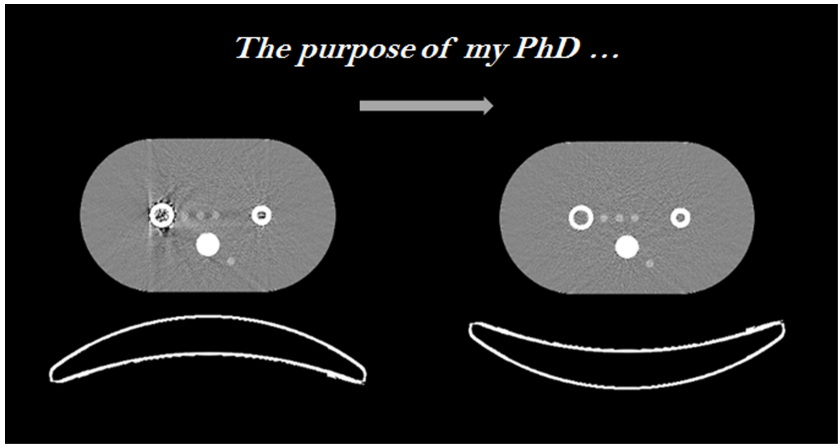
Local models and local convergence for maximum likelihood iterative reconstruction

Katrien Van Slambrouck


The purpose of my PhD …

Acknowledgments

Hier is het dan, mijn doctoraat, helemaal af. Ik ging nooit doctoreren en nu sta ik hier toch. Ik wil daarom toch graag even de tijd nemen om iedereen die mij overtuigd en gesteund heeft te bedanken.

Patrick, jij komt eerst want zonder jou zou ik nooit aan dit avontuur begonnen zijn. Als promotor van mijn masterthesis begeleidde je mijn eerste stappen in het onderzoekswerk en liet je mij kennismaken met de superinteressante wereld van de medische beeldvorming en wat een fysicus daarin allemaal kan doen. Bedankt Patrick dat je in mijn geloofde, beetje bij beetje geloofde ik ook meer in mijzelf. Tenslotte, bedankt om mij een vliegende start te geven door mij in contact te brengen met Johan en zijn groep!

Johan, bedankt voor alles. Je geeft al je studenten een eerstelijns begeleiding wat je een uitzonderlijke promotor maakt. Je deur staat op altijd open voor vragen en discussies. Het was niet altijd even gemakkelijk om je te overtuigen in een discussie maar dat maakt dat ik leerde sterkere argumenten te gebruiken. Als ik even niet langskwam, kwam je zelf kijken of alles nog ok was. Je slaagde erin mij op te peppen in dipmomentjes door steeds met nieuwe impulsen en ideeën op de proppen te komen. Bedankt voor de vele avonden en weekends dat je mijn teksten hebt nagelezen of onderdelen voor fantomen hebt gemaakt.

Bedankt *Paul* dat je mijn copromotor wilde zijn, bedankt voor je interessante suggesties en opmerkingen tijdens mijn seminaries en voor de boeiende multidisciplinaire omgeving waarin ik mocht werken. Je bent heel hard begaan met het belang van samenwerking tussen verschillende disciplines en dit is dagelijks te merken aan je inzet voor het onderzoek en de werking van het Louvre.

I would like to thank all my jury members: Prof. Hilde Bosmans, Prof. Christophe Deroose, Prof. Laurent Desbat, Prof. Frederik Maes and Prof. Jan Sijbers. Thank you for making time to read and evaluate my thesis and for all the valuable suggestions and comments that improved my manuscript.
Hilde, thank you for following my work so closely, for providing me with a direct link to the practical side of radiology. Your door was always open for discussion and practical help. Thank you for the tips on how to enjoy a week at ECR and in Vienna!

Christophe, you were my first link to the medical aspects of medical imaging. I really appreciated your interest in our methods and the occasions when you have sent us patient cases where metal artifacts troubled diagnosis in PET/CT scans. It was encouraging to see that our methods, although they are not perfect, could already improve the diagnostic process.
Laurent, many thanks for inviting us to the workshops in Lyon and Grenoble. I really enjoyed my stay as well as the interesting talks and discussions that showed me tomographical problems in a broader sense than what I was used to. I also enjoyed the pleasant evenings with interesting discussions about science, linguistics, ...
Frederik, you followed my work from the beginning. I really appreciated your profound remarks and advice during discussions on my seminars and on the Quantiviam meetings, they made me reconsider problems and solutions from another, broader point of view.
Jan, I met you right at the beginning of my PhD, when I was really a beginner entering the reconstruction field. I immediately learned that CT was so much more than scanning patients. Thank you for all questions and remarks while I was writing my very first conference abstract and at all Quantiviam meetings.

I would like to express my great gratitude to dr. Karl Stierstorfer from Siemens Erlangen. Thank you *Karl*, without your help we would never have been able to obtain the results as we have them now. For modeling CT systems in greater detail, information from the manufacturer is indispensable. Thank you for your willingness to help us, for all your valuable suggestions, for your always quick and to-the-point answers, and for your confidence in our group.

Many thanks to my colleagues: *Kathleen, Lin, Koen, Ahmad, Matthew, Anna and Selmin*: for helping me with many small and larger problems, for solving my IDL questions, for pushing the submit button ;-), for proofreading my texts, for the support when I was in a scientific dip, for the pleasant lunches and for the team drinks or other team activities!
Koen, een speciaal woordje voor jou. Als enige andere *transmissie*persoon was je mijn eerste vraag-en-klaag-adres. Heel vaak zag je mijn hoofd verschijnen en daarbij een vraag, opmerking of een 'pfff'-momentje. Bedankt om een fijne overbuur te zijn!

Koen, Gregory, Geoffrey, Hannelore, Wannes and Victor. Thank you for being around, for enlarging our group, for all the interesting discussions, for the work you did during your master thesis, for all the questions that seemed logical at first sight but turned out to be tough to answer.

Thanks to all people that 'lived' in *the Louvre* the last four years: for the short chats in the coffee room, for being 'interested' when I was filling my phantom again and again. Special thanks to *Maarten*, when I arrived you were the only one working with MLTR and IMPACT, thank you for being my help desk, for your advice and your direct answers whenever needed. Thank you *Wouter, Jean-Yves and Stijn*, for the interest in our methods, and for giving me the

opportunity to apply my work on real life problems. *Dominique and Bart*, your technical and practical support made life in the Louvre so much easier, thanks!

Een welgemeende dank u voor iedereen van de *afdeling Nucleaire Geneeskunde*. *Prof. Mortelmans* en *Prof. Van Laere*, als hoofd van de afdeling zorgden jullie voor de juiste omkadering om onderzoek te verrichten. De nauwe banden tussen onderzoek en klinische praktijk maakte dat ik kon metingen doen op een klinisiche CT scanner die bijna naast mijn bureau stond! Verder wil ik graag ook alle (staf)leden van de afdeling bedanken! Enkele onder jullie verdienen zeker nog een extra woordje van dank.
Francine, ik denk niet dat iemand zich een betere secretaresse kan indenken. Ik kon bij jou steeds terecht met al mijn praktische vragen. Je efficiëntie maakt dat vele problemen al opgelost waren nog voor ik klaar was met ze uit te leggen. Later konden we ook bij Nathalie terecht. *Nathalie* je was een goede gids door het boeiende bos van regels en administratie van onze universiteit!
Peter, dankzij jou werd ik aan het einde van elke maand beloond met enige financiële steun. Bedankt om alle financiële administratie zo efficiënt te laten verlopen, voor het snelle antwoord op alle vragen en sorry voor de keren dat ik mijn facturen en bewijsjes toch niet helemaal juist had ingevuld. Ik denk dat ik het nu min of meer snap, al is dat natuurlijk een beetje laat...
Kwinten en Mieke, dankzij jullie kon ik ook aan de slag met een echte CT. Jullie vonden steeds een gaatje voor mij in de Hirez-agenda en zorgden voor een opgewekte sfeer tijdens het vaak saaie scannen.
Kristof en Stefaan, bedankt voor de ontelbare keren dat jullie onze kleine en minder kleine IT problemen hebben opgelost. Bedankt voor alle praktische hulp bij het bedienen van camera's. Sorry voor de keren dat mijn files toch net iets groter waren dan ik dacht en de servers dat niet altijd even goed aankonden. Kristof, bedankt voor de babbels op de bus, voor de 'en hoe gaat het hier' elke keer als je mijn bureau voorbijzoefde. Als fysicus in hart en nieren breng je als geen andere je passie voor fysica en voor medische fysica over!
Tenslotte bedankt *Jenny, Sofie, Rawaha, Cindy, Michel en Andrey*, voor jullie raad en daad bij verschillende medische en praktische vragen!

Walter, je was mijn directe connectie met de radiologiepraktijk. Hoewel onze methodes nog niet klaar zijn voor routinegebruik in klinische toepassingen, was je steeds super geïnteresseerd in mijn werk. Je verruimde mijn blik van mijn computerscherm naar de ontelbare toepassingen voor metaalartefactreductie methodes. Ik apprecieer je betrokkenheid en enthousiasme heel erg. Dankzij jou kreeg onze nieuwe aanpak ook extra (internationale) aandacht. Bedankt hiervoor! *Federica*, the CT lady! Thank you for all times I could just walk in with again another question about CT systems. Thank you for your interest in my work, you made me believe I was really doing something useful.

Bedankt *Prof. Mulier* voor je interessante uiteenzetting over heupimplantaten en de problemen die je dagelijks ervaart door storende metaalartefacten. Ons bezoek aan Pellenberg maakte duidelijk dat metaalartefacten nog steeds een

belangrijk probleem zijn. De protheses die je meegaf hebben we verschillende keren gebruikt voor het maken van realitische fantomen, je vindt ze verderop in dit boekje, bedankt!

Special thanks goes to *Michel Defrise, Frederic Noo and Rolf Clackdoyle*. On many occasions I had the privilege to have discussion with you, some of the big names in the reconstruction field. Thank you for patiently listening to my problems, for suggestions and possible solutions and for your continuous interest in my work.

Nog een woordje voor de mensen van de ondersteunde diensten in het UZ Gasthuisberg. Bedankt *Marc Verbeeck* en *Joris de Ruyter* van biotechniek voor het vervaardigen van de onderdelen voor mijn fantomen en voor de technische raad. Bedankt *Jan Demey*, *Herman Van Biesen* en alle andere mensen bij informatiesystemen, dankzij jullie kon ik rekenen op een betrouwbaar en efficiënt netwerk van servers en computers.

Ik wil graag ook al mijn *vrienden en (schoon)familie* bedanken die mij steeds gesteund hebben en voor voldoende fijne afleiding hebben gezorgd. De vraag: 'leg uw doctoraat nu eens in mensentaal uit' stelden jullie ontelbare keren. Met de jaren ging die mensentaal beter en beter. Ik hoop dat jullie nu ongeveer een idee hebben wat ik gedaan heb en anders leg ik het graag nog eens in mensentaal uit bij een lekker glas ... fruitsap ;).

Een belangrijk woordje van dank is voor mijn broer Joost. *Joost*, wij hebben duidelijk andere genen gekregen, ik de meer abstracte, jij de creatieve, maar dat maakt ons een goed team en daar is dit boekje weer een mooi voorbeeld van. Bedankt voor het ontwerpen van de kaft en de uitnodiging. En, wees maar zeker dat jouw bijdrage aan dit boekje substantieel is want er zullen zeker meer mensen de kaft bestuderen dan dat er de inhoud lezen! Bedankt, *Karlien* en *Natascha*, voor de goede zorgen, de interesse in mijn werk en de nodige afleiding tijdens de laatste maanden. Bedankt ook *oma Maria* en *opa Robert* om mij steeds met open armen te ontvangen!

Bedankt *mama en papa*. Typisch staat hier dan 'om mij te laten studeren' en dat is zeker waar, maar ik wil jullie vooral bedanken omdat ik altijd mijn eigen keuzes mocht maken, keuzes die jullie misschien niet altijd verwachtten maar waar jullie mij wel altijd 100% in gesteund hebben. Bedankt voor het vele rondrijden, voor de praktische hulp, dit alles maakte de stressmomenten de afgelopen jaren minder erg. Bedankt!

De laatste dank u kan alleen maar voor jou zijn, *Stijn*. Je stond steeds achter mij bij elke stap die ik gezet heb. Ik kan ze niet tellen, de keren dat je teksten hebt nagelezen, versie na versie, blijkbaar zelfs zo goed dat je mijn doctoraat al begrijpt en kan uitleggen :). De vele smsjes en mailtjes vlak voor en na presentaties, belangrijke besprekingen of op momenten dat ik het even niet meer zag zitten hielpen mij door de bergen en dalen van dit avontuur. Sorry voor de hectische laatste maanden waarin ik soms enkel tijd had voor schrijven en slapen. Dankzij jou draaide alles en hoefde ik mij om niks anders zorgen te

maken. Men zou je misschien omschrijven als de nieuwe man, maar voor mij ben je vooral míjn geweldige, haast perfecte, man!

Katrien

Abstracts

Local models and local convergence for maximum likelihood iterative reconstruction

Iterative methods are being increasingly used for reconstruction of tomographic data as in computed tomography (CT) and positron emission tomography (PET). These methods often yield better reconstructions compared to analytical methods. An important reason for this is that iterative reconstruction has a more flexible data acquisition model which allows for a more accurate simulation of the true acquisition during the reconstruction. Moreover, a noise model can be included in the reconstruction algorithm.

This thesis investigated several remaining problems for iterative reconstruction in CT and PET based on the principles of local modeling and local convergence improvements.

For CT, iterative metal artifact reduction was studied. Metals attenuate the majority of the photons which makes that small deviations in the reconstruction model can have a large influence and introduce severe artifacts. In the current state of the art methods, data points with low photons are removed and replaced with artificial data points that do not cause artifacts. This reduces the most severe artifacts but by the removal of the data points also contrast in the regions around and between the metals is removed. Alternatively, iterative methods can be used. These methods model the effects that cause the artifacts and do not remove data points. The existing iterative methods do not reduce the artifacts sufficiently and they are often too slow for possible clinical usage.

To reduce computation time, a local model scheme was introduced. This approach uses the most accurate (and computation intensive) model for the metals in the image and a simpler model elsewhere. This local model structure is called the patchwork structure and was able to reduce the computation time without losing image quality. With the patchwork structure it was also possible to improve the convergence of the metal regions locally, leading to an improved artifact reduction. This method has been applied on simulated and measured

phantom and patient scans. The artifacts in the reconstructed images were clearly reduced without loss of contrast around the metals. Some artifacts, however reduced, are still present. The origin of these artifacts remains unclear.

The local convergence improvements that can be obtained by the patchwork structure were also evaluated for convergence improvements for the entire image volume. Hence, the volume was divided in a set of equal small image blocks that are separately updated. This alternative patchwork structure has been applied on both CT and PET data. For both applications, a considerable increase in convergence rate was obtained, yielding the same image quality in a shorter time, which is important for clinical application of iterative methods.

In clinical environments, PET is usually reconstructed with iterative reconstruction. A problem of the applied iterative method is that it tends to produce bias when the data are very noisy and in cold regions. This bias may cause problems for tracer kinetic analysis. Origins of this bias are the strong asymmetry of the employed Poisson distribution in case of low count data and the inherent nonnegativity constraint of the algorithm. A modified algorithm has been developed that switches from the Poisson distribution to the Gaussian distribution for low count data points. This new method lowers the bias significantly and gives results comparable to other bias reduction methods. An advantage of this new method is that the distribution is locally changed and that it has only a single parameter, the value of which is easily determined.

Lokale modellen en lokale convergentie voor maximum likelihood iteratieve reconstructie

Iteratieve methodes worden meer en meer gebruikt voor reconstructie van tomografische data zoals in computer tomografie (CT) en positron emissie tomografie (PET). Ze geven vaak veel betere reconstructies dan analytische methodes. Een belangrijke reden hiervoor is het feit dat iteratieve reconstructie een flexibel model van het meetproces toelaat, waardoor het meetproces accurater kan gesimuleerd worden tijdens de reconstructie. Bovendien wordt er vaak een ruismodel ingebouwd.

Deze thesis bestudeerde enkele overblijvende problemen van iteratieve reconstructie voor CT en PET op basis van de principes van lokale modellering en lokale convergentieverbetering.

Voor CT werd specifiek metaalartefactreductie bekeken. Metalen attenueren het grootste deel van de fotonen waardoor kleine afwijkingen in het reconstructiemodel grote invloed kunnen hebben en sterke artefacten veroorzaken. Standaard zal men meetpunten met heel weinig fotonen weglaten en vervangen door artificiële data om zo artefacten te voorkomen. De typische zware metaalartefacten worden zo vermeden, maar door het weglaten van de datapunten gaat

er contrast verloren in het gebied rond het metaal en tussen verschillende metalen. Een alternatief is het gebruik van iteratieve methodes die de effecten die leiden tot deze zware artefacten beter modelleren en geen data weglaten. De bestaande iteratieve methodes reduceren de artefacten echter onvoldoende en zijn zeer traag, wat klinisch gebruik haast onmogelijk maakt.

Om de rekentijd te verbeteren werd een schema met lokale modellen geïntroduceerd. Deze aanpak gebruikt het meest accurate (maar rekenintensieve) model enkel voor de metaalregio's van het beeld en elders wordt een eenvoudiger model gebruikt. Deze structuur werd de patchworkstructuur genoemd en leidde tot verkorte rekentijden zonder verlies aan beeldkwaliteit. In deze structuur was het ook mogelijk om de convergentie van de metaalregio's lokaal te verbeteren, wat leidde tot een verbeterde artefactreductie. De methode werd toegepast op gesimuleerde en gemeten fantoom- en patiëntenscans. De resulterende beelden hebben duidelijk gereduceerde artefacten zonder contrastverlies in de metaalregio's. Er zijn echter nog overblijvende artefacten, zij het sterk verminderd, waarvan de oorsprong niet duidelijk is.

De lokale convergentieverbetering die verkregen werd met de patchworkstructuur werd ook geëvalueerd voor het verbeteren van de convergentie in het gehele beeld. Daarvoor werd het beeld in verschillende kleinere beeldblokken opgedeeld die apart aangepast werden tijdens de reconstructie. Deze alternatieve patchworkstructuur werd zowel toegepast voor het reconstrueren van CT als van PET data. Voor beide toepassingen werd een aanzienlijk verhoogde convergentiesnelheid verkregen, die dezelfde beeldkwaliteit levert in een kortere rekentijd, wat belangrijk is voor de klinische toepassing van iteratieve reconstructiemethodes.

Voor PET wordt klinisch meestal reeds iteratieve reconstructie gebruikt. Een probleem met de meest gebruikte iteratieve methode is dat ze vaak niet geschikt is voor kwantificatie omdat er dikwijls een systematische afwijking, of bias, zit op de gereconstructeerde waardes. Dit komt voornamelijk voor in scans waar maar zeer weinig fotonen gemeten worden of in gebieden met weinig activiteit. Een belangrijke oorzaak hiervoor is de sterke asymmetrie van de gebruikte Poissonverdeling voor lage datawaarden en de inherente niet-negativiteitsbeperking van het algoritme. Een algoritme werd ontwikkeld dat de Poissonverdeling vervangt door een Gaussiaanse verdeling voor lage datapunten. De ontwikkelde methode vermindert de systematische fout aanzienlijk. De resultaten zijn vergelijkbaar met een andere methode voor biasreductie. Een voordeel van de nieuwe methode is dat ze lokaal werkt en dat ze slechts een parameter heeft die gemakkelijk kan bepaald worden.

English summary

The introduction of X-ray computed tomography (CT) in the early 1970s marked the onset of a continuously increasing use of tomographic, or cross-sectional imaging, as a clinical diagnostic tool. Currently, the most important modalities for three-dimensional imaging of ionizing radiation are CT, SPECT (single photon emission computed tomography) and PET (positron emission tomography).

Tomographic data and CT data in particular are typically reconstructed with analytical, filtered backprojection (FBP)-type reconstruction methods. Analytical methods apply an inverse system acquisition model on the data to reconstruct the original object. An important advantage of analytical methods is their short reconstruction time. Images can be reconstructed almost in real time. However, the inverse acquisition model is inherently based on an ideal acquisition. Although real systems are not ideal, FBP usually produces reconstructions of high quality. When the system acquisition deviates further from an ideal system, FBP starts to introduce artifacts that degrade the image quality. FBP is also particularly sensitive to noise in the data.

Iterative reconstruction is receiving increased attention. Iterative reconstruction is based on a forward acquisition model. Hence, no analytical inversion is needed and more complex (and more accurate) models can be used. When the system acquisition can be modeled more accurately, less artifacts arise. Moreover, iterative maximum likelihood reconstruction algorithms incorporate a noise model which makes them less sensitive to noise.

In CT, the standard reconstruction method is still FBP. Iterative reconstruction is gaining interest, mainly because more attention is given to limiting patient doses in CT examinations. Since iterative reconstruction can cope better with noise compared to analytical methods, the dose can be reduced while maintaining good image quality. In addition, computers are becoming faster and the computation time for a fully iterative reconstruction is decreasing which makes the application of iterative reconstruction in clinical practice more feasible.

An important remaining problem for both analytical and iterative CT reconstruction is the suppression of metal artifacts. Highly attenuating materials

such as metals that are present in the scanned object (hip implants, dental fillings, etc.) often introduce severe dark and white streaks over the image that often hinder proper diagnosis in the affected image regions. The first attempts to reduce metal artifacts already originate from the 1980s.

The techniques to reduce metal artifacts can be divided in two main groups: sinogram inpainting and dedicated iterative reconstruction. For sinogram inpainting methods, the projection rays that cross metals are considered corrupt and replaced by artificial data. This operation results in strongly reduced artifacts but also reduces the soft tissue and bone contrast around and between metals. In iterative reconstruction the effects that cause the artifacts are modeled during the reconstruction to reduce the artifacts. By using iterative methods more contrast is preserved around and between the metals. However, the metal artifact reduction is often only significant for small and low to moderately attenuating materials. Streaks and shadows are still present for larger and higher attenuating metals.

In PET and SPECT, iterative reconstruction is already used on a daily basis for almost all clinical examinations. The main reason is that the data sets for PET and SPECT are relatively small and that the data are much noisier which increases the benefit of using iterative reconstruction. The standard for PET is to use maximum likelihood expectation maximization (MLEM) reconstruction which is based on a Poisson noise model. An important disadvantage is that MLEM tends to introduce a systematic error, called bias, in the reconstruction for low count data and in low activity regions surrounded by regions with higher activity. The origin of this bias is the nonnegativity constraint of MLEM and the asymmetry of the Poisson likelihood near zero. In dynamic PET, where uptake rate constants are calculated based on the activity in several time frames, analytical methods are often still preferred for reconstruction.

Purpose

The aim of this work was the amelioration and acceleration of iterative reconstruction methods. For CT, iterative methods for metal artifact reduction were considered. Algorithm changes to decrease the calculation time and to reduce remaining artifacts were investigated. A general acceleration scheme for improved convergence for iterative reconstruction has been developed as well, both for CT and PET. Finally, for PET, a modified algorithm to reduce bias was studied.

An important concept that was used for all three topics is local adaptation of the reconstruction model and/or local increase of the convergence rate.

Metal artifact reduction in CT

The complex models used for reduction of metal artifacts increase the computation time of the iterative methods significantly. In order to limit the computation time, we proposed to apply the complex models only locally, in and around metals, and to use a simpler model for the remainder of the reconstruction volume. Therefore, the image was divided in several regions or patches: one larger nonmetal patch and several smaller metal patches, one for each metallic object. A fully polychromatic model, IMPACT, was used in the metal patches. For the nonmetal patches, a simpler polychromatic model, that only considers the energy dependent attenuation of water, was developed. Besides the energy model, the sampling of the metal patches could also be adapted per patch.

The division of the image into patches is referred to as the patchwork structure. The patchwork structure developed in this work uses sequential updates for the patches in the image. This means that all patches are separately and sequentially used in the reconstruction algorithm. When using smaller groups of voxels separately, the convergence rate for that group improves substantially. Hence, in the local model structure, the convergence of the metal patches is increased.

A last issue that needed to be considered was the tendency for inferior results of for instance IMPACT on measured data compared to simulated data. This pointed at a model mismatch in the reconstruction model with respect to the true acquisition model. Possible candidates causing this mismatch are the precorrections the camera applies on the data. The influence of several precorrection methods was studied and appropriate uncorrection algorithms have been developed. Additionally, other physical effects, like cross talk between the detector elements, have been added to the reconstruction model.

The new iterative method has been applied and evaluated on simulated and measured phantom and patient data and compared to standard iterative reconstruction and to several sinogram inpainting methods.

First, the local model schema with reduced complexity in the model for the nonmetal patches performed similarly compared to applying the complex model everywhere. However, the computation time of the algorithm was substantially reduced. Second, the improved convergence of the metal patches improved the reconstruction of the metals, resulting in higher attenuation values and sharper edges, which reduced the artifacts substantially. Finally, for application on real data some extra physical parameters had to be taken into account and uncorrection of the data for sinogram water correction and sinogram bowtie correction was required.

The local model reconstruction had strongly reduced artifacts compared to standard iterative reconstruction. Sinogram completion yielded images without dark streaks and shadows but tended to decrease contrast between and around

metals. This was not observed for the iterative methods but in these images slightly more artifacts were present. For three-dimensional metal objects with strong axial variation, some persistent artifact, with unclear origin, was still observed for the local model iterative method.

Acceleration of CT and PET reconstruction

Metal artifact reduction in CT was accelerated by varying the complexity of the reconstruction model over the image. An important reduction of the metal artifacts was achieved by improved convergence of the metal regions due to the use of the patchwork structure, where the different patches in the image were sequentially used in the reconstruction algorithm. This structure was further investigated to improve the convergence for the entire reconstruction without varying reconstruction models.

It was already known that accelerated convergence can be obtained by dividing the reconstruction voxels into several smaller groups and using these groups separately in the reconstruction. The optimal choice for a group is a set of voxels spread over the reconstruction volume. However, this is not efficient regarding access to the computer memory. Moreover, combination with ordered projection subsets, that sequentially use subsets of the projection data, is difficult.

To our knowledge it was never investigated whether updating groups of neighboring voxels still has considerable influence on the convergence of the reconstruction. This enables more efficient memory access and possibly the combination with ordered projection subsets. A modified patchwork algorithm, an image-block algorithm, that divided the image in a set of equally sized groups of neighboring voxels, was developed to accelerate the convergence both in CT and PET reconstruction by using the groups separately in the reconstruction algorithm. In both cases, a considerable gain in convergence was observed and combination with ordered projection subsets was possible.

Bias reduction in PET

Bias in iterative reconstruction of PET data was also studied. This work is an extension on the existing NEGML algorithm. NEGML reduces bias is low activity regions because it allows for the use of negative voxel values. However, for noisy, low count data, bias is still observed.

A modified NEGML algorithm was developed. This algorithm replaces the Poisson distribution by the symmetric Gaussian distribution for low data points. This change in noise model is only applied locally for low data, the Poisson distribution is still used elsewhere. The method has a parameter that defines the

number of counts below which the Poisson noise model will be switched to the Gaussian distribution. The dependence of the bias reduction on that parameter has been investigated. Further, NEGML was also compared to AML, another bias reduction method. The parameter A of the AML algorithm defines the lower bound of the reconstructed image and leads to a shift in the employed noise distribution towards higher values, which makes the distribution also less asymmetric.

Both methods reduce bias substantially when their parameter is large enough. The simulations in two dimensions indicate that an optimal value for the parameter in NEGML can be easily determined and is independent of the system geometry and the activity in the image. AML keeps the MLEM behavior for relatively small negative lower bounds especially in hot regions which leads to improved convergence in these regions. However, a relatively large negative value for A is required to reduce all bias. Unlike for NEGML this value is not independent on the geometry and the activity and an optimal values is therefore more difficult to determine.

Conclusion

In this thesis, local models and local convergence have been applied to iterative maximum likelihood algorithms. The aim was to ameliorate and accelerate iterative methods. The use of local models in the image by the patchwork structure accelerated the very slow iterative metal artifact reduction substantially. The accelerated convergence rate in the metal regions reduced the majority of the artifacts. A modified patchwork structure has been used to obtain improved convergence for the entire reconstruction for both CT and PET. Finally, an algorithm with a locally modified noise model for low count data was developed for bias reduction in low count PET data. This new algorithm yields reconstructions with substantially reduced bias.

Nederlandse samenvatting

De introductie van X-stralen computertomografie (CT) in het begin van de jaren 1970 was de start van een continue toename van het gebruik van tomografische, of cross-sectionale, beeldvorming als een klinisch diagnostisch hulpmiddel. De belangrijkste toestellen voor drie-dimensionale beeldvorming met ioniserende straling zijn momenteel CT, SPECT (enkelvoudige 'single' fotonemissie computertomografie) en PET (positron emissie tomografie).

Tomografische data en CT data meer specifiek worden typisch gereconstrueerd met analytische methodes van het type gefilterde terugprojectie (FBP). Analytische methodes passen een invers model van het meetproces toe op de data om het originele object te reconstrueren. Een belangrijk voordeel van analytische methodes is hun korte reconstructietijd. De beelden kunnen bijna simultaan gereconstrueerd worden. Het inverse model is echter inherent gebaseerd op een ideale meting. Hoewel echte systemen niet ideaal zijn, geeft FBP gewoonlijk reconstructies van hoge kwaliteit. Wanneer het systeem meer afwijkt van een ideaal systeem, introduceert FBP artefacten die de kwaliteit van het beeld verslechteren. FBP is bovendien zeer gevoelig aan ruis in de data.

Iteratieve reconstructie krijgt meer en meer aandacht. Het is gebaseerd op een voorwaarts model van het meetproces. Er is daardoor geen analytische inversie nodig en meer complexe (en meer accurate) modellen kunnen gebruikt worden. Wanneer het meetproces accurater gemodelleerd wordt, ontstaan er minder artefacten. Daarenboven hebben iteratieve maximum likelihood reconstructie algoritmes een ingebouwd ruismodel wat hen minder gevoelig maakt aan ruis.

Voor CT is de standaard reconstructie methode nog steeds FBP. Iteratieve reconstructie wint aan belang, voornamelijk omdat er meer aandacht wordt geschonken aan het beperken van de patiëntendosis bij CT onderzoeken. Iteratieve reconstructie kan beter omgaan met ruis in de data in vergelijking met analytische reconstructie en hierdoor kan de dosis verlaagd worden terwijl een goede beeldkwaliteit behouden blijft. Daarnaast worden computers sneller, waardoor de rekentijd voor volledig iteratieve algoritmes daalt. Dit maakt de toepassing van iteratieve reconstructie in de klinische praktijk beter haalbaar.

Een belangrijk resterend probleem voor zowel analytische als iteratieve CT reconstructie is de onderdrukking van metaalartefacten. Sterk attenuerende materialen zoals metalen die aanwezig zijn in het gescande object (heupimplantaten, tandvullingen, etc.), introduceren vaak ernstige donkere en witte strepen in het beeld. Deze strepen verhinderen vaak een degelijke diagnose in de geaffecteerde gebieden. De eerste pogingen om deze metaalartefacten te reduceren ontstonden reeds in jaren 1980.

De technieken voor het reduceren van metaalartefacten kunnen opgedeeld worden in twee hoofdgroepen: sinogram vervollediging en specifieke iteratieve reconstructie. Bij sinogram vervollediging worden de projecties die de metalen hebben gekruist als corrupt beschouwd en vervangen door artificiële data. Deze operatie geeft sterk gereduceerde artefacten maar ook verminderd contrast voor zacht weefsel en been in de buurt van de metalen. Iteratieve reconstructie gebruikt een model dat de effecten die aan de basis liggen van de artefacten modelleert en op deze manier de artefacten reduceert. Deze techniek is echter enkel efficiënt voor eerder kleinere metalen met lage tot middelmatige attenuatie. Voor grotere, sterker attenuerende metalen zijn er nog steeds donkere strepen en schaduwen aanwezig.

Iteratieve reconstructie wordt reeds dagelijks gebruikt voor bijna alle klinische PET en SPECT onderzoeken. De belangrijkste reden is dat de datasets voor PET en SPECT relatief klein zijn en dat de data meer ruis bevatten waardoor het gebruik van iteratieve reconstructie voordeliger wordt. PET wordt standaard gereconstrueerd met MLEM dat gebaseerd is op een Poisson ruisverdeling. Een belangrijk nadeel van MLEM is dat het de neiging heeft om een systematische fout, of bias, te introduceren voor metingen met weinig gedetecteerde fotonen of in regio's met lage activiteit omgeven door regio's met hogere activiteit. De oorzaken van deze bias zijn de inherente eis voor niet-negatieve reconstructiewaarden en de asymmetrie van de Poissonverdeling nabij nul. In dynamische PET, waar opnameconstantes berekend worden op basis van de activiteit op verschillende tijdspunten, wordt analytische reconstructie daarom vaak nog verkozen boven MLEM.

Doel

Het doel van deze thesis was het verbeteren en versnellen van iteratieve reconstructiemethodes. Voor CT hebben we iteratieve metaalartefactreductie methodes bestudeerd. Aanpassingen aan het algoritme om de berekeningstijd in te korten en de artefacten verder te reduceren werden onderzocht. Een algemeen versnellingsschema voor verbeterde convergentie in iteratieve CT en PET reconstructie werd ook beschouwd. Tenslotte werd er voor PET een methode ontwikkeld om bias in de beelden te reduceren. Een belangrijk concept dat gebruikt werd voor alle drie de onderwerpen is de lokale aanpassing van reconstructiemodellen en/of de lokale verbetering van de convergentiesnelheid.

Metaalartefactreductie in CT

De complexe modellen die gebruikt worden voor metaalartefactreductie verhogen de rekentijd van de iteratieve methodes aanzienlijk. Om deze rekentijd te beperken, stelden we voor om de complexe modellen enkel lokaal te gebruiken, in en rond de metalen, en om een eenvoudiger model toe te passen voor de rest van het reconstructievolume. Het beeld werd hiervoor in verschillende regio's of patches verdeeld: één grote niet-metaalpatch en verschillende kleinere metaalpatches, telkens één patch per metaal. Een volledig polychromatisch model, IMPACT, werd gebruikt in de metaalpatchen. Voor de niet-metaalpatchen werd een eenvoudiger polychromatisch algoritme ontwikkeld dat enkel rekening houdt met de energieafhankelijkheid van de attenuatie van water. Naast het polychromatisch model kon ook de bemonsteringsdichtheid van de metaalpatchen aangepast worden per patch.

De verdeling van het beeld in patches werd de patchworkstructuur genoemd. De patchworkstructuur zoals die toegepast werd in dit werk, maakt gebruik van seqentiële updates voor de verschillende patchen in het beeld. Dit betekent dat alle patches apart en opeenvolgend gebruikt worden in het reconstructiealgoritme. Wanneer kleinere groepen van voxels apart gebruikt worden, verhoogt dit de convergentie van die groep. Daardoor wordt in de lokale modellen structuur de convergentiesnelheid van de metaalpatchen sterk verhoogd.

Een laatste punt dat moest behandeld worden, was de neiging tot slechtere resultaten wanneer bijvoorbeeld IMPACT werd toegepast op gemeten data in plaats van op gesimuleerde data. Dit wees op een fout in het reconstructiemodel in vergelijking met de echte acquistie. Mogelijke boosdoeners waren de precorrecties die door de camera uitgevoerd worden op de data. Deze precorrecties werden dan ook onderzocht en gepaste ontcorrectie methodes werden ontwikkeld. Daarenboven werden ook andere fysische effecten, zoals cross talk tussen detector elementen, toegevoegd aan het reconstructie model.

De nieuwe iteratieve methode werd toegepast en geëvalueerd voor gesimuleerde en gemeten fantoom- en patiëntendata. Ze werd vervolgens ook vergeleken met standaard iteratieve reconstructie en met verschillende sinogram vervolledigingsmethodes.

Het lokale modellen schema met verlaagde complexiteit voor het model van de niet-metaal patch, gedroeg zich gelijkaardig in vergelijking met de toepassing van het meest complexe model voor alle patchen. Daarenboven werd de rekentijd voor het reconstructiealgoritme zo sterk verlaagd. Vervolgens bleek de verhoogde convergentie in de metaalpatchen de artefacten sterk te verminderen doordat metalen beter gereconstrueerd werden, met hogere attenuatiewaarden en scherpere randen. Tenslotte werden voor de toepassing op echte data enkele extra parameters toegevoegd aan het model en was de ontcorrectie van de data voor sinogram watercorrectie en bowtiecorrectie noodzakelijk.

De lokale modellen reconstructie had sterk verminderde artifacten in vergelij-

king met standaard iteratieve methodes. Sinogram vervolledging gaf beelden zonder donkere strepen en schaduwen maar het contrast tussen en rond metalen was vaak verlaagd. Dit was niet zo voor de iteratieve methodes, hoewel deze iets meer overblijvende artefacten hadden. Voor driedimensionale metalen objecten met een sterke variatie in de axiale richting werd een hardnekkig artefact vastgesteld waarvan de oorzaak niet kon achterhaald worden.

Acceleratie van CT en PET reconstructie

Metaalartefactreductie werd versneld door de complexiteit van het reconstructiemodel te laten variëren over het beeld. De verhoogde convergentiesnelheid van de metaalgebieden was een belangrijke oorzaak van de verbeterde artefactreductie. Dit convergentievoordeel werd verkregen door de patchworkstructuur te gebruiken. Deze structuur werd verder uitgebreid om de convergentie van een beeld in zijn geheel te verhogen zonder variatie in het reconstructiemodel.

Het was reeds bekend dat verhoogde convergentie voor het gehele beeld kan verkregen worden door de voxels van het beeld te verdelen in groepen en die groepen apart te gebruiken in het reconstructiealgoritme. De optimale manier om die groepen te kiezen is door de voxels die behoren tot een groep zoveel mogelijk te spreiden over het beeld. Dit is echter niet efficiënt voor de toegang tot het geheugen van de computer. Het kan daarenboven ook moeilijk gecombineerd worden met zogenaamde geordende projectiesubsets die delen van de projectiedata afwisselend gebruiken.

Voor zover wij weten werd het nooit onderzocht of het gebruik van groepen aaneengelinkte voxels, of beeldblokken, ook leidt tot significant verhoogde convergentiesnelheden. Dit zou een efficiëntere toegang tot het systeemgeheugen en eventueel de combinatie met projectiesubsets toelaten. Deze variant op het patchworkalgoritme, een beeldblokalgoritme, werd getest op zowel CT als PET data. In beide gevallen werd de convergentie aanzienlijk versneld en kon de methode gebruikt worden in combinatie met geordende projectiesubsets.

Biasreductie in PET

Tenslotte werd de reductie van bias in iteratieve PET reconstructie onderzocht. Dit werk is een uitbreiding van het bestaand NEGML algoritme. NEGML reduceert de bias in lage activiteitsgebieden door negatieve beeldwaarden toe te laten. Voor data met weinig gemeten fotonen geeft deze methode echter nog steeds bias.

Een aangepast NEGML algoritme werd ontwikkeld. Dit algoritme vervangt de Poissonverdeling voor lage waarden door de symmetrische Gaussiaanse verde-

ling. Deze verandering van het ruismodel gebeurt enkel lokaal in het sinogram voor lage waarden, elders wordt nog steeds het Poissonmodel toegepast. De nieuwe methode heeft een parameter die bepaald onder welke waarde de Gaussiaanse verdeling gebruikt zal worden. De mate waarin de biasreductie afhangt van deze parameter werd bestudeerd. NEGML werd ook vergeleken met een andere methode voor biasreductie: AML. De parameter A van het algoritme bepaald de ondergrens van de reconstructie en leidt tot een verschuiving van de gebruikte ruisverdeling naar hogere waarden, wat de verdeling ook hier minder asymmetrisch maakt.

Beide methodes reduceren de bias aanzienlijk wanneer hun parameter groot genoeg is. De tweedimensionale simulaties geven aan dat een optimale waarde voor de parameter in NEGML gemakkelijk kan bepaald worden en onafhankelijk zou zijn van de geometrie van het systeem en de activiteit in het beeld. AML gedraagt zich meer zoals MLEM voor kleine negatieve A, wat leidt tot betere convergentie in hete regio's. Er is echter een relatief hoge negatieve waarde nodig om alle bias te reduceren. Anders dan bij NEGML is deze waarde niet onafhankelijk van de geometrie en de activiteit en is een optimale waarde daarom moeilijker te bepalen.

Besluit

In deze thesis werden lokale modellen en lokale convergentie toegepast op iteratieve maximum likelihood algorithmes. Het doel was om de iteratieve methodes te verbeteren en te versnellen. Het gebruik van lokale modellen in het beeld via de patchworkstructuur versnelt de zeer trage iteratieve metaalartefactreductie methodes aanzienlijk. De verhoogde convergentie van de metaalgebieden reduceert een groot deel van de overblijvende artefacten. Een aangepaste patchworkstructuur werd gebruikt voor versnelde convergentie van het gehele beeld in CT en PET. Tenslotte werd een algoritme met een lokaal aangepast ruismodel ontwikkeld voor biasreductie in PET data met weinig gemeten fotonen. Dit nieuwe algoritme resulteert in een sterke reductie van de bias.

List of abbreviations

2D	Two-dimensional
3D	Three-dimensional
ABCD	Axial block coordinate descent
ABML	Maximum likelihood with lower bound A and upper bound B
AML	Maximum likelihood with lower bound A
ANMAR	Adaptive normalized metal artifact reduction
ART	Algebraic reconstruction technique
COMBOS	Patched combination of MLTRC and IMPACT
CPU	Central processing unit
CT	Computed tomography
DNA	Deoxyribonucleic acid
FBP	Filtered backprojection
FDG	Fluorodeoxyglucose
FSMAR	Frequency split metal artifact reduction
FT	Fourier transform
FOV	Field of view
FWHM	Full width half maximum
GCA	Gradient coordinate ascent
GPU	Graphical processing unit
HU	Hounsfield units
ICD	Iterative coordinate descent
IMPACT	Iterative maximum likelihood polychromatic algorithm
MAP	Maximum a posteriori
MAR	Metal artifact reduction
ML	Maximum likelihood
MLEM	Maximum likelihood expectation maximization
MLTR	Maximum likelihood for transmission
MLTRC	Maximum likelihood for transmission with polychromatic correction

MRF	Markov random field
MRI	Magnetic resonance imaging
NEGML	Maximum likelihood with negative values
NMAR	Normalized metal artifact reduction
OMAR	Metal artifact reduction for orthopedic implants
OS	Ordered subsets
OSEM	Ordered subsets expectation maximization
PC	Projection completion
PMMA	Polymethyl methacrylate
PMT	Photomultiplier tube
PET	Positron emission tomography
RAM	Random-access memory
ROI	Region of interest
SAGE	Space-alternating generalized EM algorithms
SART	Simultaneous algebraic reconstruction technique
SIRT	Simultaneous iterative reconstruction technique
SPECT	Single photon emission computed tomography
TOF	Time of flight

List of symbols

α_j	Voxel weight in voxel j
a_{ij}	Sensitivity of line of response i for activity in voxel j
A	Lower bound for reconstruction with AML or ABML
b_i	Blank value for sinogram pixel i, number of photons without attenuation
b_{ik}	Blank value for sinogram pixel i at energy k
β	Prior strength
B	Lower bound for reconstruction with ABML
c	Speed of light
c_{ij}	Sensitivity of line of response i for activity in voxel j
C_k	Correction factor per energy bin k for MLTRC
E	Energy
E_0	Initial energy
$E_{\max}$	Maximum energy
$\mathcal{F}$	Fourier transform
$\mathcal{F}^{-1}$	Inverse Fourier transform
$f(x, y)$	Two-dimensional distribution function
ϕ	Material dependence of the photo-electric effect
Φ_k	Energy dependence of photo-electric effect at energy k
i	Sinogram pixel index
I_0	Initial X-ray intensity at the source
I	X-ray intensity at the detector
j	Voxel index
k	Index of the energy bins
λ_j	Activity in voxel j
L	Poisson (log-) likelihood function
L^ψ	Modified Poisson (log-)likelihood function
l_{ij}	Intersection length of projection ray i with voxel j
μ_j	Linear attenuation coefficient in voxel j
μ_{jk}	Linear attenuation coefficient in voxel j at energy k

μ_{ref}	Linear attenuation at the reference energy
μ_{total}	Total linear attenuation
μ_{water}	Linear attenuation of water
p_i	Projection value in sinogram pixel i
$p(r, \theta)$	Projection at r for angle θ
r_i	Contamination in the measurement in sinogram pixel i by randoms and/or scatter
R	Radon transform
s_i	Contamination in the measurement in sinogram pixel i by randoms and/or scatter
τ_j	Activity or attenuation value in voxel j
s_i^S	Smoothed contamination in the measurement in sinogram pixel i by randoms and/or scatter
θ	Material dependence of Compton scattering
Θ_k	Energy dependence of Compton scattering at energy k
V	Potential function
y_i	Measured number of photons in sinogram pixel i
$\hat{y}_i$	Estimated number of photons in sinogram pixel i
ψ	Transition point from Gaussian to Poisson distribution in NEGML

Contents

Part I

Introduction

Chapter 1

Computed tomography

1.1 X-rays

It was Willem Conrad Röntgen who discovered X-ray or Röntgen radiation in 1895. He discovered a new mysterious type of invisible radiation that could go through pieces of paper, books, etc. and could be detected by photographic plates. He also found that the intensity of the beam measured after passing through an object was inversely proportional to the thickness of the material. Soon he realized that this remarkable property of the so-called X-rays could be used to picture the inside of objects.

The picture in figure 1.1 can be considered as the first radiographic image ever and is said to be the image of the hand of Anna Bertha, Röntgen's wife. From this image it is clear that X-rays interact differently with the different body tissues, i.e. soft tissue, bone and metal in the ring. This discriminating charac-

Figure 1.1: First X-ray photograph ever. It shows the hand of Röntgens wife Anna Bertha [1].

teristic made that X-ray imaging was soon introduced as a clinical diagnostic tool. Pioneers in this work were John Hall-Edwards and Marie Curie. Marie Curie encouraged the use of this new technology for diagnosis of hurt soldiers in World War I by going to the front herself with a movable X-ray imaging system [2]. Nowadays planar radiography is still based on the same principle. An X-ray source emits photons through the body to a detector. This way e.g. bone fractures can be detected. Röntgen received in 1901 the very first Nobel-prize in Physics "in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him".

This chapter discusses briefly the main aspects of computed tomography. References to more detailed information are often given in the text. A general introduction to CT is also given in [3, 4] and [5].

1.1.1 X-rays and X-ray production

X-rays are high energy electromagnetic waves (typically higher than 1 keV) that originate from interaction of free electrons with matter. X-rays are either Bremsstrahlung or so-called characteristic X-rays.

Bremsstrahlung is a German term for *braking radiation* that is produced by decelerating charged particles such as electrons. An electron is slowed down by another (larger) charged particle, e.g. an atomic nucleus. The energy released by this deceleration is emitted through high energy photons or X-rays. This results in a continuous energy range of X-rays, where the maximum energy depends on the initial energy of the electron.

When the incoming electron has sufficient energy to overcome the binding energy of bound atom electrons, it can eject an electron from the atomic energy shells of the target material. This creates a vacancy in the shell which is filled by an electron from an outer shell. The difference in binding energy between the two electron shells will be emitted as a photon. This photon is called a characteristic X-ray since the energy depends on the material of the target. Each target material has different transition energies. Figure 1.2a depicts an X-ray spectrum from a tungsten target material that is bombarded with electrons of 120 keV.

X-rays are produced in X-ray vacuum tubes, an example is depicted in figure 1.2b. The tube consists of a vacuum with an anode and cathode under a certain voltage. Free electrons are created at the cathode side by thermal excitation and accelerated towards the anode. At the anode the electron beam produces Bremsstrahlung and characteristic X-rays. However, most of the total electron beam energy is dissipated by heat due to elastic collisions with other particles. This explains why the anode needs to be cooled during X-ray production.

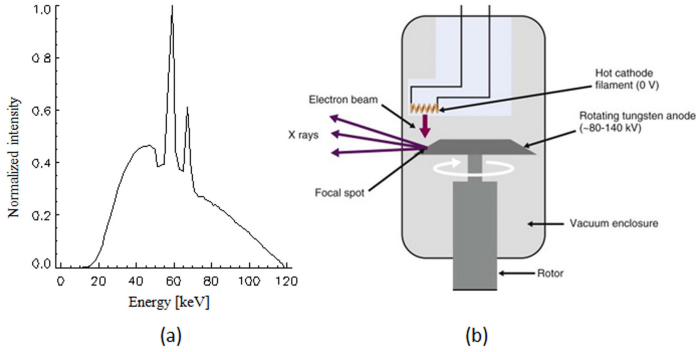


Figure 1.2: (a) X-ray spectrum for tungsten target at 120 kV. (b) X-ray vacuum tube [6].

1.1.2 Interaction with matter

X-rays and other high energy photons have multiple ways of interacting with matter. Depending on the photon energy and the atomic number of the target material, the interactions are more or less likely to happen. The most important effects at clinical X-ray energies are: Rayleigh scattering, Compton scattering and the photo-electric effect. For X-ray radiotherapy also pair production can happen.

Rayleigh scatter

Rayleigh scatter is elastic scattering where the X-ray photon interacts with an atom and is deviated from its original path. This interaction is predominant for X-rays with energies below 20 keV and is often considered as unimportant in typical clinical settings.

Compton Scatter

Compton scatter is inelastic scattering by a photon-electron interaction. The photon scatters from a nearly free outer shell electron, resulting in a scattered electron and a scattered secondary photon. The relationship between the energy of the secondary photon and the scattered angle is given by [7]:

$$E = \frac{E_0}{1 + (E_0/mc^2)(1 - \cos\theta)} \tag{1.1}$$

with E_0 the initial energy, θ the scatter angle, m the electron mass and c the speed of light. The higher the initial photon energy, the higher the probability for scattering in forward direction, i.e. with small θ. The probability

function for Compton scattering at angle θ can be determined and gives the so-called Klein-Nishina function. The probability of undergoing Compton scatter increases with incoming photon energy and slightly increases for increasing atomic number of the target material.

Photo-electric effect

In the photo-electric effect the photon is absorbed by an atom by ejection of an inner shell electron. The electron absorbs all of the photon energy, part of it is used to overcome the binding energy, the rest will be transformed into kinetic energy for the electron. In this process the photon is totally absorbed without the production of a secondary photon.

The probability of undergoing photo-electric effect increases for increasing atomic number of the target material and decreases with increasing photon energy: $P_{\text{photo}} \propto E^{-3}Z^n$ (with n between 3 and 5). The possibility for undergoing photo-electric effect has discontinuities for photons with energies just above the binding energies of the target electrons. For human tissue this energy is below the typical X-ray energy range.

Pair production

At energies above 1.022 MeV pair production becomes possible. During pair production the photon disappears and forms an electron-positron pair. For clinical CT (or PET) systems this energy level is never reached but in radiotherapy treatments this energy level can be reached. Pair production is only possible in the vicinity of heavy nuclei to preserve momentum.

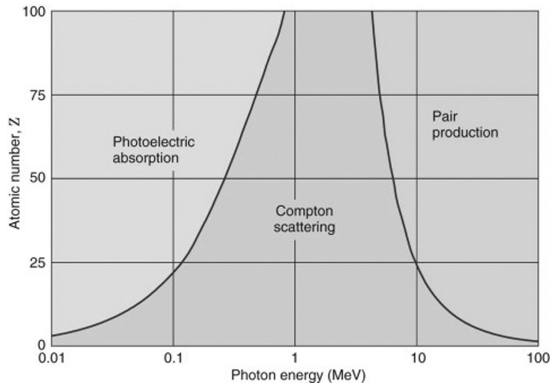


Figure 1.3: Overview of domination of Compton scattering, photo-electric effect and pair production with respect to photon energy and atomic number (Z) of the target material [6].

An overview of the dominant effects for the clinical energy range of photons is shown in figure 1.3.

Total attenuation

The attenuation of a photon beam when passing through a material follows an exponential law:

$$I = I_0 \cdot \exp(-l\mu_{\text{total}}) \tag{1.2}$$

with I_0 the initial photon intensity and I the resulting intensity, l the distance traveled through the material and μ the linear attenuation coefficient of the material expressed in cm^{-1}. μ_{total} is the sum of attenuation by Rayleigh and Compton scattering, photo-electric effect and pair production. The attenuation coefficient is material and energy dependent. The remaining intensity for a beam of photons with energies between 0 and $E_{\max}$ becomes:

$$I = \int_0^{E_{\max}} I_0(E) \cdot \exp\left(- \int_{-\infty}^{\infty} \mu_{\text{total}}(l, E) dl \right) dE \tag{1.3}$$

with $\mu_{\text{total}}(l, E)$ the attenuation coefficient at point l and energy E. In computed tomography the aim is to obtain μ for the different tissues. The attenuation can be reconstructed by measuring the ratio of the intensity of an X-ray beam before and after crossing the patient.

In clinical environments attenuation is usually not expressed by the linear attenuation coefficient μ (cm^{-1}) but in Hounsfield units. The conversion from linear attenuation to Hounsfield units (HU) is given by:

$$\text{HU} = \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}}} \cdot 1000. \tag{1.4}$$

1.2 Computed tomography scanner

X-ray planar scanning was a breakthrough in medicine but soon also its limitations became clear. Planar imaging makes two-dimensional images of the object where all positional information in the direction along the X-ray trajectory is lost. Moreover, since all structures overlay, contrast of the individual structures is reduced. Several theoretical and practical attempts were made to enable three-dimensional imaging [8]. The mathematical framework was initiated by Radon in 1917, though not at all for medical imaging purposes but for his work on gravitational fields. Radon proved mathematically that a three dimensional object can be reconstructed from an infinite number of two-dimensional projections of the object at different angles. In the early 1960s Allan Cormack was the first to develop both theoretically and practically a device to measure and reconstruct photon attenuation values of a cross section of an object. The original purpose of this work was the prediction of

radiotherapy dose distributions across the body. This work stayed relatively unnoticed at that time. In 1969, Godfrey Hounsfield developed the first clinical CT camera at EMI (Electro-Musical Instruments Ltd.). The first real clinical scan was performed in 1971, it was a brain scan where a clear frontal lobe tumor could be observed. Hounsfield did not use the work of Radon and Cormack to reconstruct the image but reconstructed the image by (iteratively) solving a gigantic set of linear equations. Cormack and Hounsfield received the Nobel prize in Medicine in 1979 "for the development of computer assisted tomography".

1.2.1 Geometry

The first CT systems consisted of a single detector-source pair [8]. The attenuation in a set of parallel lines through the object was measured by translational movement of the X-ray source. This was repeated for multiple angles (figure 1.4a). One acquisition with this model could last hours or even days. The second generation CT scanners were faster because they were using multiple pencil-beams and an array of detector elements and later on also a fan beam to illuminate multiple detector elements at the same time (figure 1.4b).

In third generation scanners, a broad detector array is illuminated at once without translation of the source (figure 1.4c). The detector can be flat or curved. The third generation with a curved detector is the basis for modern clinical CT scanners. The system as depicted in figure 1.4c is a single slice system since only one array of detectors is present. Later on also multiple detector arrays were put next to each other (figure 1.4f) [5]. Toshiba has so far the widest multislice clinical CT system with 320 adjacent detector slices [9].

Fourth generation CT scanners have a stationary ring-shaped detector with a rotating source (figure 1.4d). This system has never broken through. This is probably because the improvement in scan duration between third and fourth generation CT systems was limited and a full ring of detector elements is much more expensive than an array.

Cone beam systems are systems where a complete two-dimensional detector setup is illuminated at once (see figure 1.4e). A cone beam system is an extended fan beam system. The names fan beam and cone beam are ambiguous. Usually cone beam is used for squared flat detectors while fan beam is used for multislice curved detectors.

Recently also dual-source systems became clinically available. This system is equipped with two detector-source pairs placed at an angle of 90° with respect to each other. An important motivation for the introduction of such systems is the prevention of motion during the scan. Patients are usually asked to lay still and hold their breath during a scan. However, some patients are mentally or physically not able to lay still. Moreover, involuntary motion like heart beating

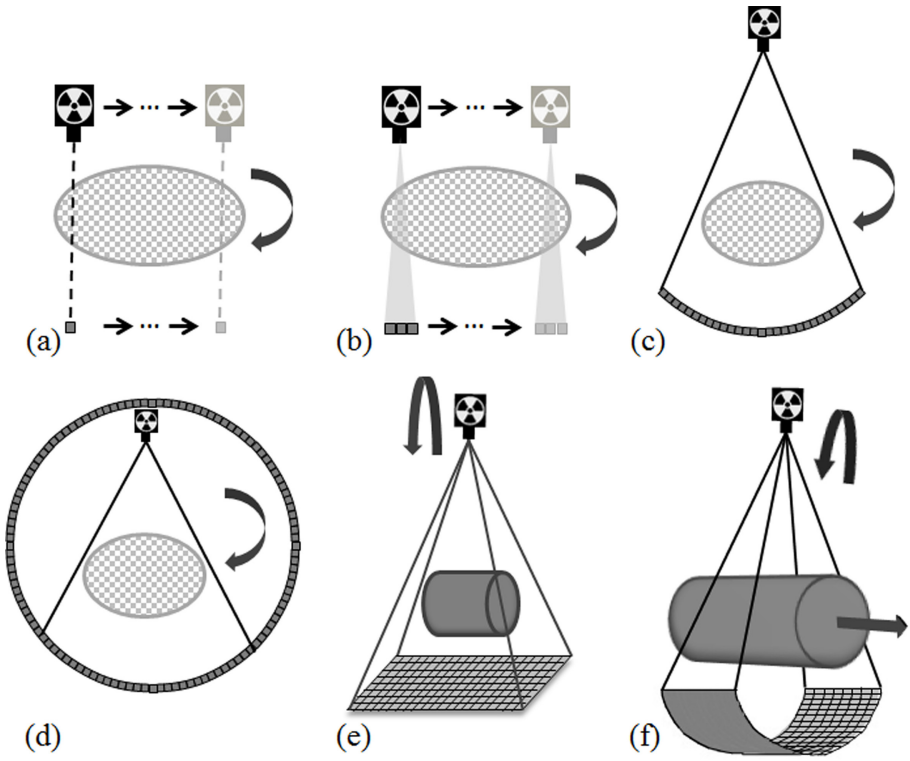


Figure 1.4: An overview of different CT systems. (a) First generation CT, (b) Second generation CT, (c) Third generation CT, (d) Fourth generation CT, (e) Cone beam CT, (f) Helical CT.

is difficult to control. The rotational speed of a CT system is limited by the weight of the systems and by the signal processing time. Nowadays, the fastest rotation time for a clinical CT system is around 0.3s per turn. Fast scanning is necessary to prevent motion during scanning. Using two sources, CT scans can be performed in half the acquisition time. In cardiac CT, these new systems can scan a full heart in one cardiac phase. Moreover the tubes can operate at different source voltages which can be advantageous for discriminating different materials (more details in section 1.2.3).

1.2.2 Scan modes

There are different ways to emit and record X-rays. In the step and shoot mode, the source-detector pair moves till a certain angle and at that point X-rays are emitted. In the continuous mode, the source-detector setup is rotating and emitting X-rays continuously. The detector recordings are split up in (short)

time frames, the recording is then assigned to the mean angle of the system during that time frame.

The detector in multislice CT is usually not wide enough to cover the whole region of interest in axial (longitudinal) direction. To cover the whole range the bed moves during the measurement (see figure 1.4f). In sequential scanning several 360° circular scans are performed. In between the different scans, the patient bed moves to the next position. In helical CT, the patient bed moves continuously at a continuous speed through the CT bore. The X-ray source describes a helical trajectory with respect to the patient. The pitch describes how far the bed moves during one 360° rotation of the source divided by the detector slice thickness or axial beam width (depending on the vendor). Higher pitch means faster bed movement or slower rotation.

1.2.3 X-ray spectrum

The X-rays in CT systems are produced with X-ray tubes (see section 1.1.1). Usually the target material is tungsten (W). For clinical CT, the voltages are set between 60V and 140V. A set of collimators is present at the source. The main reason for this collimation is to prevent unnecessary patient dose along the longitudinal axis of the patient. When scanning the thorax, the dose to nearby regions as neck and abdomen should be kept as low as possible since X-rays sent to these regions will have no diagnostic value but do give dose to the patient.

Filters Most clinical CT systems have flat filters at the source, a thin plate often made of aluminum or copper, to remove low energy X-rays from the spectrum. Low energy photons are very likely absorbed by body tissues and (almost) none of them will reach the detector after passing through a patient. These low energy photons give extra dose to the patient without giving any extra information. Besides flat filters to eliminate low energy photons, most clinical systems also have a bowtie-shaped filter. It is made of a material with attenuation close to soft tissue and has a bowtie like shape, i.e. thick at the outside and thin at the center. The rationale for this, is that the cross section through a body is the largest at the center and becomes shorter to the peripheral part of the body. Thus, generally less photons will be attenuated at the periphery compared to the center of the field of view. Filtering with a bowtie shaped filter yields detector doses with a smaller dynamic range and reduces the patient dose.

Dual energy Recently also dual energy scanners became commercially available [10, 11, 12, 13]. These systems make use of two different spectra: either by a double detector-source pair or by switching the voltage and filters in-between subsequent projection views. Since attenuation is material and energy

dependent, dual energy scanning gives information of the object at two different energies and can improve discrimination between different materials [14].

1.2.4 Detectors

Most modern CT scanners have detectors composed of scintillating crystals linked to photodiodes [3]. When an X-ray photon penetrates a scintillating crystal, the incoming energy is absorbed by excitation of the crystal's molecules. Part of this energy is released as visible or ultraviolet light. The intensity of the light is proportional to the energy of the incoming photon. The light flash is then converted to an electric signal by the photodiodes. In the photodiode the incoming light generates electron-hole pairs. Because of the (reverse) voltage over the diode, the electron-hole pairs produce an electrical current. Since the photon flux for CT measurement is high, current CT detectors work as energy integrators: not the individual photons are counted but the signal they produce in the photodiode is integrated during scanning. No information about the individual photons is left, neither the number of photons nor their energy.

Recently, photon counting detectors are being studied for CT X-ray detection [15]. In photon counting systems the X-ray energy is directly converted to an electrical signal without conversion to light photons [16]. The signal produced by the X-ray photons itself is much higher than the signal received from a scintillating material. Hence, photon counting systems suffer less from electronic noise compared to typical energy-integrating systems. Moreover, the use of direct conversion shortens the time to create an electrical signal and discrimination of individual incoming photons and their energy becomes possible. These systems are getting faster but not yet fast enough to process the high X-ray flux used in a clinical environment.

Anti-scatter grids When a photon is attenuated by Compton scattering, a secondary photon is created with a (slightly) different path. When this path is still crossing the detector area, the photon can be detected. Such scatter photons will contaminate the measurement and can introduce artifacts as discussed in section 5.2.

Scattered photons can be discriminated from primary photons by the incident angle. Most multislice CT systems are equipped with a anti-scatter grid between the different detector elements. The grids consist of small lamellae of a highly attenuating material, aligned with the direction of the X-ray beam such that photons with a deviated path (and deviated incident angles) will be absorbed. In most systems these grids are only placed along the fan-directions and not between the different detector arrays. This has both historical and practical reasons. The first CT systems had only few arrays and the probability to scatter in another detector array was rather low. Moreover, the creation

of stable two-dimensional grids is not easy because when the detector becomes larger, larger lamellae are required that are both thin (to prevent absorption of primary photons) and rigid (to prevent bending) [17].

Afterglow CT systems suffer from afterglow. Afterglow is a small signal that remains from recording the previous view. The energy recorded in the detector for one view is integrated over a certain time. However, some of the excited molecules do not scintillate in this time period and the signal will be measured during the next time period, i.e. for the next view. Moreover, the energy integrator needs to be reset after recording one view, also this resetting is not perfect and some remaining signal might be present.

Cross talk Most systems suffer from some cross talk between the different detector elements. This means that when one detector element is illuminated, some percentage of the signal will leak to neighboring detector elements and be stored as if it was detected in that element.

Noise Quantum noise is considered as the most important noise in CT measurements. It originates from the quantum nature of the attenuation process and can be modeled by a Poisson distribution. However, when using energy integrating detectors, the individual photons are not measured but the energy of the light they produce in the crystal is measured. A more accurate model that includes effects of energy integration is the Compound Poisson model [18, 19]. However, this model is quite complicated and a regular Poisson model usually performs well enough [20]. Even when no photons are arriving in the detector systems, some current or dark current can be measured e.g. due to random generation of electron-hole pairs. This current is small but can become significant when only few photons are detected and can be described as additive (Gaussian) noise.

Chapter 2

Positron emission tomography

Soon after the discovery of X-rays, also radioactive substances, as radium, that emit high energy photons or gamma rays by disintegration of the atomic nucleus, were discovered. Again, the link to study biological structures and processes was soon made [6]. In the first decade of the 20th century, molecules linked to radioactive elements were injected in plants and animals to monitor their trajectory. Important pioneers in this work were de Hevesy, Blumgart and Weiss. A breakthrough was the invention of the cyclotron in the 1930s which opened the possibility to create artificial radioactive elements that could be linked to many more molecules of interest. However, radiation detection was rather cumbersome. Over the years detection became easier with the invention a of the so-called gamma camera. The progress made for CT also opened the path for reconstructing a three-dimensional activity distribution from γ-emitting nuclei (in SPECT, single photon emission tomography) and β^+-emitting nuclei in PET (positron emission tomography).

This chapter discusses briefly the main aspects of positron emission tomography. References to more detailed literature are often given in the text. A recommended work on the physics of PET is written by Cherry and et al. [6].

2.1 Positron emission in medical imaging

2.1.1 Positron decay

In PET, radioactive pharmaceuticals or tracers are injected into the body. These tracers contain radioactive elements that undergo (β^+) or positron decay.

When a disintegration happens, a positron, the anti-particle of an electron, is released. A positron is not stable in the presence of electrons and will soon annihilate with a nearby electron. In such an annihilation process two photons, each of 511 keV (the rest mass of an electron/positron), will be emitted in opposite directions. Since these photons arise from nuclear events, they are called gamma rays.

2.1.2 PET tracers

The tracers used in PET consist of a beta-emitter linked to a molecule of interest. These molecules are typically substances that our body consumes or substances that are bound by certain receptors. The most commonly used PET tracer is fluorodeoxyglucose (FDG), a glucose analogue labeled with radioactive ^{18}F. In a FDG-PET scan, most activity will be measured at places where a lot of glucose is consumed. Overconsumption or decreased consumption can then be used to detect tumors, brain diseases, etc.

2.2 PET scanner

2.2.1 Data acquisition

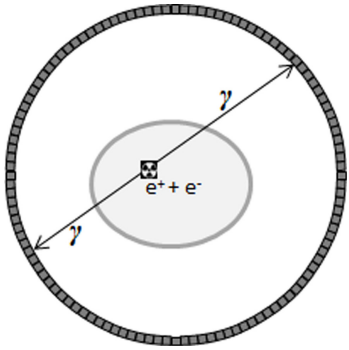


Figure 2.1: Detection of an annihilation event by a PET system.

A typical PET camera has a circular ring of detectors as shown in figure 2.1. The aim of the PET camera is to detect annihilation events based on the two gamma rays that are emitted. Since these gamma rays are emitted simultaneously in opposite directions, gamma rays detected at the same time, a coincidence detection, are assumed to originate from the same annihilation that happened somewhere on the straight line, the *line of response*, between the two detectors that detected the gamma rays.

2.2.2 Detectors

A PET detector is made of scintillating crystals linked to photo multiplier tubes (PMTs) [21]. The principle of the scintillating crystal is the same as for CT: a gamma ray excites the electrons in the crystal and when going back to their ground state, light photons are emitted. These light photons are received by PMTs. In a PMT the light photons hit a target and photoelectrons are emitted by the photoelectric effect. These photoelectrons are accelerated towards a set of dynodes that act as amplifier for the electronic signal. As in CT, the intensity of the signal is proportional to the energy of the measured gamma ray. Since photon fluxes are relatively low in PET, the electrical signal is not integrated but stored as a single detection event together with the energy.

PET detectors typically consist of a block of scintillation crystals, with each block linked to four PMTs. The location of the crystal that received the gamma ray is calculated based on the difference in signal for all four PMTS.

2.2.3 Physical processes

Attenuation and hybrid systems

The photons emitted after annihilation can interact with matter as described in section 1.1.2. If one of the two photons undergoes attenuation, no direct coincidence is measured. This means that a lot of annihilation events are not measured because one or both photons are attenuated. The typical distance a photon beam of 511 keV can travel in soft tissue before half of the photons are attenuated, is around 7 cm. Lines of response crossing denser or longer parts of the body are more likely to undergo attenuation which can result in underestimation of the activity in the center of a reconstructed object. To correct for this effect, the attenuation along a certain line of response is measured and the measured coincidence events are corrected for that. Before, PET systems were supplied with a so-called transmission source. This is a radioactive source that is rotated around the patient before or after the emission scan. Based on the emitted and detected number of photons the attenuation of the patient could be determined.

Hybrid PET/CT systems Where CT is merely an anatomical imaging device, PET gives functional information. It is very useful to combine both this functional and anatomical information for diagnosis. For example in oncology, PET can detect regions of increased glucose consumption but gives rather limited information on the exact anatomical location. When a PET image can be linked to a CT image, the suspicious region can be linked to an anatomical region. Therefore, most PET systems nowadays are combined PET/CT systems, where a CT scan is taken just before or after the PET scan, with the patient on the same table and in the same position [22].

Another advantage is that CT gives information about the attenuation in the patient and can be used to apply attenuation correction which makes a transmission point source superfluous. However, the attenuation measured in CT is not at 511 keV as in PET but for a whole spectrum with mean energy around 70 keV. Therefore, conversion of attenuation from CT energy to PET energy is necessary [23]. This requires the reconstruction of the CT scan, correlation with the PET image, energy conversion and reprojection of the reconstruction. This is distinct from the use of a transmission source where the measured attenuation can often directly be used for attenuation correction without reconstruction.

Hybrid PET/MR systems Soft tissue contrast in CT images is relatively poor, however, much more detail is obtained by MRI. This is the reason why there has been growing interest for combined PET/MRI scanners, e.g. for brain imaging. Anatomical and also functional information from MRI scans can be combined with PET. The most important disadvantage is that MRI gives no direct information about the attenuation. Different methods to obtain attenuation information based on the MRI data or the PET data are being developed [24, 25, 26].

Randoms and scatter

The photons measured by a PET camera can be divided into four main groups: true, scattered, random coincidences and single events. A true coincidence is a measurement of two gamma rays coming from the same annihilation event, traveling on one (approximately) straight line.

A scattered coincidence happens when one (or both) photons undergo Compton scattering. During Compton scattering, the initial photon is absorbed and a secondary photon is emitted with a lower energy and a deviated traveling path. The more oblique the deviation angle, the lower the energy of the secondary photon (see Eq. (1.1)). Discriminating between true and scattered coincidences in PET systems is based on the energy of the measured photons where only photons in a certain energy range are considered as part of a true coincidence. However, this discrimination is not perfect and still a significant number of scattered events will be measured, the ratio of the scattered coincidences to the true coincidences ranges from around 0.3 in brain imaging to larger than 1 for other examinations [6]. Some scatter estimation and correction methods have been developed, more details can be found in [27].

A random coincidence happens when two photons coming from a different annihilation event are measured simultaneously. The ratio of randoms to true coincidences for a PET scans lies around 0.2 for brain examinations and can increase to 2 for other examinations [6]. Single events are detections of single photons. The prevalence of single events can be used to estimate and correct for random events in the measurement [28, 29].

Resolution

Spatial resolution The spatial resolution of a PET camera is limited by several factors [21, 30]. Two factors are inherent for β^+ decay and annihilation detection.

The positrons emitted during β^+ decay have nonzero kinetic energy. This means that before annihilation with an electron, a positron travels a certain distance through the medium during which it interacts with the medium via Coulomb interactions and slows down. This distance a positron travels in a medium is called the positron range. The positron range depends on the isotope and the electron density of the material. For most isotopes the positron range is around 2 mm in soft tissue.

Secondly, when an electron and positron annihilate, their combined momentum is not zero, resulting in two photons that do not travel on one straight line but under a small angle. The resolution restriction caused by this process has a full width half maximum (FWHM) of about 0.4°.

Another resolution determining factor is related to the detection process. The PET system determines in which crystal the scintillation most likely occurred but not at what depth in the crystal, the depth of interaction. Therefore, the intrinsic resolution of the detector is about the size of the detector elements, which is around 4 mm for clinical PET systems. The crystal material and the photon energy define the average depth of interaction. However, since the scintillation is a random process, misplaced lines of response are defined when assuming all scintillations happened at the average depth of interaction. This is referred to as the parallax error and reduces the resolution of the reconstruction. This effect is large at the edge of the field of view and vanishes at the center of the field of view due to the circular geometry of the PET detector.

Temporal resolution The temporal resolution is determined by the time window wherein two photons arriving at different detector elements are considered simultaneous, i.e. coming from the same annihilation event. This time window is limited by variations in signal processing time and by the maximal difference in traveling time of the two photons coming from the same annihilation. The difference in traveling time for the two photons is used in time-of-flight (TOF) PET to determine the location of the annihilation on the line of response more precisely, which improves the image quality. Current commercially available TOF-PET systems have a temporal resolution around 500 ps, which is equivalent to around 75 mm precision on the line of response.

Energy resolution PET systems are made such that only photons within a certain energy range are recorded to reduce the detection of random and scattered events. The typical energy window for a clinical PET system is from 450 to 550 keV.

Dead time

Dead time for a PET system consists of two major components: the scintillation time and data processing time. The scintillation time is the time required for the scintillation. The scintillation is not instantaneous and dependent on the crystal. Other annihilation photons may reach the detector before the previous scintillation is fully finished resulting in a combined signal. The energy linked to this combined signal often falls outside the expected energy window and the detection is rejected. Secondly, once the scintillation is recorded some electronic processing has to be done which also takes time. The time required for both processes is the dead time, the time to process and record an incoming photon.

Sensitivity

The sensitivity describes the efficiency for a pair of detectors to detect activity on the lines of response connecting both detector elements. The inherent efficiency of the crystal and the geometrical efficiency of a detector pair are combined in the so-called normalization part of the sensitivity based on a calibration measurement. The geometrical part is dependent on the solid angle of the detector with respect to the source of activity. Changes in sensitivity due to attenuation are stored separately.

Chapter 3

Ionizing radiation and radiation dose

Both X-rays and gamma rays are ionizing radiation. This means that when they interact with matter, atoms in the object will be ionized resulting in a possibly detrimental biological effect for the human body. The early discoverers of radioactivity were mostly not aware of the detrimental effects of high energy photon radiation. Röntgen was one of the first pioneers who used radiation shielding. When his hands showed erythema (redness), he decided that the origin could be the new radiation he was working with. From then on he always used lead shields when working with X-rays [31]. However, many of the other pioneers did not protect themselves from the radiation and for example Marie Curie died from a kind of blood cancer typically seen when people are exposed for a long time to ionizing radiation. Already from the early beginning detrimental effects caused by the new radiation were reported [32]. In 1903 experiments on animals had shown that exposure to too much radiation could cause cancer and cell death. Throughout the 20th century, people became more an more careful with radiation and radiation limits were set. A lot of information is obtained from huge radiation disasters where many people were irradiated at different radiation levels as for example in Hiroshima and Nagasaki and in Chernobyl.

3.1 Biological effects

When ionizing radiation is absorbed by a tissue this might lead to biological damage, mainly DNA damage and/or cell death [33]. Radiation can cause direct ionization of cell molecules and introduce a chain of physical and chemical reactions that eventually kill the cell. This typically happens for high energy radiation as used for therapy purposes. For X- and gamma rays at energies

used in CT and PET, the most occurring damage is indirect damage. In this case water derived free-radicals (H_2O^+ and OH·) are produced by free electrons arising in the attenuation events. These molecules are extremely reactive: they can break or alter chemical bonds in cells that lead to biological damage.

Ionization in a cell can have influence on cell division. Cell division might be delayed or even made impossible due to damage in one of the components of the reproduction system of the cell. Ionization can also introduce DNA mutations. These mutations can possibly lead to malfunction or cancerous cells.

The effects of ionizing radiation can stay latent for many years, depending on the reproduction rate of the affected cells. The effects can be stochastic or deterministic and long-term or short-term. Most effects are stochastic. This means that chances for encountering effects from the radiation are statistical: they rise with dose but the severity does not increase with dose. Deterministic effects, such as tissue reactions, are effects that increase by dose (usually above a certain threshold). Not all damage induced by radiation will eventually lead to cell damage. Our bodies are already exposed to (natural) ionizing radiation since millions of years and many mechanisms for cell or DNA repair exist.

3.2 Radiation quantification and dose

There are different units to quantify radioactivity and radiation dose. The SI unit for radioactivity is the Becquerel (Bq). The Becquerel is defined as the number of nuclear disintegrations of a substance per second. Before the Curie (Ci) was used, 1Ci = 37 GBq. Radioactive decay follows an exponential law. The decay rate of a radioactive substance is described by its half-life. This is the time needed to reduce the activity of the substance to half of its initial activity. The Becquerel is a unit for radioactive decay but it does not give information about the dose delivered to the irradiated object. The Gray (Gy) is the unit for absorbed dose. It is defined as the absorbed energy by radiation per mass: one Gray equals one Joule energy delivered by radioactive radiation to one kilogram of matter. The Sievert (Sv) is the unit for the equivalent radiation dose and the effective dose. The equivalent dose is the absorbed dose multiplied by the weighting factor dependent on the type of radiation representing the damage potential of that particular radiation. For X- and gamma rays, this weighting is 1, for beta particles this is also 1 and for alpha particles this is 20. The effective dose is also expressed in Sievert and is the equivalent dose weighted for the different body tissues. Some body tissues are more sensitive to radiation than others and this is included in the effective dose.

3.3 Sources of ionizing radiation

Although most concerns are expressed about industrial use of radiation, this is only a very small amount of the total radiation received by the Belgian population. The different origins of radioactive radiation are given in table 3.1 [34].

Table 3.1: Radiation in Belgium in 2006 [34]

Origin	Percentage of Belgian radiation
Cosmic Radiation	8%
Radon	32%
Medical Applications	43%
Body	7%
Buildings and soil	9%
Other human activities	1%

An important part of the radiation we receive on earth is not man-made. Cosmic radiation are high energy particles that encounter us from space and higher atmospheric zones, most of them are primary or secondary products from reactions in the sun. The atmosphere is protecting us from most cosmic radiation, however a certain part reaches the earth surface. In higher atmospheric regions, cosmic radiation is higher. Hence, people that fly often receive a higher annual radiation dose. Moreover, some radioactive elements are already present on earth since its creation: uranium, thorium, potassium-40, etc. These elements have extremely long half-lifes and are therefore still present on earth, merely in deeper earth layers. Potassium-40 is the most important source of natural radioactivity in the human body, approximately 0.01% of potassium in the human body is radioactive. Radon is a gaseous radioactive product of uranium decay. Since it is gaseous it can escape from the soil and through buildings. Radon presence is higher in regions with rocky grounds which makes that the Southern part of Belgium receives almost double earth-related radiation compared to the Northern part.

The natural background radiation is responsible for on average 2.5 mSv per person per year in Belgium. The exposure from other sources of radiation is dependent on a persons situation (job, medical state, air traveling, ...), but is on average below 2.5 mSv for the Belgian population.

3.4 Typical doses

Table 3.2 lists typical doses for some frequently performed medical imaging examinations performed in UZ Gasthuisberg, Leuven, Belgium at the Radiology and Nuclear Medicine Department.

Table 3.2: Doses for some frequently performed examinations at UZ Gasthuisberg, Leuven, Belgium.

Examination	Effective dose
Radiography Thorax	0.06 mSv
CT Thorax	5 mSv
CT Abdomen	8 mSv
CT for attenuation correction	0.5 mSv
PET FDG Whole body	7 mSv
PET FDG Brain	3 mSv
PET/CT Whole body (diagnostic CT)	17 mSv

Chapter 4

Image reconstruction

Image reconstruction is the estimation of the attenuation (CT) or activity distribution (PET) in a three dimensional object from a CT or PET measurement. In CT and PET this reconstruction is obtained based on the measured line integrals or projections from the object at multiple angles. Two types of reconstruction exist: analytical reconstruction and iterative reconstruction. In this section both reconstruction methods are discussed, starting with an introduction of the concepts projection and backprojection. More detailed information can for example be found in [35].

4.1 Projection and backprojection

4.1.1 Projection and Radon transform

Consider a two dimensional object with attenuation or activity distribution $f(x, y)$ in a two-dimensional plane as depicted in figure 4.1. A second coordinate system $\{r, s\}$ is defined, making an angle θ with the original xy-coordinate system. A set of lines parallel to s and perpendicular to r is considered. When applying a line integration of the object along all of these parallel lines, a density profile or projection of the object for angle θ is obtained (see figure 4.1). Mathematically this can be described as:

$$p(r,\theta) = \int_{L(r,\theta)} f(x,y)dxdy \tag{4.1}$$

$$= \int_{-\infty}^{\infty} f(r\cos\theta - s\sin\theta, r\sin\theta + s\cos\theta)ds \tag{4.2}$$

where $L(r, \theta)$ describes the integration line with angle θ at position r. Applying this line integration for all possible $\{r, \theta\}$ is the two-dimensional Radon

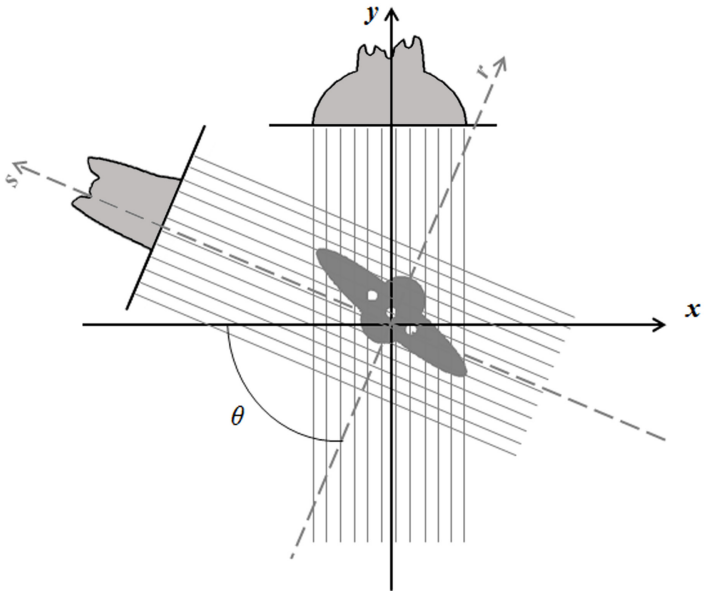


Figure 4.1: Parallel beam projection of an image

transform:

$$(Rf)(r,\theta) = p(r,\theta) \;\forall\; \{r,\theta\} \tag{4.3}$$

In tomographic imaging, $\mathbf{p}$ with elements $p(r,\theta)$ for all measured r and θ is often called the sinogram. When displayed as in figure 4.2 with horizontally index r and vertically index θ, it can be seen that the trajectory of a point through the projections has the shape of a sine.

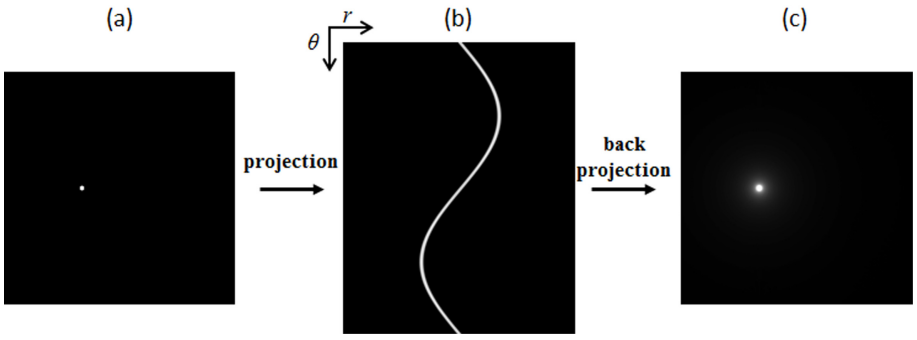


Figure 4.2: Projection and backprojection. The projection of an image (a) results in a set of projections, called a sinogram (b). The back projection of this sinogram (c) is a blurred representation of the original image.

Although CT sources are always polychromatic, a monochromatic X-ray source is usually assumed during reconstruction, giving the following approximation for $p(r, \theta)$ (see Eq. (1.3)):

$$p(r, \theta) = \ln \left(\frac{I_0(r, \theta)}{I(r, \theta)} \right) \tag{4.4}$$

The integral in the projection equation assumes a continuous reconstruction volume and continuous detector. A real detector is discrete and the reconstruction volume is discretized in pixels for two dimensions or voxels for three dimensions. The discrete form of the projection becomes:

$$p_i = \sum_j a_{ij} f_j \tag{4.5}$$

where j is the index over the image voxels, and i the index over the sinogram pixels. For CT, a_{ij} describes the intersection length of projection line i with voxel j. For PET, a_{ij} represents the sensitivity of line of response i for activity in voxel j, including all effects due to geometry, crystal efficiency and attenuation (see 2.2.2).

4.1.2 Backprojection

After measuring the projections as described above, the attenuation or activity values need to be reconstructed. An important operation in reconstruction is the backprojection. The backprojection describes the adjoint operation of the projection [35]:

$$\begin{aligned} b(x, y) &= \int_0^\pi p(x \cos\theta + y \sin\theta, \theta) d\theta \\ &= \int_0^\pi p(r, \theta) \Big|_{r = x\cos\theta + y\sin\theta} d\theta. \end{aligned} \tag{4.6}$$

Practically this means that for every angle θ the projection belonging to that angle is smeared out over the image along lines perpendicular to θ (i.e. along the projection lines that were used to produce the projection), see figure 4.2c. It can be seen from the result that backprojection is no accurate reconstruction method. The resulting image is a kind of blurred representation of the original data.

The discrete form of backprojection is given by:

$$b_j = \sum_i a_{ij} p_i. \tag{4.7}$$

For reconstruction, an inversion of the projection is required rather than the adjoint operation. In general two different reconstruction procedures exist.

In analytical reconstruction (section 4.2) a direct inversion is applied to reconstruct $f(x, y)$. In iterative reconstruction, direct inversion is avoided by replacing it with an iterative optimization process based on the forward model (section 4.3).

4.2 Analytical reconstruction

4.2.1 Central slice theorem

The central slice theorem has a key role in analytical reconstruction. Consider an image $f(x, y)$ and its 2D Fourier transform $F(k_x, k_y)$. The central slice theorem states that the 1D Fourier transform of $p(r, \theta)$ is a central profile through $F(k_x, k_y)$ at angle θ (see figure 4.3):

$$P(k, \theta) = F(k, \theta) \tag{4.8}$$

with $P(k, \theta)$ the 1D Fourier transform of $p(r, \theta)$ and $k_x = k \cos\theta$ and $k_y = k \sin\theta$.

The central slice theorem directly yields a way to reconstruct $f(x, y)$ based on $p(r, \theta)$: calculate the 1D Fourier transform of $p(r, \theta)$ for all angles θ and fill the Fourier space. The initial image $f(x, y)$ can then be obtained by applying the inverse 2D Fourier transform. However, the Fourier space is filled with radial samples and can only be filled completely when an infinite number of central slices are added. If not, interpolation of the Fourier space to a 2D grid is necessary before applying the inverse transform. This interpolation is the main reason why direct reconstruction using the central slice theorem is often not applied since there are usually not enough samples two obtain a good resampling from a radial to a rectangular grid. Especially at higher frequencies this will result in interpolation errors.

4.2.2 Filtered backprojection

An easier algorithm without the need to resample and interpolate in Fourier space can be obtained by expressing the central slice theorem in polar coordinates. The 2D inverse Fourier transform is given by:

$$f(x, y) = \mathcal{F}^{-1}\{F(k_x, k_y)\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(k_x, k_y) e^{i2\pi(xk_x + yk_y)} dk_x dk_y \tag{4.9}$$

After transformation to polar coordinates with $k_x = k \cos\theta$, $k_y = k \sin\theta$ and Jacobian $|k|$, this becomes:

$$f(x, y) = \mathcal{F}^{-1}\{F(k, \theta)\} = \int_0^{\pi} \int_{-\infty}^{+\infty} F(k, \theta) e^{i2\pi k(x\cos\theta + y\sin\theta)} |k| dk d\theta. \tag{4.10}$$

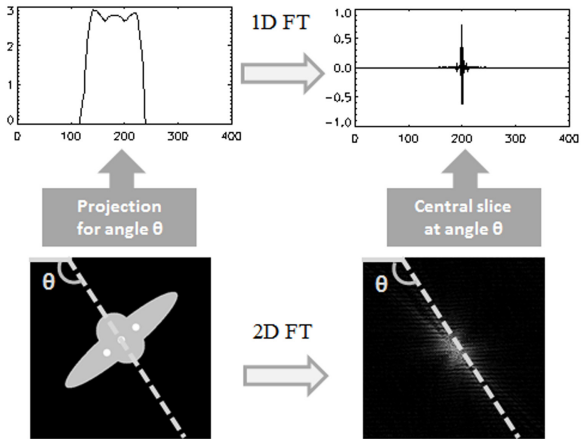


Figure 4.3: Central slice theorem (FT = Fourier transform)

Applying the definition of the backprojection in Eq. (4.6) to Eq. (4.10) gives:

$$f(x,y) = \text{Backproj}(Q) = \int_0^{\pi} Q(r,\theta)\Big|_{r=x\cos\theta+y\sin\theta} d\theta \tag{4.11}$$

with

$$Q(r,\theta) = \int_{-\infty}^{+\infty} F(k,\theta)e^{i2\pi kr}|k|dk. \tag{4.12}$$

The central slice theorem states that $F(k,\theta) = P(k,\theta)$:

$$Q(r,\theta) = \int_{-\infty}^{+\infty} P(k,\theta)e^{i2\pi kr}|k|dk \tag{4.13}$$

This results in the well-known filtered backprojection (FBP) algorithm:

$$f(x,y) = \text{Backproj}(\text{Rampfilter}(\mathbf{p})) \tag{4.14}$$

where the sinogram $\mathbf{p}$ is filtered with ramp filter $|k|$ and backprojected to obtain the reconstruction $f(x,y)$.

4.2.3 Other geometries

CT The FBP algorithm described above is geometry dependent. The reconstruction formula of (4.14) only holds for parallel beam geometries. An extension to 2D fan beam is relatively straightforward. For cone beam imaging the most commonly used algorithm is the one described by Feldkamp, Davis and Kreiss (FDK-algorithm) [36]. The problem for exact analytical reconstruction of helical CT data has remained unsolved for a long time. Only in 2002

Alexander Katsevitch has described a method that applies filtered backprojection to helical CT data [37, 38]. The algorithm has several limitations and has to be seen as a mathematical tool rather than a reconstruction tool for use in practice. For example: the algorithm is only applicable for certain combinations of detector collimations and pitches. Although most CT systems are using analytical algorithms, they do not necessarily use the dedicated reconstruction algorithm. Often the data are rebinned to 2D parallel or fan beam geometries. Rebinning is a process where the projection points are reordered and matched (by interpolation) as if they were coming from another geometry.

PET Analytical reconstruction for PET data is usually applied on data rebinned to a two-dimensional parallel beam geometry [39] and regular filtered backprojection is applied on the data.

4.3 Iterative reconstruction

4.3.1 Principles

The concept of iterative reconstruction is different from that of analytical reconstruction. Instead of applying an inverse model, it aims to find the image that after applying a forward acquisition (projection) model, best resembles the measured data. The theoretical derivation of FBP is inherently based on an infinite number of samples, iterative reconstruction directly assumes a discrete projection model.

The general scheme of iterative reconstruction is shown in figure 4.4. Iterative reconstruction starts from an initial estimate of the reconstructed data, often a uniform image is chosen. This image is projected according to a chosen model to obtain an estimate of the measured data. Based on a predefined cost function, the estimated projection and measurement are compared and an update for the image estimate is calculated. This new estimate is used to initialize a new iteration. The iteration process goes on until the estimated projection is equal (or close enough) to the measurement.

Iterative reconstruction methods differ in the acquisition/projection model used to calculate the estimated projection data and the cost function they are using. This section describes the most commonly used cost functions: least squares estimates (algebraic reconstruction) and the Poisson likelihood (maximum likelihood reconstruction) and common adaptations to the projection model.

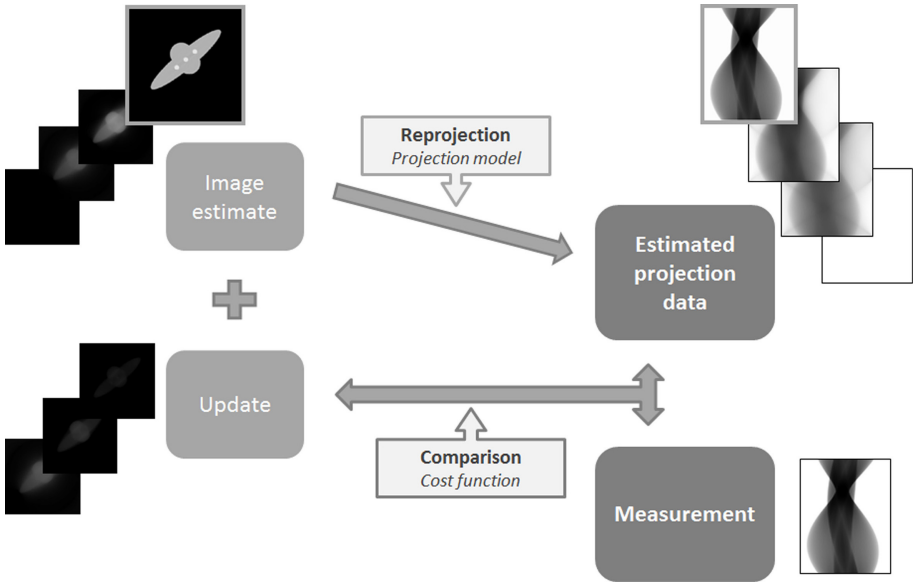


Figure 4.4: Principle of iterative reconstruction, shown here for transmission data.

4.3.2 Algebraic reconstruction

In algebraic reconstruction the problem is stated as: find the distribution $\mathbf{f}$ by solving

$$\mathbf{p} = A\mathbf{f}. \tag{4.15}$$

A is the system or projection matrix containing elements a_{ij} (see Eq. 4.5). A one-step solution to this problem would be to apply A^{-1} on $\mathbf{p}$. Since A is usually large, often singular, and usually rectangular (i.e. $\in \mathbb{R}^{N\times M}$ with $N \neq M$), it is difficult or impossible to calculate A^{-1}. A solution for the problem in Eq. (4.15) is found by applying the Landweber algorithm for solving an ill-posed inverse problem [40]. Eq. 4.15 is solved by minimizing:

$$\Psi(\mathbf{f}) = \|\mathbf{p} - A\mathbf{f}\|^2 \tag{4.16}$$

for $\mathbf{f}$. Minimizing $\Psi(\mathbf{f})$ results in the following update formula:

$$\Delta\mathbf{f} = \rho A^T(\mathbf{p} - A\mathbf{f}^{\text{old}}) \text{ with } 0 > \rho > 2 \tag{4.17}$$

where A^T is the transpose of the projection operation, i.e. the backprojection: $[A^T\mathbf{p}]_i = \sum_j a_{ij}p_i$, $\mathbf{f}^{\text{old}}$ is the current estimated image and ρ a constant. The solution of this iterative reconstruction will converge to the least squares solution of the problem in Eq. (4.15).

The reconstruction algorithm as indicated in Eq. (4.17) is SIRT [41]. By applying Eq. (4.17) all data points are used to update image $\mathbf{f}$ once. If $\mathbf{f}$ is large

as in medical image reconstruction, this might take very long. Acceleration can be obtained by dividing the projections in subsets and sequentially using the subsets to update $\mathbf{f}$ [4, 42, 43, 44]. In ART each individual data point in $\mathbf{p}$ is sequentially used to update $\mathbf{f}$. In SART, all data points measured in one projection view, i.e. for one angle θ are used simultaneously. The advantage of using these subsets of the data is a faster convergence rate. The disadvantage is that instead of reaching the least squares solution a limit cycle solution around the true solution might be obtained.

4.3.3 Maximum likelihood reconstruction

Maximum likelihood reconstruction uses a different approach. It aims to find the distribution $\mathbf{f}$ that maximizes the probability of $\mathbf{f}$ given a measurement $\mathbf{p}$: $P(\mathbf{f}|\mathbf{p})$. To calculate $P(\mathbf{f}|\mathbf{p})$ Bayes rules can be applied:

$$P(\mathbf{f}|\mathbf{p}) = \frac{P(\mathbf{p}|\mathbf{f})P(\mathbf{f})}{P(\mathbf{p})}. \tag{4.18}$$

$P(\mathbf{f})$ is prior information about $\mathbf{f}$ and $P(\mathbf{p})$ is independent of $\mathbf{f}$. Assuming no prior information is known, optimizing $P(\mathbf{f}|\mathbf{p})$ is equal to optimizing $P(\mathbf{p}|\mathbf{f})$. Since all measured data points are independent the probability can be factorized and a Poisson noise model can be included:

$$P(\mathbf{p}|\mathbf{f}) = \prod_i P(p_i|\mathbf{f}) = \prod_i \frac{\hat{p}_i^{p_i} e^{-\hat{p}_i}}{p_i!} \tag{4.19}$$

with $\hat{p}_i$ the expected value of p_i based on the reconstruction image $\mathbf{f}$ Taking the natural logarithm of both sides and ignoring all terms independent of $\hat{p}_i$ results in the (log-) likelihood used in maximum likelihood reconstruction:

$$L = \sum_i p_i \ln \hat{p}_i - \hat{p}_i. \tag{4.20}$$

Different approaches to optimize the likelihood formula are used. In emission tomography the maximum likelihood expectation-maximization (MLEM [45]) procedure is usually applied. In transmission tomography where the model for $\hat{\mathbf{p}}$ is non-linear due to the exponent, MLEM is difficult to calculate and results in a very slow algorithm. In case of transmission data usually more general, but often approximated, gradient ascent approaches are applied [46, 47]. (A more detailed description of maximum likelihood reconstruction algorithms is given in chapter 7 and 8.)

For PET the basic formulation of $\hat{p}_i$ as a function of the activity λ_j in voxel j is :

$$\hat{p}_i = \sum_j a_{ij}\lambda_j. \tag{4.21}$$

The strength of maximum likelihood reconstruction for CT is that the method can be directly applied on transmission data, instead of first applying a log-conversion to obtain attenuation data. Thus calculating $P(\mathbf{y}|\mathbf{f})$ instead of $P(\mathbf{p}|\mathbf{f})$, with $\mathbf{y}$ the transmission projection data. A basic monochromatic transmission projection model for CT projection is given by:

$$\hat{y}_i = I_0 \exp(-\hat{p}_i) = I_0 \exp(-\sum_j a_{ij}\mu_j) \tag{4.22}$$

with μ_j the attenuation coefficient in voxel j. Reconstruction with a gradient ascent approach as in MLTR [46] results in the following update equation:

$$\Delta\mu = \frac{\sum_i a_{ij}(\hat{y}_i - y_i)}{\sum_i a_{ij}(\sum_h a_{ih})\hat{y}_i} \tag{4.23}$$

More detailed update formulas for CT and PET are given in chapter 7 and 8. As for algebraic reconstruction the convergence of the reconstruction procedure can be accelerated by applying the update sequentially for subsets of the data. The ordered subsets (OS) algorithm was presented by Hudson and Larkin for MLEM [48] and divides the projections in small groups, subsets, of a few projection views. To obtain the highest convergence gain, projection views in one group should be as different as possible, i.e. not from neighboring or opposing views. Similar as in ART and SART, the use of subsets might cause limit cycle solutions.

Maximum a-posteriori reconstruction

The rule of Bayes (Eq. 4.18) yields a direct way to include prior information about the object into the reconstruction algorithm. In the previous section we assumed that no prior information about the object was known. A function describing detailed prior information is hard to define. However, the fact that body structures have relatively homogeneous attenuation or activity distributions is more easily included [49].

The prior information $P(\mathbf{f})$ can be expressed by a Markov Random Field prior:

$$P(\mathbf{f}) = \frac{1}{Z}\exp(-\beta\sum_j\sum_h w_{ih}V(\mu_j,\mu_h)) \tag{4.24}$$

where β influences the strength of the prior, Z is a normalization constant, w_{ih} is a weighting factor and $V(\mu_j,\mu_h)$ is a potential function. Weighting between different elements by w_{ih} is usually applied in the close neighborhood of voxel j. $V(\mu_j,\mu_h)$ only compares voxel values μ_j and μ_h and is often defined such that large differences between μ_j and μ_h have high potential and therefore low probability, e.g. by a quadratic function:

$$V(\mu_j,\mu_h) = (\mu_j - \mu_h)^2. \tag{4.25}$$

For reconstruction the (log-)likelihood is extended with the logarithm of the prior knowledge:

$$L = \sum_i p_i \ln \hat{p}_i - \hat{p}_i + \beta \sum_j \sum_h w_{ih} V(\mu_j, \mu_h). \tag{4.26}$$

The combined likelihood is optimized in a maximum a-posteriori (MAP) algorithm. The strength of the prior depends on the value of β. A quadratic prior gives low probability to large differences between neighboring voxels and therefore tends to lower these differences, causing a smoothing effect in the reconstruction. In reconstruction of medical images this is not always what we want since it will also smooth edges between different tissues. Therefore other prior functions with edges preserving characteristics can be used (e.g. Huber prior [50], total variation prior [51]).

Prior information or a 'penalty' term can also be included in algebraic reconstruction algorithms. Either by optimizing a combined cost function consisting of a data term and penalty term, as for MAP, or by sequentially updating the image in two separate steps: the first based on the data followed by an update based on the penalty.

4.3.4 Projection model

The projection or acquisition model plays an important role in iterative algorithms. The better the model resembles the true acquisition, the better the reconstruction. However, more accurate models usually require more complicated calculations that can prolong the calculation time per iteration substantially.

Nuyts et al. [20] gives an overview of several physical effects that can be included in a CT reconstruction model. Standard models for CT are monochromatic. Several polychromatic algorithms have been developed and are described in more detail in chapter 7 and 9. Resolution effects can be modeled by considering the finite size of the source and detector and by using a voxel footprint during projection instead of representing a voxel by a single point. The likelihood model can be modified for Compound Poisson noise statistics, detailed (Compton) scatter profiles can be included and possible patient motion can also be incorporated.

For PET, the main focus to improve the projection model considers attempts to recover resolution. An overview of geometric and physical factors that influence the resolution is given in [52, 53]. Factors that affect resolution can be detection related like the depth of interaction in the crystal and scatter in crystal, but also geometrical effects like nonuniform sampling and finite detector size or tracer related effects like the positron range influence the resolution.

4.4 Reconstruction in practice

Although the very first CT systems used an iterative reconstruction algorithm based on algebraic reconstruction, nowadays the standard reconstruction on almost all CT systems is analytical. This is mainly because analytical reconstruction is a one-step procedure which is much faster than iterative reconstruction. Analytical reconstruction assumes an ideal model. In an ideal model the reconstruction volume consists of points (that are presented as voxels for displaying images) and projection lines are infinitely thin lines connecting a point source with a point detector element. Moreover in ideal systems the measurements are assumed to be true projections of the distribution function without noise.

Under most circumstances, CT measurements are fairly close to ideal. The reconstruction volume is well sampled and the photon flux is high which results in rather low noise. Analytical reconstruction only starts to introduce artifacts in the images when these criteria are violated to a higher extent. This happens for example when the photon flux decreases and the measurements have higher noise levels. Analytical reconstruction will still treat the measurement as if it was noise free, giving equal importance to rays with high and low noise. This typically results in streaks over the image. Analytical reconstruction is also sensitive to insufficient or incomplete sampling of the Fourier space. This happens for example when the angular sampling of the object is coarse or incomplete, or when parts of the object are not visible for the detector under certain angles, often called truncation. Under such circumstances, analytical reconstruction will introduce artifacts in the image.

The performance of FBP in different situations is shown in the upper row of figure 4.5. Four different simulations were performed. The first simulation was well sampled, noiseless and free of truncation. The second simulation contained noise, in the third simulation the arms of the simulated patient were truncated during scanning and in the fourth simulation the angular sampling was strongly decreased. It can be seen that the image quality for FBP reconstruction decreases in all nonideal simulations.

In emission tomography where photon fluxes are lower and sampling is poorer, iterative reconstruction is already the standard reconstruction method for many years. The main advantage of iterative reconstruction is the forward model used in the reconstruction procedure. The model is not limited to ideal circumstances. Moreover when applying maximum likelihood reconstruction, projection lines will be weighted according to the Poisson model, thus incorporating the better signal-noise ratio of lines with high counts. Images reconstructed with a maximum likelihood cost function have fewer (streak) artifacts caused by noise. The possibility to weight projections during the reconstruction according to their intensity has gained interest for CT reconstruction. Over the last years, more and more concerns are rising about the dose delivered to the patient by a CT scan, especially when patients need a series of CT scans or

for pediatric patients. A lot of effort is done to reduce patient doses for CT examinations. Dose can be reduced by decreasing the photon flux, by decreasing the sampling rate, etc. These are typical circumstances where analytical reconstruction results in poorer image quality. Iterative reconstruction is seen as the solution to further reduce dose while maintaining image quality. This is done by including a statistical noise model but also by modeling the system geometry more accurately.

Another advantage of iterative reconstruction in CT, compared to analytical methods, is the possibility to work with transmission data rather than log-converted data. This allows to use Poisson statistics for the data. Moreover, the log-conversion can become (computationally) unstable for low intensities. To avoid this, low data points are often artificially increased to a certain point where the logarithm becomes stable. By applying the log-conversion for very low intensities, positive bias can be introduced to keep the logarithm stable.

In figure 4.5 both SART and MLTR are applied on the same data sets as FBP. As expected MLTR is the least sensitive to noise. In case of truncation MLTR and SART perform both better than FBP. In case of angular undersampling SART and MLTR give better results than FBP where SART seems a little bit better than MLTR.

The major CT vendors today have an option to install iterative CT reconstruction on one or more of their systems. However, very few of them implemented true iterative reconstruction as explained above. The main reason is that going back and forward from projection domain to image domain is extremely computationally demanding. Usually only an approximated iterative procedure that stays in the image domain is implemented, with in some cases a few feedback loops to the projection domain. As an example, a typical whole body CT scan is more than 1 Gb, the corresponding reconstruction is less than 300 Mb. Full iterative reconstruction for CT would still take one or more hours and this is not acceptable for routine clinical usage. In research environments many iterative reconstruction algorithms for CT have been developed. In this work both monochromatic and polychromatic algorithms are studied. Their advantages over analytical reconstruction and possibilities for acceleration are investigated.

As mentioned above, iterative reconstruction for PET is already well accepted. However, there is one important exception. For quantitative analysis of the reconstruction, still analytical reconstruction is preferred over MLEM. The reason is the tendency of MLEM to introduce positive bias for very low count data and in low activity regions surrounded by regions with higher activity. The origin of the bias and a possible solution to ameliorate the iterative methods for the problem are studied in chapter 10.

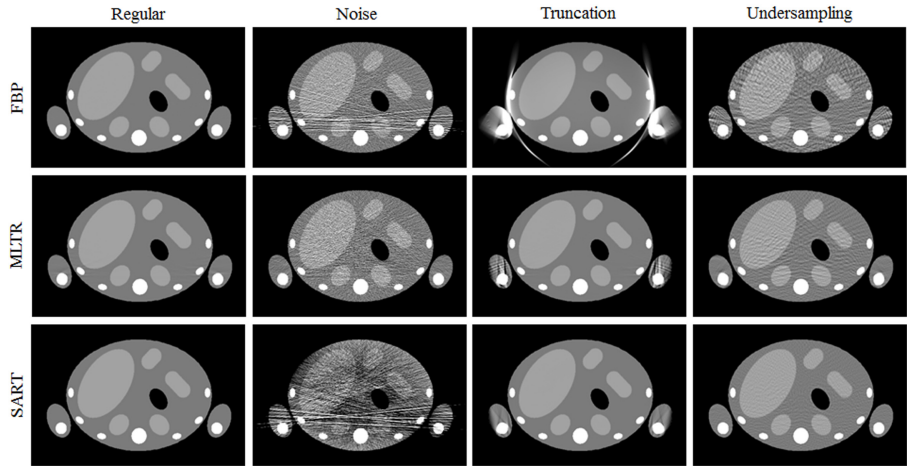


Figure 4.5: Comparison FBP, MLTR and SART in CT for a regular well sampled simulation and simulations including noise, truncation or angular undersampling.

Chapter 5

Metal artifacts in CT

Computed tomography has strongly evolved since its invention. Under most circumstances computed tomography delivers high quality images, with relatively high resolution and few artifacts. An important cause for image degradation is the presence of metals with high attenuation such as metallic implants and dental fillings. The artifacts typically appear as dark and bright streaks and shadows that mask underlying structures. This often hinders proper evaluation and diagnosis in the affected region (fig. 5.1). The most important causes of metal artifacts are beam hardening, scatter, partial volume effects, noise and motion [54, 55]. Each of these effects will be discussed in this section. In figure 5.2 an overview is given of the effect of the different causes of metal artifacts separately. The reconstructions have been performed with FBP.

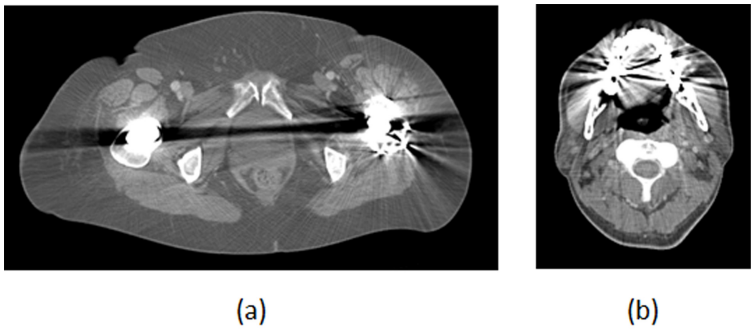


Figure 5.1: Example of metal artifacts. (a) CT scan of patient with two femoral (hip) prostheses. (b) CT scan of patient with dental fillings.

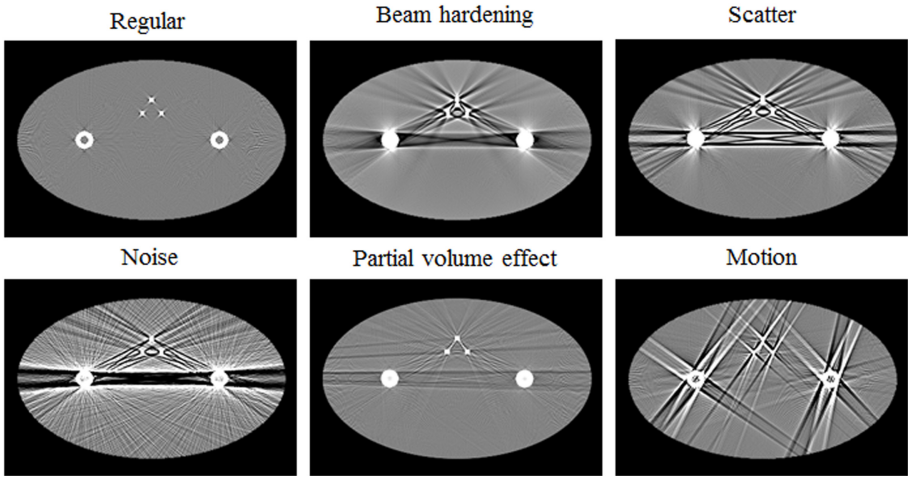


Figure 5.2: Simulation and reconstruction for the different causes of metal artifacts. The reconstructions are performed with FBP.

5.1 Beam hardening

The X-ray photons used in CT have a broad spectrum of energies. When passing through an object the lower energy photons will more likely be attenuated. This makes that the mean energy of the beam increases and the relationship between total attenuation and object length becomes nonlinear (figure 5.3). This effect is not incorporated in most reconstruction algorithms, instead a linear behavior is assumed.

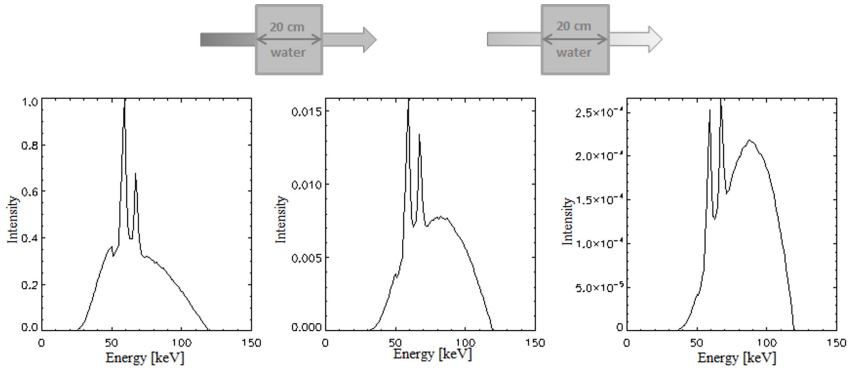


Figure 5.3: Hardening of the X-ray spectrum. A spectrum of X-rays is sent through 20 cm and 40 cm of water. The resulting spectrum is harder than before.

For objects with relatively low attenuation, this leads to cupping. Cupping

is shown in figure 5.4 where the reconstruction without sinogram correction has lower attenuation values towards the center of the object. Typically, the sinogram data are precorrected for this effect. If one assumes that the object contains only one material (in clinical applications this is water or soft tissue), a polynomial can be created that links the resulting total attenuation for a polychromatic source with the attenuation that would be obtained for a monochromatic source. In figure 5.4, a reconstruction based on sinogram data with and without precorrection is shown. For water this correction is correct, other materials are only partially corrected depending on the difference in their attenuation properties compared to water. Most metals attenuate much more than water and even when applying the water precorrection the metal value is still far from the true value. When beam hardening is not fairly incorporated in the reconstruction algorithm, serious artifacts may appear. These artifacts are typically dark streaks and shadows around and in between metals, as can be seen in figure 5.2b. Artifacts due to beam hardening can be avoided by incorporating energy dependent attenuation coefficients in the reconstruction model (see chapter 7). When the metal object attenuates so strongly that no photons remain, the term photon starvation is used. In this case no information about the projection lines through the metal is obtained. This results in incomplete tomography with stronger artifacts than for beam hardening with remaining photons.

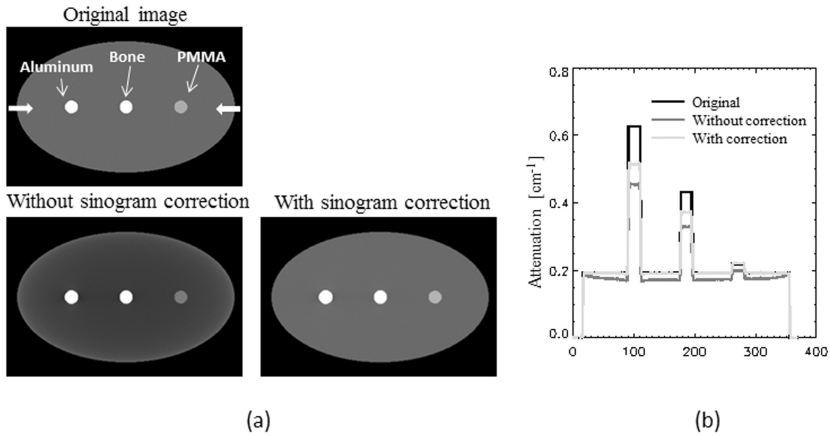


Figure 5.4: Effect of sinogram beam hardening correction on the reconstructed attenuation values. The image reconstructed based on a sinogram that is not corrected for beam hardening effects shows cupping (lower image values towards the center of the object). When the sinogram is corrected for (water) beam hardening effects, the reconstruction has no cupping anymore in the water region, other materials are only partially corrected.

5.2 Scatter

When a photon is attenuated, it is assumed to be removed from the X-ray beam by most standard reconstruction algorithms. This is correct if it is attenuated via the photo-electric effect but during Compton scattering secondary photons are produced. When the angle of the deviation is such that the photon stays within the fan, it can be detected and contaminates the measurement. For relatively low attenuating objects this again results in cupping. When highly attenuating objects, like metals, are present in the field of view more severe artifacts can arise, typically dark and white streaking. The projection through the metal receives only few photons, when photons from neighboring structures scatter into the metal projection rays, a significant increase in detected photons is achieved (see figure 5.5a).

Scatter can be reduced by hardware and software solutions. In clinical multi-slice systems a anti-scatter grid is placed on top of the detector as discussed in section 1.2.4. Scatter can also be corrected afterwards [56, 57], e.g. by object-dependent Monte Carlo simulations. Based on an initial reconstruction, the percentage of attenuation due to Compton scatter is estimated by a Monte Carlo simulation and the number of scatter photons reaching the detector is estimated. The obtained scatter profile can be subtracted from the measurement or included in a forward projection model to reduce artifacts in the image. This procedure is not applied on a routine basis, since accurate Monte Carlo simulations are very time consuming.

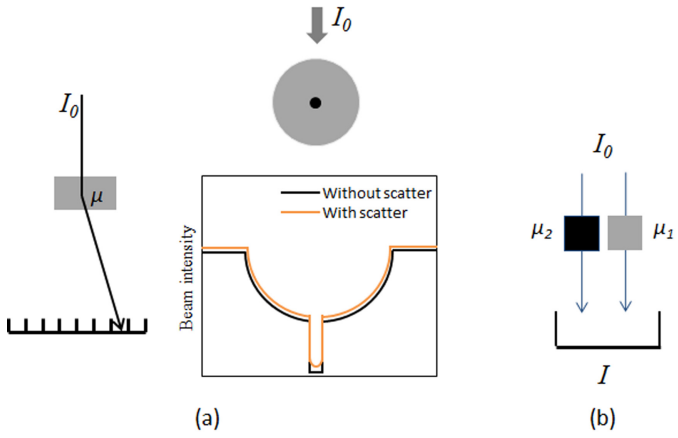


Figure 5.5: (a) Effect of Compton scatter. The relative effect of scatter will be larger in metal projection. (b) Partial volume effect.

5.3 Partial volume effects

Partial volume effects in CT are two-sided. A linear effect, also seen for other imaging modalities as in PET, is when the volume represented by a voxel is occupied by two materials with different attenuation properties. The resulting attenuation for that voxel will be a volume averaged attenuation. However the exponential nature of CT measurement makes the effect nonlinear as is shown in figure 5.5b [58]. Integration over the beam width, focal spot and projection views (in case of continuous rotation) gives rise to nonlinear partial volume artifacts. These artifacts are fine streaks connecting material boundaries and they are more pronounced when two neighboring materials have a large difference in attenuation (see figure 5.2d). The origin of nonlinear partial volume effects can be described by the following inequality:

$$I_0 \int_{\text{beam}} \exp(-l\mu_s)ds \neq I_0 \exp(-l \int_{\text{beam}} \mu_s ds) \tag{5.1}$$

where the left-hand-side is the integration as it is happening during a CT measurement and the right-hand-side is the integration assumed by most reconstruction algorithms.

5.4 Noise

Poisson photon noise caused by the quantum statistical nature of an attenuation event is considered the most dominant. It appears as small dots in the images. As with scatter, small changes in low projection values have large influence and give rise to artifacts. Projections through metals have low intensity and noise will have larger influence than for other projection rays. The effect of noise in a reconstruction with metals is shown in figure 5.2e. For small metals this results in fine streaks connecting the different metals, for larger metals the streaking is stronger causing dark bands in between the metals. This effect is less pronounced when the reconstruction model includes Poisson noise as in MLTR.

Detection related noise is often not considered. However, for measurements of only few photons it might also be important (see section 1.2.4).

5.5 Motion

When an object or patient moves during the CT acquisition the measured data become inconsistent. The same object measured in a certain position for one angle will be measured in another position for the next angle. During reconstruction this gives conflicts between the information from different projection angles and artifacts arise. When metals are present, these artifacts become

more prominent. To prevent motion patients are often asked to hold their breath and to lay as still as possible. However, some patient are mentally or physically not able to do so. Moreover, involuntary motion, like heart beating, cannot be avoided. Hence, motion tracking and correction methods are being developed. These methods record the movement of the object and include the movement in the projection and backprojection process [59, 60, 61, 62].

Part II

Objectives

Chapter 6

Objectives

Three-dimensional cross sectional imaging has become essential in a clinical environment. The most common modalities for three-dimensional imaging are CT, MRI, PET and SPECT, where CT examinations are the most performed. CT, PET and SPECT have in common that they all measure line integrals of attenuation (CT) or activity (PET and SPECT) through the patient. Progress in reconstruction software can often be applied for both types of measurements. Since their development both hardware and software components of the scanners have been continuously improved to deliver images with increasing image quality: higher resolution and fewer artifacts. Recently also more attention is given to the balance between image quality and patient dose.

Where analytical reconstruction has been the gold standard for many years, more and more interest is shown for iterative reconstruction algorithms. Iterative reconstruction performs better for noisy (low dose) data and can reduce artifacts by the incorporation of more accurate system models. The main disadvantage is that iterative reconstruction has much higher hardware requirements regarding memory and computational power. Only recently clinical CT systems are equipped with iterative reconstruction software. In PET (and SPECT) where data sets are much smaller, iterative reconstruction has currently almost completely replaced analytical reconstruction.

The purpose of this work is the amelioration and acceleration of tomographic iterative reconstruction.

The first and main part of the thesis is devoted to the study of metal artifact reduction in CT. Although CT images are usually of very high quality, metal artifacts are still an important problem. Many methods have been developed to reduce metal artifacts. The most popular methods are based on analytical reconstruction. These methods discard all information obtained from lines that went through metallic objects and replace these data with artificial data. This is very efficient for reduction of streaks and shadows. The major disadvantage

is that by removing artifacts also contrast is decreased. A second group of methods uses iterative reconstruction to reduce metal artifacts. The idea for the iterative algorithms is to include a model for the different physical processes that are the basis of the artifacts. These methods are usually only efficient for few and rather low attenuating metals. Moreover, the required computation time is often more than doubled with respect to (already slow) regular iterative reconstruction. However, these iterative methods do not result in decreased contrast.

The two important problems for iterative metal artifact reconstruction in CT: remaining artifacts and slow reconstruction, are studied. The reconstruction models applied to achieve reduction of metal artifacts accurately simulate some aspects of the acquisition process. A more accurate model usually improves artifact reduction but is often also more complex which increases the computation time for the reconstruction. These complex models are not needed for reconstruction of objects without metals. Therefore, a local model reconstruction scheme, called patchwork reconstruction, is developed that enables the use of different reconstruction models for different regions of the object. Complex models are used in and around the metals and simpler models are applied elsewhere. This will reduce the computation time considerably. Moreover, the causes for the remaining artifacts are investigated. This work is presented in chapter 7 for two-dimensional phantom studies and in chapter 9 for application on helical CT patient data.

The patchwork technique that enables varying reconstruction models over the image, can also be used to accelerate the overall convergence rate of the reconstruction. The patchwork reconstruction is designed such that all defined regions or *patches* receive a separate update in each iteration. It has been shown before by Fessler et al. [47, 83] that updating subsets of the voxels separately, improves the convergence rate of the reconstruction algorithm. However, it was suggested that voxels belonging to one subset should be chosen spread over the reconstruction volume such that they have low influence on each other. To our knowledge, it was never shown to what extent the convergence would improve when neighboring voxels are updated as one group. Choosing blocks of neighboring voxels as a group has some advantage over choosing voxels spread over the image. Neighboring voxels values are often stored at neighboring locations in the computer memory which makes updating groups of voxels spread over the image less efficient, regarding memory access, compared to updating grouped voxels. It is also worth studying whether updating blocks of image voxels can be combined with ordered projection subsets, possibly enabling a two-sided acceleration.

In chapter 8 the rate of convergence for an image-block patchwork reconstruction is studied in combination with ordered projection subsets for CT. The same mechanism can also be applied to PET reconstruction models and is investigated as well in chapter 8.

The second part of this thesis considers PET reconstruction. A first element is

the study of accelerated convergence by the use of image-blocks as described in the previous paragraph. A second element considers bias reduction for reconstruction of low count PET data. Standard PET reconstruction in the clinic is almost always done with iterative (maximum likelihood) reconstruction. The difference in noise and resolution characteristics between analytical and iterative reconstruction is large with iterative reconstruction outperforming analytical reconstruction. However, for quantitative analysis of the reconstructions, the standard iterative algorithm (MLEM) does not always result in accurate activity values. A systematic deviation from the true activity, *bias*, is sometimes observed when using MLEM. This effect increases when the data become more and more noisy. Hence, analytical reconstruction is still the method of choice when quantitative analysis is performed despite its reduced resolution and increased noise artifacts. An important reason for the bias is the asymmetry of the Poisson likelihood for low count data. This effect can be reduced by switching to a more symmetric distribution such as a Gaussian distribution. However, this switch is not necessary for data points with higher counts. A reconstruction algorithm with a modified Poisson distribution is developed in chapter 10. Similar as for the local model reconstruction, the Gaussian model will only be used where necessary, for sinogram pixels with low counts, and the Poisson model is still used for pixels with higher counts. Unlike for metal artifact reduction, the model is not chosen based on the values in the image domain but based on the value, i.e. the number of counts, in the sinogram.

Part III

Research articles

Chapter 7

Metal artifact reduction in computed tomography using local models in an image block-iterative scheme

Katrien Van Slambrouck and Johan Nuyts

A modified version of this chapter has been published in: K. Van Slambrouck and J. Nuyts (2012). Metal artifact reduction in computed tomography using local models in an image block-iterative scheme. Medical Physics, 39 (11), 7080-7093.

Purpose In iterative reconstruction, metal artifacts can be reduced by applying more accurate reconstruction models that are usually also more computationally demanding. The hypothesis of this work is that these complex models only need to be applied in the vicinity of the metals and that a less complex model can be used for the remainder of the reconstruction volume.

Method A method is described that automatically divides the reconstruction volume into metal and non-metal regions. The different regions are called patches. A different energy and resolution model can be assigned to each of the patches. The patches containing metals are reconstructed with a fully polychromatic spectral model (IMPACT) and if necessary with an increased resolution model. The patch without metals is reconstructed

with a simple polychromatic model (MLTRC) that only includes the spectral behavior of water attenuation. Comparing the computational complexity of IMPACT and MLTRC gives a ratio of 8:3. The different patches are updated sequentially as in a grouped coordinate algorithm. Two phantoms were simulated and measured: a circular phantom containing small metal cylinders and a body phantom representing a human pelvis with two femoral implants. As a first test, the sequential update of the patches was applied while using the same energy model for all patches. Secondly, the local model approach was applied using MLTRC for non-metal regions and IMPACT for metal regions. The results of different iterative reconstruction schemes are compared to the results of projection completion, another important method for the reduction of metal artifacts.

Results Reconstruction schemes including the sequential update of the patches result in images with less streak artifacts compared to a regular reconstruction. The sequential update of each of the metal regions improves the relative convergence of the metals (edges and attenuation values) against the rest of the image, which leads to an improved artifact reduction. Using the combined IMPACT+MLTRC model results in a similar image quality as using IMPACT everywhere, while providing an important benefit regarding computational complexity. Some streak and shadow artifacts were still present, but all structures present in the phantom could be observed. Projection completion results in reconstructions with less obvious streak and shadow artifacts but tends to deform or erase structures lying close to or in between metallic structures.

Conclusions Metal artifact reduction with iterative reconstruction can be achieved by using complex models only locally without losing image quality. Separately updating metal regions leads to reduced streak artifacts. Structures lying close to or in between metals are often better reconstructed, compared to projection completion results, because all available information is used.

7.1 Introduction

Image quality of computed tomography (CT) reconstructions is often seriously degraded when metals are present in the field of view. Shadows and streaks mask the object reconstruction, which often hinders good interpretation of the results in the regions close to and in between metals.

Reconstruction for CT is typically done by filtered back projection (FBP). FBP is based on an analytical inversion of the Radon transform, yielding good and fast reconstructions in (nearly) ideal conditions. However, these conditions are not met when metals are present. Several metal artifact reduction (MAR)

methods exist. Two important groups are projection completion (PC) and statistical reconstruction with advanced projection model.

In PC, the projections going through metals are considered as corrupted and are replaced by interpolated projections that cause less or no artifacts. PC starts by making an initial reconstruction, typically FBP. The metals in this reconstruction are segmented and forward projected. Hereupon, the metal projections in the measurement are removed and replaced by interpolated projections (e.g. linear interpolation [63], polynomial interpolation [64, 65, 66]). The sinogram with the interpolated projections is then used for a new reconstruction. The metals are copied from the initial reconstruction and pasted into the final reconstruction. This method is fast, because usually based on FBP-like reconstruction and often yields artifact-free images. An important disadvantage of PC is the information loss caused by removing and interpolating projection data. At first, loss of information may be caused by the segmentation: the shape, edge regions, and attenuation values of the metals might be affected. When artifacts are severe, a proper segmentation is not always straightforward. In some applications as e.g. implant follow-up in orthopedic surgery, this information can be important. Moreover, by removing and replacing projections, not only information about the metals, but about all objects contributing to the removed projections is (partly) lost, especially edge information [67]. This causes new artifacts, even far away from the metals. The more projections that need to be replaced, the more information that is lost and the higher the chance for introducing new artifacts. Some more advanced PC methods try to compensate for this information loss. The information is retrieved by using tissue segmentation during interpolation [66, 67, 68, 69], by merging images [70, 71] or projections [72] or by using information out of other non-affected projections [73, 74].

An alternative method is iterative (statistical) reconstruction with an advanced projection model. Most artifacts arise because of incorrect or incomplete modeling of the acquisition. This implies that when a more accurate model is used, artifacts should be substantially reduced. Analytical methods as FBP use an inverse model, which makes adaptations to the model very difficult. Iterative methods make use of a forward model which is more suitable for the introduction of more complex models.

A more elaborated discussion on the origin of (metal) artifacts can be found in Barrett and Neat [54] and De Man et al. [55]. The most important causes of metal artifacts are: beam hardening, (non)-linear partial volume effects, scatter, noise, and motion. Beam hardening is due to the polychromatic nature of X-rays emerging from the tube causing a non-linear behavior between absorber thickness and linear attenuation. Classical algorithms assume a monochromatic X-ray beam and consequently a linear behavior. This leads to cupping artifacts and dark shadows and streaks in between and around the metals. Non-linear partial volume artifacts stem from integrating the attenuation over the finite width of the X-ray beam. Most algorithms implicitly assume this results in

a simple averaging of the attenuation. Actually, it yields the (log-converted) intensity average. Ignoring this effect gives rise to dark and white streaks connecting the edges of highly attenuation objects, like metals. Scatter is secondary radiation coming from electron-photon interactions (Compton scatter) or atom-photon interactions (Rayleigh scatter). This secondary radiation contaminates the measurement. Usually scattered photons are ignored. Artifacts arising here are also cupping, shadows and streaks. Another source of contamination is noise, which has several origins. The quantum nature of the photons is considered as the most important. In the presence of metals this noise can lead to fine streaks connecting the metals. Motion will also lead to severe artifacts in the presence of metals but will not be considered in this study.

Iterative reconstruction with maximization of the Poisson likelihood (maximum likelihood (ML) reconstruction) inherently takes into account a Poisson model for noise in the projection data. In most of current CT systems, energy-integrating detectors are used, which results in a compound Poisson distribution of the signal. However, simple Poisson is usually a sufficient approximation [75]. ML reconstruction usually leads to reduced noise, compared to FBP, in the reconstructed image. The acquisition model can be adapted to include the effects described above that cause severe artifacts in the presence of metals. Several maximum likelihood iterative algorithms exist that have a polychromatic model to reduce beam hardening artifacts [76, 77, 78, 79]. In De Man [80] and Zbijewski and Beekman [81] ML methods reducing (non)-linear partial volume artifacts are presented. Further, if an estimate of the scatter contribution is available, inclusion in the model is straightforward. Using more accurate models strongly reduces artifacts. The main disadvantage of all iterative methods is the increased computation time. Unfortunately, when the complexity of the model increases, the computation time is enormously prolonged and this makes use in daily practice almost impossible.

In this paper we present a framework for iterative maximum likelihood reconstruction that limits the extra time caused by increasing the model complexity. Complex models are only necessary in certain regions of the object, more specifically around metals, and less complex models can be used for the remainder of the object. A patchwork structure is presented which allows local model reconstruction. Different projection models will be used in different parts of the reconstruction volume. Using complex models where needed and simpler models elsewhere should limit the computation time while keeping similar image quality. Secondly, an improved artifact reduction can be achieved by altering the convergence of the different image parts. The results of the local model iterative reconstruction schemes will be compared with FBP-based projection completions methods.

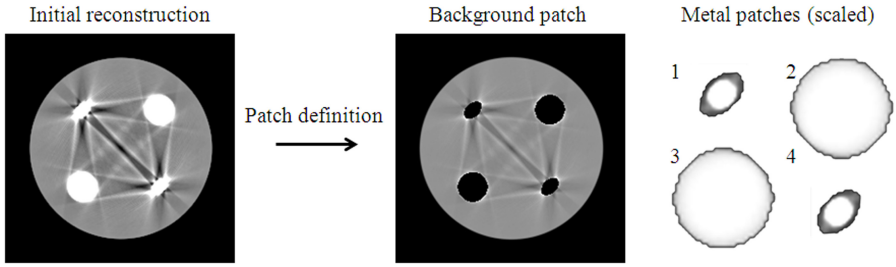


Figure 7.1: **Patch definition** Definition of the different patches based upon segmentation of an initial reconstruction (1 iteration with 116 subsets of IMPACT). Result: a background patch and four metal patches.

7.2 Methods

In this section the different steps in making a local model reconstruction are explained. The first part explains how the different regions are defined. Secondly, the different models that will be used are explained. Finally, the combination of models is described. At the end of this section, the projection completion methods used as comparison, are also briefly described.

7.2.1 Patch definition

The patchwork structure divides the reconstruction volume in non-overlapping regions, *patches*, and assigns particular reconstruction parameters to each of these regions. The patches are defined based upon segmentation of an initial reconstruction. This reconstruction can be made by FBP, but a single iteration of an iterative method is more desirable because then less artifacts are present which simplifies the segmentation. As the shape of the segmentation will not determine the final shape of the object, the segmentation accuracy is not crucial, although it should preferentially contain the whole metal[1]. In this study the metal regions are defined by a thresholding method. These regions are dilated with a few pixels to ensure inclusion of the entire metal, see figure 7.1. Each of the metal regions is considered a patch, the remainder of the object is another patch. An object with four metallic inserts, thus results in five patches. Note that thresholding segments the metals each in a separate patch. This means that bone structure would be assigned to the major patch together with soft tissue and air.

[1] A dedicated set of reconstruction parameters will be defined for regions with metals. If part of the metal is not included in this metal region, a less optimal parameter set will be assigned to it, possibly leading to (extra) artifacts.

7.2.2 Maximum Likelihood Reconstruction and the different projection models

The reconstruction is done by a maximum likelihood method, which is based on the Poisson (log-)likelihood:

$$L = \sum_i \left(y_i \ln \hat{y}_i - \hat{y}_i\right), \tag{7.1}$$

with i the index of the projection lines, y_i the measured transmission scan and $\hat{y}_i$ the estimated transmission scan, computed from the current reconstruction image $\vec{\mu} = \{\mu_j\}$, with μ_j the linear attenuation coefficient in voxel j. A gradient ascent algorithm for optimizing this likelihood, as a function of $\vec{\mu}$, has been proposed before by Nuyts et al. [46] and Fessler et al. [47]. The algorithm can be derived as follows: starting with a quadratic expansion and the condition that $\dfrac{\partial^2 L}{\partial \mu_j \partial \mu_h} \leq 0$ for all j and h (negative definite Hessian):

$$L(\vec{\mu} + \Delta\vec{\mu}) \quad \simeq \quad T_1(\vec{\mu}; \Delta\vec{\mu}) \tag{7.2}$$

$$T_1(\vec{\mu}; \Delta\vec{\mu}) \quad = \quad L(\vec{\mu}) + \sum_j \frac{\partial L}{\partial \mu_j}\Big|_{\vec{\mu}} \Delta\mu_j + \frac{1}{2} \sum_{j,h} \frac{\partial^2 L}{\partial \mu_j \partial \mu_h}\Big|_{\vec{\mu}} \Delta\mu_j \Delta\mu_h \tag{7.3}$$

$$\geq \quad L(\vec{\mu}) + \sum_j \frac{\partial L}{\partial \mu_j}\Big|_{\vec{\mu}} \Delta\mu_j + \frac{1}{2} \sum_{j,h} \frac{\partial^2 L}{\partial \mu_j \partial \mu_h}\Big|_{\vec{\mu}} {\Delta\mu_j}^2. \tag{7.4}$$

The last expression (7.4) follows from $2\Delta\mu_j\Delta\mu_h \leq {\Delta\mu_j}^2 + {\Delta\mu_h}^2$ and the fact that the second derivatives are negative. It is a surrogate function for $T_1(\vec{\mu}; \Delta\vec{\mu})$ (7.3), because it is equal to T_1 at $\Delta\mu = 0$ and below T_1 elsewhere. Maximization of this surrogate leads to the following update for $\Delta\mu_j$:

$$\Delta\mu_j = -\frac{\dfrac{\partial L}{\partial \mu_j}}{\sum_h \dfrac{\partial^2 L}{\partial \mu_j \partial \mu_h}}, \tag{7.5}$$

where all derivatives are evaluated in the current reconstruction $\vec{\mu}$. To calculate the update for μ_j, a model for $\hat{y}_i$ needs to be defined. In this work the energy and resolution properties of the model can be changed. Furthermore a scatter term can be included.

Energy model

Monochromatic model: MLTR The monochromatic model is the one used in MLTR (ML for transmission) [46]:

$$\hat{y}_i = b_i \exp\big(-\sum_j l_{ij}\mu_j\big) + s_i, \tag{7.6}$$

where b_i is the blank value for projection line i in the presence of (bowtie) filters, l_{ij} is the intersection length of projection line i and pixel j, and s_i is the scatter term for projection ray i. This yields the following update step:

$$\Delta\mu_j = \frac{\sum_i l_{ij} \frac{\hat{y}_i - s_i}{\hat{y}_i}(\hat{y}_i - y_i)}{\sum_i l_{ij} \sum_h l_{ih}(\hat{y}_i - s_i)(1 - \frac{y_i s_i}{\hat{y}_i^2})}. \tag{7.7}$$

A complete update step requires one projection and two backprojections.

Polychromatic water model: MLTRC MLTRC is a polychromatic extension of MLTR which takes into account the polychromatic behavior of the X-rays going through water. The projection model becomes:

$$\hat{y}_i = \sum_k \hat{y}_{ik} + s_i = \sum_k b_{ik} \exp\big(-CP_k \sum_j l_{ij}\mu_{j,\text{ref}}\big) + s_i, \tag{7.8}$$

$$C_k = \frac{\mu_k^{\text{water}}}{\mu_{\text{ref}}^{\text{water}}}, \tag{7.9}$$

where k is the energy bin and b_{ik} the blank value for projection line i at energy k in the presence of (bowtie) filters, $\mu_{\text{ref}}^{\text{water}}$ is the linear attenuation coefficient of water at a chosen reference energy and $\mu_{j,\text{ref}}$ is the reconstructed linear attenuation coefficient in pixel j at the same reference energy. The MLTRC update becomes:

$$\Delta\mu_j = \frac{\sum_i l_{ij} Y_i^P (1 - \frac{y_i}{\hat{y}_i})}{\sum_i l_{ij} \sum_h l_{ih} \Big[(1 - \frac{y_i}{\hat{y}_i}) Y_i^{PP} + \frac{y_i}{\hat{y}_i^2}(Y_i^P)^2\Big]} \tag{7.10}$$

with

$$Y_i^P = \sum_k C_k \hat{y}_{ik} \tag{7.11}$$

$$Y_i^{PP} = \sum_k (C_k)^2 \hat{y}_{ik}. \tag{7.12}$$

Usually a limited set of energy bins, e.g. 10, is enough to represent the spectrum. Note that the complexity (in terms of (back)projections) of this model is the same as that of regular MLTR.

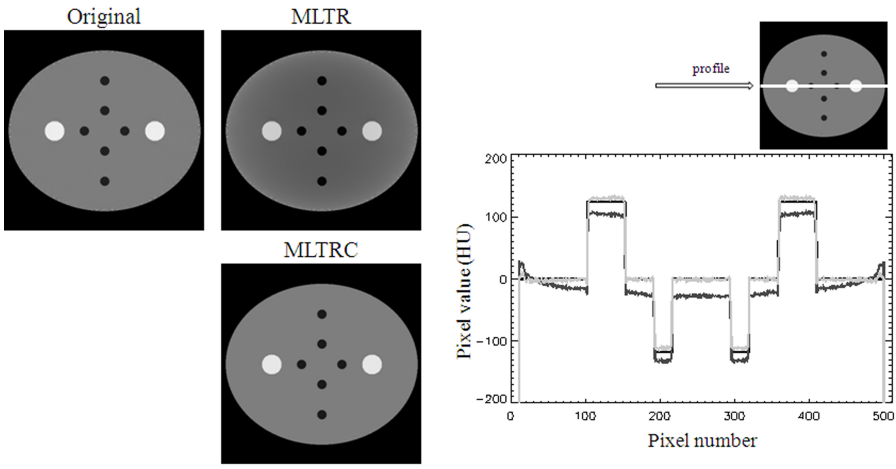


Figure 7.2: **MLTR and MLTRC** Reconstruction of (computer) water phantom with fat (darker) and PMMA (brighter) inserts, the true 'original' image is shown at the left. Reconstructions are performed for 10 iterations with 116 subsets with MLTR and MLTRC. (Window = 300 HU, Center = 0 HU). The profile through the reconstruction is compared with the true profile, where black = original, dark gray = MLTR and light gray = MLTRC.

The effect of using MLTRC compared to MLTR is shown in the next example. A simple polychromatic simulation (spectrum of 120 kV) is performed using a phantom of 50 cm with several 'fat' and PMMA (polymethyl methacrylate) details. The results for a reconstruction with MLTR and MLTRC are shown in figure 7.2. For MLTR the typical cupping is present, which is eliminated using MLTRC. Materials other than water are only partly corrected. The more the spectral properties of a material differ from those of water, the less effective this correction will be. In fact, the effect of using MLTRC is similar to using a (polynomial) water pre-correction on raw data [82]. However, with MLTRC projections for different energies can be calculated which will be necessary when combining it with fully polychromatic models.

Full polychromatic model: IMPACT The IMPACT (iterative maximum likelihood polychromatic algorithm for CT)[76] projection model uses a fully polychromatic projection estimate:

$$\begin{aligned} \hat{y}_i &= \sum_k \hat{y}_{ik} + s_i \\ &= \sum_k b_{ik} \exp\left(-\Theta_k \sum_j l_{ij}\theta(\mu_{j,\text{ref}}) - \Phi_k \sum_j l_{ij}\phi(\mu_{j,\text{ref}}) \right) + s_i \end{aligned} \quad (7.13)$$

The energy dependent linear attenuation is written as a linear combination of the Compton and the photo-electric component, with Θ_k and Φ_k the energy

dependence of respectively Compton scattering and photo-electric effect, and $\vec{\theta}$ and $\vec{\phi}$ the material dependence. We assume that $\vec{\theta}$ and $\vec{\phi}$ are unambiguously determined by the attenuation at a chosen reference energy: $\theta_j = \theta(\mu_{j,\text{ref}})$ and $\phi_j = \phi(\mu_{j,\text{ref}})$. To obtain this, we assume that a list of possible materials is available, which has been sorted according to increasing effective attenuation. We further assume that attenuation in every pixel can be well modeled as a mixture of two adjacent materials in this list. The update becomes:

$$\Delta\mu_j = \frac{\phi'_j \sum_i l_{ij} e_i Y_i^{\Phi} + \theta'_j \sum_i l_{ij} e_i Y_i^{\Theta}}{\phi'_j \sum_i l_{ij} M_i + \theta'_j \sum_i l_{ij} N_i} \tag{7.14}$$

with

$$\begin{aligned}
\phi'_j &= \frac{\partial \phi}{\partial \mu_j} & \theta'_j &= \frac{\partial \theta}{\partial \mu_j} \\
e_i &= 1 - \frac{y_i}{\hat{y}_i} & & \\
Y_i^{\Phi} &= \sum_k \Phi_k \hat{y}_{ik} & Y_i^{\Theta} &= \sum_k \Theta_k \hat{y}_{ik} \\
Y_i^{\Phi\Phi} &= \sum_k \Phi_k^2 \hat{y}_{ik} & Y_i^{\Theta\Theta} &= \sum_k \Theta_k^2 \hat{y}_{ik} \\
Y_i^{\Phi\Theta} &= \sum_k \Phi_k \Theta_k \hat{y}_{ik} & &
\end{aligned}$$

$$\begin{aligned}
M_i &= \Big(\sum_h l_{ih} \phi'_h\Big)\Big(Y_i^{\Phi\Phi} e_i + \frac{y_i (Y_i^{\Phi})^2}{\hat{y}_i^2}\Big) \\
&+ \Big(\sum_h l_{ih} \theta'_h\Big)\Big(Y_i^{\Phi\Theta} e_i + \frac{y_i Y_i^{\Phi} Y_i^{\Theta}}{\hat{y}_i^2}\Big) \\
N_i &= \Big(\sum_h l_{ih} \phi'_h\Big)\Big(Y_i^{\Phi\Theta} e_i + \frac{y_i Y_i^{\Phi} Y_i^{\Theta}}{\hat{y}_i^2}\Big) \\
&+ \Big(\sum_h l_{ih} \theta'_h\Big)\Big(Y_i^{\Theta\Theta} e_i + \frac{y_i (Y_i^{\Theta})^2}{\hat{y}_i^2}\Big)
\end{aligned}$$

The complexity for each update is 8 (back)projections. As for MLTRC, a limited number of bins gives a sufficient representation of the spectrum.

Note that the condition that the Hessian of the likelihood should be negative definite, is only proved for MLTR without or with small scatter (if $\hat{y}_i^2 \geq y_i s_i$), in other cases (substantial scatter, more complex model, ...) it cannot be proved. In our experience, this never caused convergence problems.

Resolution

The resolution is modeled by oversampling the detector elements into Q equally sized subsamples and decreasing the pixel size in the reconstruction image ($M = J * f$, the new number of pixels with f the (integer) size factor and J the original number of pixels).

$$\hat{y}_i = b_i \sum_{q=1}^{Q} \exp\big(-\sum_{m=1}^{M} l_{iqm}\mu_m\big) \tag{7.15}$$

This resolution model does not take into account crosstalk and gantry motion.

Scatter

The influence of scatter is not always negligible. Including scatter, the projection model can be represented by $\hat{y}_i^S = \hat{y}_i + Sp_i$, with p_i a known profile and S the profile factor. (This profile can be calculated e.g. by Monte Carlo simulations.) The profile factor S will be updated in each iteration and is used to scale the calculated or estimated profile p_i. The update strategy is the same as in (7.2), which gives:

$$\hat{y}_i^S = \hat{y}_i + Sp_i, \quad L = \sum_i \big(y_i \ln \hat{y}_i^S - \hat{y}_i^S\big) \tag{7.16}$$

$$\Delta S = -\frac{\frac{\partial L}{\partial S}}{\frac{\partial^2 L}{\partial S^2}} \tag{7.17}$$

$$= \frac{\sum_i (\frac{y_i}{\hat{y}_i^S} - 1)p_i}{\sum_i \frac{y_i}{(\hat{y}_i^S)^2} p_i^2}. \tag{7.18}$$

In this work, no scatter profile is available, a constant scatter profile, $p_i = 1$, will be used.

7.2.3 Local model reconstruction

For the local model reconstruction, a proper projection model $\hat{y}_{i,p}$ will be assigned to each patch p. As explained before, the models can differ in their energy and resolution properties. The combined projection estimate becomes:

$$\hat{y}_i = \sum_k b_{ik} \prod_p z_{ik,p} + Sp_i \text{ with } \quad z_{ik,p} = \frac{\hat{y}_{ik,p}}{b_{ik}} \tag{7.19}$$

Each $z_{ik,p}$ is calculated separately for each patch using the model assigned to that particular patch. During reconstruction, a patch will be treated as a

group of pixels, or image-block, in a grouped coordinate algorithm [47, 83]. This means that the patches are updated one by one, for each patch update only the voxels belonging to that particular patch are updated. Consequently, the sum over h in the denominator of eq. (7.5) is also limited to these voxels. As will be discussed later, this has a significant effect on local convergence.

Note that the summation over the energies, k, happens after combining the projections of the different patches. Therefore it is difficult to combine true monochromatic and polychromatic energy models. IMPACT and MLTRC are both polychromatic algorithms, they both produce a projection for each energy bin k and can therefore be combined in reconstructions where patches have different energy models. As mentioned before, using MLTRC instead of MLTR does not increase the computational complexity, moreover, cupping artifacts are reduced.

Updating the scatter term is a sinogram procedure which is applied after combining the contributions from different patches and different energy bins. Therefore eq. (7.18) will not be affected by the introduction of patches.

All iterative reconstructions are initialized with a non-zero image. One could use the FBP reconstruction, but in our experience artifacts present in an initial reconstruction are often hard to remove. We prefer using a contour image. This is a contour of the object filled with the attenuation value of water. This contour image can of course be based on the FBP reconstruction. Here, the contour image will be based on the initial reconstruction which is also used to define the patches.

7.2.4 Projection Completion

The results obtained with the iterative methods will be compared to three projection completion methods. These methods start by making an initial FBP-reconstruction. Based on this image a segmentation of the metals, by k-means clustering, is performed. The metal projections are determined by a forward projection of these segmented metals and removed from the original measurement. These projections are replaced by artificial projections. The new data set is then used to generate a new reconstruction with substantial artifact reduction.

The first PC method that will be used, is the classical PC as proposed by Kalender et al. [63] where the metal projections are replaced by new ones based on linear interpolation. New streaks caused by removing information can (partly) by avoided by including a normalization step as in Müller et al. [67] and Meyer et al. (NMAR) [68]. In this second approach (PC-NMAR), the measured sinogram is normalized, the interpolation step is applied and finally the interpolated sinogram is denormalized. For normalization a prior image is constructed. A segmentation of the original reconstruction defines air, tissue and bone regions. The prior image is built by assigning a particular

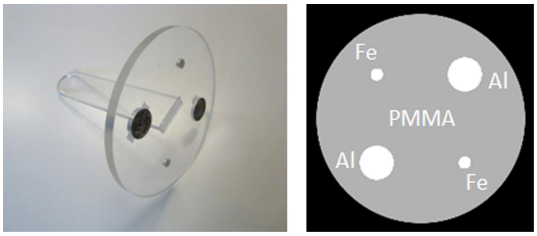


Figure 7.3: **Circular phantom** The circular phantom on the left and its computer representation on the right. This phantom is a PMMA disk with diameter 19 cm and contains two cylindrical aluminum inserts of 3 cm diameter and two cylindrical iron inserts of 1 cm diameter.

attenuation value to the different regions. The air regions become $\mu_j = 0$ (-1000 HU), the tissue region receives the typical attenuation value obtained at the particular energy settings of the scan (in HU this is around 0 HU). The bone regions retain their original values and the metal regions are set to 0 HU. The normalization is applied by dividing the measured sinogram by the projection of the prior image. Information about the intersection lengths for the different tissues is included, leading to reconstructions with less artifacts compared to regular PC. As mentioned in Meyer et al. [68], when artifacts are severe, the segmentation process can be difficult and this can affect the prior image. Therefore, the prior image will be based on the result of a regular PC reconstruction. Note that even when using the normalized algorithm, artifacts can still be present because some information about the object is irreversibly lost during the replacement of the projections, e.g. edge information between different soft-tissue structures.

The third method, PC-FSMAR [70], is an extension of the PC-NMAR method. The idea of this method is to retrieve edge information present in the FBP reconstruction by combining high frequencies of both FBP and PC-NMAR reconstructions. PC-FSMAR results in the following weighted sum:

$$\mu_j^{\text{FSMAR}} = \mu_j^{\text{NMARLow}} + w_j \mu_j^{\text{FBPHigh}} + (1 - w_j)\mu_j^{\text{NMARHigh}} \tag{7.20}$$

with μ_j^{NMARLow} the low frequencies of the PC-NMAR reconstruction, and μ_j^{FBPHigh} and μ_j^{NMARHigh} the high frequencies of the FBP and PC-NMAR reconstruction. The normalized weights w_j decrease with increasing distance of pixel j to the metal. An advantage of PC-FSMAR over other more advanced methods is the clear and rather straightforward implementation.

7.3 Simulations and Measurements

To evaluate the proposed techniques two phantoms were used. The first phan-

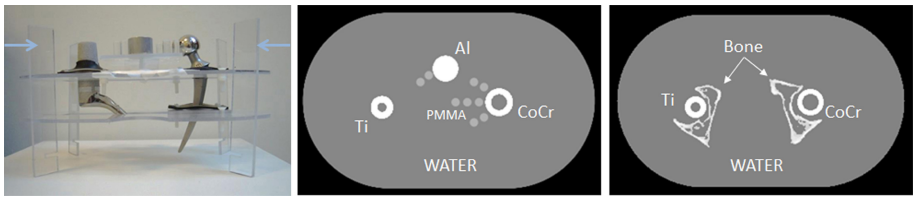


Figure 7.4: **Body phantom** On the left the scaffold for the body phantom is shown. This scaffold is put into water for scanning. The slice which is scanned is marked with arrows. The computer representation of the slice is shown in the middle. This phantom contains two implants (titanium-based and cobalt-chromium-based), an aluminum insert with diameter 3 cm and several PMMA inserts of diameter 1 cm. On the right, the variation of the computer phantom with a bone structure included, is also shown.

tom is a circular phantom made of PMMA with diameter 19 cm. It contains two aluminum inserts of 3 cm diameter and two iron inserts of 1 cm diameter as shown in figure 7.3. The phantom structure is identical for the simulation and the measurement. The second phantom represents a part of a human pelvis and hips with two femoral implants. The phantom contains a titanium based implant and a cobalt-chromium based implant. An extra aluminum insert with diameter of 3 cm is added. Further some extra PMMA cylinders of 1 cm diameter are placed around the metals. This phantom was also simulated. Besides, a second composition of the body phantom was used for simulations. Here, the two implants remain and are now surrounded by a bone structure which is based on the bone structure in a human pelvis. The real and computer phantoms are depicted in figure 7.4.

7.3.1 Simulations

The simulator is based on the CT-simulator of De Man et al. [55] extended with a distance driven projector [84]. The simulations were two-dimensional and included Poisson noise, a polychromatic spectrum (50 bins) and oversampling of the detector elements (10 times), views (5 times), source (5 times, width 0.1 cm) and image pixels of 0.01 cm. The geometry and spectrum properties were based on a Siemens Sensation 16 system at 120kV. This system has a curved detector with 672 detector elements and 1160 projection views per rotation. The spectrum was determined using Spektr software [85]. The simulation also took into account afterglow and crosstalk between the detectors.

7.3.2 Measurements

Both phantoms were scanned on a Siemens Sensation 16 (part of a Biograph 16 PET/CT system). A circular scan at 120 kV and 300 mA was performed

using collimation 2 x 1.00 mm and a rotation time of 0.5 s. Only one slice of 1.00 mm will be used for a two-dimensional reconstruction.

Data obtained from clinical scanners are typically corrected for beam hardening and other effects. The corrections are often non-linear and can interfere with our proposed models. Siemens assisted to obtain raw data without beam hardening corrections.

7.3.3 Reconstruction

All data sets are reconstructed with different reconstruction schemes:

1. Four FBP-based reconstructions are performed: regular FBP, PC, PC-NMAR and PC-FSMAR. The FBP reconstructions are performed using a typical ramp-filter.
2. A regular MLTR, MLTRC, and IMPACT reconstruction is done.
3. Different reconstructions based on a patchwork structure are performed:
   (a) **Patched MLTR, Patched MLTRC, or Patched IMPACT:** Different patches are defined by the patched structure but the same energy model (MLTR, MLTRC or IMPACT) is used in all patches.
   (b) **MLTRC + IMPACT:** The energy model for the metal patches is IMPACT, the remaining patch without metals uses the MLTRC energy model.
   (c) **Resolution:** For the patched methods, a second patched reconstruction has been done with increased resolution in the patches with the highest attenuating materials (iron and the implants). This means that three times more pixels were used and a threefold detector oversampling was applied for the projection of the particular patch.

Adaptive filtering of the projections [86] is applied to all FBP-based reconstructions to reduce noise. The iterative reconstructions schemes consist of 20 iterations with 116 subsets. For IMPACT the list of materials included all materials present in the object: for the circular phantom: PMMA, aluminum and iron, for the body phantom: water, aluminum, titanium and cobalt-chrome, for the body phantom with bone: water, bone, titanium and cobalt-chrome.

7.4 Results

7.4.1 Simulations

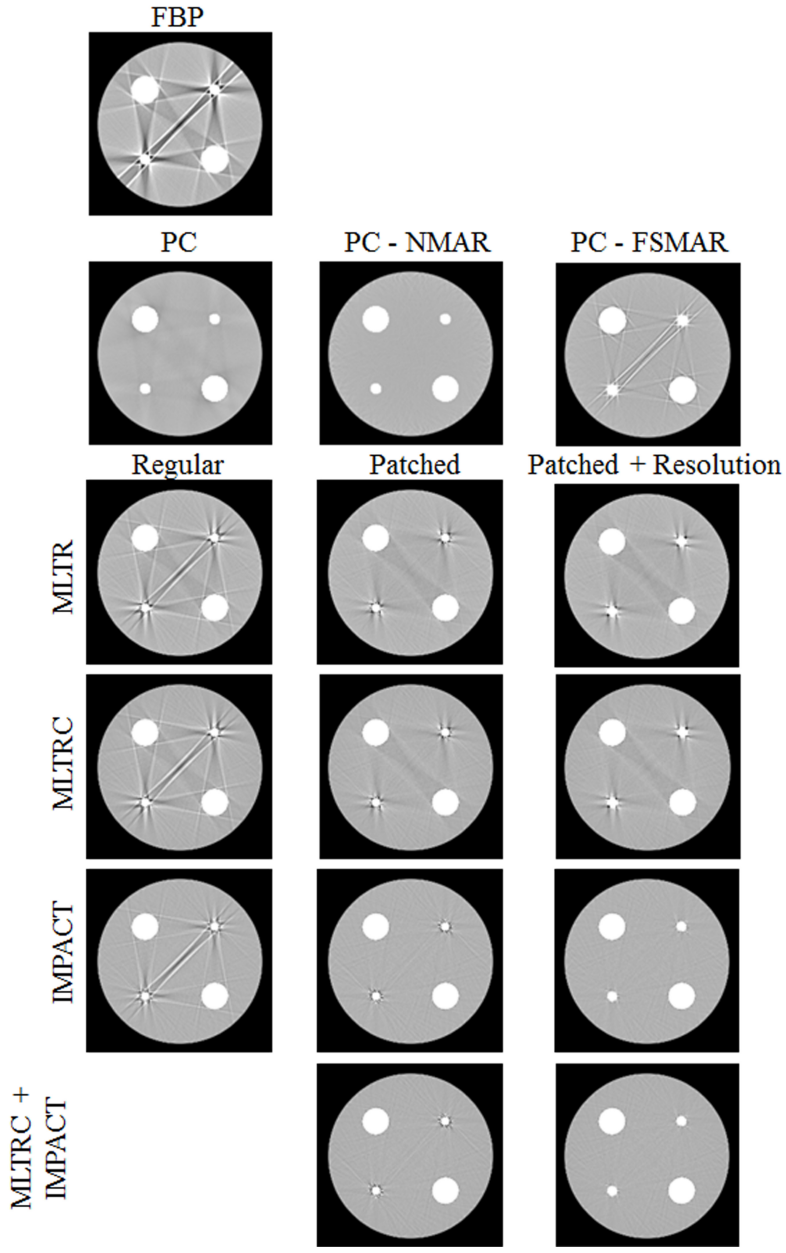


Figure 7.5: **Simulation, circular phantom** Reconstruction of the circular (computer) phantom with FBP, PC, PC-NMAR, PC-FSMAR, MLTR, MLTRC, IMPACT, Patched MLTR, Patched MLTRC, Patched IMPACT and MLTRC+IMPACT. When using patches a reconstruction with and without resolution modeling in the Fe-patches is shown. (Window = 750 HU, Center = 0 HU.)

Circular computer phantom The circular (computer) phantom is reconstructed with a pixel size of 0.1 cm. The results of the different reconstruction schemes are shown in figure 7.5.

Artifacts are reduced for MLTR, MLTRC, and IMPACT compared to FBP. Differences between MLTR, MLTRC, and IMPACT are small. In a narrow window, MLTRC and IMPACT show reduced cupping compared to MLTR, and the IMPACT reconstruction has no dark streak between the aluminum inserts. The difference between the reconstruction methods becomes more clear when using a patched update, with IMPACT having no streaks in between any of the metallic inserts. Introducing a higher resolution for the iron patches reduce the remaining streak artifacts. Note that MLTRC+IMPACT, using MLTRC for the non-metal patches, gives a very similar result as using IMPACT for the whole image.

The results for PC and PC-NMAR are almost perfect. However, this is an almost ideal situation: a symmetrical homogeneous circular phantom. PC-NMAR is performing better than regular PC. PC-FSMAR reintroduces some of the FBP artifacts in the PC reconstruction.

Body computer phantom

Body phantom with PMMA details The pixel size for this reconstruction was 0.1 cm. The overall results for the body phantom (fig. 7.6) are similar to the results of the circular phantom. The cobalt-chromium femoral prosthesis is attenuating more and consequently suffers more from beam hardening, making MLTR worse compared to MLTRC and IMPACT. Also here, the patched reconstruction methods perform better than the regular reconstruction methods. Using MLTRC+IMPACT results in the same image quality as Patched IMPACT. The influence of increased resolution by increased sampling almost negligible.

PC, and especially PC-NMAR, give a smooth, seemingly artifact-free result. However, the PMMA details are deformed or have disappeared. These details are partly recovered by PC-FSMAR. However, together with high frequency information from the object, also some of the streaks and shadows from the FBP image are reintroduced and cover some PMMA elements.

Body phantom with bone structure The reconstruction pixel size is 0.1 cm. Figure 7.7 shows the results for FBP, PC-NMAR, PC-FSMAR, IMPACT, Patched IMPACT and MLTRC+IMPACT. The results for the other methods are not shown, but they are analogous to previous examples. The edges of the bone structure are not sufficiently reconstructed with PC or PC-NMAR. PC-FSMAR regains some contrast, but also here FBP artifacts are retrieved together with edge information. The iterative methods are reconstructing the

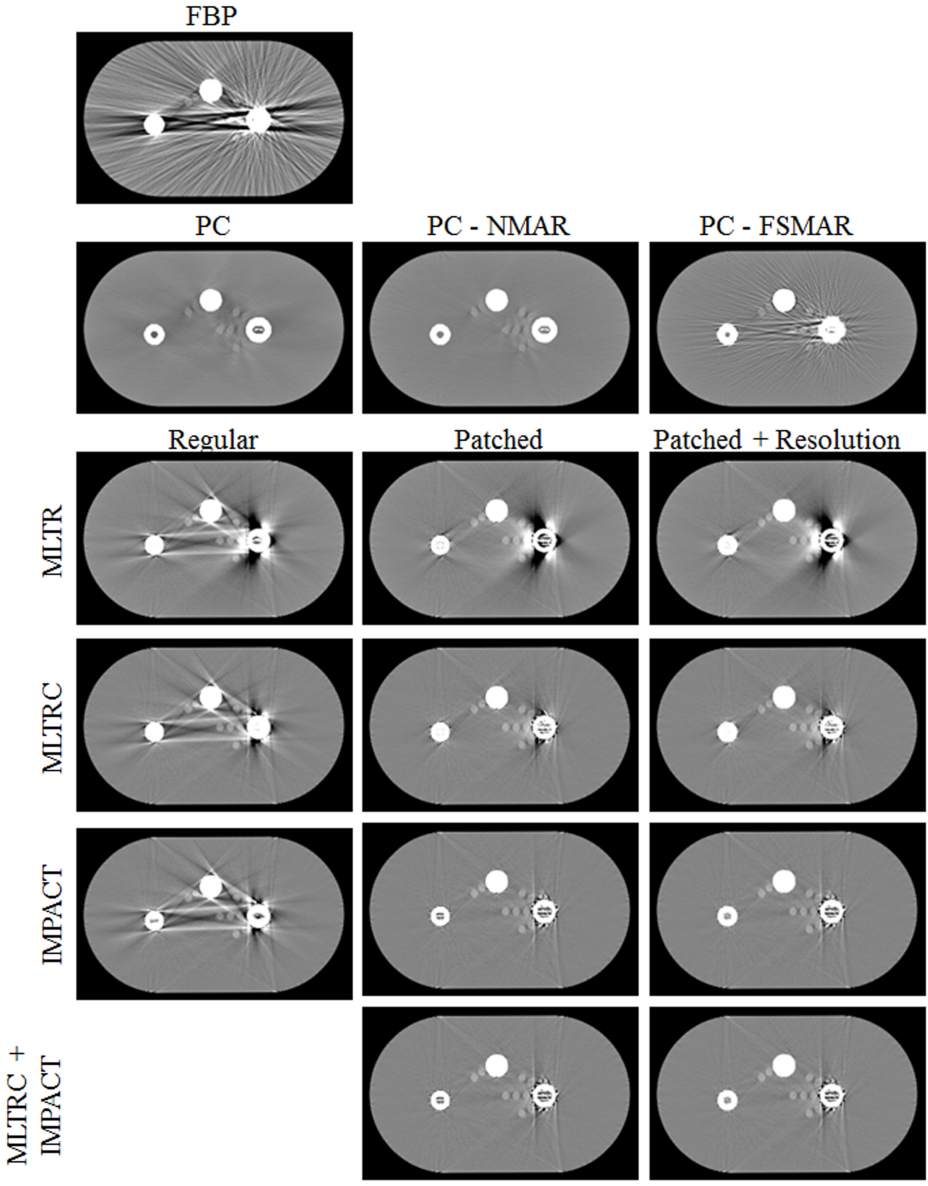


Figure 7.6: **Simulation, body phantom with PMMA details** Reconstruction of the body (computer) phantom with PMMA elements. The methods are FBP, PC, PC-NMAR, PC-FSMAR, MLTR, MLTRC, IMPACT, Patched MLTR, Patched MLTRC, Patched IMPACT and MLTRC+IMPACT. A reconstruction with and without resolution modeling in the implant-patches is shown. (Window = 750 HU, Center = 0 HU.)

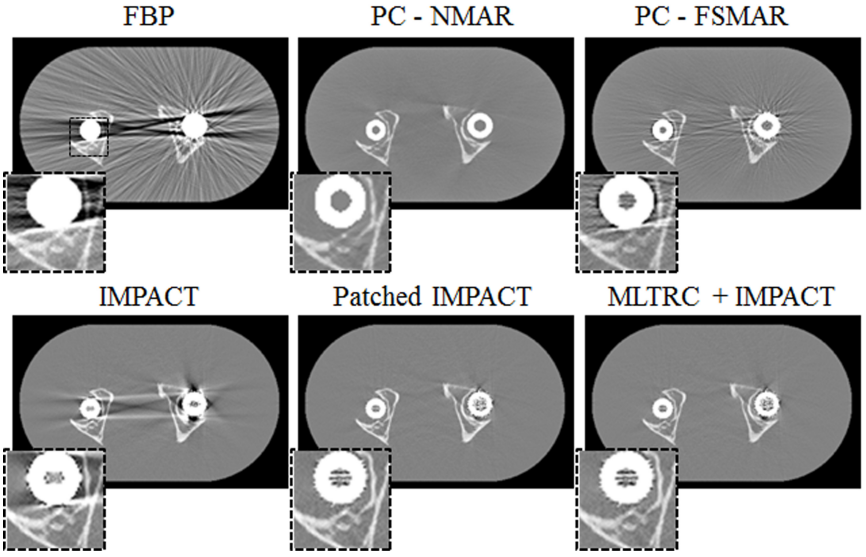


Figure 7.7: **Simulation, body phantom with bone** Reconstruction of the body (computer) phantom with bone. The methods are FBP, PC-NMAR, PC-FSMAR, IMPACT, Patched IMPACT and MLTRC+IMPACT (no resolution modeling). The region around the left implant is magnified in the inserts. (Window = 750 HU, Center = 0 HU.)

bone structure well. Note that bone is segmented into the major patch which is reconstructed with MLTRC in the combined MLTRC+IMPACT reconstruction, showing that MLTRC suffices for the reconstruction of these bone regions.

7.4.2 Measurements

Circular phantom The measurement was reconstructed on a grid with isotropic pixels of 0.098 cm. The results for the measurement of the circular phantom (fig. 7.8) are similar to those for the simulated data set. However, no clear improvement is seen for *Patched + Resolution* compared to *Patched* reconstructions. The introduction of patches in Patched IMPACT leads to a substantial artifact reduction compared to regular IMPACT. Replacing IMPACT by MLTRC in the non-metal patch does not lead to image degradation. The results for projection completion are equivalent to the simulations.

Body phantom The measurement of the body phantom was reconstructed on a grid with isotropic pixels of 0.098 cm. The results are shown in figure 7.9. They are comparable to those of the simulation. With Patched IMPACT and MLTRC + IMPACT strongly reducing the artifacts and only little influence of the use of an increased resolution for the implant patches. The details around the implants are reconstructed well. The results for PC and PC-NMAR have only mild streaks or shadows with PC-NMAR outperforming PC. The PMMA details are deformed for PC and PC-NMAR and, as in the simulation, partly retrieved in PC-FSMAR.

7.5 Discussion

Two important methods for metal artifact reduction are projection completion and iterative reconstruction with advanced projection model. Where projection completion is fast, the influence of replacing data always includes a risk. Iterative reconstruction uses all available information but is extremely slow when using a more complex model. In this study we presented a local model reconstruction scheme where complex models are only applied to a limited area of the reconstruction volume. The framework for our local model scheme is the patchwork structure and projector.

The reconstruction volume was divided in several regions or patches. For each of these patches a different energy and resolution model could be defined. This is very similar to the region of interest reconstruction proposed by Hamelin et al. [87, 88] and Yu et al. [89]. In these papers, the approach was used to obtain a high resolution reconstruction in a small region of interest, in combination with a coarser (and faster) reconstruction of the surrounding background. Here,

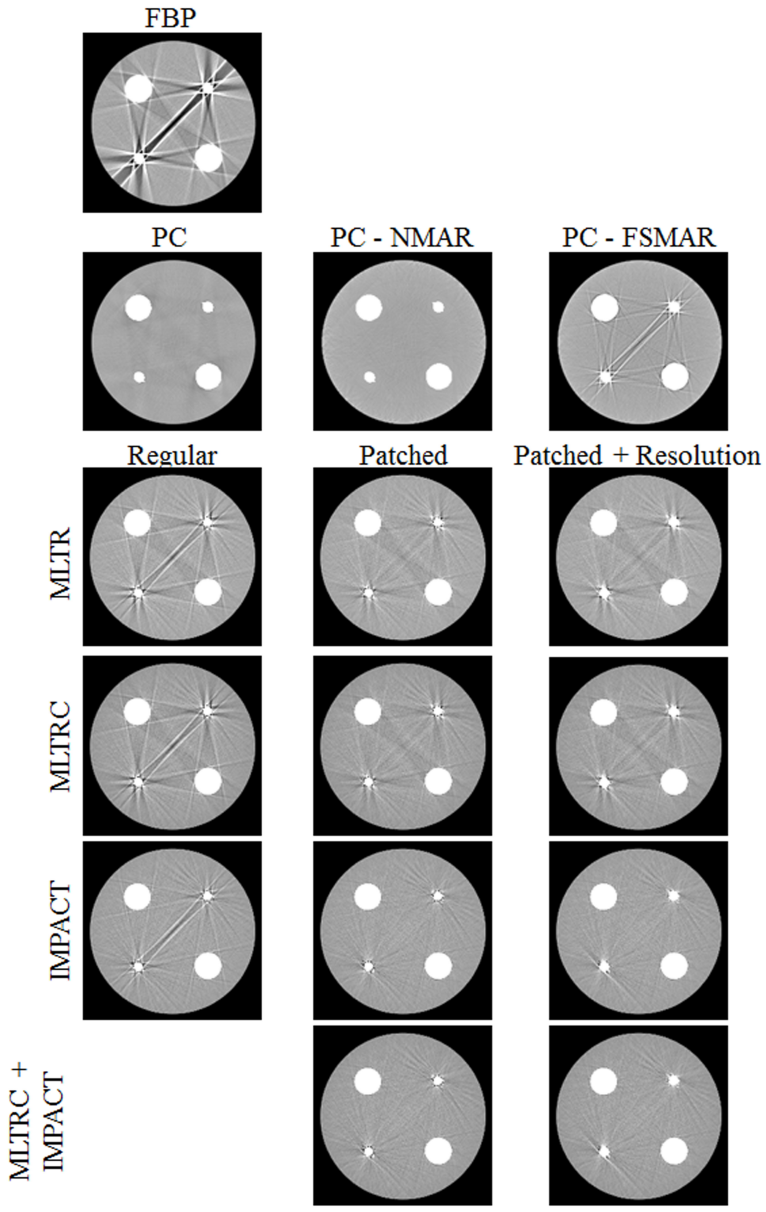


Figure 7.8: **Measurement, circular phantom** Reconstruction of the circular phantom with FBP, PC, PC-NMAR, PC-FSMAR, MLTR, MLTRC, IMPACT, Patched MLTR, Patched MLTRC, Patched IMPACT and MLTRC+IMPACT. When using patches a reconstruction with and without resolution modeling in the Fe-patches is show. (Window = 750 HU, Center = 0 HU.)

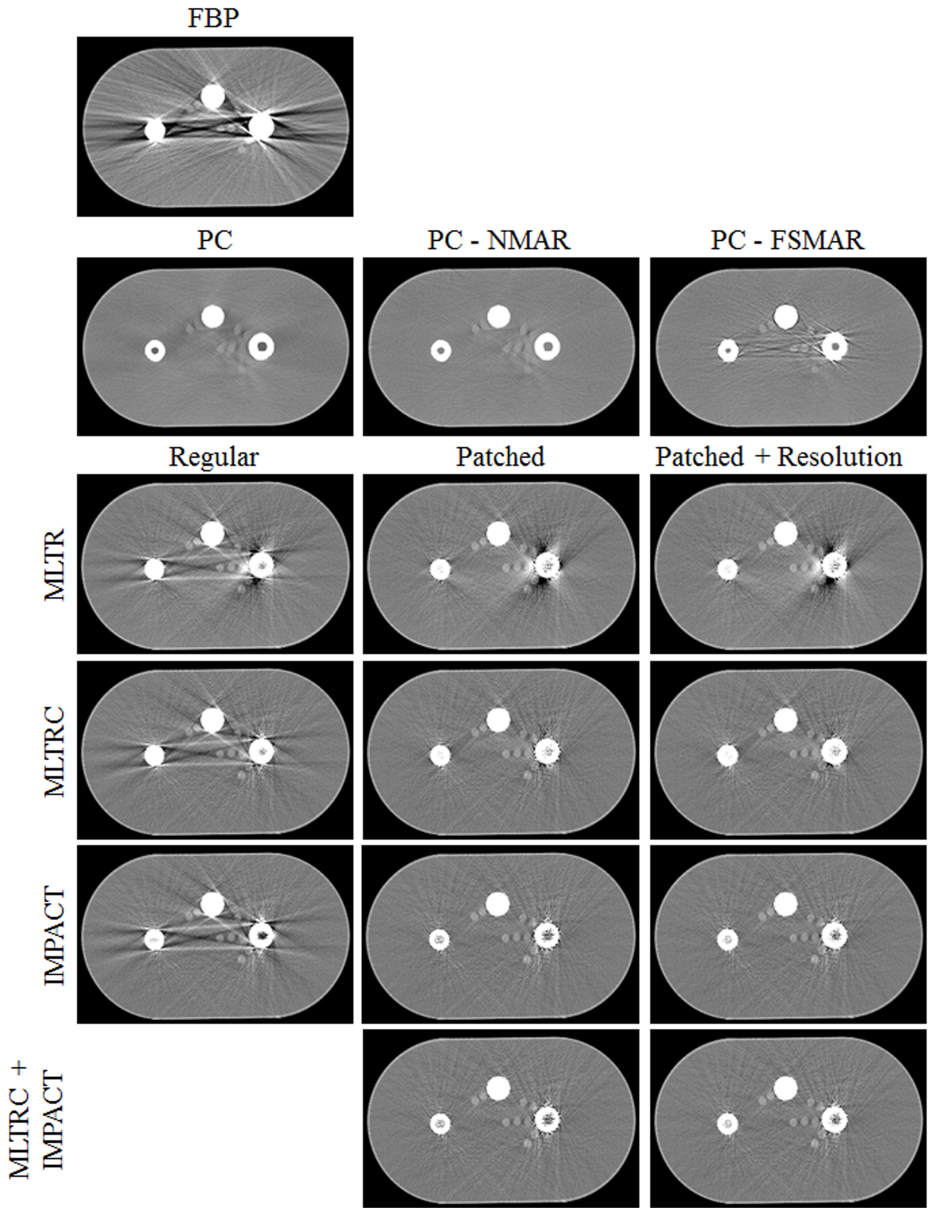


Figure 7.9: **Measurement, body phantom** Reconstruction of the body phantom with FBP, PC, PC-NMAR, PC-FSMAR, MLTR, MLTRC, IMPACT, Patched MLTR, Patched MLTRC, Patched IMPACT and MLTRC+IMPACT. A reconstruction with and without resolution modeling in the implant-patches is shown. (Window = 750 HU, Center = 0 HU.)

this approach was extended to implement a locally improved energy model and improved sampling of the detector elements. The patches are updated sequentially, each patch was considered as a group of pixels in a grouped coordinate algorithm. In a first test, the patched reconstruction was applied. Here, the volume is divided into the different patches, but the same model was used for all patches. In all examples, the application of the sequential patch update already reduced the artifacts, especially the streaks connecting the metals. The origin of this result is an improved convergence for the metal patches. Sequentially updating groups of pixels is known to improve convergence [47, 90]. The denominator of the update steps will be smaller when the area of the updated patch is smaller. For instance, in the MLTR update step (7.7), the sum $\sum_h l_{ih}$ in the denominator is proportional to the width of the updated patch. A smaller denominator makes the update larger. By giving the metal patches a larger update compared to the rest of the image, the relative convergence is altered which results in higher attenuation values and sharper edges in an earlier stage of the reconstruction process, resulting in less artifacts. A simulation experiment not shown here confirmed that for far more iterations (about 200 iterations x 116 subsets) the non-patched and patched reconstruction converge to a very similar solution. A similar reduction of streaks is obtained but both reconstructions suffer from fine streak artifacts caused by the well known noise characteristics. In Yu et al [91] is also shown that altering the convergence of different parts of the image, such as edges, can result in an improved image quality.

We hypothesized that the non-metal regions could be reconstructed with a less complex model. Therefore, we combined the energy model of MLTRC for the non-metal patch with IMPACT for the metal patches. In all examples this resulted in a very similar artifact reduction compared to Patched IMPACT. The same image quality was achieved while reducing the complexity of the largest part of the reconstruction volume from 8 (back)projections for IMPACT to 3 (back)projections for MLTRC. Artifacts originating from non-linear partial volume effects could be reduced by increasing the resolution. However, for all phantoms, the influence of an increased resolution was rather low. It is questionable whether this small improvement is worth the extra time.

As a comparison, three projection completion methods were applied to the same data sets. Regular PC and PC-NMAR reconstructions have less obvious artifacts compared to the iterative methods which all suffered from remaining (but reduced) shadow and streak artifacts. However, structures around the metals are often poorly reconstructed when using PC or PC-NMAR. PC-FSMAR tries to solve this problem by reintroducing some high frequency information from the initial FBP reconstruction. Part of the edge information is indeed retrieved, but in the phantom with the bone structure, the bone is still less sharp compared to the iterative reconstructions. Unfortunately, PC-FSMAR is not free of streaks or shadows like the other projection completion methods because some artifacts are reintroduced by using the high frequency information of the FBP reconstruction. The loss of contrast and edges, especially close to the implants,

can be important for some applications, like implant follow-up in orthopedic surgery. Other advanced projection completion methods [66, 69, 71, 72, 73, 74] could be more successful in retrieving contrast, although by replacing projection values some contrast information might still be lost.

The reconstructions with Patched IMPACT or MLTRC+IMPACT still contain artifacts: fine streaks around the metals, small brighter dots, darker regions in between the metals. The origin of fine streaks and dots is probably noise. Further, although iterative reconstruction is using all available information, some important information can be lost during the measurement itself, as for instance in case of photon starvation. The remaining artifacts could be reduced by introducing Markov priors in the reconstruction algorithm [92]. Notice that when using priors, the properties of the prior could also be changed depending on the particular patch. A second origin of artifacts might be scatter. We believe the scatter in the examples shown is relatively low and it is sufficient to use a simple scatter model, a constant factor. Depending on the system geometry and the object structure and attenuation, a more accurate (Monte Carlo simulated) scatter profile could be necessary to sufficiently reduce image artifacts.

In this study we chose to use thresholding to select the metals in the initial reconstruction. In projection completion, mistakes in the metal segmentation will affect the final reconstruction. In contrast, the segmentation used for the patch definition does not determine the shape or voxel values of the metals. The advantage of using a more sophisticated segmentation method will thus be small. Selecting a region larger than the metal will not affect the final artifact reduction. Selecting a region smaller than the metal, which can be avoided in many cases, could result in a less efficient artifact reduction because the model will not be applied to the entire metal region. To avoid selecting regions that are smaller than the metal, the metal regions were dilated after thresholding.

In this study we used local models for reconstruction with artifact reduction. The patchwork structure as presented here can also be used for other than MAR applications, e.g. reconstruction with varying resolution over the image [87, 88, 89, 93, 94, 95], and for other modalities than CT, e.g. PET or SPECT.

7.6 Conclusion

Projection completion methods are fast and efficient methods for metal artifact reduction but they tend to deform structures lying in the interpolated projection rays. Iterative reconstructions have some remaining streaks and shadows, although strongly reduced, but by using all available information a better reconstruction of structures close to and in between metals is obtained. The increased computation time caused by using more complex models can be limited by applying the more sophisticated model only locally in and around

the metals. This local model reconstruction does not result in loss of image quality. Moreover, sequentially updating the different image parts leads to an improved artifact reduction.

Acknowledgments

The authors would like to thank dr. Karl Stierstorfer from Siemens AG, Healthcare Sector, Forchheim, Germany, for his valuable suggestions and help with the data format and preprocessing and prof. dr. Michiel Mulier for the femoral implants and for the interesting discussion on the problems with implant imaging for orthopedic surgery.

This work is supported by the SBO-project QUANTIVIAM (060819) of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

Chapter 8

Reconstruction scheme for accelerated maximum likelihood reconstruction: the patchwork structure

Katrien Van Slambrouck and Johan Nuyts

A slightly modified version of this chapter has been accepted with minor revision for publication in IEEE Transactions on Nuclear Science.

Abstract Convergence of iterative algorithms can be improved by updating groups of voxels sequentially rather than updating the whole image simultaneously. The optimal way is to choose groups of uncoupled voxels, i.e. voxels spread over the reconstruction volume. While this is most efficient for convergence reasons, updating groups of distributed voxels is less efficient regarding memory access and computational burden. In this work, an image-block update scheme is presented that updates relatively large groups of voxels simultaneously while keeping a considerable gain in convergence. The sequential image-block update can also be combined with ordered subsets. This image-block or patchwork scheme is applied both to transmission and emission maximum likelihood algorithms.

8.1 Introduction

Iterative reconstruction has several advantages over analytical, filtered backprojection (FBP) type, reconstruction. A forward model, as used in iterative reconstruction, is much more suitable for model adaptations, compared to the inverse model used in analytical reconstruction methods. Consequently, with iterative reconstruction one can implement more accurate acquisition models, resulting in more accurate reconstructions. Moreover, iterative reconstruction can cope better with incomplete data and truncation. Several groups have shown that iterative reconstruction can lead to dose reduction with preserved image quality, for example in [96, 97, 98]. This is partly due to the use of non-linear constraining techniques [99, 100] and partly to the use of an accurate noise model [101]. The most important disadvantage of iterative reconstruction is the increased computation time. For CT (computed tomography), this is still an important obstacle for introduction into daily clinical practice.

Many iterative algorithms update all image voxels simultaneously, based on a computation involving all projections. Examples are: MLEM [45], ISRA [102, 103], generalized MLEM/ISRA [104], NEGML [105], ABML [106] MLTR [46], convex algorithm [107], and SIRT [41]. The convergence of iterative algorithms is often improved by using blocks of projection data sequentially instead of using all projections simultaneously. This idea is used in the ordered subsets algorithms (OSEM[48], OSTR[108] and Ordered subsets convex algorithm [109]) or in ART [42, 43] and SART [44]. Convergence can be improved as well by updating voxels sequentially rather than simultaneously as was done by Sauer and Bouman in [110]. The improved convergence of this coordinate ascent approach is (partly) offset by the increased computation time per iteration. The sequential update requires to access the system matrix column by column (retrieving all elements for a single image voxel), while conventional projector/backprojector software typically enables row by row access only (retrieving elements for a single projection line). In CT reconstruction sequential updates also cause an increased number of exponentiation operations compared to one simultaneous update. A compromise between convergence and computational burden is made by Fessler et *al.* in the space-alternating generalized EM algorithms (SAGE [111]) and the grouped coordinate ascent algorithm [47, 83] where groups of voxels are updated. The optimal way of choosing these groups is by selecting voxels that have low coupling. This means that the update for one voxel is nearly independent of the updates of other voxels in the group. The lower the coupling between voxels in a group, the faster the convergence. For most system geometries this results in groups of voxels spread over the image. The main drawback of updating groups spread over the image is inefficient memory access.

In our work on metal artifact reduction in CT we developed a reconstruction algorithm based on a patchwork structure [112]. The patchwork structure divides the reconstruction volume in several subareas or patches. In case of metal artifact reduction the patches were chosen such that every metal object

and its close surroundings were considered as one patch, the remaining volume was another nonmetal patch. The original idea of the method was to apply different reconstruction models in different patches. By updating the regions separately, a grouped coordinate algorithm was applied to the patches (one patch is one group). A substantial increase in convergence was observed for the smaller (metal) patches, although the voxels in one group were adjacent and quite strongly coupled.

Benson et *al.* present in [113] another technique for updating coupled blocks of voxels in iterative CT reconstruction. The technique is an extension of the ICD technique described in [110]. The reconstruction volume is divided in several small blocks (maximum 8×8 voxels). In the block raster strategy, the different blocks are sequentially updated while the voxels in a block are simultaneously updated. The different axial planes are updated one by one. A linear system of equations for all voxels in the block is solved by inversion of the block matrix. This block matrix includes off-diagonal elements that describe the correlation between the voxels in a group. By including this correlation, larger (but still rather small groups) of voxels can be updated with improved convergence with respect to voxel by voxel updating but at the cost of increased computational burden caused by the inversion of a matrix with off-diagonal elements.

The idea of updating relatively large groups of connected voxels as in [112] will be further explored in this work. Instead of local improved convergence, the patchwork structure presented here aims to improve the overall convergence level of the reconstruction. The framework is a rather simple and intuitive gradient ascent maximum likelihood optimization algorithm. In the next section the maximum likelihood framework is introduced, followed by a description of the patch structure. The convergence of patched updates with respect to fully simultaneous updates with and without ordered subsets will be analyzed for different CT geometries and for 3D PET (positron emission tomography).

8.2 Methods

8.2.1 Maximum likelihood gradient ascent optimization

The study presented in this work is based on a gradient ascent maximization of the Poisson (log)likelihood, which is given by:

$$L(\vec{\tau}) = \sum_i \left(y_i \ln \hat{y}_i(\vec{\tau}) - \hat{y}_i(\vec{\tau}) \right), \tag{8.1}$$

with i the index of the projection lines, y_i the measurement, $\hat{y}_i$ the estimate of y_i calculated from the current reconstructed image $\vec{\tau}$, containing the attenuation coefficients in case of transmission tomography or the activity for emission tomography.

The derivation of the patchwork update scheme is based on a gradient ascent approach [46][105]. A surrogate for the likelihood is optimized based on a quadratic approximation:

$$L(\vec{\tau} + D(\alpha)\Delta x) \simeq T_1(\vec{\tau}, \Delta x) \tag{8.2}$$

$$T_1(\vec{\tau}, \Delta x) = L(\vec{\tau}) + \sum_j \left.\frac{\partial L}{\partial \tau_j}\right|_{\vec{\tau}} \alpha_j \Delta x_j + \frac{1}{2} \sum_{j,k} \left.\frac{\partial L^2}{\partial \tau_j \partial \tau_k}\right|_{\vec{\tau}} \alpha_j \alpha_k \Delta x_j \Delta x_k. \tag{8.3}$$

D is a diagonal matrix with $D_{jj} = \alpha_j \geq 0$. The second derivatives of L are all negative and

$$2\Delta x_j \Delta x_k \leq {\Delta x_j}^2 + {\Delta x_k}^2. \tag{8.4}$$

Using (8.3) and (8.4) a surrogate function T_2 is constructed which is equal to T_1 in the current solution and smaller than T_1 elsewhere.

$$\begin{aligned} T_1(\vec{\tau}, \Delta x) &\geq T_2(\vec{\tau}, \Delta x) && (8.5) \\ &= L(\vec{\tau}) + \sum_j \left.\frac{\partial L}{\partial \tau_j}\right|_{\vec{\tau}} \alpha_j \Delta x_j \\ &\quad + \frac{1}{2} \sum_{j,k} \left.\frac{\partial^2 L}{\partial \tau_j \partial \tau_k}\right|_{\vec{\tau}} \alpha_j \alpha_k {\Delta x_j}^2. && (8.6) \end{aligned}$$

The update for $\vec{\tau}, \Delta\tau_j = \alpha_j \Delta x$, is obtained by maximizing the surrogate function T_2:

$$\Delta\tau_j = -\frac{\alpha_j \left.\frac{\partial L}{\partial \tau_j}\right|_{\vec{\tau}}}{\sum_h \alpha_h \left.\frac{\partial^2 L}{\partial \tau_j \partial \tau_h}\right|_{\vec{\tau}}} \tag{8.7}$$

This update maximizes T_2, which guarantees an increase of T_1. This in turn is expected to increase the likelihood as well, provided that the quadratic approximation in (8.4) is 'sufficiently' accurate. The new reconstruction is $\vec{\tau}_{\text{new}} = \vec{\tau} + \Delta\vec{\tau}$. When choosing an appropriate projection model this update scheme can be applied on both transmission and emission data. The models and their corresponding update will be described in the following sections.

Transmission model

In transmission tomography, attenuation values μ_j, for all voxels j are reconstructed, thus $\tau_j = \mu_j$. The simplest projection model is given by:

$$\hat{y}_i(\vec{\mu}) = b_i \exp(-\sum_j l_{ij}\mu_j) \tag{8.8}$$

with i the projection ray index, j the pixel index and b_i the blank value for projection ray i. A more complex model includes a scatter term and resolution

modeling.

$$\hat{y}_i(\vec{\mu}) = \sum_n w_{in}\hat{\psi}_n + r_i \tag{8.9}$$

$$\hat{\psi}_i(\vec{\mu}) = b_i \exp(-\sum_j l_{ij}\mu_j) \tag{8.10}$$

$$\text{with } \sum_n w_{in} = 1 \text{ and } w_{in} = w_{ni},$$

modeling blurring with convolution kernel w_{in} and scatter with a known sinogram r_i. The first and second derivative of the likelihood become:

$$\frac{\partial L}{\partial \mu_j} = \sum_i l_{ij}\hat{\psi}_i \sum_n \frac{\hat{y}_n - y_n}{\hat{y}_n} \tag{8.11}$$

$$\frac{\partial^2 L}{\partial \mu_j \partial \mu_h} = -\sum_i l_{ij}\hat{\psi}_i \sum_n w_{in} \left((1 - \frac{y_n}{\hat{y}_n}) l_{ih} + \frac{y_n}{\hat{y}_n^2} \sum_m l_{mh} w_{nm} \hat{\psi}_m \right). \tag{8.12}$$

We now apply the update given in equation (8.7):

$$\Delta\mu_j = \frac{\alpha_j \sum_i l_{ij}\hat{\psi}_i \sum_n \frac{\hat{y}_n - y_n}{\hat{y}_n}}{\sum_i l_{ij}\hat{\psi}_i \sum_n w_{in} M_n} \tag{8.13}$$

$$M_n = \left((1 - \frac{y_n}{\hat{y}_n})(\sum_h \alpha_h l_{ih}) + \frac{y_n}{\hat{y}_n^2} \sum_m (\sum_h \alpha_h l_{mh}) w_{nm} \hat{\psi}_m \right). \tag{8.14}$$

For particular choices of α_j, e.g. $\alpha_j = 1$, we can assume that $\sum_h \alpha_h l_{ih}$ is smooth and hardly modified by the blurring kernel w_{in}. This results in:

$$\Delta\mu_j = \frac{\alpha_j \sum_i l_{ij}\hat{\psi}_i \sum_n \frac{\hat{y}_n - y_n}{\hat{y}_n}}{\sum_i l_{ij}\hat{\psi}_i (\sum_h \alpha_h l_{ih}) \sum_n w_{in} \frac{\hat{y}_n^2 - y_n r_n}{\hat{y}_n^2}}. \tag{8.15}$$

For model (8.8) $w_{in} = \delta_{in}$ and $r_i = 0$ the update given in (8.15) becomes:

$$\Delta\mu_j = \frac{\alpha_j \sum_i l_{ij}(\hat{y}_i - y_i)}{\sum_i l_{ij}(\sum_k \alpha_k l_{ik})\hat{y}_i}. \tag{8.16}$$

The same algorithm is obtained by introducing an exact separable surrogate function for the likelihood, followed by a quadratic approximation, as was proposed by Fessler et *al.* in [47].

The voxel weight α_j is a design parameter that influences the convergence. If the weight α_j for voxel j is increased relative to the weight of the other voxels, the update (8.16) for voxel j is increased as well, because (8.16) has weight α_j

in the numerator and a sum over all weights in the denominator. Choosing α_j equal to 1 for all voxels results in the MLTR algorithm described in [46]. Equal update weights lead to uniform convergence for all regions in the object. When the update weights are proportional to the attenuation, Eq. (8.16) becomes the so-called convex algorithm [107], which has faster convergence for highly attenuating regions at the expense of reduced convergence for regions with low attenuation.

In the experiments below the simplified update with $w_{in} = \delta_{in}$ and $r_i = 0$ will be used. The voxel weights α_j will be set to 1 to obtain MLTR.

Emission model

In emission tomography $\tau_j = \lambda_j$, the activity in voxel j. The basic model is given by:

$$\hat{y}_i(\vec{\lambda}) = \sum_j c_{ij}\lambda_j \tag{8.17}$$

where c_{ij} represents the probability that a photon (pair) emitted in pixel j will be detected in detector (pair) i, including attenuation and sensitivity for line of response i. In a similar way as for the transmission model, also resolution effects can be modeled which yields:

$$\hat{y}_i(\vec{\lambda}) = \sum_n w_{in}\hat{\psi}_n + r_i \tag{8.18}$$

$$\hat{\psi}_i(\vec{\lambda}) = \sum_j c_{ij}\lambda_j \tag{8.19}$$

$$\text{with } \sum_n w_{in} = 1 \text{ and } w_{in} = w_{ni}.$$

where r_i are the randoms and/or scatter. Since the model in Eq. (8.17) is linear, it can be simplified by including resolution modeling in c_{ij}, which is denoted by a_{ij}:

$$\hat{y}_i(\vec{\lambda}) = \sum_n a_{ij}\lambda_j + r_i \tag{8.20}$$

The derivatives of the likelihood are:

$$\frac{\partial L}{\partial \lambda_j} = \sum_i a_{ij}\frac{y_i - \hat{y}_i}{\hat{y}_i} \tag{8.21}$$

$$\frac{\partial^2 L}{\partial \lambda_j \partial \lambda_h} = -\sum_i a_{ij}\frac{y_i}{\hat{y}_i^2}a_h. \tag{8.22}$$

Assuming $y_i \simeq \hat{y}_i$ for the second derivative, the update for activity λ_j becomes:

$$\Delta\lambda_j = \frac{\alpha_j \sum_i a_{ij}\frac{y_i - \hat{y}_i}{\hat{y}_i}}{\sum_i a_{ij}\frac{1}{\hat{y}_i}\sum_h a_{ih}\alpha_h}. \tag{8.23}$$

The separable surrogate technique in [47] could also be applied to the model described in (8.20). This would result in the same update step (also under the assumption that $y_i \simeq \hat{y}_i$ for the second derivative).

The update weight α_j has the same role as for the transmission model. Setting $\alpha_j = 1$ for all j produces the NEGML algorithm of [105][114]. When $\alpha_j = \lambda_j$, Eq. (8.23) resembles the MLEM (maximum likelihood expectation maximization [45]) algorithm and becomes identical to it when $r_i = 0$.

Note that both for transmission and emission tomography, the sinogram $\sum_j l_{ij}\alpha_j$ or $\sum_j a_{ij}\alpha_j$ could be calculated beforehand when α_j is constant over all iterations. For $\alpha_j = 1$ for all j, this sinogram can be approximated by a single value, i.e. the diameter of the field of view.

8.2.2 Patchwork reconstruction

The update weight α_j described in the previous section is the basis of the patchwork reconstruction. By choosing update weights equal to 1 for a certain subset of voxels and setting the weights to 0 for the rest the object, the convergence for that group of voxels is improved but there is no update for the voxels outside that region. One main iteration can then be defined as a number of subiterations where each time another group of voxels receives nonzero update weights. All voxels should be part of one single group. This principle was applied in the grouped coordinate algorithm [47] where each group consists of voxels spread over the reconstruction volume. This way the voxels have the lowest coupling, minimizing the denominator in Eq. (8.16) or (8.23). In the patchwork reconstruction, the group of voxels, called patches, will be groups of neighboring voxels. The different patches are determined by dividing the reconstruction volume in P equally sized image-blocks. The image volume is only subdivided in transaxial (x- and y-) direction and not for the axial or z-direction. The reason is that updating a voxel in one plane has only little or no influence on voxel values in other planes. Therefore dividing the volume in axial direction will not further improve convergence. The division of a reconstruction volume into sixteen patches is shown in figure 8.1.

The gain in convergence of using patches can be estimated when considering the sum in denominator of Eq. (8.16) and (8.23). The difference in the update step will be determined by the change in $\sum_h l_{ih}\alpha_h$ or $\sum_h a_{ih}\alpha_h$. In the described configuration, the convergence is expected to be proportional to the square root of the number of patches in the patch structure.

The patched update will be combined with ordered (projection) subsets. A reconstruction scheme using patches and projections subsets means that for one subset all patches will be updated and then the algorithm goes on to the next subset.

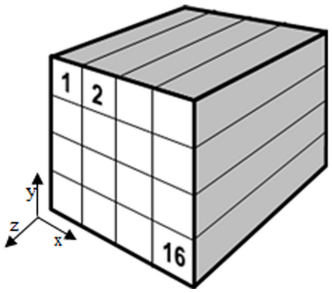


Figure 8.1: Division of three dimensional reconstruction volume into sixteen patches: x and y are the transaxial directions, z is the axial direction: CT rotational axis or PET symmetry axis.

8.3 Materials

8.3.1 Transmission tomography

Three different sets of transmission data were used: circular CT, cone beam CT and helical CT.

Circular CT

A circular scan was performed on a Siemens Sensation 16 system (part of the Biograph 16 PET/CT). A circular PMMA phantom with four aluminum inserts (two of 1 cm diameter and two of 3 cm diameter) was scanned with tube voltage 120 kV and tube current 300 mA. The collimation used for this scan was 2 x 1.00 mm. The system has a curved detector with in each row of 672 detector elements. The scan was a circular scan of 360° scan with 1160 projection angles. The field of view was 50 cm and represented by a reconstruction grid of 512 x 512 pixels. A two-dimensional reconstruction is performed by using only one of the scanned slices.

Cone beam micro CT

The second data set is a microCT scan of a mouse bone in a plastic tube filled with water. The scan is performed on a Skyscan 1172 microCT (cone beam), at 49 kV and 200μA with an 0.5 mm Al-filter. This was an 180° scan with 264 views. The detector was 1320 $\times$ 1048 pixels of 11.57 μm and this was rebinned to a pixel size of 23.14 μm during reconstruction. The reconstruction grid is 552 $\times$ 552 $\times$ 78 with isotropic voxel size of 12.0 μm.

Helical CT

The last set is a helical whole body patient scan on a Siemens Sensation 16 system (CT part of Biograph 16). The mean effective tube current was 99 mA and the tube voltage 120 kV (DLP 476.8 mGy cm, scan length 975 mm[1]. The table feed was 18 mm per rotation (pitch factor = 0.75), the rotation time 0.5s per rotation and the collimation 16 x 1.50 mm. The reconstruction grid is 512 $\times$ 512 $\times$ 48, representing 1 mm $\times$ 1 mm $\times$ 3 mm voxels.

8.3.2 Emission tomography

One emission data set was used. It was a 3D 30 min. ^{18}F-FDG PET brain study acquired on a Siemens HR+ system. The system resolution was modeled by a 2D Gaussian with FWHM of 1.78 mm in the transaxial direction and 2.25 mm in the axial direction. The reconstruction grid was 144 $\times$ 144 $\times$ 63.

Table 8.1: Reconstruction scheme to obtain the reference reconstruction

Data sets	Reconstruction scheme	Equivalent iterations
Circular CT	$20 \times 232 + 20 \times 116 + 20 \times 58 + 20 \times 29 + 20 \times 1$	8720
Cone beam CT	$50 \times 66 + 50 \times 44 + 50 \times 22 + 50 \times 11 + 50 \times 1$	7200
Helical CT	$20 \times 232 + 20 \times 116 + 20 \times 58 + 20 \times 29 + 20 \times 1$	8720
PET	$50 \times 36 + 50 \times 27 + 50 \times 18 + 50 \times 9 + 50 \times 1$	4550

8.3.3 Implementation and system information

The projector used in this study is the distance driven projector [84]. The distance driven projector was chosen because it implements the backprojector as the exact transpose of the projector, which is important for most iterative reconstruction algorithms and because it is considered as more efficient compared to pixel and ray driven methods regarding memory access. The reconstruction software is written in IDL and C and is not optimized for parallel computing on CPUs or GPUs. The system used in this study is a Solaris 64bit workstation with an Intel E5440 CPU with 12 MB of L2 cache and 32 GB RAM.

[1]The scan length is given together with the DLP (dose length product) to give an idea about the dose in this scan. During reconstruction only the lung region of the patient is reconstructed.

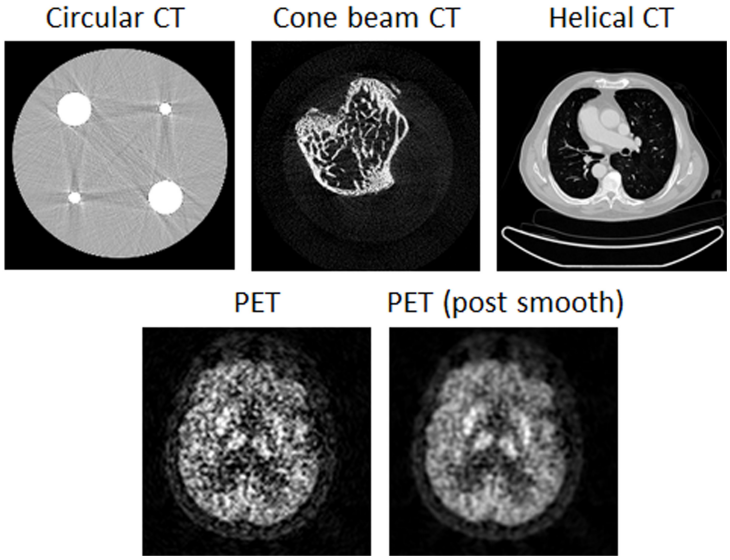


Figure 8.2: The reference images after reconstruction as indicated in table 8.1. The PET image is shown with and without post smoothing (Gaussian with full width half maximum 1.5 voxels).

8.4 Convergence analysis

This study investigates the gain in convergence by using a patched reconstruction update. The convergence for different reconstructions schemes will be compared: reconstructions with and without projection subsets (classical ordered subsets reconstruction [48]), with and without patches and combinations of projections subsets and patches. The number of patches will be 1 (regular reconstruction), 4, 16 or 64, the number of projections subsets will be 1, 20, or 40 for CT and 1, 10, or 20 for PET.

To determine the degree of convergence of a particular reconstruction, a measure based on the quadratic difference with a reference image is calculated:

$$\text{Quadratic difference} = \sum_j \left(\tau_j^{\text{ref}} - \tau_j^{\text{recon}}\right)^2 \tag{8.24}$$

with τ_j^{recon} the image value in the reconstruction under investigation and τ_j^{ref} the corresponding image value in the reference image. This reference image is obtained by a long iterated reconstruction with decreasing number of subsets to prevent limit cycle solutions [48]. The particular iteration scheme can be found in table 8.1 as well as the equivalent number of iterations without subsets. The equivalent number of iterations is calculated based on the assumption that N iterations using M projections subsets per iteration is equivalent to

N × M iterations without subsets. The degree of convergence of a certain reconstruction will be measured by the quadratic difference of that certain reconstruction with the reference image. The reference images are shown in figure 8.2.

For comparing different reconstruction schemes with and without patches and/or subsets, we calculate the convergence level obtained at 200 standard iterations. Next, we calculate when the same convergence level is reached for schemes with subsets and/or patches. This value is calculated based on linear interpolation of the convergence curves, and therefore a non-integer number of iterations may be obtained.

Sequentially updating results in faster convergence of higher frequencies [47, 110]. The more patches are used, the more pronounced this effect will be. Moreover, a convergence difference between neighboring patches can arise, depending on the order of patch updates. Both effects can give rise to artifacts in an early stage of the reconstruction. Some of these artifacts, especially those at patch borders can persist during the rest of the reconstruction. The artifacts appear as very thin lines (one or two voxels) at the patch borders and have zero means in most projections, through which it can take very long to reduce them. This effect can be avoided by underrelaxation of the first updates. A relaxation is applied to the first five updates, i.e. for the first five subsets or, if no subsets are used, the first five iterations. The underrelaxation was chosen such that the update step is the same as in a non-patched update. In practice iterative reconstruction is often started from a filtered backprojection (FBP) reconstruction. In that case no relaxation is needed since the convergence level is already relatively high. Since convergence is reached in fewer iterations when starting with FBP images, that approach was not used in this study to facilitate the observation of convergence differences. However, we verified that very similar results are obtained when starting from FDK images in the microCT study (results not shown). For the results shown below, the reconstructions were started from a uniformly filled (rough) body contour (filled with water attenuation for CT, and with the mean of the sinogram divided by the image size for PET).

8.5 Results

Table 8.2, 8.3, and 8.4 show the convergence results for the circular CT, cone beam CT and helical CT data sets respectively. Table 8.5 shows the convergence results for the PET data set. The numbers in the table represent the number of iterations necessary to achieve the same level of convergence as obtained by 200 iterations without subsets and patches. For instance, the convergence level of the microCT data at 200 iterations without subsets and patches, can be obtained in 2.8 iterations when 20 subsets and 16 patches are used. The reconstruction with 20 subsets but without patches reaches this level

in 9.8 iterations. This is a convergence gain of almost 4.

Table 8.2: Convergence results for circular CT

Patches	1 subset	20 subsets	40 subsets
1	200.0	10.0	5.0
4	104.3	5.1	2.7
16	54.5	2.8	1.5
64	29.6	1.7	0.1

Table 8.3: Convergence results for cone beam - micro CT

Patches	1 subset	20 subsets	40 subsets
1	200.0	9.8	5.0
4	103.2	5.0	2.7
16	53.1	2.8	1.6
64	29.0	1.8	1.2

Table 8.4: Convergence results for helical CT

Patches	1 subset	20 subsets	40 subsets
1	200.0	9.6	5.4
4	101.3	5.0	2.8
16	52.2	2.8	1.6
64	27.7	1.6	0.8

The data for the helical CT scan with subsets and 1, 4 or 16 patches is plotted in figure 8.3. The upper curve shows convergence per iteration, the lower curve convergence per time. Using more patches and/or subsets increases the convergence. The curves *1 patch, 40 subsets* and *4 patches, 20 subsets* coincide, also *4 patches, 40 subsets* coincides with *16 patches, 20 subsets*. Only one of these pairs also coincides in the convergence versus time curve. While *4 patches, 20 subsets* has the same convergence per iteration as *1 patch, 40 subsets*, less time per iteration is required. No improvement in time per iteration is observed when increasing the number of patches from 4 to 16. A visual example of the different convergence can be found in figure 8.4 and figure 8.5 where reconstructions with the different patch schemes are depicted. A visual example of the different convergence can be found in figure 8.4 and figure 8.5 where reconstructions with the different patch schemes are depicted.

8.6 Discussion

The popularity of iterative reconstruction in clinical practice is increasing. The main drawback is the computational burden of most iterative techniques. The

Table 8.5: Convergence results for PET

Patches	1 subset	10 subsets	20 subsets
1	200.0	20.0	10.0
4	102.5	10.3	5.1
16	53.7	5.4	2.7
64	29.4	3.0	1.6

patchwork reconstruction presented in this work aims to reduce this problem. By using a sequential update of image parts the convergence of the reconstruction is accelerated. As a measure of convergence the quadratic difference with a reference reconstruction was calculated. An acceleration of the square root of the number of patches was expected. The patchwork scheme was also combined with an ordered subset scheme where we expect an acceleration by the number of subsets. The experimental convergence results for all four modalities show a convergence gain that agrees well with these expectations. Some values are slightly higher than expected, which may be due to the linear interpolation which gives a slight overestimation for a concave monotonic decreasing function. In cases without subsets the underrelaxation of the first five updates also influences the convergence result. In practice one would often use subsets while starting from FBP which makes relaxation unnecessary. A visual perception of the improved convergence is given in figures 8.4 and 8.5. The same iteration scheme (iterations $\times$ subsets) with patches results in sharper edges, more details and fewer metal artifacts.

It might seem beneficial to further subdivide the image into more patches to increase the convergence even more. In the patch structure as described above this is most probably not advisable. By using relatively large image blocks with connected neighboring voxels, the patchwork structure enables the use of regular projector and backprojector software, which is usually optimized for dealing with large series of neighboring image voxels and projection pixels. When patches become smaller the advantage of using these projectors becomes smaller. Moreover, since the different regions are sequentially updated, the level of convergence is different for patches being updated first compared to patches that are updated later on. This difference becomes more prominent when there are more patches which also implies that it would take longer (i.e. more iterations) to obtain a relatively uniform convergence over the image. Besides, when more image parts are sequentially updated, the computational overhead becomes more important, which increases the required computation time to update the whole volume. Finally, when using relatively large patches, the patchwork structure can be combined with ordered projections subsets, giving a two-sided acceleration. The smaller the patches become, the higher the chance that the subset balance is strongly violated. More patches would eventually lead to a similar algorithm as ABCD in [83] where all axial rows are sequentially updated and which requires dedicated projection/backprojection software.

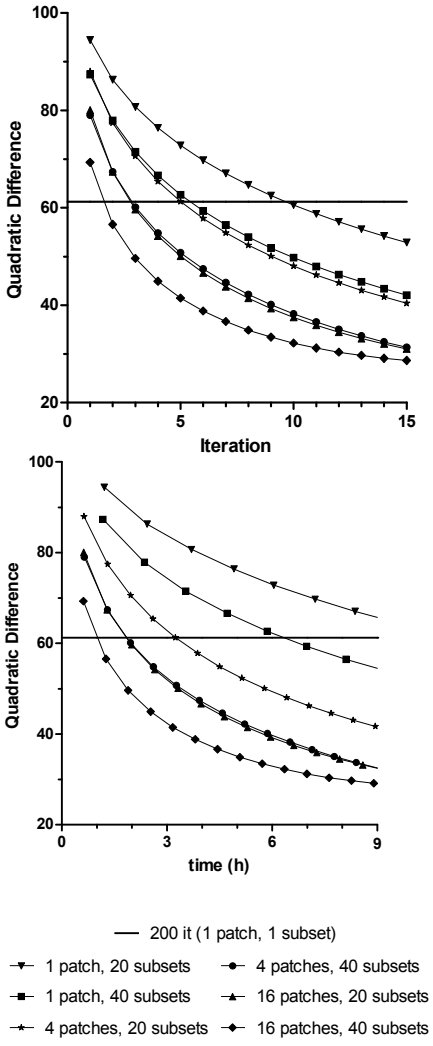


Figure 8.3: Helical CT data set: convergence per iteration (top) and per time (bottom). The horizontal line is a reference line representing the convergence level obtained after 200 iterations without patches or subsets.

When dealing with large data sets the patch reconstruction might have another advantage. By updating only a part of the reconstruction volume, more efficient data management may be achieved. For the helical CT data set (figure 8.3), the time required for one iteration with patches is only 60% of the time required for one regular iteration. This is because the helical CT data set is large and by updating only a part of the reconstruction volume at once, less data need

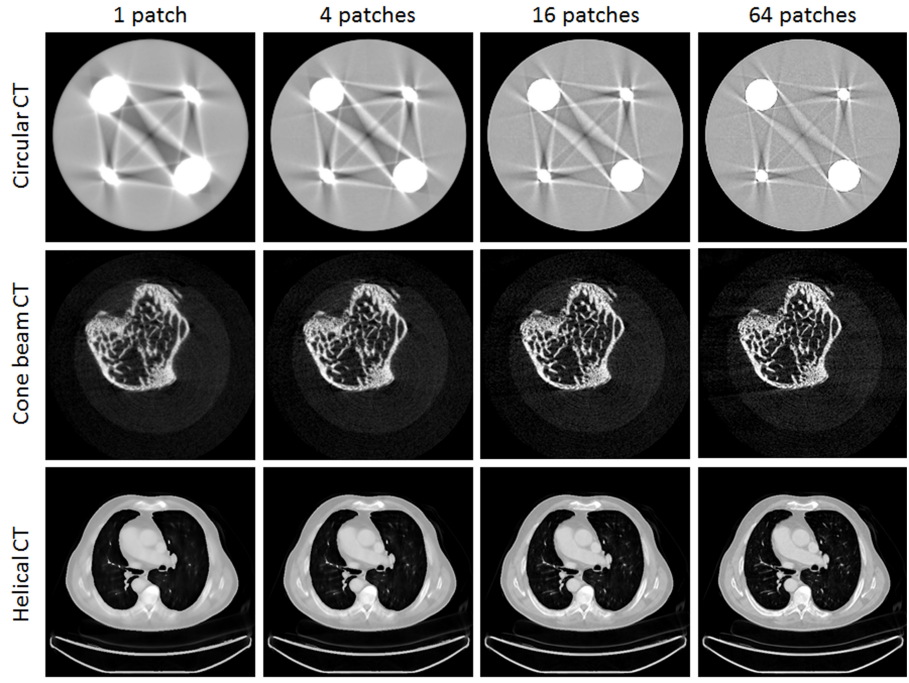


Figure 8.4: MLTR reconstructions for circular, cone beam, and helical CT for 2 iterations and 40 subsets. MLTR with 1 (= regular), 4, 16, or 64 patches.

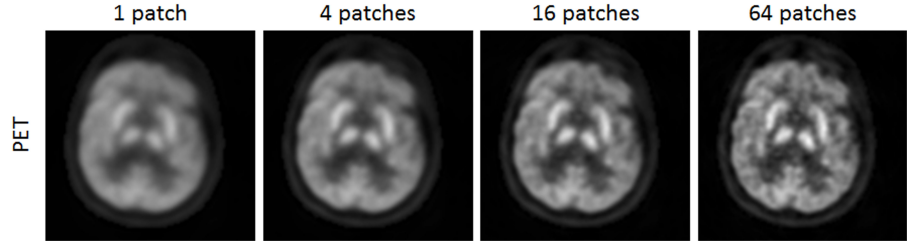


Figure 8.5: NEGML reconstructions for PET for 2 iterations and 20 subsets. NEGML with 1 (= regular), 4, 16 or 64 patches. Post-smoothed with a Gaussian kernel with a full width half maximum of 1.5 voxels.

to be handled at same time which (depending on the system) might give an acceleration in time per iteration additionally to an improved convergence per iteration. This advantage disappears when data sets are small like for the circular CT.

Nonuniform convergence of certain image parts can also be used on purpose. In our previous work on metal artifact reduction [112], the image was divided in several metal patches and one patch without metals. Since the metal patches were smaller, their update was larger, giving an increased convergence compared to the non-metal patch that receives an update with equal step size as when updating the whole image simultaneously. This resulted in faster and more efficient artifact reduction. In this work the projection model for each patch could be chosen separately. The voxel size and energy model could be different for the patches. Using different voxel sizes over the image also allows for region-of-interest reconstruction [87, 88, 89, 93, 94, 95]. The region of interest would be one patch, reconstructed with small voxels, the remaining region is another patch, reconstructed with larger pixels. This gives two advantages, the background is reconstructed with larger voxels which lowers the computation time and the patch structure is applied which will increase the convergence of the region-of-interest patch (especially when it is considerably smaller than the total volume). Besides, the patchwork reconstruction combined with non-uniform resolution modeling was used for digital breast tomosynthesis in Michielsen et *al.* [115].

8.7 Conclusion

We have presented a simple maximum likelihood gradient ascent algorithm to obtain update formulas for both emission and transmission tomography reconstruction. The algorithm includes a design parameter α_j that can alter relative convergence of particular image regions or image voxels. A very similar approach was previously used by Fessler et *al.* [47] to derive a grouped coordinate ascent algorithm for transmission tomography. By sequentially setting this parameter to 1 for a certain region and 0 for the rest of the object, a sequential 'patchwork' reconstruction can be obtained. Unlike other sequential algorithms, the different regions or patches that are sequentially updated are rather large groups of neighboring voxels. We have shown that this still gives a considerable gain in convergence while projection and backprojection can be done efficiently with conventional (back)projectors. Moreover, the larger groups allow for combining sequential voxel updates with ordered (projection) subsets. For very large data sets, the scheme also reduces the memory requirements, which may further reduce the computation time on computers with limited memory.

Chapter 9

Metal artifact reduction for three-dimensional helical CT

Katrien Van Slambrouck and Johan Nuyts

Purpose Iterative maximum likelihood based metal artifact reduction can be improved and accelerated by the use of local models in the so-called patchwork structure. The method was already applied on two-dimensional phantom data. In this work the method has been extended for three-dimensional helical CT and applied on patient data. This extension requires some specific changes to the local model scheme.

Method The patchwork structure divides the reconstruction volume in several regions or patches. Every metallic object is considered as one patch, a last patch is the remaining nonmetal region. This scheme enables the use of complex, accurate models for metal patches and simpler models elsewhere. As for the two-dimensional case, the energy and sampling properties of the forward model can be changed for the three-dimensional case. In reconstruction of helical CT data, typically anisotropic voxels are used with larger slice thickness compared to the in-slice voxel size. To prevent strong partial volume effects, the slice thickness was locally decreased. In addition, cross talk was modeled and the influence of the initial image on the final reconstruction was considered. An image-block acceleration scheme was tested for acceleration of the reconstruction in the nonmetal patch to decrease the required computation time. The method was applied on five different patient data sets and one phantom with different kinds of metal implants. The reconstructions with the

extended local model reconstruction were compared to FBP and to a sinogram inpainting method.

Results The method clearly reduces metal artifacts compared to FBP and regular iterative metal artifact reduction. More contrast is preserved between and around the metals compared to the sinogram inpainting. However, some artifacts are still present, especially at locations where there is a substantial axial variation in the shape of the metal. Scatter and partial volume effects have been investigated as the origin of these artifacts but no evidence was found. The initial image is of great importance for efficient artifact reduction. Moreover, a locally decreased slice thickness for the metal was required.

Conclusions The three-dimensional local model reconstruction scheme reduces artifacts substantially. It preserves contrast better than sinogram inpainting. However, some artifacts are present with unclear origin.

9.1 Introduction

Metal artifacts are an important remaining problem in computed tomography (CT) reconstruction. Metal artifacts are typically dark and bright streaks in between and around metallic structures such as femoral (hip) prostheses, dental fillings, metallic fixation screws, etc. These artifacts often cover important parts of the image and may hinder proper diagnosis in the affected regions.

The first method to reduce metal artifacts was already presented in the 1980s [63, 64] and over the years many new methods have been described. The methods for metal artifact reduction (MAR) can be divided into two main groups. The first and largest group contains sinogram completion or sinogram inpainting methods. The detected intensity of projection lines passing through metallic structures is often very low or even zero. The information in these projections is therefore limited. Sinogram inpainting methods assume that since the information in the metal projections is very limited, they can as well be replaced by artificial projections that cause fewer artifacts. The basic scheme of most inpainting methods is: reconstruct the data, segment the metals in the reconstruction, reproject the segmentation of the metals to determine the metal trace in the sinogram and replace the metal trace by artificial data. This new sinogram is then reconstructed and the segmented metals are pasted on top of the new reconstruction. The inpainting methods differ in the way the artificial data are created. Most methods use interpolation (linear [63], polynomial [64, 65]) between the neighboring nonmetal projections. The major disadvantage of all the inpainting methods is that by discarding all metal projections, all remaining information in these projections is also removed. This is information about contrast in the close neighborhood of metals and about edges tangent to the discarded projection lines. Solutions have been developed and

an important effort is done by Müller in [67] and Meyer et al. in NMAR[68], FSMAR[70] and ANMAR[72]. In these algorithms part of the original information is reintroduced in the interpolated rays. This limits the information loss, although the technique might also introduce part of the original artifacts. Other inpainting techniques that also use information of the materials for the creation of artificial data were developed by Bal and Spies [69] and Lemmens et al. [66]. Subtraction-based reprojection is a technique based on the same principle as inpainting where the metal traces in the sinogram are modified by subtracting inconsistencies [116]. This technique was recently implemented on some clinical CT systems [117]. An important advantage of most inpainting and subtraction methods is that they are fast since they are usually based on filtered backprojection (FBP) algorithms.

The second group is based on iterative algorithms. The origin of metal artifacts is an inaccurate modeling of the system physics. The most important causes of metal artifacts are described in detail in Barrett and Neat [54] and De Man et al. [55]. Summarized, these effects are: beam hardening, partial volume effects, scatter, noise and motion. The iterative methods try to prevent metal artifacts by using a more accurate (forward) model during the reconstruction that incorporates (some of) the listed effects. This is either done by using standard iterative reconstruction [77, 78, 79] or iterative FBP [118, 119]. The artifact reduction is usually efficient for few, small and relativity low attenuating metals but remaining artifacts are present for multiple, larger or higher attenuating metals. However, since no data are discarded better contrast preservation is achieved compared to sinogram inpainting.

In Van Slambrouck and Nuyts [112] we developed a new iterative MAR method that was tested on two-dimensional simulations and measurements. This method performed better than sinogram completion in retrieval of contrast and artifacts were clearly reduced compared to regular iterative algorithms. However, some streak artifacts were slightly more pronounced. The algorithm is based on the *patchwork structure* which has two important features compared to standard iterative reconstruction: the reconstruction model can be varied over the image and the convergence rate of the metal regions can be accelerated. The patch structure developed in [112] divides the reconstruction volume into different parts: one large nonmetal patch and one or more smaller metal patches (one for each metallic structure). A different reconstruction model was assigned to each of the patches. Fully polychromatic IMPACT [76] was chosen for the metal patches and a simpler approximate polychromatic model for the nonmetal patches. This model was MLTRC and was presented in [112]. The model only used the polychromatic behavior of water attenuation instead of modeling all materials as is done in IMPACT. Lowering the complexity of the reconstruction model for most of the image from IMPACT to MLTRC had no influence on the quality of the artifact reduction but substantially decreased the computation time for the algorithm. An image-block or grouped coordinate algorithm was applied for updating the different patches [47, 120]. This means that the patches are updated separately and sequentially. When only a small

region of the image is updated, larger step sizes can be used during image updating without compromising convergence, resulting in faster convergence of the voxel values in that region. The improved convergence for the metal patches resulted in a substantial reduction of the artifacts even for relatively large and highly attenuating metals. The reason is that the metals obtain higher attenuation values and sharper edges in an earlier stage of the reconstruction procedure and the rest of the image does not need to compensate for low metals values.

The work described in [112] was a two-dimensional phantom study. In this work, the two-dimensional method will be extended to three-dimensional helical CT and evaluated on patient data. Some specific issues related to three-dimensional reconstruction and reconstruction of patient data have been investigated to optimize the algorithm.

9.2 Materials and Methods

Section 9.2.1 gives an overview of the patchwork structure. Section 9.2.2 describes the considerations and modifications of the existing patchwork structure for helical CT and section 9.2.3 describes some topics for accelerations and optimization of the reconstruction.

9.2.1 Patch structure for local model reconstruction

The patch structure and patch reconstruction were developed in [112]. A brief overview of the different parts of the algorithm is given in this section.

Patch definition

The patch structure divides the reconstruction volume in different regions or *patches*. These patches are defined based on an initial reconstruction (see figure 9.1). This can be an FBP image or a single iteration (with subsets) of an iterative algorithm. In this work we will define the patches based on an image obtained after sinogram inpainting, since that image will also be used further on. The metals are segmented from this initial image. The segmentation is performed by a thresholding method, based on k-means clustering. Each metallic structure will be considered as a different patch. To ensure that the patch includes the complete metal, the segmentation is radially enlarged by a few voxels. The reconstruction volume without all the metal patches is considered as a last patch, the nonmetal patch.

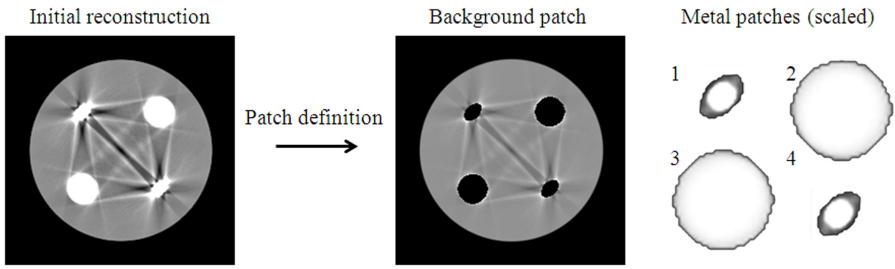


Figure 9.1: **Patch definition** Definition of the different patches based upon segmentation of an initial reconstruction. Result: one nonmetal patch and four metal patches.

Reconstruction model

A different projection model can be assigned to each of the patches. The current algorithm allows to change the energy and sampling model for each individual patch.

Energy model The reconstruction algorithms are based on the optimization of the Poisson (log)-likelihood by a gradient ascent approach described in [112, 120]. The algorithms differ in the projection model they are using, e.g. the spectrum or energy model. The basic monochromatic projection model is given by:

$$\hat{y}_i = b_i \exp\left(-\sum_j l_{ij}\mu_j\right) + r_i \tag{9.1}$$

with $\hat{y}_i$ the estimated photon intensity for projection line i, b_i the blank value, l_{ij} the intersection length of projection line i with voxel j, μ_j the linear attenuation coefficient in voxel j and r_i a known scatter term.

In this study two different polychromatic reconstruction models will be used that discretize the polychromatic spectrum by a set of energy bins. The bins have a fixed energy range with a varying intensity. The polychromatic version of Eq. (9.1) becomes:

$$\hat{y}_i = \sum_k b_{ik}\hat{y}_{ik} + r_i = \sum_k b_{ik} \exp\left(-\sum_j l_{ij}\mu_{jk}\right) + r_i \tag{9.2}$$

with k the index for the energy bins and b_{ik} the blank value for projection line i at energy k. Directly applying this projection model requires the reconstruction of K μ-values per image voxel (with K the total number of energy bins). Two

different polychromatic reconstruction models are introduced that reduce the problem to the reconstruction of only one single μ-value per image voxel.

The first algorithm is a fully polychromatic algorithm, IMPACT [76]. The algorithm uses the fact that the polychromatic attenuation is a linear combination of the attenuation caused by Compton scattering and the photo-electric effect. Both effects are decomposed in a material and energy dependent component [121]:

$$\mu_{jk} = \phi(\mu_{j,\text{ref}}) \cdot \Phi_k + \theta(\mu_{j,\text{ref}}) \cdot \Theta_k \tag{9.3}$$

which gives the following update equation:

$$\hat{y}_i = \sum_k b_{ik} \exp\left(-\Phi_k \sum_j l_{ij} \phi(\mu_{j,\text{ref}}) - \Theta_k \sum_j l_{ij} \theta(\mu_{j,\text{ref}}) \right) + r_i \tag{9.4}$$

where ϕ and θ are the material dependencies of the photo-electric effect and Compton scattering respectively and Φ_k and Θ_k are the energy dependencies for both effects. The values for ϕ and θ are calculated based on a predefined materials list. For each of the materials, ϕ and θ at a reference energy E_{ref} are calculated. When the materials are ordered for increasing attenuation this gives a piece-wise linear curve that relates ϕ and θ unambiguously to the attenuation at a reference energy $\mu_{j,\text{ref}}$. This reduces the problem to the reconstruction of one unknown per image voxel. A detailed derivation of IMPACT can be found in [76, 112].

The second polychromatic algorithm is MLTRC which was introduced in [112]. MLTRC only considers the energy dependent attenuation of water. The projection model is given by:

$$\hat{y}_i = \sum_k b_{ik} \exp\left(-C_k \sum_j l_{ij} \mu_{j,\text{ref}} \right) + r_i \text{ with } C_k = \frac{\mu_k^{\text{water}}}{\mu_{\text{ref}}^{\text{water}}}. \tag{9.5}$$

This model is correct for water but the more the attenuation values of a material differ from those of water the less accurate the reconstructed value will be. Hence, the effect of MLTRC on the reconstruction is similar to a (polynomial) water precorrection of CT data [82]. MLTRC also reduces the initial problem in Eq. (9.2) to the reconstruction of one unknown per image voxel.

More complex projection models usually require longer calculation times. The projections and backprojections are the main time consuming operation in the described algorithms. Hence, the complexity and computation time of an algorithm can be expressed by the number of (back) projections. One iteration for MLTRC results in one projection and two back projections per iteration. IMPACT is much more computationally demanding and needs four projections and four back projections per iteration. The patch structure offers a way to limit the number of projection and backprojections per iterations by allowing different reconstruction models for the different patches. In [112] we have shown

on two-dimensional examples that a similar image quality can be obtained when applying IMPACT for metal patches and MLTRC for the nonmetal patch compared to applying IMPACT for the whole reconstruction volume.

Sampling The different patches can use different sampling models. The sampling model is defined by the voxel size and the detector oversampling. For each patch the voxel size can be separately adjusted. The detector can be oversampled during the reconstruction which implies that a projection will be divided in several smaller detector elements during the projection. The degree of oversampling can be chosen for each patch separately. After projection and exponentiation the oversampled projections are again combined:

$$\hat{y}_i = b_i \sum_{q=1}^{Q} \exp\left(- \sum_{m=1}^{M} l_{iqm} \mu_m \right) \tag{9.6}$$

where each detector element i is divided into Q subsamples, a new voxel index M defines the new number of (smaller) voxels ($M = J * f$, with f the (integer) size factor and J the original number of pixels).

Increased sampling can be applied to reduce (nonlinear) partial volume effects. In [112] the sampling for the metal patches was increased both by decreased voxel size and increased detector sampling. No substantial improvement compared to a standard sampling model could be observed for 2D CT.

Reconstruction The patches are reconstructed as groups of voxels in a grouped coordinate algorithm [47, 83]. This means that the different patches are separately updated. The combined projection estimate is given by:

$$\hat{y}_i = \sum_k b_{ik} \prod_p z_{ik,p} + r_i \text{ with } z_{ik,p} = \frac{\hat{y}_{ik,p}}{b_{ik}} \tag{9.7}$$

with p the patch index and $\hat{y}_{ik,p}$ the projection for patch p at energy k. During an update of patch p only the voxels belonging to that patch will be updated. The update will be calculated based on the sampling and energy model assigned to that particular patch. A new combined estimate is calculated after a patch has been updated and then the next patch is updated. In combination with ordered subsets, all patches will be updated for one subset and then the algorithm goes on to the next subset.

9.2.2 Considerations for three-dimensional helical CT reconstruction

The previous work presented in [112] considered two-dimensional simulations and measurements. When extending the method to three-dimensional data,

some issues need to be considered, in particular the anisotropic voxel size in most CT examinations and the effect of cross talk and scatter.

Slice width Anisotropic voxel sizes are often used for CT reconstructions: the slice width is typically chosen to be larger than the in-slice voxel width. Enlarging the voxel size in axial direction enables lowering the dose, since broader slices have lower noise. Moreover, high axial resolution is often considered unnecessary, because the visual inspection is often done using only the transaxial slices. The default slice width for reconstructions of most examinations on our systems is typically 5 mm [122, 123]. Only for examinations where smaller structures need to be considers (e.g. in inner ear examinations) smaller slice widths are used by default. In our hospital, this slice width is often lowered to 3 mm, yielding voxels sizes of 1mm $\times$ 1 mm $\times$ 3mm, for most examinations.

The larger the voxels during reconstruction, the larger the non-linear partial volume effect will be. Partial volume effects are present at edges between materials and they are more pronounced when two neighboring materials have large differences in attenuation values, leading to dark and white streaks around the edges. While usually no disturbing artifacts are seen for soft tissue and bone transitions, partial volume effects are expected to be more important for transitions with metals, because metals differ substantially in attenuation compared to the surrounding tissue. When using 3 mm slices, the metal-tissue interface would be averaged over 3 mm, most probably introducing extra artifacts. A solution would be to decrease the slice thickness of the reconstruction. However, decreasing the slice thickness in FBP reconstructions often increases noise and the severeness of the metal artifacts. Iterative maximum likelihood reconstruction is less sensitive to noise and obtains higher image quality than FBP under noisy conditions. Nevertheless, unless for dedicated examinations, finer slice thickness is not required. Hence, reconstruction with finer slices, to reduce partial volume effects in the metals, would substantially increase the reconstruction time with no direct benefit for the evaluation of the nonmetal regions of the object or patient.

As explained in the previous section, the sampling model can be defined for each of the patches separately. While transaxial oversampling yielded only slightly reduced artifacts for two-dimensional examples [112], this might be different for the axial direction that is often sampled more coarsely than the transaxial direction. In this work the influence of increased sampling in the reconstruction volume (smaller voxels) and of the detector on the metal artifacts will be evaluated.

Cross talk The cross talk in CT detectors is relatively low in transverse direction, however it is higher in the axial direction (e.g. in [124]). Moreover, we have found (see figure 9.2) in a simulation experiment that the influence of cross talk is larger in the axial direction compared to transaxial direction.

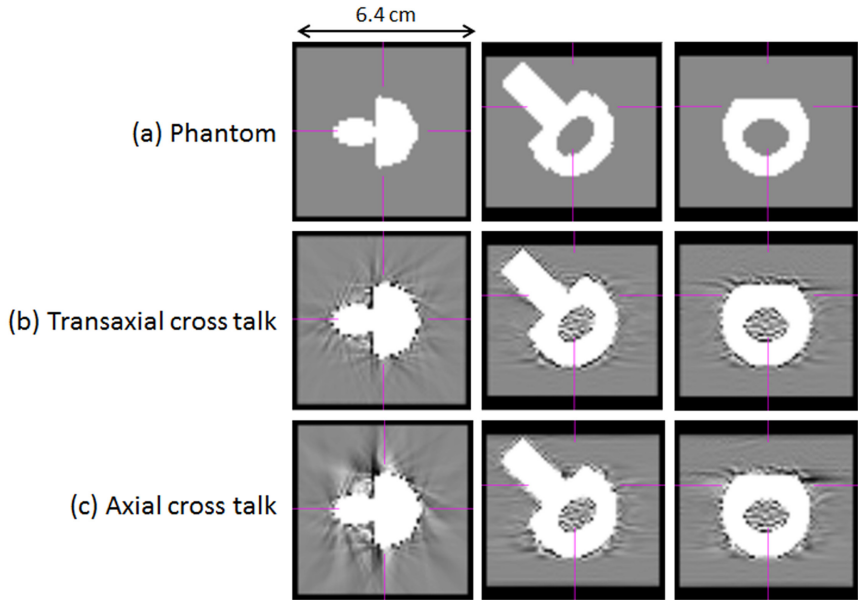


Figure 9.2: Effect of cross talk on the reconstruction. A polychromatic simulation of a spiral CT with pitch 1 at 120 kV with either transaxial cross talk or axial cross talk of 5%. The cross talk was not modeled in the reconstruction. (a) Phantom: 6.4 mm × 6.4 mm × 6.4 mm, represents head of a titanium hip implant surrounded by water. (b) IMPACT reconstruction for data with transaxial cross talk. (C) IMPACT reconstruction for data with axial cross talk. (Transverse, sagital and coronal view, W = 750HU, C = 0HU)

This is probably due to the fact that the axial direction is often more coarsely sampled.

Both MLTRC and IMPACT are extended with a convolution operation on the projection both in axial and transaxial direction to model cross talk effects[1]:

$$\hat{y}_i^C = \sum_n w_{in} \hat{y}_n \tag{9.8}$$

with $\hat{y}_n$ the combined projection model as given in Eq. (9.7) and w_{in} the cross talk between elements i and n. This is a sinogram operation that is applied on the total sinogram and cannot be changed for individual patches.

Scatter The influence of scatter in most clinical CT systems is limited by the use of an anti-scatter grid. This grid consists of very tin lamellae in between the detector elements. However, the grid is usually only present in transaxial

[1] A detailed derivation on the IMPACT and MLTRC reconstruction algorithm with a cross talk model is given in the appendix of this thesis.

direction. The detector elements in one detector slice are separated by these lamellae but there is no lamella in between the different detector slices. The field of view for most CT systems is relatively small in the axial direction and therefore the scatter contribution in axial direction is expected to be small. However, for metal projections with only few photons left, even very low scatter can become important and introduce artifacts.

The influence of axial scatter can be included in the projection model via r_i. The axial scatter will be simulated by a similar procedure as proposed by Ohnesorge et al. in [125]. It can be shown that the forward scatter is proportional to the product of the photons arriving at the detector and the attenuation along the projection line. In order to extend the estimated forward scatter to more oblique scatter to neighboring detector slices, the scatter estimate is smoothed in axial direction:

$$r_i = C \sum_i s_{in} \hat{y}_n \ln \left(\frac{b_n}{\hat{y}_n} \right) \tag{9.9}$$

where C is a factor including the fraction of the attenuation undergoing Compton scattering multiplied by the fraction of photons that are scattered under a small angle and s_{in} is a Gaussian smoothing weight.

Scatter factor C will be estimated during the reconstruction, the rest of the profile remains fixed. This can be done by optimizing the likelihood for C as proposed in [112]. The voxel updates and the update for C will be performed sequentially.

9.2.3 Acceleration and optimization

Initial image

In theory, iterative reconstruction algorithms like MLTRC and IMPACT should converge to a solution that is independent of the initial estimate for μ_j. Nevertheless, this initial estimate might have influence on the number of iterations needed to reach convergence.

In literature iterative reconstruction is often started from an FBP reconstruction [47]. Iterative reconstruction algorithms are slow and several iterations are needed to obtain diagnostic contrast and resolution. The FBP image already contains the main contrast elements, soft tissue and bone contrast, which reduces the number of iterations required for iterative reconstruction to obtain the desired image quality. However, in our experience, it is not the optimal starting image in the presence of metal artifacts because it requires many iterations to remove these artifacts.

In the phantom study described in [112], iterative reconstruction was started from a contour of the object filled with water attenuation. However, the phantom was relatively homogeneous. For the reconstruction of patient data we

propose to start the reconstruction from the result obtained by sinogram inpainting. No severe (dark) metal artifacts are present and contrast and resolution are relatively high for most regions. Unfortunately, metals might be poorly reconstructed for both FBP and inpainting methods. The metals in the inpainting methods are the same as in FBP since they are segmented from the FBP image and pasted on top op the inpainting result. The metals in FBP commonly have too low attenuation values and unsharp edges due to dark and white streak artifact at the edes of the metals.

In this work we will compare iterative reconstruction started from an FBP, NMAR and body contour image. To increase the quality of the metal reconstruction, we will introduce metal-only iterations. Within the patch structure it is possible to update the metal reconstruction without the need to update the whole image. The first five iterations for all reconstructions with patches will be metal-only, afterwards also the nonmetal patch will be updated.

Image blocks

An important requirement for introducing iterative reconstruction algorithms in clinical environments is limited computation time. Even when started from a sinogram inpainting result, it might take several iterations to improve contrast or to regain contrast that was lost during the interpolation procedure.

The reconstruction can be accelerated by the use of image blocks. This technique is related to the patchwork structure and is described in [120]. The aim is to improve the convergence rate by updating smaller image blocks separately and sequentially while all image blocks are updated with the same projection model. This image-block acceleration can be applied to the large nonmetal patch. As proposed in [120], division of the image is only applied in transaxial direction because image blocks in axial direction have low influence on the convergence. To be able to combine image blocks with ordered subsets, the size of each block should remain relatively large. By increasing the number of image blocks also the computational overhead related to the sequential update of the different regions can become nonnegligible. For balancing these two effects 16 image-blocks will be defined. The overhead for 16 image-blocks is still very low and the blocks are large enough to be combined with ordered subsets. According to the results obtained in [120], a four-fold acceleration is expected for 16 patches.

9.2.4 Measurements

Five patient data sets and one phantom data set are included in this study. All patients and the phantom have metallic implants. The phantom was added to the evaluation to provide a case for which the true object is known. A maximum intensity projection of each of the scans is given in figure 9.3. The

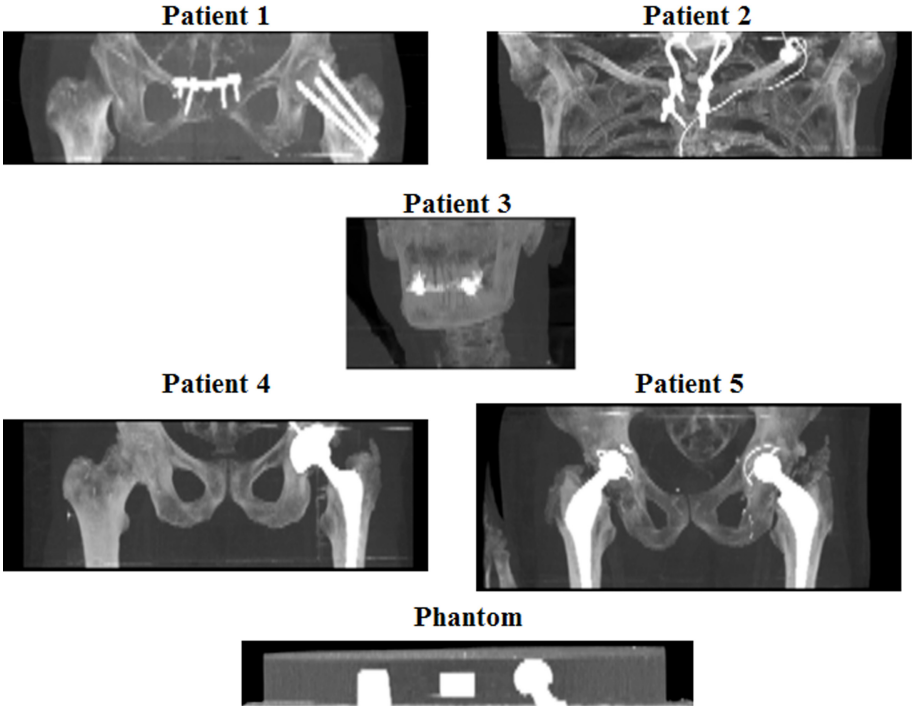


Figure 9.3: Overview of the different data sets via a maximum intensity projection.

scans are all performed with a Siemens Sensation 16 system. This systems has a curved detector with 16 rows of 672 detector elements and a field of view of 50 cm. The number of views per rotation is 1160.

Patient 1 Patient 1 is a patient with several metallic fixation screws in the left femoral head and the frontal pelvic region. The patient was scanned for a PET/CT examination on a Siemens Biograph 16 system. The CT part of this system is equivalent to a Siemens Sensation 16 system. A helical whole body CT with 16 $\times$ 1.50 mm detector collimation was performed at 120 kV with mean effective current 56 mA. The table feed was 18 mm per rotation (pitch factor = 0.75) with 0.5s per rotation. The CTDIw for this scan was 2.61 mGy.

Patient 2 Patient 2 has several metallic fixation screws in the neck region of the spine. The patient was scanned for a PET/CT examination on a Siemens Biograph 16 system. A helical whole body CT with 16 $\times$ 1.50 mm detector collimation was performed at 120 kV with mean effective current 71 mA. The

table feed was 18 mm per rotation (pitch factor = 0.75) with 0.5s per rotation. The CTDIw for this scan was 5.97 mGy.

Patient 3 Patient 3 has several metallic dental fillings. The patient was scanned for a PET/CT examination on a Siemens Biograph 16 system. A helical whole body CT with 16 × 1.50 mm detector collimation was performed at 120 kV with mean effective current 115 mA. The table feed was 18 mm per rotation (pitch factor = 0.75) with 0.5s per rotation. The CTDIw for this scan was 3.31 mGy.

Patient 4 Patient 4 has a unilateral femoral (hip) implant at the left side. The patient was scanned for an abdomen CT examination on a Siemens Sensation 16 system. A helical CT with 16 × 1.50 mm detector collimation was performed at 120 kV with mean effective current 380 mA. The table feed was 19.2 mm per rotation (pitch factor = 0.8) with 0.5s per rotation. The CTDIw for this scan was 16.63 mGy.

Patient 5 Patient 5 has bilateral femoral (hip) implants. The patient was scanned for an abdomen CT examination on a Siemens Sensation 16 system. A helical CT with 16 × 1.50 mm detector collimation was performed at 120 kV with mean effective current 350 mA. The table feed was 24 mm per rotation (pitch factor = 1.0) with 0.5s per rotation. The CTDIw for this scan was 12.25 mGy.

Phantom The phantom contains two femoral hip implants (one titanium based, left hand side in figure 9.3 and one cobalt-chromium based), a aluminum cylinder and several plastic cylinders placed close to the cobalt-chromium implant. The phantom was scanned on a Siemens Biograph 16 system. A helical CT with 16 × 0.75 mm detector collimation was performed at 120 kV with mean effective current 480 mA. The table feed was 18 mm per rotation (pitch factor = 1.50) with 0.5s per rotation. The CTDIw for this scan was 12.48 mGy.

9.2.5 Reconstruction

All data sets have been reconstructed with FBP, sinogram inpainting (NMAR) and several iterative reconstruction schemes.

FBP and NMAR

Our own FBP reconstruction software is based on two-dimensional parallel beam rebinning which might suffer from (cone beam) artifacts. The dedi-

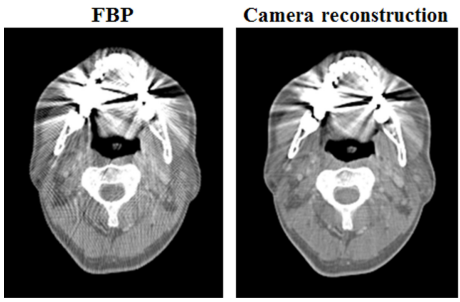


Figure 9.4: Comparison of the in-house FBP reconstruction and the dedicated camera reconstruction.

cated software from the camera uses more complex rebinning algorithms as described in Stierstorfer et al. [126] and Flohr et al. [127]. Moreover, our algorithm performs basic filtered back projection with ramp filter while dedicated, examination specific reconstruction smoothing kernels are included in the vendor's software. However, the trend and severeness of the metal artifacts are comparable for both algorithms as in shown in figure 9.4.

The sinogram inpainting method used in this work is NMAR developed by Meyer et al in [68]. This technique follows the typical inpainting scheme, but includes information of the intersection lengths of a projection line through air, soft tissue and bone. This information is obtained based on the projection of a prior image. The prior image is created by segmentation of air, soft tissue, bone and metals from an FBP image or from an image obtained by standard inpainting. In the prior image, voxels containing air are set at -1000 HU, voxels with soft tissue or metal are set at 0 HU and voxels with bone keep the attenuation values of the initial image. The prior image is then forward projected. The resulting prior sinogram is used to normalize the measured sinogram before interpolation and to denormalize the interpolated sinogram after interpolation. The introduction of the normalization step improves contrast of bone and reduces secondary streaks that may arise due to interpolation [67]. Extended versions of this technique, FSMAR [70] and ANMAR [72], try to regain contrast by adding high frequency information from the original reconstruction or sinogram. However, while regaining contrast also some streaks and shadows are reintroduced. We prefer to use NMAR for this work.

The raw data obtained from the camera are subjected (by the vendor) to certain correction procedures such as beam hardening water correction and bowtie correction. The raw data were kept unchanged for both FBP and NMAR reconstruction.

Iterative reconstruction

Different iterative reconstruction schemes were tested on patient data. An overview of the different iterative reconstruction schemes is given in table 9.1.

The standard reconstruction scheme is COMBOS$_{STA}$. The name COMBOS is used for algorithms combining IMPACT in the metal patches and MLTRC in the nonmetal patch. COMBOS$_{STA}$ starts from the NMAR reconstruction, cross talk is modeled and the voxels size for the metal patches is 1mm x 1mm x 1mm while the nonmetal patch maintains the traditional anisotropic voxels 1mm x 1mm x 3mm. No subsampling of the detector elements is applied in this standard reconstruction scheme. The algorithm starts with 4 metal-only iteration (with 116 subsets), where only metal patches are updated, followed by 5 iterations with 116 subsets where both metal and nonmetal patches are updated. No acceleration with image blocks is applied.

The algorithm includes many parameters that might have influence on the reduction of artifacts. The standard set of parameter was chosen such that it gives good results under most circumstances. The influence of each of the parameters is tested by changing each of them separately. A detailed list of the parameters for each of the algorithms is given in table 9.1.

Scatter simulation was not included in the algorithms in table 9.1. A separate test with scatter modeling in COMBOS$_{STA}$ was applied. Also the acceleration with image-blocks was evaluated in a different experiment.

An exact list of materials can be used for IMPACT, including the specific implant materials. However, this information is seldom available (patients are often operated in another hospital or the composition of the implant is not available in their file). In our experience, IMPACT is not very sensitive to the exact materials list, therefore a generic material list was used containing: water, bone, titanium and cobalt-chrome. Titanium and cobalt-chrome are popular implant materials. Another popular implant material is stainless steel but its attenuation is very close to cobalt-chrome such that it was not necessary to include it in the list.

Previous experiments demonstrated that the beam hardening and bowtie pre-correction applied on the raw data obtained from the camera, may interfere with the polychromatic model of IMPACT and MLTRC and cause artifacts. In order to be able to use the MLTRC and IMPACT model, we could undo the beam hardening and bowtie correction with the aid of the vendor of our CT systems. The bowtie filter is taken into account by including its influence in b_{ik} in Eq. (9.4) and Eq. (9.5).

Table 9.1: The parameters for the different iterative reconstruction schemes. (mod.= modeling, det.res.= detector resolution) No convergence acceleration or scatter correction is applied unless stated differently.

Name	Patch	Model	Start image	Voxel size[mm]	Cross talk mod.	Axial det.res.[mm]
IMPACT	no patches	IMPACT	NMAR	1×1×3	on	1.5
PIMPACT	metal	IMPACT	NMAR	1×1×1	on	1.5
	nonmetal	IMPACT	NMAR	1×1×3	on	1.5
COMBOS_{STA}	metal	IMPACT	NMAR	1×1×1	on	1.5
	nonmetal	MLTRC	NMAR	1×1×3	on	1.5
COMBOS_{UNI}	metal	IMPACT	UNI	1×1×1	on	1.5
	nonmetal	MLTRC	UNI	1×1×3	on	1.5
COMBOS_{FBP}	metal	IMPACT	FBP	1×1×1	on	1.5
	nonmetal	MLTRC	FBP	1×1×3	on	1.5
COMBOS_{3mm}	metal	IMPACT	NMAR	1×1×3	on	1.5
	nonmetal	MLTRC	NMAR	1×1×3	on	1.5
$\text{COMBOS}_{0.5mm}$	metal	IMPACT	NMAR	1×1×0.5	on	1.5
	nonmetal	MLTRC	NMAR	1×1×3	on	1.5
$\text{COMBOS}_{0.5/1mm}$	metal	IMPACT	NMAR	0.5×0.5×1	on	1.5
	nonmetal	MLTRC	NMAR	1×1×3	on	1.5
COMBOS_{noCRO}	metal	IMPACT	NMAR	1×1×1	off	1.5
	nonmetal	MLTRC	NMAR	1×1×3	off	1.5
COMBOS_{detSUB}	metal	IMPACT	NMAR	1×1×1	on	0.5
	nonmetal	MLTRC	NMAR	1×1×3	on	0.5

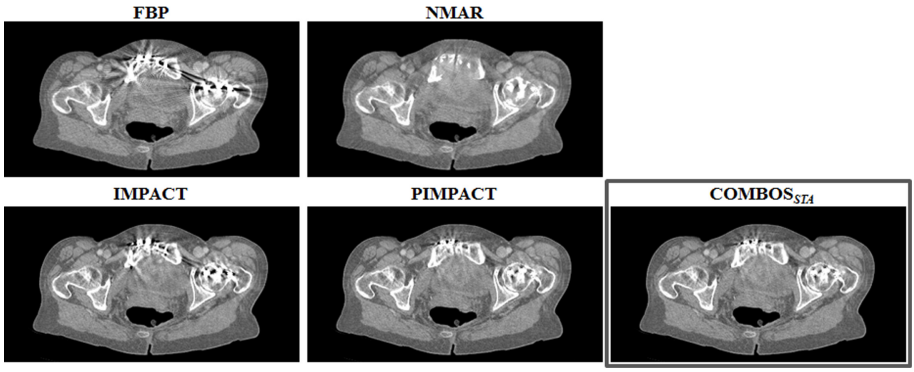


Figure 9.5: Evaluation of the influence of the patch structure on the metal artifacts for patient 1. The patched reconstructions PIMPACT and COMBOS$_{STA}$ are compared to the FBP, NMAR and IMPACT. (W = 750HU, C = 0HU)

9.3 Results

This section is composed of two different parts. In section 9.3.1 the influence of the different reconstruction parameters is evaluated. Section 9.3.2 gives an overview of the reconstructions with the optimal parameter COMBOS$_{STA}$ set deduced from section 9.3.1.

9.3.1 Evaluation of the reconstruction parameters

The section starts with the evaluation of the different reconstruction parameters. This is demonstrated by using one or two patients as an example. The result for the other patient data sets are equivalent.

Patch structure

Figure 9.5 shows the FBP, NMAR, IMPACT, PIMPACT and COMBOS$_{STA}$ reconstructions for patient 1. The FBP image has obvious steaks and shadows in between the metallic objects. The NMAR reconstruction has no remaining streaks or shadows but the frontal pelvis bone is almost completely erased. Regular IMPACT reconstruction has still severe artifacts, hardly reduced compared to FBP. The introduction of the patch structure improves the image quality substantially. However, in the region close to the metal, remaining artifacts are present. Replacing IMPACT by the MLTRC for the nonmetal patch does not affect the image quality. These results confirm the results we obtained in two dimensions in [112].

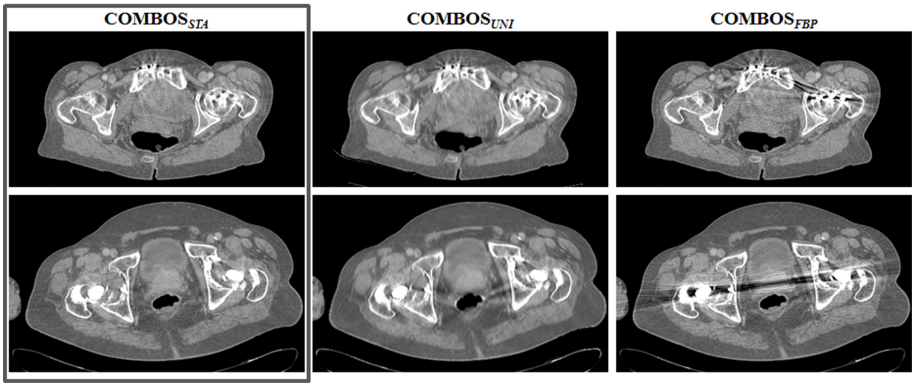


Figure 9.6: Effect of the initial image on the results of COMBOS. Top row: patient 1, bottom row: patient 5. Left : started from NMAR, middle: started from a uniform body contour, right: started from FBP. (W = 750HU, C = 0HU)

Initial image

The influence of the initial image on the result of the patched algorithms is shown in figure 9.6 for patient 1 and patient 5. The initial image clearly influences the reconstruction results. Reconstructions started from FBP have still obvious artifacts that are hardly reduced compared to FBP. The reconstructions started from a uniform image have lower contrast than the image started from NMAR. Some streak artifacts are more pronounced for COMBOS$_{UNI}$ compared to COMBOS$_{STA}$. This is more obvious for patient 5 (bottom row) where metallic structures are larger.

Resolution and sampling

Several resolution and sampling parameter can be changed in the reconstruction: the voxel size can be changed (both transaxial and axial), the detector can be oversampled and cross talk can be modeled. These parameters are evaluated in figure 9.7.

The upper row of the figure evaluates different slice thicknesses: 3mm, 1mm and 0.5 mm. The slice thickness of the metal patches has influence on the severity of the artifacts. In case of 3 mm slices in the metal patch (COMBOS$_{3mm}$) more artifacts are observed compared to 1mm slices (COMBOS$_{STA}$). Further reducing the slice thickness to 0.5 mm does not further reduce artifacts (COMBOS$_{0.5mm}$). When the voxel size in the transaxial direction is lowered (COMBOS$_{0.5/1mm}$), also no further improvement in image quality is obtained. Using oversampled detector elements during the reconstruction had also no obvious effects on the reduction of the metal artifacts (COMBOS$_{detSUB}$).

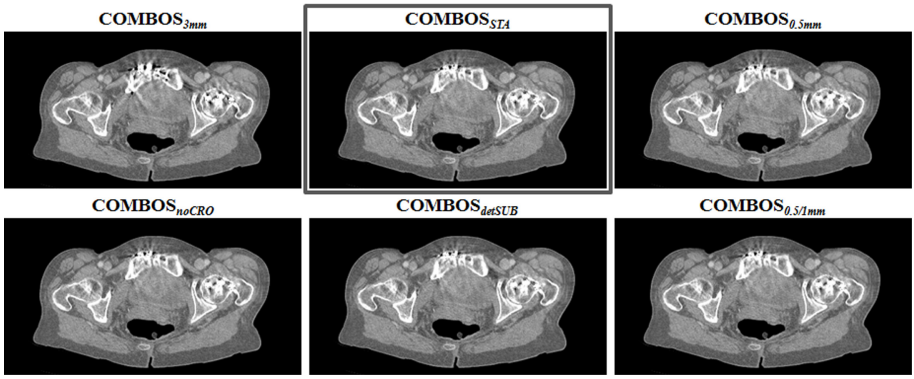


Figure 9.7: Evaluation of the different resolution parameters for COMBOS reconstruction for patient 1. Upper row: different slice thicknesses, from left to right: 3mm, 1mm and 0.5 mm. Lower row, left: no cross talk is modeled, middle: detector oversampling was applied, right: slice thickness 1 mm and tranaxial voxel size 0.5 mm. (W = 750HU, C = 0HU)

Unlike expected from the initial simulation experiment, modeling of cross talk does not improve the image quality substantially. The attenuation values in the metals are better (higher) when cross talk is modeled and when subtracting the images (not shown), it can be seen that the remaining artifacts are slightly less pronounced. A possible explanation that we do not see the effects as they were observed in the simulations could be that the raw data are in some way already corrected for cross talk by the vendor.

Scatter

Figure 9.8 shows that modeling the scatter has a negligible effect on the image quality of the reconstruction. The artifacts have changed but did not improve. The delineation of the metals is somewhat better when scatter is modeled.

Acceleration

Figure 9.9 demonstrates the result of image-blocks to accelerate the convergence in the nonmetal patch for patient 2 and compares it to regular reconstruction. It is clear that with this acceleration less iterations are needed to obtain the same contrast. The contrast for 5 accelerated iterations is again substantially improved, although with increased noise.

The reconstruction after 1 and 5 iterations for COMBOS$_{STA}$ with and without image-block acceleration are shown. The contrast in the nonmetal improves

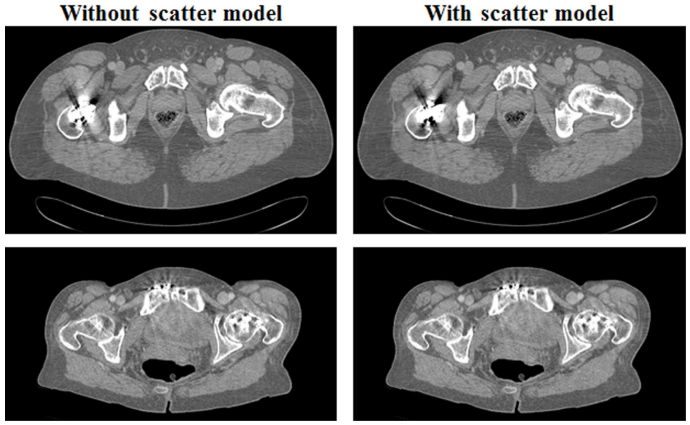


Figure 9.8: Evaluation of the influence of scatter modeling on the metal artifacts. Top row: patient 5, bottom row: patient 1. (W = 750HU, C = 0HU)

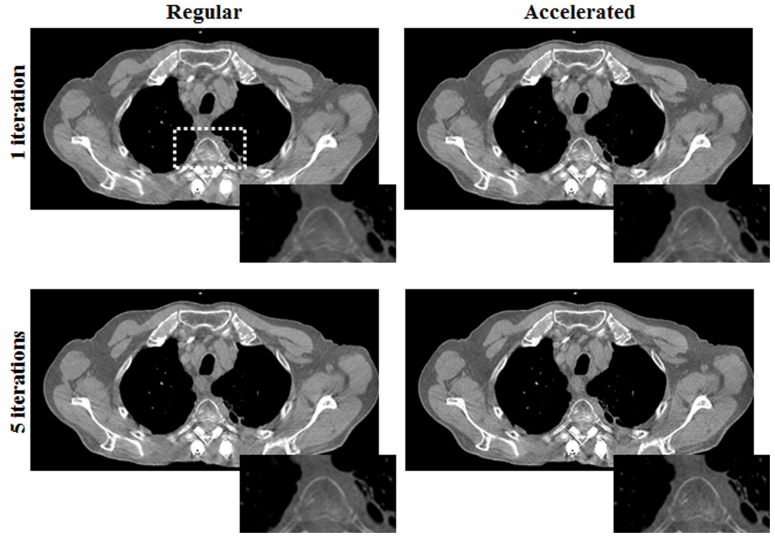


Figure 9.9: Effect of using image blocks in the nonmetal patch. One and five iteration for COMBOS$_{STA}$ without image-blocks: 'regular' and with image blocks: 'accelerated' are shown for patient 2. An insert is given that magnifies the region indicated by the box in the top left image. (Main images: W = 750HU, C = 0HU, inserts W = 2000, C = 350.)

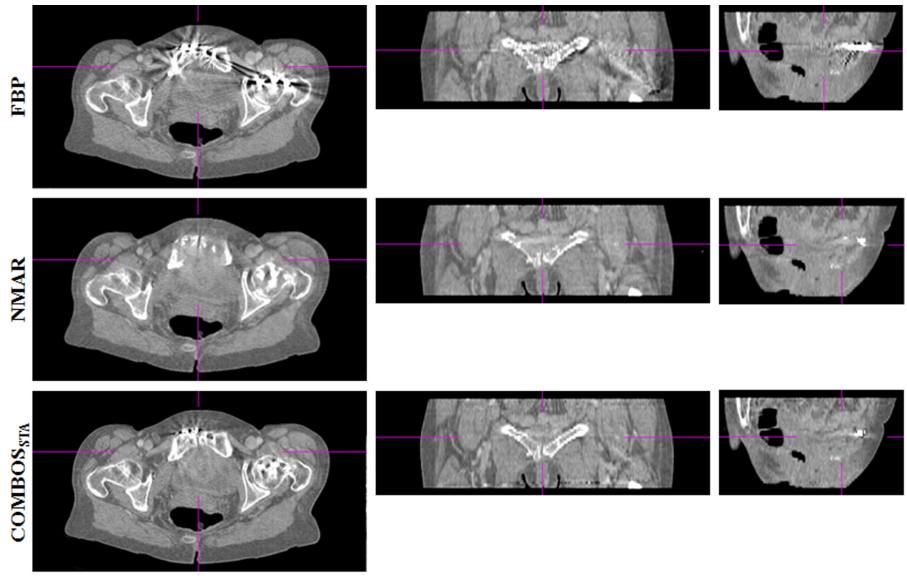


Figure 9.10: **Patient 1** Reconstruction with FBP, NMAR and COMBOS$_{STA}$: transverse, coronal and sagital view.(W = 750HU, C = 0HU)

with the iterations. The contrast achieved after 5 regular iterations is comparable to 1 iteration with image blocks. However, some of the metal artifacts are more pronounced when the accelerated scheme is used. This could be because the algorithm gives again more weight to the convergence of the background compared to the convergence of the metal regions. Care should be taken when using this accelerated scheme, perhaps more metal-only iterations are required.

9.3.2 Overview reconstructions of measured data

The reconstructions for all patient data sets and for the phantom data set with COMBOS$_{STA}$ are shown in figures 9.10-9.16. COMBOS$_{STA}$ is started from a NMAR reconstruction, and uses a decreased slice width of 1 mm for the metal patches and cross talk is modeled.

Patient 1 and patient 2 Patient 1 and 2 have many rather small metallic objects. Bone structures are almost erased for NMAR in the frontal pelvic region for patient 1 and at the end of spine for patient 2. COMBOS$_{STA}$ shows reduced artifacts compared to FBP and improved contrast compared to NMAR. The frontal pelvis bone is almost completely recovered for patient 1 and the end of the spine in between two of the fixations screws is again visible.

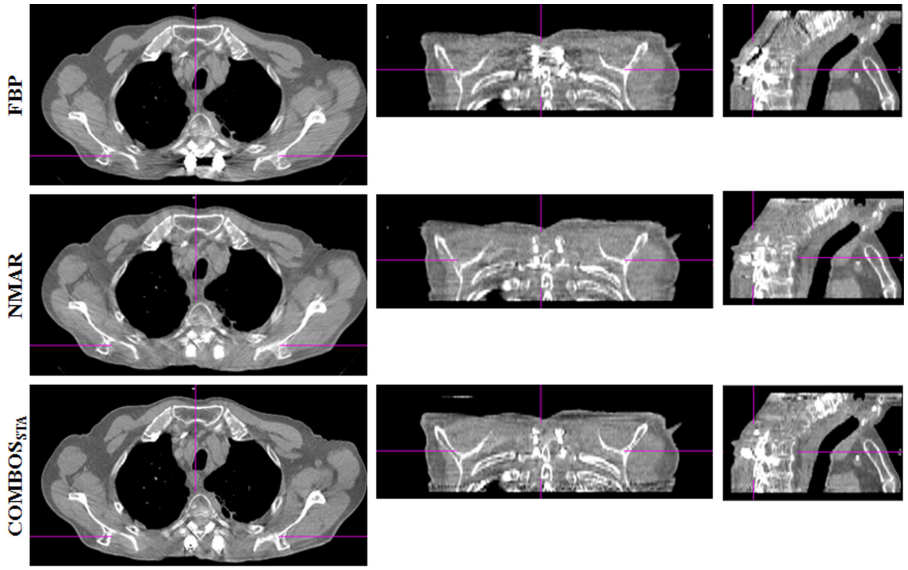


Figure 9.11: **Patient 2** Reconstruction with FBP, NMAR and COMBOS$_{STA}$: transverse, coronal and sagital view.(W = 750HU, C = 0HU)

Patient 3 Patient 3 is displayed at two different intensity scales. In figure 9.12 a typical dental window is used. COMBOS$_{STA}$ reduces the artifacts compared to FBP but has some remaining shadows. NMAR has no streaks and shadows but quite some dental contrast is lost. In figure 9.13 a window for evaluation of soft tissue contrast is used. Although COMBOS$_{STA}$ performs the best, it is still not possible to examine soft tissue close to the dental fillings.

Patient 4 and patient 5 Patients 4 and 5 both have femoral implants. The different reconstructions are shown in figure 9.14 and 9.15. The image quality for both data sets is very similar. However, the region evaluated in both cases is different. For patient 4, the reconstruction is evaluated in a region where the shape of the implant varies strongly along the longitudinal axis. COMBOS$_{STA}$ reduces the artifacts compared to FBP but has remaining artifacts. Figure 9.15 evaluates patient 5 in a region with much lower axial variation in the implant. In this region COMBOS$_{STA}$ reduces most of the artifacts. NMAR suffers only from very limited contrast lost for both data sets.

Phantom The result for the phantom with femoral implants is similar to those for the patients. Here, COMBOS$_{STA}$ performs better in retrieving the contrast of the plastic cylinders close to the implant.

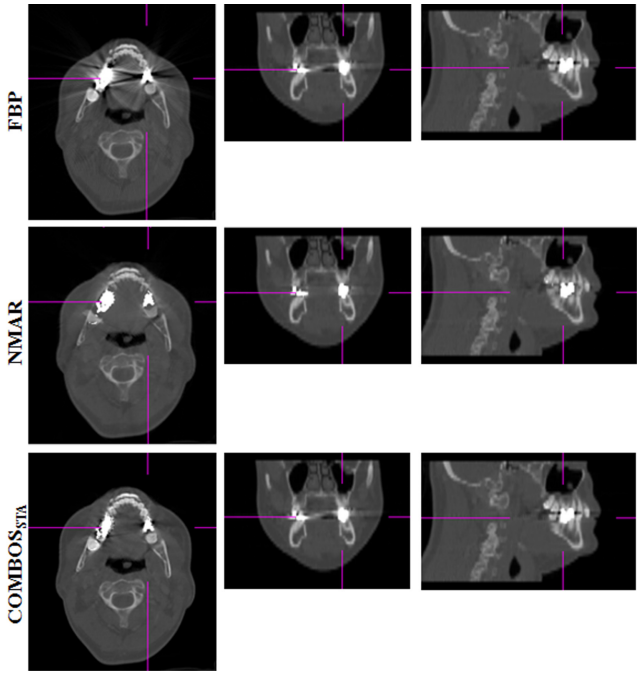


Figure 9.12: **Patient 3** Reconstruction with FBP, NMAR and COMBOS$_{STA}$: transverse, coronal and sagital view. Dental window. (W = 3000HU, C = 1000HU)

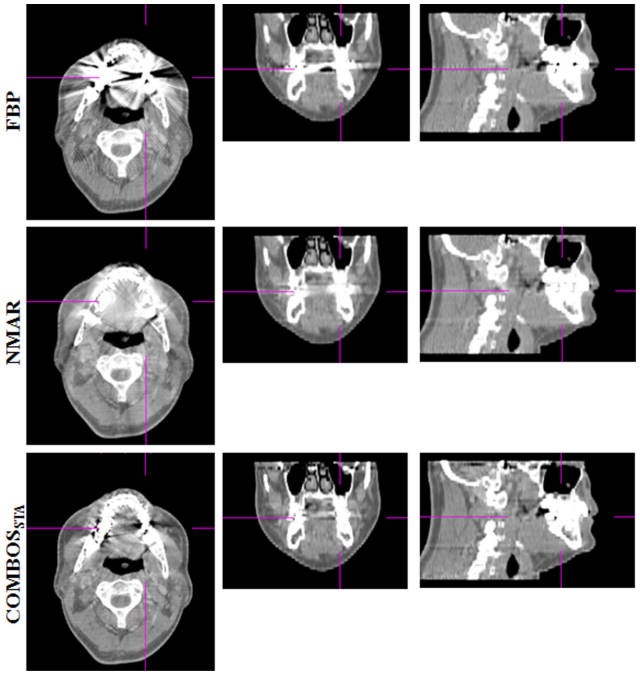


Figure 9.13: **Patient 3** Reconstruction with FBP, NMAR and COMBOS$_{STA}$: transverse, coronal and sagital view. Soft tissue window.(W = 750HU, C = 0HU)

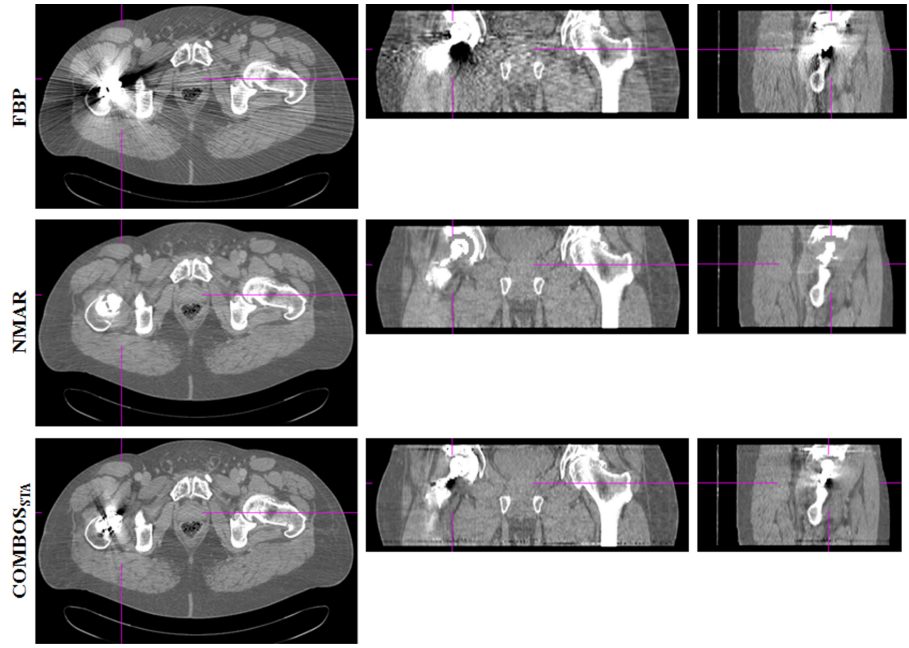


Figure 9.14: **Patient 4** Reconstruction with FBP, NMAR and COMBOS$_{STA}$: transverse, coronal and sagital view.(W = 750HU, C = 0HU)

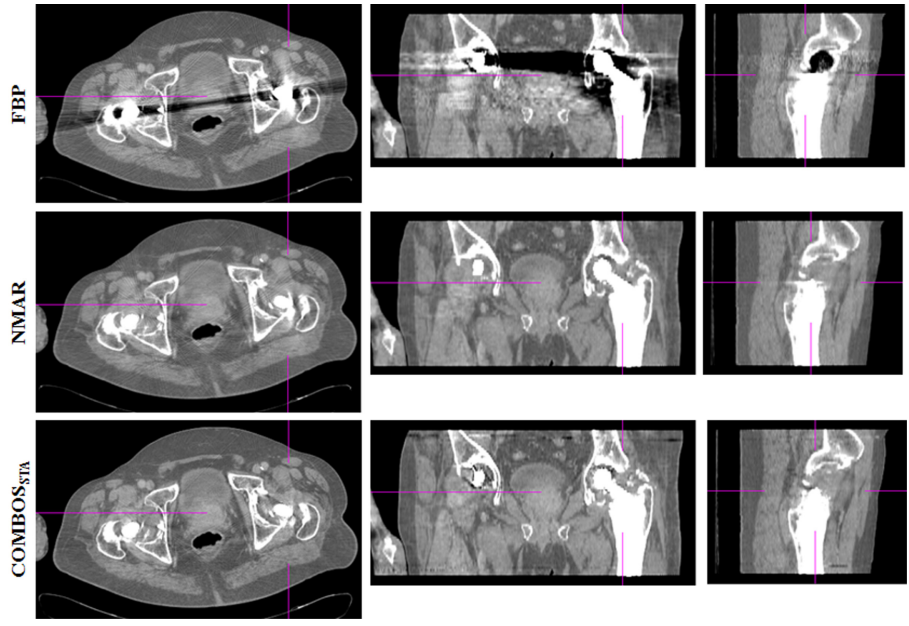


Figure 9.15: **Patient 5** Reconstruction with FBP, NMAR and COMBOS$_{STA}$: transverse, coronal and sagital view.(W = 750HU, C = 0HU)

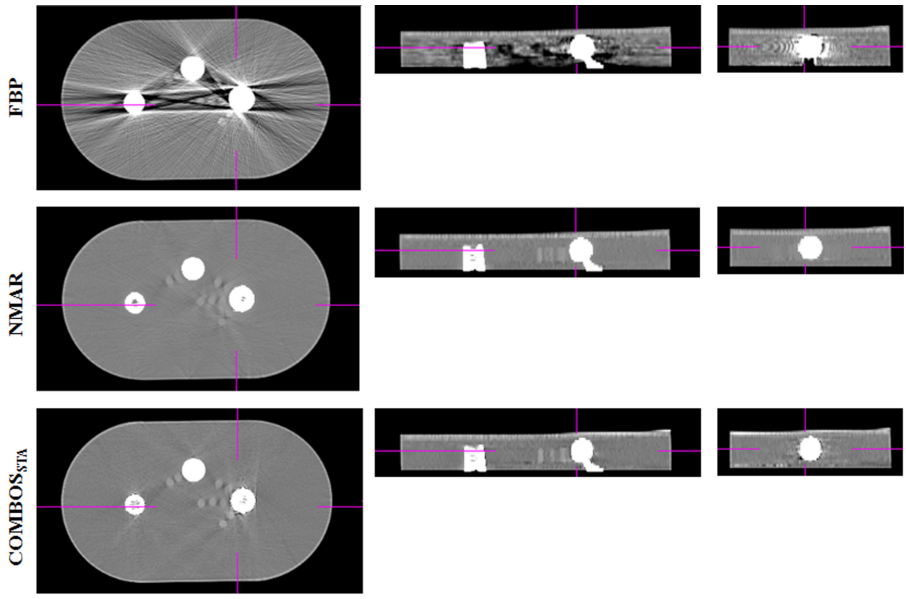


Figure 9.16: **Phantom** Reconstruction with FBP, NMAR and COMBOS$_{STA}$: transverse, coronal and sagital view.(W = 750HU, C = 0HU)

9.4 Discussion

This study presents a metal artifact reduction method based on the patch structure applied on three-dimensional helical CT patient data. It is based on the patchwork structure for local modeling and improved convergence that was developed and evaluated in Van Slambrouck and Nuyts [112] for two-dimensional CT data.

The patch structure enables the use of different reconstruction models in different regions of the reconstruction volume for iterative reconstruction. Reducing metal artifacts requires more precise modeling of the acquisition physics during the reconstruction. More precise models are usually more complex which increases the computation time of the algorithm substantially compared to the use of a simpler model. In [112] we have shown that a fully polychromatic energy model for the metal patches combined with a simpler polychromatic model for the nonmetal patch does not decrease the image quality. Moreover, when the different patches receive a separate update, the convergence rate of the metals is increased which improves the artifact reduction considerably. These effects were also observed for the patient study. An obvious reduction in artifacts compared to FBP and standard iterative reconstruction was obtained. The degree of artifact reduction is influenced by several parameters in the algorithm.

The initial image clearly affects the image quality of the iterative reconstruction. Eventually, the final result of an iterative algorithm should be independent on the initial image. However, the initial image partly determines the number of iterations needed to obtain sufficiently high image quality both in terms of resolution and artifact reduction. FBP should be avoided as initial image, because streak artifacts in the initial image are hard to remove. Starting from a uniform image requires more iterations to obtain sufficient contrast and for larger metallic objects some extra artifacts might be present compared to the use of the NMAR result as initial image.

Figure 9.9 shows clearly that, for the same number of iterations, the use of image-blocks results in improved contrast. The expected gain in computation time to obtain similar image quality is expected to be around the square root of the number of image blocks that are used, at least in the proposed configuration [120]. This acceleration scheme was not used for the overview of the patient data but might be useful for introduction of iterative methods in clinical practice since it reduces the required computation time substantially.

The standard voxel size as used by the camera is 1mm $\times$ 1mm $\times$ 3mm. This is sufficiently small to minimize partial volume effects in transaxial direction but not in axial direction. Decreasing the slice thickness for the whole reconstruction volume prolongs the computation time and results in images with denser axial sampling than required for most examinations. The patch structure enables the use of a smaller slice thickness in the metal patches while keeping the standard thickness elsewhere. As shown in figure 9.7 artifacts are much less pronounced when the metal slice thickness is decreased to 1 mm. Nevertheless, the resulting images are not free of artifacts. Remaining artifacts are mainly present in regions where the metal shape shows relativity large variation in the longitudinal direction. This might suggest that other resolution effects are the origin of the remaining artifacts. Hence, the slice thickness was further decreased, transaxial sampling was increased, oversampled detector slices were used and cross talk was modeled (figure 9.7). None of these model refinements reduced the remaining artifacts substantially.

Another cause for remaining artifacts might be scatter. In the two-dimensional study [112], scatter was expected to be very low due to the small collimation and the anti-scatter grid that is present in between the detector elements. For three-dimensional scanning, collimations are broader and multiple detector slices, not separated by anti-scatter grids, are illuminated simultaneously. A typical scatter profile is relatively smooth but a sudden increase in scatter might appear at the edge of a metallic structure. Since the remaining artifacts are present at the location where the metal shape varies, scatter is a plausible cause for these artifacts. Unfortunately, including a scatter profile in the projection model did not reduce the artifacts (see figure 9.8). Although the scatter profile was only based on a first order model, it is not very likely that more accurate scatter models would give better results since the approximate model had negligible effect.

The effect causing the remaining artifacts is not clear. In this study no regularizing operation was applied on the reconstruction. This was done on purpose to evaluated the inherent capacity for reducing metal artifacts by modeling the acquisition physics accurately. However, it is impossible to model all processes and some of the remaining artifacts might be due to subtle effects in the acquisition. In De Man et al. [92], Markov Random Field priors were used to reduce artifacts in a maximum a posteriori (MAP) reconstruction algorithm with a monochromatic reconstruction model. The patch structure can be easily extended to a MAP algorithm. Moreover, when using the patch structure, the prior properties can be changed for the different patches. Since most priors tend to smooth edges, it is probably useful to lower the prior strength in the metal patches or to turn the prior off.

9.5 Conclusion

An iterative metal artifact reduction method with local models was applied for the reconstruction of helical CT (patient) data. The method enables the use of more precise acquisition models in and around metals and simpler models elsewhere. Besides the use of a fully polychromatic model also sufficiently small voxels are required in the metal regions. Especially in the axial direction, voxels need to be smaller than typical standard settings. The initial image for the iterative algorithms has strong influence on the reconstruction results.

The proposed method has better contrast in bone and soft tissue regions compared to sinogram inpainting results and artifacts are strongly reduced. However, some remaining artifacts are present, mostly in regions where the metal shape varies strongly in the longitudinal direction. The origin of these artifacts is unclear.

Acknowledgments

The authors would like to thank Prof. F. Noo from the University of Utah, School of Medicine, Salt Lake City, Utah, USA for his valuable suggestions.

Chapter 10

Bias reduction for low-statistics PET: Maximum likelihood reconstruction with modified Poisson distribution

Katrien Van Slambrouck, Simon Stute, Claude Comtat, Merence Sibomana, Floris van Velden, Ronald Boellaard and Johan Nuyts

A paper based on this chapter has been submitted to IEEE Transactions on Medical Imaging

Abstract PET data are typically reconstructed with maximum likelihood reconstruction methods. However, these methods suffer from positive bias due to the nonnegativity constraint in the reconstruction and the strong asymmetry of the Poisson distribution near zero. This is particularly problematic for tracer kinetic modeling analysis of dynamic PET studies.

A method with modified likelihood is presented. The Poisson distribution, that is typically used, is modified by replacing the original function by a Gaussian distribution for low count data points. This modified likelihood function is optimized by a gradient ascent approach. The resulting algorithm is an extension of the so-called NEGML algorithm. Using two-dimensional simulations, the algorithm was evaluated and compared to the AML, a modification of the ABML-method proposed by Byrne, which has also been used to mitigate the bias problem. Furthermore, also the influence of different methods for randoms

handling have been investigated.

Both NEGML and AML have a single parameter that determines the effectiveness of the bias reduction. When this parameter is large in magnitude, bias-free images are obtained. The parameter for NEGML determines the transition point between the Gaussian and Poisson models. Two-dimensional simulations indicate that the value of this parameter can be determined independent of the activity in the object and of the geometry of the PET system. On the other hand, the results indicate that the parameter in AML is best selected dependent upon the activity in the reconstructed image.

It was also observed that the randoms handling has an important influence on the bias in reconstructions with MLEM and NEGML and AML with smaller parameters. Reconstruction with smoothed randoms results in lower bias compared to reconstruction with unsmoothed randoms or randoms precorrected data.

10.1 Introduction

Nowadays PET reconstruction is mainly done by applying iterative reconstruction methods. Iterative reconstruction is based on a forward model, which offers the possibility to model the true acquisition process better, e.g. by incorporation of finite resolution. Iterative reconstruction in PET is usually based on a maximum likelihood (ML) approach, to take into account the Poisson statistics of the measured data. To suppress the noise propagation, the likelihood can be combined with a prior that favors smooth reconstructions [49].

The most popular iterative ML method for PET reconstruction is maximum likelihood expectation maximization (MLEM [45, 128]). The main draw-back of MLEM is that it creates positive bias in low activity regions surrounded by regions with higher activity and for low count data, i.e. for extremely noisy data [129]. This makes that for kinetic PET studies, analytical methods, like filtered backprojection (FBP), are still the method of choice despite the fact that the images have often lower resolution and more streak artifacts due to noise [130, 131]. Dynamic PET data have often very limited numbers of counts that are sparsely divided over the lines of response. This is due to the fact that the early frames are often very short and for late frames only limited activity might be left due to the decay of the activity [131].

In every iteration of the MLEM algorithm, the current reconstruction image $\vec{\lambda}$ is updated by adding the image $\Delta\vec{\lambda}$ which is given for voxel j by:

$$\Delta\lambda_j = \frac{\lambda_j}{\sum_i c_{ij}} \sum_i c_{ij} \frac{y_i - \hat{y}_i}{\hat{y}_i} \tag{10.1}$$

with

$$\hat{y}_i = \sum_i c_{ij}\lambda_j + r_i \tag{10.2}$$

where y_i are the measured counts for detector pair i, $\hat{y}_i$ is the estimate of the measurement based on the current reconstruction $\vec{\lambda}$, c_{ij} is the sensitivity of detector pair i for activity in voxel j, and r_i is the number of scattered and/or random events.

There are several origins for the bias created by MLEM. The reconstruction formula in Eq. (10.1) shows that the update of a voxel is proportional to λ_j, the current activity estimate for voxel j. This makes that regions with low activity will converge much slower and a very high number of iterations is required to eliminate the positive bias. In practice one is not iterating long enough to avoid this incomplete convergence bias. Since there are other causes of bias, iterating longer by itself would not make the image bias-free.

Another cause is the inherent non-negativity constraint in image space. This makes that creating a region with zero mean is only possible by making all voxels zero since no negative voxel values are allowed. Under noisy circumstances MLEM will always have some remaining positive bias in low-activity regions. FBP reconstruction has no constraints on the image values and has usually both positive and negative values in a cold region.

In case of very noisy data (i.e. with very low counts), MLEM not only introduces bias in cold regions but also in regions with higher activity. Experiments with other maximum likelihood (ML) algorithms that allow negative reconstruction values (e.g.[105]) still show positive noise induced bias. We believe that this noise induced bias is a feature of the ML-solution itself. It is known that maximum likelihood estimators are only asymptotically unbiased, which means that they are only unbiased for an infinite number of counts [35, 132]. We believe this is due to the asymmetry of the Poisson distribution which becomes very strong for low counts and which also imposes a nonnegativity constraint in sinogram space. (Note that the Poisson (log)-likelihood is simply not defined for negative sinogram values). This effect causes, even at convergence, positive bias for low count measurements.

ABML proposed by Byrne in [106] can be used for bias reduction. The reconstruction values in ABML are constrained by lower bound A and upper bound B. By setting A to negative values, negative values in the image and sinogram domain are allowed, resulting in bias reduction behavior [133, 134]. The convergence of different regions is still dependent on the activity but to a lesser extent. Since reconstruction values are usually not known beforehand, upper bound B is often chosen very high and A is often set to a very low value (i.e. a negative value with high magnitude). It was not evaluated whether less extreme values for A would have influence on bias reduction or convergence.

In Nuyts et al. [105], the NEGML algorithm was proposed. It was originally developed to obtain images with a higher diagnostic value for reconstruction without attenuation correction. NEGML allows negative values in the image domain. Moreover, the convergence of the different image parts was not dependent on λ_j but a uniform weight was used for all voxels. These two charac-

teristics make that the algorithm could also be used to reduce bias, especially in cold regions surrounded by warm regions [135]. It was less successful on very low count data, moreover, its effectiveness seemed to be dependent on the implementation. The NEGML algorithm has a safety value that prevents dividing by zero. In Nuyts et al. [105] it was proposed to apply a lower limit of 1 on the denominator of the update formula. In other implementations a much smaller value was used, seemingly leading to different bias reduction capacities [134].

In this work the NEGML algorithm is extended such that it also allows for negative values in the sinogram domain. This is obtained by replacing the Poisson distribution by a symmetric distribution, the Gaussian distribution, for very small sinogram values. The influence of this new NEGML on bias reduction and convergence will be evaluated and compared to ABML.

10.2 Methods

10.2.1 NEGML

Since the Poisson distribution is in theory the correct distribution, we prefer to use it where possible and switch to another symmetric distribution for small sinogram values. The most obvious choice for a distribution which is close to Poisson but symmetric around zero is a Gaussian distribution.

The original Poisson (log-)likelihood as function of the activity is

$$\begin{aligned} L(\vec{\lambda}) &= \sum_i L_i(\vec{\lambda}) && (10.3)\\ &= \sum_i y_i \ln \hat{y}_i - \hat{y}_i. && (10.4) \end{aligned}$$

Extended with a Gaussian part, the newly proposed likelihood becomes:

$$L(\vec{\lambda}) = \sum_i L_i^{\psi}(\vec{\lambda}) \tag{10.5}$$

with

$$\begin{aligned} L_i^{\psi}(\lambda) &= y_i \ln \hat{y}_i - \hat{y}_i && \text{when } \hat{y}_i \geq \psi \\ &= -\tfrac{(y_i-\hat{y}_i)^2}{2\psi} + y_i \ln \psi - \psi + \tfrac{(y_i-\psi)^2}{2\psi} && \text{when } \hat{y}_i < \psi \end{aligned} \tag{10.6}$$

where ψ defines the point where the Poisson distribution switches to a Gaussian distribution. The last three terms in Eq. (10.6) ensure that the transition is continuous. Figure 10.1 plots the original and new likelihood with $\psi = 4$ and $\psi = 16$ for different values of y_i.

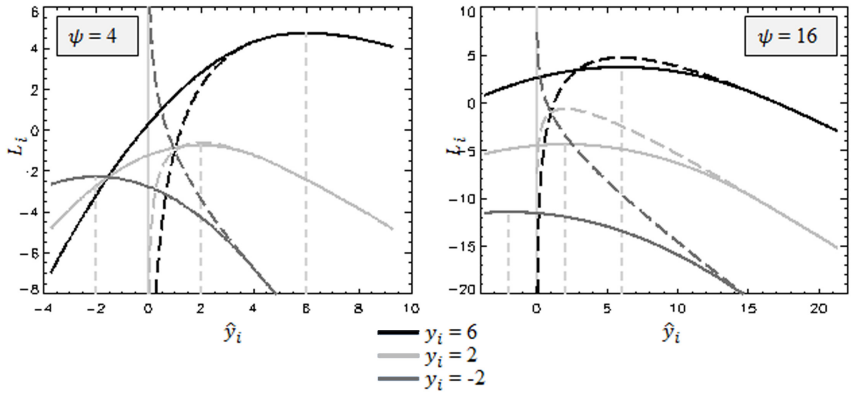


Figure 10.1: The standard likelihood function (dashed line) and the modified likelihood (solid line) for $y_i = 6$ (black), $y_i = 2$ (light grey) and $y_i = -2$ (dark grey). Left: $\psi = 4$, right: $\psi = 16$.

The NEGML algorithm is based on a gradient ascent approach, explained in more details in [46, 105, 120]. A similar idea is used by Ahn and Fessler in [136] where a surrogate for the likelihood that also extends to negative sinogram values is used to optimize the likelihood. The creation of a surrogate in the NEGML algorithm starts by approximating the likelihood as a second order truncated series expansion:

$$L(\vec{\lambda} + D(\alpha)\Delta\vec{x}) \simeq T_1(\vec{\lambda}, \Delta\vec{x})$$

$$= L(\vec{\lambda}) + \sum_j \frac{\partial L}{\partial \lambda_j}\alpha_j \Delta x_j + \frac{1}{2}\sum_{j,k} \frac{\partial^2 L}{\partial \lambda_j \lambda_k}\alpha_j\alpha_k \Delta x_j \Delta x_k \qquad (10.7)$$

$$\geq L(\vec{\lambda}) + \sum_j \frac{\partial L}{\partial \lambda_j}\alpha_j \Delta x_j + \frac{1}{2}\sum_{j,k} \frac{\partial^2 L}{\partial \lambda_j \lambda_k}\alpha_j\alpha_k (\Delta x_j)^2 \qquad (10.8)$$

where $D(\alpha)$ is a diagonal matrix with $D_{jj} = \alpha_j \geqslant 0$. To obtain the inequality in Eq. (10.8) we used the fact that $2\Delta x_j \Delta x_k \leq (\Delta x_j)^2 + (\Delta x_k)^2$ and that the second derivatives are negative. The function (10.8) is a surrogate function for T_1 since it coincides with T_1 in the current solution and is smaller than T_1 elsewhere. This function is maximized, yielding an update for λ_j, $\Delta\lambda = \alpha_j \Delta x$:

$$\Delta\lambda_j = -\frac{\alpha_j \frac{\partial L}{\partial \lambda_j}\Big|_{\vec{\lambda}}}{\sum_k \alpha_k \frac{\partial^2 L}{\partial \lambda_j \partial \lambda_k}\Big|_{\vec{\lambda}}} \qquad (10.9)$$

$$\begin{aligned}
\frac{\partial L_i^{\psi}(\lambda)}{\partial \lambda_j} &= c_{ij}\frac{y_i-\hat{y}_i}{\hat{y}_i} \quad \text{when} \quad \hat{y}_i \geq \psi \\
&= c_{ij}\frac{y_i-\hat{y}_i}{\psi} \quad \text{when} \quad \hat{y}_i < \psi \\
\frac{\partial^2 L_i^{\psi}(\lambda)}{\partial \lambda_j \partial \lambda_k} &= -c_{ij}c_{ik}\frac{y_i}{\hat{y}_i^2} \quad \text{when} \quad \hat{y}_i \geq \psi \\
&= -c_{ij}c_{ik}\frac{1}{\psi} \quad \text{when} \quad \hat{y}_i < \psi
\end{aligned} \tag{10.10}$$

The update can be simplified by assuming that $y_i \approx \hat{y}_i$ for the second derivatives in the denominator. This also ensures that the denominator is continuous and strictly positive near $\hat{y}_i = \psi$, which is not the case when the second derivatives of (10.10) are inserted in (10.9). The update becomes:

$$\Delta\lambda_j = \frac{\alpha_j \sum_i c_{ij}\frac{y_i-\hat{y}_i}{\max(\hat{y}_i,\psi)}}{\sum_i c_{ij}(\sum_k c_{ik}\alpha_k)\frac{1}{\max(\hat{y}_i,\psi)}}. \tag{10.11}$$

The original NEGML algorithm is obtained when setting $\psi = 1$ and $\alpha_j = 1$ and by constraining $\hat{y}_i$ such that $\hat{y}_i \geq 0$. Both ψ and α_j were no explicit parameters in the original NEGML work. Instead of using ψ, dividing by zero was avoided by restricting the denominator to strictly positive values, larger or equal to 1. Since $\hat{y}_i$ is a number of photons, $\hat{y}_i \geq 1$ was chosen as a reasonable value and gave good results. In other implementations of NEGML, this restriction was sometimes set at much smaller values, e.g. $\hat{y}_i \geq 10^{-4}$ [134]. The results of different implementations were different regarding bias reduction, this is probably caused by the different restrictions for the denominator. The experiments in this work will explicitly test the influence of ψ on convergence and bias values.

When only using the Poisson likelihood, i.e. $\hat{y}_i \geq \psi$ for all i, and by using $\alpha_j = \lambda_j$ and $r_i = 0$, the update becomes the MLEM update. This is an indication that the approximation made for the denominator and the truncated series expansion of the likelihood in Eq. (10.8) is fairly accurate.

The original NEGML algorithm used a mixed update step to improve the convergence in high activity regions. The update obtained for MLEM and NEGML were both calculated for voxel j and the largest of both was applied. This mixed update is equivalent to over-relaxation and might cause convergence problems. Hence, we prefer to use a pure NEGML update.

The convergence of high activity regions could be improved by altering the values for α_j. This parameter was introduced in our work on metal artifact reduction and iterative reconstruction in CT [112, 120] and serves as a voxel convergence weight during reconstruction. Similar weights are used in the grouped coordinate algorithm by Fessler et al. [47]. Choosing $\alpha_j = \lambda_j$ gives the weighting used in MLEM, choosing $\alpha_j = 1$ results in the weighting of the original NEGML algorithm. A compromise could be obtained by defining weights that have a weaker dependence on the activity but still assign higher weights to high activity regions.

NEGML can be accelerated with ordered subsets, similarly as proposed for MLEM by Hudson and Larkin in [48].

10.2.2 AML

The ABML algorithm presented by C. Byrne [106] allows to perform an MLEM-like reconstruction between an upper B and lower A bound. The reconstruction formula is given by:

$$\lambda_j^{\text{new}} = \frac{B_j P_j + A_j Q_j}{P_j + Q_j} \tag{10.12}$$

with

$$\begin{aligned} P_j &= \frac{(\lambda_j - A_j)}{\sum_i c_{ij}} \sum_i c_{ij} \frac{y_i - \sum_k c_{ik} A_k}{\hat{y}_i - \sum_k c_{ik} A_k} \\ Q_j &= \frac{(B_j - \lambda_j)}{\sum_i c_{ij}} \sum_i c_{ij} \frac{\sum_k c_{ik} B_k - y_i}{\sum_k c_{ik} B_k - \hat{y}_i} \end{aligned} \tag{10.13}$$

ABML can be used for bias reduction by setting A_j to negative values. This has been shown to reduce bias substantially [133, 134]. Often a single value for A_j is chosen, i.e. $A_j = A$ for all voxels. This allows reconstruction values with lower limit A and sinogram values $\hat{y}_i$ with lower limit $A \sum_k c_{ik}$. Since upper bound B_j is not expected to have influence on the bias, it is usually set to a value much larger than the expected maximum activity. By introducing some approximations ABML can be extended to an ordered subsets version [133] with inclusion of randoms [134].

Since very large values are usually chosen for B, we propose to use ABML with infinite upper bound, referred to as AML. The AML update step for λ_j, with uniform A, in additive form becomes:

$$\Delta\lambda_j = \frac{\lambda_j - A}{\sum_i c_{ij}} \left(\sum_i c_{ij} \frac{y_i - \hat{y}_i}{\hat{y}_i - A \sum_k c_{ik}} \right). \tag{10.14}$$

This update step can be considered as an update where the image and sinogram are shifted, with A and $A \sum_k c_{ik}$ respectively, for calculating the update step. This makes that the Poisson distribution is evaluated at higher values where it is less asymmetric. Before adding this to the current reconstruction the update is shifted back. Note that choosing $A = 0$ results in the original MLEM algorithm. AML has still an activity dependent convergence weighting: $(\lambda_j - A_j)$ for voxel j, however the weights become more uniform when $|A|$ becomes larger[1].

A is defined in image space. It is not always straightforward to determine bounds in image space. Therefore, A is usually set to a relatively extreme value such that it is low enough to avoid any bias caused by constraints on negative values.

[1] All values for A are negative in this chapter. To avoid confusion when discussing the magnitude of (negative values for) A, the absolute value will often be used.

10.2.3 Unweighted least squares and extreme values for ψ and A

The influence of the parameter ψ and A on the reconstructions of NEGML and AML will be studied in this paper. The behavior of NEGML and AML for extremely high values for ψ and extremely low values of A will be studied as well. For large ψ, all sinogram pixels will be in the Gaussian regime. It is therefore expected that NEGML will become more and more similar to unweighted least squares fitting. The same holds for AML. When $|A|$ is large, the shift from A will be so large that the difference in Poisson weighting will be very small, eliminating all weighting difference in practice.

The NEGML algorithm with $\psi \to \infty$ becomes:

$$\Delta\lambda_j = \frac{\sum_i c_{ij}(y_i - \hat{y}_i)}{\sum_i c_{ij}(\sum_k c_{ik})}. \tag{10.15}$$

When all measured sinogram values y_i are below ψ, NEGML will always use the update in Eq. (10.15).

The AML algorithm with $A \to -\infty$ yields:

$$\Delta\lambda_j = \frac{1}{\sum_i c_{ij}} \sum_i c_{ij} \frac{(y_i - \hat{y}_i)}{\sum_k c_{ik}}. \tag{10.16}$$

Unlike for NEGML this formula is theoretically only valid when $A = -\infty$ since all finite values for A will still follow the Poisson distribution at $y_i - A\sum_h c_{ih}$. However, this distribution becomes very flat for large $|A|$.

For comparison, the typical least squares algorithm has the following form:

$$\Delta\lambda_j = \frac{\rho}{\sum_i c_{ij}} \sum_i c_{ij}(y_i - \hat{y}_i) \tag{10.17}$$

with ρ a relaxation factor. The results in Eq.(10.15) and (10.16) resemble the least squares update, but are not identical to it. When $\sum_k c_{ik}$ is constant, both algorithms yield the least squares expression.

10.2.4 Additive contamination

The additive contamination r_i can be modeled as in Eq.(10.2). This is often referred to as ordinary Poisson, because it assumes that the randoms term r_i is not noisy such that the Poisson model for y_i is preserved. Since the estimated randoms are often also subjected to Poisson noise, a shifted Poisson model better preserves the photon statistics [137]. In shifted Poisson the following projection model is used:

$$\begin{array}{rcl} y_i & \to & y_i + s_i \\ r_i & = & 2s_i \end{array} \tag{10.18}$$

with r_i the randoms estimate used during reconstruction and s_i the original noisy randoms estimate. Ordinary Poisson is valid when an estimate of the randoms with low noise is available. This can be obtained by smoothing the randoms estimate [29] or by calculating them from the singles [28].

The different ways to include the randoms in the reconstruction are applied for the different reconstruction algorithms. For FBP, the randoms were always subtracted, with or without randoms smoothing. In this work a uniform randoms estimate will be used (see section 10.3) and randoms smoothing will be performed with a Gaussian function. More dedicated smoothing is applied on true PET systems [29]. For the iterative algorithms the following reconstructions were performed, where s_i is the noisy estimate of the randoms and s_i^S is a smoothed (noise-free) randoms estimate:

- Randoms precorrection with randoms smoothing: $y_i \rightarrow y_i - s_i^S$
- Regular Poisson with randoms smoothing: y_i and $r_i = s_i^S$
- Regular Poisson without randoms smoothing: y_i and $r_i = s_i$
- Shifted Poisson without randoms smoothing: $y_i \rightarrow y_i + s_i$ and $r_i = 2s_i$

10.3 Experiments

This section describes the experiments that are performed to evaluate the bias reduction capacity of NEGML compared to FBP, MLEM and AML. The experiments are based on two-dimensional simulations and consider the influence of the parameter ψ and A on the bias, variance, convergence and noise characteristics of NEGML and AML respectively for two different phantoms simulated with different settings.

10.3.1 Phantom 1

Simulations

A two-dimensional phantom with a cold, warm and hot region, as shown in figure 10.2, is simulated. The activity in the image in activity per voxel is: 0 for the cold region, 1 for the warm region and 4 for the hot region. The phantom is 46 cm x 46 cm. During simulation the image pixel grid was 920 x 920. A parallel beam simulation is performed with 920 lines of response per projection and 200 projections. Uniform water attenuation is applied. A uniform randoms contribution was simulated, assuming a Poisson distribution with the expectation equal to the mean of the uncontaminated sinogram. The system resolution was modeled by a Gaussian with FWHM of 5 mm.

The obtained sinogram is rebinned to 230 lines of response per projection and 200 projections by adding the four neighboring projections values. This means

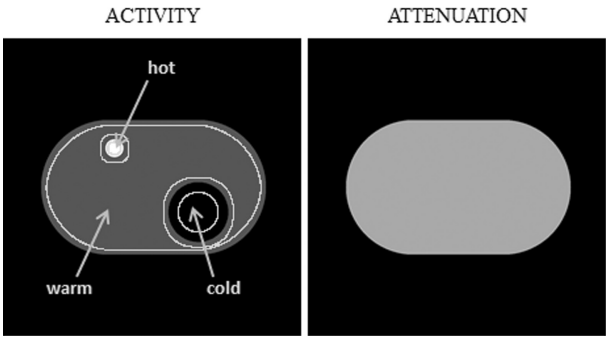


Figure 10.2: **Phantom 1** The phantom has a cold, warm and hot region. The activity for the main experiment was 0 for the cold region, 1 for the warm region and 4 for the hot region. In a second setting, the phantom's activity was multiplied by three. The phantom is assumed to have uniform water attenuation.

that the simulation was four times oversampled. This simulation was repeated for 60 noise realizations and 50 time frames of increasing duration, resulting in data sets with mean count per sinogram pixel between 0.05 and 1000 counts. In a second simulation setting for this phantom, the activity in the image was multiplied by three.

Reconstructions

Reconstructions were performed with FBP, MLEM, NEGML and AML on a 230×230 pixel grid, with pixel size 2mm $\times$ 2 mm. For NEGML different values of ψ were evaluated: $\psi = \{1, 4, 9, 16, 25, 100, 100000\}$. For AML different values of A where used: $A = \{-1, -5, -10, -50, -100, -1000, -100000\}$. For all iterative methods 200 iterations were applied without the use of ordered subsets. During reconstruction the resolution was modeled by a Gaussian with FWHM 4 mm, expressing a small mismatch with the resolution of the simulation.

The simulations were reconstructed with three different ways of randoms handling (see section 10.2.4): ordinary Poisson with randoms smoothing, shifted Poisson without randoms smoothing and precorrection without randoms smoothing. For randoms smoothing a Gaussian kernel with FWHM of 5 pixels has been used.

Evaluation

A cold, warm and hot region of interest (ROI) are defined in the phantom, as shown in figure 10.2. The bias is estimated by calculating the mean value in a

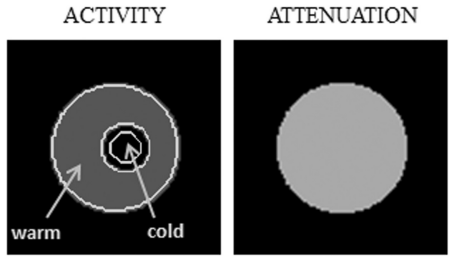


Figure 10.3: **Phantom 2** The phantom consists of a warm and cold region. The activity and attenuation are shown. The phantom is assumed to have uniform water attenuation.

ROI over the different noise realizations, the ROI mean, defined by:

$$\text{ROI Mean} = \frac{1}{NJ_{\text{ROI}}} \sum_{n}^{N} \sum_{j \in \text{ROI}} \lambda_{j,n} \tag{10.19}$$

with N the number of noise realizations and J_{ROI} the total number of pixels in the region. The ROI mean values are calculated for all 50 frames. The variance of the mean value over the noise realizations, the ROI Var, is calculated by:

$$\text{ROI Var} = \frac{1}{N} \sum_{n} (\text{ROI mean} - \sum_{j \in \text{ROI}} \lambda_{j,n})^2 \tag{10.20}$$

10.3.2 Phantom 2

This experiment is designed to further investigate whether the findings for phantom 1 and for the configuration described in the previous section, remain valid for other projections configurations. A different phantom with one cold and one warm region and different projection settings are used for this experiment.

Simulations

A simple two-dimensional phantom consisting of a uniform disk with one cold region, depicted in figure 10.3, is simulated. The activity in the warm region is set to 1, in the cold region it is set to 0. It is a 20 cm $\times$ 20 cm phantom represented by a 400 $\times$ 400 pixel grid.

To evaluate the dependence on the sinogram pixel values this phantom was simulated with three different settings. For setting 1, 100 projection angles were calculated, 500 for setting 2 and 1000 for setting 3. However, the total number of counts for each simulation was kept constant, i.e., the sensitivity of

the simulated PET system was 5 times lower for setting 2 and 10 times lower for setting 3.

A parallel beam simulation with 400 lines of response per projection, was performed, assuming that the phantom was a uniform attenuator (consisting of water). A uniform randoms estimate was set equal to the mean of the uncontaminated sinogram. This simulated estimate was subjected to Poisson noise. The simulation was rebinned for reconstruction to 100 projection lines per angle.

For all three settings, 100 noise realizations were performed for 20 time frames of increasing duration. Only relatively short time frames were considered with mean sinogram count between 0.05 and 10 counts per bin.

Reconstruction

The simulated measurements were reconstructed on a 100 x 100 pixel grid (2 mm isotropic pixel size) with FBP, MLEM, NEGML ($\psi = \{1, 4, 16\}$) and AML ($A = \{-1, -5, -50\}$). Precorretion of the randoms was used for FBP, Ordinary Poisson with smoothed randoms (FWHM = 5 pixels) was applied for the iterative methods. Iterative reconstruction was performed for 20 iterations with 10 subsets.

Evaluation

A warm and a cold ROI are defined as shown in figure 10.3. The ROI mean is calculated as in Eq. (10.19). Note that although all three settings have the same total amount of activity in the image and the sinogram, the mean per sinogram pixel is larger for settings with fewer projection views.

10.4 Results

10.4.1 Phantom 1

This section describes the main results of this work for the phantom depicted in figure 10.2. The ROI mean with respect to the frame duration is shown in figure 10.4 where the frame duration is represented by the mean number of sinogram counts in the frame. The standard error on the mean is given in figure 10.5.

$$\text{Error on the mean} = \frac{1}{\sqrt{N-1}} \sqrt{\frac{1}{N} \sum_{n} (\text{ROI mean} - \sum_{j \in \text{ROI}} \lambda_{j,n})^2} \quad (10.21)$$

The results for ordinary Poisson, shifted Poisson and randoms precorrection are shown. The upper row of the figure evaluates different values for ψ in NEGML, in comparison with FBP and MLEM. The lower row gives the evaluation of AML for different values of A.

Ordinary Poisson has the lowest bias for all methods. In general MLEM has the highest bias, especially in the cold and warm region. The bias of NEGML and AML is dependent on the model parameter. For NEGML all methods perform similarly. However, when the mean sinogram count has a value around ψ, a local increase in the bias can be observed. This is the most obvious in case of shifted Poisson. This local increase becomes less pronounced as ψ increases. AML is dependent on A for bias reduction, the bias decreases with increasing $|A|$. This effect is the most pronounced in case of shifted Poisson.

In the hot region (figure 10.4), even at relatively high counts, some remaining bias is observed for NEGML and for AML with large $|A|$. This bias reduces when the number of iterations increases (not shown). For less noisy data, MLEM performs in general better than both NEGML and AML in the high activity region.

Figure 10.6 depicts the ROI Mean as a function of the frame duration for the cold region of the same phantom with threefold increased activity reconstructed with Ordinary Poisson. In comparison with the phantom with lower activity, AML has more bias which increases for smaller $|A|$. The results for NEGML are very similar to the case with less activity in the image. The reconstructions remain almost bias free. The difference between AML and NEGML for increased activity is consistent with the fact that we defined A in the image domain and ψ in the sinogram domain. The influence of increased activity on the bias in the warm and hot region was less pronounced (not shown).

The reconstructed images for the different methods with ordinary Poisson for on average 5 counts per sinogram pixel are shown in figure 10.7. After reconstruction, the images were smoothed with a Gaussian filter with FWHM of 4 mm. The maximum number of counts in this sinogram is 19 with more than 95% of the sinogram pixels below 16 counts. The typical noise-induced streaks are the most pronounced in the FBP reconstruction. MLEM suffers the least from these streaks. NEGML and AML have only few streaks for small ψ and small $|A|$. The streaks become more pronounced with increasing ψ or increasing $|A|$. AML has voxel dependent convergence weighting which makes that it has more MLEM-like characteristics. The hot region is therefore sharper in MLEM and AML with small $|A|$. However with respect to noise streaks there is no obvious difference observed between bias-free NEGML and AML images, only that AML has somewhat more streaks in the background of the image.

The influence of the parameter ψ or A on the ROI Mean and ROI Var is shown in figure 10.8. The frames represented in this figure have on average 1 or 5 counts per sinogram pixel. The variance in the cold and warm region is lower for NEGML and AML compared to FBP but still higher than for MLEM. For

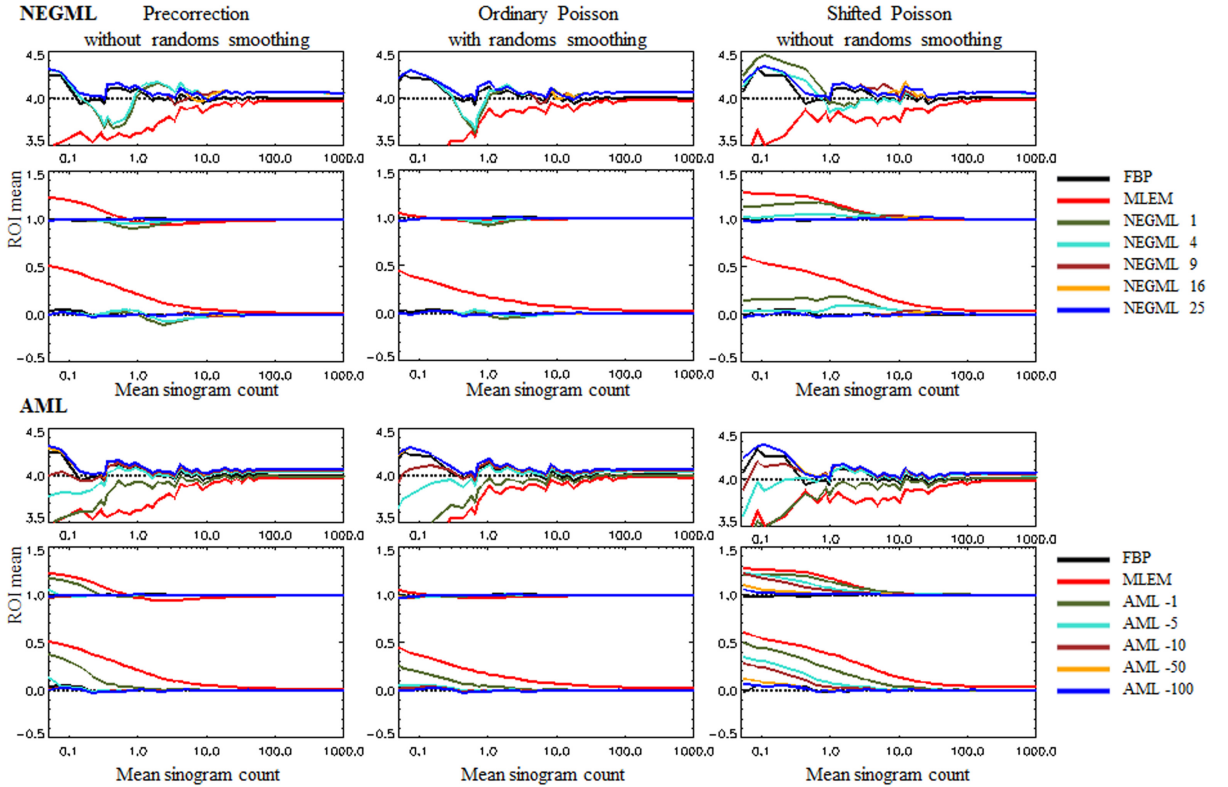


Figure 10.4: **Phantom 1** Evaluation of the ROI Mean for different frame durations. The frame duration is represented by the mean number of counts per sinogram pixel in the frame. Three different ways for randoms handling are shown. The upper row gives the results for NEGML, the lower row shows the results for AML.

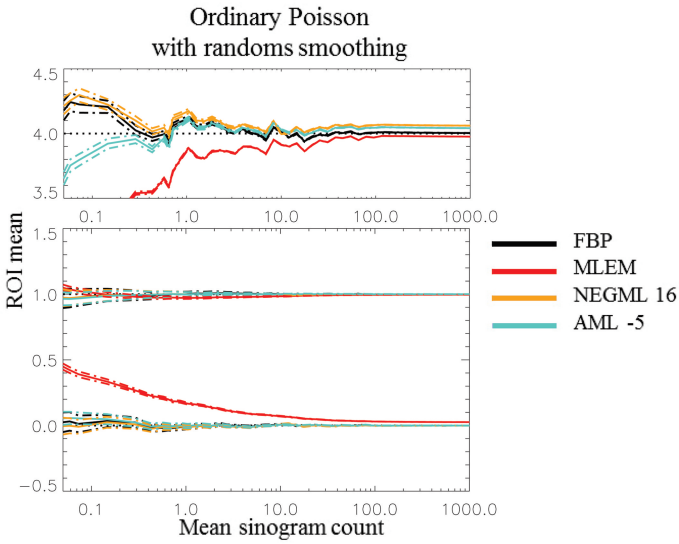


Figure 10.5: **Phantom 1** The standard error on the mean values in the plot in figure 10.4 is given for FBP, MLEM, NEGML 16 and AML -5. The solid line is the mean value, the dashed lines are the mean ± the standard error on the mean.

most parameter choices, the bias in NEGML and AML is lower than for MLEM in the cold and warm regions. Note that the mean sinogram count equal to 1 is close to ψ for NEGML1 and NEGML4, which explains the increased bias for these parameters (compare to figure 10.4). The bias due to incomplete convergence in the hot region, for NEGML and AML with large negative A, was expected from previous results.

The influence of ψ and A on the convergence of NEGML and AML for all three regions is given in figure 10.9, for a measurement with on average 5 counts per sinogram pixel. The result is shown for ordinary Poisson with randoms smoothing and for parameters: $\psi = \{4, 25\}$ and $A = \{-1, -10, -10^5\}$. For this frame duration NEGML with $\psi = 25$ worked completely in the Gaussian regime. The convergence for the cold and warm region is relatively fast. At 50 iterations all methods converged. In the hot region, MLEM and AML with small negative $|A|$ converge faster than NEGML and AML with larger $|A|$. An experiment with more iterations (not shown) indicates that most of the algorithms were not yet converged in the hot region, another 100 to 200 iterations are needed to obtain complete convergence and a hot region without bias.

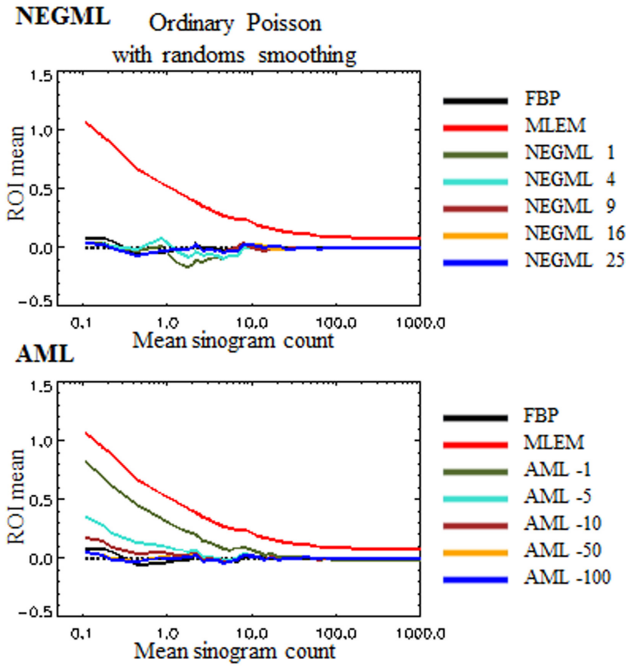


Figure 10.6: **Phantom 1 with threefold increased activity** Evaluation of the ROI Mean for the cold regions for different frame durations with ordinary Poisson. The frame duration is represented by the mean number of counts per sinogram pixel in the frame. The upper row gives the results for NEGML, the lower row shows the results for AML.

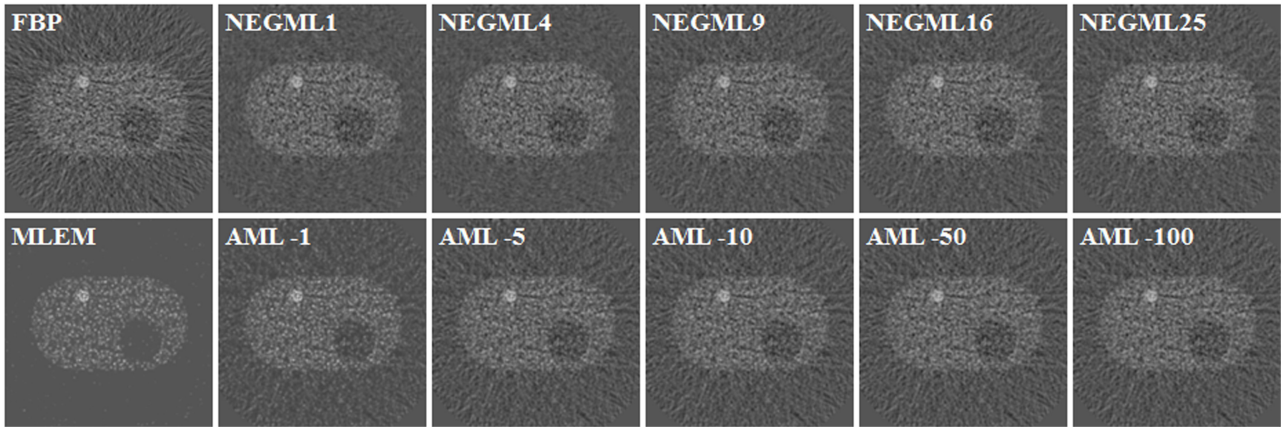


Figure 10.7: **Phantom 1** FBP, MLEM, NEGML ($\psi = \{1, 4, 9, 16, 25\}$) and AML reconstruction ($A = \{-1, -5, -10, -50, -100\}$). Ordinary Poisson with randoms smoothing for a frame with mean sinogram count 5. The reconstructions are postsmoothed with a Gaussian kernel of FWHM 4mm.

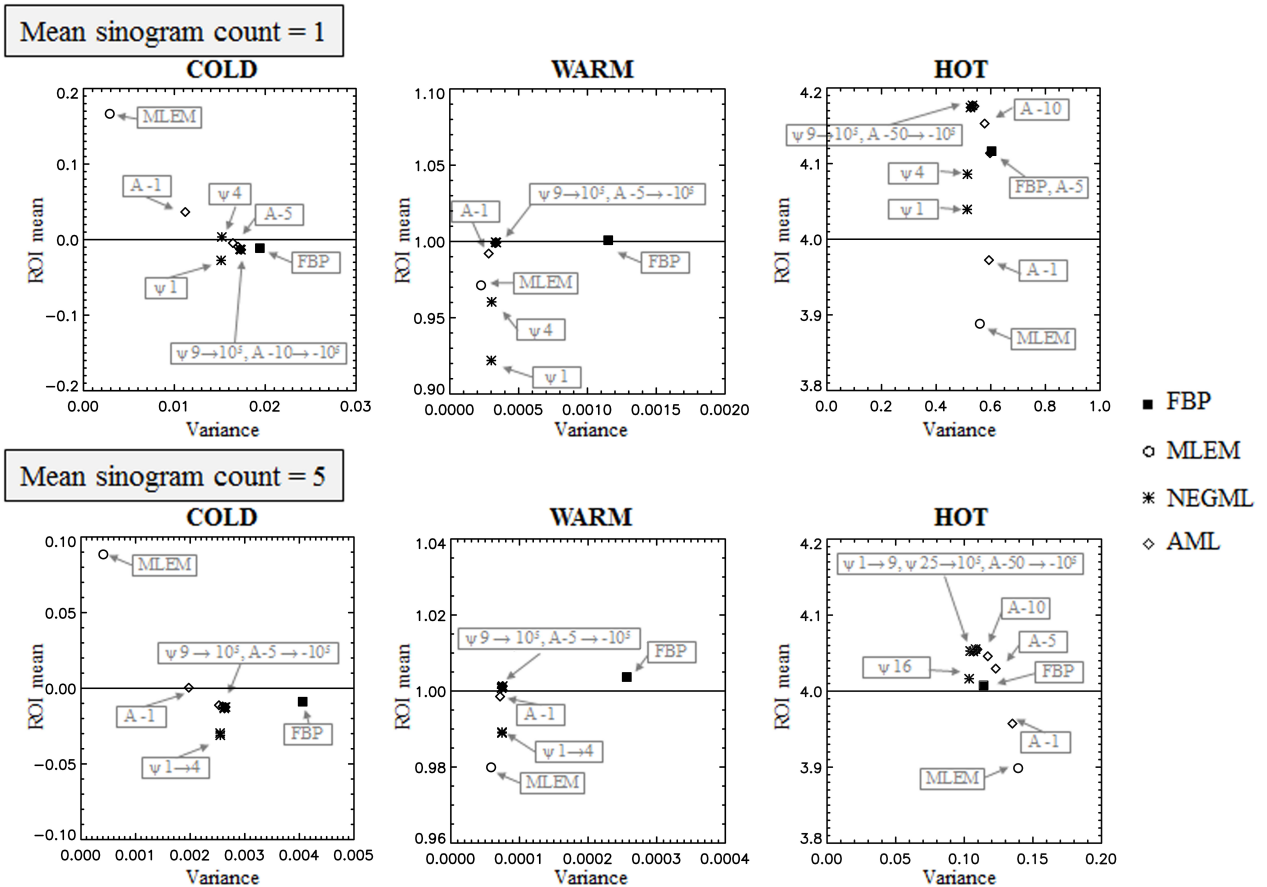


Figure 10.8: **Phantom 1** Mean with respect to variance in the cold regions, warm region and hot region for different values of ψ and A. The chosen frame duration correspond to on average 1 (top) and 5 (bottom) counts per sinogram pixel. The solid line represent the true value for each region. Reconstructed using ordinary Poisson with randoms smoothing for the iterative methods and precorrection with smoothed randoms for FBP.

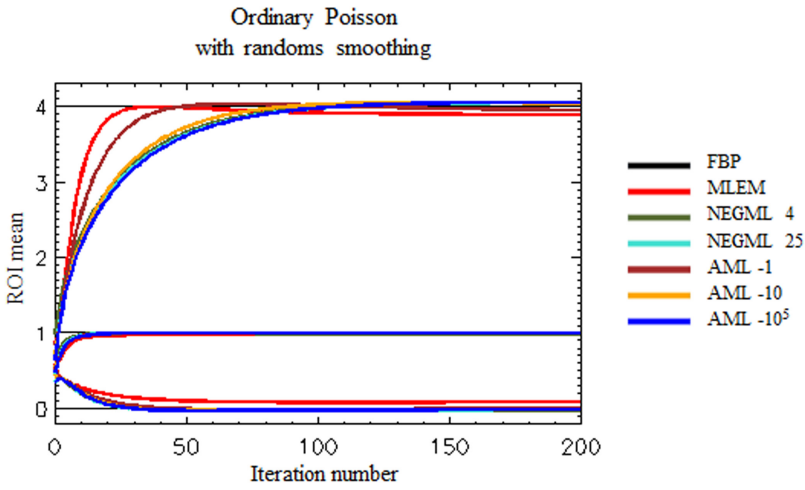


Figure 10.9: **Phantom 1** Convergence of MLEM, NEGML (with $\psi = \{4, 25\}$) and AML (with $A = \{-1, -10, -10^5\}$). The value for FBP is given as a reference

10.4.2 Phantom 2

The graphs in figure 10.10 depict the ROI mean in the cold and warm region of phantom 2 for a set of short frame durations for FBP, MLEM, NEGML and AML. Three different projection settings with a different number of projection angles were used.

FBP and NEGML show similar bias reduction for all three settings. For the smaller parameters, $\psi = \{1, 4\}$, again a local increase is observed, this is the most obvious for the warm region. For $\psi = 16$ no bias is observed for all three settings. The bias for MLEM and AML is not the same for all three settings. Although the activity in the image is the same for all settings, the same value for A results in different bias reduction depending on the setting. The largest value of $|A|$, $A = -50$, results in bias-free reconstruction for all three settings.

10.5 Discussion

A new method for bias reduction has been introduced. The method, NEGML, is based on a modified Poisson distribution instead of the Poisson distribution as usually used for PET reconstruction. The Poisson distribution is strongly asymmetric near zero and constrained to positive values, which is an important cause of bias in reconstructions of noisy, low count PET data. An obvious sym-

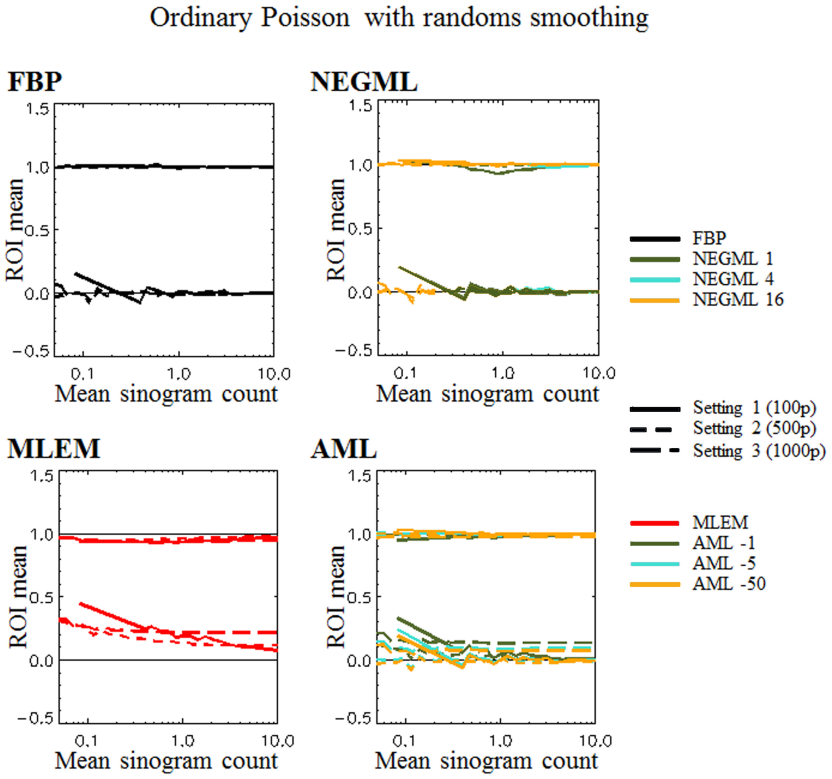


Figure 10.10: **Phantom 2** ROI mean in the cold and warm region for three settings. Setting 1: 100 projections angles (solid line), setting 2: 500 projection angles (small dashes), setting 3: 1000 projection angles (long dashes).

metric function that is defined for negative values is the Gaussian distribution. The parameter ψ defines at which point, i.e. for which number of counts, the distribution switches from Poisson to Gaussian. The method allows for negative values both in image and sinogram domain. The influence of ψ on the bias in the reconstruction is evaluated. It can be shown that when all values in the sinogram lie below ψ, NEGML reduces to an unweighted least squares algorithm.

NEGML was compared to another bias reduction method AML, a simplified version of ABML [106]. Instead of constraining the reconstructed image between lower bound A and upper bound B, AML only uses a lower bound. When this lower bound is set to very large (negative) 1, AML also converges to an unweighted least squares algorithm.

The bias reduction by NEGML and AML comes at the cost of increased variance. The parameters ψ of NEGML and A of AML select a compromise between reduced bias and increased variance. From figure 10.4 it is clear that the way that randoms are handled has a strong influence on the bias. Ordinary Poisson copes best with noisy low count data. Shifted Poisson is the worst, this is mostly due to the fact that it is using unsmoothed noisy randoms estimates. In a test with Ordinary Poisson with noisy randoms (not shown) increased bias was observed as well. When randoms precorrection is applied, the estimated randoms are subtracted from the measurement and might cause negative values in the sinogram. Methods that can handle negative values as NEGML and AML, introduce less bias for these precorrected sinogram data compared to MLEM where all negative sinogram values are set to zero.

MLEM has a considerable amount of bias for relatively low count data, especially in a cold region. This bias is reduced by NEGML and AML. The larger ψ and the smaller $|A|$ the better the bias reduction. For NEGML all values of ψ reduce the bias, but when the mean sinogram count lies around the value of ψ, increased bias is observed. This local increase in bias is less strong for higher values of ψ and vanishes almost completely for $\psi = 16$ or 25. The difference in bias for different values of A is relatively limited for Ordinary Poisson but can become important when using shifted Poisson where smaller $|A|$ show significant bias.

The hot regions do have more bias than the cold and warm region for all methods. For NEGML and AML with high $|A|$ it can be seen that some bias is left even for relatively high count data. This bias is merely a convergence issue. The evaluated hot region is rather small with much higher activity than the rest of the image. Voxel-dependent convergence weighting as in MLEM and AML (for small $|A|$), increases the convergence substantially. This effect decreases when $|A|$ becomes larger or for uniform weights as in NEGML. When iterating longer this bias disappears. The difference in convergence for the other regions is much smaller as shown in figure 10.9.

The convergence can be increased by the use of ordered subsets. Further,

for NEGML, α_j can be chosen non-uniformly but care has to be taken since convergence might be slowed down in lower count regions. We tested some values for α_j (not shown), the results were comparable to the results for small $|A|$ in AML, improved convergence in hot regions at the cost of risking more bias for cold regions.

The noise characteristics in the image change by using different parameters for AML and NEGML. The images in figure 10.7 show that increased ψ increases noise streaks in the image. However, although NEGML operates in full least squares mode, the streaks are still milder than for FBP. A similar trend is seen for AML, although AML keeps more of the typical MLEM look for small $|A|$ due to the weighting with the activity. However, no obvious difference could be observed between bias-free NEGML and AML images.

The plots in figure 10.8 demonstrate that both NEGML and AML have lower variance than FBP and the bias is usually better than for MLEM, as mentioned before (small) ψ values around the mean sinogram count have increased bias and should be avoided.

Figure 10.6 and 10.10 suggest that the bias reduction for NEGML does not depend on the particular simulation or measurement settings. For a different number of projection angles, NEGML with $\psi = 16$ always results in bias-free reconstructions. The same result was also obtained for both settings of phantom 1 (figures 10.4 and 10.6). This behavior was expected since a large part of the bias is due to the asymmetry of the Poisson distribution which is an effect in the sinogram. When choosing $\psi = 16$, NEGML will only switch to the Gaussian regime when necessary to reduce bias and keep the Poisson distribution, and therefore the noise behavior of typical maximum likelihood methods for high counts. NEGML replaces the Poisson distribution by a Gaussian distribution, where the choice of the distribution is solely based on the counts in a sinogram pixel, not on the activity. Hence, we expect that this effect will also hold for other settings as three dimensional PET.

Figure 10.6 and 10.10 show that a single value of A is not sufficient to reduce the bias under all circumstances, unless A is set to a value that is much lower than the expected activity in the image. In that situation, AML becomes similar to unweighted least squares reconstruction, which may not be preferable for dynamic studies where the Poisson behavior is still desired when the counts are high enough. This makes the determination of A less straightforward. Some tuning needs to be done to obtain a value of A that is sufficiently low, such that a sufficiently large shift is applied and such that sufficiently low image values can be obtained. Unlike for NEGML, lowering A changes the bias and image properties gradually. An optimized value for A might result in bias-free images with MLEM appearance. Studying the existence of an optimal value for A or more general, optimizing the image A_j with $j = 1..J$, was beyond the scope of this work.

10.6 Conclusion

A modified NEGML algorithm, based on a combination of a Poisson and Gaussian distribution, was proposed and has shown to reduces the bias significantly. An optimal value for switching from Poisson to a Gaussian distribution was found and it was shown to be independent on the activity in the image in two-dimensional simulations. NEGML was compared to AML. AML keeps the MLEM behavior for relatively small negative lower bounds especially in hot regions, leading to improved convergence for these regions. The larger the negative lower bound A becomes in absolute value, the more effective the bias reduction, giving images closer to least squares solutions. However, the optimal choice of the lower bound in AML is less straightforward. When bias is present in MLEM or NEGML and AML with small parameter, the experiments show that the way of random handling has an important influence on the amount of bias in the reconstruction. The use of unsmoothed randoms or randoms precorrected data increases the bias substantially.

Part IV

Concluding discussion

Chapter 11

Concluding discussion

This thesis considered improvements for iterative reconstruction in CT and PET. The reconstruction algorithms for both modalities are very similar since both techniques are based on the measurement of line integrals of attenuation or activity in the object. Consequently, approaches developed for one modality can often be translated to the other. The iterative algorithms used in this thesis are all based on the same gradient ascent optimization approach with (sometimes implicitly) adaptable voxel weights. This enabled the combination of different reconstruction models depending on the location in the image and improved local convergence. The gradient ascent approach was then also applied to a modified likelihood function. The concept of local models in the image and sinogram and local convergence improvement yielded better iterative reconstructions, regarding image quality and required computation time, in both CT and PET.

The three topics that have been considered are iterative metal artifact reduction in CT, acceleration of iterative reconstruction in general, applied to CT and PET, and PET bias reduction.

11.1 Metal artifact reduction

11.1.1 Main contributions

The image quality of CT images seriously degrades when there are metals present in the field of view. Metallic objects possibly present in the human body are dental fillings, (hip) implants, fixation screws, etc. The first methods to reduce metal artifacts were already presented in the 1980s. However, only recently the first dedicated metal artifact reduction method has been implemented on a clinical CT system (Philips, OMAR [117]). The main reason for

this reluctance is probably that most metal artifact reduction methods modify the measured data by sinogram inpainting or subtraction-based sinogram correction. The reconstruction after applying such techniques might be seemingly artifact-free but more subtle contrast in the region around metal and from tissue transitions lying in the modified projection rays might be lost. Iterative reconstruction could be a valuable alternative since it aims to reduce artifacts by a more accurate modeling of the acquisition. The data are not altered and all measured contrast information is still present in data.

IMPACT is a polychromatic algorithm for metal artifact reduction developed by B. De Man et al. in [76]. The algorithm reduces metal artifacts substantially when the metals are small and have low to moderate attenuation. For larger, highly attenuating metals still severe dark streaks and dark shadows are present. Moreover, the method was typically less effective for real measurements than for the corresponding simulated data and it increases the computation time per iteration substantially compared to simpler monochromatic iterative algorithms. This elongated computation is certainly an important obstacle for possible clinical usage of the method.

The local model scheme as developed in this work reduces several remaining problems for IMPACT and more general for iterative metal artifact reduction. The first aim was to limit the time-consuming calculations by the introduction of location dependent projection models that are more advanced for metal regions and simpler for other regions. A new more simple polychromatic algorithm was developed: MLTRC. Unlike IMPACT that takes into account the polychromatic attenuation of a list of materials, MLTRC considers only one material, the main material of the object, which in clinical applications is soft tissue (or water). Furthermore, the MLTRC model only requires the same number of (back)projection operations as the monochromatic MLTR model, and less than half of IMPACT. The advantages of MLTRC are that it results in fewer artifacts compared to MLTR (with the same reconstruction time) and that it is compatible with IMPACT for combined usage in a local model scheme. The results and the underlying idea of MLTRC are similar compared to sinogram precorrection (see section 5.1), however, the Poisson statistics of the data are better preserved when used in a maximum likelihood method. (It should be noted, however, that the use of Poisson statistics is already an approximation, because most CT systems use energy integrating detectors, resulting in a different, more complicated noise distribution [20].)

The second objective was the reduction of the artifacts for large and highly attenuating materials. It was found that the metals require many iterations to reach their true (high) attenuation values. This process was accelerated by the use of individual image updates for each of the metal regions or patches, which improves the convergence rate in these regions significantly. The lack of convergence in the metal regions turned out to be the main cause of the reduced efficiency of regular IMPACT in the presence of large and highly attenuation metals. Consequently, the remaining severe artifacts are largely reduced by the

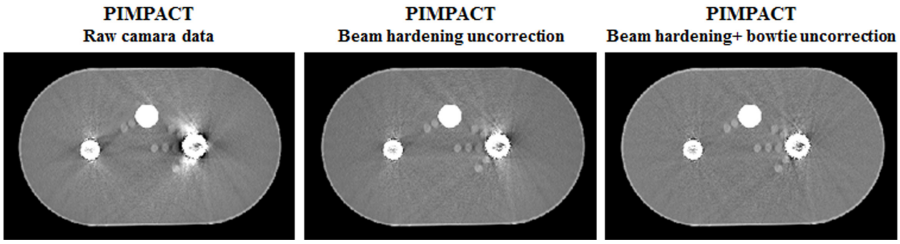


Figure 11.1: Effect of precorrection applied by the camera software on reconstructions with PIMPACT (as used in chapter 7). Left: reconstruction from sinogram data as obtained from the camera. Middle: reconstruction from sinogram data that are uncorrected for the beam hardening sinogram correction. Right: reconstruction from the sinogram data that are uncorrected for the beam hardening and bowtie correction.

use of individual updates of the metal regions.

In simulations only few artifacts were left after the introduction of the patch update. Nevertheless, the iterative methods still performed worse for measured data. This was an indication that the IMPACT and MLTRC model did not include all important acquisition physics.

An essential difference between measured and simulated data is that raw data obtained from clinical CT systems are usually not truly raw. Several correction mechanisms are applied to make the data usable for (clinical) reconstruction e.g. corrections for varying detector efficiency or for variations in the tube output. Together with the vendor of our CT systems we evaluated many of the corrections that are applied on the data and found that two corrections certainly interfere with the reconstruction models of MLTRC and IMPACT: the beam hardening and bowtie correction. The sinogram beam hardening correction is applied to suppress cupping artifacts (see section 5.1). It maps the polychromatic attenuation to a monochromatic attenuation assuming that the measured object is only composed of a single material, usually water or soft tissue. The polychromatic model of IMPACT or MLTRC is not correct when the data are precorrected for beam hardening. Similarly, correcting for position dependent filtering of the spectrum in the bowtie correction also interferes with the polychromatic model.

As mentioned in chapter 7 and chapter 9 these corrections need to be canceled to allow correct modeling of the polychromatic spectrum. Cancellation of these corrections required inversion of the correction operations which was done in close collaboration with the vendor of our CT systems. The different uncorrection steps for the phantom data used in chapter 7 are depicted in figure 11.1. Figure 11.1a is the PIMPACT (patched IMPACT) reconstruction from the data as obtained from the camera and figure 11.1b from the data after cancellation of the beam hardening correction. The data in figure 11.1c were

in addition also modified to undo the bowtie correction. It was found that cancellation of the corrections had an important influence on the performance of IMPACT and MLTRC.

This cancellation of corrections on the 'raw' data is probably only necessary in research environments. However, many corrections are applied and only few of them really interfere with the algorithms. The raw data with the desired 'rawness' are, to our knowledge, not available on most of the current CT systems. Implementation of our method (by the vendor) on a clinical system could probably allow to use the data before correction.

Besides the inclusion of the described uncorrection steps, the implementation of the IMPACT algorithm was also refined, e.g. it was found that the way the spectrum is binned has an influence on the data (equal energy range per bin performs better than equal amount of photons), also more adequate interpolation of the available attenuation tables improved the results.

The optimized algorithm with inclusion of the uncorrections resulted in ameliorated reconstructions for measured data both in image quality and in computation time after introduction of local models and the patch structure. The reconstructed images have only limited artifacts when the metal that is scanned has smooth axial boundaries. For sudden changes in the metal shape in axial direction some artifacts are still observed. So far, the origin of these artifacts is unclear, several possible causes (like scatter and partial volume effects) have been investigated but none of them reduced the remaining artifacts.

The results presented in this work suggest that the likelihood around the maximum is relatively flat. A first indication is that the regular and patch reconstruction yield very different artifacts for the same number of iterations. As mentioned in chapter 7, when iterating much longer, around ten times as long, both algorithms converge to a similar reconstruction. However, the reconstruction duration for that many iterations is clinically unacceptable. Moreover, for many iterations the reconstructions becomes unacceptably noisy. A second indication is the impact of the initialization on the artifacts. FBP images are typically used to initialize iterative reconstruction. However, to our knowledge, this has never been evaluated before for metal artifact reduction algorithms. We have seen that images that already contain very strong artifacts, like the FBP images, are bad initializations for iterative methods, since it can take very long to reduce the original artifacts. This is not an intricsic problem since, at full convergence all initialization will lead to the same image. However, the optimal initialization leads to the required image quality in the shortest time. For the initial tests with the patch structure in chapter 7, we preferred to use a uniform initialization to allow for unbiased comparison with FBP and NMAR without influence from one method on the other. In chapter 7 our experience on the influence of each of the initializations was explicitly presented.

11.1.2 Future perspectives

The reconstructions of the patient data in chapter 9 still contain artifacts. The origin of these artifacts remains unclear. The artifacts are the most pronounced when the metallic implant varies strongly in the axial direction. Partial volume effects and scatter have been investigated and attempts to compensate for these effects had negligible result on the reconstructions. Afterglow might be another cause of the artifacts. This effect can be modeled as an additive component similar to scatter. We performed some small experiments (not shown) where the afterglow term was calculated as a fraction of the measured data (shifted in angular direction) and used this as additive term r_i (similar treatment as for the scatter contribution). This again had negligible influence on the reconstruction. Possibly further data corrections or a combination of small effects during the measurement or during data processing may be the origin of the remaining artifacts. Further research on this topic will probably require to unravel more detailed information about the acquisition process and data processing.

Another effect that has not been discussed is the existence of wind mill artifacts in multi-slice helical CT. Wind mill artifacts are caused by incomplete sampling of the axial direction of the object when using a relatively high pitch [138]. The artifacts look somewhat similar to the remaining artifacts we observed. In case of missing information, multiple solutions are compatible with the data, resulting in several local maxima in the cost function. When constraining the reconstruction or starting from a good initial image, the algorithm can be guided towards the correct maximum. We have investigated incomplete sampling as the cause of the remaining artifacts. Several attempts to improve the initial image were made, none of them could avoid the introduction of the artifacts. Further, we have tried to constrain the reconstruction in the first iterations such that the artifacts could not be introduced. However, when turning off this constraining after several iterations, the reconstruction drifted away from the obtained artifact-free reconstruction and reintroduced the artifacts. Those experiments are another indication that the optimal reconstruction is not an element of the null-space of our problem and that the model that is currently used, is not accurate enough.

Such inaccuracies in the model might be caused by small deviations in the spectrum or attenuation tables we are using with respect to the true spectrum and to the true attenuation. In Depypere et al. [139] also a modification of the IMPACT model was proposed that allows to adapt the Compton and photo-electric component of the attenuation for each of the base materials of IMPACT during the reconstruction. The goal was to compensate for possible deviations in the spectral properties of the theoretical model. The context of the paper was the reconstruction of microCT dual energy data. This method has been tested on the data shown in this work to evaluate if compensation for possible spectral deviations would reduce the remaining artifacts. (This experiment has not been included in the text.) However, the application of this method did not reduce artifacts in the mono-energetical reconstruction

for clinical CT applications, as been used on this work. This result could be expected since we had already observed that the reconstruction algorithm was not very sensitive to the particular metal material in the IMPACT materials list. The use of either iron, stainless steel or cobalt-chrome in the IMPACT materials list for the reconstruction of cobalt-chrome prostheses had almost no influence on the artifacts in the reconstruction (see also chapter 9 section 9.2.5). Likewise, replacing water by PMMA had also no effect. This can be explained by the different energy range that is used in this work and in the work presented in [139]. In [139] microCT applications were considered, the photon energies for microCT applications are typically lower compared to clinical applications (in [139] a beam at 70kV and 100 kV was used, in this work a beam at 120kV was used). The curves for attenuation as a function of energy are steeper at lower energies and flatter for higher energies. This makes that small spectral deviations at lower energies will have more influence on the attenuation compared to the same deviation at higher energy. Moreover, the study of [139] considered dual energy scans and showed artifacts with a different appearance compared to those in this work. Although dual energy scans should provide more information about the object, contradictory information from both scans might introduce artifacts and then methods to compensate for these inconsistencies by slightly changing the attenuation properties of a predefined material will become more relevant.

The reconstruction algorithms in this work never used any form of regularization (e.g. by the use of prior information). This was done on purpose to study the intrinsic capacity to reduce metal artifacts by modeling the physics of the acquisition process. The remaining artifacts can probably be further reduced by application of a maximum a-posteriori (MAP) algorithm (section 4.3.3). De Man et al. studied in [92] the application of MAP reconstruction with MLTR for metal artifact reduction. The artifact reduction was substantial, although this was only tested for phantom data with relatively small metallic structures. However, MAP in combination with the local model scheme should most probably further improve the reconstruction. Furthermore, some projection rays going through metals might suffer from photon starvation and under these circumstances no information is present in these projection rays and priors can serve as regularization. Within the patch structure it is possible to make the prior parameters location dependent, i.e. different for each patch. Applying a prior function might smooth the metal-tissue transitions and possibly introduce new partial volume artifacts. This can be avoided by altering the prior strength for the metal and nonmetal patches or by switching the prior off in the metal patches.

The optimal algorithm for metal artifact reduction is probably application dependent. In cases where detailed contrast information is not needed, as for PET attenuation correction, sinogram inpainting methods certainly suffice. Under circumstances where contrast is crucial and information about the close metal surroundings is desired, as in follow-up of orthopedic surgery, iterative reconstruction gives much more accurate reconstructions.

The evaluation of the artifacts in this work was mainly done by visual inspection of the reconstruction. This is a very common procedure in the evaluation of artifacts in CT and more particularly for metal artifacts (e.g. [67, 68, 71, 72]). For instance, The Radiation Protection report 162 [140] of the European Union that describes the acceptability of medical radiological equipment does not contain any guideline to evaluate artifacts in CT and for 2D radiography images, visual inspection is the criterion for evaluation of artifacts. Attempts to evaluate metal artifacts and metal artifact reduction methods are mainly based on errors on the reconstructed Houndsfield unit (e.g. [69, 70, 75, 141, 142]). The evaluation is usually applied on relatively simple phantoms and hardly on patient data (also because of lack of ground truth). Methods based on sinogram inpainting usually perform well with respect to the Hounsfield units they reconstruct (although metal attenuation values are often underestimated). For instance in radiotherapy applications where treatment planning and dose calculations are based on the CT images, a metric considering errors on Houndsfield units is probably very relevant. However, possible more subtle reduction in contrast is usually not considered. This is often not relevant for dose calculation but it is important for diagnostic imaging. In our opinion, a metric for evaluation of diagnostic CT reconstructions should be based on the influence different reconstruction methods have on diagnosis. To our knowledge no automated algorithm exists to evaluate diagnosis relevance (either application specific or in general). A study with physicians scoring the quality of the reconstruction of multiple patient data sets, for a well defined application, would be of interest as for instance proposed in [73].

The local model technique is independent of the geometry of the CT system and can be extended to other geometries like cone beam. Cone beam systems are for example used in radiotherapy, dentistry, interventional cardiac CT and micro CT.

Although the general local model scheme can be directly applied on cone beam data, probably some cone beam specific issues need to be included in the projection model. First, most cone beam systems have no anti-scatter grids which makes the influence of scatter on the metal artifacts larger and dedicated scatter estimation may be necessary. Second, cone beam systems are often open systems that rotate much slower than typical multislice systems. Multislice CT systems have a source-detector pair on a ring rotating in a closed gantry. Under these circumstances very fast and stable rotation can be established. For open cone beam systems the source and detector are installed on two opposing arms. This reduces the maximum (safe) rotation time of the systems and might result in small geometrical deviations on the source and detector position (e.g. by gravitational effects). When the scan time increases some motion is almost inevitable, e.g. breathing and heart beating, and as shown in section 5.5 this can have a large influence on metal artifacts. Moreover, deviations in the position of the source and detector might introduce extra partial volume effects. Finally, in dentistry and cardiac applications often truncated data are used since only a limited region in the patient is of interest, i.e. the mouth or the heart.

Iterative reconstruction algorithms can cope better with truncation than FBP (see figure 5.4) but the influence of substantial truncation on the performance of dedicated metal artifact reduction techniques is not known.

The application of the technique presented in this work requires a lot of information from the vendor about the geometry of the system and the processing of the raw CT data. This information is often confidential and difficult to obtain, especially for clinical systems. This makes the application of the technique less evident for research purposes where this information is often not available. However, this is not an intrinsic but rather a practical problem of the proposed method.

The introduction of dual energy CT offers new possibilities for metal artifact reduction. In a dual energy scan, the object is scanned with two different spectra. This enables better tissue discrimination but also better modeling of the metal attenuation. A dual energy extension of IMPACT has been presented by Depypere et al. in [139]. The dual energy information can be used to estimate the photo-electric and Compton component of the attenuation directly (without linking them to a reference attenuation value) or by keeping the link between the components and adding an extra parameter to model the attenuation for materials with K-edges situated in the applied energy range. K-edges cannot be modeled by regular IMPACT because it assumes a monotonic attenuation function with respect to the energy, composed of a Compton scattering and photo-electric component. Application of the patch structure to dual energy reconstruction schemes can be investigated to broaden the application of the technique and to further improve metal artifact reduction.

The concept for local modeling by the use of a patch structure can also be applied for other applications than MAR. The work of Michielsen et al. [115] uses the patch structure to model the position dependent resolution in tomosynthesis by a patch dependent convolution operation.

The local model framework is not specific for transmission tomography applications. The basics for reconstruction are very similar in PET and SPECT reconstruction which makes that local modeling could also be used for PET or SPECT problems.

11.2 Acceleration for iterative CT and PET reconstruction

11.2.1 Main contributions

In order to introduce a reconstruction algorithm in clinical practice, it needs to be fast enough. Time requirements for CT reconstruction are somewhat higher than for PET since it is sometimes used in emergency departments and for (almost) real time imaging, e.g. for catheter placement.

Iterative reconstruction is much slower than analytical reconstruction. In CT, data sets are large and iterative reconstruction with a feedback loop to the projection domain is still too slow for routinely use. In PET, iterative reconstruction is already used, but data sets are becoming larger, which increases the required computation time.

Mechanisms to improve the convergence rate per iteration typically use subsets of the data. Ordered projection subsets use the measured data sequentially to update the entire image. In grouped-coordinate algorithms, a single voxel or a small group of voxels is sequentially updated. When only a small region of the image is updated, larger step sizes can be used without compromising convergence, resulting in faster convergence of the voxel values in that region. The optimal choice is choosing the voxels in a group spread over the image such that they have low influence on each other. This has disadvantages regarding memory access and combination with ordered subsets is often not possible.

Our work on metal artifact reduction with local models, showed that when updating a block of neighboring voxels, a substantial convergence improvement could be achieved. An image-block acceleration method was investigated in chapter 8. The entire reconstruction volume is divided in equally sized image-blocks. Updating image blocks of neighboring voxels has the advantage that standard (well optimized) projectors can be used, memory access is more efficient and that it can be combined with ordered subsets when the image blocks are sufficiently large.

This image-block update scheme with the patchwork structure is applicable to both CT and PET models and is independent of the specific geometry of the system. It was applied for circular, cone beam and helical CT and for 3D PET in combination with ordered subsets. The acceleration in convergence was significant and equal to the square root of the number of patches. Although not tested, this technique could also be used for SPECT reconstruction, since similar algorithms are used for both PET and SPECT.

This acceleration technique was also applied in [115] for digital breast tomosynthesis (DBT) reconstruction where the patches were chosen as planes parallel to the detector surface.

11.2.2 Future perspectives

The major focus of this work was the acceleration of the iterative algorithms by means of software related, system-independent acceleration. Growing attention is given to parallel implementation of iterative reconstruction using multiple CPUs (central processing units) or GPUs (graphical processing units) simultaneously [143]. Projection and backprojection operations are relatively easy applied in parallel. This has been investigated by K. Michielsen [115] for the work on tomosyntesis reconstruction, where data sets are even larger than for CT. OpenCL software was used for both CPU and GPU based multithread-

ing. Reconstruction of a tomosynthesis data set is reduced from several hours to less than half an hour. This work could be extended to helical and/or cone beam CT geometries.

Acceleration by parallelization will be an important factor for enabling fast iterative reconstruction such that it can be used in clinical procedures. Moreover, also other operations, such as convolutions, can be implemented in parallel. Many groups are working on multithreaded (CPU and/or GPU) tomography reconstruction as e.g. [144, 145] and [146].

11.3 Bias reduction in PET

11.3.1 Main contributions

For most PET applications, iterative reconstruction with MLEM outperforms analytical reconstruction regarding image quality. Nevertheless, in kinetic modeling where the quantification of reconstructed activity is very important, FBP is still used because MLEM tends to introduce bias for low count data and low activity regions. The (original) NEGML reconstruction algorithm [105] can reduce bias by allowing negative voxel values and by activity independent convergence weighting of the voxels. This reduces bias in low count regions but is insufficient for low count data.

A modified NEGML algorithm was proposed. An important reason for the bias caused in MLEM is caused by the cost function it optimizes, the Poisson likelihood. Maximum likelihood estimators are known to be biased for low count data. Therefore, the likelihood has been modified such that the asymmetry inherent to the Poisson distribution is removed. Hence, a modified likelihood was used that switches from a Poisson to a Gaussian distribution for low count data. The algorithm has one single sinogram-based parameter, the transition point for the distribution. The applied distribution is determined for each sinogram pixel individually based on the number of counts.

This modified NEGML algorithm performs better than MLEM and than the original algorithm in terms of bias reduction. The images look somewhat different than for MLEM. In MLEM, the weighting of the projection rays with the Poisson likelihood prevents significant streaking as observed for FBP reconstructions and the voxel dependent activity convergence of MLEM improves the reconstruction of high activity regions. NEGML has uniform voxel convergence which makes the reconstruction of high activity regions slower. When most of the data are treated with the Gaussian distribution, some streaking is present but still much less pronounced than for FBP.

The modified NEGML method has been compared to AML. In this method a shift is applied on the projection data towards higher counts such that the Poisson weighting becomes less asymmetric. The shift is based on the activity in

the image and is applied on the whole sinogram. AML has a more gradual transition from MLEM-like characteristics towards more Gaussian characteristics. However, to obtain bias-free images, a relatively large shift is needed. Bias-free reconstructions with AML are very similar to those with NEGML.

NEGML is somewhat more intuitive due to the sinogram based parameter. One single parameter can be used for a series of reconstructions with different activity and different frame durations. For AML this requires a little more tuning depending on the activity distribution and frame duration.

11.3.2 Future perspectives

The evaluation in chapter 10 was mainly two-dimensional. The experiments indicated that bias reduction with NEGML is independent on the geometry and the activity in the image. This suggests that NEGML would have comparable results for three-dimensional PET. A collaboration with the group of C. Comtat et al. (Orsay, France) and R. Boellaard et al. (Amsterdam, The Netherlands) is initiated to evaluate NEGML and AML for bias reduction in 3D PET data. The first results suggest that also in three dimensions, NEGML is unbiased for switching from Poisson to Gauss at 16 counts.

Our group has also started to study the influence of the reconstruction algorithm (FBP, MLEM or NEGML) on the parameters of kinetic tracer uptake modeling in a rat micro (3D) PET study.

Part V

Appendices

Appendix A

MLTRC with cross talk model

The basic model for MLTRC is presented in chapter 7 and in Van Slambrouck et al. [112]. In chapter 9, this model was extended by adding cross talk modeling to the algorithm. We modeled the cross talk by a convolution operation on the projection values. In this appendix a detailed derivation for MLTRC with convolution operation is given. The derivation is not specific for cross talk modeling but for a general convolution operation.

Recall that the Poisson (log-) likelihood is given by:

$$L = \sum_i y_i \ln \hat{y}_i - \hat{y}_i. \tag{A.1}$$

The general update for voxel j based on the gradient ascent approach developed in chapter 7 and 8 is given by (with $\alpha_j = 1$):

$$\Delta\mu_j = -\frac{\frac{\partial L}{\partial \mu_j}}{\sum_h \frac{\partial^2 L}{\partial \mu_j \partial \mu_h}} \tag{A.2}$$

A.1 Model

The standard projection model for sinogram pixel i in MLTRC is given by:

$$\hat{y}_i = \sum_k b_{ik} \exp\left(-C_k \sum_j l_{ij}\mu_j\right). \tag{A.3}$$

When including a convolution operation this becomes:

$$\hat{y}_i = \sum_n w_{in} \sum_k \hat{\psi}_{nk} \tag{A.4}$$

with w_{in} the convolution weights for sinogram pixel i and

$$\hat{\psi}_{ik} = b_{ik} \exp\left(-C_k \sum_j l_{ij} \mu_j \right). \tag{A.5}$$

A.2 Initial derivative calculation

$$\frac{\partial \hat{y}_i}{\partial \mu_j} = -\sum_n w_{in} l_{nj} Y_n^{\mathrm{I}} \tag{A.6}$$

$$\frac{\partial Y_i^{\mathrm{I}}}{\partial \mu_j} = -l_{ij} Y_i^{\mathrm{II}} \tag{A.7}$$

with

$$Y_i^{\mathrm{I}} = \sum_k C_k \hat{\psi}_{ik} \tag{A.8}$$

$$Y_i^{\mathrm{II}} = \sum_k C_k^2 \hat{\psi}_{ik} \tag{A.9}$$

A.3 Optimization of the likelihood

First derivative

$$\begin{aligned} \frac{\partial L}{\partial \mu_j} &= \sum_i \left(\frac{y_i}{\hat{y}_i} - 1 \right) \frac{\partial \hat{y}_i}{\partial \mu_j} \\ &= \sum_i \left(\frac{y_i}{\hat{y}_i} - 1 \right) \left(-\sum_n w_{in} l_{nj} Y_n^{\mathrm{I}} \right) \\ &= \sum_i l_{ij} Y_i^{\mathrm{I}} e_i \end{aligned} \tag{A.10}$$

with

$$e_i = \sum_n w_{ni} \left(1 - \frac{y_n}{\hat{y}_n} \right) \tag{A.11}$$

Note that the summation indexes i and n are switched in between the second and third step of (A.10).

Second derivative

$$\begin{aligned}\frac{\partial^2 L}{\partial \mu_j \partial \mu_h} &= \sum_i l_{ij}\left[e_i \frac{\partial Y_i^{\mathrm{I}}}{\mu_h} + Y_i^I \frac{\partial e_i}{\partial \mu_h}\right] \\ &= -\sum_i l_{ij}\left[e_i l_{ih} Y_i^{\mathrm{II}} + Y_i^{\mathrm{I}} \sum_n w_{ni} \frac{y_n}{\hat{y}_n^2} \sum_m w_{nm} l_{mh} Y_m^{\mathrm{I}}\right] \quad \text{(A.12)}\end{aligned}$$

$$\begin{aligned}\sum_h \frac{\partial^2 L}{\partial \mu_j \partial \mu_h} &= -\sum_i l_{ij}\left[\left(\sum_h l_{ih}\right) e_i Y_i^{\mathrm{II}}\right. \\ &\quad \left. + Y_i^{\mathrm{I}} \sum_n w_{ni} \frac{y_n}{\hat{y}_n^2} \sum_m w_{nm} \left(\sum_h l_{mh}\right) Y_m^{\mathrm{I}}\right] \\ &= -\sum_i l_{ij}\left(Y_i^{\mathrm{II}} V_i + Y_i^{\mathrm{I}} W_i\right) \quad \text{(A.13)}\end{aligned}$$

with

$$V_i = \left(\sum_h l_{ih}\right) e_i \quad \text{(A.14)}$$

$$W_i = \sum_n w_{ni} \frac{y_n}{\hat{y}_n^2} \sum_m w_{nm} Y_m^{\mathrm{I}} \left(\sum_h l_{mh}\right) \quad \text{(A.15)}$$

Update equation

$$\Delta \mu_j = \frac{\sum_i l_{ij} Y_i^{\mathrm{I}} e_i}{\sum_i l_{ij}\left(Y_i^{\mathrm{II}} V_i + Y_i^{\mathrm{I}} W_i\right)} \quad \text{(A.16)}$$

Appendix B

IMPACT with cross talk model

The basic model for IMPACT is given in chapter 7 and was initially developed by B. De Man et al. in [76]. The model presented in [76] did not include cross talk. Cross talk modeling has been added to the algorithm in chapter 9. Similar to MLTRC, we modeled cross talk by a convolution operation on the projection values. In this appendix a detailed derivation for IMPACT with convolution operation is given. The derivation is not specific for cross talk modeling but for a general convolution operation.

Recall that the Poisson (log-) likelihood is given by:

$$L = \sum_i y_i \ln \hat{y}_i - \hat{y}_i. \tag{B.1}$$

The general update for voxel j based on the gradient ascent approach developed in chapter 7 and 8 is given by (with $\alpha_j = 1$):

$$\Delta\mu_j = -\frac{\frac{\partial L}{\partial \mu_j}}{\sum_h \frac{\partial^2 L}{\partial \mu_j \partial \mu_h}} \tag{B.2}$$

B.1 Model

The projection model for IMPACT with resolution modeling for sinogram pixel i becomes:

$$\hat{y}_i = \sum_n w_{in} \sum_k \hat{\psi}_{nk} \tag{B.3}$$

with w_{in} the smoothing weight for sinogram pixel i and

$$\hat{\psi}_{ik} = b_{ik} \exp\left(-\Phi_k \sum_j l_{ij}\phi(\mu_j) - \Theta_k \sum_j l_{ij}\theta(\mu_j) \right). \tag{B.4}$$

B.2 Initial derivative calculation

$$\begin{aligned} \frac{\partial \hat{y}_i}{\partial \mu_j} &= \sum_n w_{in} \sum_k \frac{\partial \hat{\psi}_{nk}}{\partial \mu_j} \\ &= \sum_n w_{in} \sum_k b_{nk} \exp\left(-\Phi_k \sum_j l_{nj}\phi(\mu_j) - \Theta_k \sum_j l_{nj}\theta(\mu_j) \right) \\ &= -\sum_n w_{in} l_{nj} \left(\phi'_j Y_n^{\Phi} + \theta'_j Y_n^{\Theta} \right) \end{aligned} \tag{B.5}$$

with $\phi'_j = \frac{\partial \phi_j}{\partial \mu_j}\Big|_{\mu_j}$ and $\theta'_j = \frac{\partial \theta_j}{\partial \mu_j}\Big|_{\mu_j}$. At the transition points between materials the derivatives are calculated as follows for ϕ_j:

$$\phi_j = \frac{1}{2} \lim_{\epsilon \to 0} \left(\frac{\partial \phi_j}{\partial \mu_j}\Big|_{\mu_j - \epsilon} + \frac{\partial \phi_j}{\partial \mu_j}\Big|_{\mu_j + \epsilon} \right) \tag{B.6}$$

and analog for θ_j.

$$\begin{aligned} \frac{\partial Y_i^{\Phi}}{\partial \mu_j} &= \sum_k \Phi_k \frac{\partial \hat{\psi}_{ik}}{\partial \mu_j} \\ &= \sum_k \Phi_k \hat{\psi}_{ik} \left(-\Phi_k l_{ij} \phi'_j - \Theta_k l_{ij} \theta'_j \right) \\ &= -l_{ij} \left(\phi'_j Y_i^{\Theta\Theta} + \theta'_j Y_i^{\Phi\Theta} \right) \end{aligned} \tag{B.7}$$

$$\frac{\partial Y_i^{\Theta}}{\partial \mu_j} = -l_{ij} \left(\phi'_j Y_i^{\Phi\Theta} + \theta'_j Y_i^{\Theta\Theta} \right) \tag{B.8}$$

with

$$
\begin{aligned}
Y_i^{\Phi} &= \sum_k \Phi_k \hat{\psi}_{ik} && \text{(B.9)}\\
Y_i^{\Theta} &= \sum_k \Theta_k \hat{\psi}_{ik} && \text{(B.10)}\\
Y_i^{\Phi\Phi} &= \sum_k \Phi_k^2 \hat{\psi}_{ik} && \text{(B.11)}\\
Y_i^{\Theta\Theta} &= \sum_k \Theta_k^2 \hat{\psi}_{ik} && \text{(B.12)}\\
Y_i^{\Phi\Theta} &= \sum_k \Phi_k \Theta_k \hat{\psi}_{ik} && \text{(B.13)}
\end{aligned}
$$

B.3 Likelihood derivatives

First derivative

$$
\begin{aligned}
\frac{\partial L}{\partial \mu_j} &= \sum_i \left(\frac{y_i}{\hat{y}_i} - 1\right) \frac{\partial \hat{y}_i}{\partial \mu_j}\\
&= -\sum_i \left(\frac{y_i}{\hat{y}_i} - 1\right) \sum_n w_{in} l_{nj} \left(\phi_j' Y_n^{\Phi} + \theta_j' Y_n^{\Theta}\right)\\
&= \sum_i l_{ij} \left(\phi_j' Y_i^{\Phi} + \theta_j' Y_i^{\Theta}\right) e_i \qquad \text{(B.14)}
\end{aligned}
$$

with

$$
e_i = \sum_n w_{ni} \left(1 - \frac{y_n}{\hat{y}_n}\right) \qquad \text{(B.15)}
$$

Note that the summation indexes i and n have been switched between the second and third step of (B.14).

Second derivative

$$
\begin{aligned}
\frac{\partial^2 L}{\partial \mu_j \partial \mu_h} &= \sum_i l_{ij} \frac{\partial}{\partial \mu_h} \left(\phi_j' Y_i^{\Phi} + \theta_j' Y_i^{\Theta}\right) e_i\\
&+ \sum_i l_{ij} \left(\phi_j' Y_i^{\Phi} + \theta_j' Y_i^{\Theta}\right) \frac{\partial e_i}{\partial \mu_h} \qquad \text{(B.16)}
\end{aligned}
$$

The second derivative of ϕ_j and θ_j is zero on the linear parts of the curves. At the transition point the same assumption as in (B.6) is used.

$$\frac{\partial e_i}{\partial \mu_j} = -\sum_n w_{ni} \frac{y_n}{\hat{y}_n^2} \sum_m w_{nm} l_{mj} \left(\phi_j' Y_m^{\Phi} + \theta_j' Y_m^{\Theta} \right) \tag{B.17}$$

$$\begin{aligned} \frac{\partial^2 L}{\partial \mu_j \partial \mu_h} &= \sum_i l_{ij} \left[-\phi_j' \left(\phi_h' Y_i^{\Phi\Phi} + \theta_h' Y_i^{\Phi\Theta} \right) l_{ih} - \theta_j' \left(\phi_h' Y_i^{\Phi\Theta} + \theta_h' Y_i^{\Theta\Theta} \right) l_{ih} \right] e_i \\ &- \sum_i l_{ij} \left(\phi_j' Y_i^{\Psi} + \theta_j' Y_i^{\Theta} \right) \sum_n w_{ni} \frac{y_n}{\hat{y}_n^2} \\ &\qquad \cdot \sum_m w_{nm} l_{mh} \left(\phi_h' Y_m^{\Phi} + \theta_h' Y_m^{\Theta} \right) \end{aligned} \tag{B.18}$$

$$\begin{aligned} \sum_h \frac{\partial^2 L}{\partial \mu_j \partial \mu_h} &= -\sum_i l_{ij} \left(\sum_h \phi_h' l_{ih} \right) \left(Y_i^{\Phi\Phi} e_i \right) \phi_j' \\ &\quad -\sum_i l_{ij} \left(\sum_h \theta_h' l_{ih} \right) \left(Y_i^{\Phi\Theta} e_i \right) \phi_j' \\ &\quad -\sum_i l_{ij} \left(\sum_h \phi_h' l_{ih} \right) \left(Y_i^{\Phi\Theta} e_i \right) \theta_j' \\ &\quad -\sum_i l_{ij} \left(\sum_h \theta_h' l_{ih} \right) \left(Y_i^{\Theta\Theta} e_i \right) \theta_j' \\ &\quad -\sum_i l_{ij} \left(\phi_j' Y_i^{\Phi} + \theta_j' Y_i^{\Theta} \right) \sum_n w_{ni} \frac{y_n}{\hat{y}_n^2} \\ &\qquad \cdot \sum_m w_{nm} \left(Y_m^{\Phi} \left(\sum_h l_{mh} \phi_h' \right) + Y_m^{\Theta} \left(\sum_h l_{mh} \theta_h' \right) \right) \\ &= \phi_j' \sum_i l_{ij} M_i + \theta_j' \sum_i l_{ij} N_i \end{aligned} \tag{B.19}$$

with

$$M_i = \left(\sum_h l_{ih}\phi'_h\right)\left(Y_i^{\Phi\Phi}e_i\right) + \left(\sum_h l_{ih}\theta'_h\right)\left(Y_i^{\Phi\Theta}e_i\right) \tag{B.20}$$
$$+ Y_i^{\Phi}\sum w_{ni}\frac{y_n}{\hat{y}_n^2}\sum_m w_{nm}\left(\left(Y_m^{\Phi}\sum_h l_{ih}\phi'_h\right) + Y_m^{\Theta}\left(\sum_h l_{ih}\phi'_h\right)\right)$$

$$N_i = \left(\sum_h l_{ih}\phi'_h\right)\left(Y_i^{\Phi\Phi}e_i\right) + \left(\sum_h l_{ih}\theta'_h\right)\left(Y_i^{\Theta\Theta}e_i\right) \tag{B.21}$$
$$+ Y_i^{\Theta}\sum w_{ni}\frac{y_n}{\hat{y}_n^2}\sum_m w_{nm}\left(\left(Y_m^{\Phi}\sum_h l_{ih}\phi'_h\right) + Y_m^{\Theta}\left(\sum_h l_{ih}\phi'_h\right)\right)$$

The update for voxel j becomes:

$$\Delta\mu_j = \frac{\phi'_j\sum_i l_{ij}Y_i^{\Phi}e_i + \theta'_j\sum_i l_{ij}Y_i^{\Theta}e_i}{\phi'_j\sum_i l_{ij}M_i + \theta'_j\sum_i l_{ij}N_i} \tag{B.22}$$

Appendix C

On the implementation of the patchwork structure

The patchwork structure is a new concept that was developed and implemented during this work. Some aspects regarding the implementation of the patchwork projector have been mentioned in chapters 7, 8 and 9. In this appendix an overview of my implementation of the patchwork structure is given.

A patchwork reconstruction requires a patchwork image, a patchwork projector and patched reconstruction algorithms. The main element in the structure is the projector, this will be the basis for all other patch elements.

C.1 Patched volume and patchwork image

Patch structure Patchwork images can be composed in two different ways: they can contain rectangular patches that lie next to each other, as in chapter 8, or irregular smaller patches lying in other (larger) patches as in chapter 7 and chapter 9. Currently only two levels of patches are supported. This means that no patch can be defined in a patch that already lies in another patch.

Each patch in the patchwork image represents a part of the reference reconstruction volume, however, not necessarily at the same voxel size and there is no overlap allowed between different patches. To allow a proper patch keeping, some information needs to be linked to each of the patches:

- The location of the patch in the reference volume
- Are some parts of its volume occupied by other patches and which patches are these?
- Is this patch lying in another patch and which patch is this?

- The voxel size of the patch with respect to a reference voxel size

The reference voxel size is usually the voxel size that would be used for a regular nonpatched image. The voxel sizes of the patches are chosen such that they are always an integer factor times smaller or larger than the reference voxel size.

Patch position definition The composition of the patchwork image can be calculated manually or by one of the two patch definitions scripts that have been written. Currently two different parameter sets are allowed to define a patch in a volume: begin and end point of each of the patches or a list with voxels belonging to each of the patches.

The first automated procedure for a patch definition is made for acceleration purposes and divides a volume into N^2 (without axial patches) or N^3 (with axial patches) equally sized patches. Of course, the number of voxels in the reference volume in the row and column (and plane) direction should be dividable by N.

The second procedure is made for metal artifact reduction and defines patches for each of the metals based on a threshold and a margin. Voxels with a value exceeding the threshold belong to a (small) patch, all other voxels will belong to the the main (larger) patch. The border range defines the extension of the patch beyond the segmented border, e.g. to make the segmentation of the metal a little larger than just the metal. To prevent the creation of many patches of a few voxels, a minimum number of voxels per patch is set. The patch can be defined irregularly shaped following the shape of the metal or a rectangular or cuboid patch can be drawn around the voxels that exceeded the given threshold. In our experience, the option with rectangular patches might suffer more from edge artifacts in the corners of the rectangle. Moreover, irregular patches also limit the number of voxels in a patch to the minimum.

The information obtained from the patch volume definition procedures is part of the information used to create a patchwork projector. Since the projector contains all geometrical information about the patchwork structure it is used as main object to derive other patchwork structures from (see figure C.1). For instance, patchwork images can only be defined based on a patchwork projector or based on another patchwork image. When creating a patchwork image, the volume can be left empty or it can be filled with a given regular image.

Displaying images A procedure has been written that converts the patchwork image into a regular image. For displaying the whole reconstruction volume, all patches should have the same voxel size. By default the reference voxel size is used. However, smaller or larger voxel sizes can be selected too and the different patches of the images can be displayed separately at their own resolution.

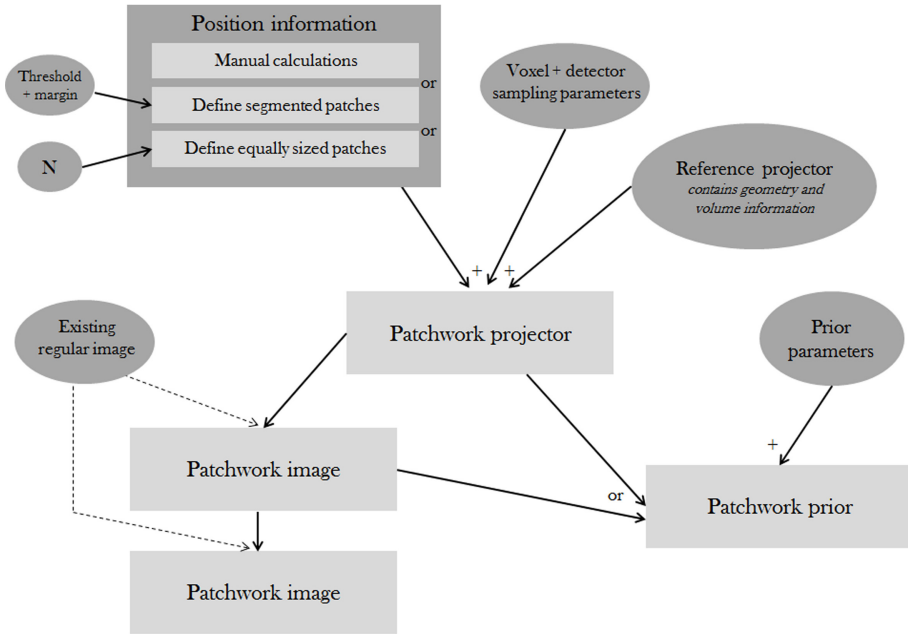


Figure C.1: Definition of the patchwork structure. The position information, voxel and detector sampling and the reference projector are the input for the definition of a patchwork projector. A patchwork image can be created based on a patchwork projector or another patchwork image. The image can be filled with the values of an existing image. A patchwork prior can be defined based on a patchwork projector or based on a patchwork image.

C.2 Patchwork projector

The patchwork projector is a combination of projectors for each of the patches. These projectors are defined based on a reference projector that contains, besides general geometrical information about the system, the following projection information per patch:

- Which volume needs to be (back)projected?
- Where in the sinogram does the patch project?
- Should this projection be oversampled?

It is calculated beforehand where in the sinogram a patch projects because this limits the calculations during projection and backprojection. The oversampling is only applied in the forward projection and not in the backprojection.

Projector definition and patch structure The definition of all patch structures is based on the patchwork projector. The creation of a patchwork projector is based on (see figure C.1):

- Position information about the patches (that can be obtained from the patch definition procedures)
- A reference projector
- Voxel and detector sampling parameters per patch

This structure is the main element for patchwork reconstruction and contains all information about the patches in image and sinogram domain. Hence, to keep all patchwork structures compatible with each other, all other patchwork structures are derived from the patchwork projector.

Combined projection The set of all patch projections is used to compute the (combined) forward projection of the (patched) image.

When combining different reconstruction models a separate projection is saved for each of the patches. Ideally only the part of the sinogram containing information about the trace of that particular patch should be stored per patch, however this was not yet implemented.

The different (stored) projections are combined to form the combined projection estimate (also given in Eq. (7.19)):

$$\hat{y}_i = \sum_k b_{ik} \prod_p z_{ik,p} + Sp_i \text{ with } \quad z_{ik,p} = \frac{\hat{y}_{ik,p}}{b_{ik}} \tag{C.1}$$

where each $z_{ik,p}$ is calculated separately for each patch using the model assigned to that particular patch. The projections of the patches are combined by ordering them from highest to lowest detector resolution. The ones with the

highest resolution are first mutually combined (by a multiplication since this is done after exponentiation), rebinned to the next-to-lowest resolution and combined with projections at the next-to-lowest resolution, etc. (Usually only two different detector resolutions were used.)

C.3 Patchwork prior

Besides a patchwork image and a patchwork projector also a patchwork prior has been implemented. This patchwork prior has not been used for the studies described in this thesis.

The patchwork prior can be defined based on a patchwork projector or based on a patchwork image and with additional information for the prior. The prior is implemented mainly for being used in the local model structure. Currently a different prior can be used in each of the patches and the prior can be switched on or off for each of the patches separately. For instance, a prior could be used to suppress noise or artifacts in the main nonmetal patch and switched off in the metal patch(es) to prevent edge smoothing of the metal which introduces partial volume artifacts.

The patchwork prior is not yet ready for more general usage because currently it is locally defined in the patch. This means that at the patch borders no information about the neighboring patches can be used. However, an implementation that enables 'over-the-edge' smoothing is not straightforward when different voxel sizes are used in different patches.

C.4 Patched reconstruction

The principle of patched reconstruction is depicted in figure C.2. It consists of three major parts: a script that runs over the iterations and the subsets, a script that runs over the patches and the reconstruction algorithm itself.

The main script contains two loops: over the iterations and over the subsets. For each subset a projection estimate based on the current patchwork image I is calculated. The following is passed to the patch keeping script:

- The current reconstruction I
- The patchwork projector R
- (if required) The patchwork prior G
- The projection estimate $\hat{y}_{i \in s}$ for subset s
- The subset of the measured data $y_{i \in s}$
- The energy models for each patch

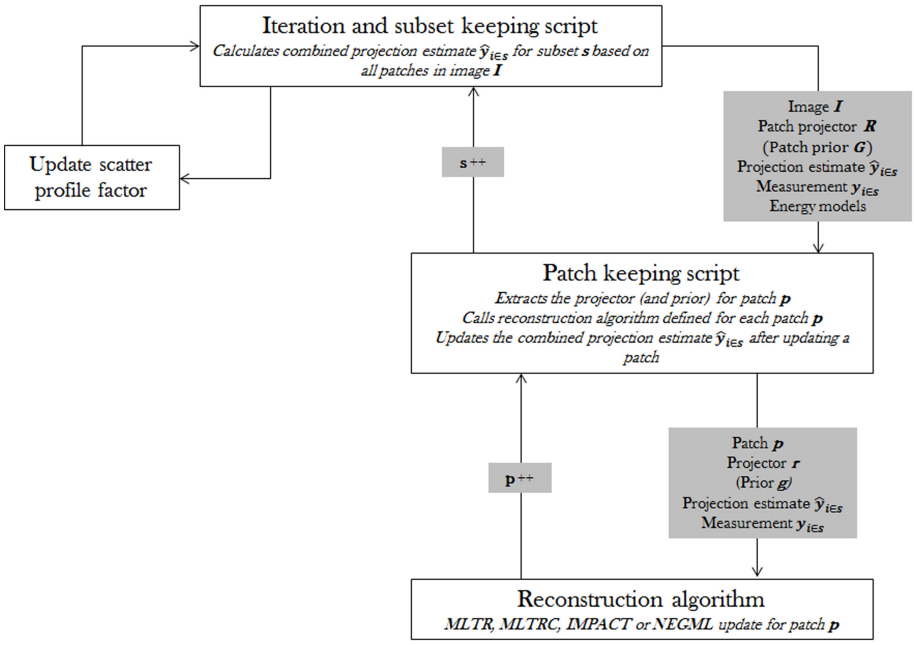


Figure C.2: An overview of a patched reconstruction

The energy model was not included in the patchwork definition, such that the same patch definition can be used for reconstruction with different reconstruction models or different combinations of reconstruction models. However, the resolution model is (by default) included in the projector because it is based on image and projector specific parameters.

The patch keeping script extracts the patch volumes one by one out of the image I and sends them (also one by one) to the correct reconstruction algorithm to calculate an update.

Combined projection estimate After updating one of the patches, the combined projection update needs to be updated as well. This has been implemented in two different manners.

For circumstances where the patches can have different models, a projection for each patch is saved. As said before this could be optimized by saving only the trace of the patch in the sinogram. For MLTR and MLTRC the projections of the attenuation values (at the reference energy) are stored. For IMPACT two projections need to be stored: a first projection for the Compton component and a second projection for the photo-electric component. Only the projection(s) of the patch that has been updated will be recalculated. Next, a new combined update for the subset under consideration is calculated.

The projections of all patches separately do not have to be stored when a single monochromatic model is used, as used in chapter 8 for acceleration of MLTR and NEGML. In that case, the calculated $\Delta\mu$, the difference between the previous and new reconstruction, can be forward projected and added to the existing combined estimate.

The combined estimate for an MLTR-model with three patches $\{p_1, p_2, p_3\}$ is given by:

$$\begin{aligned}\hat{y}_i &= b_i \exp\Big(-\sum_{j\in p_1} l_{ij}\mu_j\Big)\exp\Big(-\sum_{j\in p_2} l_{ij}\mu_j\Big)\exp\Big(-\sum_{j\in p_3} l_{ij}\mu_j\Big)\\ &= b_i \exp\Big(-\sum_{j\in p_1} l_{ij}\mu_j - \sum_{j\in p_2} l_{ij}\mu_j - \sum_{j\in p_3} l_{ij}\mu_j\Big) \qquad \text{(C.2)}\end{aligned}$$

If separate projections are stored $\sum_{j\in p} l_{ij}\mu_j$ is stored for each patch p and recalculated after updating a patch to be combined again as in Eq. (C.2) to form the combined projection estimate.

When the same monochromatic model is used this process can be simplified, without the need to store all projections separately. The combined projection estimate, before updating a new patch is given by:

$$\hat{y}_i^{\text{old}} = b_i \exp\Big(-\sum_{j\in p_1} l_{ij}\mu_j\Big)\exp\Big(-\sum_{j\in p_2} l_{ij}\mu_j\Big)\exp\Big(-\sum_{j\in p_3} l_{ij}\mu_j\Big) \qquad \text{(C.3)}$$

For instance after updating patch 3, p_3, the new estimate can be calculated by:

$$\begin{aligned}\hat{y}_i^{\text{new}} &= b_i \exp\Big(-\sum_{j\in p_1} l_{ij}\mu_j\Big)\exp\Big(-\sum_{j\in p_2} l_{ij}\mu_j\Big)\exp\Big(-\sum_{j\in p_3} l_{ij}\mu_j^{\text{new}}\Big)\\ &= \hat{y}_i^{\text{old}} \exp\Big(\sum_{j\in p_3} l_{ij}\mu_j^{\text{old}}\Big)\exp\Big(-\sum_{j\in p_3} l_{ij}\mu_j^{\text{new}}\Big)\\ &= \hat{y}_i^{\text{old}} \exp\Big(-\sum_{j\in p_3} l_{ij}\Delta\mu_j\Big) \qquad \text{(C.4)}\end{aligned}$$

This means that only $\exp\Big(-\sum_{j\in p_3} l_{ij}\Delta\mu_j\Big)$ needs to be calculated to update the combined projection. However, this is not possible for polychromatic models because there the sum over the energy bins k needs to be applied after combining the patches. In this situation, the combined estimate is recalculated based on the stored projections of the attenuation values at the reference energy (for MLTRC) or the projections of the Compton and photo-electric component of the attenuation (for IMPACT).

Scatter profile update After each subset the main script can also call a procedure that updates the scatter profile factor (see chapter 7).

List of publications

International peer-reviewed papers

- ***K. Van Slambrouck*** and J. Nuyts. "Metal artifact reduction for three-dimensional helical CT." (In preparation.)
- ***K. Van Slambrouck***, S. Stute, C. Comtat, M. Sibomana, F.H.P. van Velden, R. Boellaard and J. Nuyts. "Bias reduction for low-statistics PET: Maximum likelihood reconstruction with modified distribution." (Submitted to *IEEE Transactions on Medical Imaging*.)
- ***K. Van Slambrouck*** and J. Nuyts (2013). "Reconstruction scheme for accelerated maximum likelihood reconstruction: the patchwork structure." *IEEE Transactions on Nuclear Science*. (Accepted with minor revision.)
- M. Cona, Y. Feng, Y. Li, F. Chen, K. Vunckx, L. Zhou, ***K. Van Slambrouck***, A. Rezaei, O. Gheysens, J. Nuyts, A. Verbruggen, R. Oyen, and Y. Ni (2013). "Comparative study of Iodine-123-labeled-hypericin and Tc-99m-labeled-hexakis [2-methoxyisobutylisonitril] in a rabbit model of myocardial infarction." *Journal of Cardiovascular Pharmacology* (in press).
- A. Mouton, N. Megherbi, ***K. Van Slambrouck***, J. Nuyts and T. P. Breckon (2013). "An experimental survey of metal artefact reduction in computed tomography." *Journal of X-ray Science and Technology*, 21 (2), pp. 193-226.
- K. Michielsen, ***K. Van Slambrouck***, A. Jerebko and J. Nuyts (2013). "Patchwork Reconstruction with Resolution Modeling for Digital Breast Tomosynthesis." *Medical Physics*, 40 (3), 0311505.
- ***K. Van Slambrouck*** and J. Nuyts (2012). "Metal artifact reduction in computed tomography using local models in an image block-iterative scheme." *Medical Physics*, 39 (11), 7080-7093.
- G. Van Gompel, ***K. Van Slambrouck***, M. Defrise, K.J. Batenburg, J. de Mey, J. Sijbers, J. Nuyts (2011). "Iterative correction of beam hardening artifacts in CT." *Medical Physics*, 38, S36-S49.

Conference Proceedings (Full paper)

- K. Michielsen, ***K. Van Slambrouck***, A. Jerebko, and J. Nuyts (2012). "Patchwork reconstruction with resolution modeling for digital breast tomosynthesis." *Proc. 2nd Intl. Mtg. on Image Formation in X-Ray CT*, Salt Lake City, Utah, USA, pp. 21-24.

- J. Nuyts, S. Stute, ***K. Van Slambrouck***, F.H.P. van Velden, R. Boellaard and C. Comtat (2011). "Maximum-likelihood reconstruction based on a modified Poisson distribution to reduce bias in PET." *IEEE Nucl Sci Symp Conf Record*, Valencia, Spain, pp. 4337-4341.

- S. Stute, J. Nuyts, ***K. Van Slambrouck***, F. van Velden, R. Boellaard and C. Comtat (2011). "Image properties of various ML-based reconstructions of very noisy HRRT data." *IEEE Nucl Sci Symp Conf Record*, Valencia, Spain, pp. 4311-4315.

- ***K. Van Slambrouck*** and J. Nuyts (2011). "Local models for artefact reduction in iterative CT reconstruction." *IEEE Nucl Sci Symp Conf Record*, Valencia, Spain, pp. 4346-4350.

- ***K. Van Slambrouck*** and J. Nuyts (2011). "Accelerated convergence with image-block iterative reconstruction." *IEEE Nucl Sci Symp Conf Record*, Valencia, Spain, pp. 2576-2580.

- ***K. Van Slambrouck***, and J. Nuyts (2010). "A patchwork (back)projector to accelerate artifact reduction in CT reconstruction." *IEEE Nucl Sci Symp Conf Record*, Knoxville, TN, USA, pp. 2625-2629.

- ***K. Van Slambrouck***, G. Van Gompel, M. Defrise, K.J. Batenburg, J. Sijbers and J. Nuyts (2010). "Iterative correction of beam hardening artifacts in CT." *Proc. 1st Intl. Mtg. on Image Formation in X-Ray CT*, Salt Lake City, Utah, USA, pp. 244-247.

Conference Abstracts (not published as full paper)

- A. Mouton, N. Megherbi, ***K. Van Slambrouck***, J. Nuyts and T. Breckon (2013). "A distance weighted extension to an intensity limiting method for metal artefact reduction in CT." IEEE International Conference on Image Processing, September 15-18, Melbourne, Australia.

- ***K. Van Slambrouck***, K. Michielsen and J. Nuyts (2013). "Comparison of iterative reconstruction and sinogram inpainting for the reduction of metal artefacts in CT." European Congress of Radiology, March 7-11, Vienna, Austria.

- ***K. Van Slambrouck*** and J. Nuyts (2012). "Local Model Reconstruction for Metal Artifact Reduction." Symposium of Belgian Hospital Physicists Association, February 10-11, Brussels, Belgium.

Presentations on workshops

- ***K. Van Slambrouck***, S. Stute, C. Comtat, M. Sibomana, F.H.P. van Velden, R. Boellaard and J. Nuyts. Numerical Modeling and Simulation of Inverse Problems in Medical Imaging (NMSIPMI) workshop, 10 and 11 October 2013, Grenoble. "Bias reduction for low-statistics PET: Maximum likelihood reconstruction with modified distribution."
- ***K. Van Slambrouck*** and J. Nuyts. DROITE workshop, 24 and 25 October 2012, CPE, Lyon. "Metal artifact reduction in computed tomography."
- J. Nuyts and ***K. Van Slambrouck***. DROITE workshop: Maths of Tomography, Medical Imaging and HPC, 8 June 2013, MESR, Paris. "The use of local physical models for iterative reconstruction of CT."
- ***K. Van Slambrouck*** and J. Nuyts. IUAP workshop on medical image processing, 28 May 2010, Ghent, Belgium. "A patchwork (back)-projector for CT: preliminary results."

Award

Young Physicist Award 2012, Belgian Hospital Physicists Association (BHPA) Symposium, 10-11 February 2012.

Curriculum Vitae

Katrien Van Slambrouck was born on the 13th of October in 1986 in Leuven, Belgium. She graduated from secondary school (Greek-mathematics, Paridaens, Leuven) in June 2004. In September 2004 she started her studies Bachelor in Physics at the Department of Physics and Astronomy at the KU Leuven, Leuven, Belgium and graduated cum laude in July 2007. From September 2007 she studied Master in Cellular and Molecular Biophysics at the KU Leuven Science Faculty and graduated magna cum laude in July 2009.

In September 2009 she started her PhD studies at the Faculty of Medicine, Department of Imaging and Pathology, Division of Nuclear Medicine and the Medical Imaging Research Center of the KU Leuven on the topic: Local models and local convergence for maximum likelihood iterative reconstruction. In February 2012 she received the Young Physicists Award from the Belgian Hospital Physicists Association (BHPA) for her work on metal artifact reduction with local reconstruction models.

Bibliography

[1] NASA: http://science.hq.nasa.gov/kids/imagers/ems/xrays.html. (Accessed 5th June 2013)

[2] American Institute of Physics: http://www.aip.org/history/curie/war1.html. (Accessed 11th June 2013.)

[3] J. Hsieh. *Computed tomography: principles, design, artifacts and recent advances.* SPIE Press, Bellingham/Washington, 2003.

[4] A. Kak and M. Slaney. *Principles of computerized tomographic imaging.* SIAM, Philadelphia, 2001.

[5] T.G. Flohr, S. Schaller, K. Stierstorfer, H. Bruder, B.M. Ohnesorge and U. Joseph Schoepf. "MultiDetector Row CT Systems and Image-Reconstruction Techniques." *Radiology*, **235**, pp. 756-773, 2005.

[6] S.R. Cherry, J.A. Sorenson and M.E. Phelps. *Physics in Nuclear Medicine, fourth edition.* Elsevier-Saunders, Philadelphia, 2012.

[7] K. Krane. *Introductory Nuclear Physics.* John Wiley & Sons, Hoboken/New Jersey, 1988.

[8] W.R. Hendee. "Cross sectional medical imaging: a history." *RadioGraphics*, **9** (6), pp. 1155-1180, 1989.

[9] Toshiba Healthcare: http://medical.toshiba.com/products/ct/aquilion-one-vision/index.php. (Accessed 11th June 2013.)

[10] Siemens Healthcare: https://www.medical.siemens.com/siemens/de_DE/gg_ct_FBAs/files/Options_Portal/Case_Studies/Dual_Energy_CT.pdf. (Accessed 21 July 2013.)

[11] Toshiba Medical Systems Europe: http://www.toshiba-medical.eu/en/Our-Product-Range/CT/Technologies/Dual-Energy/. (Accessed 21 July 2013.)

[12] Philips Healthcare: http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/CT/

Exploring-the-spectrum-Advances-and-potential-of-spectral-CT. (Accessed 21 July 2013.)

[13] GE Healthcare: http://www3.gehealthcare.com/en/Products/Categories/Computed_Tomography/Discovery_CT750_HD/Spectral_Imaging. (Accessed 21 July 2013.)

[14] T. Johnson, B. Krauß, M. Sedlmair, M. Grasruck, H. Bruder, D. Morhard, C. Fink, S. Weckbach, M. Lenhard, B. Schmidt, T. Flohr, M. Reiser, C. Becker. "Material differentiation by dual energy CT: initial experience." *European Radiology*, **17** (6), pp. 1510-1517, 2007.

[15] K. Taguchi, S. Srivastava, H. Kudo, and W. Barber. "Enabling Photon Counting Clinical X-ray CT." *Conference record IEEE Medical Imaging Conference, Orlando, FL, USA*, pp. 3581-3585, 2009.

[16] P. Suetens. *Fundamentals of Medical Imaging, second edition.* Cambridge University Press, Cambridge, 2009.

[17] G. Vogtmeier, R. Dorscheid, K.J. Engel, R. Lutha, R. Mattson, B. Harwood, M. Appleby, R. Randolph and J. Klinger. "Two-dimensional anti-scatter-grids for computed tomography detectors." *Proceedings of SPIE*, **6913**, 2008.

[18] B.R. Whiting. "Signal statistics in x-ray computed tomography." *Proceedings of SPIE*, **4682**, pp. 53-60, 2002

[19] I. Elbakri and J. Fessler. "Efficient and accurate likelihood for iterative image reconstruction in x-ray computed tomography." *Proc. SPIE*, **5032**, pp. 1839-1850, 2003.

[20] J. Nuyts, B. De Man, J.A. Fessler, W. Zbijewski and F.J. Beekman. "Modelling the physics in the iterative reconstruction for transmission computed tomography." *Physics in Medicine and Biology*, **58**, pp. R63-R96, 2013.

[21] J. Humm, A. Rosenfield and A. Del Guerra. "From PET detectors to PET scanners." *European Journal of Nuclear Medicine and Molecular Imaging*, **30** (11), pp. 1574-1597, 2003.

[22] O. Mawlawi and D. Townsend. "Multimodality imaging: an update on PET/CT technology." *European Journal of Medical Molecular Imaging*, **36** (S1), pp. S15-S29, 2009.

[23] P. Kinahan, D. Townsend, T. Beyer and D. Sashin. "Attenuation correction for a combined 3D PET/CT scanner." *Medical Physics*, **25** (10), pp. 2046-2053, 1998.

[24] M. Hofmann, B. Plicher, B. Schölkopf and T. Beyer. "Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques." *European Journal of Nuclear Medicine and Molecular Imaging*, **36** (S1), pp. S96-S104, 2009.

[25] V. Keereman, P. Mollet, Y. Berker, V. Schulz, S. Vandenberghe. "Challenges and current methods for attenuation correction in PET/MR." *MAGMA*, **26** (1), pp. 81-98, 2013.

[26] A. Rezaei, M. Defrise, G. Bal, C. Michel, M. Conti, C. Watson and J. Nuyts. "Simultaneous reconstruction of activity and attenuation in time-of-flight PET." *IEEE Transactions Medical Imaging*, **31** (12), pp. 2224-2233, 2012

[27] H. Ziadi. "Comparative evaluation of scatter correction techniques in 3D positron emission tomography." *European Journal of Nuclear Medicine*, **27** (12), pp. 1813-1826, 2000.

[28] C. Stearns, D. McDaniel, S. Kohlmeyer, P. Arul, B. Geiser and V. Shanmugan. "Random Coincidence Estimation from Single Event Rates on the Discovery ST PET/CT Scanner." *Conference record IEEE Medical Imaging Conference, Portland, Oregan, USA*, **5**, pp. 3067-3069, 2003.

[29] C. Watson. "Co-fan-sum ratio algorithm for randoms smoothing and detector normalization in PET". *Conference record IEEE Medical Imaging Conference, Knoxville, TN, USA*, pp. 3326-3329, 2010.

[30] C. Levin and E. Hoffman. "Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution." *Physics in Medicine and Biology*, **44**, pp. 781-799, 1999.

[31] A. G. Ghom. *Textbook of oral radiology.* Elsevier, Noida/India, 2008.

[32] C. Meinhold, J. Taschner and W. Inkret. "Radiation and Risk - A hard look at the data: A brief History of radioation." *Los Alamos Science*, **23**, pp. 116-123, 1995.

[33] N. Suntharalingam, E. Podgorsak and J. Hendry. *Review of radiation oncology physics: a handbook for teachers and students. Chapter 14: Basic Radiobiology.* IAEA, Vienna, 2005.

[34] SCK-CEN: http://www.sckcen.be/nl/Wetenschap-Samenleving/Gezondheid-Veiligheid/Gezondheid-Straling/Blootstelling-aan-straling. (Accessed 6th June 2013).

[35] H. Barrett and K. Myers. *Foundations of Image Science.* John Wiley and Sons, Hoboken/New Jersey, 2004.

[36] L. A. Feldkamp, L. C. Davis, and J. W. Kress. "Practical cone-beam algorithm.", *Journal of the optical society*, **1**(6), pp. 612-619, 1984.

[37] A. Katsevich. "Theoretically Exact Filtered Backprojection-Type Inversion Algorithm for Spiral CT." *SIAM Journal in Applied Mathematics*, **62**(6), pp. 2012-2026, 2012.

[38] F. Noo, J. Pach and D. Heuscher. "Exact helical reconstruction using native cone-beam geometries." *Physics in Medicine and Biology*, **48**, pp. 3787-3818, 2003.

[39] M. Defrise, P. Kinahan, D. Townsend, C. Michel, M. Sibomana and D. Newport. "Exact and approximate rebinning algorithms for 3-D PET data." *IEEE Transactions Medical Imaging*, **16** (2), pp. 145-158, 1997.

[40] M. Defrise. *Short course: medical image reconstruction, part 2.* NSS/MIC, Knoxville, 2010

[41] P. Gilbert. "Iterative methods for the three-dimensional reconstruction of an object from projections." *Journal Theoretical Biology*, **36**, pp. 105-177, 1971.

[42] R. Gordon, R. Bender, and G. Herman. "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and x-ray photography." *Journal Theoretical Biology*, **29**, pp. 471-481, 1970.

[43] G. Hounsfield. "A method and apparatus for examination of a body by radiation such as X or Gamma radiation." Patent Specification 1283915, London, England, 1972.

[44] A.H. Andersen and A.C. Kak. "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm." *Ultrasonic Imaging*, **6**, pp. 81-94, 1984.

[45] L.A. Shepp and V. Vardi. "Maximum likelihood reconstruction for emission tomography." *IEEE Transactions Medical Imaging*, **11**(2), pp. 113-122, 1982.

[46] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, L. Mortelmans. "Iterative reconstruction for helical CT: a simulation study." *Physics Medicine Biology*, **43**, pp 729-737, 1998.

[47] J.A. Fessler, E.P. Ficaro, N.H. Clinthorne and K. Lange. "Grouped-coordinate ascent algorithm for penalized-likelihood transmission image reconstruction." *IEEE Transactions on Medical Imaging*, **16** (2), pp. 166-175, 1997.

[48] H. Hudson and R. Larkin. "Accelerated image reconstruction using ordered subsets of projection data." *Journal Theoretical Biology*, **13** (4), pp. 601-609, 1994.

[49] J. Qi and R. Leahy. "Iterative reconstruction techniques in emission computed tomography", *Physics in Medicine and Biology*, **51** (15), pp. R541-R578, 2006.

[50] E. Mumcuoğlu, R. Leahy and S. Cherry. "Bayesian reconstruction of PET images: methodology and performance analysis." *Physics in Medicine and Biology*, **41**, pp. 1777-1807, 1996.

[51] E. Sidky, Y. Duchin, X. Pan and C. Ullberg. "A constrained, total-variation minimization algorithm for low-intensity x-ray CT." *Medical Physics*, **38**, pp. S117-S125, 2011.

[52] E. Mumcuoğlu, R. Leahy, S. Cherry and E. Hoffman. "Accurate geometric and physical response modelling for statistical image reconstruction in high resolution PET." *Conference record IEEE Medical Imaging Conference, San Francisco, CA, USA*, **3**, pp. 1569-1573, 1993.

[53] A. Rahmin. "Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls." *Medical Physics*, **40** (6), pp. 064301 (1-16), 2013.

[54] J.F. Barrett and N. Keat. "Artifacts in CT: Recognition and avoidance." *RadioGraphics*, **24**, pp. 1679-1691, 2004.

[55] B. De Man, J. Nuyts, P. Dupont, G. Marchal and P. Suetens. "Metal streak artifacts in X-ray computed tomography: a simulation study." *IEEE Transactions Nuclear Science*, **46** (3), pp. 691-696, 1999.

[56] E.-P. Rührnschopf and K. Klingenbeck. "A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. Part 1: Scatter compensation approaches." *Medical Physics* **37** (7), pp. 4296-4311, 2011.

[57] E.-P. Rührnschopf and K. Klingenbeck. "A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. Part 2: Scatter estimation approaches." *Medical Physics* **38** (9), pp. 5186-5199, 2011.

[58] P. Joseph and R. Spital. "The exponential edge-gradient effect in x-ray computed tomography." *Physics in Medicine and Biology*, **26** (3), pp. 473-487, 1981.

[59] J.-H. Kim, J. Nuyts, Z. Kuncic and R. Fulton. "The feasibility of head motion tracking in helical CT: A step toward motion correction." *Medical Physics*, **40** (4), pp. 041903 (1-4), 2013.

[60] C. Rohkohl, H. Bruder, K. Stierstorfer and T. Flohr. "Improving best-phase image quality in cardiac CT by motion correction with MAM optimization." *Medical Physics*, **40** (3), pp. 031901 (1-15), 2013

[61] A. Isola, M. Grass1 and Wiro J. Niessen. "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction." *Medical Physics*, **37** (3), pp. 1093-1109, 2010.

[62] S. Rit, D. Sarrut, L. Desbat. "Comparison of Analytic and Algebraic Methods for Motion-Compensated Cone-Beam CT Reconstruction of the Thorax." *IEEE Transactions on Medical Imaging*, **28** (10), pp. 1513-1525, 2009.

[63] W. Kalender, R. Hebel and J. Ebersberger. "Reduction of CT artifacts caused by metallic implants." *Radiology*, **164**, pp. 567-568, 1987.

[64] G. Glover and N. Pelc. "An algorithm for the reduction of metal clip artifacts in CT reconstructions." *Medical Physics*, **8** (6), pp. 799-807, 1981.

[65] A. Mahnken, R. Raupach, J.E. Wildberger, B. Jung, N. Heussen, T.G. Flohr, R.W. Günther and S. Schaller. "A new algorithm for metal artifact reduction in computed tomography, in vitro and in vivo evaluation after total hip replacement." *Investigative Radiology*, **38** (12), pp. 769-775, 2003.

[66] C. Lemmens, D. Faul and J. Nuyts. "Suppression of metal artifacts in CT using a reconstruction procedure that combines MAP and projection completion." *IEEE Transactions Medical Imaging*, **28** (2), pp. 250-260, 2009.

[67] J. Müller and T.M. Buzug. "Spurious structures created by interpolation-based CT metal artifact reduction." *Proceedings of SPIE*, **7258** (1), pp. Y1-Y8, 2009.

[68] E. Meyer, R. Raupach, M. Lell, B. Schmidt and M. Kachelrieß. "Normalized metal artifact reduction (NMAR) in computed tomography." *Medical Physics*, **37** (10), pp. 5462-5493, 2010.

[69] M. Bal and L. Spies. "Metal artifact reduction in CT using tissue-class modeling and adaptive prefiltering." *Medical Physics*, **33** (8), pp. 2852-2559, 2006.

[70] E. Meyer, R. Raupach, M. Lell, B. Schmidt and M. Kachelrieß. "Frequency split metal artifact reduction in computed tomography." *Medical Physics*, **39** (4), pp. 1940-1916, 2012.

[71] O. Watzke and W.A. Kalender. "A pragmatic approach to metal artifact reduction in CT: merging of metal artifact reduced images." *European Radiology*, **14**, pp. 849-856, 2004.

[72] E. Meyer, R. Raupach, B. Schmidt, A.H. Mahnken and M. Kachelrieß. "Adaptive normalized metal artifact reduction (ANMAR) in computed tomography." *Conference record IEEE Medical Imaging Conference, Valencia, Spain*, pp. 2560-2565, 2011.

[73] S. Tohnak, A.J.H. Mehnert, M. Mahoney and S. Crozier. "Dental CT metal artefact reduction based on sequential substitution." *Dentomaxillofacial Radiology*, **40**, pp. 184-190, 2011.

[74] M. Yazdi, M.A. Lari, G. Bernier and L. Beaulieu. "An opposite view data replacement approach for reducing artifacts due to metallic dental objects." *Medical Physics*, **38** (4), pp. 2275-2281, 2011.

[75] J. Williamson, B. Whiting, J. Benac, R. Murphy, G. Blaine, J. O'Sullivan, D. Politte and D. Snyder. "Prospects for quantitative computed tomo-

graphy imaging in the presence of foreign metal bodies using statistical image reconstruction." *Medical Physics*, **29** (10), pp. 2404-2418, 2002.

[76] B. De Man, J. Nuyts, P. Dupont, G. Marchal and P. Suetens. "Iterative maximum-likelihood polychromatic algorithm for CT." *IEEE Transactions Medical Imaging*, **20** (10), pp. 999-1008, 2001.

[77] I. Elbakri and J. Fessler. "Statistical image reconstruction for polyenergetic X-ray computed tomography." *IEEE Transactions Medical Imaging*, **21**(2), pp. 89-99, 2002.

[78] N. Menvielle, Y. Goussard, D. Orban and G. Soulez. "Reduction of beam-hardening artifacts in X-ray CT." *Proceeding on the IEEE Engineering in Medicine and Biology Conference, Shanghai, China*, pp. 1865-1868, 2005.

[79] I. Elbakri and J. Fessler. "Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation." *Physics Medicine Biology*, **48**, pp. 2453-2477, 2003.

[80] B. De Man. *Iterative reconstruction for reduction of metal artifacts in computed tomography.* PhD thesis, Katholieke Universiteit Leuven, Belgium, 2001.

[81] W. Zbijewski and F.J. Beekman. "Characterization and suppression of the edge and aliasing artifacts in iterative x-ray CT reconstruction." *Physics Medicine Biology*, **49** (1), pp. 145-157, 2004.

[82] G. Herman. "Correction for beam hardening in computed tomography." *Physics Medicine Biology*, **24** (1), pp. 81-106, 1979.

[83] J.A. Fessler and K. Donghwan. "Axial block coordinate descent (ABCD) algorithm for X-ray CT image reconstruction." *Proceedings on the Meeting on Fully 3D image Reconstruction in Radiology and Nuclear Medicine, Potsdam, Germany*, pp. 262-265, 2011.

[84] B. De Man and S. Basu. "Distance-driven projection and backprojection in three dimensions." *Physics Medicine Biology*, **49** (11), pp. 2463-2475, 2004.

[85] J.H. Siewerdsen, A.M. Waese, D. J. Moseley, S. Richard and D.A. Jaffray. "Spektr: A computational tool for x-ray spectral analysis and imaging system optimization." *Medical Physics*, **3** (1), pp. 3057-3067, 2004.

[86] M. Kachelrieß, O. Watzke and W.A. Kalender. "Generalized multi-dimensional adaptive filtering for convectional and spiral single-slice, multi-slice, and cone-beam CT." *Medical Physics*, **28** (4), pp. 475-490, 2001.

[87] B. Hamelin, Y. Goussard, J. Dussault, G. Cloutier, G. Beaudoin and G. Soulez. "Design of iterative ROI transmission tomography reconstruction

procedures and image quality analysis." *Medical Physics*, **37** (9), pp. 4577-4589, 2010.

[88] B. Hamelin, Y. Goussard, D. Gendron, J. Dussault, G. Cloutier, G. Beaudoin and G. Soulez. "Iterative reconstruction of real data with metal artifact reduction." *IEEE Symposium on Biological Imaging, Paris, France*, pp. 1453-1456, 2008.

[89] Z. Yu, J. Thibault, C.A. Bouman, K.D. Sauer and J. Hsieh. "Edge-localized iterative reconstruction for computed tomography." *Proceedings on the Meeting on Fully 3D image Reconstruction in Radiology and Nuclear Medicine, Beijing, China*, pp. 255-258, 2009.

[90] K. Van Slambrouck and J. Nuyts. "Accelerated convergence with image-block iterative reconstruction." *Conference record IEEE Medical Imaging Conference, Valencia, Spain*, pp. 2576-2580, 2011.

[91] Z. Yu, J. Thibault, C.A. Bouman, K.D. Sauer and J. Hsieh. "Fast model-based X-ray CT reconstruction using spatially non-homogeneous ICD optimization." *IEEE Transactions Image Processing*, **20** (1), pp. 161-175, 2011.

[92] B. De Man, J. Nuyts, P. Dupont, G. Marchal, P. Suetens. "Reduction of metal streak artefacts in x-ray computed tomography using a transmission maximum a posteriori algorithm." *IEEE Transactions Nuclear Science*, **47** (3), pp. 977-981, 2000.

[93] L.J. Meng and N. Li. "Non-uniform object-space pixelation (NUOP) for penalized maximum-likelihood image reconstruction for a single photon emission microscope system." *IEEE Transactions Nuclear Science*, **56** (5), pp. 2777-2788, 2009.

[94] J.S. Maltz. "Optimal time-activity basis selection for exponential spectral analysis: application to the solution of large dynamic emission tomographic reconstruction problems." *IEEE Transactions on Nuclear Science*, **49** (4), pp. 1452-1464, 2001.

[95] J.G. Brankov and M.N. Wernick. "Tomographic image reconstruction using content-adaptive mesh modeling." *Proceeding on the IEEE Conference on Image Processing, Thessaloniki, Greece*, pp. 690-693, 2001.

[96] M.J. Willemink, T. Leiner, P.A. de Jong, L.M. de Heer, R.A.J. Nievelstein, A.M.R. Schilham and R.P.J. Budde. "Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality." *European Radiology* - published online on 16 January 2013.

[97] A.K. Hara, R.G. Paden, A.C. Silva, J.L. Kujak, H.J. Lawder and W. Pavlicek. "Iterative Reconstruction Technique for Reducing Body Radiation Dose at CT: Feasibility Study." *Gastrointestinal Imaging*, **193**, pp. 764-771, 2009.

[98] M. Katsura, I. Matsuda, M. Akahane, J. Sato, H. Akai, K. Yasaka, A. Kunimatsu and K. Ohtomo. "Model-based iterative reconstruction technique for radiation dose reduction in chest CT: comparison with the adaptive statistical iterative reconstruction technique." *European Radiology*, **22**, pp. 1613-1623, 2012.

[99] E.Y. Sidky, Y. Duchin, X. Pan and C. Ullberg. "A constrained, total-variation minimization algorithm for low-intensity x-ray CT." *Medical Physics*, **38** (S1), pp. S117-S125, 2011.

[100] J. Wang, T. Li and L. Xing. "Iterative image reconstruction for CBCT using edge-preserving prior." *Medical Physics*, **36** (1), pp. 252-260, 2009.

[101] P.J. La Rivière. "Penalized-likelihood sinogram smoothing for low-dose CT." *Medical Physics*, **32** (6), pp. 1676-1683, 2005.

[102] M.E. Daube-Witherspoon and G. Muehllehner. " An iterative image space reconstruction algorithm suitable for volume ECT." *IEEE Transactions Medical Imaging*, **5** (2), pp. 61-66, 1986.

[103] G.E.B. Archer and D.M. Titterington. "The iterative image space reconstruction algorithm (ISRA) as an alternative to the EM algorithm for solving positive linear inverse problems. *Statistica Statica*, **5**, pp. 77-96, 1995.

[104] A.R. De Pierro. "In the relation between ISRA and the EM algorithm for positron emission tomography, *IEEE Transactions on Medical Imaging*, **12** (2), pp. 328-333, 1993.

[105] J. Nuyts, S. Stroobants, P. Dupont, S. Vleugels, P. Flamen and L. Mortelmans. "Reducing loss of image quality due to the attenuation artifact in uncorrected PET whole body images." *Journal of Nuclear Medicine*, **43**, pp. 1054-1062, 2002.

[106] C. Byrne. "Iterative algorithms for deblurring and deconvolution with constraints. *Inverse Problems*, **14**, pp. 1455-146, 1998.

[107] K. Lange and F. A. Fessler. "Globally convergent algorithms for maximum a posteriori transmission tomography." *IEEE Transactions in Image Processing*, **4** (10), pp. 1430-1438, 1995.

[108] H. Erdoğan and J. A. Fessler. "Ordered subsets algorithms for transmission tomography." *Physics in Medicine and Biology*, **44**, pp. 2835-2851, 1999.

[109] C. Kamphuis and F. J. Beekman. "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm." *IEEE Transaction on Medical Imaging*, **17** (6), pp. 1101-1105, 1998.

[110] K. Sauer and C. Bouman. "A local update strategy for iterative reconstruction from projections." *IEEE Transactions Signal Processing*, **41** (2), pp. 534-548, 1993.

[111] J. A. Fessler and A. O. Hero. "Space-alternating generalized expectation maximization algorithm." *IEEE Transactions on Signal Processing*, **42** (10), pp. 2664-2677, 1994.

[112] K. Van Slambrouck and J. Nuyts. "Metal artifact reduction in computed tomography using local models in an image block-iterative scheme." *Medical Physics*, **39** (11), pp. 7080-7093, 2012.

[113] T. Benson, B. De Man, L. Fu, J.-B. Thibault. "Block-based Iterative coordinate descent." *Proceeding of the IEEE Medical Imaging Conference, Knoxville, TN, USA*, pp. 2856-2859, 2010.

[114] J. Nuyts, S. Stute, K. Van Slambrouck, F. van Velden, R. Boellaard, and C. Comtat. "Maximum-likelihood reconstruction based on a modified Poisson distribution to reduce bias in PET." *Conference record IEEE Medical Imaging Conference, Valencia, Spain*, pp. 4337-4341, 2011.

[115] K. Michielsen, K. Van Slambrouck, A. Jerebko and J. Nuyts. "Patchwork Reconstruction with Resolution Modeling for Digital Breast Tomosynthesis." *Medical Physics*, **40** (3), pp. 031105 (1-10), 2012.

[116] Y. Takashi, S. Mori, T. Kozuka, K. Gomi, T. Nose, T. Tahara, M. Ogichi and T. Yamashita. "Preliminary study of correction of original metal artefacts due to I-125 seeds in postimplant dosimetry for prostate permanent implant brachytherapy." *Radiation Medicine*, **24** (2), pp. 133-138, 2006.

[117] Philips Healthcare: Metal Artifact Reduction for Orthopedic Implants (O-MAR) http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/CT/Metal-Artifact-Reduction-for-Orthopedic-Implants-%28O-MAR%29.

[118] T. Fluchs. "Strahlaufhärtungskorrekturen in der Computer-Tomographie." PhD thesis, Friedrich-Alexander Universität Erlangen-Nürnberg, 1998.

[119] G. Van Gompel, K. Van Slambrouck, M. Defrise, K.J. Batenburg, J. de Mey, J. Sijbers and J. Nuyts. "Iterative correction of beam hardening artifacts in CT", *Medical Physics*, **38** Suppl 1: S36-S49, 2011

[120] K. Van Slambrouck and J. Nuyts. "Reconstruction scheme for accelerated maximum likelihood reconstruction: the patchwork structure." (Submitted to *IEEE Transactions on Nuclear Science*: minor revision), 2013.

[121] R. Alvarez and A. Macovski. "Energy-selective reconstructions in X-ray computerized tomography." *Physics in medicine and biology*, **21** (5), pp. 733-744, 1976.

[122] Siemens Healthcare, Application Guide Somatom Sensation 10/16: http://www.medical.siemens.com/siemens/en_GB/gg_ct_FBAs/files/CIP/appl_guides/sensation/ApplicationGuide_Sensation16.pdf.

[123] Siemens Healthcare, Application Guide Somatom Sensation 64: http://www.medical.siemens.com/siemens/en_GB/gg_ct_FBAs/files/CIP/appl_guides/sensation/ApplicationGuide_Sensation64.pdf.

[124] R. Carmi, O. Shapiro and D. Braunstein. "Resolution enhancement of x-ray ct by spatial and temporal MLEM deconvolution correction"." *Conference record IEEE Medical Imaging Conference, Rome, Italy*, **5**, pp. 2765-2768, 1994.

[125] B. Ohnesorge, T. Flohr and K. Klingenback-Regn. "Efficient object scatter correction algorithm for third and fourth generation CT scanners." *European Radiology*, **9**, pp. 563-569, 1999.

[126] K. Stierstorfer, T. Flohr and H. Bruder. "Segmented multiple plane reconstruction: a novel approximate reconstruction scheme for multi-slice spiral CT." *Physics in Medicine and Biology*, **47**, pp. 2571-2581, 2002.

[127] T. Flohr, K. Stierstorfer, H. Bruder, J. Simon, A. Polacin, and S. Schaller. "Image reconstruction and image quality evaluation for a 16-slice CT scanner." *Medical Physics*, **30** (5), pp. 832-845, 2003.

[128] K. Lange and R. Carson. "EM reconstruction for emission and transmission tomography." *Journal of Computed Assisted Tomography*, **8** (2), pp. 306-316, 1984.

[129] F.H.P. van Velden, R.W. Kloet, B.N.M. van Berckel, S.P.A. Wolfensberger, A.A. Lammertsma and R. Boellaard. "Comparison of 3DOP-OSEM and 3D-FBP reconstruction algorithms for High-Resolution research tomograph studies: effects of randoms estimation methods." *Physics in Medicine and Biology*, **53**, pp. 3217-3230, 2008.

[130] R. Boellaard, A. van Lingen and A. A. Lammertsma. "Experimental and Clinical Evaluation of Iterative reconstruction (OSEM) in dynamic PET: quantitative characteristics and effects on kinetic modeling. *Journal of Nuclear Medicine*, **42** (5), pp. 808-817, 2001.

[131] A. Reilhac, S. Tomeï, I. Buvat, C. Michel, F. Keheren and N. Costes. "Simulation-based evaluation of OSEM iterative methods in dynamic brain PET studies." *NeuroImage*, **39**, pp. 359-268, 2008.

[132] C. Cloquest and M. Defrise. "MLEM and OSEM deviate from the Cramer-Rao bound at low counts." *IEEE Transactions on Nuclear Science*, **60** (1), pp. 134-143, 2013.

[133] K. Erlandsson, D. Visvikis, W. Waddington, I. Cullum, P. Jarritt and L. Polowsky. "Low-statistics reconstruction with AB-EMML." *Conference record IEEE Medical Imaging Conference, Lyon*, **2**, pp. 15/249-253, 2000

[134] J. Verhaeghe and A. Reader. "AB-OSEM reconstruction for improved Patlak kinetic parameter estimation: a simulation study." *Physics in Medicine and Biology*, **55**, pp. 6739-6757, 2010.

[135] L. Grèzes-Besset, J. Nuyts, R. Boellaard, I. Buvat, C. Michel, C. Pierre, N. Costes and A. Reilhac. "Simulation-based evaluation of NEG-ML iterative reconstruction of low count PET data." *Conference record IEEE Medical Imaging Conference, Hawaii*, pp. 3009-3014, 2007.

[136] S. Ahn and J. Fessler. "Emission image reconstruction for randoms-precorrected PET allowing negative sinogram values." *IEEE Transactions Medical Imaging*, **23** (5), pp. 591-601, 2004.

[137] M. Yavuz and J. Fessler. "Statistical image reconstruction methods for randoms-precorrected PET scans." *Medical Image Analysis*, **2** (4), pp. 369-378, 1999.

[138] M. D. Silver, K. Taguchi, I. A. Hein, B. Chiang, M. Kazama, I. Mori. "Windmill artifact in multislice helical CT." *Proceeding of SPIE*, **5032**, pp. 1918-1927, 2003.

[139] M. Depypere, J. Nuyts, N. van Gastel, G. Carmelier, F. Maes and P. Suetens. "An iterative dual energy CT reconstruction method for K-edge contrast material." *Proceeding of SPIE*, **7961** (79610), M1-M7, 2011.

[140] European Commission: Directorate-General for Energy, Directorate D, Nuclear Safety and Fuel Cycle, Unit D4, Radiation Protection. *Radiation Protection 162: Criteria for Acceptability of Medical Radiological Equipment used in Diagnostic Radiology, Nuclear Medicine and Radiotherapy*, European Union, Luxembourg, 2012.

[141] R. Pauwels, H. Stamatakis, H. Bosmans, R. Bogaerts, R. Jacobs, K. Horner, K. Tsiklakis and The SEDENTEXCT project consortium. "Quantification of metal artifacts on cone beam tomography images." *Clinical Oral Implants Research*, **24** (A100): pp. 94-99, 2013.

[142] F.E. Boas, D. Fleischmann. "Evaluation of two iterative techniques for reducing metal artifacts in computed tomography." *Radiology*, **259**(3), pp. 894-902, 2011.

[143] G. Pratx and L. Xing. "GPU computing in medical physics: a review." *Medical Physics*, **38** (5), pp. 2685-2697, 2011.

[144] W.J. Palenstijn, K J. Batenburg, and J. Sijbers. "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs).", *Journal of Structural Biology*, **176** (2), pp. 250-253, 2011.

[145] P. Noël, A. Walczaka, J. Xua, J. Corsoa, K. Hoffmann and S. Schafera. "GPU-based cone beam computed tomography.", *Computer Methods and Programs in Biomedicine*, **9** (8), pp. 271-277, 2010.

[146] J. Agulleiro, F. Vazquez, E.M. Garzon, J. Fernandez."Hybrid computing: CPU+GPU co-processing and its application to tomographic reconstruction." *Ultramicroscopy*, **115**, pp. 109-114, 2012.

[147] S. Stute, J. Nuyts, K. Van Slambrouck, M. Sibomana, F. van Velden, R. Boellaard, C. Comtat. "Image properties of various ML-based reconstructions of very noisy HRRT data." *Conference record IEEE Medical Imaging Conference, Valencia, Spain*, pp. 4311-4315, 2011.