Universiteit Gent
Faculteit Wetenschappen
Vakgroep Fysica en Sterrenkunde
Proeftuinstraat 86, B-9000 Gent, België
Tel: +32-(0)9-264.66.11

UNIVERSITEIT GENT

Improved and practically feasible reconstruction methods for high resolution X-ray tomography

Yoni De Witte

Promotor: Prof. dr. Luc Van Hoorebeke

Examencommissie:
Prof. dr. Jan Ryckebusch (voorzitter)
Prof. dr. Luc Van Hoorebeke (promotor)
Dr. Peter Cloetens (leescommissie)
Prof. dr. Michel Defrise (leescommissie)
Prof. dr. Veerle Cnudde
Dr. ir. Manuel Dierick
Prof. dr. ir. Stefaan Vandenberghe
Dr. ir. Jelle Vlassenbroeck

Faculteit Wetenschappen

Proefschrift tot het behalen van de graad van
Doctor in de Wetenschappen: Fysica
Mei 2010

Voorwoord

I can't believe it! Reading and writing actually paid off!
– Homer Simpson

Boven alles is het schrijven van een doctoraat vooral een gevecht tegen je eigen persoon. Elke dag opnieuw moet je jezelf oppeppen en trachten je vooropgesteld aantal pagina's te schrijven. Er zijn dagen waarop je geen letter op papier krijgt en doelloos naar je scherm zit te staren, fantaserend over hoe geweldig het wel niet zou zijn om je PC gewoon in de vijver te keilen. Op andere dagen vliegen de pagina's je rond de oren en moet je alleen af en toe even pauzeren om te vermijden dat je toetsenbord wegsmelt. En zo gaat het maar door..., dagen, weken, maanden aan een stuk, tot die magische dag waarop je dat laatste '.' zet en de wereld doet verstommen met die bevrijdende woorden: "Mij thesis is af!!!".

Hoewel een goede voorbereiding en een sterk doorzettingsvermogen belangrijk zijn, zou het halen van een doctoraat onmogelijk zijn zonder de nodige hulp en steun van collega's, vrienden en familie. Ik ben hen dan ook veel dank verschuldigd.

Luc, jij hebt me vier jaar geleden de kans geboden om aan een doctoraat te beginnen in jouw groep. Ook al was er oorspronkelijk maar uitzicht op een positie voor één jaar, ik heb nooit spijt gehad van mijn keuze en ben heel blij dat ik het risico genomen heb. Die klimaatkamer zal wellicht nog even op zich moeten laten wachten, maar ik ben blij dat je me de kans hebt gegeven om mijn eigen ding te doen. Dankzij jou heb ik mijn interesses kunnen laten aanwakkeren op tal van congressen en tegelijkertijd ook een stapje in de wereld mogen zetten. Bedankt ook voor het luisterende oor en de altijd aanwezige steun, je bent meer dan alleen een promotor.

Bedankt aan al mijn collega's en ex-collega's, zowel anciens als jonkies. Bert, bedankt voor je enthousiasme en de kennismaking met de helix geometrie. Bedankt Manu voor de lessen Labview voor beginners, ook al ben ik eigenlijk een C++ ketter. Veerle, jouw oppeppende woorden na een vernietigende review zijn altijd een sterke motivatie geweest om ermee door te gaan, misschien meer dan je zelf beseft. De fijne kneepjes van het programmeren heb ik geleerd van Jelle, bedankt ook voor de inspirerende gesprekken over reconstructie en voor het schenken van mijn eerste GPU. Bert, Manu, Veerle, Jelle, jullie hebben mij ingewijd in tomografie en mij samen met Luc over de streep gehaald om in deze fantastische groep te komen werken. Jullie enthousiasme bij onze eerste kennismaking was overweldigend. Dank ook aan de enige echte 'phase-man' Matthieu om mij in te wijden in fase-contrast, je had waarschijnlijk nooit gedacht dat ik zo geïntrigeerd

zou raken. Denis, bedankt voor de technische uitwerking van de helix en voor het aanbrengen van vele interessante samples en problemen. Patric, Jan DW, Tim, Marijn, zonder de geologie zou UGCT nooit staan waar we nu staan, doe zo verder. Pieter, bedankt om terug te keren en ons te voorzien van technische ondersteuning. Loes, veel succes met je verdere onderzoek, of het nu analyse of reconstructie is. Ik wens alle doctoraatsstudenten heel veel succes met hun thesis. Joris, Jan VDB, Dries, ik ben ervan overtuigd dat UGCT nog sterker zal groeien nu jullie labo zich vervoegd heeft. Verder ook dank aan mijn overige collega's van het INW voor de leuke kerstfeestjes, barbecues en andere aangelegenheden. Dank ook aan de shotters van het INW en de S9 voor ons wekelijkse potje mentale ontspanning, daar heb ik altijd naar uitgekeken. Niet te vergeten zijn natuurlijk de mensen van de ateliers en het secretariaat, jullie houden immers onze speeltuin draaiende.

Het leven bestaat natuurlijk uit meer dan werken alleen. Bedankt aan alle vrienden en (schoon)familie voor jullie steun en om af en toe eens te vragen hoe het nu zit met dat ding dat ik moet schrijven. De meesten onder jullie hebben er waarschijnlijk geen idee van waarover het echt gaat, maar ik vind het toch enorm sympathiek dat jullie er steeds werkelijk geïnteresseerd naar vroegen. Bedankt ook aan al mijn voetbalkameraden om mijn fysieke en geestelijke gezondheid van de afgrond te redden.

Een bijzondere dank gaat uit naar mijn Saartje, gewoon omdat je er altijd bent voor mij. Toen ik je leerde kennen moest ik mijn middelbaar nog afmaken en nu, zoveel jaren later zijn we nog altijd gelukkig samen en ben ik eindelijk afgestudeerd geraakt (ik heb het nu echt wel gehad). Gedaan dus met mopjes te maken als "Werken? Gij studeert toch nog."

Als laatste wil ik natuurlijk mijn ouders en mijn broer bedanken. Zij hebben mij de kans gegeven om te studeren en hebben mij altijd gesteund in elke keuze die ik heb gemaakt. Mijn doctoraat heb ik aan hen te danken. Ma, een 'dokter' zoals ik in het lagertje altijd wou worden is er dus niet van gekomen, maar ik denk dat ik je nu toch ook wel trots gemaakt heb. Pa, wees gerust, nu stop ik definitief met studeren en ga ik gewoon werken hoor. Broerie, ook aan u ben ik veel dank verschuldigd. Jij hebt me al leren rekenen toen ik nog maar in het kleuterklasje zat, en zie, het heeft opgebracht hé. Als laatste wil ik nog alle mensen bedanken die ik schandelijk vergeten ben in mijn lijstje, waarvoor mijn excuses, en simpelweg nog eens aan iedereen die het aanbelangd zeggen: Mercikes!

Yoni

Samenvatting
–Summary in Dutch–

Sinds hun ontdekking, meer dan een eeuw geleden, worden X-stralen gebruikt om een beeld te vormen van de interne structuur van allerlei voorwerpen. In tegenstelling tot de 2-dimensionale beeldvorming door middel van radiografie, biedt X-stralen computer tomografie (CT) een volledige 3-dimensionale voorstelling van het onderzochte voorwerp. Tomografie is wellicht best gekend vanwege zijn toepassing in de medische wereld. In de voorbije jaren is deze techniek echter ook doorgedrongen in tal van industriële en wetenschappelijke toepassingen en is hij geëvolueerd tot een volwaardig instrument dat gebruikt wordt in verschillende onderzoeksdomeinen. De huidige meest vooruitstrevende toestellen kunnen voorwerpen bestuderen met een resolutie van minder dan één micrometer.

Aan het Centrum voor X-stralen Tomografie van de Universiteit Gent (UGCT) wordt reeds jaren onderzoek gevoerd naar computer tomografie. Het vroegere onderzoek was gericht op neutronen tomografie en op tomografie gebruik makende van een lineaire versneller. Daarna verschoof het onderzoek in de richting van de hoge resolutie X-stralen tomografie. Om de gebruikelijke beperkingen van commerciële systemen te omzeilen, werd er besloten om zelf een hoge resolutie toestel te ontwerpen en te bouwen. Reeds vanaf het ogenblik dat de bouw voltooid werd, levert dit ultramoderne toestel schitterende resultaten in tal van toepassingen. Het behalen van deze indrukwekkende beeldkwaliteit is echter slechts mogelijk door het gebruik van geschikte reconstructie software. Daarom wordt binnen UGCT ook onderzoek verricht naar reconstructie algoritmes, die geïmplementeerd worden in een uiterst efficiënt pakket genaamd Octopus. Omwille van deze directe toegang tot de reconstructie, is het mogelijk om steeds de best mogelijke beeldkwaliteit te bekomen. Dankzij deze intense samenwerking tussen hardware en software is UGCT op korte tijd uitgegroeid tot één van de leiders op het gebied van hoge resolutie X-stralen tomografie.

Ondanks de reeds indrukwekkende resultaten ligt het in de aard van de mens, en van onderzoekers in het bijzonder, om steeds meer te willen. In het geval van X-stralen tomografie betekent dit hogere resolutie, betere beeldkwaliteit en meer flexibiliteit. In dit werk worden enkele van de meest voorkomende problemen in hoge resolutie CT aangepakt, door het zoeken naar en het ontwikkelen van reconstructie methodes die meer geschikt zijn dan de algoritmes die momenteel als standaard gelden. Eerder dan op de theoretische achtergrond ligt de nadruk van

dit onderzoek vooral op de praktische toepassing van zulke verbeterde technieken. Dit betekent dat enkel die methodes beschouwd worden die haalbaar zijn in de praktijk en die onmiddellijk toegepast kunnen worden op de huidige beschikbare toestellen. De onderzochte technieken moeten dus op een efficiënte manier geïmplementeerd kunnen worden en ze mogen geen aanpassingen vereisen aan de opstelling van de CT scanner of aan de routine van de CT scan.

Het onderzoek dat voorgesteld wordt in dit werk werd uitgevoerd binnen UGCT. Alhoewel het onderzoek naar de hardware en de software in feite redelijk gescheiden zijn, is er bij UGCT steeds een sterke wisselwerking tussen deze twee domeinen. Het uiteindelijke doel van beide onderzoeksgebieden is namelijk hetzelfde: de continue verbetering van hoge resolutie X-stralen tomografie. Kort samengevat kunnen de doelstellingen van het onderzoek uitgevoerd in het kader van dit werk als volgt geformuleerd worden:

> **Het verbeteren van de kwaliteit en het uitbreiden van het toepassingsgebied van hoge resolutie X-stralen tomografie, door het onderzoeken en ontwikkelen van verbeterde reconstructie algoritmes, rekening houdend met de praktische haalbaarheid in alledaagse toepassingen.**

In hoofdzaak worden in dit werk drie aspecten behandeld met betrekking tot de tomografische reconstructie van hoge resolutie X-stralen projectiedata, opgenomen met een toestel dat gebruik maakt van een X-stralen buis. Deze drie onderwerpen zijn: de reconstructie in de helische conische-bundel geometrie, iteratieve reconstructiealgoritmes en de reconstructie van projectiedata waarin fasecontrast aanwezig is.

In hoofdstuk 3 worden de analytische reconstructiealgoritmes besproken voor de parallelle-, de waaiervormige- en de conische-bundel geometrie. Aan hoge resolutie X-stralen systemen gebruik makend van een X-stralen buis wordt voornamelijk de conische geometrie toegepast. Deze geometrie heeft echter twee grote nadelen: hij introduceert fouten vanwege de conische bundel en is niet geschikt om langgerekte voorwerpen te bestuderen. Als oplossing hiervoor hebben we de helische conische-bundel geometrie ingevoerd in hoge resolutie CT. Voor ons toepassingsgebied blijkt het Katsevich algoritme het meest geschikte reconstructiealgoritme te zijn. De snelheid van het reconstructieproces kan aanzienlijk verbeterd worden door het toepassen van verschillende technieken, die zowel op hardware als op software gebaseerd kunnen zijn. Wij hebben onze reconstructiesoftware versneld door gebruik te maken van grafische kaarten. Dit hoofdstuk wordt afgesloten door enkele resultaten te tonen waarin de helische conische-bundel geometrie werd toegepast.

De belangrijkste facetten van iteratieve reconstructiealgoritmes worden uitgebreid besproken in hoofdstuk 4. De toepassing van deze technieken in hoge resolutie CT, waarin typisch heel grote volumes moeten verwerkt worden, wordt voornamelijk bemoeilijkt door twee praktische problemen, zijnde de reconstructiesnel-

heid en de zware geheugenvereisten. De reconstructie kan opnieuw aanzienlijk versneld worden door gebruik te maken van grafische kaarten. Voor het tweede probleem stellen we een multiresolutie benadering voor, die ons in staat stelt om het vereiste geheugen aanzienlijk te beperken. Gebruik makend van deze twee oplossingen, stellen we dan een uiterst efficiënte implementatie voor van de *simultaneous algebraic reconstruction technique* (SART), die ons in staat stelt om een iteratief reconstructiealgoritme toe te passen op data van hoge resolutie X-stralen CT. Daarna tonen we aan dat de iteratieve methodes een aanzienlijke verbetering van de reconstructiekwaliteit kunnen bieden ten opzichte van de analytische reconstructie, zeker wanneer om bepaalde redenen de projectiedata enigszins beperkt is, bijvoorbeeld wanneer slechts een klein aantal projecties genomen werden, wanneer de projecties veel ruis bevatten of wanneer the projectiehoeken onregelmatig gekozen werden. We tonen ook hoe we, door de voorwaartse projectie van onze iteratieve methode aan te passen, erin kunnen slagen om in de reconstructie fouten te verminderen die afkomstig zijn van *beam hardening* door sterk absorberende insluitsels. Om af te sluiten tonen we resultaten van alledaagse scans waarbij we onze iteratieve reconstructie gebruikten om de beeldkwaliteit te verbeteren.

In het laatste deel, hoofdstuk 5, bespreken we het optreden van fasecontrast, wat een gevolg is van de refractie van X-stralen. Fasecontrast in de projecties leidt tot storende fouten in de gereconstrueerde beelden wanneer we hier niet op een gepaste manier mee omgaan. We tonen aan hoe we een methode, genaamd *modified Bronnikov algorithm* (MBA), kunnen gebruiken om het fasesignaal uit de projecties te halen. Hiermee kunnen we dan de refractiefunctie van het voorwerp reconstrueren, wat ons veel betere beelden oplevert dan de gebruikelijke reconstructie. Dit algoritme faalt echter wanneer de attenuatie door het voorwerp te groot wordt. Om dit probleem op te lossen hebben we een alternatieve oplossing ontwikkeld die gericht is op de reductie van het fasesignaal. Deze methode, genaamd *Bronnikov-aided correction* (BAC), stelt ons in staat om de attenuatiefunctie van middelmatig en sterk absorberende fasevoorwerpen te reconstrueren, waarbij de typische fouten geïntroduceerd door fasecontrast verwijderd of toch minstens sterk gereduceerd zijn. De beide technieken MBA en BAC bieden ons twee complementaire oplossingen voor het reconstrueren van projecties waarin zowel attenuatie- als fasecontrast aanwezig is. Beide methodes werden uitgebreid geëvalueerd door ze toe te passen in de verwerking van data afkomstig van de scanners van UGCT, maar ook van data die opgenomen werden aan andere hoge resolutie opstellingen. Als gevolg van de indrukwekkende kwaliteit van de reeds bekomen beelden, zijn beide technieken ingeschakeld in de alledaagse werking van UGCT.

Publications

Journal papers

[1] Yoni De Witte, Jelle Vlassenbroeck, and Luc Van Hoorebeke. *A multiresolution approach to iterative reconstruction algorithms in X-ray computed tomography*. IEEE Transactions on Image Processing, In Press.

[2] Joris Soons, Anthony Herrel, Annelies Genbrugge, Peter Aerts, Jeffrey Podos, Dominique Adriaens, Yoni De Witte, Patric Jacobs, and Joris Dirckx. *Mechanical stress, fracture risk and beak evolution in Darwin's ground finches (Geospiza)*. Philosophical Transactions of the Royal Society B - Biological Sciences, 365(1543):1093–1098, 2010.

[3] Yoni De Witte, Matthieu Boone, Jelle Vlassenbroeck, Manuel Dierick, and Luc Van Hoorebeke. *Bronnikov-aided correction for x-ray computed tomography*. Journal of the Optical Society of America A, 26(4):890–894, 2009. *Also published in* Virtual Journal for Biomedical Optics, 4(6), 2009.

[4] Matthieu Boone, Yoni De Witte, Manuel Dierick, Jan Van den Bulcke, Jelle Vlassenbroeck, and Luc Van Hoorebeke. *Practical use of the modified Bronnikov algorithm in micro-CT*. Nuclear Instruments and Methods in Physics Research B, 267(7):1182–1186, 2009.

[5] Veerle Cnudde, Geert Silversmit, Matthieu Boone, Jan Dewanckele, Björn De Samber, Tom Schoonjans, Denis Van Loo, Yoni De Witte, Marlina Elburg, Laszlo Vincze, Luc Van Hoorebeke, and Patric Jacobs. *Multi-disciplinary characterisation of a sandstone surface crust*. Science of the Total Environment, 407(20):5417–5427, 2009.

[6] Jan Dewanckele, Veerle Cnudde, Matthieu Boone, Denis Van Loo, Yoni De Witte, Koen Pieters, Jelle Vlassenbroeck, Manuel Dierick, Bert Masschaele, Luc Van Hoorebeke, and Patric Jacobs. *Integration of X-ray micro tomography and fluorescence for applications on natural building stones*. Journal of Physics: Conference Series, 186(1):12082–12084, 2009.

[7] Yoni De Witte, Veerle Cnudde, Koen Pieters, Bert Masschaele, Manuel Dierick, Jelle Vlassenbroeck, Luc Van Hoorebeke, and Patric Jacobs. *X-ray micro-CT applied to natural building materials and art objects*. X-ray Spectrometry, 37(4):383–387, 2008.

[8] Veerle Cnudde, Bert Masschaele, Hilde De Cock, Kristin Olstad, Lieven Vlaminck, Jelle Vlassenbroeck, Manuel Dierick, Yoni De Witte, Luc Van Hoorebeke, and Patric Jacobs. *Virtual histology by means of high-resolution X-ray CT*. Journal of Microscopy-Oxford, 232(3):476–485, 2008.

[9] Manuel Dierick, Luc Van Hoorebeke, Patric Jacobs, Bert Masschaele, Jelle Vlassenbroeck, Veerle Cnudde, and Yoni De Witte. *The use of 2D pixel detectors in micro- and nano-CT applications*. Nuclear Instruments and Methods in Physics Research A, 591(1):255–259, 2008.

Conference contributions

[1] Yoni De Witte, Matthieu Boone, Jelle Vlassenbroeck, Manuel Dierick, Bert Masschaele, Luc Van Hoorebeke, and Veerle Cnudde. *Correcting phase contrast artefacts in X-ray CT imaging*. In 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pages 574–577. Boston, 2009.

[2] Yoni De Witte, Jelle Vlassenbroeck, Kjeld Vandeputte, Jan Dewanckele, Veerle Cnudde, Luc Van Hoorebeke, Gerald Ernst, and Patric Jacobs. *3D image analysis of a volcanic deposit*. In Geophysical Research Abstracts, EGU 2009, volume 11. Vienna, 2009.

[3] Matthieu Boone, Ana Almeida, Yoni De Witte, Manuel Dierick, Jean Paul Remon, Chris Vervaet, and Luc Van Hoorebeke. *Detection of very small density fluctuations using the modified Bronnikov algorithm on micro-CT data*. In Microscopy and Microanalysis 2009 abstracts, pages 386–387. Richmond, 2009.

[4] Veerle Cnudde, Matthieu Boone, Jan Dewanckele, Yoni De Witte, Denis Van Loo, Koen Pieters, Jelle Vlassenbroeck, Manuel Dierick, Bert Masschaele, Luc Van Hoorebeke, and Patric Jacobs. *Multi-resolution X-ray CT research applied on geomaterials*. In Geophysical Research Abstracts, EGU 2009, volume 11. Vienna, 2009.

[5] Jan Dewanckele, Matthieu Boone, Veerle Cnudde, Geert Silversmit, Yoni De Witte, Denis Van Loo, Laszlo Vincze, Luc Van Hoorebeke, and Patric Jacobs. *Contribution of X-ray based techniques to the structural and chemical characterisation of igneous rocks*. In Proceedings 12th Euroseminar on Microscopy Applied to Building Materials. Dortmund, 2009.

[6] Jan Dewanckele, Marijn Boone, Tim De Kock, Veerle Cnudde, Matthieu Boone, Yoni De Witte, Koen Pieters, Denis Van Loo, Luc Van Hoorebeke, and Patric Jacobs. *Disintegration of sedimentary rocks used as building material: evaluation and quantification in 4D*. In Geophysical Research Abstracts, EGU 2009, volume 11. Vienna, 2009.

[7] Denis Van Loo, Robert Speijer, Bert Masschaele, Manuel Dierick, Veerle Cnudde, Matthieu Boone, Yoni De Witte, Jan Dewanckele, Luc Van Hoorebeke, and Patric Jacobs. *The evaluation of Computed Tomography hard- and software tools for micropaleontologic studies on foraminifera*. In Geophysical Research Abstracts, EGU 2009, volume 11. Vienna, 2009.

[8] Yoni De Witte, Matthieu Boone, Veerle Cnudde, Jan Dewanckele, Manuel Dierick, Bert Masschaele, Koen Pieters, Denis Van Loo, Jelle Vlassenbroeck, Patric Jacobs, and Luc Van Hoorebeke. *Application of X-ray micro CT in multi-disciplinary research*. In Book of Abstracts Workshop on X-Ray Micro Imaging of Materials, Devices, and Organisms, pages 52–52. Dresden, 2008.

[9] Yoni De Witte, Jelle Vlassenbroeck, Manuel Dierick, Bert Masschaele, and Luc Van Hoorebeke. *Octopus: high-performance software for reconstruction and visualisation of CT-data*. In Book of Abstracts Workshop on X-Ray Micro Imaging of Materials, Devices, and Organisms, pages 103–103. Dresden, 2008.

[10] Jan Dewanckele, Veerle Cnudde, Yoni De Witte, Matthieu Boone, Denis Van Loo, Koen Pieters, Jelle Vlassenbroeck, Manuel Dierick, Bert Masschaele, Luc Van Hoorebeke, and Patric Jacobs. *A powerful technique for micro structural analyses on natural building stones*. In Abstract CD-rom 33rd International Geological Congress. Oslo, 2008.

[11] Jan Dewanckele, Veerle Cnudde, Yoni De Witte, Matthieu Boone, Denis Van Loo, B. Schouenborg, Luc Van Hoorebeke, and Patric Jacobs. *Integrated study of high resolution CT and XRF applied to natural building stones: preliminary results*. In Book of Abstracts 9th International Conference on X-ray Microscopy (XRM), pages 87–87. Zürich, 2008.

[12] Jan Dewanckele, Veerle Cnudde, Jelle Vlassenbroeck, Manuel Dierick, Yoni De Witte, Denis Van Loo, Matthieu Boone, Koen Pieters, Bert Masschaele, Luc Van Hoorebeke, and Patric Jacobs. *X-ray CT: a powerful analysis tool for assessing the internal structure of valuable objects and for constructing a 3D database*. In Proceedings of the 14th International Conference on Virtual Systems and Multimedia, Project Papers, pages 357–363. Limassol, 2008.

[13] Denis Van Loo, Bert Masschaele, Veerle Cnudde, Manuel Dierick, Jelle Vlassenbroeck, Yoni De Witte, Matthieu Boone, and Jan Dewanckele. *X-ray micro-tomography under altered environmental conditions with EMCT-scanner*. In Book of Abstracts 9th International Conference on X-ray Microscopy (XRM), pages 233–233. Zürich, 2008.

[14] Yoni De Witte, Veerle Cnudde, Manuel Dierick, Bert Masschaele, Koen Pieters, Jelle Vlassenbroeck, Luc Van Hoorebeke, and Patric Jacobs. *Micro-CT applied to natural building materials and art objects*. In TECHNART 2007 Non-destructive and Microanalytical Techniques in Art and Cultural Heritage, pages 62–62. Lisboa, 2007.

[15] Jelle Vlassenbroeck, Bert Masschaele, Manuel Dierick, Veerle Cnudde, Yoni De Witte, Koen Pieters, Luc Van Hoorebeke, and Patric Jacobs. *Recent developments in the field of X-ray nano- and micro-CT at the Centre for X-ray Tomography of the Ghent University*. In Microscopy and Microanalysis, Volume 13, Supplement S02, pages 184–185, 2007.

Table of Contents

List of Figures

List of Tables

If we knew what it was we were doing,

it would not be called research, would it?

Albert Einstein, 1879° – 1955†

1

Introduction

Ever since their discovery, more than a century ago, X-rays are used to visualize the internal structure of a variety of objects in a noninvasive way. Many years later, X-ray computed tomography (CT) was developed, which provides a complete 3-dimensional representation of the object, instead of the 2-dimensional projections that are acquired in a regular radiography. Although tomography is mostly known from the medical world, it is also used intensively in industrial and scientific applications and has evolved into an extremely valuable tool for various research fields. The current high resolution systems can image objects with a resolution of less than one micrometer.

At the Centre for X-ray Tomography of the Ghent University (UGCT), research on computed tomography is performed for several years now. In earlier years, this research concerned neutron tomography and X-ray tomography using a linear accelerator. Later, the focus shifted towards high resolution X-ray tomography. To overcome the typical limitations of commercially available systems, the development of a custom designed high resolution CT scanner was started at UGCT. Ever since its completion, this state-of-the-art device produces images of very high quality for a wide range of applications. Obtaining these impressive results is however not possible without proper reconstruction software. Therefore, reconstruction algorithms are also studied and developed at UGCT, and these are implemented in a highly efficient reconstruction package, called Octopus. Because of this direct access to the reconstruction, it is possible to always obtain the best possible image quality. It is due to this synergy between hardware and software development that UGCT has become a centre of excellence in high resolution X-ray

tomography.

Despite these already impressive results, it is the inherent nature of people, especially researchers, to constantly demand for more. In case of high resolution tomography, this means higher resolution, improved quality and increased flexibility. This work tackles some of the most frequently encountered problems in high resolution CT by studying and developing reconstruction methods that are more advanced than the current standard reconstruction algorithms. The focus of this research lies on the practical application of such improved methods, rather than on the theoretical background. This means that only those methods are considered which are practically feasible and which can be immediately applied to the currently available hardware. The investigated methods thus need to allow an efficient implementation and they should not require any changes in the scanner set-up or complicate the scanning process.

1.1 Put briefly

The research presented in this work was performed at UGCT. Even though research in the fields of hardware and software is quite different, at UGCT there is always a very strong interaction between both aspects of CT, since both areas of research are geared towards the same goal: the continuous improvement of high resolution X-ray CT. The aspirations of this work can be stated as follows:

> **To improve the quality and expand the application area of high resolution X-ray computed tomography, by investigating and developing improved reconstruction algorithms, bearing in mind the practical feasibility in every-day applications.**

Essentially, the following three aspects are handled in this work: helical cone-beam geometry, iterative reconstruction and phase contrast. Each of these subjects is presented and discussed in a separate chapter. At the end of each chapter, several applications are presented which illustrate the gain in reconstruction quality that can be attained by using the improved reconstruction algorithm. Even though these application sections are relatively brief compared to the discussions of the applied methods, they in fact constitute the backbone of the presented work. After all, the ultimate goal of this work is the practical application of these methods in order to improve the quality of the reconstruction of every-day scans.

1.2 Main contributions

The main contributions of the author to the field of high resolution X-ray computed tomography, which are presented in this work are:

- The search, implementation and validation of the Katsevich algorithm.
- The application of the helical cone-beam geometry in HR[1] X-ray CT.
- The reconstruction of the four most common geometries[2] using a GPU[3].
- The development of a projection simulator.
- The study of iterative reconstruction methods.
- The efficient, GPU-based implementation of SART[4].
- The development and validation of the multiresolution reconstruction.
- The evaluation of SART, compared to the analytical reconstruction.
- The introduction of the forward model incorporating beam hardening.
- The application of SART in HR X-ray CT.
- The validation of the phase retrieval method MBA[5].
- The development and validation of the phase reduction method BAC[6].
- The application of MBA and BAC in HR X-ray CT.
- The validation of MBA and BAC at various HR X-ray CT systems.

1.3 Outline

This work opens in chapter 2 with some historical background regarding X-ray imaging, from the discovery of X-rays to the development of high resolution tomography. Next, the fundamental concepts of X-ray generation, interaction and detection are briefly explained, followed by a word on the resolution of an X-ray imaging system and a short but important notice on the flux of the X-ray beam. We proceed by reciting the typical steps that constitute the complete process of a CT scan. Besides the obvious acquisition and reconstruction of the projection data,

[1] High Resolution
[2] Parallel-, fan-, cone- and helical cone-beam geometry
[3] Graphical Processing Unit
[4] Simultaneous Algebraic Reconstruction Technique
[5] Modified Bronnikov Algorithm
[6] Bronnikov-Aided Correction

this usually also includes the visualization and analysis of the reconstructed volume. This chapter finishes by presenting the Centre for X-ray Tomography and by introducing the different high resolution scanners that are available at our facility.

Chapter 3 is the first of the three major chapters in this work. We start by introducing the fundamental concepts of analytical reconstruction and present the standard algorithms for the parallel-, fan- and cone-beam geometry. Cone-beam is the most commonly used geometry in high resolution CT and also the one used in all the scanners available at UGCT. As will be shown, under certain conditions this geometry is no longer adequate and the helical cone-beam geometry may be more beneficial. Application of this geometry is however very rare in high resolution tomography and is by no means a standard solution. In this chapter, we will briefly discuss different possible reconstruction algorithms for the helical cone-beam geometry and motivate our choice of the Katsevich algorithm, which provides the most appropriate solution for our applications. Afterwards, the Katsevich algorithm and its implementation and benefits are explained in more detail. Next, some practical issues such as beam hardening and region of interest scans are discussed. We also present our module to simulate projection images, which was used extensively in the evaluation of the reconstruction methods presented in this and the following chapters. We proceed by discussing an important aspect regarding the use of reconstruction algorithms in practice: the reconstruction speed. Several methods and devices that can be used to accelerate this process are elucidated, with a special focus on graphical processing units (GPUs). We started by implementing the parallel-, fan- and cone-beam reconstruction algorithm on the GPU. Recently, we also ported the helical cone-beam algorithm to the GPU. Next, we shortly discuss our further development of the reconstruction software Octopus, including a report on the performance of the package. To conclude this chapter, we present some applications in which the object under study was scanned using the helical trajectory. In these examples this geometry was chosen over the regular cone-beam geometry in order to avoid the typical cone-beam artefacts and to be able to scan elongated objects at the highest possible resolution.

As an alternative to the standard analytical reconstruction, iterative algorithms are proposed in chapter 4. After introducing the basic concepts of algebraic and statistical reconstruction algorithms, the focus turns towards the practical implementation of such algorithms. Several important aspects of this implementation are clarified: the projection access order, the selection of an appropriate voxel basis and the different models that can be used for the forward and backprojection of the reconstructed volume. Next, we discuss the convergence of iterative methods and take a look at the initialization of the reconstruction volume and the constraints that can be placed on the reconstructed values. We proceed by presenting simple solutions to practical problems regarding region of interest scans and region of interest reconstructions. In the following, we turn our attention towards the issue of the huge memory requirements of iterative methods, which poses one of the two biggest obstacles to applying such methods in the reconstruction of data from high resolution CT systems. To overcome this problem, we developed a multiresolu-

tion approach, which allows us to divide the total reconstruction volume in smaller subvolumes, hence decreasing the required memory. The second major obstacle, the reconstruction speed, is handled by presenting an efficient, GPU-based implementation of the simultaneous algebraic reconstruction technique. Subsequently, we present the results of a series of tests that was performed to evaluate the performance of our implementation as compared to the standard analytical reconstruction. These tests are performed using projection data that are somehow limited, e.g. a low number of projections, significant contributions of noise, irregular sampling of the projection angles, etc. We also propose a modification to the forward projection model, which allows reducing artefacts resulting from beam hardening due to the presence of highly attenuating inclusions. We conclude this chapter by presenting results of applications where our algebraic algorithm was used to improve the quality of the reconstructed images.

The last major chapter 5 concerns the appearance of phase contrast. We start by introducing the basic concepts of phase contrast imaging. Two different theoretical approaches of phase contrast generation are discussed, the ray optical and the wave optical approach, followed by an explanation of the concepts of spatial and temporal coherence. Next, we address the different methods that can be used to generate and measure phase contrast. We then illustrate the detrimental effect of reconstructing projections containing both phase and attenuation contrast using conventional reconstruction methods. In the remainder of this chapter, we elucidate two methods which allow us to appropriately process phase contrast data. The first method is based on phase retrieval and is extensively evaluated using simulated projection data. We then discuss our development of a second, complementary method, based on phase reduction, which provides a solution to cases where our phase retrieval method is no longer adequate. Next, some practical concerns for both our phase retrieval and phase reduction are handled. At the end of this chapter, several applications are shown which illustrate the potential of both our phase processing methods. Besides our own data, we also evaluated the validity of these methods using data acquired at other high resolution facilities.

We conclude this work in chapter 6 by summarizing our obtained results and by providing an outlook on possible future research.

Science is the belief in

the ignorance of the experts.

Richard Feynman, 1918° – 1988†

2

X-ray Computed Tomography

For centuries, researchers have been developing various techniques to acquire information about the fascinating world that surrounds mankind. These tools range from the radio telescopes used to study distant stars and the universe, to the particle accelerators used to investigate the fundamental constituents of matter. Many of these devices are based on the emission and/or detection of electromagnetic waves, of which the wavelength depends on the size of the object(s) under study, as illustrated by figure 2.1.

X-ray computed tomography was originally developed for medical purposes, as a tool to investigate the interior of the human body without surgically opening it. Apart from X-ray CT, there exist a range of other noninvasive imaging modalities. In ultrasound computed tomography (UCT), ultrasonic waves are used to image acoustic properties of the human body. Techniques like positron emission tomography (PET) and single photon emission computed tomography (SPECT) are based on the measurement of gamma-rays emitted inside the body. Magnetic resonance imaging (MRI) uses the difference in magnetic spin properties of tissues to investigate the interior of the body.

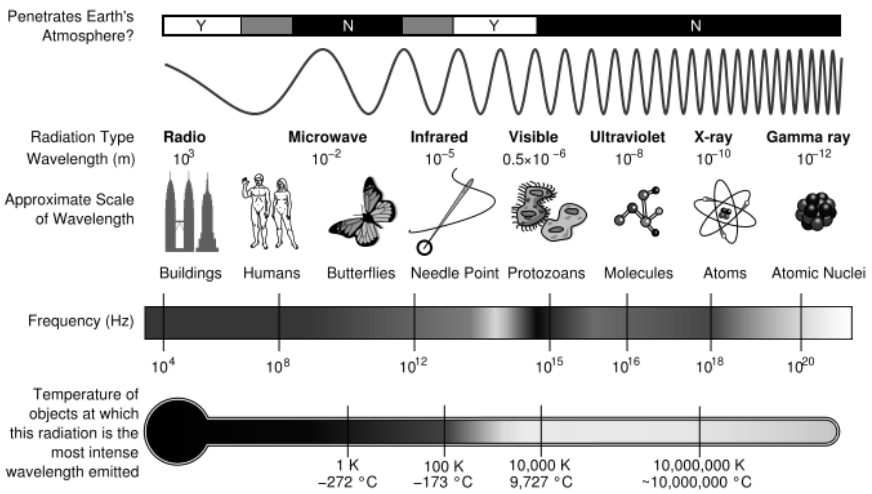


Figure 2.1: Electromagnetic waves with different wavelengths are used to study objects of different sizes. (From [1])

2.1 A brief history

2.1.1 Radiography

The history of CT goes back to 1895, the year in which the German physicist Wilhelm Conrad Röntgen discovered X-rays while he was investigating the external effects of various types of vacuum tubes. He originally published his results in the paper "Über eine neue art von Strahlen" ("On a new kind of rays") [2]. Soon after their discovery, X-rays were put to use in diagnostic imaging and they are still a very important tool up to this day, accompanied by a whole range of other imaging techniques like fluoroscopy, nuclear medicine, ultrasound and magnetic resonance imaging. For discovering this new type of radiation, Röntgen was awarded the Nobel Prize in Physics in 1901 and he is considered the father of diagnostic radiology.

A *radiography* contains information on the attenuation of X-rays integrated along their path from the source in which they are generated to the detection plane where the image is recorded. It provides a two-dimensional (2D) representation of a three-dimensional (3D) structure, and hence is often referred to as a *shadow* image or *projection* image.

2.1.2 Tomography

Contrary to radiography, *computed tomography* (CT)[1] provides a complete 3D representation of body parts by taking a series of projection images at different angles, which is achieved by rotating the X-ray source and detector around the patient. The mathematical foundation for tomography was already proposed in 1917, by the Austrian mathematician Johann Karl August Radon, who introduced the Radon transform which could be used to obtain the reconstruction of an object based on its projection data. However, tomography was only put to practice much later when computers became readily available. In 1963 and 1964, Allan MacLeod Cormack, an American physicist, published two papers containing the theoretical foundations for CT. It was Sir Godfrey Hounsfield, an English electrical engineer, who independently took these theoretical calculations into a real application and invented the first CT scanner in 1972, called the EMI-scanner after his employer. For their pioneering work on CT, Cormack and Hounsfield received the Nobel Prize in Physiology or Medicine in 1979.

The first clinical CT scanners were manufactured in 1974 and were dedicated to scanning the head of a patient. In 1976, the first whole body scanners became available. This first generation uses a source generating a thin pencil beam of X-rays and a single detector to measure the transmitted X-ray intensity. They use the *translate-rotate* method to acquire the projection data, by translating the source and detector simultaneously over the patient, repeating this for every projection angle. This projection data is used to reconstruct one cross-section through the patient's body. This process is repeated for several cross-sections at different positions, which takes hours to complete.

In the second generation of CT scanners, multiple detectors are installed and the source generates a fan shaped beam. They again use the *translate-rotate* motion, but a significant decrease in scanning time is achieved.

Scanning times are reduced to below 10 seconds per cross-section by eliminating the translational movement. In these third generation scanners, the fan-beam is directed onto an array of detectors that are fixed relative to the X-ray source.

The fourth generation scanners use a fixed ring of detectors, completely encircling the patient. Since now only the X-ray tube needs to be rotated, the complexity of the design is greatly reduced. However, scanning times are still limited since the X-ray tube and other components are interfaced by cable, so the gantry always needs to return to its initial position after rotating along a full circle. This set-up is still restricted to scanning separate cross-sections through the body.

This problem is solved by the introduction of the *slip ring* technology, by which the tube can spin nonstop within the detector ring. This enables the use of helical (or spiral) CT, in which the patient table smoothly moves through the scanner while the scanner gantry rotates continuously around the patient. This method

[1] Also known as *computed axial tomography* (CAT)

significantly reduces acquisition time and allows the reconstruction of a true 3D volume, instead of a series of cross-sections. Modern CT scanners use multiple rows of detectors, decreasing scanning time even further. These multirow scanners have grown to 2, 6, 8, 16, 32, 64, 128, etc. detector rows and are evolving towards real 2D detectors. Some systems even contain multiple X-ray tubes to speed up the scanning process.

All of these technical innovations, which are aimed towards decreased scanning time and dose deposition and increased quality and patient comfort, would not be possible without advances in computer science. As each of these developments poses higher demands from a computational point of view, their practical feasibility is heavily dependent on the evolution in high performance computing. Currently, the fastest state-of-the-art devices can scan parts of a patient in real-time, meaning both acquisition and reconstruction are performed almost instantly.

Given the historical background, it is clear that, even after all these years, the development of CT is only possible by a close collaboration between mathematicians, physicists, engineers and computer scientists. Without any of these, CT would have never evolved into the powerful instrument it is today.

2.1.3 High resolution

In addition to the medical field, X-ray CT was also quickly introduced in industrial applications for quality control and nondestructive testing. Using the right equipment, the interior of large and heavily attenuating objects such as cars, engines and turbine blades can be investigated.

Both industrial and medical CT however only offer a limited resolution, since they are used to study rather large objects. The development of CT with a resolution within the micrometer scale quickly caught the attention of researchers from a wide variety of fields such as biology, pharmacy, engineering, palaeontology, geology and many more. When the resolution of the CT system is in the order of micrometers, it is referred to as microCT, also known as μCT, microtomography and similar terminologies. The term nanoCT is used once a resolution below one micrometer is obtained, although the term submicroCT is more appropriate. For simplicity, both microCT and submicroCT (or nanoCT) are referred to as high resolution CT in this work. Unlike medical CT, in industrial and high resolution CT it is often more feasible to rotate the object and keep the X-ray source and detector fixed.

The first high resolution X-ray CT system was built in the early 1980s by Elliot and Dover [3]. Like the first medical scanners, it used a pencil beam with a low flux and the *translate-rotate* method. This made the acquisition for a complete volume very slow, from several hours to days. Similar systems at that time suffered from the same problems. A newer generation of scanners used the fan-beam geometry to reduce scanning time, but it was not until the development of third

generation scanners that high resolution CT became a viable research tool. These scanners combined an X-ray tube with a small focal spot size with a 2D detector, resulting in the detection of a conical X-ray beam. Although the flux of such a tube is still rather low, the scanning time was significantly decreased since the projection data of the whole volume can now be measured by a single rotation. An advantage of the fan-beam and cone-beam geometries is that the projection image of an object can be magnified by moving the object closer to the X-ray source. Thus, by increasing the magnification, the resolution of the scanner can be improved. In practice, the use of high resolution cone-beam scanners only proceeded after Feldkamp, David and Kress presented their algorithm that allowed for fast and efficient reconstruction of cone-beam data [4]. As with medical CT, the use of the helical cone-beam geometry can offer several advantages and is currently making its way into the field of high resolution CT as well. High resolution CT can also be performed at certain synchrotron beamlines. Although these provide several advantages over an X-ray tube (higher X-ray flux, coherent beam, etc.), such installations are more difficult to access and are typically used in one-off applications.

2.2 Fundamental X-ray concepts

X-rays and γ-rays are a form of high energetic electromagnetic waves or photons (wave-particle duality). Their energy E is related to the wavelength λ or frequency ν by:

$$E = \frac{hc}{\lambda} = h\nu \,, \tag{2.1}$$

where h is Planck's constant, equal to $4.136 \cdot 10^{-15}$ eV s and c the speed of light in vacuum, equal to $3 \cdot 10^8$ m / s. X-rays are defined as electromagnetic radiation emitted by charged particles, while γ-rays are emitted by the nucleus in processes of radioactive decay or are created in annihilation processes[2].

2.2.1 Generation

X- and γ-rays can be created by a number of different processes. The processes relevant for tomography are briefly discussed here.

Radioactive decay

When a nucleus decays, it can be left in an excited state. To return to a more stable state, the nucleus emits high energy photons, γ-rays, which carry away the

[2] Older literature distinguishes between X- and γ-rays based on the wavelength, nowadays they are distinguished by their origin.

excessive energy. Since the difference in energy between the excited and stable state is constant for a certain decay process, the emitted photons always have the same energy, so a γ-ray source can be used as a mono-energetic source. Due to the production of a monochromatic beam containing high energy photons, γ-sources can be an alternative to X-ray tubes. However, using γ-sources results in a relative low flux and it is difficult to obtain a small focal spot size. Furthermore, safety issues have to be taken into account when using radioactive sources. Application of γ-ray sources is usually limited to industrial tomography.

Synchrotron radiation

In a synchrotron[3], charged particles (usually electrons) are accelerated to very high energies and injected in a quasi-circular storage ring, consisting of straight sections and bending magnets. When a relativistic particle is deflected from its path by a magnetic field, it loses some of its energy by emitting high energy photons. When it was first observed, synchrotron radiation was seen as a nuisance, causing unwanted energy loss. Soon afterwards however, it was found that this radiation can be very useful in numerous experiments, and hence it is deliberately produced ever since.

In the first generations of synchrotrons, the electromagnetic radiation is only created by the bending magnets, creating radiation with a wide spectral distribution. To obtain monochromatic photons, crystalline monochromators are added to the experimental set-up to extract a monochromatic beam, resulting in a decrease in flux.

In the current third generation synchrotrons, devices are installed in the straight sections which are specially designed to produce higher photon fluxes and/or monochromatic radiation. These insertion devices consist of a number of periodically positioned magnets, forcing the electrons to following a sinusoidal trajectory. Two types of magnetic devices are commonly used: *undulators*, which produce an almost monochromatic photon beam and *wigglers*, which produce a polychromatic beam of high intensity.

Synchrotron radiation offers some unique advantages, making it highly suitable for X-ray tomography. It consists of an almost parallel beam of high intensity, providing sufficient statistical information at relative small scanning times even at a position far away from the source. The radiation is also spatially coherent, as it originates from a very small area. This allows imaging using wave-related effects, which can be used to increase resolution and contrast, especially in low attenuating materials. Furthermore, synchrotron radiation can be used to produce an almost monochromatic beam with an energy that is tunable to some extent. The downside is that synchrotron installations are quite expensive and accessibility is limited. Furthermore, magnification of the imaged object can only be achieved using complex X-ray optics.

[3] A synchrotron is a type of particle accelerator.

X-ray tube

Although many different types of X-ray tubes exist, they are all based on the same principles. This is illustrated by figure 2.2, which schematically represents the functioning of the famous Coolidge tube. An electric current is used to heat a filament[4], the cathode (K), which emits electrons due to the thermionic effect. These electrons are accelerated in a vacuum tube towards a target plate[5], the anode (A), by applying a high voltage U_a between cathode and anode, resulting in a current. These fast electrons collide with the target and deposit their energy in it. A small amount of the deposited energy is used to generate X-rays, which emit from the target and escape the tube through an exit window. The rest of the energy is released as heat.

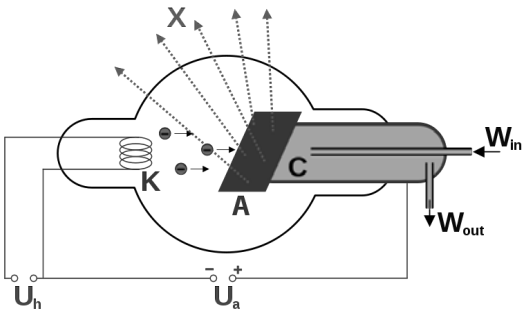


Figure 2.2: Schematic representation of a Coolidge tube. (From [5])

By using electromagnetic lenses, the electrons can be focused onto the target such that the X-rays are generated within a small area of the target, which is called the *focal spot* of the tube. Since the spot heats up due to the energy deposit of the electrons, the electron current needs to be limited to prevent the target from melting. A smaller spot size thus requires a lower current, which implies a decrease in the number of photons that are generated in the target, thus a lower X-ray flux. Depending on the design, the target plate can be cooled (C) which allows for a higher electron current and X-ray flux.

Two types of targets exist, *transmission* and *directional* targets. In a transmission target, the generated X-rays escape the tube along the same direction as the incoming electrons. This makes it possible to achieve a very small spot size, but prevents direct cooling of the target, thus only allowing for a small flux. In a directional tube, the target plate is tilted with respect to the incoming electrons, as illustrated by the Coolidge tube in figure 2.2. The X-rays now emanate from the tube at a certain angle relative to the incoming electrons. Using this type of target the spot can be cooled directly from the back of the target, allowing for a much higher X-ray flux, at the expense of an increased focal spot size.

[4] Usually tungsten
[5] Usually tungsten or molybdenum

When an electron beam strikes the target material, X-rays can be created by two processes:

Characteristic radiation: An incoming electron can collide with a shell electron, transferring a part of its energy to the struck electron, which is dissipated into heat. The majority of the incoming electrons interacts by this process, which accounts for the heating of the target. A fraction of these collisions results in the removal of the shell electron, leaving a hole in the shell. This gap is immediately filled by a higher shell electron dropping into the hole while emitting a photon of a specific energy. The energy of the photon is well-defined and equal to the difference in energy between the two electron states, yielding a characteristic peak in the emitted spectrum.

Bremsstrahlung: An electron can also interact with the nuclei of the target material by Coulomb interaction, losing a significant amount of energy by emitting a photon. Bremsstrahlung yields a continuous X-ray spectrum, where the energy of the emitted photons lies between 0 and $E_{max} = q\,U$, with q the electric charge of an electron and U the high voltage of the tube.

2.2.2 Interaction

Particle properties

X-rays can interact with matter in a number of different ways:

Photo-electric absorption: A photon can transfer all of its energy to a shell electron, ejecting it from the atom. The free electron rapidly loses its energy and move only a short distance from its original location. This interaction is only possible when the energy of the photon is higher than the binding energy of the electron. The remaining energy is converted into kinetic energy of the ejected electron. The interaction probability for photo-electric absorption τ can, for the typical energies encountered in X-ray CT (5 to 150 keV[6]), be approximated by:

$$\frac{\tau}{\rho} \propto \left(\frac{Z}{E}\right)^3 , \tag{2.2}$$

where ρ is the mass density, Z the atomic number of the element and E the energy of the photon.

Compton scattering: A photon can also interact with an atomic electron, transferring some of its energy. Since the photon leaves the interaction site in a different direction, this interaction is classified as a scattering process. Compton scattering in the object can be an undesirable effect in imaging, as some of the deviated photons reach the detector, making the object a radiation scatterer, which

[6]kilo electron volt

may distort the image quality. The interaction probability for Compton scattering σ can be described by:

$$\frac{\sigma}{\rho} \propto \left(\frac{Z}{A}\right) f(E) \,, \tag{2.3}$$

where A is the mass number and $f(E)$ is an energy dependent factor that can be calculated using the Klein-Nishina approximation.

Rayleigh scattering: In this process, the photon is scattered by the whole electron cloud instead of a single electron. This is an elastic scattering process, in which no energy is transferred. Although this interaction occurs at low energies and results in relative large scattering angles, it generally poses no significant contribution.

Pair production: In this process the photon is absorbed creating an electron and a positron. Since this interaction only occurs for photon energies above 1.022 MeV, it is not relevant in high resolution CT.

Nuclear reactions: At even higher photon energies (several MeV), photons may induce nuclear reactions. Again, considering the related energies, this interaction is irrelevant in the present context.

The photo-electric absorption and the Compton scattering are the most relevant interactions for CT. The relative contribution of both processes depends on several factors. Photo-electric absorption depends mainly on the atomic number of the elements constituting the object. Compton scattering is largely independent of atomic number and can only be used to detect variations in the electron density of the composition. The scattered photons are redistributed in space over a relatively large solid angle. Hence, when the distance between the object and the detector is large, only a small fraction of the scattered photons reaches the detector and the distribution in the detector plane is nearly uniform. Therefore, in high resolution CT Compton scattering can be considered as an attenuation process and the distortion of the projection image by scattering can usually be ignored. When this scattering distortion is no longer negligible, the use of collimators can prevent scattered radiation from reaching the detector and degrading the image quality. However, the use of collimators is not feasible when using 2D detectors and a conical bundle. Therefore, in high energy CT, in which Compton scattering is the dominant effect, usually line detectors are used.

Wave properties

Due to the wave-particle duality, all photons, including X-rays, are subject to wave-related effects. Every wave can be described by its amplitude, wavelength (or frequency) and phase. The interaction of a wave with a medium is determined by the complex refraction index $n = 1 - \delta + i\beta$, where β is responsible for the attenuation of the wave and δ for the phase shift, which is due to a difference in propagation speed between the medium and vacuum. This phase shift causes a

deformation of the wavefront as the part traversing the medium moves at a different speed. When the studied features are much larger than the wavelength, the ray optical approach can be used to represent the waves. Using this approach, each part of the wavefront can be represented by a ray perpendicular to the wavefront. The introduction of a phase shift can then be seen as a change in the direction of the incoming ray when it goes from one medium to another, an effect which is called *refraction*. In addition, at distances further away from the medium, interference between the original and the deformed wavefront results in a complex pattern of intensities, called *diffraction*, for which the ray optical approach is no longer valid. Both refraction and diffraction effects are inevitably encountered in high resolution X-ray imaging systems and should be appropriately accounted for.

Law of Lambert-Beer

Consider a monochromatic X-ray beam of intensity I, which is proportional to the number of photons per unit time and unit area, and an infinitesimally thin slab of thickness ds. The slab consists of a material with linear attenuation coefficient $\mu = \tau + \sigma$, which combines both contributions of photo-electric absorption and Compton scattering. The change in intensity of the beam after passing through the slab is then given by [6]:

$$\frac{dI}{ds} = -\mu I \tag{2.4}$$

Integrating this along the path L from the source to the detector position yields the law of Lambert-Beer:

$$I = I_0 \cdot \exp\left[-\int_L \mu(s)\, ds\right] , \tag{2.5}$$

where I_0 is the unattenuated beam intensity and where the linear attenuation coefficient $\mu(s)$ depends on the material composition at position s along the path L. Even though the individual interactions of photons with matter are of statistical nature, the macroscopic intensity of the beam can thus be described using a deterministic exponential law.

Note that this formula, which is the basic equation in CT, is only valid for attenuation processes and for a monochromatic beam. Some of the most important artefacts that arise in high resolution CT are due to a violation of these conditions.

2.2.3 Detection

The detection of X-rays can be direct or indirect. Although direct detection offers several advantages, due to technical issues the most commonly used detectors in practical high resolution systems are based on indirect detection.

Indirect detection

In an indirect detector, incoming X-rays interact with a scintillating material, converting their deposited energy into visible light, which is guided to a sensor and converted into an electric signal. As this signal is roughly proportional to the deposited energy, it contains combined information regarding the incident flux and its spectrum. Disadvantages of this detection mechanism are that spectral information is typically lost and the scattering of light inside the scintillator limits the resolution of the detector. However, indirect detectors are predominantly used in X-ray tomography due to the following advantages.

First, by varying the scintillator material and thickness, the energy response of the detector can be optimized to fit the current application. Second, visible light detectors are widely available so it is possible to select the best suited sensor. Third, X-rays passing through the scintillator are usually stopped in the light guiding system, minimizing radiation damage to the sensor.

Direct detection

In a direct detector, incoming X-rays interact in the sensitive layer of the detector, creating electron-hole pairs which are collected by an electric field. The generated charge can then be integrated on the pixel or can be immediately processed for each interaction event. This makes it possible to count every single photon and derive its energy, offering interesting possibilities in imaging. However, in practical applications, a direct photon detector needs to overcome some problems.

First, the charge deposited by a photon needs to be read out before a second photon interacts with the detector. This means that the incoming X-ray flux needs to be sufficiently low in order to match the frame rate of the detector. Second, to determine the energy, all of the photon's energy needs to be deposited into the detector, so the photon needs to interact through photo-electric absorption. As most detectors are composed of silicon, only low energy X-rays can be detected. Therefore, direct detectors are currently limited to application in high resolution tomography of low absorbing objects. Third, as usually a pixellated detector is used, the charge needs to be contained within one pixel, otherwise the interaction could be interpreted as several lower energy events.

Recently, several direct detectors were developed that are capable of overcoming these problems to a certain extent. These detectors have already been used at some experimental set-ups, offering promising results. It is believed that they will become common use in the near future, thereby expanding the possibilities of high resolution X-ray tomography.

2.2.4 Resolution

In a high resolution scanner, the object under investigation is usually installed on a translation stage, which is used to position the object. When using a conical X-ray beam, the projection images of small objects can be magnified by moving the object closer to the X-ray source. Large objects on the other hand are positioned close to the detector, in order to capture the projection image completely. The object magnification M is then given by:

$$M = \frac{SDD}{SOD} \, , \tag{2.6}$$

where SOD is the distance between the source and the object, and SDD is the distance between the source and the detector. In high resolution CT, M is rather large, typically between 10 and 100. The resolution of the imaging system is mainly limited by the focal spot size d_s of the X-ray source. Since the spot size is magnified as well, at high magnification the finiteness of the spot size becomes perceptible and the projection image becomes blurred, as illustrated in figure 2.3. Furthermore, the resolving power d of the detector[7] also limits the resolution of the system. A general expression to determine the best achievable resolution R of the imaging system is then given by:

$$R = \frac{d}{M} + \left(1 - \frac{1}{M}\right) d_s \, . \tag{2.7}$$

This formula complies with the fact that the achievable resolution of a system can never be better than the focal spot size, as even for very high magnification $M \to \infty$, $R \geq d_s$.

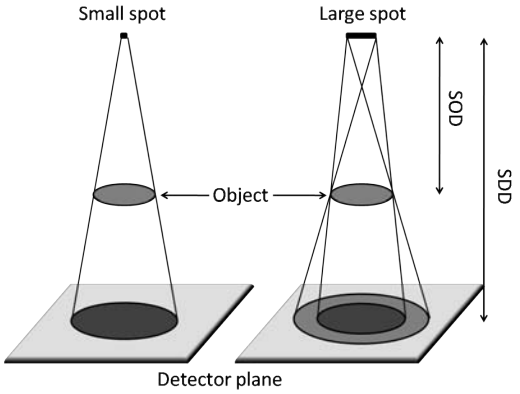


Figure 2.3: Illustration of the influence of the focal spot size on the projection image.

[7]The resolving power is defined as the smallest separation distance between two features (lines or points) at which they can still be distinguished separately.

2.2.5 Flux

Another important concept in X-ray imaging is the X-ray flux. The number of photons that pass through the object and are detected within a certain time frame depends on the generated flux, which thereby determines the amount of statistical information. A higher flux offers the advantage of either obtaining an improved signal to noise ratio of the reconstructed object, or requiring a shorter scanning time while maintaining the same statistical information. In high resolution CT using an X-ray tube, the flux is typically very low, in order to obtain a small enough spot size without melting the target plate. The filament usually poses additional constraints to the attainable flux. Therefore, scanning times are typically much longer than in medical or industrial CT. Currently, high resolution CT using a high flux is only possible at synchrotron installations.

As an X-ray tube emits a conical bundle, the intensity of the bundle decreases quadratically when moving further away from the source. This means that, within a given timespan, a detector far from the source detects less photons per pixel than the same detector positioned close to the source. In positioning the detector, a well considered trade-off has to be made. Moving the detector further away decreases statistical information (increases scanning time), decreases the cone angle (thus reducing cone-beam artefacts) and allows for higher magnification.

2.3 Process of a CT scan

A tomographic scan basically consists of two processing steps: the acquisition, in which the object is rotated and a series of projection images is taken, and the reconstruction, which calculates the 3D representation of the object based on the projection images. The 3D volume that is obtained can then be visualized, processed and analysed further.

A typical X-ray CT scanner consists of three main components, as illustrated in figure 2.4: an X-ray source that generates X-rays in a conical bundle, an X-ray detector which is used to acquire the projection images and a motor system on which the object is placed, usually consisting of a rotation motor stacked on a set of translation stages.

2.3.1 Acquisition

The actual CT scan consists of the acquisition of three sets of projection images:

- *Dark fields*: These are images acquired when the X-ray tube is turned off. They contain possible pixel offsets and dark current contributions. The exposure time to record these images is best chosen equal to the exposure time of the projection images.

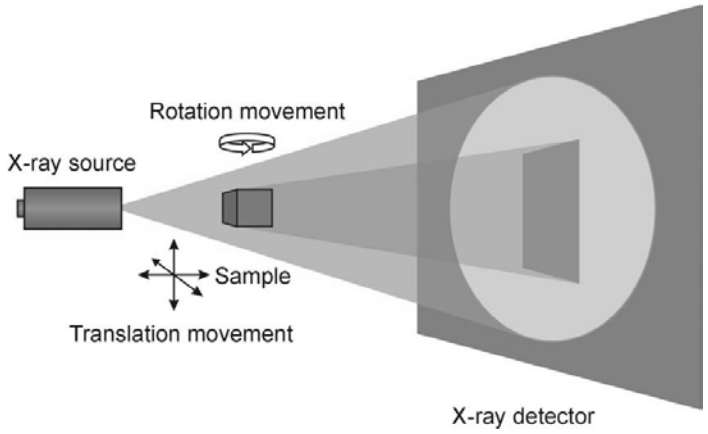


Figure 2.4: Schematic representation of a typical X-ray scanner.

- *Flat fields*: These are taken when the X-ray tube is turned on, but with the object positioned outside the field of view of the detector, to correct for inhomogeneities in the X-ray beam profile.
- *Projections*: A set of projections is taken by rotating the object over a fixed angular step between two consecutive projections. The range of this rotation can be 180° plus the fan angle of the beam for a short scan or 360° for a full scan. More advanced acquisition geometries may require rotation over more than 360° and/or additional movements.

The spectrum of the X-rays can be altered to better suit the current experiment by varying the high voltage inside the X-ray tube. Furthermore, filters (thin plates of e.g. copper, aluminum, etc.) can be installed to remove low energy photons and harden the spectrum of the incoming X-ray beam.

2.3.2 Reconstruction

Based on these three sets of projection images, the *normalised* projection images can be obtained, which contain values between 0 and 100% transmission, by:

$$N(i,j) = \frac{P(i,j) - D(i,j)}{F(i,j) - D(i,j)}\,, \tag{2.8}$$

where $N(i,j)$ corresponds to the normalised intensity value of pixel (i,j), and where D, F and P are the dark field, flat field and projection images, respectively. The acquired images can also be filtered by applying a spot filter or a ring filter, to avoid certain artefacts that would otherwise arise during reconstruction.

Using an appropriate reconstruction algorithm, the projection images are then used to calculate a 3D representation of the scanned object, consisting of a number

of volume elements (voxels), where each voxel contains information on the linear attenuation coefficient of the object at that position.

Reconstruction can be performed using a number of different reconstruction methods such as analytical, algebraic or statistical routines. Many different approaches exist, each with their benefits and shortcomings.

2.3.3 Visualization and analysis

The acquisition and reconstruction steps make up the actual process of a CT scan. However, in most cases scanning the object is only a part of the intended study and the resulting 3D volume is used in further research. Using appropriate software, the 3D volume can be visualized and one can for instance render certain parts of the object transparent or cut through it virtually. This allows the user to observe the internal structure of the object in a noninvasive way. In engineering applications, this 3D volume is often used to generate a geometric model of the object, which can then be used in finite element simulations. The volume can also be analysed using certain routines to obtain quantitative information of the object, e.g. measurements of distances, the size and orientation of grains, the complexity of a pore network, etc.

2.4 The Centre for X-ray Tomography

The Centre for X-ray Tomography of the Ghent University (UGCT) was founded in 2006 as a collaboration between two research groups: the *Radiation Physics* group of the Department of Physics and Astronomy and the *Sedimentary Geology and Engineering Geology* group of the Department of Geology and Soil Science. Recently, a third research group has joined the collaboration, the *Laboratory of Wood Technology* of the Department of Forest and Water Management.

UGCT is an open facility, offering scientists and engineers access to nondestructive X-ray imaging and image analysis for visualizing features in the interior of objects in 2D and/or 3D. It is involved in research projects in the fields of biology, geology, palaeontology, pharmacy, engineering and several other domains. UGCT develops the reconstruction software package *Octopus* and the 3D analysis package *Morpho+*. The spin-off company *inCT* was founded in 2008, which offers X-ray services to companies and which is responsible for the commercial distribution of Octopus.

The main research field of UGCT is the development and application of *state-of-the-art*, high resolution CT scanners. To overcome the limitations of standard desktop systems, all components are carefully selected taking into account the need for optimal performance and flexibility. The research on high resolution tomography also requires the study of reconstruction algorithms, in order to obtain

highly efficient reconstruction methods that are capable of reducing a number of typically encountered artefacts. The development of analysis software is driven by the need for an accessible and modifiable tool, containing specific routines for each application. It is this synergy between hardware and software developments that has made UGCT a centre of excellence in the world of high resolution X-ray tomography.

At this moment, there are three X-ray scanners available at the facility: the *900nmCT-*, *400nmCT-* and *EMCT-scanner*. Each of these scanners was designed for a specific application area, while maintaining a high degree of flexibility.

2.4.1 900nmCT-scanner

This device, shown in figure 2.5, is the first high resolution scanner developed at UGCT. The X-ray source is a Feinfocus dual head system, using a transmission head for high resolution applications and a directional head for larger objects. The best achievable focal spot size (and thus resolution) of this set-up is about 900 nm.

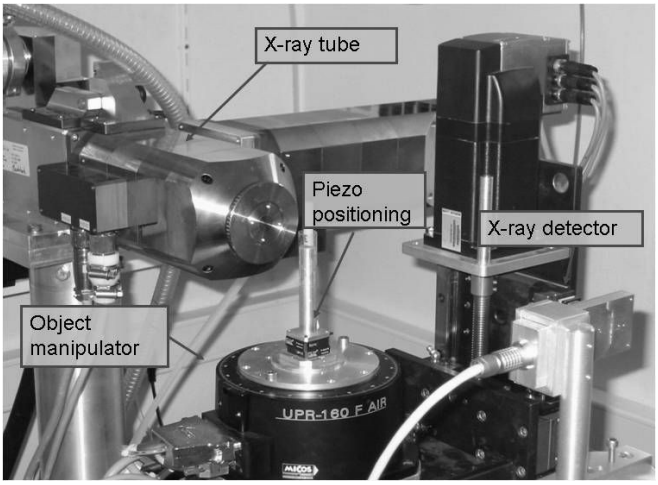


Figure 2.5: Picture of UGCT's 900nmCT-scanner.

Different motors are used to manipulate the object. A high precision, air-bearing motor stage, from Micos, is used to rotate the object during the scan. As the object would possibly rotate outside the field of view of the detector during the scan if positioned manually, a 2-axis micropositioning system (Piezo positioning) is mounted on the rotation stage, allowing accurate centring of the object onto the axis of rotation. The object itself is usually mounted on a rod, allowing it to be positioned very close to the source. The rotation stage is mounted on a stacked 3-axis translation system, which is used to position the object in and out of the beam and to control the magnification. In addition, the translation stage is also used to

alter the vertical position of the object during a helical cone-beam scan. This geometry was introduced following the results on helical cone-beam presented later in this work, in order to avoid the appearance of cone-beam artefacts and to enable the scanning of long objects. However, since the vertical stage was selected well before helical cone-beam was considered, its range and accuracy limit the applicability of this new geometry at this scanner.

The X-ray detector shown here is a Rad-Eye CMOS flat panel, which is just one of many detectors available at the facility, others are for instance the Photonic Science VHR, the Varian 2520V Paxscan and the Medipix2. Switching the detector is simply done by manually replacing the detector and adjusting a single setting in the control software of the scanner.

2.4.2 400nmCT-scanner

Based on the experience obtained by building the 900nmCT-scanner and driven by the quest for higher resolution, a new scanner was developed at UGCT (figure 2.6), with a resolution down to 400 nm. It uses two separate Hamamatsu X-ray tubes mounted on a translation stage, which allows to quickly switch between a nanofocus tube and a high power microfocus tube. Two different detectors, a high resolution CCD camera and a large flat panel are installed on another translation stage, allowing to switch between detectors without any manual manipulation.

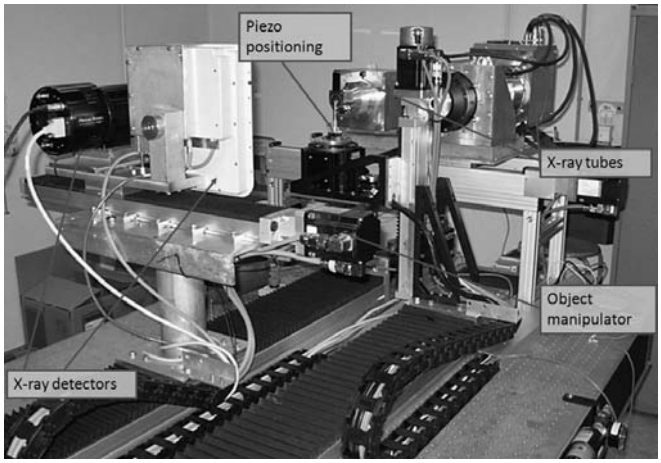


Figure 2.6: Picture of UGCT's 400nmCT-scanner.

As this scanner was specifically designed for wood research, where often long objects need to be scanned, the helical cone-beam geometry was incorporated into the design of the scanner and the vertical translation motor was chosen to have a much larger range and better accuracy than the one in the 900nmCT-scanner.

Furthermore, the entire scanner is mounted on a damped optical table, to reduce the influence of environmental vibrations.

2.4.3 EMCT-scanner

An emerging demand in high resolution CT is the installation of additional hardware in the scanner, which can be used to control the object's environment (temperature, atmospheric pressure and composition, etc.) or to apply external forces to it (pressure, stretch, torsion, etc.) during the scan. Since such devices require all kinds of external connectors, it is not feasible to install them on the rotation stage. Therefore, a new type of scanner, the Environmental CT scanner (EMCT), is under development at UGCT, where the X-ray source and detector now rotate around a stationary object, comparable to the medical scanners, but with much higher resolution.

To achieve this, the source and detector are mounted on a high precision gantry (figure 2.7). Source and detector are installed on a translation stage, allowing control over the magnification. The gantry contains a central opening, centred above a vertical translation motor, on which the object (and an optional manipulation device) is mounted. This vertical motor is used to bring the object into the X-ray beam, and can also be used during the scan to achieve the helical cone-beam geometry. A small computer is also installed on the rotating part of the gantry, which runs the control software of the scanner and is used to store the projection images during acquisition. Sending instructions to and receiving feedback from this PC is done using wireless communication. Power to these devices is supplied through a custom designed slip ring.

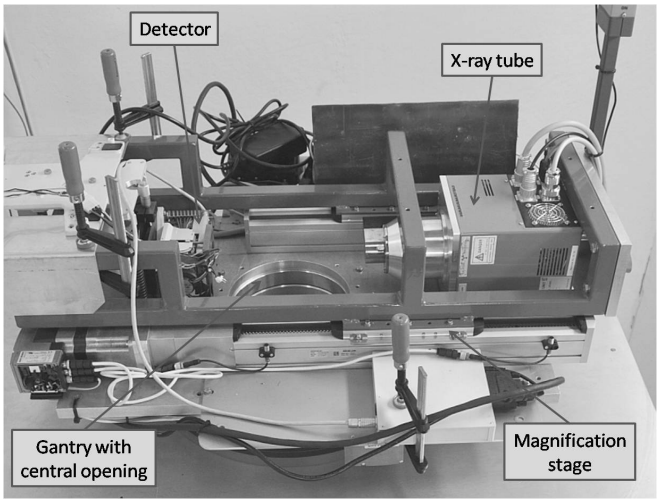


Figure 2.7: Picture of UGCT's EMCT-scanner.

References

[1] *http://en.wikipedia.org/wiki/Electromagnetic_spectrum*.

[2] W. C. Röntgen. *On a new kind of rays*. Würzburg Physical and Medical Society, 1895.

[3] J. C. Elliot and S. D. Doover. *X-ray microtomography*. Journal of Microscopy, 126:211–213, 1982.

[4] L. A. Feldkamp, L. C. Davis, and J. W. Kress. *Practical cone-beam algorithm*. Journal of the Optical Society of America A, 1(6):612–619, 1984.

[5] *http://en.wikipedia.org/wiki/X-ray_tube*.

[6] M. M. Ter-Pogossian. *The physical aspects of diagnostic radiology*. Harper and Row, New York, 1967.

I have had my results for a long time,

but I do not yet know how I am to arrive at them.

Karl Friedrich Gauss, 1777° – 1855†

3

Analytical Reconstruction

This chapter introduces the basic concepts of the analytical reconstruction algorithms that are generally used to recover the cross-sections of an object from its projection data. Different geometries can be used for the acquisition of the data. Four geometries are discussed that are most relevant for this work, being parallel-, fan-, cone- and helical cone-beam. Parallel-beam is used in X-ray tomography set-ups at synchrotron beamlines. Fan-beam is used in the older generations of medical CT scanners and also in high energy systems, where a line detector is used with a collimator attached to exclude the contribution of photons scattered in the object or the scanner. The most common geometry used in high resolution X-ray CT systems based on an X-ray tube is cone-beam. Almost every high resolution scanner, both commercial and academic, uses this geometry. The fourth geometry is the helical (or spiral) cone-beam, which can offer several benefits over the regular cone-beam. This geometry is steadily making its way into the world of high resolution tomography and is already included in some of the latest high resolution X-ray CT scanners. The high resolution scanners developed at UGCT mainly use cone-beam geometry but were also equipped with the possibility to perform a helical cone-beam scan, which can be used to improve the image quality in certain applications.

3.1 Fourier slice theorem

This chapter starts by explaining the reconstruction of a two-dimensional object. In tomography, a set of line integrals $P_\theta(t)$ given by:

$$P_\theta(t) = \int_L f(x,y)\, ds\ , \tag{3.1}$$

is used to reconstruct the object function $f(x,y)$ which represents some physical property of the object. The parameter θ is the angle under which the projection is taken and t is the position on the detector, the integration is performed along the line L, which is defined by:

$$x\ \cos\ \theta + y\ \sin\ \theta = t\ . \tag{3.2}$$

In X-ray CT, the function $f(x,y)$ represents the X-ray linear attenuation coefficient of the material, which will be noted as $\mu(x,y)$. For a monochromatic beam, the relation between the intensity of the X-ray beam before $I^0_\theta(t)$ and after $I_\theta(t)$ transmission through the object is given by the law of Lambert-Beer:

$$I_\theta(t) = I^0_\theta(t) \exp\left[-\int_L \mu(x,y)\, ds\right]\ , \tag{3.3}$$

or equivalently:

$$\int_L \mu(x,y) ds = -\ln \frac{I_\theta(t)}{I^0_\theta(t)}\ . \tag{3.4}$$

Thus the measurement of both intensities results in a set of line integrals that can be used to reconstruct the object function $\mu(x,y)$.

The basic concept on which the analytical reconstruction algorithms are based is the Fourier Slice Theorem, which states that [1]:

> The 1D Fourier transform of a parallel projection of a 2D object function $f(x,y)$ at an angle θ with respect to the X-axis, gives a slice of the 2D Fourier transform $F(u,v)$ of the function $f(x,y)$ at an angle θ with respect to the u-axis.

This theorem indicates that by taking projections of an object at different angles and Fourier transforming each of these, one can determine the values $F(u,v)$ on radial lines, as illustrated in figure 3.1. Then, by using the inverse Fourier transform, the object function $f(x,y)$ can be obtained. This results in a simple method that can be used to reconstruct the object. However, this approach has certain flaws that limit its applicability. As in practice only a limited number of projection images can be acquired, the function $F(u,v)$ is only known along a finite number of radial lines (figure 3.2). The inverse Fourier transform, needed to obtain $f(x,y)$,

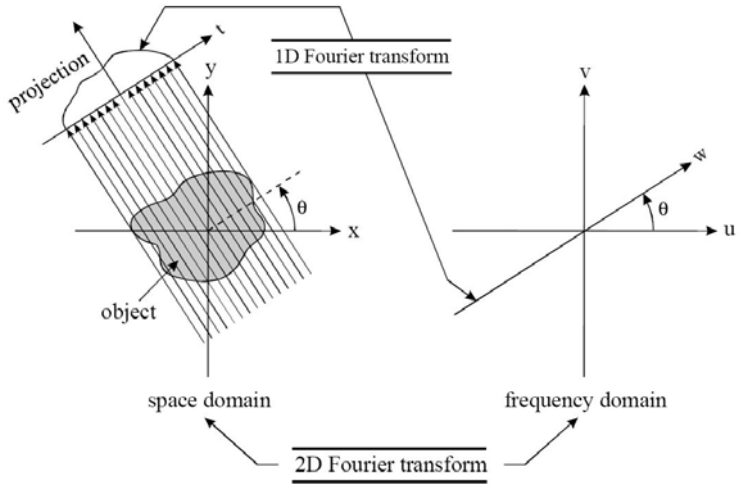


Figure 3.1: The Fourier slice theorem. (From [2])

requires the function to be sampled on a square grid, meaning one needs to use some kind of interpolation to convert from the radial to the square grid. However, simple interpolation schemes fail when used in the Fourier space. It is clear that the error of this conversion becomes larger when moving away from the centre, as the density of the radial points becomes sparser. This means there is a greater error in the high frequency components of the image than in the low frequency ones. This leads to severe degradation of the reconstructed images. Several solutions, like improved interpolation schemes, exist that tackle this problem, but these Direct Fourier Methods [3–5] have a varying degree of success when applied to real data.

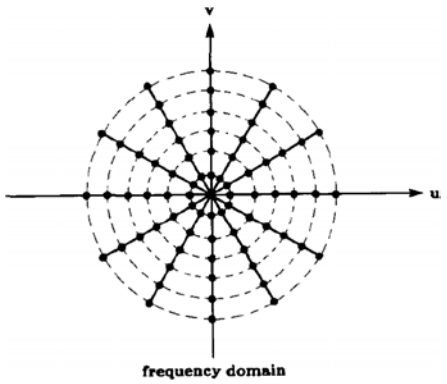


Figure 3.2: The object function in the Fourier domain, sampled along radial lines. (From [2])

From the Fourier slice theorem, another reconstruction method can be derived which is known as the filtered backprojection. Due to its accuracy and possibility of fast implementation, this algorithm has become the standard reconstruction method for straight ray tomography in parallel-, fan- and cone-beam geometry. The filtered backprojection will be introduced for the reconstruction of parallel-beam data and will then be extended to fan- and cone-beam.

3.2 Parallel-beam

As the name states, in parallel-beam the paths of the X-rays that pass through the sample are parallel to each other as shown in figure 3.3. This geometry can be accomplished for instance by moving an X-ray source and a single detector element along parallel lines on opposite sides of an object while acquiring data. Another possibility is provided by a synchrotron beamline, which delivers a parallel X-ray beam of a certain width and height.

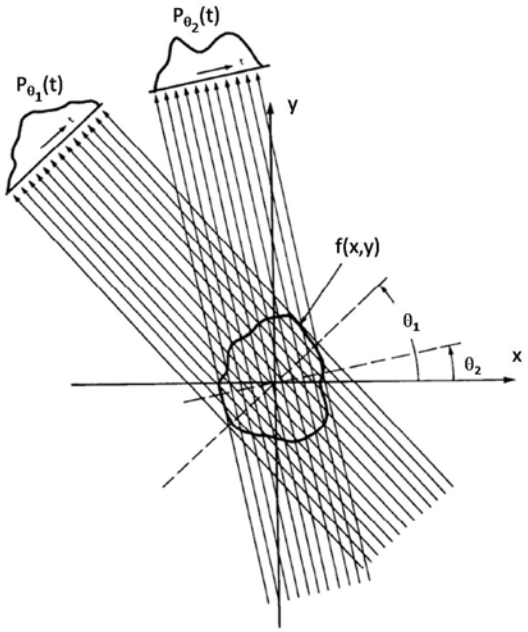


Figure 3.3: In parallel-beam geometry a projection contains the line integrals along a set of parallel rays through the object at a certain angle. (From [6])

The inverse Fourier transform of the object function $f(x, y)$ can be expressed by:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{i2\pi(ux+vy)}\,du\,dv\,. \tag{3.5}$$

By switching from the rectangular coordinate system (u, v) to a polar coordinate system (w, θ) with:

$$\begin{aligned} u &= w \cos\theta \\ v &= w \sin\theta \,. \end{aligned} \tag{3.6}$$

This can be rewritten as:

$$f(x,y) = \int_0^{\pi} \left[\int_{-\infty}^{\infty} F(w,\theta)|w| e^{i2\pi wt}\, dw \right] d\theta \,, \tag{3.7}$$

where

$$t = x \cos\theta + y \sin\theta \,. \tag{3.8}$$

According to the Fourier slice theorem, the 2D Fourier transform $F(w, \theta)$ can be replaced by the Fourier transform of the projection at angle θ, $S_\theta(w)$. Using this, the object function $f(x, y)$ can be reconstructed from the set of line integrals $P_\theta(t)$ by:

$$f(x,y) = \int_0^{\pi} Q_\theta(t)\, d\theta \,, \tag{3.9}$$

$$Q_\theta(t) = \int_{-\infty}^{\infty} S_\theta(w)|w| e^{i2\pi wt}\, dw \,, \tag{3.10}$$

$$S_\theta(w) = \int_{-\infty}^{\infty} P_\theta(t) e^{-i2\pi wt}\, dt \,. \tag{3.11}$$

The last equation (3.11) simply represents the 1D Fourier transform of the projection data at an angle θ. Equation (3.10) is a filtering operation, where the frequency response is given by $|w|$. The first equation (3.9) represents the backprojection of the filtered projections. For a certain projection angle θ, the filtered projection $Q_\theta(t)$ contributes to every point (x, y) in the image plane its value at t given by (3.8). It is easily seen that the contribution of $Q_\theta(t)$ is the same for every point (x, y) on the line defined by (3.8), thus the filtered projection $Q_\theta(t)$ is smeared out, or backprojected over the image plane. By summing the contributions of an infinite amount of filtered projections over $180°$, the object function $f(x, y)$ in the image plane is thus obtained. This process of first filtering the projection data and then backprojecting them to retrieve the object function is called the *filtered backprojection*.

In practice, it is obviously impossible to acquire an infinite amount of projections or to integrate over all frequencies. It is thus necessary to formulate discrete versions of the continuous functions in equations (3.9-3.11). The continuous projection angle θ is replaced by a set of K projection angles θ_k for which the projection data $P_\theta(t)$ are acquired. The line integrals from the projection data are sampled at positions $n\Delta t$, where Δt is the detector pixel pitch[1]

[1]The pixel pitch of the detector is the distance between the centres of two adjacent pixels

and $n = -(N-1)/2, \ldots, (N-1)/2$, where N is the number of pixel columns in the detector. This is under the assumption that the projection data are zero outside the index range of n. The projection data can thus be represented by the discrete function $P_{\theta_k}(n\Delta t)$.

For practical purposes, it is safe to assume that the energy contained in the Fourier transform component above a certain frequency is negligible, meaning that the projection data may be considered band limited. The maximum frequency contained in the projection data is given by the Nyquist frequency $W = \frac{1}{2\Delta t}$. The Fourier transform $S_\theta(w)$ of the projections is then sampled at frequencies $m\Delta w$, where $\Delta w = \frac{2W}{N}$ and $m = -(N-1)/2, \ldots, (N-1)/2$, thus yielding $S_{\theta_k}(m\Delta w)$.

Using these practical sampling schemes, the formulas (3.9-3.11) can be made discrete:

$$f(x,y) = \frac{\pi}{K} \sum_{k=1}^{K} Q_{\theta_k}(x \cos\theta_k + y \sin\theta_k) , \tag{3.12}$$

$$Q_{\theta_k}(n\Delta t) = \Delta w \sum_{m=-\frac{N-1}{2}}^{\frac{N-1}{2}} S_{\theta_k}(m\Delta w)|m\Delta w| e^{i2\pi mn\Delta w\Delta t} , \tag{3.13}$$

$$S_{\theta_k}(m\Delta w) = \Delta t \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} P_{\theta_k}(n\Delta t) e^{-i2\pi mn\Delta w\Delta t} . \tag{3.14}$$

This set of equations constitutes an implementation of the filtered backprojection algorithm and can be explained by the following three steps. First, the discrete Fourier transform of the projection data is calculated (3.14). Second, the filtering operation is applied to the data and the inverse Fourier transform is calculated (3.13). Finally, for each projection angle θ_k, the value of the filtered projection $Q_{\theta_k}(t)$ at detector position t, given by (3.8), is added to $f(x,y)$. It is obvious that the detector position t, for a point (x,y) and angle θ_k, usually does not correspond to one of the positions $n\Delta t$ at which the filtered projection is known. Therefore, an approximation to the required value is obtained by using some form of interpolation. Simple linear interpolation is often adequate.

Typically, the object function $f(x,y)$ is sampled on a square grid of $N \times N$ voxels[2], with a sampling distance of Δt. Although the size of this reconstruction grid can be chosen arbitrarily, it is a rational choice to use the same number of voxels as there are pixel columns in the detector. From the circular path of the scanner, it also follows that the region that can be reconstructed exactly is not a square, but a circle inscribed in the square region. Therefore, the reconstruction is limited to this circular region. Theoretically, the number of projections K necessary to obtain a reconstruction without undersampling in the angular direction is

[2]Even though the current derivation is performed in 2D, the building blocks of the cross-sections are referred to as voxels instead of pixels, in order to avoid confusion with the detector pixels.

given by $\frac{\pi}{2}N$. Thus the number of projections should be chosen in the same order as the number of detector columns. According to these remarks, it is easily seen that the computational complexity of the backprojection process is of the order $O(N^3)$.

For the same reasons, the filtering operation is also of the order of $O(N^3)$: $O(N)$ for the number of projections and $O(N^2)$ for the Fourier transform. A fast implementation of the Fourier transform, known as the Fast Fourier Transform (FFT), reduces this operation to $O(N\ \log N)$, thus the total filtering operation is then of the order $O(N^2\ \log N)$. Some remarks are in place regarding the filtering operation. First, equation (3.13) is only valid when the projection data are of finite bandwidth and of finite order. Since these conditions are actually never satisfied, artefacts are created as the aperiodic convolution in (3.10) is implemented as a periodic convolution. Second, the implementation in (3.13) implies that every frequency in the cell $m = 0$, thus the frequency interval $\left[-\frac{\Delta w}{2}, \frac{\Delta w}{2}\right]$, is zeroed, whereas according the theory in (3.10) only one frequency at $w = 0$ should be zeroed. Because of this, the low frequencies are removed which results in a decrease in function values when moving away from the centre and causes a DC shift in the image. These effects can be reduced (but not eliminated) by increasing the number of points used in the Fourier transform, as this will decrease the size of the sampled cells in the frequency domain. This is done by zero-padding the projection data prior to the convolution. In the practical implementation used in this work, the projection data are zero-padded to a length of $(3N)_2$, which is the smallest integer that is a power of 2 and that is greater than $3N$.

The filter that is used in the convolution is called the *ramp filter*, which, due to the fact that the highest frequency in the projection is finite, can be expressed as:

$$H(w) = \begin{cases} |w| & \text{if } |w| < W \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

In addition, one can modify the filtering operation by modifying the ramp filter with some kind of smoothing function, like the Hamming window or the cosine filter. These modifications tend to suppress the higher frequencies, which results in lower noise levels in the reconstructed images, but add a certain degree of smoothing as well. A noise filter which works very well in practice is given by:

$$H(w) = \begin{cases} |w| - \frac{\alpha}{W}w^2 & \text{if } |w| < W \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

where α can be used to tune the amount of noise suppression. This modified ramp filter is illustrated in figure 3.4 for some values of α.

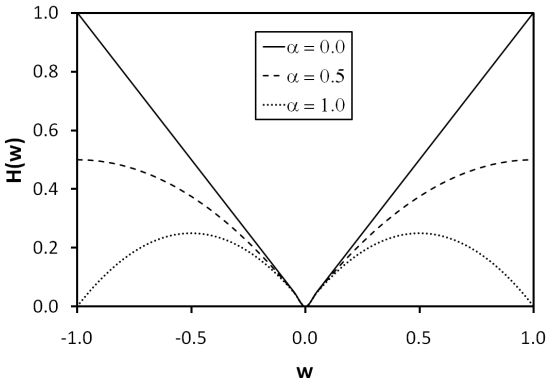


Figure 3.4: The modified ramp filter in the frequency domain for $W = 1$, for different values of the noise suppression parameter α.

3.3 Fan-beam

A fan-beam geometry can be accomplished by using a point source of X-rays and placing a line of detector elements at the other side of the object. As can be expected, this geometry requires a modification of the simple backprojection algorithm presented in the previous section. There exist two types of fan-beam, based on either equidistant or equiangular intervals (figure 3.5). Equidistant intervals are achieved when the detector pixels are placed on a straight line with an equal distance in between. This type of detector is called a planar detector. For the second type the detector elements are placed on the arc of a circle with its centre positioned at the X-ray source, by which the angle between two adjacent rays is constant. This is referred to as a circular detector.

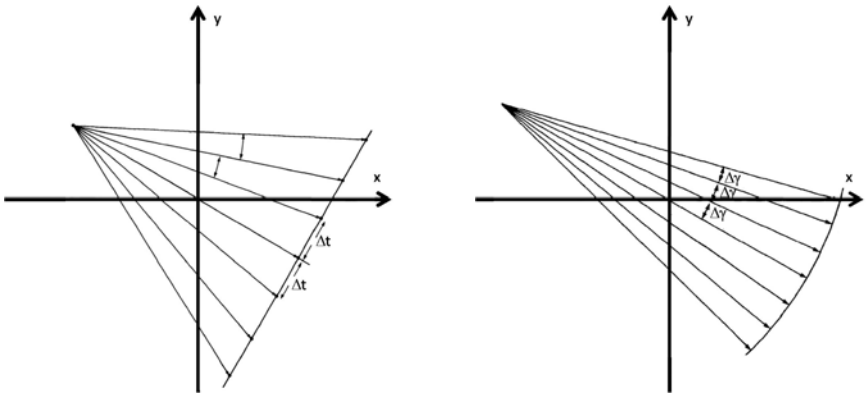
3.3.1 Planar detector

A complete derivation of the reconstruction formula is outside the scope of this work. The resulting algorithm can be summarized by the following set of equations [1]:

$$f(x,y) = \int_0^{2\pi} \frac{R^2}{U(x,y,\theta)^2}\, Q_\theta(t)\, d\theta\,, \tag{3.17}$$

$$Q_\theta(t) = \int_{-\infty}^{\infty} S_\theta(w)|w| e^{i2\pi w t}\, dw\,, \tag{3.18}$$

$$S_\theta(w) = \int_{-\infty}^{\infty} \frac{R}{\sqrt{R^2+t^2}}\, P_\theta(t) e^{-i2\pi w t}\, dt\,, \tag{3.19}$$



(a) Planar detector (b) Circular detector

Figure 3.5: In fan-beam geometry a projection contains the line integrals along a set of rays emanating from a point source. These rays can be sampled at equiangular or at equidistant intervals, depending on the type of detector that is used. (From [6])

with

$$U(x,y,\theta) = R - x\ \sin\theta + y\ \cos\theta\,, \tag{3.20}$$

$$t = R\,\frac{x\ \cos\theta + y\ \sin\theta}{R - x\ \sin\theta + y\ \cos\theta}\,, \tag{3.21}$$

where R is the distance from the source to the origin of the coordinate system[3] and $U(x,y,\theta)$ represents the distance from the source to the point (x,y) projected on the central ray. As opposed to the formulas for parallel-beam (3.9-3.11), the fan-beam reconstruction thus requires two additional factors. The pre-weighting factor $R/\sqrt{R^2+t^2}$ that is applied to the projection data (3.19) corrects for the divergence of the bundle at the detector plane. The second factor is the weight $R^2/U(x,y,\theta)^2$ that is inserted in the backprojection (3.17).

3.3.2 Circular detector

In case of a circular detector, the projection data can be represented by the function $P_\theta(\gamma)$, where γ is the angle between the central ray and the ray hitting the detector

[3]The origin is located at the centre of the object.

pixel. The reconstruction can then be formulated as:

$$f(x,y) = \int_0^{2\pi} \frac{R^2}{L(x,y,\theta)^2}\, Q_\theta(\gamma)\, d\theta\,, \tag{3.22}$$

$$Q_\theta(\gamma) = \int_{-\infty}^{\infty} S_\theta(w)|w| e^{i2\pi w\gamma}\, dw\,, \tag{3.23}$$

$$S_\theta(w) = \int_{-\infty}^{\infty} P_\theta(\gamma) e^{-i2\pi w\gamma}\, d\gamma\,, \tag{3.24}$$

with

$$L(x,y,\theta) = \sqrt{(R - x\ \sin\theta + y\ \cos\theta)^2 + (x\ \cos\theta + y\ \sin\theta)^2} \tag{3.25}$$

$$\gamma = \arctan \frac{x\ \cos\theta + y\ \sin\theta}{R - x\ \sin\theta + y\ \cos\theta}\,. \tag{3.26}$$

For the circular detector, there is no divergence of the bundle in the curved detector plane, so there is no need for an additional pre-weighting of the projection data. The weight in the backprojection however remains necessary.

3.3.3 Short scan

It is obvious that in parallel-beam, projections that are 180° apart are mirror images of each other:

$$P_\theta(t) = P_{\theta+180^\circ}(-t)\,. \tag{3.27}$$

This means it is only necessary to acquire projections for angles θ from 0° to 180°. For fan-beam, it can be found that acquisition is required over an angle of $180^\circ + 2\gamma_m$ [1], where γ_m is half the opening angle of the fan-beam. This additional angular range results in overlapping regions in the data sampling, which leads to artefacts in the reconstruction. This can be prevented by adding an additional window to the filtering operation. In order to avoid high frequency artefacts, the data should not simply be truncated by a one-zero window, but instead a smooth window like the *Parker filter* [7] should be applied:

$$w_\theta(\gamma) = \begin{cases} \sin\left[\frac{\pi}{4}\ \frac{\theta}{\gamma_m - \gamma}\right]\,, & 0 \le \theta \le 2\gamma_m - 2\gamma\,, \\ 1\,, & 2\gamma_m - 2\gamma \le \theta \le \pi - 2\gamma\,, \\ \sin\left[\frac{\pi}{4}\ \frac{\pi + 2\gamma_m - \theta}{\gamma_m + \gamma}\right]\,, & \pi - 2\gamma \le \theta \le \pi + 2\gamma_m\,, \end{cases} \tag{3.28}$$

where θ is the angle under which the projection is taken and γ is the angle between the central ray and the ray hitting the detector pixel.

3.4 Cone-beam

The algorithms described in the previous sections were developed to reconstruct a single slice of a measured object. In order to reconstruct a volume segment of the object, one needs to repeat this process slice by slice for different positions of the sample or of the source and detector system. A more efficient acquisition can be obtained by using a 2D detector. Using an X-ray source, the object is now illuminated by a cone of X-rays, the main advantage being a reduction of the scanning time as the entire object can be scanned at once. The projection data, $P_\theta(t, r)$, are now a function of the projection angle θ and the horizontal and vertical positions t and r of the detector pixel. This cone-beam geometry is shown in figure 3.6.

Although in practical applications both planar and circular 2D detectors are used, only the algorithm for a planar detector will be discussed here as it is the most commonly used set-up in high resolution CT.

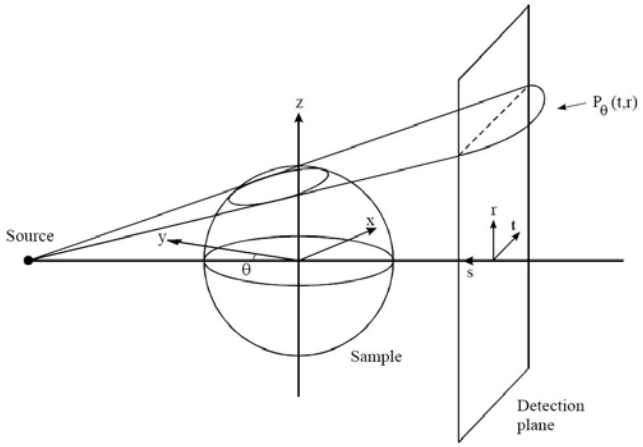


Figure 3.6: In cone-beam geometry a 2D detector is used to measure the line integrals along a set of rays emanating from a point source. (From [1])

3.4.1 Exact reconstruction

Exact reconstruction of the 3D object function $f(\vec{x})$ with $\vec{x} = (x, y, z)$ is based on the *3D Radon transform*, which can be defined as follows [8]:

> Consider a plane defined by its normal $\vec{\alpha}$ and the distance s to the origin. The 3D Radon transform $\mathcal{R}f(\vec{\alpha}, s)$ is the integral of $f(\vec{x})$ over this plane.

Each plane is represented by a unique point in the object space, being the intersection of the plane and its normal passing the origin. The 3D Radon space now consists of all Radon values $\mathcal{R}f(\vec{\alpha}, s)$ placed at these points. The 3D Fourier slice theorem can then be stated as:

> The 1D Fourier transform $F\{\mathcal{R}f\}(\vec{\alpha}, s)$ of the Radon values with respect to s is equal to a central slice at direction $\vec{\alpha}$ of the 3D Fourier transform $F\{f\}(\vec{x})$ of the object function.

This means that, if the 3D Radon values in the object space can be obtained, the 3D Fourier transform of the object function can be calculated and the object function $f(\vec{x})$ can be derived. The Tuy-Smith condition [9] states that an exact reconstruction is possible if all planes intersecting the object also intersect the source trajectory at least once. Intuitively, this condition can be clarified by noting that the source must be positioned in a plane in order to measure its integral. It is clear that the Tuy-Smith condition is not satisfied for a circular trajectory. Consider for instance a plane parallel to the source trajectory. This plane may intersect the object but not the trajectory. For an exact reconstruction, the source path should therefore be extended with an additional circle [10] or line [11] out of the plane of the original circular trajectory or by following a helical path. For a circular path, the available Radon data are confined to a torus, the missing data are referred to as the *shadow zone* (see figure 3.7).

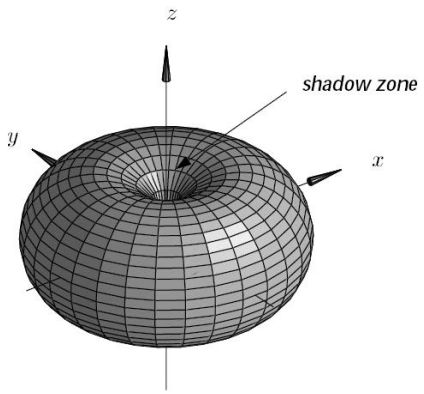


Figure 3.7: The values of the 3D Radon transform that can be retrieved using the cone-beam geometry. (From [12])

The derivation above assumes that the integrals are taken over the entire plane. In practice this is only possible if the object has a limited support, the detector covers the entire projection of the sample and the object function outside the sample can be considered to be zero, this is called *untruncated* data.

Several implementations based on the 3D Radon transform have been presented [13, 14], of which a thorough presentation can be found in [15].

3.4.2 The FDK algorithm

The most commonly used method for the reconstruction of cone-beam data in practical applications is the algorithm described by Feldkamp, David and Kress [16]. Although this algorithm is only an approximate one, for small cone angles the associated errors are rather small and often acceptable. In return, unlike exact methods, FDK can handle data truncated in the longitudinal direction, which is often encountered in real scans. Additionally, FDK is an algorithm that is very amendable to highly efficient implementation and can thus offer very fast reconstruction speeds.

The algorithm is a natural extension to 3D of the filtered backprojection algorithm. For a planar detector, it can be given by the following set of equations:

$$f(x,y,z) = \int_0^{2\pi} \frac{R^2}{U(x,y,\theta)^2}\, Q_\theta(t,r)\, d\theta\,, \tag{3.29}$$

$$Q_\theta(t,r) = \int_{-\infty}^{\infty} S_\theta(w,r)|w| e^{i2\pi wt}\, dw\,, \tag{3.30}$$

$$S_\theta(w,r) = \int_{-\infty}^{\infty} \frac{R}{\sqrt{R^2+t^2+r^2}}\, P_\theta(t,r) e^{-i2\pi wt}\, dt\,, \tag{3.31}$$

with

$$U(x,y,\theta) = R - x\ \sin\theta + y\ \cos\theta\,, \tag{3.32}$$

$$t = R\,\frac{x\ \cos\theta + y\ \sin\theta}{R - x\ \sin\theta + y\ \cos\theta}\,, \tag{3.33}$$

$$r = z\,\frac{R}{R - x\ \sin\theta + y\ \cos\theta}\,, \tag{3.34}$$

where R is again the distance from the source to the origin, $U(x,y,\theta)$ is the distance from the source to the point (x,y) projected on the central ray. The 2D projections under an angle θ are given by $P_\theta(t,r)$ and t and r are the coordinates of the pixel in the detector plane.

The number of cross-sections that can be reconstructed in cone-beam depends on the number of pixel rows M in the detector. Due to the cone angle, only a volume described by a truncated double cone (see figure 3.8) is exposed from all directions and only this volume can be reconstructed without introducing additional errors. The number of detector rows M, and thus the number of cross-sections that can be reconstructed, is typically of the same order as the number of detector columns N, which in turn is of the same order as the number of projections K. The required computations to perform the backprojection are thus of the order $O(N^4)$, whereas the filtering operation is of the order $O(N^3 \log N)$ when using a fast Fourier transform.

As was discussed for the fan-beam reconstruction, it is sufficient to acquire projection data over the range of $180^\circ + 2\gamma_m$, where γ_m is half the fan angle of

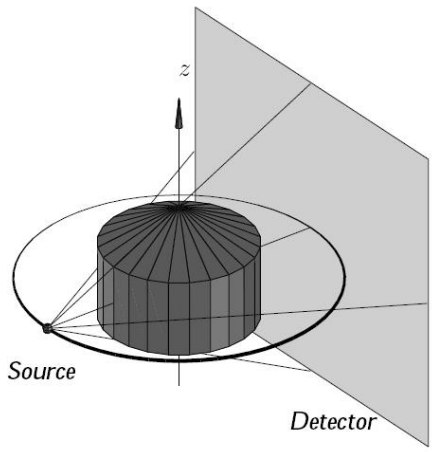


Figure 3.8: The volume that can be reconstructed in cone-beam geometry. (From [12])

the beam, by applying the Parker filter to the projections. However, due to the cone angle the off-centre detector rows will contain slightly different information for projection angles within the overlapping region. Limiting the angular range of the scan will thus deteriorate the sampling of the 3D Radon space and enhance the artefacts that arise from this missing data.

For FDK, the following three important properties were observed:

- Reconstruction is exact in the mid-plane, $z = 0$, since it is identical to the fan-beam reconstruction. The errors in the reconstruction will become larger in planes further away from the mid-plane.
- It is exact for objects that are homogeneous in the Z-direction, thus when $f(x, y, z) = f(x, y)$.
- The integral value $\int f(x, y, z)dz$ is preserved.

These three properties imply that a distortion is expected, manifested as a blurring along the Z-axis. This blurring only occurs in parts of the object that are not homogeneous in the Z-direction and it deteriorates when moving away from the central plane of the geometry. Several modifications of the FDK method exist that were developed to tackle some of its errors. For an elaborate discussion on these variations, the reader is referred to literature, e.g. [12].

3.4.3 Cone-beam artefacts

As was discussed above, the cone-beam geometry does not provide a complete sampling of the 3D Radon transform of the object function. Reconstructions from

a cone-beam scan thus contain artefacts that are related to the cone angle of the beam, known as *cone-beam artefacts* or *cone artefacts*. Intuitively, this can be understood as follows. A transition in the object function can only be completely resolved if projection data are available for rays perpendicular to this transition. Except for the horizontal central plane (corresponding to the XY-plane), the rays are always oblique in a horizontal plane. Therefore, information on the transitions along the Z-axis is lost to some extent. In figure 3.9, the reconstruction of a simulated Defrise phantom (appendix A) for a typical cone angle is shown. Instead of retrieving a set of clearly separated discs, the phantom is heavily distorted due to cone artefacts. Several modifications to the FDK method exist, which were specifically developed in order to reduce cone artefacts, e.g. [12, 17].

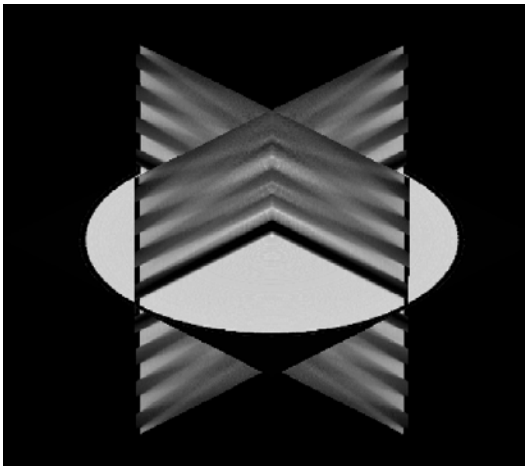


Figure 3.9: Reconstruction result for a simulated Defrise phantom using cone-beam geometry. A combined image containing the central cross-sections along each of the main axes is shown.

One could argue that the cone angle can be reduced by simply moving the detector further away from the source, while simultaneously changing the distance from the object to the source by the same factor, in order to maintain the same magnification. However, as the source emits a conical bundle, the intensity of the X-ray beam drops quadratically with the distance from the source. Moving the detector away from the source would thus require a longer integration time in order to obtain the same level of statistical information. Besides obvious practical and economical considerations, increased scanning times are best avoided as this can lead to a loss of resolution due to instabilities in the scanner (e.g. drift of the X-ray spot, etc.) or object (e.g. thermal expansion, dehydration, etc.). Following these practical limitations, simply increasing the source to detector distance to reduce cone artefacts is often not considered a feasible solution.

Producing a spot that is sufficiently small for high resolution scans implies using an X-ray bundle of low intensity. Due to this low flux, the cone angle of the

set-up is usually rather large in order to obtain acceptable scanning times. Therefore, cone-beam artefacts are often encountered in a typical X-ray tube based high resolution system.

3.5 Helical cone-beam

The cone-beam geometry discussed in the previous section is the most commonly used geometry in high resolution CT. However, it was shown that its circular trajectory does not allow an exact reconstruction and artefacts can arise due to the incomplete sampling of the Radon space. An alternative geometry that does allow for an exact reconstruction is the helical cone-beam (or spiral cone-beam), which can be accomplished by introducing an additional translation along the axis of rotation between every projection (figure 3.10). The helical path is also better suited to scan elongated objects[4]. When scanning an elongated object in cone-beam, the resolution is limited, as the magnification of the geometry needs to be limited in order for the detector to cover the whole object. Using the helical geometry, the magnification can be increased, thus maximizing resolution, as the detector does not need to cover the entire vertical range within one projection. The helical trajectory however requires a vertical movement that needs to be very precise, which further complicates the scanner set-up. Especially in high resolution systems, this additional technical requirement often poses an obstacle.

It needs to be noted that the helical cone-beam is predominantly used in medical imaging, where cone angles are typically rather small since only a small number of detector rows is used as compared to the number of columns. In high resolution CT, the number of detector rows is usually about the same as the number of columns and the cone angle is relatively large, therefore the choice of an appropriate reconstruction algorithm will vary from the ones typically used in medical imaging. The concept of the helical path is only very recently introduced in high resolution CT. Several manufacturers are nowadays developing and introducing helical high resolution scanners and it is believed they will become more and more common due to the advantages of the helical trajectory. As both have their shortcomings and benefits, helical scanners will probably never replace cone-beam scanners, instead they should be seen as complementary systems. The two high resolution CT scanners developed at UGCT are both equipped with the possibility to perform a helical scan trajectory.

3.5.1 Overview of helical cone-beam algorithms

Before presenting an overview of the most common reconstruction algorithms, the concepts of the *long-object* and *short-object* need to be clarified [19]. In order to

[4] An elongated object is an object which is considerably larger in one dimension than in the other two dimensions

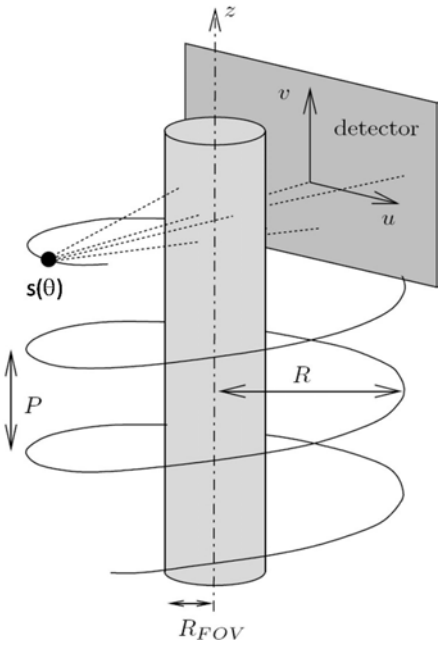


Figure 3.10: Acquisition geometry and the source path in helical cone-beam. (From [18])

reconstruct a region of the object using a short-object algorithm, one needs to scan the whole object. Contrary, long-object algorithms require only a part of the object to be scanned. In the following, only long-object algorithms will be discussed as they are obviously more suited in practical applications. An elaborate discussion of several helical algorithms can be found in [12].

Approximate algorithms

The following methods have the advantage that they are inexpensive in terms of computational complexity. However, due to their approximate character, they are only acceptable for small cone angles. Although this is a common feature for medical scanners, high resolution scanners usually do not meet this requirement. As the approximate algorithms introduce artefacts with increasing cone angle, they are not suited for high resolution CT.

Rebinning algorithms: The acquired projections are rebinned to a stack of fan-beam projections corresponding to each slice. For small cone angles, 2D filtered backprojection combined with axial interpolation provides images of sufficient quality.

Generalised FDK: This solution is based on the FDK algorithm which was generalized to a helical scan. It is a heuristic derivation from the 2D fan-beam algorithm.

PI-methods: In this family of algorithms, the detector fits exactly in the Tam-Danielsson window[5] and the data capture is complete and almost nonredundant. These methods are based on the observation that the rebinned ray geometry guarantees that the backprojection step delivers the same number of contributions to all object points.

Exact algorithms

Contrary to the approximate algorithms, exact algorithms avoid intrinsic errors and therefore provide superior quality when dealing with large cone angles. The disadvantage of these methods is that they are computationally intensive, thus leading to rather low reconstruction speeds which limits their practical use.

Tam's algorithm: The idea is to calculate the 3D Radon transform of a plane from a combination of several projections that provide some kind of triangulation of the plane, instead of just using a single projection like the Grangeat formula. This however requires an additional circular scan of the top and bottom of the region that is to be reconstructed, which is not very attractive in practical applications.

Kudo's algorithm: This is a filtered backprojection type algorithm, meaning the projection data can be processed parallel. It is actually a quasi-exact method, as it introduces a minor approximation.

PHI-methods: These methods introduce virtual objects $f_\phi(\vec{x})$ for each value of the azimuthal angle ϕ, which are equal to the real object $f(\vec{x})$ in a certain region. The exact Radon data for a 2D parallel-beam projection of $f_\phi(\vec{x})$ onto the meridian plane of angle ϕ can be calculated for every angle ϕ. The object function $f(\vec{x})$ can then be reconstructed in that certain region where it is equal to $f_\phi(\vec{x})$.

Katsevich algorithm: This rather new algorithm further improves the effectiveness of the exact long-object algorithm and uses the projection data contained in the PI-lines to reconstruct the object function, offering better resolution than the other algorithms described above.

3.5.2 Katsevich algorithm

For the reconstruction of helical projection data, it was chosen to implement the algorithm derived by Katsevich [20], as it contains several features that are important for practical use at a high resolution set-up. First of all, since it is an exact algorithm, it can be used to reconstruct data acquired using a large cone angle. Second, it is a long-object method, so reconstruction of a region of an object can

[5]The concept of the Tam-Danielsson window will be elucidated in the following section on the Katsevich algorithm.

be obtained without scanning the whole object. Finally, it allows for efficient implementation and thus high reconstruction speed. The reconstruction formula will only be shown for a planar detector, similar results can be derived for a circular detector.

Helical source path

The position of the X-ray source in a helical path, illustrated in figure 3.10, can be given by:

$$s(\theta) = [R\cos\theta, R\sin\theta, P\frac{\theta}{2\pi}]^T \ . \tag{3.35}$$

R is the distance from the source to the origin and is thus equal to the radius of the helical path. The pitch P is the displacement of the source along the axis of rotation for one full turn of the source. The angle θ is the angle under which the projection is taken. Note that this angle is not limited to the interval $[0, 2\pi[$ as a helical scan may consist of several turns.

PI-lines

A *PI-line* is defined as a line segment that connects two points on the helical path that are separated by less than one helical turn (figure 3.11). For a constant helical pitch, there exists a unique PI-line through every point $\vec{x}$ inside the helix [21]. Let the endpoints of a PI-line be defined by $s(\theta_b)$ and $s(\theta_t)$, then the parametric interval corresponding to the unique PI-line passing through $\vec{x}$ is denoted by $I_{PI} = [\theta_b, \theta_t]$. By definition, $\theta_t - \theta_b < 2\pi$. Given a fixed point expressed in cylindrical coordinates $\vec{x} = [r\cos\gamma, r\sin\gamma, z]^T$, the PI-line interval can be found by solving θ_b from the following equation [22]:

$$z = \frac{P}{2\pi}\left\{[\pi - 2\alpha]\cdot\left[1 + \frac{r^2 - R^2}{2R(R - r\cos(\gamma - \theta_b))}\right] + \theta_b\right\} \ , \tag{3.36}$$

where

$$\alpha = \arctan\left(\frac{r\sin(\gamma - \theta_b)}{R - r\cos(\gamma - \theta_b)}\right) \ , \tag{3.37}$$

and

$$\frac{2\pi z}{P} - \pi \leq \theta_b \leq \frac{2\pi z}{P} \ . \tag{3.38}$$

Additionally, θ_t can be found by:

$$\theta_t = \theta_b + \pi - 2\alpha \ . \tag{3.39}$$

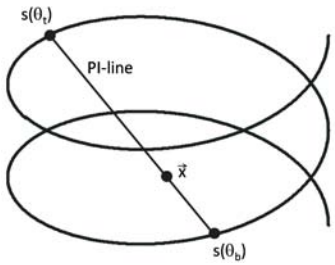


Figure 3.11: Illustration of the PI-line for the point $\vec{x}$. (From [23])

Local detector coordinates

A local coordinate system (u, v, w) can be attached to the planar detector, with unit vectors that depend on the projection angle θ:

$$\begin{aligned} \vec{e}_u(\theta) &= [-\sin\theta, \cos\theta, 0]^T \\ \vec{e}_v(\theta) &= [-\cos\theta, -\sin\theta, 0]^T \\ \vec{e}_w(\theta) &= [0, 0, 1]^T \ . \end{aligned} \tag{3.40}$$

The vector $\vec{e}_v$ points from the source point $s(\theta)$ to the centre of the detector. The vectors $\vec{e}_u$ and $\vec{e}_w$ span the detector, respectively representing the pixel columns and rows. For a projection angle θ, the ray through the point $\vec{x} = (x, y, z)$ hits the detector in the local coordinate (u^*, w^*) given by:

$$\begin{aligned} v^*(s, \vec{x}) &= R - x\ \cos\theta - y\ \sin\theta \ , \\ u^*(s, \vec{x}) &= \frac{D}{v^*(s, \vec{x})} \cdot (-x\ \sin\theta - y\ \cos\theta) \ , \\ w^*(s, \vec{x}) &= \frac{D}{v^*(s, \vec{x})} \cdot (z - \frac{P\theta}{2\pi}) \ , \end{aligned} \tag{3.41}$$

with D the distance from the source to the detector.

κ-lines

Consider a virtual detector curved along the cylinder containing the helical source trajectory, limited by two consecutive helix turns in height and by two vertical lines in width (figure 3.12). This detector is called a *PI-detector* and its projection on the real detector is known as the *Tam-Danielsson window* [24, 25]. A *κ-plane* is defined as any plane that has three intersections with the helix such that one intersection lies half-way between the other two. The κ-plane denoted by $\kappa(\theta, \psi)$ has three intersects $s(\theta)$, $s(\theta+\psi)$ and $s(\theta+2\psi)$, where $\psi \in\ ]-\pi/2, \pi/2[$. A κ-line is now defined as the intersecting line between the detector and the κ-plane. It is shown that for $\psi \in\ ]-\pi/2, \pi/2[$ the κ-plane, and thus the κ-line, through a given

point $\vec{x}$ is unique when the projection of this point lies inside the Tam-Danielsson window. The κ-line for an angle ψ is given by:

$$w_\kappa = \frac{DP}{2\pi R} \cdot \left(\psi + \frac{\psi}{D \tan \psi} u \right) . \tag{3.42}$$

In figure 3.13 a number of κ-lines is shown along with the Tam-Danielsson window for a certain detector.

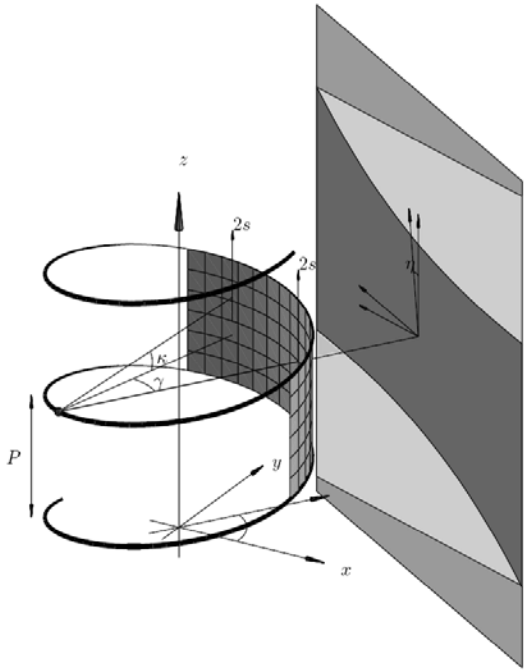


Figure 3.12: The PI-detector and its projection on the detector, defining the Tam-Danielsson window. (From [12])

Reconstruction formula

Let the object function be denoted by $f(\vec{x})$ and the projection data by $P_\theta(u, w)$. The Hilbert kernel is defined as:

$$k_H(t) = \frac{1}{\pi t} . \tag{3.43}$$

Finally, using all of the above, the Katsevich reconstruction formula for a pla-

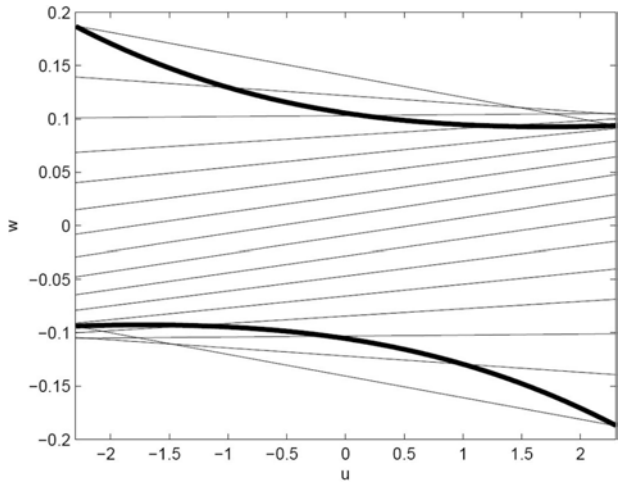


Figure 3.13: Illustration of the κ-lines, the broad curved lines represent the edges of the Tam-Danielsson window. (From [26])

nar detector can be given by the following set of equations [26]:

$$f(\vec{x}) = \frac{1}{2\pi}\int_{I_{PI}} \frac{1}{v^*(s,\vec{x})} P_\theta^F(u^*,w^*)\, d\theta\,, \tag{3.44}$$

$$P_\theta^F(u,w_\kappa) = \int_{-\infty}^{\infty} k_H(u-u')\, \frac{D}{\sqrt{(u')^2+D^2+w_\kappa^2}} P_\theta^D(u,w_\kappa)\, du'\,, \tag{3.45}$$

$$P_\theta^D(u,w) = \frac{\partial P_\theta(u,w)}{\partial\theta} + \frac{u^2+D^2}{D}\,\frac{\partial P_\theta(u,w)}{\partial u} + \frac{uw}{D}\,\frac{\partial P_\theta(u,w)}{\partial w}\,, \tag{3.46}$$

The function $P_\theta^D(u,w)$ is the derivative of the projection data, which can be found using (3.46) by taking the partial derivatives along the projection angle, the pixel columns and the pixel rows. Equation (3.45) denotes filtering of the derived projection data by convolution with the Hilbert kernel (3.43). The filtering is again a 1D operation, but unlike the previous geometries, it is not performed along parallel rows in the detector, but along the sloped κ-lines. Additional smoothing can be accomplished by adding an extra window function to the filtering, for instance a Hamming window. The object function $f(\vec{x})$ is then calculated by applying the backprojection operator (3.44), which performs a weighted sum of the filtered projection values along the PI-line corresponding with the point $\vec{x}$. According to the properties of the PI interval, the number of steps and the range of this integration depends on the position of the point $\vec{x}$. The number of backprojected contributions to the object function $f(\vec{x})$ thus varies, in contrast with the geometries discussed in the previous sections, where the integration range is the same for every point $\vec{x}$.

The actual implementation of the Katsevich algorithm can vary. In Noo *et al.* [27] and Yu *et al.* [23] the reader can find two possible implementations, along with a more detailed description regarding some practical issues. The implementation used in this work is based on the paper by Yu *et al.*.

3.5.3 In practice

Setting up the acquisition protocol for a helical cone-beam scan is a bit more complicated than for a regular cone-beam scan. Three additional questions need to be answered:

1. What is the step size of the translation between two consecutive projections?
2. What are the start and end positions of this translation?
3. How many projections need to be taken?

First, the reader is reminded that in the helical reconstruction the number of projections used in the backprojection differs for every voxel. This number of projections, and the corresponding angular range of these projections, is determined by the voxel's PI-line. Each PI-line covers a certain angular range of roughly 180°. Thus, for comparison with the cone-beam geometry, the helical cone-beam geometry can be considered to work in a short scan regime.

To answer the questions regarding the scanning protocol, consider a detector with N pixel rows and M columns and a pixel pitch d. The detector is placed at a distance SDD from the source, while the distance from the object to the source is given by SOD. Using this geometry, the field of view FOV[6] of the detector is given by:

$$FOV = 2\,SOD\,\sin\left[\tan^{-1}\left(\frac{Md}{2\,SDD}\right)\right]\,. \tag{3.47}$$

The maximum helical pitch P_{max} of the helical trajectory depends on the geometrical settings and the size of the detector, and can be calculated by [23]:

$$\Delta \;=\; 2\,\cos^{-1}\left(\frac{FOV}{2\,SOD}\right)\,, \tag{3.48}$$

$$P_{max} \;=\; 2\pi\,SOD\,\frac{N}{M}\,\frac{\sin\Delta}{2\pi-\Delta}\,. \tag{3.49}$$

It is obvious that the maximum pitch strongly depends on the number of pixel rows in the detector, such that the allowed pitch increases with an increasing number of rows. When choosing a pitch larger than this, the Tam-Danielsson window is not

[6]The FOV is defined as the diameter of the cylindrical volume that is covered by the detector, which depends on the geometrical configuration of the scan.

completely covered by the projection data and an exact reconstruction is no longer possible, which results in artefacts in the reconstruction. When a smaller pitch is used the reconstruction is exact, but some of the acquired data are redundant. As with the other geometries, the quality of the reconstruction depends on the angular step size between two consecutive projections. In the helical cone-beam geometry, it is common to express this angular sampling by the number of projections S that is taken per turn, thus for a rotation of one full circle. The maximum step size of the translation between two projections is then simply given by $p_{max} = P_{max}/S$. In practice, the step size p is usually determined by rounding down the maximum step size such that it complies with the accuracy of the motor system, which only results in a minimum amount of redundant data.

The start and end position of the translation can be calculated approximately based on geometrical considerations. The start position z_{start} is obtained by noting that the object can only be reconstructed exactly when its top is projected entirely below the lowest detector row in the first projection. Similar, the end position z_{stop} is found when the bottom of the object is projected above the highest detector row in the last projection. For an object of height[7] H, the total vertical translation Δz can then be given by:

$$\Delta z = z_{stop} - z_{start} = H + 2 \left(SOD + \frac{FOV}{2} \right) \tan \alpha_{cone} \,, \tag{3.50}$$

where half the cone angle α_{cone} of the X-ray beam is given by:

$$\alpha_{cone} = \tan^{-1} \left(\frac{Nd}{2\, SDD} \right) \,. \tag{3.51}$$

The number of projections K_{HB} is then easily found by:

$$K_{HB} = \frac{\Delta z}{p} = S\, \frac{\Delta z}{P} \,. \tag{3.52}$$

The number of projections required for a helical scan is thus a linear function with respect to the height of the object and with an offset depending on the scan geometry and the detector.

As was mentioned, the helical cone-beam geometry can be considered a short scan regime. Thus, when comparing both geometries, the acquisition for the cone-beam geometry should also be a short scan[8]. Keeping the angular step size fixed, this means that the number of projections required for the cone-beam scan K^0_{CB} is equal to:

$$K^0_{CB} = S\, \frac{\pi + 2\, \alpha_{fan}}{2\pi} \,, \tag{3.53}$$

[7] Since both the FDK and the Katsevich algorithm can process vertically truncated data, the height of the object is irrelevant as the scanned region can always be reconstructed exactly. The height of the object thus in fact refers to the height of the region of interest for the object.

[8] As was mentioned before, in cone-beam the object can never be reconstructed exactly. Since the sampling of the data is even worse for a short scan, it is advisable to always perform a full scan. However, for this comparison the short scan is more suitable.

where half the fan angle α_{fan} of the X-ray beam is given by:

$$\alpha_{fan} = \tan^{-1}\left(\frac{Md}{2\,SDD}\right) . \tag{3.54}$$

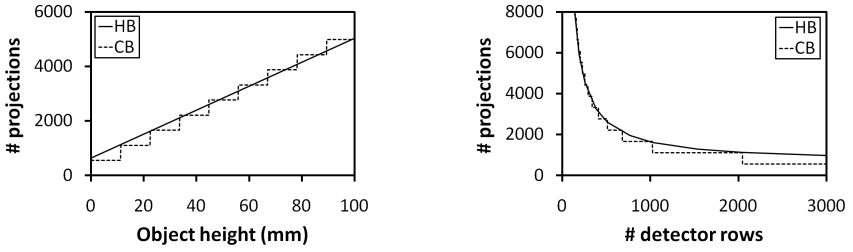
The maximum height of the object that can be fully reconstructed using the cone-beam acquisition can again be found using geometrical considerations by:

$$H_{CB} = 2\left(SOD - \frac{FOV}{2}\right)\frac{N\,d}{2\,SDD} . \tag{3.55}$$

Objects with a height smaller or equal to H_{CB} can be reconstructed from one scan, for heights larger than H_{CB}, additional scans are required to cover the entire object. In general, the required number of projections for the cone-beam scan of an object of height H is given by:

$$K_{CB} = a \cdot K^0_{CB} \text{ , for } (a-1)\,H_{CB} < H \leq a\,H_{CB} \text{ , } a \in \mathbb{N} . \tag{3.56}$$

Now, to illustrate the differences in the acquisition of a helical and a regular cone-beam scan, consider the following example of a typical scan with N = 1024, M = 1024, d = 0.25 mm, SOD = 50 mm and SDD = 1000 mm. The number of projections per turn is S = 1024, thus the projections are taken with an angular step of 0.35°. For the single cone-beam scan, the number of projections is $K^0_{CB} = 553$ and the maximum height of the object is H_{CB} = 11.17 mm. The required number of projections for both cone-beam (CB) and helical cone-beam (HB) is shown as function of the total height of the scanned object in figure 3.14(a). Figure 3.14(b) illustrates the number of projections required to scan an object of height $2{\cdot}H_{CB}$ with a varying number of detector rows. By evaluating the reconstructions of simulated phantom data, it was found that the reconstruction quality does not change when varying the number of detector rows.

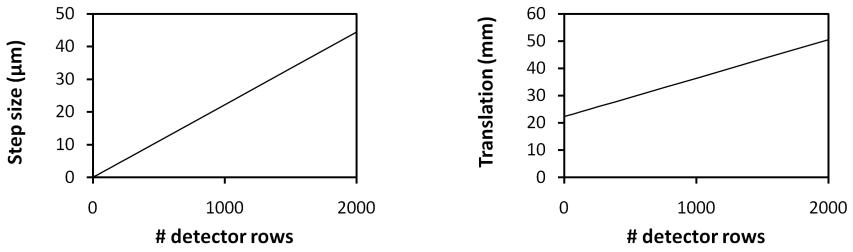


(a) (b)

Figure 3.14: Illustration of the number of required projections for a helical (HB) and a regular cone-beam (CB) scan, as function of the object height using 1024 rows (a) and as function of the number of rows for a height of 2·H_{CB} (b).

Another practical aspect of a helical scan concerns the translation movement. Depending on the application, the translation movement needs to be performed

with a certain accuracy and needs to cover a specified range. In practice, combining a high accuracy and a large travel range is not obvious. Therefore, it is important to know how the step size and the range vary with respect to the geometry, but also how they can be modified by limiting the number of detector rows that is used for the acquisition of the data. In figure 3.15, the maximum step size p_{max} and the total range of the translation Δz are given as function of the number of detector rows. From this relation, it can be seen that by decreasing the number of detector rows that is used, the necessary range of the translation can be reduced at the expense of decreasing the step size and thus increasing the required accuracy of the translation movement. Using more rows allows using a less accurate translation, but requires a larger range. For a small number of detector rows, the total translation range converges to the height of the object, which is to be expected as in the limit of a single detector row the helical cone-beam geometry defaults to the fan-beam geometry.



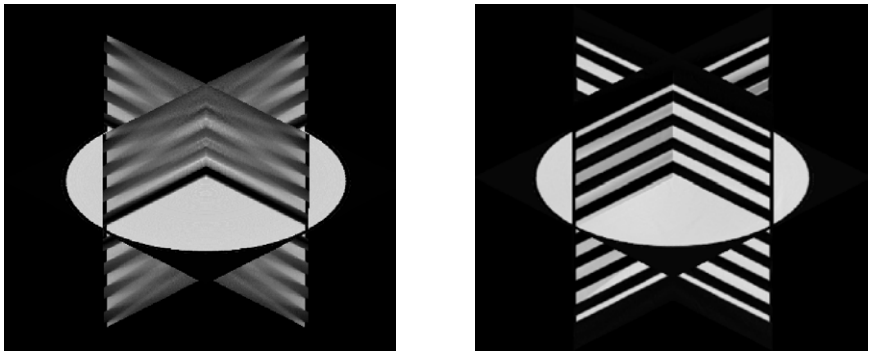
(a) (b)

Figure 3.15: The maximum step size (a) and the necessary total range (b) of the translation movement for helical cone-beam.

3.5.4 Cone-beam artefacts

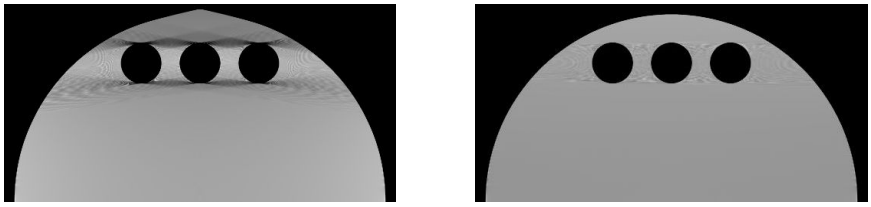
Contrary to cone-beam, the helical cone-beam geometry does provide a complete sampling of the 3D Radon transform. This allows for an exact reconstruction of the object function and thus cone-beam artefacts will not be present. This can be observed in figure 3.16, where the cross-sections along the three main axes are shown of the reconstructed object function of a Defrise phantom, for both a cone-beam scan and a helical cone-beam scan. One can clearly see the presence of cone-beam artefacts, appearing as a certain vertical blending that becomes worse when moving away from the mid-plane and moving towards the rotation axis, which was to be expected considering the shape of the shadow zone (fig. 3.7). The reconstruction of the helical scan does not show any of these artefacts at all. A second phantom containing several spherical pores is shown in figure 3.17. In the cone-beam geometry, the top of the phantom and the top and bottom of the voids are severely distorted. These cone-beam artefacts are not present in the helical cone-beam reconstruction. The full cone angle used in these simulations is 14°

for the Defrise phantom and 28° for the second phantom, which are typical angles used in a practical high resolution X-ray CT system.



(a) Cone-beam (b) Helical cone-beam

Figure 3.16: Comparison of cone-beam artefacts for a simulated Defrise phantom using cone-beam (a) and helical cone-beam geometry (b).



(a) Cone-beam (b) Helical cone-beam

Figure 3.17: Comparison of cone-beam artefacts for a phantom containing several spherical pores using cone-beam (a) and helical cone-beam geometry (b).

Choosing a helical path is thus shown to provide an adequate solution to the problem of cone-beam artefacts. One major limitation of the helical trajectory is that the calibration and alignment of the system has to be very accurate. Especially the vertical movement was found to be very crucial, even the slightest deviation can result in a serious deterioration of the reconstruction quality.

3.6 Region of interest scan

It was shown that the reconstruction algorithms discussed in the previous sections are capable of handling data truncated in the vertical direction, thus handling the long object problem. However, the projection data can also be truncated in the horizontal direction, meaning the detector does not always cover the entire projection of the object, thereby violating the assumption of untruncated data that was posed

in the derivation of the algorithms. Reconstruction of horizontally truncated data is often required, especially in high resolution applications when the object cannot be made sufficiently small due to limitations in the object preparation. Such cases are often referred to as *local* or *region of interest* tomography, as only a certain region of the entire object is scanned.

When processing such truncated projections, the zero-padding of the filtering operation introduces steep edges, resulting in high frequency components which generate artefacts in the reconstruction, typically observed as a bright edge surrounding the sample. The arising of these artefacts can also be explained intuitively in the following way. Looking at the projection data, the total amount of attenuation is larger than what can be expected from the attenuation values contained inside the circular region of the object that is reconstructed. The reconstruction algorithm, which assumes that the object must be confined to this circular region, sees this higher attenuation and tries to compensate it by introducing a strongly attenuating border in the outer layer of the region.

The truncated data reconstruction can be improved substantially by padding the projection data with a $\cos^2$-function [28]. For a detector containing N pixel columns, both sides of the data are expanded by adding a $\cos^2$-profile that falls to zero after N_{pad} pixels, as shown in figure 3.18. The resulting smooth transition no longer introduces high frequency effects and removes the bright border near the edge of the circular region. The result of this region of interest filter is illustrated in figure 3.19, where reconstruction results are shown for a cropped region of a full view scan (a), an unfiltered region scan (b) and a filtered region scan (c).

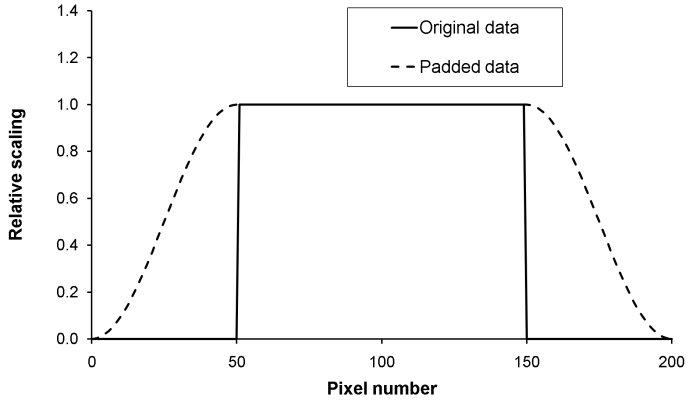
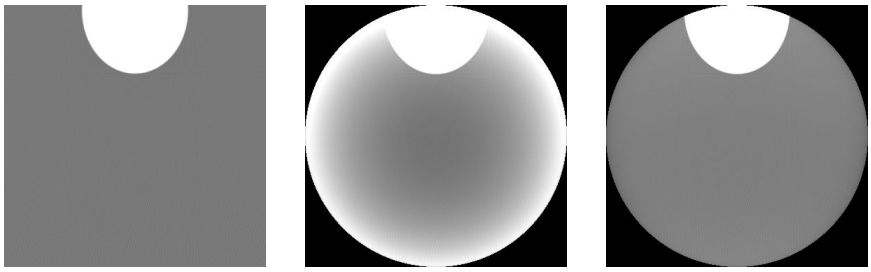


Figure 3.18: Padding of the projection data for a region of interest scan, with $N_{pad} = N/2$.

In theory, application of this filter is only valid for data truncated by the same amount on both sides and for all projections. In practice, it turns out that the filter also works well when the object only moves outside the field of view of the detector for a limited amount of projections. Results are not satisfactory when scanning



(a) (b) (c)

Figure 3.19: Reconstruction results of a region of interest scan, a comparison is shown for a cropped region of a full view scan (a), an unfiltered region scan (b) and a filtered region scan (c).

an off-centre part of a noncylindrical sample, in such cases, more advanced region-of-interest processing needs to be applied. One such proposed approach could be to first scan and reconstruct the entire object at a lower resolution. The reconstruction of the low resolution scan can now be used to correct the projection images of the high resolution scan, which could then be reconstructed without the typical error.

3.7 Beam hardening

Most X-ray sources used in practical set-ups do not generate a monochromatic, but a polychromatic X-ray beam. Within the typical energy range used in CT imaging, the linear attenuation coefficient of most materials decreases with energy. Low energy photons are thus attenuated more frequently than higher ones, causing the transmitted beam to contain proportionally more high energy photons. The mean energy of the spectrum of the transmitted beam is thus higher than the one of the incoming beam, this effect is called *beam hardening*. Additionally, each type of detector has a different response with regard to a certain X-ray energy.

Since the imaging process thus depends on the energy spectrum of the generated X-ray bundle, equation (3.3) is no longer valid and severe nonlinearities are introduced in the reconstruction. An artefact that typically results from this beam hardening is so called *cupping*. As only a small piece of the object suffices to stop the low energy photons, the outer layers of the object are attributed an attenuation coefficient which is too high. Additionally, beam hardening causes smearing of the reconstructed images, by which pores in the object appear to be partially filled.

A polychromatic beam also creates streaks and flares in the vicinity of strongly attenuating regions in the object, for instance metallic inclusions. Additionally, wide dark streaks appear between such different strongly attenuating inclusions.

These artefacts, which are due to extreme beam hardening, are more often referred to as *metal artefacts*.

A simple correction to the cupping artefact can be derived as follows. When the X-ray beam is monochromatic and passes through a homogeneous object, the measured quantity $\ln(I/I_0)$ is proportional to the length of the X-ray path through the object. For a polychromatic beam, this relation is obviously not linear but follows a certain curve (figure 3.20). By experimentally finding the expression of this curve, one can correct the measured attenuation values and thus obtain a reconstruction without cupping. Two of such methods will be discussed.

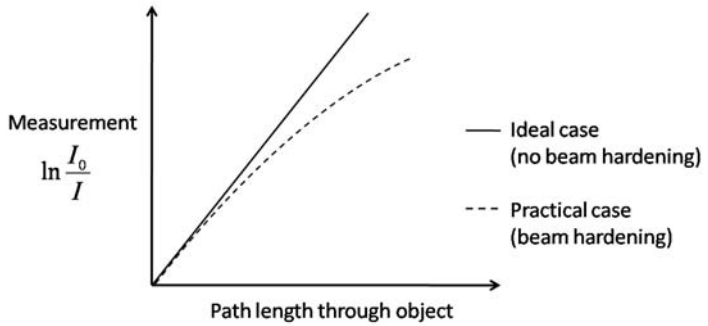


Figure 3.20: An illustration of the beam hardening effect for a homogeneous object. For a polychromatic beam the measured intensity depends nonlinearly on the path length through the object.

The first method consists of simply applying a polynomial correction to the measured attenuation values. The parameters of the correction are found through trial and error by evaluating the reconstruction of a single slice. It is found that, in practical situations it is often sufficient to add a second order term to correct the cupping.

The second method uses a smarter, but more elaborate approach. First, the central cross-section of the object is reconstructed without any beam hardening correction applied. This slice is thresholded, to separate the material from the air. Next, for different projection angles, the path lengths of the X-rays through the object are calculated. These path lengths are then compared against the corresponding measured attenuation values and a curve is fitted to the plotted points. Finally, the attenuation values of all projections are corrected using the obtained curve.

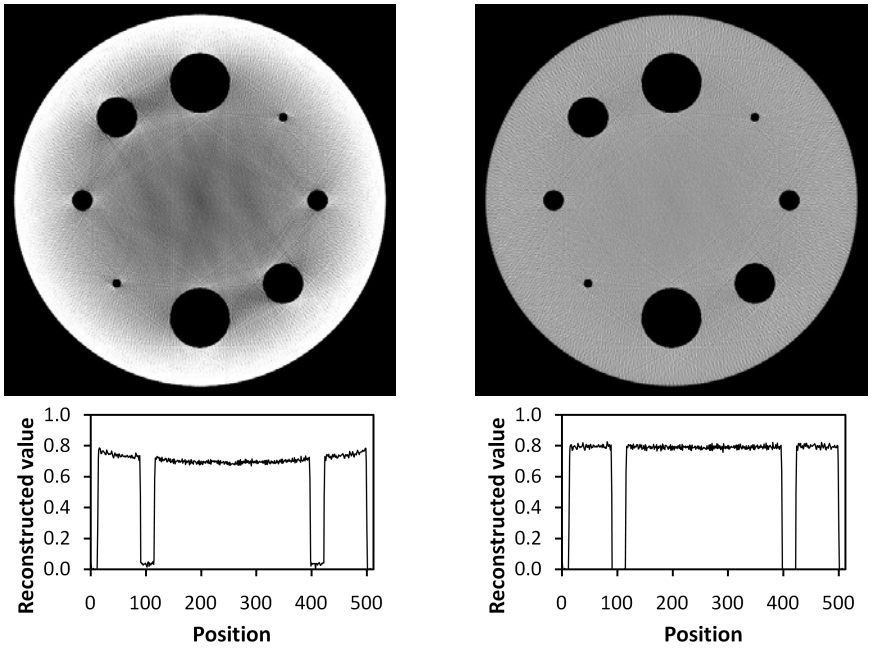
The fitted curve of the second method can be a simple polynomial function. Vandecasteele *et al.* [29] derived a more funded representation of this curve, based on a bimodal energy model. The energy response of the total system is assumed to be composed of a low and a high energy component, each of which is characterised by a mean energy E_i and a linear attenuation coefficient μ_i. By decomposition of

the law of Lambert-Beer (3.4) into a linear combination of these two components, the relation between the measured projection values and the material thickness d can be expressed as:

$$-\ln \frac{I}{I_0} = \mu_2 + \ln \left[\frac{1+\alpha}{1+\alpha\, e^{(\mu_2-\mu_1)d}} \right], \tag{3.57}$$

where μ_1, μ_2 and α are the parameters that need to be fitted.

The effect of such a correction method can be observed in figure 3.21. For objects that mainly consist of a rather homogeneous matrix containing several pores and inclusions, these methods of beam hardening correction work very well. However, when the object contains some larger, strongly attenuating inclusions (for instance metal parts), these simple methods fail. A correction for such metal artefacts will be evaluated later. It is an iterative process that requires a reprojection of the reconstructed volume, which will be discussed in detail in chapter 4.



(a) Uncorrected (b) Corrected

Figure 3.21: Reconstructed cross-sections and line profiles of projection data suffering from beam hardening. The resulting cupping effect was reduced by applying a polynomial correction.

3.8 Projection simulator

In order to verify the quality of the reconstruction algorithms described in this chapter and the next, a program was developed to simulate projection data for a monochromatic X-ray beam. These projection data can be simulated in either parallel-, fan-, cone- or helical cone-beam geometry. A whole range of parameters can be adapted to produce all kinds of misalignments of the simulated scanner, for instance vertical centre and centre of rotation, tilt, skew and slant of the detector, etc.

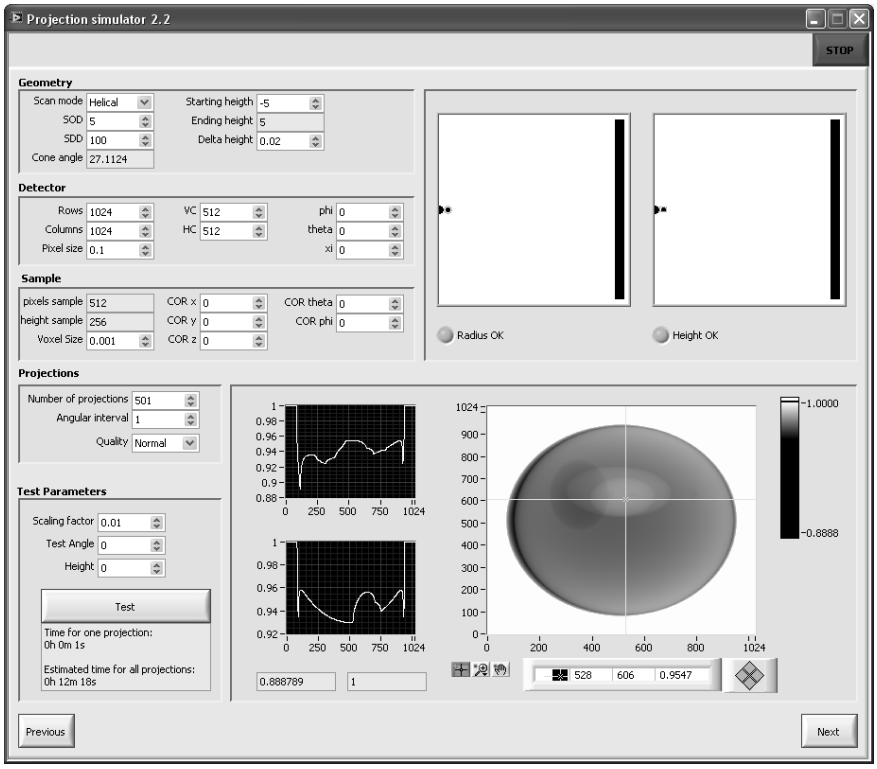


Figure 3.22: Screenshot of the projection simulator.

Simulated data can be produced for three kinds of phantom. The first phantom is any combination of ellipsoids, like the Shepp-Logan phantom [30] (see appendix A) which is the most common used phantom in quality assessment of reconstruction algorithms. The attenuation of this type of phantom can be calculated analytically by adding the attenuation of each individual ellipsoid, which is simply given by the intersection length of the ray and the ellipsoid multiplied by its linear attenuation value. For the second type, a grid pattern can be simulated. This pattern can be composed of either a collection of dots located at the grid points, lines parallel to the X-, Y- and Z-axis, or planes parallel to the XY-, YZ-

and XZ-plane. Another possible pattern that can be selected is the Defrise phantom (appendix A), which consists of a stack of discs, alternated by areas without any material (air). The third possibility is a stack of bitmap images, which can be used to represent about any object shape. Calculating the projection image for a custom object is done by *ray tracing*, a process which is commonly used in 3D visualization. Ray tracing will be explained in more detail in chapter 4, as it is the basic step in any iterative reconstruction algorithm. A screenshot of the projection simulator is shown in figure 3.22.

3.9 Improving reconstruction speed

As was mentioned above, the number of calculations required for the 3D reconstruction of the object function is of the order $O(N^4)$. It is clear that the reconstruction process poses high computational demands. In medical applications, reconstruction is required almost instantaneously, especially for the upcoming 4D scanners that provide real-time visualization of parts of the human body, which is used to guide surgeons during operations. In most nonmedical applications real-time reconstruction is not a prerequisite, however fast reconstruction is highly desirable to allow for an acceptable throughput of the system. Several routes can be chosen to increase the performance of reconstruction algorithms, of which the most common methods are discussed below.

3.9.1 Fast backprojection

The most elegant way to speed up the reconstruction is obviously a decrease in the computational complexity of the algorithm, in particular the backprojection step which claims the majority of the processing time. Several fast backprojection algorithms have been developed and presented, independent from each other, among which are the multilevel inversion of the Radon transform [31], the $O(N^3 \log N)$ backprojection [32] and the link method [12]. All of these methods are based on some decomposition scheme, breaking down the backprojection to a series of iteratively smaller processing steps until some basic step is reached. This decomposition is of an approximate nature, but it was shown that under certain conditions the image quality is not significantly influenced.

3.9.2 Cylindrical reconstruction

A significant increase in speed can be realized by reconstructing the object function on a grid in cylindrical coordinates (r, φ, z) [33]. The angular sampling step $\Delta\varphi$ of the cylindrical grid should be chosen such that $\Delta\theta = N\Delta\varphi$, where $\Delta\theta$ is the angle between two consecutive projections and N is an integer number.

Combined with a sinogram based reconstruction, the number of calculations in the backprojection step can be reduced to a minimum while usage of the CPU cache memory is optimised, resulting in a very fast backprojection routine. When the backprojection is complete, the cylindrical grid can be interpolated to a Cartesian grid with only minimal errors. Note that, unlike the direct Fourier methods, this conversion from cylindrical to Cartesian coordinates is performed in the spatial domain for which a simple interpolation method is sufficient.

3.9.3 Server-client architecture

The reconstruction process is highly suited for parallel implementation. It is therefore straightforward to use multiple CPU cores simultaneously to decrease the reconstruction time. Additional scaling can be obtained by distributing the computational work load over several (multicore) client PC's, by using a server-client architecture [34]. The volume to be reconstructed is divided in several batches of cross-sections, which are assigned to the different clients. Using an efficient implementation, the reconstruction speed can be scaled almost linearly with the number of available processing cores.

3.9.4 Specialized hardware

Many routines in image processing are very amendable to parallel implementation, CT reconstruction being one of them. For such applications, specialized hardware was designed providing a high amount of parallel computational power. The most important types will be briefly discussed.

FPGA

An *FPGA* or *Field-Programmable Gate Array* is an integrated circuit that can be reconfigured by a hardware designer. It consists of programmable logic components and interconnects which can be programmed to perform basic logic and simple math functions. Its internal structure provides resources for building high performance data processing systems and allows programmers to hardcode a specific algorithm without incurring the overheads of instruction-style programming. Several impressive results were obtained using FPGA's to speed up the reconstruction of CT data [35–37]. Efficient usage of FPGA's is however not straightforward. Furthermore, an FPGA is not a very flexible device. While it can be optimised to process a certain routine, it needs to be reconfigured when it is used to execute a different algorithm.

Cell BE

The *Cell Broadband Engine* consists of one main processor element and eight tightly coupled processing elements [38, 39]. It was specifically designed to be used in the Sony Playstation 3, but was also made available for common use by Mercury Computer Systems. An application of the Cell BE can be found in [40], where it was used in the reconstruction of cone-beam data using the FDK algorithm. At this point however, it is unclear if the development of this type of hardware will be continued in the future. Combined with its high cost, this solution does not seem suitable for common use.

GPU

In recent years, the booming gaming industry has pushed manufacturers to produce high end *graphical processing units* or *GPUs* at an acceptable price. In figure 3.23, a comparison is shown of the evolution in performance[9] between the latest Nvidia GPUs and Intel CPUs. The latest generations of GPUs contain hundreds of processing cores on a single card, providing a massive amount of parallel processing power. Although these GPUs are originally designed to produce renderings of complex 3D scenes at an impressive frame rate, their vast computational power can be exploited to perform other operations as well. It appears that the reconstruction process in computed tomography is very much suited to be accelerated using an GPU [41–43].

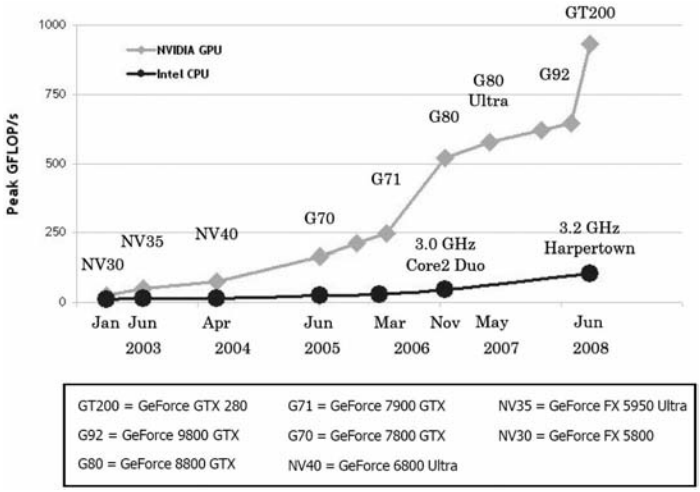


Figure 3.23: Comparison of the evolution in performance between the latest Nvidia GPUs and Intel CPUs. (From [44])

[9] Expressed in giga flops per second (GFLOPS/s), a flop being short for *floating point operation*

Until a few years ago, tomographic reconstruction could be performed by translating the algorithm to the OpenGL (Open Graphics Library) language, which is a standard language developed in order to control GPUs of different manufacturers. This translation is however not straightforward, as the OpenGL language is designed specifically for computer graphics and not for performing custom operations. In 2007, Nvidia released its *Compute Unified Device Architecture (CUDA)*, which allows performing generic calculations on the GPU using a programming language with a C-interface. The coming of CUDA enabled usage of the massively parallel computational power of GPUs by any researcher or developer with some programming skills. As the development of GPUs is thriven by the huge market of today's gaming industry, it can safely be assumed that this evolution will be continued in the future. Newer generations will emerge at an impressive rate, providing researchers with more computational power than ever dreamed of. Furthermore, due to the size and the competitive nature of the market, the prices of GPUs are and will remain rather low, in contrast with other high performance platforms like the Cell BE processor and FPGA technologies.

Larrabee

It is worth noting that Intel has been working on a new type of hardware called *Larrabee* [45], which is considered a hybrid between CPU and GPU. It has 32 cores and is created in order to compete with the latest high performance GPUs from ATI and Nvidia. Undoubtedly, this latest development holds great promise for increasing reconstruction speed and should certainly be considered in any future research.

OpenCL

OpenCL (Open Computing Language) is a new framework created to develop applications executing across heterogeneous platforms like CPUs, GPUs, Larrabee and others. It introduces a standard language that can be used to exploit any of these devices using the same code, so there will no longer be a need to write different code depending on the device one intends to use. At the point when this work was written, OpenCL was only very recently released and was still in its beta-testing stage. Therefore it was not considered for implementation yet. As with the Larrabee however, it should definitely be considered in future work.

3.9.5 GPU implementation

Before this work was started, the CPU implementations for the parallel-, fan- and cone-beam geometry were already highly optimised by reconstructing on a cylindrical grid, resulting in very fast reconstructions. The helical reconstruction however, which was implemented during this work, does not profit as much from this

cylindrical approach as the other geometries. The reason being that minimizing calculations and optimizing cache cannot be made very efficient, due to the fact that by using a helical path a voxel is projected at a different height for every projection. As a result, CPU-based reconstruction of a helical scan remains very time consuming, thereby limiting its practical use even though its benefits are rather obvious.

During this work, the reconstruction algorithms for parallel-, fan-, cone- and helical cone-beam described in the previous sections were implemented on the GPU in order to increase their performance. A Cartesian grid was used as it was found that the cylindrical grid did not allow for an efficient implementation due to the GPU's completely different architecture. A further improvement was achieved by executing the backprojection step on the GPU in parallel with the remaining operations of the reconstruction process on the CPU (file input and output, weighting and filtering, etc.). The most significant gain was achieved for the helical algorithm, thereby eliminating one of the major practical obstacles in using the helical geometry in high resolution CT.

The experience that was obtained by porting these algorithms to the GPU also proved to be extremely useful for the research on iterative algorithms in chapter 4. As these families of reconstruction methods pose an even higher computational burden, implementing them on the GPU was imperative in order to obtain reconstruction speeds that are acceptable in a practical context.

3.10 Octopus

Octopus is a software package for the reconstruction of tomographic data. It is developed in cooperation between UGCT and inCT. The package is developed to be independent of the scanner set-up and to provide flexible and high performance data processing options. It contains a server-client architecture for increased performance. Originally, Octopus was created to reconstruct projection data from parallel neutron beam tomography [46]. After the development of the high resolution scanner at UGCT, it was necessary to introduce the processing of cone-beam data in Octopus [47]. Following the improvements that can be obtained using the helical cone-beam geometry and the introduction of this new scanning path to the UGCT set-up, the Katsevich reconstruction algorithm was added to Octopus in the course of this work. Additionally, Octopus was extended with the GPU-based implementations for all common geometries (parallel, fan, cone and helical cone), enabling any user to benefit from the increase in reconstruction speed.

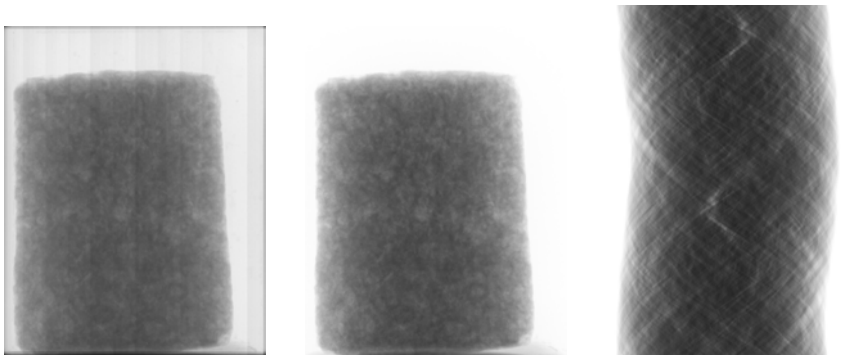
A more detailed description about Octopus can be found in [34]. Here, only the basic processing steps will be briefly treated, as to explain the reader how one goes from a set of measured projection data to a complete 3D reconstruction of the object. A discussion on the performance of Octopus is also given. It is noted that every reconstruction in this work is performed using Octopus.

3.10.1 Basic processing steps

The raw projection images coming from the detector are first filtered using a spot filter. A spot filter removes broken pixels (dark or bright) from the images by applying a thresholded median filter. These filtered images are then normalised by flat field and offset images. Flat field images are acquired by moving the sample out of the detector's field of view while the X-ray source is on. Flat field images, denoted by I_0, thus contain the intensity profile of the X-ray beam. Offset images I_d are taken when the X-ray bundle does not hit the detector (by turning of the source or shielding the detector) and correct for the detector's dark current. The normalised projection images I_N are then obtained from the measured projections I using:

$$I_N = \frac{I - I_d}{I_0 - I_d} \,. \tag{3.58}$$

Additionally, if a region on the detector can be selected which is never covered by the object, the average value of this region can be used to correct the normalised images for fluctuations in the bundle intensity. Next, a ring filter is applied to correct for systematic pixel deviations, which would otherwise result in ring artefacts in the reconstructed slices. The obtained images are now reordered into *sinograms*. A sinogram contains the projection data of a single detector row for all projection angles. Sinograms originate from the parallel- and fan-beam geometry, where each sinogram can be used to reconstruct one cross-section through the object. The optimised structure of the current implementation of the FDK algorithm in Octopus also requires sinograms. The helical algorithm on the other hand does not require sinograms but uses the normalised images. Figure 3.24 illustrates the obtained images for some of the processing steps.



(a) Original image (b) Normalised image (c) Sinogram

Figure 3.24: Illustration of the basic processing steps.

After these steps have been completed, the actual reconstruction of the object can begin. In Octopus, it is possible to reconstruct only a single slice at any location in the object. This allows the user to evaluate the effect of every parameter

that can be adjusted in the reconstruction module (figure 3.25). These parameters include corrections for misalignment of the system, application of a noise filter, beam hardening correction, etc. After each of these parameters is selected with care, the reconstruction of the 3D object function is performed.

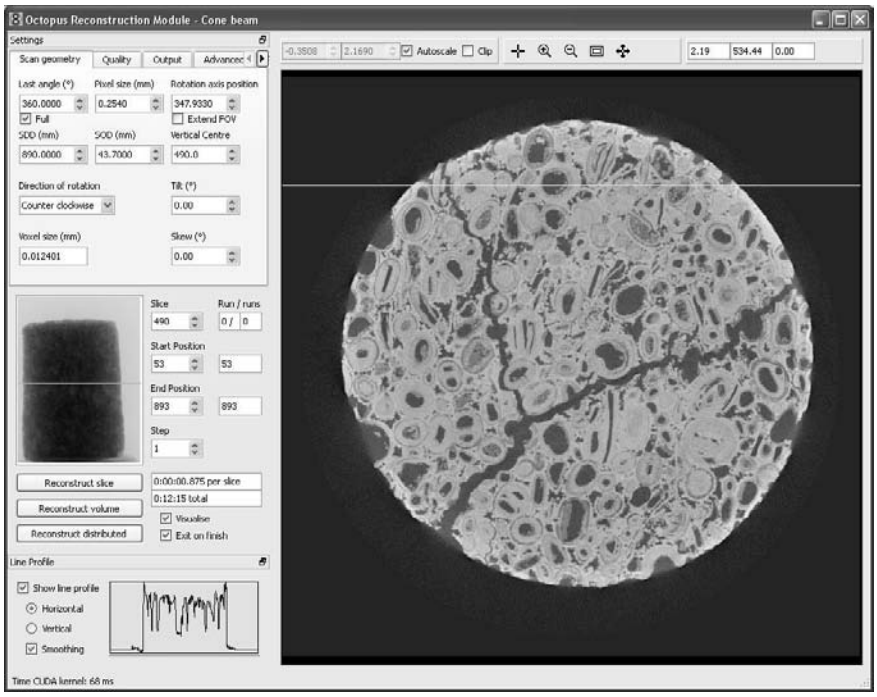


Figure 3.25: Screenshot of the reconstruction module of Octopus.

3.10.2 Performance

The performance of the Octopus package can be illustrated by providing some timing results for a typical cone-beam data set obtained at UGCT, given in table 3.2. The CPU-based reconstruction (using the optimization of the cylindrical grid) on a PC with two Intel Quad-Core Xeon E5430 (2.66 GHz) processors (8 cores) took about 36 minutes. With the server-client architecture, using 4 client PC's each housing two Intel Quad-Core Xeon E5405 (2.00 GHz) processors, (32 cores in total) this was reduced to nearly 12 minutes. The GPU-version was tested on the workstation described in table 3.1, which was used for all tests presented in this work, unless stated otherwise. On this test PC, the same data set was reconstructed in only 12 minutes, making it as fast as the cluster at only a fraction of its price.

For the helical cone-beam geometry, a volume of 1024 cross-sections of 1024 $\times$ 1024 voxels was reconstructed from 1158 projections of 1024 $\times$ 1024 pixels.

The reconstruction on the CPU using all four processing cores took 104 minutes. Using the GPU this only took 25 minutes, which corresponds to a speed-up of a factor of 4. Since the reconstruction time scales almost linear with the number of processing cores, the GPU-version is thus 16 times faster than the CPU-version using only one core. For comparison, a similar scan of the same object in cone-beam geometry requires 554 projections in short scan mode. The reconstruction time for this scan was 12 minutes for the CPU and 4 minutes for the GPU. For cone-beam, using the GPU thus speeds up the reconstruction by a factor 3 compared with the CPU using 4 cores, and by a factor 12 when using only one core.

Table 3.1: Specifications of the workstation used for the tests presented in this work.

Processor	Intel Core 2 Quad Q9550, 4 cores at 2.83 GHz
Graphics card	Nvidia Tesla C1060, 240 cores at 1.3 GHz
RAM memory	Corsair 4096 MB DDR3 XMS3-1600
Hard disk	Western Digital RE2, 7200 RPM

Table 3.2: Parameters of a typical data set.

Detector	Varian Paxscan 2520V (1880 rows, 1496 columns)
Cross-section size	1496 x 1496
# Cross-sections	1880
# Projections	1000
Source-object distance	170 mm
Source-detector distance	890 mm
Vertical centre (pixels)	890
Centre of rotation (pixels)	710.3
Last angle	360°
Detector sampling	Bilinear interpolation
Filter	Regular ramp-filter
Tilt angle	-0.20°
Beam hardening correction	None

It has to be noted that the increase in reconstruction speed of the GPU implementation is seemingly not nearly as impressive as results presented in other work, where improvements up to 10 to 100 times are claimed. This can simply be explained by noting that such comparisons are usually obtained by evaluating the same implementation on both CPU and GPU. However, due to the different architecture of the hardware, this implementation will most likely not be optimal for both platforms at the same time, as was already noted regarding the cylindrical approach. These timings are thus most often obtained by comparing an optimised GPU-version with a nonoptimised CPU-version, hence the large differences. Furthermore, such speed-ups are usually reported compared to a CPU version using

only a single processing core. In this work, timings were compared against an already highly optimised cylindrical reconstruction code. Timings were also obtained for a CPU implementation using a Cartesian grid, using only one processor core. Under these conditions, the reported improvement of 10-100 times was easily achieved. For this reason, one should always look at absolute reconstruction timings. Doing so, it was found that the reconstruction times obtained in this work are very competitive with other high end commercial reconstruction packages.

3.11 Applications

To conclude this chapter, some examples of scans of real objects are shown, where the helical cone-beam geometry was chosen instead of the common cone-beam geometry. There are two main reasons why a helical cone-beam scan may be preferred over a cone-beam scan. The first is that the reconstruction of a helical cone-beam scan does not contain cone-beam artefacts. The second advantage is that using a helical trajectory it is possible to scan very long objects in one scan at the highest resolution. In cone-beam geometry one needs to scan such an object with a lower magnification so that the required region can be completely covered by the detector, which results in a loss of resolution. Alternatively, one could perform several scans of the object at different heights until the whole range is covered. Each of these scans then needs to be reconstructed separately and stitched afterwards, which is a cumbersome and difficult task. The application of the helical cone-beam geometry is a relatively new development in high resolution X-ray CT and it is yet only rarely used in practice. However, as will be illustrated by the following examples, it has the potential of becoming an excellent research tool in a wide range of applications. Helical cone-beam CT is therefore believed to become a standard technique complementary to cone-beam CT in the near future.

3.11.1 Soda can

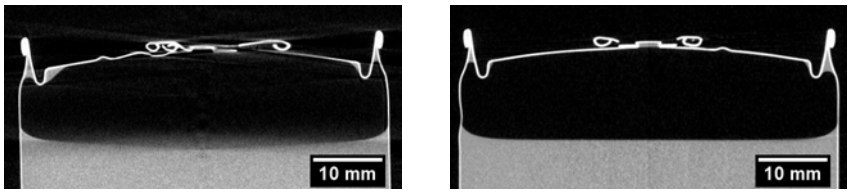
The first example is the CT scan of a small soda can of height 88.5 mm and diameter 53 mm. The can was scanned twice using the same geometrical configuration, acquiring projections following a regular cone-beam trajectory for the first scan and following a helical path for the second scan. The geometry of the scans is described in table 3.3 and corresponds to a full cone angle of 17.3°. The voxel size of both reconstructed volumes is 88.4 µm. For the helical scan, the pitch was 80 mm, so the step size of the translation movement between two consecutive projections was 0.1 mm.

The reconstruction results for both scans are presented in figure 3.26. A cone-beam artefact is clearly observed at the interface between the liquid and the gas inside the top of the can. Instead of a sharp edge, this artefact produces a smooth transition between the liquid and the vacuum. Further cone artefacts are found at

Table 3.3: Scan settings for the helical cone-beam scan of the soda can.

# Projections:	700	
# Detector rows:	920	
# Detector columns:	748	
Pixel pitch:	254	µm
Source to object:	270	mm
Source to detector:	770	mm
Cone angle:	17.3	°
Angular step:	0.45	°
Helical pitch:	80	mm
Voxel size:	88.4	µm

the interfaces between the top of the can and the gas inside, and between the top and the air surrounding the can. In the helical reconstruction, these artefacts are not present. Hence, the helical cone-beam scan provides a reconstruction of much higher quality than the cone-beam scan.



(a) Cone-beam (b) Helical cone-beam

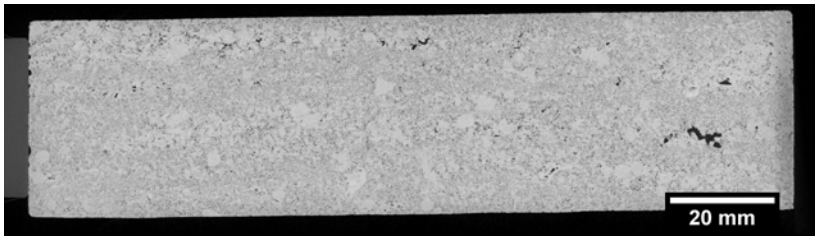
Figure 3.26: The reconstructions of the scans of a soda can, acquired in cone-beam (a) and in helical cone-beam geometry (b). The images represent vertical cross-sections parallel to the XZ-plane, through the centre of the can. Note that both images represent a similar, but not an identical part of the can.

3.11.2 Worm tunnels

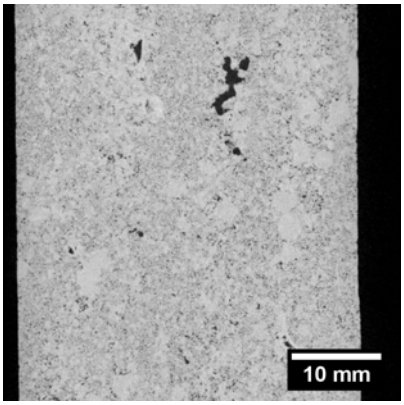
In the second application, a piece of rock was scanned in which several tunnels were dug by small worms. High resolution X-ray CT is an excellent tool for investigating these tunnels. Since the tunnels usually run in the same direction, normal to the surface out of which the sample was retrieved, studying them is best performed using an elongated sample. Hence, helical cone-beam is obviously the most suited scanning geometry. The parameters of this scan are given in table 3.4. The reconstruction volume consists of 3230 slices of 925×925 voxels. The reconstruction time of this volume was 60 minutes using the workstation described in table 3.1. Les Brownlow (XRT®) is acknowledged for providing this sample.

Table 3.4: Scan settings for the rock containing the worm tunnels.

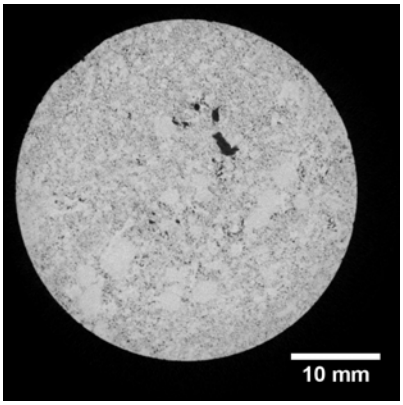
# Projections:	3167	
# Detector rows:	710	
# Detector columns:	925	
Pixel pitch:	254	µm
Source to object:	125	mm
Source to detector:	650	mm
Cone angle:	17.3	°
Angular step:	0.36	°
Helical pitch:	60.7	mm
Voxel size:	48	µm



(a) Full length



(b)



(c)

Figure 3.27: Reconstruction of a piece of rock containing several worm tunnels. Cross-sections (a) and (b) are taken parallel to the XZ-plane, section (c) is parallel to the XY-plane.

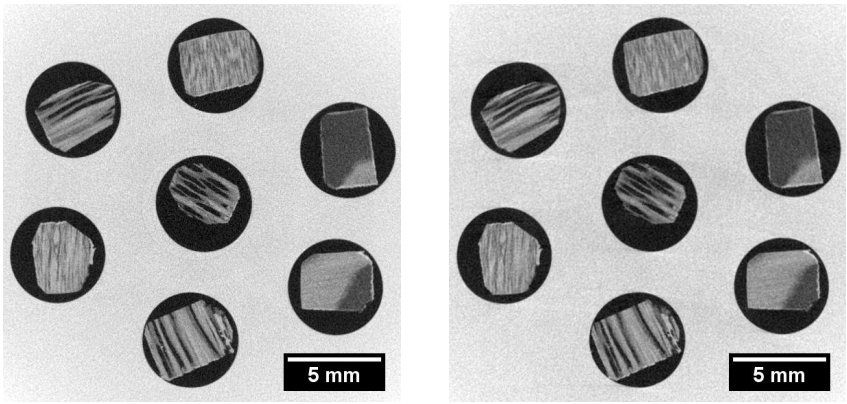
3.11.3 Wood densitometry

The last application example of the helical cone-beam geometry concerns the field of wood densitometry. In wood densitometry, a thin cylinder is drilled out of a tree trunk. The drilling is performed in the direction normal to the trunk's surface, such that the retrieved drill core contains a large number of growth rings. This drill core is then used to determine the variations in the radial density of the trunk, which offers information on the growth pattern of the tree. This growth pattern can then be used to study the growth of a type of tree under different conditions (location, soil, climate, etc.) or to compare different types of tree. Dries Vansteenkiste (Laboratory of Wood Technology, Universiteit Gent) is acknowledged for providing this sample.

Table 3.5: Scan settings for the wood densitometry.

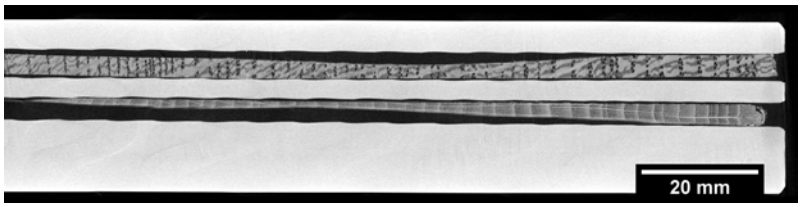
# Projections:	3579	
# Detector rows:	710	
# Detector columns:	925	
Pixel pitch:	254	µm
Source to object:	95	mm
Source to detector:	650	mm
Cone angle:	8.9	°
Angular step:	0.36	°
Helical pitch:	45.8	mm
Voxel size:	36.5	µm

The conventional techniques to measure this density are quite cumbersome, requiring many hours (sometimes days) of manual labour. Using X-ray CT, the density of the pieces of wood can be measured with a much higher efficiency. Since the drilled pieces of wood are usually very long as compared to their diameter, it is obvious that helical cone-beam is much more beneficial than the regular cone-beam geometry. For this application, seven holes were drilled in a plastic cylinder. A piece of wood was placed in each of these holes, after which the entire cylinder was scanned. To evaluate the obtained image quality, the cylinder was scanned both in cone-beam and in helical cone-beam. The parameters of these scans are given in table 3.5. Reconstructed cross-sections for both scans are shown in figure 3.28. At first sight, almost no difference in quality can be observed between the two scans. When looking closely a both cross-sections, it seems that the cone-beam reconstruction is slightly sharper than the helical cone-beam scan. However, this small difference is negligible compared to the major advantage of using a helical trajectory. Results for the full reconstruction of the helical scan are presented in figure 3.29. The total reconstruction volume contains 3720 cross-sections of 925×925 voxels and was reconstructed in 80 minutes on the workstation. The obtained image quality is really impressive, certainly regarding the small scale of the wood pieces. It is thus shown that high resolution X-ray CT using the helical trajectory can be an extremely valuable tool for wood densitometry.

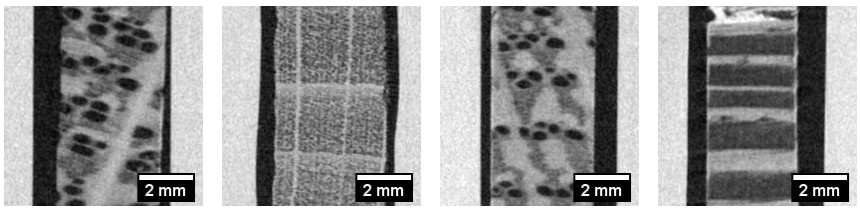


(b) Cone-beam (b) Helical cone-beam

Figure 3.28: Reconstructed cross-sections (parallel to the XY-plane) of the cone-beam (a) and the helical cone-beam scan (b) of the seven pieces of wood from the wood densitometry sample.



(a) Full length



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

Figure 3.29: Reconstruction results of the wood densitometry sample, showing cross-sections parallel to the XZ-plane along the full length of the scanned region (a), and zoomed in on a small region of four of the pieces of wood (b-e).

References

[1] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, 1988.

[2] S. Pan and A. Kak. *A computational study of reconstruction algorithms for diffraction tomography: interpolation versus filtered backprojection*. IEEE Transactions on Acoustics, Speech and Signal Processing, 31(5):1262–1275, 1983.

[3] H. Stark, J. W. Woods, I. Paul, and R. Hingorani. *Direct Fourier reconstruction in computer tomography*. IEEE Transactions on Acoustics, Speech and Signal Processing, 29(2):237–245, 1981.

[4] J. O'Sullivan. *A fast sinc function gridding algorithm for Fourier inversion in computer tomography*. IEEE Transactions on Medical Imaging, 4(4):200–207, 1985.

[5] M. Magnusson-Seger. *Linogram and other direct Fourier methods for tomographic reconstruction*. Phd dissertation, Linköping University, 1993.

[6] A. Rosenfeld and A. C. Kak. *Digital Picture Processing*. Academic Press, New York, 1982.

[7] D. L. Parker. *Optimal short-scan convolution reconstruction for fan beam CT*. Medical Physics, 9(2):254–257, 1982.

[8] A. Averbuch, R. Coifman, M. Israeli, I. Sedelnikov, and Y. Shkolnisky. *Advances in signal transforms: theory and application*. Hindawi Publishing Corporation, New York, 2007.

[9] H. Tuy. *An inversion formula for cone-beam reconstruction*. SIAM Journal of Applied Mathematics, 43(3):546–552, 1983.

[10] A. Katsevich. *Image reconstruction for the circle-and-arc trajectory*. Physics in Medicine and Biology, 50(10):2249–2265, 2005.

[11] G. L. Zeng and G. T. Gullberg. *A cone-beam tomography algorithm for orthogonalcircle-and-line orbit*. Physics in Medicine and Biology, 37(3):563–577, 1992.

[12] H. Turbell. *Cone-beam reconstruction using filtered backprojection*. Phd dissertation, Linköping University, 2001.

[13] P. Grangeat. *Analyse d'un système d'imagerie 3D par reconstruction a partir de radiographies x en géométrie conique*. Phd dissertation, Ecole Nationale Supérieure des Télécommunications, 1987.

[14] M. Defrise and R. Clack. *A cone-beam reconstruction algorithm using shift-variant filtering and cone-beam backprojection*. IEEE Transaction on Medical Imaging, 13(1):186–195, 1994.

[15] C. Jacobson. *Fourier methods in 3D-reconstruction from cone-beam data*. Phd dissertation, Linköping University, 1996.

[16] L. A. Feldkamp, L. C. Davis, and J. W. Kress. *Practical cone-beam algorithm*. Journal of the Optical Society of America A, 1(6):612–619, 1984.

[17] M. Grass, Th. Köhler, and R. Proksa. *3D cone-beam CT reconstruction for circular trajectories*. Physics in Medicine and Biology, 45(2):329–347, 2000.

[18] H. Kudo, T. Rodet, F. Noo, and M. Defrise. *Exact and approximate algorithms for helical cone-beam CT*. Physics in Medicine and Biology, 49(13):2913–2931, 2004.

[19] K. Zeng and Z. Chen. *Review of recent developments in cone-beam CT reconstruction algorithms for long-object problems*. Image Analysis and Stereology, 23(2):83–87, 2004.

[20] A. Katsevich. *Analysis of an exact inversion algorithm for spiral cone-beam CT*. Physics in Medicine and Biology, 47(15):2583–2597, 2002.

[21] M. Defrise, F. Noo, and H. Kudo. *A solution to the long-object problem in helical cone-beam tomography*. Physics in Medicine and Biology, 45(3):623–643, 2000.

[22] K. Champley. *Katsevich inversion formula for 3D-D helical tomography*. Math 656 presentation notes (unpublished), Oregon State University, 2006.

[23] H. Yu and G. Wang. *Studies on implementation of the Katsevich algorithm for spiral cone-beam CT*. Journal of X-Ray Science and Technology, 12(2):97–116, 2004.

[24] P. E. Danielsson, P. Edholm, J. Eriksson, and S. M. Magnusson. *Towards exact reconstruction for helical cone-beam scanning of long objects*. In Proceedings of Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, pages 141–144, Pittsburgh, 1997.

[25] K. C. Tam, S. Samarasekera, and F. Sauer. *Exact cone beam CT with a spiral scan*. Physics in Medicine and Biology, 43(4):1015–1024, 1998.

[26] A. J. Wunderlich. *The Katsevich inversion formula for cone-beam computed tomography*. Phd dissertation, Oregon State University, 2006.

[27] F. Noo, J. Pack, and D. Heuscher. *Exact helical reconstruction using native cone-beam geometries*. Physics in Medicine and Biology, 48(23):3787–3818, 2003.

[28] M. M. Seger. *Rampfilter implementation on truncated projection data. Application to 3D linear tomography for logs*. In Proceedings Symposium on Image Analysis, pages 33–36, Lund, 2002.

[29] E. Van De Casteele, D. Van Dyck, J. Sijbers, and E. Raman. *An energy-based beam hardening model in tomography*. Physics in Medicine and Biology, 47(23):4181–4190, 2002.

[30] L. S. Shepp and B. F. Logan. *The Fourier reconstruction of a head section*. IEEE Transactions on Nuclear Science, 21(3):21–43, 1974.

[31] A. Brandt, J. Mann, M. Brodski, and M. Galun. *A fast and accurate multilevel inversion of the radon transform*. SIAM Journal of Applied Mathematics, 60(2):437–462, 1999.

[32] S. Basu and Y. Bresler. *$O(N^3 logN)$ backprojection algorithm for the 3-D Radon transform*. IEEE Transactions on Medical Imaging, 21(2):76–88, 2002.

[33] N. Sorokin. *An FPGA-based 3D backprojector*. Phd dissertation, Universität des Saarlandes, 2003.

[34] J. Vlassenbroeck. *Advances in laboratory-based X-ray microtomography*. Phd dissertation, Ghent University, 2009.

[35] I. Goddard and M. Trepanier. *High-speed cone-beam reconstruction: an embedded systems approach*. In SPIE Medical Imaging 2002: Visualization, Image-Guided Procedures and Display, volume 4681, pages 483–491, 2002.

[36] M. Leeser, S. Coric, E. Miller, H. Yu, and M. Trepanier. *Parallel-beam backprojection: an FPGA implementation optimized for medical imaging*. In Proceedings Tenth International Symposium on Field Programmable Gate Arrays, pages 217–226, Monterey, 2002.

[37] J. Li, C. Papachristou, and R. Shekhar. *An FPGA-based computing platform for real-time 3D medical imaging and its application to cone-beam CT reconstruction*. Journal of Imaging Science and Technology, 49(3):237–245, 2005.

[38] D. Pham, S. Asano, M. Bolliger, M. N. Day, H. P. Hofstee, C. Johns, J. Kahle, A. Kameyama, J. Keaty, Y. Masubuchi, M. Riley, D. Shippy, D. Stasiak, M. Suzuoki, M. Wang, J. Warnock, S. Weitzel, D. Wendel, T. Yamazaki, and K. Yazawa. *The design and implementation of a first-generation CELL processor*. In IEEE International Solid-State Circuits Conference, volume 1, pages 184–185, San Francisco, 2005.

[39] B. Flachs, S. Asano, S. H. Dhong, P. Hotstee, G. Gervais, R. Kim, T. Le, P. Liu, J. Leenstra, J. Liberty, B. Michael, S. M. Mueller, O. Takahashi, A. Hatakeyama, Y. Watanabe, and N. Yano. *A streaming processing unit for*

a CELL processor. In IEEE International Solid-State Circuits Conference, volume 1, pages 134–135, San Francisco, 2005.

[40] M. Kachelriess, M. Knaup, and O. Bockenbach. *Hyperfast perspective cone beam backprojection*. In IEEE Nuclear Science Symposium Conference Record, volume 3, pages 1679–1683, San Diego, 2006.

[41] F. Xu and K. Mueller. *Ultra-fast 3D filtered backprojection on commodity graphics hardware*. In IEEE International Symposium on Biomedical Imaging: Nano to Macro, volume 1, pages 571–574, Arlington, 2004.

[42] K. Mueller, F. Xu, and N. Neophytou. *Why do commodity graphics hardware boards (GPUs) work so well for acceleration of computed tomography?* In Proceedings of SPIE, Computational Imaging V, volume 6498, San Jose, 2007.

[43] B. Jang and D. Kaeli. *Multi GPU implementation of iterative tomographic reconstruction algorithms*. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 185–188, Boston, 2009.

[44] *NVIDIA CUDA™ programming guide, version 2.3.1*. 2009.

[45] L. Seiler, D. Carmean, E. Sprangle, T. Forsyth, M. Abrash, P. Dubey, S. Junkins, A. Lake, J. Sugerman, R. Cavin, R. Espasa, E. Grochowski, T. Juan, and P. Hanrahan. *Larrabee: a many-core x86 architecture for visual computing*. ACM Transactions on Graphics, 27(3):1–15, 2008.

[46] M. Dierick, B. Masschaele, and L. Van Hoorebeke. *Octopus, a fast and user-friendly tomographic reconstruction package developed in LabView*. Measurement Science and Technology, 15(7):1366–1370, 2004.

[47] J. Vlassenbroeck, M. Dierick, B. Masschaele, V. Cnudde, L. Van Hoorebeke, and P. Jacobs. *Software tools for quantification of X-ray microtomography*. Nuclear Instruments and Methods in Physics Research A, 580(1):442–445, 2007.

I am ashamed to tell you to how many figures

I carried these computations,

having no other business at the time.

Isaac Newton, 1642° – 1727†

4

Iterative Reconstruction

Besides the analytical reconstruction algorithms, there exist alternative approaches such as iterative reconstruction methods. These consider the reconstruction process as the optimization of a discrete representation of the object function in order to satisfy a system of equations that describes the imaging modality. Iterative methods can provide reconstructions of higher quality, e.g. when only a limited number of projections is available, when the sampling of the projections is not equiangular or when certain orientations are missing. Other advantages are their better noise handling and the possibility to improve the modelling by incorporating certain physical effects like scattering, beam hardening, etc. The disadvantage of iterative algorithms is that they pose a high computational burden, which is the main reason they are not or only rarely used in X-ray CT. In this chapter, all important aspects of iterative algorithms are covered and an efficient implementation is derived, which enables the iterative reconstruction of high resolution CT data in a practical context.

4.1 Algebraic reconstruction

The object function $f(x, y, z)$ of the object is represented on a 3D grid of N voxels corresponding to function values f_j with $j = 1...N$. The line integral corresponding to a detector pixel can be written in a discrete form, which is called a *ray sum*. The relationship between these ray sums p_i and the function values f_j is expressed

by:

$$\sum_{j=1}^{N} w_{ij} f_j = p_i \ , \ i = 1 \dots M \ , \tag{4.1}$$

where $M = R\,M_p$ is the total number of measured ray sums in all the projection images, R being the number of projections and M_p the number of detector pixels in a projection image P_θ. The weights w_{ij} represent the contribution of the j-th voxel to the i-th ray sum. The ray sums p_i can be retrieved from the projection images by:

$$p_i = -\ln\left(\frac{I_i}{I_i^0}\right) \ , \tag{4.2}$$

where I_i is the X-ray intensity measured by the detector pixel corresponding to p_i and I_i^0 the measured, unattenuated intensity. The imaging process can then be seen as a system of linear equations:

$$\begin{aligned} w_{11}f_1 + w_{12}f_2 + \cdots + w_{1N}f_N &= p_1 \\ w_{21}f_1 + w_{22}f_2 + \cdots + w_{2N}f_N &= p_2 \\ &\cdots \\ w_{M1}f_1 + w_{M2}f_2 + \cdots + w_{MN}f_N &= p_M \ . \end{aligned} \tag{4.3}$$

In matrix notation, this can be written as:

$$\mathbf{W} \cdot \mathbf{F} = \mathbf{P} \ . \tag{4.4}$$

Here, $\mathbf{P}$ is the $(M \times 1)$ column vector containing the ray sums p_i, $\mathbf{F}$ is the $(N \times 1)$ column vector representing the function values f_j and $\mathbf{W}$ is the $(M \times N)$ matrix of the weights w_{ij}.

In theory, the object function may be obtained by inverting the system of equations in (4.3). In practice however, M is usually not equal to N and the system is under- ($M < N$) or overdetermined ($M > N$). Furthermore, the presence of noise in the projections yields an inconsistent system of equations. Nevertheless, even when $M = N$ and when noise is not present, the huge size of the matrix $\mathbf{W}$ excludes any kind of matrix inversion due to computational reasons. Least squares methods can be used to overcome the problems related to noise and M not being equal to N, but these are again very impractical due to the size of N. Consider the 2D reconstruction of a data set of 256 projections with 256 detector pixels on a grid of 256×256 voxels. The matrix $\mathbf{W}$ then has dimensions of 65.536×65.536, which already poses a severe computational challenge. In practice, data sets are typically much larger and the reconstruction is performed in 3D, so it is obvious that the matrix $\mathbf{W}$ would quickly become unmanageable.

4.1.1 The Kaczmarz method

For large values of M and N, the set of equations (4.3) can be solved by an iterative method introduced by Kacmarz [1] and further elucidated by Tanabe [2]. Let the

image be represented by the point $\vec{f} = (f_1, f_2, \ldots, f_N)$ in an N-dimensional space. Each of the equations in (4.3) then defines a hyperplane. When a unique solution exists, the hyperplanes intersect in a single point, which is the solution to the system of equations. To illustrate this method, consider the case where $M = N = 2$. The hyperplanes are then lines and the system of equations is given by:

$$\begin{aligned} w_{11}f_1 + w_{12}f_2 &= p_1 \\ w_{21}f_1 + w_{22}f_2 &= p_2 \, . \end{aligned} \tag{4.5}$$

This system is now solved as follows, as illustrated by figure 4.1. Project the initial solution $\vec{f}^{(0)}$ orthogonally onto the first line, reproject the resulting point $\vec{f}^{(1)}$ on the second line, then reproject $\vec{f}^{(2)}$ on the first line again and so on. When a unique solution exists, the projected point converges towards the intersection $\vec{f}$ of the two lines, which is the solution to the system.

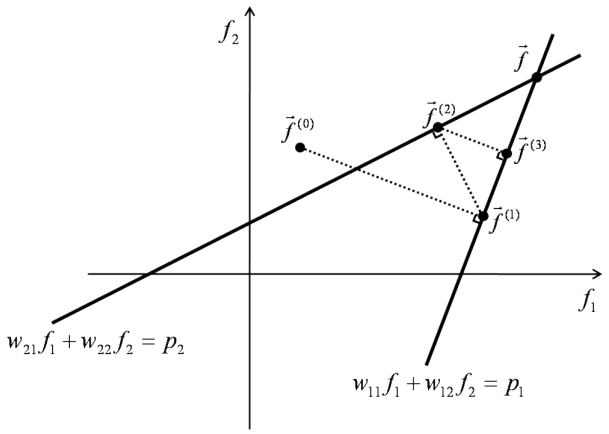


Figure 4.1: Illustration of the Kaczmarz method for two unknowns.

Turning back to the N-dimensional case, the initial solution $\vec{f}^{(0)}$ is projected onto the first hyperplane, given by the first equation in (4.3), resulting in the solution $\vec{f}^{(1)}$. This new solution is then projected onto the hyperplane represented by the second equation in (4.3), giving the solution $\vec{f}^{(2)}$, and so forth. Mathematically, this process can be expressed by [3]:

$$\vec{f}^{(k)} = \vec{f}^{(k-1)} + \frac{p_k - \vec{w}_k \cdot \vec{f}^{(k-1)}}{\vec{w}_k \cdot \vec{w}_k} \, \vec{w}_k \, , \tag{4.6}$$

where $\vec{w}_k = (w_{k1}, w_{k2}, \ldots, w_{kN})$. After each of the M equations in (4.3) is handled, the whole procedure can be repeated for another iteration and so on, until some condition of convergence is reached.

In practical situations, due to noise and computational errors, the hyperplanes will not necessarily intersect in one point and many approximate solutions exist. In

contrast with filtered backprojection methods, there is much freedom in the actual implementation of iterative methods and each of the choices made will have an effect on which of the approximate solutions is recovered. In sections 4.3 to 4.13, the different aspects of the implementation are discussed, justifying the choices that were made for the practical implementation presented in this work.

4.1.2 Algebraic reconstruction technique

The algebraic reconstruction technique (ART) can now be described by rewriting equation (4.6):

$$f_j^{(k+1)} = f_j^{(k)} + \lambda \frac{p_i - \sum_{n=1}^{N} w_{in} f_n^{(k)}}{\sum_{n=1}^{N} w_{in}^2} w_{ij} \ , \tag{4.7}$$

where an extra *relaxation parameter* λ was added. This parameter is usually chosen within the range $]0.0, 1.0]$ and is used to dampen the contribution of noise to the reconstruction. The update step of the volume given by equation (4.7) is performed for every equation in (4.3), thus for every pixel of every projection image. The initial solution is thus updated M times, each time based on a different detector pixel. If the solution has not yet converged sufficiently after processing every pixel in every projection image, the entire routine can be performed again starting with the solution obtained by the first iteration, and this can be repeated for a number of iterations.

The reconstruction process in (4.7) can be separated into three general processing steps. The first step is the *forward projection* or *reprojection* of the current solution, given by the calculation of the ray sum:

$$\hat{p}_i = \sum_{n=1}^{N} w_{in} f_n^{(k)} \ . \tag{4.8}$$

The second step is the *correction computation*, which computes the correction factor:

$$c_i = \lambda \frac{p_i - \hat{p}_i}{\sum_{n=1}^{N} w_{in}^2} \ . \tag{4.9}$$

Finally, the *backprojection* step is performed, in which the correction is distributed onto the reconstruction grid and an improved solution is obtained:

$$f_j^{(k+1)} = f_j^{(k)} + c_i \, w_{ij} \ . \tag{4.10}$$

In other words, the algebraic reconstruction procedure can be described as follows. Calculate the forward projection of the current reconstruction volume.

Compare this simulated projection value to the corresponding measured value and determine the correction of the volume. Apply this correction factor to the volume and obtain the updated reconstruction volume. Repeat this procedure for every detector pixel in every projection image for a number of iterations, until the solution converges sufficiently. Schematically, ART is described by the following routine:

- Initialise the reconstruction volume.
- Repeat until convergence:
  - Select a detector pixel p_i from projection P_θ.
  - Forward projection: Compute the ray sum corresponding to pixel p_i.
  - Correction computation: Subtract the ray sum from the measured value in p_i.
  - Backprojection: Distribute the correction into the volume.

In addition to ART, which applies an additive correction, alternatively one can use a multiplicative correction. This method, known as the Multiplicative Algebraic Reconstruction Technique (MART) [4] is described by:

$$f_j^{(k+1)} = \lambda \frac{p_i}{\sum_{n=1}^{N} w_{in} f_n^{(k)}} f_j^{(k)} . \tag{4.11}$$

This method inherently prohibits the object function values f_j to become negative, which is an advantage over ART. However, such multiplicative methods have several disadvantages and are rarely used in practice.

4.1.3 Simultaneous algebraic reconstruction technique

Instead of correcting the volume for every detector pixel in every projection image, one can also accumulate the corrections corresponding to a certain range of pixels before updating the reconstruction volume. This is usually done by gathering the contributions of every pixel in a certain number of projection images. The Simultaneous Iterative Reconstruction Technique (SIRT) [5] only applies the update after a full iteration, thus by accumulating the corrections from all pixels of every projection image. However, this requires a large number of iterations and thus a long processing time before the solution converges. Faster convergence is achieved using the Simultaneous Algebraic Reconstruction Technique (SART) [6], which updates the volume after accumulating the corrections from every detector pixel of one projection image. This method has the advantage that it was proven to be the better method for use in cone-beam reconstruction and that it is best suited for increasing performance using a graphical processing unit [3, 7].

Recently, the use of *Ordered Subsets* (OS), which originates from the reconstruction of single photon and positron emission tomography (SPECT and PET)

using expectation maximization (EM) methods [8], was introduced in the field of algebraic techniques [9] which provides a more general approach. Using OS, the update is performed after accumulating the correction for all pixels belonging to a certain set of S projection images. When $S = R$, this defaults to SIRT whereas $S = 1$ is equivalent to SART. The application of OS can thus be seen as the more general approach for simultaneous algebraic reconstruction techniques. Mathematically, the update process is described by:

$$f_j^{(k+1)} = f_j^{(k)} + \lambda \frac{\sum_{p_i \in OS_s} \left[\frac{p_i - \sum_{n=1}^{N} w_{in} f_n^{(k)}}{\sum_{n=1}^{N} w_{in}} \right] w_{ij}}{\sum_{p_i \in OS_s} w_{ij}}, \tag{4.12}$$

where the summation range $p_i \in OS_s$ indicates every detector pixel contained in the ordered subset of projections OS_s.

Two differences are observed compared to the original ART formula (4.7). First, the difference between the measured and the calculated ray sum is now divided by the sum of the weights, instead of by the sum of squared weights. Intuitively, it means that this difference is now normalised by the physical length $\sum_{n=1}^{N} w_{in}$ of the corresponding ray. Second, the update is now performed by calculating the weighted average of all corrections from every ray in the ordered subset. Looking at the two special cases $S = 1$ and $S = R$, it is obvious that the number of projections S used in the subsets determines the convergence speed of the solution. Higher S will result in slower convergence, but the reconstruction will suffer less from noise as the update is averaged over more values.

Simultaneous ART using ordered subsets can be described by the following scheme:

- Initialise the reconstruction volume.
- Repeat until convergence:
  - Select an ordered subset OS_s from the projection images.
  - Forward projection: Compute the ray sums corresponding to the pixels $p_i \in OS_s$.
  - Correction computation: Subtract the ray sums from the corresponding measured values for all $p_i \in OS_s$.
  - Backprojection: Distribute the corrections into the volume using a weighted average.

4.2 Statistical reconstruction

Another type of iterative reconstruction algorithms is provided by the statistical methods, which are based for instance on the maximum likelihood approach. In statistical reconstruction methods, the forward model of the acquisition is based on the expected number of transmitted photons $\hat{t}_i$ which can be expressed by:

$$\hat{t}_i = b_i \cdot \exp\left(-\sum_{j=1}^{N} w_{ij} f_j \right) , \tag{4.13}$$

which can be derived directly from the ray sums calculated in the algebraic techniques. The actual number of transmitted photons, given by t_i is retrieved from the measured X-ray intensity I_i. The values b_i represent the number of photons that are detected when the X-ray beam is not attenuated and are obtained from the measured X-ray intensities I_i^0.

Maximum likelihood for transmission tomography

The basic idea behind the statistical reconstruction is to find, for a set of measured transmitted numbers $\vec{t}$, the values of the object function $\vec{f}$ that maximize the probability $P(\vec{f}\,|\,\vec{t}\,)$, which by Bayes' rule can be given by:

$$P(\vec{f}\,|\,\vec{t}\,) = \frac{P(\vec{t}\,|\,\vec{f}\,) \cdot P(\vec{f}\,)}{P(\vec{t}\,)} \tag{4.14}$$

Since the term $P(\vec{t}\,)$ does not depend on $\vec{f}$, it can be neglected. The term $P(\vec{f}\,)$ depends on prior information about the object function and can be ignored when no prior information is available, which is generally the case. Optimization of the probability $P(\vec{f}\,|\,\vec{t}\,)$ is now thus equal to maximizing $P(\vec{t}\,|\,\vec{f}\,)$. This reduction of the reconstruction problem is referred to as the *maximum likelihood* approach.

The resulting probability can be factorized to:

$$P(\vec{t}\,|\,\vec{f}\,) = \prod_{i=1}^{M} P(\,t_i\,|\,\vec{f}\,) \,, \tag{4.15}$$

since statistical variations in the measured transmitted numbers t_i are not correlated. Maximization of this probability is equivalent to optimizing its logarithm:

$$\ln P(\vec{t}\,|\,\vec{f}\,) = \sum_{i=1}^{M} \ln P(\,t_i\,|\,\vec{f}\,) \,, \tag{4.16}$$

as this is a monotonic function.

Due to the quantum nature of photons, X-ray imaging systems can be assumed to follow a Poisson distribution. For an average count rate $\hat{t}$, the Poisson distribution t is described by:

$$P(\, t \,|\, \hat{t}\,) = \frac{\hat{t}^{\,t}\; e^{-\hat{t}}}{t!}\;. \tag{4.17}$$

This implies that the probability of measuring a transmission value t_i is solely determined by the expected transmission value $\hat{t}_i$, which depends on the object function $\vec{f}$. This means that:

$$P(\, t_i \,|\, \hat{t}_i\,) = P(\, t_i \,|\, \vec{f}\,)\;. \tag{4.18}$$

Combining equations (4.16 - 4.18) and ignoring constant terms, the problem of optimizing the probability $P(\,\vec{f}\,|\,\vec{t}\,)$ is reduced to maximizing the log-likelihood function *L* given by:

$$L = \sum_{i=1}^{M} (t_i \cdot \ln \hat{t}_i - \hat{t}_i)\;. \tag{4.19}$$

Since the transmitted numbers of photons $\hat{t}_i$ are calculated from $\vec{f}$ using a certain forward projection model, maximization of *L* optimizes the object function $\vec{f}$.

Optimization of the function *L* can be performed by using the *expectation maximization* method (EM) [10], which is an iterative method. The maximization step of this method leads to a transcendental equation that can be solved approximately by expanding the function as a truncated series. Using a simple gradient ascent algorithm [11], the optimization of the object function $\vec{f}$ can be formulated as [12]:

$$f_j^{(k+1)} = f_j^{(k)} - \frac{\left.\frac{\partial L}{\partial f_j}\right|_{\vec{f}^{\,(k)}}}{\displaystyle\sum_{i=1}^{M} \left.\frac{\partial^2 L}{\partial f_j\, \partial f_i}\right|_{\vec{f}^{\,(k)}}}\;. \tag{4.20}$$

As with the algebraic methods, the convergence of the solution can be accelerated by using ordered subsets. The summation $\sum_{i=1}^{M}$ over all detector pixels in every projection image can then be replaced by the summation $\sum_{t_i \in OS_s}$ which runs over every detector pixel in a certain set of projection images OS_s. Using the simple forward model in (4.13) and applying ordered subsets, equation (4.20) results in the following iterative update procedure:

$$f_j^{(k+1)} = f_j^{(k)} + \lambda\, \frac{\displaystyle\sum_{t_i \in OS_s} w_{ij}\; (\hat{t}_i - t_i)}{\displaystyle\sum_{t_i \in OS_s} w_{ij} \left[\sum_{n=1}^{N} w_{in}\right]\; \hat{t}_i}\;. \tag{4.21}$$

with $\hat{t}_i$ given by (4.13) and where a relaxation parameter λ was added. This algorithm is called the *maximum likelihood for transmission tomography* (ML-TR).

Although the basic ideas behind algebraic and statistical methods are quite different, the derived formalism for the iterative update routine is quite similar. In fact, only two major differences can be observed. The first is that the forward projection step in the algebraic methods uses the integrated attenuation values, whereas in the statistical methods it uses the expected number of photons. Secondly, in algebraic methods, the correction factors are obtained by normalizing the difference between the simulated and the measured values by the length of the ray. In statistical methods this normalization is not performed, however in the backprojection step two additional factors are used in the calculation of the denominator of the weighted average of the correction factors: the length of the ray and the expected number of transmitted photons.

4.3 Towards a practical implementation

Iterative reconstruction algorithms can be implemented in various different ways. First of all, a choice has to be made whether one prefers an algebraic or a statistical method. Furthermore, each iterative method consists of three major processing steps: forward projection, correction computation and backprojection, each of which can be executed using several alternative schemes. Next, one has to choose when the update of the volume is performed, either on a ray per ray basis or after accumulating corrections for a certain number of projection angles using ordered subsets. Finally, the number of iterations and the value of the relaxation parameter λ will have a definite influence on the reconstruction result.

It is obvious that combining all of these options yields a wide range of possible implementations of an iterative method. In the subsequent sections, the important aspects of a practical implementation are covered and an attempt is made to justify the implementation presented in this work. For application in high resolution cone-beam X-ray CT, in which volumes typically consist of a high number of voxels, the choices made to achieve a practical implementation are mainly dictated by the need for acceptable reconstruction times and memory requirements. In order to obtain a decent reconstruction speed, the implementation details are selected such that the algorithm allows for a highly efficient execution on a GPU. In perusing the following discussion, the reader should bear in mind these practical requirements.

4.4 Projection access order

A first practical issue that arises is the question in which order the projection images should be accessed during reconstruction. Already in the earliest literature on ART it was noticed that this access order has a definite influence on the performance and quality [3, 13]. A naive approach is to handle the projections sequentially. However, since the angular distance between two consecutive projections

is minimal, convergence of the solution is quite slow. Additionally, the lack of information from rather different angular positions tends to bias the reconstruction and may cause the solution to drift away from the desired solution.

A more well-considered approach is to access the projections such that two succeeding projections are only minimally correlated. Some commonly applied schemes are the prime number decomposition (PND) [14], the multilevel (MLS) [15] and the random access scheme (RAS) [16]. A more general and well-founded approach is given by the weighted distance scheme (WDS) [17]. This scheme minimizes the correlation between consecutive projections by arranging them such that:

1. A series of subsequently applied projections is evenly distributed across a wide angular range.
2. At no time is there an angular range that is covered more densely than others.

In words, the WDS can be described as follows. From a pool of unused projections, select the projection that optimizes both angular spacing and spread with respect to a subset of recently applied projections, taking into account that more recently applied projections have a stronger influence on the selection process.

The weighted distance scheme

Originally, the WDS was presented for a set of projections with a constant angular interval between 0° and 180°. Here, a slightly different derivation of the method is given which can also be applied when the projections are not equally spaced and when they cover a span of up to 360°. Consider a set of R projections P_i with angular positions θ_i, $1 \leq i \leq R$ and $0 \leq \theta_i < 360^\circ$. The pool Λ holds the L projections that are not yet used in the current iteration. The projection ordering is implemented as a circular queue Γ of length S. The current amount of used projections in Γ is denoted by Q, which can never become greater than S. The size of S determines the influence of the previously accessed projections to the selection process. The weight of each previous projection is given by a linearly decaying function such that the influence of early projections diminishes over time. Although S can be chosen arbitrary, $S = R$ seems to be a good choice, meaning that a projection's weight drops to zero after selecting R projections.

The queue Γ is empty ($Q = 0$) at the beginning of the reconstruction. At the start of each iteration, the pool Λ is filled with all R projections. When a projection P_l is selected, it is removed from the pool Λ and inserted into Γ. For $Q < S$, the projection is simply added to the queue. When $Q = S$, the circular queue is full and the oldest projection in Γ is replaced by the newly selected projection, which means that this old projection ceases to influence the selection process. This selection procedure is repeated until the list Λ is empty, after which it is refilled

for the next iteration, if necessary. The access order used in the reconstruction is now determined by the consecutive selection of the projections P_l.

The first projection to occupy the queue Γ can be chosen arbitrarily. Let this be P_0, which usually corresponds to $\theta_0 = 0^\circ$. Each succeeding projection $P_l \in \Lambda$ is now selected such that it minimizes the norm:

$$D_l = \tilde{\mu}_l^2 + \alpha \cdot \tilde{\sigma}_l^2 \ . \tag{4.22}$$

Choosing $\alpha = 0.5$ turns out to be a good choice for a large range of values for R.

It is now shown how the norm D_l is calculated. The *angular distance* between two projections $P_l \in \Lambda$ and $P_q \in \Gamma$ is given by:

$$d_{lq} = \text{MIN}(|\theta_l - \theta_q|, 360^\circ - |\theta_l - \theta_q|) \ , \tag{4.23}$$

and the corresponding *repulsive force* by:

$$r_{lq} = |180^\circ - d_{lq}| \ . \tag{4.24}$$

The repulsive force thus linearly decays with increasing angular distance. It is noted that these definitions differ somewhat from the ones given in [17]. In the original presentation of the WDS, the projection indices l and q are used to calculate the repulsive force instead of the angles θ_l and θ_q, which can only be done when the angular interval is constant. Furthermore, projection angles are limited to the interval 0° and 180°, implying that the projections P_θ and $P_{\theta+180^\circ}$ are equal. Although this is true for parallel-beam CT, in other geometries this is not correct due to the fan angle of the beam. Therefore, in the derivation presented in this work it was assumed that projections that are 180° apart are minimally correlated. Although this assumption is not entirely valid (neither is the assumption that they are maximally correlated), the derived accessing scheme yields satisfactory results. However, if one wants to exploit the fact that P_θ and $P_{\theta+180^\circ}$ are maximally correlated instead (because they are equal or at least very similar), equation (4.24) simply needs to be replaced by:

$$r_{lq} = |90^\circ - d_{lq}| \ . \tag{4.25}$$

The projection P_l experiences a weighted repulsive force μ_l exerted by the projections $P_q \in \Gamma$:

$$\mu_l = \frac{\sum\limits_{P_q \in \Gamma} w_{P_q} \, r_{lq}}{\sum\limits_{P_q \in \Gamma} w_{P_q}} \ , \tag{4.26}$$

where the weight w_{P_q} depends on the influence of the previously accessed projection P_q on the projection P_l, which decreases linearly over time. When the projections in the queue Γ are ordered from old (index 1) to new (index Q) and

projection P_q has the index I_{P_q} in this reordered queue, then this weight is given by:

$$w_{P_q} = \frac{I_{P_q}}{Q} \; . \tag{4.27}$$

The value μ_l is now normalised to the interval $[0, 1]$:

$$\tilde{\mu}_l = \frac{\mu_l - \text{MIN}_{1 \leq k \leq L}(\mu_k)}{\text{MAX}_{1 \leq k \leq L}(\mu_k) - \text{MIN}_{1 \leq k \leq L}(\mu_k)} \; . \tag{4.28}$$

Using only the distance criterion, a newly selected projection can maximize the distance to some projections in Γ, while being close to others. This leads to a situation where projections are selected from one of several clusters in a cyclic fashion, resulting in a situation that clearly does not satisfy the requirements for an improved scheme. To eliminate the large distance fluctuations that give rise to this behaviour, the second term in equation (4.22) is added, which represents the weighted standard deviation of the angular distances:

$$\sigma_l = \sqrt{\frac{\sum\limits_{P_q \in \Gamma} w_{P_q} \cdot (d_{lq} - \bar{d}_l)^2}{\sum\limits_{P_q \in \Gamma} w_{P_q}}} \; , \tag{4.29}$$

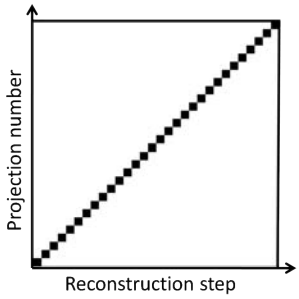
with:

$$\bar{d}_l = \frac{1}{Q} \sum_{P_q \in \Gamma} d_{lq} \; . \tag{4.30}$$

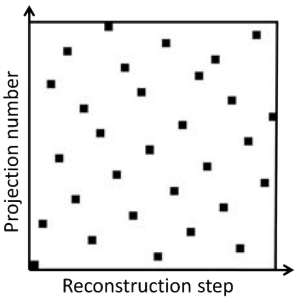
Maintaining a small value of σ_l prevents projections from clustering into groups of angular ranges. Normalization of σ_l to $\tilde{\sigma}_l$ is done as in equation (4.28).

In figure 4.2, the access order pattern is illustrated for a set of $R = 30$ projections. The WDS provides a very uniform sampling of the projections without causing any clustering into angular groups. In figure 4.3[1], results are shown for the reconstruction of a phantom using both sequential access and the WDS, using a single iteration. The sequential access reconstruction is very distorted due to the bias introduced by inadequate angular sampling, since at any time during the reconstruction only a limited angular range of projection data is used to correct the solution. Application of the WDS on the same projection data does not suffer from this artefact and results in images of much higher quality.

[1] In this chapter, all reconstructed images of the modified Shepp-Logan phantom (appendix A) are displayed using a grey scale of [0.00-0.04], meaning that a reconstructed linear attenuation coefficient of 0.00 mm^{-1} corresponds to black and 0.04 mm^{-1} to white.



(a) Sequential



(b) WDS

Figure 4.2: The projection access order for $R = 30$ projections, according the sequential scheme (a) and the weighted distance scheme (b).

(a) Sequential

(b) WDS

Figure 4.3: The impact of the projection access order on the reconstruction quality.

4.5 Voxel basis

The original object function $f(x, y, z)$ is a continuous function, meaning it has a well-defined value at every point (x, y, z) in the 3D space. Due to obvious practical limitations, this function is represented on a discrete grid of N voxels. The coordinates of the j-th grid point, which represents the discrete function value f_j, is denoted by $\vec{x}_j = (x_j, y_j, z_j)$. A uniformly distributed, Cartesian grid is used and the grid points are sampled at an interval Δ. Function values at positions other than the grid points can be retrieved by using a process called *interpolation*, which depends on the *basis function* b that is chosen to represent the volume. The

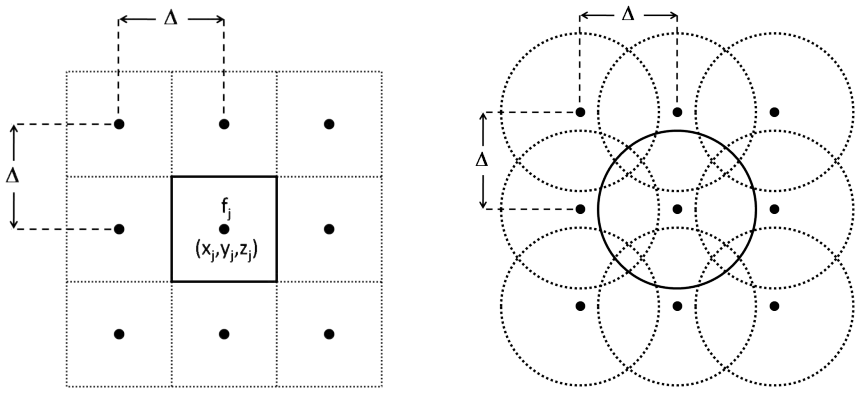
function value at a random location can now be obtained by:

$$f(x,y,z) = \sum_{j=1}^{N} f_j \; b(x-x_j, y-y_j, z-z_j) \; . \tag{4.31}$$

4.5.1 Cubic voxels

The simplest basis function, which was already used in the early days of ART, is to use cubic voxels (figure 4.4(a)). In this representation, the object function is constant and equal to f_j in the cubic volume centred at the grid point (x_j, y_j, z_j). The function value $f(x,y,z)$ is then equal to the value f_j of the grid point nearest to (x,y,z). Mathematically, this can be described by:

$$b(x,y,z) = \begin{cases} 1 & \text{if } |x|,|y|,|z| \leq \Delta/2 \\ 0 & \text{otherwise .} \end{cases} \tag{4.32}$$



(a) Cubic basis (b) Spherical basis

Figure 4.4: 2D Representation of the object function on a discrete grid with interval Δ, for both cubic (a) and spherical (b) basis functions.

4.5.2 Spherical basis functions

A more advanced representation is provided by using spherically symmetric basis functions [18] (figure 4.4(b)), for which the function value $b(r)$ only depends on the distance r from the centre of the sphere. The continuous object function can then be retrieved by:

$$f(x,y,z) = \sum_{j=1}^{N} f_j \; b(r_j) \; , \tag{4.33}$$

where

$$r_j = \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2}\,, \tag{4.34}$$

is the distance from the point (x, y, z) to the centre (x_j, y_j, z_j) of the j-th grid point. In order to cover the entire volume, the basis functions need to overlap in space. Using spherical basis functions results in a continuous representation of the object function which, unlike the cubic function, does not contain straight edges, sharp transitions or any other artificial structures.

An interesting family of spherically symmetric basis functions is given by the generalized Kaiser-Bessel window function [19]. These functions have a convenient analytical form and their smoothness properties can be controlled by two parameters: m (nonnegative integer) is the order of the modified Bessel functions and α (real number) is the taper parameter which determines the degree of localization of the Fourier transform of the window function. A third parameter a determines the extent of the function. The basis function is then given by the following expression:

$$b^{(m,\alpha,a)}(r) = \begin{cases} \dfrac{\left(\sqrt{1-(r/a)^2}\right)^m \cdot\ I_m\left(\alpha\sqrt{1-(r/a)^2}\right)}{I_m(\alpha)} & \text{if } 0 \le r \le a \\ \\ 0 & \text{otherwise}\,, \end{cases} \tag{4.35}$$

where I_m is the modified Bessel function of order m.

An important quality of the Kaiser-Bessel functions is that the X-ray projection through a spherical element, calculated by integrating the function values along the intersection of the ray and the sphere, has an analytical expression. For a spherically symmetric element, this X-ray projection only depends on the distance s between the ray and the centre of the sphere and is given by the Abel transform:

$$p(s) = 2\int_0^{\sqrt{a^2-s^2}} b(\sqrt{r^2+s^2})\,dr\,, \tag{4.36}$$

with a the radius of the sphere. For the generalized Kaiser-Bessel functions, this becomes [19]:

$$p^{(m,\alpha,a)}(s) \quad = \quad \frac{a}{I_m(\alpha)}\sqrt{\frac{2\pi}{\alpha}}\left(\sqrt{1-(s/a)^2}\right)^{m+1/2} \cdot I_{m+1/2}\left(\alpha\sqrt{1-(s/a)^2}\right) \tag{4.37}$$

Each ray is determined by two points: the source location $\vec{x}_s$ and the location of the corresponding detector pixel $\vec{x}_d$. The distance s between a ray and the centre of the sphere located at the grid point $\vec{x}$ can then be calculated by:

$$s = \frac{|(\vec{x} - \vec{x}_d) \times (\vec{x} - \vec{x}_s)|}{|\vec{x}_d - \vec{x}_s|}\,, \tag{4.38}$$

where $\times$ denotes the vector product.

The parameters m, α, and a should be chosen with care in order to accurately describe the imaging process [7]. Setting the order m to 2 results in a continuous derivative at the border of the spherical volume and everywhere inside it. The extent a is chosen to be 2Δ. Parameter α can be calculated to be equal to 10.8 for the 2D reconstruction and to 10.4 in 3D [7]. The profile of the generalized Kaiser-Bessel basis function for these values is illustrated in figure 4.5.

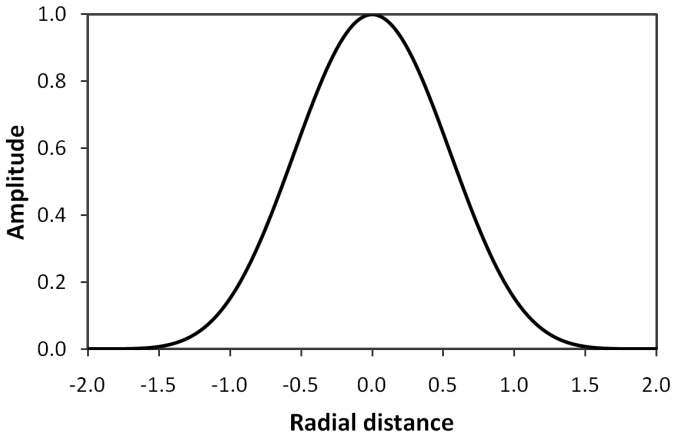


Figure 4.5: Illustration of the generalized Kaiser-Bessel basis function. The amplitude $b^{(m,\alpha,a)}(r)$ is plotted against the radial distance r for the parameters $m = 2$, $\alpha = 10.4$ and $a = 2$.

4.5.3 Linear interpolation

Another approach is to use linear interpolation to retrieve the function values $f(x, y, z)$. In 1D, the linearly interpolated value $f(x)$ is obtained by finding the largest point x_j smaller than x and the smallest point x_{j+1} which is larger than x. With these two points, $f(x)$ is retrieved by using the equation of the straight line through the two points:

$$f(x) = f(x_j) + \frac{x - x_j}{x_{j+1} - x_j}(f(x_{j+1}) - f(x_j)) \,. \tag{4.39}$$

Linear interpolation for multiple dimensions[2] is simply achieved by successively applying 1D linear interpolation in all dimensions. The basis function is now given by:

$$b(x, y, z) = b^{LI}(x)\, b^{LI}(y)\, b^{LI}(z) \,, \tag{4.40}$$

[2]Linear interpolation is often referred to as bi-linear and tri-linear interpolation in 2D and 3D, respectively.

with

$$b^{LI}(\alpha) = \begin{cases} 1 - \frac{|\alpha|}{\Delta} & \text{if} \quad |\alpha| < \Delta \\ 0 & \text{if} \quad \Delta \leq |\alpha| \end{cases} \tag{4.41}$$

4.5.4 Cubic interpolation

Instead of linear interpolation, one can also use cubic interpolation which uses four adjacent function values to determine the intermediate value. Using cubic B-splines [20], the basis function is represented as:

$$b(x, y, z) = b^{CI}(x)\, b^{CI}(y)\, b^{CI}(z)\,, \tag{4.42}$$

with

$$b^{CI}(\alpha) = \begin{cases} \frac{2}{3} - \frac{1}{2}\left(\frac{|\alpha|}{\Delta}\right)^2 \left(2 - \frac{|\alpha|}{\Delta}\right) & \text{if} \quad |\alpha| < \Delta \\ \frac{1}{6}\left(2 - \frac{|\alpha|}{\Delta}\right)^3 & \text{if} \quad \Delta \leq |\alpha| < 2\Delta \\ 0 & \text{if} \quad 2\Delta \leq |\alpha| \end{cases} \tag{4.43}$$

4.5.5 Cylindrical grid

Alternatively, instead of using a Cartesian (cubic) grid, one can also use a cylindrical grid on which the discrete object function is represented [21], similar to what was introduced in filtered backprojection to speed up the reconstruction. All previously presented methods for retrieving the function value at a random point can also be applied on a cylindrical grid, except for the cubic voxels which are now replaced by cylindrical voxels (see figure 4.6). The possible benefits of employing a cylindrical grid are illustrated later.

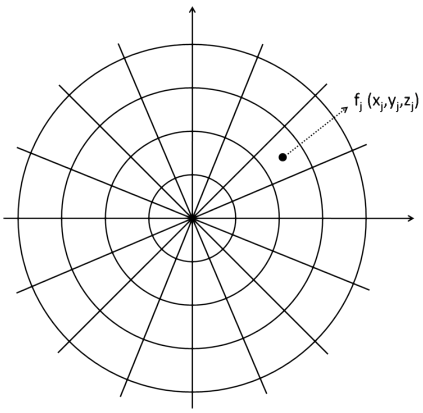


Figure 4.6: Representation of the discrete object function values on a cylindrical grid (polar in 2D). The values f_j now correspond to the central points $\vec{x}_j$ of the cylindrical voxels.

4.6 Forward and backprojection

One of the most important and time-consuming steps in any iterative reconstruction method is the calculation of the weights w_{ij} introduced in equation (4.1). These determine the contribution of the j-th voxel to the i-th ray, which refers to the value corresponding to the i-th detector pixel (be it the ray sum p_i or the expected number of photons t_i). As was discussed in the previous section, there exist many different ways of representing the object function by a set of discrete function values (expressed in terms of the basis functions), each of which requires a different calculation of the weights. The schemes that are most commonly encountered in research on iterative reconstruction algorithms will be briefly presented.

4.6.1 Splatting

The technique of *splatting* was first introduced for applications in volume rendering [22] and consists of determining the analytical footprint of a voxel on the detector. An extensive demonstration of its applicability to iterative reconstruction is given in [7]. Although originally it was presented for spherical elements, it can be applied to general volume elements. Splatting can be implemented using two different approaches: *ray driven* and *voxel driven*.

Ray driven

In the ray driven approach, the ray sum for each detector pixel is accumulated separately. For each ray p_i, the contribution $w_{ij}f_j$ of the j-th voxel to the i-th

ray sum is calculated by integrating the function value f_j over the line segment corresponding to the intersection between the ray and that voxel[3]. The ray sum is then simply obtained by intersecting the ray with each voxel and adding all contributions. Using the cubic voxels, in which $f(x, y, z)$ is constant, the weights are given by:

$$w_{ij} = l \,, \tag{4.44}$$

in which l is the intersection length (figure 4.7(a)). For spherically symmetric basis functions, the integrated value $p(s)$ only depends on the distance s between the ray and the centre of the voxel (figure 4.7(d)). Using the Kaiser-Bessel functions, this integral has an analytical expression and the weight can be expressed as:

$$w_{ij} = p^{(m,\alpha,a)}(s) \,, \tag{4.45}$$

with $p^{(m,\alpha,a)}(s)$ given by equation (4.37). For increased performance, these integrated values are usually calculated and stored in a look-up table before the reconstruction. Improved quality can be obtained by averaging the ray sums for several positions within each detector pixel, instead of using only the centre of the pixel (figure 4.7(b)). Using the cubic voxels, one could consider the rays as being of finite extent, such that the weights are proportional to the intersecting volume of the ray and the voxel (figure 4.7(c)).

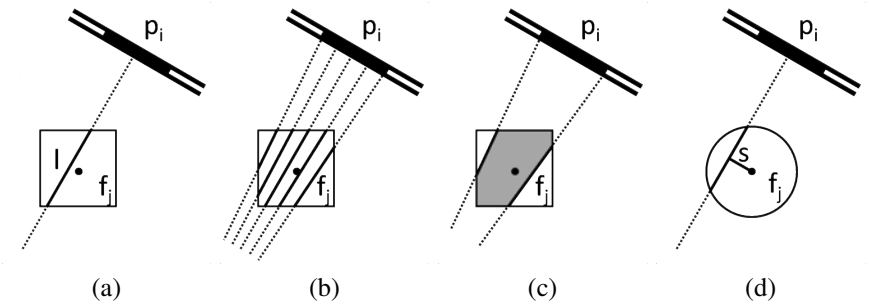


Figure 4.7: Ray driven splatting for cubic (a-c) and spherical (d) volume elements. In (a) and (d), only the ray corresponding to the centre of the detector pixel is used to determine the weights whereas in (b) the average is taken over several rays sampled at equidistant points in the pixel. In (c), the weight is proportional to the intersecting volume between the voxel and a ray of finite extent.

Voxel driven

The voxel driven approach maps the footprint of every voxel onto the detector and accumulates these to construct the projection image. For each voxel, the inter-

[3]The weight w_{ij} can then simply be obtained by dividing this contribution by f_j.

section with every ray is calculated and the contribution $w_{ij} f_j$ is calculated by integrating the function value f_j over this line segment, as was shown for the ray driven approach (figure 4.8). It is noted that both approaches will result in exactly the same weights, which are determined ray per ray for the ray driven approach and voxel per voxel for the voxel driven approach.

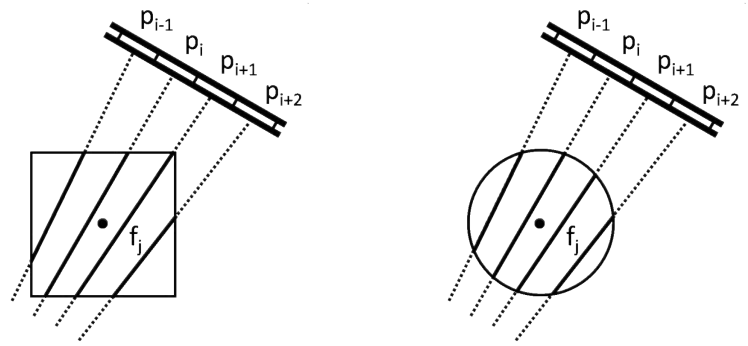


Figure 4.8: Voxel driven splatting for cubic (a) and spherical (b) volume elements.

4.6.2 Ray sampling

Another approach for determining the weights and calculating the ray sums is achieved by sampling a number of equidistant points along each ray. At each sampling point (x, y, z), the function value $f(x, y, z)$ is obtained by interpolation. This value is then multiplied by the distance h between successive sampling points and added to the ray sum. The weights w_{ij} are thus obtained by calculating, for each sampling point, the contribution of every voxel to that point and multiplying it by h. The quality of the ray sampling approach obviously depends on the sampling distance s and improves as s decreases. The oversampling factor S can be used to indicate the quality of the sampling and is given by:

$$S = \frac{\Delta}{s} , \tag{4.46}$$

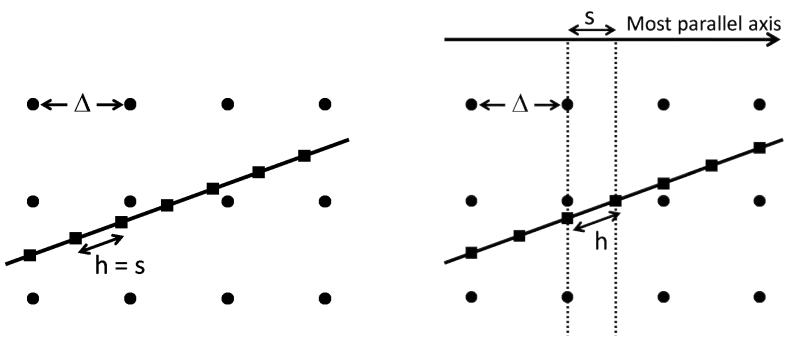
where Δ is again the interval of the grid points. The sampling of the points can be achieved in several ways, of which the two most common methods are now presented.

Simple method

The simplest way of determining the sampling points is by choosing a fixed interval $h = s$. Using this method, the number of sampling points depends on the orientation of the ray and hence for a conical beam this number varies from ray to ray.

Joseph's method

Using the sampling method presented by Joseph [23], first it is determined which of the three main axes is most parallel to the central ray of the projection that is currently being processed. For each ray, the sampling interval is chosen such that its projection onto the most parallel axis is of length s. This implies that the number of sampling points is always equal for every ray while the interval h varies according to the ray's orientation. It was shown in [24] that Joseph's method requires a much lower oversampling factor than the simple method in order to obtain similar accuracy and as such provides higher performance. In a practical implementation, it is convenient to let the oversampling factor S be an integer value so that every S-th sampling point can be chosen to coincide with the grid lines perpendicular to the most parallel axis, as illustrated in figure 4.9.



(a) Simple method (b) Joseph's method

Figure 4.9: Forward projection through ray sampling, illustrated for an oversampling factor $S = 2$.

4.6.3 Other

In addition to the methods presented above, many other approaches exist to calculate the weights and the forward projection.

Siddon's method

Siddon [25] presented an efficient method for calculating the intersection lengths between a ray and a grid of cubic voxels, which can be used in the implementation of the ray driven splatting using cubic voxels. This method considers the voxels as the intersection volumes of three orthogonal sets of equally spaced, parallel planes, rather than as independent elements. Using the parametric equation of the ray, the intersections between the ray and each of these sets of planes can then be calculated separately in a very efficient way. The three sets of intersection points are then merged and reordered, after which the intersection lengths can be easily retrieved.

Köhler's method

The method presented by Köhler *et al.* [26] first determines the function values at the two intersection points between the ray and a cubic voxel using bi-linear interpolation. Next, the function value at the point halfway between the two intersection points is retrieved using tri-linear interpolation. These three function values are then combined using Simpson's rule and added to the ray sum.

Distance-driven method

In the distance-driven method proposed by De Man and Basu [27] every voxel and every detector pixel are mapped on a common plane. The weight w_{ij} is then obtained by calculating the overlapping area between the projection of the j-th voxel and the i-th detector pixel, and applying a proper weighting.

4.6.4 Fast backprojection

When using an iterative method where the update of the volume is based on at least one entire projection, the performance of the backprojection step in terms of speed can be increased significantly using interpolation, as will be explained for SART in the following.

After all ray sums for a certain projection image in the current subset are obtained, a correction image is calculated containing the values c_i for each detector pixel, given by:

$$c_i = \frac{p_i - \sum_{n=1}^{N} w_{in} f_n}{\sum_{n=1}^{N} w_{in}} \,. \tag{4.47}$$

For a given grid point (voxel) f_j, the update corresponding to one projection image P_θ is then obtained by projecting the grid point onto the detector plane and determining the value $\tilde{c}_{(f_j,P_\theta)}$ in the projection point from the correction values c_i using interpolation. The backprojection, and thus the update step, is then performed by adding the average of the interpolated correction values for every projection in the subset multiplied by λ:

$$f_j^{(k+1)} = f_j^{(k)} + \frac{\lambda}{S} \sum_{P_\theta \in OS_s} \tilde{c}_{(f_j,P_\theta)} \ , \tag{4.48}$$

with S the number of projections in the subset OS_s.

4.6.5 Performance

As was shown, several forward and backprojection methods exist, each of which performs differently in terms of quality and speed. To obtain an efficient implementation of an iterative reconstruction algorithm, it is thus imperative to carefully select the projection method.

In figure 4.10 and table 4.1, results are shown and compared for several projection methods. The *normalised root mean squared error* (NRMSE) is defined by [13]:

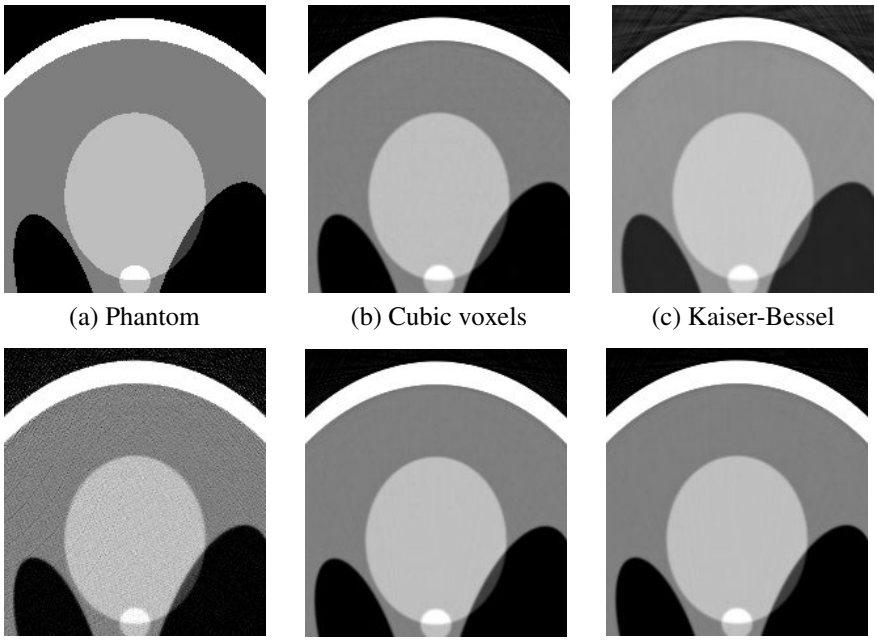
$$NRMSE = \sqrt{\frac{\sum_{j=1}^{N}(o_j - f_j)^2}{\sum_{j=1}^{N}(o_j - \bar{o})^2}} \ , \tag{4.49}$$

where o_j is the object function value in the j-th voxel of the original phantom and $\bar{o}$ is the average of these values. The NRMSE measures a combination of the noise and the bias of the reconstruction as compared to the phantom and is used to quantify the error of the reconstruction.

Splatting was performed using the voxel driven approach. Joseph's method was used for the ray sampling and different values for the sampling interval s were evaluated. From these tests, the optimal value for s was found to be equal to the grid interval Δ. No increase in quality was observed when a smaller interval and thus a more accurate sampling was chosen. The ray sampling method was also tested using cubic interpolation and was found to be comparable in quality to the linear interpolation, while decreasing reconstruction speed.

The different methods were evaluated based on the reconstruction of a single cross-section (in fan-beam) of 512 $\times$ 512 voxels, reconstructed from 512 projections of 512 $\times$ 512 detector pixels (of which only one row is used). From now on, a reconstruction of these dimensions will be referred to as a 512^4 volume. The

geometry for this simulated scan can be described by a pixel pitch[4] of 0.5 mm, a source to object distance of 50 mm and a source to detector distance of 1000 mm, resulting in a full cone angle of 14.6°. The timings were obtained using only one processor of the workstation (table 3.1). It is noted that, although a certain amount of time was spent on optimizing each of these different methods, their timings can probably be improved by spending more time on developing a more efficient implementation. However, the presented timings should provide some indication regarding the computational efficiency of each method.



(a) Phantom (b) Cubic voxels (c) Kaiser-Bessel

(d) Nearest neighbour (e) Linear (f) GPU-implementation

Figure 4.10: Comparison of reconstructions obtained using different projection methods: original phantom (a), voxel driven splatting using cubic basis functions (b), voxel driven splatting using Kaiser-Bessel window ($m = 2$, $\alpha = 10.8$ and $a = 2$) (c), ray sampling with nearest neighbour interpolation (d), ray sampling with linear interpolation (e) and finally the GPU-based implementation (f), which uses ray sampling with linear interpolation and the fast backprojection.

For increased performance, the weights were calculated and stored during the forward projection, to avoid having to calculate them a second time during the backprojection. Although this approach increases speed, storage of these weights requires a large amount of memory, which becomes very impractical during volume reconstruction. By using symmetries of the circular trajectory, the calculation and storage of the weight matrix can be reduced significantly, up to a factor of

[4]the *pixel pitch* is the distance between the centre of two adjacent pixels

Table 4.1: Reconstruction timings and the normalised root mean squared error for different projection methods. The splatting methods were performed voxel driven. The ray sampling uses Joseph's method, with the sampling interval s equal to the grid interval Δ.

Method	time (sec.)	NRMSE
Splatting, cubic voxels	73	0.1839
Splatting, Kaiser-Bessel ($m = 2$, $\alpha = 10.8$, $a = 2$)	52	0.1944
Ray sampling, nearest neighbour	9	0.2259
Ray sampling, linear	26	0.1805
Ray sampling, linear, fast backprojection	5.8	0.1805
GPU-implementation: ray sampling, linear, fast backprojection	0.410	0.1805

8 [28]. As was mentioned before, the object function can also be represented on a cylindrical grid. When the projection data are sampled at equiangular intervals, the cylindrical symmetry can then be exploited optimally and the weights only need to be calculated and stored for one projection angle [21]. Using this approach, the processing time for splatting using the Kaiser-Bessel window was reduced to merely 9 seconds, as compared to the 52 seconds for the Cartesian reconstruction. However, even then the memory requirements for a volume reconstruction are not feasible. Consider a volume consisting of $N = n^3$ grid points, with n the number of points in one dimension. The number of relevant grid points contained inside the cylindrical reconstruction volume is then roughly $0.8\,n^3$. For a square footprint of extent 4, an average of 16 rays will pass through each grid point, thus for every projection about $13\,n^3$ weights need to be stored. For $n = 512$, which still represents only a rather small data set, this requires an amount of 7 GB of storage (in floating point precision) per projection image.

From these presented results, it follows that the ray sampling method using linear interpolation matches the image quality of the more advanced projection methods, while providing higher reconstruction speed. Furthermore, it was found that the fast backprojection method does not result in any apparent reduction in image quality. As the weights can be calculated on-the-fly during the forward projection and no longer need to be recalculated or stored for the backprojection, the fast backprojection method significantly increases performance.

The combination of the linearly interpolated ray sampling with the fast backprojection also provides a very intuitive solution, which is closely related to the actual physics of the imaging process. In the forward projection, rays emanating from the source are traced through the object, accumulating the attenuation values in the object volume along their path, until they hit the detector plane. These simulated, integrated attenuation values are then compared to the measured values obtained from the projection images and the difference, normalised by the length of each ray, is smeared out over the object during the backprojection. Furthermore, this approach is also very suited for the implementation of additional

physical effects (e.g. beam hardening, refraction, etc.), in order to improve the forward modelling and hence reduce certain artefacts in the reconstruction, as will be illustrated later.

4.7 Convergence

The convergence behaviour of iterative reconstruction algorithms is quite complex and a lot of research is performed in order to investigate these properties, e.g. [29–32]. In a more comprehensive framework regarding iterative methods, SART is found to be a special case of the *Landweber scheme*, which in turn belongs to the more general class of *simultaneous block-iterative* (SimBI) algorithms. Since an extensive discussion of these theoretical proofs of convergence does not fit into the scope of this work, only the most important results will be given.

Under the condition that the relaxation parameter λ satisfies $0 < \lambda \leq 2$ and that the matrix $\mathbf{F}$ containing the reconstructed function values is positive definite, the solution obtained by SART converges to the least-square solution of $\mathbf{W} \cdot \mathbf{F} = \mathbf{P}$ (equation (4.4)), while minimizing the norm $\left\|\mathbf{F} - \mathbf{F}^0\right\|$, where $\mathbf{F}^0$ is the initial solution. Furthermore, the rate of the convergence is linear when λ is kept constant during the entire iterative process.

Despite the proven theoretical convergence however, the solution does not necessarily improve with increasing number of iterations. Due to the fact that the reconstruction is an ill-posed problem, an exact solution that satisfies all ray equations can never be achieved. In combination with computational inaccuracies, the projection error never goes to zero, but converges towards a small, nonzero value, as shown in figure 4.11. The projection error is calculated by averaging the squared values of the correction image c_i corresponding to each detector pixel. With an increasing number of iterations, these errors accumulate and distort the reconstructed image. This means that optimal image quality is obtained after a certain number of iterations, after which quality deteriorates [33, 34]. The problem of apparent divergence of iterative algorithms can be handled by the introduction of regularization methods, which have recently been developed in order to improve reconstruction quality of iterative algorithms, e.g. [35, 36]. The study and application of such methods however is outside the scope of this work.

The convergence problem is illustrated in figure 4.12, where the reconstruction error is given as function of the iteration number, for different values of the relaxation parameter λ. For small λ = 0.1, the error decreases slowly when the number of iterations increases. For λ = 0.3, the error reduces at a faster rate, but after a certain iteration number, it steadily increases, so instead of converging to a stable solution, the quality of the reconstructed image deteriorates. For higher λ = 0.5, this effect is even more apparent and for high λ = 0.7, the optimal solution is already reached after one iteration and using more iterations only increases the reconstruction error. The rate of the convergence strongly depends on the value

chosen for the relaxation parameter λ, in that higher values result in faster convergence. When comparing the absolute values of the reconstruction error at the optimal number of iterations, it seems that using smaller λ, and thus more iterations, results in slightly better quality than when using higher λ, the downside being that reconstruction speed increases linearly with the number of iterations. In case the reconstruction is stopped after a number of iterations smaller than the optimal number, the algorithm has not had sufficient time to converge and the solution looks blurred when compared to the optimal solution. When the number of iterations is larger, the reconstruction accumulates the remaining projection errors resulting in images that seem to contain a lot of noise.

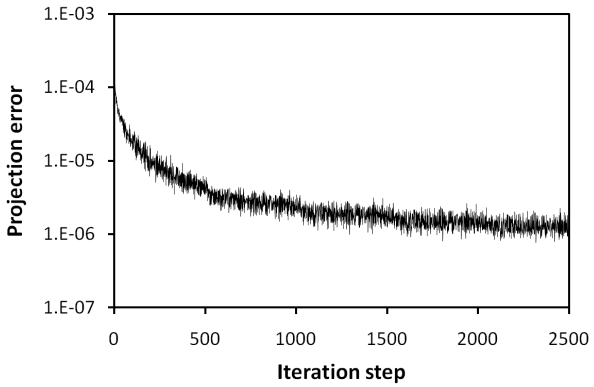


Figure 4.11: The mean projection error as function of the iteration step. The term iteration step refers to the update of the reconstruction volume after using one projection of a full iteration.

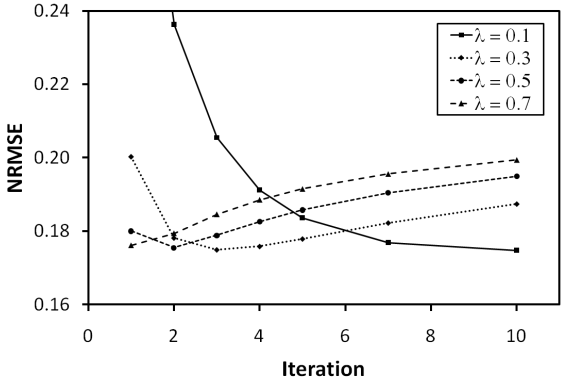


Figure 4.12: Convergence of the reconstruction as function of the number of iterations for different values of λ.

It is noted that results were rather different when the projection images of the phantom were simulated by creating a discrete representation of the phantom and

applying the forward projector of the SART implementation. When these projections were reconstructed, SART always converged for any value of λ between 0 and 1, without increasing the reconstruction error for higher numbers of iterations. This proves that the implementation of SART that was used for these tests is consistent and it indicates that the apparent divergence of the algorithm may be caused by a difference between the analytical projection of the continuous phantom and the algebraic projection of the discrete phantom. In real applications, data is never consistent so one will always experience the divergence of the algorithm. Therefore, it was decided to continue this discussion using the analytical projections (causing divergence), since the conclusions drawn can then also be applied to real data.

From a practical point of view, it is interesting to look at things from a slightly different perspective by considering the convergence as a function of λ while keeping the number of iterations fixed. This evolution is shown in figure 4.13 and is similar to the one in figure 4.12: the reconstruction error decreases with increasing λ to a certain optimal value, after which it steadily increases. This means that, for a given number of iterations, an optimal value of λ can be found which minimizes the reconstruction error. The rate of the convergence and the subsequent increase of the error strongly depends on the value of the relaxation parameter λ. Similar to the previous discussion, choosing a value for λ lower than the optimal value results in blurry looking images, whereas for higher values the reconstructed images become noisy.

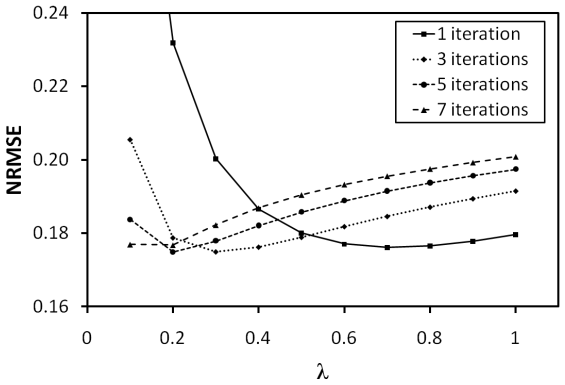
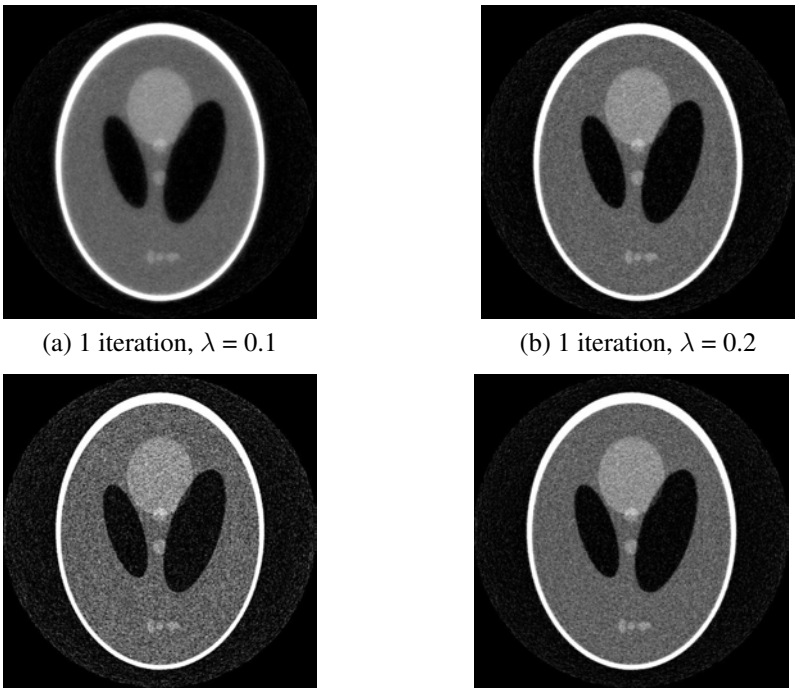


Figure 4.13: Convergence of the reconstruction as function of λ for a different number of iterations.

A major problem regarding the convergence is that for a given number of iterations there is no such thing as a *best* relaxation parameter, which always results in the optimal reconstruction for different projection data. Instead, the optimal value depends on the object, the geometry of the set-up, the noise characteristics and many other factors [16]. This is illustrated by figure 4.14, which shows reconstruction results for the same phantom where noise has now been added to the

projection images. The minimal error now corresponds to $\lambda = 0.2$ for 1 iteration and to $\lambda = 0.1$ for 3 iterations, which is clearly different from the optimal values obtained for the phantom without noise. Although much research is performed on this subject, no general framework yet exists that allows determining the optimal settings for different projection data.



(a) 1 iteration, $\lambda = 0.1$ (b) 1 iteration, $\lambda = 0.2$

(c) 1 iteration, $\lambda = 0.5$ (d) 3 iterations, $\lambda = 0.1$

Figure 4.14: Comparison of reconstructions using a different number of iterations and a different value of the relaxation parameter. Poisson noise was added to the simulated projection data of the phantom to illustrate the trade-off between smoothness and noise.

In most practical applications encountered in high resolution CT, information regarding the internal structure of the scanned object is not available and so the reconstruction error cannot be calculated, which makes it very difficult to select the optimal values of the relaxation parameter and the number of iterations. Quantitatively, one may determine the convergence by measuring the projection error and stopping the reconstruction when this error converges. Alternatively, the user may assess the quality of a single cross-section reconstructed using different settings and determine the optimal parameters visually. Using this approach, with a fixed number of iterations, the user can alter λ to vary between a more blurred or a more noisy reconstruction. In this regard, λ can be seen as a noise parameter and the user has to make a trade-off between smoothness and noise. Furthermore, the number of iterations may be increased in order to improve reconstruction quality,

at the expense of increased reconstruction time.

Regarding the application of iterative algorithms in high resolution CT presented in this work, the number of iterations was usually chosen rather small, from 1 to 3, as no apparent improvements in quality could be obtained using more iterations. The relaxation parameter λ is typically set between 0.3 and 0.7, as for most applications 0.5 seems to provide the best images. For the remainder of this work, when no values are mentioned a single iteration with $\lambda = 0.5$ is used to perform the reconstruction.

4.8 Initialization

In principle, the initial solution of an iterative algorithm can be chosen arbitrarily. One could attempt to increase convergence by initializing the volume with the result of a reconstruction obtained by another, faster, reconstruction method, such as the filtered backprojection. However, the features from the initial image will be retained for a long time during the iterative process and it may become very difficult to remove artefacts that were introduced by the initial reconstruction. Furthermore, as it is the intention to apply an iterative method to increase reconstruction quality, it seems irrational to start with a less accurate method which may have already generated certain errors.

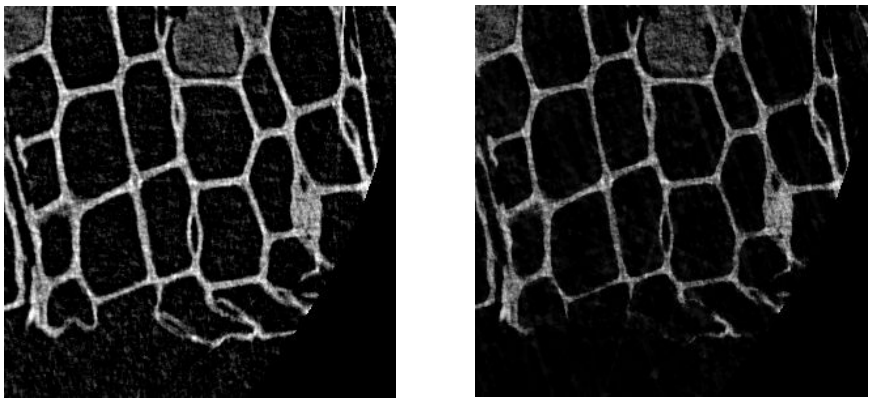
It seems that the only rational starting point is to use a uniform distribution of the object function [37]. A plausible initial value can be obtained by determining the average linear attenuation coefficient of the entire scanned region from all projection images. However, after evaluating reconstructions using different initial values, no increase in image quality was found (figure 4.15). The only observation that was made is that quality may decrease when the initial value is chosen too high. Therefore, the best initial value for the entire reconstruction volume seems to be zero.

(a) 0 (b) 0,0115 (c) 1

Figure 4.15: Reconstruction results for uniform initialization with different values: 0 (a), 1 (c) and the average linear attenuation value 0.0115 obtained from all projection images (b).

4.9 Constraints

One of the properties of iterative methods is that constraints can be introduced to limit the amount of possible solutions. For conventional X-ray CT, a simple constraint is that every linear attenuation coefficient has to be positive, as a negative value is physically not possible. This condition can be implemented directly in the backprojection step of the iterative method by a simple truncation. The application of this constraint results in a reduction of noise in the reconstructed image, especially in the background. However, it was also found that this constraint may cause certain small features to disappear, which is obviously unacceptable. This is illustrated in figure 4.16, where one can clearly see how several features in the reconstructed cross-section can no longer be observed when applying the nonnegative constraint. This artefact was observed in the reconstruction of various objects and it seems that small and thin features near the edge of the object are most susceptible to this effect.



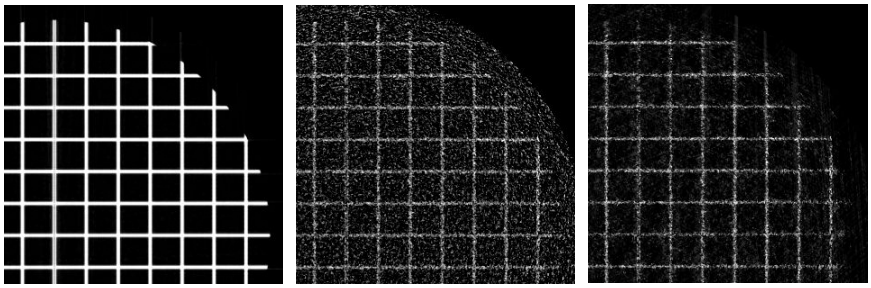
(a) No constraints (b) Non-negative

Figure 4.16: Reconstructions of a piece of wood without constraints (a) and with the condition that the object function values can never become negative (b).

Currently, it is still unclear which mechanism produces these artefacts. A possible cause is that due to the contribution of noise, phase contrast or some other effect in the imaging process, at some points in the projection image the measured intensity is higher than it should be. At the edge of the object, where the attenuation is rather low due to the short path length, this may cause the normalized measured intensity to become larger than 1. Taking the logarithm of this intensity then yields a negative ray sum for the corresponding ray path. Without constraints, the reconstructed value for certain voxels can become negative, even though this is physically impossible. When calculating a ray sum, negative background voxels can compensate for the positive value of other voxels, making it possible for the total ray sum to become negative and hence comply with the measured negative

ray sums. When the nonnegative constraint is enforced, the background voxels can no longer be used to obtain negative sums. Therefore, for certain features the reconstructed values iterate to zero, since this is the only way by which the calculated ray sums can comply as best as possible with the negative ray sums. This mechanism is consistent with the fact that typically only features at the edge of the object disappear.

To verify this assumption, projection data was simulated of a square grid consisting of straight lines along the X- and Y-axis. The grid was confined within the circular field of view, such that it contains several protrusions of different sizes and orientations. Noise was added to the simulated images such that the projections yield negative ray sums. The reconstructions of the projections with and without the nonnegative constraint are shown in figure 4.17, the (unconstrained) reconstruction of the noiseless data is also given for comparison. The NRMSE for the reconstruction with the constraint is 0.66, compared to 1.73 for the unconstrained reconstruction. Enforcing the constraint thus seems to provide a better reconstruction quality. This can also be seen in the reconstructed images, where the constrained cross-section suffers much less from noise. However, when looking closely at the edges of the grid, it can be seen that some of the protrusions tend to disappear. Contrary, all protrusions are fully reconstructed in the unconstrained cross-section. Thus, despite the significantly higher noise level and reconstruction error, the unconstrained reconstruction seems to be more reliable.



(a) Noiseless (b) No constraints (b) Non-negative

Figure 4.17: Reconstruction of the simulated projections of a grid (a). After adding noise to the projections, the data were reconstructed without using any constraints (b) and with the condition that the object function values can never become negative (c).

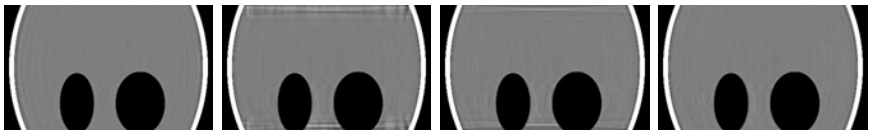
It is noted that, even though the proposed mechanism is plausible and complies with the observations, a more elaborate verification of this effect is advisable. Since the observed artefacts have a considerable influence on the image quality, future research regarding such constraints is indispensable for the further development of iterative reconstruction algorithms.

4.10 Region of interest scan

As with analytical algorithms, the reconstruction of truncated data no longer leads to an exact solution and using the iterative methods as such introduces artefacts in the reconstruction. Since in practical situations truncated data are often encountered, it is imperative to modify the implementation such that the resulting artefacts are removed or at least reduced. As was already explained in the previous chapter, projection data can be truncated in the horizontal and vertical direction. Ideally, one would first scan the entire sample at a lower resolution and use this result to obtain a reconstruction of the region of interest scan which is free of truncation artefacts. However, due to practical concerns taking two scans is often not feasible and thus it is important to investigate the processing of truncated data.

4.10.1 Vertically truncated

In chapter 3, it was already discussed that most analytical reconstruction methods require untruncated data. Both the FDK (for cone-beam) and the Katsevich algorithm (for helical cone-beam) can handle vertically truncated data without introducing any inherent artefacts, which offers them a great advantage over other algorithms. Iterative algorithms on the other hand do suffer from vertically truncated data, as is shown in figure 4.18, where a part of the full reconstruction is compared to the reconstruction of the truncated data. Without any modifications to the algorithm, artefacts can be found at the top and bottom of the reconstruction volume, which are due to the fact that any information regarding the regions outside the measured range is absent.



(a) Full (b) Truncated (c) Extended (d) Infinite

Figure 4.18: Cross-sections along the XZ-plane of the reconstruction of a region of interest scan where the data are vertically truncated. The full reconstruction (from untruncated projections) is shown in (a), as compared to the reconstruction of the truncated data (b). In (c), the reconstruction volume was extended and the projection data were padded (only the fully covered volume is shown). Reconstruction (d) was obtained by assuming that the reconstruction volume and the projection data are vertically homogeneous outside the measured range.

A modification to the iterative algorithms is now explained which allows the reconstruction of vertically truncated data without introducing artefacts. Usually only the reconstruction volume that is fully covered by the projection data for each angle is considered for reconstruction. When using a conical beam, certain rays

will also pass through regions below or above this volume, thereby collecting information outside the fully covered volume. Using the FDK or Katsevich algorithm, this is not a problem as each voxel can be reconstructed independently from the rest of the volume. However, when an iterative method is used, the reconstruction of each voxel is correlated and the entire volume should be reconstructed simultaneously. Therefore, artefacts are introduced when the entire object is not fully covered. A first modification is to extend the reconstruction volume, such that every projection ray always passes through the entire field of view of the detector (figure 4.19). However, when the data are vertically truncated, it is impossible to reconstruct the entire object and even using an extended volume is not sufficient to prevent artefacts (figure 4.18(c)).

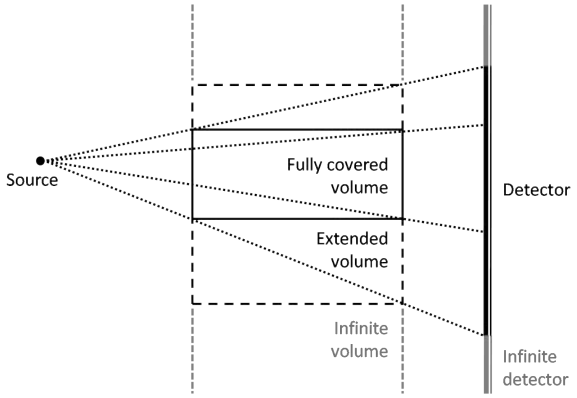


Figure 4.19: For an iterative reconstruction, instead of only reconstructing the volume that is fully covered for each projection angle, the reconstruction volume is extended such that every ray in the projection data always passes through the entire field of view of the detector. Vertically truncated data can then be handled by virtually extending the reconstruction volume and the projection data infinitely along the vertical direction.

Fortunately, it was found that this problem can be resolved by assuming that the volume is infinitely extended and vertically homogeneous outside the extended volume. The cross-sections outside the extended volume are then supposed to be equal to the last cross-section of the extended volume, i.e. the cross-sections above are equal to the top slice of the extended volume and the ones below are equal to the bottom slice. In accordance, the detector is given a similar infinite extent as well. To be exact, for the infinite detector one should take into account the increased path length of the rays when moving away from the horizontal centre of the geometry. However, for most practical geometries this increase is very small and can thus safely be neglected. Results are shown in figure 4.18(d), which clearly show that the presented solution provides reconstructions of vertically truncated data without introducing additional artefacts.

In order to reduce the influence of noise in practical situations, instead of using only the last pixel row, the infinite extent of the detector is determined by aver-

aging a certain number of pixel rows at the edge of the detector. It is noted that in a practical implementation the introduction of the extended volume increases the required amount of memory to store the volume during reconstruction and also increases reconstruction time. Fortunately, the addition of the infinite extension of the volume and projection data is only virtual. It can be introduced by making a minor adaptation to the code, which poses no additional burden towards performance.

4.10.2 Horizontally truncated

Data that are horizontally truncated also lead to artefacts in the reconstruction, typically observed as a cupping effect where the edges of the reconstructed field of view (FOV) appear to be more attenuating than the centre. Different methods were evaluated to reduce these artefacts. The best solution seemed to be an extension of the reconstruction volume outside the field of view of the detector by a certain amount, as illustrated in figure 4.20. Although for this additional volume a decent reconstruction cannot be obtained due to a lack of information, it appears to be sufficient in order to reduce the truncation artefacts. When the reconstruction finishes, the extension is simply removed to obtain the final result. Alternatively, methods based on horizontal padding of the projections (zero, linear, cosine and other padding methods) were also evaluated. Results of these tests were at best similar to the FOV extension, while increasing reconstruction time as a larger number of rays need to be simulated for each projection.

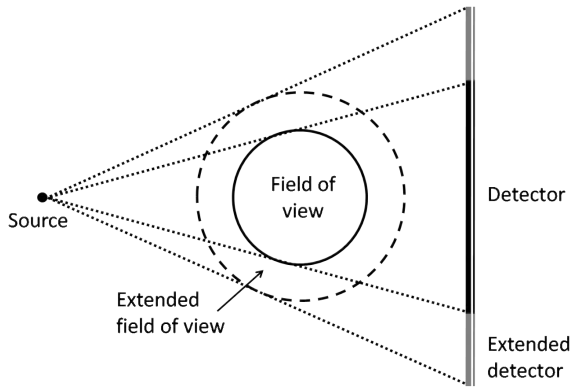
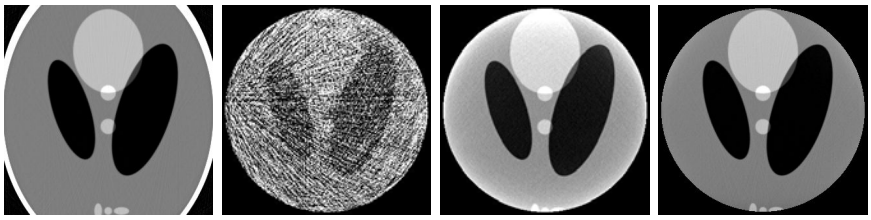


Figure 4.20: The reconstruction volume is extended outside the field of view of the detector by a certain amount, in order to reduce the artefacts due to horizontal truncation of the data.

In figure 4.21, results are shown for the reconstruction of the phantom where 100 pixels were removed at each side of the detector, resulting in projections containing 312 pixel columns. An extension of 20% provided the best result. Smaller values introduced the expected cupping artefact and also generated more streak

artefacts. Higher values led to an inverse cupping artefact (the attenuation values in the centre of the FOV became higher than in the outer regions), indicating over-correction. The same test was also performed for different degrees of truncation. When 50 pixels were removed at each side, results were optimal for an extension of 5%. For a removal of 150 pixels, an extension of 50% provided the best reconstruction. However, in this last case the cupping artefact could never be sufficiently reduced, even for larger extensions, indicating that this type of correction is only valid up to a certain level of truncation.



(a) Full scan (b) Extension 0% (c) Extension 2% (d) Extension 20%

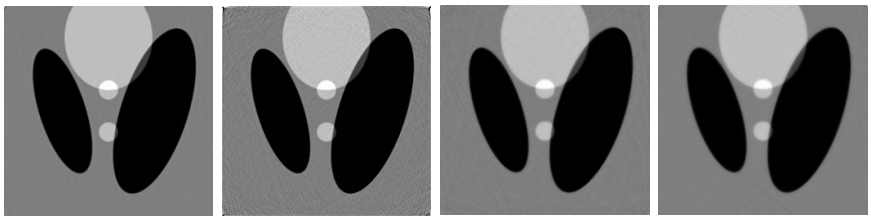
Figure 4.21: Reconstructions of a region of interest scan, for different values of the extension of the FOV: 0% (b), 2% (c) and 20% (d). The reconstruction from the full scan (a) is shown for comparison.

4.11 Region of interest reconstruction

Instead of reconstructing the entire volume, it is often feasible to only reconstruct a region of interest (ROI) of the whole object as this reduces both computation time and memory usage. As was already mentioned earlier, this is not a problem when using an algorithm like FDK or Katsevich, as these allow every voxel to be reconstructed independently from the rest of the volume. When using an iterative method however all voxels are correlated and the problem of a ROI reconstruction is not that straightforward.

An adequate solution to the ROI reconstruction problem for iterative algorithms can be found by the following consideration. In order to reconstruct a certain ROI, one needs to dispose of its contribution to the projection data. The measured data however do not only contain the contribution of the ROI but of the entire object. If the reconstruction of the entire volume is already available, one could simulate the forward projections of the entire volume excluding the ROI and subtract them from the measured projections. The corrected projections thus obtained now only contain the projections of the ROI and can thus be used in the iterative reconstruction of the ROI.

The reconstruction of the entire volume could be obtained by using a faster reconstruction method, such as the filtered backprojection. This approach was already presented in [38] for the reconstruction of a two dimensional cross-section.



(a) Full (b) ROI (c) Padding 8% (d) Padding 20%

Figure 4.22: Region of interest reconstruction where the entire volume was first reconstructed using the same iterative method at half of the original resolution (b-d), compared to the full reconstruction (a). Artefacts can be reduced by expanding the ROI for reconstruction and removing the border afterwards. The number of voxels that was added at each side of the ROI was 20 and 50, corresponding to a padding of 8% (c) and 20% (d) of the number of voxels in one direction of the ROI, respectively.

Application in three dimensions however immediately implies an important decrease in memory requirements as only the ROI needs to be stored during reconstruction instead of the whole volume. However, as was already mentioned earlier, it does not seem rational to use an iterative algorithm to try to improve quality while starting the reconstruction based on another method, which may have already introduced artefacts. Another possibility is to reconstruct the entire volume using the desired iterative method at a low resolution (thus requiring less voxels to cover the entire volume), which requires much less processing time. Results for this last approach are presented in figure 4.22, where the ROI is the central region of 256×256 voxels from a total of 512×512 voxels. The entire volume was reconstructed at half the resolution, so it contains four times (eight times for the 3 case) less voxels than the entire volume at the normal resolution. It can be seen that the ROI reconstruction introduces minor artefacts at the edge of the ROI and increases the number of streaks. These can be reduced by marginally expanding the ROI for the reconstruction and removing the expanded border afterwards. A similar idea of a *multiresolution* reconstruction was also used by Niinimaki *et al.* [39] for 2D local tomography in dental applications. They implemented this idea based on wavelets for a Bayesian inversion method [40] using a prior model of the imaged tissues.

4.12 Memory requirements

It was already noted that in an iterative reconstruction the entire volume needs to be processed simultaneously. To obtain an acceptable reconstruction speed, this requires the entire volume to be kept in memory during reconstruction. In high resolution CT, volumes typically range from 1024 to 2048 voxels per dimension. Storing this in floating point precision (4 bytes) requires about 4 GB ($1024^3 \times 4$

bytes) to 32 GB ($2048^3 \times 4$ bytes) of memory. Although nowadays such amounts of CPU-RAM memory are no longer exceptional, to obtain a decent reconstruction speed it is highly desirable to execute the iterative method on a GPU, which requires an equally high amount of GPU-RAM memory. Currently the highest amount of on-board memory on a single GPU is found on the Nvidia Tesla C1060, which has 4 GB of RAM. It is obvious that this large memory requirement poses a major obstacle in the practical application of iterative reconstruction algorithms in high resolution CT.

A possible solution would be to use the CPU-RAM memory as a buffer and constantly send parts of the volume to the GPU for processing. Although this could by-pass the memory problem, the large amount of data transfer would cause a dramatic decrease in reconstruction speed, thereby eliminating any gain that is achieved by using a GPU.

4.12.1 Multiple GPUs

A more interesting possibility is to use multiple GPUs in the reconstruction, where every GPU stores a part of the entire volume. Since the simulation of the forward projection is a cumulative process, for a given projection angle one can simply calculate the rays-sums for each subvolume separately and add them together afterwards. Once the correction image is calculated in the correction step, the fast backprojection can be performed independently for each voxel. This method, illustrated in figure 4.23, not only allows processing of larger volumes, it also significantly increases reconstruction speed, almost linearly with the number of GPUs used[5]. However, this solution quickly becomes rather expensive and is by no means a flexible one, as additional hardware is required when facing a data set that is still larger than what can be handled at a given moment.

4.12.2 Multiresolution reconstruction

A more generic solution to the memory problem is to find a way to somehow reduce the size of the volume that needs to be kept in memory during reconstruction. A first observation that can be made in this regard is that in cone-beam geometry, the reconstruction volume can be divided in two parts, corresponding to the object volume above and below the horizontal symmetry plane (optical axis) of the scanner set-up. Due to the geometry of the cone-beam, a single X-ray will never travel through both parts of the volume, at least when effects like scattering and refraction are neglected. Therefore, both subvolumes are not correlated and can thus be reconstructed independently, which immediately results in a memory reduction by a factor 2. It is noted however that it is recommended to include some overlap

[5] Assuming all GPUs are of the same type

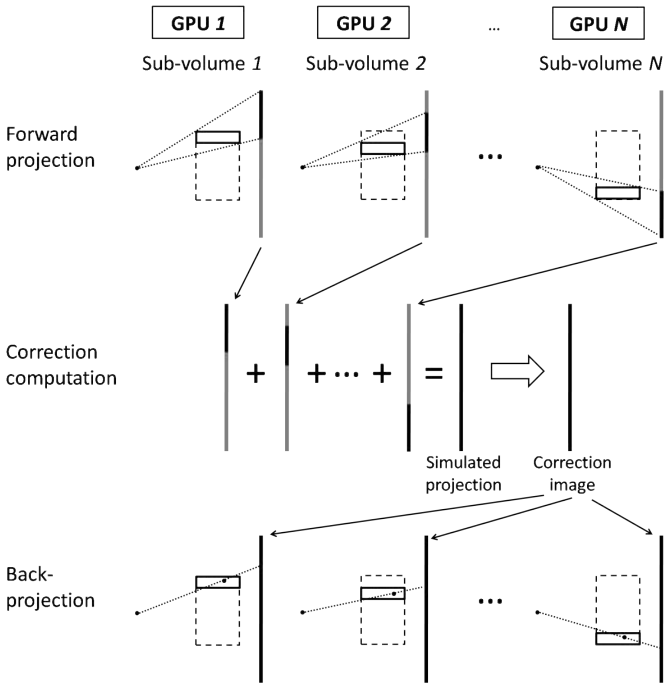


Figure 4.23: Reconstruction scheme for using multiple GPUs in an iterative reconstruction in order to cope with the required memory and to decrease the reconstruction time.

between the two subvolumes, which is removed when the reconstruction finishes, to prevent the introduction of artefacts due to edge effects.

A more comprehensive solution to reduce the required amount of memory is given by the multiresolution approach, presented by De Witte *et al.* in [41], which can be seen as an extension of the method that was discussed for the ROI reconstruction. The idea behind this method is that the iterative reconstruction of the entire volume on a small and rather coarse grid can be used to retrieve the contributions (to the projections) of a number of adjacent ROIs, which can be reconstructed independently on a larger and more detailed grid.

Before proceeding, note the following definitions. *Low resolution* means that the size of the voxels used in the reconstruction is chosen rather large (i.e. a large spacing between the grid points), so the volume can be covered by a relatively low number of voxels. Similar, *high resolution* denotes using small voxels, meaning a high amount of voxels is required to cover the volume.

First, the entire volume is reconstructed at a low resolution. Next, a certain subvolume is selected and represented at high resolution. The contribution of this subvolume to the measured projections can be obtained by calculating the reprojections of the entire low resolution volume excluding the subvolume, and subtracting

these reprojections from the measured projections. By repeating this process using different subvolumes and combining them afterwards, the entire volume can be reconstructed at high resolution while only requiring a fraction of its total memory needs.

The presented method is applicable to any iterative algorithm that requires a forward projection of the reconstructed volume. The processing steps are elucidated by presenting a possible implementation scheme for SART, which is illustrated in figure 4.24:

1) Reconstruct the entire volume at a low resolution and keep it in memory during the entire reconstruction process.

2) Reconstruct a part of the volume at high resolution:

   I) Create and initialize the high resolution subvolume in memory

   II) For every projection:

      i) Calculate the reprojection of the low resolution volume excluding the contribution from the low resolution subvolume.

      ii) Subtract the reprojection from the measured projection, this processed projection now only contains the contribution of the subvolume.

      iii) Reproject the high resolution subvolume and calculate the difference with the processed projection.

      iv) Update the high resolution subvolume based in this difference.

   III) Iterate the previous step for every projection and for a given number of iterations.

3) Store this subvolume on disk.

4) Repeat steps 2 and 3 using different subvolumes until the whole volume is covered and then combine them to obtain the entire reconstruction volume at high resolution.

It is noted that alternative schemes may be considered, which could provide a difference in performance depending on the available hardware. One could for instance develop a scheme that uses temporary dumps to the hard disk or CPU-RAM memory. Another possibility would be to pre-calculate and store every reprojection of the low resolution volume, so these do not need to be calculated again for every subvolume.

The quality of the multiresolution reconstruction is again evaluated using the modified Shepp-Logan phantom with the settings described in section 4.6.5, thus with a size of 512^4 and a full cone angle of 14.6°. Results are compared with the normal reconstruction of the entire volume. The voxel size of the reconstructed

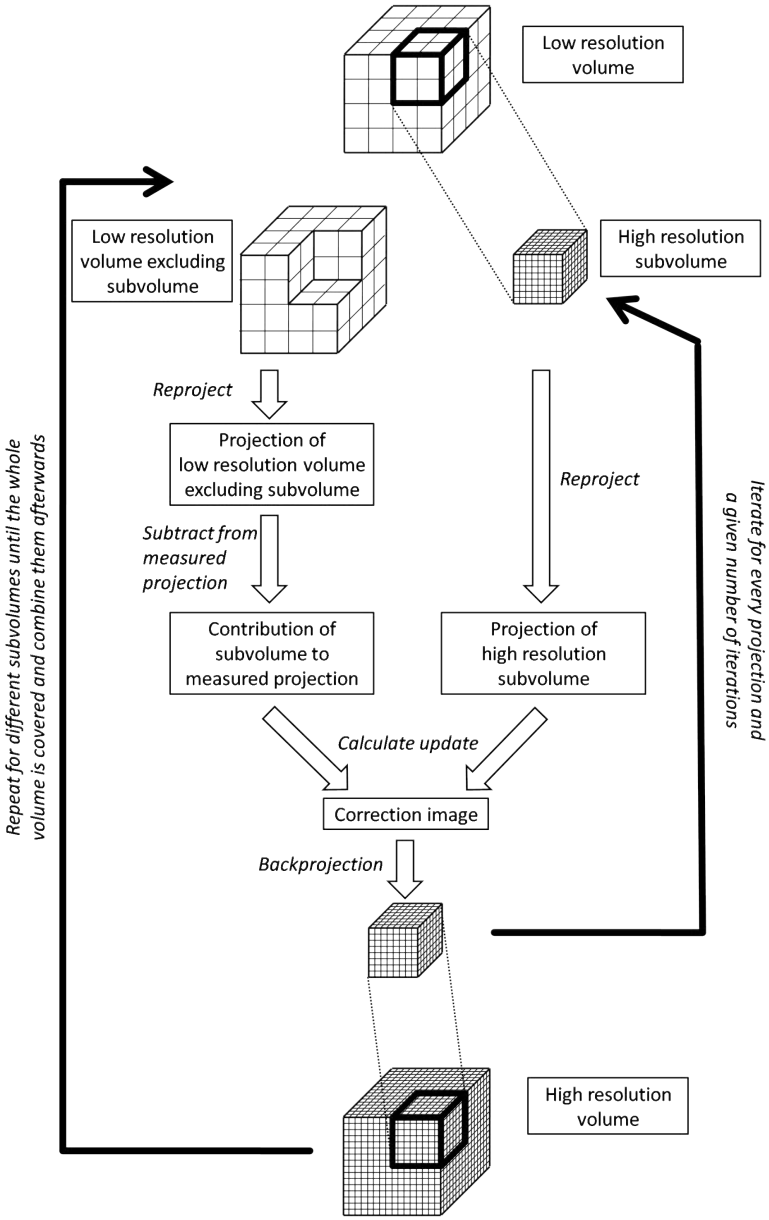
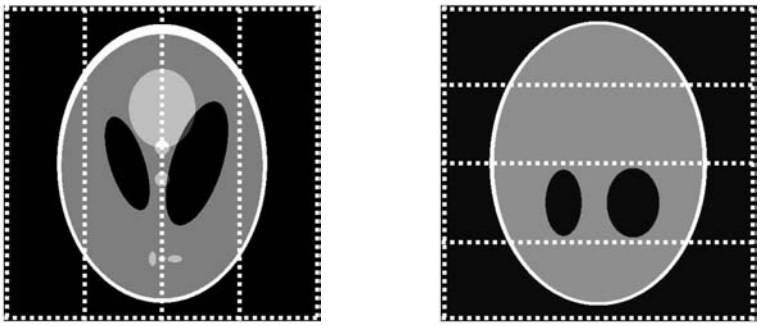


Figure 4.24: Possible implementation scheme for the multiresolution reconstruction.

high resolution volume is 25 µm. For the low resolution volume, the voxel size depends on the factor N_{DS} at which the high resolution volume was downsampled and is given by $N_{DS} \times 25\,\mu m$. For a downsampling factor of 2, the voxel size is thus 50 µm. The number of subvolumes in which the full volume is divided is denoted by N_{SV}. A quantification of the reconstruction quality is again provided by evaluating the normalized root mean squared error (NRMSE), given by equation (4.49). By evaluating the reconstructed images and calculating the NRMSE it was found that high image quality could be obtained using only a single iteration in the reconstruction. Increasing the number of iterations only resulted in a marginal increase of quality, while the reconstruction time increased linear. Therefore the presented results were obtained by using a single iteration in the reconstruction of both the low and high resolution volumes. Using a fixed number of one iteration, the optimal value of the relaxation parameter λ was determined by reconstructing the central slice of the object for different values of λ and selecting the one that yields the lowest NRMSE. This optimal value of λ depends on the experimental conditions and therefore varies for each test.

Splitting along X- or Y-axis

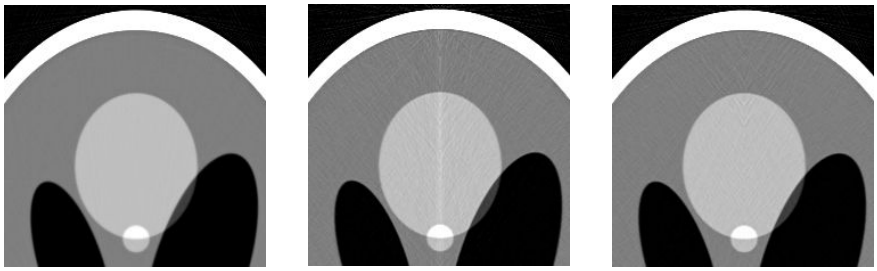
The reconstruction quality is first investigated by dividing the volume into subvolumes by splitting it along the X-axis, thus by dividing it by planes parallel to the YZ-plane, as illustrated in figure 4.25(a). As the XY-plane of the reconstruction volume corresponds to the rotation plane of the object, results will be similar when the volume is split along the Y-axis. The low resolution volume is sampled down with a factor of N_{DS} = 2 per dimension as compared to the full high resolution volume, causing a decrease in memory requirements by a factor of 8.



(a) X-splitting (b) Z-splitting

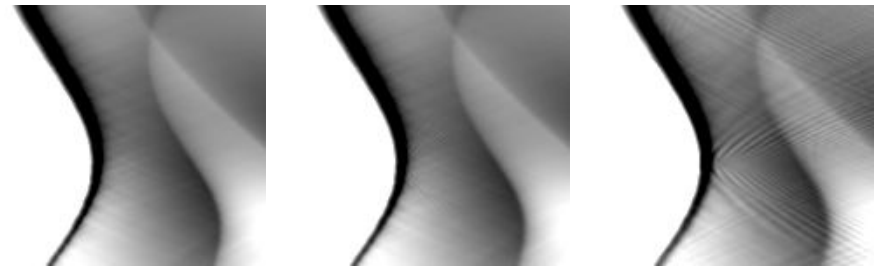
Figure 4.25: Illustration of the division into subvolumes of the reconstruction volume along the X-axis (a) and the Z-axis (b), for $N_{SV} = 4$.

When splitting the volume into two regions, one can see a bright separation line between the two regions (figure 4.26(b)). This can be prevented by marginally



(a) Full (b) No overlap (c) N_{OV} = 5%

Figure 4.26: The full reconstruction of the volume (a) is compared with the multiresolution reconstruction, where the volume is split into two regions along the X-axis, without overlap (b) and with 5% overlap (c), with $N_{DS} = 2$ and $N_{SV} = 2$.



(a) N_{DS} = 1 (b) N_{DS} = 2 (c) N_{DS} = 4

Figure 4.27: Part of the sinograms created by reprojection of the low resolution volume, as function of the downsampling factor N_{DS}, with $N_{OV} = 5\,\%$ and $N_{SV} = 2$.

(a) N_{DS} = 1 (b) N_{DS} = 2 (c) N_{DS} = 4

Figure 4.28: Reconstructed cross-sections as function of the downsampling factor N_{DS}, , with $N_{OV} = 5\,\%$ and $N_{SV} = 2$.

(a) N_{SV} = 2 (b) N_{SV} = 4 (c) N_{SV} = 8

Figure 4.29: Reconstructed cross-sections as function of the number of subvolumes N_{SV} in which the volume was split, with $N_{OV} = 5\,\%$ and $N_{DS} = 2$.

Table 4.2: The NRMSE of the reconstructed cross-sections as function of different settings for splitting along the X-axis: the overlap N_{OV}, the number of subvolumes N_{SV} and the downsampling factor N_{DS}. The optimal relaxation parameter λ is given for completeness. The full reconstruction was performed using λ = 0.8 and yielded a NRMSE of 0.1514. The increase of the NRMSE relative to the one of the full SART reconstruction is given by δ.

$N_{OV}(\%)$	N_{SV}	N_{DS}	λ	NRMSE	$\delta(\%)$
0	2	2	0.6	0.1631	7.7
5	2	2	0.7	0.1569	3.6
10	2	2	0.7	0.1549	2.3
20	2	2	0.8	0.1530	1.0
5	2	2	0.7	0.1569	3.6
5	4	2	0.6	0.1610	6.3
5	8	2	0.5	0.1678	10.8
5	2	2	0.7	0.1569	3.6
5	2	4	0.6	0.1648	8.8
5	2	8	0.5	0.1809	19.4

expanding the subvolumes and by removing their overlap (thus only using the original subvolume) in the final image (figure 4.26(c)). An overlap of N_{OV} = 5% of the entire volume on both sides of the subvolume was found to be sufficient. When compared to the full reconstruction of the phantom (figure 4.26(a)), splitting the volume has resulted in an increase of streaking artefacts inside the object (figure 4.26(c)). These streaks appear because the reprojection of the low resolution volume is only an approximation of the real projection. In figure 4.27, parts of the sinograms are shown that are created by reprojection of the central slice of the low resolution volume, as function of the size of this volume. It is clear that downsizing this volume has an impact on the quality of the reprojection, which in turn has an influence on the reconstruction quality of the high resolution volume. In figure 4.28, the reconstructed cross-sections are compared for different values of

the downsampling. The number of regions in which the volume is split also has an influence on the streaks, as can be seen in figure 4.29. A quantitative analysis of the reconstruction quality is provided by the measurement of the reconstruction error NRMSE for different settings, shown in table 4.2. These errors indicate that the reconstruction quality of the multiresolution method improves with increasing overlap and deteriorates with an increasing number of subvolumes and downsampling factor.

Splitting along Z-axis

The division of the high resolution volume can also be performed by splitting the volume along the Z-axis, which is equal to the rotation axis of the object. This division into subvolumes is illustrated in figure 4.25(b). In figure 4.30, results are shown for a volume that was split into 8 regions, with a downsampling factor of 2. The images shown are cross-sections through the volume parallel to the XZ-plane. Without overlap, artefacts can again be seen at the edges of the regions (figure 4.30(b)). By applying an overlap of 10%, these artefacts are removed and the resulting image (figure 4.30(c)) does not show any apparent differences compared to the image of the full reconstruction (figure 4.30(a)). The reconstruction errors as function of the overlap are presented in table 4.3, showing that the error can be neglected for a 10% overlap. The artefacts observed at the bottom of the object are cone-beam artefacts and are thus not due to the application of the multiresolution method. Quantitative results are presented in table 4.3 and show a similar effect of the settings on the reconstruction quality as the splitting along the X-axis. However, when a sufficient amount of overlap is used, the increase of the NRMSE is much smaller. Therefore, in contrast with the X- or Y-splitting, dividing the volume along the Z-axis thus provides the ability of reconstructing the volume with only a marginal loss of quality.

(a) Full (b) No overlap (c) N_{OV} = 10%

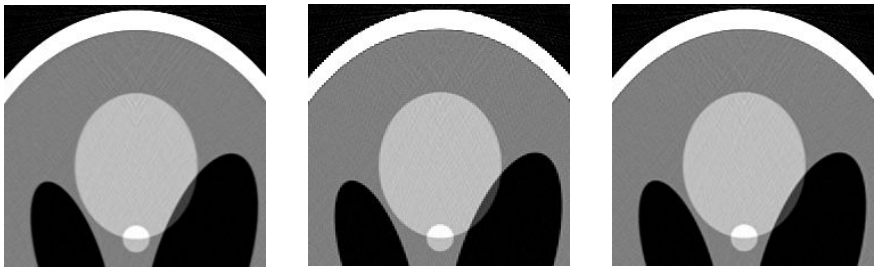
Figure 4.30: Reconstructed cross-sections parallel to the XZ-plane, for the full reconstruction (a), splitting into 8 subvolumes along the Z-axis without overlap (b) and with a 10% overlap (c).

Table 4.3: The NRMSE of the reconstructed cross-sections as function of different settings for splitting along the Z-axis. The optimal relaxation parameter was now found to be the same for all setting and equal to λ = 0.8. For the full reconstruction, the NRMSE was 0.1915.

$N_{OV}(\%)$	N_{SV}	N_{DS}	NRMSE	$\delta(\%)$
0	4	2	0.2194	14.542
5	4	2	0.1940	1.289
10	4	2	0.1915	0.004
10	4	4	0.1916	0.029
10	8	2	0.1918	0.159
10	8	4	0.1935	1.035

Initialization

As the low resolution volume already contains the correct reconstruction values (although at a lower resolution), it seems rational to initialize each of the subvolumes using these values in order to increase convergence of the solution. Results for this approach are presented in figure 4.31 and in table 4.4. These images show that the low resolution initialization results in very rough and chequered images. It seems that the reconstruction algorithm is unable to significantly alter the initial solution, even when the relaxation parameter λ is increased. Results can be improved by using linear interpolation instead of nearest neighbour interpolation in the resampling from low to high resolution. However, using initialized subvolumes only seems to degrade image quality while offering no apparent advantages.



(a) Zero initialization (b) Nearest neighbour (c) Linear interpolation

Figure 4.31: Reconstructed cross-section using 4 regions, by splitting the volume in 2 along the X- and Y-axis. Results are shown for the case where the subvolumes were initialized to 0 (a) and to the values retrieved from the low resolution reconstruction by nearest neighbour sampling (b) and by interpolation (c).

Table 4.4: The NRMSE for the multiresolution method, for different ways of initializing the high resolution subvolumes. The relative increase of the error compared to the normal reconstruction is given by δ. The volume was split in 2 subvolumes along the X-axis and the volume was sampled down with a factor of 2.

Settings	NRMSE	$\delta(\%)$
zero initialization	0.182	
low resolution init	0.210	16.5
interpolated low resolution	0.183	1.7

Performance

To conclude this section, an indication of the expected memory reduction and the speed of the multiresolution reconstruction is provided. These timings were obtained using the workstation described by table 3.1. First, consider the data set presented in the previous tests. The volume was split in 2 regions along each axes (thus 8 subvolumes in total) and a 5% overlap between the subvolumes was added. The low resolution volume was downsampled by a factor 2. This resulted in a decrease of the required memory from 541 MB (for the full reconstruction) to 120 MB (for the multiresolution reconstruction), while the reconstruction time increased from 111 seconds to 245 seconds. The increase in reconstruction time is due to the initial reconstruction of the low resolution volume and to the required overlap of the different subvolumes. Furthermore, the calculation of the forward projection of the low resolution volume has to be repeated for every subvolume, which poses an additional burden. Redundancy of this repeated forward projection may be avoided by considering alternative implementation schemes, for instance by using memory dumps to the CPU RAM-memory.

Now, consider the same phantom, but with a size of 1500^4, which is typically encountered in high resolution CT. The pixel pitch was decreased to 0.17 mm, resulting in a similar full cone angle of 14.53°. Reconstructing this volume requires 13.5 GB of memory, which is far more than the 4 GB available on the Tesla C1060. This data set was reconstructed by first splitting the volume in half at the horizontal symmetry plane. Each of these halves was then reconstructed separately by splitting the volume in 2 along each of the three main axes. Using a 5% overlap and a downsampling of a factor 2, the memory requirements were reduced to a mere 2.75 GB. The total reconstruction of the entire volume (both halves) took 5 hours and 10 minutes.

It has to be noted that, due to the necessity of the overlap to prevent artefacts, the multiresolution approach is not able to reconstruct any volume regardless of its size. Although a great decrease in required memory is achieved, the overlap stills require a certain minimum amount of memory. In practical applications, the size of the reconstruction volume is currently limited to about 2048^3 voxels for a GPU with 4 GB of RAM-memory.

4.13 Algebraic or statistical

Although the derivations of the algebraic and the statistical reconstruction algorithms are based on different concepts, the resulting formulas (equations (4.12) and (4.21)) and processing routine are quite similar. The major difference is that the calculation of the iterative update of the reconstruction volume for an algebraic routine is based on the differences in the ray sums between the measured and the simulated projection, whereas for a statistical routine this is based on the differences in the number of photons, which is simply obtained by taking the exponential of minus the calculated ray sums. Furthermore, the update step in a statistical reconstruction also contains an additional normalization by the number of photons, which can easily be introduced. In a practical implementation, these differences only require a minor adaptation to the code making it very easy to switch between an algebraic or a statistical method.

4.14 GPU implementation

Based on the previous discussions in this chapter, a highly efficient implementation of an iterative reconstruction algorithm can now be presented. The choices made for the implementation are aimed specifically at application in high resolution X-ray CT and execution on a GPU to increase performance.

Before proceeding, the reader should halt here for a moment and return to the Kaczmarz method introduced at the beginning of this chapter. Note that, even though at first glance any connection seems absent, in essence the practical implementation presented here is still directly derived from the original formula for the algebraic reconstruction technique as given by equation (4.7). Making this reflection clearly illustrates the major contrast between a proposed theoretical solution to a problem and its actual elaboration for every-day use in a practical situation.

As no apparent differences in quality[6] or performance were observed between algebraic and statistical methods, it was decided to use an algebraic algorithm. SART seemed to be the best choice, as it combines relatively high reconstruction speed with low memory requirements. ART was discarded since it requires a high number of updates of the volume (one for every detector pixel at each projection angle), making it computationally rather inefficient, even more so when a GPU is used, as its inherent structure does not allow it to fully exploit the GPU's parallel architecture. On the other hand, using ordered subsets in the reconstruction requires either every simulated projection image in the subset to be kept in memory or storing an additional volume in which the contribution of every projection in

[6] Statistical methods provide improved image quality for data with poor statistical information. In high resolution X-ray CT the statistical information is usually quite good, so it was expected that statistical and algebraic methods provide similar quality.

the subset to the update is accumulated, which is not suitable when reconstructing large data sets.

The forward projection of the volume is calculated using ray sampling, where the function values at the sampling points are retrieved using tri-linear interpolation. Due to its architecture, a GPU provides optimal performance when all parallel processors execute the same number of steps in a loop, which directly implies that Joseph's method is the optimal choice to perform the ray sampling. The sampling interval was chosen equal to the grid spacing. The update of the volume is performed using the fast backprojection with bi-linear interpolation. Using the CUDA language, the reconstruction volume and the correction image can be stored as a 3D and 2D array, respectively. By applying a texture to these arrays, the retrieval of the linearly interpolated values is very fast due to the GPU's architecture, which makes it significantly faster than a similar CPU-based implementation. Since the calculation and storage of the weights is thus no longer required, the combination of Joseph's method with the fast backprojection results in maximized performance and minimized memory requirements.

For the 512^4 volume, the reconstruction time for the central cross-section was reduced from 5.8 seconds on the CPU (using one processing core at 2.83 GHz) to only 0.414 seconds on the GPU (compared to about 0.200 seconds on the GPU using FDK), which is an increase in reconstruction speed by a factor of 14. For the larger 1500^4 volume, times decreased from 160 seconds to 5.182 seconds (compared to about 0.850 seconds on the GPU using FDK), which is a speed-up of 31 times. For a volume reconstruction, this increase in speed becomes even more prominent, as the interpolation for the forward projection then requires tri-linear instead of bi-linear interpolation, which is performed much more efficiently by a GPU than a CPU.

Intuitively, this forward and backprojection method is very closely related to the actual physical imaging process: X-rays are generated at the source, fired towards the detector and their path is tracked through the object while being attenuated by the objects composing materials. The simulated projection is then simply compared to the measured projection and the difference is smeared out over the current solution to obtain a new estimate of the imaged object. As will be shown in the next section, using ray sampling the forward projection can be easily modified to include other physical effects besides attenuation. This allows the simulated projection to be calculated more accurately with respect to the complete physical imaging process (which inherently includes attenuation combined with all other physical effect), which can be used to reduce certain disturbing artefacts.

The access order of the projections is determined by the weighted distance scheme. The volume is initialized to zero and no constraints are placed on the possible values of the reconstructed object function. When a conical beam is used, the volume is extended at the top and bottom during reconstruction. Vertical truncation of the projection data is handled by assuming the infinite extension, horizontal truncation by expanding the reconstruction volume outside the field of view of the

detector. Region of interest reconstructions are based on the reconstruction of a low resolution volume using the same iterative method. Reconstructions of volumes that are too large are achieved using the multiresolution approach, where the subvolumes are initialized to zero. Splitting along the Z-axis is preferred, unless the reduction in memory is still insufficient in which case the volume is also split along the X- and Y-axis.

Overall, the presented routine can be summarized by three important properties: it is fast, it has low memory requirements and its implementation is straightforward, both in a computational and a physical sense. This implementation is able to overcome the minor and major obstacles posed when using an iterative reconstruction algorithm in a practical application. It allows reconstructing the large volumes encountered in high resolution X-ray CT at an acceptable speed using easily accessible and affordable resources.

4.15 Artefact Reduction

In the previous sections, an extensive discussion was provided on the application of iterative reconstruction algorithms. A computationally efficient and practical implementation was derived, which enables the iterative reconstruction of high resolution CT data. In this section, a series of tests is presented, which illustrate the possible improvements in image quality that can be achieved by using an iterative algorithm instead of the commonly used filtered backprojection algorithm. This improvement is obtained by reducing a number of artefacts which are typically encountered in practical situations. For some types of artefacts, the inherent nature of iterative methods already provides improved reconstruction. For other types, the forward projection model is extended and improved to resemble more the actual physical process of X-ray imaging, in order to reduce the artefacts due to physical effects other than monochromatic attenuation.

Unless stated otherwise, each of these tests is performed using the modified Shepp-Logan phantom (see appendix A) with the settings described in section 4.6.5, thus with a size of 512^4 and a full cone angle of 14.6°.

4.15.1 Number of projections

A first test is to evaluate the reconstruction quality in terms of the number of projections that are taken from the scanned object. The possibility of achieving similar reconstruction quality with fewer projections is a very appealing thought, since this reduces the deposited dose in the object and decreases the scanning time.

First, consider the number of projections for the filtered backprojection (FBP) R_{FBP} and algebraic reconstruction technique (ART) R_{ART} that are required for

the reconstruction of 2D fan-beam data according theoretical motivations, a problem which was originally examined in [42]. The number of detector pixels is n, which is used to reconstruct a square grid of n^2 voxels. Since the reconstruction region is usually limited to the inscribed circle, the number of unknown voxels to be reconstructed is $\frac{\pi}{4}n^2$. The system of $R_{ART} \cdot n$ equations (4.3), which describes the algebraic reconstruction process, is determined when:

$$\frac{\pi}{4}n^2 = R_{ART} \cdot n \,, \tag{4.50}$$

thus the number of required projections for the algebraic method is:

$$R_{ART} = \frac{\pi}{4}n = 0.785 \cdot n \,. \tag{4.51}$$

For voxels of size d, the sampling interval in Fourier space is at least $\Delta\omega = 1/(d \cdot n)$ and the highest spatial frequency $\omega_{max} = 1/(2d)$. To completely sample the Fourier space, the projection data need to be sampled at an angular interval:

$$\Delta\theta = \frac{\Delta\omega}{\omega_{max}} = \frac{2}{n} \,. \tag{4.52}$$

The required angular range of the projection data for fan-beam reconstruction is about π (neglecting the additional range equal to the fan angle), thus the number of projections is given by:

$$R_{FBP} = \frac{\pi}{\Delta\theta} = \frac{\pi}{2}n = 1.57 \cdot n \,. \tag{4.53}$$

These results indicate that theoretically ART only requires half the number of projections than FBP, a fact which can be generalized to all iterative methods.

In figure 4.32, results are shown for the reconstruction of the 512^4 modified Shepp-Logan phantom, for an increasing number of projections. For the algebraic reconstruction, the relaxation parameter was set to 0.7, which generally provided the best results for all tests. Using only a single iteration seemed to be sufficient, as increasing the number of iterations did not result in any apparent improvement in quality. From these results, it is obvious that quality increases with an increasing number of projections. However, looking at the reconstructed cross-sections, it is difficult to tell which algorithm performs best, as each suffers from its own artefacts. For a low number of projections, FBP contains a lot of streak artefacts, whereas SART tends to be blurred. For a higher number of projections, both methods provide similar results.

The quantification of the reconstruction error in figure 4.33 provides a rather different perspective. From these results, it is clear that SART performs much better for a low number of projections, with an error down to half the error obtained using FBP. For higher numbers of projections, the errors become comparable which is to be expected when looking at the cross-sections.

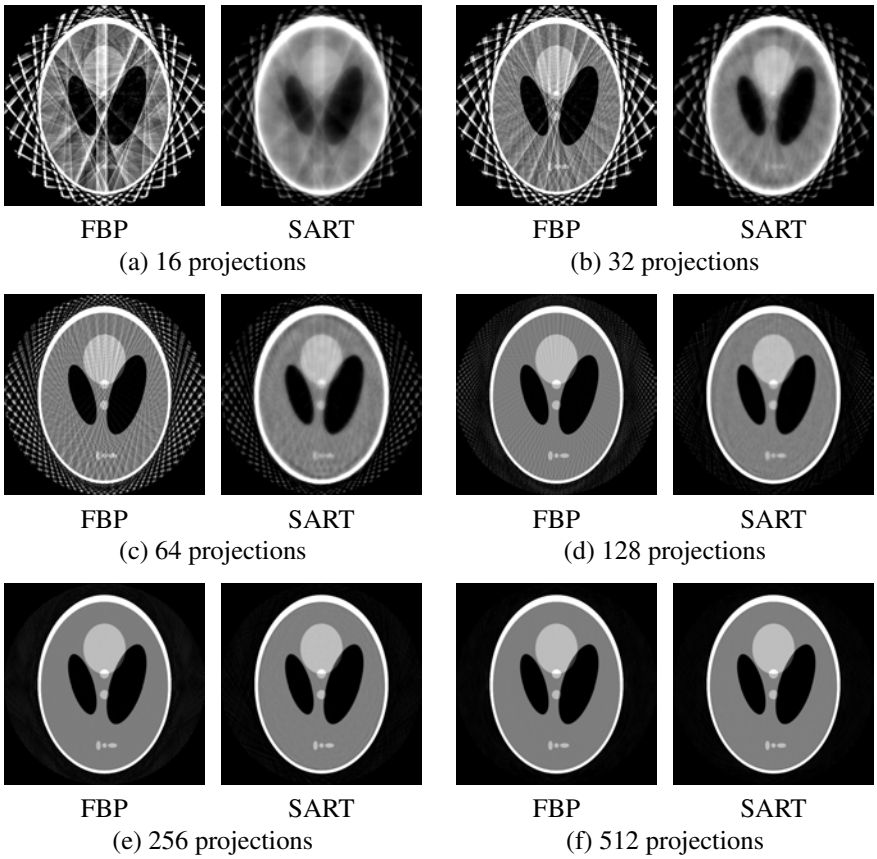


Figure 4.32: Comparison of reconstructions with an increasing number of projections.

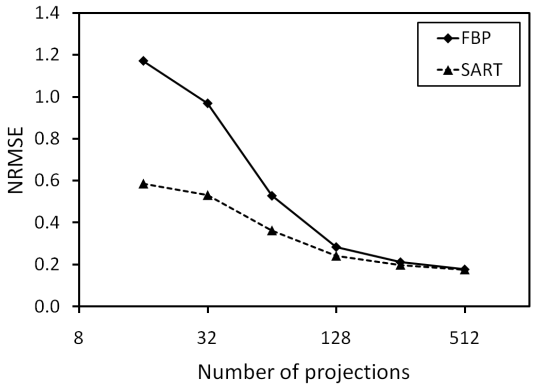


Figure 4.33: The normalized root mean squared error (NRMSE) as function of the number of projections used in the reconstruction. For SART, 1 iteration was used and λ was set to 0.7.

Regarding the theoretical derivation, it can be seen that the SART error for $R_{SART} = 32$ is similar to the FBP error for $R_{FBP} = 2\ R_{SART} = 64$. However, the difference between these errors becomes larger for lower numbers and smaller for higher numbers. For the theoretical values $R_{SART} = 0.785 \cdot 512 = 402$, the SART reconstruction error was found to be 0.1807 compared to an FBP error of 0.1721 for $R_{FBP} = 1.57 \cdot 512 = 804$.

It can be concluded from this evaluation that SART provides better reconstruction quality than FBP when a low number of projections is available. However, it is obvious that when too little projections are taken, even SART provides reconstructions of poor quality. So although results can be improved using an appropriate algorithm, it is advisable to always take a number of projections in the order of the number of detector pixel columns, as long as this is practically feasible.

4.15.2 Noise

The imaging process can be described adequately to the first order by Poisson statistics. For a Poisson distribution with an expected number of events λ, the probability of detecting k events is given by:

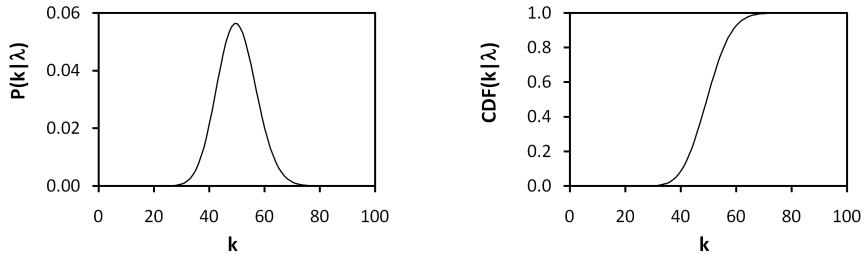
$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}\ , \tag{4.54}$$

and the cumulative distribution function $CDF(k|\lambda)$ by:

$$CDF(k|\lambda) = e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}\ . \tag{4.55}$$

In figure 4.34, these functions are illustrated for $\lambda = 50$. The noise behaviour is studied as function of the expected number of events Λ per detector pixel when the X-rays are not attenuated. For each detector pixel, the expected number of events λ_i is then given by $\Lambda \cdot I_i$, where I_i is the normalized intensity measured in that pixel. Noise is now added to the projection images as follows. For every pixel, a random number r is generated between 0 and 1. The inverse of the cumulative distribution function is then used to determine the number of events k_i such that $CDF(k_i|\lambda_i) = r$. The normalized intensity with Poisson noise in the pixel is now given by k_i/Λ.

In figure 4.35, results are shown for the reconstruction of projection data for an increasing number of photon counts, thus a decreasing noise level. In the FBP reconstruction the additional noise filter described by equation (3.16) was used with parameter $\alpha = 1.0$. For SART, the relaxation parameter λ was chosen differently for each noise level, such that the reconstruction error is minimized. Again, increasing the number of iterations did not provide any significant improvement in quality, so only a single iteration was used. It can be seen that SART decreases the



(a) (b)

Figure 4.34: The probability $P(k|\lambda)$ (a) and the cumulative distribution function $CDF(k|\lambda)$ (b) as function of the number of events k, for a Poisson distribution with an expected number of events $\lambda = 50$.

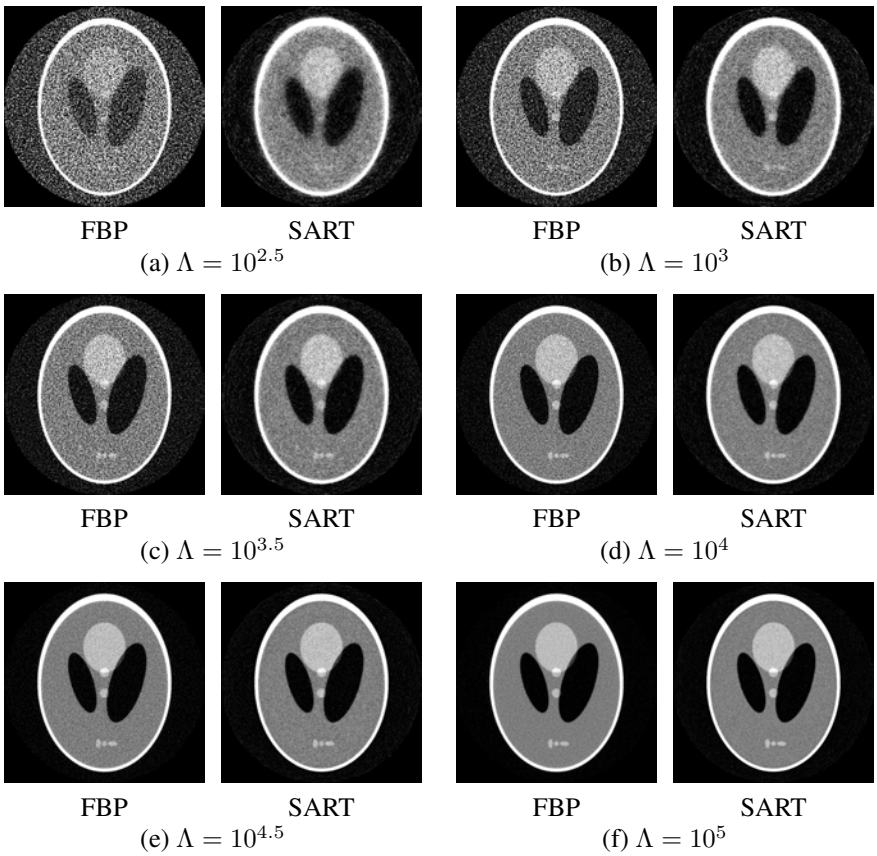


Figure 4.35: Comparison of reconstructions from projections with different noise levels.

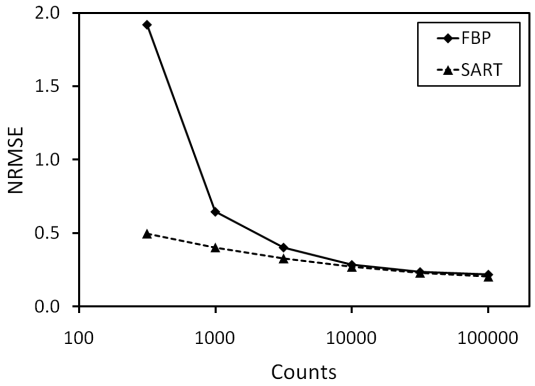


Figure 4.36: The NRMSE as function of the noise level (determined by the expected number of counts in the background) in the reconstruction. For SART, 1 iteration was used and for λ the value was chosen that minimizes the error.

level of noise in the reconstruction as compared to FBP. Consequently, the reconstruction error, shown in figure 4.36, is significantly lower for SART, especially at high noise levels.

4.15.3 Limited angle

Due to several practical limitations (installation of peripheral devices, extended range of the object, etc.), it may not be possible to rotate the object over the required angular range, not even for a short scan. In this situation, a series of projections in a certain angular range will be missing. The term *limited angle tomography* is often used to describe this situation. For these tests, it will be assumed that information is missing along two opposite circle segments in the projection angles, described by the angular intervals $[\alpha_b,\ \alpha_e]$ and $[\alpha_b + 180°,\ \alpha_e + 180°]$. This results in a bow-shaped segment in the projection angles that is missing, as illustrated in figure 4.37. In short, such an interval will be denoted by the opening angle $\alpha = \alpha_e - \alpha_b$ of one segment.

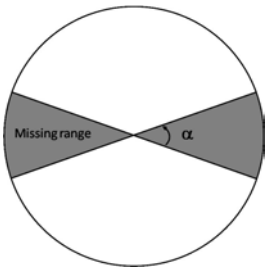


Figure 4.37: The missing angular range in a limited angle tomography.

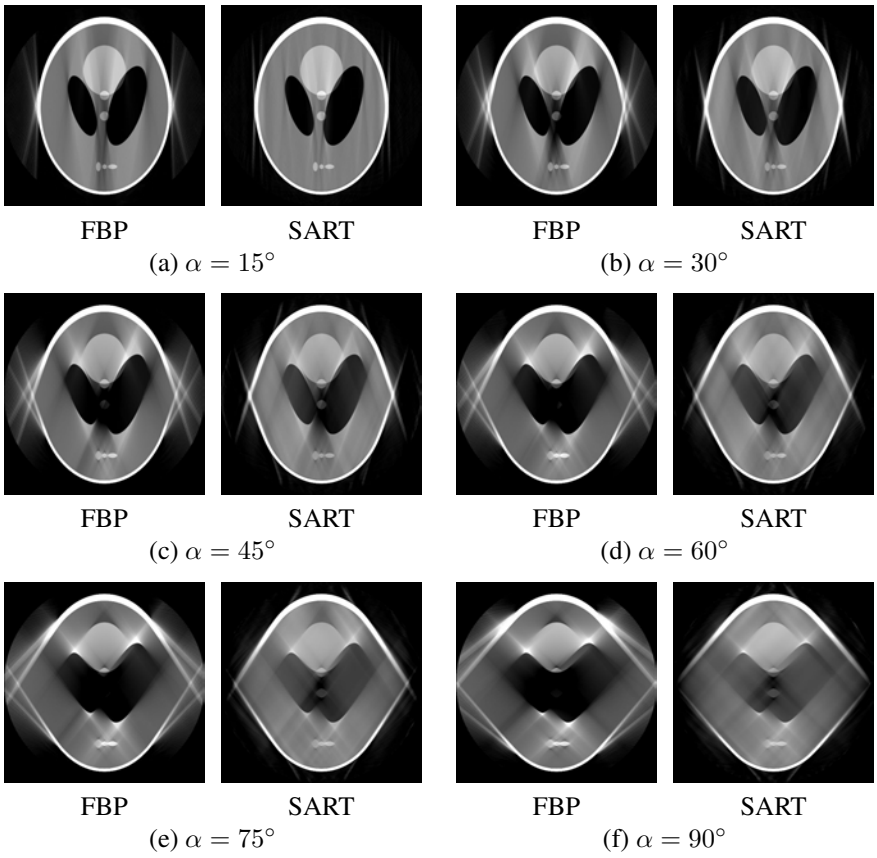


Figure 4.38: Reconstruction results for limited angle tomography as function of the opening angle α of the missing range.

This test was performed for missing intervals corresponding to α = 15°, 30°, 45°, 60°, 75° and 90°. Reconstruction results are presented in figure 4.38. As was to be expected, the image quality of the reconstructed cross-sections degrades with increasing α. Due to the lack of projection data along angles close to the X-axis, information on transitions along the Y-axis in the cross-section is lost to a certain extent, which leads to distorted reconstructions in which the ellipsoids are open, instead of being nicely closed. Another consequence of the missing information is that the reconstruction values in certain regions are no longer correct, which can clearly be seen in figure 4.39. For both FBP and SART, the reconstructed values are much lower than the correct values of the original phantom, an effect which deteriorates towards the centre of the cross-section. Using FBP, this artefact is much stronger, leading to negative reconstruction values. As a result, the circular feature in the middle of the phantom, between the two void ellipsoids, consists only of negative values, making it very hard to distinguish. The feature seems to

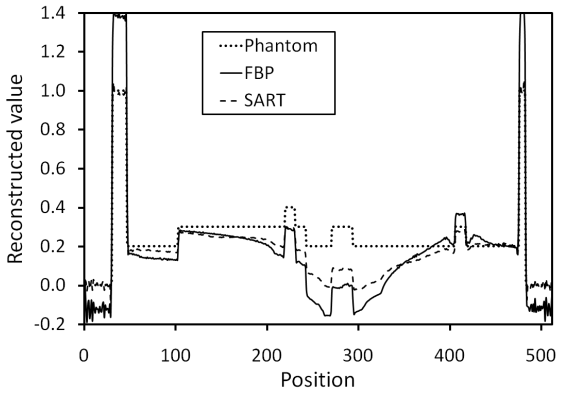


Figure 4.39: A line profile along the central vertical line of the reconstructed cross-section, for both the original phantom and the FBP and SART reconstructions of a limited angle tomography with $\alpha = 60°$.

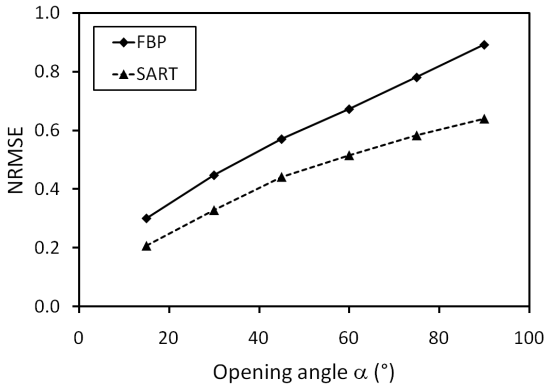


Figure 4.40: The NRMSE as function of the opening angle α of the segment of missing projection angles.

have disappeared in the two voids. Furthermore, in FPB the reconstructed values of the outer ellipsoid are much higher. The reconstruction error, shown in figure 4.40, is also much lower for SART than for FBP, even for relative small missing angles.

4.15.4 Missing angular intervals

Another possibility regarding missing information is the situation in which at several angular positions projection data are missing within a certain angular interval. This is illustrated in figure 4.41, where a series of projection angles is shown in which projection data are missing for an interval of 14° at 3 angular positions.

Missing data of this kind are sometimes encountered in a scanner using a fast scan regime. In a fast scan, the rotation of the object is continuous and projection images are acquired during the rotation. This is in contrast with the more common step and shoot routine, in which the rotation is stepwise and is performed each time between two consecutive projections. When applying the fast scan routine, at some times during the scan a reset of the detector's framebuffer may be required. During this reset, no images will be recorded, but since the rotation of the object is continuous, there will be a gap in the projection data and any information corresponding to those missing projection angles will be missing.

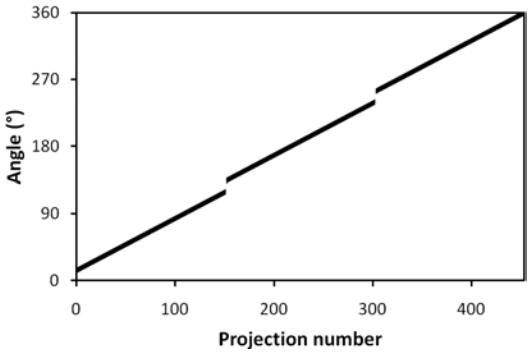


Figure 4.41: Projection angle as function of the projection number. In total, 3 intervals containing each 20 projections are missing, corresponding to 14° of missing information per interval.

The reconstruction quality in case of missing angular intervals was evaluated as function of two different parameters: the angular interval over which the projection data are missing and the number of intervals that are missing. The intervals were distributed equally over 360° and the number of intervals was chosen to be odd, as an even number would correspond to limited angle tomography. It is noted that, when a certain angular interval is missing, information corresponding to that interval plus 180° is available, a fact which is not true in limited angle tomography. One missing interval simply corresponds to a short scan regime, so the tests start with 3 missing intervals.

Some of the reconstructed cross-sections from these two tests are presented in figure 4.42. It is obvious from these images that reconstruction quality degrades with increasing size of the intervals and an increasing number of intervals, at least for the FBP reconstructions. The SART results on the other hand do not seem to suffer from this loss of information and appear to be of similar quality than the normal reconstruction (without the missing intervals). This can be seen more clearly in the measurement of the reconstruction error, shown in figure 4.43. While for FBP the error rises quickly, for SART the error only increases slightly. Using FBP, slightly improved results were obtained by interpolating the missing data from the known projections, but reconstruction quality was still much lower compared to SART.

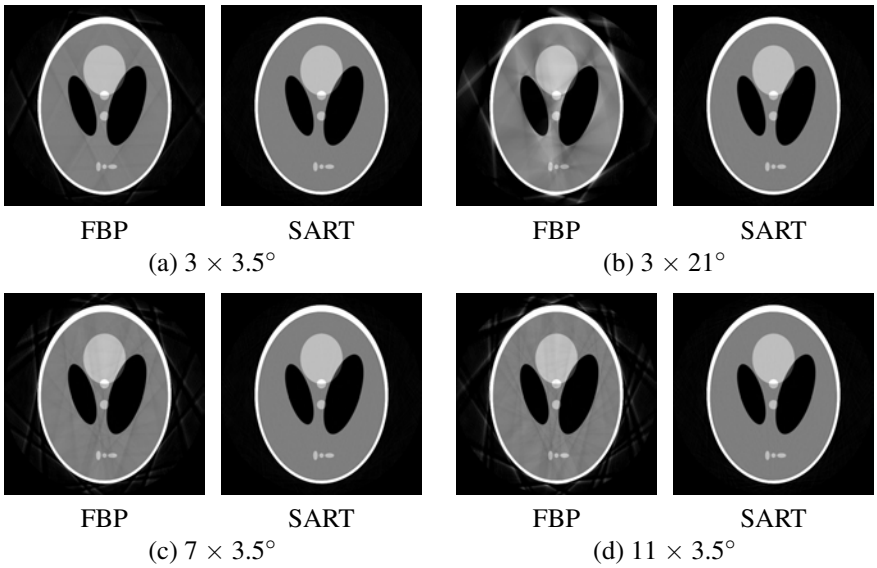


Figure 4.42: Reconstruction of data where projections from a number of angular intervals are missing.

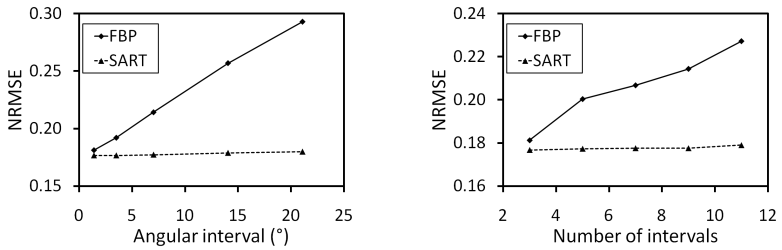
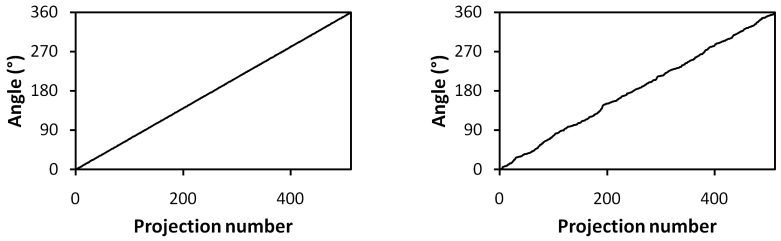


Figure 4.43: Reconstruction error as function of the size of the angular interval using 3 intervals (a), and as function of the number of intervals for intervals of angular size 3.5° (b).

4.15.5 Non-equiangular sampling

As a last test regarding the angular sampling, the case of nonequiangular sampling is considered. This situation again may occur when using a fast scan routine. Due to several technical causes, like delayed communication, the projection data may not be acquired at a fixed angular interval. This results in a nonequiangular sampling of the projection data, which needs to be compensated in the FBP by an appropriate angular weighing. This weighing is determined for each projection by calculating the angular range that is covered by that projection. This type of

sampling was achieved in two different ways: by random sampling of the projection angles and by applying a random move to the equiangular positions. The projection angles for both sampling methods are shown in figure 4.44.



(a) Random move of angles (b) Random angles

Figure 4.44: Projection angle as function of the projection number sampled nonequiangular by applying a random move to the equiangular positions (a) and by random sampling of the angles (b).

FBP SART FBP SART

(a) Random move of angles (b) Random angles

Figure 4.45: Reconstructed cross-sections of projection data acquired at nonequiangular positions.

The reconstruction results for the nonequiangular sampling are presented in figure 4.45. The reconstruction errors for the random movement are 0.185 and 0.180 for FBP and SART, respectively, and 0.223 and 0.189 for the random sampling. In both cases, the reconstruction quality obtained using SART is better than FBP, resulting in less streaking and shading artefacts. As was to be expected, the difference in quality becomes more apparent as the sampling deviates more from the fixed angular step.

4.15.6 Cone-beam artefacts

As was already discussed in section 3.4.3, the acquisition of projections in the regular cone-beam geometry does not yield a complete sampling of the 3D Radon transform, resulting in cone-beam artefacts. In addition, the FDK algorithm (section 3.4.2) is only an approximate method and has the property of introducing an amount of blurring along the Z-axis, which increases as the cone angle increases. Although alternative geometries exist which do not suffer from these cone artefacts, like the helical cone-beam geometry (section 3.5.4), the standard cone-beam geometry is still predominantly used in high resolution X-ray CT due to several practical considerations. Therefore, it is very important to investigate the appearance of cone-beam artefacts when iterative methods are used, since a reduction of cone-beam artefacts is highly favourable. It is noted that in [43], the application of SART was already evaluated in the reduction of cone-beam artefacts in off-centre cone-beam CT. This study showed that SART provided better quantitative reconstructions than the modified FDK methods.

The evaluation of cone-beam artefacts was performed using the Defrise phantom, which is described in appendix A. The phantom contains 19 discs of diameter 84 mm and linear attenuation value 0.1 cm^{-1}, separated by a distance of 14 mm (between the centre of two consecutive discs). The thickness T of the discs was varied between simulations in order to investigate its influence on the artefacts and was chosen equal to 1.4 mm, 7.0 mm and 12.6 mm. The central disc was centred at the horizontal symmetry plane (optical axis) of the scanner geometry. The distance from the X-ray source to the object was 1000 mm and 2000 mm to the detector. The detector contains 1200 rows and 364 columns of pixels, with a pixel pitch of 0.508 mm. By choosing a high number of pixel rows (compared to the number of pixels columns), a large range of values for the cone angle can be evaluated in a single simulation. Only the results for the upper half and the central disc of the phantom are shown, results for the lower half are similar due to symmetry.

For a certain location along the axis of rotation of the geometry, the cone angle α is easily obtained by:

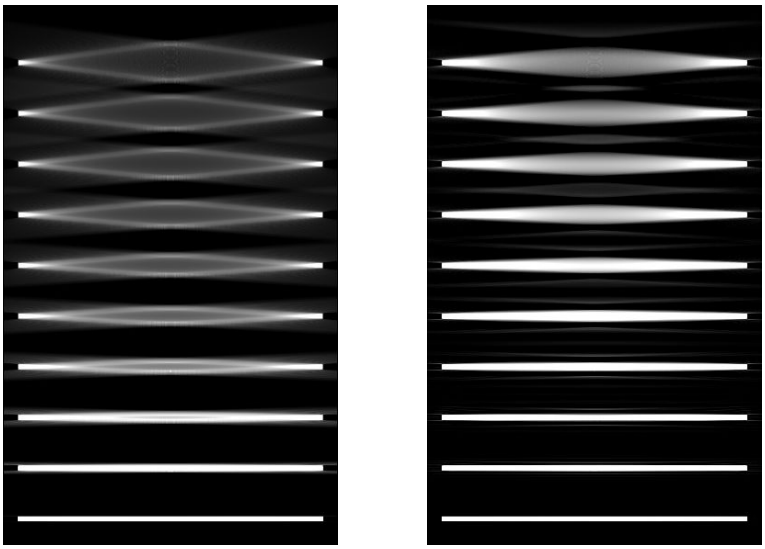
$$\alpha = \tan^{-1}\left(\frac{z}{S}\right) , \tag{4.56}$$

with z the distance from that location to the horizontal symmetry plane and S the distance from the source to the rotation axis. Using this expression, each reconstructed disc can be identified by the cone angle at its centre.

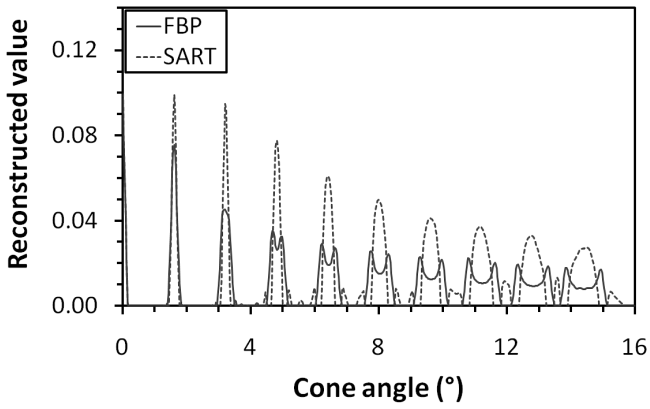
In figures 4.46 to 4.48, results are shown for both the FDK and SART reconstruction of the Defrise phantom, for different thicknesses of the discs. For the thin discs (thickness 1.4 mm), FDK seems to split every disc into two thinner, distorted surfaces, with a lower reconstructed value in between. These two surfaces connect at the edges, and the angle between them at this location corresponds to the full cone angle at that height. Using SART, the discs are widened towards the centre with a slope corresponding to the cone angle. The reconstructed values decrease

as the widening increases. Additionally, stripes appear above and below the discs, which at larger cone angles combine to some kind of smaller disc in between the actual discs of the phantom, containing a low reconstructed value. In the FDK reconstruction of the medium discs (thickness 7.0 mm), the actual disc decreases in thickness with increasing cone angle, while a halo of lower reconstructed values appears that increases in thickness. For the most upper discs, the actual disc starts to disappear, while the halos from different discs start to overlap. In SART, the discs contain several stripes, which combine as the disc size decreases. The decrease of the disc thickness and the increase of the halo are far less than in the FDK reconstruction. Finally, for the thick discs (thickness 12.6 mm), the same artefacts are visible as with the medium discs. However, because of the smaller gap of air in between consecutive discs, the blending between the halos in the FDK image already starts at relatively small cone angles. At higher cone angles, the gap of air almost completely disappears. For SART, the halos also overlap at the higher cone angles and a higher number of stripes can be noticed than with the medium discs.

Quantitative results are presented in figures 4.49 to 4.51, where the reconstruction error is given for each disc separately as function of the cone angle at the centre of that disc. The error is determined by calculating the normalized root mean squared error as compared to the original phantom. For each disc, this error is measured in the region that starts halfway between the disc and the previous disc, and stops halfway between the disc and the next disc. So both the disc and half of the void interval above and below it are considered in the calculation of the reconstruction error. In these plots it can be seen how the error increases with increasing cone angle, which visualizes the deterioration of the cone-angle artefacts when moving away from the horizontal symmetry plane. Furthermore, it is clear that the reconstruction error becomes significantly smaller when using SART instead of FDK.

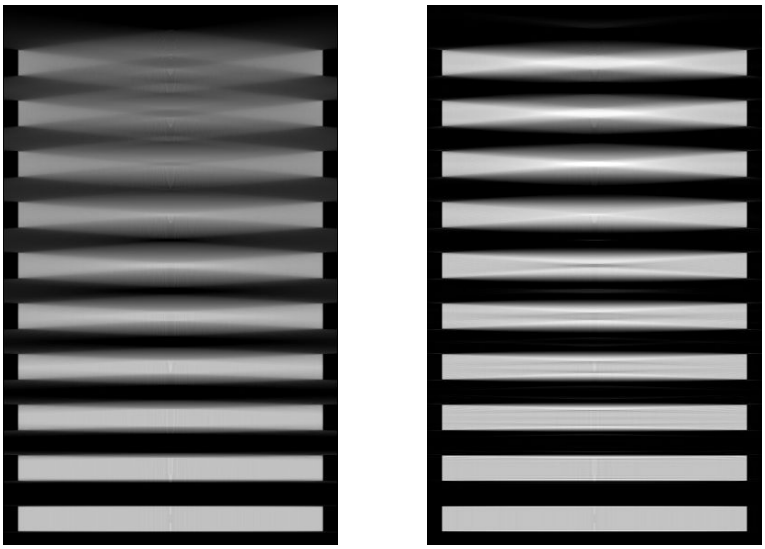


(a) FDK (b) SART

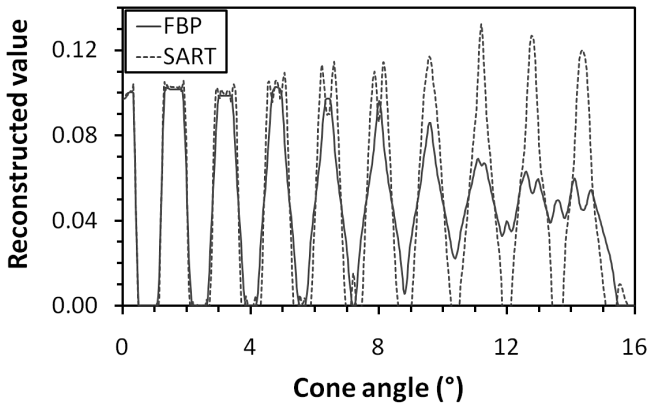


(c) Line profile

Figure 4.46: Reconstructed cross-sections of the simulated Defrise phantom parallel to the XZ-plane, using FDK (a) and SART (b). The thickness of the discs is 1.4 mm. Starting at the bottom, the horizontal centre of each disc corresponds to a full cone angle of 0.0°, 1.6°, 3.2°, 4.8°, 6.4°, 8.0°, 9.6°, 11.2°, 12.8° and 14.4° respectively. The line profile (c) is taken across the vertical central line of the cross-sections and is plotted as function of the corresponding cone angle at that position.

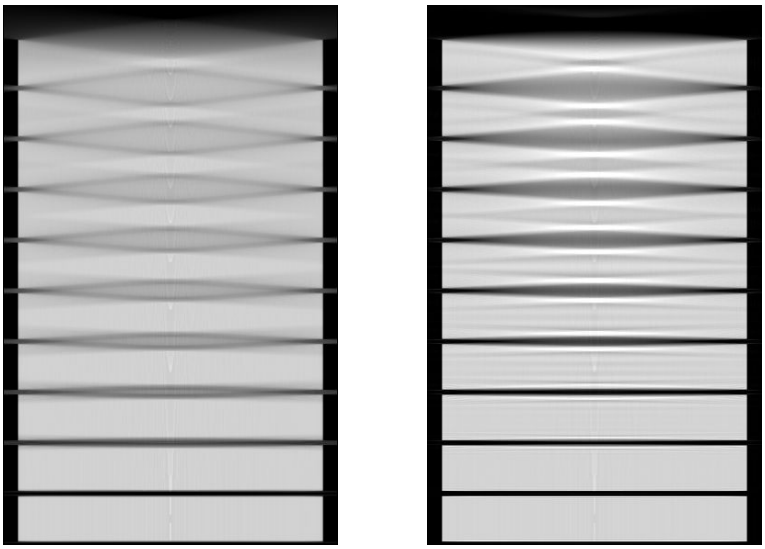


(a) FDK (b) SART

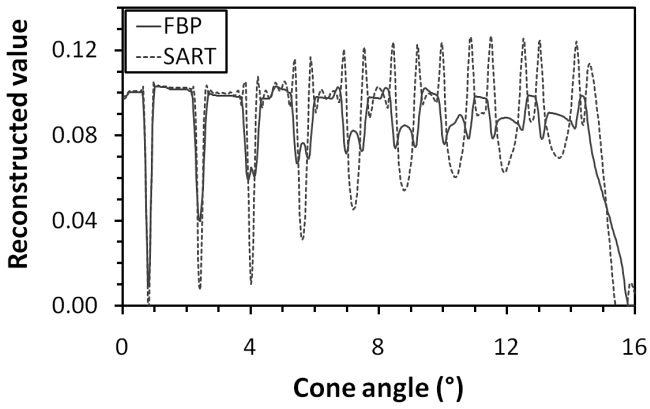


(c) Line profile

Figure 4.47: Reconstructed cross-sections of the simulated Defrise phantom parallel to the XZ-plane, using FDK (a) and SART (b). The thickness of the discs is 7.0 mm. Starting at the bottom, the horizontal centre of each disc corresponds to a full cone angle of 0.0°, 1.6°, 3.2°, 4.8°, 6.4°, 8.0°, 9.6°, 11.2°, 12.8° and 14.4° respectively. The line profile (c) is taken across the vertical central line of the cross-sections and is plotted as function of the corresponding cone angle at that position.



(a) FDK (b) SART



(c) Line profile

Figure 4.48: Reconstructed cross-sections of the simulated Defrise phantom parallel to the XZ-plane, using FDK (a) and SART (b). The thickness of the discs is 12.6 mm. Starting at the bottom, the horizontal centre of each disc corresponds to a full cone angle of 0.0°, 1.6°, 3.2°, 4.8°, 6.4°, 8.0°, 9.6°, 11.2°, 12.8° and 14.4° respectively. The line profile (c) is taken across the vertical central line of the cross-sections and is plotted as function of the corresponding cone angle at that position.

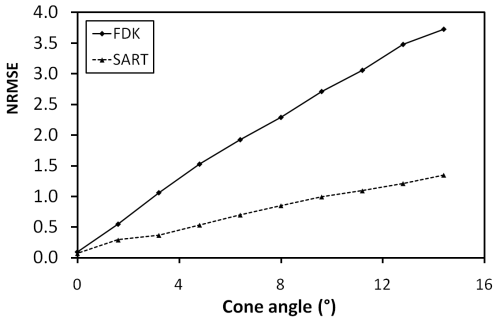


Figure 4.49: The normalized root mean square error (NRMSE) of each reconstructed thin disc as function of the cone angle at its centre.

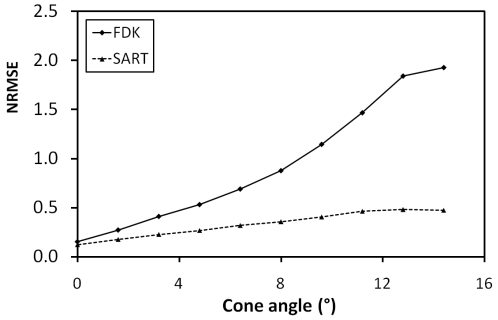


Figure 4.50: The normalized root mean square error (NRMSE) of each reconstructed medium disc as function of the cone angle at its centre.

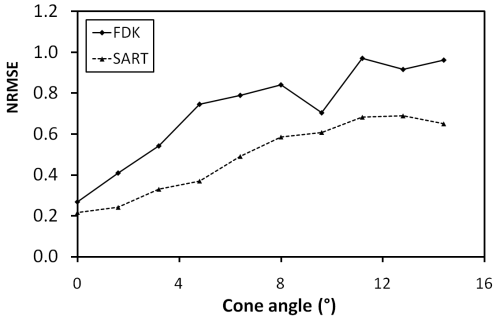


Figure 4.51: The normalized root mean square error (NRMSE) of each reconstructed thick disc as function of the cone angle at its centre.

4.15.7 Beam hardening

In section 3.7, the problem of beam hardening and the resulting artefacts were already introduced and a simple correction method to reduce the cupping artefact was discussed. This method however does not provide a solution to the typical dark streaks that appear in regions connecting several strongly attenuating objects. Beam hardening artefacts are a result of the fact that the linear attenuation coefficient depends on the energy of the X-rays. A polychromatic beam thereby violates the exponential attenuation described by the law of Lambert-Beer (equation (3.3)). The reduction of beam hardening artefacts is especially critical in medical imaging, and many studies have been performed to understand the cause of these artefacts and to find correction schemes. In this regard, the introduction of iterative algorithms allowed the development of a whole new range of correction methods, e.g. [44, 45] and many others, and it is still a very popular field of ongoing research. Most of these methods however are based on some kind of prior information and require knowledge about the composition of the object, which is perfectly feasible in medical applications, but not in high resolution CT where the size, shape and composition of objects varies over a wide range.

During this work, a more general method for reducing beam hardening artefacts was evaluated, inspired by earlier work presented by Hsieh *et al.* [46]. In their paper, a general beam hardening correction approach was presented which is applicable to any reconstruction algorithm and which can briefly be described by the following steps: First, a reconstruction is performed without any beam hardening correction. Next, the strongly attenuating regions in the reconstruction are selected by applying a threshold. These regions are then forward projected and the calculated ray sums are used to correct the original projection images using some polynomial function, of which the parameters can be tuned. The tomographic reconstruction is then reperformed using the corrected projections, to arrive at a reconstruction with a beam hardening correction. This method can be performed iteratively for additional corrections using multiple components.

Many methods similar to the one described above exist, which all use the same principle: the original projection data are corrected based on the reprojection of a temporary solution, and then reconstructed to obtain an improved solution. The correction of the projection data can be performed by application of a polynomial function, interpolation of the data, or some other method. Nonetheless, these methods usually require at least two backprojections and one forward projection of the data, which can become quite cumbersome.

Using an algebraic reconstruction algorithm however, it is possible to integrate such corrections directly in the reconstruction by modifying the forward projection step in the routine, thereby eliminating the need for multiple reconstructions and additional forward projections. The following correction method provides a practically feasible solution applicable in a variety of experimental situations without requiring any prior knowledge, thereby satisfying the requirements for application in high resolution CT. In this correction model, the ray sums p_i are no longer

calculated by simply accumulating the attenuation values f_i while tracing a ray through the reconstruction volume. Instead, L different intervals $[f_-^{(l)}, f_+^{(l)}]$ are created, defined by a minimum $f_-^{(l)}$ and maximum $f_+^{(l)}$ attenuation value. Partial ray sums $p_i^{(l)}$ are now calculated for each interval, adding only attenuation values within that specific interval:

$$p_i^{(l)} = \sum_{j=1}^{N} w_{ij} f_j^{(l)} \ , \tag{4.57}$$

with

$$f_j^{(l)} = \begin{cases} f_j & \text{if } f_-^{(l)} \leq f_j < f_+^{(l)} \\ 0 & \text{otherwise .} \end{cases} \tag{4.58}$$

The total ray sum is then calculated as the sum of polynomial functions of the partial ray sums:

$$p_i = \sum_{l=1}^{L} \left[\sum_{j=1}^{\infty} \alpha_j^{(l)} \cdot \left(p_i^{(l)} \right)^j \right] \ , \tag{4.59}$$

where the $\alpha_j^{(l)}$ are the coefficients of the polynomial function.

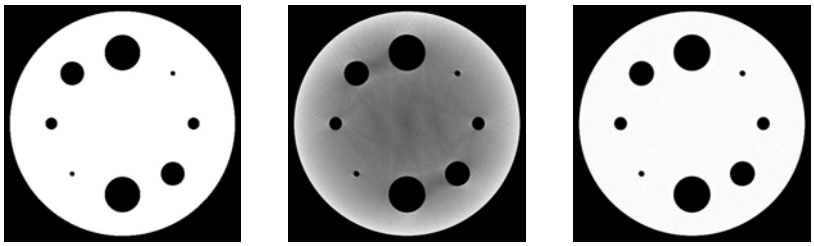
Although this may seem to be an extensive calculation, which would definitely have its impact on the reconstruction time, this method can be implemented with only little effort and a minimal decrease in speed. In practice, using two different intervals is usually sufficient: one weakly attenuating and one strongly attenuating interval, which are separated by a simple threshold value. The weakly attenuating interval corresponds to the matrix of the object which gives rise to the cupping artefacts. The strongly attenuating interval contains the strongly attenuating inclusions trapped in the matrix that are the cause of the dark streaks. During the forward projection of one ray, two separate ray sums are now kept, one weakly attenuating p_i^w and one strongly attenuating ray sum p_i^s. At each sampling step, the retrieved attenuation value is compared to the threshold value and added to the appropriate ray sum. When the tracing of the current ray is finished, the corresponding total ray sum can be calculated from the weakly and strongly attenuating ray sums using second order polynomial functions with an offset of zero ($a_0 = b_0 = 0$):

$$p_i = a_1 \cdot p_i^w + a_2 \cdot \left(p_i^w\right)^2 + b_1 \cdot p_i^s + b_2 \cdot \left(p_i^s\right)^2 \ , \tag{4.60}$$

The threshold value can easily be determined by reconstructing a single cross-section through the object[7]. The optimal parameters for the polynomial correction of the weakly attenuating matrix (a_1 and a_2) and of the strongly attenuating inclusions (b_1 and b_2) can be found experimentally by evaluating a single reconstructed cross-section using different values.

[7] In cone-beam, the cross-section at the horizontal symmetry plane of the set-up can be reconstructed correctly in fan-beam geometry. Therefore, it is possible to evaluate parameters on a single cross-section, instead of each time having to reconstruct the entire volume.

A qualitative and quantitative evaluation of the beam hardening correction method is performed using the metal phantom described in appendix A, with a pixel pitch of 0.5 mm, source to object distance of 20 mm and source to detector distance of 1000 mm. Beam hardening was introduced in the simulated projection images by applying a polynomial function to the attenuation of the phantom. Results are shown in figures 4.52 and 4.53 for the reconstruction of the phantom without the 8 strongly attenuating inclusions, illustrating the effect of the correction regarding the cupping artefact. Using the modified forward projection, the cupping artefact was removed from the reconstruction. The optimal values for the polynomial correction were a_1 = 1.0 and a_2 = -0.5. The measured errors of the reconstruction were 0.20 for the uncorrected and 0.08 for the corrected SART reconstruction. Reconstructions obtained using the filtered backprojection algorithm are not shown here, but quality was very similar to the uncorrected SART reconstruction.



(a) Phantom (b) SART (c) SART corrected

Figure 4.52: Reconstructed cross-sections of a phantom containing a cupping artefact due to beam hardening, without any correction (b) and using the polynomial correction included in the forward projection (a_1 = 1.0, a_2 = -0.5) (c), as compared to the original phantom (a).

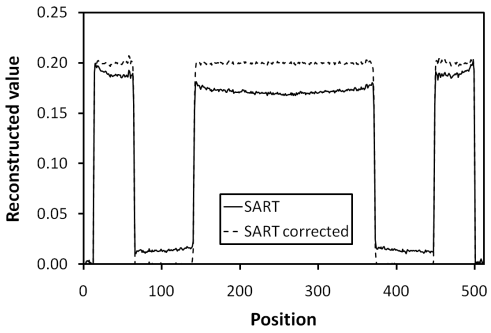
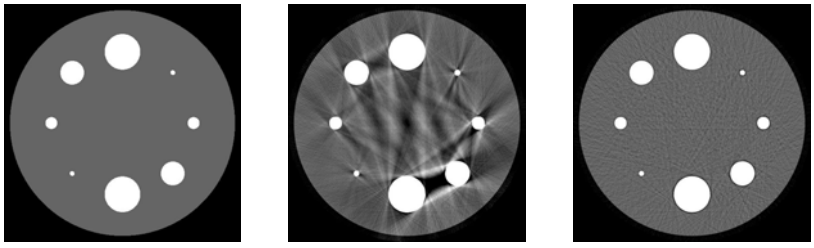


Figure 4.53: Line profiles across the vertical central line of figures 4.52(b) and (c), showing the uncorrected and corrected reconstruction for the cupping artefact.

In figures 4.54 and 4.55, reconstruction results are shown of the phantom containing the strongly attenuating inclusions, which give rise to the dark streak artefacts, in addition to the cupping artefact in the matrix. The dark streaks appear between each combination of two strongly attenuating features, which is a typical feature of this type of beam hardening artefact. By using the modified forward projection as discussed above, both the cupping and the dark streak artefacts can be removed by choosing appropriate parameters for the polynomial correction. The optimal parameters for this simulation were a_1 = 1.0, a_2 = -0.5, b_1 = 2.0 and b_2 = -0.8, with the threshold level set at 0.3. The reconstruction errors were 0.77 and 0.10 for the uncorrected and corrected SART reconstruction, respectively.

It is thus shown that the reconstruction quality can be improved significantly by introducing a simple modification to the forward projection model. Using the general framework for iterative methods presented in this chapter, the path is cleared for the development of more advanced forward models. Future research on such improved models holds great promise regarding the further increase in reconstruction quality obtained by iterative algorithms.



(a) Phantom (b) SART (c) SART corrected

Figure 4.54: Reconstructed cross-sections of a phantom containing a cupping artefact and dark streaks due to beam hardening, without any correction (b) and using the polynomial correction included in the forward projection (a_1 = 1.0, a_2 = -0.5, b_1 = 2.0, b_2 = -0.8) (c), as compared to the original phantom (a).

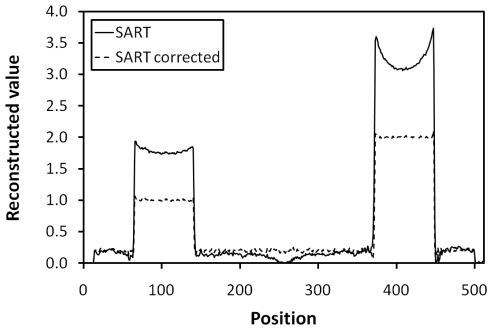
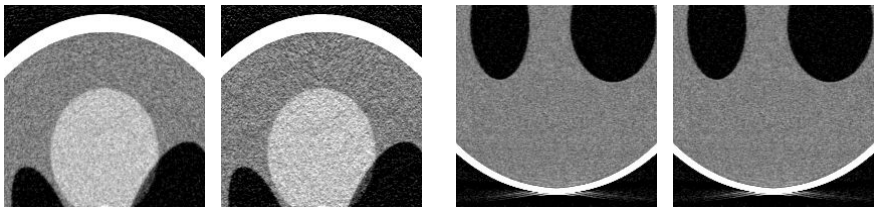


Figure 4.55: Line profiles across the vertical central line of figures 4.54(b) and (c), showing the uncorrected and corrected reconstruction for the cupping and dark streak artefacts.

4.15.8 Multiresolution

The benefit of using an iterative reconstruction algorithm is that it can provide superior image quality over the filtered backprojection when the projection data are somehow limited (low number of projections, noise, limited angle, etc.). It is therefore important to validate the quality of the multiresolution approach under such conditions. In this regard, four different cases of limited projection data are considered: a low number of projections, projections containing significant noise, projections covering a limited angular range and projections acquired at irregular angular sampling. The test for the low number of projections is performed by using only one in four (thus at total of 128 projections instead of 512) of the original projections. The noisy data are generated by adding noise to the simulated projection data according to a Poisson distribution, with the average number of counts in the background given by $\Lambda = 10^5$. For the limited angle case, the projection data were obtained by removing the projections acquired at angles in the intervals [67.5°,112.5°] and [247.5°,292.5°] from the original data set. This means that projection data are missing over a range of 45°. The irregular angular sampling of the data is obtained by randomly selecting 512 angles between 0°and 360°and calculating the projections at these angles, which results in a nonuniform distribution of the projections. For each of these five situations (the normal data and the four cases of limited data), reconstructions were performed using the full FDK algorithm, the full SART and the multiresolution SART (MR-SART) method with different settings. These evaluations were done using both the X- and Z-splitting and results are presented in tables 4.5 and 4.6, respectively. Some of the reconstructed images for the case of noise in the projections are shown in figure 4.56 for illustration.



SART MR-SART SART MR-SART

(a) X-splitting (b) Z-splitting

Figure 4.56: Comparison of reconstruction results for projections containing noise, using SART and MR-SART with X-splitting ($N_{SV} = 4$ and $N_{DS} = 4$) (a) and with Z-splitting ($N_{SV} = 8$ and $N_{DS} = 4$) (b).

For the X-splitting, the NRMSE is measured for the cross-section parallel to the XY-plane, at z = -0.25, which represents the same cross-section as was used in the previous evaluation of the X-splitting. It can be seen that for the normal data the full SART reconstruction provides a slightly better result than the full FDK. When using the multiresolution SART, quality deteriorates and the FDK

Table 4.5: The NRMSE of the reconstructed cross-sections and the relative increase as function of different settings for splitting along the X-axis. Results for the normal projection data are compared with results from limited projection data. The optimal relaxation parameter λ is given for completeness. The multiresolution SART reconstructions (MR-SART) are performed using a 10% overlap between the subvolumes. The increase of the NRMSE relative to the one of the full SART reconstruction is given by δ.

Settings	λ	NRMSE	$\delta(\%)$
Normal data			
Full FDK reconstruction		0.1533	1.23
Full SART reconstruction	0.8	0.1514	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.8	0.1550	2.34
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.8	0.1770	16.93
Low number of projections			
Full FDK reconstruction		0.2698	18.37
Full SART reconstruction	0.9	0.2280	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.9	0.2296	0.71
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.9	0.2386	4.65
Noise			
Full FDK reconstruction		0.2676	30.98
Full SART reconstruction	0.4	0.2043	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.4	0.2186	6.98
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.4	0.2400	17.44
Limited angle			
Full FDK reconstruction		0.5637	30.10
Full SART reconstruction	0.9	0.4332	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.9	0.4327	-0.12
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.9	0.4472	3.22
Irregular angular sampling			
Full FDK reconstruction		0.2989	75.76
Full SART reconstruction	0.9	0.1701	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.9	0.1786	5.02
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.9	0.2050	20.52

reconstruction becomes the better candidate. Results are different for the limited projection data. In all four cases, it can be seen that the NRMSE again increases when using the MR-SART (except for the limited angle case with $N_{SV} = 2$ and $N_{DS} = 2$) instead of the full SART. However, since the reconstruction error of the FDK is now much higher than that of SART, it appears that the MR-SART can still provide a better reconstruction quality.

In case of Z-splitting the NRMSE is calculated using the central slice in the XZ-plane, the same slice that was used in the evaluation of the Z-splitting previously. Results for these tests are rather different than the ones for the X-splitting.

Table 4.6: The NRMSE of the reconstructed cross-sections and the relative increase as function of different settings for splitting along the Z-axis. Results for the normal projection data are compared with results from limited projection data. The optimal relaxation parameter λ is given for completeness. The multiresolution SART reconstructions (MR-SART) are performed using a 10% overlap between the subvolumes. The increase of the NRMSE relative to the one of the full SART reconstruction is given by δ.

Settings	λ	NRMSE	$\delta(\%)$
Normal data			
Full FDK reconstruction		0.2119	10.646
Full SART reconstruction	0.8	0.1915	
MR-SART: $N_{SV} = 4$, $N_{DS} = 2$	0.8	0.1915	0.004
MR-SART: $N_{SV} = 8$, $N_{DS} = 4$	0.8	0.1935	1.035
Low number of projections			
Full FDK reconstruction		0.3925	37.945
Full SART reconstruction	0.9	0.2846	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.9	0.2846	0.001
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.9	0.2847	0.040
Noise			
Full FDK reconstruction		0.3143	32.104
Full SART reconstruction	0.4	0.2380	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.4	0.2380	0.000
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.4	0.2380	0.000
Limited angle			
Full FDK reconstruction		0.9220	12.819
Full SART reconstruction	0.9	0.8173	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.9	0.8173	0.001
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.9	0.8174	0.014
Irregular angular sampling			
Full FDK reconstruction		0.3535	67.450
Full SART reconstruction	0.9	0.2111	
MR-SART: $N_{SV} = 2$, $N_{DS} = 2$	0.9	0.2111	0.005
MR-SART: $N_{SV} = 4$, $N_{DS} = 4$	0.9	0.2115	0.170

First of all, for the normal data the NRMSE for FDK is considerably higher than for the full SART. This is because SART allows some reduction of cone-beam artefacts as compared to FDK [43, 47], hence the smaller error. More importantly, it can be seen that the reconstruction error hardly increases when the multiresolution approach is used, which means that the reconstruction quality is maintained despite the division into subvolumes. Consequently, MR-SART thus provides a significantly higher reconstruction quality than FDK.

All together, it is not that unexpected that the Z-splitting performs better than the X- or Y-splitting. In parallel- or fan-beam geometry, cross-sections can be

reconstructed independently from each other as the X-rays move parallel to the XY-plane, which means there is no blending of information along the Z-axis. In cone-beam, there is a certain amount of blending, but this is far less than what occurs along the X- or Y-plane due to the rotation of the object. When splitting along the Z-axis, the amount of blending and the resulting artefacts obviously depend on the cone angle of the geometry.

4.16 Applications

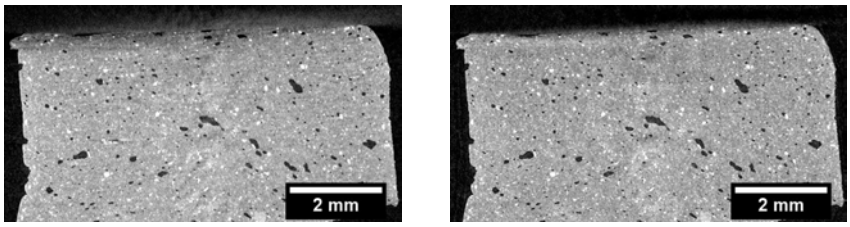
The potential improvement in image quality when using an iterative reconstruction algorithm is illustrated by the following examples of applications. Each of these samples was scanned using the cone-beam geometry where the projection data are acquired over 360°. The GPU implementation of SART presented in section 4.14 was used to reconstruct the volumes and the results are compared to the standard FDK reconstruction.

4.16.1 Local building material

The first sample is a building material created by the reaction of a kaolinitic soil (Jordan Red Ghoul) with a sodium hydroxide (NaOH) solution. The plastic reaction mixture is pressed and cured at 80°C for one day. The kaolonite grains, which have dimensions of the order of 10 µm, are dissolved during the reaction. The purpose of the scan was to see whether the morphology of the reaction product can be visualized. This material is studied in an attempt to produce low-cost building materials with local resources in the Middle-East. Hubert Rahier (Physical Chemistry and Polymer Science, Vrije Universiteit Brussel) is acknowledged for providing this sample.

Table 4.7: Scan settings for the local building material.

# Projections:	1000	
# Detector rows:	940	
# Detector columns:	748	
Pixel pitch:	254	µm
Source to object:	46	mm
Source to detector:	890	mm
Cone angle:	15.3	°
Angular step:	0.36	°
Voxel size:	13	µm



(a) FDK (b) SART

Figure 4.57: Results for the local building material using both FDK and SART. The images show the upper half of the reconstructed sample, with the cross-sections taken parallel to the XZ-plane.

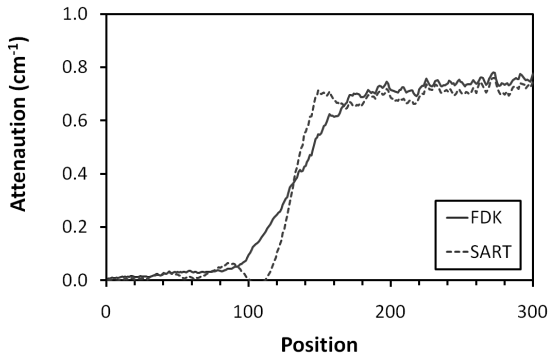


Figure 4.58: Vertical line profiles of the images in figure 4.57, where the profiles were averaged over the width of the sample.

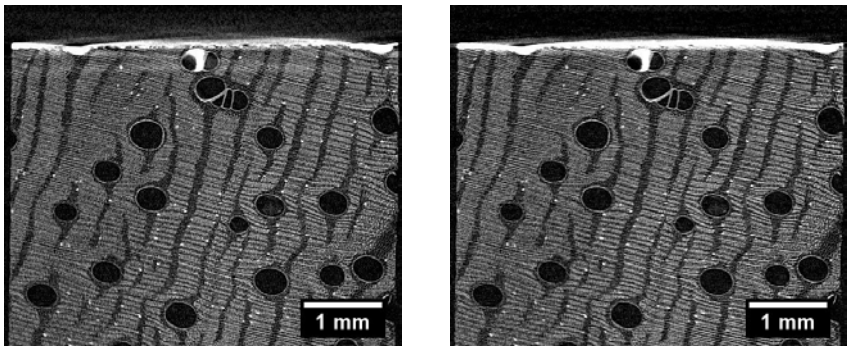
The parameters of the scan are given in table 4.7. The sample was reconstructed in a volume of 940 cross-sections of 748×748 voxels, which took 3 minutes using the FDK algorithm and 12.5 minutes using SART. The resulting reconstructions using both FDK and SART are presented in figure 4.57. As one can see, the FDK reconstruction suffers severely from cone-beam artefacts as the reconstructed values in the top layers of the sample gradually decrease. Instead of a sharp edge, this decrease results in a smooth transition between the top of the sample and the air. As the strongly attenuating inclusions remain visible while the bulk material is dissolving, it looks as if the inclusions are floating in the air. The SART reconstruction suffers less from this decrease in reconstructed values and therefore provides a more realistic representation of the object. In figure 4.58, the vertical line profiles are shown for both reconstructions. It can clearly be seen that the SART reconstruction has a much steeper slope at the transition between the object and the air.

4.16.2 Coated African Padauk

The second application is a piece of African Padauk (wood), which is coated by a layer of paint on top. The paint, mainly containing titanium dioxide, has a much higher attenuation value than the wood and should therefore be clearly visible. The aim of this study consisted of the following aspects: determine the thickness of the paint layer, the intrusion depth of the paint into the wood, the adhesion surface between the paint and the wood and the roughness of the paint surface. Jan Van den Bulcke (Laboratory of Wood Technology, Universiteit Gent) is acknowledged for providing this sample.

Table 4.8: Scan settings for the coated African Padauk.

# Projections:	800	
# Detector rows:	940	
# Detector columns:	748	
Pixel pitch:	254	µm
Source to object:	32	mm
Source to detector:	890	mm
Cone angle:	15.3	°
Angular step:	0.45	°
Voxel size:	9.1	µm



(a) FDK (b) SART

Figure 4.59: FDK and SART reconstructions of the piece of coated African Padauk. The cross-sections are taken parallel to the XZ-plane.

The scan parameters are shown in table 4.8. The results of the reconstructions are shown in figure 4.59, along with the averaged line profiles in figure 4.60. Due to the cone artefacts in the FDK reconstruction the reconstructed values of the paint layer decrease towards the centre of the sample. Below the paint layer, there seems to be a horizontal region with higher reconstructed values, while above the

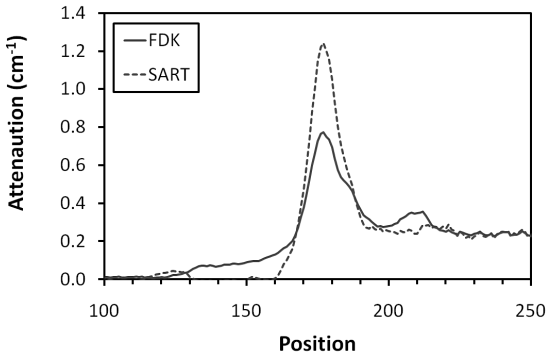


Figure 4.60: Vertical line profiles of the images in figure 4.59, where the profile were averaged over the width of the sample.

layer there is a halo of lower reconstructed values. These regions with different reconstructed values obviously do not represent physical properties of the sample, but are due to cone artefacts. In the SART reconstruction, the reconstructed values of the paint are much higher and more equally spread. The region containing high reconstruction values below the layer is not present and the halo above the layer is much smaller.

4.16.3 Mineral rock

Table 4.9: Scan settings for the mineral rock.

# Projections:	1200	
# Detector rows:	1880	
# Detector columns:	1496	
Pixel pitch:	127	µm
Source to object:	35	mm
Source to detector:	890	mm
Cone angle:	15.3	°
Angular step:	0.3	°
Voxel size:	5	µm

The following example is a piece of rock containing several small, strongly attenuating inclusions and one large mineral inclusion, as can be seen in the projection image in figure 4.61. This large and strongly attenuating inclusion decreases the reconstruction quality by introducing dark streaks in the cross-sections. These streaks are a direct result of the extreme beam hardening that occurs along the ray

paths traversing the inclusion. For such strongly attenuating objects, applying a simple beam hardening correction is not sufficient. Hu Dong (Numerical Rocks) is acknowledged for providing this sample.

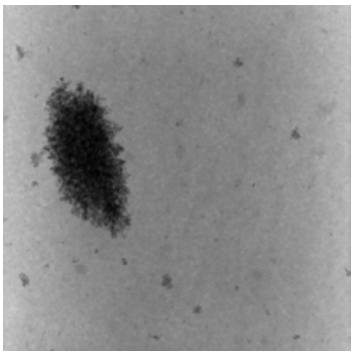
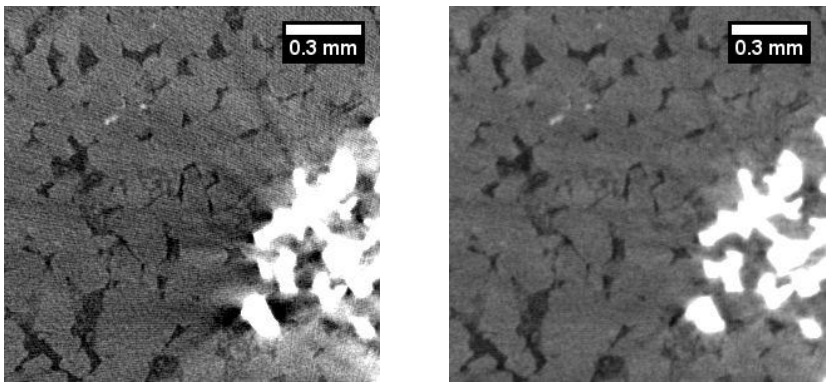


Figure 4.61: Projection image of the mineral rock.



(a) FDK (b) SART

Figure 4.62: Reconstruction of the fragment of mineral rock using the FDK algorithm and SART using the correction for the extreme beam hardening introduced by the strongly attenuating mineral inclusion.

The scan parameters for this scan are given in table 4.9. The results for the reconstruction are shown in figure 4.62. In the SART reconstruction, the correction method presented in section 4.15.7 was used to reduce the dark streaks. In the FDK cross-section, the area around the large inclusion contains a multitude of dark streaks which severely distorts the image. Due to these streaks, it is impossible to detect any of the smaller pores near the inclusion. In the SART cross-section, the dark streaks are significantly reduced and the pores can be detected, allowing for a much more accurate analysis of the scanned sample.

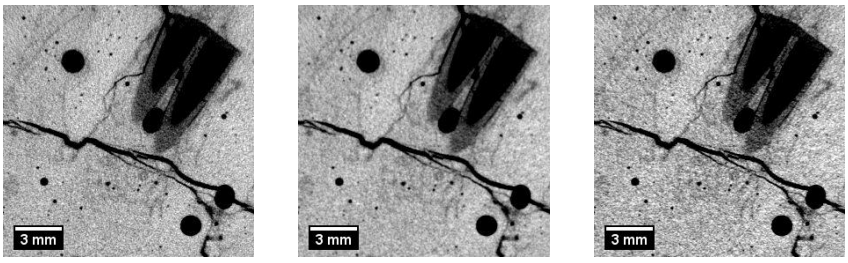
4.16.4 Self-healing cement

This application concerns a piece of self-healing cement. The piece contains two pairs of glass tubes, each filled with one component of a two component glue. By placing the cement in a pressure bench, cracks are induced. Due to the applied pressure, the glass tubes break, allowing the two components to mix, leak into the cracks and set. The aim of the scan was to see which of the cracks are filled by the glue and thus repaired. Kim Van Tittelboom (Magnel Laboratory for Concrete Research, Universiteit Gent) is acknowledged for providing the sample.

Table 4.10 contains the parameters of this scan. Since the reconstruction volume required about 10 GB of storage, it did not fit in the GPU's memory, so for SART it was reconstructed using the multiresolution method presented in section 4.12.2. In order to validate the presented multiresolution approach in practical applications, the projection data and the reconstruction volume were first downsampled by a factor 2. As such, the object can be reconstructed using SART without requiring subvolumes, thus allowing a comparison of the reconstruction quality between FDK, SART and MR-SART. Results of these reconstructions are presented in figure 4.63. It can be seen that the X-splitting introduces additional noise and streaking artefacts in the slice, which is in agreement with the previous phantom study. Contrary, the Z-splitting does not seem to have any influence on the reconstruction quality. The quality of the reconstructed slices can be evaluated quantitatively by measuring the signal to noise ratio (SNR) of a homogeneous region inside the object. This is calculated by SNR = $\bar{f} \,/\, \sigma_f$, where $\bar{f}$ is the average of the reconstructed coefficients f inside the region and σ_f the standard deviation of these values. A higher SNR means that the relative contribution of noise is smaller, which indicates better reconstruction quality. For the X-splitting, the SNR is 17.87 for the FDK reconstruction, 22.48 for SART and 14.08 for MR-SART. The noise level in the reconstruction is thus reduced using SART, while no apparent loss in detail can be observed. The application of the multiresolution approach using X-splitting increases the noise level in the reconstructed cross-sections. For the Z-splitting, the SNR is 22.19 for the FDK, 23.49 for SART and 23.48 for MR-SART, thus showing a similar reconstruction quality for the multiresolution approach.

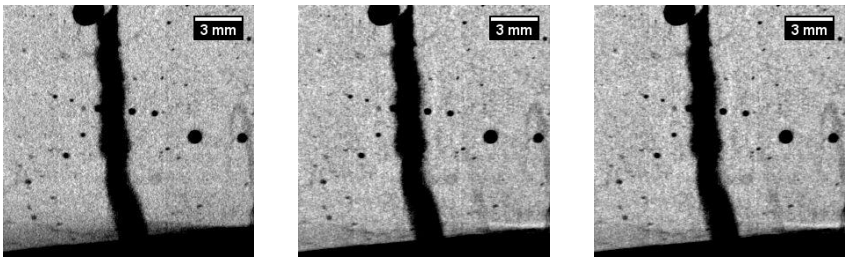
Table 4.10: Scan settings for the self-healing cement.

# Projections:	1000	
# Detector rows:	1880	
# Detector columns:	1496	
Pixel pitch:	127	µm
Source to object:	197	mm
Source to detector:	890	mm
Cone angle:	15.3	°
Angular step:	0.36	°
Voxel size:	28	µm



FDK SART MR-SART

(a) X-splitting



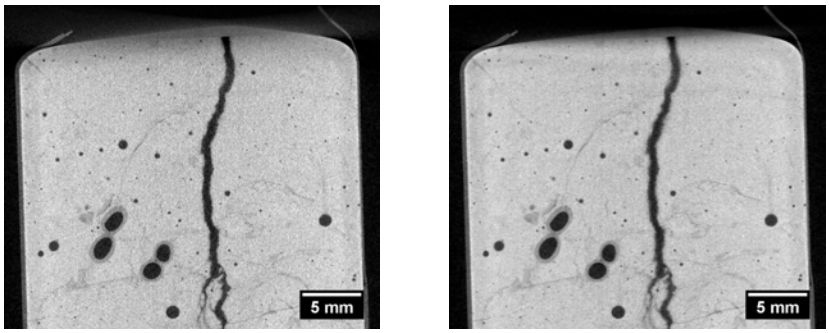
FDK SART MR-SART

(b) Z-splitting

Figure 4.63: Reconstructed cross-sections of a piece of self-healing cement, using FDK, SART and MR-SART with X-splitting (N_{OV} = 10%, N_{SV} = 4, N_{DS} = 2) (a) and Z-splitting (N_{OV} = 10%, N_{SV} = 8, N_{DS} = 2) (b). The images in (a) are taken parallel to the XY-plane, the image in (b) are taken parallel to the XZ-plane.

The object was then reconstructed at the full resolution by applying the multiresolution approach using Z-splitting. The resulting cross-sections for both FDK and MR-SART are shown in figure 4.64. As with previous samples, using FDK the sharp edge of the sample is reconstructed as a smooth transition. This transition is much steeper in the MR-SART reconstruction, making it much more reliable to track the crack up to the surface of the cement. This reduction of the cone-beam artefact can also be clearly observed in figure 4.63(b). In addition, the MR-SART reconstruction also suffers less from noise. To quantify this, the signal to noise ratio (SNR) was measured in a homogeneous region in the upper left corner of the sample in figure 4.64. For the FDK reconstruction the SNR was only 9.68, compared to 19.6 for SART, which means SART provides a much more favourable noise behaviour, while there appears to be no loss of resolution.

It is noted that the SNR for the reconstruction at full resolution is considerably smaller than the one for the downsampled reconstruction, especially for FDK. The downsampling of the projection data is achieved by averaging the measured intensity value over 2×2 neighbouring pixels. The statistical information in the downsampled projections is thus four times larger than the original projections,



FDK MR-SART

Figure 4.64: Reconstructed cross-sections parallel to the XZ-plane of the piece of self-healing cement, using FDK and MR-SART with Z-splitting (N_{OV} = 10%, N_{SV} = 8, N_{DS} = 2).

yielding a reduced noise level. Hence, the SNR of the reconstructed cross-sections increases, be it at a loss of resolution. Since SART was shown to provide a better noise handling especially at higher noise levels, the difference in the SNR between FDK and SART now becomes more apparent in the full resolution reconstruction than in the rebinned reconstruction.

References

[1] S. Kaczmarz. *Angenaherte auflosung von systemen linearer gleichungen*. Bulletin International de l'Academie Polonaise des Sciences Lettres A, 6-8A:355–357, 1937.

[2] K. Tanabe. *Projection method for solving a singular system of linear equations and its applications*. Numerical Mathematics, 17(3):203–214, 1971.

[3] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, 1988.

[4] R. Gordon, R. Bender, and G. T. Herman. *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography*. Journal of Theoretical Biology, 29(3):471–481, 1970.

[5] P. Gilbert. *Iterative methods for the three-dimensional reconstruction of an object from projections*. Journal of Theoretical Biology, 36(1):105–117, 1972.

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

[7] K. Mueller. *Fast and accurate three-dimensional reconstruction from cone-beam projection data using algebraic methods*. Phd dissertation, The Ohio State University, 1998.

[8] M. H. Hudson and R. S. Larkin. *Accelerated image reconstruction using ordered subsets of projection data*. IEEE Transactions on Medical Imaging, 13(4):601–609, 1994.

[9] F. Xu, K. Mueller, M. Jones, B. Keszthelyi, J. Sedat, and D. Agard. *On the efficiency of iterative ordered subset reconstruction algorithms for acceleration on GPUs*. Computer Methods and Programs in Biomedicine, In Press.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. *Maximum likelihood for incomplete data via the EM algorithm*. Journal of the Royal Statistical Society: Series B, 39(1):1–38, 1977.

[11] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans. *Iterative reconstruction for helical CT: a simulation study*. Physics in Medicine and Biology, 43(4):729–737, 1998.

[12] B. De Man. *Iterative reconstruction for reduction of metal artifacts in computed tomography*. Phd dissertation, Katholieke Universiteit Leuven, 2001.

[13] G. T. Herman. *Image reconstruction from projections: the fundamentals of computerized tomography*. Academic Press, New York, 1980.

[14] G. T. Herman and L. B. Meyer. *Algebraic reconstruction techniques can be made computationally efficient*. IEEE Transactions on Medical Imaging, 12(3):600–609, 1993.

[15] H. Guan and R. Gordon. *A projection access order for speedy convergence of ART (algebraic reconstruction technique): a multilevel scheme for computed tomography*. Physics in Medicine and Biology, 39(11):2005–2022, 1994.

[16] M. C. van Dijke. *Iterative methods in image reconstruction*. Phd dissertation, Rijksuniversiteit Utrecht, 1992.

[17] K. Mueller, R. Yagel, and J. F. Cornhill. *The weighted distance scheme: a globally optimizing projection ordering method for ART*. IEEE Transactions on Medical Imaging, 16(2):223–230, 1997.

[18] R. M. Lewitt. *Alternatives to voxels for image representation in iterative reconstruction algorithms*. Physics in Medicine and Biology, 37(3):705–716, 1992.

[19] R. M. Lewitt. *Multidimensional digital image representations using generalized Kaiser-Bessel window functions*. Journal of the Optical Society of America A, 7(10):1834–1846, 1990.

[20] D. Ruijters, B. M. ter Haar Romeny, and P. Suetens. *Efficient GPU-based texture interpolation using uniform B-splines*. Journal of Graphics, GPU, and Game Tools, 13(4):61–69, 2008.

[21] L. Jian, L. Litao, C. Peng, S. Qi, and W. Zhifang. *Rotating polar-coordinate ART applied in industrial CT image reconstruction*. NDT & E International, 40(4):333–336, 2007.

[22] L. Westover. *Footprint evaluation for volume rendering*. Computer Graphics, 24(4):367–376, 1990.

[23] P. M. Joseph. *An improved algorithm for reprojecting rays through pixel images*. IEEE Transactions on Medical Imaging, 1(3):192–196, 1982.

[24] H. Turbell. *Cone-beam reconstruction using filtered backprojection*. Phd dissertation, Linköping University, 2001.

[25] R. L. Siddon. *Fast calculation of the exact radiological path for a three-dimensional CT array*. Medical Physics, 12(2):252–255, 1984.

[26] T. Köhler, H. Turbell, and M. Grass. *Efficient forward projection through discrete data sets using tri-linear interpolation*. In IEEE Nuclear Science Symposium Conference Record, volume 2, pages 113–115, Lyon, 2000.

[27] B. De Man and S. Basu. *Distance-driven projection and backprojection in three dimensions*. Physics in Medicine and Biology, 49(11):2463–2475, 2004.

[28] T. M. Benson and J. Gregor. *Framework for iterative cone-beam micro-CT reconstruction*. IEEE Transactions on Nuclear Science, 52(5):1335–1340, 2005.

[29] Y. Censor and T. Elfving. *Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem*. SIAM Journal on Matrix Analysis and Applications, 24(1):40–58, 2002.

[30] M. Jiang and G. Wang. *Convergence of the simultaneous algebraic reconstruction technique (SART)*. IEEE Transactions on Image Processing, 12(8):957–961, 2003.

[31] J. Wang and Y. Zheng. *On the convergence of generalized simultaneous iterative reconstruction algorithms*. IEEE Transactions on Image Processing, 16(1):1–6, 2007.

[32] G. Qu, C. Wang, and M. Jiang. *Necessary and sufficient convergence conditions for algebraic image reconstruction algorithms*. IEEE Transactions on Image Processing, 18(2):435–440, 2009.

[33] F. Natterer. *The Mathematics of Computerized Tomography*. Wiley, New York, 1986.

[34] F. Boschen, A. Kummert, H. Herzog, and K. Galkowski. *On the convergence of algebraic reconstruction methods for computer tomography*. In The Fourth International Workshop on Multidimensional Systems, pages 172–177, Wuppertal, 2005.

[35] E. Y. Sidky, C.-M. Kao, and X. Pan. *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT*. Journal of X-Ray Science and Technology, 14(2):119–139, 2006.

[36] W. Xu and K. Mueller. *Accelerating regularized iterative CT reconstruction on commodity graphics hardware (GPU)*. In IEEE International Symposium on Biomedical Imaging, pages 1287–1290, Boston, 2009.

[37] L. Kaufman. *Implementing and accelerating the EM algorithm for positron emission tomography*. IEEE Transactions on Medical Imaging, 6(1):37–51, 1987.

[38] A. Ziegler, T. Nielsen, and M. Grass. *Iterative reconstruction of a region of interest for transmission tomography*. Medical Physics, 35(4):1317–1327, 2008.

[39] K. Niinimakï, S. Siltanen, and V. Kolehmainen. *Bayesian multiresolution method for local tomography in dental x-ray imaging*. Physics in Medicine and Biology, 52(22):6663–6678, 2007.

[40] T. Frese, C. A. Bouman, and K. Sauer. *Adaptive wavelet graph model for Bayesian tomographic reconstruction*. IEEE Transactions on Image Processing, 11(7):756–770, 2002.

[41] Y. De Witte, J. Vlassenbroeck, and L. Van Hoorebeke. *A multiresolution approach to iterative reconstruction algorithms in X-ray computed tomography*. IEEE Transactions on Image Processing, Accepted March 4, 2010.

[42] H. Guan and R. Gordon. *Computed tomography using algebraic reconstruction techniques (ARTs) with different projection access schemes: a comparison study under practical situations*. Physics in Medicine and Biology, 41(9):1727–1743, 1996.

[43] S. Valton, F. Peyrin, and D. Sappey-Marinier. *Analysis of cone-beam artifacts in off-centered circular CT for four reconstruction methods*. International Journal of Biomedical Imaging, 2006:1–8, 2006.

[44] I. A. Elbakri and J. A. Fessler. *Statistical image reconstruction for polyenergetic X-ray computed tomography*. IEE Transactions on Medical Imaging, 21(2):89–99, 2002.

[45] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. *Reduction of metal streak artifacts in X-ray computed tomography using a transmission maximum a posteriori algorithm*. IEEE Transaction on Nuclear Science, 47(3):977–981, 2000.

[46] J. Hsieh, R. C. Molthen, C. A. Dawson, and R. H. Johnson. *An iterative approach to the beam hardening correction in cone beam CT*. Medical Physics, 27(1):23–29, 2000.

[47] K. Mueller, R. Yagel, and J. J. Wheller. *Anti-aliased three-dimensional cone-beam reconstruction of low-contrast objects with algebraic methods*. IEEE Transactions on Medical Imaging, 18(6):519–537, 1999.

The most exciting phrase to hear in science,

the one that heralds new discoveries,

is not ’Eureka!’ but ’That’s funny ...’.

Isaac Asimov, 1920° – 1992†

5

Phase Contrast

Conventional X-ray tomography is based on the difference in the attenuation of radiation by different materials. However, when an X-ray beam passes through an object, not only its amplitude but also the phase of the X-rays is modified. Using an appropriate experimental set-up, this phase shift can be measured, which allows obtaining information regarding the refractive index distribution of the object. This phase contrast imaging can offer several advantages over attenuation contrast imaging. At hard X-ray energies, the variations in the refractive index of weakly or moderately attenuating materials are generally several orders of magnitude higher than variations in the attenuation. This means that phase contrast imaging can be used to visualize and quantify objects containing poor attenuation contrast. Furthermore, due to the weak attenuation at higher X-ray energies, the application of phase contrast imaging can be used to reduce the total deposited dose, which is of high importance in medical and in-vivo imaging.

Depending on the experimental conditions, e.g. object composition, X-ray energy, etc., phase contrast is often observed in high resolution projection images. In radiography, the presence of phase contrast can be beneficial since it provides edge enhancement, which improves the detectability of small features. In tomography however, reconstructing such mixed projection images[1] with conventional reconstruction algorithms introduces artefacts in the obtained cross-sections. These artefacts are typically observed as negative attenuation coefficients and wrongly emphasized edges. In this regard, phase contrast can be seen as an artefact that needs

[1]The term *mixed projections* refers to the presence of both an observable attenuation and phase contrast in the projection images.

to be dealt with in order to improve the image quality of conventional high resolution CT systems. This goal can be achieved in two different ways. One way is to extract the phase signal from the mixed projections and use it to reconstruct the refractive index of the object, which corresponds to phase contrast CT. Alternatively, one can remove the phase signal from the projections and then reconstruct the attenuation signal without the appearance of phase artefacts in the cross-sections, which corresponds to conventional attenuation CT. Both approaches are discussed in this chapter, resulting in two complementary techniques that can be applied at any high resolution CT system.

The reader should note that the research presented in this chapter was not performed in order to investigate or improve phase contrast tomography. Instead, the goal was to find a practical solution to the appearance of phase artefacts in attenuation tomography, as this poses one of the main problems in achieving higher resolution. Furthermore, only those solutions were considered that require neither changes in the scanning protocol (e.g. images taken at multiple distances for each rotation angle) nor in the scanner hardware (e.g. analyser crystals, gratings, etc.). This restriction follows from practical considerations, as it is undesirable to increase the acquisition time of a scan or the complexity of the set-up.

5.1 Theory

The technique of phase contrast imaging was first proposed by Gabor [1] back in 1948. Due to its possible advantages over attenuation contrast imaging, it has become a very popular research field in X-ray imaging, especially since the availability of highly coherent synchrotron radiation and the fabrication of X-ray tubes with a small focal spot size. In this section, the basic theory behind phase contrast is given, which should allow the reader to comprehend the phase contrast imaging techniques discussed later on.

5.1.1 Complex refractive index

The complex refractive index of a material for waves with a wavelength λ is given by:

$$n(\lambda) = 1 - \delta(\lambda) + i\beta(\lambda) \; . \tag{5.1}$$

The real part $\Re[n(\lambda)] = 1 - \delta(\lambda)$ corresponds to the phase shift of the propagating wave. It is presented by the deviation $\delta(\lambda)$ from unity[2], which is called the refractive index decrement. The imaginary part $\Im[n(\lambda)] = \beta(\lambda)$ describes the attenuation of the waves and is called the attenuation index. According to [2], $n(\lambda)$ can be given by:

$$n(\lambda) = 1 - \frac{r_e}{2\pi}\lambda^2 \sum_q n_q f_q(0) \; , \tag{5.2}$$

[2]The refractive index of vacuum is 1.

where r_e is the classical electron radius and n_q the number of atoms of type q per unit volume. The complex atomic scattering factor for the forward scattering direction is:

$$f(0) = f_1 + if_2 \ , \tag{5.3}$$

of which the real part f_1 is thus related to the refraction $\delta(\lambda)$ and the imaginary part f_2 to the attenuation $\beta(\lambda)$. These atomic scattering factors have been measured and tabulated for most elements [3]. In figure 5.1, these factors are shown as function of the photon energy for carbon. For higher photon energies[3] the real factor, related to the refractive index and thus the phase, is significantly higher than the imaginary factor, related to the attenuation. In case of graphite, which has an atomic density of $1.136 \cdot 10^{29}$ m^{-3}, for an X-ray energy of 10 keV the refractive index decrement δ is $4.71 \cdot 10^{-6}$ and the attenuation index β is $4.622 \cdot 10^{-9}$.

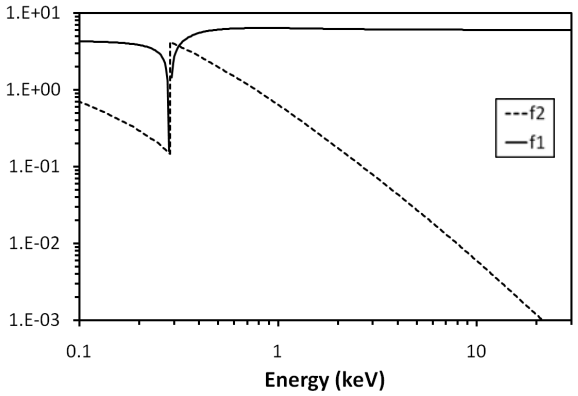


Figure 5.1: The real (f_1) and imaginary part (f_2) of the atomic scattering factor (in arbitrary units) for carbon, as function of the photon energy.

5.1.2 Wave interaction and propagation

Consider a monochromatic plane wave $U_i(x, y)$ of wavelength λ that propagates along the positive Z-axis and is incident upon an object located at $z = z_o$. The distribution of the complex refractive index in this object is given by $n(x, y, z) = 1 - \delta(x, y, z) + i\beta(x, y, z)$. Due to the relative small refractive index decrement for X-rays, the propagation path in the object can be assumed to be straight. The transmitted wave directly behind the object, $U_d(x, y)$ with the propagation distance $d = 0$, is then given by:

$$U_{d=0}(x, y) = T(x, y)U_i(x, y) \ , \tag{5.4}$$

where the interaction of the wave with the object is described by a transmittance function [4]:

$$T(x, y) = \exp[-B(x, y) + i\varphi(x, y)] \ . \tag{5.5}$$

[3]Typical energies for scanning low Z materials at high resolution are between 5 and 30 keV

The attenuation $B(x, y)$ and the phase shift $\varphi(x, y)$ can be considered as projections of the attenuation and refractive index of the object respectively:

$$B(x,y) = \frac{2\pi}{\lambda} \int \beta(x,y,z;\lambda)\, dz\ , \tag{5.6}$$

$$\varphi(x,y) = -\frac{2\pi}{\lambda} \int \delta(x,y,z;\lambda)\, dz\ . \tag{5.7}$$

After transmission through the object, the wave is attenuated by $\exp[-B(x, y)]$ and has experienced a phase shift $\exp[i\varphi(x, y)]$, as illustrated by figure 5.2.

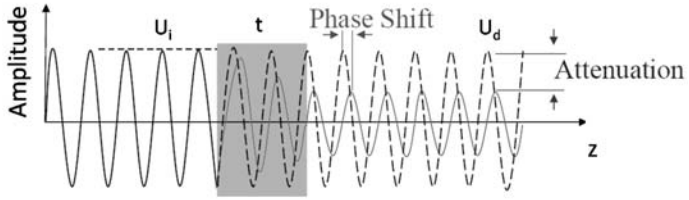


Figure 5.2: The tranmission of a wave U_i through an object of thickness t and constant refractive index $n = 1 - \delta + i\beta$. The transmitted wave U_d is attenuated by a factor $\exp[-(2\pi/\lambda)\beta t]$ and has experienced a phase shift $[-(2\pi/\lambda)\delta t]$. (From [5])

After passing through the object, the wave propagates in free space over a distance d and its intensity forms a so-called Fresnel diffraction pattern. The intensity is given by the squared modulus of the exit wave:

$$I_d(x,y) = I_i(x,y)\, |T(x,y) * P_d(x,y)|^2\ , \tag{5.8}$$

where $*$ denotes convolution and $I_i(x, y) = |U_i(x, y)|^2$ is the initial intensity of the wave. The *Fresnel propagator* $P_d(x, y)$ is defined as:

$$P_d(x,y) = \frac{1}{i\lambda d} \exp\left[i\frac{\pi}{\lambda d}(x^2 + y^2)\right]\ . \tag{5.9}$$

Phase contrast is mainly described by two mathematical models. The first formalism is the simplest one and uses the *ray optical approach* to describe the effect as refraction of X-rays in an object [6, 7]. The second model, the *wave optical approach*, is more rigorous and describes the diffracted X-ray wave field according to the paraxial Fresnel diffraction theory [4, 8]. It is noted that recently a third formalism was presented [9, 10], but this theory will not be discussed here.

5.1.3 Ray optical approach

Tranport of Intensity Equation

Consider a coherent wavefield described by the function:

$$U(\vec{r}) = [I(\vec{r})]^{1/2} \exp[i\varphi(\vec{r})] \,, \tag{5.10}$$

where $\vec{r}$ is the position vector, $I(\vec{r})$ is the intensity of the wave and $\varphi(\vec{r})$ its phase. The surfaces at which $\varphi(\vec{r})$ is constant are represented by wavefronts, the local propagation direction is given by the phase gradient $\nabla\varphi(\vec{r})$, as illustrated in figure 5.3.

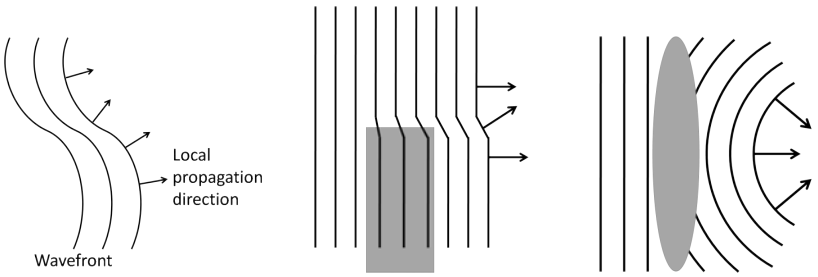


Figure 5.3: Illustration of the wavefronts and local propagation directions (a), for a wave passing through a straight interface (b) and an optical lens (c).

When this wave is propagating along the positive Z-axis of an orthogonal coordinate system XYZ, the wave amplitude can be given by:

$$U_z(\vec{r}_\perp) = [I_z(\vec{r}_\perp)]^{1/2} \exp[i\varphi_z(\vec{r}_\perp)] \,, \tag{5.11}$$

where $\vec{r}_\perp = (x, y)$ is a two-dimensional vector in the transverse direction. The amplitude $U_z(\vec{r}_\perp)$ then satisfies the parabolic equation [11]:

$$\left(i\frac{\partial}{\partial z} + \frac{\nabla_\perp^2}{2k} + k \right) U_z(\vec{r}_\perp) = 0 \,, \tag{5.12}$$

where $\nabla_\perp^2 = [(\partial^2/\partial x^2) + (\partial^2/\partial y^2)]$ and the angular wavenumber $k = 2\pi/\lambda$.

Now let equation (5.12) be multiplied by U_z^* on the left-hand side and let the complex conjugate of equation (5.12) be multiplied by U_z on the left-hand side. By subtracting the two resulting equations, the following expression, known as the *Transport of Intensity Equation (TIE)* [12] is derived:

$$\frac{2\pi}{\lambda}\frac{\partial}{\partial z} I_z(\vec{r}_\perp) = -\vec{\nabla}_\perp \left[I_z(\vec{r}_\perp) \vec{\nabla}_\perp \varphi_z(\vec{r}_\perp) \right] \,, \tag{5.13}$$

This equation directly relates the variations in intensity of the wave in the direction of the propagation to the intensity and phase in a plane perpendicular to the propagation. Since the phase changes when the wave passes through a medium, variations in the refractive index of the object result in a change of intensity with the wave propagation. Intuitively, the connection between the phase and intensity of a wave can be explained as follows, illustrated by figure 5.3. When a wave passes through a medium, the obtained phase difference deforms the wavefront[4]. Due to this deformation, the wave intensity is redistributed on a vertical plane at some distance behind the object.

As an example, consider the two-dimensional case of a circular object with radius R and a homogeneous refractive index $n = 1 - \delta + i\beta$. According to equation (5.7), the phase shift of the transmitted wave relative to the wave in vacuum is given by:

$$\varphi(x) = -\frac{4\pi}{\lambda}\delta\sqrt{R^2 - x^2}\ . \tag{5.14}$$

The phase gradient is then given by:

$$\frac{\partial\varphi(x)}{\partial x} = \frac{4\pi}{\lambda}\delta\frac{x}{\sqrt{R^2 - x^2}}\ , \tag{5.15}$$

which contains singularities at $x = R$, thus at the edge of the object, which are due to the rapid change in the refractive index. This is illustrated in figure 5.4. Looking at the TIE (equation (5.13)), high values of $\vec{\nabla}_\perp\varphi_z(\vec{r}_\perp)$ result in strong variations in the intensity along the propagation direction.

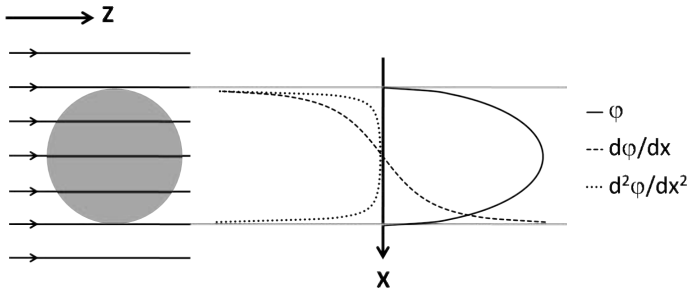


Figure 5.4: The transmission of a coherent wave through a circular object. The displayed profiles correspond to the phase difference φ after transmission, the gradient $\partial\varphi/\partial x$ of the phase in the plane transversal to the propagation and the second derivative $\partial^2\varphi/\partial x^2$ of the phase.

[4]As opposed to visible light, the refractive index for X-rays is smaller than 1 for most materials. Consequently, the part of the wavefront passing through the material moves ahead of the undisturbed part.

Edge enhancement

It is now shown how the variations in intensity are visualized. The local propagation vector $\vec{k'}(\vec{r})$, which is perpendicular to the wavefront at position $\vec{r}$, is in the paraxial approximation given by:

$$\vec{k'}(\vec{r}) = \left[\frac{\partial \varphi(\vec{r})}{\partial x}, \frac{\partial \varphi(\vec{r})}{\partial y}, k_z \right] . \tag{5.16}$$

The angular deviation between $\vec{k'}$ and the original propagation vector $\vec{k}$ can then be expressed as:

$$\Delta\alpha \approx \frac{\lambda}{2\pi} |\vec{\nabla}_\perp \varphi(\vec{r})| . \tag{5.17}$$

For typical X-ray energies and light materials used in high resolution CT, $\Delta\alpha$ is of the order of a few microradians. The interpretation of equation (5.17) is that areas which contain a strong phase gradient, such as edges in the object, introduce a deformation of the wavefront. Consequently, rays close to an edge are deflected from their original path, resulting in a loss of intensity in the forward direction, while in other directions intensity increases (see figure 5.5). This refraction process causes so-called edge enhancement in the measured projection image.

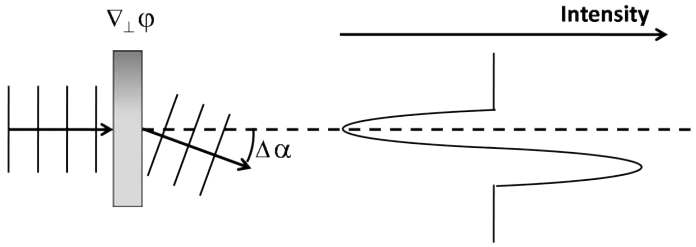


Figure 5.5: Deflection of a coherent wave from its propagation direction due to a phase gradient. The measured intensity shows a deficit along the forward direction of the deflected rays, while showing an increase elsewhere. (From [13])

To illustrate this, consider again the 2D example of the circular object. Using equation (5.17), the angular deviation then becomes:

$$\Delta\alpha = \frac{\lambda}{2\pi} \left| \frac{\partial \varphi(x)}{\partial x} \right| = \delta \frac{2x}{\sqrt{R^2 - x^2}} . \tag{5.18}$$

Since the phase gradient diverges at $x = R$, the rays deviate by a large angle, even though δ is small. This leads to an observable redistribution of the intensity in the corresponding forward direction, thus at boundaries or edges. This explains why a radiography of an object looks like a direct image of the contours of the object and the features within. The edge enhancement effect is typically observed as an upward and downward peak in the intensity profile, as illustrated in figure 5.6 for the circular object. In general, any rapid variation in the refractive index or

the thickness of an object is imaged by the edge effect. Furthermore, the position of the loss in intensity is essentially independent of the wavelength, so the edge enhancement even occurs when a polychromatic X-ray beam is used.

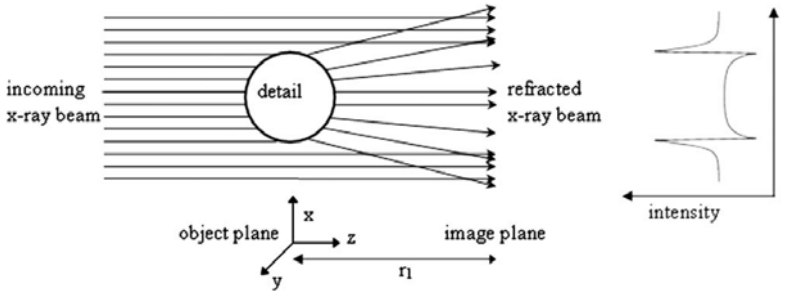


Figure 5.6: Schematic display of the mechanism for phase contrast formation for a circular object. The intensity is measured at a distance r_1 from the object. The intensity profile only represents the variations due to phase contrast, attenuation of X-rays by the object was ignored. (From [7])

TIE for weak attenuation

The TIE (equation (5.13)) can be simplified in case the attenuation by the object is weak. The TIE for weak attenuation forms the basis for many phase contrast reconstruction methods, including the phase retrieval and reduction methods discussed in this work. Even though this derivation is usually performed using wave equations, it can be derived from the ray optical approach as well, using the angular deviation [7].

First, assume that the free-space propagation distance d (the distance between the object and the image plane) is large compared to the object size. At first, attenuation of the X-rays by the object is ignored. The X-ray beam is a monochromatic, parallel beam of initial intensity $I_i(x, y)$, with each X-ray propagating parallel to the Z-axis. After impinging on the object in the point $(x, y, 0)$, each single X-ray undergoes an angular deviation $\alpha(x, y)$. Hence, it reaches the image plane at the point (x_d, y_d, d), given by:

$$\begin{aligned} x_d &\approx x + d\,\alpha_x(x, y) \\ y_d &\approx y + d\,\alpha_y(x, y) \end{aligned} \tag{5.19}$$

where it is assumed that $\sin\alpha \approx \alpha$ since the magnitude of the deviations is of the order of microradians. The angles $\alpha_x(x, y)$ and $\alpha_y(x, y)$ represent the projection of $\alpha(x, y)$ on the XZ-plane and the YZ-plane, respectively, and are given by:

$$\begin{aligned} \alpha_x(x, y) &\approx \frac{\lambda}{2\pi}\frac{\partial\varphi(x, y)}{\partial x} \\ \alpha_y(x, y) &\approx \frac{\lambda}{2\pi}\frac{\partial\varphi(x, y)}{\partial y} \end{aligned} \tag{5.20}$$

Equation (5.19) has only one solution, meaning there is a unique ray connecting the points $(x, y, 0)$ and (x_d, y_d, d). This allows the intensity in the image plane $I_d(x_d, y_d)$ to be written as [14]:

$$\begin{aligned} I_d(x_d, y_d) &= I_i(x, y) \left| \frac{\partial(x_d, y_d)}{\partial(x, y)} \right|^{-1} \\ &= I_i(x, y) \left| \begin{matrix} 1 + d\frac{\partial \alpha_x(x,y)}{\partial x} & d\frac{\partial \alpha_x(x,y)}{\partial y} \\ d\frac{\partial \alpha_y(x,y)}{\partial x} & 1 + d\frac{\partial \alpha_y(x,y)}{\partial y} \end{matrix} \right|^{-1} \end{aligned} \tag{5.21}$$

Since $d\lambda \ll 1$[5], products of the partial derivatives can be neglected, so the following expression is obtained:

$$I_d(x_d, y_d) \approx I_i(x, y) \left[1 + \frac{d\lambda}{2\pi} \nabla_\perp^2 \varphi(x, y) \right]^{-1} . \tag{5.22}$$

The positions x_d and y_d can be replaced by x and y, since $d\alpha_x/x \ll 1$ and $d\alpha_y/y \ll 1$. For small variations of the phase, $(d\lambda/2\pi)\nabla^2\varphi(x, y) \ll 1$, this then becomes:

$$I_d(x, y) \approx I_i(x, y) \left[1 - \frac{d\lambda}{2\pi} \nabla_\perp^2 \varphi(x, y) \right] . \tag{5.23}$$

Next, consider the case of a monochromatic, conical beam emanating from a point source, located at $x = 0$, $y = 0$ and at a distance s from the object. This results in an image magnification $M = (s+d)/s$. The equations (5.19) describing the X-ray path to the image plane are now written as:

$$\begin{aligned} x_d &\approx M\, x + d\, \alpha_x(x, y) \\ y_d &\approx M\, y + d\, \alpha_y(x, y) \end{aligned} \tag{5.24}$$

By a derivation similar to the one above, the following expression can then be obtained:

$$I_d(x_d, y_d) \approx \frac{I_i(x, y)}{M^2} \left[1 + \frac{d\lambda}{2\pi M} \nabla_\perp^2 \varphi(x, y) \right]^{-1} . \tag{5.25}$$

The positions x_d and y_d can be replaced by $M\, x$ and $M\, y$, since $(d\, \alpha_x)/(M\, x) \ll 1$ and $(d\, \alpha_y)/(M\, y) \ll 1$. The intensity $I_{d,i}(x, y) = I_i(x, y)/M^2$ is the intensity that would have reached the image plane in the absence of the object. By approximation to the first order in $(d\lambda/2\pi M)\nabla^2\varphi(x, y) \ll 1$, this becomes:

$$I_d(Mx, My) \approx I_{d,i}(x, y) \left[1 - \frac{d\lambda}{2\pi M} \nabla_\perp^2 \varphi(x, y) \right] . \tag{5.26}$$

The previous expression was obtained for a pure phase object. In order to include attenuation by the object, the unattenuated intensity $I_{d,i}(x, y)$ is replaced

[5] In high resolution CT, $d\lambda \approx 10^{-10}$ for typical X-ray energies of 12 keV and a propagation distance of 1 m.

by the attenuated intensity $I_{d,i}(x,y)\exp[-2B(x,y)]$ [7], with $B(x,y)$ given by equation (5.6). Finally, the following expression is derived:

$$I_d(Mx,My) \approx I_{d,i}(x,y)e^{-2B(x,y)}\left[1-\frac{d\lambda}{2\pi M}\nabla_\perp^2\varphi(x,y)\right] . \tag{5.27}$$

This equation (5.27) is known as the *Transport of Intensity Equation for weak attenuation* and will be denoted as the *TIEW*. The TIEW describes the redistribution of the transmitted intensity as the product of the attenuation term and a phase sensitive term, determined by the Laplacian of the phase shifts. From equation (5.27), it can be seen that the phase contrast signal increases as the propagation distance d increases. The signal also depends on the rapidity of the phase shift variations induced by the object. Furthermore, it is obvious that in the contact plane (the plane just behind the object, $d = 0$) the phase signal is not present and only the attenuation of the X-ray beam by the object is detected. The phase contrast signal only becomes observable at nonzero distances.

5.1.4 Wave optical approach

The phenomenon of phase contrast can also be addressed in terms of Fresnel diffraction. This diffraction pattern is due to the interference of the refracted and the undisturbed wave. The Fresnel diffraction theory [4] is more elaborate than the ray optical approach and provides a more rigorous description of the phase contrast effect.

TIEW

First, it is shown how the Transport of Intensity Equation can be derived using the wave approach. Consider again equations (5.8) and (5.9):

$$I_d(\vec{r}_\perp) = I_i(\vec{r}_\perp)|T(\vec{r}_\perp) * P_d(\vec{r}_\perp)|^2 , \tag{5.28}$$

$$P_d(\vec{r}_\perp) = \frac{1}{i\lambda d}\exp\left[i\frac{\pi}{\lambda d}|\vec{r}_\perp|^2\right] , \tag{5.29}$$

with $\vec{r}_\perp = (x,y)$. These equations describe the recorded intensity $I_d(\vec{r}_\perp)$ after propagation of the wave over a distance d, expressed by the Fresnel propagator $P_d(\vec{r}_\perp)$. The Fourier transform of a function $g(\vec{r}_\perp)$ is defined as:

$$\tilde{g}(\vec{f}) = F\{g\}(\vec{f}) = \int g(\vec{r}_\perp)e^{-i2\pi\vec{r}_\perp\cdot\vec{f}}\,d\vec{x} , \tag{5.30}$$

with $\vec{f} = (u,v)$ the spatial frequencies and $\cdot$ the scalar product. The Fourier transform of the Fresnel propagator can then be written as:

$$\tilde{P}_d(\vec{f}) = \exp(-i\pi\lambda d|\vec{f}|^2) . \tag{5.31}$$

The convolution in equation (5.28) can then be calculated in the Fourier domain. Consequently, the Fresnel diffraction pattern can be calculated in the Fourier space as [15]:

$$\tilde{I}_d(\vec{f}) = \int I_i(\vec{r}_\perp)\, T\left(\vec{r}_\perp - \frac{\lambda d}{2}\vec{f}\right)\, T^*\left(\vec{r}_\perp + \frac{\lambda d}{2}\vec{f}\right)\, e^{-i2\pi\vec{r}_\perp\cdot\vec{f}}\, d\vec{x}\,. \quad (5.32)$$

In the contact plane $d = 0$, the intensity is determined only by the attenuation of the object and can be written as:

$$I_{d=0}(\vec{r}_\perp) = I_i(\vec{r}_\perp)|T(\vec{r}_\perp)|^2 = I_i(\vec{r}_\perp)\exp[-2B(\vec{r}_\perp)]\,. \quad (5.33)$$

This means that by measuring the intensity in the contact plane, the attenuation of the X-ray beam by the object is fully determined, since no phase signal is present. At distances $d > 0$, the phase contrast becomes apparent, resulting in a mixed intensity profile, containing both the attenuation and the phase shift introduced by the object.

The Taylor expansion of $T(\vec{r}_\perp \pm \lambda d\vec{f}/2)$ with respect to d and retaining only the first order terms results in:

$$T\left(\vec{r}_\perp \pm \frac{\lambda d}{2}\vec{f}\right) = T(\vec{r}_\perp) \pm \frac{\lambda d}{2}\vec{f}\cdot\vec{\nabla}_\perp T(\vec{r}_\perp)\,. \quad (5.34)$$

After substitution in equation (5.32), the intensity can be expressed as [15]:

$$I_d(\vec{r}_\perp) = I_{d=0}(\vec{r}_\perp) - \frac{\lambda d}{2\pi}\,\vec{\nabla}_\perp[I_{d=0}(\vec{r}_\perp)\,\vec{\nabla}_\perp\varphi(\vec{r}_\perp)]\,. \quad (5.35)$$

Under the assumption that $B(\vec{x})$ only varies insignificantly, thus if it is close to constant, $I_{d=0}(\vec{x})$ can be moved outside the gradient operator, which leads to:

$$I_d(\vec{r}_\perp) = I_{d=0}(\vec{r}_\perp)\left[1 - \frac{\lambda d}{2\pi}\,\nabla^2_\perp\varphi(\vec{r}_\perp)\right]\,. \quad (5.36)$$

In case of a conical beam, the magnification M is introduced in the Fresnel propagator [7]:

$$P_d(\vec{r}_\perp) = \frac{M}{i\lambda d}\exp\left[i\frac{\pi}{\lambda d}|\vec{r}_\perp|^2\right]\,, \quad (5.37)$$

which after a similar derivation yields:

$$I_d(\vec{r}_\perp) = I_{d=0}(\vec{r}_\perp)\left[1 - \frac{\lambda d}{2\pi M}\,\nabla^2_\perp\varphi(\vec{r}_\perp)\right]\,, \quad (5.38)$$

where $I_{d=0}(\vec{r}_\perp) = I_i(\vec{r}_\perp)\exp[-2B(\vec{r}_\perp)]$ is the intensity in the image plane when no phase contrast is present. This formula again represents the Transport of Intensity Equation for Weak attenuation, and is equal to the result (equation (5.27)) derived using the ray optical approach.

Diffraction pattern

It is now shown how a diffraction pattern can be calculated using the wave optical approach. Consider a coherent point source, located at the coordinates (x_s, y_s). This source emits radiation in the form of spherical waves, which according to Maxwell's equation can be written as:

$$E_0(x,y) = \frac{1}{s+d}\,\exp\left[\frac{2\pi i}{\lambda}\,\left(s+d+\frac{(x-x_s)^2}{2(s+d)}+\frac{(y-y_s)^2}{2(s+d)}\right)\right]\,, \quad (5.39)$$

where (x, y) is a point in the image plane. The distance between the source and the object is given by s, d is the distance between the object and the image plane. After the wave is transmitted through the object, it has experienced a phase shift. Now consider the case where the change of phase only takes place in the X-direction. This corresponds to linear objects, such as a cylinder oriented along the Y-axis. Using the standard Kirchoff integral [4], the spherical wave in the image plane can then be expressed by [16]:

$$\begin{aligned} E(x,y) &= \frac{1}{\sqrt{i\lambda sd(s+d)}}\exp\left[\frac{2\pi i}{\lambda}\,\left(s+d+\frac{(y-y_s)^2}{2(s+d)}\right)\right] \\ &\quad\cdot\int d\chi\,\exp\left[\frac{2\pi i}{\lambda}\left(\frac{(\chi-x_s)^2}{2s}+\frac{(\chi-x)^2}{2d}\right)\right] \\ &\quad\cdot\,\exp[i\varphi(\chi)]\,. \end{aligned} \quad (5.40)$$

This formula allows calculating the diffraction pattern of an arbitrary phase object. Now consider the case of an object of finite size in the range $-R < \chi < R$, thus $\varphi(\chi) = 0$ if $|\chi| > R$. For convenience, equation (5.40) is usually rewritten in the form:

$$E(x,y) = E_0(x,y)[1+c(x)]\,, \quad (5.41)$$

with:

$$\begin{aligned} c(x) &= \sqrt{\frac{s+d}{i\lambda sd}}\exp\left[\frac{2\pi i}{\lambda}\,\left(-\frac{(x-x_s)^2}{2(s+d)}\right)\right] \\ &\quad\cdot\int_{-R}^{R} d\chi\,\exp\left[\frac{2\pi i}{\lambda}\left(\frac{(\chi-x_s)^2}{2s}+\frac{(\chi-x)^2}{2d}\right)\right] \\ &\quad\cdot\,\{\exp[i\varphi(\chi)]-1\}\,. \end{aligned} \quad (5.42)$$

The intensity of the wave in the image plane, normalized to the background, is then given by:

$$I(x) = \left|\frac{E(x,y)}{E_0(x,y)}\right|^2 = 1 + 2\,\Re[c(x)] + |c(x)|^2\,. \quad (5.43)$$

Thus, the diffraction pattern can be simulated by calculating the values of $c(x)$ for every point x in the image plane. From equation (5.43), it is possible to determine the distance $d_{optimal}$ between the object and the image plane that maximizes the phase contrast [17]:

$$d_{optimal} = \frac{R^2}{4\lambda} \,. \tag{5.44}$$

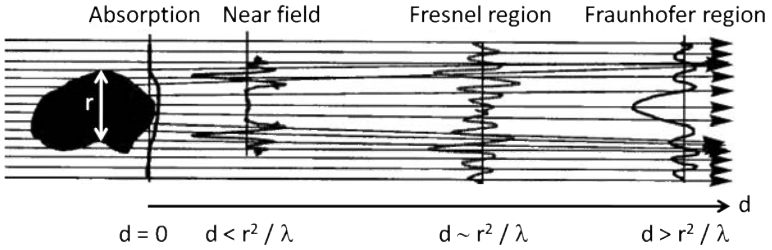
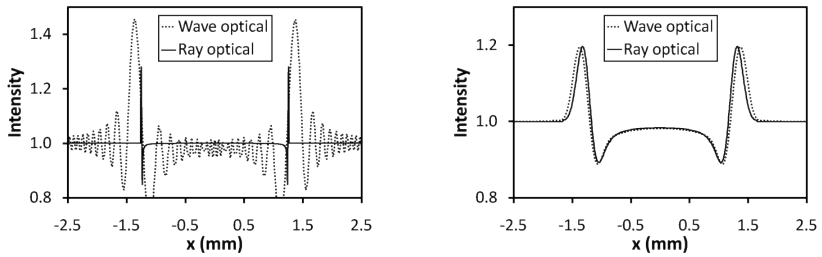


Figure 5.7: The diffraction pattern as function of the propagation distance of the transmitted wave. (From [18])

The diffraction pattern strongly depends on the propagation distance of the transmitted wave, as illustrated in figure 5.7. Consider an object of size r and assume that the coherence of the incoming wave is sufficient. In the contact plane $d = 0$ right behind the object, the detected intensity is only due to attenuation. In the near-field $d < r^2/\lambda$ phase effects are observable as edge enhancement. With increasing propagation distance, in the Fresnel region $d \sim r^2/\lambda$, the refracted waves start to interfere and the imaged object is no longer recognisable. Even further from the object, in the Fraunhofer region $d > r^2/\lambda$, the image becomes the squared modulus of the Fourier transform of the object. In high resolution systems based on an X-ray tube, the projection images are usually detected in the near field, thus in the edge enhancement regime.

In practice, the ideal intensity profile calculated by equation (5.42) is never observed, due to the influence of the finite spot size of the X-ray source, the spectral distribution of the X-rays, limited resolution of the detector, etc. To account for a polychromatic beam, the simulated intensity can be calculated as a weighted sum of the intensities simulated for the monochromatic components of the beam. The finite spot size and imperfect detector can be introduced by convolution with a point spread function [5, 7].

In figure 5.8, results are shown for the simulation of the phase contrast signal using the wave optical and the ray optical approach. The wave optical simulation was performed using equation (5.42). For the ray optical approach, results were obtained by applying the TIEW (equation (5.27)). The imaged object is a homogeneous cylinder with a radius of 50 µm. The source to object and source to detector distances were 1 m and 25 m, respectively. The profiles were sampled at an interval of 8 µm, with a photon energy of 9 keV. The simulation was performed using cone-beam geometry. In figure 5.8(a), the case of an ideal point source is



(a) Ideal point source (b) Finite spot size

Figure 5.8: Simulation of phase contrast using the wave optical and the ray optical approach. The object is a homogeneous cylinder of radius 50 µm. Attenuation of the X-rays by the object was ignored. The relative intensity is plotted as function of the position in the image plane. Results are shown for an ideal point source (a) and for a finite spot size (b) by convolving the result with a Gaussian point spread function with a 10 µm FWHM.

assumed. For the wave optical approach, the phase contrast signal consists of a succession of minima and maxima, with the most intense maxima corresponding to the edges of the cylinder. Using the ray optical approach, only one minimum and maximum is found for each edge. It is obvious that the profiles obtained using the two different models are quite different. In practice, the X-rays emerge from a spot of finite size and the intensity profiles look rather different, since the high frequency components of the profiles are now cancelled out. This is illustrated in figure 5.8(b), which shows the same profiles after convolution with a Gaussian point spread function with a 10 µm FWHM[6]. Both resulting profiles now contain only one minimum and maximum for each edge and are in good agreement.

5.1.5 Spatial coherence

An important requirement for phase contrast imaging is using an X-ray beam with high spatial coherence, meaning that the amplitudes of the waves are highly correlated between different points transverse to the direction of propagation. The *transverse coherence length* $d_\perp$ is given by [19]:

$$d_\perp = \lambda \frac{s}{\sigma} \,, \tag{5.45}$$

where λ is the wavelength of the X-rays, s is the distance from the source to the object and σ is the spot size of the X-ray source, which is assumed to be incoherent. The transverse coherence length can thus be increased by increasing the distance between the source and the object or by decreasing the spot size, provided the flux of the X-ray beam reaching the detector plane is still sufficient. For a typical high resolution set-up such as those used at UGCT, the transverse

[6]Full width at half maximum

coherence length is of the order of 1 µm. At a synchrotron facility, it is typically of the order 100 µm, which explains why phase contrast imaging is usually done using synchrotron radiation.

5.1.6 Temporal coherence

Temporal coherence is related to how monochromatic the emitted radiation is. Even though a certain temporal coherence is required to observe phase contrast, this restriction is much less severe than the spatial coherence. In most X-ray imaging systems, temporal coherence does not pose a problem, meaning that the use of polychromatic radiation is allowed.

The influence of a polychromatic beam can be evaluated by returning to equation (5.17). A polychromatic beam leads to a range of angular deviations, depending on the energy of the X-rays. As a result, the loss in intensity in the corresponding forward direction is still present and clearly outlined, so the downward peak in the intensity profile remains sharp. The upward peak, corresponding to a gain in intensity, on the other hand is broader, but less high[7] (see figure 5.9). So even though a polychromatic beam is used, the variations in refractive index of the object can still be observed in the intensity image.

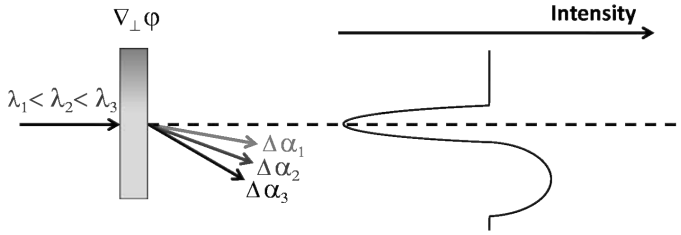


Figure 5.9: Deflection of a polychromatic wave from its propagation direction due to a phase gradient. The wave consists of a superposition of three waves with different wavelengths. The measured intensity shows a sharp deficit along the forward direction of the deflected rays, while showing a broad increase elsewhere. (From [13])

Next, consider the influence of polychromatic radiation on the TIEW. By using the formulas describing the atomic scattering factors (5.2) and (5.3), equation (5.7) can be rewritten as [8]:

$$\varphi(x,y) = -\lambda r_e \int \rho(x,y,z)\, dz \,, \tag{5.46}$$

where the electron density $\rho(x,y,z)$ represents the real part of the summation $\sum_q n_q f_q(0)$ of the atomic scattering factors. The atomic scattering factor f_1 for each element can generally be considered a constant within the typical energy

[7] The total areas comprised by the upward and downward peak are equal.

range used in high resolution imaging (see e.g. figure 5.1). Therefore, $\rho(x, y, z)$ can be assumed to be independent of the wavelength λ. Inserting equation (5.46) in (5.27) results in:

$$I_d(x, y) \approx I_{d,i}(x, y) e^{-2B(x,y)} \left[1 + \lambda^2 \frac{d\, r_e}{2\pi M} \nabla_\perp^2 \left(\int \rho(x, y, z)\, dz \right) \right] . \quad (5.47)$$

The structure of the intensity image is thus proportional to the projected electron density and is to the first order independent of the wavelength [8], since for a polychromatic beam the factor λ^2 is simply replaced by a weighted sum of the spectrum.

Combining this with the discussion on the transverse coherence length, the following can be concluded. In a first approximation, the TIEW can be used to represent the intensity image of a weakly attenuating object, acquired using the polychromatic beam of an X-ray tube with a sufficiently small focal spot size. This conclusion and its experimental validation were first reported by Wilkins *et al.* [8], which heralded the start of phase contrast imaging using X-ray tubes.

5.2 Phase contrast generation

Several methods exist to generate contrast from the phase shift of the X-rays. These can be classified into methods using an interferometer [20, 21], analyser methods [6, 22], grating-based methods [23, 24] and free-space propagation methods [8, 16, 25, 26].

Interferometer: In a set-up using an interferometer, any phase modulation is directly converted to an intensity modulation after interference with the reference beam. This requires a highly parallel and monochromatic beam. Although being one of the oldest techniques for phase contrast imaging, its practical application is limited due to the very delicate construction of the interferometer.

Analyser: An analyser crystal is placed between the object and the detector, which only transmits plane waves propagating in a predetermined direction. Since the object introduces distortions to the waves, only those planar sections of the wave propagating in the right direction are transmitted. By changing the angle of the analyser, different planar sections of the distorted wavefront can be imaged, leading to a series of images containing contrast. This technique again requires a highly parallel and monochromatic beam.

Gratings: A grating consists of a thin plate with transmitting slits placed at a fixed interval. Typically, one phase grating is used as a beam splitter and a second attenuation grating as an analyser. By bringing these two grids into the beam, the angular deviations introduced by the object can be translated into changes of the locally transmitted intensity. This detected intensity can be used as a direct measure of the object's local phase gradient. The grating-based methods can be applied

to polychromatic radiation from an incoherent source by installing an additional third grating. Furthermore, they yield separate attenuation and phase contrast images. Acquisition is performed using the phase-stepping approach, which requires one of the gratings to be scanned along the transverse direction over one period of the grating, taking an image for every point of the scan.

Free-space propagation: As was already mentioned in the theoretical discussion, phase contrast can be achieved by simply letting the beam propagate in free-space, provided the spatial coherence of the X-ray beam is sufficient. This method is equivalent to in-line holography [1]. After allowing for sufficient propagation, the phase modulation is transformed in an amplitude modulation, which can be detected as intensity contrast. This results in mixed projection images, which contain both attenuation and phase information. Phase contrast in these mixed projection images is typically observed as enhancement of the edges of the object, thus where the phase gradient is large. Conversely, this means that phase contrast may appear at any high resolution set-up, even when no additional optics are installed.

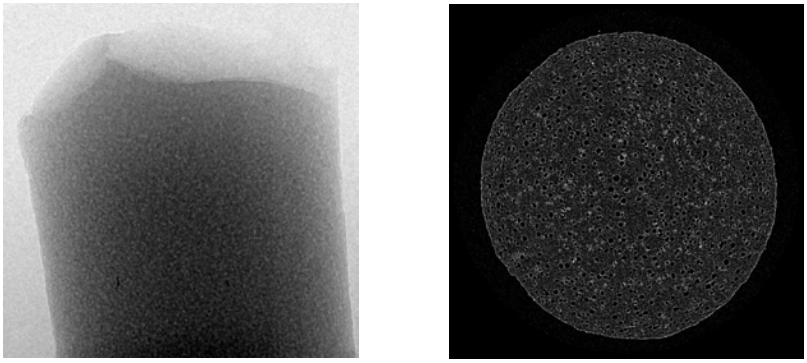
In the case of high resolution X-ray CT, phase contrast thus emerges from free-space propagation. Therefore, the remainder of this chapter will only handle phase contrast due to free-space propagation. Even though other methods can be applied to an X-ray tube as well, these will not be considered since the goal of this work was to find solutions to the appearance of phase contrast in conventional high resolution scans, thus without modifying the scanner protocol or hardware.

5.3 Reconstruction of mixed projections

When using an X-ray source with a small focal spot size, the transverse coherence length is sufficiently large for phase contrast, due to free-space propagation, to be observed in the acquired projection images. This can happen at a synchrotron facility as well as at a tube based set-up. Even though they are not specifically constructed to utilize phase contrast, every high resolution system may inevitably detect a phase signal: *nature does not provide us with a magical button to switch off phase contrast*. As a result, depending on the experimental conditions, every high resolution CT set-up will acquire mixed projection images, containing both attenuation and phase contrast.

Although this phase signal may be beneficial in radiography to enhance the detectability of small features, the tomographic reconstruction of mixed projection images introduces artefacts in the cross-sections when the phase signal is not properly dealt with. Typical manifestations of these phase contrast artefacts are negative attenuation coefficients and wrongfully emphasized edges. This is illustrated in figure 5.10, showing a (mixed) projection image of a small tablet and a cross-section obtained using a standard reconstruction algorithm. The effect of the phase signal can clearly be observed at the edges of the tablet, showing a strong variation in the reconstructed attenuation. This variation consists of a steep increase in

attenuation, followed by a decrease with attenuation coefficients dropping below zero. This profile, visualized as a black and white border, is a typically observed phase artefact. Similar artefacts can be found surrounding the pores in the tablet. Although these artefacts result in edge enhancement in the cross-sections, the reconstructed coefficients no longer represent the attenuation accurately. Instead, the white border may be interpreted as a strongly attenuating coating or as some kind of deposit, which leads to an erroneous interpretation of the internal structure of the tablet. In this regard, the phase contrast signal is considered an artefact that needs to be dealt with in order to improve the quality and the reliability of the reconstructed cross-sections. This can be achieved in two ways, either by *phase retrieval* or by *phase reduction*, as will be discussed in the subsequent sections.



(a) Mixed projection image (b) Reconstructed cross-section

Figure 5.10: A mixed projection image (a) and a reconstructed cross-section (b) of a tablet (diameter ≈ 2.6 mm). The projection images contain strong phase contrast, which introduces artefacts in the reconstruction when not properly dealt with.

In attenuation contrast CT, the reconstructed values represent the linear attenuation coefficients. Mixed projections however contain both attenuation and phase contrast, so an interesting question is raised: which physical quantity is actually reconstructed from the mixed projections? The answer to this question can be given by returning to the TIEW (equation (5.27)). Analysing this equation learns that the recorded intensity is the product of an attenuation and a phase contribution. It is now shown how the reconstructed quantity can be derived in case the filtered backprojection algorithm is used for the reconstruction, according to [19].

For simplicity, assume the object is exposed to a monochromatic, parallel X-ray beam. First, recall equation (5.4) describing the transmitted wave function $U_{d=0}(x, y)$ right behind the object as function of the wave function $U_i(x, y)$ before passing through the object:

$$U_{d=0}(x, y) = T(x, y)\, U_i(x, y)\ . \tag{5.48}$$

The transmission function $T(x, y)$ describing the object is given by equation (5.5), with $B(x, y)$ and $\varphi(x, y)$ given by equations (5.6) and (5.7), respectively. In a

more convenient form for this discussion, this can be written as:

$$T(x,y) = A(x,y)\, e^{i\varphi(x,y)} \,, \tag{5.49}$$

with:

$$A(x,y) = \exp\left[-\frac{1}{2}\int \mu(x,y,z)\, dz\right] \,, \tag{5.50}$$

and:

$$\varphi(x,y) = \frac{2\pi}{\lambda}\int [n_r(x,y,z) - 1]\, dz \,. \tag{5.51}$$

The attenuation function A is now written in terms of the linear attenuation coefficient μ, which is proportional to the imaginary part β of the complex refractive index: $\mu = (4\pi/\lambda)\beta$. The phase term φ is now expressed in terms of the real part of the complex refractive index: $n_r = 1 - \delta$. The intensity of the incident wave is $I_i(x,y) = |U_i(x,y)|^2$. At a propagation distance of $d = 0$ directly behind the object, the intensity is only due to attenuation and is given by:

$$I_{d=0}(x,y) = I_i(x,y)\, A^2 \,. \tag{5.52}$$

Since for a parallel beam $M = 1$ and $I_{d,i} = I_i$, the intensity in the image plane according to the TIEW (equation (5.27)) is given by:

$$I_d(x,y) \approx I_i(x,y) A^2 \left[1 - \frac{d\lambda}{2\pi}\nabla_\perp^2 \varphi(x,y)\right] \,. \tag{5.53}$$

If $(\lambda d/2\pi)\nabla_\perp^2 \varphi(x,y) \ll 1$, this can be written in a more convenient form:

$$I_d(x,y) \approx I_i(x,y) A^2 \exp\left[-\frac{d\lambda}{2\pi}\nabla_\perp^2 \varphi(x,y)\right] \,. \tag{5.54}$$

The reconstruction in attenuation CT is based on the logarithm of the relative intensity recorded in the image plane. Applying this to the previous equation yields:

$$\begin{aligned} -\ln\left[\frac{I_d(x,y)}{I_i(x,y)}\right] &= -2\ln A + \frac{d\lambda}{2\pi}\nabla_\perp^2 \varphi(x,y) \\ &= \int \mu(x,y,z)\, dz + d\,\nabla_\perp^2 \left[\int n_r(x,y,z) dz\right] \,. \end{aligned} \tag{5.55}$$

The tomographic reconstruction of the first term provides the distribution of the linear attenuation coefficients $\mu(x,y,z)$. By extension from 2D to 3D, it is known that second derivatives of the projection images allow the recovery of the Laplacian of the projected distribution. This means that the reconstruction of the second term yields the Laplacian of $n_r(x,y,z)$. The quantity obtained through tomographic reconstruction of mixed projection images is thus given by:

$$q(x,y,z) = \mu(x,y,z) + d\,\nabla^2 n(x,y,z) \,, \tag{5.56}$$

which is a sum of the attenuation contribution, equal to the one obtained using conventional attenuation CT, and a phase contribution, proportional to the Laplacian of the phase.

5.4 Phase retrieval

When a projection image contains only attenuation contrast, as in conventional X-ray tomography, the linear attenuation coefficients $\mu(x, y, z)$ can be reconstructed directly from the projections. Reconstruction of the phase function, presented by the refractive index decrement $\delta(x, y, z)$, however requires an intermediate step. In this step, the phase shift $\varphi(x, y)$ is calculated from the mixed projections for every rotation angle, a problem which is known as phase retrieval. Afterwards, the refraction function can be reconstructed in a similar way as the attenuation. A range of different phase retrieval methods exists, e.g. [15], most of which require projections taken at a limited number of propagation distances for each angle.

Due to practical limitations, acquiring multiple projections per angle is not feasible when using a tube based system. Since such a system uses a conical beam, the size of the projected image changes with different propagation distances. This not only means that the different images need to be matched, but also that the resolution of the images varies. In practice, one typically uses either a large detector with low resolution (large pixel pitch) or a small detector with high resolution (small pixel pitch). The acquisition of multiple images per angle requires a large detector with high resolution, which is very difficult to achieve[8]. Furthermore, such methods usually require a projection image taken in the contact plane $d = 0$, for which the magnification is $M = 1$. Since the image in the contact plane is acquired without magnification, the resolution of this image will never be sufficient. Finally, since the total scanning time for a conventional scan is already rather high, resulting in problems with tube and object stability, the necessity of multiple images is not a very comforting thought. Since a synchrotron source provides a highly parallel and intense beam, the application of multiple image methods is far less problematic. Hence, such methods are usually performed using synchrotron radiation instead of an X-ray tube.

In high resolution CT, a data set typically contains around a thousand projections, consisting of a thousand by a thousand pixels. Consequently, since the phase retrieval has to be calculated for every projection image, any practically feasible algorithm needs to be very efficient. Combined with the restriction of requiring only a single image per projection angle, only a very limited number of solutions is available.

5.4.1 Simultaneous phase and attenuation extraction

An approach that satisfies these practical requirements is given by Paganin *et al.* [27]. Although the derivation of this method is based on the assumption that the object is homogeneous, it turns out to work very well for more general objects

[8] Even though the acquisition of the multiple images may be achieved using several different detectors, this is usually impractical due to their different energy response.

as well [26]. This method simultaneously retrieves the phase and the attenuation from a mixed projection, which allows reconstructing both the attenuation and the phase function of the object separately.

For simplicity, this derivation is first given for a parallel beam. Assume that the object is homogeneous and can be described by its attenuation μ and the refractive index decrement δ. The incident plane-wave radiation is monochromatic and of uniform intensity I_i. The intensity in the contact plane is given by:

$$I_{d=0}(\vec{r}_\perp) = I_i e^{-\mu T(\vec{r}_\perp)} \,, \tag{5.57}$$

where $T(\vec{r}_\perp)$ is the thickness of the object projected onto the image plane, or equally the length of the ray path through the object. For a sufficiently thin object, the phase of the beam in the contact plane is proportional to the projected thickness:

$$\varphi_{d=0}(\vec{r}_\perp) = -\frac{2\pi}{\lambda}\delta T(\vec{r}_\perp) \,. \tag{5.58}$$

Consider again the Transport of Intensity Equation (5.13):

$$\frac{2\pi}{\lambda}\frac{\partial}{\partial z} I_z(\vec{r}_\perp) = -\vec{\nabla}_\perp \left[I_z(\vec{r}_\perp) \vec{\nabla}_\perp \varphi_z(\vec{r}_\perp) \right] . \tag{5.59}$$

Substituting equations (5.57) and (5.58) into the TIE and making use of the identity:

$$\delta \nabla_\perp \left[e^{-\mu T(\vec{r}_\perp)} \, \nabla_\perp T(\vec{r}_\perp) \right] = -\frac{\delta}{\mu} \nabla_\perp^2 e^{-\mu T(\vec{r}_\perp)} \,, \tag{5.60}$$

yields:

$$-\frac{\delta}{\mu} I_i \nabla_\perp^2 e^{-\mu T(\vec{r}_\perp)} = \frac{\partial}{\partial z} I_{d=0}(\vec{r}_\perp) \,. \tag{5.61}$$

The partial derivative of the intensity can be estimated by using two intensity measurements in closely spaced planes: the contact plane $z = 0$ and the image plane $z = d$:

$$\frac{\partial}{\partial z} I_{d=0}(\vec{r}_\perp) \approx \frac{I_d(\vec{r}_\perp) - I_i e^{-\mu T(\vec{r}_\perp)}}{d} \,. \tag{5.62}$$

Substitution in equation (5.61) results in the following expression:

$$\left(1 - \frac{d\delta}{\mu} \nabla_\perp^2 \right) e^{-\mu T(\vec{r}_\perp)} = \frac{I_d(\vec{r}_\perp)}{I_i} \,, \tag{5.63}$$

which is in fact equal to the TIEW (equation (5.27)) when the attenuation and phase are expressed in terms of δ, μ and $T(\vec{r}_\perp)$ for the assumed homogeneous object. The problem of retrieving the phase and attenuation is now reduced to solving equation (5.63) as function of the projected thickness $T(\vec{r}_\perp)$.

To solve this equation, represent the intensity in the contact and the image plane in terms of Fourier transforms, denoted by $F\{\}$:

$$\begin{aligned} I_i e^{-\mu T(\vec{r}_\perp)} &= \frac{I_i}{2\pi} \iint F\{e^{-\mu T(\vec{r}_\perp)}\} e^{i\vec{k}_\perp \cdot \vec{r}_\perp} \, d\vec{k}_\perp \,, \\ I_d(\vec{r}_\perp) &= \frac{1}{2\pi} \iint F\{I_d(\vec{r}_\perp)\} e^{i\vec{k}_\perp \cdot \vec{r}_\perp} \, d\vec{k}_\perp \,. \end{aligned} \tag{5.64}$$

Substitution in equation (5.63) yields:

$$F\{e^{-\mu T(\vec{r}_\perp)}\} = \mu \frac{F\{I_d(\vec{r}_\perp)\} \;/\; I_i}{d\delta\; \vec{k}_\perp^{\,2} + \mu} \;. \tag{5.65}$$

Now, by taking the inverse Fourier transform $F^{-1}\{\}$ the following solution is obtained for $T(\vec{r}_\perp)$:

$$T(\vec{r}_\perp) = -\frac{1}{\mu} \ln \left[F^{-1} \left\{ \mu \frac{F\{I_d(\vec{r}_\perp)\} \;/\; I_i}{d\delta\; \vec{k}_\perp^{\,2} + \mu} \right\} \right] \;. \tag{5.66}$$

In case of illumination by a conical beam with the source located at a distance s before the object, the intensity $I_d(\vec{r}_\perp)$ is related to the intensity $I_d^\infty(\vec{r}_\perp)$ resulting from parallel illumination by [28]:

$$I_d(\vec{r}_\perp) = \frac{1}{M^2} I_d^\infty(\vec{r}_\perp / M) \;, \tag{5.67}$$

with $M = (s + d)/s$ again the magnification. Using this relation in equation (5.66), a final expression for a conical beam is obtained:

$$T(\vec{r}_\perp) = -\frac{1}{\mu} \ln \left[F^{-1} \left\{ \mu \frac{F\{M^2\; I_d(M\vec{r}_\perp)\} \;/\; I_i}{\frac{d\delta}{M}\; \vec{k}_\perp^{\,2} + \mu} \right\} \right] \;. \tag{5.68}$$

This formula allows solving the TIE as function of the projected thickness of a homogeneous object from a single projection image. Both attenuation and phase can be easily derived using equations (5.57) and (5.58). After applying this method to the projection images at all rotation angles, a tomographic reconstruction method can be used to retrieve the structure of the object.

In Mayo *et al.* [26], this method is applied to data acquired at a tube based CT scanner, thus using a polychromatic X-ray beam. An important observation is that the phase retrieval method provides a significant reduction of noise, producing a cleaner reconstructed image. Furthermore it was found that, although being derived for a homogeneous object, the method seems to work surprisingly well for multicomponent objects as well. A critical parameter in this phase retrieval method is the ratio between the attenuation μ and the refractive index decrement δ of the object. If an incorrect value for this ratio is used, the retrieved phase images may still contain traces of diffraction fringes, or else be oversmoothed. Since for a multicomponent material the ratio will only be correct for one component, features composed of other materials may show such artefacts.

5.4.2 Bronnikov algorithm

Another solution is presented by Bronnikov in [29, 30], in which a reconstruction formula for phase contrast tomography is proposed. The reconstruction was written in a form that resembles the filtered backprojection algorithm. The difference

lies in the filtering operation, whereas FBP typically uses a (modified) ramp filter, a new type of filter is introduced to process the phase signal. In this approach, the intermediate step of phase retrieval is redundant and the phase function can be reconstructed directly from the phase contrast images.

Fundamental theorem

The Cartesian coordinate system of the object will be denoted by (x_1, x_2, x_3). The rotation of the object is described by the angle θ, with x_3 the rotation axis. The object is described by the linear attenuation coefficients $\mu(x_1, x_2, x_3)$ and by the refractive index decrement[9] $f(x_1, x_2, x_3) = \Re[n(x_1, x_2, x_3)] - 1$. The object is illuminated by a monochromatic, coherent wave field of wavelength λ. For simplicity of the derivation, the incident beam is assumed to be parallel, thus $M = 1$. The object's complex transmission function at projection angle θ is given by:

$$T_\theta(x, y) = \exp\left[-\frac{1}{2}\mu_\theta(x, y)\right] \cdot \exp[i\varphi_\theta(x, y)] \,, \tag{5.69}$$

with:

$$\mu_\theta(x, y) = \int_{\mathbb{R}^2} \mu(x_1, x_2, y)\, \delta(x - x_1 \cos\theta - x_2 \sin\theta)\, dx_1 dx_2 \,, \tag{5.70}$$

$$\varphi_\theta(x, y) = \frac{2\pi}{\lambda} \int_{\mathbb{R}^2} f(x_1, x_2, y)\, \delta(x - x_1 \cos\theta - x_2 \sin\theta)\, dx_1 dx_2 \,. \tag{5.71}$$

By assuming that μ_θ only varies slowly and that the propagation distance d is in the near field, the TIEW can again be used to describe the intensity in the image plane by (recall equation(5.27)):

$$I_{\theta,d}(x, y) = I_{\theta,d,a}(x, y)\left[1 - \frac{d\lambda}{2\pi M}\nabla_\perp^2 \varphi_\theta(x, y)\right] \,, \tag{5.72}$$

with $I_{\theta,d,a}(x, y)$ the intensity which is only due to attenuation. This intensity is given by:

$$I_{\theta,d,a}(x, y) = I_{\theta,d,i}(x, y)\, e^{-\mu_\theta(x,y)} \,, \tag{5.73}$$

where $I_{\theta,d,i}(x, y)$ is the intensity that would have reached the image plane in the absence of the object. For the derivation, recall the 2D Radon transform of the function $g(x, y)$ over lines defined by $s = x\ \sin\omega + y\ \cos\omega$ in the XY-plane:

$$\hat{g}(s, \omega) = \int_{\mathbb{R}^2} g(x, y)\, \delta(s - x\ \sin\omega - y\ \cos\omega)\, dxdy \,. \tag{5.74}$$

[9] The refractive index decrement is now represented by f to avoid confusion with the Dirac delta function δ.

The 3D Radon transform of the function $f(x_1, x_2, x_3)$ describes integration over planes with normal vector $(\cos\theta\ \sin\omega, \sin\theta\ \sin\omega, \cos\omega)$ and distance s from the origin, and is given by:

$$\begin{aligned}\hat{f}(s,\theta,\omega) \quad &= \quad \int_{\mathbb{R}^3} f(x_1,x_2,x_3)\ \delta[s-(x_1\ \cos\theta + x_2\ \sin\theta) \\ &\quad \cdot \sin\omega - x_3\ \cos\omega]\ dx_1 dx_2 dx_3\ . \end{aligned} \tag{5.75}$$

The Bronnikov algorithm is based on the following theorem, of which a proof can be found in [30]:

Theorem

Let:

$$g_\theta(x,y) = \frac{I_{\theta,d}(x,y)}{I_{\theta,d=0}(x,y)} - 1\ , \tag{5.76}$$

where $I_{\theta,d}(x,y)$ is the intensity distribution at at sufficiently small distance d and where $I_{\theta,d=0}(x,y)$ is the intensity in the contact plane, then:

$$\frac{\partial^2}{\partial s^2}\hat{f}(s,\theta,\omega) = -\frac{1}{d}\ \hat{g}_\theta(s,\omega)\ . \tag{5.77}$$

□

Reconstruction algorithm

The theorem thus describes a relation between the Radon transform of the refractive index decrement and the Radon transform of the measured intensity data, which allows deriving a reconstruction algorithm. This is achieved by using an explicit formula for the inversion of the Radon transform:

$$f(x_1,x_2,x_3) = -\frac{1}{4\pi^2}\int_0^\pi \sin\omega\ d\omega \int_0^\pi \left[\frac{\partial^2}{\partial s^2}\hat{f}(s,\theta,\omega)\right] d\theta\ , \tag{5.78}$$

where following the differentiation, s is replaced by:

$$s' = (x_1\ \cos\theta + x_2\ \sin\theta)\cdot\sin\omega + x_3\ \cos\omega\ . \tag{5.79}$$

Since this formula readily contains the second derivative of the Radon transform, the reconstruction problem is immediately solved by inserting equation (5.77):

$$f(x_1,x_2,x_3) = \frac{1}{4\pi^2 d}\int_0^\pi \sin\omega\ d\omega \int_0^\pi \hat{g}_\theta(s',\omega)\ d\theta\ . \tag{5.80}$$

This formula can be used directly in the numerical implementation of the reconstruction algorithm. However, in the current form it requires intensive computations. Therefore, it is more convenient to calculate the integral over the angle ω, which results in a more practical algorithm. After some calculations and by using:

$$q(x,y) = \frac{|y|}{x^2+y^2} \,, \tag{5.81}$$

the following expression can be derived:

$$f(x_1,x_2,x_3) = \frac{1}{4\pi^2 d}\int_0^\pi [q * g_\theta] d\theta \,, \tag{5.82}$$

where following the convolution the arguments x and y of $q * g_\theta$ are replaced by $x = x_1 \cos\theta + x_2 \sin\theta$ and $y = x_3$. Equation (5.82) provides a reconstruction algorithm in the form of a filtered backprojection. The convolution can be computed in the Fourier domain, with the Fourier transform of the filter function $q(x,y)$ given by:

$$Q(\xi,\eta) = \frac{|\xi|}{\xi^2+\eta^2} \,, \tag{5.83}$$

where ξ and η are the spatial frequencies. From this expression, one may recognize the filter as being a combination of two well-known filters: the ramp filter $|\xi|$ which follows from the impulse response of the backprojection, and the 2D integrator $1\ /\ (\xi^2+\eta^2)$ which inverts the Laplacian operator applied to the phase shift.

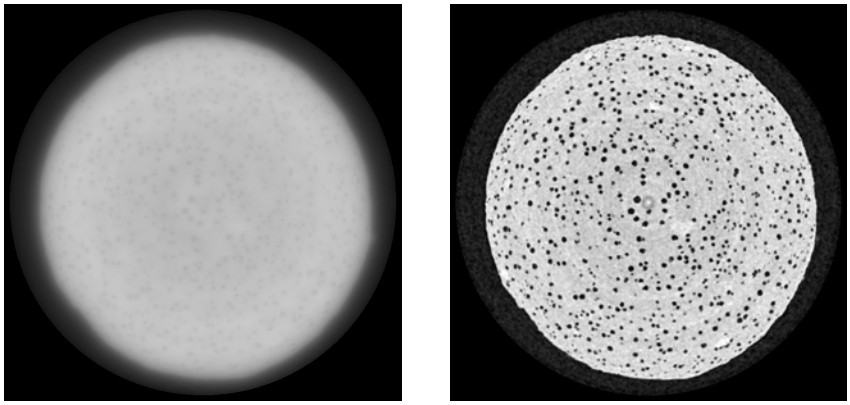
Attenuation correction

The reconstruction formula given by equation (5.82) thus allows retrieving the distribution of the refraction function of the object, provided the function $g(x,y)$ is known. According to equation (5.76), this function can be calculated based on the measured intensity in the image plane and in the contact plane. For a pure phase object, the intensity in the contact plane simplifies to unity, so the phase distribution can be reconstructed. However, when the object also introduces attenuation, even only a small amount, the reconstructed images are severely distorted (see figure 5.11(a)). This means that, for a mixed object, two projection images are required at each rotation angle. Consequently, this approach resolves into a multiple image method and thereby no longer fits the imposed requirements for a practical solution.

An alternative approach was presented by Groso *et al.* [31], who modified the Bronnikov algorithm in order to correct the reconstruction artefacts due to the remaining attenuation (see figure 5.11(b)). The correction consists of adding a correction factor α to the denominator of the filter function:

$$Q(\xi,\eta) = \frac{|\xi|}{\xi^2+\eta^2+\alpha} \,, \tag{5.84}$$

The resulting method, which is called the *modified Bronnikov algorithm* (MBA), thus allows reconstructing the refraction function based on single projection images, provided the attenuation by the object is weak and only slowly varying. Although for the derivation of this method the incident X-ray beam was assumed to be parallel, this reconstruction method can be applied directly to a conical beam as well.



(a) Bronnikov algorithm (b) Modified Bronnikov algorithm

Figure 5.11: The reconstruction of the phase function of the tablet from a set of single, mixed projections. The maximum attenuation of the X-rays by the tablet is about 8%. The reconstruction was performed using the Bronnikov algorithm (a) and using the modified Bronnikov algorithm which corrects for the remaining attenuation (b).

5.4.3 Practical

In practice, it is found to be much more computationally efficient [15] to subdivide the filter function of MBA into two separate filters:

$$Q(\xi,\eta) = p(\xi,\eta)r(\xi)\ , \tag{5.85}$$

where $r(\xi)$ is the ramp filter:

$$r(\xi) = |\xi|\ , \tag{5.86}$$

which is already a part of the filtered backprojection algorithm. The phase filter $p(\xi,\eta)$ is given by:

$$p(\xi,\eta) = \frac{1}{\xi^2+\eta^2+\alpha}\ . \tag{5.87}$$

Reconstruction is then performed by first applying the phase filter $p(\xi,\eta)$ to the mixed projection images, and then reconstructing the obtained phase projections using a standard reconstruction algorithm. Thus, even though MBA was developed

as a direct reconstruction method, in a practical implementation it may actually become a phase retrieval method.

Considering this separation and looking back at the phase retrieval algorithm described by equation (5.68), it turns out that both methods are in fact very similar. This is a rather surprising conclusion, as both methods are derived from a different perspective using rather different assumptions. In both methods, the phase projection is retrieved from the mixed projection by taking the Fourier transform of the measured projection image, applying a filtering function and taking the inverse Fourier transform. The resulting images are then reconstructed using a standard tomographic reconstruction method. The filtering functions for both methods are similar and can be described using general parameters a en b by:

$$p(\xi, \eta) = \frac{a}{\xi^2 + \eta^2 + b} \,, \tag{5.88}$$

Due to this uniformity of the filtering function for both methods, it follows that the correction parameter α will have a similar influence on the reconstruction as the ratio between μ and δ. This means that selecting α incorrectly may result in either the reappearance of fringes or blurring of the images. The optimal value of the correction factor for MBA can be found using a semi empirical approach. This is done by comparing simulated with experimental data, as was shown in [31]. In a practical, tube based set-up however, the magnification, the spectrum of polychromatic radiation and the object composition can all vary significantly, resulting in a wide range of combinations. This huge amount of possible experimental conditions makes the determination of the optimal value of the correction parameter α far less evident.

5.5 Modified Bronnikov algorithm

As was discussed above, both phase retrieval methods are in fact similar. For historical reasons, from this point on the applied phase retrieval method will be referred to as MBA. The application of the MBA reconstruction method was introduced at UGCT by M. Boone, as part of his master thesis regarding phase contrast formation at a high resolution X-ray tube [32].

In practice, a certain amount of attenuation is always present in the acquired projection images. Therefore, the attenuation parameter α is always required and as was mentioned by Mayo *et al.* [26] its value should be chosen with care. If not done properly, the retrieved phase projections and the resulting reconstructed cross-sections are distorted, either by oversmoothing or by reappearance of the fringes. Due to the variety of experimental conditions however, determination of the correct α is not obvious. To obtain some insight into the choice of an appropriate value for α, it is interesting to study the quality of MBA with regard to some of the experimental parameters. This is done by evaluating the reconstruction of the

phase phantom described in table A.3. The mixed projection images are simulated based on the following expression:

$$I_d = I_{d,i}\, e^{-\mu_\theta} \left[1 - \frac{d\lambda}{2\pi M} \nabla_\perp^2 \varphi + \frac{d\lambda}{2\pi M} \left(\frac{\partial \mu_\theta}{\partial x} \frac{\partial \varphi_\theta}{\partial y} + \frac{\partial \mu_\theta}{\partial y} \frac{\partial \varphi_\theta}{\partial x} \right) \right] , \quad (5.89)$$

which is derived in [30] and holds in the near field of the Fresnel region. The integrals μ_θ and φ_θ are calculated by formulas (5.70) and (5.71). In these simulations, the size of the object is about $r \approx 1$ mm and the energy of the X-rays is 10 keV (λ = 0.124 nm). Thus, for typical propagation distances d in the order of 1 m, the image plane lies in the near field region ($d < r^2/\lambda$). Under the assumption that the attenuation μ_θ varies insignificantly, equation (5.89) simplifies to the TIEW (equation (5.27)). Compared to the TIEW, equation (5.89) thus includes first order terms of the variance in μ_θ, which provides an appropriate approximation when the attenuation can no longer be ignored.

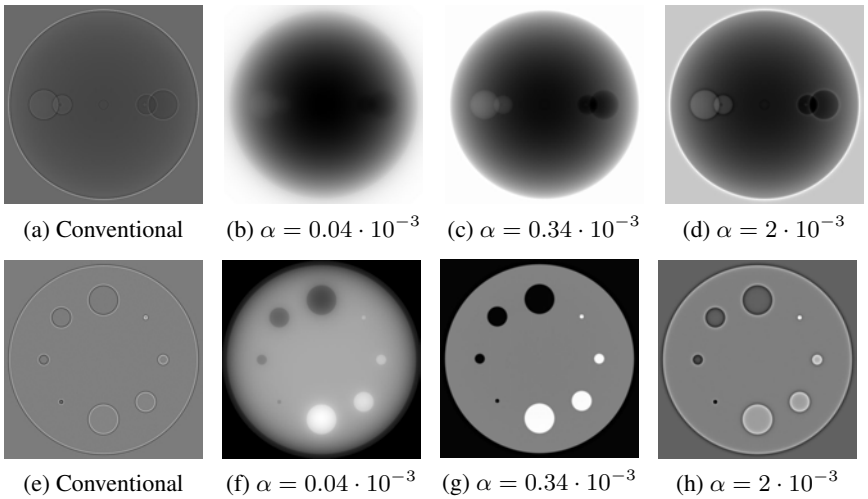
The phase is retrieved from the mixed projection images using the MBA filter (equation (5.84)). Afterwards, the exponential function is applied to the retrieved phase images, such that the resulting images are similar to conventional attenuation projections. Finally, these phase projections are reconstructed using the FDK algorithm.

5.5.1 Attenuation correction parameter

For the first test, the object is defined by $\mu_0 = \mu_1 = 0.05\,\text{mm}^{-1}$ and $\delta_0 = \delta_1 = 1.0 \cdot 10^{-6}$. The maximum attenuation by the bulk of the object is then about 4.7%. In practice, these parameters describe a phantom consisting of a large sphere with $\mu = \mu_0$ and $\delta = \delta_0$, containing four pores with $\mu = 0$, $\delta = 0$ and four inclusions with $\mu = 2\ \mu_0$ and $\delta = 2\ \delta_0$. This means that for the two composing materials the ratio between μ and δ is the same.

In the simulation, the source to object distance is 10 mm and the source to detector distance 250 mm, thus the propagation distance d is 240 mm. The pixel pitch of the detector is 50 µm. In figure 5.12, a mixed projection image (a) of the phantom is shown together with phase projections obtained by the MBA filter with different values of the attenuation correction parameter α. When α is chosen too low (b), the resulting phase projection images are oversmoothed. When it is too high (d), the phase fringes reappear in the projections. For an appropriate value of α (c), the projections are quite sharp and contain no traces of the fringes. The reconstructed cross-sections of the obtained projection images clearly illustrate the effect of an incorrect value of α, resulting in images that are either too blurry (f) or that still contain phase artefacts (h). Using the correct α, the object can be reconstructed very well (g) and MBA thus provides a major improvement in image quality compared to the conventional reconstruction (e).

In this and the following tests, the optimal value was chosen by reconstructing a cross-section of the object for different values of α and selecting the value that



(a) Conventional (b) $\alpha = 0.04 \cdot 10^{-3}$ (c) $\alpha = 0.34 \cdot 10^{-3}$ (d) $\alpha = 2 \cdot 10^{-3}$

(e) Conventional (f) $\alpha = 0.04 \cdot 10^{-3}$ (g) $\alpha = 0.34 \cdot 10^{-3}$ (h) $\alpha = 2 \cdot 10^{-3}$

Figure 5.12: The effect of the attenuation correction parameter α on the quality of the phase retrieval (b-d) and the subsequent reconstructions (f-h), as compared to the conventional reconstruction (e) of the mixed projection images (a).

resulted in the lowest reconstruction error, the NRMSE, as compared to the original phantom. This is illustrated in figure 5.13, where the reconstruction error is shown as a function of α. The error reaches a minimum at $\alpha = 0.34 \cdot 10^{-3}$, which is the optimal value for the attenuation correction given the current experimental conditions. For completeness, the error using the conventional reconstruction is 6.23, compared to 0.09 for the optimal α.

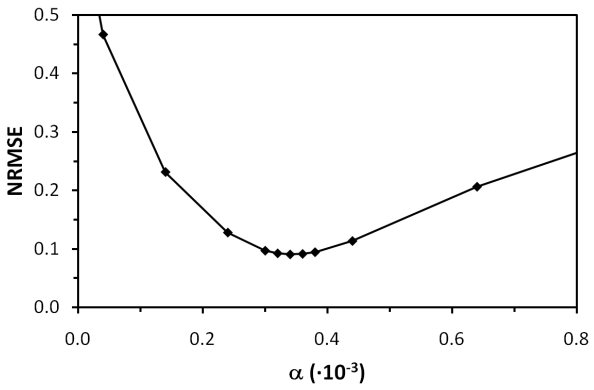


Figure 5.13: The normalised root mean squared error (NRMSE) of the MBA reconstruction of a mixed phase and attenuation object, as function of the attenuation correction parameter α.

5.5.2 Refractive index decrement

A second test is the evaluation of MBA for different values of the refractive index decrement. The object is now described by $\mu_0 = \mu_1 = 0.05\ \text{mm}^{-1}$ and the parameter $\delta = \delta_0 = \delta_1$, which is varied from $0.25 \cdot$ to $4.0 \cdot 10^6$. The highest positive peak of the corresponding phase signal then varies from 1.025 to 1.4, respectively, in the mixed projection images. As can be seen from the results in figure 5.14, the optimal value for α strongly depends on the magnitude of the observed phase signal. The reconstruction error however only increases slightly with increasing δ (from 0.0898 to 0.0911), so MBA is capable of retrieving the phase signal accurately regardless of its amplitude.

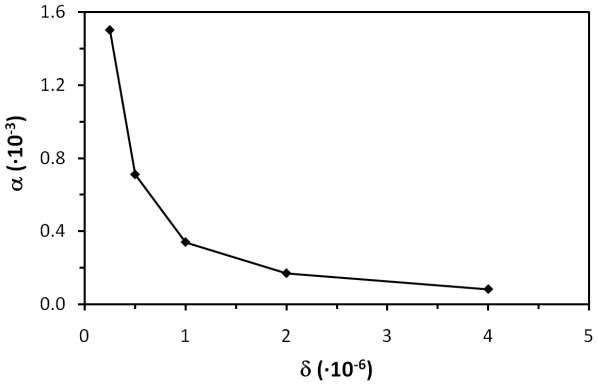


Figure 5.14: The optimal attenuation correction parameter α as function of the refractive index decrement δ.

By looking at equation (5.89) in combination with (5.70) and (5.71), it follows that the contribution of the phase signal is determined by the term:

$$\frac{d}{M}\left(\nabla_{\perp}^{2}\delta_{\theta} + \frac{\partial\mu_{\theta}}{\partial x}\frac{\partial\delta_{\theta}}{\partial y} + \frac{\partial\mu_{\theta}}{\partial y}\frac{\partial\delta_{\theta}}{\partial x}\right), \tag{5.90}$$

where δ_θ is now given by the integral of δ over the ray path, similar to equation (5.71). Since this term is proportional to d, $1/M$ and δ, for the current simulations varying d or $1/M$ will have a similar effect on α as varying δ. Even though λ has disappeared in equation (5.90), the phase term still depends on the wavelength due to the dependence of $\delta(\lambda)$ (and $\mu(\lambda)$). Therefore, a change of wavelength can simply be described by changing δ, which means that the influence of λ on α can again be understood by the results presented above.

5.5.3 Remaining attenuation

Since MBA was derived directly from the TIEW, the attenuation of the X-rays by the object is assumed to be small and only slowly varying. In fact, in case of a

single image acquisition the original derivation is only valid when no attenuation is present. This problem was resolved by introducing the attenuation correction parameter α, but it is yet to be determined how rigid this correction is. Since in a practical situation a certain amount of attenuation is always present, it is very important to validate MBA as function of this remaining attenuation.

In contrast to the TIEW, the phase term of the simulation model represented by equation (5.89) also contains a contribution of the derivative of the attenuation. Hence, there is an inconsistency between the simulation model and the reconstruction method, which becomes apparent when the conditions of small and slowly varying attenuation are not met. Evaluation of the reconstruction quality as function of the remaining attenuation thus allows quantifying the error introduced by this approximation.

For these tests, the refractive index decrement of the object was kept constant at $\delta_0 = \delta_1 = 1.0 \cdot 10^{-6}$, while the attenuation $\mu = \mu_0 = \mu_1$ is varied from 0.02 to 0.80 mm^{-1}. This corresponds to a maximum attenuation by the bulk of the object ranging from 2 to 54%. Results for this test are presented in figure 5.15, which seems to suggest a linear relation between the attenuation correction parameter α and the remaining attenuation.

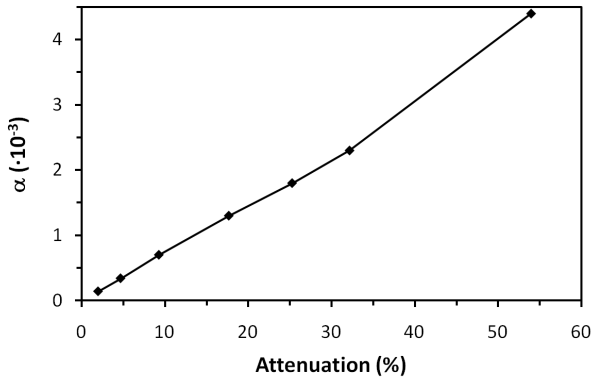


Figure 5.15: The attenuation correction parameter α as function of the remaining attenuation by the object.

However, the possible existence of such a simple relation does not mean that MBA remains valid for increasing attenuation. As can be seen in figure 5.16, the reconstruction quality is seriously affected by the remaining attenuation, as was to be expected. With increasing attenuation, the reconstructed cross-sections are distorted by the appearance of a cupping artefact, as shown in figure 5.17. The introduction of this artefact results from the violation of the conditions of a weak and slowly varying attenuation, which was assumed in the derivation of MBA.

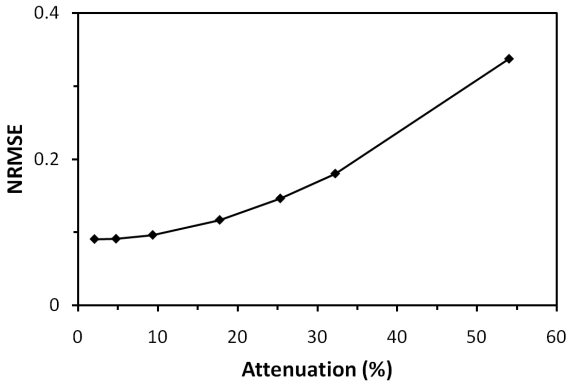
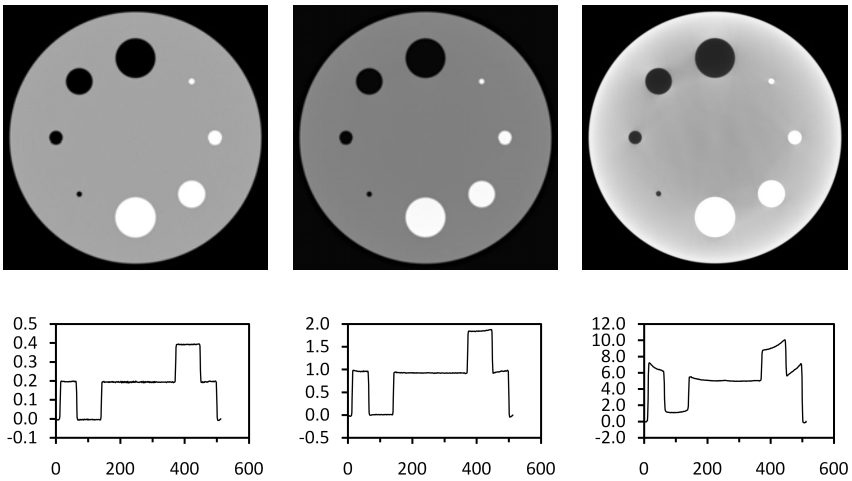


Figure 5.16: The NRMSE as function of the remaining attenuation by the object.



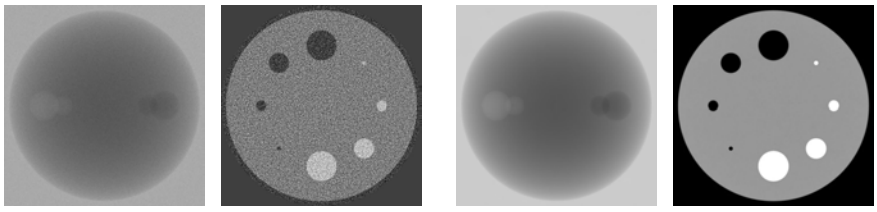
(a) Attenuation 2% (b) Attenuation 9% (c) Attenuation 54%

Figure 5.17: The remaining attenuation by the object introduces a cupping artefact in the reconstructed cross-sections obtained by using MBA. The line profiles are taken along the vertical central line of the cross-sections and are given in arbitrary units.

5.5.4 Noise

It was already shown in [26] and [30] that this type of phase retrieval reduces noise significantly. In high resolution CT, the noise handling properties of the reconstruction algorithm are very important due to the low flux and thus limited statistical information. It is therefore interesting to verify the noise behaviour of MBA and compare it with conventional reconstruction algorithms for attenuation CT. For this test, projection images of the phase phantom are simulated both as attenuation only and as mixed projections, which allows comparison between attenuation and phase CT. Both sets of projections were calculated with different noise levels. The noise is generated using a Poisson distribution and can be described by the expected number of counts Λ per detector pixel when the X-rays are not attenuated, as was explained in section 4.15.2.

The object settings for this test are $\mu_0 = \mu_1 = 0.05\ \mathrm{mm}^{-1}$ and $\delta_0 = \delta_1 = 1.0 \cdot 10^{-6}$. The maximum attenuation by the bulk of the object is then about 4.7%. In practice, these parameters describe a phantom consisting of a large sphere with $\mu = \mu_0$ and $\delta = \delta_0$, containing four pores with $\mu = 0$, $\delta = 0$ and four inclusions with $\mu = 2\ \mu_0$ and $\delta = 2\ \delta_0$. This means that for the two composing materials the ratio between μ and δ is the same. The difference between phase and attenuation imaging can clearly be seen in figure 5.18. Looking at the projection images, the smaller features in the phantom can be distinguished much more easily in the mixed projection, whereas they can hardly be observed in the attenuation projection. Even though both projections are simulated using the same noise level, the phase reconstruction provides a cross-section that contains only little noise, whereas the attenuation cross-section is heavily distorted. This different noise behaviour can be understood by noting that the phase retrieval method in fact integrates the observed peaks, making it much more stable with regard to noise.



Projection Reconstruction Projection Reconstruction

(a) Attenuation (b) Phase

Figure 5.18: Comparison of the noise behaviour between attenuation (a) and phase contrast CT (b), for $\Lambda = 10^{2.5}$ counts. The projection image shown in (b) is the phase projection obtained after applying the MBA filter on the mixed projection.

A quantitative comparison between the noise behaviour of phase and attenuation CT is given in figure 5.19, where the reconstruction error is displayed as function of the noise level. For attenuation CT, the error rises quickly with increasing

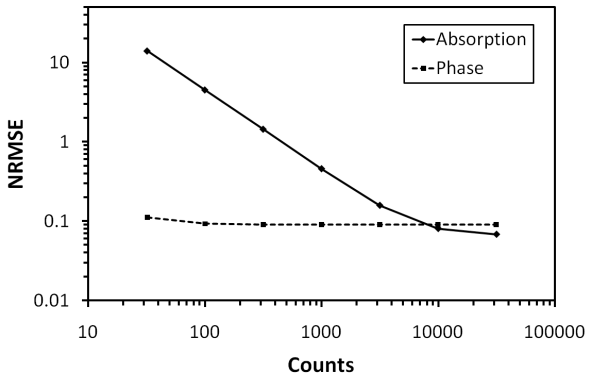


Figure 5.19: The reconstruction error for attenuation and phase contrast CT as function of the noise level, described by the expected number of counts Λ.

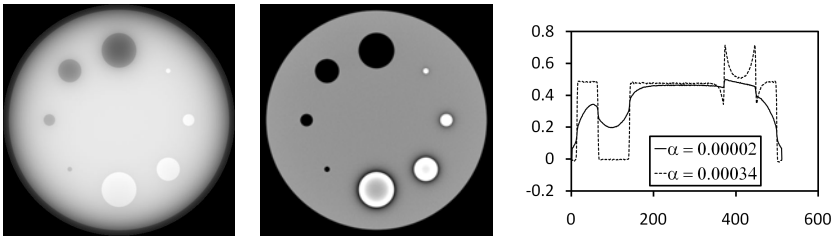
noise. For phase contrast CT however, the reconstruction error is almost constant, with only a slight increase towards a really low number of counts. These results show that the phase retrieval is very stable and only suffers mildly from noise. This suggests that phase contrast imaging can be performed using lower exposure times than conventional attenuation imaging, which would provide a major advantage. Experimental verification of this behaviour should definitely be performed in further research, as this could have important consequences for the further development of high resolution X-ray CT.

5.5.5 Different ratio

In the discussion on the simultaneous phase and attenuation extraction (section 5.4.1) it was already mentioned that the phase retrieval fails if the projected object consists of multiple components with a different ratio between the attenuation μ and the refractive decrement δ. This problem is not encountered in the derivation of the Bronnikov algorithm. However, MBA was only proposed as a correction to the original algorithm and is not theoretically justified. Since both phase retrieval methods were in fact found to be similar, it can be assumed that MBA will suffer from different μ/δ ratios as well.

The influence of this different ratio was evaluated by considering the projections of the phase phantom with $\mu_0 = 0.05\,\text{mm}^{-1}$, $\mu_1 = 0\,\text{mm}^{-1}$ and $\delta_0 = \delta_1 = 1.0 \cdot 10^{-6}$. Choosing these parameter means that the four inclusions now have the same attenuation coefficient as the bulk of the object (presented by the surrounding sphere), while their refractive index is different from the bulk. The ratio μ/δ of the inclusions is thus half the ratio of the bulk. The results for this test are presented in figure 5.20. In case of a constant ratio, the optimal value for the attenuation correction parameter was found to be $\alpha = 0.34 \cdot 10^{-3}$. Using this value,

the phase retrieval for the bulk of the object and the pores inside is again optimal. However, retrieval of the inclusions is inadequate as the phase artefacts reappear in the reconstructed cross-section. These artefacts can be removed by setting α to $0.02 \cdot 10^{-3}$, but since this value is too low for the phase retrieval of the bulk it now becomes heavily smoothed. It is thus impossible to find a correct value for α at which both components are reconstructed well. This reveals the failure of the phase retrieval in case of a varying ratio μ/δ for a multicomponent object.



(a) $\alpha = 0.02 \cdot 10^{-3}$ (b) $\alpha = 0.34 \cdot 10^{-3}$ (c) Line profile

Figure 5.20: The reconstructed cross-sections of the phase phantom with a different ratio between μ and δ for the two components. The line profile is taken along the vertical central line of the images.

5.5.6 Polychromatic beam

The last test, which concludes this section, is how MBA performs when a polychromatic X-ray beam is used. This evaluation is of high importance regarding the practical application of MBA, since nearly every tube based high resolution CT system uses a polychromatic beam. Although this beam can be made monochromatic using the right optics, the resulting loss in intensity is not feasible. When simulating the phase contrast signal according to equation (5.89) with different wavelengths[10], the shape of the phase signal changes but the position of the positive and negative peak remains in place. From the discussion on the temporal coherence using equation (5.17) however, it was noted that the position of the positive peak of the phase signal changes with the wavelength. Since in the current simulation model of the phase signal this position is stationary with respect to the wavelength, it does not allow evaluation of the phase retrieval for a polychromatic beam.

An improved model that does contain the shift of the positive peak can be obtained by returning to the derivation of the TIEW in the ray optical approach, of which equation (5.27) is the final result. In this derivation, the position in the image plane (x_d, y_d), given by equation (5.24), was approximated by assuming that the deviation of the positions in the image plane due to refraction is small. For

[10]Note that the refractive index decrement $\delta(\lambda)$ also depends on the wavelength λ

the phase phantom used in the previous tests, at the edges of the phantom this deviation goes up to about 0.2 times the pixel pitch, which means that the limit of the approximation is reached. By dropping this approximation, the phase signal can thus be simulated more accurately using equation (5.25). In practice, this new simulation model is achieved by simulating the intensity distribution using equation (5.89) and then remapping this distribution from the positions (x, y) to the image plane positions (x_d, y_d) using equation (5.24). Using this modified model, the position of the phase signal now changes with a different wavelength. This is illustrated in figure 5.21, which shows two simulations of the phase signal using a different X-ray energy. It can be seen that the negative peak of the signal remains in place while the positive peak moves further away with decreasing energy, which is in agreement with the earlier discussion on the energy dependence of the refraction. This shift is however quite small, due to the very small deviation of the X-rays using the current settings. For 10 keV X-rays, the maximum shift is only 0.23 pixels, compared to a maximum shift of 0.66 pixels for 6 keV X-rays.

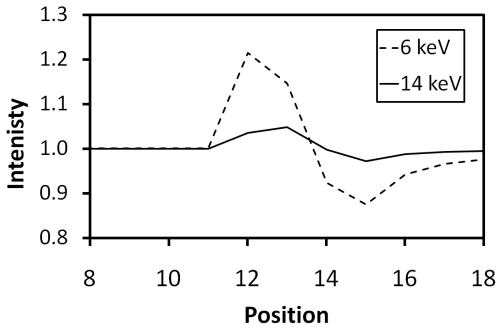


Figure 5.21: The phase signal at the edge of the phase phantom for two different energies.

According to equations (5.2) and (5.3), the refractive index decrement δ and the linear attenuation μ (which is proportional to β) are proportional to f_1/E^2 and f_2/E^2, respectively, where E is the X-ray energy and f_1 and f_2 are the atomic scattering factors. Within the typical energy range used in high resolution CT, it is a valid approximation to assume that the factor f_1 is constant. This means that the refractive index decrement δ can be assumed to be proportional to $1/E^2$. The factor f_2 on the other hand is not constant, but can roughly be described as being inversely proportional to the energy. Using this assumption, the linear attenuation μ is found to be proportional to $1/E^3$. The intensity for a polychromatic beam can now be simulated by splitting the energy spectrum into discrete bins. Using the two relations $\delta \propto 1/E^2$ and $\mu \propto 1/E^3$, the intensity distributions can be calculated separately for each energy bin. The total intensity for the polychromatic beam is then obtained by combining these separate intensity distributions using a weighted sum, where the weights are given by the relative contributions of the corresponding energy bins to the total photon energy spectrum.

The influence of a polychromatic beam is investigated by comparing the ob-

tained results with the ones obtained using a monochromatic beam. The monochromatic beam is simulated with an energy of E = 10 keV and the phantom is described by the linear attenuation $\mu_0 = \mu_1 = 0.05\,\text{mm}^{-1}$ and refractive index decrement $\delta_0 = \delta_1 = 1.0 \cdot 10^{-6}$. For the polychromatic beam, the X-ray spectrum is defined by a Gaussian distribution with mean energy 10 keV and a full width at half maximum of 6 keV. This spectrum is separated into 11 energy bins of width 1 keV, ranging from 4.5 to 15.5 keV. In order to obtain a valid comparison, the attenuation of the phantom at 10 keV was chosen to be $\mu_0 = \mu_1 = 0.0345\,\text{mm}^{-1}$, such that the mean attenuation by the object is similar to the attenuation for the monochromatic beam. Similarly, the refractive index decrement of the phantom was chosen to be $\delta_0 = \delta_1 = 0.83 \cdot 10^{-6}$, which results in a similar height of the positive peaks of the phase signal.

The resulting reconstructions for the mono- and polychromatic beam are very similar and almost no apparent differences can be found. For the monochromatic beam, the optimal attenuation correction parameter was found to be $\alpha = 0.36 \cdot 10^{-3}$, resulting in a reconstruction error of 0.0940, compared to $\alpha = 0.30 \cdot 10^{-3}$ and an error of 0.1034 for the polychromatic beam. The line profiles along the edges of both reconstructed phantoms are plotted in figure 5.22. From these profiles, the edge of the phantom seems to be slightly less sharp for the polychromatic beam, which means that the MBA results in slightly blurrier images. However, using these typical experimental settings the difference between using a mono- or polychromatic beam appears to be only very small, which means that this type of phase contrast imaging can be performed properly using a tube based set-up.

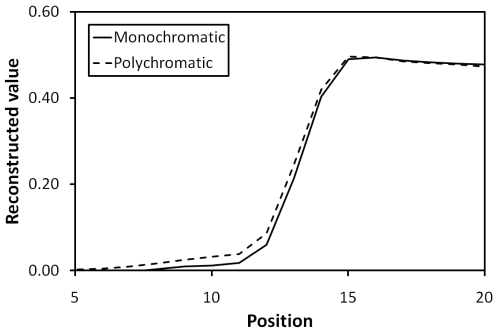


Figure 5.22: The reconstruction of the phase function for a monochromatic beam of 10 keV and a polychromatic beam with mean energy of 10 keV and FWHM of 6 keV.

5.5.7 Practical

Due to both its high quality and fast and simple implementation, MBA serves as an excellent practical method for retrieving the phase signal from mixed projection images. At UGCT, MBA is frequently applied in the reconstruction of every-

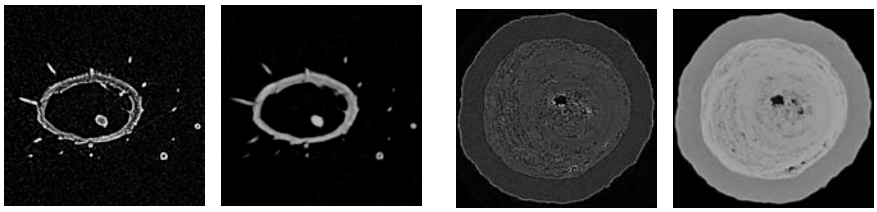
day high resolution scans. Using MBA, many impressive results were already obtained, providing a much higher quality than the conventional reconstruction (e.g. [33, 34]), as will be shown at the end of this chapter.

However, as is the case with any technique, MBA also has its shortcomings when used in a practical context. In the simulations it was shown that MBA is no longer valid when the remaining attenuation by the object becomes too large, in which case a cupping artefact appears in the reconstructed slices. The remaining attenuation can be decreased by using an X-ray beam of higher energy, which is more favourable for the phase retrieval. However, this requires an increase of the propagation distance in order to detect the phase contrast signal, which is usually not feasible due to the decreased X-ray intensity in the detector plane. Furthermore, MBA fails when the ratio between the attenuation and the refractive index decrement varies for the different components in the object, resulting in a possible blurring of the image. The use of a polychromatic beam, which implies a varying μ/δ ratio for different energies, also seems to introduce some blurring in the reconstructions. These effects deteriorate the image quality and are simply a result of the violation of the conditions under which MBA was derived. Since in practice these ideal conditions are never met, artefacts are expected to appear in the reconstructed images. Both the blurring and cupping artefact are often encountered, depending on the experimental conditions. Examples of these artefacts are shown in figure 5.23, in which MBA is compared to the conventional reconstruction of the projection data[11]. In the conventional reconstruction, the cross-sections are severely distorted due to the appearance of phase contrast artefacts. MBA on the other hand retrieves and reconstructs the phase signal, resulting in images of much higher quality. However, the MBA reconstruction of the fly leg seems to be severely blurred compared to the conventional reconstruction. The MBA reconstruction of the bead contains a strong cupping artefact, which can be clearly observed as an increase in the reconstructed value towards the edge of the bead.

Finally, in a practical context it is vital to note the importance of proper normalisation. If the projections can not be normalised adequately and the background intensity within a single projection image varies [12], MBA is not able to retrieve the phase signal correctly. This is due to the fact that MBA involves a double integration of the function $g_\theta(x, y)$ (equation (5.76)), which should be zero outside the projection of the object. Any deviation in the background intensity therefore introduces a dc-shift which severely distorts the retrieved images and the resulting reconstruction.

[11] The reconstruction of both examples is discussed in more detail in the application section of this chapter. For now, the focus will be limited to the blurring and cupping due to the application of MBA.

[12] E.g. when the background intensity decreases towards the centre of the projection or varies as function of the height.



Conventional MBA Conventional MBA

(a) Fly leg (b) Bead

Figure 5.23: Illustration of the blurring (a) and cupping (b) artefacts that are introduced when using MBA in practice, as compared to the conventional reconstruction of the projection data.

5.6 Phase reduction

Application of MBA already allows reconstructing a wide variety of weakly attenuating objects which produce both an attenuation and a phase contrast signal. However, there still remains a large collection of objects for which MBA fails, especially when a large amount of attenuation remains present. Such moderately or strongly attenuating objects[13] are in fact quite commonly encountered in high resolution X-ray CT. Therefore, there still exists a need for a reconstruction algorithm that can handle these objects appropriately.

The search for such a method is performed from a rather different perspective than the methods presented before. Instead of retrieving and reconstructing the phase contrast signal, the object is again reconstructed based on its attenuation contrast. Since the conventional reconstruction of mixed projections inevitably introduces phase contrast artefacts in the cross-sections, attempts are now made to reduce these phase artefacts. After all, it seems quite rational to try to reconstruct strong phase objects based on their attenuation contrast, as this signal is usually more accurate than the phase signal. This is especially true for conventional high resolution set-ups, as these are designed to optimize attenuation and not phase contrast. For such systems, the appearance of phase contrast is usually considered as being a disturbing effect, which is unavoidably present at high resolutions and generates severe artefacts in the reconstructed images.

As part of this work, research was performed on the development of such a phase artefact reduction algorithm, based on iterative reconstruction methods. The idea behind this is that, if one can accurately calculate the phase signal generated by the reconstructed object, this simulation can be inserted into the forward projection step of the iterative algorithm, which would allow the object to be reconstructed without phase artefacts. However, as simple as this idea may sound,

[13]To discriminate between the two types of objects, from now on a *weak phase object* refers to a weakly attenuating object that creates a phase signal, whereas a *strong phase object* denotes a moderately or strongly attenuating object which also creates a phase signal.

bringing it into practice is far from obvious.

Unfortunately, the attempts to introduce phase contrast into an iterative reconstruction have not (yet) led to any satisfying results. In this section, some of the attempts that were made for this study are presented. Each of these attempts is briefly explained and a short discussion is provided regarding the performance and the reason of failure. Although this approach may seem a bit odd to some readers, these ideas and concerns were included since others may benefit from it in future research, either by avoiding certain errors or by finding a solution to some problems. Furthermore, this research resulted in the development of an efficient phase artefact reduction algorithm know as the *Bronnikov-Aided correction* (BAC), which is presented in the next section. Although BAC no longer has any connection with iterative algorithms, it may be interesting to elucidate the train of thought that has eventually led to this method.

5.6.1 Iterative reconstruction

As was shown in the previous sections, several methods exist to simulate the phase contrast signal. To this end, both the ray and wave optical approach provide appropriate simulation methods, based either on the transport of intensity equation or on the Kirchoff integral. An additional method is provided by applying *Snell's law*[14], which describes the relation between the angle of the incident and refracted ray at the interface between two materials of different refractive index. For small deviations, this method corresponds to the expression of the angular deviation given by equation (5.17). For a method to be practically feasible, it has to obey three important requirements: accuracy, speed and discretization. It is quite obvious that the simulation method needs to produce the phase signal with a certain accuracy, otherwise inconsistencies with the real phase signal result in severe reconstruction artefacts. Since the simulation needs to be executed for every forward projection step, it needs to be fast since otherwise the reconstruction time quickly gets out of proportion. The last requirement is that the simulation works using a discrete input. Unlike the previous simulations, which use analytical phantoms, the reconstructed object is now represented on a discrete grid, so the simulation method needs to work on a discrete object. Considering these practical limitations, simulations based on the Kirchoff integral were discarded, even though this model is the most accurate.

One of the practical issues for such forward projection models is that now both the attenuation and the refraction function of the object need to be known, thus every voxel in the reconstruction volume should now contain two different values. In ultrasound CT, refraction effects can be corrected by simultaneously reconstructing both attenuation and refraction [35]. This is possible because during the acquisition both the amplitude and the time-of-flight of the emitted waves are

[14] Also known as the *law of refraction*

measured. In conventional X-ray CT, acquisition consists of a single projection containing the combined attenuation and phase contrast signal. This means that in theory, it is impossible to reconstruct both functions since only one combined measurement is available.

5.6.2 Model based on Snell's law

For the method using Snell's law or the angular deviation, the following problem is encountered. Normally, in the forward projection a ray is generated for each detector pixel. Each ray is then traced through the reconstruction volume while accumulating the interpolated values of the discrete object function. In conventional attenuation CT, these rays follow a straight path from the source to the detector. When refraction is introduced in this model, the rays may deviate from their original path. Consequently, some detector pixels are hit by several rays, while others are never hit. In the pixels that are never hit, the simulated intensity is calculated to be zero, which does not correspond to the real projection. To solve this problem, one could generate additional rays at intermediate positions or even generate a whole bunch of rays taking into account the finite spot size of the X-ray source. However, such an approach, which is much like a Monte-Carlo simulation, requires a large amount of calculations and is therefore not practically feasible.

An alternative solution is found by estimating for a certain detector pixel the direction of the corresponding incident ray at the source, where the ray is emitted. Using this approach, for every pixel again only one ray needs to be traced, which is computationally much more efficient. Finding this initial direction starting from the final position is known as the ray linking problem, for which a solution can be found for instance by using the method of successive approximation [36]. Even though this forward projection model is obviously slower than the original, attenuation-only model, it can be implemented rather efficiently. The problem with this model however is the accuracy when applied to a discrete grid. For the simulation of shapes that can be described analytically, this method yields rather accurate results. In case of a discrete object, the rays are deviated at each interface between two voxels, since every voxel usually contains a different reconstructed value for the refractive index. Due to this large number of calculations, a certain computational error accumulates. Even though this error is relatively small, since it is an angular deviation it is enlarged significantly in the projection image due to the large propagation distance. This problem of accuracy is a common problem for ray tracing methods of discrete objects when refraction is included. Simulations of several phantom objects using this model were performed, but the quality of the calculated phase signal was found to be insufficient. Hence, further evaluation of this forward projection model was not continued.

5.6.3 Model based on the TIE

A forward projection method based on the transport of intensity equation seems to be the most promising model for the development of a phase reduction algorithm, since it is rather straightforward and can be implemented efficiently. The simplest model of this type is based on the TIEW (equation (5.72)). In the forward projection step, two ray sums are now calculated simultaneously for each ray: the conventional sum for the attenuation and a second sum to accumulate the coefficients of the refraction, which are discrete approximations of the integrals μ_θ and φ_θ given by equations (5.70) and (5.71), respectively. Using the TIEW (equation (5.72)), the simulated intensity distribution can then be calculated from the two ray sums. By comparing this simulated projection image with the acquired projection, the correction image can be derived, which is then used to update the reconstruction volume. Alternatively, one can use the higher order approximation of the TIEW given by equation (5.89) or the remapping described by equation (5.24) to obtain a more accurate calculation of the intensity. However, no apparent improvements were observed using such simulations.

The issue of the simultaneous reconstruction of attenuation and refraction can be approximately solved by assuming that both coefficients are proportional. Another approach is to first apply a phase retrieval algorithm to the mixed projection images. The obtained phase projections can then be used to reconstruct the refraction function of the object. Meanwhile, the original, mixed projections are used to reconstruct the attenuation simultaneously, using the phase correction calculated from the forward projection of the refraction. It was shown previously that phase retrieval using MBA is not appropriate when applied to strong phase objects. However, when it is applied in the proposed phase reduction algorithm, it was found to perform well enough, since the calculation of the phase correction only requires a rough approximation of the actual refraction function.

In practice, the best results are obtained when MBA is applied using a small value for the attenuation correction parameter. Consequently, the phase signal and the resulting reconstruction of the refraction function are blurred. However, this blurred function seems to be much more suited for the calculation of the phase correction than a sharp version. This is because the simulation of the phase contrast intensity requires application of the Laplacian operator, which enhances sharp transitions. Hence, a blurred refraction function results in a nice and smooth phase contrast signal, whereas a sharp function results in a rough signal with strong variations. This sharp signal is very much subject to computational errors and to errors due to the approximations that are required and therefore the consequent phase correction is very unstable. Contrary, the blurred signal is much less error-prone, resulting in a more stable correction.

The problem with this method is that in the early steps of the iterative reconstruction, the reconstruction of the refraction function is not yet appropriate to allow for an accurate simulation of the phase signal. As a result, severe artefacts

are introduced into the reconstruction of the attenuation function, which are very hard to correct for in the later steps. Better results are obtained when the refraction function is reconstructed before starting the reconstruction of the attenuation function[15], instead of reconstructing both simultaneous.

An observant reader may have already noticed a certain redundancy in this approach. After retrieval of the phase signal, the refraction function is reconstructed. This function is then reprojected during the reconstruction of the attenuation function. Since for the proposed correction in practice only the projection of the phase signal is required, it is obvious that the whole process of back- and forward projection of the phase signal is unnecessary. In fact, all that is required for the attenuation reconstruction is two sets of projection images: the original, mixed projections and the retrieved phase projections. The mixed projections are then simply corrected using the phase projections, after which the attenuation function can be reconstructed. After making this simplification, it follows that the proposed algorithm has become independent of the iterative reconstruction algorithm for which it was originally developed. This is of high importance for practical applications, since this phase artefact reduction method can now be used in combination with any reconstruction algorithm. The phase reduction has simply become a preprocessing operation, which is applied directly to the projection data.

5.7 Bronnikov-aided correction

The Bronnikov-aided correction (BAC), presented by De Witte *et al.* in [37], was developed specifically to handle mixed projection images for which MBA is no longer adequate, the most important reason being the failure for relatively large remaining attenuation. Consider again the transport of intensity equation for an object with weak and almost homogeneous attenuation:

$$I_{\theta,d}(x,y) = I_{\theta,d,a}(x,y)\left[1 - \frac{d\lambda}{2\pi M}\nabla_{\perp}^{2}\varphi_{\theta}(x,y)\right] , \tag{5.91}$$

which expresses the total intensity distribution $I_{\theta,d}(x,y)$ as a product of the attenuation contrast $I_{\theta,d,a}(x,y)$ and the phase contrast $[1-(d\lambda/2\pi M)\nabla_{\perp}^{2}\varphi_{\theta}(x,y)]$. When applying the phase filter from MBA, given by equation (5.87):

$$p(\xi,\eta) = \frac{1}{\xi^2+\eta^2+\alpha} . \tag{5.92}$$

the phase contrast signal $\tilde{\varphi}_{\theta}(x,y)$ is retrieved from the mixed projection images. Since theoretically the phase signal of an attenuating object cannot be obtained

[15]Note that, even though quality is sufficient to simulate the phase signal for the correction, the quality of the reconstructed refraction function is rather poor. It only serves as an intermediate result, used to reduce the phase artefacts in the subsequent reconstruction of the attenuation function, which is of much higher quality

from a single in-line projection image, the function $\tilde{\varphi}_\theta(x,y)$ only represents an approximation of the real phase function $\varphi_\theta(x,y)$. Furthermore, this phase retrieval is only valid using certain assumptions such as weak and almost homogeneous attenuation. Since in the intended application area these conditions are no longer satisfied, $\tilde{\varphi}_\theta(x,y)$ is only a very rough approximation. Nonetheless, it appears to serve its purpose well enough. Instead of using the phase signal to reconstruct the object's refraction function as in MBA, this information is now used to correct the mixed projection images and to obtain the pure attenuation images by:

$$I_{\theta,d,a}(x,y) = \frac{I_{\theta,d}(x,y)}{1 - \frac{d\lambda}{2\pi M}\nabla_\perp^2 \tilde{\varphi}_\theta(x,y)} \;, \tag{5.93}$$

These processed projections can now be reconstructed using a conventional reconstruction algorithm, without introducing phase contrast artefacts in the cross-sections.

5.7.1 Practical

The BAC method can be described by the following procedure, as illustrated by figure 5.24. First, the phase filter (equation (5.87)) derived from the modified Bronnikov algorithm is applied to the mixed projection (fig. 5.24(a)) to retrieve the approximate phase function $\tilde{\varphi}_\theta(x,y)$ (fig. 5.24(b)). Next, the correction function is calculated as:

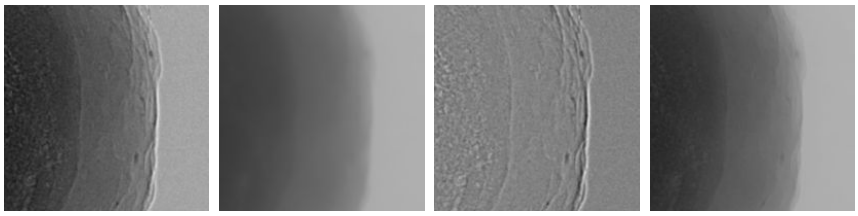
$$C_\theta(x,y) = 1 - \frac{d\lambda}{2\pi M}\,\nabla_\perp^2 \tilde{\varphi}_\theta(x,y) \;, \tag{5.94}$$

which represents an estimation of the phase contrast (fig. 5.24(c)). The original, mixed projection $I_{\theta,d}(x,y)$ (fig. 5.24(a)) is then divided by the correction function $C_\theta(x,y)$ to obtain the corrected projection $I_{\theta,d,a}(x,y)$ (fig. 5.24(d)) which should now only contain the attenuation signal. Finally, the set of corrected projections is reconstructed using a conventional algorithm. Note that the final result of this phase reduction method is the reconstructed attenuation function of the object, contrary to the previous phase retrieval methods from which the object's refraction function is reconstructed.

In practice, the correction function is calculated by:

$$C_\theta(x,y) = 1 - \gamma\,\nabla_\perp^2 \tilde{\varphi}_\theta(x,y) \;, \tag{5.95}$$

where the coefficient $\lambda d/(2\pi M)$ is replaced by the parameter γ. In theory, this parameter can be obtained directly from the experimental settings. However, when BAC is applied at a set-up using a polychromatic X-ray tube, determination of this parameter is not evident, as one needs to know both the emitted spectrum and the energy response of the detector quite accurately. It was therefore found more feasible to replace this value with a tunable parameter. This parameter γ is referred to as the *phase reduction parameter*.



(a) Mixed (b) Phase (c) Correction (d) Attenuation

Figure 5.24: Application of the BAC method to a bead, illustrating the different processing steps: the original, mixed projection image (a), the retrieved, approximate phase signal (b), the correction image (c) and finally the corrected attenuation image (d).

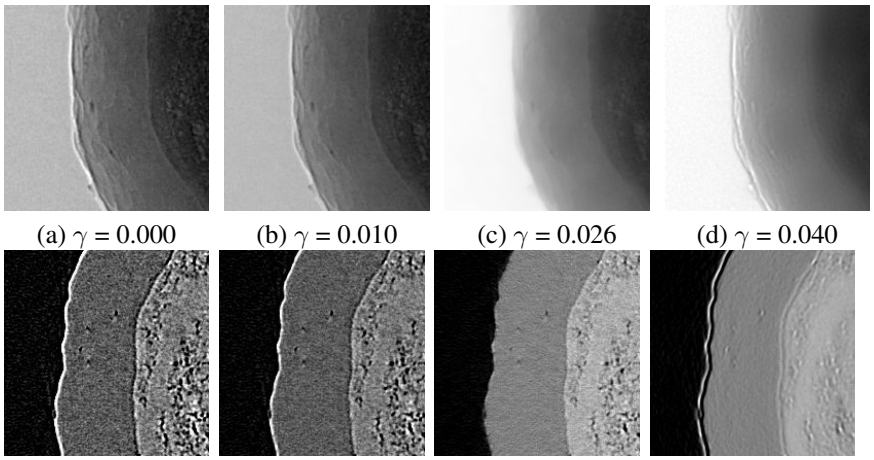
5.7.2 Phase reduction parameter

The optimal value of the phase reduction parameter γ can be found by evaluating both the corrected projections and the reconstructed images for different values of the parameter. The influence of γ is illustrated in figures 5.25 and 5.26, showing the intensity distribution and the reconstructed attenuation for different values of γ. Using the value $\gamma = 0$, BAC obviously simplifies to the conventional reconstruction of the uncorrected, mixed projection images. With increasing γ, the phase signal and the resulting artefacts are reduced, until at some point, the signal inverts and becomes larger again. The point just before this inversion corresponds to the optimal value of γ. For lower values, the phase signal is insufficiently reduced and phase artefacts still appear in the cross-sections. When γ is chosen too large, both the projections and the reconstructed images are severely distorted due to the apparent inversion of the phase signal. When using the optimal value of $\gamma = 0.026$, the phase signal almost completely disappears from the projection images. Consequently, the resulting cross-section can be reconstructed without introducing artefacts due to the phase contrast.

5.7.3 Reconstruction of strong phase objects

As was shown before, application of MBA on objects with a relatively high remaining attenuation does not yield satisfactory results. In contrast with this phase retrieval method, BAC was developed to retrieve the attenuation contrast from the mixed projection images. As BAC reconstructs the attenuation function of the object, it is expected that BAC provides better results than MBA when reconstructing projection data from phase objects with moderate or strong remaining attenuation.

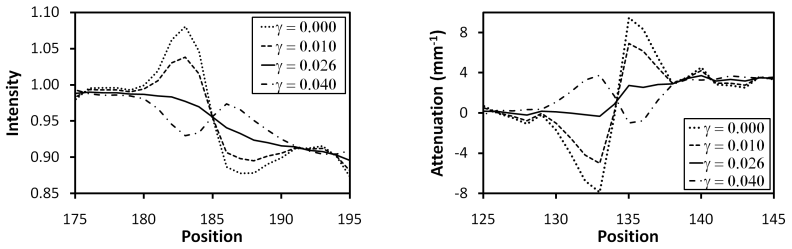
An evaluation of BAC as function of the attenuation by the object is provided by a test similar to the one for MBA in section 5.5.3. The approximate phase function $\tilde{\varphi}_\theta(x, y)$ was retrieved by applying MBA to the mixed projections, where the attenuation correction parameter α was chosen equal to the optimal value that was found for the MBA reconstructions. The optimal value of γ was found by



(a) γ = 0.000 (b) γ = 0.010 (c) γ = 0.026 (d) γ = 0.040

(e) γ = 0.000 (f) γ = 0.010 (g) γ = 0.026 (h) γ = 0.040

Figure 5.25: The influence of the phase reduction parameter γ of BAC on the corrected projection images (b-d) and the resulting cross-sections (f-h), for different values of γ. The conventional reconstruction (e) of the mixed projection images (a) corresponds to $\gamma = 0.000$.



(a) Projection intensity (b) Reconstructed attenuation

Figure 5.26: The influence of the phase reduction parameter γ of BAC. Line profiles of the intensity distribution in the projection images (a) and of the reconstructed attenuation function (b) are shown for different values of γ, compared to the conventional reconstruction of the mixed projections (corresponding to γ = 0.000).

reconstructing the central cross-section for several values of the parameter and selecting the one that yields the lowest reconstruction error.

The results of this test are presented in figure 5.27, where measurements of the reconstruction error are compared between the conventional, MBA and BAC reconstruction. It can be seen that the NRMSE for BAC decreases with increasing attenuation, whereas the opposite is true for MBA. As a result, BAC provides better quality for higher remaining attenuation and is thus better suited for the reconstruction of strong phase objects. The point at which MBA becomes the better candidate is not predetermined, but depends on the experimental conditions.

For completeness, the NRMSE of the conventional reconstruction is also given, clearly showing that either MBA or BAC can provide a significant improvement in quality when applied in the reconstruction of mixed projection images.

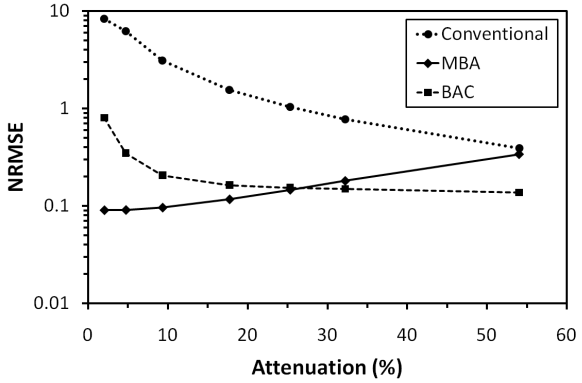


Figure 5.27: The NRMSE of BAC as function of the attenuation by the object, as compared to MBA and the conventional reconstruction.

5.8 MBA and BAC in practice

The increase in resolution of today's state-of-art high resolution X-ray CT scanners is mainly due to the development of X-ray sources with a decreasing focal spot size. Phase contrast due to free space propagation can occur at tube based systems when the focal spot size is sufficiently small. The appearance of phase contrast is thus not limited to synchrotron facilities but also arises at many other high resolution systems. As was shown before, reconstruction of the resulting mixed projection images using conventional reconstruction algorithms yields severe artefacts. Consequently, phase contrast artefacts are encountered at many high resolution systems, posing a limitation to the obtained quality.

Fortunately, the reconstruction quality can be significantly increased by applying either MBA or BAC. Following these important results, both complementary solutions were added to the reconstruction software Octopus. This allows users of other high resolution systems to improve their reconstruction results as well. Both methods are implemented as a preprocessing routine which acts on the normalised projection images. A screenshot of the module, that incorporates both phase filters in a user-friendly way, is shown in figure 5.28.

Since in practice a quantification of the reconstruction quality based on the NRMSE is not possible, optimization of the parameters α and γ needs to be performed in another way. Currently, the best way to find these parameters, which vary for each scan, is a qualitative comparison of the retrieved projections, which can usually be performed using only a single projection image.

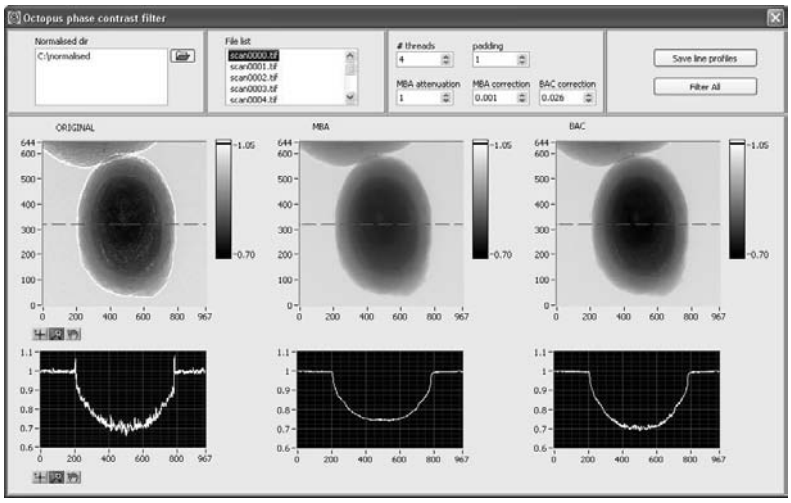


Figure 5.28: Screenshot of the Octopus Phase Contrast Filter module.

For MBA, only one parameter needs to be found, the attenuation correction parameter α. When using lower values α, the retrieved phase projections are blurred. For higher values, the phase edges reappear in the retrieved images which lead to artefacts in the reconstruction. Since in practice most objects consist of multiple components, it is impossible to find the correct α due to the different ratios between the attenuation and phase coefficients of the different materials. Hence, a compromise needs to be made between blurring and phase artefacts. This compromise is subjective and depends on how the user intends to further process the reconstructed cross-sections. If the image quality is most important, it may be best to choose a slightly higher α since otherwise too much detail may be lost due to the blurring. For quantitative analysis on the other hand it may be beneficial to allow some degree of blurring if this removes the phase artefacts completely.

Application of BAC seems more complicated, as this requires optimization of both the attenuation correction parameter α and the phase reduction parameter γ. Fortunately, in practice both parameters can be chosen almost independently from each other. Parameter α is best found by evaluating the retrieved phase projections. In contrast with MBA however, there is no need for a compromise. As was mentioned earlier, BAC performs much better when using a smooth version of the phase projections. Thus, α is best chosen slightly smaller than its optimal value for MBA, such that the retrieved phase function no longer contains any phase edges and the image is slightly blurred. The second parameter γ can be obtained by starting at 0 and increasing the value until the phase signal inverts. The optimal value of γ is found right before this point of inversion. However, in some applications using the optimal value results in a slight blurring of the edge of the object. If this is the case, it may be beneficial to use a slightly smaller γ. Even though the phase artefacts are then not completely removed, they are significantly reduced while

the edges of the object are preserved. In fact, by making a marginal modification to the reconstruction algorithm, it is possible to witness the influence of γ in real time, based on the reconstruction of a single cross-section through the object. This modification obviously makes it much easier to find the optimal value for γ.

The computational burden for both MBA and BAC can be minimized by using an efficient implementation of the fast Fourier transform (FFT). On the workstation described in table 3.1, typical processing times for either of these methods, including reading and writing of the images, are only a couple of minutes for small data sets (typically 800 projections of 1000×1000 pixels) and 10-15 minutes for larger data sets (typically 1000 projections of 1500×1500 pixels).

5.9 Applications

In this section the practical use of both the phase retrieval and phase reduction methods is illustrated by applying them to real projection data. The resulting reconstructions are compared with the conventional reconstruction of the mixed projections. This clearly demonstrates the improvement in image quality that can be obtained by simply applying either of the two methods, without having to modify the scanner set-up or the scanning protocol. The projection data for the first examples were acquired using the high resolution CT scanners available at UGCT. In the last four examples results are presented for data from other high resolution systems. The reconstructions of both unprocessed and processed projections are performed using the standard FDK algorithm.

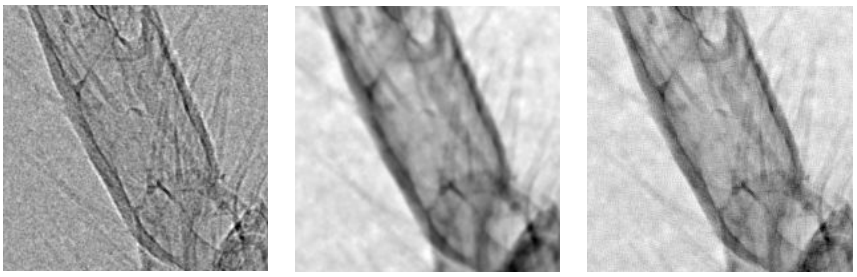
5.9.1 Leg of a fly

The first application example is the scan of the leg of a fly, which is used to illustrate the effect of both MBA and BAC on a typical biological sample. The results for this evaluation were published in [37], where BAC was first presented. The leg has a diameter of about 100 µm. The parameters for this scan are given in table 5.1. In figure 5.29, an original projection of the leg is shown, in which one can clearly observe the presence of a rather strong phase contrast, in addition to the conventional attenuation contrast. By applying MBA to this projection, the phase signal is retrieved and converted to a phase projection, which can be reconstructed using a conventional algorithm to obtain the refraction function of the object. It can be seen that the retrieved phase projection is slightly blurred compared to the original projection. Alternatively, BAC can be applied to the mixed projection to reduce the phase signal, which allows reconstruction of the attenuation function where the typical phase artefacts are significantly reduced. The corrected projection appears to be as sharp as the original image.

After applying MBA and BAC to all projections one can obtain the MBA and BAC reconstruction of the sample using a conventional reconstruction algorithm.

Table 5.1: Scan settings for the leg of a fly.

# Projections:	800	
# Detector rows:	624	
# Detector columns:	963	
Pixel pitch:	38	µm
Source to object:	4	mm
Source to detector:	241	mm
Cone angle:	5.6	°
Angular step:	0.45	°
Voxel size:	0.63	µm

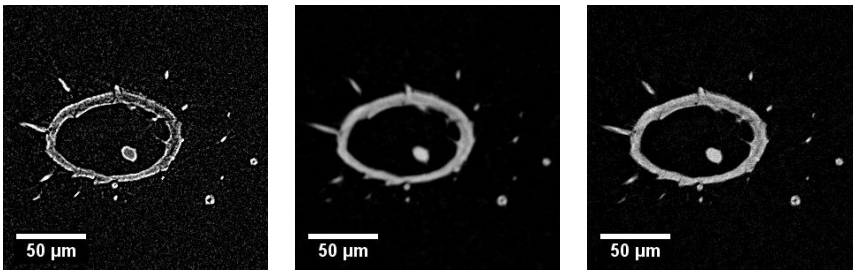


(a) Original (b) MBA (c) BAC

Figure 5.29: An original projection image of the leg of a fly (a) and the projections that are retrieved from it using MBA (b) and BAC (c).

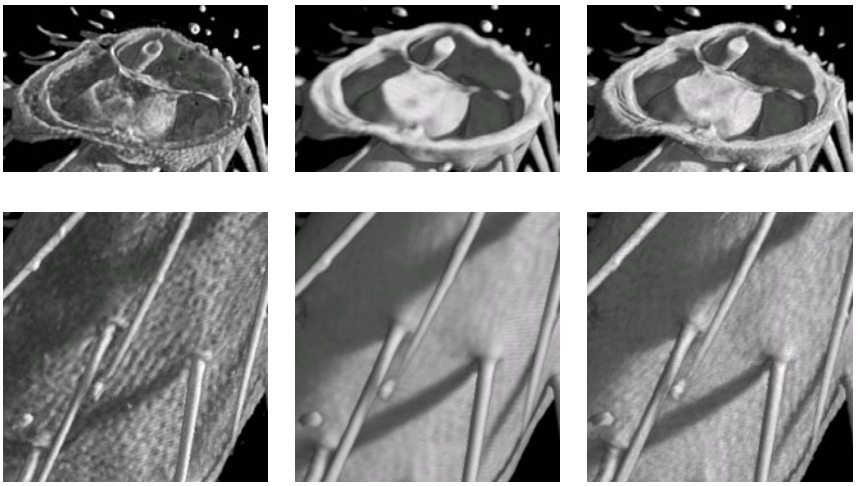
The central cross-section of the reconstructed fly leg is presented in figure 5.30, comparing results between the different reconstruction methods. The three resulting images are clearly different from each other. The conventional cross-section is very sharp, but it suffers from severe phase artefacts. Because of these artefacts, structures that should be rather homogeneous appear to consist of a strongly attenuating edge and an inner region of very weak attenuation that is hardly distinguishable from the background noise. In the MBA cross-section, the phase artefacts have almost completely disappeared, but the image is blurred and some detail is lost. Looking back at figure 5.29, this slight blurring was expected since the retrieved phase projection appears less sharp than the original projection. The BAC cross-section on the other hand does not suffer from this blurring and remains very sharp, while the phase artefacts have been greatly reduced. Both MBA and BAC also result in lower relative noise levels than the conventional reconstruction.

In figure 5.31, 3D renderings of the three reconstructed volumes are shown. The three upper images are created by making a virtual cut through the 3D volumes, revealing the internal structure of the leg. Due to the phase artefacts, the conventional reconstruction seems to suggest that the outer structure of the leg



(a) Conventional (b) MBA (c) BAC

Figure 5.30: Reconstruction of the fly leg using the conventional reconstruction (a), as compared to the reconstruction of the refraction function after application of MBA (b) and the attenuation reconstruction after applying the phase correction using BAC (c).



(a) Conventional (b) MBA (c) BAC

Figure 5.31: 3D renderings of the reconstructed volumes using the conventional (a), MBA (b) and the BAC (c) reconstruction.

consists of a hollow region between strongly attenuating surfaces, which obviously does not correspond to the physical reality. For both MBA and BAC reconstructions, these structures are nicely filled. The BAC reconstruction also reveals several details in the leg's structure, which cannot be observed in the MBA reconstruction due to the blurring. An example of this can be seen on the left side of the upper images, in the outer structure of the leg, where the BAC image reveals gaps in the outer structures of the leg that cannot be detected in the MBA image. A second example can be seen in the lower images, in which it can be seen that the leg is covered by a bubble-like texture[16]. Due to the blurring, this texture can

[16]This texture is in fact due to small hairs on the leg.

hardly be distinguished in the MBA reconstruction, whereas it is much clearer in the conventional and the BAC reconstruction.

To conclude, it is found that both MBA and BAC provide a result of much higher quality than the conventional reconstruction, which fails due the strong presence of phase artefacts. Using MBA, the phase artefacts are completely removed, but the result is slightly blurred. In the cross-sections resulting from the BAC reconstruction, the phase artefacts are greatly reduced while the details in the internal structure of the sample remain intact. It is important to note that, although both methods produce similar images, they actually reconstruct different physical quantities: the MBA method reconstructs the refraction function, whereas BAC reconstructs the attenuation function of the sample. This means that any comparison between the two methods should always be interpreted with caution.

5.9.2 Bead

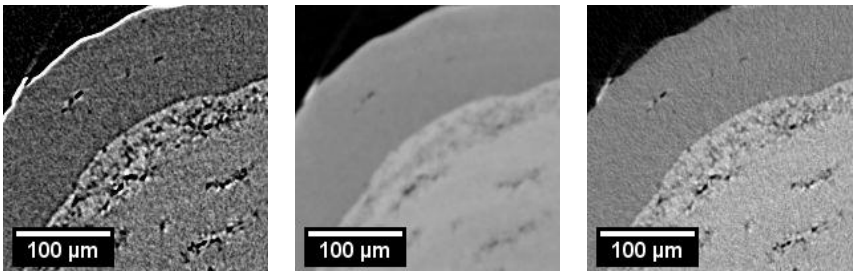
The second example is a pharmaceutical bead, with a diameter of about 850 µm. It consists of a dense core surrounded by a less dense mantle. Both the core and the mantle consist of homogeneous material with air inclusions trapped inside. The structure of the core and mantle are important in pharmaceutical product development, since they determine the solubility of the bead. This application example was also published in [37]. Ruxandra Govoreanu (Johnson & Johnson) is acknowledged for providing this sample.

Table 5.2: Scan settings for the bead.

# Projections:	800	
# Detector rows:	644	
# Detector columns:	967	
Pixel pitch:	38	µm
Source to object:	10	mm
Source to detector:	270	mm
Cone angle:	5.2	°
Angular step:	0.45	°
Voxel size:	1.4	µm

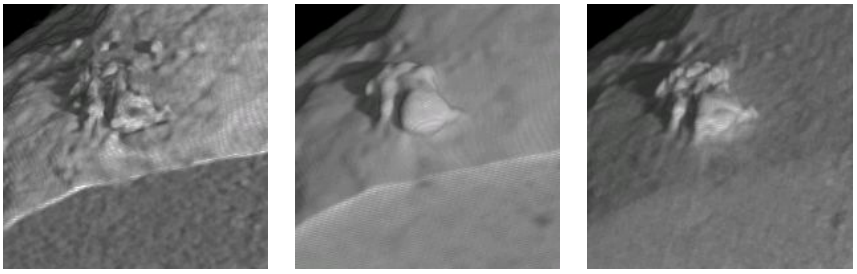
Table 5.2 summarizes the parameters of the scan. Resulting cross-sections for each of the three reconstruction methods are shown in figure 5.32. The mantle and core can easily be distinguished in all three images. However, in the conventional cross-section, two more components can be detected. The first is a strongly attenuating border that surrounds the beads mantle. The second is a faint black and white borderline between the mantle and the core. Neither of these components represents an actual physical feature, but they are easily identified as phase artefacts. Other phase artefacts can be observed in the form of a thin, white border

surrounding the different air inclusions. In the MBA cross-section, these artefacts have been completely removed, but the image is severely blurred when compared to the conventional and BAC results. Also, in the MBA cross-section, the mantle is not homogeneous, but seems to become more attenuating when moving away from the bead's centre. This cupping is also observed in the core of the bead, albeit in a more subtle way. As was shown previously, this cupping results from the remaining attenuation by the bead, which goes up to 25%. In the BAC cross-section, the phase artefacts around the mantle and between the mantle and core have vanished. Phase artefacts surrounding the inclusions have been heavily reduced. Neither mantle or core show any sign of cupping and they both present a very homogeneous attenuation, while all details are preserved.



(a) Conventional (b) MBA (c) BAC

Figure 5.32: The conventional reconstruction of the bead (a), compared to the MBA (b) and BAC reconstruction (c).



(a) Conventional (b) MBA (c) BAC

Figure 5.33: 3D renderings of the reconstructed volumes of the bead.

Figure 5.33 shows 3D renderings of the three reconstructed volumes which have been virtually cut. It can again be seen that the conventional reconstruction suggests the existence of a strongly attenuating border surrounding the surface. A similar border can also be seen in the MBA image, but here it is due to the cupping artefact in the mantle. This wrongly emphasized border is not present in the BAC reconstruction, which instead shows a highly homogeneous mantle with a very clean transition between the mantle and the air around the bead. The BAC image also reveals much more detail on the mantle's surface than the MBA image does.

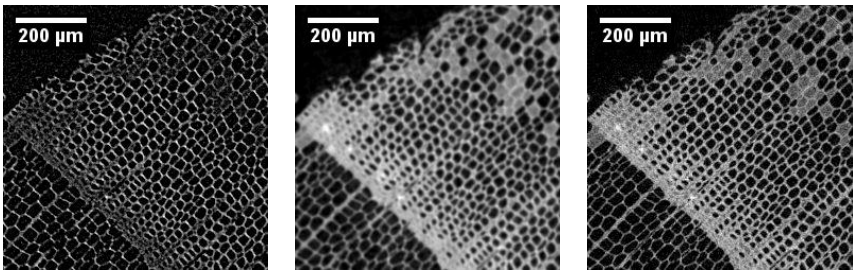
This example clearly illustrates the failure of MBA when applied to a strongly attenuating phase object. Due to the remaining attenuation, the resulting cross-sections are severely distorted by the introduction of a cupping artefact. The BAC method was shown to deal very well with phase artefacts, without reducing the sharpness and level of detail of the reconstructions.

5.9.3 Fungus in wood

A third application which is used to illustrate both MBA and BAC is the reconstruction of the scan of a piece of pine wood. This piece was taken from a tree trunk at the edge of a growth ring, so the wall thickness of the vessels inside the wood is different at different locations in the sample. A fungus was grown on this sample, filling some of the wood vessels. The scientific interest of this sample lies in the determination of the growth rate of the fungus inside the wood. Jan Van den Bulcke (Laboratory of Wood Technology, Universiteit Gent) is acknowledged for providing this sample.

Table 5.3: Scan settings for the piece of wood with fungus inside.

# Projections:	1000	
# Detector rows:	940	
# Detector columns:	748	
Pixel pitch:	254	µm
Source to object:	10	mm
Source to detector:	860	mm
Cone angle:	15.8	°
Angular step:	0.36	°
Voxel size:	3.0	µm



(a) Conventional (b) MBA (c) BAC

Figure 5.34: The conventional (a), MBA (b) and BAC (c) reconstructions of the piece of wood containing fungus.

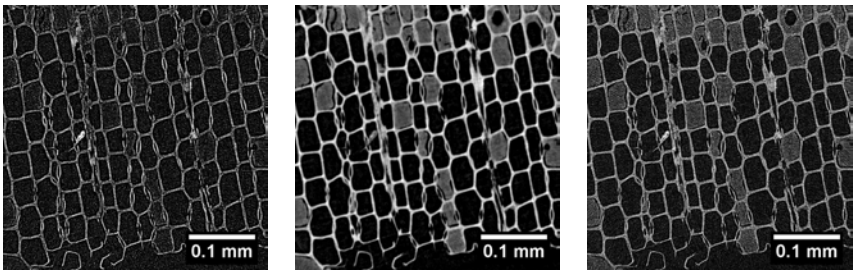
The scan parameters for this sample are given in table 5.3. The central cross-sections of the reconstructions are shown in figure 5.34. Using the conventional reconstruction, only the edges of the wood vessels, which are enhanced by the phase contrast, are reconstructed. Consequently, the reconstructed vessel walls appear very thin and there seems to be only little difference in wall thickness between the two regions defined by the growth ring. Furthermore, the fungus is practically invisible so it is impossible to determine its growth. Both MBA and BAC provide a much more accurate representation of the scanned sample. In both reconstructions, the fungus can now clearly be observed. The vessel walls are also reconstructed with a certain thickness, which is different in the two different regions. The resulting MBA cross-section is slightly blurred as compared to the conventional and BAC cross-sections. In contrast with the MBA image, BAC seems to reveal some of the internal structure of the fungus. However, it was found that this internal structure is below the resolution of this CT-scan. Combined with the higher noise level in the BAC image, this apparent presence of structure is inconclusive.

Table 5.4: Scan settings for the second piece of wood with fungus inside.

# Projections:	800	
# Detector rows:	800	
# Detector columns:	970	
Pixel pitch:	29.6	µm
Source to object:	3.3	mm
Source to detector:	164	mm
Cone angle:	8.3	°
Angular step:	0.45	°
Voxel size:	0.6	µm

Results for a similar example of the growth of fungus inside wood vessels are shown in figure 5.35. The parameters for this scan are given in table 5.4. These results originate from one of the first scans using the new 400nmCT-scanner developed at UGCT. As with the previous sample, when no phase processing is applied, only the edges introduced by the phase contrast are reconstructed. For the thinnest walls of the wood vessels, this is not much of a problem as for such thin structures reconstructing only the edges gives almost the same results as when reconstructing the bulk as well. For the thicker walls however, reconstruction of the bulk is necessary, as several of these walls contain holes and without the bulk it is not possible to determine which is an actual hole and which is due to an artefact. Furthermore, the fungus that has grown inside some of the vessels cannot be detected as it is very difficult to determine whether the grey values inside the vessel are due to noise or the fungus. Although with the naked eye one may locate the fungus, any segmentation or automated processing is bound to fail. As one can see, in the MBA reconstruction the fungus is clearly visible and the walls of the vessels all come out very nice. Although a slight loss of resolution can again

be observed, this reconstruction is very suitable for further analysis. The BAC reconstruction also clearly reveals the fungus and fills the vessel walls. It does not suffer any loss in resolution, but contains more noise than the MBA reconstruction.



(a) Conventional (b) MBA (c) BAC

Figure 5.35: The conventional (a), MBA (b) and BAC (c) reconstructions of the second piece of wood with fungus grown inside the wood vessels.

5.9.4 Horse biopsy

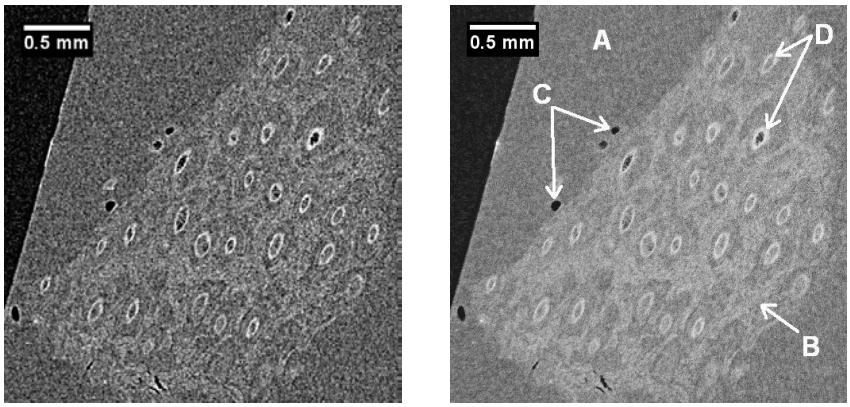
A nice application of the BAC method is the reconstruction of a horse biopsy. Since biological tissue typically has a low X-ray attenuation coefficient, it is very susceptible to phase artefacts, especially when it concerns a small sample. The sample is a skin punch biopsy obtained from a Belgian draft horse at the left lateral neck region. Prior to biopsy, the horse was injected with Lipiodol to increase contrast of skin lymphatic vessels. After biopsy, the sample was fixed in formalin, dehydrated and embedded in paraffin. In a paper by Cnudde *et al.* [38], a similar sample of equine origin was used to discuss the advantages of using high resolution X-ray CT for virtual histology. In that paper, the results were obtained using the conventional reconstruction, since the MBA and BAC methods were not yet available at that time. Application of BAC on this data was shown to yield a more accurate and reliable reconstruction. The following results and discussion were published in [39]. Hilde De Cock (Department of Veterinary Medicine, Universiteit Antwerpen) is acknowledged for providing the sample.

The parameters for the scan of the biopsy are summarized in table 5.5. Results of the BAC reconstruction are compared with those of the conventional reconstruction in figure 5.36. As is indicated in the images, the skin consists of hair follicles surrounded by thick and randomly crossing collagen bundles. The area of weak attenuation surrounding the biopsy is the paraffin, which contains several air inclusions. The hair shaft and the root sheath of the follicles can be clearly distinguished from each other, because the root sheath has become much more attenuating due to the contrast agent.

When looking closely at the images, important differences can be observed between both cross-sections. First of all, the BAC cross-section seems to suffer

Table 5.5: Scan settings for the horse biopsy.

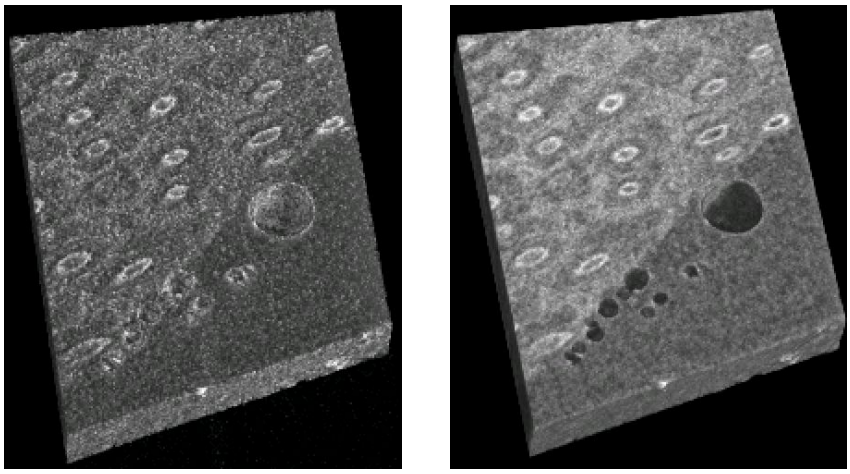
# Projections:	1000	
# Detector rows:	1880	
# Detector columns:	1496	
Pixel pitch:	127	µm
Source to object:	33.8	mm
Source to detector:	890	mm
Cone angle:	15.3	°
Angular step:	0.36	°
Voxel size:	6.0	µm



(a) Conventional (b) BAC

Figure 5.36: Conventional (a) and BAC (b) reconstruction of the horse biopsy. The letters in figure (b) denote the different features of the biopsy: (A) paraffin, (B) collagen bundles, (C) air inclusions and (D) hair follicles.

much less from noise than the conventional cross-section and has a better contrast. Therefore, there is a much clearer distinction between the paraffin and the collagen bundles, which improves the general image quality. Because of this, segmentation of the sample based on a simple threshold can be performed more easily and more accurately, which improves the quantitative analysis of the sample by further image processing. In the conventional cross-section, one can clearly notice phase artefacts at the transition from the paraffin to the surrounding air. Typical phase artefacts are also seen around the air inclusions in the paraffin. Although the enhancement of such edges can be beneficial in some applications, in this sample it is highly undesirable since it causes a degradation of the general image quality. Furthermore, it can also lead to faulty interpretations about the structure of the sample.



(a) Conventional (b) BAC

Figure 5.37: 3D renderings of the reconstructed horse biopsy.

An example of such possible mistakes is the air trapped in the paraffin. Due to edge enhancement, the air inclusions have a strongly attenuating border surrounding them. This causes some of them to look similar to the hair follicles, especially when the inclusion is noisy. While most of these fake follicles can be identified as inclusions by looking at their position and size, others might be misinterpreted and classified as hair follicles. Using BAC, the phase artefacts in the resulting cross-sections are significantly reduced, thus providing a more accurate description of the interior of the biopsy. Also, since the air inclusions no longer contain a strongly attenuating border, they are far less likely to be interpreted as being hair follicles and thus analysis of the reconstructed slices is more reliable.

In figure 5.37, 3D renderings are shown for both reconstructions. The renderings reveal the same differences as the cross-sections, so the same remarks can be made here as well. Due to the reduced noise and the better contrast, the different regions and features can be identified and displayed more easily in the BAC volume. The figure also emphasizes how difficult it is to separate the air inclusions from the hair follicles in the conventional reconstruction, which is far more obvious using the BAC method.

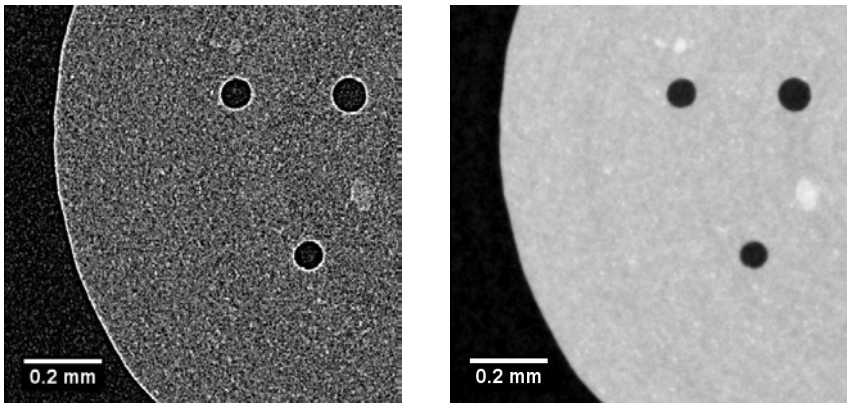
5.9.5 Polymer

The MBA method was used in an application regarding sustained drug release. A drug was inserted into a polymer matrix, consisting of ethylene vinyl acetate, through a process known as *hot-melt extrusion*. High resolution tomography was used in this research to quantify the porosity of a small tablet of this fabricate

before and after a dissolution experiment. Ana Almeida (Laboratory of Pharmaceutical Technology, Universiteit Gent) is acknowledged for providing the sample.

Table 5.6: Scan settings for the polymer.

# Projections:	1200	
# Detector rows:	1880	
# Detector columns:	1496	
Pixel pitch:	127	µm
Source to object:	14.3	mm
Source to detector:	890	mm
Cone angle:	15.3	°
Angular step:	0.3	°
Voxel size:	2.0	µm



(a) Conventional (b) MBA

Figure 5.38: Reconstructed cross-sections of the polymer matrix, comparing the conventional (a) with the MBA (b) reconstruction.

The scan parameters are shown in table 5.6. The central cross-sections for the resulting conventional and MBA reconstructions are presented in figure 5.38. There are several features in the sample that contain a slightly higher reconstructed value than the rest of the matrix. Due to the phase contrast, in the conventional reconstruction the air inclusions in the polymer matrix are surrounded by a strongly attenuating border, which is physically not present. In a cross-section, one can observe several regions of stronger attenuation. However, it is not possible to determine whether this stronger attenuation represents a feature or whether it corresponds to the top or bottom of an air inclusion. Therefore, the phase artefacts preclude any accurate segmentation of the strongly attenuating features. In the MBA cross-section, these phase artefacts surrounding the air inclusions are nowhere to

be seen. Hence, the reconstruction using MBA allows for a reliable segmentation of the features.

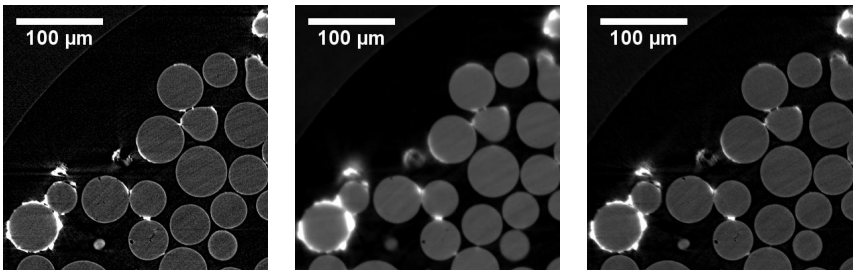
5.9.6 Sintered glass balls

A further validation of MBA and BAC is provided by applying both methods to projection data acquired at other high resolution systems. The first such system is a synchrotron based system, the Hard X-Ray Micro Tomography Beamline 8.3.2 of the Advanced Light Source at the Lawrence Berkeley National Laboratory. The object under study is a collection of sintered glass balls with potassium iodide deposited non uniformly. This sample is used as a model sample for pore flow measurements. Alastair MacDowell, Jamie Nasiatka (Advanced Light Source, Lawrence Berkeley National Laboratory) and Jonathan Ajo-Franklin (Earth Sciences Division, Lawrence Berkeley National Laboratory) are acknowledged for providing the projection data of this sample.

This tomography installation uses a highly parallel, monochromatic beam, which was operated at an X-ray energy of 35 keV. Further parameters for this scan are given in table 5.7. The central cross-sections of each of the three reconstructed volumes is shown in figure 5.39. In the conventional reconstruction, each of the glass balls seems to be surrounded by a highly attenuating border. Due to this phase artefact, these borders cannot be separated from the deposit, which also has a higher attenuation value. This makes segmentation of the deposit for further quantitative analysis almost impossible. The reconstruction using MBA does not contain any of these phase artefacts, enabling a clear separation between the glass balls and the deposit. However, the reconstruction seems to be slightly blurred, which causes the deposit to be increased in size. In the BAC reconstruction, phase artefacts can still be observed, but they have been reduced sufficiently to allow for an accurate segmentation. By removing or reducing the artefacts resulting from phase contrast, both MBA and BAC thus provide images of higher quality and allow for an easier and much more reliable analysis of the deposit.

Table 5.7: Scan settings for the sintered glass balls.

# Projections:	1800	
# Detector rows:	250	
# Detector columns:	2600	
Pixel pitch:	4.43	µm
Source to detector:	70	mm
Angular step:	0.1	°
Voxel size:	4.43	µm



(a) Conventional (b) MBA (c) BAC

Figure 5.39: The central cross-sections of the conventional (a), MBA (b) and BAC (c) reconstructions of the sintered glass balls.

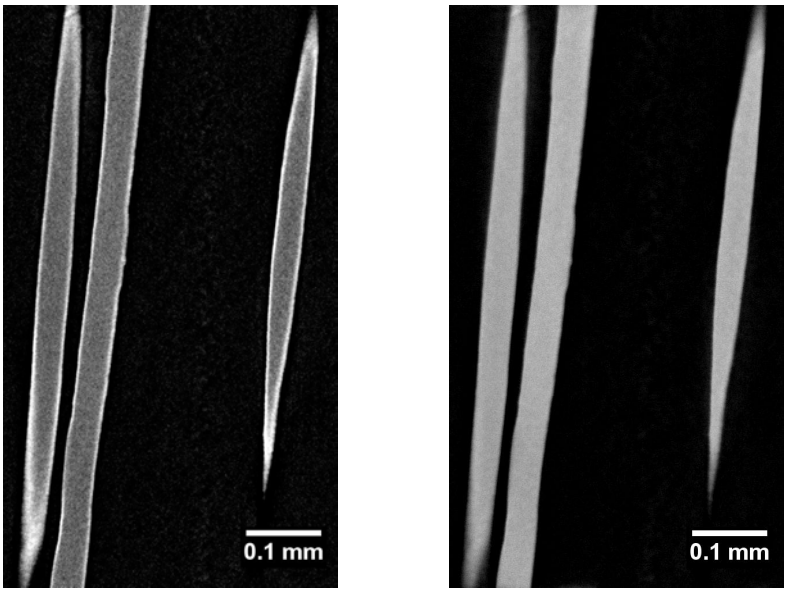
5.9.7 Human hair

The next application considers the scan of human hair. This scan was performed at the High Resolution X-ray Computed Tomography Facility at The University of Texas at Austin, using a commercially available device, the Micro-XCT-400 from Xradia. Richard Ketcham and Jessica Maisano (High Resolution X-ray Computed Tomography Facility, The University of Texas at Austin) are acknowledged for sharing this data.

Table 5.8: Scan settings for the human hair.

# Projections:	666	
# Detector rows:	1024	
# Detector columns:	1024	
Pixel pitch:	1	µm
Source to object:	40.5	mm
Source to detector:	53	mm
Angular step:	0.277	°
Voxel size:	0.76	µm

The projection data were acquired in short scan mode, covering a range of 184°. The scan parameters are summarized in table 5.8. Results for both the conventional and the BAC reconstruction are shown in figure 5.40. As was to be expected, the conventional reconstruction contains phase artefacts at the transition from the hairs to the surrounding air. Using BAC, these artefacts are almost completely removed, providing a reconstruction that represents the actual physical structure much more accurately. The MBA reconstruction, which is not shown here, resulted in an similar, but slightly blurred image as compared to the BAC image.



(a) Conventional (b) BAC

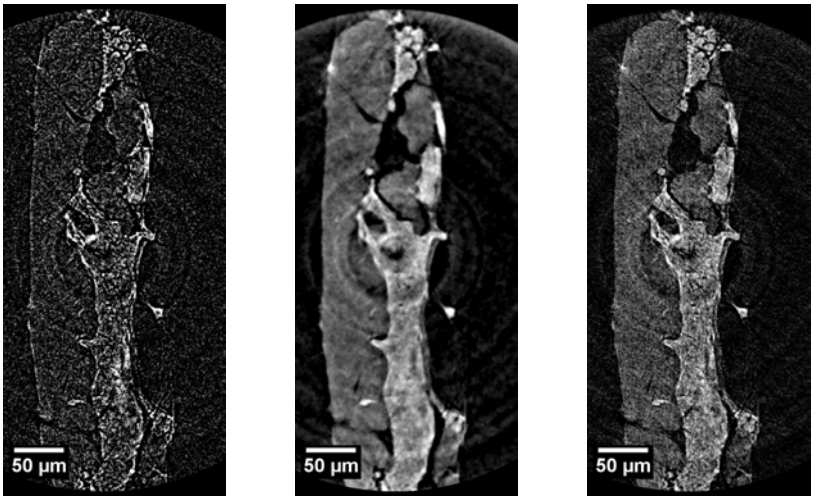
Figure 5.40: Reconstruction of the human hair obtained by the conventional (a) and the BAC (b) reconstruction. The cross-sections are taken parallel to the XZ-plane.

5.9.8 Carbon filter

In the following application, the projections were acquired using a Skyscan 2011 nanotomograph. Elke Van de Casteele and Bart Pauwels (SkyScan) are acknowledged for providing the data. The scan parameters are given in table 5.9, projections were taken over a range of 185.6°. The object under study is a carbon filter, which mainly consists of two components. Results for the three reconstruction methods are presented in figure 5.41. In the conventional reconstruction, the weakly attenuating component of the filter can hardly be separated from the background noise. The only visual separation is the phase edge around the component. The strongly attenuating component also does not appear as a solid structure, but instead consists of a collection of strongly attenuating phase edges. The MBA reconstruction reveals both components to be solid structures and although losing some resolution, provides a much better representation of the object. However it seems to deteriorate the ring artefacts, especially at the outer edge of the reconstructed volume. This is because the normalisation of the projections could not be done properly and the background intensity within a single projection image varies. As was already mentioned, the application of MBA then introduces a dc-shift in the retrieved images, which distorts the quality of the reconstructed images. The BAC reconstruction shows both weakly and strongly attenuating component clearly. It contains more noise than the MBA reconstruction but in return hardly

Table 5.9: Scan settings for the carbon filter.

# Projections:	464	
# Detector rows:	1024	
# Detector columns:	1280	
Pixel pitch:	10.6	µm
Source to object:	5.4	mm
Source to detector:	149.4	mm
Cone angle:	4.2	°
Angular step:	0.4	°
Voxel size:	0.39	µm



(a) Conventional (b) MBA (c) BAC

Figure 5.41: Reconstructed cross-sections of the carbon filter using the conventional (a), MBA (b) and BAC (c) reconstruction.

suffers from any loss in detail detectability.

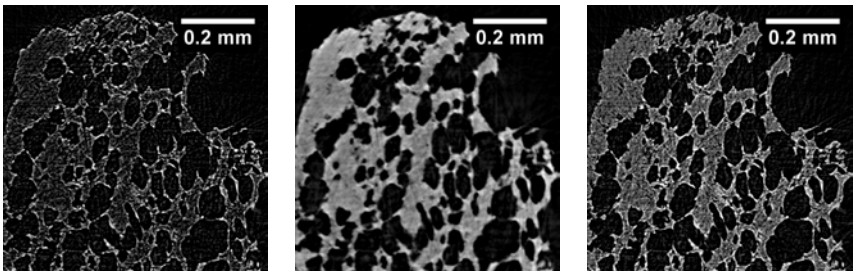
5.9.9 Biscuit

The last example presents the application of MBA and BAC in the reconstruction of a piece of biscuit. The projections were taken with an experimental PCX system from XRT. Les Brownlow (XRT®) is acknowledged for his support and for providing this data. The projections were acquired covering an angular range of 185°. The scan parameters are summarized in table 5.10. Results of the reconstructions are shown in figure 5.10. Again, due to the appearance of phase contrast in the

Table 5.10: Scan settings for the biscuit.

# Projections:	1552	
# Detector rows:	1730	
# Detector columns:	1780	
Pixel pitch:	9	µm
Source to object:	35	mm
Source to detector:	350	mm
Cone angle:	2.5	°
Angular step:	0.4	°
Voxel size:	0.9	µm

projection images, only the outer edges of the sample are reconstructed when no appropriate phase processing is applied, while the bulk area of the structure is hardly distinguishable from the noise in the background. This makes it look as if the biscuit is actually a hollow structure, which is obviously incorrect. This also makes it very difficult to analyse the sample afterwards, as this usually requires a binarisation of the volume, which is nearly impossible. The MBA reconstruction does not contain any phase edges and results in a realistic representation of the biscuit where the bulk area is now clearly visible. Despite the slight loss of resolution the volume is very suitable for further analysis. Using BAC, the phase edges are reduced significantly and the bulk area of the object is reconstructed rather well. Unlike the MBA cross-section, the phase artefacts are not completely removed in the BAC images, but in return no loss of detail is observed.



(a) Conventional (b) MBA (c) BAC

Figure 5.42: Central cross-sections of the biscuit obtained using the conventional (a), MBA (b) and BAC (c) reconstruction.

References

[1] D. Gabor. *A new microscopic principle*. Nature, 161(4098):777–778, 1948.

[2] B. L. Henke, E. M. Gullikson, and J. C. Davis. *X-ray interactions: photoabsorption, scattering, transmission and reflection at E = 50-30,000 eV, Z = 1-92*. Atomic Data and Nuclear Data Tables, 54:181–342, 1993.

[3] `http://henke.lbl.gov/optical_constants/asf.html`.

[4] M. Born and E. Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Pergamon, London, 1959.

[5] M. El-Ghazaly, H. Backe, W. Lauth, G. Kube, P. Kunz, A. Sharafutdinov, and T. Weber. *X-ray phase contrast imaging at MAMI*. European Physical Journal A, 28(1):197–208, 2006.

[6] T. J. Davis, D. Gao, T. E. Gureyev, W. Stevenson, and S. W. Wilkins. *Phase-contrast imaging of weakly absorbing materials using hard X-rays*. Nature, 373(6515):595–598, 1995.

[7] A. Peterzol, A. Olivo, L. Rigon, S. Pani, and D. Dreossi. *The effects of the imaging system on the validity of the ray-optical approach to phase contrast imaging*. Medical Physics, 32(12):3617–3627, 2005.

[8] S. W. Wilkins, T. E. Gureyev, D. Gao, A. Pogany, and A. W. Stevenson. *Phase-contrast imaging using polychromatic hard X-rays*. Nature, 384(384):335–338, 1996.

[9] X. Wu and H. Liu. *A new theory of phase-contrast X-ray imaging based on Wigner distributions*. Medical Physics, 31(9):2378–2384, 2004.

[10] E. F. Donnelly, R. R. Price, and D. R. Pickens. *Experimental validation of the Wigner distributions theory of phase-contrast imaging*. Medical Physics, 32(4):928–931, 2005.

[11] M. R. Teague. *Deterministic phase retrieval: a Green's function solution*. Journal of the Optical Society of America, 73(11):1434–1441, 1983.

[12] T. E. Gureyev, A. Roberts, and K. A. Nugent. *Partially coherent fields, the transport-of-intensity equation, and phase uniqueness*. Journal of the Optical Society of America A, 12(9):1942–1946, 1995.

[13] N. Kardjilov. *Furhter developments and applications of radiography and tomography with thermal and cold neutrons*. Phd dissertation, Technical University Munich, 2003.

[14] T. E. Gureyev and S. W. Wilkins. *On x-ray phase imaging with a point source*. Journal of the Optical Society of America A, 15(3):579–585, 1998.

[15] M. Langer, P. Cloetens, J. P. Guigay, and F. Peyrin. *Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography*. Medical Physics, 35(10):4556–66, 2008.

[16] A. Snigirev, I. Snigireva, V. Kohn, S. Kuznetsov, and I. Schelokov. *On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation*. Review of Scientific Instruments, 66(12):5486–5492, 1995.

[17] F. Arfelli, M. Assante, V. Bonvicini, A. Bravin, G. Cantatore, E. Castelli, L. Dalla Palma, M. Di Michiel, R. Longo, A. Olivo, S. Pani, D. Pontoni, P. Poropat, M. Prest, A. Rashevsky, G. Tromba, A. Vacchi, E. Vallazza, and F. Zanconati. *Low-dose phase contrast x-ray medical imaging*. Physics in Medicine and Biology, 43(10):2845–2851, 1998.

[18] A. Bronnikov. *Phase-contrast CT: Fundamental theorem and fast image reconstruction algorithms*. In Proceedings of SPIE, Developments in X-Ray Tomography V, volume 6318, pages Q3180–Q3180, San Diego, 2006.

[19] P. Cloetens, M. Pateyron-Salomé, J. Y. Buffière, G. Peix, J. Baruchel, F. Peyrin, and M. Schlenker. *Observation of microstructure and damage in materials by phase sensitive radiography and tomography*. Journal of Applied Physics, 81(9):5878–5886, 1997.

[20] U. Bonse and M. Hart. *An x-ray interferometer*. Applied Physics Letters, 6(8):155–157, 1965.

[21] M. Ando and S. Hosoya. *An attempt at x-ray phase-contrast microscopy*. Proceedings of the 6th International Conference on X-ray Optics and Microanalysis, pages 63–68, 1972.

[22] D. Chapman, W. Thomlinson, R. E. Johnston, D. Washburn, E. Pisano, N. Gmür, Z. Zhong, R. Menk, F. Arfelli, and D. Sayers. *Diffraction enhanced x-ray imaging*. Physics in Medicine and Biology, 42(11):2015–2025, 1997.

[23] T. Weitkamp, A. Diaz, B. Nohammer, F. Pfeiffer, T. Rohbeck, P. Cloetens, M. Stampanoni, and C. David. *Hard x-ray phase imaging and tomography with a grating interferometer*. In Proceedings of the SPIE, Developments in X-Ray Tomography IV, volume 5535, pages 137–142, Denver, 2004.

[24] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David. *Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources*. Nature Physics, 2(4):258–261, 2006.

[25] P. Cloetens, R. Barrett, J. Baruchel, J.-P. Guigay, and M. Schlenker. *Phase objects in synchrotron radiation hard x-ray imaging*. Journal of Physics D: Applied Physics, 29(1):133–146, 1996.

[26] S. C. Mayo, T. J. Davis, T. E. Gureyev, P. R. Miller, D. Paganin, A. Pogany, A.W. Stevenson, and S.W. Wilkins. *X-ray phase-contrast microscopy and microtomography*. Optics Express, 11(19):2289–2302, 2003.

[27] D. Paganin, S. C. Mayo, T.E. Gureyev, P. R. Miller, and S. W. Wilkins. *Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object*. Journal of Microscopy, 206(1):33–40, 2002.

[28] A. Pogany, D. Gao, and S. W. Wilkins. *Contrast and resolution in imaging with a microfocus x-ray source*. Review of Scientific Instruments, 68(7):2774–2782, 1997.

[29] A. Bronnikov. *Reconstruction formulas in phase-contrast tomography*. Optics Communications, 171(4-6):239–244, 1999.

[30] A. Bronnikov. *Theory of quantitative phase-contrast computed tomography*. Journal of the Optical Society of America A, 19(3):472–480, 2002.

[31] A. Groso, R. Abela, and M. Stampanoni. *Implementation of a fast method for high resolution phase contrast tomography*. Optics Express, 14(18):8103–8110, 2006.

[32] M. Boone. *Fasecontrast beeldvorming aan een nanofocus X-stralenbuis*. Master thesis, Ghent University, 2007.

[33] M. Boone, Y. De Witte, M. Dierick, J. Van den Bulcke, J. Vlassenbroeck, and L. Van Hoorebeke. *Practical use of the modified Bronnikov algorithm in micro-CT*. Nuclear Instruments and Methods Physics Research B, 267(7):1182–1186, 2009.

[34] J. Van den Bulcke, M. Boone, J. Van Acker, and L. Van Hoorebeke. *High-resolution X-ray imaging and analysis of coatings on and in wood*. Journal of Coatings Technology and Research, 2009.

[35] S. Li, K. Mueller, M. Jackowski, D. Dione, and L. Staib. *Fast Marching Method to correct for refraction in ultrasound computed tomography*. In Proceedings of IEEE International Symposium on Biomedical Imaging, pages 896–899, Arlington, 2006.

[36] S. J. Norton. *Computing ray trajectories between two points: a solution to the ray-linking problem*. Optical Society of America A, 4(10):1919–1922, 1987.

[37] Y. De Witte, M. Boone, J. Vlassenbroeck, M. Dierick, and L. Van Hoorebeke. *Bronnikov-aided correction for x-ray computed tomography*. Journal of the Optical Society of America A, 26(4):890–894, 2009.

[38] V. Cnudde, B. Masschaele, H. E. V. De Cock, K. Olstad, L. Vlaminck, J. Vlassenbroeck, M. Dierick, Y. De Witte, L. Van Hoorebeke, and P. Jacobs. *Virtual histology by means of high-resolution X-ray CT*. Journal of Microscopy, 232(3):476–485, 2008.

[39] Y. De Witte, M. Boone, J. Vlassenbroeck, M. Dierick, B. Masschaele, L. Van Hoorebeke, and V. Cnudde. *Correcting phase contrast artefacts in X-ray CT imaging*. In Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 574–577, Boston, 2009.

That theory is worthless.

It isn't even wrong!

Wolfgang Pauli, 1900° – 1958†

6

Conclusions and Outlook

In this work, we elaborately studied three major aspects regarding the tomographic reconstruction of high resolution, X-ray projection data, acquired at an X-ray tube based CT scanner: the reconstruction in the helical cone-beam geometry, iterative reconstruction algorithms and the reconstruction of projections containing phase contrast. Above all, the emphasis of our research was on the practical feasibility of the presented solutions.

In chapter 3, we discussed the standard, analytical reconstruction of projection data acquired in the parallel-, fan-, and cone-beam geometry. We then introduced the alternative helical cone-beam geometry in high resolution CT, as a solution to two common problems encountered in the cone-beam geometry, which is currently the standard geometry in tube based systems: the appearance of cone-beam artefacts and the scanning of elongated objects. The reconstruction algorithm that is most appropriate to our intended application area was found to be the Katsevich algorithm. In accordance with the first positive results, the helical path was immediately incorporated into the design of UGCT's new high resolution scanner. To speed up the reconstruction process, we looked at different acceleration methods, both software and hardware based, and decided to increase the performance of our reconstruction software by porting the algorithms to the GPU. We completed the chapter by presenting some practical applications using the helical trajectory. We conclude that the helical cone-beam geometry provides a complementary solution to the cone-beam geometry. It can be efficiently applied in high resolution X-ray CT, thereby further extending the application area and the quality of this technique. By applying the helical path frequently to every-day scans, we may further validate

the quality and the practical feasibility of this technique. Considering the possibilities and the already impressive results, we believe that the helical cone-beam will become a standard geometry in X-ray tube based, high resolution CT, which will be used in addition to the cone-beam geometry.

An in-depth study of iterative algorithms was provided in chapter 4. After discussing several important details of the implementation, we addressed the two major problems regarding the practical application of such algorithms: the reconstruction speed and the memory requirements. Regarding the large data sets typically acquired in high resolution CT, these problems seemed to pose an insurmountable computational burden. We were able to overcome these problems by porting the algorithm to the GPU and by proposing a multiresolution approach. Furthermore, we discussed solutions to several minor practical issues. Combining this knowledge, we presented our highly efficient implementation of SART, which can be readily applied in practice. We then evaluated our implementation, and iterative algorithms in general, by comparing reconstruction results with the conventional filtered backprojection. It was shown that iterative methods provide higher quality, especially when applied to projection data that are somehow limited. We also presented a modification to the forward projection model, which allows us to correct artefacts due to extreme beam hardening by strongly attenuating inclusions. We finished this chapter by showing results of the application of our SART implementation to every-day scans, illustrating the potential increase in reconstruction quality. To conclude, this research has provided us with the possibility of using an iterative algorithm in the reconstruction of high resolution X-ray CT data. Future work may consist of a further evaluation of the application of our SART implementation in every-day scans. The efficiency of iterative algorithms may undoubtedly be improved by keeping a close eye on the future development of high performance computer hardware. We also raised the issue of the application of the nonnegative constraint in the reconstruction. The influence of such constraints on the image quality needs to be investigated more thoroughly. Another promising direction for future research is the improvement of the forward projection step, by incorporating additional physical effects other than the exponential attenuation of X-rays. Such improved models may be used to correct several disturbing artefacts that typically occur in X-ray CT. Due to the significant improvement in quality and the many potential benefits, iterative reconstruction algorithms may someday become the new standard.

In the last major chapter 5, we discussed the appearance of phase contrast and showed how this effect introduces severe artefacts in the reconstruction when not properly dealt with. We investigated possible solutions to this problem, with the restriction that such solutions may not require any changes in the scanner hardware or in the acquisition protocol. First, we discussed and evaluated a solution based on phase retrieval, known as MBA. We showed that this algorithm provides a very efficient reconstruction method, producing a high quality representation of the refraction function of the object. By our simulations, it was also found that MBA proves to be very stable regarding noise. This behaviour should definitely

be verified by future experiments, as it could provide a significant advantage in the further development of high resolution CT. However, we also illustrated the failure of MBA when the remaining attenuation by the object was too high. In order to overcome this problem, we developed an alternative solution based on phase reduction, BAC. This method allows us to reconstruct the attenuation function of medium or strongly attenuating phase objects, where the phase artefacts in the reconstruction are removed or at least heavily reduced. MBA and BAC provide us with a complementary solution to appropriately reconstruct projections containing both phase and attenuation contrast. We evaluated these methods in the reconstruction of several data sets, acquired both at our own UGCT scanners and at other high resolution facilities. The resulting reconstructions showed that the application of either solution provides a much higher image quality than the conventional reconstruction. Currently, both methods are frequently applied in the reconstruction of every-day scans at UGCT, which already led to several impressive results, as was shown in the applications. We conclude that, by investigating and developing solutions to the appearance of phase contrast, we were able to significantly increase the reconstruction quality of very high resolution scans, thereby pushing the limits of X-ray CT towards even higher resolutions. Future research may consist of the further application of MBA and BAC at UGCT, but possibly also at other facilities. For general application, it will be imperative to develop a more rigid method to determine the optimal parameters required for both algorithms, preferably by an automated routine. Furthermore, even though we were yet unable to find an appropriate solution, we believe there lies great potential in the incorporation of phase contrast generation in iterative reconstruction algorithms.

As a general conclusion, we think it is appropriate to state that we were able to considerably extend the application area of high resolution X-ray CT and that we achieved a significant increase in reconstruction quality for a wide range of applications.

Research is what I'm doing

when I don't know what I'm doing.

Wernher Von Braun, 1912° – 1977†

A

Phantom Description

A.1 Modified Shepp-Logan phantom

The Shepp-Logan phantom consists of a collection of 10 ellipsoids, determined by the following parameters: the centre $\vec{x}_0 = (x_0, y_0, z_0)$, the lengths of the half axes a, b and c along the X-, Y- and Z-axis, respectively, the rotation angle α and finally the attenuation value μ, which is homogeneous inside each ellipsoid (see figure A.1). The phantom was slightly modified to resemble a more typically encountered object in high resolution CT, by decreasing the contrast between the phantom's outer shell and the internal features. The modified Shepp-Logan phantom is described by the parameters in table A.1. The cross-section at z = -0.25 is illustrated in figure A.2.

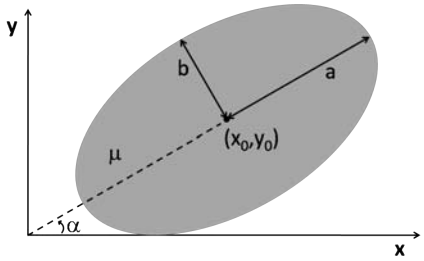


Figure A.1: The parameters describing an ellipse in 2D.

Table A.1: Parameters of the modified Shepp-Logan phantom.

a	b	c	x_0	y_0	z_0	α	μ
0.69	0.92	0.90	0	0	0	0	0.10
0.6624	0.874	0.88	0	-0.184	0	0	-.08
0.11	0.31	0.21	0.22	0	0	-18	-0.02
0.16	0.41	0.22	-0.22	0	-0.25	18	-0.02
0.21	0.25	0.35	0	0.35	-0.25	0	0.01
0.046	0.046	0.046	0	0.10	-0.25	0	0.01
0.046	0.046	0.02	0	-0.10	-0.25	0	0.01
0.046	0.023	0.02	-0.08	-0.605	-0.25	0	0.01
0.023	0.023	0.10	0	-0.605	-0.25	0	0.01
0.023	0.046	0.10	0.06	-0.605	-0.25	0	0.01

Figure A.2: The modified Shepp-Logan phantom at z = -0.25.

A.2 Metal phantom

The metal phantom consists of one large sphere, in which several smaller, highly attenuating spheres are embedded. The phantom can be described in the same way as the Shepp-Logan phantom, with the parameters given in table A.2. The central cross-section of this phantom is shown in figure A.3.

Table A.2: Parameters of the metal phantom.

a	b	c	x_0	y_0	z_0	α	μ
0.95	0.95	0.95	0	0	0	0	0.02
0.15	0.15	0.15	0	0.6	0	0	0.08
0.10	0.10	0.10	0.4242	0.4242	0	0	0.08
0.05	0.05	0.05	0.6	0	0	0	0.08
0.02	0.02	0.02	0.4242	-0.4242	0	0	0.08
0.15	0.15	0.15	0	-0.6	0	0	0.18
0.10	0.10	0.10	-0.4242	-0.4242	0	0	0.18
0.05	0.05	0.05	-0.6	0	0	0	0.18
0.02	0.02	0.02	-0.4242	0.4242	0	0	0.18

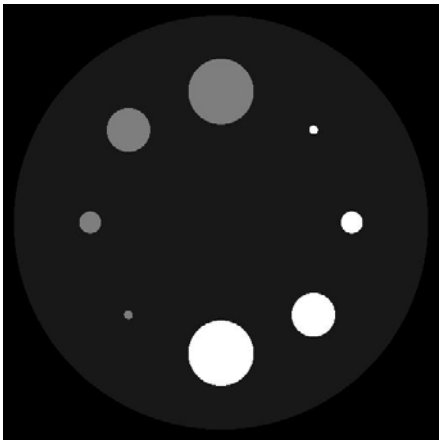


Figure A.3: The central cross-section of the metal phantom.

A.3 Phase phantom

The phantom used in the phase contrast simulations consists of a large sphere containing a collection of smaller spherical pores and inclusions. Each sphere is now described by an additional parameter, the refractive index increment δ. The parameters for this phantom are given in table A.3, figure A.4 shows the central cross-section.

Table A.3: Parameters of the phase phantom.

a	b	c	x_0	y_0	z_0	α	μ	δ
0.95	0.95	0.95	0	0	0	0	μ_0	δ_0
0.15	0.15	0.15	0	0.6	0	0	-μ_0	-δ_0
0.10	0.10	0.10	0.4242	0.4242	0	0	-μ_0	-δ_0
0.05	0.05	0.05	0.6	0	0	0	-μ_0	-δ_0
0.02	0.02	0.02	0.4242	-0.4242	0	0	-μ_0	-δ_0
0.15	0.15	0.15	0	-0.6	0	0	μ_1	δ_1
0.10	0.10	0.10	-0.4242	-0.4242	0	0	μ_1	δ_1
0.05	0.05	0.05	-0.6	0	0	0	μ_1	δ_1
0.02	0.02	0.02	-0.4242	0.4242	0	0	μ_1	δ_1

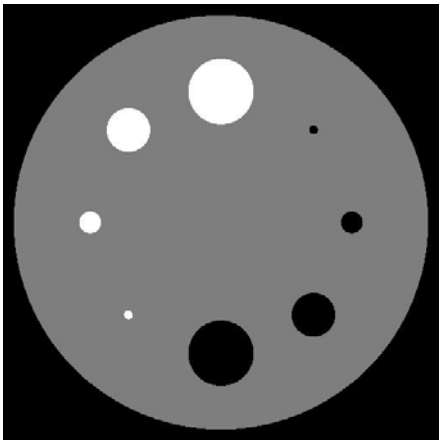


Figure A.4: The central cross-section of the phase phantom.

A.4 Defrise phantom

The Defrise phantom consists of a number of N discs, which are separated from each other by air. The attenuation value μ inside the discs is homogeneous and equal for all discs. All discs are of diameter D and thickness T and are separated from each other by a distance I between their centres, as illustrated in figure A.5.

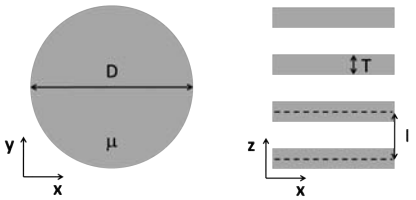


Figure A.5: The parameters describing a Defrise phantom.

A.5 X-ray transform

The projection images can be obtained by calculating the integrated attenuation value $P_\theta(\vec{x}_d)$ along the ray path corresponding to each pixel at position $\vec{x}_d$ by:

$$P_\theta(\vec{x}_d) = \int_0^1 f(\vec{x}_s + t \cdot (\vec{x}_d - \vec{x}_s))\, dt \,, \tag{A.1}$$

where $\vec{x}_s$ is the position of the source and where both the pixel and the source position are rotated over the projection angle θ. For a phantom composed of N geometrical shapes with a constant linear attenuation value μ_i, this can be expressed by:

$$P_\theta(\vec{x}_d) = \sum_{i=1}^{N} \mu_i \cdot L_i \,. \tag{A.2}$$

with L_i the intersection length between the ray determined by $(\vec{x}_s, \vec{x}_d)$ and the i-th shape.

Discs

When the phantom consists of a collection of discs, these intersection lengths can be easily derived using simple geometrical calculations for the intersection between a line and a cylinder.

Ellipsoids

For a unit sphere with constant attenuation μ, the attenuation function is given by:

$$f(\vec{x}) = \begin{cases} \mu & \text{if } ||\vec{x}|| \leq 1 \\ 0 & \text{otherwise .} \end{cases} \tag{A.3}$$

An ellipsoid determined by the parameters $(\vec{x}_0, a, b, c, \alpha)$ can be transformed into a unit sphere by the following transformation matrix:

$$\mathbf{A} = \begin{pmatrix} \frac{\cos\alpha}{a} & \frac{\sin\alpha}{a} & 0 \\ \frac{-\sin\alpha}{b} & \frac{\cos\alpha}{b} & 0 \\ 0 & 0 & \frac{1}{c} \end{pmatrix} \tag{A.4}$$

Let the direction $\vec{\phi}$ of a ray be given by:

$$\vec{\phi} = \frac{\vec{x}_d - \vec{x}_s}{||\vec{x}_d - \vec{x}_s||} , \tag{A.5}$$

and the transform D by:

$$Df(\vec{x}, \vec{\phi}) = \int f(\vec{x} + t\vec{\phi})\, dt, \tag{A.6}$$

Denote the following properties:

$$g(\vec{x}, \vec{\phi}) = f(\vec{x} - \vec{x}_0, \vec{\phi}) \quad \Rightarrow \quad Dg(\vec{x}, \vec{\phi}) = Df(\vec{x} - \vec{x}_0, \vec{\phi}) \tag{A.7}$$

$$g(\vec{x}, \vec{\phi}) = f(\mathbf{B}\vec{x}, \vec{\phi}) \quad \Rightarrow \quad Dg(\vec{x}) = \frac{Df(\mathbf{B}\vec{x}, \omega)}{||\mathbf{B}\phi^T||} , \; \omega^T = \frac{\mathbf{B}\phi^T}{||\mathbf{B}\phi^T||} , \tag{A.8}$$

where **B** is a nonsingular matrix. The intersection length can now be calculated by:

$$\begin{aligned} \vec{x} &= \vec{x}_s - \vec{x}_0 , \\ S &= ||\mathbf{A}\vec{x}^T||^2 - \frac{(\mathbf{A}\vec{x}^T \cdot \mathbf{A}\vec{\phi}^T)^2}{||\mathbf{A}\vec{\phi}^T||^2} , \\ L &= \begin{cases} 2 \cdot \dfrac{\sqrt{1-S}}{||\mathbf{A}\vec{\phi}^T||} & \text{if } S \leq 1 \\ 0 & \text{otherwise .} \end{cases} \end{aligned} \tag{A.9}$$