Graz University of Technology

Master's Thesis

Scattered radiation reduction in digital volume tomography for dental applications

conducted at the
Institute of Medical Engineering
Graz University of Technology, Austria

in cooperation with
Sirona Dental Systems GmbH
Bensheim, Germany

by
Lukas Franz Muhr

Supervisors:
Univ.-Prof. Dipl.-Ing. Dr. Rudolf Stollberger, Graz University of Technology
Dipl.-Phys. Dr. Kai Stannigel, Sirona Dental Systems GmbH

Graz, January 19, 2017

Abstract

Cone-beam computed tomography (CBCT) typically suffers from scattered X-rays leading to inaccuracies of CT numbers and artifacts, such as cupping or streaks. A vast variety of scattered radiation estimation and compensation techniques exist. The improved primary modulator scatter estimation (iPMSE) method is based on a spatial modulation of the primary intensity, shifting this desired signal towards higher spatial frequencies. In contrast, the scattered radiation generated in the object is composed of mainly low spatial frequencies. With the modulator in the beam path, seperation of the two signals is possible in either the Fourier- or the image space. In this thesis, the iPMSE method has been implemented and tested on a CBCT device for dental applications. While the iPMSE method has many advantages (e.g. robustness, short computation time), its patch-wise constant approach for the scattered radiation profile may lead to inaccuracies close to sharp object edges in the projection images. To avoid those, a novel convolution-based ansatz was developed in this thesis, in which the scattered radiation profile is estimated by convolving a scatter potential derived from measured intensities with a physically motivated kernel, yielding a hybrid iPMSE (hiPMSE) method. The method was implemented and tested on both simulation and measurement data and showed promising results in both cases where scatter artifacts could be substantially reduced.

Keywords: Cone-beam computed tomography, scattered radiation, artifacts, primary modulator, scatter kernel

Kurzfassung

In der Kegelstrahl-Computertomographie treten typischerweise gestreute Röntgenstrahlen auf, die zu Ungenauigkeiten in den rekonstruierten Grauwerten und zu Cupping- und Streifen-Artefakten führen. Es existieren viele verschiedene Streustrahlschätzungs- und Kompensierungstechniken. Eine davon ist die Improved Primary Modulator Scatter Estimation (iPMSE) Methode, die auf der räumlichen Modulation der Primärintensität basiert und so zu einer Verschiebung dieses Signals hin zu höheren Frequenzen führt. Im Gegensatz dazu besteht die Streustrahlung, die erst im Objekt entsteht, aus hauptsächlich niedrigen räumlichen Frequenzen. Befindet sich der Modulator im Strahlengang, so ist eine Trennung der beiden Signale nach der Messung entweder im Fourier- oder im Bildraum möglich. In dieser Arbeit wurde die iPMSE-Methode auf einem Kegelstrahl-Computertomographie-Gerät für dentale Anwendungen implementiert und getestet. Die iPMSE Methode hat viele Vorteile (z.B. Robustheit, kurze Berechnungszeit), jedoch führt die Annahme, dass die Streustrahlung auf kleinen Teilbildern konstant ist, zu Ungenauigkeiten an starken Objektkanten in den Projektionsbildern. Um diese Ungenauigkeiten zu vermeiden wurde eine neue hybride Methode entwickelt (hiPMSE). Mit dieser Methode wird die Streustrahlung mittels Faltung eines Streupotentials, das aus den gemessenen Intensitäten abgeleitet wurde, mit einem Streukern geschätzt. Die neue Methode wurde sowohl auf Simulations- als auch auf Messdaten implementiert und getestet, und führte in beiden Fällen zu einer wesentlichen Reduktion von Streustrahlartefakten.

Schlüsselwörter: Kegelstrahl-Computertomographie, Streustrahlung, Artefakte, Primärmodulation, Streukern

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

date	(signature)

Sperrvermerk

Die vorgelegte Masterarbeit mit dem Titel 'Scattered radiation reduction in digital volume tomography for dental applications' basiert auf internen, vertraulichen Daten und Informationen der Sirona Gruppe. Diese Masterarbeit darf nur vom Erst- und Zweitgutachter sowie berechtigten Mitgliedern des Prüfungsausschusses eingesehen werden. Eine Vervielfältigung und Veröffentlichung dieser Masterarbeit ist auch auszugsweise nicht erlaubt. Die Weitergabe des Inhalts der Arbeit im Gesamten oder in Teilen sowie das Anfertigen von Kopien oder Abschriften - auch in digitaler Form - sind grundsätzlich untersagt. Dritten darf dieser Bericht nur mit der ausdrücklichen schriftlichen Genehmigung der Sirona Dental Systems GmbH zugänglich gemacht werden. Diese Regelung hat bis zum 26.01.2022 Bestand.

Acknowledgements

I am grateful for Prof. Dr. Rudolf Stollberger's decision to let me conduct a thesis with an industrial partner and I am thankful for his supervision of this thesis.
I owe thanks to Dr. Kai Stannigel for his expert guidance and many interesting discussions and suggestions during the course of this thesis.
I also want to thank my colleagues from the 'research and innovation' and 'image quality and algorithms' groups at Sirona Dental Systems for incorporating me in their team and many shared meals and discussions.
Prof. Dr. Marc Kachelrieß and Joscha Maier from the German Cancer Research Center (DKFZ) in Heidelberg has to be thanked for suggesting to combine primary modulator scatter estimation with a convolution-based method and for many fruitful discussions. I also owe thanks to Joscha Maier for conducting the Monte-Carlo simulations.
I would like to thank my parents, Roswitha and Franz, for their financial and emotional support during the last years. Thank you for giving me the opportunity to study. I am very grateful. I also would like to thank my sister Julia, my brother Maximilian and my grandmother Rosa for their support.

Contents

List of Figures

1 Introduction

1.1 Introduction

Computed tomography (CT) is a non-invasive imaging modality employing X-ray projection images from different angles to reconstruct sectional images of the attenuation coefficient of the imaged patient or object. Since its invention in the 1960s, several generations of CT machines have been developed, all with the aim of reducing scan time and patient dose and increasing accuracy and image quality. CT imaging for dental diagnostics and radiation therapy employs cone-beam CT (CBCT) with two dimensional detectors with the objective of acquiring a complete volume in a single rotation [1].

In contrast to classical single-slice CT with a fan beam, CBCT with a flat-panel imager or an image amplifier typically suffers from larger amounts of scattered radiation due to greater irradiated volumes. The extent of scatter is much larger than in the single-slice case and can even exceed the primary signal behind strongly absorbing structures [2]. Scatter in turn violates the assumed linear relationship between measured projection values and the line integrals of attenuation coefficients along an intersection path and leads to decreased image quality and therefore to a loss of information and diagnostic usability. Although different in physical origin, scatter artifacts resemble those known from beam hardening. In projection images scatter leads to a loss of contrast and in volume data the presence of scatter manifests itself in cupping and streak artifacts and inaccuracies of CT numbers [3][4].

In the dental community CBCT is known as digital volume tomography (DVT) and it has manifold applications in dentistry and orthodontics. The main indication is implantology but a DVT scan is justified whenever sectional images may offer new insights that were not available with conventional X-ray imaging [5]. Scatter is, along with beam hardening, among the biggest sources for degraded image quality. Thus, the prevention and correction of artifacts due to scatter is a major challenge in CBCT and DVT and a fast and accurate method to do so is desirable [2].

Scatter correction methods can generally be divided into five approaches:

- physical scatter rejection
- analytical modeling
- Monte Carlo (MC) simulations

- scatter measurements
- primary modulation

Physical scatter rejection techniques aim at lowering the amount of scatter impinging on the detector by means of an anti-scatter grid (ASG), an air gap, collimation or a bow-tie filter [2]. These techniques can significantly lower the amount of scattered radiation reaching the detector, but usually additional correction is needed. The use of an ASG lowers both the scatter and primary radiation reaching the detector and thus radiation dose must be raised to achieve the same noise level as without ASG. In analytical modeling methods a scatter potential derived from measured intensities is convolved with a scatter kernel to yield the desired scatter distribution [6–10]. As with all scatter estimation techniques, the estimated scatter is then simply subtracted from the total signal. From a physical point of view, MC simulations are the most accurate scatter estimation techniques since the complete imaging process is modeled [11]. However, this high accuracy comes at the cost of a high computational burden that is clinically not acceptable. MC simulations are therefore often implemented with a low number of simulated photon trajectories and combined with other approaches to lower the amount of calculation time [9][12]. Scatter can also be directly measured behind strong absorbers where no primary signal can reach the detector. With the use of a beam-absorber array (BAA) the scatter signal can partly be measured and partly be interpolated under the assumption that scatter is predominately a low-frequency background signal with little to no spatial information about the object being imaged [13]. An excellent overview of current approaches in scatter correction is given by Rührnschopf and Klingenbeck in [2] and [14].

Recently, a promising hybrid method named the Primary Modulator Scatter Estimation (PMSE) technique has been proposed [15]. This method aims at simultaneously measuring the scatter and primary intensities without increased patient dose or scan time and, compared to other methods, low computational cost. The key element of PMSE is a primary modulator of spatially varying attenuation. The modulator is mounted between the X-ray source and the object of interest, resulting in a shift of the incident signal to higher spatial frequencies. While the primary signal, which is party of the signal that is not scattered, retains this modulation until it impinges on the detector, the scatter signal, which originates in the object, loses the high frequency modulation due to the scatter process and displays a mainly low-frequency background signal. Figure 1.1 illustrates this situation. Separation of the two desired signals has been demonstrated using either a Fourier-based approach [15] or an image-based approach. The present work is based on an image-based approach which is named the Improved Primary Modulator Scatter Estimation technique (iPMSE). This method has the advantage of being able to cope with highly irregular modulator patterns [16][17].

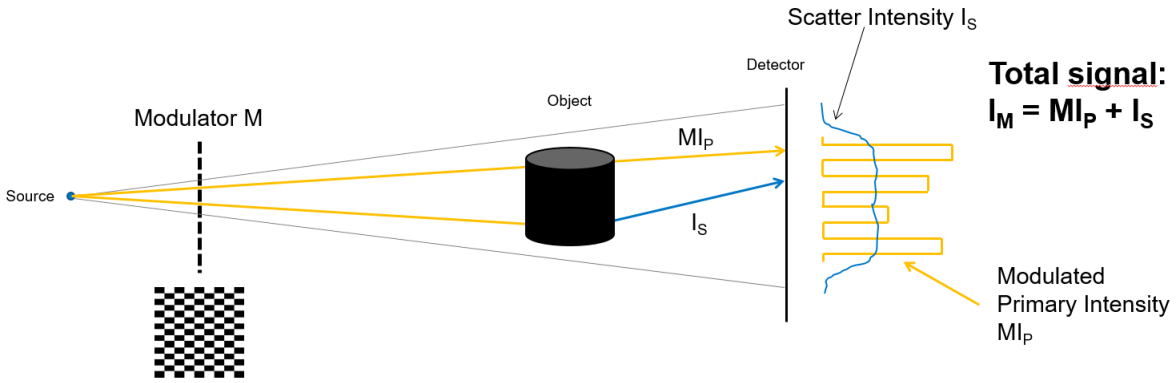


Figure 1.1: The primary modulator is mounted between the X-ray source and the object, resulting in a shift of the incident primary signal to higher spatial frequencies. The scatter signal I_S does not show this modulation when it impinges on the detector due to the braod point spread function of the underlying scatter process. The measured signal on the detector is the sum of both the modulated primary radiation and the scattered radiation: $I_M = MI_P + I_S$ [16].

1.2 Objective

The main goal of this thesis is to implement and further enhance the iPMSE technique in the context of dental DVT and to assess the achievable image quality enhancements of this method. Furthermore, a general quantification of scatter magnitudes and scatter effects for typical dental DVT geometries and objects should be carried out.

While the iPMSE method has many advantages, its patch-wise constant approach for the scattered radiation profile may lead to inaccuracies close to sharp object edges in the projection images. In order to avoid those and to incorporate the physics of scatter generation, a convolution-based approach has been developed. This approach estimates the scattered radiation profile by convolving a scatter potential derived from measured intensities with a physically motivated kernel. This novel method is titled Hybrid Primary Modulator Scatter Estimation technique (hiPMSE).

This thesis is structured as follows: In section 2, a brief introduction to the principles of CT with a focus on scattered radiation is given. Section 3 introduces the concepts of the iPMSE and hiPMSE methods and explains their practical implementation. Subsequently, the capability of scatter correction of both algorithms is assessed for both simulation and measured data in section 4. Section 5 concludes with a discussion of achieved image quality improvements, practical problems and possible directions for future work.

2 Background

This chapter gives a short overview on the interaction of X-rays with matter and especially the formation of scatter. Furthermore, the basic principles of acquiring sectional images from projection images and the occurence of artifacts due to scatter are explained. The purpose of this chapter is to provide a brief introduction necessary for the understanding of this thesis. The interested reader is advised to consult standard references, for example from Buzug [1] or Hsieh [18].

2.1 Interaction of X-rays with matter

In medical X-ray imaging the energy of photons ranges from about 20keV up to about 140keV [18], depending on the application and the interaction properties of the imaged materials. In this energy range there are three fundamental ways in which X-rays can interact with matter: photoelectric absorption, Rayleigh scattering and Compton scattering. These interactions result in an attenuation of the incident beam during penetration of a medium. Although not relevant for the medical imaging energy range, but mentioned for the sake of completeness, pair production contributes to the attenuation of X-rays for photon energies above $2 \cdot 511$keV. Lambert-Beer's law of attenuation considers the combined effect of these processes to predict the radiation intensity after X-rays have passed through layers of known thickness and materials. During some of these afore-mentioned processes electrons are freed from the shell of atoms, leaving these atoms ionized. The use of ionizing radiation is a clear disadvantage of X-ray CT compared to other imaging modalities (e.g. magnetic resonance imaging (MRI) or ultrasound imaging) as it has both deterministic (i.e. tissue reactions) and stochastic (i.e. increased cancer risk) effects.

2.1.1 Lambert-Beer's law of attenuation

When travelling through a slice of matter with a known thickness $\mathrm{d}x$, a number of photons $\mathrm{d}N$ are removed from the initial photon number N due to absorption or scattering. The number of interacting photons at a given position x, $N_{\mathrm{int}}(x)$, is proportional to the number of incident photons N and to the thickness $\mathrm{d}x$ [4].

$$\mathrm{d}N(x) = -N_{\mathrm{int}}(x) = -\mu(x)N(x)\mathrm{d}x \tag{2.1}$$

μ is called the linear attenuation coefficient and describes the propability of a photon beeing removed from the incident beam. Its unit is m^{-1}.

Integration of Equation 2.1 yields the famous law of Lambert-Beer (Equation (2.2)), which can be used to determine the number of photons after a beam has travelled a certain length x [4].

$$N(x) = N_0 \cdot e^{-\int_0^x \mu(x')\,\mathrm{d}x'} \tag{2.2}$$

As mentioned before, the attenuation coefficient is composed of the sum of the different interaction probabilities for photoelectric absorption, Rayleigh scattering and Compton scattering. μ is a material characteristic and thus a function of position x and also the energy of the incident X-rays [4].

$$\mu_{\mathrm{total}} = \tau_{\mathrm{Photo}} + \sigma_{\mathrm{Rayleigh}} + \sigma_{\mathrm{Compton}} \tag{2.3}$$

Figure 2.1 displays the probabilities of interaction per unit length for the three different interaction processes for water as well as the total attenuation coefficient as a function of X-ray energy. It is apparent that in the diagnostic energy range above 50keV Compton scattering is the most prevalent form of X-ray-matter interaction.

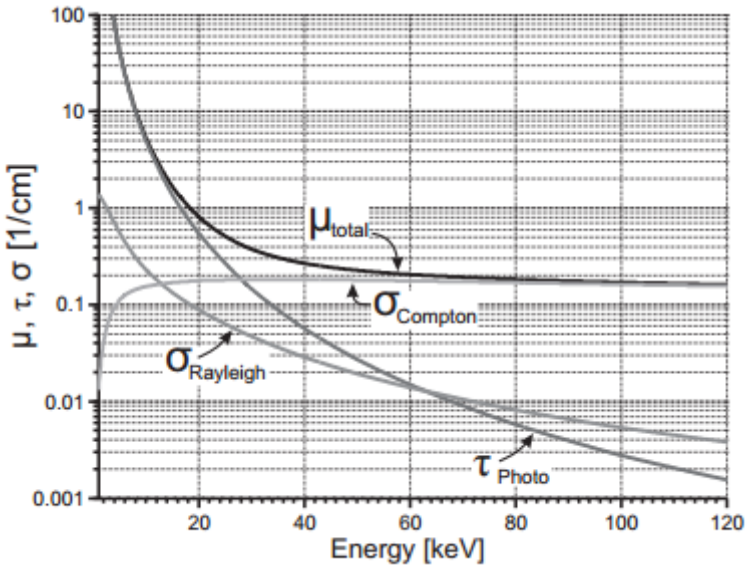


Figure 2.1: Interaction propabilities per unit length for water for photoelectric absorption (τ_{Photo}), Rayleigh scattering ($\sigma_{Rayleigh}$) and Compton scattering ($\sigma_{Compton}$). μ_{total} depicts the total linear attenuation coefficient for water. With increasing energy, the total linear attenuation coefficient declines; in other words, low energy photons are attenuated more than high energy ones. Figure taken from [4].

2.1.2 Photoelectric absorption

When the energy of an X-ray photon exceeds the binding energy of an electron in a certain absorber atom shell, E_B, the entire energy of this photon can be absorbed by an atom. In this process an electron of a lower shell is kicked off the atom and can then travel as a free electron through the material [1]. This so-called photoelectron, with an energy of $T = E - E_B$, is ejected from the shell leaving behind a ionized atom. The free spot in the atomic shell is filled with an electron either from an outer shell or with free electrons in the absorber medium. An outer shell has a higher binding energy and thus characteristic X-ray photons may also be generated which can in turn again ionize absorber atoms [4]. The photoelectric effect, discovered by Albert Einstein in 1905, produces a positive ion, a photoelectron and characteristic radiation [18]. The incident X-ray photon transfers all of its energy to the freed electron and is therefore destroyed. This process is illustrated in Figure 2.2.

The linear absorption coefficient τ_{Photo} depends on the energy E of the incoming photon and the absorbers atomic number Z [1].

$$\tau_{\text{Photo}} = k\frac{\rho}{A}\frac{Z^4}{E^3} \tag{2.4}$$

where k is a constant and depends on the atomic shells involved in this process, ρ is the absorber material density and A is the atomic weight. Absorption shows strong Z^4 dependence [1]. This dependence is the reason for the usage of high Z materials such as lead for X-ray shielding or ASGs [4].

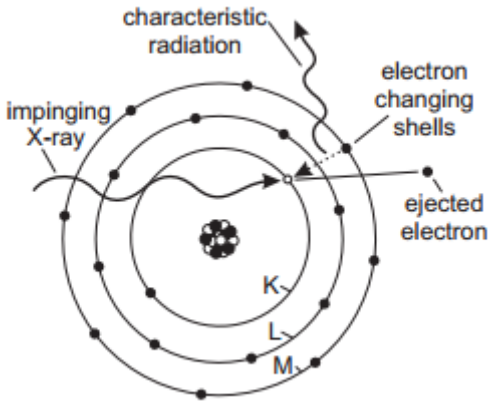


Figure 2.2: The process of photoelectric absorption illustrated. An electron is liberated from an inner shell and this vacancy is quickly filled with an electron from a higher shell with higher binding energy. The difference energy between those shells is emitted as characteristic radiation. The incident photon is completely destroyed. Figure taken from [4].

2.1.3 Rayleigh scattering

Rayleigh scattering is also known as elastic or coherent scattering. It occurs with particles with significantly smaller diameters than the incident radiations wavelength. The incident and the

scattered waves have equal wavelength, but the direction of the scattered wave differs from the direction of the incident wave [1]. The incident radiation sets the electrons of a particle in oscillation, causing them to emit radiation of the same wavelength as the incident wave. The two waves combine and form the scattered wave. The incident X-ray photon keeps all of its energy and no ionization occurs [4]. This process is illustrated in Figure 2.3.

The differential Rayleigh scatter cross sections for water at room temperature are shown in Figure 2.4. One can observe that Rayleigh scattering occurs in direction of the deflected incoming radiation producing a slightly broadened X-ray beam [18]. With increasing photon energy the probability for forward scatter also increases.

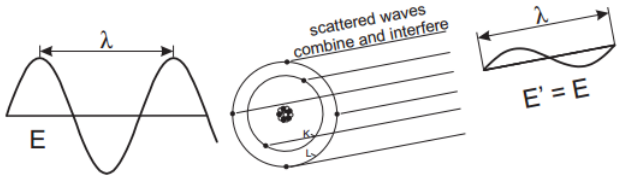


Figure 2.3: The process of Rayleigh scattering illustrated. The resulting scattered wave oscillates at the same frequency as the incident wave, no transition of energy occurs. Figure taken from [4].

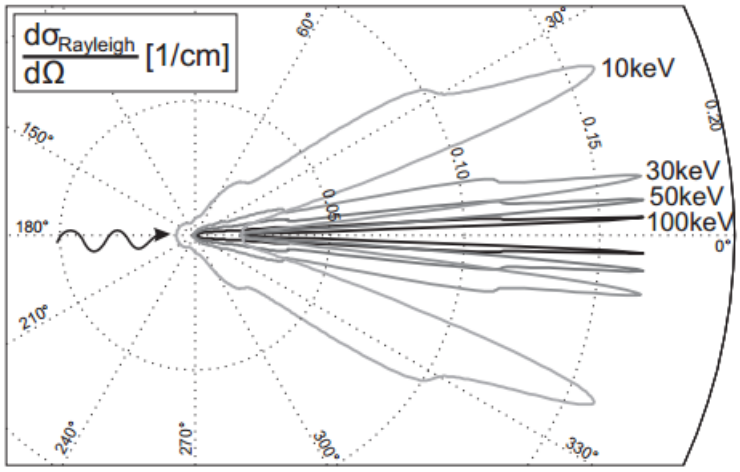


Figure 2.4: Differential Rayleigh scatter cross sections for water at room temperatur as a function of scattering angle θ and various photon energies. Figure taken from [4].

2.1.4 Compton scattering

Compton scattering can be seen as a mixture of scattering and photoelectric absorption [1]. It is the most important interaction mechanism in tissue-like materials for diagnostically relevant radiation [18]. Compton scattering is an inelastic process in which the wavelength and energy of the incident photon are altered. The incoming photon, which has considerable higher energy than the binding energy of an electron it interacts with, transfers a portion of its energy to the electron, liberating it from its shell. These electrons are often called secondary electrons or recoil

electrons. Thus, a Compton interaction produces a positive ion, a recoil electron and a scattered photon with less energy than the incoming one (see Figure 2.5). The recoil electron and the scattered photon may again ionize atoms if their energy is sufficient high. In contrast to Rayleigh scattering, where only forward scatter occurs, Compton scattered photons may be deflected in an arbitrary direction. Low-energy photons are primarily backscattered, while high-energy photons have a higher propability for forward scatter. Because of this wide deflection angle, the scattered photon provides little to no spatial information about the location of interaction [18]. Figure 2.6 shows the differential cross sections for Compton scatter as a function of scatter angle and various photon energies.

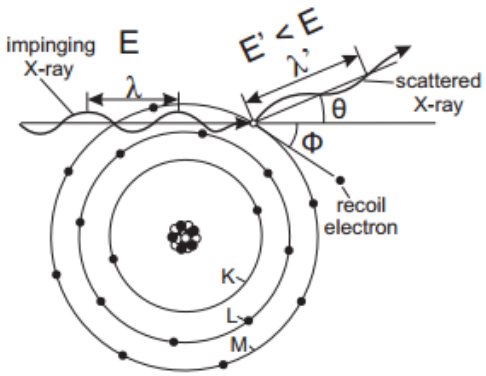


Figure 2.5: The process of Compton scattering illustrated. An incident photon ejects an electron from it's shell. The resulting scattered wave oscillates at a lower frequency as the incident wave, indicating that a transition of energy has occured. Figure taken from [4].

2.1.5 Scatter analysis of the human head

In [4] Wiegert performed a detailed scatter analysis of the human head conducting Monte-Carlo simulations. These results may be used to develop tailored scatter correction algorithms for neurological and dental applications. A voxelized CT data set from a helical CT scan has been used to compute simulated projection images of scatter and primary radiation of the human head. A flat-panel C-arm system with a source-detector distance (SDD) of 117 cm and a source-object distance (SOD) of 76.5 cm was assumed. The simulated tube voltage was set to 100kV, after prefiltration to remove low-energy photons, the mean photon energy has been found to be 68keV. No ASG was used in this study. From now on the primary signal will be associated with the symbol I_P while the scatter signal will be depicted by I_S. Together, these signals form the total measured intensity:

$$I_M = I_P + I_S \tag{2.5}$$

The results from this study can be seen from three different viewing directions in Figure 2.7. The images (d) to (f) depict the resulting normalized primary radiation I_P and the images (g)

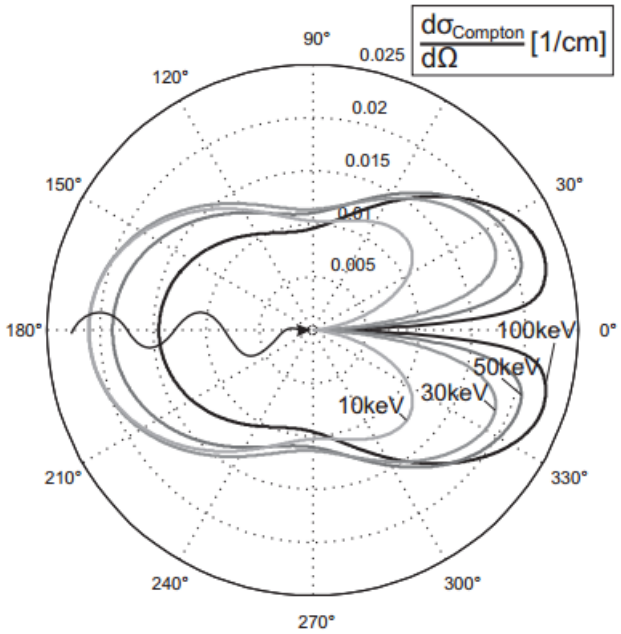


Figure 2.6: Differential Compton scatter cross sections for water at room temperature as a function of scatter angle θ and various photon energies. Figure taken from [4].

to (i) show the detected normalized scatter I_S. As can be seen from the scatter images, the scatter distribution is somewhat smaller in the middle of the head phantom. Although scatter arises in this region, the increased penetration path length leads to increased attenuation in these regions. Therefore, the highest amount of scatter can be found near the boundaries of the head. These boundaries are indicated by the red lines in all images. In subfigure (i) it can be seen that in the region behind the air cavities of the nose the amount of detected scatter is slightly larger than in the surrounding areas due to a lack of absorption. A general observation valid for all three different views is that the detected spatial distribution of the scatter is flat and smooth with very limited information about the imaged object.

Figure 2.8 shows an important quantity to assess the problem of scatter in X-ray projection images. This is the scatter-to-primary ratio (SPR), defined as the quotient of detected scatter to detected primary in each pixel:

$$SPR_i = \frac{I_{S,i}}{I_{P,i}} \tag{2.6}$$

In regions with high SPR the interesting primary signal lies in the range of the scatter signal or can even be exceeded by the scatter signal. This usually results in severe artifacts due to violation of certain assumptions made in the reconstruction. In contrast to the scatter distribution, the SPR shows strong spatial variation because of its dependency on the reciprocal primary radiation. Behind strongly absorbing structures (e.g. bones or in the middle of the head), where only little primary radiation but a comparatively high amount of scatter reaches the detector, the SPR can be as high as 1 or even above this value. In more lateral regions the SPR drops to less than 50%. It can be noted that the effect a certain amount of scatter will have on

reconstructed volumes strongly depends on the amount of primary detected within an area. A certain scatter level may not cause any harm in regions of low SPR, while the same amount of scatter may lead to strong artifacts in regions of high SPR. This means that in areas of high SPR accurate scatter correction is more important than in areas with lower SPR. It may also be noted, and this is especially interesting in the context of dental DVT, that the SPR reaches its highest values in the area of the mouth, more precisely behind teeth and the chin.

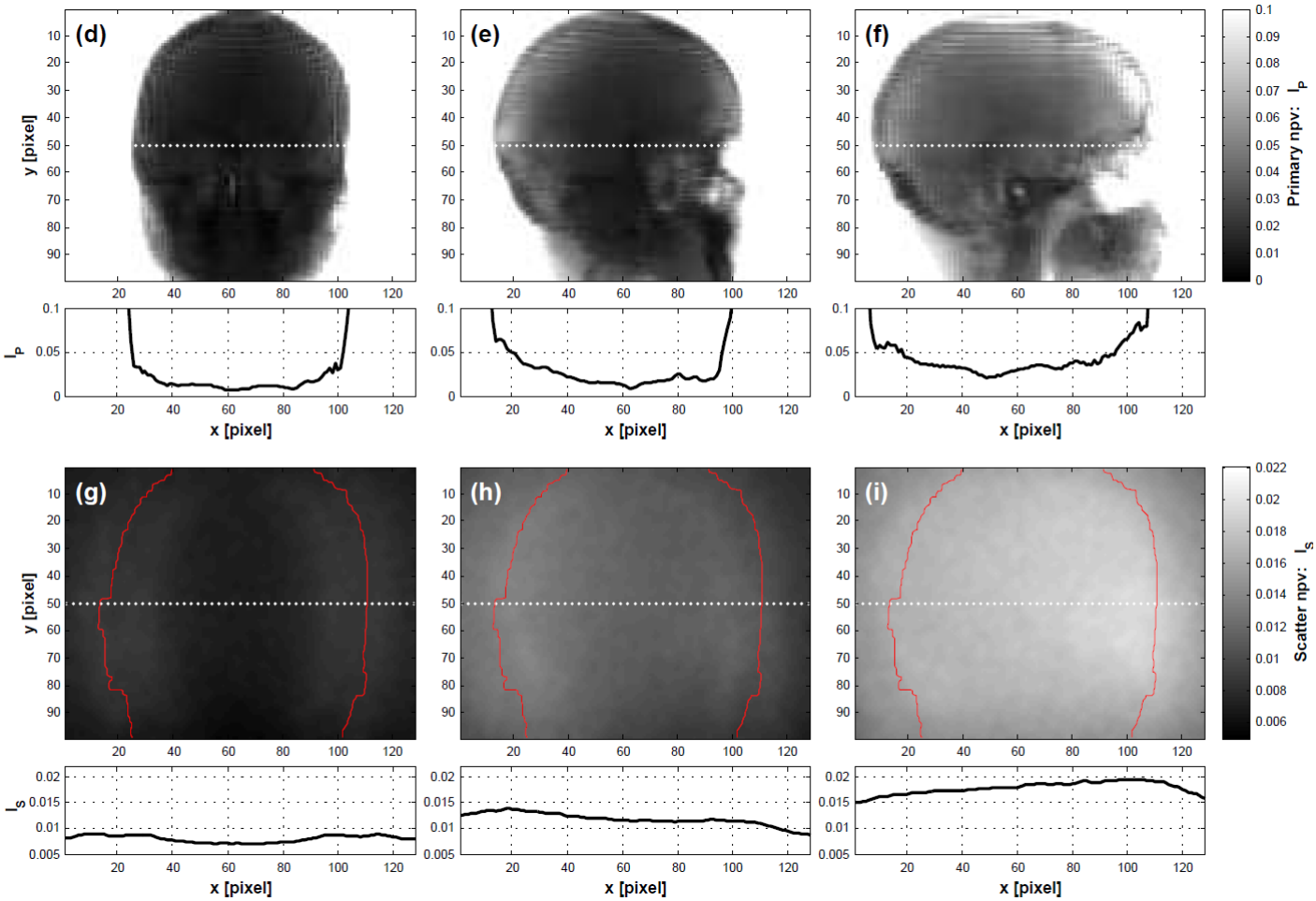


Figure 2.7: Results of a high-photon MC simulation study of a human head phantom. Images (d) to (f) depict the resulting normalized primary radiation I_P and the images (g) to (i) show the detected normalized scatter I_S. Figure taken from [4].

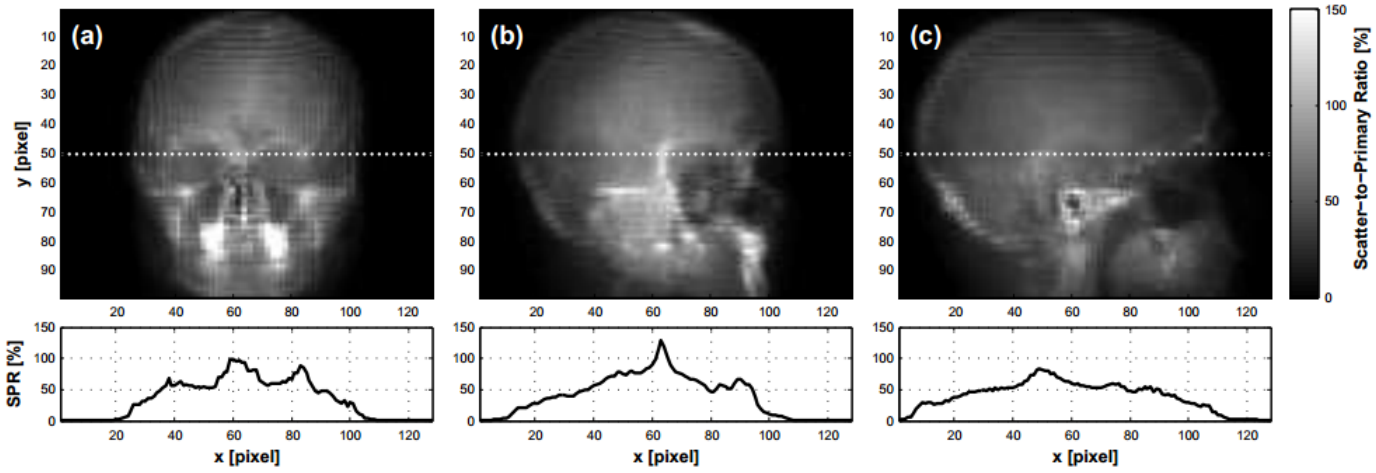


Figure 2.8: Scatter-to-primar ratio (SPR) for three exemplary views of a human head phantom. In contrast to the smooth scatter distribution, the SPR shows strong spatial variation. Figure taken from [4].

2.2 Fundamentals of Computed Tomography

2.2.1 Measuring of line integrals

Equation (2.2) rewritten in X-ray intensities instead of photon numbers reads

$$I = I_0 \cdot \mathrm{e}^{-\int_{-\infty}^{\infty} \mu(s)\,\mathrm{d}s}. \tag{2.7}$$

The X-ray intensity after passing through an object on a straight line decreases exponentially, whereby the exponent is the negative sum of all attenuation coefficients μ along this line. In Equation (2.7) I denotes the X-ray intensity in a single detector pixel, I_0 the intensity prior to attenuation and s parametrizes the straight photon trajectory. This describes an ideal situation where only photons of the same energy occur. X-ray tubes in practical CT machines always emit a polychromatic spectrum $I_0(E)$ and one has to integrate over all occuring energies to account for the energy-dependence of the linear attenuation coefficient

$$I = \int_0^{E_{max}} I_0(E) \cdot \mathrm{e}^{-\int_{-\infty}^{\infty} \mu(E,s)\,\mathrm{d}s}\,\mathrm{d}E. \tag{2.8}$$

In practice only Equation (2.7) is used and this leads to one of the biggest sources of degraded image quality, the beam hardening effect. In X-ray CT one is interested in the spatial distribution of the attenuation coefficient. Normalizing both sides of Equation (2.7) with I_0 and computing the logarithm yields the line integral of the linear attenuation coefficient along a straight X-ray trajectory through the object or patient being imaged.

$$p = -\log\left(\frac{I}{I_0}\right) = \int_{-\infty}^{\infty} \mu(s)\,\mathrm{d}s \tag{2.9}$$

p is then often called a projection measurement. Between $-\log\left(\frac{I}{I_0}\right)$ and $\int_{-\infty}^{\infty} \mu(s)\,\mathrm{d}s$ now exists a linear relationship and this is the basis for image reconstruction. Scatter represents an additional signal that adds to the primary signal and this leads to non-linearity in the afore-mentioned relationship. Equation (2.9) no longer holds and artifacts arise.

2.2.2 Cone-beam computed tomography scanner setup

In CBCT and DVT an X-ray tube and an X-ray detector rotate isocentrically around an object of interest to acquire a number of projection images in small angular increments. The goal of image reconstruction techniques is to compute stacks of sectional images, often called a volume, from these projection images. A typical CBCT setup can be seen in Figure 2.9. Figure 2.10 shows a state-of-the art DVT device with a flat-panel detector.

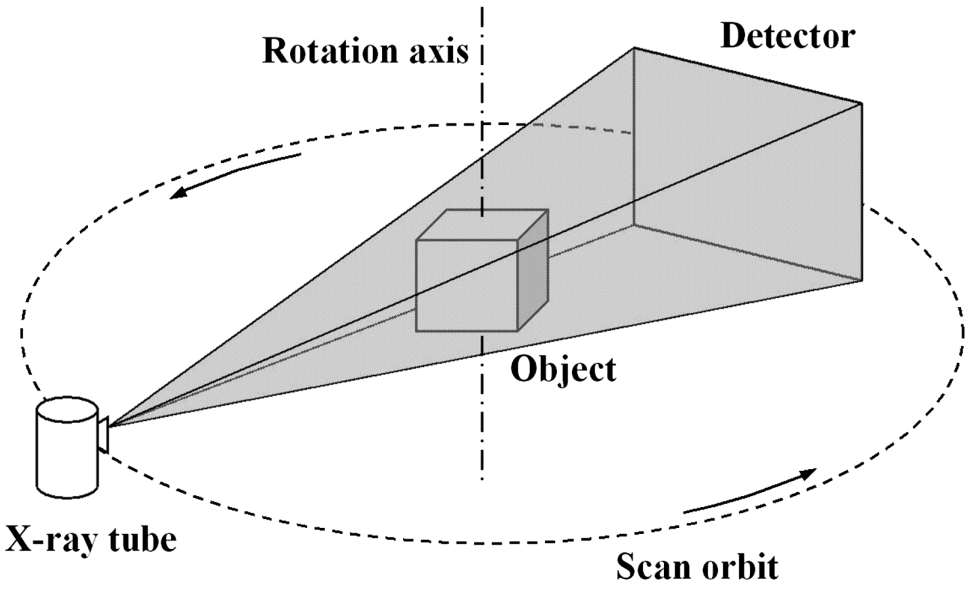


Figure 2.9: A typical CBCT setup. Both the detector and the X-ray source rotate around the object. Figure taken from `http://opticalengineering.spiedigitallibrary.org/data/Journals/OPTICE/22109/037004_1_1.png`.

2.2.3 Image reconstruction

CT aims at reconstructing the 2D distribution of the linear attenuation coefficient within an object. This is then often called a slice[1]. A vast variety of different image reconstruction techniques exists. These techniques can be divided into analytical and iterative reconstruction algorithms. This section covers the very basics of analytical image reconstruction. Starting with the most common method in commercial CT scanners, the filtered backprojection (FBP), additional extensions will be introduced in order to expand the classical FBP method from parallel-beam geometry to cone-beam geometry. To this day, the majority of commercial CBCT

[1] See the word tomography: from Greek *tomos* (slice) and *graphein* (write).

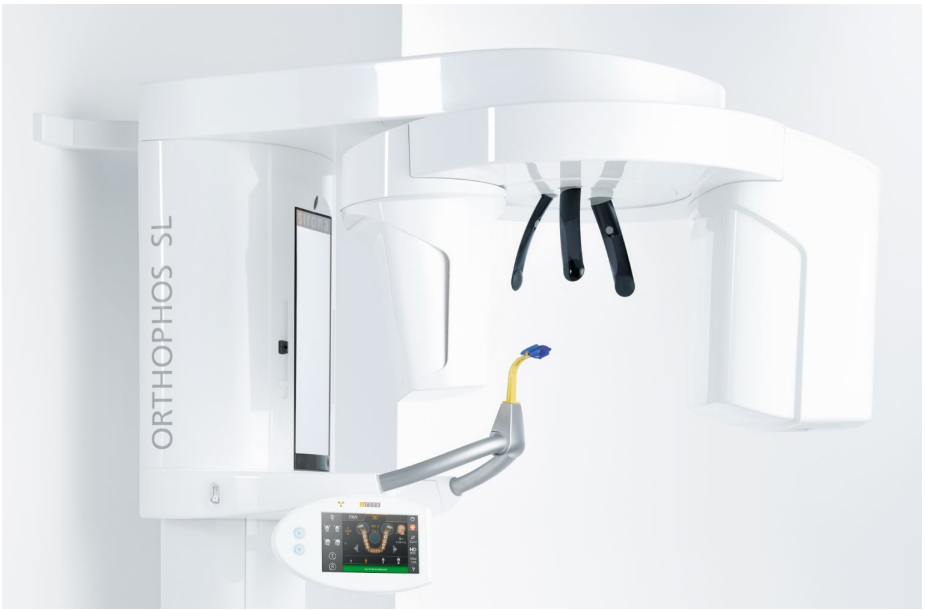


Figure 2.10: A state-of-the-art DVT device. The patient is positioned between the rotating source-detector assembly and bites a block to prevent motion during data acquisition. Figure with kind permission by Sirona Dental Systems GmbH, Bensheim, Germany.

and DVT devices employ variants of the FDK algorithm published by Feldkamp, David and Kress in 1984 [19]. The following derivations are based on the presentation in [20].

2D fan-beam reconstruction

For the following derivation parallel-beam geometry with a single row detector is assumed. A single measurement from a specific angle θ, called a projection, contains a number of transmitted normalized intensities at different detector pixel positions t. If sorted in a single matrix, these measurements from all angles are called a *sinogram*.

Consider a single line out of a set of lines, as depicted in Figure 2.11. At a certain projection angle θ and at detector position t, Equation (2.9) may be rewritten as

$$p^P(\theta, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y)\delta(y\cos\theta - x\sin\theta - t)\,\mathrm{d}x\,\mathrm{d}y \tag{2.10}$$

where the Dirac delta function is used to parametrize this line and the superscript P indicates that this projection is parallel. This function $p^P(\theta, t)$ is known as the *Radon transform*[2] of the function $f(x, y) = \mu(x, y)$. The *Fourier slice theorem* states that the values of the 1D Fourier transform of a parallel projection of an object $f(x, y)$ taken at an angle θ, $\mathcal{F}_1 p^P(\theta, t)$, correspond to a radial line in the 2D Fourier transform $\mathcal{F}_2 f(u, v)$ of said object at the corresponding angle θ (see Figure 2.11).

$$\mathcal{F}_1 p^P(\theta, \rho) = \mathcal{F}_2 f(-\rho\sin\theta, \rho\cos\theta) \tag{2.11}$$

[2] The Radon transform is an integral transform and was introduced by the Austrian mathematician Johann Radon in 1917. Original German paper title: *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten.*

As can be seen in Figure 2.11, it is sufficient to acquire projections in an angular interval of

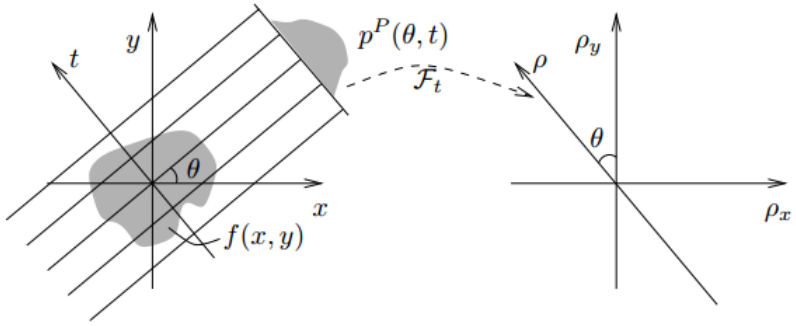


Figure 2.11: The Fourier slice theorem. Left: a parallel projection $p^P(\theta,t)$ of the object $f(x,y)=\mu(x,y)$. Right: The 1D Fourier transform of the projection corresponds to a radial line of values in the 2D Fourier space of the object. Figure taken from [20].

length π to cover the whole 2D Fourier space of the object. With the Fourier slice theorem and with the usage of polar coordinates instead of Cartesian coordinates, the object function of interest, $\mu(x,y)$, can be obtained by employing the 2D inverse Fourier transform

$$\mu(x,y)=\frac{1}{2}\int_0^{2\pi}\int_{-\infty}^{\infty}\mathcal{F}_1 p^P(\theta,\rho)\mathrm{e}^{j2\pi\rho(y\cos\theta-x\sin\theta)}|\rho|\,\mathrm{d}\rho\,\mathrm{d}\theta. \tag{2.12}$$

Equation (2.12) can be further simplified:

$$\begin{aligned}
\mu(x,y)&=\frac{1}{2}\int_0^{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p^P(\theta,t)\mathrm{e}^{-j2\pi\rho t}\,\mathrm{d}t\,\mathrm{e}^{j2\pi\rho(y\cos\theta-x\sin\theta)}|\rho|\,\mathrm{d}\rho\,\mathrm{d}\theta && (2.13)\\
&=\frac{1}{2}\int_0^{2\pi}\int_{-\infty}^{\infty}p^P(\theta,t)\int_{-\infty}^{\infty}|\rho|\,\mathrm{e}^{j2\pi\rho(y\cos\theta-x\sin\theta-t)}\,\mathrm{d}\rho\,\mathrm{d}t\,\mathrm{d}\theta && (2.14)\\
&=\int_0^{2\pi}\int_{-\infty}^{\infty}p^P(\theta,t)g_\infty^P(y\cos\theta-x\sin\theta-t)\,\mathrm{d}t\,\mathrm{d}\theta && (2.15)\\
&=\underbrace{\int_0^{2\pi}\underbrace{(p^P*g_\infty^P)}_{\text{Filtering}}(\theta,y\cos\theta-x\sin\theta)\,\mathrm{d}\theta}_{\text{Backprojection}} && (2.16)
\end{aligned}$$

g_∞^P is often called a ramp filter, referring to its shape in the Fourier domain

$$g_\infty^P(t)=\frac{1}{2}\int_{-\infty}^{\infty}|\rho|e^{j2\pi\rho t}\,\mathrm{d}\rho. \tag{2.17}$$

This filtering step prevents blurring in the resulting 2D distribution $\mu(x,y)$ since it accentuates higher frequencies in the Fourier domain and damps lower frequencies. To conclude, the projection data are first ramp-filtered and then back-projected onto the image plane. Backprojection can be understood as the adjoint operation to projection, i.e. a projection value is smeared out over the image points along a ray. While continuous derivations are useful for understanding the matter, in a practical system one needs to discretize the above equations.

Modern CT devices do not emit parallel X-rays, but divergent rays that emerge in a single point, namely the X-ray source. Figure 2.12 displays the projection setup for fan-beam geometry. The X-ray source rotates around the object in a circular trajectory with radius R. An emitted ray is characterized by the projection angle β and the fan angle γ, which is spanned by the respective ray and the central ray of a projection. The original FBP algorithm has to be adapted to meet this changed geometry setup. With flat detectors X-rays are sampled equidistantly along the a-axis. This is referred to as a virtual detector and it is aligned perpendicular to the central ray in the center of rotation. The true detector coordinates further away from the center of rotation are obtained by simply scaling a. With this geometry setup a projection ray is described by a duplet (β, a).

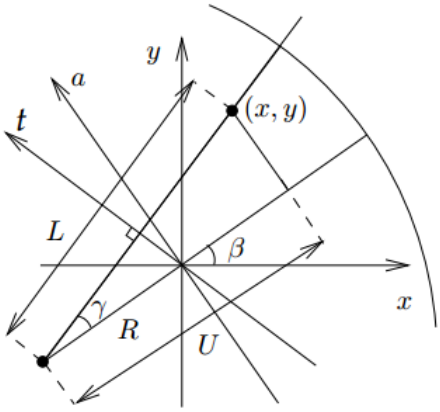


Figure 2.12: Fan-beam geometry. Figure taken from [20]

The filtered projection $\tilde{p}^F(\beta, a)$ with ^{F} referring flat detector geometry is computed by using the same filter kernel as for the parallel-beam case and a pre-weighting factor $\cos\gamma$

$$\tilde{p}^F(\beta, a) = (p(\beta, a)\cos\gamma) * g_\infty^P(\gamma) \tag{2.18}$$

Backprojection is achieved by

$$\mu(x, y) = \int_0^{2\pi} \frac{R^2}{U(x, y, \beta)^2} \tilde{p}^F(\beta, a(x, y, \beta))\,\mathrm{d}\beta \tag{2.19}$$

whereby $U(x, y, \beta)$ is the distance from the X-ray source to a point (x, y) projected onto the central ray.

3D cone-beam reconstruction

In traditional fan-beam CT adjacent slices are reconstructed consecutively, whereas with CBCT with a 2D detector a complete volume can be reconstructed directly from the 2D projection data. This situation is illustrated in Figure 2.13. It is important to understand that reconstruction with the FDK algorithm is only exact in the central plane of the cone-beam [19]. According to the Tuy-Smith-condition, exact reconstruction of an object point requires that all planes

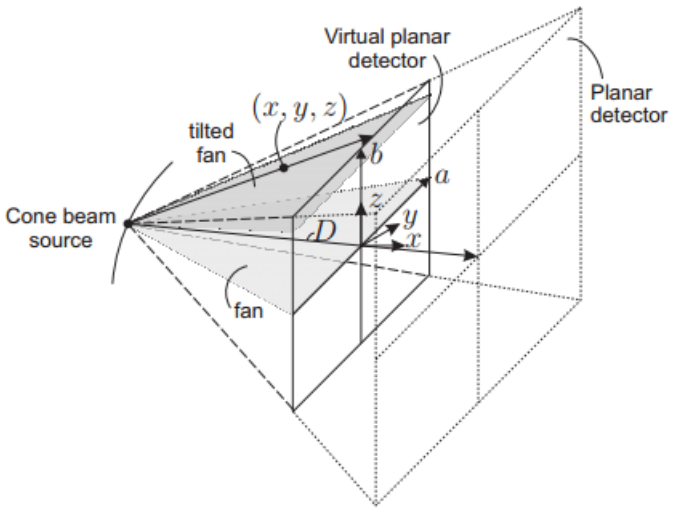


Figure 2.13: Cone-beam projection. Note the variable change for the radius of the circular trajectory from R to D. Figure taken from [4].

intersecting the object at the position of this point also intersect the source trajectory [21]. This condition is clearly violated for circular trajectories and therefore reconstruction is only exact for the central plane. As can be seen in Figure 2.13, the FDK method makes use of a virtual detector positioned parallel to the original detector in the coordinate origin. A single ray can now be parametrized by the projection angle β and the virtual detector coordinates a and b. The FDK method is also a filtered backprojection method. The pre-weighting factor is

$$\cos\gamma\cos\kappa = \frac{D}{\sqrt{D^2 + a^2 + b^2}} \tag{2.20}$$

Filtered projections are then computed according to

$$\tilde{p}^F(\beta, a, b) = \left(\frac{D}{\sqrt{D^2 + a^2 + b^2}} p^F(\beta, a, b)\right) * g^P(a) \tag{2.21}$$

The filtering operation is therefore still one-dimensional. A point (x,y,z) is then reconstructed by

$$\mu(x, y, z) = \int_0^{2\pi} \frac{D^2}{U(x, y, \beta)^2} \tilde{p}^F(\beta, a(x, y, \beta), b(x, y, z, \beta))\,\mathrm{d}\beta \tag{2.22}$$

Equipped with the above derivations, one is now capable of reconstructing a whole volume from 2D CBCT projection data. Note that in practice corrections for non-circular trajectories (e.g. elliptical) as well as for scan protocols that do not cover an angular range of 2π have to be applied. Figure 2.14 summarizes the process of reconstruction with the FDK method.

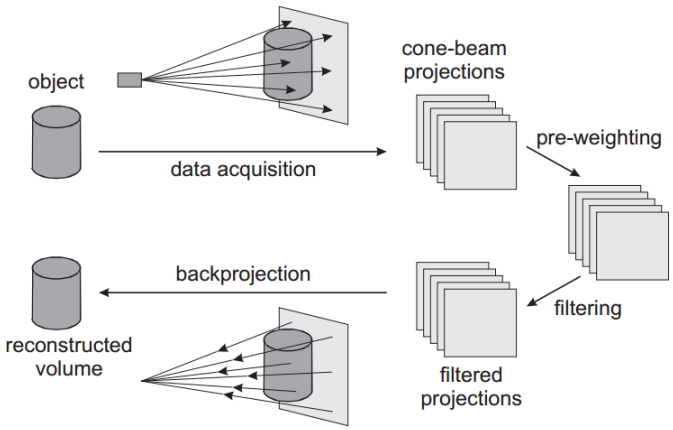


Figure 2.14: Cone-beam filtered backprojection with the FDK algorithm. Figure taken from [4].

Pre- and post-reconstruction processing

The measured data have to be pre-processed to represent the intensity values $\frac{I}{I_0}$ needed as input for the reconstruction process. Before computing the negative logarithm in Equation (2.9), sensor defects must be repaired. Furthermore, the data must be detruncated, i.e. extrapolated at the projection image borders. After applying the filtered backprojection, slices are often corrected for noise with e.g. a bilateral filter and, for viewing purposes, values are clipped to a certain range (windowing).

Most scatter correction algorithms, also the ones implemented and developed in this thesis, work on the pre-corrected images before computation of the negative logarithm.

2.3 Artifacts due to scattered radiation

Not all of the x-ray photons that reach the detector are primary photons. A significant portion of the signal is generated from scattered radiation. Scattered radiation adds a low-frequency bias to the true attenuation measurements. The minus logarithm operation that is necessary to convert x-ray photon flux into the line integrals transforms the linear summation into a non-linear operation. This is due to the fact that $\log(x+y) \neq \log(x) + \log(y)$ [18]. The larger the detector, the higher the probability that scattered photons incite it. Thus, the image-degrading effect of scattered radiation will affect CBCT machines more than classical highly collimated fan-beam CTs [22]. Scatter represents a major source of image degradation in CBCT systems. In projection images, scatter leads to a loss of contrast. In volume data artifacts such as cupping in homogeneous regions of material and streaks between regions of high contrast arise. Additionally, the signal-to-noise ratio (SNR) decreases when scattered radiation is detected.

2.3.1 Contrast loss

In projection images, the additional scatter signal I_S leads to a detoriation of contrast (see figure 2.15). Assume two objects of homogeneous material, one with an extra high contrast insert. Let the values of the line integrals of only primary radiation be $-\log(I_P)$ and $-\log(I_{PS})$ with scatter added to the primary, and $-\log(I_{P,C})$ and $-\log(I_{PS,C})$ respectively for the case with a high contrast insert inside the otherwise homogeneous material. In the presence of scatter contrast in line integral values $LI_{C'}$ is reduced compared to the scatter free case (LI_C) [4].

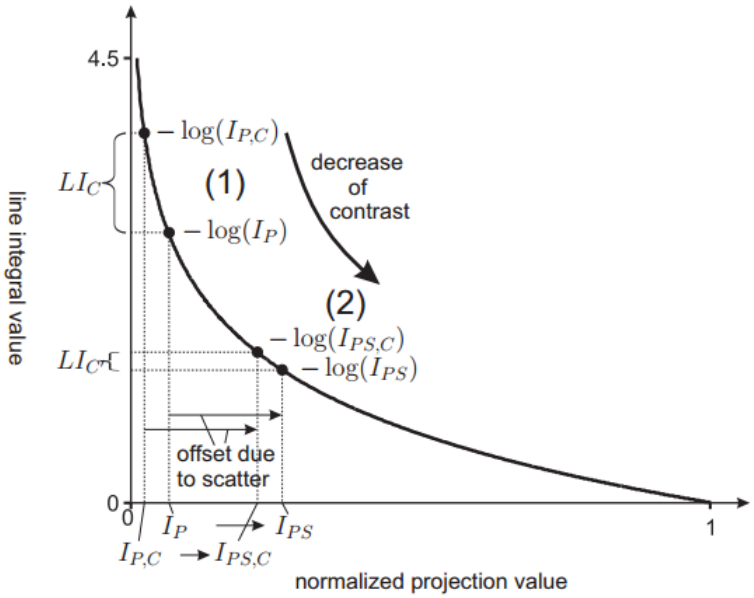


Figure 2.15: Curves for line integrals over normalized projection values. Compare the reference point (1) without scatter and point(2) with scatter added to the primary. Figure taken from [4].

2.3.2 Cupping and streak artifacts

In homogeneous objects the detection of scatter leads to measured signals that are higher than they should be, and this corresponds to an underestimation of attenuation coefficients, which leads to the well-known cupping artifact (see figure 2.16).

Figure 2.17 displays two exemplary projections through a homogeneous object with two high density rods. In view 1 the primary intensity at the detector I_P is low because of the incoming X-ray flux penetrating both high density rods, while the detected scatter is relatively uniform across the field-of-view (FOV). This leads to a SPR that is greater than 1. In view 2 the measured I_P through the rods is much higher, whereas the detected scatter intensity remains about the same as in view 1. The scatter adds a larger fraction to the detected total signal in view 1 and therefore the occurrence of nonisotropic artifacts will be present in reconstructions. Less attenuation is measured through the rods in view 1 than in view 2 due to the scatter contribution, and the reconstructed attenuation coefficients will show a dark streak connecting the rods [3] (see figure 2.18).

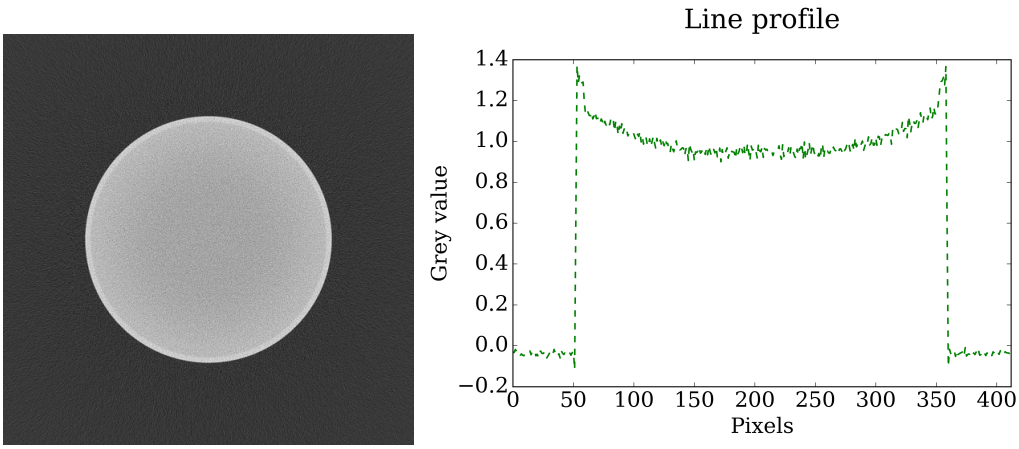


Figure 2.16: Left: Reconstruction of a simulated water cylinder with scatter from monochromatic projections. An underestimation of attenuation coefficients in the middle of the object occurs due to the presence of scatter. Right: a line profile through the middle of the water cylinder shows a cup shape.

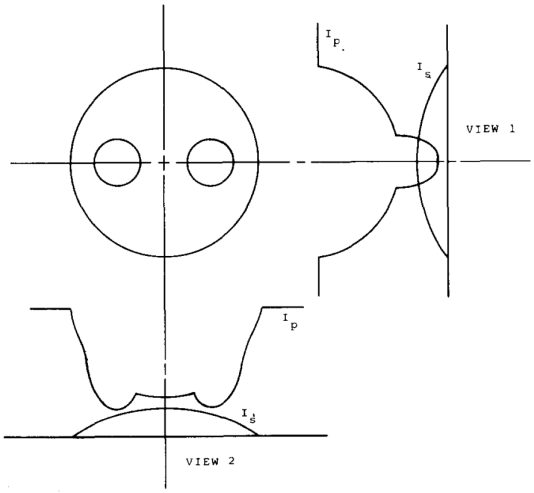


Figure 2.17: Illustrated primary and scatter signals for a cylinder with two strongly attenuating rods. After the penetration of both rods, the incident X-ray flux may be so weak such that it can be exceeded by the scatter signal, leading to a high scatter-to-primary ratio. The reconstructed attenuation coefficients will show a dark streak connecting the rods. Figure taken from [3].

2.3.3 Impact of scatter on noise

Two types of noise occur in X-ray projection images. The first noise contribution, which is rooted in the physics of the image formation process, is image dependent and reflects the statistical fluctuation of the X-ray photons. It can be considered as Poisson noise. The second is independent from the image and due to electrical and roundoff error and is often modelled as Gaussian noise. In many situations, the Gaussian noise is assumed to be negligible small and therefore only the Poisson part is considered [23].

The presence of scatter also influences the noise in the acquired projection images and reconstructed slices. Due to the fact that scatter contributes to the overall detected signal, but carries little to no spatial information of the object, it contributes to the noise but not to the

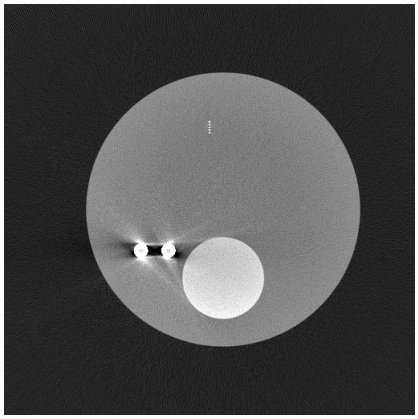


Figure 2.18: Reconstruction of a simulated water cylinder with two highly attenuating titanium inserts, a bone sphere and an aluminium grid from monochromatic projections. Between the titanium inserts, a dark streaks appears due to scatter. Also, the cupping artifact can be observed in both the water cylinder and the bone sphere.

wanted primary signal:

$$\mathrm{SNR}_i = \frac{I_{P,i}}{\sqrt{I_{P,i} + I_{S,i}}} = \sqrt{\frac{I_{P,i}}{1 + SPR_i}} \tag{2.23}$$

The index i here stands for a single detector pixel and SNR denotes the signal-to-noise ratio. In regions with high SPR, the SNR can be significantly lowered by the scatter signal [18]. The use of Equation 2.23 is valid only for projection images.

3 Methods and Materials

This chapter covers the theoretical derivations and implementation details of the methods and algorithms for scatter estimation and reduction implemented and developed during this thesis. First, the Improved Primary Modulator Scatter Estimation (iPMSE) technique will be explained. Afterwards, a framework for convolution-based scatter estimation will be introduced. The third section explains the novel Hybrid Primary Modulator Estimation Scatter Estimation (hiPMSE) algorithm, a combination of iPMSE and convolution-based scatter estimation. Due to the fact that all algorithms mentioned above are formulated as optimization problems, the fourth section briefly explains the minimization algorithms that were used in this thesis. The last section concludes with a short description of hardware and devices used for the measurements.

3.1 Improved primary modulator scatter estimation (iPMSE)

With Primary Modulator Scatter Estimation (PMSE) techniques the primary and scatter signals can be measured simultaneously without increased patient dose or scan time. To achieve this, a primary modulator, e.g. a checkerboard pattern with alternating attenuation properties, is inserted between the X-ray source and the object (see figure 3.1). The modulator pattern is superimposed on the incident primary radiation, shifting this desired signal to higher spatial frequencies. The key hypothesis states that the high-frequency components of the incident X-ray spatial distribution do not result in strong high-frequency signals in the scatter due to the broad point spread function (PSF) of the scatter process. This was previously valiadated with MC simulations [15]. Under the assumption that scatter is comprised of mainly low-frequency components separation of the interesting primary and scatter is possible using either a Fourier- or an image-based approach.

The Fourier-based approach has been proven to work well under ideal conditions (e.g. table-top CBCT system, perfectly regular modulator pattern, small blurring due to the nearly infinitesimal focal spot size), where high scatter correction accuracy can be achieved [15]. In devices that are used patient-side such as C-arm CT or DVT systems deviations from these ideal conditions occur. Both the X-ray source and the detector move around the object and this movement is subject to mechanical inaccuracies, the penumbra effect is present and often the modulators show manufacturing irregularities, resulting in a violation of the underlying model assumptions. To overcome these drawbacks, Ritschl et al. developed an image-based approach

that can deal with nearly any kind of static high-frequency modulation [16].

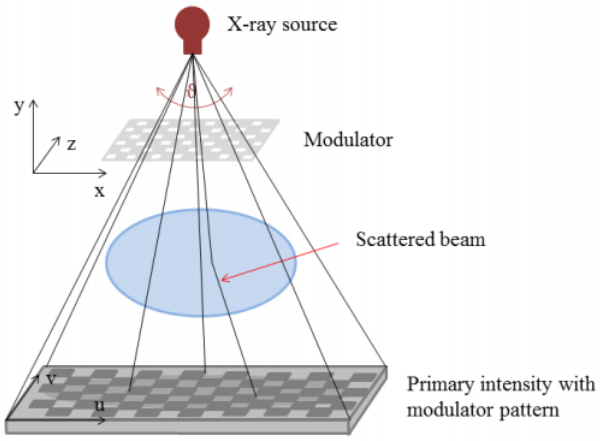


Figure 3.1: The primary modulator is mounted between the X-ray source and the object, and results in a spatially high-frequency pattern superimposed on the incident primary radiation. The scatter, which originates in the object, looses this high-frequency modulation due to the scatter process and can then afterwards be separated from the total detected signal. Figure taken from [24].

With the modulator in place the total measured signal $\boldsymbol{I}_M$ can be modeled as

$$\boldsymbol{I}_M = \boldsymbol{M} \cdot \boldsymbol{I}_P + \boldsymbol{I}_S \tag{3.1}$$

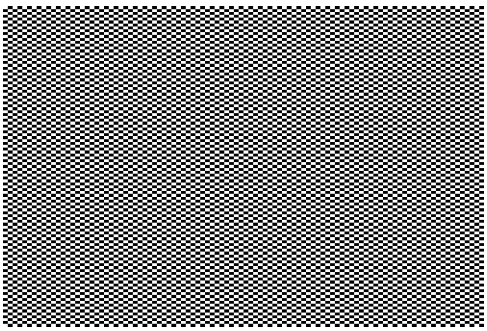
where $\boldsymbol{I}_M, \boldsymbol{I}_P$ and $\boldsymbol{I}_S$ are image matrices flattened into a vector and $\boldsymbol{M}_{i,i} = m_i$ is a diagonal matrix with the transmission values of the modulator. The matrix $\boldsymbol{M}$ can easily be obtained by acquiring a projection image of the modulator with no object in place and flattening this image into a vector $\boldsymbol{m}$. In Figure 3.2 simulated images for an artificial jaw phantom for the four afore-mentioned components are shown. These images serve as references throughout the next sections and can be considered whenever $\boldsymbol{I}_M, \boldsymbol{M}, \boldsymbol{I}_P$ or $\boldsymbol{I}_S$ are mentioned. Both $\boldsymbol{I}_M$ and $\boldsymbol{M}$ are assumed to be normalized by the unattenuated intensity $\boldsymbol{I}_0$ which in turn is assumed to be constant.

The interesting primary signal $\boldsymbol{I}_P$ can now be calculated by inverting $\boldsymbol{M}$ and subtracting the scatter signal $\boldsymbol{I}_S$ from the total signal $\boldsymbol{I}_M$:

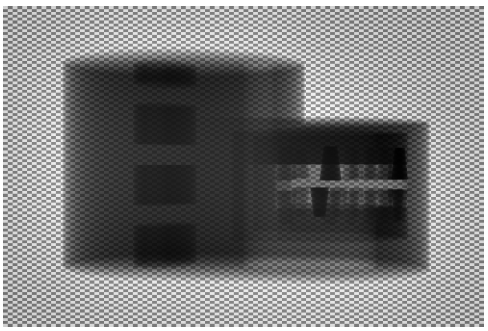
$$\boldsymbol{I}_P = \boldsymbol{M}^{-1} \cdot (\boldsymbol{I}_M - \boldsymbol{I}_S) \tag{3.2}$$

Assume that there exists an estimation for the scatter signal $\boldsymbol{I}_S^{\text{est}}$. The error of this scatter estimate can be denoted by $\Delta s = \boldsymbol{I}_S^{\text{est}} - \boldsymbol{I}_S$ and the estimated primary $\boldsymbol{I}_P^{\text{est}}$ amounts to

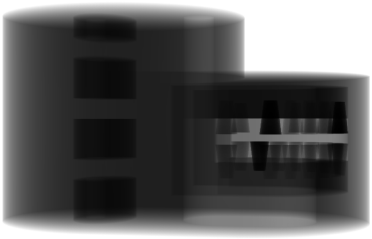
$$\begin{aligned} \boldsymbol{I}_P^{\text{est}} &= \boldsymbol{M}^{-1} \cdot \left(\boldsymbol{I}_M - \boldsymbol{I}_S^{\text{est}}\right) && (3.3)\\ &= \boldsymbol{M}^{-1} \cdot (\boldsymbol{I}_M - \Delta s - \boldsymbol{I}_S) && (3.4)\\ &= \boldsymbol{I}_P - \boldsymbol{M}^{-1} \cdot \Delta s && (3.5) \end{aligned}$$



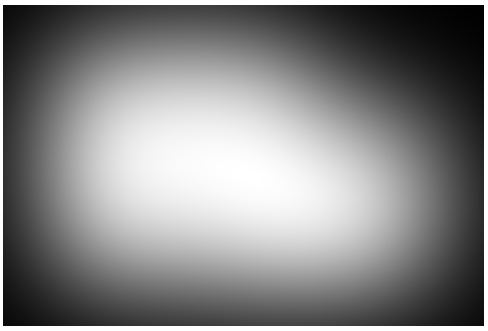
(a) $\boldsymbol{M}$



(b) $\boldsymbol{I}_M$



(c) $\boldsymbol{I}_P$



(d) $\boldsymbol{I}_S$

Figure 3.2: Individual components of the iPMSE and the hiPMSE method. Only $\boldsymbol{I}_M$ and $\boldsymbol{M}$ are directly measurable whereas $\boldsymbol{I}_P$ and $\boldsymbol{I}_S$ are initially unknown and are to be computed by the algorithm.

As long as there exists a non-zero error Δs the modulator pattern will remain visible in the estimated primary signal. The problem of finding the correct scatter estimate can be formulated as an optimization problem. Due to the fact that the modulator pattern contains strong edges the spatial gradient of the estimated primary signal $\nabla \boldsymbol{I}_P^{\text{est}}$ is an obvious choice for measuring the visibility of the modulator pattern in the primary estimate. Thus, it can be used as the cost function in the optimization problem. This cost function $C_{TV}(\boldsymbol{I}_S^{\text{est}})$ should reach a minimum when the optimal scatter values are found and no modulator pattern remains visible in the estimated primary.

$$C_{TV}(\boldsymbol{I}_S^{\text{est}}) = \|\nabla \boldsymbol{I}_P^{\text{est}}\|_1 = \|\nabla \boldsymbol{M}^{-1} \cdot \left(\boldsymbol{I}_M - \boldsymbol{I}_S^{\text{est}}\right)\|_1 \tag{3.6}$$

where the subscript TV stands for total variation. Previously, it has been shown that the use of a L1-Norm leads to convex behavior of the cost function [16][25]. This has also been validated in the present work. The resulting cost function does not show the typical kink one would expect when using a L1-Norm due to the fact that this L1-Norm is effectively a sum of gradient magnitudes. ∇ is an operator giving the magnitude of the gradient of $\boldsymbol{I}_P^{\text{est}}$ as $\sqrt{\left(\frac{\partial \boldsymbol{I}_P^{\text{est}}}{\partial u}\right)^2 + \left(\frac{\partial \boldsymbol{I}_P^{\text{est}}}{\partial v}\right)^2}$, where u and v are image coordinates.

The optimization problem is solved patch-wise by calculating one optimal scatter estimation value for each patch, assuming that the scatter is constant within a subimage. Similar down-sampling operations are commonly applied in scatter correction methods and justified by the

fact that the scatter signal is smooth and mainly comprised of low spatial frequencies [6]. The minimization problem is solved with Brent's algorithm to find the global minimum (see section 3.4). After the optimal scatter values for all subimages have been found the scatter signal is upsampled to the original resolution by bilinear interpolation and smoothed by a Gaussian filter with a kernel width in the range of the patch size. The patch size must be chosen large enough to contain at least on period of the modulator pattern. One patch has a size of $(2d_u+1)\times(2d_v+1)$ pixels and the patches overlap by d_u pixels in u-direction and by d_v pixels in v-direction. In this thesis $d_u = d_v \in [27, 43]$ with a modulator period of approximately 25 pixels has been used.

3.2 Convolution-based scatter estimation

Convolution methods assume that scatter can be estimated by a spatially invariant convolution of the primary signal with a scatter kernel. This concept was first introduced by Seibert and Boone in 1988 and has since served as the basis for many publications [26].

Throughout this thesis, a modified version of the model proposed by Ohnesorge et al. in [6] is used. The scatter estimation is obtained by convolving a forward scatter fraction $\Phi = \Phi(I_P)$ with a scatter kernel K:

$$I_S^{CB}(\vec{c}) = \Phi(I_P, \vec{c}) * *K(\vec{c}) \tag{3.7}$$

where $**$ denotes the 2D convolution that runs over the detector coordinates u and v and the superscript CB stands for convolution-based. The scatter kernel K is dependent on both scanner and object geometry and the scatter properties of the object. While convolution based methods offer the advantage of relatively low computational cost, which is a prerequisite for a clinically usable scatter estimation algorithms, a drawback of these models is that they contain a number of free parameters $\vec{c}$ that determine the shape of the scatter kernel and of the scatter potential. These parameters have to be calibrated in advance by means of dedicated calibration measurements or by fitting the open parameters to pencil-beam simulations [6][7][27].

The forward scatter fraction, or scatter potential, is given by

$$\Phi(I_P, c_0, c_1) = c_0 + c_1 \cdot I_P \log\left(\frac{I_P}{I_0}\right) \tag{3.8}$$

This scatter potential indicates how many photons are scattered on their way to the detector and do not contribute to the primary intensity [28]. The scatter intensity can be divided into an approximately constant fraction c_0, which is mainly associated with Compton scattering, and a fraction $c_1 \cdot I_P \log\left(\frac{I_P}{I_0}\right)$ that varies with the detector coordinates, and models Rayleigh scattering [29]. Note that in practice most commonly the intensity containing scattered and primary radiation is used instead of I_P, which is unknown. Figure 3.3 shows an exemplary scatter potential for $c_0 = 0$ and $c_1 = 1$. Regions with high attenuation and regions with no

attenuation do not contribute to the total scatter signal, such that $\Phi(0) = \Phi(I_0) = 0$.

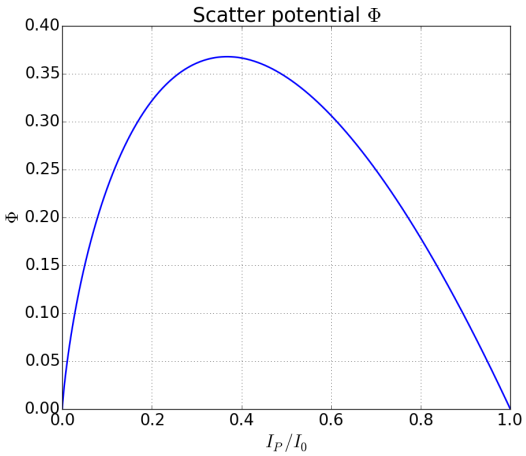


Figure 3.3: An exemplary scatter potential Φ with $c_0 = 0$ and $c_1 = 1$. The boundary conditions are $\Phi(0) = \Phi(I_0) = 0$ according to Equation 3.8.

A vast variety of different scatter kernels has been developed. These kernels have in common that they all exhibit low-pass characteristic since the resulting scatter estimation should be smooth and comprised of low spatial frequencies. The most straight-forward kernel is a simple 2D Gaussian kernel with only one open parameter σ:

$$K_1(r, \sigma) = \frac{1}{N_1} e^{-\frac{1}{2}(\frac{r}{\sigma})^2} \tag{3.9}$$

where $r = \sqrt{u^2 + v^2}$ and u and v are detector coordinates. Often the sum of two shifted Gaussians is used as a scatter kernel [10][28]:

$$K_2(r, \sigma, b) = \frac{1}{N_2} \sum_{\pm} e^{-\frac{1}{2}(\frac{r \pm b}{\sigma})^2} \tag{3.10}$$

with normalization constants

$$N_i = \frac{K_i}{\int K_i(u, v)\, \mathrm{d}u\, \mathrm{d}v} \tag{3.11}$$

The kernels presented above are rotationally symmetric. Once all free parameters in the kernel and the scatter potential have been determined they can be stored in the parameter vector $\vec{c}$. Note that the difficulty of the fitting procedure to obtain the unknown parameters increases with an increasing number of parameters. The final convolution-based scatter estimate I_S^{CB} is then obtained by a 2D convolution of the scatter potential Φ with the scatter kernel K (see Equation 3.7).

3.3 Hybrid primary modulator scatter estimation (hiPMSE)

The algorithm introduced here is a combination of the previously presented iPMSE method and a convolution-based scatter estimation technique. Thus, it is called Hybrid Primary Modulator Scatter Estimation (hiPMSE). hiPMSE aims at iteratively fitting the unknown parameter vector $\vec{c}$ of a convolution based scatter model such that the estimated scatter shows as small as possible deviation from the true scatter distribution. The error in the scatter estimation is, as with iPMSE, again assessed by the visibility of the modulator pattern in the estimated primary intensity. The gradient-based cost-function that evaluates said visibility of the modulator pattern is now a function of the open parameter vector $\vec{c}$, indicating that the hiPMSE method is again formulated as an optimization problem. The parameters are varied in a way that decreases the cost function until the modulator in the estimated primary signal vanishes. The cost function can be expressed as follows:

$$C_{TV}^{CB}(\boldsymbol{I}_S^{\text{est}}(\vec{c})) = \|\nabla \boldsymbol{I}_P^{\text{est}}\|_1 = \|\nabla \boldsymbol{M}^{-1} \cdot \left(\boldsymbol{I}_M - \boldsymbol{I}_S^{\text{est}}(\vec{c})\right)\|_1 \tag{3.12}$$

where the scatter estimate $\boldsymbol{I}_S^{\text{est}}$ is a function of the parameter vector $\vec{c}$.

The hiPMSE algorithm consists of the following steps:

1. Set iteration counter to $i = 0$
2. Calculate initial primary estimate without scatter consideration as $\boldsymbol{I}_P^{\text{est},0} = \boldsymbol{M}^{-1}\boldsymbol{I}_M$ and set it as best known estimate: $\boldsymbol{I}_P^{\text{est}*} = \boldsymbol{I}_P^{\text{est},0}$
3. Calculate value of cost function as $C_{TV}^{CB}(0) = \|\nabla \boldsymbol{I}_P^{\text{est}*}\|_1$ and initialize the minimum known cost function value: $C_{TV}^{CB,min} = C_{TV}^{CB}(0)$
4. Define initial parameter vector $\vec{c}_i$ such that the start parameters lie within a physically reasonable range
5. Repeat steps a) to e) until $|C_{TV}^{CB,i-1} - C_{TV}^{CB,i}| < \epsilon$
   a) Calculate scatter potential with best known primary estimate $\boldsymbol{I}_P^{\text{est}*}$ and current set of parameters $\vec{c}_i$: $\Phi(\boldsymbol{I}_P^{\text{est}*}, \vec{c}_i) = c_{i,0} + c_{i,1} \cdot \boldsymbol{I}_P^{\text{est}*} \cdot \log\left(\frac{\boldsymbol{I}_P^{\text{est}*}}{\boldsymbol{I}_0}\right)$
   b) Calculate scatter estimate $\boldsymbol{I}_S^{\text{est}}(\vec{c}_i) = \Phi(\boldsymbol{I}_P^{\text{est}*}, \vec{c}_i) * *K(\vec{c}_i)$
   c) Calculate primary estimate $\boldsymbol{I}_P^{\text{est},i} = \boldsymbol{M}^{-1} \cdot \left(\boldsymbol{I}_M - \boldsymbol{I}_S^{\text{est}}(\vec{c}_i)\right)$
   d) Evaluate cost function:

      IF $C_{TV}^{CB}(\boldsymbol{I}_S^{\text{est}}(\vec{c}_i)) < C_{TV}^{CB,min}$:

      $\boldsymbol{I}_P^{\text{est}*} \leftarrow \alpha \cdot \boldsymbol{I}_P^{\text{est},i} + (1-\alpha) \cdot \boldsymbol{I}_P^{\text{est}*}$, $C_{TV}^{CB,min} \leftarrow C_{TV}^{CB}(\boldsymbol{I}_S^{\text{est}}(\vec{c}_i))$, update $\vec{c}$

      IF $C_{TV}^{CB}(\boldsymbol{I}_S^{\text{est}}(\vec{c}_i)) \geq C_{TV}^{CB,min}$:

      update $\vec{c}$

   e) Increase iteration counter: $i \leftarrow i + 1$

The first primary estimate is calculated with $\boldsymbol{I}_S^{\text{est}} = 0$ and $\boldsymbol{I}_P^{\text{est},0}$ represents a simple demodulation of the measured data without scatter consideration and will therefore show a superimposed modulator pattern. Alternatively, an initial demodulation with a constant scatter value for one projection is conceivable in step 2. such that $\boldsymbol{I}_P^{\text{est},0} = \boldsymbol{M}^{-1} \cdot \left(\boldsymbol{I}_M - I_S^{const}\right)$. To this point $\boldsymbol{I}_P^{\text{est},0}$ is the best known primary estimate and is therefore saved as such. Afterwards the cost function is evaluated giving a measure of modulator visibility in the initial primary estimate (step 3.). After the initialization of the parameter vector $\vec{c}$ (step 4.) an optimization loop is carried out as long as the changes in the cost function are greater than a certain threshold ϵ. In this thesis $0.001 < \epsilon < 0.1$ has been used.

The goal is to find an optimal set of parameters $\vec{c}^*$ that is capable of approximating the true scatter distribution. The ideal scatter distribution $\boldsymbol{I}_S^{\text{est}*}$ is found when no modulator pattern remains visible in the estimated primary intensity. Finding the ideal scatter estimate $\boldsymbol{I}_S^{\text{est}*}$ is an iterative procedure in which both the parameter vector $\vec{c}$ and the primary estimate $\boldsymbol{I}_P^{\text{est}*}$ are altered. In each iteration a scatter potential Φ and a scatter estimate $\boldsymbol{I}_S^{\text{est}}(\vec{c}_i)$ must be computed in order to asses the influence of the current parameter vector $\vec{c}_i$ on the primary estimate $\boldsymbol{I}_P^{\text{est},i}$ (steps a)-c)). Note that it may be necessary to correct for invalid or negative values in the scatter potential in case a bad primary estimate is used.

Whenever a better set of parameters is found less modulator pattern is superimposed on the current primary estimate, the cost function decreases and both the best known primary estimate $\boldsymbol{I}_P^{\text{est}*}$ and the minimum known cost function value $C_{TV}^{CB,min}$ are updated. $\boldsymbol{I}_P^{\text{est}*}$ is updated as a weighted sum of the new best primary estimate that was found in the current iteration and the overall best known estimate to this point. This is necessary to prevent oscillation of the primary estimate between two states in the first few iterations due to wrong initial scatter estimation. $0 < \alpha < 1$ is the weighting parameter that controls the speed of the overall cost function decrease and the rate of convergence of the hiPMSE algorithm. In this study $\alpha \in [0.5; 0.7]$ has been found to work well. The parameter vector $\vec{c}$ is then updated such that the costfunction will potentially further decrease. The last best known parameter vector is used here as a start value in the next iteration.

In case that the value of the cost function is equal or greater to the minimal known value $C_{TV}^{CB,min}$, implying that the current set of parameters $\vec{c}_i$ does not lead to less modulator pattern in the estimated primary $\boldsymbol{I}_P^{\text{est},i}$, no changes in the best known primary estimate $\boldsymbol{I}_P^{\text{est}*}$ are made and only the parameter vector $\vec{c}$ is updated. With this approach the modulator pattern in $\boldsymbol{I}_P^{\text{est}*}$ vanishes more and more as the iteration counter increases. The algorithm terminates when the value of the cost function does not change much from one iteration to another. This implies that also the parameter vector $\vec{c}$ and $\boldsymbol{I}_P^{\text{est}*}$ do not change much anymore.

The update step of the parameter vector c described in the context of hiPMSE utilizes a multidimensional Nelder-Mead downhill simplex algorithm (see section 3.4). Since a huge number of 2D convolutions with kernels of the same size as the data matrix (e.g. 648 x 633 pixels) must be employed during the hiPMSE procedure the convolutions were implemented in the Fourier domain for faster computation. Borders were extended by zero-padding and the convolution

result was cropped to the original image dimensions. In case the X-ray beam is perfectly collimated to the detector, zero-padding at the image borders may be the method of choice. It can be noted that in cases of different collimations other border handling techniques can lead to more precise scatter estimations.

After the final scatter estimate has been computed potentially zero or negative pixels in the final primary estimate have been replaced by their neighborhood's mean value in order to prevent errors in the following logarithm operation. Furthermore a Gaussian filter with $\sigma = 3$ was used in regions of strong attenuation where the primary signal fell under a certain threshold t to prevent noise streaks in the final reconstructions. In this thesis $t = e^{-6}$ was used. These last two corrections are part of both the iPMSE and the hiPMSE routine.

3.3.1 Validation procedure

To develop and to verify the hiPMSE algorithm polychromatic MC simulations have been carried out in order to generate a ground truth for both the scatter and primary intensities ($\boldsymbol{I}_S^{MC}$ and $\boldsymbol{I}_P^{MC}$). For simulation data the free parameters of the scatter potential in Equation 3.8 and the parameters of the kernel in Equation 3.10 were fitted to the MC results in a least-squares sense:

$$C(\vec{c}_{fit}) = \|\boldsymbol{I}_S^{MC} - \boldsymbol{I}_S^{CB}(\vec{c}_{fit})\|_2^2 \tag{3.13}$$

where

$$\boldsymbol{I}_S^{CB}(\vec{c}_{fit}) = \boldsymbol{I}_P^{MC} * *K(\vec{c}_{fit}) \tag{3.14}$$

The obtained parameters $\vec{c}_{fit}$ were then used to generate scatter with the convolution based model and this scatter signal was added to the simulated and modulated primary intensity:

$$\boldsymbol{I}_M^{\text{sim}} = \boldsymbol{M}\boldsymbol{I}_P^{\text{sim}} + \boldsymbol{I}_S^{CB} \tag{3.15}$$

The primary intensity $\boldsymbol{I}_P^{\text{sim}}$ was simulated with a ray tracing method. $\boldsymbol{I}_M^{\text{sim}}$ served as the "measured signal" on which the hiPMSE method was implemented and tested. This strategy has been chosen in order to minimize computation time for generating the simulation data and to demonstrate the scatter correction capability of convolution based scatter estimation methods. For real measurement data, where no ground truth was available, correctness of found parameter values was assessed by inspection of visible modulator remains in projection data and by artifact reduction capability in volume data.

3.4 Optimization

All optimization algorithms this thesis utilizes are part of the `scipy.optimize` package for `Python` 3.4.3.5 running under Windows 7 [30]. The methods of Brent and Nelder & Mead find the minimum of a function of one or more variables without the use of derivatives of the objective function. Derivative-free optimization methods are particularly useful if the derivatives of an object function are hard to calculate or do not even exist.

3.4.1 Brent's method

With the iPMSE method only one scatter value I_S^{est} that leads to a modulator free primary patch must be found. This is equivalent to finding the minimum of a function (the cost function $C_{TV}(I_S^{\text{est}})$) of one variable. The method of Brent [31] is an appropriate algorithm to address this problem. Brent's method uses inverse parabolic interpolation whenever possible to speed up the convergence of a golden section method. The golden section method to find a minimum of a function successively brackets the range of values in which the minimum must lie. A minimum is bracketed when a triplet of points $(a < b < c)$ exists such that $f(b) < f(a) < f(c)$. The minimum then clearly lies within the interval (a, c). f is evaluated on a new point x to construct a new bracketing interval from the observed function value at point x. Figure 3.4 exemplifies the successive bracketing of a minimum [32].

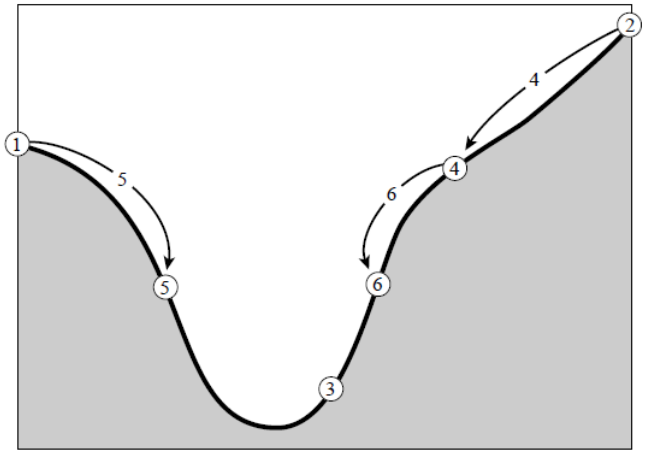


Figure 3.4: Golden section search: successively bracketing of a minimum. The original bracketing triplet consists of points 1,3,2. The function is then evaluated at 4, and this point replaces point 2; then at 5, replacing 1; then at 6 which replaces 4. After these steps the minimum is bracketed by points 5,3,6. Figure taken from [32].

With golden section search the minimum is bracketed with greater precision in each iteration. However, if the objective function is smooth near the minimum (which can be assumed on a well-behaved function) Brent's method speeds up the convergence of the golden section algorithm by fitting a parabola through any three points near the minimum. The minimum of the parabola is computed to determine a new point of the bracketing interval. Figure 3.5 displays inverse parabolic interpolation.

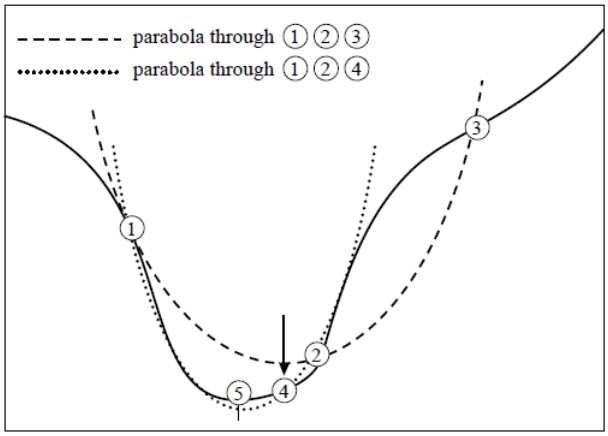


Figure 3.5: Convergence to a minimum by inverse parabolic interpolation. The first parabola (dashed line) is fitted through the points 1,2,3 which lie on the original function (solid line). The objective function is then evaluated on the parabolas minimum (point 4), which replaces point 3. The next parabola (dotted line) is fitted through the points 1,4,2. The minimum of this parabola is at point 5, which is close to the minimum of the objective function. Figure taken from [32].

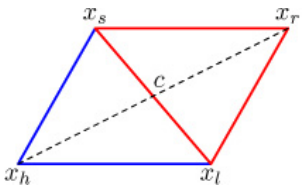
3.4.2 Nelder-Mead downhill simplex algorithm

With hiPMSE a number of free parameters, depending on the used scatter potential and the associated scatter kernel, shall be determined. To this day, the downhill simplex method proposed by Nelder and Mead in 1965 [33] is one of the most commonly used derivative-free optimization algorithms for multidimensional optimization, that is, finding the minimum or maximum of a function of more than one independent variable. While this method is robust it is not ideal in terms of iteration steps [32]. This method has been chosen because it is easy to apply and does not require much computational power.

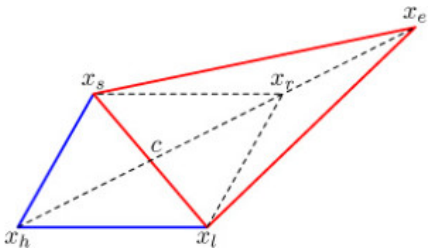
The method defines $N+1$ vertices in the N-dimensional search space. Together these vertices are called a simplex. Thus, in 2D space a simplex is a triangle and in 3D space a simplex is equal to a tetrahedron and so on. It is not possible to bracket the optimum and one is therefore advised to pick an initial guess that is close to the optimum. The starting guess is used to form the first simplex. In each iteration the algorithm evaluates the objective function at each vertex of the simplex. These values are ordered from best (lowest score) to worst (highest score). The simplex is continually refined in each iteration by means of reflection, contraction, expansion of the worst point or shrinkage of the entire simplex around the best known value. The reflected, expanded or contracted new vertex lies on a straight line defined by the highest vertex (worst point) and the centroid for the best side of the current simplex.

Consider a simplex comprised by points (x_h, x_l, x_s). The point with the highest score x_h is reflected through the centroid of the remaining vertices to a point x_r (see Figure 3.6 a)). If x_r is less than the second best point x_s and greater than the lowest vertex x_l, accept this point x_r an move on to the next iteration. If the reflected point x_r is lower than the lowest point x_l, then compute the expansion point x_e (see Figure 3.6 b). If this point x_e is lower than the reflected point x_r, accept x_e and go to the next iteration. If the expanded point x_e is equal or greater

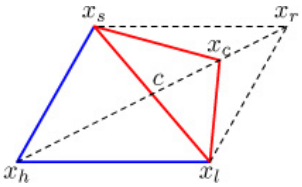
than the reflected point x_r, then accept x_r and move on to the next iteration. If the reflected point x_r is equal or greater than x_s compute the contraction point x_c that can either lie in- or outside of the current simplex. For this, either x_h or x_r are used, depending on which of these points has the lowest score. If $x_s \leq x_r < x_h$ compute x_c and if $x_c \leq x_r$ accept x_c and move to the next iteration. Otherwise, perform a shrink operation. If $x_r \geq x_h$ compute x_c and if $x_c < x_h$ accept x_c and move to the next iteration. Otherwise, perform a shrink operation. The shrink operation contracts the entire simplex towards the best known vertex. The optimization terminates if either a maximum number of iterations has been reached or all vertices gathered around the optimum point [32][33].



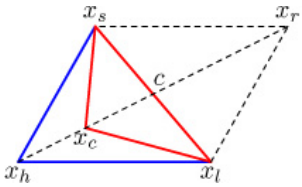
(a) Reflection: the highest point x_h is reflected through the centroid of the remaining vertices to a point x_r.



(b) Expansion: If the reflected point x_r is lower than the lowest point x_l, then the expansion point x_e is computed.



(c) Contraction: If $x_s \leq x_r < x_h$ compute x_c and if $x_c \leq x_r$ accept x_c that lies outside of the current simplex.



(d) Contraction: If $x_r \geq x_h$ compute x_c and if $x_c < x_h$ accept x_c that lies inside of the current simplex.

Figure 3.6: Nelder-Mead downhill simplex algorithm: reflection, expansion and contraction. Figures taken from `http://docs.chejunkie.com/amoeba-method-algorithm/`.

3.5 Hardware

3.5.1 Devices and measurements

All measurements were executed with a laboratory DVT system that is based on a state-of-the-art DVT machine. The X-ray tube operated at a tube voltage of 85kV and a tube current of 7mA. The tube's inherent filtration consisted of 2.5 mm Al and 0.3 mm Cu. The tube's anode angle is 5° with a nominal focal spot value of 0.5. Neither a bow-tie filter nor an anti-scatter-grid

were used. The device uses a flat-panel CsI detector of an approximate size of $16x16$ cm. One pixel has a size of $240x240$ μm. The source-object-distance is 33.3 cm and the source-detector-distance is 52.4 cm. The primary modulator was mounted to the case of the X-ray tube with duct tape resulting in a distance from the focal spot to the modulator of about 15 cm. Each measurement consists of 200 projection images covering an angular range of about 200°. All projection images were pre-processed with the device's inherent correction software.

For each object of interest four independent measurements are necessary. The imaged object must not be moved between measurement 1. and 2. and the same applies for the modulator in measurements 2. and 3.:

1. Measurement of the object only: to asses scatter-affected reconstructions
2. Measurement of the object with the modulator in place: input for iPMSE and hiPMSE methods
3. Measurement of the modulator only
4. Air scan: for normalization purposes

3.5.2 Modulator

A modulator was designed in order to conduct measurements on a state-of-the-art DVT device that can be used in the iPMSE and hiPMSE methods. The pattern of the modulator, the modulator period, modulator material and modulator thickness are the main design aspects one must consider.

Although essentially not required with iPMSE or hiPMSE a checkerboard pattern was the pattern of choice to generate as little deviation to the theory of primary modulator scatter estimation. This pattern could possibly also be used with the classical Fourier-based approach of separating primary and scatter intensities. The side length of one blocker was chosen to be equal to $d_u = d_v = 1$ mm or a modulator period of $p = 2$ mm respectively. The modulator period mapped on the detector amounted to about 25 pixels or approximately 7 mm. The modulator material must have a high density or high mass attenuation coefficient such that the modulator can be thin [34]. The modulator was manufactured on a circuit board using a conventional etching method with a copper thickness of 210 μm (see Figure 3.7).

The total measured transmission of the modulator was approximately 85%, which is in the recommended range of transmission values [35].

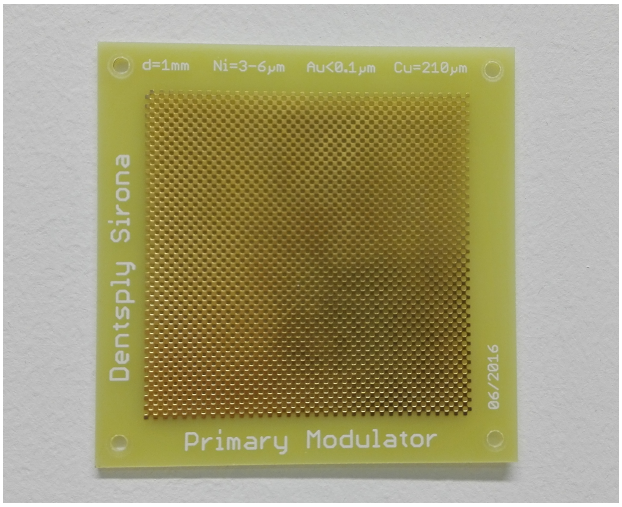


Figure 3.7: The copper checkerboard modulator. The total transmission amounted to about 85% of the incident radiation.

4 Results

4.1 Monte-Carlo simulations of two phantoms

MC simulations were conducted in order to generate a ground truth for both the scatter and primary intensities. For one, realistic scatter-to-primary-ratios (SPR) must be available in order to possibly asses the correctness of the results of both the iPMSE and the hiPMSE method. On the other hand, for simulation data scatter must be generated manually in a realistic way. For this, the free parameters of the convolution-based model introduced in section 3.2 were fitted to the MC-generated scatter intensities such that the obtained parameters $\vec{c}_{fit}$ recreate the underlying scatter distribution with as little deviation as possible.

Two different models were used in this study:

- **Model 1** consists of the scatter potential (see Equation (3.8)) and the scatter kernel from Equation (3.10), i.e.

$$\boldsymbol{I}_S^{\text{est}}(\vec{c}) = \Phi(\boldsymbol{I}_P^{\text{est}*}, \vec{c}) * *K_2(\vec{c}) \tag{4.1}$$

- **Model 2** is a simplified version of model 1. The parameters of the scatter potential were set to $c_0 = 0$ and $c_1 = 1$ for ease of computation. Also, a simple 2D Gaussian function was used as a scatter kernel (see Equation (3.9)), i.e.

$$\boldsymbol{I}_S^{\text{est}}(\vec{c}) = \Phi(\boldsymbol{I}_P^{\text{est}*}) * *K_1(\vec{c}) \tag{4.2}$$

Utilization of the second model results in a fitting problem with only one free parameter (the width σ of the kernel). Both fits can also be seen as a test to verify that the used scatter potentials and kernels are principally capable of estimating true scatter distributions.

Two different objects were of interest in this study: first, a simple water cylinder was examined. Secondly, a water phantom with three different inserts (air, aluminium and teflon) was simulated. All MC simulations were carried out together with the German Cancer Research Center (DKFZ, Heidelberg, Germany) utilizing an in-house MC simulation environment. The geometry of a state-of-the-art DVT machine was accurately modeled by employing projection matrices that were calibrated on a real device.

4.1.1 Water cylinder

Model 1

Figure 4.1 depicts the results of the MC simulations for the simple water cylinder. The projection of the cylinder is not exactly upright due to slight errors in the projection matrices from the device. This does, however, not diminish the validity of these results.

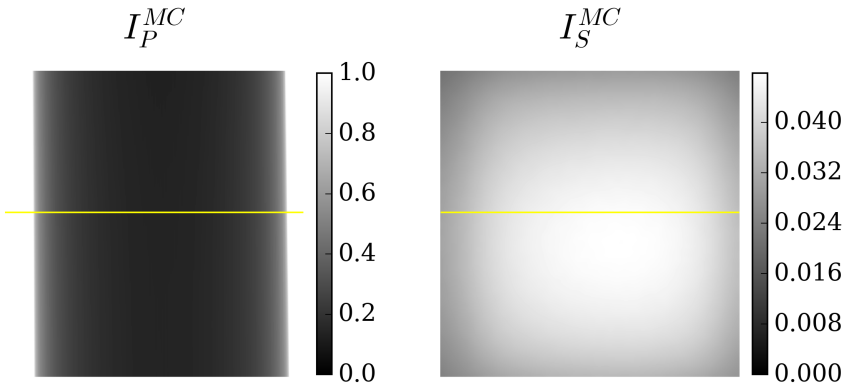


Figure 4.1: Water cylinder: MC-simulated intensities $\boldsymbol{I}_P^{MC}$ and $\boldsymbol{I}_S^{MC}$.

$\boldsymbol{I}_P^{MC}$ and $\boldsymbol{I}_S^{MC}$ were used to calibrate the open parameters in a convolution-based scatter model utilizing the cost function in formula 3.13 to obtain $\boldsymbol{I}_S^{CB}$. Specifically, the scatter potential from formula 3.8 and the scatter kernel K_2 from formula 3.10 have been used. This results in a model with 4 open parameters, two of them (c_0 and c_1) defining the shape and magnitude of the scatter potential Φ and the other two (b and σ) influencing the shape of the scatter kernel. On the left side of Figure 4.2 the resulting scatter distribution $\boldsymbol{I}_S^{CB}$ that was generated with the convolution-based model and the fitted parameters $\vec{c}_{fit}$ can be seen. The right image in Figure 4.2 depicts the deviations of $\boldsymbol{I}_S^{CB}$ to $\boldsymbol{I}_S^{MC}$. The total mean error was 7.63% and mainly stems from the outer image areas. This may be due to the fact that the scatter potential that was convoluted with the scatter kernel was extended by zero-padding at the image borders. In table 4.1 the fitted parameter values can be seen. c_0 and c_1 are dimensionless scaling factors whereas b and σ are given in units of pixels. The negative sign of the parameter b may look odd, but does in fact not matter as the kernel is rotationally symmetric. The profile through the middle of the resulting normalized convolution kernel can be seen in Figure 4.3.

Table 4.1: Individual components of $\vec{c}_{fit}$ for the water cylinder for model 1.

c_0	c_1	b	σ
0.027	0.068	-29.459	63.571

In Figure 4.4 line profiles through the MC-simulated scatter and primary estimates, the convolution based scatter estimate and the scatter potential can be seen. The convolution based scatter estimate shows very little deviation to the MC-based scatter profile. This indicates that

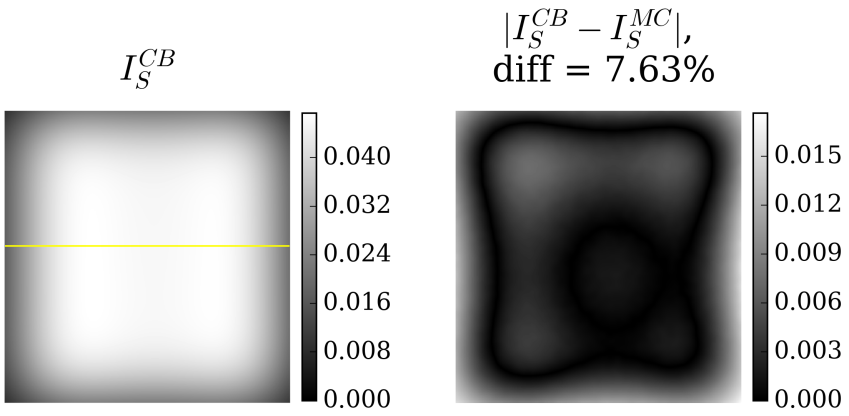


Figure 4.2: Water cylinder - Model 1: Left: Convolution-based scatter estimate $\boldsymbol{I}_S^{CB}(\vec{c}_{fit})$. Right: Difference to the scatter reference image $\boldsymbol{I}_S^{MC}$. The difference given above the right image is the total overall percental difference.

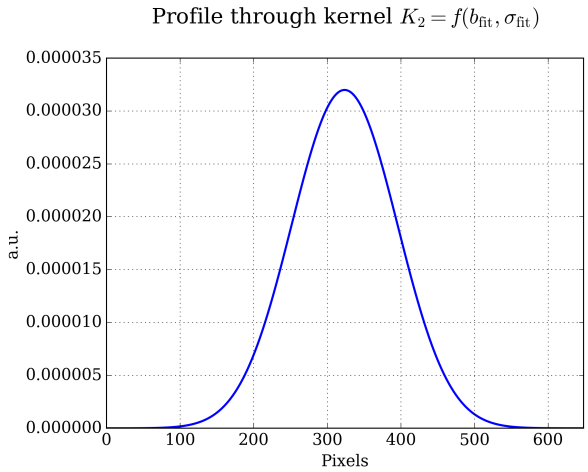


Figure 4.3: Water cylinder - Model 1: Profile through the normalized scatter kernel $K_2 = f(b_{fit}, \sigma_{fit})$.

the convolution-based model is in fact capable of accurately estimating scatter.

Model 2

For the model 2 the originally introduced parameters in the scatter potential were set to $c_0 = 0$ and $c_1 = 1$. In Figure 4.5 the resulting scatter distribution from the simplified Gaussian model with only one free parameter, σ, is illustrated. The total error to the MC generated scatter image is 6.99% and thus a bit smaller than in the above case with model 1.

Figure 4.6 shows line profiles through MC-simulated primary and scatter intensities, through the convolution-based scatter estimate and the scatter potential for the model 2. The fitted parameter of the 2D Gaussian distribution of the scatter kernel was $\sigma = 608$ pixels. While with both models the MC scatter distribution can be modeled accurately, the used scatter potential in each model greatly differs. With the model 2 the overall intensity in the scatter potential is much higher in object regions, and the resulting kernel is about 10 times wider than with model 1. The scatter potential in model 2 does however drop rapidly to the image borders where the

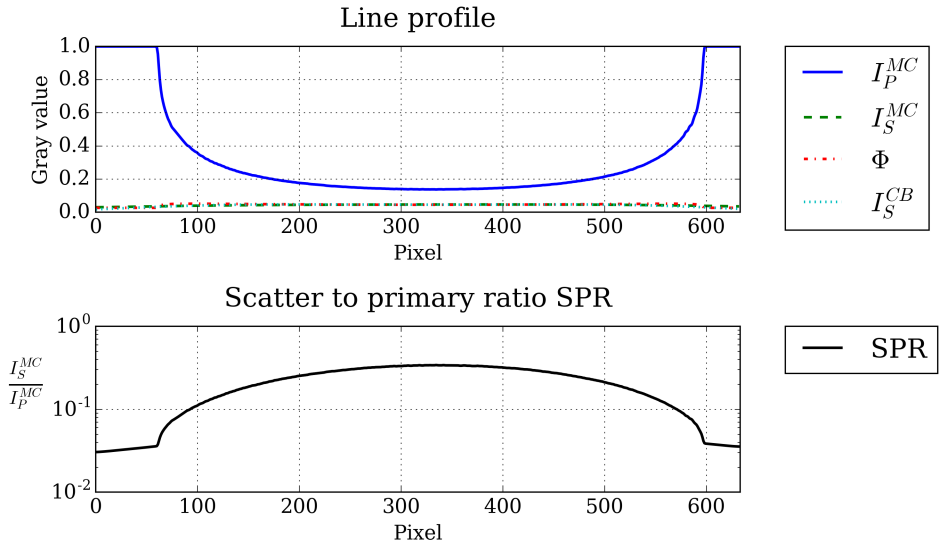


Figure 4.4: Water cylinder - Model 1: Top: Line profiles through the MC-simulated primary and scatter intensities, a line profile through the convolution-based scatter estimate generated with least-square fitted parameters and through the scatter potential. Bottom: Scatter-to-primary ratio from MC simulations.

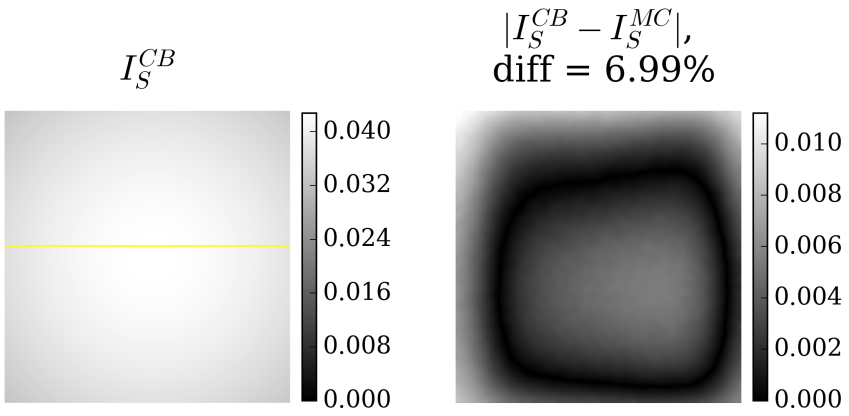


Figure 4.5: Water cylinder - Model 2: Left: Convolution-based scatter estimate $\boldsymbol{I}_S^{CB}(\sigma_{fit})$ and difference to the scatter reference image $\boldsymbol{I}_S^{MC}$.

X-rays did not penetrate the object but air. The lack of scaling factors c_0 and c_1 in the scatter potential may result in a relatively wide kernel compared to model 1 in order to smooth out the distinct edge in the scatter potential and to lead a smooth and flat scatter distribution.

4.1.2 Water phantom with inserts

Model 1

Figure 4.7 depicts the results of the MC simulations for the water phantom with the three different inserts.

On the left side of Figure 4.8 the resulting scatter distribution $\boldsymbol{I}_S^{CB}$ that was generated with model 1 can be seen. The right image in Figure 4.8 depicts the deviations of $\boldsymbol{I}_S^{CB}$ to $\boldsymbol{I}_S^{MC}$. It

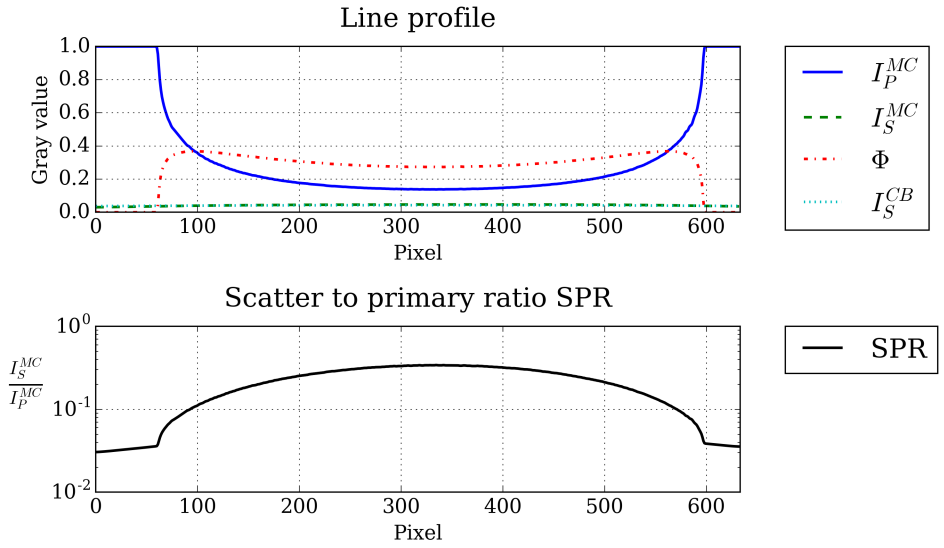


Figure 4.6: Water cylinder - Model 2: Top: Line profiles through the MC-simulated primary and scatter intensities, through the convolution-based scatter estimate generated with model 2 and through the scatter potential. Bottom: Scatter-to-primary ratio from MC simulations.

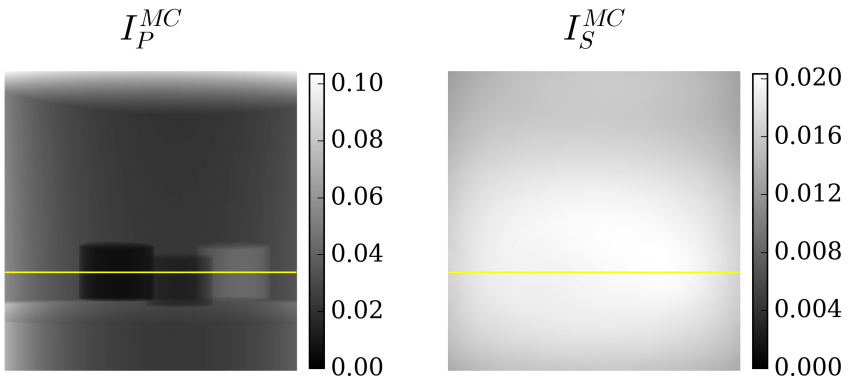


Figure 4.7: Water phantom: MC-simulated intensities $\boldsymbol{I}_P^{MC}$ and $\boldsymbol{I}_S^{MC}$ for the water phantom with inserts.

can be noted that both scatter signals are flat and smooth. The main differences occur again at the image borders and in the top image area. The total mean error was 9.32%. In table 4.2 the fitted parameter values can be seen. A profile through the resulting normalized convolution kernel is illustrated in Figure 4.9. Compared to the previous kernel parameters of the water cylinder in section 4.1.1 the scatter kernel is wider and does show effectively no shift in the two summed Gaussian distributions.

Table 4.2: Individual components of $\vec{c}_{fit}$ for the water phantom for model 1.

c_0	c_1	b	σ
0.009	0.132	0.002	96.477

In Figure 4.10 line profiles through the MC-simulated scatter and primary estimates, the convolution based scatter estimate and the used scatter potential can be seen. As noted before,

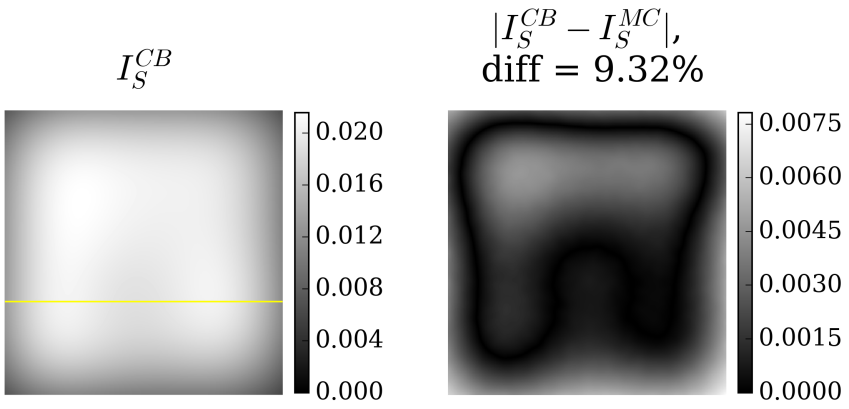


Figure 4.8: Water phantom - Model 1: Left: Convolution-based scatter estimate $\boldsymbol{I}_S^{CB}(\vec{c}_{fit})$. Right: difference to the scatter reference image $\boldsymbol{I}_S^{MC}$.

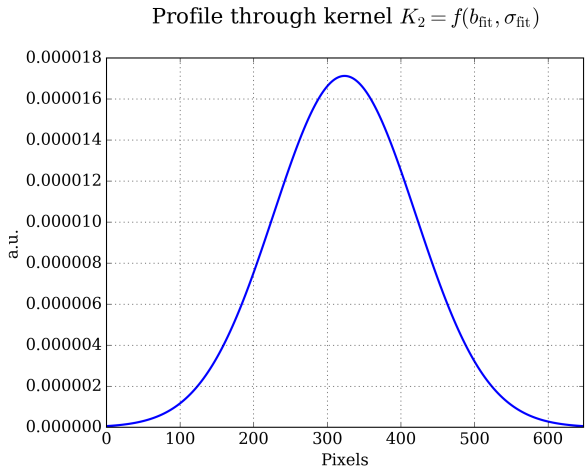


Figure 4.9: Water phantom - Model 1: Profile through the normalized scatter kernel $K_2 = f(b_{fit}, \sigma_{fit})$.

the biggest deviations in the convolution-based scatter distribution to the MC scatter estimate occur at the image borders.

Model 2

Figure 4.11 shows the resulting scatter distribution of the simplified model 2. The total error was 5.12% and, as with the water cylinder, again smaller than with the more complex model 1. At the bottom of the difference image the borders of two inserts can be recognized. Also, the inserts are bordered by a region of close-to-zero error.

Figure 4.12 depicts line profiles through the MC-simulated scatter and primary estimates, the convolution based scatter estimate and the used scatter potential. The fitted parameter of the 2D Gaussian distribution of the scatter kernel was $\sigma = 572$ pixels and thus a little smaller but in the same range as above for the water cylinder. Again, the scatter potential is much higher than the corresponding MC primary intensity and the broad 2D Gaussian kernel is used to flatten this potential.

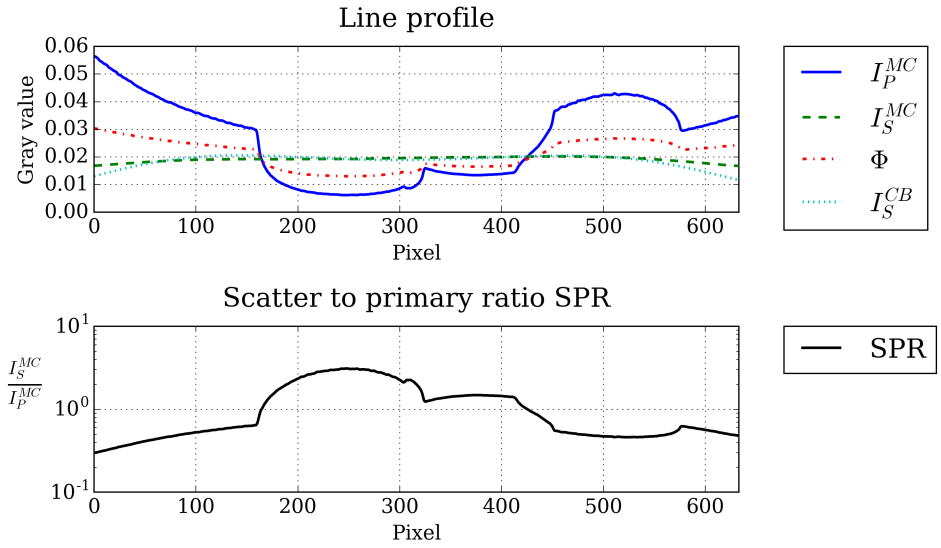


Figure 4.10: Water phantom - Model 1: Top: Line profiles through MC-simulated primary and scatter intensities, through the convolution-based scatter estimate generated with model 1 and the scatter potential. Bottom: Scatter-to-primary ratio from MC simulations.

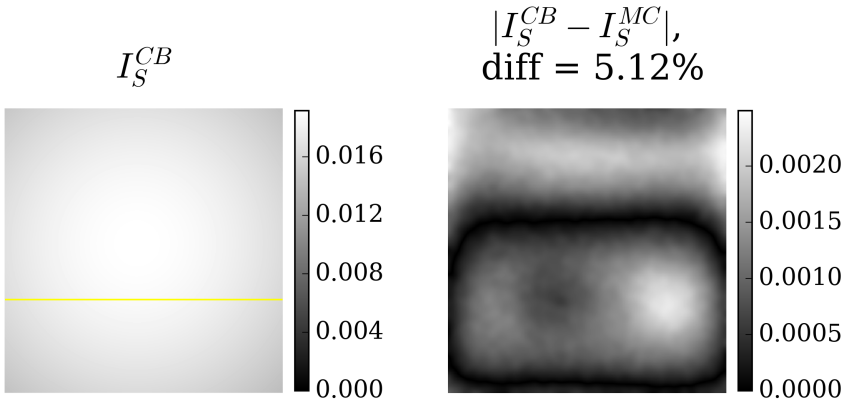


Figure 4.11: Water phantom - Model 2: Left: Convolution-based scatter estimate $\boldsymbol{I}_S^{CB}(\vec{c}_{fit})$ and difference to the scatter reference image $\boldsymbol{I}_S^{MC}$.

It has been shown that both used models are in fact capable of recreating a realistic scatter distribution from measured data. For both the water cylinder and the water phantom the obtained scatter parameters of the first model were used to generate scatter for non-MC simulated monochromatic projections. The resulting image was then used to verify both the iPMSE as well as the hiPMSE methods. Since both scatter models are interchangeable and with comparable scatter recreation potential the second model was used to estimate scatter from real measurement data due to ease of computation.

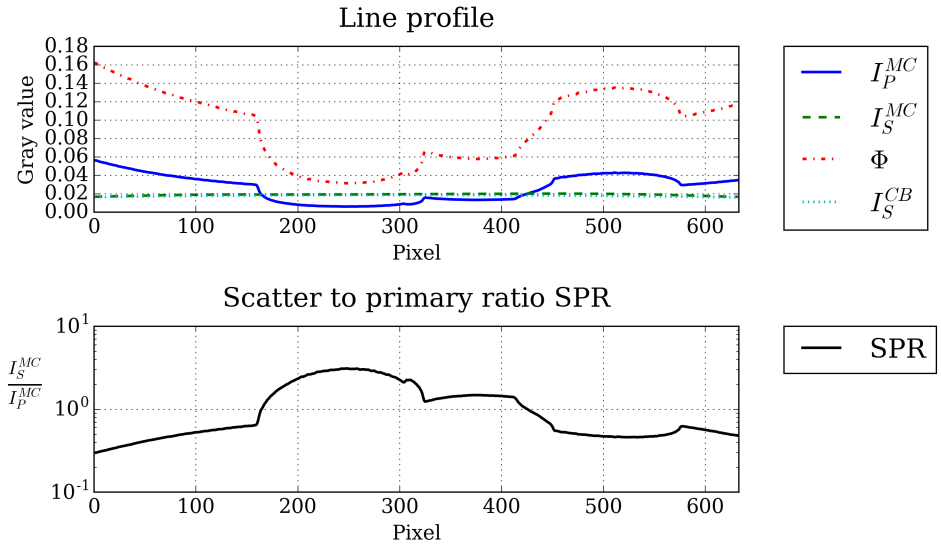


Figure 4.12: Water phantom - Model 2: Top: Line profiles of MC-simulated primary and scatter intensities through the convolution-based scatter estimate generated with model 2 and through the scatter potential. Bottom: Scatter-to-primary ratio from MC simulations.

4.2 Improved Primary Modulator Scatter Estimation (iPMSE)

The iPMSE technique was first implemented and tested on monochromatic simulation data. When the algorithm was stable and proven to work well under ideal conditions measurements were carried out in order to verify the method for real data.

4.2.1 Simulations

Jaw phantom

The iPMSE method was tested on different phantoms and for different amounts of scatter and worked well independently of the imaged object. The influence of strong object edges on the estimated scatter value was found to be less significant for simulation data than for real measurements. This may be due to the fact that in simulation data the modulator edges are always perfectly sharp. In this section an exemplary simulated measurement of a jaw phantom will be shown and discussed. The jaw phantom consists of a head-shaped water form (tissue) with several vertebral bodies, lower and upper jaws and teeth. Most of the teeth are composed of dentin but some were replaced with titanium in order to emulate dental implants. The simulated primary signal $\boldsymbol{I}_P$ was generated by simply forward projecting a geometrical object, the jaw phantom, whereby no scatter and only monochromatic X-rays occurred. The scatter was in this case simulated with the convolution based model that was introduced in section 3.2, specifically the kernel proposed by Zhao (model 1, see formula 3.10), and then added to the modulated primary intensity. For this, the parameters from table 4.2 were used. The modulation of the primary signal was achieved by multiplication of the primary image $\boldsymbol{I}_P$ with a modulator of alternating blockers of air ($m_{ii} = 1$) and a blocker material ($m_{ii} = 0.6$). For all simulation data Poisson noise with $N = 10000$ was added to the modulated primary signal, leading to a

signal-to-noise ratio of about 80 in the object. Figure 4.13 shows the simulated images for $\boldsymbol{I}_M$, $\boldsymbol{I}_P$ and $\boldsymbol{I}_S$.

The estimated primary and scatter signals $\boldsymbol{I}_P^{\text{est}}$ and $\boldsymbol{I}_S^{\text{est}}$ can be seen in Figure 4.14. The patch size was chosen as $d_u = d_v = 43$ pixels with a modulator period of approximately 30 pixels. Figure 4.15 displays the behavior of the cost function $C_{TV}(I_S^{\text{est}})$ (see Equation 3.6) for different values of I_S^{est} for a patch located in a tissue area of the jaw phantom. The usage of a L1-norm leads to a cost function that is convex, such that only one optimal solution per patch exists. This is confirmed by the fact that the strength of edges introduced by the modulator is proportional to the error Δs [16].

The patch-wise ansatz for the solution of the minimization problem leads to a somewhat stained appearance of the scatter signal that is not exactly as smooth as the simulated scatter image. The overall shape of the scatter intensity is, however, reproduced. Figure 4.16 attests high accuracy to the iPMSE method in this case since the line profiles of the estimations precisely follow the profiles of the simulated intensities. The SPR reaches values as high as approximately 1.5 behind the strongly absorbing titanium inserts.

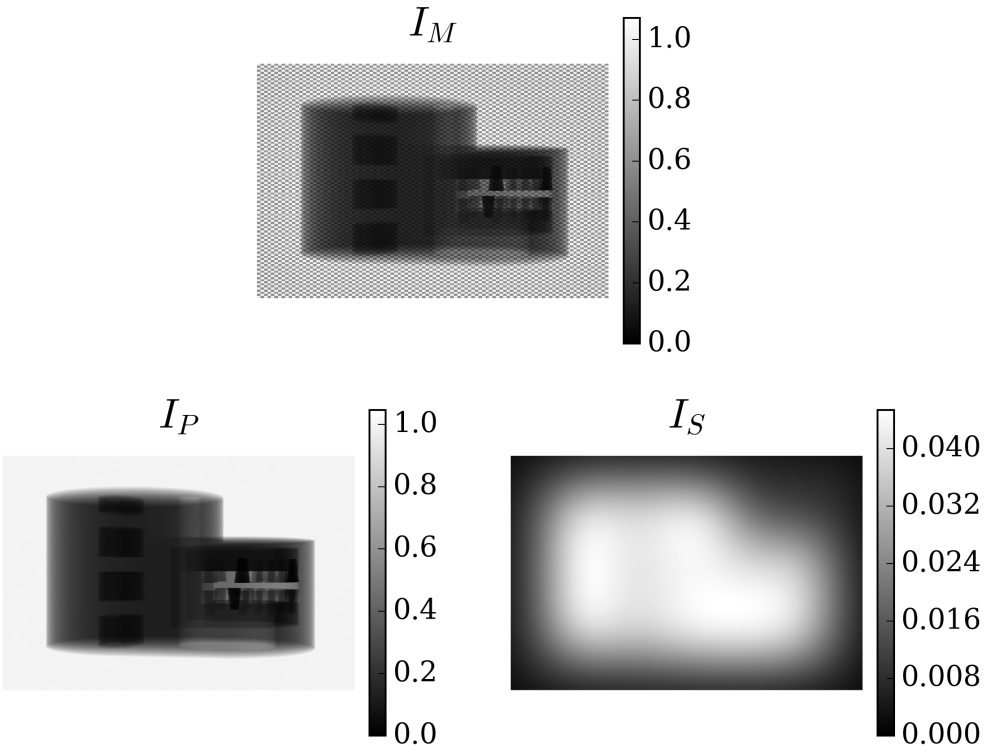


Figure 4.13: Jaw phantom: Simulated intensities $\boldsymbol{I}_M$, $\boldsymbol{I}_P$ and $\boldsymbol{I}_S$. The primary intensity was generated from monochromatic forward projections whereas the scatter intensity was generated with a convolution-based model.

Figure 4.17 displays a scatter affected and a scatter corrected slice of the reconstructed volume. Note that for all displayed slice images in this work the gray values are arbitrary scaled and do net represent Houndsfield units. As expected, both the cupping artifact in the homogeneous water body and streaking artifacts between the titanium insert and teeth are present in the

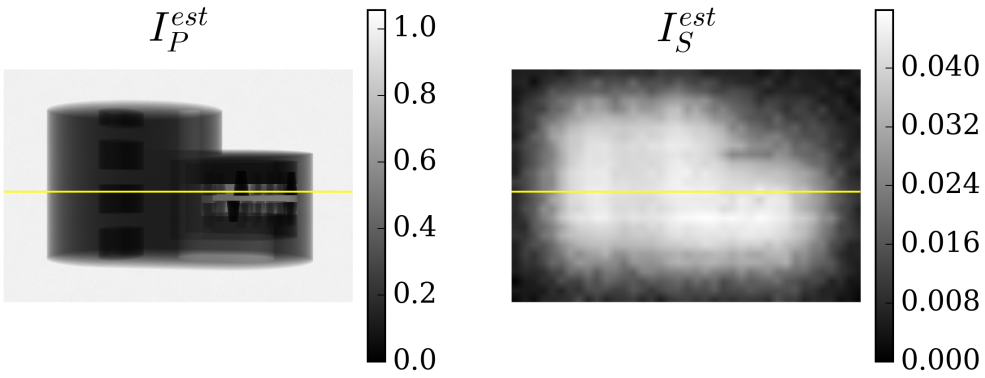


Figure 4.14: Jaw phantom - iPMSE: Estimated intensities $\boldsymbol{I}_P^{est}$ and $\boldsymbol{I}_S^{est}$. The yellow line indicates the location of the line profiles in Figure 4.16.

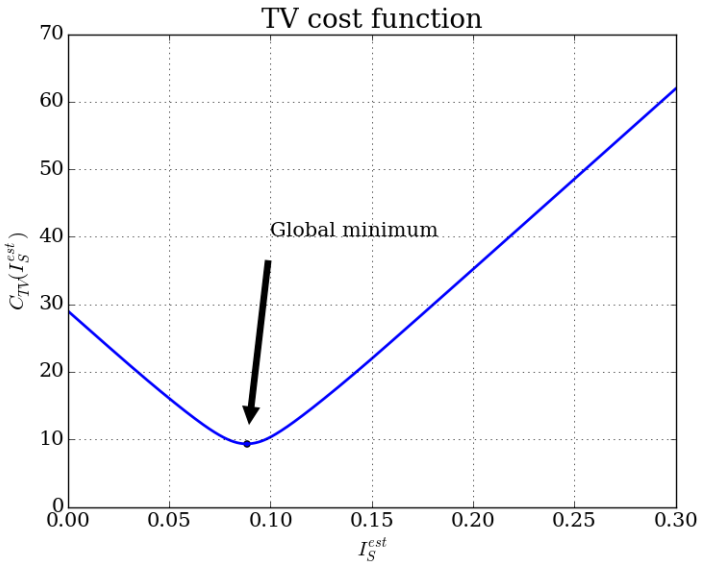


Figure 4.15: Jaw phantom - iPMSE: Behavior of the cost function for a patch in the tissue-like area for different scatter values. A global minimum is found at $I_S^{est} = 0.088$.

scatter-affected reconstructions. The scatter corrected slice is almost free of scatter artifacts. Compared to the scatter-free reference on the left in Figure 4.18 only little deviations occur as streaks emerging from the titanium teeth. This is also evident from the difference image on the right side of Figure 4.18 where directed noise structures are visible. The overall noise level in the scatter-corrected slice is, as with all subtraction-based scatter correction algorithms, increased. This is due to the fact that the scatter noise remains in the image after scatter correction [23]. Line profiles through the water shape that are again indicated by the yellow lines can be seen in Figure 4.19. The scatter-corrected profile shows very little deviations to the scatter free reference. The cupping artifact along with the deviations from the ideal rectangular profile shape due to streaks could be almost completely eliminated.

It has been shown that the iPMSE method works well on simulation data and under ideal conditions. Both the prominent cupping and streaking artifacts due to the occurrence of scattered

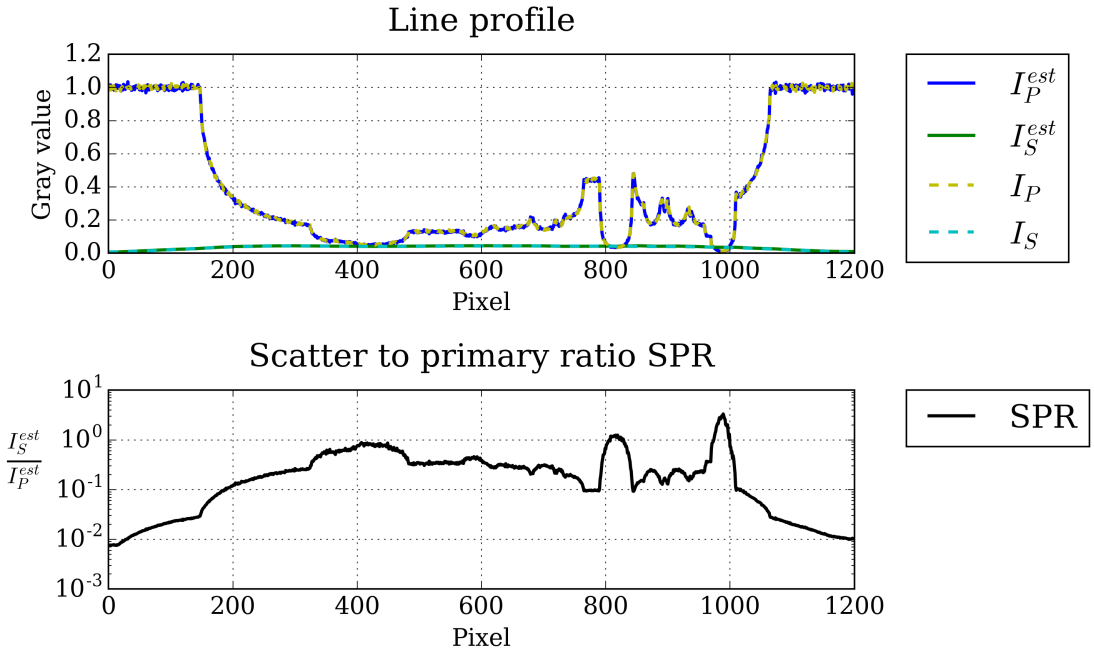


Figure 4.16: Jaw phantom - iPMSE: Top: Line profiles of simulated and estimated primary and scatter intensities. The estimated profiles (solid lines) show very little deviation to the true profiles (dashed lines). Bottom: Estimated scatter-to-primary ratio.

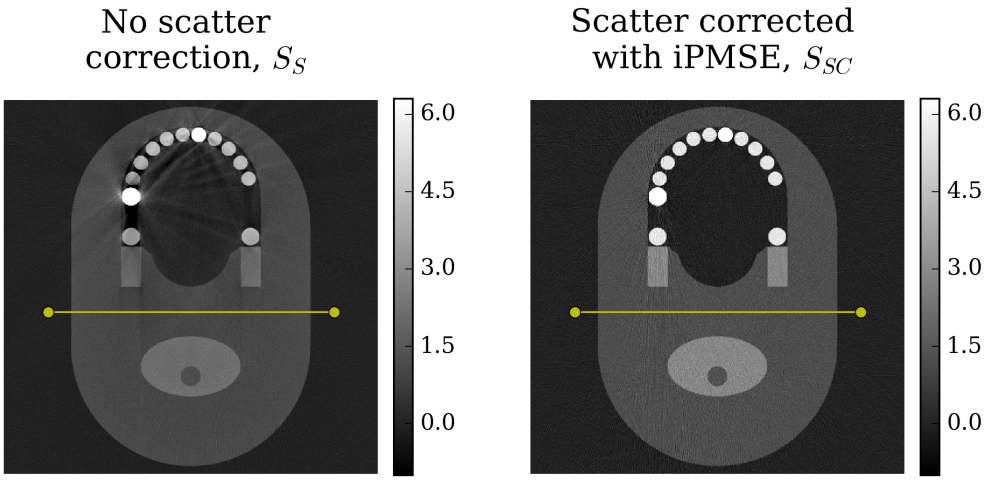


Figure 4.17: Jaw phantom - iPMSE: A scatter affected and a scatter corrected slice of the reconstructed volume.

X-rays could be reduced substantially to improve homogeneity and overall visual appearance. However, the patch-wise approach used in the estimation of the scatter signal leads to a patchy visual appearance of the estimated scatter intensities. Although the patches overlap by half the patch size in u- and v-direction, sometimes even neighboring patches show divergent scatter values. This may be due to the influence of noise or due to the fact that the cost function does not only detect modulator edges but also sharp object edges. Consequently the estimated scatter values are influenced by the presence of strong object edges in the projection images.

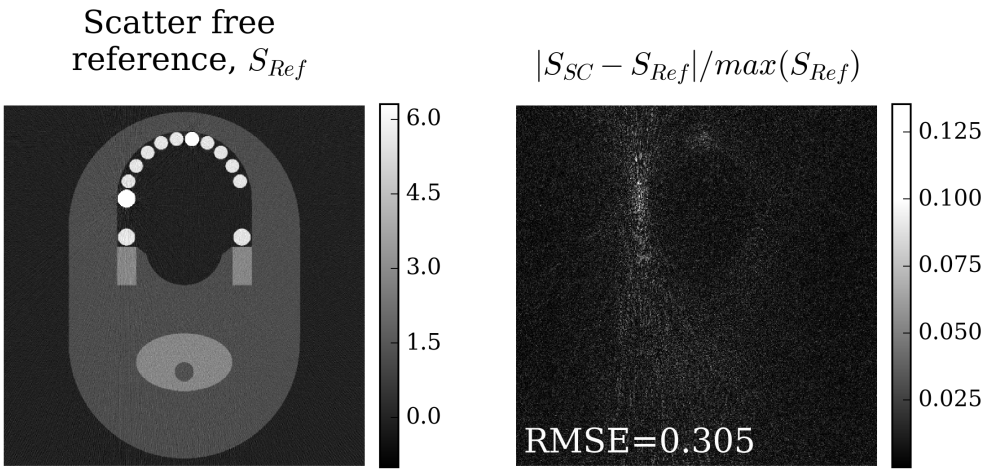


Figure 4.18: Jaw phantom - iPMSE: Left: Scatter free reference. Right: Difference of scatter-corrected and scatter-free images in relation to the maximum of the scatter free reference. The root mean squared error (RMSE) is defined as $RMSE = \sqrt{mean\left(\left(\frac{S_{SC}-S_{Ref}}{max(S_{Ref})}\right)^2\right)}$*, where mean() computes the average value over all pixels of an image.*

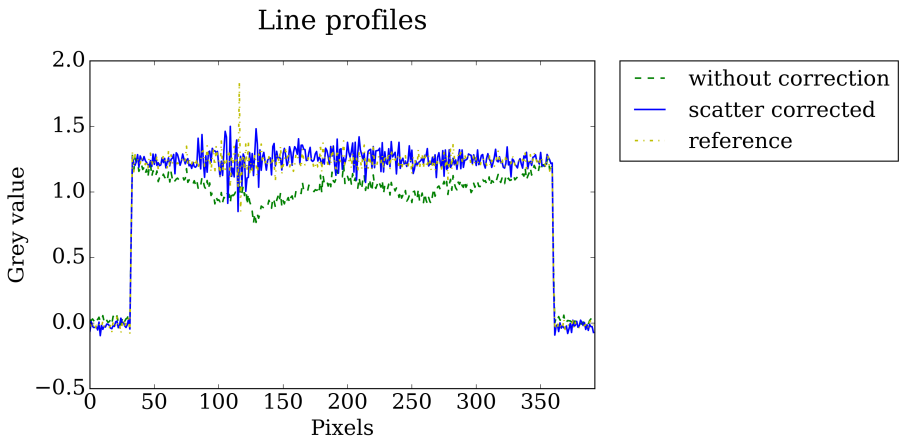


Figure 4.19: Jaw phantom - iPMSE: Line profiles through the scatter-affected, the scatter-corrected and the reference slice of the three respective reconstructions. RMSE stands for Root Mean Square Error and is defined as

Klein addressed this issue in [24] and proposes to overcome this problem with a modified cost function that considers the estimated gradient of the object.

4.2.2 Measurements

Water cylinder

Figure 4.20 shows measured projection images for a water cylinder with the modulator in place ($\boldsymbol{I}_M$) and the modulator $\boldsymbol{M}$ only. The water cylinder occupies almost the complete FOV except for a small area of air on both sides of the cylinder. The projection image of the modulator $\boldsymbol{M}$ shows a typical artifact that originates from the relatively small anode angle of 5° in combination with the finite size of the focal spot: a spatially variant blurring of the checkerboard pattern

can be observed. On the left side of the modulator projection a clear differentiation between the copper and the air areas of the modulator is possible. The projected modulator pattern exhibits sharp edges. On the right side of the projection image of the modulator the situation is different. To the top and to the bottom the projection image of the modulator changes from the ideal checkerboard pattern to a diagonal stripes pattern. In the middle, however, the checkerboard pattern can still be observed but with much more blurring and less distinct edges than on the right side. This problem is inherent to all measurements indifferent of the imaged object.

Figure 4.21 shows the estimated primary and scatter intensities. In this case, the iPMSE algorithm does not lead to the desired results. First, the scatter estimate $\boldsymbol{I}_S^{\text{est}}$ has a very patchy appearance. In can be noted that both edges of the water cylinder have great influence on the scatter estimation. Oscillating estimated scatter values along the left edge of the cylinder can be observed. In the area of the right image border side of the scatter estimation scatter is overestimated. Presumably this is due to the fact that the modulator edges are strongly blurred in this region and the cost function fails to correctly estimate a scatter values due to a lack of sharp edges. The error in the scatter estimation directly translates into the primary estimation $\boldsymbol{I}_P^{\text{est}}$. In regions were the scatter was overestimated the modulator pattern is still visible in the estimated primary. The line profile (see Figure 4.22) through the scatter image however shows, except for the area in the far right, a flat behavior of the scatter estimated as expected. The SPR reaches a maximum of approximately 0.5 which lies, compared to the MC simulations, in a realistic range.

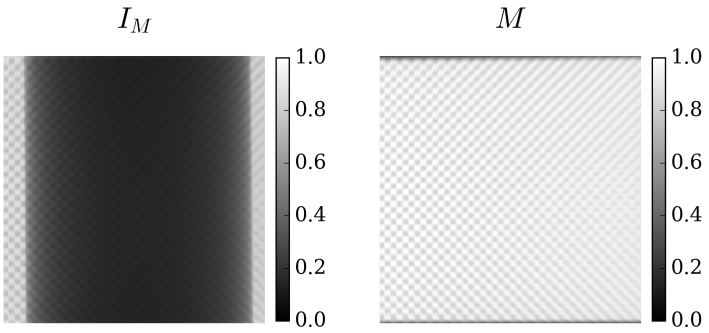


Figure 4.20: Water cylinder: Measured intensities $\boldsymbol{I}_M$ and $\boldsymbol{M}$.

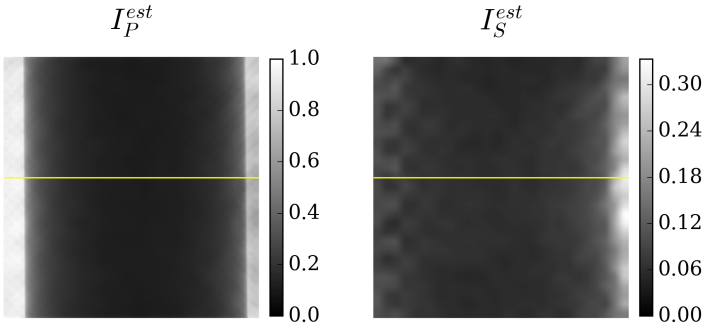


Figure 4.21: Water cylinder - iPMSE: Estimated intensities $\boldsymbol{I}_P^{est}$ and $\boldsymbol{I}_S^{est}$.

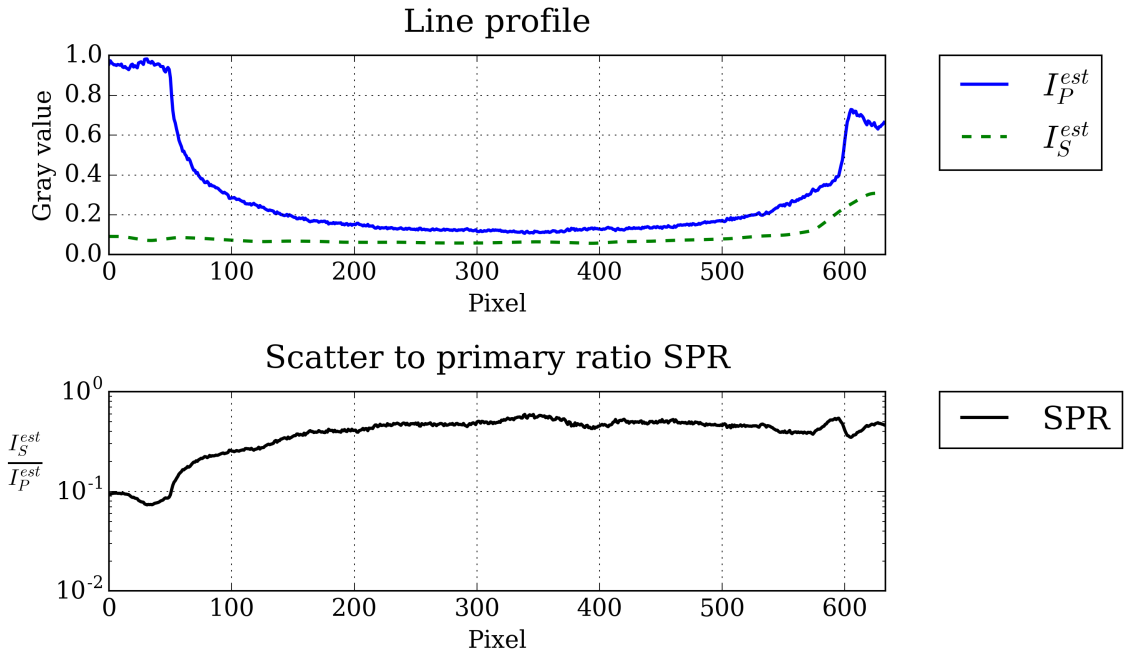


Figure 4.22: Water cylinder - iPMSE: Top: Line profiles of estimated primary and scatter intensities. Bottom: Estimated scatter-to-primary ratio.

Figure 4.23 shows scatter-affacted and scatter-corrected slices of a reconstruction. As expected, the patchy appearance of the primary estimation greatly deteriorates image homogeneity. The modulator remains in the primary estimate lead to ring artifacts in a reconstructed slice. The potential benefit from the iPMSE scatter correction is annihilated by the errors in the scatter- and primary estimations. Comparison of profiles through the reconstructed slices still suggests that the cupping artifact that is present in the uncorrected slice could be reduced. This comes at the cost of greatly reduced image homogeneity.

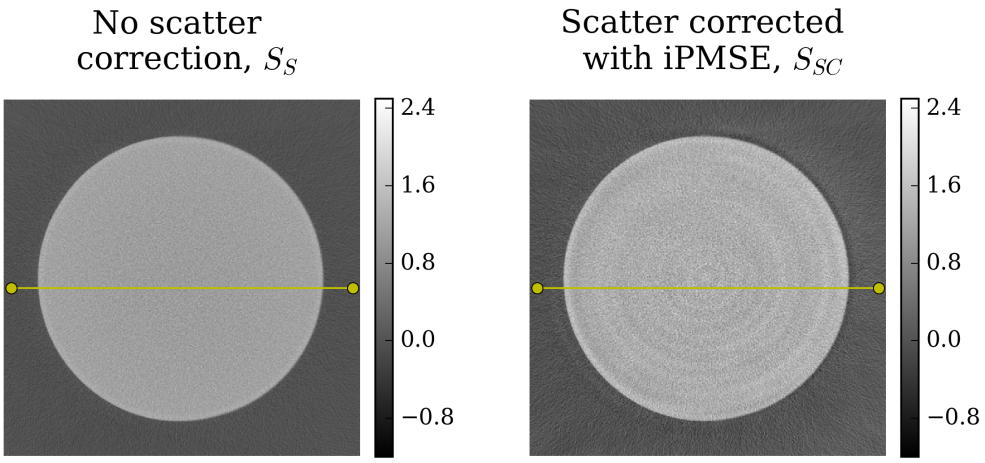


Figure 4.23: Water cylinder - iPMSE: A scatter affected and a scatter corrected slice of the reconstructed volume.

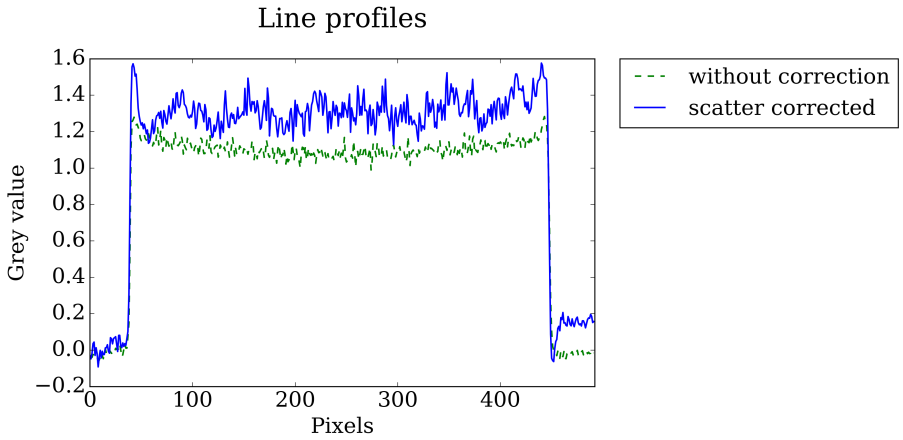


Figure 4.24: Water cylinder - iPMSE: Line profiles through the scatter-affected and the scatter-corrected reconstructions.

In can be noted that in regions were distinct modulator edges are present the TV cost function works as expected. Problems arise when strong object edges are mistaken for modulator edges by the cost function and the estimated scatter value tries to compensate for these strong edges. Also, the strong blurring especially on the right side of the modulator images leads to wrong scatter values in these regions.

4.3 Hybrid Primary Modulator Scatter Estimation (hiPMSE)

4.3.1 Simulations

In this section the results of the novel hiPMSE method are presented. The method worked well on simulation data where scatter was generated with the same convolution-based model that was used in the estimation of the scatter in the hiPMSE method. As for the iPMSE method,

the simulation data has again been generated according to the procedure described in section 3.3.1. In particular, Model 1 was employed with the obtained parameters from sections 4.1.1 and 4.1.2 respectively to generate a realistic scatter distribution that was subsequently added to the modulated primary intensity. This total signal $\boldsymbol{I}_M$ served as the input for the hiPMSE method.

Water cylinder

Figure 4.25 shows the simulated scatter and primary intensities as well as the total simulated intensity $\boldsymbol{I}_M$. The scatter intensity was generated with the parameters from table 4.1. Figure 4.26 shows the corresponding scatter and primary estimations. No modulator is visible in $\boldsymbol{I}_P^{\text{est}}$ which indicates that the error in the scatter estimation Δs is relatively small. The corresponding line profiles can be seen in Figure 4.27. The estimated profiles follow the simulated ones precisely. The SPR reaches a maximum of about 0.3 in the center of the water cylinder.

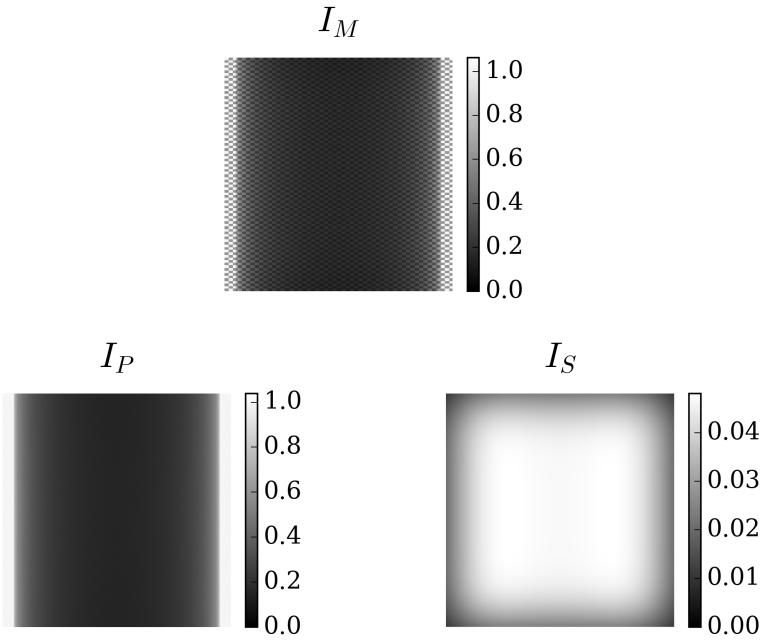


Figure 4.25: Water cylinder: Simulated intensities $\boldsymbol{I}_M$, $\boldsymbol{I}_P$ and $\boldsymbol{I}_S$.

Reconstructions from both the scatter-affected and the scatter-corrected sinograms are shown in Figure 4.28. The scatter-affected slice exhibits an underestimation of attenuation values in the middle of the object due to scatter, which is also evident from the line profile in Figure 4.29. The scatter-corrected slice does not show this cupping artifact and the profile through this slice follows the scatter-free reference accurately.

Water phantom

Figure 4.30 shows slices from the scatter-affected and the scatter-corrected reconstructions of the water phantom with the three inserts. The scatter-affected slice shows a noticeable cupping

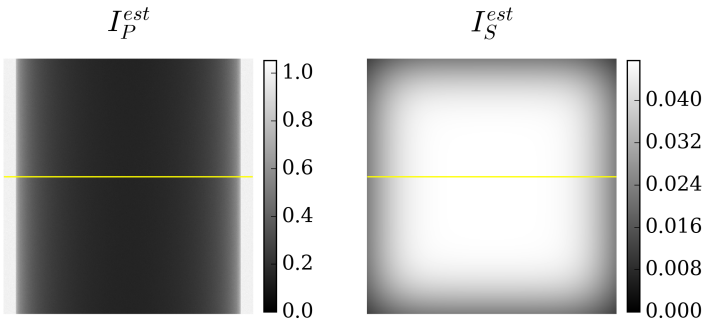


Figure 4.26: Water cylinder - hiPMSE: Estimated intensities $\boldsymbol{I}_P^{est}$ and $\boldsymbol{I}_S^{est}$.

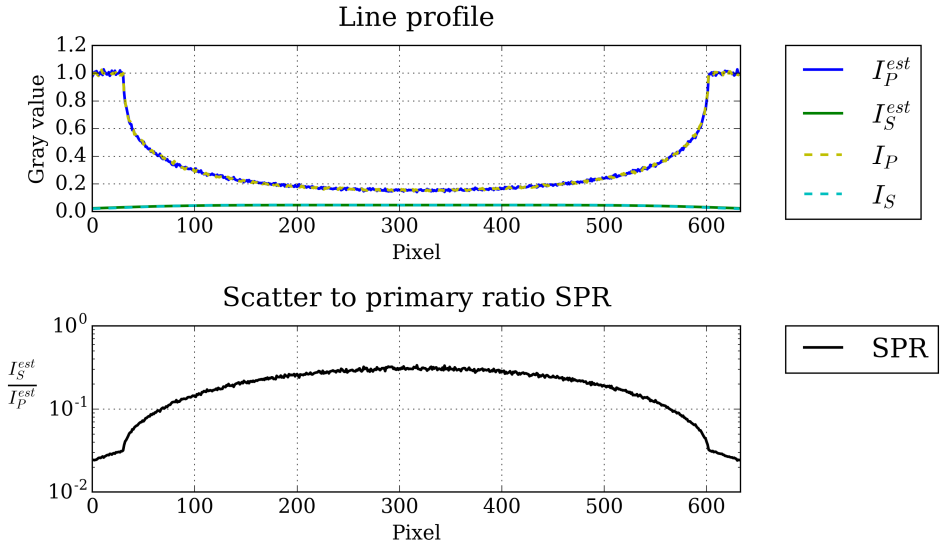


Figure 4.27: Water cylinder - hiPMSE: Top: Line profiles of simulated and estimated primary and scatter intensities. Bottom: Estimated scatter-to-primary ratio.

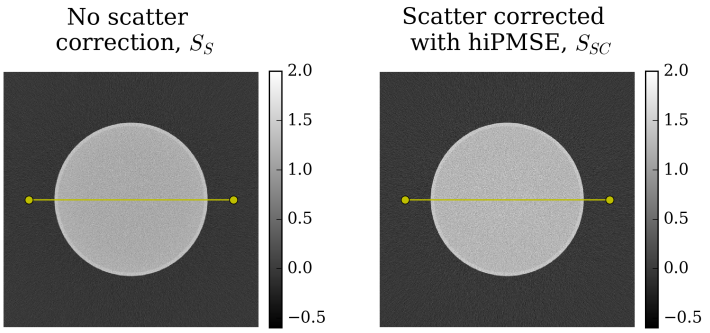


Figure 4.28: Water cylinder - hiPMSE: A scatter affected and a scatter corrected slice of the reconstructed volume.

artifact in the aluminium insert as well as dark streaks between the aluminium and the teflon insert and between the air and the aluminium insert. In the scatter-corrected slice, none of these artifacts can be observed. Figure 4.31 depicts a scatter-free reference slice and the deviation of the scatter-corrected slice to this reference. The line profiles in Figure 4.32 visualize the

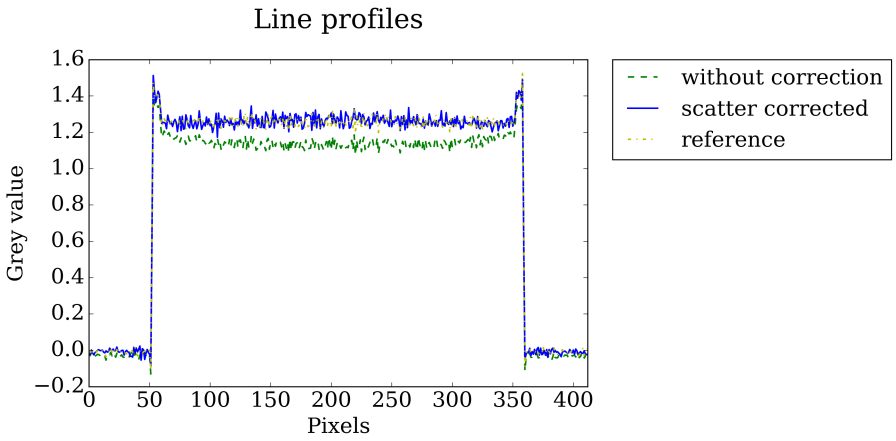


Figure 4.29: Water cylinder - hiPMSE: Line profiles through the scatter-affected, the scatter-corrected and the reference slice of the three respective reconstructions.

artifact correction potential of the hiPMSE method. The cupping artifact in the aluminium insert is greatly reduced. The noise level inside the insert increases, which is also evident from the difference image in Figure 4.31.

Figure 4.33 compares estimated values of the parameter vector $\vec{c}^*$ to the simulated reference values for every projection image. The mean values of the estimated parameters are compared to the reference values in table 4.3. The parameters c_0 and c_1 were estimated with great precision on average, but both show variations over the 200 independent projections. The parameter b that determines the shift in the summed Gaussians in Equation 3.10 is greatly overestimated, while the actual width of the Gaussian distributions σ is slightly underestimated. Together these over-and underestimations seem to cancel each other as the sum of two smaller but displaced Gaussians exhibits effectively the same kernel width as a Gaussian with larger sigma. The resulting kernel width is comparable to the simulated one. It must in fact be noted that model 1 is relatively robust and it was observed that different combinations of parameters lead to a scatter distribution that was capable of completely removing the modulator pattern. These findings suggest that, for this exact phantom, the cost function does not have a unique minimum.

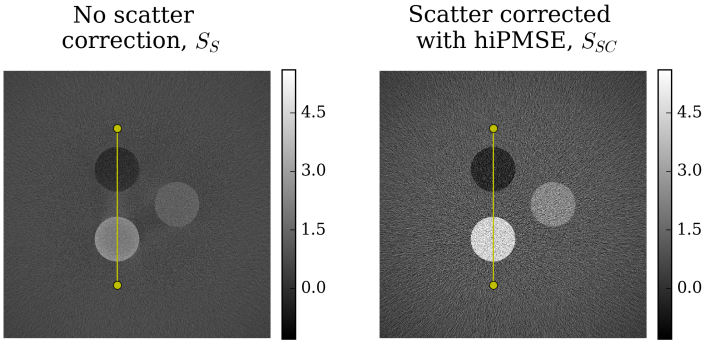


Figure 4.30: Water phantom - hiPMSE: A scatter affected and a scatter corrected slice of the reconstructed volume.

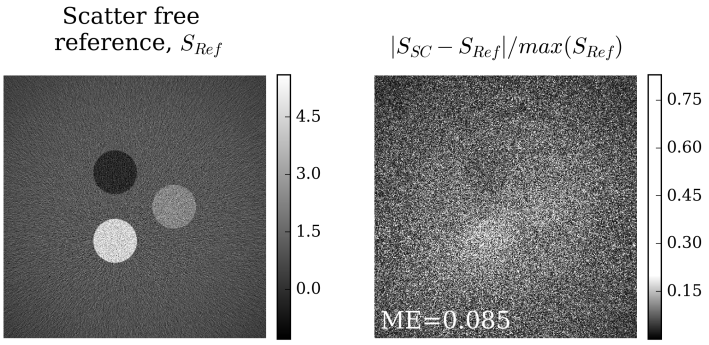


Figure 4.31: Water phantom - hiPMSE: Left: Scatter free reference. Right: Difference of scatter-corrected and scatter-free images in relation to the maximum of the scatter free reference. The mean error (ME) is defined as $ME = mean\left(|\frac{S_{SC}-S_{Ref}}{max(S_{Ref})}|\right)$*, where mean() computes the average value over all pixels of an image.*

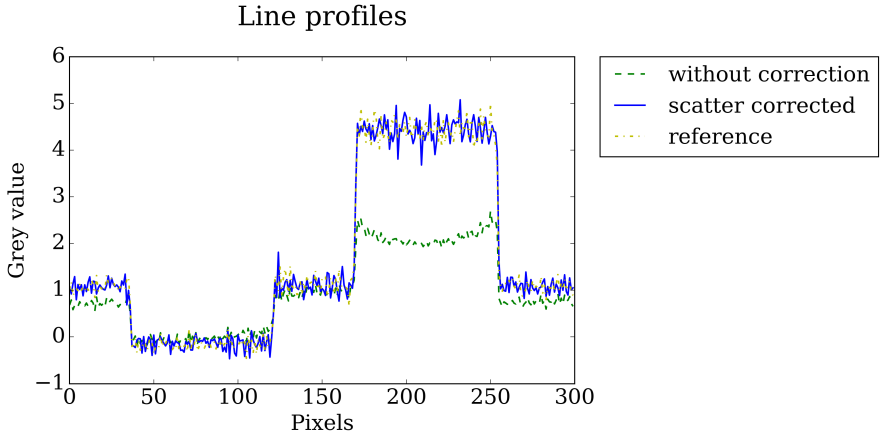


Figure 4.32: Water phantom - hiPMSE: Line profiles through the scatter-affected, the scatter-corrected and the reference slice of the three respective reconstructions.

Table 4.3: Mean estimated parameters of $\vec{c}*$ *for the water phantom compared to reference values used for the simulation.*

	c_0	c_1	b	σ
ref	0.009	0.132	0.002	96.477
est	0.009	0.135	25.481	91.272

4.3.2 Measurements

For real measurement data model 2 was used in the hiPMSE method. This is partly to lower computation times, and partly due to non-convergence of the algorithm on certain phantoms. As shown in Section 4.1, this model nevertheless exhibits accuracy comparable or even better than model 1 and is therefore well suited to estimate scatter.

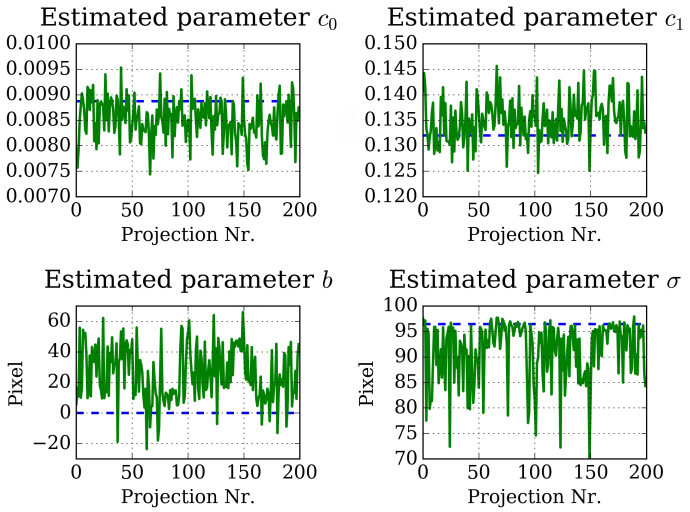


Figure 4.33: Water phantom - hiPMSE: Estimated parameters (solid line) over projection number compared to simulated parameters (dashed line).

Water cylinder

The measurement data used here is the same that served as input for the iPMSE method in Section 4.2.2. Please refer to Figure 4.20 for projection images of the object with the modulator in place and a projection of the modulator. The water cylinder had approximately the same diameter and was placed in the same geometry setup in both simulation and real measurements.

Estimated scatter and primary signals can be seen in Figure 4.34. There are no modulator remains visible in $\boldsymbol{I}_P^{\text{est}}$. In contrast to the classical iPMSE method the resulting scatter image $\boldsymbol{I}_S^{\text{est}}$ is flat and does not show this patchy appearance. Figure 4.35 shows the estimated parameter of the scatter kernel σ_{est} over all projections. It is not constant due to the fact that the water cylinder was placed slightly off center and wanders around in the measured projection images. This implies that also the scatter distribution changes on a projection-to-projection basis. The mean estimated width of the 2D Gaussian function was $\bar{\sigma}_{\text{est}} = 487$. This is a slightly smaller value than suggested by the MC fitted parameter in section 4.1.1 but still in the same range. It can be noted from Figure 4.36 that the estimated SPR is greater than the SPR in the MC simulation study of the same object (see section 4.1.1). This may be due to the fact that scatter is overestimated towards the middle of the water cylinder. Although this overestimation is not large enough to generate visibly noticeable modulator remains in the projection image, it is large enough to produce a slight capping artifact in the scatter-corrected reconstructions (see Figure 4.38). Slight ring artifacts in the slice images, which stem from visibly not noticeable modulator remains, are visible in Figure 4.37 in the center of the water cylinder and these findings support the theory of slightly overestimated scatter towards the center.

Teeth and implants in small bucket - phantom 1

In lack of a suitable phantom for dental situation imitation a simple phantom was constructed. The phantom consisted of a small bucket slightly larger than the FOV filled with water with

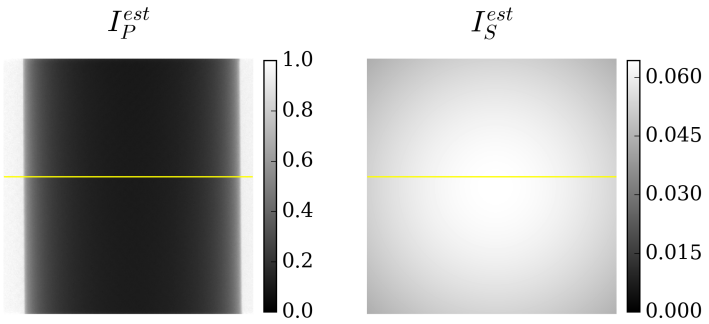


Figure 4.34: Water cylinder (measured) - hiPMSE: Estimated intensities $\boldsymbol{I}_P^{est}$ and $\boldsymbol{I}_S^{est}$.

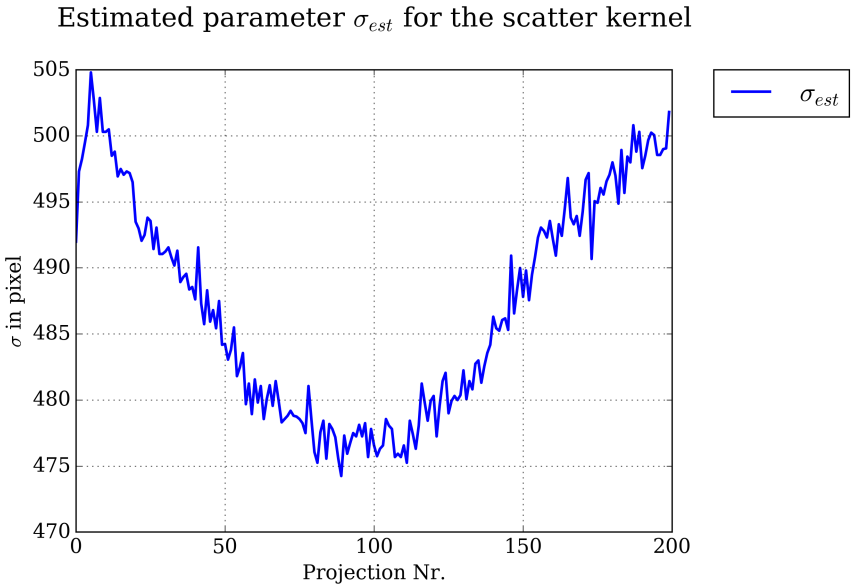


Figure 4.35: Water cylinder (measured) - hiPMSE: Estimated kernel parameter σ_{est}.

three teeth that are embedded in plastic and two titanium rods that are typically used in implantology. In Figure 4.39 the measured projection image $\boldsymbol{I}_M$ with the modulator in place and the projection image of the modulator $\boldsymbol{M}$ are shown. In Figure 4.40 the estimated primary and scatter intensities can be seen. Again, the estimated width of the 2D Gaussian function was projection angle dependent due to the changing underlying attenuation situation. The mean estimated width of the Gaussian function was $\bar{\sigma}_{\text{est}} = 630$. The estimated primary signal shows slight modulator remains. Figure 4.41 shows profiles through the scatter and primary estimation and the estimated SPR. As expected, the SPR reaches highest values behind the titanium implants where the scatter signal exceeds the primary signal. This in turn leads to a dark streak connecting the implants in the reconstructions (see Figure 4.42). Slight streak artifacts are also evident between the titanium implants and the teeth and the teeth only. The scatter-corrected slice in Figure 4.42 does not show any of these streaking artifacts due to a lack of scatter in the projection images. This is a prime example of the scatter correction capability of the novel hiPMSE algorithm introduced in this thesis. The reconstructions do however show rings in the middle of the bucket due to the modulator remains visible in the

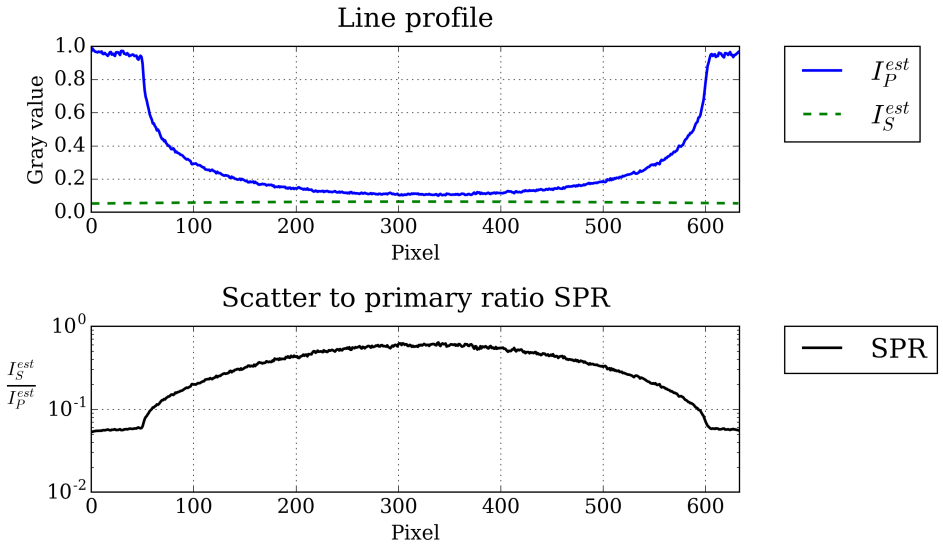


Figure 4.36: Water cylinder (measured) - hiPMSE: Top: Line profiles of estimated primary and scatter intensities. Bottom: Estimated scatter-to-primary ratio.

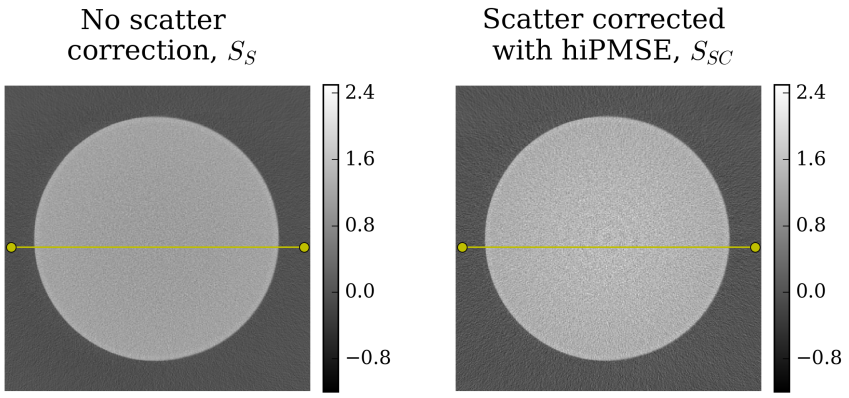


Figure 4.37: Water cylinder (measured) - hiPMSE: A scatter affected and a scatter corrected slice of the reconstructed volume.

corrected projection images. These modulator remains might be of course ascribed to a slightly over- or underestimation of scatter or to modulator motion. This motion can occur within a measurement on a projection-to-projection basis or between measurements. Clearly a modulator image $\boldsymbol{M}$ that does not exactly match the modulator pattern in the measurement $\boldsymbol{I}_M$ will not be able to completely demodulate said measurement after scatter was subtracted. The line profiles visible in Figure 4.43 show that the dark streak vanishes in the scatter-corrected slice. It must also be noted that a slight brightening occurs between the implants.

Teeth and implants in bucket - phantom 2

This measurement is very similar to the phantom in section 4.3.2 except for the considerably more voluminous bucket and a changed arrangement of teeth and implants. The bucket with greater diameter resulted in a overall weaker detected signal and a lower SNR due to more

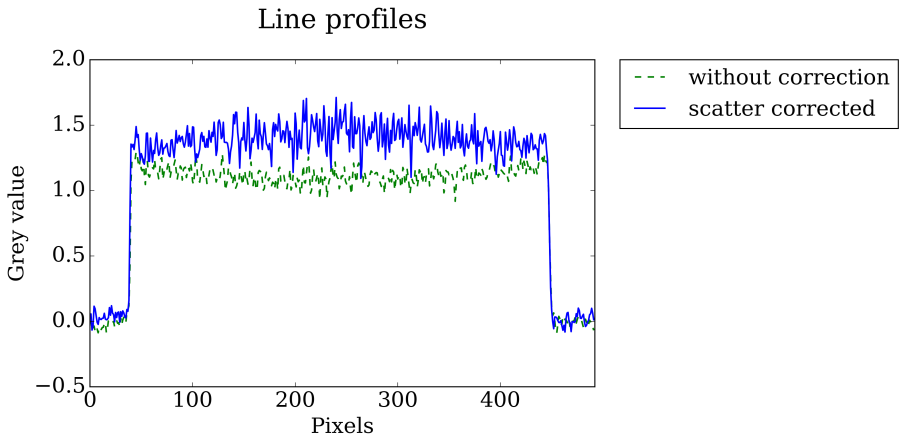


Figure 4.38: Water cylinder (measured) - hiPMSE: Line profiles through the scatter-affected and the scatter-corrected reconstructions.

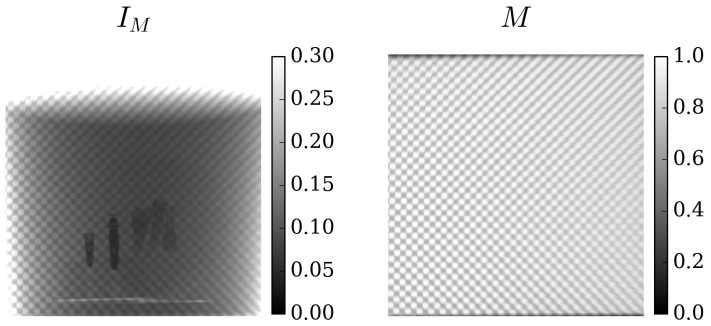


Figure 4.39: Bucket 1 with teeth and implants: Measured intensities $\boldsymbol{I}_M$ and $\boldsymbol{M}$.

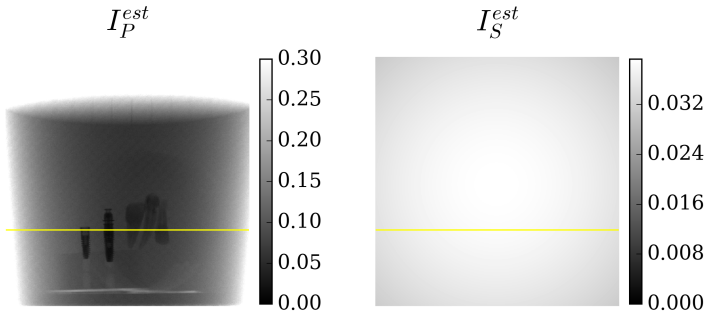


Figure 4.40: Bucket 1 with teeth and implants - hiPMSE:Estimated intensities $\boldsymbol{I}_P^{est}$ and $\boldsymbol{I}_S^{est}$.

attenuation (see Figure 4.44). The estimated scatter and primary images can be seen in Figure 4.45. Very little modulator remains are visible which implies a good scatter estimation. The scatter-corrected projection image $\boldsymbol{I}_P^{\text{est}}$ appears to be much darker as the uncorrected image $\boldsymbol{I}_M$ with identical windowing, which implies a large amount of detected scatter in the uncorrected case. This is also evident from Figure 4.46 where the scatter signal exceeds the primary signal except for the outer image areas. The SPR reaches a maximum of about 11 behind the teeth. Slices of the scatter-corrected and scatter-affected volumes are shown in Figure 4.47. The streak artifacts between the implants vanish. Also, slight ring artifacts due to modulator remains are

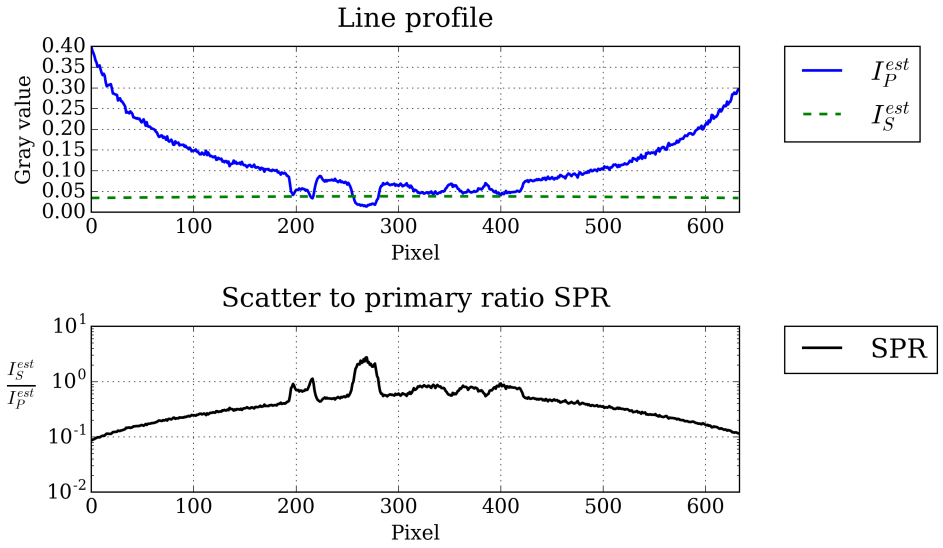


Figure 4.41: Bucket 1 with teeth and implants - hiPMSE: Top: Line profiles of estimated primary and scatter intensities. Bottom: Estimated scatter-to-primary ratio.

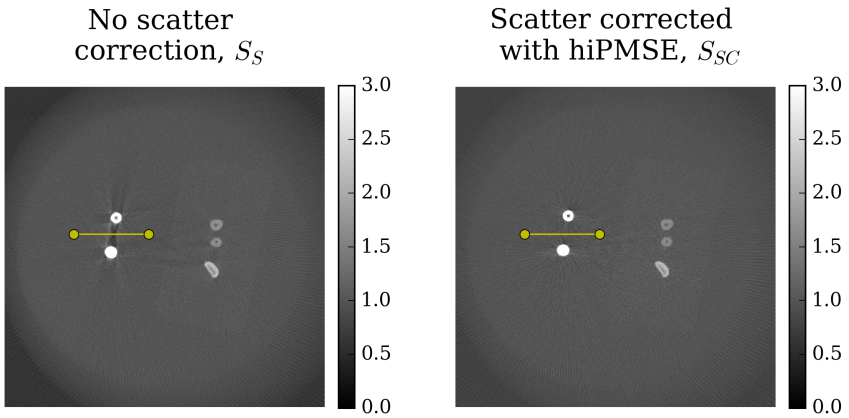


Figure 4.42: Bucket 1 with teeth and implants - hiPMSE: A scatter affected and a scatter corrected slice of the reconstructed volume. A linear windowing function was applied to both images in order to create a visually comparable situation. In both images, the gray value of water has been mapped to 1 and the gray value of teeth to 2.

visible. The mean estimated width of the Gaussian function was $\bar{\sigma}_{\text{est}} = 503$.

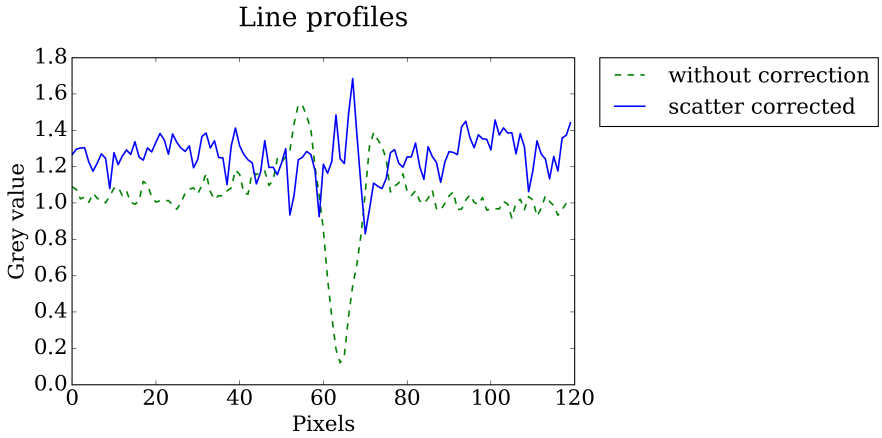


Figure 4.43: Bucket 1 with teeth and implants - hiPMSE: Line profiles through the scatter-affected and the scatter-corrected reconstructions.

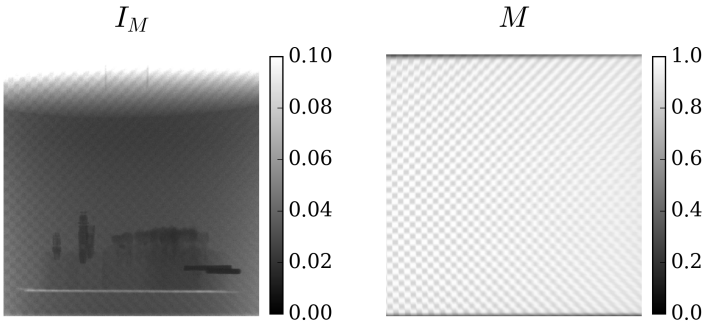


Figure 4.44: Bucket 2 with teeth and implants: Measured intensities $\boldsymbol{I}_M$ and $\boldsymbol{M}$.

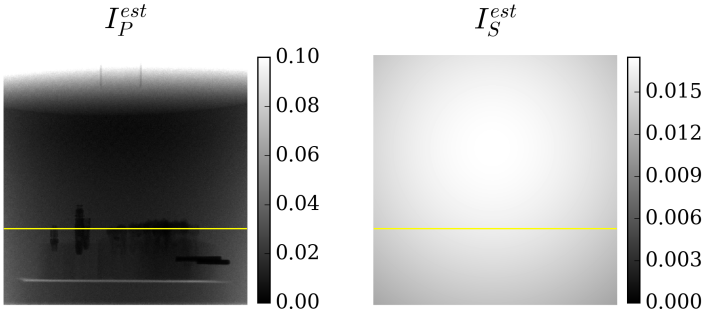


Figure 4.45: Bucket 2 with teeth and implants - hiPMSE: Estimated intensities $\boldsymbol{I}_P^{est}$ and $\boldsymbol{I}_S^{est}$.

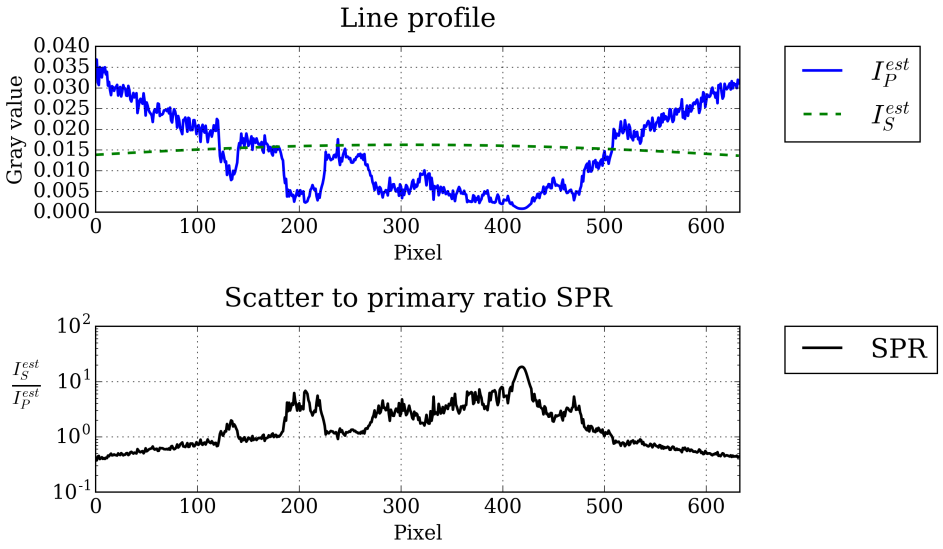


Figure 4.46: Bucket 2 with teeth and implants - hiPMSE: Top: Line profiles of estimated primary and scatter intensities. Bottom: Estimated scatter-to-primary ratio.

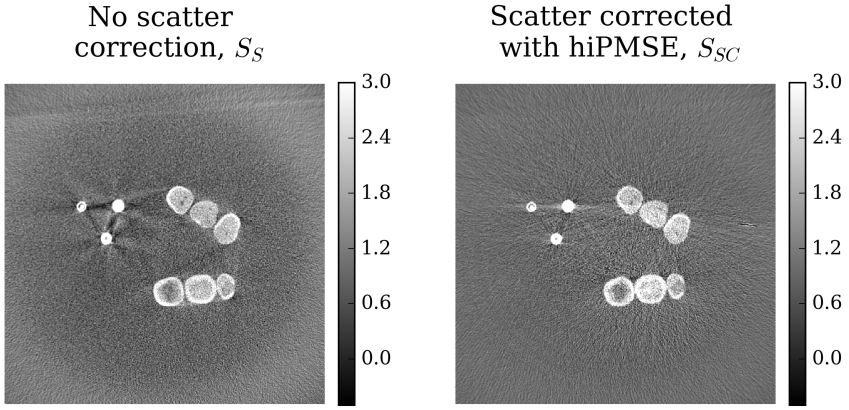


Figure 4.47: Bucket 2 with teeth and implants - hiPMSE: A scatter affected and a scatter corrected slice of the reconstructed volume. A linear windowing function was applied to both images in order to create a visually comparable situation. In both images, the gray value of water has been mapped to 1 and the gray value of teeth to 2. The region-of-interest (ROI) for teeth was set in the uppermost tooth, the water ROI was set directly over the aforementioned tooth and was of the same size.

5 Discussion and Outlook

This last section critically discusses the achieved results, calls the remaining problems and suggests solutions to the aforementioned. Moreover, a short outlook to possible future work is given and other ideas to further improve the methods are presented.

5.1 Summary and Conclusion

The present work deals with the implementation and further development of primary modulator-based scatter estimation techniques, namely the Improved Primary Modulator Scatter Estimation (iPMSE) and the Hybrid Primary Modulator Scatter Estimation (hiPMSE). The performances of the methods was assessed with simulation and real measurement data. The measurements were generated with a laboratory system that is based on a state-of-the-art DVT system.

With regard to the results that were discovered in this thesis, scattered radiation reduction for dental DVT devices is found to be a crucial point for image quality improvements.

5.1.1 Improved Primary Modulator Scatter Estimation (iPMSE)

The iPMSE method worked well under ideal conditions, i.e. simulation data (see Figure 4.17). On real data, problems arose due to modulator blurring. In regions where sharp modulator edges are present, the method worked as expected and was able to find the scatter estimation value that leads to a removal of the modulator pattern. However, with too much blurring, the gradient-based cost function could not find a correct scatter estimation value and overestimated scatter in these regions (see Figure 4.21). This overestimation leads to modulator remains and therefore to ring artifacts in the final volume (see Figure 4.23). It is crucial to completely remove the modulator pattern to really benefit from the achievable scatter correction.

5.1.2 Hybrid Primary Modulator Scatter Estimation (hiPMSE)

The hiPMSE method was first implemented on simulation data. For this, the scatter distribution was generated with the same model with which it was estimated. To further strengthen the statement of scatter correction potency, a MC simulation of a full sinogram with the modulator in place must be carried out. The hiPMSE method's scatter kernel and potential must then be calibrated to this truly unknown scatter distribution. In this way, over-fitting of the model

parameters could be prevented. Nevertheless, the hiPMSE algorithm worked well on both simulated data with the model 1 and on real data with model 2. The discovered results motivate for further investigation of the technique. The lack of patchiness and the inherent smoothness of the obtained scatter distributions are a clear advantage over the classical iPMSE technique. It was shown that the hiPMSE method has potential especially for dental CBCT.

Traditionally, the acquisition of volume data is a necessary step in implantology [5], and these metal implants cause, aside from metal artifacts, severe streak artifacts due to scatter. With the hiPMSE technique, these streaks could be reduced or even completely removed (see Figures 4.42 and 4.47).

Regarding the hiPMSE algorithm itself, several changes could be made in order to further improve the algorithm. First, more sophisticated scatter kernels and potentials could be used to increase scatter estimation accuracy [27][36]. Furthermore, the usage of generic optimization algorithms could be avoided by developing a custom minimizer. This would further speed up the parameter searching procedure.

Also, a combination of the iPMSE and the hiPMSE methods could be realized. The scatter corrected and demodulated primary estimate from an iPMSE run would herein serve as the first primary estimate in the hiPMSE algorithm. As an inital set of parameters in the hiPMSE algorithm, parameters fitted to the iPMSE generated result could be used. This would further improve the rate of convergence.

5.1.3 Computation time

For the iPMSE method computation time amounted to about 5 seconds for one projection and to about 20 minutes for a full sinogram, consisting of 200 projection images on a standard desktop PC with an Intel Core i5 CPU, operating at 2.8 GHz and with 16 GB of RAM for non-optimized Python code, employing out-of-the-box optimization routines. For the hiPMSE method computation time amounted to about 1 to 2 minutes for one projection, and to about 3 to 6 hours for a full sinogram on the same PC. These times are way over clinically and medically acceptable computation times for a scatter correction algorithm. This is partly due to the fact that the Python programming language does not make use of multiple CPU cores inherently. It must however be noted, that these times can be significantly lowered if the code is optimized. Firstly, a compiled programming language like C or C++ could be used. Secondly, since each projection is independent of another, multiple projection images could be processed at the same time. The code can even be parallelized by using modern GPU programming techniques.

5.2 Possible directions for future work

The spatially variable modulator blurring in the projection images leads to wrong scatter estimation and therefore modulator remains in the scatter corrected images. These modulator remains in turn introduce ring artifacts in the reconstructed volumes. Relief could be provided

by redefining the cost function in such that not only edges are utilized as a measure of scatter estimation correctness, but also raw data coverage:

$$C_{TV,DF}(\boldsymbol{I}_S^{est}) = \alpha\|\nabla \boldsymbol{M}^{-1} \cdot \left(\boldsymbol{I}_M - \boldsymbol{I}_S^{est}\right)\|_1 + \frac{\beta}{2}\|(\boldsymbol{M}\boldsymbol{I}_P^{est} + \boldsymbol{I}_S^{est}) - \boldsymbol{I}_M\|_2^2 \tag{5.1}$$

The subscript DF indicates that also a data fidelity term is considered in the cost function. It is conceivable that the parameters α and β are spatially variant and adjusted to the underlying situation. When strong modulator edges are present, the first term of Equation 5.1 is given more weight and in regions where modulator edges are not reliable, the data fidelity term is more accentuated. The data fidelity term could also be used in a refined cost function within the hiPMSE algorithm. Furthermore, boundary constraints could be added such that the primary must always be greater than zero ($I_P > 0$) and also the scatter must fulfill the same condition ($I_S > 0$).

With real measurement data a simplified scatter model 2 was used in the hiPMSE algorithm. This is partly to lower computation times, and partly due to non-convergence of the algorithm on certain phantoms. Future work should assess this problem and expand the hiPMSE method for measurement data to the more general model 1.

It must be accepted, that a perfect estimation of a scatter distribution with no deviation (i.e. no modulator remains) to the true scatter distribution is not easy to achieve in general. From the results presented in this thesis, it is evident that if only slight modulator remains are present in the primary estimates, ring artifacts will be introduced in the reconstructed volumes. Also, the spatially dependent beam hardening of the modulator must be accounted for. Both tasks could be resolved by the Empircal Cupping Correction (ECCP) algorithm [37].

All subtraction-based scatter removal algorithms suffer from increased noise in the corrected projection images. With iPMSE and hiPMSE, a very flat scatter distribution is subtracted from the measured signal. This leads to removal of the overall scatter signal, but not of the scatter noise which is left in the images. To truly benefit from the scatter correction a dedicated noise suppression algorithm must be applied. Zhu et al. suggest such an algorithm in [23].

On some measurements, motion of the modulator on a projection-to-projection basis or between independent measurements could be observed. This is due to the fact that both the detector and the X-ray source rotate around the object. This motion must be corrected in order to achieve error free primary estimates and reconstructions [17]. Based on the findings of this work, a simple registration where only translation is considered, should be sufficient.

The insights of this thesis suggest, that primary modulator based techniques offer an elegant way of scatter estimation. The simultaneous measurement of scattered and primary radiation without increased patient dose or scan time is a clear advantage compared to other methods. With further development and refinement of the algorithms presented above, the image-degrading effect of scatter can be further reduced.

Bibliography

[1] T. M. Buzug, *Computed tomography: from photon statistics to modern cone-beam CT*. Springer Science & Business Media, 2008.

[2] 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*, vol. 38, no. 7, pp. 4296–4311, 2011.

[3] G. Glover, "Compton scatter effects in ct reconstructions," *Medical physics*, vol. 9, no. 6, pp. 860–867, 1982.

[4] J. Wiegert, "Scattered radiation in cone-beam computed tomography: analysis, quantification and compensation". PhD thesis, RWTH Aachen, 2007.

[5] A. Fuhrmann, *Zahnärztliche Radiologie*. Georg Thieme Verlag, 2013.

[6] B. Ohnesorge, T. Flohr, and K. Klingenbeck-Regn, "Efficient object scatter correction algorithm for third and fourth generation ct scanners," *European radiology*, vol. 9, no. 3, pp. 563–569, 1999.

[7] H. Li, R. Mohan, and X. R. Zhu, "Scatter kernel estimation with an edge-spread function method for cone-beam computed tomography imaging," *Physics in medicine and biology*, vol. 53, no. 23, p. 6729, 2008.

[8] J. Star-Lack, M. Sun, A. Kaestner, R. Hassanein, G. Virshup, T. Berkus, and M. Oelhafen, "Efficient scatter correction using asymmetric kernels," in *SPIE Medical Imaging*, pp. 72581Z–72581Z, International Society for Optics and Photonics, 2009.

[9] M. Baer and M. Kachelrieß, "Hybrid scatter correction for ct imaging," *Physics in medicine and biology*, vol. 57, no. 21, p. 6849, 2012.

[10] W. Zhao, S. Brunner, K. Niu, S. Schafer, K. Royalty, and G.-H. Chen, "A patient-specific scatter artifacts correction method," in *SPIE Medical Imaging*, pp. 903310–903310, International Society for Optics and Photonics, 2014.

[11] W. Zbijewski and F. J. Beekman, "Efficient monte carlo based scatter artifact reduction in cone-beam micro-ct," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 7, pp. 817–827, 2006.

[12] A.-P. Colijn and F. J. Beekman, "Accelerated simulation of cone beam x-ray scatter projections," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 5, pp. 584–590, 2004.

[13] R. Ning, X. Tang, and D. L. Conover, "X-ray scatter suppression algorithm for cone-beam volume ct," in *Medical Imaging 2002*, pp. 774–781, International Society for Optics and Photonics, 2002.

[14] E.-P. Rührnschopf and K. Klingenbeck, "A general framework and review of scatter correction methods in cone beam ct. part 2: Scatter estimation approaches," *Medical physics*, vol. 38, no. 9, pp. 5186–5199, 2011.

[15] L. Zhu, N. R. Bennett, and R. Fahrig, "Scatter correction method for x-ray ct using primary modulation: Theory and preliminary results," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 12, pp. 1573–1587, 2006.

[16] L. Ritschl, R. Fahrig, M. Knaup, J. Maier, and M. Kachelrieß, "Robust primary modulation-based scatter estimation for cone-beam ct," *Medical physics*, vol. 42, no. 1, pp. 469–478, 2015.

[17] B. Bier, K. Müller, M. Berger, J. Choi, R. Fahrig, and A. Maier, "Scatter correction for c-arm ct using primary modulation," in *Proc. The fourth international conference on image formation in x-ray computed tomography*, pp. 383–386, 2016.

[18] J. Hsieh, "Computed tomography: principles, design, artifacts, and recent advances," SPIE Bellingham, WA, 2009.

[19] L. Feldkamp, L. Davis, and J. Kress, "Practical cone-beam algorithm," *JOSA A*, vol. 1, no. 6, pp. 612–619, 1984.

[20] H. Turbell, *Cone-beam reconstruction using filtered backprojection*. PhD thesis, Linköping University, 2001.

[21] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 546–552, 1983.

[22] R. Schulze, U. Heil, D. Groβ, D. Bruellmann, E. Dranischnikow, U. Schwanecke, and E. Schoemer, "Artefacts in cbct: a review," *Dentomaxillofacial Radiology*, 2014.

[23] L. Zhu, J. Wang, and L. Xing, "Noise suppression in scatter correction for cone-beam ct," *Medical physics*, vol. 36, no. 3, pp. 741–752, 2009.

[24] V. Klein, "Robust primary modulation-based scatter estimation and correction for x-ray-imaging and computed tomography," undergraduate honors thesis, University of Heidelberg, Department of Physics and Astronomy, 2015.

[25] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[26] J. A. Seibert and J. M. Boone, "X-ray scatter removal by deconvolution," *Medical physics*, vol. 15, no. 4, pp. 567–575, 1988.

[27] M. Sun and J. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Physics in medicine and biology*, vol. 55, no. 22, p. 6695, 2010.

[28] M. Kachelrieß, *Reduktion von Metallartefakten in der Röntgen-Computer-Tomographie*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 1998.

[29] Y. Kyriakou and W. A. Kalender, "X-ray scatter data for flat-panel detector ct," *Physica Medica*, vol. 23, no. 1, pp. 3–15, 2007.

[30] Scipy, "Python-based ecosystem of open-source software for mathematics, science, and engineering." `http://www.scipy.org/`.

[31] R. P. Brent, *Algorithms for minimization without derivatives*. Courier Corporation, 2013.

[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C*, vol. 2. Cambridge university press Cambridge, 1996.

[33] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[34] H. Gao, L. Zhu, and R. Fahrig, "Modulator design for x-ray scatter correction using primary modulation: Material selection," *Medical physics*, vol. 37, no. 8, pp. 4029–4037, 2010.

[35] H. Gao, L. Zhu, and R. Fahrig, "Optimization of system parameters for modulator design in x-ray scatter correction using primary modulation," in *SPIE Medical Imaging*, pp. 76222A–76222A, International Society for Optics and Photonics, 2010.

[36] C. Kim, M. Park, Y. Sung, J. Lee, J. Choi, and S. Cho, "Data consistency-driven scatter kernel optimization for x-ray cone-beam ct," *Physics in medicine and biology*, vol. 60, no. 15, p. 5971, 2015.

[37] R. Grimmer, R. Fahrig, W. Hinshaw, H. Gao, and M. Kachelrieß, "Empirical cupping correction for ct scanners with primary modulation (eccp)," *Medical physics*, vol. 39, no. 2, pp. 825–831, 2012.