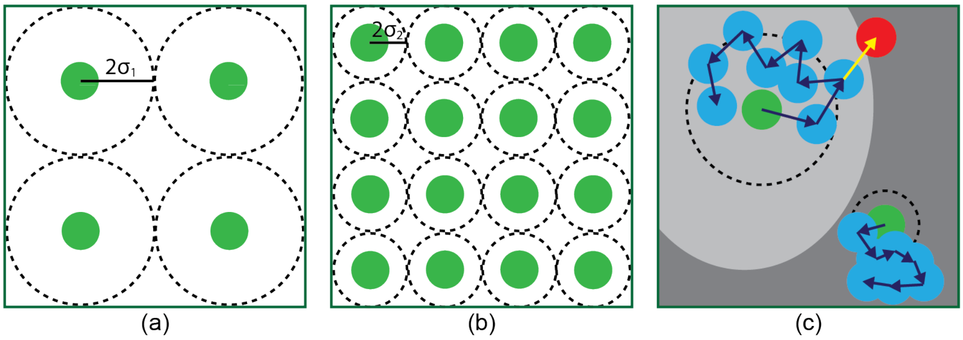


Figure 5.2: Scheme of seed distribution. The amount of seeds is influenced only by the size of the reconstruction area and transition width σ_g. (a) Distribution for transition width σ_1. (b) Distribution for the same reconstruction area dimensions but with smaller transition width σ_2. The positions of seeds in both images are just for illustration purposes - in the reconstruction the seeds are not placed on a regular grid but are distributed using a Halton sequence. (c) A random walk scheme. The accepted samples (blue) are more likely to remain within the $2\sigma_g$ radius (dashed line) of a seed (green). The samples placed in the region with different density are more likely to be rejected (red disc) and the random walk continues in the region with a similar density. A random walk is more dense if the σ_g is low, as shown in the right corner of the image.

size leads to noisy reconstructions and regularization is necessary to achieve a quality comparable to other state-of-the-art methods.

We define one iteration g of the algorithm as a processing of all seeds with the same diameter d_g together with their entire random walks.

As we use samples with Gaussian energy distribution, the footprint of a sample onto residuals is not homogeneous. As a result, Eq. 5.1 has to be changed in the following way

$$\Delta^{\pm} F_j(x_k) = r_j^2 - (r_j \pm e_g * w_{jg})^2, \tag{5.3}$$

where e_g denotes the sample energy in the g^{th} iteration and w_{jg} denotes the weighting due to the sample's decaying Gaussian footprint.

For each iteration we have to set an amount of seeds S_g and length of their corresponding random walk L_g. For a given iteration g, S_g is computed as follows:

$$S_g = \frac{dim_x dim_y dim_z}{(4\sigma_g)^3} \tag{5.4}$$

where $dim_{x,y,z}$ are dimensions of a reconstructed area and σ_g is the width of the Gaussian transition function. The width σ_g corresponds to the standard deviation in the Gaussian distribution. Since the transition function is used to generate a new sample within a random walk, the width directly influences the size of the region which might be reached within one random walk (and which might differ for each iteration). Since the σ_g parameter corresponds to the standard deviation in the Gaussian distribution, a new sample is going to be placed in a region with a diameter 4σ with probability of 95.45%. Since the Gaussian distribution ensures that every point in the volume can be selected with non-zero probability, the random walk can strongly diverge from the seed. However, for highly heterogeneous data is such a divergence highly unlikely due to the Metropolis-Hastings sampling strategy and it is more likely that the random walk stays approximately in the region of 4σ. Thus dividing the size of the reconstructed area by the diameter gives a good estimate for a reasonable amount of seeds for a given iteration, see Fig. 5.2. Note that the seeds are generated using Halton sequence and thus have a uniform distribution.

A random walk of each seed should cover the region of $(4\sigma_g)^3$ with sufficient amount of samples. Thus, given a number of Samples Per Voxel (SPV) for an iteration V_g (note that $V = \sum_g V_g$), we compute the amount of samples for a random walk for the g^{th} iteration as

$$L_g = V_g \frac{(4\sigma_g)^3}{{d_g}^3}.$$

At the end of each iteration, we decrease the sample's diameter d_g and energy e_g (for details see Sec. 6.2.1 and Sec. 6.3, respectively). As will be explained in Sec. 6.2.1, with decreasing diameter increase the Gaussian weights within the sample and thus we have to decrease the energy in order to keep the highest value within the sample constant. Consequently, our approach does not qualify as simulated annealing (see Sec. 4.2.1).

5.3 Memory Efficient Solution for Specimen-Level Interior Problem

The interior problem in cryoET is caused by a specimen being larger than the scanned area. It produces strong artifacts at the edges of the reconstructed volume that not only make the edges of the tomogram unreliable but also change the histogram of the tomogram as they cause strong oversaturation of the edges. This can both limit the contrast as well as make the subsequent processing like thresholding or windowing more difficult. To suppress them, one should ideally reconstruct the whole specimen, i.e. extend the volume size to correspond to the size of the scanned specimen. This not only increases the computation time but also significantly increases memory requirements, making this strategy practically infeasible. Within our framework, we can however easily achieve the required extension without increasing the memory requirements. We simply extend the area to in which seeds and samples can be placed and process them normally, except in the case of sample's acceptance. Here we do not store the sample, but only update the affected residuals correspondingly. So the reconstruction runs as if the complete specimen were present but the samples in the extended area are never stored. In other words, although the computation requirements increase, the memory requirements remain unchanged.

5.4 Locality of PSRT

One of the building blocks of all state-of-the-art methods is the backprojection step, where values from the projections are distributed (smeared) back into to the volume along a corresponding ray, i.e. each value changes a large part of the volume (see Fig. 3.3). In contrast, in the PSRT, each random walk performs a local reconstruction of the volume, whereas the areas with a need for higher energy are automatically sampled more densely. In case of a sample's acceptance, the sample changes only the small part of the volume it is covering which makes it possible for PSRT to focus on more interesting areas during the reconstruction. This behavior is especially important with respect to the progressive scheme as it allows the smaller samples to refine only those areas where larger samples could no longer bring any benefit, i.e. high-frequency regions.

5.4.1 Region of Interest Reconstruction

The global character of state-of-the-arts methods is cause for their inability to easily focus the reconstruction process only on specific region within the volume. In order to perform a Region Of Interest

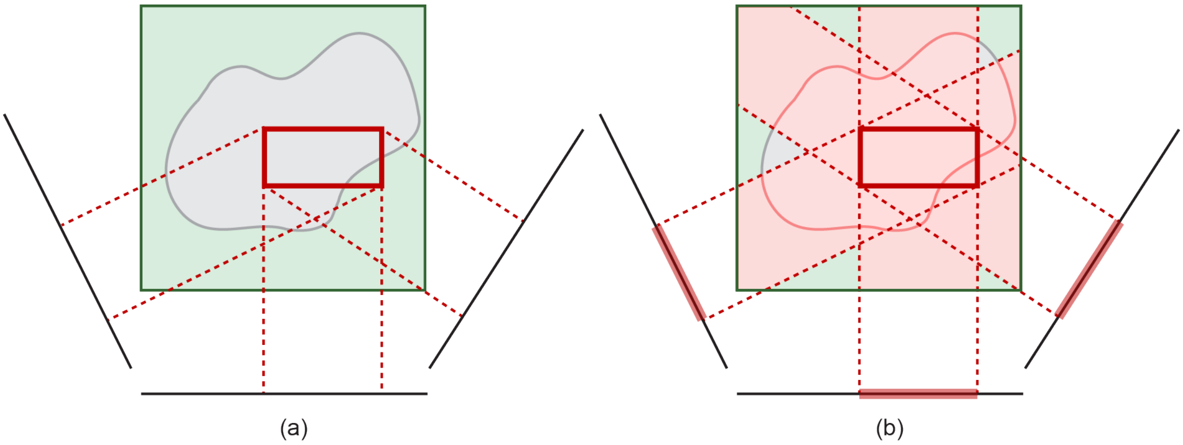


Figure 5.3: (a) Projection of the region of interest (red rectangle) on the input projections to localize the parts that correspond to the Region of Interest (ROI). (b) Indication of the scope of information contained in the input projections. The values do not correspond only to the region of interest but to the whole slab through the volume. Outside the region of interest is the information from the input projections inconsistent, resulting in artifacts on the edges of the ROI.

(ROI) reconstruction, it is necessary to localize the respective parts in the input projections that correspond to the ROI, cut them out from the input projections, and perform the reconstruction using those. Each cut-out projection still contains information from the entire volume slab that contains the ROI (see Fig. 5.3. Consequently, if we use the cut-out projections directly, we put the information corresponding to the whole tomogram into only a small part of it. It will demonstrate itself in form of strong artifacts on the edges of the reconstructed area. To overcome this, one can either reconstruct bigger ROI and then trim it [Chen and Förster, 2013] or preprocess the projections prior to the reconstruction. The preprocessing typically consists of weighting based on the relative sizes of the whole tomogram and the ROI, i.e. in case of geometric methods we weight the values by the ratio between the ray-ROI intersection length and ray-volume intersection length [Xu et al., 2010]. It is worth noting that the values in ROI reconstructions usually differ from those obtained with whole-tomogram reconstruction – the difference between high and low-intensity structures is decreased, which might lead to misinterpretation of ROI tomograms.

In contrast, the local character of PSRT allows us to perform ROI reconstruction easily and without any need for additional weighting. We use the positions and shapes of ROIs to place the seeds only to those regions and then perform a series of random walks as usual. By setting the transition width reasonably small, we ensure that the random walks stay roughly in our ROIs. Note however that we do not set any boundaries for the random walks. If a random walk gets outside its region, it progresses normally. It is not stopped or forced to return to the designated region. This prevents any artifacts to arise on the edges of the ROIs. So in essence, we direct the algorithm to process the localized areas with higher priority but it is up to the MH procedure to determine, whether such additional processing improves the objective function and only if it does, are such corrections accepted.

As we will show in Sec. 7.3, PSRT can recover the same values in ROIs as those obtained by whole-tomogram reconstruction. However, it is important to point out that we do not start the ROI reconstruction on an empty volume. Instead we perform one or more iterations with bigger samples and then perform ROI reconstructions within the blurred tomogram to obtain high-frequency details in ROIs. This ensures, that we do not concentrate the information from the whole tomogram only in the ROIs and thus helps to avoid the artifacts. Moreover, the reconstructions in the localized areas are not performed sequentially one after another. If we performed the reconstruction region per region, most of the energy from the residuals would be concentrated in the regions processed as first. To prevent that, we mix the order in which the series of random walks are performed. This is possible, since each region typically contains hundreds of random walks. In that way, all regions are reconstructed simultaneously. Finally, in PSRT, we do not try to zero out the residuals with one sample - the sample's energy is independent from the actual values in the residuals and in comparison to them is very small which prevents a single random walk to concentrate all the information from the residuals in only one region.

Note that the region of interest reconstruction problem is also known as the interior problem and will be discussed in more detail in Ch. 10.

CHAPTER 6

IMPLEMENTATION

Currently, the typical dimensions of cryoET reconstruction are around 4096×4096×500 voxels. As we will see in the following chapters, in PSRT we need at least 1000 samples per voxel to obtain a reconstruction with reasonable quality, making PSRT very demanding w.r.t. computation time. It is thus important to optimize every step of the algorithm in order to achieve the best efficiency both in terms of computation time and memory requirements.

In this chapter we present details of our PSRT implementation. First, we discuss volume representation and effective evaluation of our objective function. Then we explain how we compute the sample energy for each iteration. We describe the way we ensure the proper behavior of the random walks near the boundaries of the reconstructed area and review the implementation of the memory efficient solution for the interior problem. Lastly, we describe the Region Of Interest (ROI) reconstruction and discuss various possibilities for ROI specification. The possible parallelization will be discussed in Ch. 9.

6.1 Volume representation

As mentioned in Sec. 5.1.1, our reconstructed area is continuous, i.e. we do not place samples onto a grid, which allows us to compute more accurate projections of the samples onto the residuals. However, storing the final reconstruction in standardized file formats requires a grid representation and thus samples have to be discretized into voxels at some point. This can be done either right before the write-out of the volume or each sample can be discretized right after it has been accepted.

In the first case, we have to store all accepted samples during the whole reconstruction process. The only advantage of this approach is the ability to define a grid resolution at the end of the reconstruction process [Gregson et al., 2012]. However, in cryoET settings this approach would lead to unfeasible memory requirements. Storing the accepted sample means storing its coordinates and the diameter. Each accepted sample would thus occupy 13 bytes of memory as the coordinates would have to be stored as floats and the diameter as unsigned char. As we will see in the following chapters, we use at least 1000 samples per voxel for the reconstruction. Thus, even if only half of the samples are accepted, the memory requirements per voxel would be around 6500× higher than in the case of their immediate discretization, where only one float value has to be stored. Thus, for the volume of size

4096×4096×500 voxels we would need at least 50TB of space.

In the second case, we have to define the voxel size, and thereby the resolution, prior to the reconstruction process. Each accepted sample can be then discretized into voxels immediately. As shown above, this approach has lower memory requirements and it is therefore possible to perform the reconstruction in memory. As a result, we prefer the voxel representation of the volume. We set the voxel size to be the same as is the pixel size of the input projections. However, it is possible to further subdivide both voxels and pixels into smaller regions if we want to achieve higher precision (see Fig. 6.1a and Fig. 6.1b). Due to the immediate discretization, this subdivision can be done effectively using the precomputed weighting kernels as described in Sec. 6.2.1.

Note that in our implementation the volume is fixed with its center in the origin of our coordinate system and the input projections are rotated around it. This is done only once, prior the reconstruction.

6.2 Evaluation of the Objective Function

Evaluation of the change in the objective function as defined in Eq. 5.3 is very time consuming and it is thus important to perform it efficiently. Especially, if samples of larger diameters are involved. There are two crucial parts of the evaluation: Projection of the sample onto the residuals and computation of the change in the residual values. We present in this section an efficient way for sample representation based on precomputed weighting kernels that speeds up both the sample's projection and subsequent residual change evaluation.

6.2.1 Sample Representation

The energy within a sample is not homogeneously distributed. We model the energy as a 3D Gaussian distribution cropped to the sample's radius. In the discretized settings, this distribution is represented as a Gaussian weighting kernel. As the weights are the same for all samples of a given size, the weighting kernel can be precomputed at the beginning of each iteration. The weighting kernel is stored as

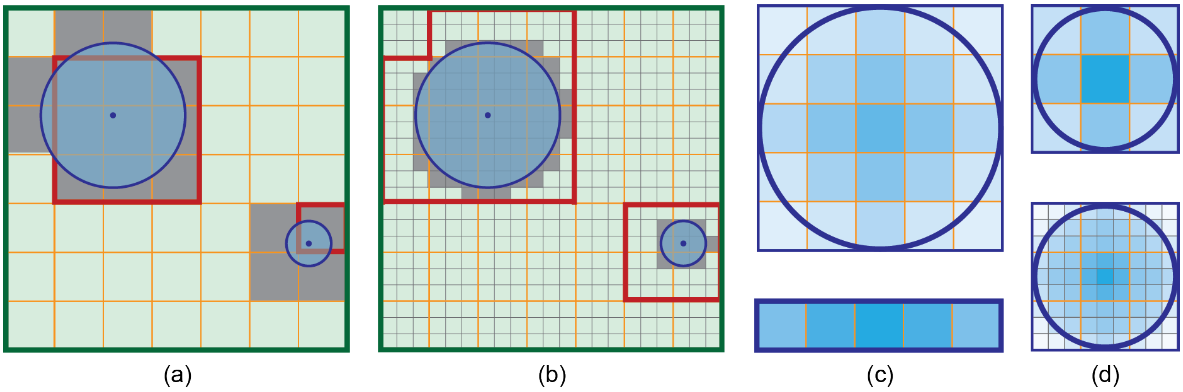

(a) (b) (c) (d)

Figure 6.1: The difference between sample processing in case of (a) no voxel subdivision and (b) subdivision of size 3. The darker area represents (sub)voxels affected by the sample and the red border represents the voxels into which the sample is stored. The affected area is based on the position of the sample's center. As it can be seen in (a), without the sudivision we obtain slightly inaccurate and incomplete set of voxels, whereas with subdivision (b) the set of affected voxels is complete. (c) The weighting kernel of a sample with diameter 5 (top) with the corresponding projection kernel (bottom). (d) Comparison between the weighting kernels for sample with diameter 3 without (top) and with (bottom) further subdivision.

a list of weights with their spatial offsets to the kernel's center. This allows for fast processing when updating the volume.

As the projection of 3D Gaussian distribution onto a 2D plane is guaranteed to be Gaussian as well, we may handle the projections of a sample onto residual images in a similar way. For a given sample size we precompute a 2D weighting kernel as well as coordinates' offsets. The projection of the sample onto a residual image then consists only of projecting its center. The rest of the pixels influenced by the sample is found using the precomputed offsets. This optimization allows us to evaluate ΔF in a very efficient way.

In case of subdivision, we create one 3D weighting kernel for each subvoxel position and one 2D kernel for each subpixel position. Once a sample is accepted, we find out the subvoxel in which its center belongs to, pick the corresponding weighting kernel and update the volume accordingly. We proceed in the same way for sample's projection. Note that a subdivision size is given for one dimension, i.e. if the size is s, we subdivide each voxel into s^3 subvoxels and each pixel into s^2 subpixels (see Fig. 6.1c).

Since we use the Gaussian weighting kernels for discretization, the diameter of the sample (given in voxels/pixels) is always an odd integer and we decrease it after each iteration by 2 until we reach the number of iterations specified by user with the smallest possible diameter being of course 1. Thus, given a starting diameter D_{start}, the maximum number of iterations is $\left\lfloor \frac{D_{start}}{2} \right\rfloor$.

Note that the weighting kernel is normalized to sum up to 1, which leads to increasing weights with decreasing sample diameter.

6.2.2 Projection Computation

Thanks to the weighting kernels, we have to project only the sample's center onto the residuals. For a sample with center c and a residual image m, the projection can be computed as a dot product:

$$x = h_m \cdot (c - r_m)$$
$$y = v_m \cdot (c - r_m)$$

where x, y denotes projected coordinates in the image (residual) space, h_m, v_m are 3D vectors for horizontal and vertical pitch of the rotated residual image m, and r_m is the 3D position of the upper-left corner of the rotated residual image. We can precompute h, v and rc for every residual image prior the reconstruction and further speed up the computation.

6.3 Energy Computation

At the beginning of g^{th} iteration, the values in both 2D and 3D kernels are multiplied with the energy e_g. Since the values in the weighting kernels are increasing after each iteration, we have to decrease the energy accordingly in order to keep the maximum value within the sample constant through the whole reconstruction. The energy for the first iteration is set as $e_1 = e_{start}$ and for the subsequent iterations, it is computed as

$$e_g = e_{g-1} \frac{w_{g-1}}{w_g} \tag{6.1}$$

where w_g denotes the central (and thus the maximal) weight in the weighting kernel corresponding to the g^{th} iteration. Decreasing the energy in each iteration is essential, as otherwise the overall sample's energy would be increasing during the reconstruction process, leading to lower acceptance probability at the later iterations and thus slower convergence of the algorithm.

In Sec. 4.2.1 we briefly mentioned simulated annealing as a version of the MH algorithm with progressively decreasing energy that is suitable for optimizing non-convex functions. If needed (i.e. in case of more complex formulation of the reconstruction problem), we could switch to simulated annealing by adding additional scaling factor to Eq. 6.1.

6.4 Random Walks Behavior on the Boundaries

If we constrain random walks only to the reconstructed area, the areas near its edges might be undersampled as the probability of generating a sample near the boundaries is generally lower (especially in case of bigger samples for which we typically choose relatively high value of the transition width σ). The undersampling poses a problem in cryoET settings where the whole reconstructed region contains meaningful information.

To ensure sufficient sampling, we implement two measures. First, we virtually, i.e. without changing the actual dimensions of the volume, extend the area where samples (but not seeds) can be placed (see Fig. 6.2a) and wrap those samples that are generated outside of the extended area. Note that this extension is minimal. We found empirically that the starting sample diameter is sufficient.

The second measure is normalizing the change of the objective function ΔF. The change ΔF is defined as the sum over the changes of the residual values affected by the current sample x_k (see Eq. 5.1). If the projection of the sample x_k onto residuals does not exceed the boundaries of any of the residuals then the number of affected pixels j is $d_g^2 * M$ with d_g being the diameter of the sample in the g^{th} iteration and M denoting the total number of residual images. For the sample x_k that is placed near the edges it is likely that its projection onto the residuals may exceed boundaries of some of the residuals. In such case, the number of affected pixels j would be lower than $d_g^2 * M$. Consequently, $\Delta F(x_k)$ would be smaller than for other samples from the random walk which would decrease the acceptance probability for the sample x_k. To avoid that, we normalize ΔF in the following way:

$$\Delta F(x_k) = \Delta F(x_k) * \frac{M d_g^2}{j}.$$

6.5 Memory Efficient Solution for Specimen-Level Interior Problem

As mentioned in Sec. 5.3, the implementation of a Memory Efficient Solution for the specimen-level Interior Problem (MESIP) is simple. We virtually extend the reconstructed area in x dimension and allow not only samples but also seeds to be placed there (see Fig. 6.2b). We process the samples in the extended area normally. However in case of acceptance we only update the affected residuals without storing the samples. Thus, the extension does not increase memory requirements.

The size of the extension w_{ext} is computed in the following way:

$$w_{ext} = min\{w_{scanned}, w + 2 * d * \tan(\alpha_{max})\}$$

where d is the depth and w the width of the reconstructed area (in voxels), α_{max} is the maximum angle for a given tilt-series and $w_{scanned}$ is an approximate width of the scanned specimen. For the derivation of this formula see Sec. 10.3. As the latter is generally not known, we approximate this value by $2 * w$. The derivation of this formula can be found in Sec. 10.3. Note that y and z still have to be extended by the starting sample diameter to avoid undersampling near the edges (see Sec. 6.4).

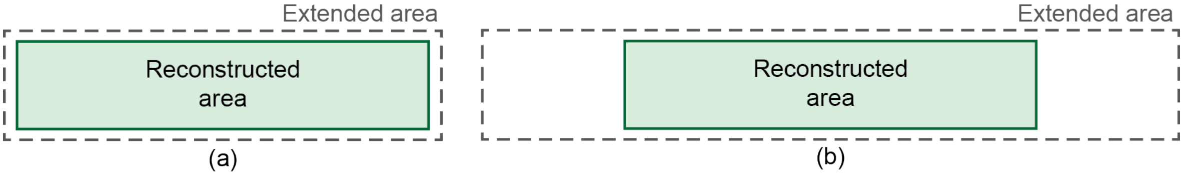


Figure 6.2: Extension scheme for ensuring ergodicity at the edges of the reconstructed volume. (a) Extension in case of no interior problem. The dashed line shows the extended area for the samples. (b) Extension when taking the interior problem into consideration. The extension is larger in x dimension as the specimen is wider than the detector. In this case, not only samples but also the seeds are placed in the extended area.

6.6 Region of Interest Reconstruction

We can regard Region of Interest (ROI) reconstruction as Importance Sampling (IS). We simply sample the regions we want to focus on more densely by putting more seeds into them. Note however that unlike IS described in Sec. 4.1.2, the placement of the seeds is not random as we again use the Halton sequence for their distribution within a region. Assuming a ROI is a bounding box, we compute the number of seeds for each region according to Eq. 5.4, except that we replace the dimensions of the volume by the dimensions of the region. The calculation of the length of random walks remains the same as the random walks are independent from the shape or size of the ROI.

ROI can be specified in two ways: either by its center and dimensions or as a binary mask. The former works well for convex objects while the binary mask is preferable in case of more complex structures (e.g. membranes or filaments). The mask has to be of the same size as the reconstructed volume and can contain ROIs of different shapes. Since our current implementation requires a ROI to be a bounding box, we have to process the mask prior to reconstruction. We create tight bounding boxes by splitting the mask into subregions with dimensions $(v,1,1)$, where v corresponds to a set of neigbouring voxels (in x dimension) that are all marked as ROIs (see Fig. 6.3).

Although this approach works, there might be more efficient ways how to specify ROIs of certain shapes. For instance, for a centerline of the filament mask, the ROI could be parametrized by a primary dimension that runs along the centerline with the two other dimensions running along the primary directions in the plane perpendicular to the centerline. In this case, the number of seeds could be computed in the same way as we do now. For an extremely complex ROI, the most appropriate approach for seed placement would be to employ a “real” importance sampling that interprets the ROI mask as a distribution to draw the samples from. The estimation of number of seeds would be

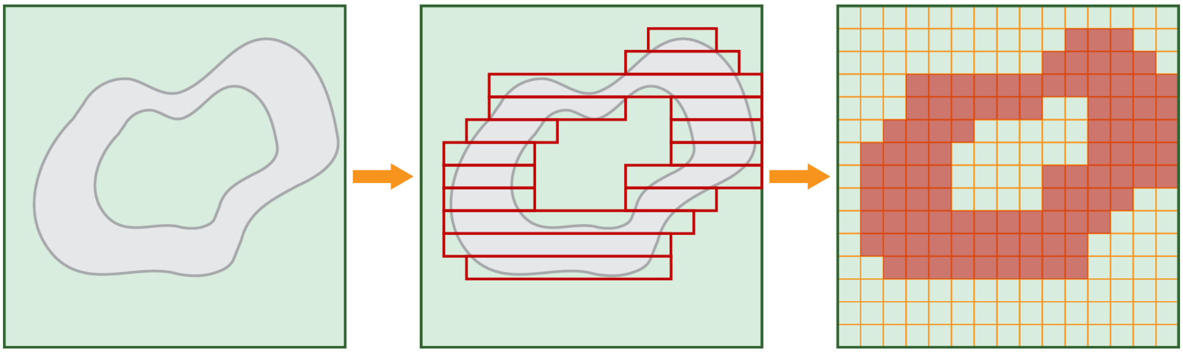


Figure 6.3: Creation of ROIs from a binary mask. The binary mask is split into subregions as indicated by the boxes with red borders.

however more complicated in this case.

As mentioned in Sec. 5.4.1, we do not perform the ROI reconstruction on an empty volume but on a partially reconstructed tomogram. We can either use an existing tomogram and its corresponding residuals or we can start the reconstruction from scratch and specify at which iteration we want to switch from the whole tomogram reconstruction to the ROI reconstruction. In the latter case, we refer to the whole tomogram reconstruction as to the initial reconstruction or initial phase and to the ROI reconstruction as to the refinement phase or importance sampling.

CHAPTER

7

PROPERTIES OF PSRT

In this section we present the results obtained with PSRT from both 2D and 3D synthetic datasets as well as evaluate its performance on real experimental cryoET data. We show the crucial importance of the coarse-to-fine scheme w.r.t. the reconstruction quality, especially in terms of noise reduction and contrast enhancement. We demonstrate that our memory efficient solution for the specimen level interior problem removes the artifacts associated with this problem. We evaluate the locality character of PSRT and show that the region of interest reconstruction is able to recover more accurate values than alternative approaches. We discuss memory requirements as well as computational time needed for different types of reconstructions. Finally, we review the current parameter setup and discuss possibilities for lowering the number of parameters.

7.1 Coarse-to-fine Reconstruction

In this section we first show the necessity of the coarse-to-fine scheme by performing a set of experiments on 2D datasets and compare the results to Weighted Back Projection (WBP) and Simultaneous Algebraic Reconstruction Technique (SART). We then validate our approach on 3D synthetic dataset by reconstructing the structure of a yeast ribosome. Finally, we discuss the influence of this progressive approach on the noise levels and contrast in tomograms.

All WBP reconstructions were made using the IMOD package [Kremer et al., 1996]. The SART reconstructions were made using the Ettention framework [Dahmen et al., 2016]. To obtain a fair comparison, the SART reconstructions were run using different relaxation factors (0.05, 0.3, 0.5, 1.0) and different number of iterations (1, 10, 15, 20, 25, 30) and the reconstructions that provided the best result in terms of error were chosen. As the WBP implementation from IMOD significantly changes the range of the reconstructed values, a normalization of all tomograms was necessary in order to compare them. The normalization was done using the EMAN2 package [Ludtke et al., 1999]. The comparison was done by measuring the Relative Root Mean Square Error (RRMSE) as well as by comparing features relevant for a given dataset. Settings of the user-defined parameters are given in the second column of Tab. 7.1.

Parameters	2D Synthetic Data			3D Synthetic Data			3D Synthetic Data with Noise		
SPV	350	350	300	150	300	800	150	300	800
Transition Width	5.0	3.0	1.0	3.5	1.0	0.5	3.5	1.0	0.5
Diameters	5	3	1	7	5	3	7	5	3
Scaling Factor	50			4000			Depends on SNR		

Table 7.1: Parameter setup for experiments on synthetic datasets.

7.1.1 2D Synthetic Dataset

We performed two 2D reconstructions of the well-known Shepp-Logan Phantom (SLP) [Shepp and Logan, 1974] from 1D projections. The first dataset consisted of 160 projections, each of size 512×1 pixels, covering the full angular range of 360°. The second dataset consisted of 65 projections, of the same size, with tilt angles from -64.5° to 62.5°, introducing the missing wedge problem into the reconstruction process. For each reconstruction, we measured the RRMSE and compared the line profiles that go through the tumors present in the tomogram. The line profiles show us whether a reconstruction method is able to recover the tumors well enough for them to be detected automatically.

First, we discuss our results on the complete dataset. We compare our coarse-to-fine approach to the non-progressive approach and to WBP and SART. An initial experiment was performed by using a constant sample size during the whole reconstruction. The results are shown in Fig. 7.1. As expected, using the diameter of size 1 produced very noisy reconstruction. Increasing the sample's diameter to 5 lead to significant noise reduction but at the same time introduced ringing artifacts which can be observed not only visually but also on the corresponding graph. This effect got more apparent with further increasing the sample's diameter, as can be seen in the reconstruction with the sample size 9. The relative RMSE was 24.4%, 18.9%, and 39.8% respectively, reflecting the visual quality of the reconstructions.

To correct for the ringing artifacts, samples with smaller diameter than the starting one are necessary. We performed the reconstruction using our coarse-to-fine approach with starting sample diameter 5 and 9. The final reconstructions did not contain any artifacts and in comparison to the reconstructions with fixed diameter showed better contrast and less noise, see Fig. 7.2. Moreover, the tumors were not only clearly visible but more likely to be found using a contrast-based automatic detection method, as can be seen from the corresponding graphs. While there were only small differences between the reconstructions with diameters 5 and 9, the latter needed around 20% less computation time.

The reconstruction with the starting diameter 5 was compared to WBP and SART (see Fig. 7.3 and Fig. 7.4). In case of the complete projection set, the results are visually comparable, with all methods providing very good tumor recovery. SART provided the best reconstruction in terms of RRMSE with only very little artifacts, while PSRT showed the best contrast with slightly more artifacts in the homogeneous areas. WBP, while generally comparable, was trailing in all measured aspects.

In the missing wedge dataset, PSRT recovered the tumors better than WBP or SART, provided better contrast as well as the best relative RMSE score (see Fig. 7.4). Only the artifacts within the homogeneous part of the skull were better suppressed by SART.

We conclude that PSRT correctly reconstructs the 2D SLP and that the results are better or comparable to the state-of-art methods used in cryoET.

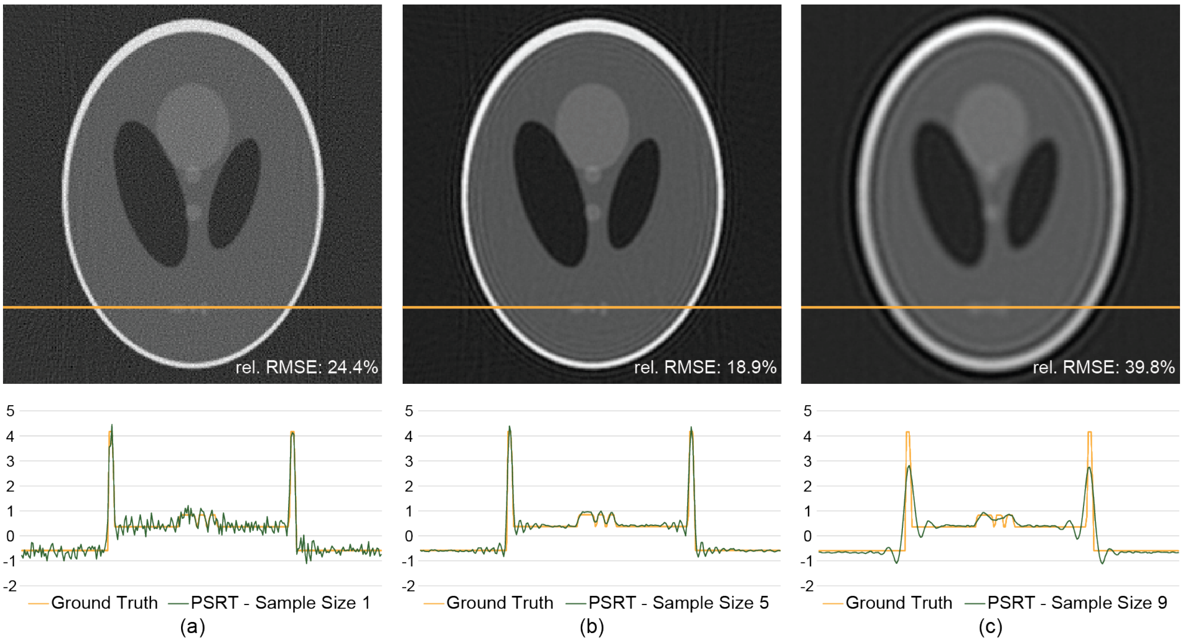


Figure 7.1: 2D reconstructions of size 256×256 of Shepp-Logan Phantom obtained using PSRT with constant sample size and corresponding graphs showing the values along the indicated lines: a) Sample size 1 pixel, relative RMSE 24.4%. b) Sample size 5 pixels, relative RMSE 18.9%. c) Sample size 9 pixels, relative RMSE 39.8%.

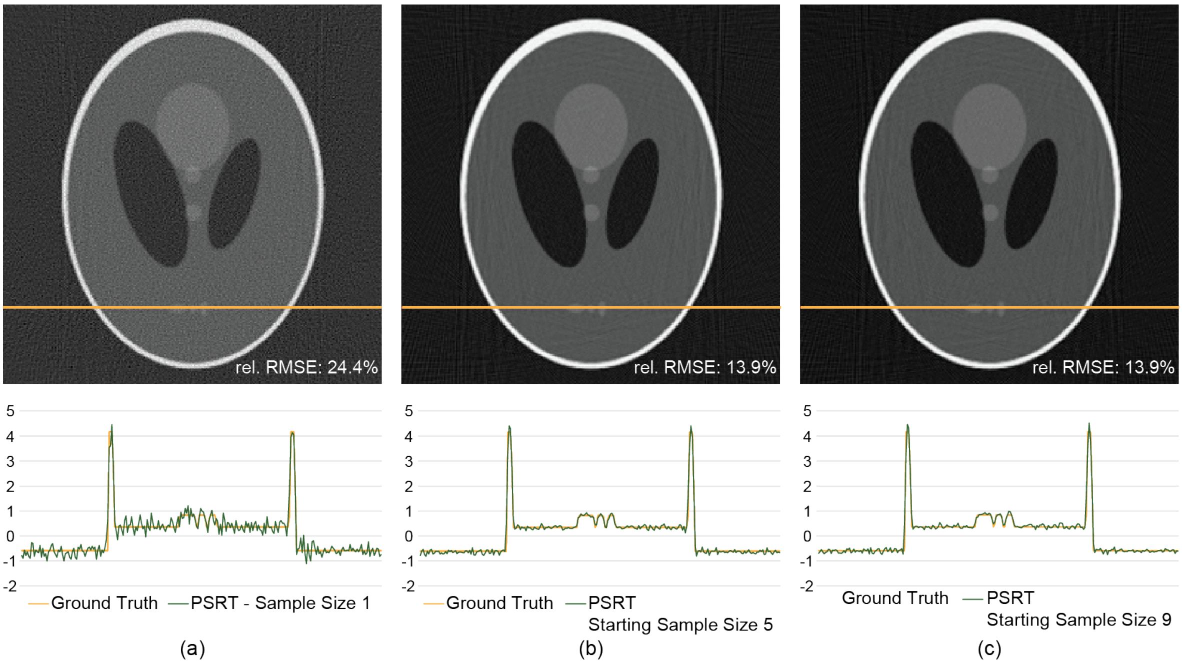
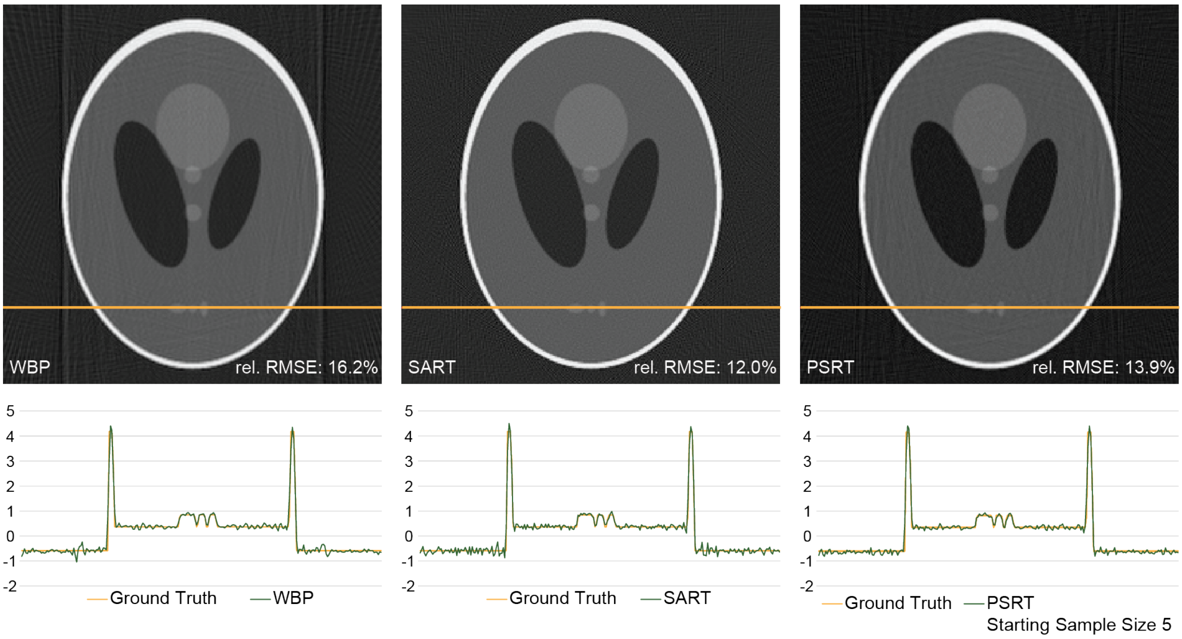
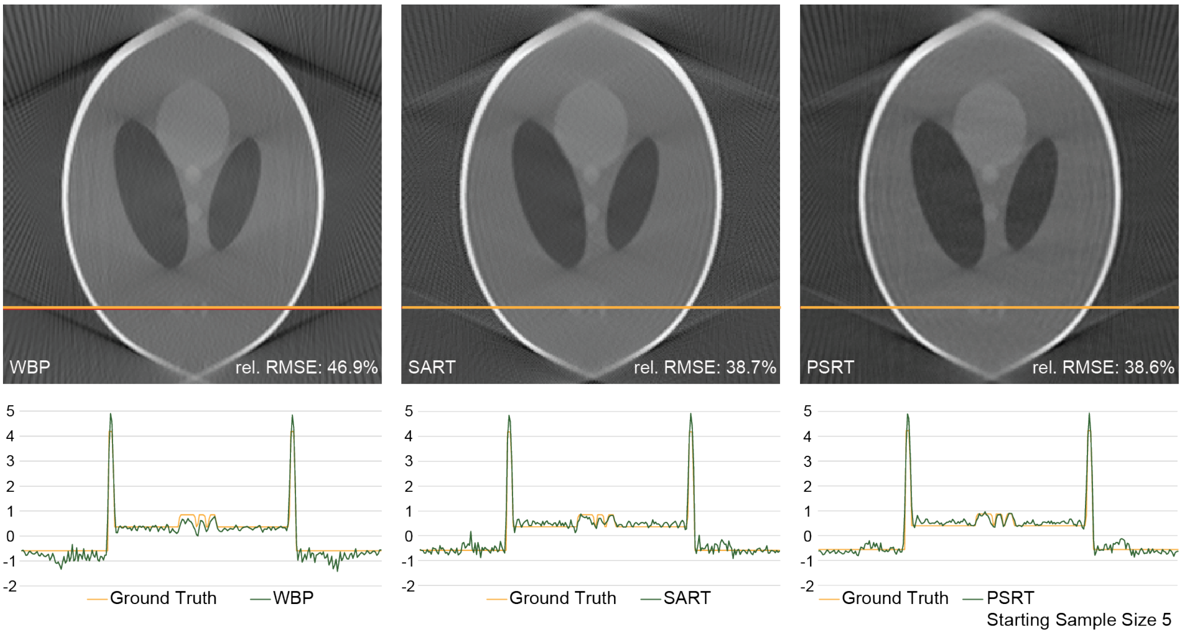


Figure 7.2: 2D reconstruction of size 256×256 of Shepp-Logan Phantom obtained using PSRT and coarse-to-fine scheme and corresponding graphs showing the values along the indicated lines: a) Sample size 1 pixel, relative RMSE 24.4%. b) Starting sample size 5 pixels, relative RMSE 13.9%. c) Starting sample size 9 pixels, relative RMSE 13.9%.


WBP
rel. RMSE: 16.2%
SART
rel. RMSE: 12.0%
PSRT
rel. RMSE: 13.9%
5
4
3
2
1
0
-1
-2
Ground Truth
WBP
Ground Truth
SART
Ground Truth
PSRT
Starting Sample Size 5

(a) (b) (c)

Figure 7.3: 2D reconstruction of size 256×256 of Shepp-Logan Phantom from 160 projections covering range of 360°: a) WBP, relative RMSE 16.2%. b) SART after 20 iterations and relaxation factor 0.5, relative RMSE 12.0%. c) PSRT with starting sample diameter of 5 pixels, relative RMSE 13.9%.


WBP
rel. RMSE: 46.9%
SART
rel. RMSE: 38.7%
PSRT
rel. RMSE: 38.6%
5
4
3
2
1
0
-1
-2
Ground Truth
WBP
Ground Truth
SART
Ground Truth
PSRT
Starting Sample Size 5

(a) (b) (c)

Figure 7.4: 2D reconstruction of size 256×256 of Shepp-Logan Phantom from 65 projections covering range of 127°: a) WBP, relative RMSE 46.9%. b) SART after 20 iterations and relaxation factor 0.5, relative RMSE 38.7%. c) PSRT with starting sample diameter of 5 pixels, relative RMSE 38.6%.

7.1.2 3D Synthetic Dataset

As a ground-truth for 3D synthetic experiments we used a model of the Saccharomyces cerevisiae (yeast) ribosome 80S from EMDB database (EMD-1669). Prior to use, we filtered the the model from 8.8Å down to slightly above 30Å. Besides being a well-known test model system, the experiments performed on the synthetic model provided an optimal parameter setup for the experimental dataset we will discuss in the next chapter, as it contains the same complex. The size of the input projections was 32×32 pixels and we performed reconstruction on a complete projection set consisting of 160 projections as well as on an incomplete projection set with 40 projections and tilt range from −60° to 57°.

Using a setup similar to the 2D synthetic tests (see the second column of Tab. 7.1), i.e. Samples Per Voxel (SPV) 1000, the width for the transition function of the same size as the sample's diameter, and starting diameter 7, we achieved the lowest RRMSE of 9% for the complete projection set and 40% for the missing wedge set. Fig. 7.5 depicts the resulting structure for the complete projection set as well as its evolution with the progression of the algorithm. The resulting structure closely matches the ground-truth, showing that most of the information was recovered during the reconstruction.

Later on we observed that on experimental datasets the setup using samples with diameter 1 introduces excessive noise and that a setup without samples of size 1 is needed, which is given in Tab. 7.1, the third column. To compensate for the absence of samples of size 1, we had to decrease the width of the transition function. This enables the algorithm to recover the high-frequency information by placing the larger samples very densely. Also the SPV had to be changed non-uniformly, putting more emphasis to the smaller samples. The resulting structures seemed to recover the same amount of detail as those obtained with samples of size 1 for both complete and incomplete projection set. However, the RRMSE for the complete projection set increased to 11% while RRMSE for the incomplete projection set decreased to 32%.

We performed an additional test for convergence behavior and stability of the method with respect to the overall complex structure and reported RRMSE. Even a large number of additional samples with

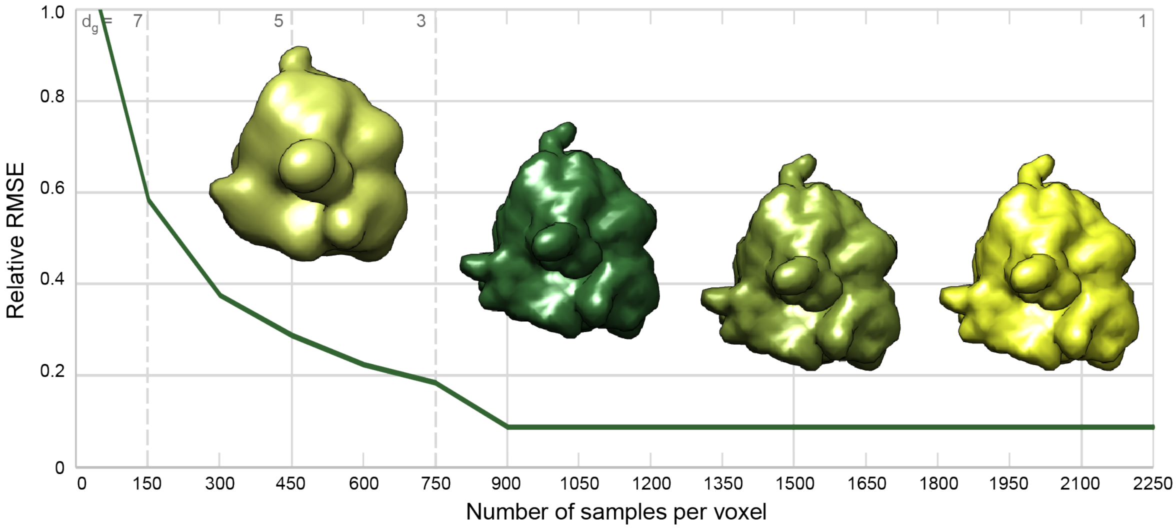


Figure 7.5: The progress of the reconstruction of the synthetic ribosome with starting sample size 7. The minimum RRMSE of 9% was achieved after 1000 SPV and remained stable even with additional 1200 SPV of size 1. The structure of the ribosome was also not altered by the additional samples, showing the stability of the method.

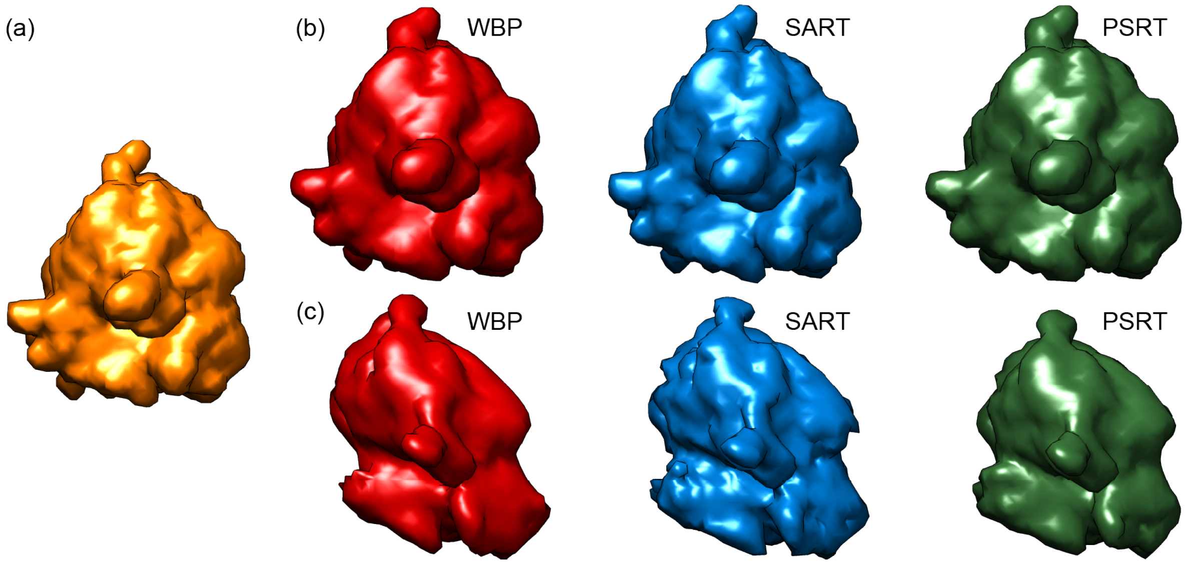


Figure 7.6: Comparison of the synthetic ribosome obtained with WBP, SART, and PSRT. (a) Ground-truth. (b) Structures obtained from the projection set covering whole 360° with RRMSE reaching 12%, 11%, and 9% respectively. (c) Structures obtained from the incomplete projection set with rRMSE of 52%, 44%, and 32%.

diameter 1 did not improve neither the relative error nor the structure of the ribosome (see Fig. 7.5), which shows stability of the method.

The results of the comparison of PSRT to WBP and SART, both for the full projection set and the missing wedge set, are shown in Fig. 7.6. With the full projection set all three methods delivered comparable reconstructions, both in terms of recovered structure and measured error. In the case of the incomplete set, the distortion of the structure caused by the missing wedge was especially strong for WBP, whereas both iterative methods seemed to suppress it reasonably well. This is also reflected by the measured RRMSE values that are 52% for WBP and 44% for SART. Although PSRT reached the lowest error of 32%, it seems that SART managed to recover more details of the structure.

7.1.3 Noise

To better simulate cryoET conditions, we added Gaussian noise to the incomplete projection set of the 80S ribosome and performed tests for the following Signal-to-Noise Ratios (SNR) in input projections: 20, 10, 5, 1, 0.5, 0.1 and 0.05.

For the tests on noisy data we kept the same parameter setup except for the scaling factor α, which had to be adapted for each SNR level, as the added noise changed the range of the values in the input projections. The measured RRMSE for the SNR 20 was approximately 37% and for the SNR 0.05 we achieved a RRMSE of 81%. We compared our results to WBP and SART. As can be seen in Fig. 7.7, PSRT outperformed both WBP and SART by achieving lower RRMSE and higher SNR in tomograms for all noise levels. The recovered ribosome structures of the ribosomes at different noise levels are shown in Fig. 7.8. One might object that the results for SNR 0.05 no longer resemble the structures of the ribosome. It is thus interesting to know that they are very similar to the subtomograms one usually has to process within subtomogram averaging, as the SNR for the experimental datasets is usually similarly low [Chen et al., 2013].

To evaluate the stability under different noise conditions, we performed reconstructions with addi-

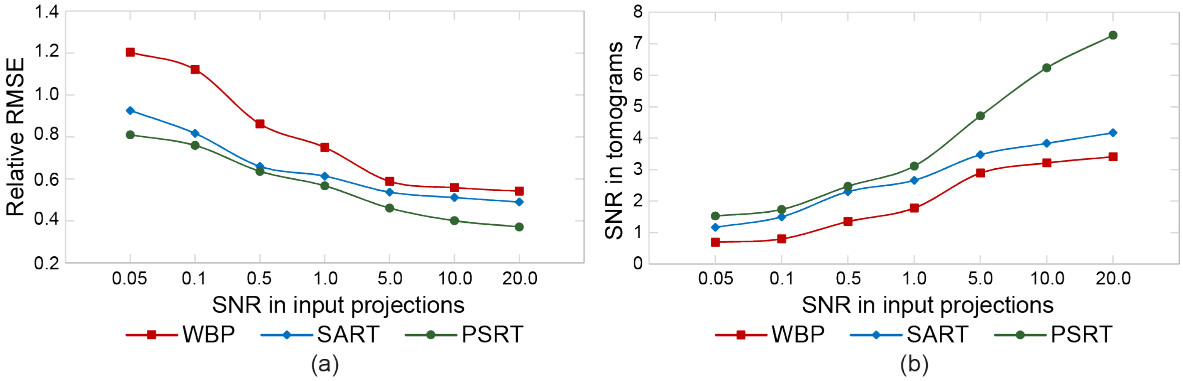


Figure 7.7: Comparison of WBP, SART, and PSRT in connection with different noise levels present in the input projections. (a) Comparison of the relative RMSE. (b) Comparison of noise levels in the reconstructed tomograms.

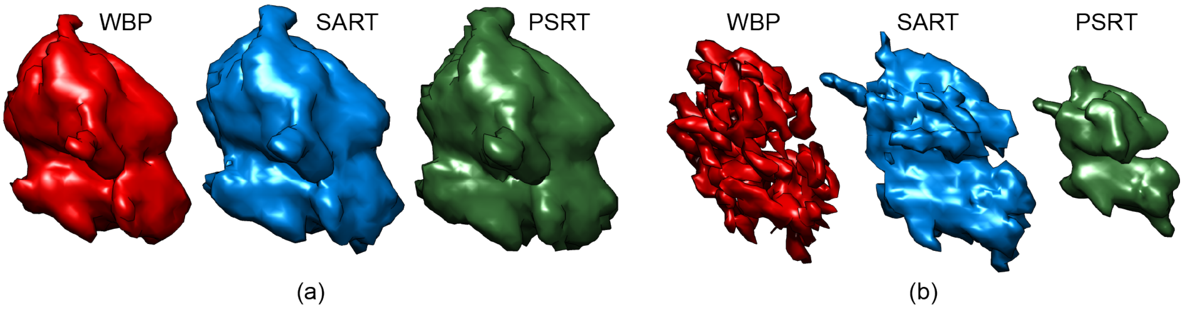


Figure 7.8: Ribosome structures obtained with WBP, SART, and PSRT from the incomplete projection set with SNR (a) 20 and (b) 0.05.

tional 1200 SPV of size 1. While the recovered structure seemed to remain stable, we observed increase in the relative RMSE and decrease in SNR in the final tomograms. This corresponds to the visual quality of the tomograms: Their contrast slightly decreased and they became more noisy. We observed similar behavior on the experimental datasets, where samples with size 1 introduced excessive noise but the resolution remained stable. This leads us to the conclusion that in order to keep higher contrast under noisy conditions the samples of size 1 should be avoided.

7.1.4 Enhanced Contrast for Low-Defocus Data

As mentioned in Sec. 2.2.2, the specimens are typically imaged at higher defocus as it promotes the lower frequencies and thereby the contrast. The data which is more in focus contains better information in terms of high frequencies but the structures are barely visible in the reconstructed tomograms due to very poor contrast.

We performed experiment on data with defocus $-2\mu\text{m}$ to test PSRT performance in case of low-defocus data. The data contains HIV-1 ΔMACANC virus like particles [Gross et al., 2000] and was kindly provided by Florian Schur and John Briggs from EMBL institute in Heidelberg. The data was collected on a Titan KRIOS microscope equipped with a 4k Gatan Quantum K2 direct electron detector at voltage of $300kV$ with defocus $-2\mu\text{m}$, and total dose under $72\,\text{e}^-/\text{Å}^2$. The acquired tilt-series contained 29 projections of size 3710×3710 with tilt angles ranging from $-42°$ to $42°$.

We performed two reconstructions: One of the full size dataset and one on the dataset binned by factor 4 and compared our results to SART and WBP. For SART we tested different number of iterations

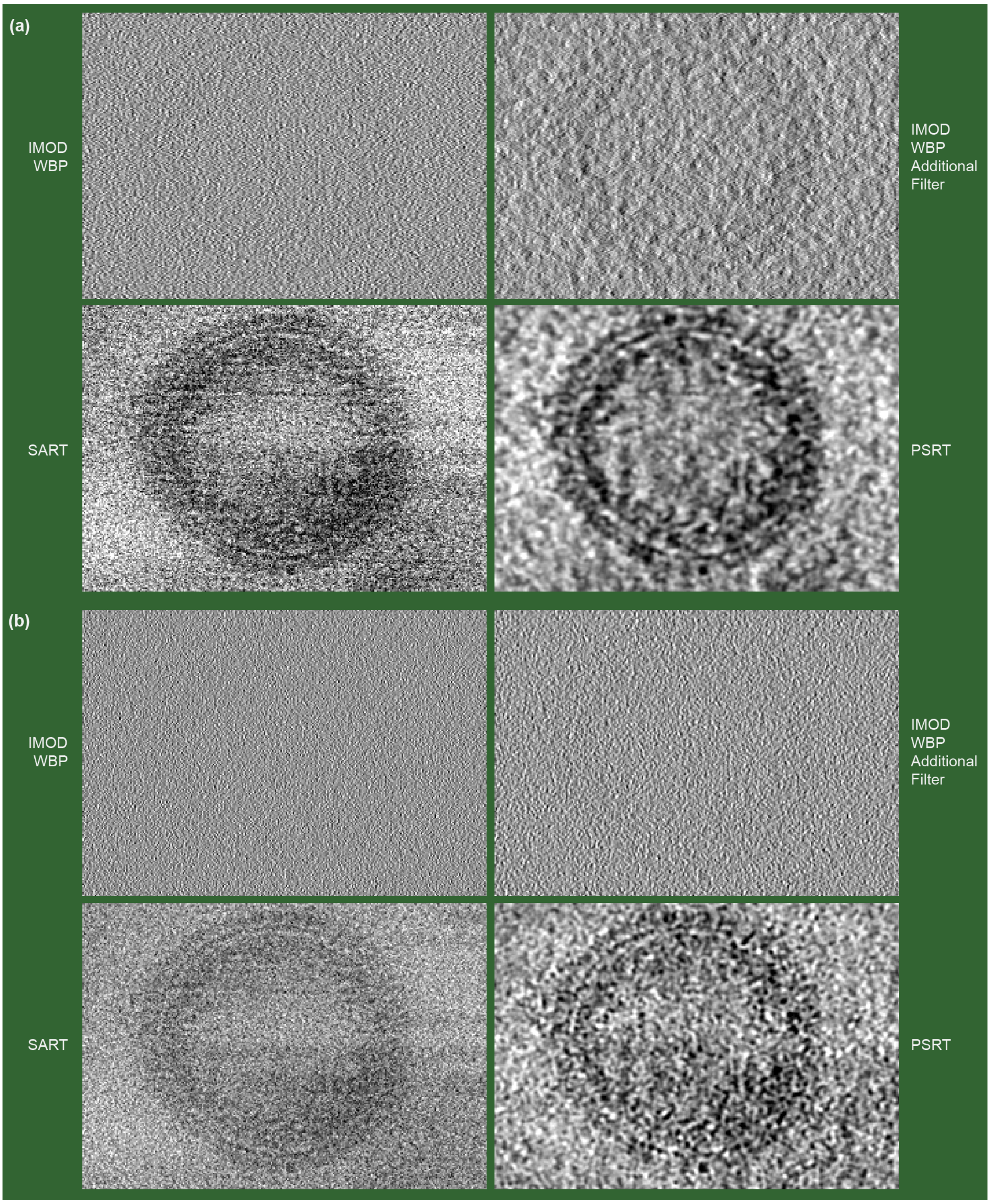

(a)
IMOD WBP
IMOD WBP Additional Filter
SART
PSRT
(b)
IMOD WBP
IMOD WBP Additional Filter
SART
PSRT

Figure 7.9: Comparison of reconstructions obtained with different techniques from low-defocus input projections. Shown is a cut-out from a slice in the xy plane that contains one HIV-1 ΔMACANC virus like particle. (a) Reconstructions from the binned tilt-series with size 928×928. (b) Reconstructions from the full-sized tilt-series with size 3710×3710.

in combination with different relaxation parameters. We obtained the best result in terms of contrast for 1 iteration and $\lambda = 0.05$. Due to rather poor contrast properties of the WBP reconstructions, we decided to use an additional low-pass filter as a preprocessing step. For the PSRT parameter setup see Tab. 7.2. Note that the low-defocus data has poor contrast and thus, in order enhance the contrast in the tomograms, we put more emphasis on larger samples to promote more low-frequencies.

Parameters	Projection size 3710×3710					Projection size 928×928					
SPV	400	400	400	400	400	400	400	400	200	200	200
Transition Width	10.5	9.5	8.5	5.0	4.5	6.5	5.5	4.5	3.5	2.5	1.0
Diameters	21	19	17	15	13	13	11	9	7	5	3
Scaling Factor	60000					7200					

Table 7.2: Parameter setup for low-defocus data.

The results are shown in Fig. 7.9. Despite the additional filtering, the contrast in WBP reconstructions is very poor and in the full-sized tomograms the particles are essentially not visible. The contrast in the reconstructions obtained with SART and PSRT is significantly better. The coarse noise in the PSRT tomograms is caused by the relatively large stopping diameter. Despite that, it seems that PSRT recovered more details from the internal structure, however this is hard to judge, as we do not have any quality metric for this dataset. Nevertheless, the achieved contrast shows that PSRT is able to enhance contrast also for the low-defocus data.

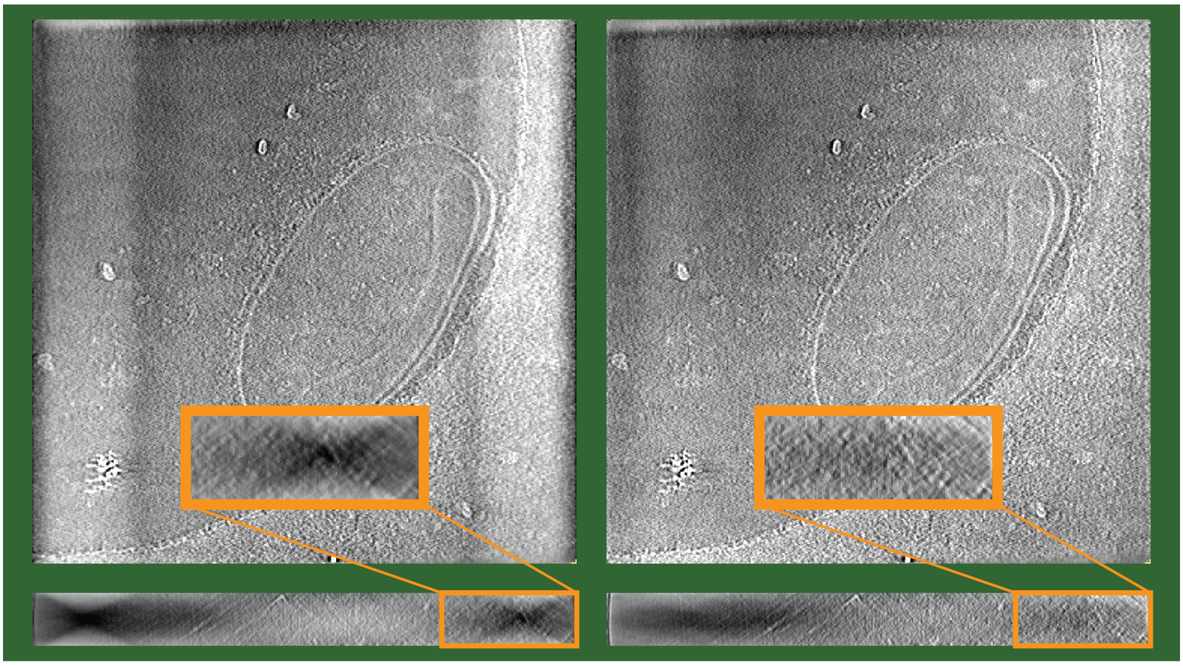
Currently, the low-defocus data is acquired mostly to provide missing frequencies for CTF correction (see Sec. 2.3.2 for more details) and is typically not used for data analysis as the contrast is too low. We believe that the promising contrast we obtained with PSRT could be used to perform segmentation or Subtomogram Averaging (SA) even on low-defocus data, although additional experiments are necessary to confirm that.

7.2 Memory Efficient Solution to Interior Problem

We focus on the interior problem in more detail in Ch. 10 where we also discuss the artifacts associated with it and possible solutions as well as review the results obtained with different reconstruction techniques. However, for a sake of completeness, we present here the results for PSRT reconstruction on experimental dataset showing the artifacts caused by the interior problem and their suppression using the suggested Memory Efficient Solution to Interior Problem (MESIP). The experimental dataset was obtained from EMPIAR databse (DOI: 10.6019/EMPIAR-10009) and consists of 51 projections from $-60.39°$ to $60.97°$. We aligned the tilt-series and downsampled it to size 512×512. The reconstructed tomogram had dimensions 512×512×50. Based on that, the extension for x dimension was computed to be 692. Both tomograms are shown in Fig. 7.10. The MESIP clearly suppressed the artifacts associated with the interior problem.

7.3 Region of Interest Reconstruction

We performed set of experiments to show the performance of different approaches towards Region Of Interest (ROI) reconstruction. The test object was 3D phantom containing nine spheres with different intensities (see Fig. 7.11a) and we reconstructed the whole tomogram as well as three center spheres separately. To be able to reconstruct ROI with WBP and SART we had to create a tilt-series for each sphere by cutting out the respective parts from the original input projections. As described



(a) No MESIP

(b) MESIP

Figure 7.10: Suppresion of the interior problem artifacts using the proposed PSRT method. In the PSRT reconstruction without MESIP (a) we can observe strong artifacts on the edges of the volume both from the front and the side view. These artifacts are completely suppressed if MESIP is used (b). Note that the dark regions in the left parts of the bottom images are caused by the missing wedge problem and the high intensities resulting from the interior problem only enhance these regions even more as can be seen in (a).

in Sec. 5.4.1, to avoid high intensity artifacts on the edges of ROIs, we have to either weight the cut-out projections or cut out and reconstruct larger regions to prevent the artifacts from interfering with the objects of interest. We chose the latter approach and reconstructed larger regions to prevent the artifacts from interfering with the spheres (i.e. for the spheres with radius 5 voxels, we cut out projections of size 20×20 and reconstructed cubic volumes of corresponding dimensions). For PSRT we first reconstructed one tomogram with 3 ROIs and then we reconstructed each of them separately, i.e. we performed three reconstructions, each with one ROI (see Fig. 7.11b). In both cases was the starting diameter for the ROI reconstruction 3 (for the complete parameter setup of all PSRT reconstructions see Tab. 7.3).

The comparison of the line profiles between the ground-truth, whole tomogram, and ROI reconstructions can be found in Fig. 7.11c. For both WBP and SART the ROI reconstructions were significantly worse in terms of recovered values than the whole-volume reconstructions. Moreover, the differences between the intensities of individual spheres decreased, making them less distinguishable. The PSRT ROI reconstructions have significantly smaller differences w.r.t. the whole tomogram reconstruction, with essentially no difference between the tomogram with 3 ROIs and the tomograms with single ROI. We also tested the stability of our method in case of ROI reconstruction – we added additional 1000 SPV of size 1 to our ROIs. As can be seen on the graphs, the values remains practically unchanged, showing the stability of PSRT even in case of the local reconstructions.

Our results confirm that the local character of PSRT can be efficiently used for ROI reconstructions and that it outperforms both WBP and SART in terms of recovering values of the structures within the ROIs.

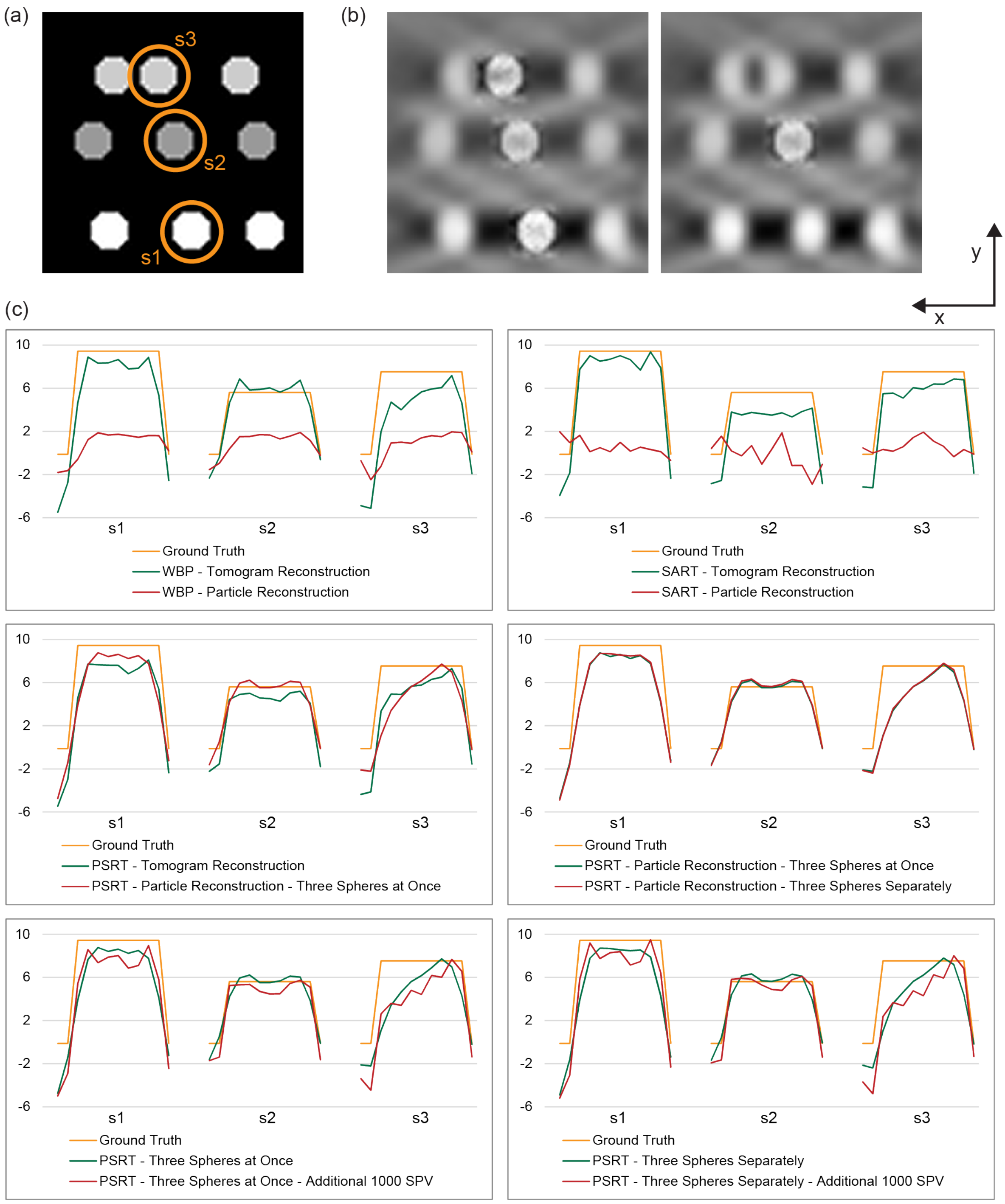


Figure 7.11: (a) A slice in xy from the ground-truth containing nine spheres of radius 5 and labels for the three reconstructed spheres. (b) PSRT ROI reconstructions with three spheres (left) and one sphere (right). (c) Graphs with line profiles through the centers of the spheres (in x direction) for all performed reconstructions. The diameter of the spheres was 10 voxels.

Parameters	Whole Tomogram				ROI Reconstruction (starting at diameter 3)				
SPV	400	400	400	400	400	400	**400**	**400**	(1400 for stability test)
Transition Width	7.0	5.0	3.0	1.0	7.0	5.0	**0.5**	**0.1**	
Diameters	7	5	3	1	7	5	**3**	**1**	
Scaling Factor	3000				3000				

Table 7.3: Parameter setup for the whole-tomogram and for the ROI reconstructions of the spheres.

7.3.1 ROI Using a Binary Mask

In Sec. 6.6, we discussed the possibility of specifying a ROI by a binary mask. We created such a mask for the dataset described in Sec. 7.2 and performed two ROI reconstructions. We used the same setup for both tomograms, only the transition width for ROI was halved for the second tomogram (see Tab. 7.4). The impact of this change can be seen in Fig. 7.12c and Fig. 7.12d). While for the greater transition width, the algorithm reconstructed area also in the proximity of the mask, the random walks with the smaller transition width recovered only the regions corresponding to the mask. In both cases, the ROI reconstructions correctly recovered the structure indicated by the mask without introducing any artifacts into the ROIs. Interesting insight on the amount of recovered information provide the residuals. Despite the initial reconstruction, the structure was still present in the corresponding residuals. The subsequent ROI reconstruction recovered even the high-frequency details, making the structure “invisible” in the residuals (see Fig. 7.12e and Fig. 7.12f). We thus conclude that the masking approach is suitable for ROI reconstructions.

Parameters	Initial Reconstruction				Importance Sampling 1			Importance Sampling 2		
SPV	160	160	160	160	200	200	50	200	200	50
Transition Width	6.5	5.5	4.5	3.5	**2.0**	**1.0**	**1.0**	**1.0**	**0.5**	**0.5**
Diameters	13	11	9	7	5	3	1	5	3	1
Scaling Factor	3000									

Table 7.4: Parameter setup for ROI reconstructions using a binary mask. The initial phase was the same for both reconstructions. The importance sampling started at diameter 5.

7.4 Memory Requirements

Similar to the state-of-the-art methods, we keep the residuals and the voxel representation of the volume (both stored in floats) in memory. In case that the volume does not fit into memory, we have to split the input projections into smaller parts and perform a set of reconstructions of smaller volumes (as is the case also for the state-of-the-art techniques). Due to the parallel image acquisition scheme, we can easily split the projections along the y-axis. While this approach works for other techniques, in PSRT we might observe discontinuities on the edges between neigbouring tomograms due to the use of bigger samples (i.e. at the early stages of the iteration process). As this concerns also possible parallelization schemes for PSRT, we will discuss this in more detail in Sec. 9.1.

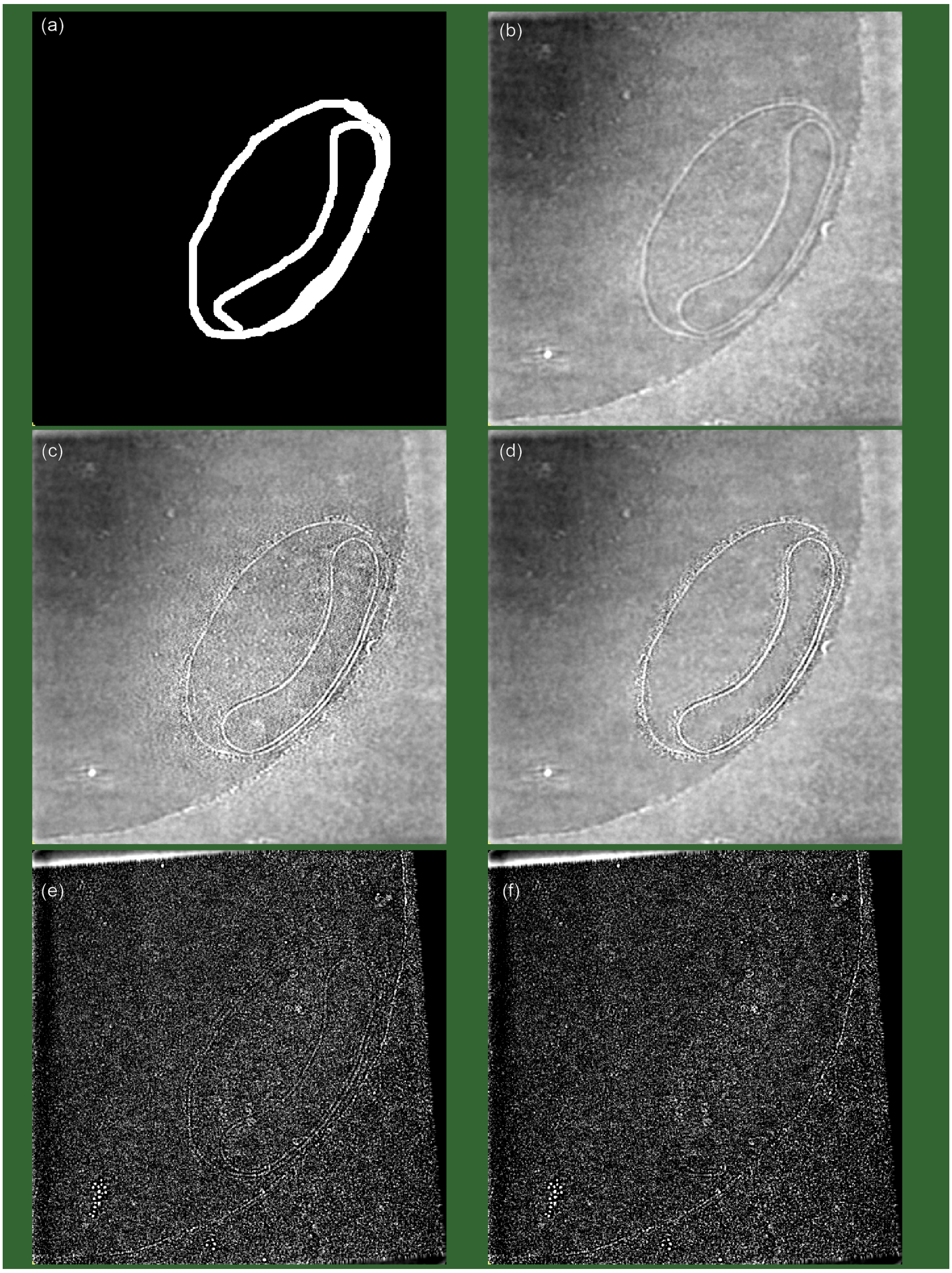


Figure 7.12: (a) Slice of the binary mask used for Importance Sampling (IS). (b) The corresponding slice of the initial reconstruction (same for both IS schemes). (c) Final reconstruction with the first IS scheme. (d) Final reconstruction with the second IS scheme. (e) One of the residual images that correspond to the initial reconstruction. (f) The same residual image after the importance sampling with the second scheme. The structure of the object of interest is hardly visible which shows that the most of the structural information was recovered during the IS reconstruction. Note that for illustration purposes we enhanced the contrast of the residual images, as they were to noisy to be easily interpreted otherwise.

7.5 Computation Time

To test the performance of PSRT, we performed series of reconstructions on the experimental dataset described in Sec. 7.2. We used five different parameter setups to reconstruct tomograms with dimensions $512\times512\times50$ and $1024\times1024\times100$. The reconstructions of the whole tomogram are denoted as PSRT and PSRT MESIP. We denote as IS1 and IS2 the importance sampling schemes from Sec. 7.3.1. Finally, the reconstruction of the whole tomogram without the use of samples with the diameter 1 is denoted as NS1. The complete parameter setup can be found in Tab. 7.5 for the whole-tomogram reconstructions and in Tab. 7.4 for the importance sampling reconstructions. We compared our results to non-parallel version of IMOD WBP and non-parallel version of Ettention SART (running 5 iterations on the CPU). All reconstructions were carried out on Intel(R) Core(TM) i7-3770K CPU (3.50GHz) with 32.0GB RAM. The results are shown in Tab. 7.6.

Parameters	PSRT							PSRT NS1					
SPV	160	160	160	160	200	200	50	160	160	160	160	200	300
Transition Width	6.5	5.5	4.5	3.5	2.5	1.5	1.0	6.5	5.5	4.5	3.5	2.5	0.5
Diameters	13	11	9	7	5	3	1	13	11	9	7	5	3
Scaling Factor	3000							3000					

Table 7.5: Parameter setup for whole-tomogram reconstructions. The parameter setup for PSRT with MESIP is identical to the one in the second column, as the necessary amount of seeds and length of random walks is computed automatically.

Running time of the basic non-progressive algorithm on the experimental data with tomogram dimensions of $1024\times1024\times100$ and sample size 1 would be close to two months. Introducing the coarse-to-fine scheme speeds up the whole process by a factor of 60. Despite that, PSRT is still significantly slower than WBP and SART as can be seen in Tab. 7.6. According to our expectations, the MESIP needs more computational time with the overhead being around 25%. As can be seen in Fig. 7.13, the processing time increases with decreasing sample size, making the use of samples with size 1 very expensive. Avoiding them using more samples with size 3 and smaller transitions width proved very effective as well as using the importance sampling. The worse time for IS1 setup shows that the transition width has a significant impact on the reconstruction speed of ROIs. Note that using IS is not always possible and that the speed-up we can gain depends on the size of ROIs and the chosen parameter setup (i.e. at which sample size we start to perform IS reconstruction).

PSRT is a Monte Carlo approach and thus its computational time is admittedly worse than by most of the state-of-the art methods. In cryoET, the typical size of the projections lies between $2k\times2k$ and $4k\times4k$. Given the computational time necessary for obtaining reasonable reconstruction from $1k\times1k$ data, running a reconstruction on 4k dataset, even with the employment of IS, would be still inconvenient, making parallelization of PSRT a necessity for real world applications. However, the character of PSRT makes any parallelization a challenging task, as we will see in Ch. 9.

Dataset Size	WBP IMOD	SART Ettention	PSRT	PSRT MESIP	PSRT IS1 MESIP	PSRT IS2 MESIP	PSRT NS1 MESIP
512×512×50	1.89s	72.5s	4h 9m 9s	5h 13m 39s	1h 49m 03s	53m 15s	2h 38m 07s
1024×1024×100	9.51s	8m 24.14s	33h 41m 27s	42h 2m 33s	10h 02m 20s	6h 53m 19s	22h 06m 37s

Table 7.6: Computational times for different techniques on the experimental dataset. Note that both PSRT reconstructions with IS as well as NS1 were performed using MESIP as well. All reconstructions were carried out on Intel(R) Core(TM) i7-3770K CPU (3.50GHz) with 32.0GB RAM.

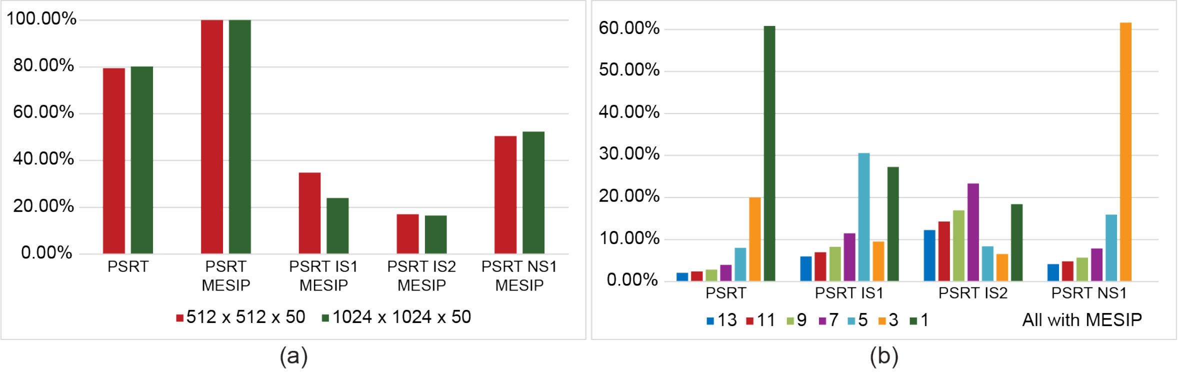


Figure 7.13: (a) Graphs showing reconstruction times for both datasets using different approaches within PSRT. (b) The distribution of a reconstruction time among different sample sizes.

7.6 Parameter Setup

Prior a reconstruction, we have to specify four parameters: The starting diameter of a sample D_{start}, the total number of samples per voxel V, the width of the transition function σ, and the scaling factor for the initial energy computation α. In this section we suggest their possible setup and review their influence on the reconstruction quality as well as on the convergence speed.

Starting Diameter

The starting diameter should be chosen based on the noise properties of the input projections and the size of the reconstructed volume. Starting with larger diameter speeds up the convergence and reduces noise. If we want to localize the Regions Of Interest (ROIs) in the reconstructed tomograms and subsequently perform Importance Sampling (IS) on them we should set the starting diameter to be approximately of the same size as the smallest features we want to recover. In that way we can quickly obtain the initial reconstruction with rough structures of the objects of interest and start the IS to recover the finer details. Starting with large diameter would in this case prolong the initial phase and thereby the overall reconstruction time as it would require more iterations to reveal the ROIs.

Samples per Voxel

We found empirically that 1000 – 1500 Samples Per Voxel (SPV) are sufficient to obtain the reconstruction with reasonable quality. If the uniform distribution of the total number of samples over the number of iterations is not suitable, we can specify SPV for every iteration separately. In case of importance sampling, we set less SPV per iteration for the initial phase and increase SPV for the refinement phase which speeds up the reconstruction.

Transition Width

The width of the transition function should be chosen with respect to the sample's diameter and properties of the input projections, e.g. SNR or heterogeneity/smoothness. Tests on datasets from cryoET showed that a width around 50% of the sample's diameter is optimal for the initial reconstruction.

Setting it too small (i.e. around 1 for samples larger than 5) increases the acceptance probability but suppresses the advantages of using bigger samples and thus effectively slows down the convergence, as the covered region is very small. On the other hand, setting the width equal to the diameter leads to higher rejection rates as the differences between subsequent ΔF might be too high for highly heterogeneous data. However, for piecewise homogeneous volumes (such as phantoms) we can set the width to be as wide as the sample's diameter, since the acceptance probability does not decrease with distance so rapidly in the homogeneous parts. For ROI reconstructions the transition width should be small to keep the random walks within the ROIs. For less accurate masks or bounding boxes we should set the transition width to a large enough value that would allow us to recover also the area in the vicinity of a given ROI.

Energy Scaling Factor

The energy scaling factor should compensate for rejected samples and samples accepted with negative energy. As those are hard to predict, it is difficult to choose an optimal value for the scaling factor. This is further complicated by the fact that it is data dependent, as the starting energy is based also on the average intensity of the input projections (see Eq. 5.2). In other words, while there is a logic behind setting the first three parameters, the scaling factor has to be guessed. Currently, we choose the factor based on the mean value of the input projections and the starting diameter and then manually adjust the factor based on the result of the first iteration. It is obvious that this approach is far from optimal and finding a better way of setting the energy scaling factor is necessary.

7.6.1 Automatization

We believe, it is possible to compute the first three parameters automatically based on additional input from a user. For instance, if we know the size of the smallest feature, that should be recovered, the estimate of the noise level in the input projections (i.e. on the scale from 1 to 5), defocus, and the accuracy of the ROI specification, we can compute optimal values for all three parameters automatically. In essence, that is the approach we currently use to set up the parameters manually. What remains as an open issue is the scaling factor. Although some experiments suggest that a standard deviation of the input projections could be used to compute at least the order of the factor, more experiments are necessary to design a reliable derivation of the scaling factor.

CHAPTER 8

INTEGRATION TO SUBTOMOGRAM AVERAGING PIPELINE

Subtomogram Averaging (SA) is a novel protocol for obtaining high-resolution macromolecular structures from cryoET tomograms. It is based on averaging of subtomograms that contain structures at the same conformation state. The averaging not only leads to enhanced contrast and reduced noise but also removes the distortions caused by the missing wedge problem, as it combines information from structures with different orientations. The achievable resolution of the final structure as well as the reliability of the localization and alignment strongly depend on the quality of the subtomograms.

In this chapter we demonstrate the integration of PSRT into the standard SA pipeline. We show how the enhanced contrast delivered by PSRT improves the performance of template matching-based localization compared to both WBP and SART. The locality of PSRT and Region Of Interest (ROI) nature of SA are used together as an efficient importance sampling scheme to significantly speed up the reconstruction of subtomograms. Finally, we validate on experimental cryoET data that PSRT in combination with SA leads to correct reconstruction of high-resolution structures of macromolecular complexes.

8.1 Subtomogram Averaging Pipeline

The SA protocol consists of the localization of the instances, their classification, alignment, and averaging (see Fig. 8.1). The localization and subsequent classification produce positions and rough orientations of instances with the same conformation that should be aligned and averaged together. Typically, the whole SA pipeline is performed on the same volume. This however has several major disadvantages. First, the complete volume is reconstructed to the same level-of-detail (disregarding the level-of-detail anisotropy caused by the missing wedge), while only a small fraction of it in the end contributes to the resulting structure. Second, the properties of the reconstructed volume that are necessary for reliable localization do not necessarily correspond to the properties needed for the final high-resolution structures. In PSRT, we split the SA pipeline in two distinct steps. We first perform an initial reconstruction using only larger samples. This high-contrast and very smooth volume is then used to perform the initial localization step. From this, we take only the positions of the particles and

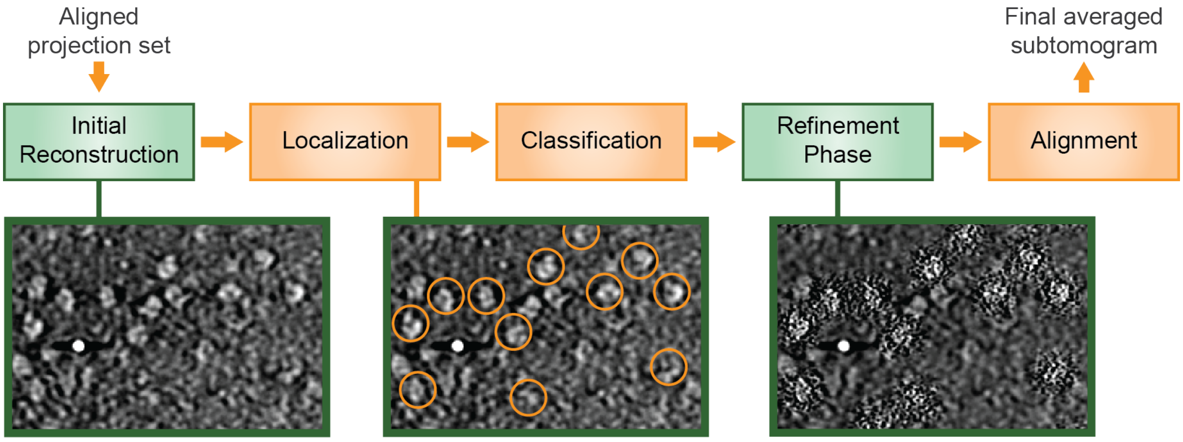


Figure 8.1: Scheme of the proposed PSRT-SA pipeline. First few iterations of PSRT are performed to obtain high contrast but blurry reconstruction on which the localization of the structures is made. Found positions are used for additional importance sampling for subsequent PSRT iterations, during which only the regions we want to focus on are further refined. The final refined reconstruction is used for SA alignment and averaging.

use them as importance sampling information to drive the refinement phase of PSRT reconstruction. Once the reconstruction is finished, we cut out the particles and perform alignment and averaging on them to obtain the final high-resolution structure.

8.2 Experimental Data

To evaluate the performance of PSRT on experimental data, we reconstructed a macromolecular complex with a well-known structure, the 80S Saccharomyces cerevisiae (yeast) ribosome. The dataset [Chen et al., 2013] was kindly provided by Dr. Friedrich Förster from the Max Planck Institute for Biochemistry in Martinsried. The sample was prepared as detailed in [Brodsky, 2010] and the tilt-series was collected on a Titan KRIOS microscope equipped with a 4k Falcon I direct electron detector at a voltage of $300kV$ with a defocus of $-4\mu m$, pixel size of 2.88 Å, and total dose under $60\,e^{-}/Å^{2}$. The acquired tilt-series contained 40 projections with equidistant tilt angles from −60° to 57°. To obtain the reconstruction in reasonable time, we downsampled the tilt-series to the size of 1024×1024 with pixel size 11.52 Å. Our goal was to demonstrate the ability of the algorithm to reconstruct a correct biological structure, not to achieve the highest possible resolution.

8.3 Reconstruction

The reconstruction was performed using the combined PSRT-SA protocol described in Sec. 8.1. As a baseline, we used the parameters from the 3D synthetic reconstruction (see Tab. 7.1, third column). However, we increased the starting diameter to 13 to fill in the larger volume faster. This also resulted in an increase of SPV from 1250 to 1900 and also in an increase of the scaling factor by factor 1250. Tab. 8.1 summarizes all the parameter values. The initial reconstruction was performed on the samples with diameters from 13 to 7 voxels. We localized 1000 possible candidates within the initial tomogram from which we chose a reliable subset for further processing. We then used their positions to perform ROI reconstruction to obtain the final refined tomogram. The size of each ROI was 36^2 and

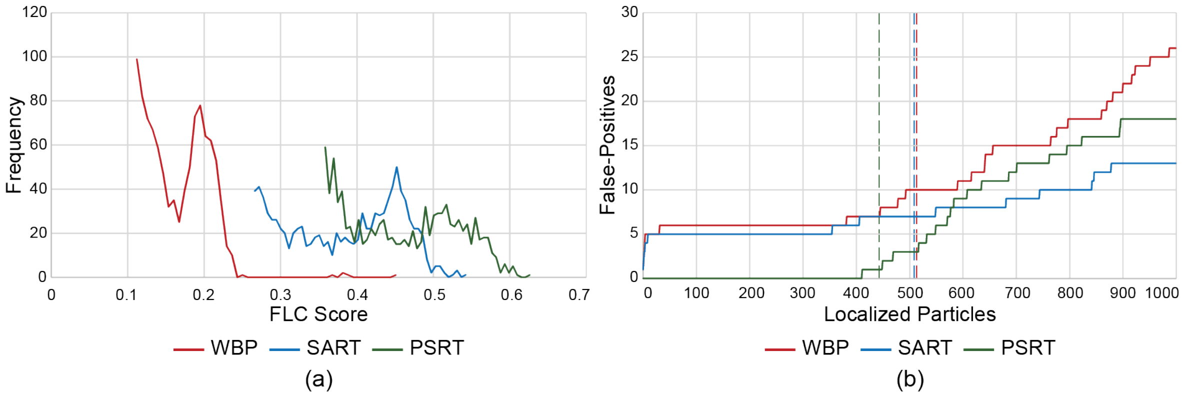


Figure 8.2: (a) Histograms of localization scores of 1000 candidates found in WBP, SART (with relaxation parameter 0.1 and 1 iteration), and PSRT reconstructions. The FLC score is low for WBP with values mostly around 0.15 – 0.25, the small amount of structures with values near 0.4 are gold markers. For SART most values are between 0.4 – 0.5, with values above 0.5 containing also gold markers. For PSRT reconstruction, more than one third of structures has FLC score over 0.5 and all of them are ribosomes. (b) Graph showing the increase of false-positive candidates with the decreasing cross-correlation score for all three methods. The dashed lines indicate the number of particles that were suggested for further processing by Gaussian fitting with threshold σ. There was only 1 false-positive result with relatively low FLC score within the suggested amount of particles for PSRT.

was determined based on the ribosome size. Finally, we cut out the particles from the refined regions and performed final alignment and averaging on them.

The localization, aligning, and averaging were carried out using the PyTom toolbox [Hrabe et al., 2012]. The localization was done by a template-based matching in real space as described in [Förster et al., 2010]. As the template we again used the model of 80S ribosome obtained from EMDB database (EMD-1669) downsampled to roughly 35Å. The alignment was performed by Fast Rotational Matching (FRM) described in [Chen et al., 2013]. The Spider package [Frank et al., 1981] was used to compute Fourier Shell Correlation (FSC) curves. We compared our results to IMOD WBP [Kremer et al., 1996] and SART from the Ettention package [Dahmen et al., 2016]. Similar to the previous chapter, we performed multiple SART reconstructions with different relaxation factors (0.05, 0.3, 0.5, 1.0) and different number of iterations (1, 10, 15, 20, 25, 30) and chose the one that provided the best result in terms of achieved resolution.

Parameters	Initial Tomogram				Importance Sampling		Stability Test
SPV	150	150	150	150	300	1000	1000
Transition Width	6.5	5.5	4.5	3.5	1.0	0.5	0.5
Diameters	13	11	9	7	5	3	1
Scaling Factor	5000000						

Table 8.1: Parameter setup for PSRT reconstructions.

8.4 Localization

The localization result should contain list of particles' positions together with rough rotations sorted according to decreasing cross-correlation score with the template. However, due to noise and low contrast, gold markers or contamination present in the specimen are usually found as well. They have

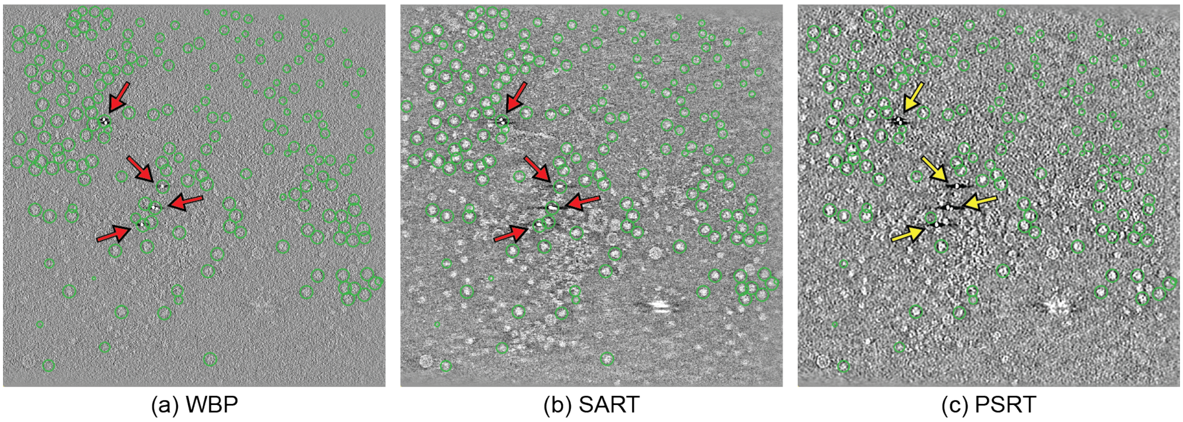

(a) WBP (b) SART (c) PSRT

Figure 8.3: Comparison of localization of (a) 509 particles for WBP, (b) 507 for SART (with relaxation parameter 0.1 and 1 iteration), and (c) 447 for PSRT. As indicated by the red arrows, the localization made on WBP and SART included also gold markers falsely marked as ribosomes. The yellow arrows show that in PSRT none of the gold markers present in this slice were falsely interpreted. While in WBP and SART all gold beads present in the tomogram were falsely marked as ribosomes, in PSRT there was only one false-positive candidate within the 447 particles (not visible this slice).

to be removed from the list manually and thus prevent a fully automatic execution of the SA pipeline. To reduce the number of false-positive candidates is thus of the utmost importance.

The number of particles that should be used for further alignment is usually determined by a histogram with cross-correlation scores (see Fig. 8.2a). One fits a Gaussian, that should indicate the distribution of true-positives, to the tail of each curve and chooses the cut-off threshold [Hrabe et al., 2012]. Typical threshold value corresponds to a standard deviation σ of the Gaussian. Our localization results yields 509 particles for WBP, 507 for SART and 447 for PSRT. As shown in Fig. 8.2b, the suggested amounts of particles for both WBP and SART contain also number of false-positive candidates (mostly consisting of gold markers and their ghosts, see Fig. 8.3), while for PSRT there is only 1 candidate that should be excluded from further processing. Note that by setting a slightly higher cut-off threshold we would have zero false-positive candidates for PSRT, while for both SART and WBP the number of false-positives would remain practically the same.

More suitable properties of the PSRT reconstruction for automatic localization and analysis of the tomograms are indicated also by the ribosomes' profiles shown in Fig. 8.4. Despite the good contrast of the SART reconstruction, the ribosomes are still hard to distinguish from their surroundings automatically. In the PSRT reconstruction, on the other hand, the ribosomes' profiles are clearly visible as the level of noise is low.

Note that while the best ribosome resolution using SART combined with SA was achieved with relaxation parameter 0.3 and 20 iterations, the best properties w.r.t. localization were obtained using relaxation parameter 0.1 with only 1 iteration. Therefore, similar to PSRT, we used one SART tomogram for localization and a different one for the final alignment to obtain the best results. Yet, PSRT provided much better localization properties on the smoothed intermediate volume.

8.5 Final Alignment

We obtained a correct structure of the 80S ribosome with a resolution of 33.8Å, as given by the FSC 0.5 gold-standard criterion. We did not perform any classification and determined only one average structure. Comparing the resulting structure to the single particle reference (EMD-1669), we conclude

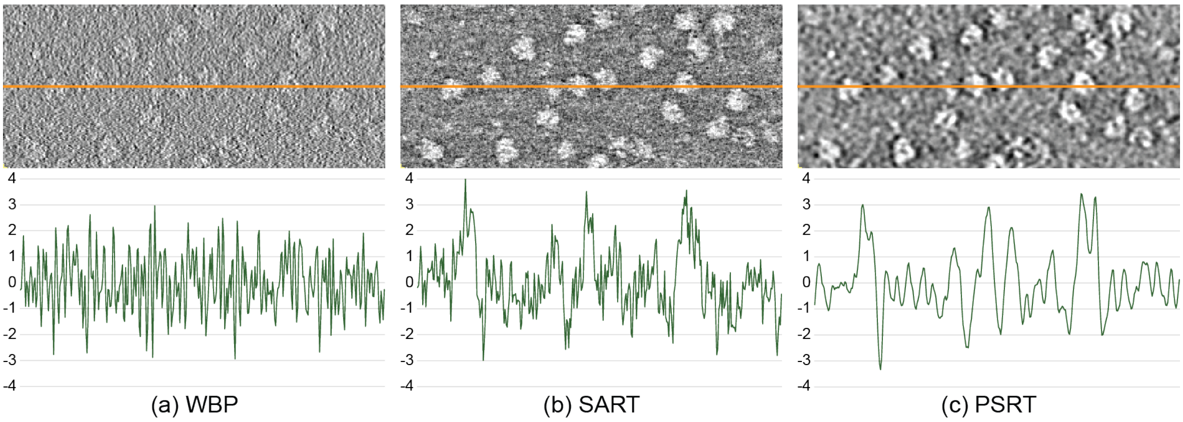


Figure 8.4: Comparison of the reconstruction contrast in (a) WBP, (b) SART (with relaxation parameter 0.1 and 1 iteration), and (c) PSRT reconstructions. The corresponding graphs show reconstructions' profiles on the indicated lines. The WBP provides very low-contrast and noisy reconstruction, complicating the localization of the particles. In the SART reconstruction, the contrast is better, however the particles' profiles are still hard to distinguish from their surroundings. The particles are clearly visible on the graph corresponding to the PSRT reconstruction, indicating more suitable properties for an automated localization.

that all the relevant features as appropriate for this resolution level were correctly reconstructed. The final structure as well as its evolution with the algorithm progression is shown in Fig. 8.5. A snapshot of one of the slices from the final volume is shown in Fig. 8.1.

Next, we ran the standard SA protocol on the reconstructions obtained with WBP and SART (with 20 iterations and relaxation parameter 0.3). All the SA parameters were identical to the PSRT protocol, except that the localization for WBP was performed in the standard way on the final volume. The resulting resolution of WBP reconstruction was 33.8 Å and for SART 34.0 Å. The results with the corresponding FSC graphs are shown in Fig. 8.6.

For sake of completeness, we also performed alignment on WBP and SART using the 446 positions and rotations found by localization on PSRT. While for WBP the obtained resolution was similar, around 33.9 Å, it got slightly worse for SART, yielding a structure with resolution of 34.5 Å.

8.5.1 Stability

As mentioned early, we observed that on real datasets the setup using samples with diameter 1 introduces excessive noise and thus the smallest diameter we used during the refinement reconstruction was of size 3. To show that the samples with size 1 are indeed not needed if the setup of transition widths is optimal, we ran an importance sampling with additional 1000 SPV with diameter 1. As can be seen in Fig. 8.7, we did not gain any improvement (or distortion) of the structure even with the additional samples. This was also confirmed by the measured resolution which worsened at first about 0.2 Å but then remained stable as well. This experiment thus verifies the stability of the method w.r.t. additional samples under the noisy experimental conditions.

8.6 Reconstruction on 4k Data

Using the parallelization scheme, which will be described in the next chapter, we were able to perform a reconstruction on the full-sized 4k projections. We had to adapt the parameter setup to cope with

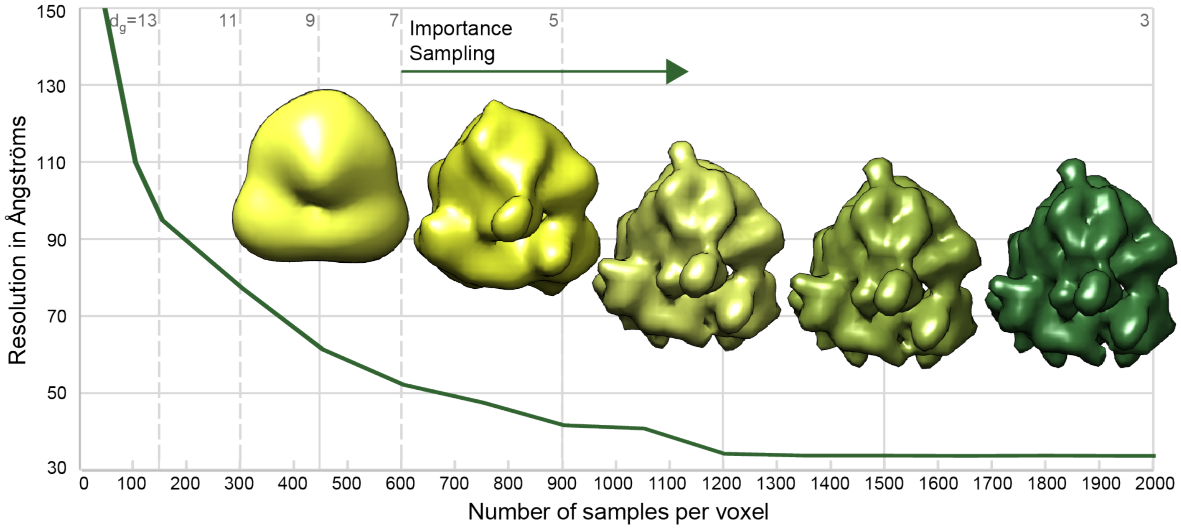
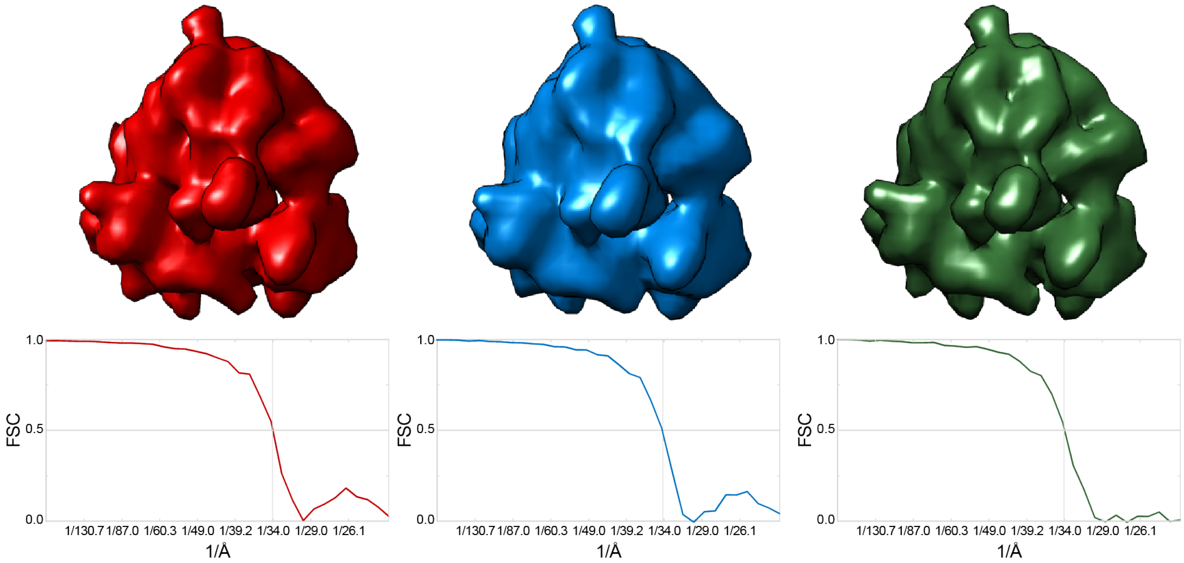


Figure 8.5: Progress of the reconstruction of the proposed PSRT-SA protocol on the experimental dataset. As expected, the structure of the complex gradually and monotonically improves as the samples with smaller sizes are processed, i.e. in line with the improvement to the objective function. The structures shown were obtained by subtomogram averaging using FRM alignment implemented within the PyTom toolbox. For all of them we used the same set of 446 ribosomes obtained by localization after samples with diameter $d_g = 7$ (marked by the Importance Sampling arrow).



(a) WBP (b) SART (c) PSRT

Figure 8.6: Reconstruction results of PSRT-SA compared to WBP and SART. In the top are the reconstructed structures, in the bottom their corresponding FSC graphs. There are no visible differences in the structure of the ribosomes. This is consistent with the achieved resolution: 33.8Å for WBP, 34.0Å for SART, and 33.8Å for PSRT.

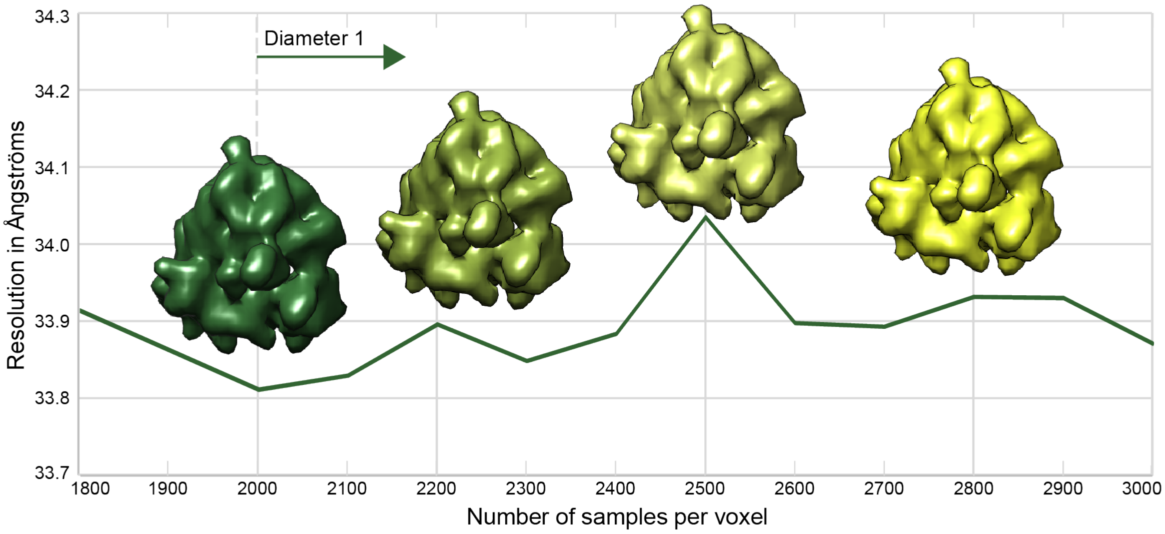


Figure 8.7: Stability of the ribosome structure and resolution with additional iterations of the PSRT-SA protocol. The additional 1000 SPV with diameter 1 did not change the structure of the ribosome and the resolution remained stable as well. The high peak in the graph corresponds to a decrease of only 0.2 Å.

larger dimensions: We initiated the reconstruction with samples of size 21 and started the importance sampling already at diameter 15. The complete set of parameters can be found in Tab. 8.2.

In a traditional SA pipeline, the automatic localization is never done on full 4k reconstruction. There are two reasons for that. First, the computational time necessary to compute cross-correlation between the template and each voxel in the volume would be enormous. Secondly, a downsampling of tomograms removes a high-frequency noise that prevents a reliable template-based matching. Therefore, the tomograms are typically downsampled 4 or even 8 times for the purposes of localization. This reduces not only the processing time but also enhances low frequencies in the downsampled tomograms. The positions obtained during the localization have to be multiplied by the downsampling factor, while the rotations can be used directly. In line with this approach, we used the positions obtained from the 1k dataset to drive the importance sampling on the 4k tomogram and to identify the particles in 4k reconstructions from WBP and SART.

After the alignment and averaging, we obtained ribosomes with the same structure and similar resolution as we got from 1k tomograms. Although this result might seem surprising, it can be explained by the fact that we did not perform CTF correction on the input projections. It is well possible that CTF limits the achievable resolution to roughly 33 Å.

Parameters	Initial Reconstruction			Importance Sampling						
SPV	150	150	150	150	150	150	150	150	300	1000
Transition Width	10.5	9.5	8.5	5.0	4.5	4.0	3.5	3.0	1.0	0.5
Diameters	21	19	17	15	13	11	9	7	5	3
Scaling Factor	35000000									

Table 8.2: Parameter setup for reconstruction on 4k data.

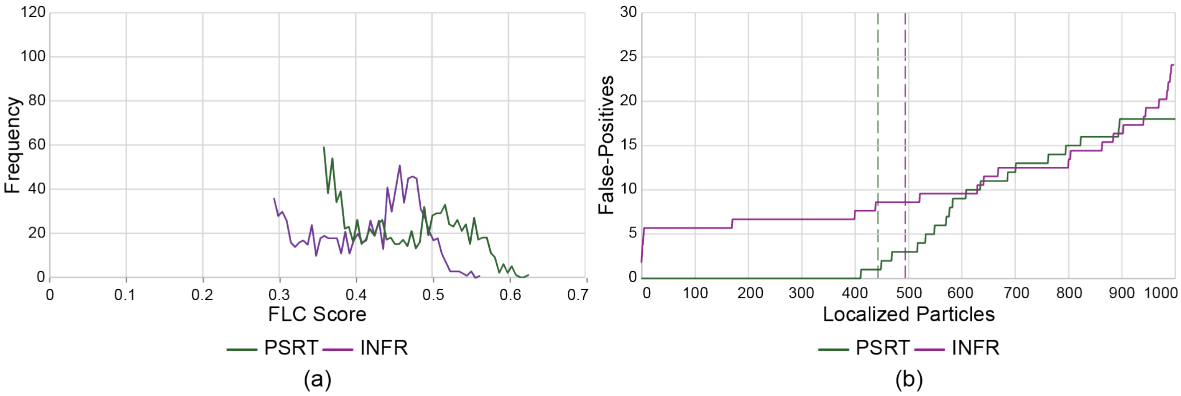


Figure 8.8: (a) Histograms of localization scores of 1000 candidates found in PSRT and INFR (after 25 iterations) 1k reconstructions. For INFR most values are between 0.4 – 0.55, with values above 0.5 containing also gold markers. (b) Graph showing the increase of false-positive candidates with the decreasing cross-correlation score for PSRT and INFR. The dashed lines indicate the number of particles that were suggested for further processing by Gaussian fitting with threshold σ. For INFR the number was 493.

Comparison to INFR

The highest resolution achieved on this dataset was 22.8Å and it was obtained using a novel technique called Iterative Nonuniform fast Fourier transform based Reconstruction method (INFR) [Chen and Förster, 2013]. As its implementation is publicly available, we tried to reproduce the reported resolution. However, the ribosome we recovered using INFR was similar to those obtained with other methods both in terms of structure and resolution which was 33.8Å. As the paper on INFR is rather sparse on details, it is hard to judge if our inability to achieve the same resolution was caused by the lack of CTF correction or by different settings for localization and alignment (or both).

It is worth noting that the resolution of 22.8Å was obtained also from WBP tomogram if the rotations found during the localization on INFR tomogram were used for the final alignment. This suggests that the real advantage of INFR lies in its good localization properties on a downsampled tomogram rather than in the quality of a reconstructed 4k tomogram. It is therefore interesting that the localization on INFR tomogram was worse than on the initial PSRT tomogram both in terms of the cross-correlation score and the false-positives (see Fig. 8.8). It is thus not unreasonable to assume that by using CTF corrected data and/or different setup for localization and alignment we could achieve similar or even better resolution with PSRT.

CHAPTER 9

PARALLELIZATION

Given the computation times for PSRT reconstructions presented in the previous chapters, it is obvious that an additional speed-up is necessary in order to make PSRT a viable alternative to the state-of-the-art reconstruction techniques. It is reasonable to assume that even further optimization on the algorithmic level would not gain us the necessary speed-up and thus, assuming that we have an optimal parameter setup, the only way how we can further accelerate the convergence is to parallelize the computation in some way. There are two levels at which we can consider to do that. We can either split the reconstructed area along the y-axis into smaller parts and process them in parallel or we can perform the random walks simultaneously which, as mentioned early, is not straightforward. In ideal case we could combine both schemes to gain even more speed-up.

In this chapter, we review both schemes together with possible obstacles that might be encountered during their execution. We also validate the first scheme on experimental dataset and show its performance both in terms of reconstruction quality and achieved speed-up.

9.1 Parallelization on the Reconstructed Area

The 3D reconstruction problem can be divided into smaller parts by splitting both the reconstructed area and the corresponding input projections along the y-axis. In the case of parallel rays the splitting can go as far as to reduce the 3D reconstruction into a set of 2D reconstructions out of 1D projections. The final volumes are then stacked together to yield the final tomogram. In fact, many of the currently available implementations of reconstruction techniques prefer the set of 2D reconstructions, as it can lower memory requirements and facilitate the Fourier transform (in case of Fourier based methods) and geometric computations (for reconstructions performed in real space). As the reconstructions are independent, they can be carried out simultaneously and thus this scheme is widely used for parallelization.

To parallelize PSRT, we adapt the division of the reconstructed area and the corresponding residuals to accommodate for the specific properties of the algorithm. As described in Sec. 6.4, we have to ensure proper behaviour of random walks near the boundaries and thus we have to virtually extend a reconstructed area of each part. For the x and z dimension, which corresponds to the width and depth of the reconstructed area respectively, this extension remains the same as for the whole

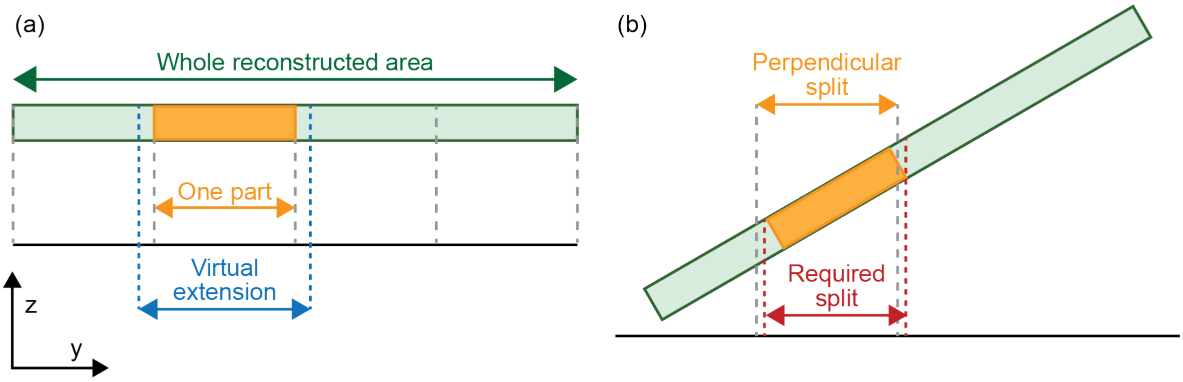


Figure 9.1: (a) Splitting of the reconstructed area along y-axis in case of the parallel acquisition scheme. (b) Effect of the non-perpendicularity of the electron beam on the splitting. The areas in the reconstructed volumes no longer correspond to the parts of projections obtained with the perpendicular split.

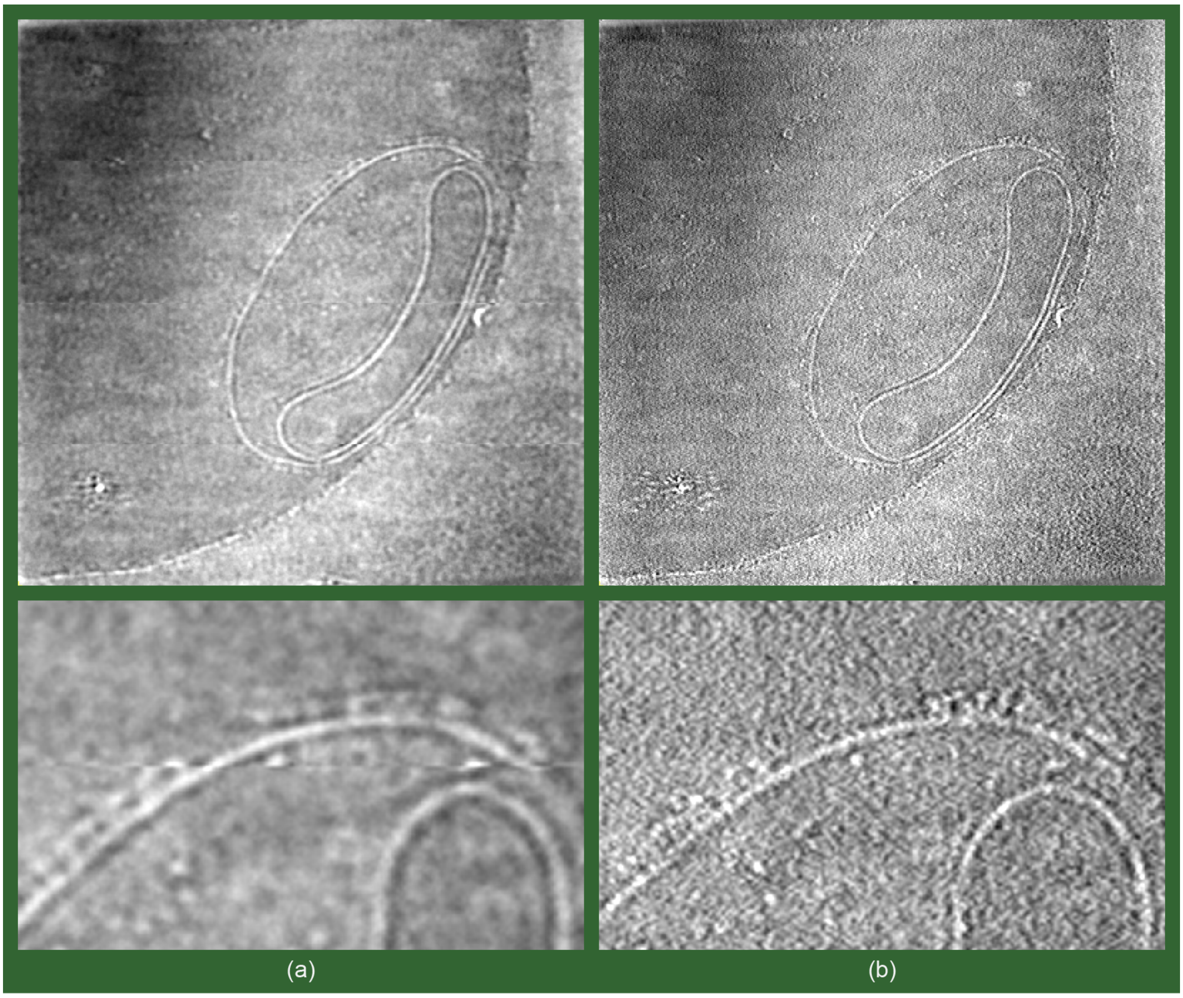
tomogram reconstruction (i.e. we extend the area into which we can place samples by the size of the starting diameter). In y dimension we virtually extend the area by the same size but allow also for seeds to be placed there (similarly as in the case of the interior problem, see Sec. 6.5 for details). The scheme of the division along y-axis is depicted in Fig. 9.1a. Note that the extension in y dimension poses a limitation on achievable improvement on computational time. While for other reconstruction techniques, the time needed for reconstruction of one part is roughly t/n (plus some minimal overhead for the splitting and merging), where n is number of parts and t denotes the time needed for the whole-tomogram reconstruction, in case of PSRT the speed-up does not scale linearly to the number of parts. Nevertheless, as we will see in Sec. 9.1.2, the speed-up over the non-parallelized version is still significant.

Note that despite the sufficient sampling near the edges, we can still observe artifacts on the splitting boundaries, especially at the early stages of the reconstruction process (see Fig. 9.2). The artifacts are caused by the use of bigger samples. They create patterns which are not fine enough and thus result in discontinuities on the splitting boundaries. With decreasing the diameter of the samples, the artifacts become less apparent and for the samples of size 1 they disappear completely.

Currently, in our implementation the parallel computation is carried out using Open Multi-Processing (OpenMP) [Dagum and Menon, 1998], which allows for easy employment on multicore systems. Implementations using Message Passing Interface (MPI) [Forum, 1993] should be considered in order to use the parallel version also on the cluster. Note that parallelization on this coarse level is not suitable for General Processing Unit (GPU) employment, as the number of parts, even for large volumes, is too low to fully exploit the advantages of GPUs.

Non-Perpendicularity of the Electron Beam

The parallelization scheme described above is based on the assumption that the electron beam is perpendicular to the specimen stage. However, due to technical imperfections of TEM, the stage can be tilted along the x-axis which, if neglected, can lead to artifacts in tomograms [Castaño-Díez et al., 2006]. Although the x-tilt is usually small (typically up to 5°), it complicates the splitting along y-axis as the input projections are no longer aligned w.r.t. the corresponding parts of the reconstructed area (see Fig. 9.1b). In case of a 3D reconstruction, the perpendicularity of the beam is not used in the computation of a volume. Therefore, in order to correct for the x-tilt it is only necessary to ensure the correct correspondence among the splitting parts and the corresponding input projections. Thus,



(a)

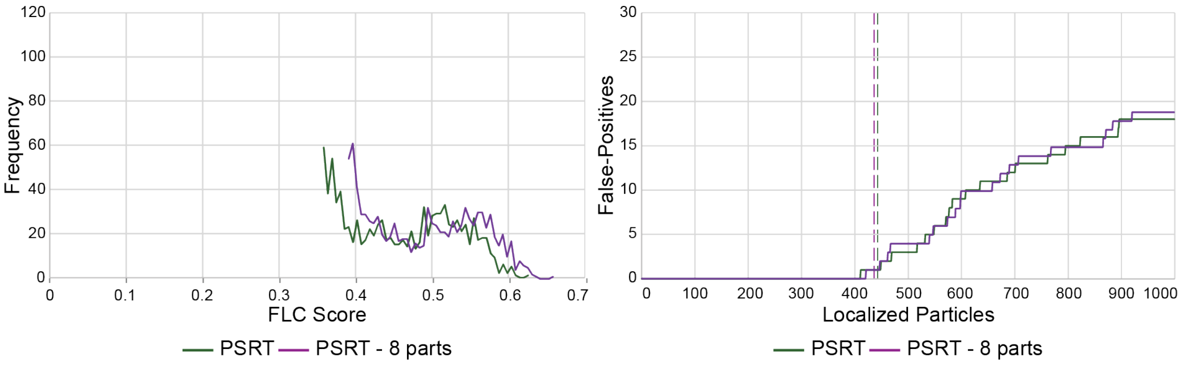
(b)

Figure 9.2: The reconstructions obtained with the parallel-scheme and division of the reconstructed area into 4 parts. (a) Tomogram after the iteration with sample size 7 showing the artifacts on the splitting boundaries (top) with a close-up of the structure influenced by the artifacts (bottom). (b) Tomogram after the last iteration with sample size 1. The artifacts are hardly visible.

the correction for the x-tilt requires only the computation of the shifts and the computation has to be done only once, prior to the splitting. On the other hand, to correct for x-tilt in the case we perform a set of 2D reconstructions is more demanding and typically requires elaborate interpolation schemes [Sandberg et al., 2003].

9.1.1 Validation on Experimental Data

To evaluate the influence of the artifacts along the splitting boundaries we performed the same experiment as described in Sec. 8.3, only that we now use our parallel scheme for the reconstruction. We reconstructed the initial tomogram by splitting the reconstructed area into 8 parts. Since the initial part ended with samples with diameter 7, the merged tomogram contained artifacts along the splitting boundaries. We then performed the localization on this tomogram and obtained results similar to those obtained with non-parallel scheme (see Fig. 9.3). The positions found during the localization were used for subsequent importance sampling of the initial tomogram and alignment was performed



(a) (b)

Figure 9.3: Comparison of localization results on 1k tomograms obtained with non-parallel and parallel version of PSRT. The parallel version split the reconstruction into 8 parts. The parallelization, despite the artifacts on the splitting boundaries, had no significantimpact on quality of the localization both in terms of the achieved cross-correlation score and the false-positive candidates.

to obtain the final averaged structure. The obtained ribosome had the same resolution and structure as the one recovered with non-parallel version. We thus conclude that the artifacts do not have significant impact on the localization properties or the quality of the final averaged structure.

The reconstruction on 4k dataset, as described in Sec. 8.6, was performed using the parallel scheme with subdivision into 16 parts. We again used the parallel scheme for both the initial reconstruction and for the Region Of Interest (ROI) reconstructions. Note, however, that the positions used for importance sampling were found in the initial 1k tomogram obtained with non-parallel scheme. As the final ribosome was similar to those obtained with other methods, both in terms of the structure and achieved resolution, we conclude that the parallel scheme can be used to carry out the reconstruction on large datasets without jeopardizing the quality of the information contained in the reconstructed tomograms. This is of special importance not only in the terms of parallelization. Current datasets can easily reach 8k dimensions and thus the reconstruction might no longer fit into memory, making it essential to have the ability to split the volume into smaller parts without risking the quality of the outcome.

9.1.2 Computation Times

We performed series of experiments to evaluate the speed-up w.r.t. the number of parts as well as dimensions of the reconstructed volumes. We use the experimental dataset obtained from EMPIAR database (for detailed description see Sec. 7.10) and reconstructed tomograms with three different dimensions: 512×512×50, 1024×1024×100, and 2048×2048×200. We tested our implementation on system with two CPUs (Xeon DP Haswell-EP E5-2699, 2.3GHz 45MB 145W, 18 cores each). Since the system contains up 36 cores, we could evaluate the computational times for 1, 2, 4, 8, 16, and 32 parts. The results are presented in Tab. 9.1, with the visualization of the speed-up factors shown in Fig. 9.4. Note that the measured times are without the intermediate volume write-outs. If we store the intermediate volume after each iteration (e.g. to observe reconstruction progress), the computation time would increase as we would have to merge the volumes. This overhead is however negligible w.r.t. the overall time needed for the reconstruction.

As expected, the speed-up does not scale linearly with the number of parts and although the improvement on computation time over the non-parallel scheme is major, the overall times needed for the reconstructions are still not competitive. We can also observe that the scaling gets worse once the

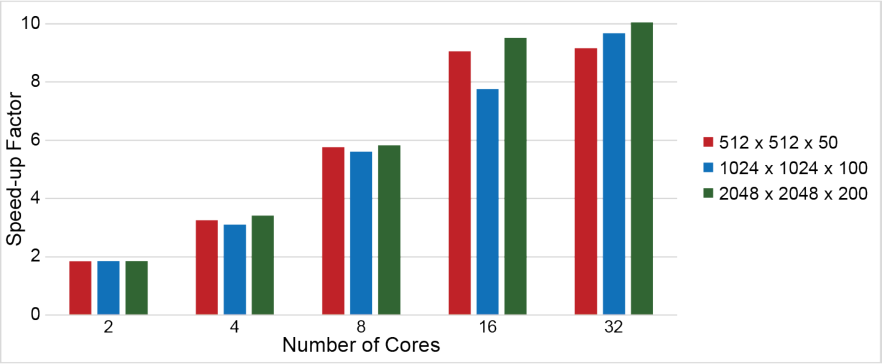


Figure 9.4: Graphs showing the speed-up factors for reconstructions with different dimension obtained with the parallelization scheme.

parts are to small w.r.t. the sample size. For instance, the computation time remains practically the same for 16 and 32 parts on the smallest dataset which is not surprising, as the size of one part in y dimension is 16 for 32 parts and the starting sample's diameter is 13. Thus the number of parts we can use and still gain significant speed-up is limited.

We conclude that despite the major improvement over the non-parallel scheme, it still takes too long to finish a reconstruction and the current performance cannot compete with state-of-the-art methods.

	512×512×50		1024×1024×100		2048×2048×200	
Cores	Time	Speed-up	Time	Speed-up	Time	Speed-up
1	4h 03m 58s		33h 49m 03s		289h 52m 46s	
2	2h 12m 22s	1.8×	18h 15m 32s	1.9×	156h 30m 50s	1.9×
4	1h 15m 01s	3.3×	10h 53m 14s	3.1×	84h 57m 14s	3.4×
8	42m 18s	5.8×	6h 01m 51s	5.6×	49h 46m 24s	5.8×
16	26m 57s	9.0×	4h 21m 26s	7.8×	30h 26m 57s	9.5×
32	26m 37s	9.1×	3h 29m 44s	9.7×	28h 50m 04s	10.1×

Table 9.1: Computational times on multicore PC with two CPUs (Xeon DP Haswell-EP E5-2699, 2.3GHz 45MB 145W, 18 cores each).

9.2 Parallelization of Random Walks

To make PSRT a viable alternative to state-of-the-art methods used in cryoET, parallelization on the level of sample processing might be necessary. Given the character of random walks in PSRT, it should be obvious that the parallelization within a single random walk is not possible. However, we might consider to perform the series of random walks not in a sequential but in a simultaneous manner. As there might be thousands of random walks, this type of parallelization would be especially well suited for GPU. Although we discuss in this section possible approaches for the simultaneous processing of random walks, their implementation is subject for the future work.

By parallelization of random walks, the main problem is the evaluation of a change in the objective function ΔF. It could happen that during a computation of ΔF for one sample, the values in the residuals would be changed by acceptance of another sample and thus the evaluation of ΔF would be based on old residual information. There are two ways how to approach this situation. The first one is the use of extensive locking to prevent overlapping of read and write operations. Since locking is expensive and since it is hard to minimize the overlap of the samples due to the random character of their placement, this approach might not be very practical. The second possibility is to simply ignore that and allow for evaluation of ΔF from old residual information. Although this would potentially jeopardize the convergence property of PSRT from the theoretical point of view, in practice the sample energy is very low w.r.t. the values in the residuals and thus the impact of possible simultaneous changes would be low as well.

Note that the difference between the two approaches is similar to ART and SIRT. In ART, we update the volume after we compute the error for a single pixel and thus the information we have is always up-to-date. By contrast, in SIRT, we compute the error for all pixels in all projections and update the volume simultaneously. ART converges fast but the quality of results is worse. SIRT needs more iterations but the results are better w.r.t. noise and contrast. In case of PSRT we do not have to worry about the reconstruction quality but about the speed-up. In the first approach, the locking might be so expensive that it cancels out the advantages of using GPU. We assume that the second approach might, similar to SIRT, slower the convergence and thus in end effect lower the speed-up gained by using GPU. Currently, it is hard to judge the amount of the speed-up (if any) that could be gained by parallelizing the PSRT on the level of random walks and this topic remains subject of future work.

CHAPTER 10

GEOMETRIC ARTIFACTS IN CRYOET

In Sec. 2.2.2, we briefly mentioned that single-tilt acquisition scheme introduces artifacts into the reconstructed volumes. More precisely, the artifacts are caused by very-well known problems: The missing wedge problem, the missing information problem, and finally the interior problem. These artifacts and their effects on the uniqueness and stability of the inverse problem solution have been studied separately to various extents, both by the mathematical as well as the electron microscopy community (see e.g. [Louis, 1989, 1986; Louis and Rieder, 1989; Frikel and Quinto, 2013; Xu et al., 2010; Palmer and Loewe, 2014]). However, to our knowledge, no study has been performed that describes those artifacts systematically across various reconstructions methods and with nomenclature and focus on their manifestation in cryoET reconstructions. In this chapter, we fill this gap by presenting all the problems in a unified geometrical framework and characterizing both their individual as well as the combined influence on quality of cryoET tomograms obtained with different reconstruction techniques.

We explain how each of the problems arises and when it can be avoided and how. In our description we first focus only on whole-tomogram reconstruction and summarize the influence of individual artifacts on subtomogram averaging later in the separate section. Then we demonstrate the artifacts on a set of experiments performed on numerical phantom volumes, simulating and evaluating each of the problems separately. The numerical experiments are followed by illustrating the combined effects on an experimental dataset and disentangling the contributions of the individual artifacts. In the experiments, we compare different reconstruction techniques: In addition to WBP (both real space and Fourier space approach), SART, and PSRT we also evaluate Iterative Nonuniform fast Fourier transform based Reconstruction method (INFR) [Chen and Förster, 2013] as it represents a rather rare combination of an iterative approach with a Fourier based approach. For all methods we discuss their ability to suppress the geometric artifacts. Finally, we draw conclusions and recommendations on numerous points, regarding the mutual influence of the artifacts and their significance to the interpretation both in electron tomography and subtomogram averaging experiments.

For illustration purposes we reduce the 3D reconstruction problem to 2D reconstruction out of 1D projections with the 2D reconstruction lying in xz plane. This simplification does not have any influence on the artifacts.

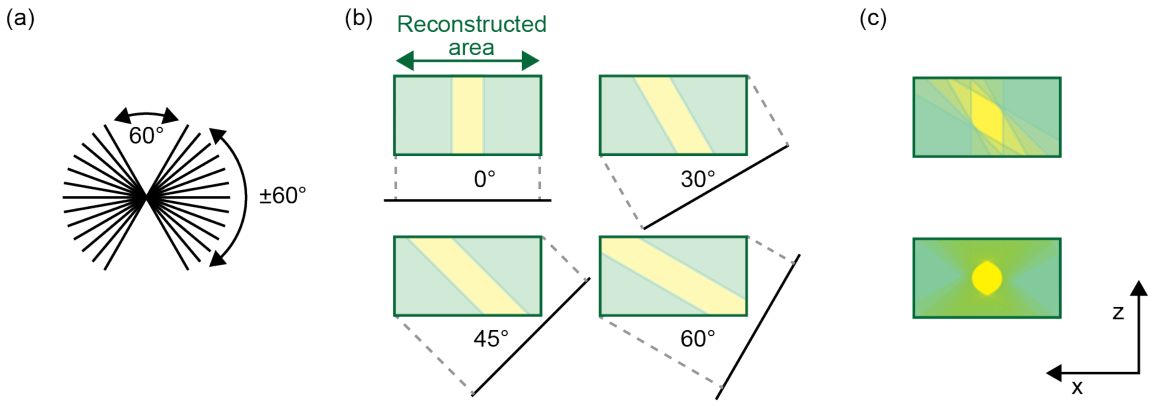


Figure 10.1: (a) 2D scheme of the Missing Wedge (MW) problem. (b) Illustration of backprojection from four projections containing projected sphere values. (c) Reconstruction of a sphere from four projections taken under 0°, 30°, 45°, and 60° tilt angle (top) and from many projections taken within a tilt range of ±60° (bottom). The missing wedge causes the sphere to be distorted in z direction.

10.1 Missing Wedge

In the single-tilt scheme of TEM, the rotation of the scanned specimen is limited and the input projections do not cover the whole angular range of 180° (see Fig. 10.1), causing the Missing Wedge (MW) problem, which is also known as the *incomplete projection set problem* or the *limited (tilt) angle problem* [Louis, 1986]. The main reason for the limited tilting is the increase of the effective thickness of the scanned specimen at higher tilt angles where the path of the electron beam through the scanned specimen becomes longer, leading to a decrease of the Signal-to-Noise Ratio (SNR).

The MW problem affects a whole tomogram and demonstrates itself through two types of artifacts. The first one is distortion of the structures in the direction of the MW [Radermacher, 1988]. In general, the shape of the distortion differs for various structures based on their shape and orientation [Frank, 2006; Palmer and Loewe, 2014]. Moreover, the distortion increases with an increasing size of the MW and in some cases might lead to complete loss of thin features, such as membranes, that are oriented perpendicularly to the MW direction. The second type of artifacts are dark halos and streaking artifacts near objects with high contrast (in cryoET this is most apparent near gold markers). Recently, it has been shown in [Palmer and Loewe, 2014] that some projections with high tilt angle might not provide any useful information, as their SNR is too low, and that the effective missing wedge might be actually larger than indicated by the tilt angles. This influences any computations that use the tilt angles range for missing wedge representation (e.g. resolution estimation).

Few approaches have been suggested to eliminate or at least reduce the MW problem. The most common approach in cryoET is dual-axis (or double-tilt) tomography [Penczek et al., 1995; Mastronarde, 1997] where the specimen is tilted around two orthogonal axes reducing the missing wedge into a missing-pyramid. Although recent development of both hardware and software methods allow for this approach to become more common [Iancu et al., 2005; Zheng et al., 2009; Myasnikov et al., 2013], the double-tilt does not eliminate the problem completely. A different approach is the use of cylindrical specimen, the thickness of which does not increase during the tilting. This approach was successfully applied in ET in both material science [Fukuda et al., 2004; Kawase et al., 2007; Hernández-Saz et al., 2012] and biology [Turner et al., 1991; Barnard et al., 1992]. Recently, a special holder for cylindrical specimen for frozen-hydrated biological material was developed by [Palmer and Loewe, 2014], making it possible to use this approach also in cryoET. In the field of material science, a-priori information was successfully used to reduce or eliminate the MW problem [Batenburg et al., 2009; Saghi et al., 2011; Leary et al., 2013]. Currently, there are no software-based approaches to completely prevent or eliminate the MW artifacts in cryoET tomograms.

10.2 Missing Information

The second cause for geometric artifacts is Missing Information (MI) in the input projections. As depicted in Fig. 10.2, if the width of the reconstructed area is as large as the width of the input projections, then the projections at higher tilt angles do not contain information from corners of the reconstructed area. Consequently, the reconstructed tomogram has anisotropic resolution with the corners containing less information than the rest of the tomogram. The corners appears darker than the rest of the reconstruction. Moreover, the edges of the input projections cause vignetting artifacts that are formed on the boundaries between the reliable part of the tomogram and its corners. This becomes more apparent with increasing depth of the reconstructed area. If the depth is the same as the width, the artifacts form a ring in the reconstruction (see Fig. 10.11d). The brightness of the ring and of the corners is influenced by different filtering properties, such as cut-off frequencies, boundary conditions handling, and/or interpolation in case of Fourier based techniques. Depending on their brightness, the artifacts might significantly change the histogram of a reconstruction and thereby limit its contrast and make any subsequent analysis more difficult. As depicted in Fig. 10.2b, the size of the area affected by the MI artifacts depends on the reconstructed depth.

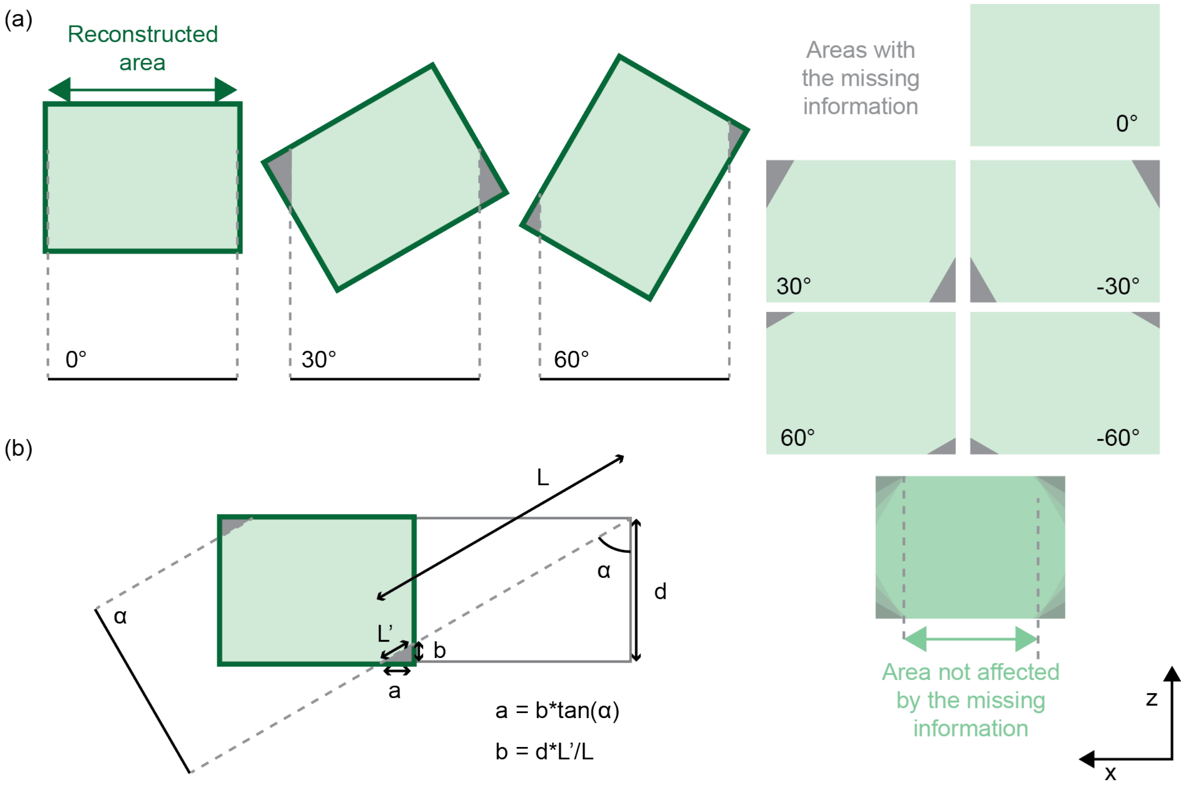


Figure 10.2: 2D scheme of the Missing Information (MI) problem. (a) The reconstructed area has the same width as the input projections. For larger depth, the projections at higher tilt angles do not cover the whole reconstructed area causing the corners of the area to contain less information than the rest of the tomogram. (b) The extend to which the tomogram is affected by the missing information depends on the depth of the reconstructed area.

For geometric methods, there are only two ways how to avoid the artifacts completely and both lead to unnecessary loss of useful information. The first approach is to reconstruct an area which is smaller than the width of the input projections. The second approach is to reconstruct an area with the same width as the input projections and then trim it to get rid off the parts affected by the artifacts. The latter one is the only solution applicable also to Fourier based methods.

10.3 Interior Problem

The specimen-level Interior Problem (IP) in cryoET refers to the presence of redundant information contained in the input projections. It is also known as the *long object problem*, the *truncated projections problem* or *region-of-interest reconstruction problem* [Louis and Rieder, 1989]. As depicted in Fig. 10.3, the scanned specimen is larger than the detector and since in cryoET we have parallel rays, every value of a given input projection integrates the values along the ray of the same length (see Fig. 10.3a left). However, during the back projection step the values are not distributed over the rays of the same length as the reconstructed area is smaller than the scanned specimen and some ray segments are thus shorter. This results in increased values at the edges of the tomogram (Fig. 10.3a right). For the reconstructed area with width w and depth d and for the maximal tilt angle α_{max}, the width of the area affected by IP on one side of the tomogram is $d * \tan\alpha_{max}$. This reduces the reliable area in the tomogram in x dimension by $2 * d * \tan\alpha_{max}$. Moreover, the strong artifacts also change the tomogram's histogram, which can both limit the contrast as well as complicate any subsequent processing.

For geometric based methods there are four options how to avoid the artifacts caused by the IP. The first one is to extend the reconstructed area in x dimension so it contains all the information from the input projections (see Fig. 10.3c right) and then to trim the volume to the original dimensions. This not only adds to the computational time but also significantly increases memory requirements, making such strategy impractical.

The second option is to extend the reconstructed area in x dimension only by an amount necessary to avoid the artifacts in the region of interest and then to trim the volume to the original dimensions. As depicted in Fig. 10.3c, the width of the extended reconstructed area w_{ext} is given by:

$$w_{ext} = min\{w_{scanned}, w + 2 * d * \tan(\alpha_{max})\} \tag{10.1}$$

where $w_{scanned}$ is the width of the scanned specimen. However, this solution could also lead to the significant increase of both the memory requirements as well as the processing time as even this minimal extension can double the size of a tomogram. The increased memory requirements were overcome by the Memory Efficient Solution to IP (MESIP) designed within the PSRT method. It extends the reconstructed area by $2 * d * \tan(\alpha_{max})$ but this extension is only virtual, i.e. the extended parts are processed normally but they are not stored and hence the memory requirements do not increase. The MESIP approach prevents the artifacts but is based on a local character of PSRT and thus its integration into other geometric based methods is not straightforward.

The third approach is to trim the reconstructed area to get rid off the affected parts which would lead to an unnecessary loss of the information present in the projections.

The fourth and last method is an approximate solution called Long Object Compensation (LOC), which was suggested in Xu et al. [2010]. It uses additional weighting of the input projections that is based on the ratio of ray's length passing through a specimen to a ray's length passing through a reconstructed area. This weighting compensates for the shorter rays and thereby suppresses the artifacts.

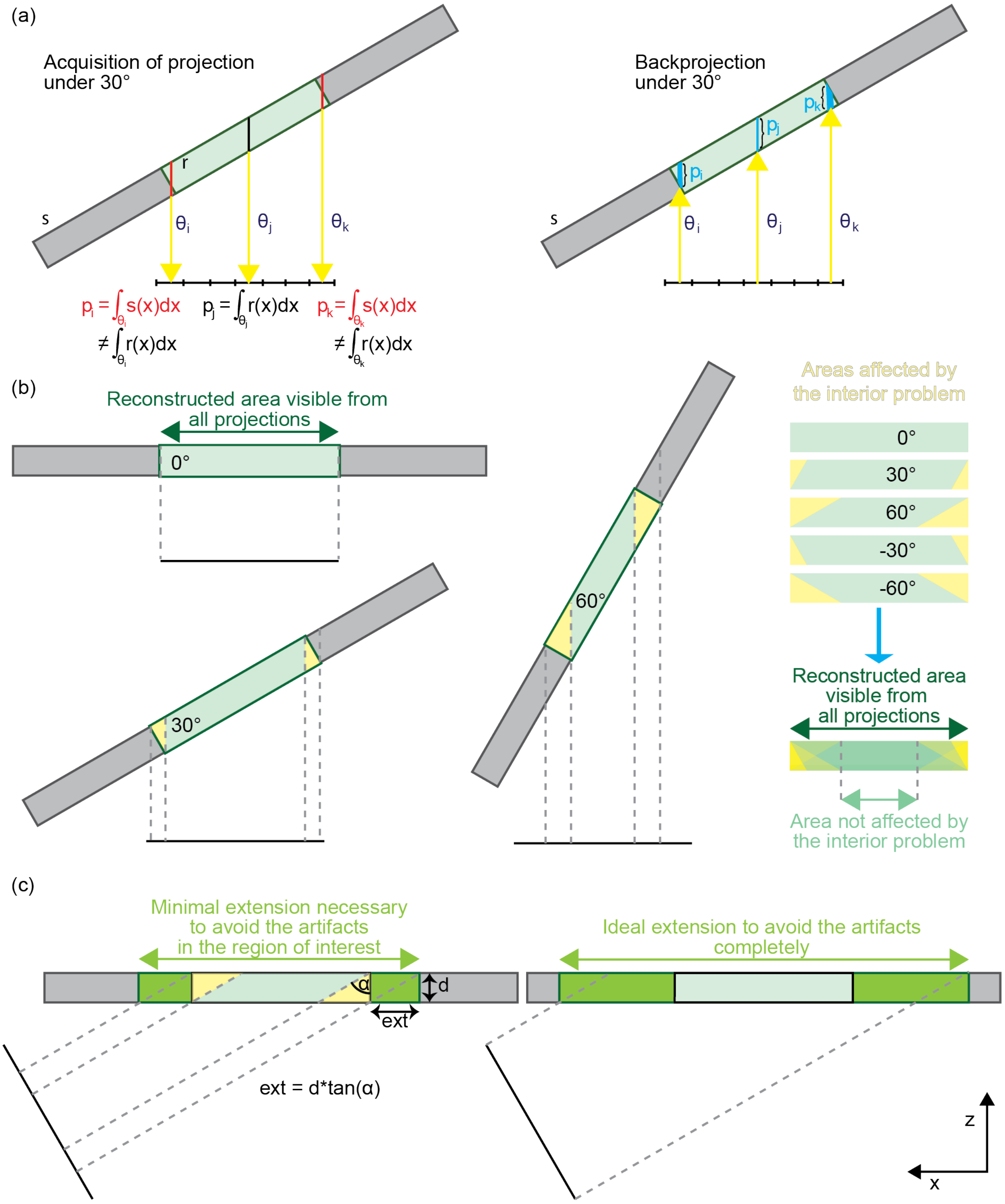

(a)
Acquisition of projection under 30°
s
r
θi
θj
θk
$p_i = \int_{\theta_i} s(x)dx \neq \int_{\theta_i} r(x)dx$
$p_j = \int_{\theta_j} r(x)dx$
$p_k = \int_{\theta_k} s(x)dx \neq \int_{\theta_k} r(x)dx$
Backprojection under 30°
s
pi
pj
pk
θi
θj
θk
(b)
Reconstructed area visible from all projections
0°
30°
60°
Areas affected by the interior problem
0°
30°
60°
-30°
-60°
Reconstructed area visible from all projections
Area not affected by the interior problem
(c)
Minimal extension necessary to avoid the artifacts in the region of interest
α
d
ext
ext = d*tan(α)
Ideal extension to avoid the artifacts completely
z
x

Figure 10.3: 2D scheme of the Interior Problem (IP). (a) The difference between the acquisition process and backprojection. The values are distributed during the backprojection into a smaller area than they were obtained from. (b) The coverage of the reconstructed area affected by the interior problem for different tilt angles. Note that even for large angles the reconstructed area is completely covered by the projection and thus no MI problem occurs. These areas are overlapped to show the amount of reliable information in the tomogram (right). (c) The computation of minimal extension sufficient to avoid the artifacts in the reconstructed area for the maximal tilt angle α_{max} (left) and an illustration of an ideal extension that would cover all values from all projections (right).

As mentioned in Sec. 3.2.2, the Fourier based methods are not flexible w.r.t. changing the dimensions of the reconstructed area and the implicit size of 2D tomogram reconstructed from 1D projections of the size w is fixed to $w \times w$. There are two possibilities for obtaining a tomogram with the size $w \times d$ where $d < w$. In the first approach, one performs IFT directly on the image spectrum of the size $w \times w$ and trims the final tomogram to have the dimensions $w \times d$. As this is similar to the extension in the geometric approaches, the final tomogram does not contain any IP artifacts but the memory requirements might be very large. In the second approach, the image spectrum of the size $w \times w$ is first transformed to have the size $w \times d$ and only after that IFT is performed. In this case, the IP artifacts can appear in the form of aliasing as observed in [Sandberg et al., 2003]. To avoid that, the authors suggest to extend the image spectrum in x direction prior the IFT. Despite a different derivation, the proposed extension is consistent with Eq. 10.1.

It is important to realize that in cryoET the interior problem occurs also w.r.t. to the z-dimension, as the reconstructed depth is usually significantly smaller than the original thickness of the scanned specimen. The decrease of the ray length w.r.t. to the z-dimension is same for all values in one projection and thus does not cause any visible artifacts. However, the overall values in the whole tomogram are affected by this and therefore it should be considered while interpreting the reconstructed values.

10.4 Subtomogram Averaging

The influence and impact of the geometric artifacts differs for Subtomogram Averaging (SA) as one does not work with the tomogram as a whole but is interested only in individual instances of a structure that are distributed within it. These instances have to be localized, aligned, classified and finally averaged to obtain one or few high-resolution structures. Often multiple tomograms are used to increase the number of instances and thereby improving the resolution of the final structure.

The SA protocol itself provides solution to MW problem - averaging of many subtomograms with different orientations eliminates the shape distortions due to MW in the final averaged structure. Nevertheless, these distortions do complicate many steps within the SA procedure such as template-based localization or the alignment step itself. Specific procedures have to be devised to perform correct subtomogram alignment with the presence of MW and even with these procedures, the amount of information available to perform the alignment is significantly reduced due to the presence of MW artifact.

The vignetting artifacts caused by MI in the corners of a tomogram might, based on their intensities, complicate localization and further processing of instances in those regions. Moreover, due to the lower resolution of the corners, the instances bear less reliable information. For SA, the simplest solution is to avoid the affected regions. As MI influences only relatively small part of a tomogram, it should not impact that many instances.

Similarly to MI, if present the disproportional intensities at the edges of a tomogram caused by IP w.r.t. the x dimension might complicate the localization and any further processing of the affected instances. One solution is to completely avoid the areas affected by the artifacts. However, based on a type of the structure and the dimensions of a tomogram, the IP artifacts might have impact on a significant number of instances and thus eliminating the artifacts using one of the possibilities described above is better option than the complete avoidance of affected regions. As stated previously, the IP w.r.t. the z dimension does not cause visual artifacts but influences the values in the tomogram. In SA this puts additional pressure on a correct normalization step, as the range of the values might be different for each tomogram used and without a robust normalization the subtomogram alignment will not provide reliable results.

10.5 Experimental Evaluation

We performed 2D reconstructions of three phantoms, each simulating one of the discussed geometric cases, both with and without the noise in the input projections. We performed an additional set of 2D reconstructions to study the combined effects of the geometric problems on values within a tomogram. Finally, we performed 3D reconstruction on an experimental dataset where all three geometric problems appear. Different reconstruction techniques were used: WBP from IMOD (IMOD WBP), which is a Fourier based method, WBP with both filtering and back projection done in the Real Space (RS WBP), and SART from the Ettention framework [Dahmen et al., 2016]. For the filtering in IMOD WBP we used 0.35 for the radial filter cutoff and 0.05 for falloff (for details on the filter used in IMOD, see [Sandberg et al., 2003]). In the RS WBP we implemented a simple Ram-Lak filter [Ramachandran and Lakshminarayanan, 1971] with no window function and with the size identical to the input projections width. For SART reconstructions on phantoms we used relaxation parameter $\lambda = 0.7$ and 3 iterations for noiseless input projections, $\lambda = 0.1$ and 8 iterations for the input projections with noise. SART reconstructions on experimental dataset were performed using $\lambda = 0.3$ and 5 iterations.

In addition to the state-of-the-art methods, we also performed reconstructions using PSRT and the implementation of Iterative Nonuniform fast Fourier transform based Reconstruction method (INFR) from PyTom, which belongs to the Fourier based methods. Currently, the PyTom's implementation of INFR allows only for cubic reconstructions, making this method rather inefficient for larger volumes. To save space, the tomograms presented here were trimmed in the z dimension to correspond to the size of volumes obtained with the other techniques. For all INFR reconstructions presented here we used 25 iterations. The parameter setup for all PSRT reconstructions can be found in Tab. 10.1 and Tab. 10.2.

In order to compare the values on the 2D phantoms, we normalized them using the EMAN2 package [Ludtke et al., 1999].

Parameters	MW				MW with Noise				MI				MI with Noise			
SPV	200	200	200	100	200	200	200	100	200	200	200	100	300	200	200	10
Transition Width	7.0	5.0	3.0	1.0	7.0	5.0	3.0	1.0	7.0	5.0	3.0	1.0	7.0	5.0	3.0	1.0
Diameters	7	5	3	1	7	5	3	1	7	5	3	1	7	5	3	1
Scaling Factor	300				300				200				300			

Table 10.1: Parameter setup for PSRT reconstructions.

Parameters	IP				IP with Noise				Spheres				Experimental dataset					
SPV	200	200	200	100	300	300	200	10	300	300	300	100	160	160	160	160	200	200
Transition Width	7.0	5.0	3.0	1.0	7.0	5.0	3.0	1.0	7.0	5.0	3.0	1.0	6.5	5.5	4.5	3.5	2.5	0.5
Diameters	7	5	3	1	7	5	3	1	7	5	3	1	13	11	9	7	5	3
Scaling Factor	400				1000				200				3000					

Table 10.2: Parameter setup for PSRT reconstructions.

10.5.1 Missing Wedge

The phantom was a 150×80 rectangle with intensity 0.7 and we reconstructed areas of the size 300×200 and 300×80. The input projections were of size 300×1 and we used three different sets of projections: noiseless complete projection set with 40 projections covering ±90°, noiseless incomplete projection

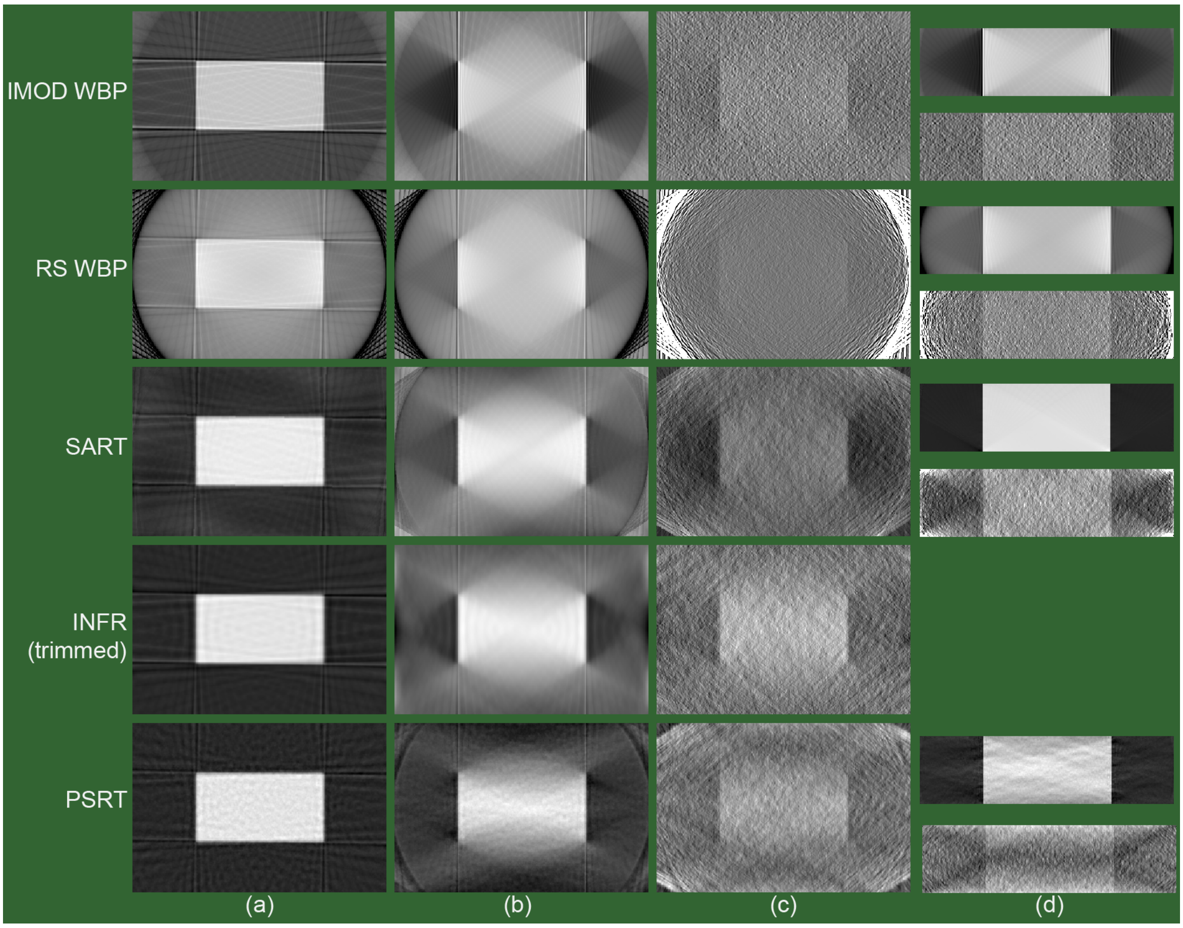


Figure 10.4: Comparison of different reconstruction techniques in the presence of the Missing Wedge (MW) problem. The phantom was a 150×80 rectangle with intensity 0.7. (a) 300×200 reconstruction from the complete projections set consisting of 40 projections covering ±90°. (b) 300×200 reconstruction from the incomplete noiseless projection set with 30 projections with angles from −60° to +56°. (c) 300×200 reconstruction from the incomplete projection set with Gaussian noise with SNR 10. (d) 300×80 reconstruction from the incomplete projection set without noise (top) and with Gaussian noise with SNR 10 (bottom).

set with 30 projections from −60° to +56° and the same incomplete set with Gaussian noise with SNR 10.

All the resulting tomograms are shown in Fig. 10.4 (corresponding Fourier spectra can be found in Supplementary material, Sec. A.1). Regardless of the reconstruction method, all tomograms of the size 300×200 obtained from the noiseless incomplete projection set contain the artifacts associated with the MW problem: The dark halos in the form of two triangles on the outer vertical boundaries of the phantom, the lost horizontal edges of the phantom and also the distortion of the phantom in the direction of the MW. In the presence of noise, the artifacts are less distinct and only the vertical edges of the phantom are clearly visible. If the reconstructed area has the same depth as the phantom, the distortions in MW direction are suppressed, leading to the better results. This is especially apparent by the SART reconstruction from the noiseless projection set – the dark halos on the outer vertical boundaries of the phantom are hardly visible. Although the dark halos are also barely visible in the PSRT reconstruction, we can observe that the method does not recover the homogeneous parts as well as SART. This is caused by the Monte Carlo character of the PSRT method.

10.5.2 Missing Information

The phantom was a 200×150 rectangle with intensity 0.7. Three tilt-series were produced, the first one covering the whole object with projections of size 500×1, the second one with projections of size 200×1 (the same size as is the width of the phantom), and the third one the same as the second tilt-series only with Gaussian noise with SNR 10. The experiments were done with complete projection set (40 projections with angles from ±90°).

The results are shown in Fig. 10.5. While the tomograms from projections of the size 500×1 do not contain any strong artifacts in the phantom area, the artifacts described in Sec. 10.2 are clearly visible on all reconstructions (except for PSRT) from the noiseless projection set of the size 200×1. Note that enlarging the reconstructed area would not help to avoid the artifacts in this case – as long as

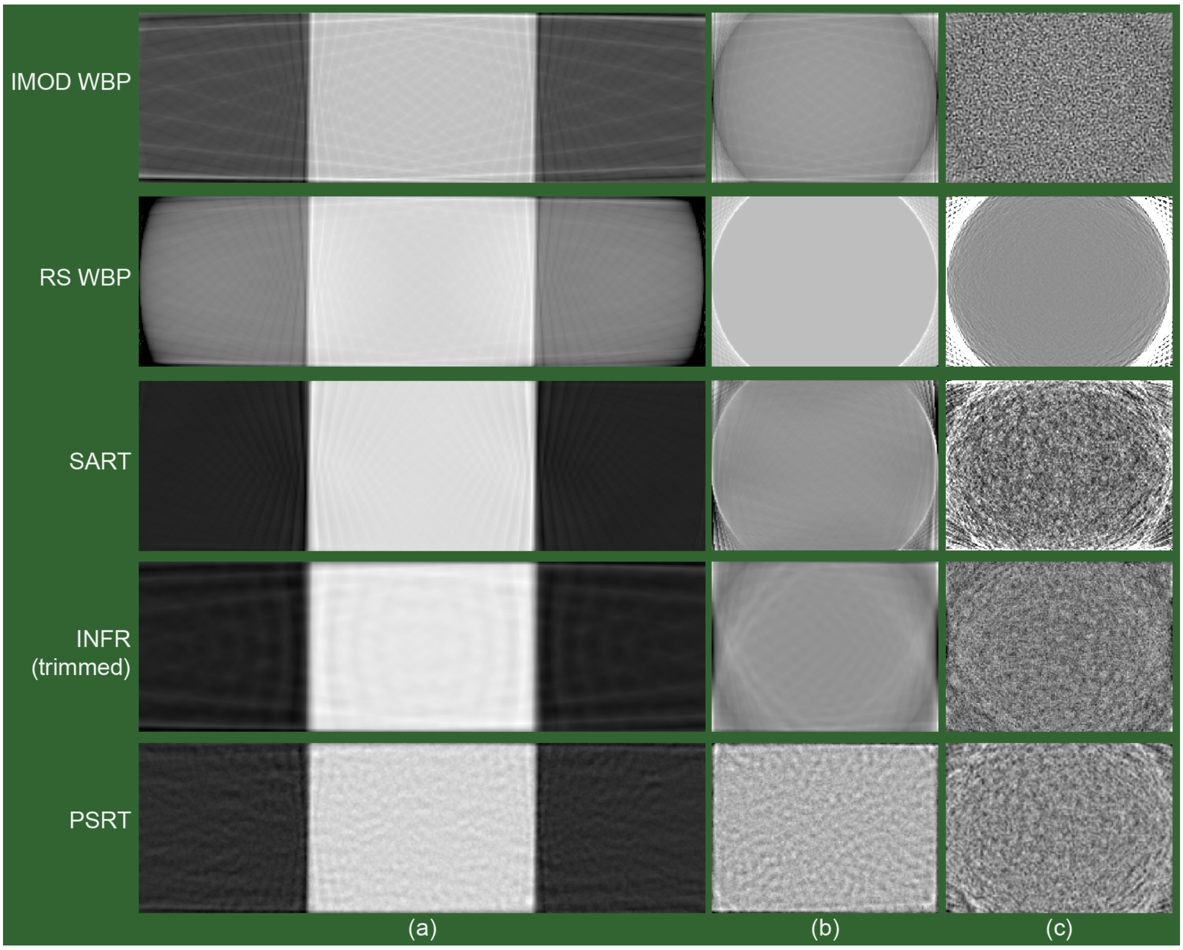


Figure 10.5: Comparison of different reconstruction techniques in the presence of the Missing Information (MI) problem. The phantom was a 200×150 rectangle with intensity 0.7. (a) Reconstructions of size 500×150 from projections of size 500×1. The input projections cover the whole object and thus the MI artifacts are not present in the reconstructed object. However, in case of RS WBP method one can observe the ring artifact at the edges of the reconstructed area. (b) Reconstructions of size 200×150 from noiseless projections of size 200×1. The input projections have the same size as the reconstructed object leading to missing information in the corners of the tomograms. The difference in the intensity of the characteristic ring like artifact is caused by different filtering (or its lack of) and whether the input projections contain meaningful information or are empty at the edges. (c) Reconstructions of size 200×150 from projections of size 200×1 with additional Gaussian noise with SNR 10.

the size of the projections is smaller than or equal to the size of the phantom, the artifacts will arise within the phantom area. The intensity of the ring artifacts as well as the appearance of the corners differ due to the use of a different filtering techniques or the lack of them. These differences are even more prominent in the presence of noise. The artifacts in IMOD WBP and INFR are barely visible and their intensity corresponds to the rest of the tomogram. Further experiments showed that this is not a property of Fourier based methods but it is caused by filtering done in the Fourier space (see Supplementary material, Sec. A.4). Interesting are the results delivered by PSRT – the vignetting artifacts are hardly visible both in the noiseless tomogram as well as in the presence of noise. We assume that this is caused by the different reconstruction approach that, unlike the other methods, does not back project values from the input projections along the rays but inserts them locally into the reconstructed volume and thus avoids emphasizing edges of the input projections that cause the vignetting.

We can also observe that the ring artifacts can have different gray values for the same method, depending on whether the boundaries of the projections contain meaningful information (bright ring) or empty space (dark ring), which can be seen if we compare the results from RS WBP for MW and MI.

10.5.3 Interior Problem

The phantom was a 300×20 rectangle with intensity 0.7. Three tilt-series were produced: the first one covers area larger than than the object with projections of size 500×1, the second one covers only part of the object with projections of size 100×1, the third one is the same only with Gaussian noise with SNR 10. All three tilt-series consist of 40 projections with angles from ±90°. In addition to the techniques used in the previous sections, we performed reconstructions using SART method with the LOC (for more details see Sec. A.3) and PSRT with MESIP (see Sec. 5.3) to show performance of both approaches.

The reconstructed areas were of size 500×20 and 100×20. The results are shown in Fig. 10.6. As expected, the artifacts described in Sec. 10.3 are clearly visible in all geometric based methods that do not use any compensation w.r.t. IP. The IMOD WBP, SART with LOC, and PSRT with MESIP suppress/avoid the artifacts reasonably well. The results delivered by INFR show that despite the squared reconstruction done in Fourier space, the artifacts appear in case of the iterative approach.

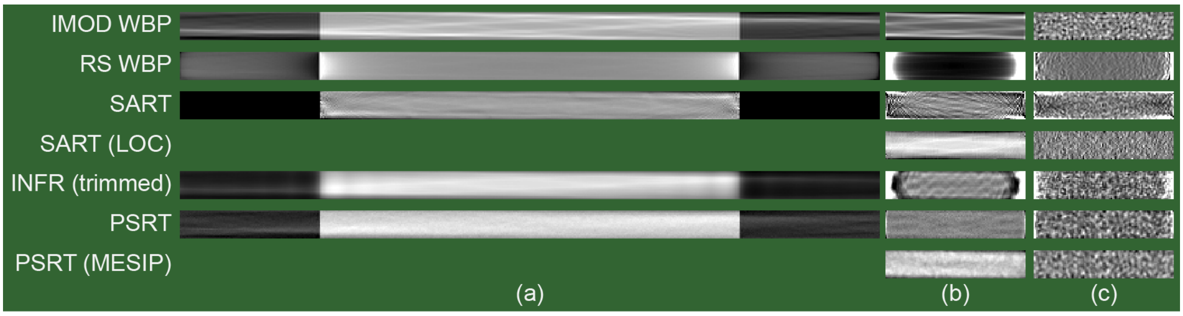


Figure 10.6: Comparison of different reconstruction techniques in the presence of the Interior Problem (IP). The phantom was a 300×20 rectangle with intensity 0.7. (a) Reconstructions of size 500×20 from projections of size 500×1. The input projections cover the whole object and thus the IP does not occur. (b) Reconstructions of size 100×20 from noiseless projections of size 100×1. The input projections are smaller than the phantom and thus the reconstructed areas contain artifacts caused by the IP except for IMOD WBP, SART with LOC, and PSRT with MESIP reconstructions. (c) Reconstructions of size 100×20 from projections of size 100×1 with additional Gaussian noise with SNR 10.

10.5.4 Influence on Reconstructed Values

To study effects of the artifacts on the values within a tomogram, we performed an additional reconstruction of the phantom with size 300×150 that contained 12 spheres. The phantom had intensity 0.7 (with empty space having intensity 0.0), the spheres had radius 6 and intensity 2.0 with one sphere having intensity 10.0 (see Fig. 10.7 top). We generated two tilt-series: the first one had size 500×1 with 30 projections from −60° to +56°, the second tilt-series covered the same angular range and its projections had size 150×1. Tomograms generated from the first series contained only MW artifacts while tomograms generated from the second tilt-series contained in addition to MW also IP and, based on the depth, also MI artifacts. We reconstructed tomograms with depth 150 and 80 to study the influence of the depth on the reconstructed values.

First, we compared the voxel values of the spheres to each other within each tomogram. We cut out selected areas and evaluated their values along horizontal and vertical lines through their centers (see Fig. 10.7 and Fig. 10.8). We also compared the unprocessed subtomograms and the normalized ones to study the influence of the normalization on the subtomograms. For the MW series, the distortion due to the MW is clearly visible on the values corresponding to the vertical lines through the particles. These values are not recovered very well in comparison to the values along the line in the x dimension. Moreover, the strong intensities from the high intensity particle disrupted the structure of particle 1 and 3. While the distortion in particle 3 was partially suppressed by the normalization, the strong artifacts within particle 1 remained (see Fig. 10.7c and Fig. 10.7d). In the second tilt-series with size 150×1, the particles 2, 4, and 5 are affected by IP and MI problem. In tomograms with strong intensities in the corners and at the edges (WBP RS, SART, PSRT), the particles suffer from distortions that were present even after the normalization. The application of LOC for SART and MESIP for PSRT improved the values in all three particles. The particle 5, which is completely in the MI region, was recovered surprisingly well in tomograms without the strong vignetting artifacts. For both tilt-series the influence of a tomogram's depth did not seem to have significant impact on the recovered values with the exception for PSRT and SART where superior results were achieved by reconstructing the original thickness of the specimen. Especially in the case of the first tilt-series, PSRT pronounced the artifacts associated with the MW problem very strongly when the reconstructed depth was smaller than the original thickness of the specimen.

We also compared tomogram values w.r.t. to the ground truth. We observed that, regardless of the type of artifacts, reconstructing the original thickness of a specimen is necessary for recovering correct values in the geometric based methods. We further observed that the strong intensities caused by MI and IP negatively influenced the values also outside the affected areas. The use of LOC for SART and MESIP for PSRT lead to major improvement in the recovered values. The graphs are showns in Fig. 10.9 and Fig. 10.10, respectively.

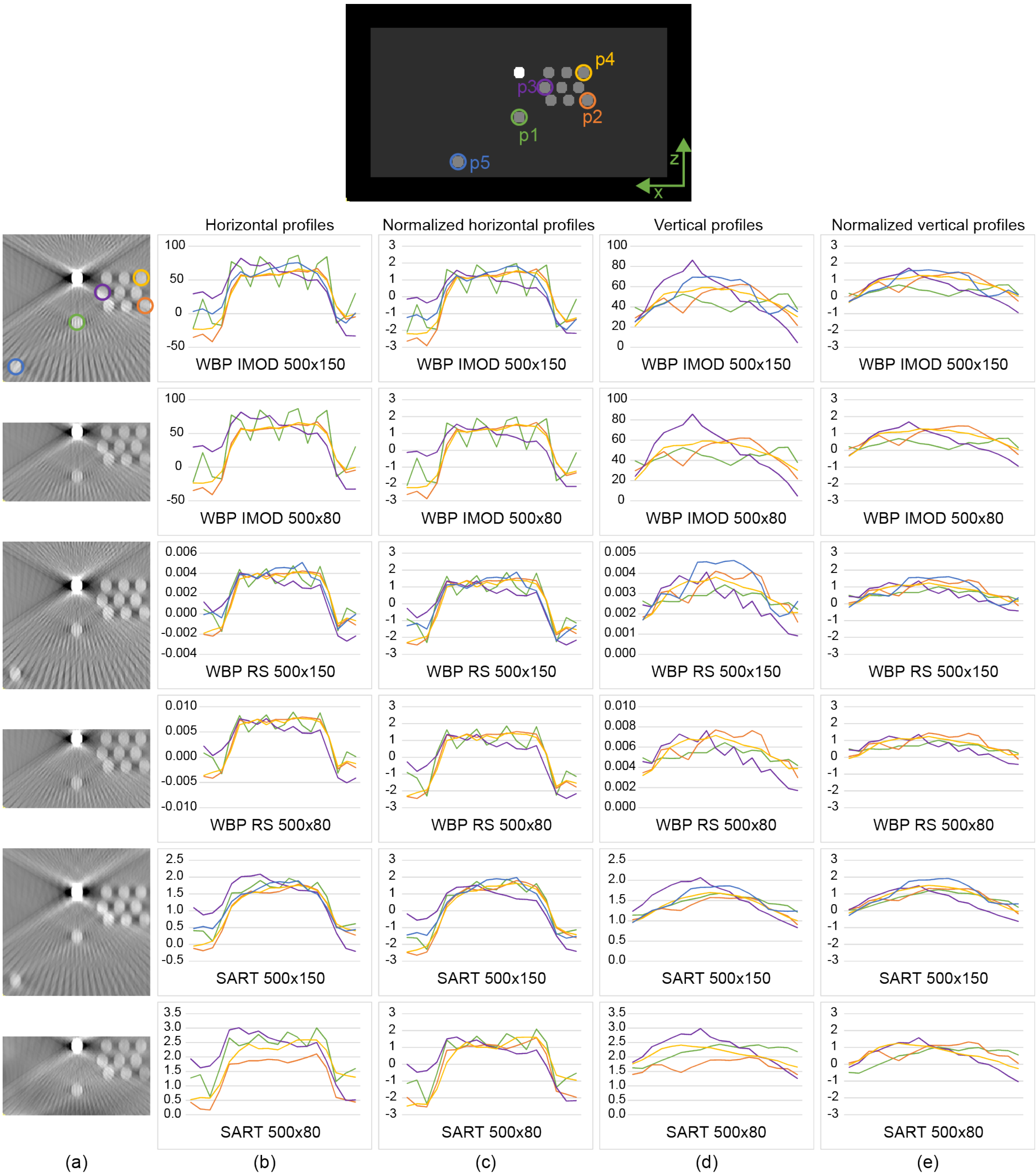
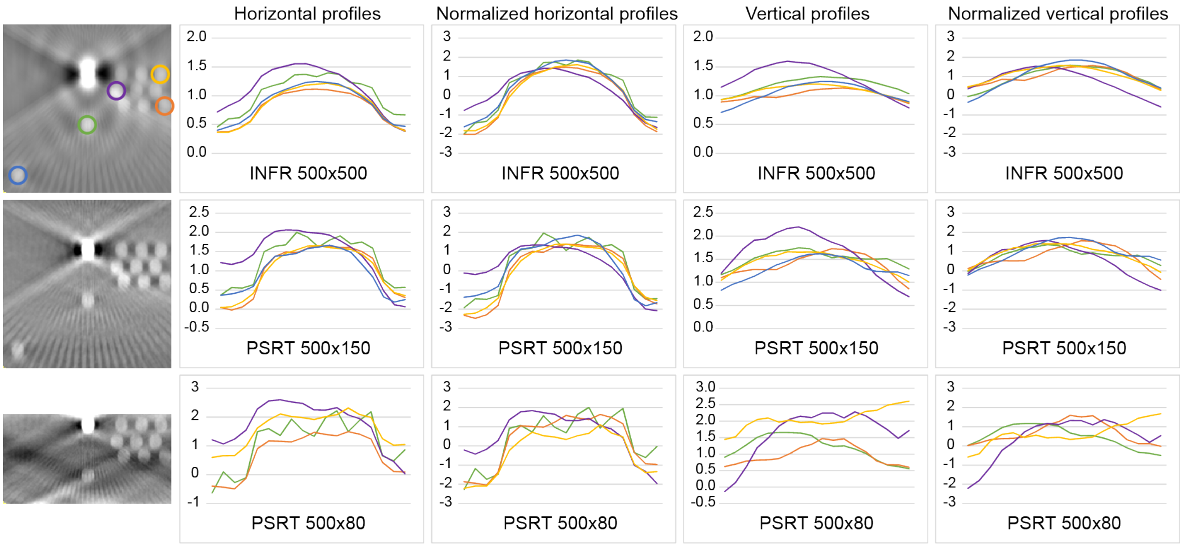


Figure 10.7: Comparison of particle values for different techniques in the presence of MW artifacts. The phantom of size 300×150 with background intensity 0.7, 12 spheres with radius 6 and intensity 2.0 with one sphere having intensity 10 (top). (a) Reconstructed tomograms (trimmed to size 150 in x dimension to save space). (b-e) Graphs comparing values from the five particles marked in the phantom. The values correspond to the horizontal lines through particle centers from both unprocessed subtomograms (b) as well as normalized subtomograms (c) and to the vertical lines through particle centers from both unprocessed subtomograms (d) as well as normalized subtomograms (e). Note that the ranges for unprocessed subtomograms differ while for the normalized subtomograms have the same range.



(a) (b) (c) (d) (e)

Figure 10.7: (Continued)

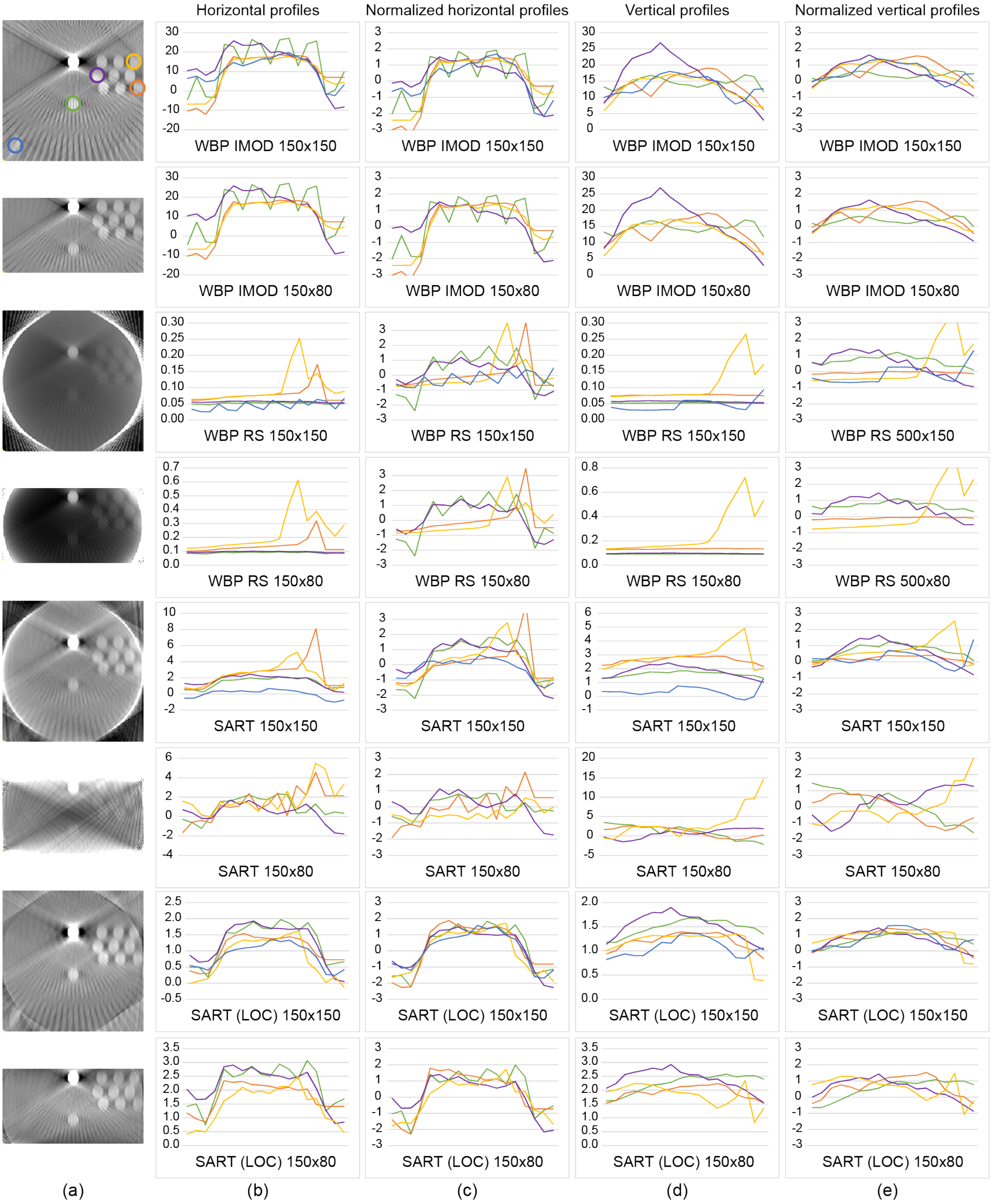
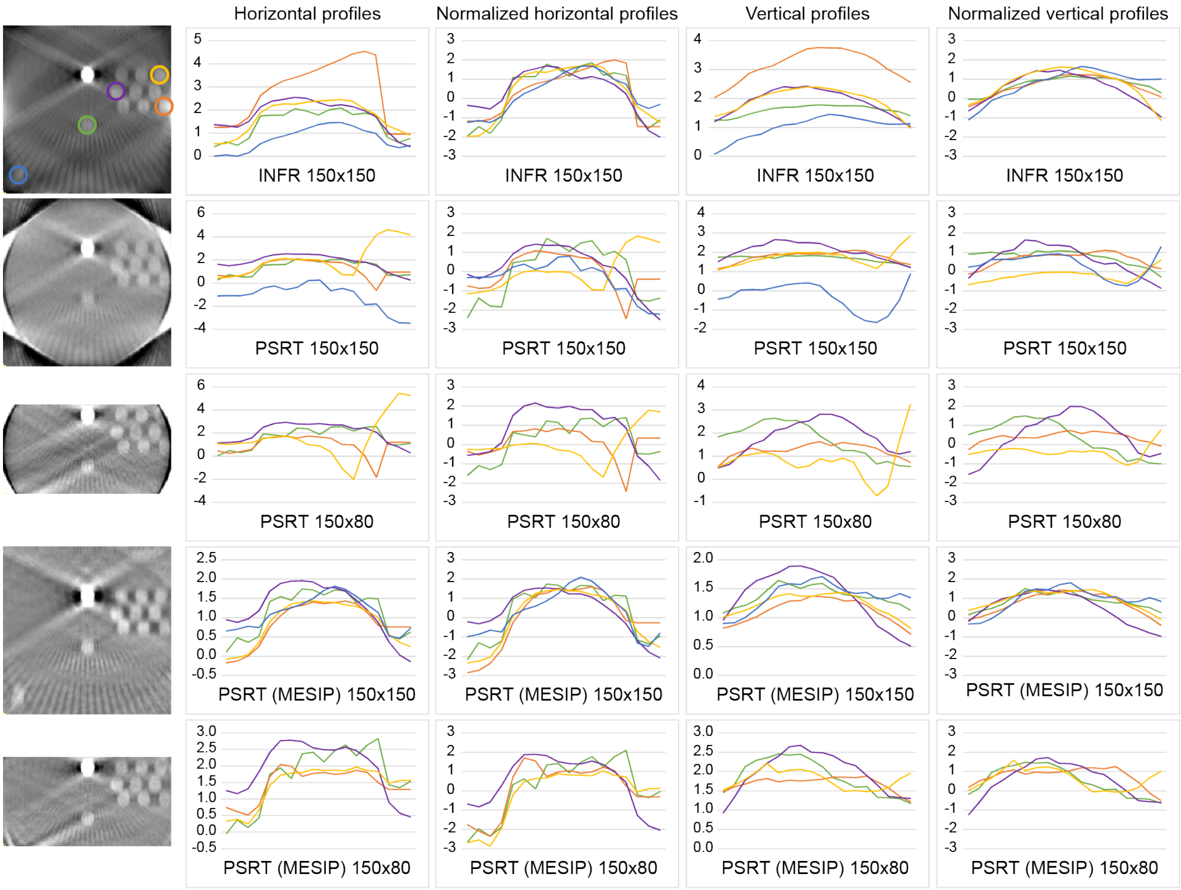


Figure 10.8: Comparison of particle values for different techniques in the presence of MW, IP, and (based on the depth) MI artifacts. (a) Reconstructed tomograms (trimmed to size 150 in x direction to save space). (b-e) Graphs comparing the voxel values from the five particles marked in the WBP IMOD reconstruction. The values correspond to the horizontal lines through particle centers from both unprocessed subtomograms (b) as well as normalized subtomograms (c) and to the vertical lines through particle centers from both unprocessed subtomograms (d) as well as normalized subtomograms (e). Note that the ranges for unprocessed subtomograms differ while for the normalized subtomograms have the same range.



(a) (b) (c) (d) (e)

Figure 10.8: (Continued)

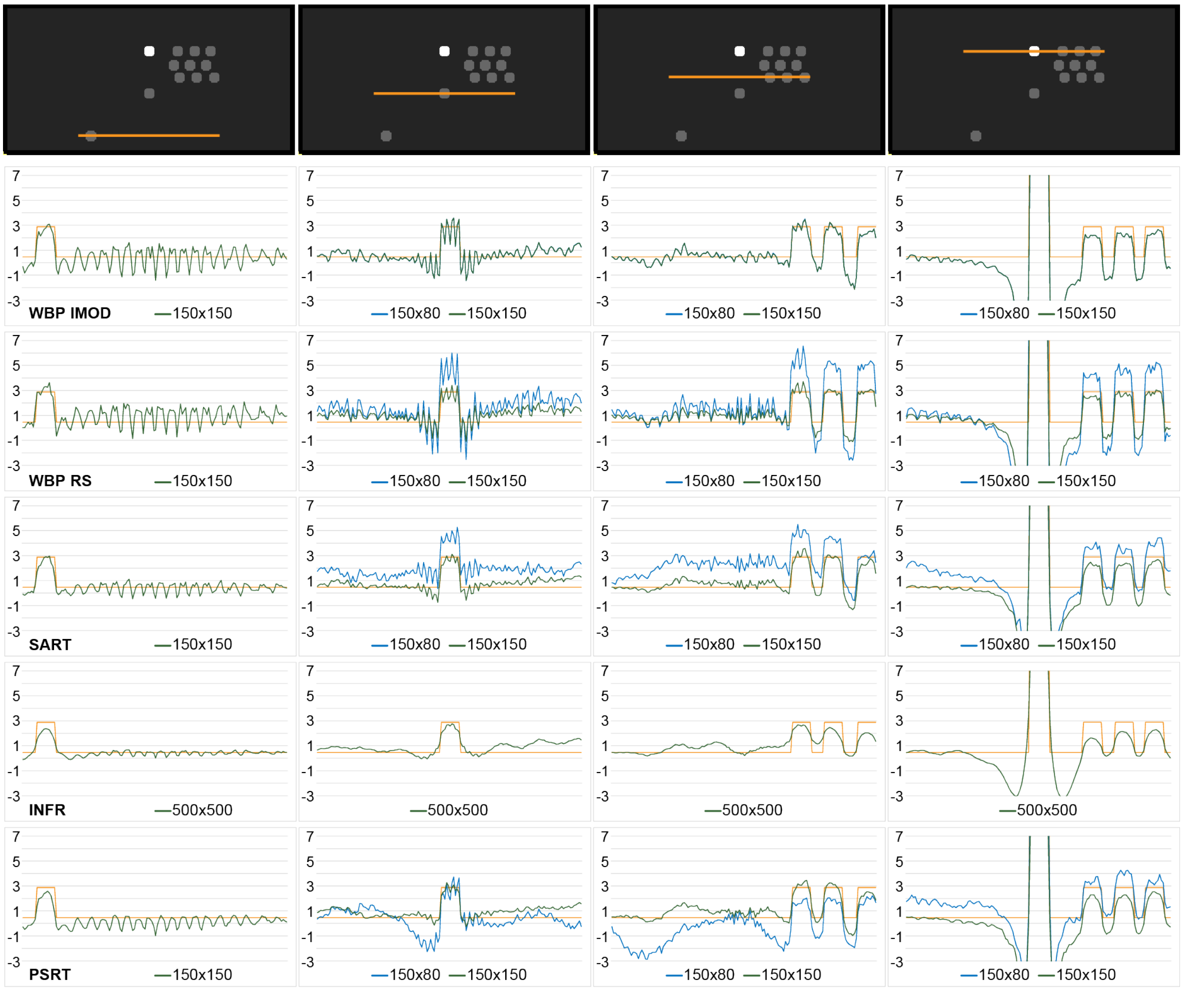


Figure 10.9: Comparison of different techniques w.r.t. reconstructed values in the presence of MW problem. Each graph compares values from the ground-truth (shown in the first row) to the values in reconstructed tomograms with sizes 500×80, represented by blue color, and 500×150 (500×500 for INFR), represented by green color. The values correspond to the respective lines depicted in the ground-truth in the first row.

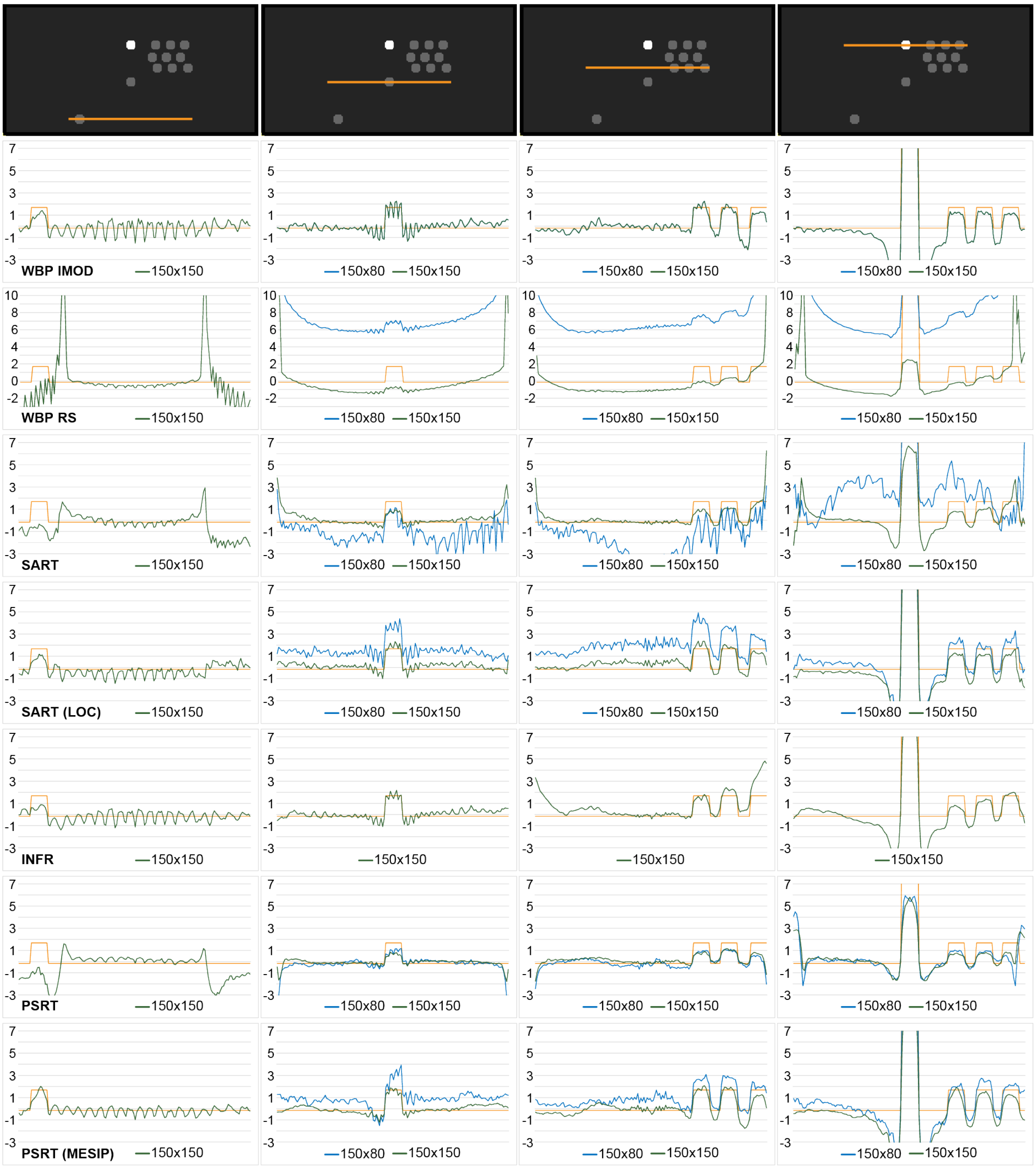


Figure 10.10: Comparison of different techniques w.r.t. reconstructed values in the presence of MW, MI, and IP problem. Each graph compares values from the ground-truth (shown in the first row) to the values in reconstructed tomograms with depth 150×80, represented by blue color, and 150×150, represented by green color. The values correspond to the respective lines depicted in the ground-truth in the first row.

10.5.5 Experimental Data

We used experimental tilt-series from EMPIAR databse (DOI: 10.6019/EMPIAR-10009) consisting of 51 projections from −60.39° to 60.97°. This dataset demonstrates all three geometric errors described in the previous sections, which is typical for most data in cryoET. We aligned the tilt-series and downsampled it to size 512×512. We performed additional reconstructions using SART with LOC and PSRT with MESIP to show their effect on the experimental dataset. To test filtering effects on geometric based methods, we also performed SART reconstructions from the input projections filtered in Fourier space prior the reconstruction.

We reconstructed volumes of sizes 512×512×50 and 512×512×512 to show the difference between artifacts caused by the interior problem and the missing information problem. The results are shown in Fig. 10.11. For SART with LOC and PSRT with MESIP see Fig. A.3 in Supplementary material and for the filtered SART see Fig. A.7 in Supplementary material. The bright artifacts present on the reconstructions with depth 50 are caused by the interior problem (Fig. 10.11a and Fig. 10.11b). SART reconstruction has particularly strong artifacts, which is to be expected as the error is accumulating in each iteration. The bright interior problem artifacts are completely suppressed by the dark vignetting artifacts caused by the MI problem in the reconstructions with depth 512 (Fig. 10.11c and Fig. 10.11d). The filtering in the Fourier space applied in IMOD WBP suppresses even these vignetting artifacts, making them hardly visible. On the other hand, the filtering in real space in RS WBP makes the artifacts look brighter than the rest of the tomogram. This is caused by the pronounced edges of filtered input projections that are bright after applying the convolution with Ram-Lak filter. Similar effect can be observed, rather surprisingly, on SART reconstruction with LOC, where the vignetting ring is brighter as well.

Both types of artifacts further complicate automatic processing of the tomograms, since they have to be either trimmed or their intensity windows manually adjusted. While the interior problem artifacts can be successfully suppressed by using a proper compensation or extension (see Supplementary material, Sec. A.2 and Sec. A.3), the missing information problem cannot be entirely avoided, especially for tomograms with larger sizes in the z dimension. However, it seems that filtering in Fourier space can make these artifacts less visible in the terms of intensity/contrast and that even for the geometric based methods (see Supplementary material, Sec. A.4).

10.6 Conclusions

In this chapter we presented an overview of three different geometric problems that arise in cryoET and showed their impact on the quality of reconstructed tomograms. We demonstrated that even if some artifacts resulting from these problems cannot be avoided there are ways to decrease their negative influence on tomograms.

The MW artifacts affects the whole tomogram and currently there are no SW based solutions how to avoid them on the tomogram level without additional prior information. Although in SA we can average out the effects of MW in the final structure, the distortions still complicate automatic steps within the SA pipeline such as automatic localization and alignment. Additionally, our results emphasize that not only the shape distortions are problematic for SA, but also the intensity changes caused by the presence of objects with high intensities (such as gold markers) that might cause severe changes in the voxel values of the surrounding particles (see Fig. 10.7c and Fig. 10.7d)). Avoiding the particles in close proximity of high intensity objects might improve the resolution of the final structure.

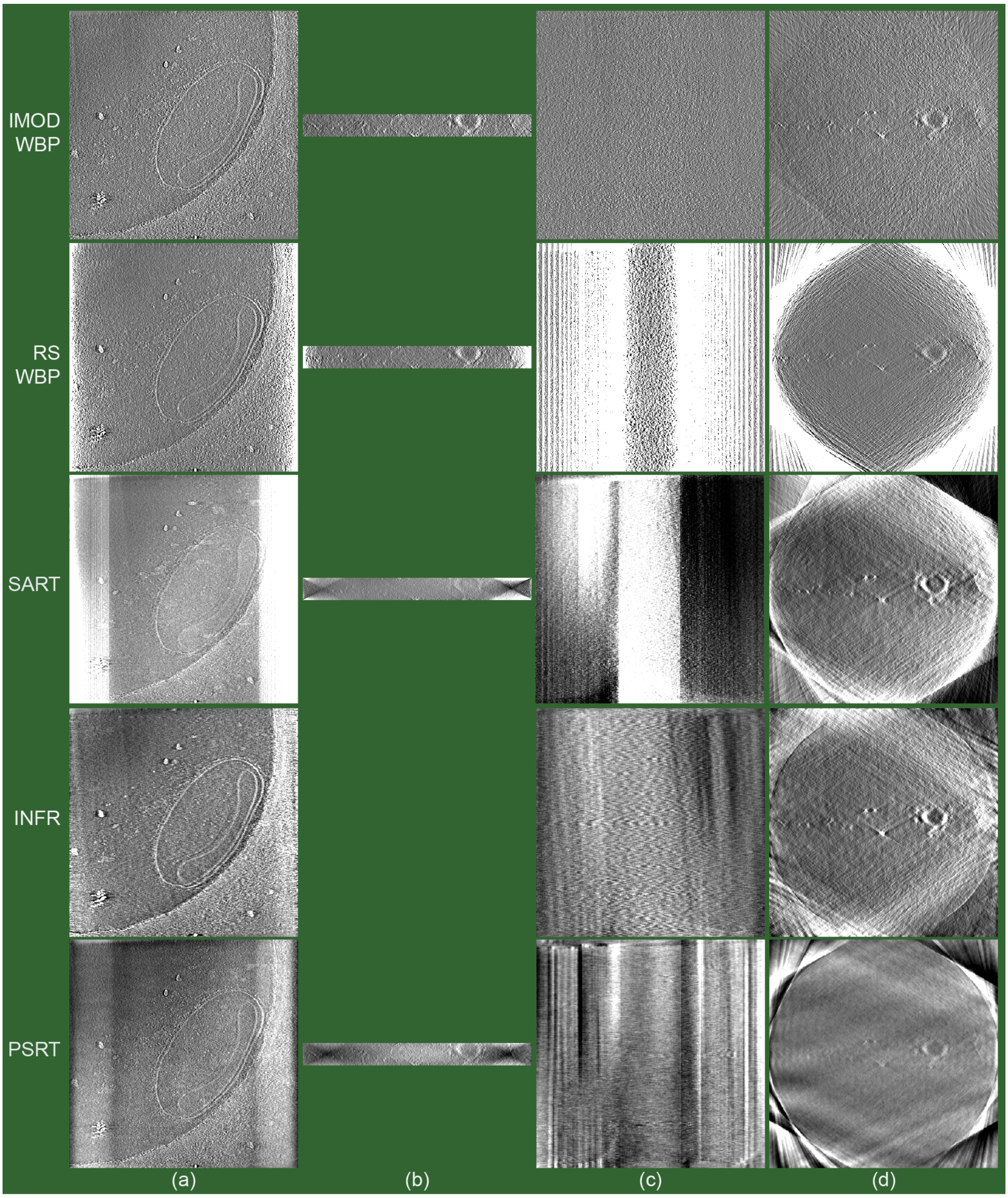


Figure 10.11: Comparison of different reconstruction techniques on the experimental dataset. The input projections were of the size 512×512 and the reconstructed area had size 512×512 with depth 50 (a-b) and 512 (c-d). For the depth of 50, the MI artifacts are covering only a very small portion of the tomogram and can be neglected. All the artifacts visible in some of the tomograms are caused by the IP. These artifacts are drowned out by the artifacts caused by MI in the reconstructions with the depth of 512. (a) Slice 50 in xy plane. (b) Slice 118 in xz plane. (c) Slice 512 in xy plane. (d) Slice 118 in xz plane.

Similarly to MW, the MI artifacts cannot be entirely avoided. However, they influence only a relatively small part of a tomogram causing vignetting artifacts and lower resolution in the affected regions. Depending on the reconstruction method, the vignetting artifacts might have relatively high intensities and we showed in Sec. 10.5.4 that in that case they have a negative influence also on the values outside the affected regions. This poses problems both for general tomography, where it complicates any subsequent processing of a tomogram, and for SA, where it might complicate the localization of particles. On the other hand, the tomograms without the strong vignetting do not pose significant constraints on tomogram processing and it seems that even the particles within the areas of lower resolution (i.e. the corners of the tomograms) can be recovered reasonably well in that case (see Fig. 10.8). However, it is important to keep in mind that the affected areas do contain less reliable information since fewer projections contributed to their reconstruction. While this might not seem as much of a problem for SA, where one basically averages it out, it is important to keep in mind that the resolution of the final structure is limited by the quality of individual subtomograms. Therefore, it might be better to ignore the low-resolution regions in further processing.

The strong intensities caused by IP have similar effect on tomograms and their values as the strong vignetting caused by MI and therefore also similar consequences. However, there are ways how to suppress them. We showed in Sec. 10.6 and Sec. 10.5.5 that although the artifacts caused by IP are visible in all geometric based methods, they are more pronounced in the case of iterative approaches (including the Fourier-based INFR from PyTom), as the high values are accumulating at the tomogram's edges in each iteration. We tested two approaches for suppressing/avoiding the IP artifacts called LOC and MESIP and we confirmed they both perform very well on phantoms as well as on experimental dataset. Although MESIP provides an accurate solution to IP, it prolongs the reconstruction time and its integration to different methods is not straightforward. LOC, on the other hand, performs fast and can be easily integrated as a preprocessing step to any method – geometric or Fourier-based.

We showed that, regardless of the reconstruction technique, filtering in Fourier space makes artifacts caused by MI and IP less pronounced, thereby improving the overall contrast in tomograms. The filtering in Fourier space can be done as a preprocessing step for any geometric-based method, especially if the goal is to reconstruct tomogram with larger depth or if no compensation/solution for IP is available. In those cases, the filtering might lead to significant brightness reduction of the MI and IP artifacts. However, the filtering does not remove the artifacts completely and even if they are hardly visible, it is important to be aware of them as they make certain parts of the tomogram less reliable (i.e. the corners in case of MI).

We showed in Sec. 10.5.4 how the depth of the reconstructed area influences the values in tomograms for geometric-based methods and that choosing an accurate depth leads to more accurate values of objects in tomogram. This is of special importance for SA, as varying values of same/similar objects across different tomograms cause problems if one wants to combine the information from these tomograms. In order to reconstruct values that faithfully represent the particles within the specimen, the reconstruction depth should correspond to the actual thickness of the scanned specimen. This might lead to almost cubic reconstructions making this approach not feasible. A viable alternative is to choose the depth of each tomogram one wants to combine so that the values of the structures within the tomograms are within the same range.

In addition to the state-of-the-art methods, we tested two new approaches: INFR (PyTom's implementation) and PSRT. For INFR the contrast of the experimental dataset was among the best. Currently, the main disadvantage of this method, as implemented in PyTom, is its ability to reconstruct only cubic volumes. PSRT showed some interesting properties on the MW (Fourier spectra) and MI (hardly visible vignetting artifacts on phantoms even without filtering) and also provides MESIP for IP. The tomograms from experimental dataset has contrast comparable to INFR. Currently, the main disadvantage of PSRT is its computational time and complicated parameter setup. However, both these methods are new and their shortcomings can be overcome in the future.

Finally, we showed that geometric as well as Fourier based methods have both their advantages and disadvantages w.r.t. different problems present in cryoET and specifically that both these types of algorithms may require different settings to perform optimally. We believe that using/combining both approaches based on specific needs is the best way how to achieve the best possible results in the field of cryoET.

CHAPTER 11

CONCLUSIONS

As mentioned numerous times, cryoET plays an important role in the field of Structural Biology as it is able to study the intracellular distributions of macromolecules in their close to native environment. It is often combined with high-resolution protocols such as Subtomogram Averaging (SA) to obtain structures of individual complexes from the tomograms. Essential part of cryoET is the reconstruction process that provides us with the 3D model of a specimen's scattering potential. As shown, the reconstruction in cryoET faces many challenges as the input projections suffer from very low Signal-to-Noise Ratio (SNR) and the final tomograms contain numerous artifacts caused by the single-tilt acquisition geometry. The state-of-the-art methods deliver noisy and low-contrast reconstructions, which complicates visual analysis of the tomograms as well as their automatic processing.

The aim of this work was to improve the quality of tomograms used in cryoET. We presented a novel iterative approach for tomographic reconstruction that was designed to match the specific properties present in cryoET and showed that it surpasses the state-of-the-art methods by delivering tomograms of better quality both in terms of SNR and automatic processing during the execution of the SA pipeline. Moreover, we performed a systematic study on geometric artifacts that are present in cryoET tomograms and on their influence on the quality of recovered information. In this final chapter, we summarize our main contributions, discuss possible follow-ups to our work and contemplate perspectives of tomographic reconstruction in cryoET.

11.1 Progressive Stochastic Reconstruction Technique

In Chapter 5 we presented basic principles of our novel iterative reconstruction approach called Progressive Stochastic Reconstruction Technique (PSRT). The method is based on Monte Carlo minimization and a sampling strategy inspired by the Metropolis-Hastings sampling strategy and thus considerably differs from the state-of-the-art methods used in the field. We tailored the method to fit the specific conditions present in cryoET datasets by designing a progressive scheme that helps to reduce noise and enhances the contrast, by proposing the Memory Efficient Solution to the specimen-level Interior Problem (MESIP) and finally, by exploiting the locality of our method to perform region of interest (ROI) reconstruction using additional importance sampling.

We revealed the implementation details of PSRT in Chapter 6, where we also discussed numerous points that have to be considered in order to ensure the correctness of the method as well as its efficiency.

We showed the properties of PSRT on both synthetic and real datasets in Chapter 7. We first proved the necessity of the progressive scheme for cryoET reconstructions both in terms of quality and computational time and showed its positive influence on noise reduction as well as on contrast improvement. We then validated our method on a synthetic dataset and compared it to state-of-the-art methods showing that the results are comparable or better, both in terms of visual quality and measured error. Furthermore, we demonstrated how the use of the MESIP removes all artifacts associated with the interior problem. We designed an experiment to show that locality of PSRT can be efficiently used to perform ROI reconstructions and demonstrated that the results obtained using our ROI scheme are superior to those obtained with the state-of-the-art methods. Finally, we discussed the computation time of PSRT as well as its rather complex parameter setup and reviewed the possibilities for their improvement.

In Chapter 8, we successfully integrated the method into a SA pipeline and demonstrated on experimental data not only that the recovered structure of the ribosome is correct but also showed that improved contrast at the early stages of the reconstruction process significantly improves the automatic template-based localization of the structures present in the tomograms. Thus, in terms of localization properties, our results surpassed those obtained with state-of-the-arts methods. Furthermore, we again demonstrated the ability of PSRT to focus more on specific parts of the volume during the reconstruction process by using the localization results to drive the importance sampling over the identified ROIs. This is of special importance in case of SA, as one is interested in individual structures that are distributed within the whole volume but represent only a small portion of it. Thus the ability to perform more local reconstructions can simplify further analysis of the tomograms as well provide significant speed-up. Note that the state-of-the-art methods completely lack this ability and every part of the volume is reconstructed with the same procedure.

Finally, in Chapter 9 we proposed the parallelization of PSRT based on splitting the reconstruction problem into smaller parts. We demonstrated the speed-up achieved by this approach and successfully validated its correctness on experimental datasets. We also briefly discussed the possibility of additional parallelization of PSRT that could be done by simultaneous processing of the random walks.

Based on our promising results, we believe that PSRT could become a viable option for tomographic reconstruction in cryoET especially, in conjunction with subsequent segmentation and/or automated analysis. We believe that its good localization properties could lead to fully-automatic processing during the SA. This is of a special importance, as the amount of tomograms that have to be processed to obtain a single structure can easily go up to hundreds and thus the possibility to process them without any need for manual intervention can significantly reduce the processing time.

Currently, PSRT suffers from two major drawbacks. The first one is the relatively long time that is needed to obtain a tomogram. We have to use both the parallelization scheme and the importance sampling in order to obtain the reconstruction in a reasonable time. Despite that, the computational times of PSRT are currently not competitive and additional speed-up is thus necessary. One possibility would be to perform the random walks simultaneously as discussed in Sec. 9.2. The second drawback is the parameter setup that requires setting four parameters prior to any reconstruction. Although there is a logic behind them, it might be difficult for an inexperienced user to choose the optimal values in the first few attempts. Moreover, our experiments showed that the optimal setting depends on many factors and thus it would be hard to provide a fully automatic computation of all the parameters. Nevertheless, we believe it is possible to compute three of them automatically if a user provides additional information about the data (see Sec. 7.6.1 for more details). What remains an open issue is the scaling factor which is currently very hard to predict and although is seems there might be a connection between its optimal value and the standard deviation of the input projections, we were so far unable to derive it exactly.

PSRT is a new approach and thus there are many options for possible follow-up work. For instance, we currently use simple Gaussian function to guide the placements of the samples during the random walks. We might consider to use more complex, even non-symmetric functions that would better reflect the underlying structures within a tomogram. We might also consider using more elaborate seeding scheme based e.g. on a distribution obtained from some initial volume estimate or projection residuals. We believe, this could further speed-up the convergence and enhance the contrast of the areas with high energy. More enhanced contrast could allow for a template-free localization of particles within a tomogram based on a simple thresholding algorithm.

Currently, we formulate the reconstruction problem as a convex function. However, the Monte Carlo approach can be used to minimize also nonlinear non-convex functions as it is able to overcome the multiple-minima problem. This property allows as to use more accurate and possibly non-convex formulation of the forward operator within the same framework.

Finally, we shall note that our approach can be directly used for any tomographic reconstruction based on the parallel acquisition scheme and can be also extended to different geometric schemes albeit it might require more complex approach for the precomputation of the weighting kernels.

11.2 Study on Geometric Artifacts in CryoET

In Chapter 10 we provided a systematic study of the major artifacts caused by the Missing Wedge (MW), the Missing Information (MI), and the specimen-level Interior Problem (IP). Focusing on their manifestation in cryoET reconstructions, we illustratively described these artifacts, using a unified terminology, and explained how they artifacts arise and when they can be avoided. We performed a series of experiments to demonstrate their effects on the reconstruction quality w.r.t. classic cryoET as well as the SA. These experiments were performed both on synthetic and experimental data and different reconstruction methods were used to test their ability to suppress the artifacts.

Our results showed that although some artifacts cannot be avoided, there are ways how to decrease their negative influence on the reconstruction quality. The exact possibilities depend on the purpose of a tomogram (e.g. visual analysis or automatic processing) and on the used reconstruction method. We tested two approaches for suppressing/avoiding IP artifacts, one of them being the MESIP implemented within the PSRT framework, and we confirmed that they both perform very well on phantoms as well as on experimental datasets. We showed that, regardless of the reconstruction technique, filtering in Fourier space makes artifacts caused by MI and IP less pronounced and thereby improves the overall contrast in the reconstructions. We observed that for geometric based methods the depth of the reconstructed area influences the values in tomograms and that choosing an accurate depth leads to more accurate values of reconstructed objects. This is especially important for SA, where varying values across different tomograms might cause problems during their combination. Finally, we showed that geometric as well as Fourier based methods may require different settings to perform optimally and that both approaches have their advantages and disadvantages w.r.t. different problems present in cryoET. We concluded that combining both approaches based on specific needs might be the optimal way how to achieve the best possible results in the field of cryoET and SA.

We believe that our study could lead to improvement of cryoET tomograms by helping to better interpret the geometric artifacts and to choose an optimal way for dealing with them. We also hope that our study will bring to attention the different requirements on tomogram quality posed by the primary purpose of its intended use. One interesting follow-up to this study would be designing a series of experiments that would test the influence of varying values of the same/similar objects across a set of tomograms on different steps of the SA pipeline, ideally both on synthetic and experimental data. We believe that such a study would help to further optimize the reconstruction methods w.r.t. SA and would consequently increase the chances for a fully automatic execution of the pipeline.

11.3 Perspectives on Tomographic Reconstructions in CryoET

In Chapter 2 we reviewed the basic principles of cryoET pipeline and outlined the main challenges we face in order to obtain valuable information on the structures of the studied macromolecules. In the past few years, major technical advances of TEM as well as specimen preparation techniques led to significant quality improvement of the acquired images and consequently to higher resolution of the reconstructed tomograms. The application of high-resolution protocols such as SA pushed the achievable resolution even higher, with the best resolution being currently around 8Å. As mentioned previously, the results of SA and similar protocols strongly depend on the quality of the provided tomograms. It is therefore rather surprising that there were only relatively few attempts in the recent years to improve the quality of tomograms either by adapting the existing reconstruction methods to specific cryoET conditions or by designing new approaches. The current algorithmic development focuses more on optimizing the preprocessing of the input projections (especially in form of CTF correction) and on improving the postprocessing steps, such as automatic segmentation or individual steps of the SA pipeline.

In this work, we showed that there is still space for improvement of tomographic reconstruction itself in cryoET and that the enhanced quality, especially in terms of contrast, allows for more reliable automatic processing of tomograms. To have tomograms that would allow for a reliable and fully automatic execution of the SA pipeline is of utmost importance, since the best resolutions in the field were achieved by averaging together not hundreds but thousands of subtomograms. It is thus our believe that more effort should be put into the development of new reconstruction approaches or the improvement of the existing ones. Currently, the evaluation of a new method is very difficult as there are no reliable metrics to asses the quality of reconstructed tomograms or their resolution. Consequently, any new method has to be validated by a comparison to the state-of-the-art methods which is very time consuming and in the end provides us only with relative results. A partial solution to this problem could be establishing a database with various phantoms, designed to test different properties of tomograms, as well as with representative set of experimental data. The results obtained by different reconstruction techniques would be part of the database as well, which would allow for easy and fast comparisons towards existing techniques. It would be useful not only for the quality assessment of newly developed techniques but could also help a user to choose an optimal reconstruction method based on its requirements. We believe that such a database would contribute to faster development of new techniques and more objective assessment of their quality and thus ultimately to better resolution of the recovered macromolecular structures.

APPENDIX A

SUPPLEMENTAL MATERIAL FOR GEOMETRIC ARTIFACTS IN CRYOET

A.1 Fourier Spectra for MW Reconstructions

We performed the Fourier transform of tomograms presented in Sec. 10.5.1 to demonstrate the missing wedge problem also in Fourier space. The resulting Fourier spectra are shown in Fig. A.1.

A.2 Extensions for IP

To show the effects of the extension approaches described in Sec. 10.3, we performed SART and PSRT reconstructions with the following dimensions: 512×512×50, 692×512×50 (this extension corresponds to Eq. 10.1) and 1024×512×50 (we consider 1024 to represent the ideal extension as depicted in Fig.10.3c right). We also perform one reconstruction of size 512×512×50 using LOC/MESIP 10.3. The Fig. A.2 shows 257th slice in xz plane of the tomograms. It is clearly visible, that the minimal extension is sufficient to suppress the artifacts out of the area of interest, and with even greater extension the artifacts disappear. It is also visible that the quality of the information in the extended areas is low but this fact does not influence the quality of the region of interest. Finally, we can observe that the LOC and MESIP approaches deliver results very similar to the extended reconstructions. The problem with the extension lies that both processing time and memory requirements increase with the increasing size of the reconstructed area.

A.3 Additional Interior Problem Results

We compared two different approaches for IP that were mentioned in Sec. 10.3. The first approach was designed for iterative methods by Xu et al. [2010] and is called Long Object Compensation (LOC). This approximate solution to IP uses additional weighting of the input projections that is based on the ratio of ray's length passing through a sample to a ray's length passing through the reconstructed

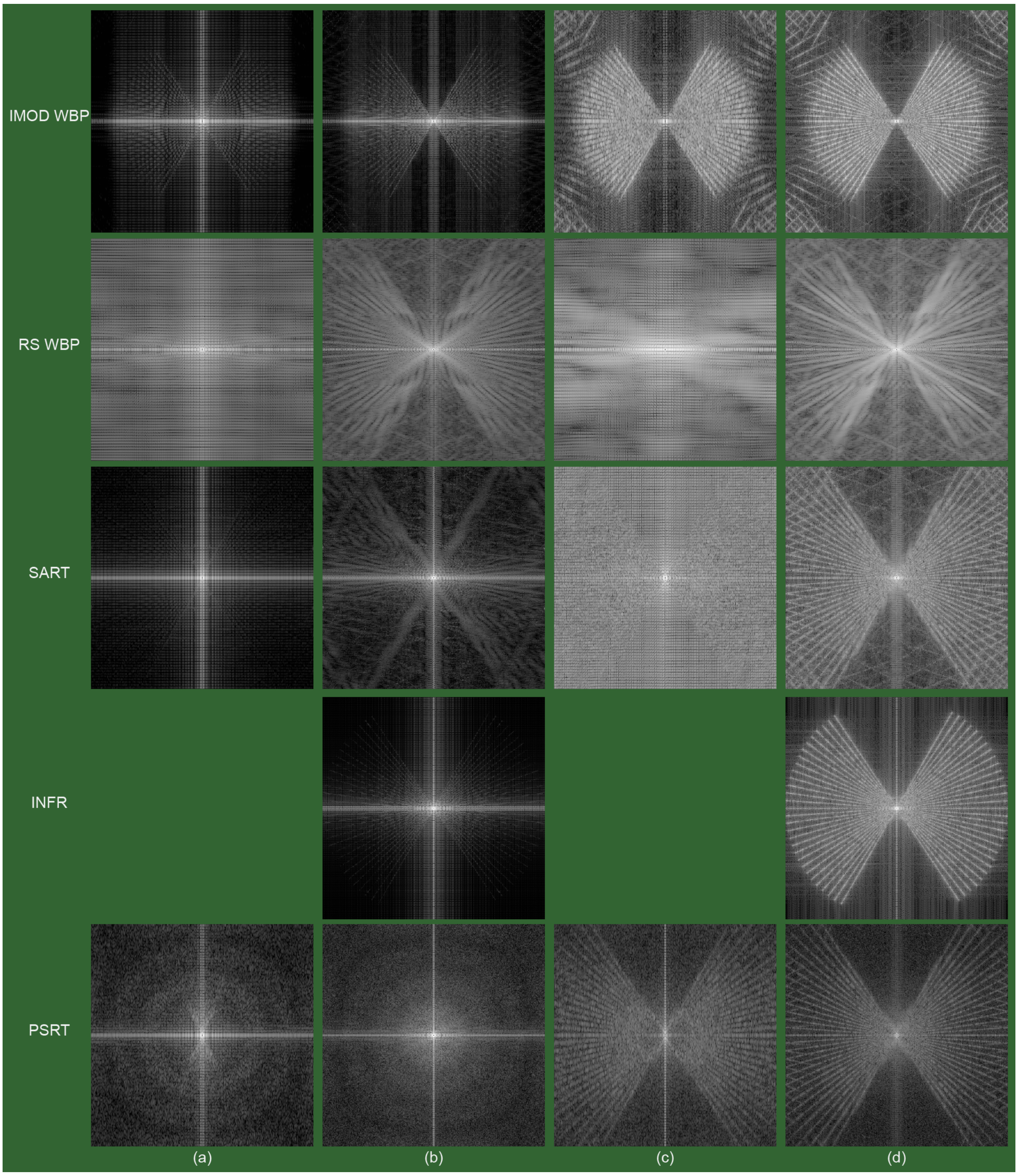


Figure A.1: Fourier spectra of tomograms showed in Fig. 10.4. (a) 300×80 reconstruction from incomplete noiseless projection set with 30 projections with angles from −60° to +56°. (b) 300×200 (300×300 for INFR) reconstruction from the same projection set. (c) 300×80 reconstruction from the incomplete projection set with Gaussian noise with SNR 10. (d) 300×200 (300×300 for INFR) reconstruction from the incomplete projection set with Gaussian noise with SNR 10.

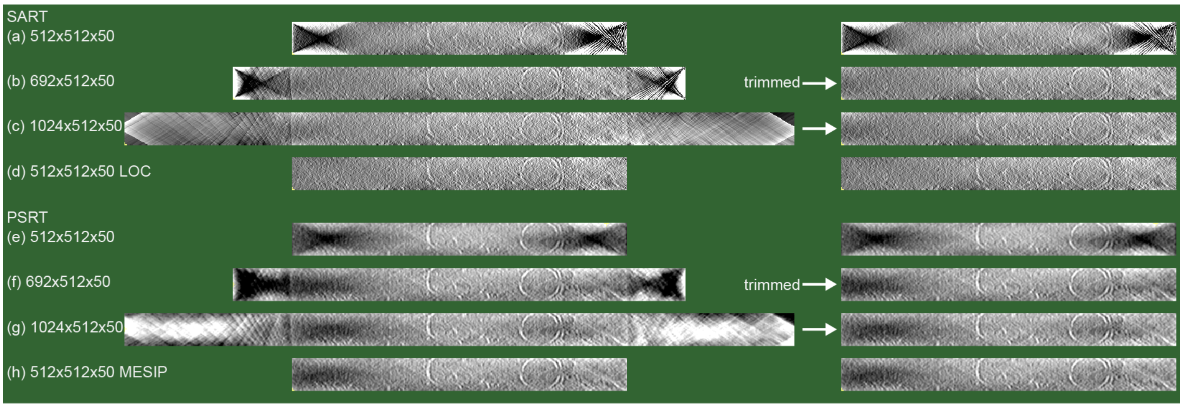


Figure A.2: Comparison of different approaches for suppressing IP artifacts on tomogram edges for SART and PSRT. Each figure depicts central slice in xz plane. (a) SART tomogram with IP artifacts on the edges. (b) SART reconstruction extended in x dimension by the minimal amount necessary to avoid the artifacts according to Eq. 10.1. (c) SART tomogram with the extension in x dimension that approximates the ideal extension from Fig. 10.3c right. (d) SART tomogram with applied LOC. (e-g) Equivalent reconstructions obtained by PSRT. (h). SART tomogram with applied MESIP.

area. The LOC is a part of the Ettention framework and although it is similar to the one suggested by Xu et al. [2010], here the compensation is done in each iteration on the virtual projections instead of on the input projections. The second approach was designed within the PSRT method and is called Memory Efficient Solution for IP (MESIP). It extends the reconstructed area by an amount necessary to avoid IP artifacts but this extension is only virtual, i.e. the extended parts are processed normally but they are not stored and hence the memory requirements do not increase. The MESIP approach delivers more accurate reconstructions than LOC but is based on a local character of PSRT method and thus its integration to other geometric based methods is not straightforward.

A.4 Filtering

The filtering plays a significant role in the visual appearance of the MI and IP artifacts. To show the influence of filtering on the quality of reconstructions, we performed additional experiments using simple Back Projection (BP), i.e. without any filtering, and Real Space WBP (RS WBP) with filtering done both in Fourier and real space. The filtering in Fourier space was done by multiplication of the Fourier transform (FT) of input projections with the FT of the Ram-Lak filter. The filtering in real space was done by convolution of the input projections with the Ram-Lak filter. We also compared the results to the reconstructions obtained with IMOD WBP (i.e. both filtering and reconstruction performed in Fourier space). The reconstructions for different geometric problems are shown in Fig. A.4, Fig. A.5, and Fig. A.6. It is clearly visible that the MI and IP artifacts are less pronounced if the filtering is done in Fourier space. It also shows that the difference between the intensities of RS WBP and IMOD WBP as presented in Fig. 10.5, Fig. 10.6, and Fig. 10.11 was caused by filtering in different space and not by performing reconstruction in different space.

To show the effect of filtering in Fourier space in combination with iterative methods, we performed additional SART reconstruction on experimental dataset that was filtered with a Ram-Lak filter in Fourier space. The comparison to non-filtered SART (both with and without LOC) is shown in Fig. A.7. The filtering alone partly suppresses the IP artifacts and also helps with the contrast both in the case of the IP and the MI artifacts. To remove the IP artifacts completely, LOC had to be applied as well.

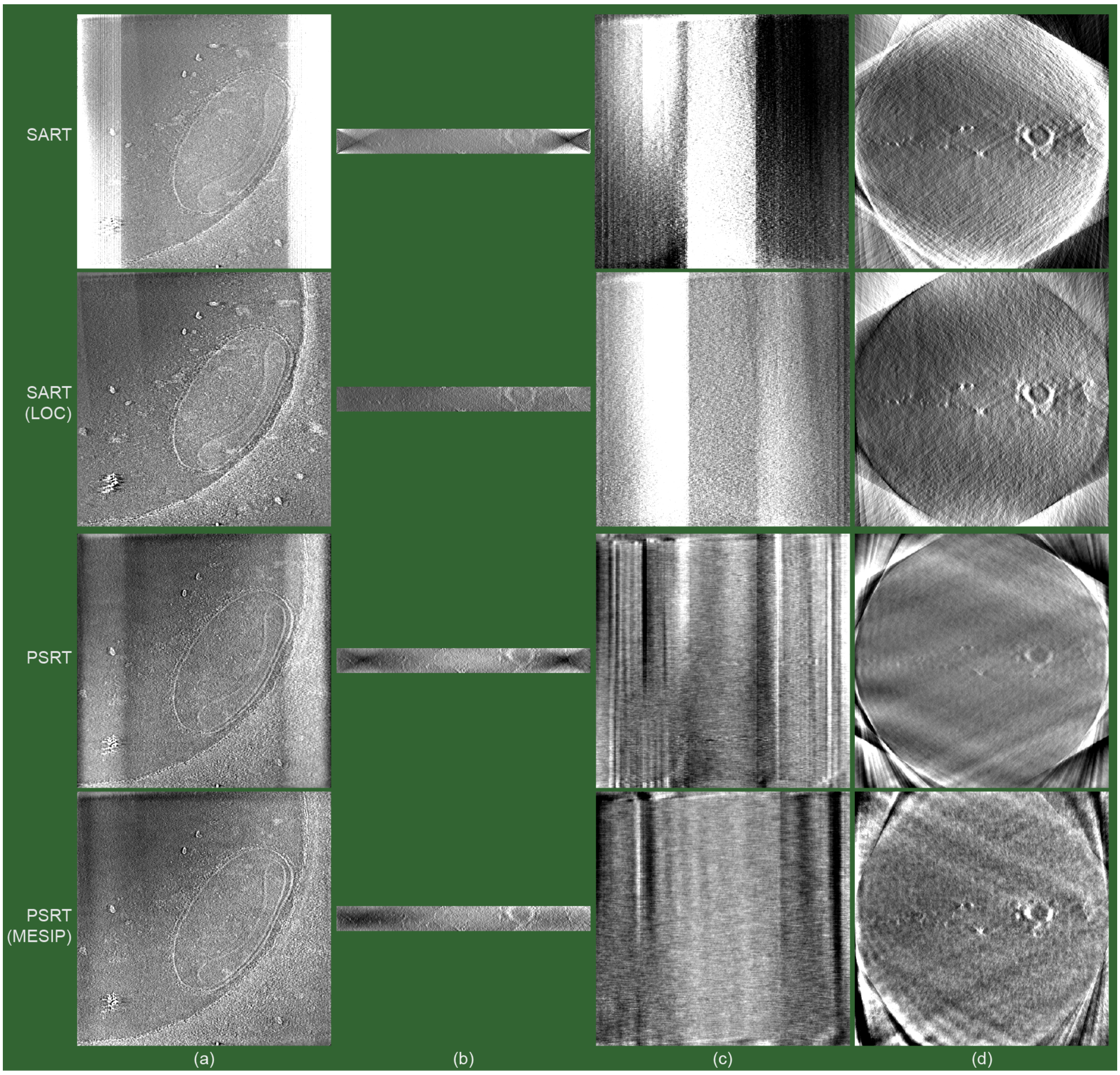


Figure A.3: Comparison of SART and PSRT approaches to IP performed on the experimental dataset. The input projections were of the size 512×512 and the reconstructed area had size 512×512 with depth 50 (a-b) and 512 (c-d). The first two rows compare SART with and without the Long Object Compensation (LOC). Although IP is only an approximate solution to IP, the artifacts associated with the IP problem are suppressed very well. The PSRT employs Memory efficient solution to IP (MESIP) which allows to extend a reconstructed area by an amount necessary to avoid the artifacts without increasing memory requirements. As shown in the last row, using this approach the artifacts caused by IP are completely avoided. (a) Slice 50 in xy plane. (b) Slice 118 in xz plane. (c) Slice 512 in xy plane. (d) Slice 118 in xz plane.

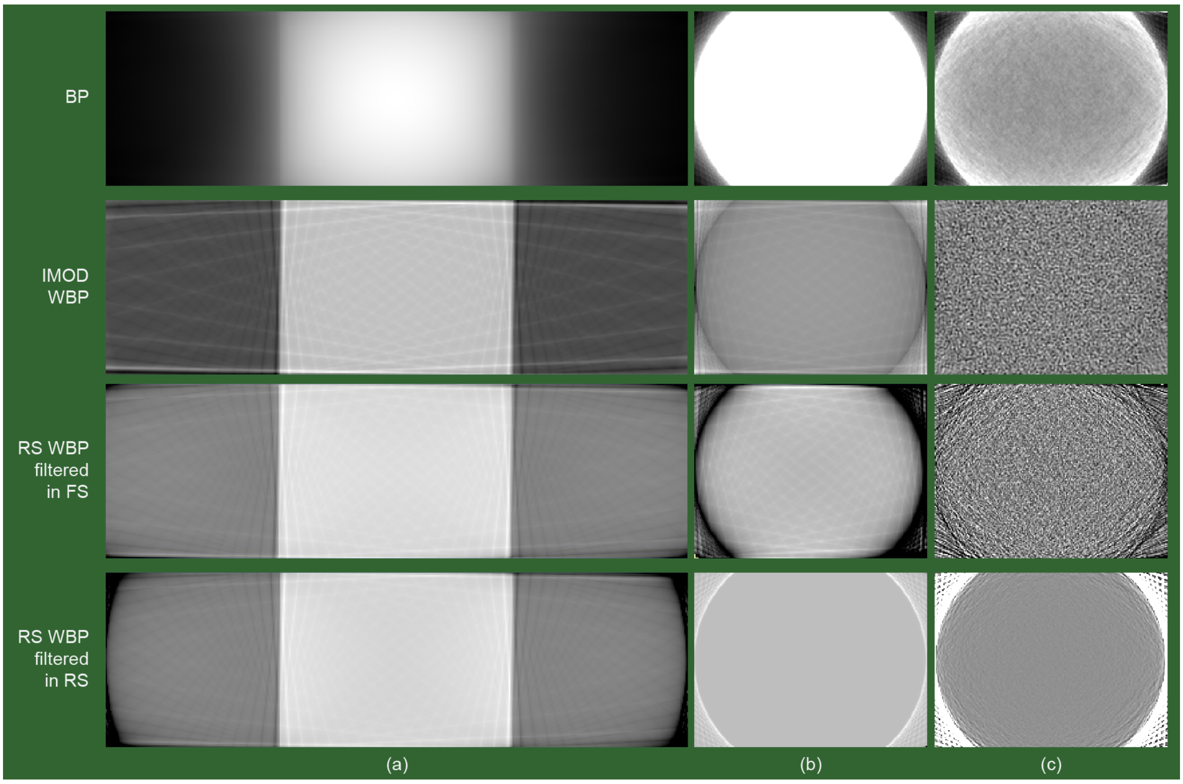

BP
IMOD WBP
RS WBP filtered in FS
RS WBP filtered in RS
(a)
(b)
(c)

Figure A.4: Comparison of reconstructions obtained by simple Back Projection (BP), IMOD WBP, and RS WBP from projections filtered in Real Space (RS) and Fourier Space (FS) in the presence of MI problem. (a) Reconstructions of size 500×150 from projections of size 500×1. (b) Reconstructions of size 200×150 from noiseless projections of size 200x1. (c) Reconstructions of size 200×150 from projections of size 200×1 with additional Gaussian noise with SNR 10.

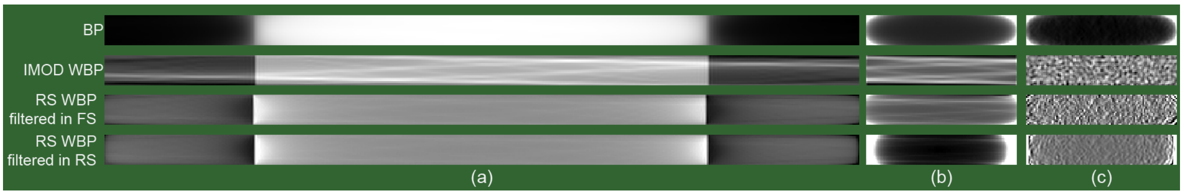

BP
IMOD WBP
RS WBP filtered in FS
RS WBP filtered in RS
(a)
(b)
(c)

Figure A.5: Comparison of reconstructions obtained by simple Back Projection (BP), IMOD WBP, and RS WBP from projections filtered in Real Space (RS) and Fourier Space (FS) in the presence of IP problem. (a) Reconstructions of size 500×20 from projections of size 500×1. (b) Reconstructions of size 100×20 from noiseless projections of size 100x1. (c) Reconstructions of size 100×20 from projections of size 100×1 with additional Gaussian noise with SNR 10.

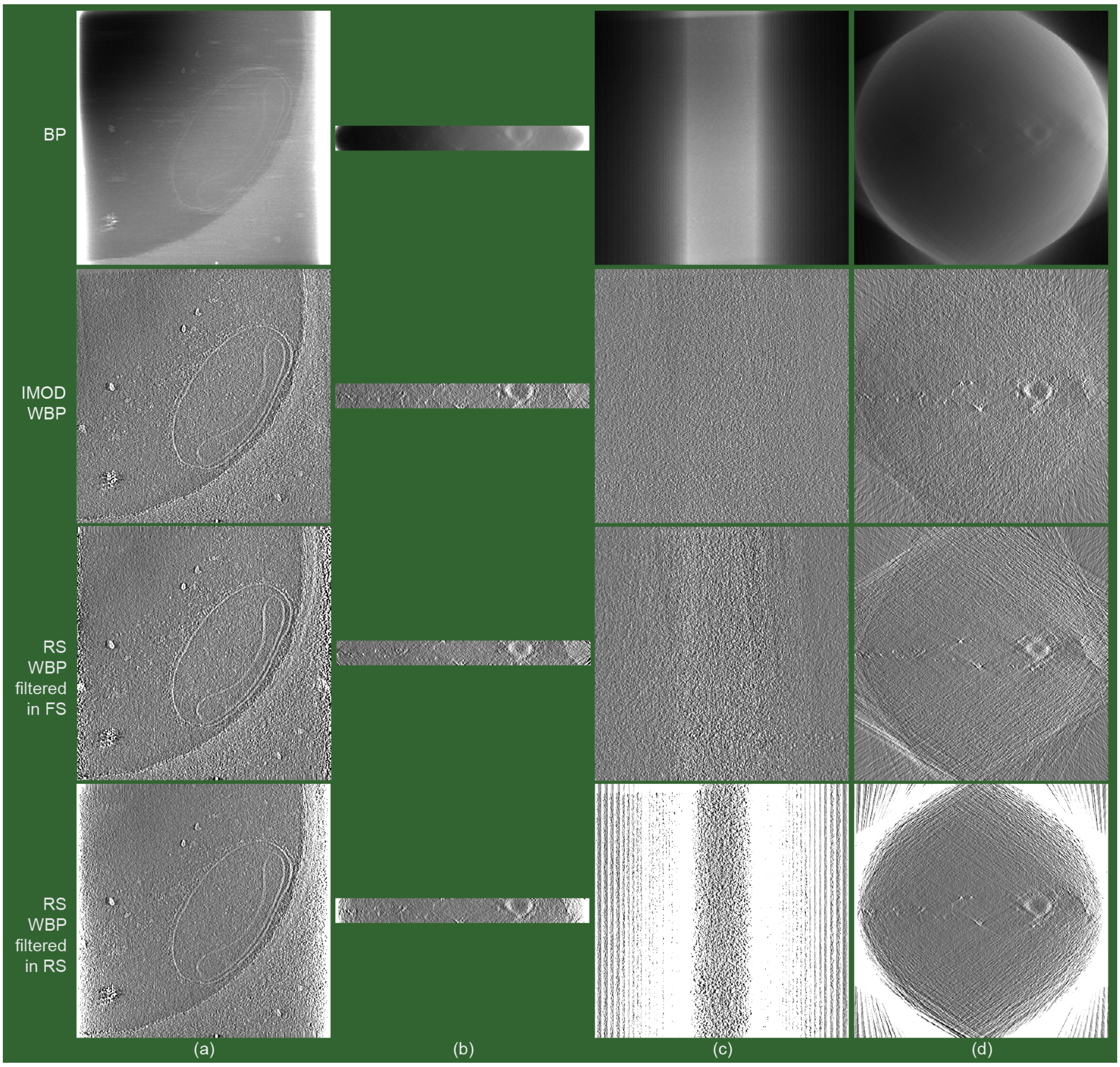

BP
IMOD WBP
RS WBP filtered in FS
RS WBP filtered in RS
(a)
(b)
(c)
(d)

Figure A.6: Comparison of reconstructions from the experimental dataset obtained by simple Back Projection (BP), IMOD WBP, and RS WBP from projections filtered in Real Space (RS) and Fourier Space (FS). The input projections were of the size 512×512 and the reconstructed area had size 512×512 with depth 50 (a-b) and 512 (c-d). (a) Slice 50 in xy plane. (b) Slice 118 in xz plane. (c) Slice 512 in xy plane. (d) Slice 118 in xz plane.

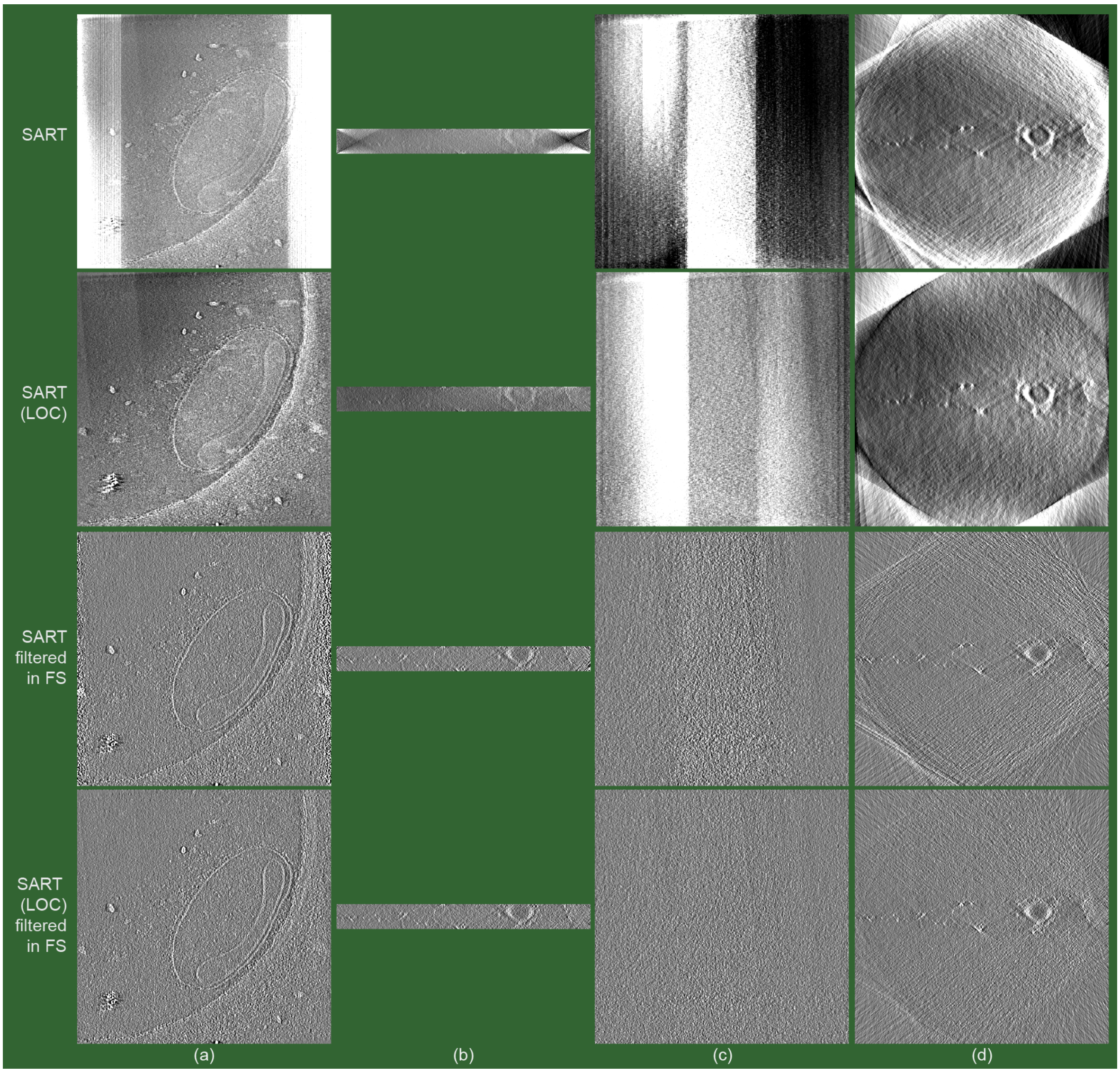


Figure A.7: Comparison of different SART reconstructions on the experimental dataset. The input projections were of the size 512×512 and the reconstructed area had size 512×512 with depth 50 (a-b) and 512 (c-d). The first two rows compare SART with and without the Long Object Compensation (LOC). The third row shows reconstruction without the LOC but with the input projections filtered in the Fourier space with the Ram-Lak filter prior the reconstruction. The filtering alone partly suppresses the IP artifacts and also helps with the contrast both in the case of the IP and the MI artifacts. To remove the IP artifacts completely, the LOC had to be applied as shown by the results in the last row where SART reconstruction is done using both filtered input projections and the LOC. (a) Slice 50 in xy plane. (b) Slice 118 in xz plane. (c) Slice 512 in xy plane. (d) Slice 118 in xz plane.

BIBLIOGRAPHY

J. I. Agulleiro and J. J. Fernández. Fast tomographic reconstruction on multicore computers. *Bioinformatics*, 27(4):582–583, 2011. 3.2.1

A. Al-Amoudi, L.P. Norlen, and J. Dubochet. Cryo-electron microscopy of vitreous sections of native biological cells and tissues. *J. Struct. Biol.*, 148:131–135, 2004. 2.1

A. Al-Amoudi, D. Studer, and J. Dubochet. Cutting artefacts and cutting process in vitreous sections for cryo-electron microscopy. *J. Struct. Biol.*, 150:109–121, 2005. 2.1

F. Amat, F. Moussavi, L. R. Comolli, G. Elidan, K. H. Downing, and M. Horowitz. Markov random field based automatic image alignment for electron tomography. *J. Struct. Biol.*, 161:260–275, 2008. 2.3.1

F. Amat, L. R. Comolli, F. Moussavi, J. Smit, K. H. Downing, and M. Horowitz. Subtomogram alignment by adaptive Fourier coefficient thresholding. *J. Struct. Biol.*, 171:332–344, 2010. 2.5.3

A. H. Andersen and A. C. Kak. Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrason. Imaging*, 6(1):81–94, 1984. 3.2.1, 3.3

R. Barbuzza and A. Clausse. Metropolis Monte Carlo for tomographic reconstruction with prior smoothness information. *IET-IPR*, 5(2):198–204, April 2011. 4.2.1, 5

R. Barbuzza, M. Vénere, and A. Clausse. Tomographic reconstruction using heuristic Monte Carlo methods. *J. Heuristics*, 13(3):227–242, June 2007. 4.2.1

D. P. Barnard, J. N. Turner, J. Frank, and B. F. McEwen. A 360° single-axis tilt stage for the high-voltage electron microscope. *J. Microsc.*, 167(1):39–48, 1992. 10.1

A. Bartesaghi, P. Sprechmann, J. Liu, G. Randall, G. Sapiro, and S. Subramaniam. Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.*, 162:436–450, 2008. 2.5.2, 2.5.3

K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. Van Tendeloo. 3D imaging of nanomaterials by discrete tomography. *Ultramicroscopy*, 109(6):730–740, 2009. 10.1

T.ăA. M. Bharat, C. J. Russo, J. Löwe, L. A. Passmore, and S. H. W. Scheres. Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure*, 23(9):1743–1753, 2015. 2.3.2, 2.5, 2.5.1

R. N. Bracewell. Strip integration in radio astronomy. *Aust. J. Phys.*, 9:198–217, 1956. 3.2.2

R. N. Bracewell and A. C. Riddle. Inversion of fan-beam scans in radio astronomy. *Astrophys. J.*, 150:427–434, 1967. 3.2.1, 3.2.2

S. Brandt, J. Heikkonen, and P. Engelhardt. Automatic alignment of transmission electron microscope tilt series without fiducial markers. *J. Struct. Biol.*, 136:201–213, 2001a. 2.3.1

S. Brandt, J. Heikkonen, and P. Engelhardt. Multiphase method for automatic alignment of transmission electron microscope images using markers. *J. Struct. Biol.*, 133:10–22, 2001b. 2.3.1

S. S. Brandt and U. Ziese. Automatic TEM image alignment by trifocal geometry. *J. Microsc.*, 222:1–14, 2006. 2.3.1

J. A. Briggs. Structural biology in situ – the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.*, 23 (2):261–267, 2013. 2.5

J.A. Briggs, J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H. G. Krausslich. Structure and assembly of immature hiv. *Proc. Natl. Acad. Sci. U. S. A.*, 106:11090–11095, 2009. 2.5

J. L. Brodsky. The use of in vitro assays to measure endoplasmic reticulum- associated degradation. *Methods Enzymol.*, 470:661–679, 2010. 8.2

G. Cardone, K. Grunewald, and A. C. Steven. A resolution criterion for electron tomography based on cross-validation. *J. Struct. Biol.*, 151:117–129, 2005. 3.3

D. Castaño-Díez, A. Seybert, and A. S. Frangakis. Tilt-series and electron microscope alignment for the correction of the non-perpendicularity of beam and tilt-axis. *J. Struct. Biol.*, 154:195–205, 2006. 2.3.1, 9.1

D. Castaño-Díez, A. Al-Amoudi, A. M. Glynn, A. Seybert, and A. S. Frangakis. Fiducial-less alignment of cryo-sections. *J. Struct. Biol.*, 159:413–423, 2007. 2.3.1

D. Castaño-Díez, M. Scheffer, A. Al-Amoudi, and A. S. Frangakis. Alignator: A GPU powered software package for robust fiducial-less alignment of cryo tilt-series. *J. Struct. Biol.*, 170:117–126, 2010. 2.3.1

D. Castaño-Díez, M. Kudryashev, M. Arheit, and Stahlberg H. Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.*, 178(2):139–151, 2012. 2.5

Y. Chen and F. Förster. Iterative reconstruction of cryo-electron tomograms using nonuniform fast Fourier transforms. *J. Struct. Biol.*, 185(3):309–316, 2013. 3.2.2, 5.4.1, 8.6, 10

Y. Chen, T. Hrabe, S. Pfeffer, O. Pauly, D. Mateus, N. Navab, and F. Förster. Detection and identification of macromolecular complexes in cryo-electron tomograms using support vector machines. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium*, pages 137 3– 1376, May 2012. 2.5.1

Y. Chen, S. Pfeffer, T. Hrabe, J. M. Schuller, and F. Förster. Fast and accurate reference-free alignment of subtomograms. *J. Struct. Biol.*, 182(3):235–245, 2013. 2.5, 2.5.3, 7.1.3, 8.2, 8.3

Y. Chen, S. Pfeffer, J. J. Fernández, C. O. S. Sorzano, and F. Förster. Autofocused 3D classification of cryoelectron subtomograms. *Structure*, 22(10):1528–1537, 2014. 2.5.2

R. A. Crowther, D. J. DeRosier, and A. Klug. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, 317(1530): 319–340, 1970. 3.2.1, 3.2.2

L. Dagum and R. Menon. Openmp: an industry standard api for shared-memory programming. *Comput. Sci. & Eng., IEEE*, 5(1):46–55, 1998. 9.1

T. Dahmen, L. Marsalek, N. Marniok, B. Turoňová, S. Bogatchev, P. Trampert, S. Nickels, and P. Slusallek. The ettention software package. *Ultramicroscopy*, 161:110–118, 2016. 3.2.1, 7.1, 8.3, 10.5

R. Danev, S. Kanamaru, M. Marko, and K. Nagayama. Zernike phase contrast cryo-electron tomography. *J. Struct. Biol.*, 171:174–181, 2010. 2.3.2

T. de Castro Martins, E. D. L. B. Camargo, R. G. Lima, M. B. P. Amato, and M. S. G. Tsuzuki. Image reconstruction using interval simulated annealing in electrical impedance tomography. *IEEE Trans. Biomed. Eng.*, 59(7):1861–1870, 2012. 4.2.1

D. J. De Rosier and A. Klug. Reconstruction of three dimensional structures from electron micrographs. *Nature*, 217:130–134, 1968. 3.2.2

K. Dierksen, D. Typke, R. Hegerl, A. J. Koster, and W. Baumeister. Towards automatic electron tomography. *Ultramicroscopy*, 40:71–87, 1992. 2.2.2

J. Dubochet, M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowall, and P. Schultz. Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophysics*, 21:129–228, 1988. 2.1

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. *J. Struct. Biol.*, 180:488–496, 2012. 2.5

D. Fanelli and O. Öktem. Electron tomography: A short overview with an emphasis on the absorption potential model for the forward problem. *Inverse Probl.*, 24(1):013001, 2008. 3.1

J. J. Fernández. High performance computing in structural determination by electron cryomicroscopy. *J. Struct. Biol.*, 164(1):1–6, 2008. 3.2.1

J. J. Fernández. Computational methods for electron tomography. *Micron*, 43(10):1010–1030, 2012. 2.3.2, 3.2.1, 3.2.2, 3.3

J. J. Fernández, S. Li, and R. A. Crowther. Ctf determination and correction in electron cryotomography. *Ultramicroscopy*, 106:587–596, 2006. 2.3.2

J. A. Fessler and B. P. Sutton. Nonuniform fast Fourier transforms using minŰmax interpolation. *IEEE Trans. Signal Process.*, 51:560–574, 2003. 3.2.2

F. Förster and R. Hegerl. Structure determination in situ by averaging of tomograms. In J. R. McIntosh, editor, *Cellular Electron Microscopy*, volume 79 of *Methods in Cell Biology*, pages 741–767. Academic Press, 2007. 1, 2.5, 2.5.3

F. Förster, S. Pruggnaller, A. Seybert, and A. S. Frangakis. Classification of cryo- electron sub-tomograms using constrained correlation. *J. Struct. Biol.*, 161:276–286, 2008. 2.5.2

F. Förster, B. G. Han, and M. Beck. Visual proteomics. *Methods Enzymol.*, 483:215–243, 2010. 2.5.1, 8.3

The MPI Forum. Mpi: A message passing interface, 1993. URL `http://www.mpi-forum.org`. 9.1

J. Frank. *Electron Tomography: Methods for Three-dimensional Visualization of Structures in the Cell.* Springer, New York, 2nd edition, 2006. 1, 2, 3.1, 3.3, 10.1

J. Frank, B. Shimkin, and H. Dowse. Spider a modular software system for electron image processing. *Ultramicroscopy*, 6(4):343–357, 1981. 8.3

J. Frikel and E. T. Quinto. Characterization and reduction of artifacts in limited angle tomography. *Inverse Probl.*, 29(12):125007, December 2013. 10

M. Fukuda, S. Tomimatsu, K. Nakamura, M. Koguchi, H. Shichi, and K. Umemura. A new FIB fabrication method for micropillar specimens for three-dimensional observation using scanning transmission electron microscopy. *J. Electron Microsc.*, 53(5):479–483, 2004. 10.1

Y. Fukuda, U. Laugks, V. Lucić, W. Baumeister, and R. Danev. Electron cryotomography of vitrified cells with a volta phase plate. *J. Struct. Biol.*, 190(2):143–154, 2015. 2.3.2

P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor. Biol.*, 36(1):105–117, 1972a. 1, 3.2.1

P. F. C. Gilbert. The Reconstruction of a Three-Dimensional Structure from Projections and Its Application to Electron Microscopy. II. Direct Methods. *Proc. R. Soc. B Biol. Sci.*, 182(1066):89–102, July 1972b. 3.2.1, 3.2.2, 3.2.2

W. R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1999. ISBN 0412055511. 4.1

R. M. Glaeser. Retrospective: Radiation damage and its associated "information limitations". *J. Struct. Biol.*, 163(3):271–276, 2008. 2.1

R. Gordon and G. T. Herman. Reconstruction of pictures from their projections. *Commun. ACM*, 14(12): 759–768, 1971. 4.2.1

R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.*, 29(3):471–81, 1970. 3.2.1

J. Gregson, M. Krimerman, M. B. Hullin, and W. Heidrich. Stochastic tomography and its applications in 3D imaging of mixing fluids. *ACM Trans. Graph.*, 31(4):52:1–10, 2012. 4.2.1, 5, 5.1.2, 5.1.2, 5.2, 6.1

N. Grigorieff. Three-dimensional structure of bovine nadh: ubiquinone oxidoreductase (complex i) at 22 a in ice. *J. Mol. Biol.*, 277:1033–1046, 1998. 3.2.2

I. Gross, H. Hohenberg, T. Wilk, K. Wiegers, M. Grättinger, B. Müller, S. Fuller, and H.-G. Kräusslich. A conformational switch controlling hiv-1 morphogenesis. *EMBO J.*, 19:103–113, 2000. 7.1.4

J. H. Halton. Algorithm 247: Radical-inverse quasi-random point sequence. *Commun. ACM*, 7(12), 1964. 4.1.1

J. M. Hammersley and D. C. Handscomb. *Monte {Carlo} Methods*. Methuen's monographs on applied probability and statistics. Methuen, 1964. ISBN 0-416-52340-4. 4, 4.1

R. Han, F. Zhang, X. Wan, J. J. Fernández, F. Sun, and Z. Liu. A marker-free automatic alignment method based on scale-invariant features. *J. Struct. Biol.*, 186(1):167–180, 2014. 2.3.1

H. Haneishi, T. Masuda, N. Ohyama, and T. Honda. Three-dimensional blood vessel reconstruction by simulated annealing. *Opt. Lett.*, 14(20):1095–1097, 1989. 4.2.1

G. Harauz and M. van Heel. Exact filters for general geometry three dimensional reconstruction. *Optik*, 73: 146–156, 1986. 3.3

W. K. Hastings. Monte Carlo sampling methods using markov chains and their applications. *Biometrika*, 57 (1):97–109, 1970. 4.1.3

R. Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl. Acad. Sci. U. S. A.*, 110(45):18037–18041, 2013. 2.5.1

R. Henderson, A. Sali, M. L. Baker, and et al. Outcome of the first electron microscopy validation task force meeting. *Structure*, 20:205–214, 2012. 3.3

G. T. Herman. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. Springer Publishing Company, Incorporated, 2nd edition, 2009. ISBN 185233617X, 9781852336172. 3.1

J. Hernández-Saz, M. Herrera, and S. I. Molina. A methodology for the fabrication by FIB of needle-shape specimens around sub-surface features at the nanometre scale. *Micron*, 43(5):643–650, 2012. 10.1

J. B. Heymann, G. Cardone, D. C. Winkler, and A. C. Steven. Computational resources for cryo-electron tomography in Bsoft. *J. Struct. Biol.*, 161(3):232–242, 2008. 3.2.2

T. Hrabe, Y. Chen, S. Pfeffer, L. K. Cuellar, A. V. Mangold, and F. Förster. PyTom: A python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.*, 178(2):177–188, 2012. 2.5, 2.5.2, 2.5.3, 3.2.2, 8.3, 8.4

C.-E. Hsieh, A. Leith, C. A. Mannella, J. Frank, and M. Marko. Towards high-resolution three-dimensional imaging of native mammalian tissue: Electron tomography of frozen-hydrated rat liver sections. *J. Struct. Biol*, 153:1–13, 2006. 2.1

C. V. Iancu, E. R. Wright, J. Benjamin, W. F. Tivol, D. P. Dias, G. E. Murphy, R. C. Morrison, J. B. Heymann, and G. J. Jensen. A "flip-flop" rotation stage for routine dual-axis electron cryotomography. *J. Struct. Biol.*, 151 (3):288–297, 2005. 10.1

F. John. The ultrahyperbolic differential equation with four independent variables. *Duke Math. J.*, 4(2): 300–322, 06 1938. 1, 3

S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Int. Acad. Pol. Sci. Let. Cl. Med.*, A35:355–357, 1937. 3.2.1

A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging.* IEEE Press, New York, 1988. 3.2.1

N. Kawase, M. Kato, H. Nishioka, and H. Jinnai. Transmission electron microtomography without the "missing wedge" for quantitative structural analysis. *Ultramicroscopy*, 107(1):8–15, 2007. 10.1

K. J. Kearfott and S. E. Hill. Simulated annealing image reconstruction method for a pinhole aperture single photon emission computed tomograph (SPECT). *IEEE Trans. Med. Imaging*, 9(2):128–143, 1990. 4.2.1

J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.*, 23(3):462–466, 09 1952. 4.1

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598): 671–680, 1983. 4.1, 4.2.1

G. P. Kishchenko, R. Danev, R. Fisher, J. He, C. Hsieh, M. Marko, and H. Sui. Effect of fringe-artifact correction on sub-tomogram averaging from Zernike phase-plate cryo-TEM. *J. Struct. Biol.*, 191(3):299–305, 2015. 2.3.2

T. Knopp, S. Kunis, and D. Potts. A note on the iterative mri reconstruction from nonuniform k-space data. *Int. J. Biomed. Imaging*, 2007:24727, 2007. 3.2.2

H. Kohr and A. K. Louis. Fast and high-quality reconstruction in electron tomography based on an enhanced linear forward model. *Inverse Probl.*, 27(4):045008, 2011. 3.2.1

A. J. Koster, H. Chen, J. W. Sedat, and D. A. Agard. Automated microscopy for electron tomography. *Ultramicroscopy*, 46:207–227, 1992. 2.2.2

A. J. Koster, R. Grimm, D. Typke, R. Hegerl, A. Stoschek, J. Walz, and W. Baumeister. Perspectives of molecular and cellular electron tomography. *J. Struct. Biol.*, 120:276–308, 1997. 2.2.2

J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.*, 116(1):71–76, 1996. 7.1, 8.3

M. Kudryashev, H. Stahlberg, and D. Castaño-Díez. Assessing the benefits of focal pair cryo-electron tomography. *J. Struct. Biol.*, 178:88–97, 2012. 2.3.2

L. Landweber. An iteration formula for fredholm integral equations of the first kind. *Am. J. Math.*, 73(3): 615–624, 1951. 3.2.1

R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland. Compressed sensing electron tomography. *Ultramicroscopy*, 131(0):70–91, 2013. 10.1

A. Leis, B. Rockel, L. Andrees, and W. Baumeister. Visualizing cells at nanoscale. *Trends Biol. Sc.*, 34:60–70, 2009. 3.3

A. P. Leis, M. Beck, M. Gruska, C. Best, R. Hegerl, W. Baumeister, and J. W. Leis. Cryo-electron tomography of biological specimens. *Signal Processing Magazine*, 23(3):95–103, 2006. 3.2.1

S. Li, J. J. Fernández, W. F. Marshall, and D. A. Agard. Three-dimensional structure of basal body triplet revealed by electron cryo-tomography. *EMBO J.*, 31:552–562, 2012. 2.3.2

Z. Li and H. A. Scheraga. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci.*, 84(19):6611–6615, 1987. 4.2.1

Y. Liu, P. A. Penczek, B. F. McEwen, and J. Frank. A marker-free alignment method for electron tomography. *Ultramicroscopy*, 58:393–402, 1995. 2.3.1

A. K. Louis. Incomplete data problems in X-ray computerized tomography: I. Singular value decomposition of the limited angle transform. *Numer. Math.*, 48:251–262, 1986. 10, 10.1

A. K. Louis. *Inverse und schlecht gestellte Probleme (Teubner Studienbucher) (German Edition)*. B.G. Teubner, 1989. ISBN 0-416-52340-4. 3.1, 3.1, 10

A. K. Louis and A. Rieder. Incomplete data problems in X-ray computerized tomography: II. Truncated projections and region-of-interest tomography. *Numer. Math.*, 56:371–383, 1989. 10, 10.3

V. Lucić, F. Förster, and W. Baumeister. Structure studies by electron tomography: From Cells to Molecules. *Annu. Rev. Biochem.*, 74:833–865, 2005. 1, 3.2.1

V. Lucić, A. Rigort, and W. Baumeister. Cryo-electron tomography: The challenge of doing structural biology in situ. *J. Cell Biol.*, 202(3):407–419, 2013. 1.2.1, 2.1, 2.2

S. J. Ludtke, P. R. Baldwin, and W. Chiu. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.*, 128(1):82–97, 1999. 7.1, 10.5

M. Marko, C. Hsieh, C. W. Moberlychan, C. A. Mannella, and J. Frank. Focused ion beam milling of vitreous water: Prospects for an alternative to cryo-ultramicrotomy of frozen-hydrated biological samples. *J. Microsc.*, 222:42–47, 2006. 2.1

M. Marko, C. Hsieh, R. Schalek, J. Frank, and C. Mannella. Focused-ion-beam thinning of frozen-hydrated biological specimens for cryo-electron microscopy. *Nat. Methods*, 4:215–217, 2007. 2.1

D. N. Mastronarde. Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the imod package. *J. Microsc.*, 230:212–217, 2008. 2.3.1

D.N. Mastronarde. Dual-axis tomography: An approach with alignment methods that preserve resolution. *J. Struct. Biol.*, 120:343–352, 1997. 10.1

S. Matej, J. Fessler, and I. Kazantsev. Iterative tomographic image reconstruction using Fourier-based forward and back-projectors. *IEEE Trans. Med. Imaging*, 23:401–412, 2004. 3.2.2

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953. 4.1.3

A. G. Myasnikov, Z. A. Afonina, and B. P. Klaholz. Single particle and molecular assembly analysis of polyribosomes by single- and double-tilt cryo electron tomography. *Ultramicroscopy*, 126(0):33–39, 2013. 10.1

F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction.* SIAM, Philadelphia, PA., 2001. 3.1, 3.1

D. Nicastro, X. Fu, T. Heuser, A. Tso, M. E. Porter, and R. W. Linck. Cryo-electron tomography reveals conserved features of doublet microtubules in flagella. *Proc. Natl. Acad. Sci. U. S. A.*, 108:E845 – E853, 2011. 2.3.2

S. J. Norton. Iterative reconstruction algorithms: Convergence as a function of spatial frequency. *J. Opt. Soc. Am. A*, 2(1):6–13, 1985. 3.2.2

J. O. Ortiz, F. Brandt, V. R. F. Matias, L. Sennels, J. Rappsilber, S. H. W. Scheres, M. Eibauer, F. U. Hartl, and W. Baumeister. Structure of hibernating ribosomes studied by cryoelectron tomography in vitro and in situ. *J. Cell Biol.*, 190:613–621, 2010. 2.3.2

C. M. Palmer and J. Loewe. A cylindrical specimen holder for electron cryo-tomography. *Ultramicroscopy*, 137(0):20–29, 2014. 10, 10.1

P. Penczek, M. Marko, K. Buttle, and J. Frank. Double-tilt electron tomography. *Ultramicroscopy*, 60:393–410, 1995. 10.1

P. A. Penczek, R. Renka, and H. Schomberg. Gridding-based direct Fourier inversion of the three-dimensional ray transform. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.*, 21:499–509, 2004. 3.2.2

A. Philippsen, H. A. Engel, and A. Engel. The contrast-imaging function for tilted specimens. *Ultramicroscopy*, 107:202–212, 2007. 2.3.2

S. A. Qureshi, S. M. Mirza, and Arif M. Hybrid simulated annealing image reconstruction for transmission tomography. *Inverse Probl. Sc. Eng.*, 17(5):647–664, 2009. 4.2.1

M. Radermacher. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J. Electr. Micr. Technique*, 9:359–394, 1988. 10.1

M. Radermacher. Weighted backprojection methods. In J. Frank, editor, *Electron Tomography*, pages 91–115. Plenum Press, New York, 1992. 1, 3.2.1

M. Radermacher, T. Wagenknecht, A. Verschoor, and J. Frank. Three-dimensional reconstruction from a single-exposure random conical tilt series applied to the 50s ribosomal subunit of escherichia coli. *J. Microsc.*, 146:113–136, 1987. 3.2.1

J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. *Math. Phys. Klasse*, 69:262–277, 1917. 1, 3, 3.1

G. N. Ramachandran and A. V. Lakshminarayanan. Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of Fourier transforms. *Proc. Natl. Acad. Sci.*, 68:2236–2240, 1971. 10.5

D. Ress, M. L. Harlow, M. Schwarz, R. M. Marshall, and U. J. McMahan. Automatic acquisition of fiducial markers and alignment of images in tilt series for electron tomography. *J. Electron Microsc.*, 48:277–287, 1999. 2.3.1

A. Rigort, F. J. Bauerlein, E. Villa, M. Eibauer, T. Laugks, W. Baumeister, and J. M. Plitzko. Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography. *Proc. Natl. Acad. Sci. U. S. A.*, 109:4449–4454, 2012. 2.1

H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statist.*, 22(3):400–407, 09 1951. 4.1

C. P. Robert and G. Casella. *Monte Carlo statistical methods.* Springer, New York, 2nd edition, 2004. ISBN 0-387-21239-6. 4

R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method.* Springer-Verlag, 2004. 4.1

Z. Saghi, D. J. Holland, R. Leary, A. Falqui, G. Bertoni, A. J. Sederman, L. F. Gladden, and P. A. Midgley. Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. A compressed sensing-electron tomography (CS-ET) approach. *Nano Letters*, 11(11):4666–4673, 2011. 10.1

K. Sandberg, D. N. Mastronarde, and G. Beylkin. A fast reconstruction algorithm for electron microscope tomography. *J. Struct. Biol.*, 144(1-2):61–72, 2003. 3.2.2, 3.3, 9.1, 10.3, 10.5

S. H. Scheres, R. Melero, M. Valle, and Carazo J. M. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Structure*, 17:1563–72, 2009. 2.5.2

F. K. M. Schur, W. J. H. Hagen, A. de Marco, and J. A. G. Briggs. Determination of protein structure at 8.5 Å resolution using cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.*, 184(3):394–400, 2013. 2.5

L. Shepp and B. F. Logan. The Fourier reconstruction of a head section. *IEEE Trans. Nucl. Sci.*, 21(3):21–43, 1974. 7.1.1

U. Skoglund, L. G. Öfverstedt, R. M. Burnett, and G. Bricogne. Maximum-entropy three-dimensional reconstruction with deconvolution of the contrast transfer function: A test application with adenovirus. *J. Struct. Biol.*, 117(3):173–188, 1996. 3.2.1

C. O. S. Sorzano, C. Messaoudi, M. Eibauer, J. R. Bilbao-Castro, R. Hegerl, S. Nickell, S. Marco, and J. M. Carazo. Marker-free image registration of electron tomog- raphy tilt-series. *BMC Bioinformatics*, 10:124, 2009. 2.3.1

J. C. Spall. *Introduction to Stochastic Search and Optimization.* Wiley, 2003. ISBN 0-471-33052-3. 4.1

M. Stölken, F. Beck, T. Haller, R. Hegerl, I. Gutsche, J.-M. Carazo, W. Baumeister, S.H.W. Scheres, and S. Nickell. Maximum likelihood based classification of electron tomographic data. *J. Struct. Biol.*, 173(1):77–85, 2011. 2.5.2

D. Studer, B. M. Humbel, and M. Chiquet. Electron microscopy of high pressure frozen samples: Bridging the gap between cellular ultrastructure and atomic resolution. *Histochem. Cell Biol.*, 130:877–889, 2008. 2.1

S. Subramaniam. The siv surface spike imaged by electron tomography: One leg or three? *PLoS Pathog.*, 2: e91, 2006. 2.5.3

E. Sundermann and I. Lemahieu. Pet image reconstruction using simulated annealing. *Proc. SPIE*, 2434: 378–386, 1995. 4.2.1

J. N. Turner, D. P. Barnard, P. McCauley, and D. L. Dorset. Diffraction and imaging from all perspectives: Unlimited specimen tilting in the high-voltage electron microscope. *Proc. 49th Annu. Meet. Electron Microsc. Soc. Am.*, pages 994–995, 1991. 10.1

B. Turoňová, L. Marsalek, T. Davidovič, and P. Slusallek. Progressive stochastic reconstruction technique (psrt) for cryo electron tomography. *J. Struct. Biol.*, 189(3):195–206, 2015. 1.1

B. Turoňová, L. Marsalek, and P. Slusallek. On geometric artifacts in cryo electron tomography. *Ultramicroscopy*, 163:48–61, 2016. ISSN 0304-3991. 1.1

M. van Heel and M. Schatz. Fourier shell correlation threshold criteria. *J. Struct. Biol.*, 151(3):250–262, 2005. 3.3

L. M. Voortman, S. Stallinga, R. H. Schoenmakers, L. J. van Vliet, and B. Rieger. A fast algorithm for computing and correcting the ctf for tilted, thick specimens in TEM. *Ultramicroscopy*, 111:1029–1036, 2011. 2.3.2

S. Webb. SPECT reconstruction by simulated annealing. *Phys. Med. Biol.*, 34(3):259–281, 1989. 4.2.1

D. B. Williams and C. B. Carter. *Transmission electron microscopy*. Springer, 2009. 1, 2.2.1, 2.2.2

H. Winkler and K. A. Taylor. Multivariate statistical analysis of three-dimensional cross-bridge motifs in insect flight muscle. *Ultramicroscopy*, 77(3-4):141–152, 1999. 1, 2.5

H. Winkler and K. A. Taylor. Focus gradient correction applied to tilt series image data used in electron tomography. *J. Struct. Biol.*, 143:24–32, 2003. 2.3.2

H. Winkler and K. A. Taylor. Accurate marker-free alignment with simultaneous geometry determination and reconstruction of tilt series in electron tomography. *Ultramicroscopy*, 106:240–254, 2006. 2.3.1

H. Winkler and K. A. Taylor. Marker-free dual-axis tilt series alignment. *J. Struct. Biol.*, 182(2):117–124, 2013. 2.3.1

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. *J. Struct. Biol.*, 168:378–387, 2009. 2.3.2

M. Xu, M. Beck, and F. Alber. Template-free detection of macromolecular complexes in cryo electron tomograms. *Bioinformatics*, 27:69–76, 2011. 2.5.1

M. Xu, M. Beck, and F. Alber. High-throughput subtomogram alignment and classification by Fourier space constrained fast volumetric matching. *J. Struct. Biol.*, 178:152–164, 2012. 2.5.3

W. Xu, F. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard, and K. Mueller. High-performance iterative electron tomography reconstruction with long-object compensation using graphics processing units (GPUs). *J. Struct. Biol.*, 171(2):142–153, 2010. 3.2.1, 5.4.1, 10, 10.3, A.3

L. Yu, R. R. Snapp, T. Ruiz, and M. Radermacher. Probabilistic principal component analysis with expectation maximization (PPCA-EM) facilitates volume classification and estimates the missing data. *J. Struct. Biol.*, 171(1):18–30, 2010. 2.5.2

Z. Yu and A. S. Frangakis. Classification of electron sub-tomograms with neural networks and its application to template-matching. *J. Struct. Biol.*, 174:494–504, 2011. 2.5.2

G. Zanetti, J. D. Riches, S. D. Fuller, and J. A. Briggs. Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.*, 168:305–312, 2009. 2.3.2

W. Zhang, M. Kimmel, C. M. T. Spahn, and P. A. Penczek. Heterogeneity of large macromolecular complexes revealed by 3D cryo-em variance analysis. *Structure*, 16:1770–1776, 2008. 3.2.2

S. Q. Zheng, A. Matsuda, M. B. Braunfeld, J. W. Sedat, and D. A. Agard. Dual-axis target mapping and automated sequential acquisition of dual-axis {EM} tomographic data. *J. Struct. Biol.*, 168(2):323–331, 2009. 10.1