Dissertation

submitted to the
Combined Faculties for the Natural Sciences
and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

M.Sc. Joscha Maier

Born in: Stuttgart, Germany

Oral examination: 19.06.2019

Artifact Correction and Real-Time Scatter Estimation for

X-Ray Computed Tomography in Industrial Metrology

Referees: Prof. Dr. Peter Bachert

Prof. Dr. Marc Kachelrieß

Artifact Correction and Real-Time Scatter Estimation for X-Ray Computed Tomography in Industrial Metrology

Artifacts often limit the application of computed tomography (CT) in industrial metrology. In order to correct these artifacts, the so-called simulation-based artifact correction (SBAC) was developed in this thesis. For this purpose, analytical and Monte Carlo (MC) based models were set up to simulate the CT measurement process for a given component as accurately and efficiently as possible. Calculating the difference between this simulation and an ideal one yields an estimate of the present artifacts that can be used to correct the corresponding CT measurement. The potential of this approach was demonstrated for the correction of the most common CT artifacts, i.e. beam hardening, x-ray scattering, off-focal radiation, partial volume effects, and cone-beam artifacts. In any case, the SBAC provided CT reconstructions that showed almost no artifacts and whose quality was clearly superior to state-of-the-art reference approaches. In this context, the problem of long runtimes of scatter simulations was solved by another novel approach, the so-called deep scatter estimation (DSE). The DSE uses a deep convolutional neural network which was trained to map the acquired projection data to given MC scatter estimates. Once the DSE network is trained, it can be used to process unknown data in real-time. In different simulation studies and measurements, it could be shown that DSE generalizes to various acquisition conditions and components while providing scatter distributions that differ by less than 2 % from MC simulations. Thus, the two developed approaches make an important contribution to correct CT artifacts efficiently and to extend the applicability of CT in the field of industrial metrology.

Artefaktkorrektur und Echtzeit Streustrahlschätzung für die Röntgen-Computertomographie in der industriellen Messtechnik

Artefakte schränken häufig die Anwendung der Computertomographie (CT) in der industriellen Messtechnik ein. Um diese zu korrigieren, wurde hier die sogenannte simulationsbasierte Artefaktkorrektuur (SBAC) entwickelt. Dazu wurden analytische und Monte Carlo (MC) basierte Modelle aufgestellt, die den CT Messprozess für ein gegebenes Bauteil möglichst genau und effizient simulieren. Aus der Differenz zwischen dieser und einer idealen Simulation kann anschließend ein Korrekturterm abgeleitet werden, der die Artefakte der entsprechenden CT-Messung korrigiert. Das Potential dieses Verfahrens wurde hier für die Korrektur der häufigsten CT-Artefakte, d.h. Strahlaufhärtung, Streustrahlung, Extrafokalstrahlung, Partialvolumeneffekte und Kegelstrahlartefakte, demonstriert. In jedem Fall lieferte die SBAC CT-Rekonstruktionen, die fast keine Artefakte mehr zeigten und deren Qualität die gängiger Referenzverfahren deutlich übertraf. In diesem Zusammenhang wurde das Problem der langen Laufzeiten von genauen Streustrahlsimulationen durch ein weiteres neu entwickeltes Verfahren, der sogenannten „deep Scatter Estimation" (DSE), gelöst. Dazu verwendet die DSE ein tiefes faltendes neuronales Netz, welches darauf trainiert wurde gemessene Projektionsdaten auf gegebene MC Streustrahlschätzungen abzubilden. Sobald das Training abgeschlossen ist, erlaubt DSE unbekannte Daten in Echtzeit zu prozessieren. In verschiedenen Simulationsstudien und Messreihen konnte gezeigt werden, dass DSE dabei auf verschiedene Aufnahmebedingungen und Bauteile generalisiert und Streustrahlverteilungen liefert, die um weniger als als 2 % von MC Simulationen abweichen. Damit leisten die beiden entwickelten Verfahren einen wichtigen Beitrag zur effizienten Korrektur von CT-Artefakten und zur Erweiterung der Anwendungsmöglichkeiten der CT im Bereich der industriellen Messtechnik.

Contents

List of Acronyms

1D 1-dimensional

2D 2-dimensional

3D 3-dimensional

ANN . . . . Artificial neural network

CAD . . . . Computer-aided design

CBCT . . . . Cone-beam CT

CDF . . . . Cumulative distribution function

CE . . . . Cross entropy

CPU . . . . Central processing unit

CS . . . . Compton scattering

CT . . . . Computed tomography

DCNN . . . . Deep convolutional neural network

DSE . . . . Deep scatter estimation

EADL . . . . Evaluated atomic data library

EPDL . . . . Evaluated photon data library

ESF . . . . Edge-spread function

FBP . . . . Filtered backprojection

FOM . . . . Field of measurement

GPU . . . . Graphics processing unit

HSE . . . . Hybrid scatter estimation

KSE . . . . Kernel-based scatter estimation

LSF . . . . Line-spread function

MAPE . . . . Mean absolute percentage error

MAR . . . . Metal artifact reduction

MC . . . . Monte Carlo

MLCG . . . . Multiplicative linear congruential generator

MLP . . . . Multilayer perceptron

MSE . . . . Mean squared error

NMAR . . . . Normalized metal artifact reduction

PDF . . . . Probability density function

PE . . . . Photoelectric effect

PMMA . . . . Poly methyl methacrylate

ReLU . . . . Rectified linear unit

RS . . . . Rayleigh scattering

SBAC . . . . Simulation-based artifact correction

TV . . . . Total variation

List of Figures

List of Tables

1 | Introduction

Since its introduction by Godfrey N. Hounsfield in 1972 [1], x-ray CT has become an indispensable imaging modality in several fields of application. The first applications of CT, however, were restricted to medical imaging and led to the award of the 1979 Nobel Prize in Physiology or Medicine to Hounsfield and Allan M. Cormack who laid several of the theoretical foundations of CT [2]–[4]. Being able to recover the spatial distribution of the x-ray attenuation coefficient from a set of transmission measurements (i.e. line integrals over the attenuation coefficient), the potential of CT has soon been recognized for other applications apart from medicine. In the early 1980s first adaptions of medical CT started to appear for industrial purpose, mainly in the field of nondestructive material inspection [5]–[7]. The emergence of large area planar detectors and the development of the corresponding cone-beam CT (CBCT) reconstruction algorithms further pushed this trend [8]. With the rising variety and complexity of industrial components and the need for tolerance and geometrical quality control, the use of CT for dimensional metrology started to be investigated in the 1990s [9]–[11]. The development continued until the introduction of the first commercial dimensional CT system in 2005 [12]. This enabled the accurate measurement of hundreds of tolerances of a given workpiece simultaneously. Furthermore, CT allows to assess internal features nondestructively which is a big advantage compared to other modalities such as tactile coordinate measurement machines or optical scanners. Nowadays, dimensional metrology and flaw detection are still the primary applications of industrial CT [12]. While it is used routinely e.g. in plastics industry, the investigation of highly attenuating or multi-material components, however, remains a major challenge. As depicted in figure 1.1, the CT reconstructions of these components are often corrupted by severe artifacts that appear as cupping, shading or dark streaks. The degradation of image quality that comes along with the presence of artifacts impairs or even precludes an appropriate metrological assessment, and therefore, limits the applicability of metrological CT [13]. To overcome this drawback, this thesis aims at developing efficient approaches to correct for these artifacts.

In general, CT artifacts arise whenever the physics of the CT data acquisition is not modeled appropriately within the reconstruction process or if sampling criteria are not met [14]. Analytic reconstruction algorithms that are used most commonly in metrological CT, for instance, assume a linear relationship between the acquired projection data and the intersection length through the sample. However, practically
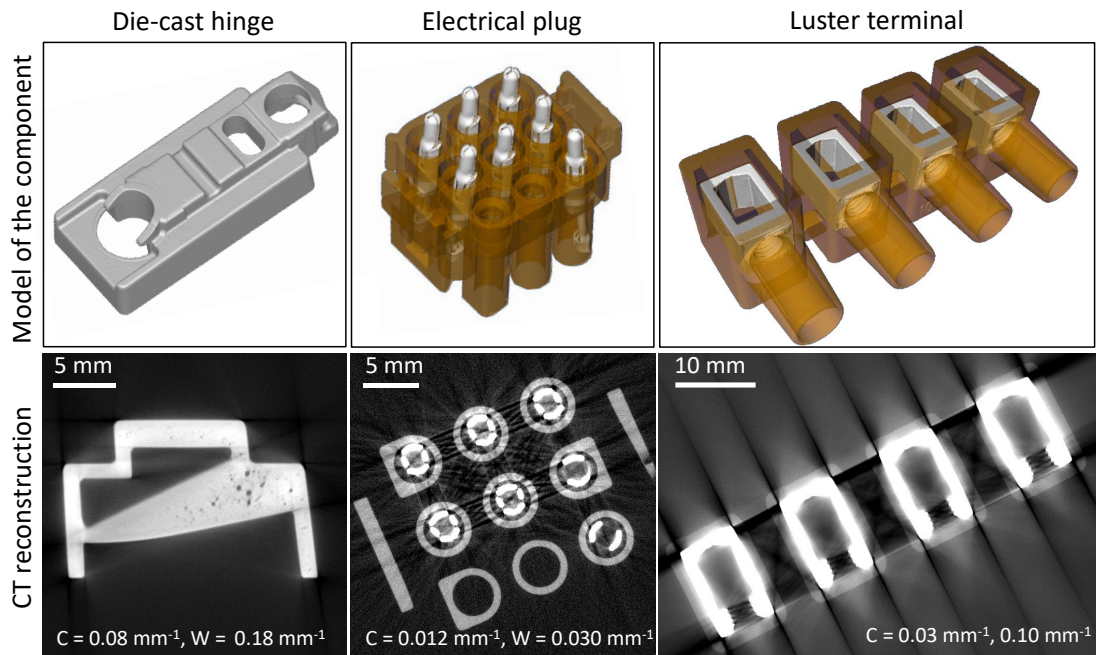
Figure 1.1: Models (top row) and CT reconstructions (bottom row) of typical highly attenuating (left column) and multi-material components (middle and right column). The CT reconstructions clearly show the severe impact of CT artifacts that prevent an appropriate metrological assessment.

there are several effects that lead to a violation of this assumption. Here, beam hardening that is caused by the increasing x-ray attenuation for decreasing x-ray energies, x-ray scattering, partial volume effects and off-focal radiation are the most prominent ones [15]. Another common source of artifacts is the use of circular scan trajectories in combination with a cone-beam setup. Since this trajectory only allows for an exact CT reconstruction within the mid-plane, so-called cone-beam artifacts are introduced, especially in the periphery of the reconstructed volume.

There are several approaches that have been proposed for the correction of CT artifacts [16]–[25]. Most of them can be classified as either iterative, pre- or postcorrection approaches. The former solve the reconstruction problem iteratively while incorporating dedicated models of the interaction between x-rays and matter within the reconstruction process. Pre- and postcorrection approaches, in contrast, apply empirically or physically motivated correction terms prior or post to an analytic reconstruction. However, existing approaches are typically optimized for a certain artifact or can be applied to either single- or multi-material components only. Therefore, a novel approach that applies to arbitrary components and accounts for all of the CT artifacts mentioned above is presented in this thesis. The basic idea of this approach, referred to as simulation-based artifact correction (SBAC), is to correct for artifacts using simulations of the CT data acquisition process. Given a prior model of the component, two simulations

are performed: one simulation that models all the physics leading to artifacts and a second one that fulfills all assumptions of analytic reconstruction algorithms, and thus, represents an ideal CT measurement. The difference of these simulations is an estimate for the present artifacts and can be used subsequently to correct the measurement. Since the performance of the SBAC highly relies on the accuracy of the simulations to reproduce measured data, precise models have been developed or existing models have been refined to fit this purpose. Namely, these models that are described and validated in chapter 3 of this thesis cover the generation of x-rays, the effect and the determination of the focal spot distribution including off-focal radiation, the interaction of x-ray photons within the measured object as well as the x-ray detection process. The application of these models within the framework of the SBAC as well as a dimensional evaluation of the SBAC can be found in chapter 4.

Besides accuracy, computational performance is another important issue to be considered. Typically, an industrial CT acquisition is performed within the order of several minutes [12]. Thus, any correction approach should not exceed this time considerably to be applied within the workflow routinely. Here, x-ray scatter correction turns out to be the main bottleneck. Due to the complexity of x-ray scatter interactions, accurate scatter prediction approaches usually rely on Monte Carlo (MC) methods which are, however, very time consuming and require dedicated prior knowledge [26]. While several more efficient scatter estimation approaches have been proposed, they are by far less accurate than MC methods [27]. Thus, up to date there is always a trade-off between accuracy and computational performance. To combine both of these properties, a novel scatter estimation approach based on a deep convolutional neural network (DCNN) is proposed as a third major development in this thesis. Inspired by the outstanding performance of DCNNs in several areas of CT imaging such as denoising [28]–[32], registration [33], artifact reduction [34]–[36], segmentation [37]–[40] or sparse view CT [41]–[44], the so-called deep scatter estimation (DSE) is presented. The basic idea of DSE is to set up a DCNN that is trained to reproduce MC scatter estimates given only a function of the measured projection data as input. Presenting the DSE network a sufficiently large amount of training samples, i.e. pairs of projection images and MC scatter estimates, it is able to learn a suitable mapping. Once that mapping has been learned, DSE can be applied to unknown projection images and yields scatter estimates with similar accuracy as MC simulations. Since DCNNs can be implemented efficiently on a graphics processing unit (GPU), DSE performs in real-time. An extensive evaluation of DSE for metrological CT, demonstrating its potential as well as its limitations, is given in chapter 5 of this thesis.

Although the focus of this work is set on CT metrology, it has to be noted that none of the approaches presented here is restricted to that field. Since all of them address a general problem of CT imaging, their use might also be beneficial for other applications such as clinical CT or interventional CT, for instance.

2 | Fundamentals

2.1   X-Ray Imaging and CT

X-rays have the ability to traverse objects or patients while being attenuated (absorbed or scattered) with a certain probability. As this probability depends on the material as well as the density distribution, x-rays are used by several imaging modalities to investigate the object's or patient's morphology. A brief overview of the underlying physics and their application to x-ray imaging is given in the following.

2.1.1   X-Ray Generation

Electromagnetic radiation in the energy range of 100 eV up to several MeV is referred to as x-ray radiation [45]. X-rays used in medical and industrial imaging applications are usually generated using reflection or transmission x-ray tubes. The basic working principle of these tubes relies on the generation of free electrons, their acceleration in an electric field and their subsequent deceleration within a metal target. The basic setup is depicted in figure 2.1. Electrons are emitted from a hot cathode (glow filament), which is heated to approximately 2400 K to overcome the binding energy of the electrons in the filament [14]. An acceleration voltage U_A (also referred to as tube voltage), applied between the hot cathode and a ring anode, accelerates the free electrons towards the anode. In that process the electrons gain kinetic energy T that is given by $T = eU_A$, with e being the elementary charge. To focus the electron beam as well as to control its intensity, a grid cap (also known as Wehnelt cylinder) is placed in between the hot cathode and the ring anode. The grid cap is biased to a negative voltage U_C relative to the hot cathode, thus, creating a repulsive electric field that suppresses the emission of electrons from the outer areas of the cathode. The effective size of the electron source can be adjusted by varying the bias voltage. However, as an increasing bias also decreases the beam current, there is always a trade-off between beam quality and beam intensity in practice. On its way to the target, the electron beam passes an additional system of centering and focusing coils. While the centering coils control the incidence of the electron beam on the target, the focusing coils serve as convergent magnetic lens to adjust the focal spot size.

When the fast electrons hit the x-ray target they loose energy by interactions with the target atoms. Here it is convenient to distinguish between collision energy loss and radiative energy loss. The collision energy loss is caused by inelastic Coulomb

collisions with bound atomic electrons leading to excitation and ionization [46], [47]. This may lead to vacancies in inner electronic shells which are subsequently filled by outer shell electrons. The corresponding transition energy is given off as electromagnetic radiation which is also referred to as characteristic radiation. The transitions are termed with the letters K, L, M, ...denoting the final electronic state, and Greek indices $\alpha, \beta, \gamma,\ldots$ denoting the difference between the initial electron shell and the final electron shell. Thus, a transition from L to K shell is termed K_α, a transition from M shell to K shell is termed K_β, and so on.
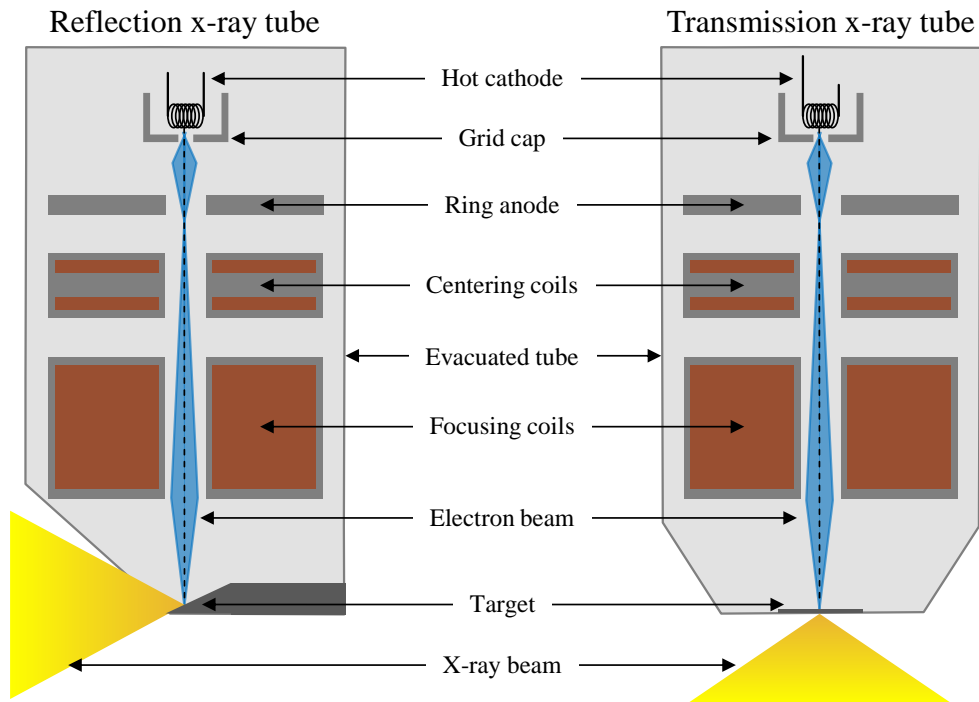
Figure 2.1: Basis setip of reflection (left) and transmission x-ray tubes (right). Here, reflection x-ray tubes can usually be operated with higher tube currents as there is a better heat dissipation. Therefore, they are used whenever a high power is required. Transmission x-ray tubes are used if small focal spot sizes are needed.

In contrast, the radiative energy loss is caused by the deflection of the incident electrons by the Coulomb field of atomic nuclei and is associated with emission of bremsstrahlung [48]. As the initial momentum of the electron is shared between three bodies (the residual electron, the emitted photon, and the atomic nucleus), the emitted bremsstrahlung spectrum is continuous with a maximum energy that equals the kinetic energy of the incident electron $h\nu_{\mathrm{max}} = T$. For electrons, the ratio between collision energy loss per path length and radiative energy loss per path lengths can be

approximated as [49]:

$$\frac{dT/ds|_{\text{rad}}}{dT/ds|_{\text{col}}} = \frac{Z \cdot T}{1400 \cdot m_0 c^2}, \tag{2.1}$$

with T being the electron's kinetic energy, Z being the atomic number and $m_0 c^2 \approx$ 511 keV being the energy of a free electron at rest. Therefore, materials with a high atomic number such as tungsten (Z=74) usually serve as x-ray target. However, according to equation (2.1) only about 1 % of the electron's energy is converted to bremsstrahlung assuming typical tube voltages in the order of 100 kV. The remaining 99 % are absorbed as heat which poses a limitation on the maximum power or the tube current, respectively.

2.1.2 Interaction of X-Rays with Matter

X-rays can interact with matter in several ways. Possible interactions can be summarized systematically as follows [50]:

Type of interaction	*Interaction effect*
1. Interaction with atomic electrons	(a) Absorption
2. Interaction with nucleons	(b) Elastic scattering
3. Interaction with the electric field of the nuclei / electrons	(c) Inelastic scattering
4. Interaction with the meson field of the nucleons	

Theoretically, there are 12 possibilities to combine both columns. However, many of these processes are quite infrequent or have not been observed yet. Considering energy ranges of medical and industrial CT which are typically below 1 MeV, there are only three relevant processes: photoelectric absorption (1a), Rayleigh scattering (1b) and Compton scattering (1c). These effects are commonly described in terms of the linear attenuation coefficient $\mu(E)$ which corresponds to the interaction probability per path lengths P/dl for $dl \ll 1/\mu$. Here, the total attenuation coefficient is simply the sum of the attenuation coefficients of the photoelectric effect μ_{PE}, Rayleigh scattering μ_{RS} and Compton scattering μ_{CS}:

$$\mu(E) = \mu_{\text{PE}}(E) + \mu_{\text{RS}}(E) + \mu_{\text{CS}}(E). \tag{2.2}$$

The contribution of these three effects to the total attenuation coefficient is shown in figure 2.2. At low energies the photoelectric effect typically makes up the main contribution while Compton scattering becomes dominant at higher energies. The latter poses a major problem to CT imaging which will be discussed in more detail in chapter 4.

From a physical point of view, x-ray matter interactions are rather described by the (differential) interaction cross section σ. The cross section is proportional to the interaction probability but is given in units of an area. Thus, it can be interpreted as

effective target area for a certain interaction. Having a sample with N_T target particles, an interaction probability P and an area F that is irradiated homogeneously, σ is given as:

$$\sigma(E) = P(E)\frac{F}{N_T}. \tag{2.3}$$

The relation to the attenuation coefficient is given by:

$$\sigma(E) = \frac{A}{\rho N_A}\mu(E), \tag{2.4}$$

with ρ being the density, N_A being the Avogadro constant and A being the atomic mass. A more detailed description of the photoelectric effect, Rayleigh scattering and Compton scattering as well as the phenomenological Beer-Lambert law that is commonly used in CT to describe x-ray matter interactions is given in the following sections.
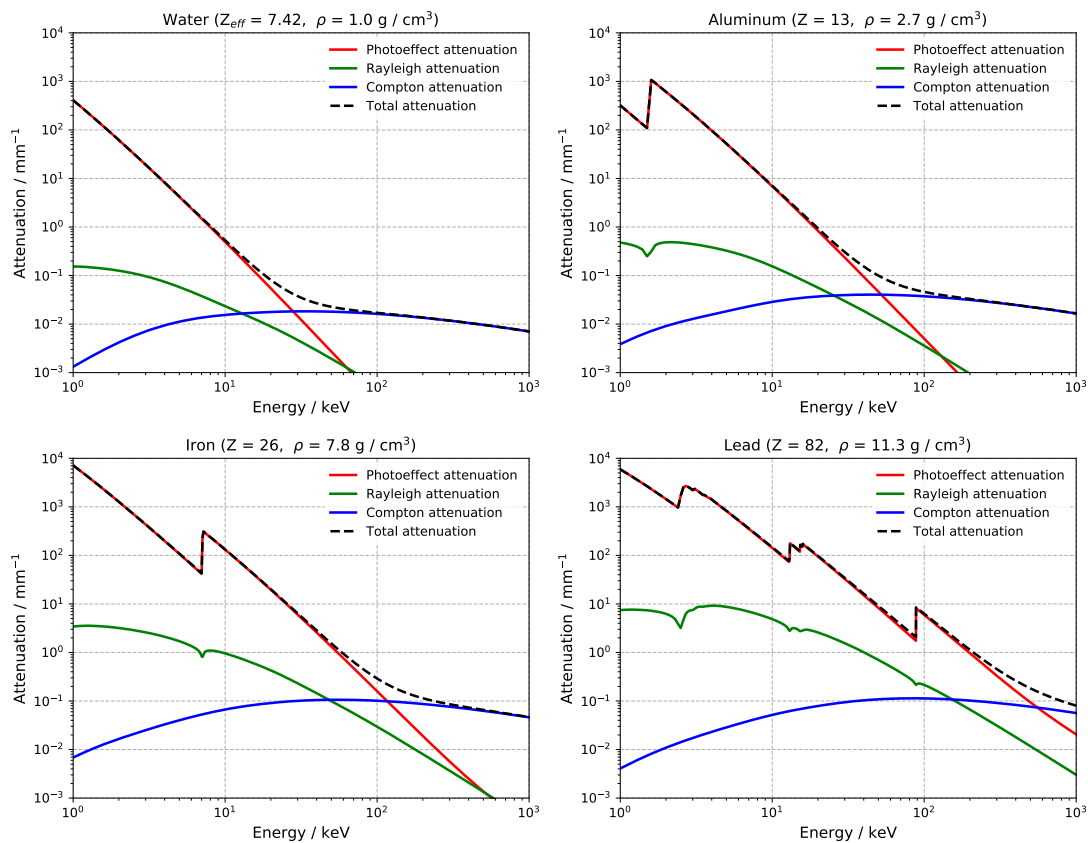
Figure 2.2: Contribution of the photoelectric effect, Rayleigh scattering and Compton scattering to the total attenuation coefficient for water, aluminum, iron and lead at energies between 1 keV and 1000 keV.

i.) Beer-Lambert Law of Attenuation

X-ray matter interactions, in terms of the attenuation of a well-collimated x-ray beam, can be described by the Beer-Lambert law [51]–[53]. In contrast to charged particles, the attenuation shows a truly exponential behavior as x-rays are absorbed or scattered in a single event. The corresponding law can be derived by assuming an infinitesimally thin absorber of thickness dl that is irradiated with a large number of N x-rays. With the interaction probability of the absorber being $\mu\, dl$, the change dN of the number of x-rays N is given by:

$$dN = -N\mu\, dl. \tag{2.5}$$

This is justified by the fact that any interaction leads to the complete removal of the corresponding x-ray from the (collimated) beam. Solving the corresponding differential equation with boundary conditions $N(0) = N_0$ and $N(\infty) = 0$ yields the Beer-Lambert law:

$$N(l) = N_0 \cdot e^{-\mu l}, \tag{2.6}$$

where μ is the linear attenuation coefficient (units of an inverse length). Equation (2.6) describes the attenuation of an homogeneous absorber for a single monochromatic energy. However, x-ray imaging applications usually deal with inhomogeneous objects ($\mu \to \mu(\boldsymbol{r})$) and polychromatic x-ray beams ($N_0 \to N_0(E)$, $\mu \to \mu(E)$). Accordingly, the polychromatic Beer-Lambert law for inhomogeneous objects is given by:

$$N(l, E) = N_0(E) \cdot e^{-\int_0^1 \mu(\boldsymbol{s}+\lambda\cdot\boldsymbol{t}, E)d\lambda}, \tag{2.7}$$

where $\boldsymbol{s}$ and $\boldsymbol{t}$ are the emission point and the directional vector parametrizing the x-ray beam. However, common x-ray detectors do not provide an appropriate energy resolution to measure $N(l, E)$ according to equation (2.7) directly. They rather measure the integral over a certain energy range:

$$I(l) = \int_{E_b}^{\infty} \mathcal{E} \cdot \eta(E) \cdot N_0(E) \cdot e^{-\int \mu(\boldsymbol{s}+\lambda\cdot\boldsymbol{t}, E)d\lambda} dE, \tag{2.8}$$

with $\eta(E)$ being the detector efficiency, i.e. the fraction of x-rays that is absorbed within the detector, and

$$E_b = \begin{cases} \text{Discriminator threshold} & \text{Ideal photon counting detectors} \\ 0 & \text{Energy integrating detectors,} \end{cases} \tag{2.9}$$

and

$$\mathcal{E} = \begin{cases} 1 & \text{Ideal photon counting detectors} \\ E & \text{Energy integrating detectors.} \end{cases} \tag{2.10}$$

As can be seen by equation (2.9) and (2.10), there are currently two different detector technologies used in x-ray imaging which will be discussed in more detail in section 2.1.3.

ii.)   Photoelectric Effect

The transition of atomic electrons between a bound state and a continuum state induced by the absorption of electromagnetic radiation is referred to as photoelectric effect. Considering the interaction between the electron and the photon field as first-order perturbation (which is valid for low photon density fields) implies that the interaction is restricted to one photon and one electron. Thus, the photon is absorbed by an electron in a certain shell b with the corresponding binding energy E_b. Obviously, photoionisation of a given shell requires the photon energy E_γ to exceed the binding energy ($E_\gamma > E_b$) which gives rise to the edges in the photoelectric cross section (see figure 2.2). The remaining energy is transferred to kinetic energy T of the electron:

$$T = E_\gamma - E_b. \tag{2.11}$$

Theoretically, the photoelectric cross section σ_{PE} for a certain electronic shell can be derived from quantum mechanics by calculating the corresponding transition matrix element. It can be shown that the important region for the matrix element is large compared to nuclear size but small compared to the dimension of the atomic cloud. Consequently, the photoelectric interactions occur predominately with K-shell electrons [54]. First attempts to derive an analytic expression for the photoelectric cross section have been performed in 1930 by Stobbe using dipole approximation with exact non-relativistic hydrogen-like electron wave functions for the 1s state [55]:

$$\sigma_{\text{PE},1s}^{\text{Stobbe}} = \left(4\sqrt{2}Z^5\alpha^4\frac{m_0c^2}{E_\gamma}\varphi_0\right)\left[2\pi\left(\frac{\epsilon_K}{E_\gamma}\right)^{\frac{1}{2}}\frac{e^{-4n_1\cot^{-1}n_1}}{1-e^{-2\pi n_1}}\right], \tag{2.12}$$

with Z being the atomic number, α the finestructure constant, m_0 the electron rest mass, c the speed of light, E_γ the energy of the incident photon, φ_0 the Thomson cross-section that holds for the elastic scattering of photons at free charged particles, ϵ_K the K-shell binding energy and $n_1 = \sqrt{\epsilon_K/(E_\gamma - \epsilon_K)}$. Although this result relies on several approximations it yields a roughly correct dependence of the cross section on the atomic number and the energy. For small photon energies close to the absorption edge equation (2.12) yields $\sigma_{\text{PE}} \propto Z^5/E^{\frac{8}{3}}$, while $\sigma_{\text{PE}} \propto Z^5/E^{\frac{7}{2}}$ for high photon energies. With the availability of high speed computers the calculation of more accurate results became possible using numerical methods. An extensive discussion of these models that consider relativistic electron wave functions, all contributing multipoles as well as shielding effects can be found in reference [54]. Today, accurate photoelectric cross sections are available for all shells of the elements $Z = 1 - 100$ and photon energies up to 1000 GeV. In this work, photoelectric cross sections are obtained by interpolation of tabulated values from the evaluated photon data library (EPDL) [56] that are based on theoretical shell cross sections of Scofield and total cross sections of Hubbell [57], [58].

iii.)   Compton Scattering

The inelastic scattering of electromagnetic waves at charged particles, usually electrons, is referred to as Compton scattering. It is named after Arthur H. Compton who first

came up with a theoretical explanation of the wavelength shift $(\lambda' - \lambda)$ of scattered photons (wavelength λ') as a function of the scatter angle ϑ [59]. Assuming inelastic scattering at quasi-free electrons at rest, the corresponding relation can be deduced from energy and momentum conservation as:

$$\lambda' - \lambda = \lambda_c(1 - \cos\vartheta), \tag{2.13}$$

where $\lambda_c = h/(m_0 c)$ is the so called Compton wavelength. Using the relation $E = hc/\lambda$ in equation (2.13), the energy E'_γ of the scattered photon is given by:

$$E'_\gamma = \frac{E_\gamma}{1 + \frac{E_\gamma}{m_0 c^2}(1 - \cos\vartheta)}. \tag{2.14}$$

In case of free electrons at rest, the differential Compton scattering cross section is given by the Klein-Nishina formula [60]:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\bigg|_{\text{Klein-Nishina}} = \frac{1}{2}\frac{\alpha^2}{m_0^2}\left(\frac{E'_\gamma}{E_\gamma}\right)^2\left[\frac{E'_\gamma}{E_\gamma} + \frac{E_\gamma}{E'_\gamma} - \sin^2\vartheta\right], \tag{2.15}$$

where α is the fine-structure constant and E'_γ/E_γ is defined by equation (2.14). Considering the scattering at atomic electrons, which are neither free nor at rest, deviations from equation (2.15) can be observed. These deviations that are caused by the electron binding energy as well as interference effects with bound electrons are commonly accounted for by a so-called incoherent scatter function $S(q, Z)$ that depends on the momentum transfer q and the atomic number Z. Thus, the differential cross section for incoherent scattering at atomic electrons is given by:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = S(q, Z)\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\bigg|_{\text{Klein-Nishina}}. \tag{2.16}$$

A detailed description of the evaluation of the incoherent scatter function and a comparison to measurements is given in reference [61]. In this work, the differential as well as total Compton scatter cross sections are obtained by interpolation of tabulated values from the EPDL [56] that are calculated according to equation (2.16) using the incoherent scatter function of Hubbell et al. [61], [62].

iv.) Rayleigh Scattering

Rayleigh scattering refers to the elastic scattering of photons at bound atomic electrons. Therefore, Rayleigh scattering only leads to a change of the initial flight direction while the energy of the scattered photon remains unchanged. The atomic Rayleigh scattering differential cross section can be calculated using non-relativistic perturbation theory. For energies much higher than the K-shell binding energy it is given as [63]:

$$\frac{d\sigma}{d\Omega} = |F(q, Z)|^2\frac{d\sigma}{d\Omega}\bigg|_{\text{Thomson}}, \tag{2.17}$$

where

$$\frac{d\sigma}{d\Omega}_{\text{Thomson}} = \frac{1}{2}r_e^2(1 + \cos^2\vartheta) \tag{2.18}$$

is the Thomson differential cross section that describes scattering at free charged particles by classical electromagnetism and $F(q, Z)$ is the atomic form factor that corresponds to the Fourier transform of the electron density distribution with q being the magnitude of the momentum transfer and Z being the atomic number. For energies in the range of the K-edge binding energy, equation (2.17) is commonly modified as follows [64]:

$$\frac{d\sigma}{d\Omega} = |F(q, Z) + f' + i \cdot f''|^2 \frac{d\sigma}{d\Omega}_{\text{Thomson}}, \tag{2.19}$$

where the complex term $f' + i \cdot f''$ is the so-called anomalous scatter factor that is associated with scattering via absorption and reemission. In this work, the differential cross sections for Rayleigh scattering are obtained by interpolation of tabulated values from the EPDL [56]. The EPDL cross sections are based on equation (2.19) using the non-relativistic form factors of Hubbell [61] and the anomalous scattering factors of Cullen [65].

2.1.3 Detection of X-Rays

The detection of x-rays relies on their attenuation within the detector material and the subsequent conversion of the x-ray energy to an electrical signal that can be further processed by the detector's electronics. Here it has to be distinguished between indirect conversion and direct conversion as depicted in figure 2.3. The two different detector technologies are briefly described in the following.

Indirect Converting Detectors Indirect conversion refers to a two-step process in which x-rays are converted to optical photons that are subsequently detected by photomultiplier or a photo diode [66]. In medical and industrial CT, the corresponding detectors are usually based on inorganic scintillators. Here, the scintillator has a band gap between the valence band and the conduction band of energy E_{gap} [67]. Incident x-rays lead to multi-step interactions with the lattice of the scintillator material. In a first step, incident x-rays transfer energy to bound electrons by the photoelectric effect or Compton scattering as described in section 2.1.2. Subsequently, these electrons can transfer their kinetic energy by elevating further electrons from the valence band to the conduction band, leaving holes in the valence band. The whole process requires less than 1 ps [66]. The corresponding number of electron-hole pairs can be approximated as [67]:

$$N_{\text{ehp}} = \frac{E_\gamma}{E_{\text{ehp}}} = \frac{E_\gamma}{\beta \cdot E_{\text{gap}}}, \tag{2.20}$$

where E_γ is the energy of the incident x-ray and E_{ehp} is the energy that is required to generate an electron-hole pair. This energy is typically a factor $\beta \approx 3 - 7$ higher
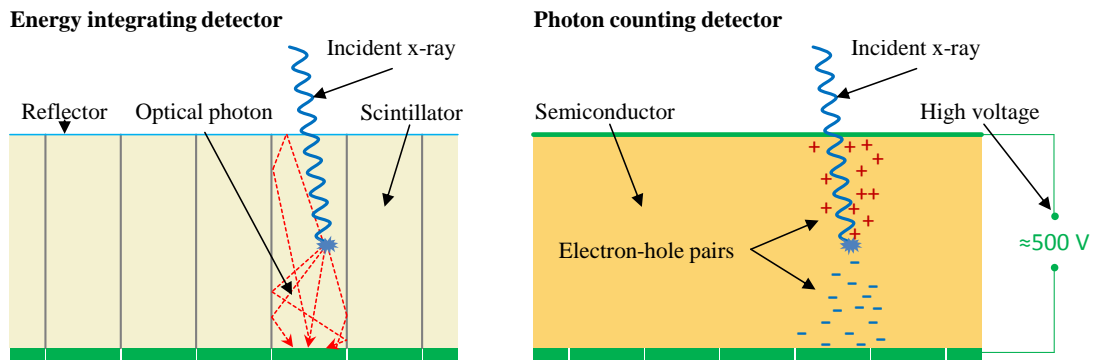
Figure 2.3: Layout of a conventional scintillator-based energy integrating detector (indirect conversion) and a semiconductor-based photon counting detector (direct conversion).

than the energy of the band gap as there is an energy loss e.g. due to lattice vibrations [68]. Considering the pure crystal, the return of electrons from the conduction band to the valence band by the emission of photons would be an inefficient process. Another drawback would be the typical band gap energies which resulted in photons outside the optical spectrum when electrons deexcite to the ground state. To increase the probability of optical photon emission, impurities are commonly added to the scintillator crystal. These impurities, i.e. dopants like Eu, Ti or Tb create special sites within the lattice (luminescence centers) that distort the regular band structure. Thus, compared to the pure crystal, there are additional energy levels within the band gap. Once an electron-hole pair is created, the positive hole drifts to the location of the luminescence center and ionizes it as the ionization energy is lower than the one of a lattice site. The corresponding electron moves freely in the conduction band until it encounters an ionized luminescence center. If the dopant is appropriately chosen, the subsequent recombination leads to the emission of optical photons. Using the formula of equation (2.20), their number can be estimated as [67]:

$$N_{\text{ph}} = N_{\text{ehp}} \cdot S \cdot Q = \frac{E_\gamma}{\beta \cdot E_{\text{gap}}} \cdot S \cdot Q, \tag{2.21}$$

with S being the transport efficiency of electron-hole pairs and Q being the quantum yield of the luminiscence process. Typically, $10^4 - 10^5$ photons are created per MeV which are then converted to an electrical signal by the photo diode [66]. As the half-life time of the excited states is in the order of 50-500 ns, indirect converting detectors are not able to register single x-rays assuming typical fluxes in CT [69]. The measured signal corresponds to the integrated charge over the readout time. Since the charge is proportional to the energy, these detectors are also referred to as energy integrating detectors.

Direct Converting Detectors Direct converting detectors are based on semiconductor sensors that are operated to measure the electron-hole pairs, generated by incident

x-rays. Here, the first steps of the detection process until the formation of electron-hole pairs are as described in the previous section. To detect the electron-hole pairs, the semiconductor sensor is equipped with electrodes on its surface. Applying a bias voltage, typically in the order of a few hundred volts, allows to accumulate the charge $Q = eN_{\mathrm{ehp}}$ in a capacitor with capacity C. Subsequently, the capacitor can be discharged over a resistance R which yields a time-dependent voltage pulse:

$$U(t) = U_0\, e^{-t/(RC)}, \tag{2.22}$$

where $U_0 = Q_0/C$ corresponds to the total collected charge Q_0. The pulse train $U(t)$ is then processed by a pulse-shaping amplifier followed by a pulse height analysis. In that process, the incoming signal $U(t)$ is compared to a reference voltage. That voltage can be related to a certain energy threshold E_b by an appropriate calibration. Once the signal exceeds the reference voltage or the energy threshold, respectively, the event is counted as an incident photon with energy $E_\gamma > E_b$. Since the readout process is much faster than for energy integrating detectors, single photons can be counted. Furthermore, this detector technology allows to sort incident x-rays by their energy while suppressing electronic noise by an appropriate choice of the threshold.

2.1.4 Imaging

X-ray imaging applications make use of the material specificity of x-ray matter interactions. X-rays that traverse different materials are attenuated differently which allows to get insight in the object's morphology. Here, the acquisition of x-ray projection images corresponds to the measurement of line integrals over the attenuation coefficient (see equation (2.7)). Thus, a single projection cannot resolve the morphology of an object in beam-direction but only to some extend in lateral direction. While this information is sufficient in several applications such as conventional radiography, 3-dimensional (3D) information can be reconstructed by acquiring several projections over an angular range of at least 180°. The principle setup as well as the mathematical basics of CT image reconstructions are briefly explained in the following.

i.) CT Setup and Data Acquisition

The basic setup of a CT system consists of an x-ray source and an opposed detector measuring the x-rays that traversed the scanned object without interaction. To acquire x-ray projection images from at least 180°, the source and the detector are either mounted on a rotating gantry or the object itself is rotated. The first CT system developed by Sir Godfrey N. Hounsfield had only two detector elements aligned in z-direction [70]. As only two rays could be acquired simultaneously, the source and the detector had to be moved step by step across the object to derive a complete projection for a given view angle. Thus, the data are acquired in parallel beam geometry and can be reconstructed as described in the following section. Later generations of CT systems used a fan beam geometry with multiple laterally aligned detector elements to acquire a complete projection in one shot. Today's CT systems typically acquire several

z-slices at once using a cone-beam geometry and detectors with up to several thousands rows and columns. While clinical CT systems use cylindrical detectors, industrial CT systems are usually equipped with flat detectors. Each detector element measures the x-ray intensity I which is assumed to follow the monochromatic Beer-Lambert law as given by equation (2.6). This intensity is normalized to a flat field image I_0, i.e. an acquisition without object to calculate projection values p as:

$$p = -\ln\frac{I}{I_0} = \int \mu(\boldsymbol{s} + \lambda \cdot \boldsymbol{t})d\lambda, \tag{2.23}$$

which evaluates to $p = \mu \cdot L$ in case of a homogeneous object with intersection length L. Thus, for ideal data, there is a linear relationship between the projection value and the line integral over the attenuation coefficient.

ii.)   CT Reconstruction

The mathematical basics of 2D CT reconstruction are based on the work of Johann Radon who showed that a continuous function $f(x,y)$ with $\int\int dx\,dy\,\frac{f(x,y)}{\sqrt{x^2+y^2}} < \infty$ can be recovered from an infinite set of line integrals along straight lines in the 2-dimensional (2D) Euclidean space [71]. This section briefly discusses the application to CT imaging but is, for the sake of simplicity, restricted to the most simple case of 2D reconstruction in parallel beam geometry.

The CT Reconstruction Problem in 2D Parallel Beam Geometry   In 2D parallel beam geometry rays are usually parametrized as shown in figure 2.4. Here, ϑ describes the angle of the ray with respect to the x-axis and ξ denotes the distance of the ray from the isocenter which is assumed to be the origin of the coordinate system. Thus, any ray is described by the following equation of a line:

$$x\cos\vartheta + y\sin\vartheta - \xi = \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\vartheta - \xi = 0, \tag{2.24}$$

with $\hat{\boldsymbol{n}}_\vartheta = (\cos\vartheta, \sin\vartheta)^T$ being the unit vector normal to the projection direction. Using equation 2.24, the line integrals given by equation (2.23) can be expressed as:

$$\begin{aligned}
p(\vartheta, \xi) &= \int \mu(\boldsymbol{s} + \lambda \cdot \boldsymbol{t})d\lambda \\
&= \int \mu(x,y)\delta(x\cos\vartheta + y\sin\vartheta - \xi)dxdy \\
&= \int \mu(\boldsymbol{r})\delta(\boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\vartheta - \xi)d^2r \\
&= \mathcal{R}\mu(\boldsymbol{r}),
\end{aligned} \tag{2.25}$$

and is also referred to as Radon transform (or x-ray transform in case of 3D data) with $\mathcal{R}$ being the radon transform operator. Given the line integrals over an angular range of at least 180°, the CT reconstruction aims at recovering $\mu(\boldsymbol{r})$ by inverting equation (2.25), i.e. $\mu = \mathcal{R}^{-1}p$.
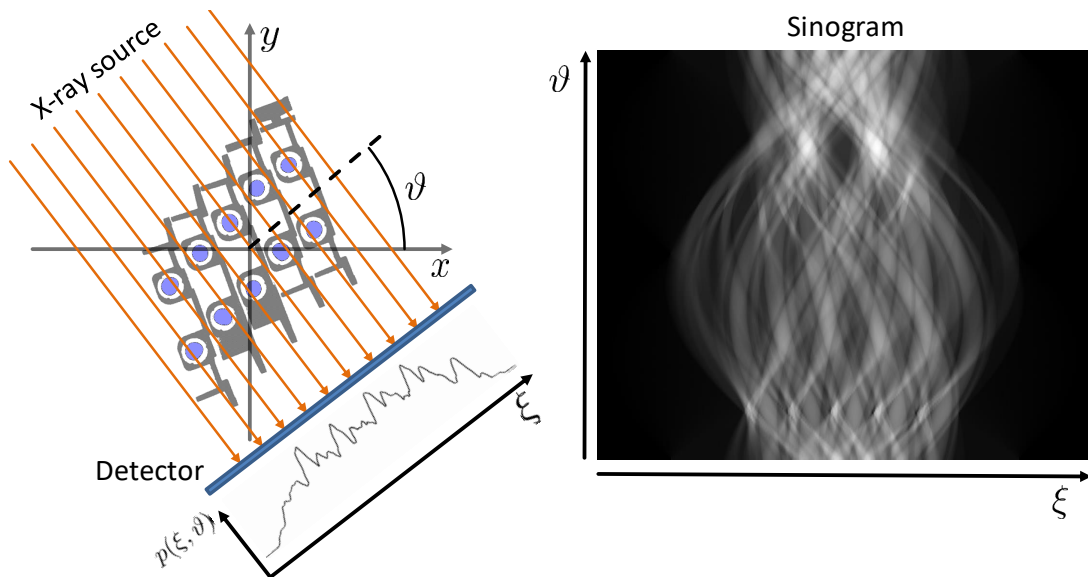
Figure 2.4: Left: Schematic of an ideal 2D pencil beam geometry. Parallel rays, emitted by the x-ray source, are described by their distance ξ to the origin and the angle ϑ between the x-axis. Right: Corresponding sinogram. Each pixel of the sinogram corresponds to a certain ξ and ϑ and represents the value of the line integral given in equation (2.23).

Backprojection Backprojection refers to the "smearing" of the projection values back to image domain. For a given angle ϑ, the image $f(\boldsymbol{r})$ is assigned the projection value that corresponds to the ray through $\boldsymbol{r}$:

$$f_\vartheta(\boldsymbol{r}) = p(\vartheta, \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\vartheta) \tag{2.26}$$

Performing the backprojection for all angles yields:

$$\begin{aligned} f(\boldsymbol{r}) &= \int_0^\pi p(\vartheta, \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\vartheta) d\vartheta \\ &= \mathcal{B}p(\vartheta, \xi), \end{aligned} \tag{2.27}$$

where $\mathcal{B}$ is the backprojection operator. It can be easily shown that the application of the backpropagation operator does not yield the desired solution to the reconstruction problem, i.e. $\mathcal{B} \neq \mathcal{R}^{-1}$.

Substituting equation (2.25) into equation (2.27) yields:

$$
\begin{aligned}
f(\boldsymbol{r}) &= \int_0^\pi \int \mu(\boldsymbol{r}')\delta((\boldsymbol{r}-\boldsymbol{r}')\hat{\boldsymbol{n}}_\vartheta)d^2r'd\vartheta \\
&= \int \mu(\boldsymbol{r}') \left[\int_0^\pi \delta((\boldsymbol{r}-\boldsymbol{r}')\hat{\boldsymbol{n}}_\vartheta)d\vartheta \right] d^2r' \\
&= \int \mu(\boldsymbol{r}') \left[\int_0^\pi \delta(|\boldsymbol{r}-\boldsymbol{r}'|\cos(\varphi-\vartheta)d\vartheta \right] d^2r' \\
&= \int \mu(\boldsymbol{r}') \left[\int_0^\pi \frac{\delta(\vartheta-\vartheta_0)}{|\boldsymbol{r}-\boldsymbol{r}'|\sin(\pm\pi/2)}d\vartheta \right] d^2r' \\
&= \int \mu(\boldsymbol{r}')\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}d^2r' \\
&= \mu(\boldsymbol{r}) * h(\boldsymbol{r}).
\end{aligned}
\tag{2.28}
$$

In step 2 the relation $\boldsymbol{a}\cdot\boldsymbol{b} = |\boldsymbol{a}|\cdot|\boldsymbol{b}|\cos\angle_a^b$ was used with φ being the angle between $|\boldsymbol{r}-\boldsymbol{r}'|$ and the x-axis. In step 3, the relation $\delta(g(x)) = \sum_i |g'(x_i)|^{-1}\delta(x-x_i)$ was used, with $g'(x)$ being the derivative of $g(x)$ and x_i being the simple zero points of $g(x)$. The argument of the δ-distribution has exactly one zero point ϑ_0 within the interval $0 \le \vartheta < \pi$: $\vartheta_0 = \varphi + \pi/2$ or $\vartheta_0 = \varphi - \pi/2$ which evaluates to the expression given in the fourth line.

As shown by equation (2.28), the backprojection of the projection data p does not yield the distribution of the attenuation coefficient $\mu(\boldsymbol{r})$ but its convolution with the point spread function $h(\boldsymbol{r}) = 1/|\boldsymbol{r}|$.

Filtered Layergram: Deconvolution of the Backprojection One possibility to recover $\mu(\boldsymbol{r})$ from $f(\boldsymbol{r})$ as given in equation (2.28) is to perform a deconvolution. Applying the convolution theorem which states that a convolution in spatial domain corresponds to a multiplication in Fourier domain yields:

$$
\mathcal{F}_2 f(\boldsymbol{r}) = \mathcal{F}_2\mathcal{B}p(\vartheta,\xi) = \mathcal{F}_2(\mu(\boldsymbol{r}) * h(\boldsymbol{r})) = \mathcal{F}_2\mu(\boldsymbol{r})\cdot\mathcal{F}_2h(\boldsymbol{r}),
\tag{2.29}
$$

where $\mathcal{F}_2$ is the 2D Fourier transform operator. With the 2D Fourier transform of the point spread function $h(\boldsymbol{r})$

$$
\mathcal{F}_2 h(\boldsymbol{r}) = \mathcal{F}_2|\boldsymbol{r}|^{-1} = |\boldsymbol{k}|^{-1},
\tag{2.30}
$$

equation (2.29) can be rearranged to:

$$
\mathcal{F}_2\mu(\boldsymbol{r}) = |\boldsymbol{k}|\mathcal{F}_2\mathcal{B}p(\vartheta,\xi).
\tag{2.31}
$$

Applying the inverse 2D Fourier yields $\mu(\boldsymbol{r})$ as:

$$
\mu(\boldsymbol{r}) = \mathcal{F}_2^{-1}|\boldsymbol{k}|\mathcal{F}_2\mathcal{B}p(\vartheta,\xi)
\tag{2.32}
$$

Thus, a solution to the reconstruction problem is to perform a backprojection of the projection data p followed by a filtering in Fourier domain with a 2D ramp function $|\boldsymbol{k}|$

and an inverse Fourier transform to get back to image domain. This method, which is also known as filtered layergram, has the drawback of requiring all projections before Fourier domain filtering can be started. Thus, practical applications are usually based on the so-called filtered backprojection (FBP) that applies a filtering operation to every acquired projection prior to the backprojection. The FBP algorithm can be derived from equation 2.32 using the central slice theorem.

Central Slice Theorem   The central slice theorem relates the 1-dimensional (1D) Fourier transform of the projection data $p(\vartheta, \xi)$ to the 2D Fourier transform of the distribution $\mu(\boldsymbol{r})$. The 1D Fourier transform of the projection data with respect to ξ is given as:

$$
\begin{aligned}
P_\vartheta(k) &= \mathcal{F}_1\{p(\vartheta, \xi)\}_\xi \\
&= \int \left[\int \mu(\boldsymbol{r})\delta(\boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\vartheta - \xi)d^2r\right] e^{-2\pi i k\xi}d\xi \\
&= \int \mu(\boldsymbol{r}) \left[\int \delta(\boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\vartheta - \xi)e^{-2\pi i k\xi}d\xi\right] d^2r \\
&= \int \mu(\boldsymbol{r})e^{-2\pi i k(\boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\vartheta)}d^2r
\end{aligned}
\tag{2.33}
$$

Comparing equation (2.33) to the 2D Fourier transform of $\mu(\boldsymbol{r})$

$$
\begin{aligned}
M(\boldsymbol{k}) &= \mathcal{F}_2\{\mu(\boldsymbol{r})\} \\
&= \int \mu(\boldsymbol{r})e^{-2\pi i(\boldsymbol{r} \cdot \boldsymbol{k})}d^2r
\end{aligned}
\tag{2.34}
$$

shows:

$$
P_\vartheta(k) = M(k\hat{\boldsymbol{n}}_\vartheta).
\tag{2.35}
$$

Equation (2.35) is referred to as Fourier slice theorem. It states that the 1D Fourier transform of a projection corresponds to a straight line l through the origin of 2D Fourier space of $\mu(\boldsymbol{r})$ that is given as $l : \boldsymbol{k} \cdot \hat{\boldsymbol{n}}_\vartheta = 0$. Considering all angles, it equals the 2D Fourier transform of $\mu(\boldsymbol{r})$ in polar coordinates. Using the operator notation yields:

$$
\mathcal{F}_1\mathcal{R} = \mathcal{F}_2.
\tag{2.36}
$$

Applying simple operator algebra, equation (2.36) can be rearranged to $\mathcal{R}^{-1} = \mathcal{F}_2^{-1}\mathcal{F}_1$, which is obviously a solution to the reconstruction problem and referred to as direct Fourier reconstruction. However, as measured data are sampled on a discrete polar grid a resampling to a Cartesian grid is required before applying the inverse Fourier transform. As this resampling includes a complex interpolation, this approach is usually not the method of choice in practice.

Filtered Backprojection (FBP) Starting with equation (2.32), the FBP can be derived by replacing the 2D Fourier transform as well as their inverse according to the central slice theorem (see equation (2.36)):

$$\mu(\boldsymbol{r}) = \mathcal{F}_2^{-1} |\boldsymbol{k}| \mathcal{F}_2 \mathcal{B} p(\vartheta, \xi)$$
$$= \mathcal{R}^{-1} \mathcal{F}_1^{-1} |\boldsymbol{k}| \mathcal{F}_1 \mathcal{R} \mathcal{B} p(\vartheta, \xi),$$

with $|\boldsymbol{k}| = |k \cdot \hat{\boldsymbol{n}}_\vartheta| = |k|$ this can be rearranged to:

$$\mu(\boldsymbol{r}) = \mathcal{B} \mathcal{F}_1^{-1} |k| \mathcal{F}_1 p(\vartheta, \xi)$$
$$= \mathcal{B} \mathcal{F}_1^{-1} |k| P_\vartheta(k)$$
$$= \mathcal{B}(\mathcal{F}_1^{-1} |k|) * p(\vartheta, \xi). \tag{2.37}$$

In the last step the convolution theorem according to equation (2.29) was applied such that the filter operation is performed in spatial domain. Equation (2.37) is referred to as FBP. It states that the distribution of the attenuation coefficient $\mu(\boldsymbol{r})$ can be reconstructed by filtering the acquired projections $p(\vartheta, \xi)$ with the reconstruction kernel

$$\mathcal{F}_1^{-1} |k| = \int |k| e^{2\pi i k \xi} dk = \frac{-1}{2\pi^2 \xi^2} \tag{2.38}$$

and the subsequent backprojection of the filtered projections to image domain as described by (2.27).

2.1.5 Industrial CT

In contrast to conventional optical or tactile inspection techniques, CT is able to reconstruct a complete virtual 3D model of the sample in terms of a voxel volume or a surface mesh, respectively. This offers several benefits: internal structures can be investigated nondestructively, all features of the sample are measured at once and the evaluation of the measurement is mostly independent of the data acquisition. Furthermore, CT provides a high spatial resolution up to a few tens of nanometers [72], comparably short scan times and only minor restrictions with respect to the shape, the size or the material of the sample. Therefore, CT has become an important tool in several areas of industry such as flaw detection, failure analysis, reverse engineering, assembly analysis or metrology [12].

In contrast to clinical CT systems in which the detector and the x-ray source are mounted on a gantry, industrial CT systems are usually based on a table-top setup in which the object itself is rotated. Using a rotary table that is displaceable in source-detector direction allows to change the geometric magnification i.e. the spatial resolution as well as the field of measurement. Thus, the geometry can be adjusted to fit samples with varying dimensions. Besides the geometry, the spatial resolution is mainly determined by the focal spot size as well as the detector pixel size. Therefore, industrial CT systems are preferably equipped with high resolution flat detectors and micro-focus transmission x-ray tubes that are typically operated between 40 kV and 450 kV [12].

2.2   Monte Carlo Methods

MC methods refer to a broad class of approaches that make use of repeated random sampling to solve deterministic or stochastic problems. The earliest calculations date back to Compte du Buffon's needle tossing experiments in 1777 to calculate the number π [73] [1]. However not until the availability of digital computers, MC methods became widely applicable. Today, MC finds various applications ranging from social science to quantum chemistry. In this thesis, however, the main purpose is the simulation of radiation transport. As several results are derived using MC methods, the basic principles are briefly reviewed in the following.

2.2.1   Basics of Probability Theory

i.)   Random Variables

A random variable is a variable that results from a repeatable process with at least two possible outcomes which, however, cannot be predicted with certainty. While the actual realization of random variables is unknown in advance, their distribution is typically well known. Considering a continuous random variable X, it is given in terms of a probability density function (PDF) $p(x)$ that describes the probability P of finding X in an interval $[x, x + dx]$:

$$p(x)dx = P[x < X < x + dx]. \tag{2.39}$$

Since the probability of finding X within a given interval cannot be negative and the probability of finding X anywhere has to equal 1, a PDF has to have the following properties:

$$p(x) \geq 0, \tag{2.40}$$

and

$$\int_{-\infty}^{\infty} p(x)dx = 1. \tag{2.41}$$

In case of discrete random variables the PDF can be defined similarly using delta distributions $\delta(x - x_n)$. Considering a random variable that can take N discrete values $x = x_0, x_1, \ldots, x_{N-1}$ with the corresponding probabilities $p_0, p_1, \ldots, p_{N-1}$, it is given as:

$$p(x) = \sum_n p_n \delta(x - x_n). \tag{2.42}$$

---

[1] Assume to have a needle of length d tossed on a plane ruled with parallel lines of distance d. For a certain angle ϑ the probability to hit the line is $P = d \cdot |\cos\vartheta|/d$. Averaging over all angles $((1/\pi) \cdot \int_0^\pi |\cos\vartheta| d\vartheta)$ yields a probability of $2/\pi$. Thus, π can be estimated as twice the number of tries divided by the number of misses.

Sometimes it is more convenient to characterize random variables by their cumulative distribution function (CDF) $C(x)$ that describes the probability $P(X \leq x)$ of X being equal or smaller than x. Thus, $C(x)$ is given as the following integral over the PDF:

$$C(x) = \int_{-\infty}^{x} p(x')dx' \tag{2.43}$$

ii.)   Expected Value, Variance and Covariance

The expected value $E(X)$ of a random variable X with the PDF $p(x)$ is defined as:

$$E(X) = \int_{-\infty}^{\infty} x \, p(x) \, dx. \tag{2.44}$$

Similarly, the expected value $E(f)$ of a function $f(X)$ is defined as:

$$E(f) = \int_{-\infty}^{\infty} f(x) \, p(x) \, dx. \tag{2.45}$$

The variance $\mathrm{Var}(X)$ of a random variable X is defined as the expected value of the squared deviation from the mean:

$$\mathrm{Var}(X) = E[(X - E(X))^2]. \tag{2.46}$$

Similarly, the variance $\mathrm{Var}(X)$ of a function $f(X)$ is defined as:

$$\mathrm{Var}(f) = E[(f - E(f))^2]. \tag{2.47}$$

Given two random variables X and Y and a constant c, the following properties can be derived using equation (2.44) and equation (2.46), respectively:

$$E(cX + Y) = cE(X) + E(Y) \tag{2.48}$$
$$\mathrm{Var}(cX + Y) = c^2 \, \mathrm{Var}(X) + \mathrm{Var}(Y) + 2c \, E[(Y - E(Y)) \cdot (X - E(X))]. \tag{2.49}$$

The last term in equation (2.49) is referred to as covariance and describes the correlation between the random variables X and Y. In case of X and Y being independent, the covariance equals zero.

iii.)   Two-Dimensional Random Variables

Similar to section i.) the PDF $p(x, y)$ of a 2D random variable (X, Y) must have the following properties:

$$p(x, y) \geq 0, \tag{2.50}$$

and

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} p(x, y)dy = 1. \tag{2.51}$$

The PDF over a subset of the sample space, i.e. the probability that X takes a value in the interval $[x, x + dx]$ and Y any other value is referred to as marginal PDF $q(x)$:

$$q(x) = \int_{-\infty}^{\infty} p(x, y) dy. \tag{2.52}$$

The joint probability density can be expressed as:

$$p(x, y) = q(x)p(y|x) = q(y)p(x|y), \tag{2.53}$$

where $p(x|y)$ is the conditional PDF, i.e. the probability of X taking a value in $[x, x + dx]$ given $Y = y$:

$$p(x|y) = \frac{p(x, y)}{q(y)}. \tag{2.54}$$

It has to be noted that using the definition of the conditional PDF, any joint PDF of many random variables can be decomposed as follows:

$$p(x_1, \ldots, x_N) = q(x_1) \cdot p(x_2|x_1) \ldots p(x_N|x_{N-1}, \ldots, x_1) \tag{2.55}$$

iv.) Law of Large Numbers & Monte Carlo Integration

The basic idea of MC methods relies on the law of large numbers which states that the mean $\bar{f}_N$ of a sequence of samples $\{f(X_n)\}$, with X_n being drawn from the same distribution $p(x)$, converges to the expected value:

$$\lim_{N \to \infty} \bar{f}_N = \lim_{N \to \infty} \frac{1}{N} \sum_n f(X_n) = \int_{-\infty}^{\infty} f(x) \, p(x) \, dx, \tag{2.56}$$

Note that the this holds only true as long as the function f is integrable, finite and piecewise continuous. In that case $\bar{f}$ is referred to as a consistent estimator of the integral on the right-hand side of equation (2.56).

Reading equation (2.56) the other way round can be applied to solve integrals using random samples. This so-called Monte Carlo integration shall be briefly explained by considering the following integral:

$$I = \int_a^b g(x) dx. \tag{2.57}$$

Introducing an arbitrary PDF $p(x)$ with $p(x) = 0 \ \forall \ x \notin [a, b]$, the function $g(x)$ to be integrated can be expressed as $g(x) = G(x)p(x)$, with $G(x) = g(x)/p(x)$. Substitution yields:

$$
\begin{aligned}
I &= \int_a^b G(x)p(x) dx \\
&= \lim_{N \to \infty} \frac{1}{N} \sum_n G(X_n) \overset{N \gg 1}{\approx} \frac{1}{N} \sum_n G(X_n),
\end{aligned} \tag{2.58}
$$

where the second step applies the law of large numbers. Thus, an estimator of the integral I is given as the average of $\{G(X_n)\}$, with X_n being random variables drawn from the distribution $p(x)$. Given $\mathrm{Var}(G)$, the unbiased estimate of the variance $\mathrm{Var}(I)$ can be calculated as:

$$\mathrm{Var}(I) = \frac{1}{N^2} \sum_n \mathrm{Var}(G)) = \frac{\mathrm{Var}(G)}{N}. \tag{2.59}$$

It has to be noted that equation (2.59) also holds true for integrals of higher dimension D, which is a major advantage over other numerical integration approaches, e.g. first-order numerical quadrature with a variance proportional to $N^{-2/D}$.

v.) Central Limit Theorem

While the law of large numbers states that the Monte Carlo estimate converges to the expected value for increasing sample numbers N, the central limit theorem applies to the distribution of that estimate. It basically states that the average $\bar{X}$ of a sequence $\{X_n\}$ of independent and identically distributed random variables, with expected values $\mathrm{E}(X) = \mu$ and variance $\mathrm{Var}(X) = \sigma^2$, converges in distribution to a normal $N(0, \sigma^2)$, i.e.:

$$\sqrt{N}(\bar{X} - \mu) \to N(0, \sigma^2). \tag{2.60}$$

Here, convergence in distribution means that the CDF of $\sqrt{N}(\bar{X} - E(X))$ converges pointwise to the CDF of $N(0, \sigma^2)$. Thus, the central limit theorem enables us to turn knowledge of the expected value and the variance, that are given according to equation (2.48) and equation (2.49), into statements on the probability of μ being the result of a given trial.

2.2.2 Random Sampling

Monte Carlo methods require the generation of random numbers that are distributed according to a given PDF. Random sampling, in that sense, refers to the process of drawing samples from that PDF. In most cases the corresponding random numbers are rather so-called pseudo-random numbers as they are generated by a deterministic approach. Thus, in contrast to true random numbers that may result from a random physical process such as radioactive decay or thermal noise, they are reproducible and not at all random in the mathematical sense. However, true randomness is not necessary and not even wanted as long as a sequence of pseudo-random numbers is indistinguishable from a sequence of true random numbers. The basics of generating such sequences are briefly discussed in the following. For a more comprehensive review of random number generation the reader is referred to [74].

i.) Generation of Uniform Random Numbers

The generation of arbitrarily distributed random numbers is usually based on the use of uniform random numbers $\xi \in [0, 1)$ as shown in the next section. The earliest computer-based method of generating such numbers has been proposed by von Neumann in 1946

and is referred to as mid-squares method [75]. Given any number as seed, this approach squares it and removes the middle digits of the resulting number as the random number. Subsequently, that number is used as seed for the next iteration and so on. Running into cyclic behavior with a very short periods, the mid-square method was soon abandoned. In the following decades so-called multiplicative linear congruential generators (MLCGs) were widely used [76]. Given a seed R_0, they produce a sequence of random numbers as:

$$R_{n+1} = (aR_n + c) \bmod m, \ \ \xi_{n+1} = R_{n+1}/m, \tag{2.61}$$

with mod referring to the modulo operation, and a, c and m being constants to be properly chosen. Good random properties are e.g. achieved using $a = 7^5$, $c = 0$, and $m = 2^{31} - 1$ [77]. However, the period of this generator is only in the order of 10^9 which might not be sufficient for several MC calculations. Modern uniform random number generators are therefore often based on so-called Xorshift-generators that produce random sequences with high period by a small number of XOR and bitshift operations [78] or combined generators that combine MLCG and Xorshift-generators. If not stated otherwise, all random numbers used in this work are generated by a combined generator as described in reference [79].

ii.)   Inverse Transform Method

The inverse transform method represents an approach to draw samples from a given PDF $p_X(x)$ using random numbers ξ distributed uniformly in the interval $[0, 1)$. The basic principle is depicted in figure 2.5 and relies on the properties of the CDF of $p(x)$. By definition (see (2.43)) the CDF $C_X(x)$ of X is a monotonically increasing function of x, and thus, has an inverse function $C_X^{-1} : [0, 1] \to \mathbb{R}$. The equation

$$Y = C_X^{-1}(\xi) \tag{2.62}$$

defines a new random variable Y. The CDF of Y is given as:

$$
\begin{aligned}
C_Y(x) \quad &\equiv \quad P(Y \leq x) \\
&\stackrel{\partial_x C_X \geq 0}{=} P(C_X(Y) \leq C_X(x)) \\
&\stackrel{\xi = C_X(Y)}{=} P(\xi \leq C_X(x)) \\
&\stackrel{p_\xi(x)=1}{=} C_X(x),
\end{aligned}
\tag{2.63}
$$

where $P(a \leq b)$ denotes the probability of a being smaller or equal to b. To derive equation (2.63) the fact that $C_X(x)$ is a monotonically increasing function was used in the first step, the definition of ξ according to equation (2.62) was used in the second step, and the definition of the PDF of a uniform random variable $p_\xi(x) = 1$ was used in the last step. As $C_Y(x) = C_X(x)$, it is proven that the random variable Y is distributed according to the PDF $p(x)$. Thus, random samples from $p(x)$ can be drawn by sampling
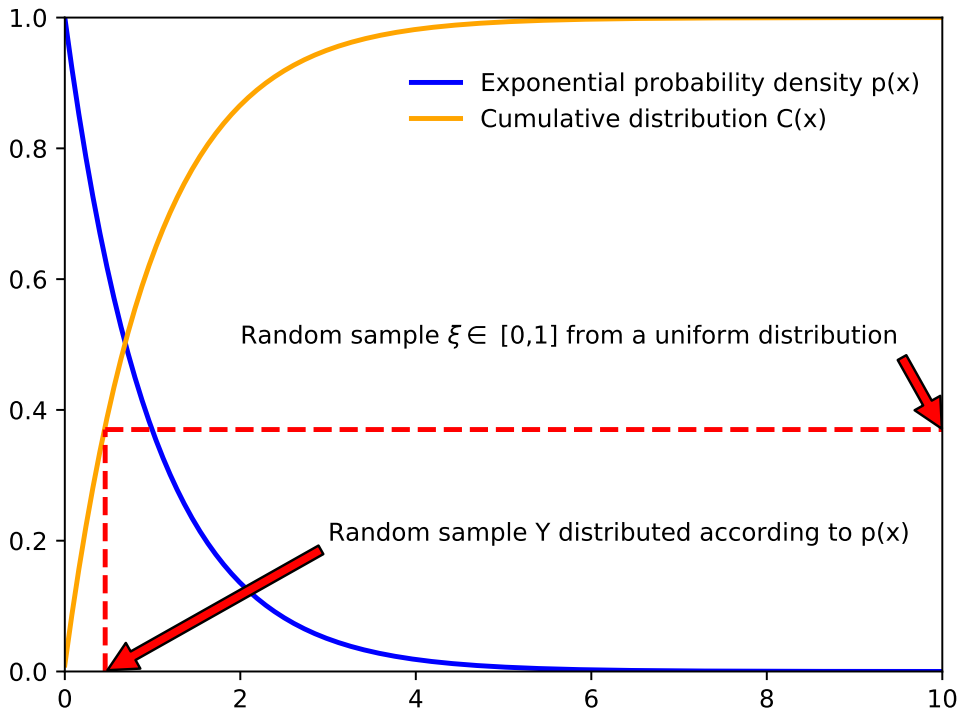
Figure 2.5: Basic principle of the inverse transform method. Given a PDF $p(x)$, it's corresponding CDF $C_X(x)$, and a uniform random variable $\xi \in [0, 1)$, the inverse transform $C_X^{-1}(\xi)$ defines a new random variable Y that is distributed according to $p_X(x)$.

a uniform random number and applying the inverse transform as given in equation (2.62).

It has to be noted that the inverse transform method does not require the CDF $C_X(x)$ to have an analytic inverse. Rewriting equation (2.62) as

$$\xi = \int_{-\infty}^{x} p(x')\,dx', \tag{2.64}$$

defines the sampling equation of a variable from an arbitrary PDF $p(x)$. In any case, this integral can be solved numerically.

iii.) Discrete Distributions

Considering the special case of discrete distributions with point probabilities $p_0, \ldots, p_{N-1}$ corresponding to the discrete points $x_0, x_1, \ldots, x_{N-1}$, the CDF is given as:

$$C(x) = \int_{-\infty}^{x} \sum_{n=0}^{N-1} p_n \delta(x - x_n) = \sum_{n=0}^{n \leq x} p_n. \tag{2.65}$$

As depicted in figure 2.6, the random sample $X \in x_0 x_1, \ldots, x_{N-1}$ can be determined by finding the index n such that

$$C(x_n) < \xi \leq C(x_{n+1}). \tag{2.66}$$

Thus, most easily X can be found using some sort of search algorithm. However, even fast approaches such as a binary search (complexity $\mathcal{O}(\log\ n)$) might require too long to be applicable in high speed MC codes. An optimal sampling approach for discrete distributions is described in the following section.
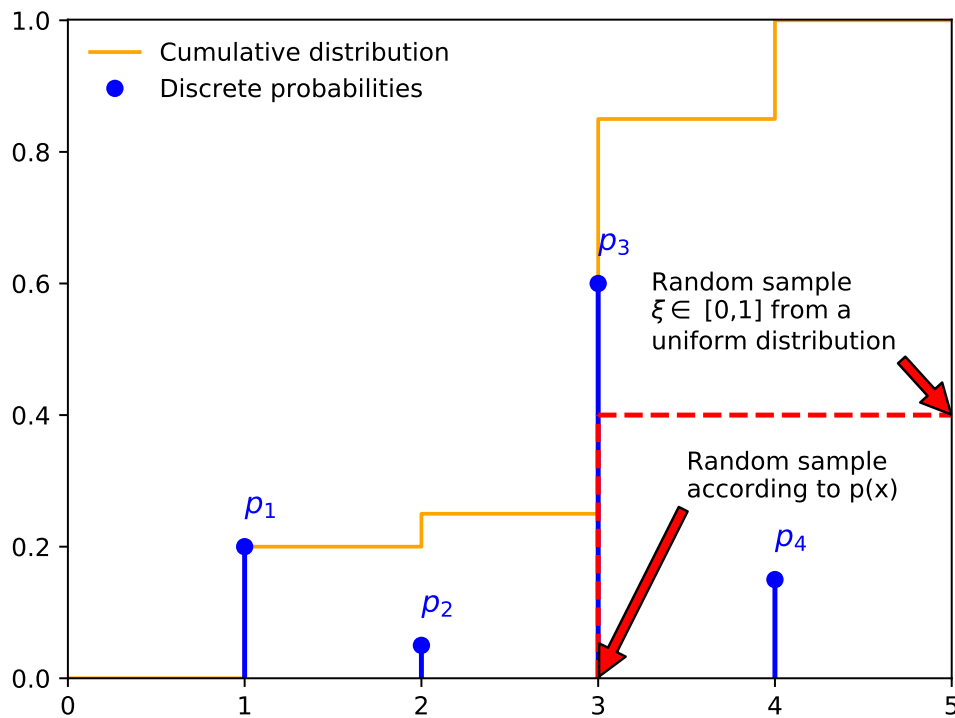
Figure 2.6: Basic principle of the inverse transform method for discrete distributions. Using a uniform random variable $\xi \in [0, 1]$, random samples distributed according to a given PDF, with the corresponding CDF $C(x)$ can be drawn by finding the index n such that $C(x_n) < \xi \leq C(x_{n+1})$.

iv.)   Walkers's Discrete Sampling Method

To avoid the need of a search algorithm to determine the discrete random variable X, it was desirable to have equal probabilities ($p_i = 1/N$) for each of the N possible realizations $x_0, x_1, \ldots, x_{N-1}$ of X. In that case, these realizations could be distributed on a regular grid with the grid points $i \in \{1, 2, \ldots, N-1\}$ corresponding to the realization x_i. Given a (continuous) uniform random number $\xi \in [0, 1)$, a certain x_i can

be sampled as the realization corresponding to the index

$$i = \lfloor \xi \cdot N \rfloor, \tag{2.67}$$

where $\lfloor \, \rfloor$ denotes the cast to an integer number by cutting the decimal places of the argument. However, as depicted in figure 2.7, in case of $p_i \neq p_j$ for $i \neq j$ the possible realizations of X are distributed on an irregular grid. To make use of a fast sampling scheme as given by equation (2.67), Walker's method introduces a regular grid with N points and assigns each grid point at most two possible realizations x_i and x_{ij} of X as well as the conditional probability $p(i|x_{ij})$, i.e. a 3-tuple $(x_i, x_{ij}, p(i|x_{ij}))$, with i being the index of the grid point. Thus, there are N tuples $(x_0, x_{0j}, p(0|x_{0j})), (x_1, x_{1j}, p(1|x_{1j})), \ldots, (x_{N-1}, x_{(N-1)j}, p(N-1|x_{(N-1)j}))$. Now the sampling process consists of the following steps:

1. Sampling of a uniform random number ξ.

2. Set the index $i = \lfloor N\xi \rfloor$, and $r = N\xi - \lfloor N\xi \rfloor$ (note that r is again a uniform random number in $[0, 1)$).

3. If $r \leq p(i|x_{ij})$ return x_{ij}.

4. Else return x_i.
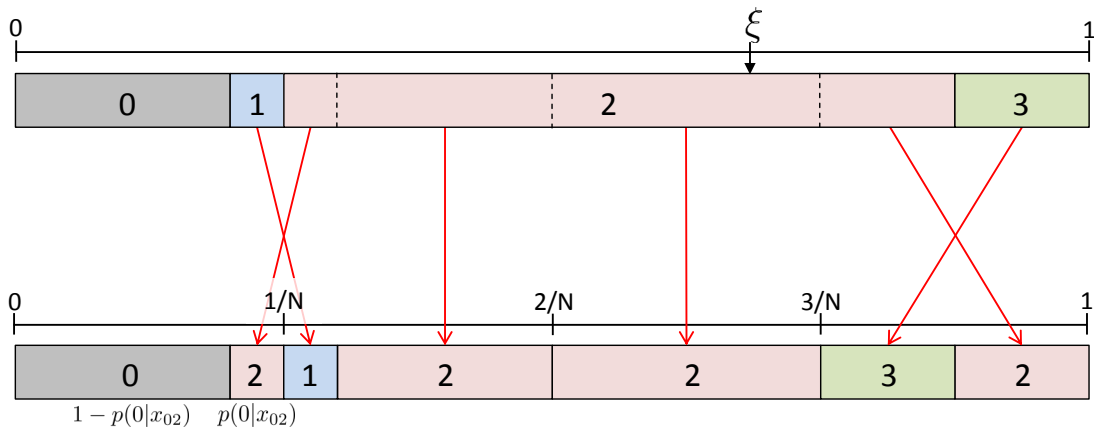
Figure 2.7: Basic principle of Walker's method. The point probabilities p_i of a discrete distribution form an irregular grid if $p_i \neq p_j$ for $i \neq j$ (see top row). Thus, a given uniform random number ξ cannot be directly mapped to a sample point i but requires a search algorithm to determine the corresponding point according to equation (2.66). Therefore, Walker's method rearranges the probabilities onto a regular grid. Each grid point is assigned at most two different realizations of the discrete random variable X that occur with the conditional probability $1 - p(i|x_{ij})$ and $p(i|x_{ij})$, respectively.

Considering this sampling routine, the probability $P(X = x_i)$ of drawing the sample x_i can be calculated as:

$$P(X = x_i) = \frac{1}{N} \left((1 - p(i|x_{ij})) + \sum_{j \neq i} p(j|x_{ij}) \right). \tag{2.68}$$

Thus, given the point probabilities $p_0, p_1, \ldots, p_{N-1}$ of a discrete distribution, samples can be drawn using Walker's method by setting the realizations x_{ij} as well as the conditional probabilities $p(i|x_{ij})$ such that $P(X = x_i) = p_i \, \forall \, i \in \{0, 1, \ldots, N-1\}$.

v.) Rejection Methods

Rejection methods, introduced by von Neumann [80], represent another class of approaches to draw random samples from a given PDF p(x). The basic idea of rejection methods is to draw samples from a PDF $\pi(x)$ (different from $p(x)$) and to reject them subsequently with a certain probability such that the accepted samples are distributed according to $p(x)$ (see figure 2.8). The corresponding rejection function $r(x)$ is defined by the following equation:

$$p(x) = C\pi(x)r(x), \tag{2.69}$$

with C being a constant such that $C\pi(x) > p(x)$. This choice of C ensures that $0 \leq r(x) < 1$. Given $p(x), \pi(x)$ and C, the rejection method consists of the following four steps:

1. Sample a random variable X from $\pi(x)$.

2. Sample a uniform random number $\xi \in [0, 1)$.

3. If $\xi > r(x)$ reject X and start over at 1.

4. Else return X.

The efficiency η of rejection methods can be defined as the expected value of the rejection function $r(x)$, i.e.

$$\eta = \int r(x)\pi(x)dx = \frac{1}{C}. \tag{2.70}$$

Thus, the PDF $p(x)$ should be chosen such that the constant C can be set as low as possible while fulfilling the requirement $C\pi(x) > p(x)$. However, it has to be noted that the efficiency as defined above does not consider the computational cost of sampling X from $\pi(x)$. Distributions that can be sampled efficiently might be preferred over more complex distributions that are more similar to $p(x)$.
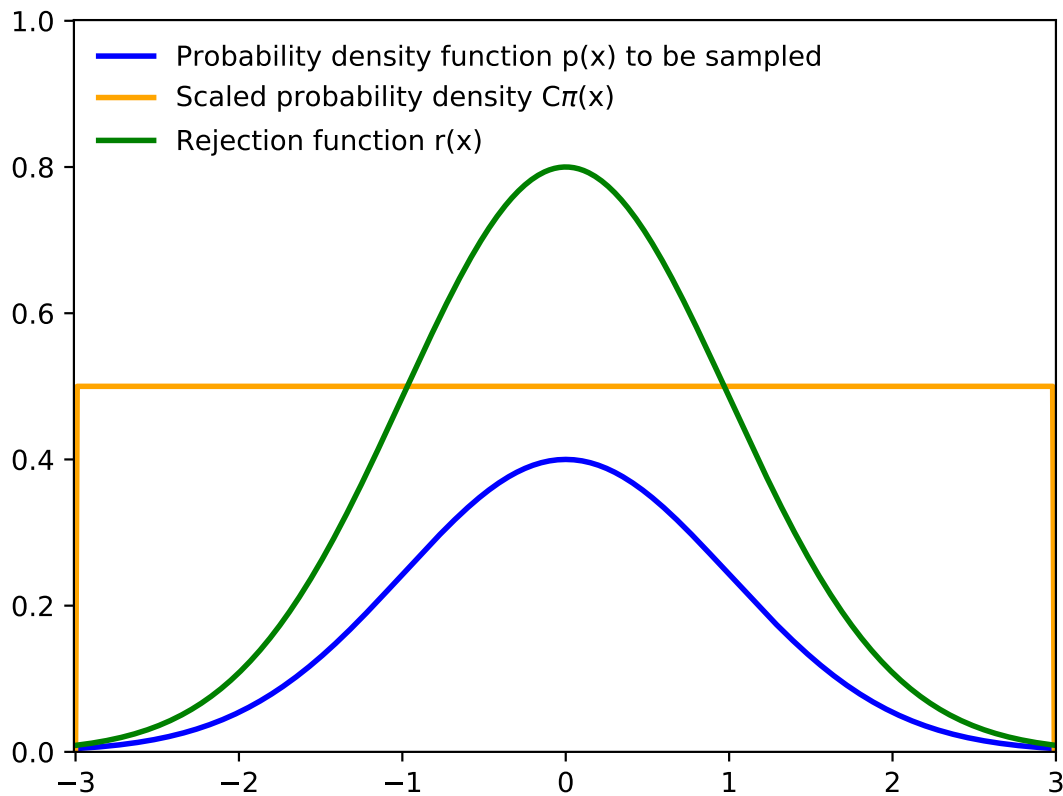
Figure 2.8: Basic principle of rejection methods. Samples are drawn from a PDF $\pi(x)$ (different from $p(x)$) with $C\pi(x) \geq p(x)$ and rejected with a certain probability $P = 1 - r(x)$ such that the accepted samples are distributed according to $p(x)$.

vi.) Multivariate Random Variables

As stated in equation (2.55) the PDF of a multivariate distribution can be expressed as the product of a marginal PDF and conditional PDFs. Since marginal and conditional PDFs depend only on one variable, samples can be drawn as described in the previous sections. Finally, the random variable is composed as the product of the individual samples.

2.2.3 Application to X-Ray Imaging

Considering x-ray imaging, the state $\boldsymbol{S} = (r_x, r_y, r_z, d_x, d_y, d_z, E)^T$ of an x-ray photon is typically determined by it's position $\boldsymbol{r} = (r_x, r_y, r_z)^T$, it's flight direction $\boldsymbol{d} = (d_x, d_y, d_z)^T$ as well as it's energy E. According to equation (2.44) the expected value of any quantity of interest $f(\boldsymbol{S})$, e.g. the absorbed dose at a given position or the number of x-rays reaching a certain detector element, can be calculated as:

$$\mathrm{E}(f) = \int f(\boldsymbol{S}) p(\boldsymbol{S}) d\boldsymbol{S}. \tag{2.71}$$

As stated by equation (2.58), an MC estimate of $\mathrm{E}(f)$ can be derived by averaging f over a sequence of random samples $\{\boldsymbol{S}\}_N$ distributed according to $p(\boldsymbol{S})$. Unfortunately, the joint PDF $p(\boldsymbol{S})$ is usually unknown or too complex to be calculated. However, the simulation of random walks of single photons, where each step is sampled from a conditional PDF, provides a practical approach to draw samples from the joint PDF. For a typical CT setup, a straightforward implementation of an MC simulation consists of the following steps:

 1: Define the margins of the volume to be simulated in terms of a bounding box
 2: **for** Number of x-ray photons N to be simulated **do**
 3:     Sample the position $\boldsymbol{r}_0$ of the x-ray emission
 4:     Sample the flight direction $\boldsymbol{d}_0$ of the x-ray photon
 5:     Sample an x-ray energy E_0 according to the x-ray source spectrum
 6:     **while** $E_n > 0$ and $\boldsymbol{r}_n$ inside bounding box **do**
 7:         Sample path length λ to the next interaction point: $\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + \lambda \boldsymbol{d}_n$
 8:         Sample interaction effect I_n
 9:         Sample new flight direction $\boldsymbol{d}_{n+1}$ according to the interaction effect and the flight direction $\boldsymbol{d}_n$ of the incident x-ray photon
10:         Update energy E_{n+1} of the x-ray photon according to the interaction effect and the scatter angle $\varphi = \arccos(\boldsymbol{d}_n \cdot \boldsymbol{d}_{n+1})$
11:         New state of the x-ray photon is given as $\boldsymbol{S}_{n+1} = (\boldsymbol{r}_{n+1}, \boldsymbol{d}_{n+1}, E_{n+1})^T$
12:         **if** X-ray photon is in region of interest **then**
13:           Score the contribution of the current trial: $q \mathrel{+}= f(\boldsymbol{S}_{n+1})$
14:         **end if**
15:     **end while**
16: **end for**
17: MC estimate of $E(f)$ is given as $E(f) \approx q/N$

---

i.) Sampling the Initial State of the X-Ray Photon

Considering x-ray imaging applications, x-ray photons are generated using x-ray tubes as described in section 2.1.1. Thus, the initial state of the x-ray photon is determined by the emission characteristics of the x-ray source. Assuming a point like focal spot at position $\boldsymbol{s} = (s_x, s_y, s_z)^T$, it is completely described by the differential photon fluence

$\frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)$, with N denoting the number of x-ray photons, Ω denoting the solid angle, φ denoting the azimuth angle, and ϑ denoting the polar angle. The corresponding PDF $p(\varphi, \vartheta, E)$ can be calculated as:

$$p(\varphi, \vartheta, E) = \frac{\sin(\vartheta)\frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)}{\int_0^{2\pi} d\varphi \int_0^{\pi} \sin(\vartheta)d\vartheta \int_0^{E_{\max}} dE \frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)}. \tag{2.72}$$

Expressing the joint PDF as

$$p(\varphi, \vartheta, E) = q(\varphi)p(\vartheta|\varphi)p(E|\varphi, \vartheta), \tag{2.73}$$

samples (φ, ϑ, E) can be drawn according to the marginal PDF $q(\varphi)$ or the conditional PDFs $p(\vartheta|\varphi)$ and $p(E|\vartheta, \varphi)$ which are given as:

$$q(\varphi) = \frac{\int_0^{\pi} \sin(\vartheta)d\vartheta \int_0^{E_{\max}} dE \frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)}{\int_0^{2\pi} d\varphi \int_0^{\pi} \sin(\vartheta)d\vartheta \int_0^{E_{\max}} dE \frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)} \tag{2.74}$$

$$p(\vartheta|\varphi) = \frac{\sin(\vartheta) \int_0^{E_{\max}} dE \frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)}{\int_0^{\pi} \sin(\vartheta)d\vartheta \int_0^{E_{\max}} dE \frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)} \tag{2.75}$$

$$p(E|\vartheta, \varphi) = \frac{\frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)}{\int_0^{E_{\max}} dE \frac{d^2N}{d\Omega dE}(\varphi, \vartheta, E)} \tag{2.76}$$

Typically, the photon fluence is not available as a continuous function but rather evaluated on a discrete grid. Thus, the PDFs above are given by their discrete equivalents and can be sampled e.g. using Walker's method. In the special case of an isotropic emission which is often assumed due to the lack of more appropriate models, the photon fluence has no solid angle dependency, and therefore $\frac{d^2N}{d\Omega dE}(\vartheta, \varphi, E) = \frac{dN}{dE}(E)$. Thus, the PDFs are given as:

$$p(\varphi) = \frac{1}{2\pi} \tag{2.77}$$

$$p(\vartheta) = \frac{\sin(\vartheta)}{2}. \tag{2.78}$$

Applying the inverse transform method yields the following sampling equations:

$$\varphi = 2\pi\xi \tag{2.79}$$

$$\vartheta = \cos^{-1}(2\xi - 1). \tag{2.80}$$

In any case a realization of the initial state $\boldsymbol{S}_0$ is given as:

$$\boldsymbol{S}_0 = (s_x, s_y, s_z, \sin(\vartheta_0)\cos(\varphi_0), \sin(\vartheta_0)\sin(\varphi_0), \cos(\vartheta_0), E_0)^T, \tag{2.81}$$

with φ_0, ϑ_0 and E_0 being random samples of the azimuth angle, the polar angle and the energy.

It has to be noted that the assumption of a point like focal spot does not hold true for real CT systems. To be precise, it should be considered within the PDF given in equation (2.73). However, as the focal spot dimensions are usually small compared to the volume to be simulated, it can be neglected in several applications such as x-ray scatter prediction which is the main purpose of the MC simulation in this thesis.

ii.)   Sampling the Path Length

Given the actual state $\boldsymbol{S}_n$ of an x-ray photon, the probability $p(\lambda|\boldsymbol{S}_n)d\lambda$ of having the next interaction after a path length λ, i.e. in the interval $[\boldsymbol{r}_n + \lambda\boldsymbol{d}_n, \boldsymbol{r}_n + (\lambda + d\lambda)\boldsymbol{d}_n]$, corresponds to the relative number of x-ray photons absorbed in that interval. Thus, it can be calculated using the Beer-Lambert (see equation (2.7)) as follows:

$$
\begin{aligned}
p(\lambda|\boldsymbol{S}_n)d\lambda &= \frac{N(\lambda, E_n) - N(\lambda + d\lambda, E_n)}{N_0(E_n)} \\
&= e^{-\int_0^\lambda \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda'} - e^{-\int_0^{\lambda+d\lambda} \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda'} \\
&= e^{-\int_0^\lambda \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda'}\left(1 - e^{-\int_\lambda^{\lambda+d\lambda} \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda'}\right) \\
&= e^{-\int_0^\lambda \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda'}\left(1 - e^{-\mu(\boldsymbol{r}_n + \lambda\boldsymbol{d}_n, E_n)d\lambda}\right) \\
&= e^{-\int_0^\lambda \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda'}\left(\mu(\boldsymbol{r}_n + \lambda\boldsymbol{d}_n, E_n)d\lambda\right).
\end{aligned}
\tag{2.82}
$$

The corresponding CDF $C(\lambda|\boldsymbol{S}_n)$ can be calculated by integrating equation (2.82) and is given as:

$$
C(\lambda|\boldsymbol{S}_n) = 1 - e^{-\int_0^\lambda \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda'}.
\tag{2.83}
$$

Applying the inverse transform method yields the following sampling equation for the path length λ:

$$
\ln(\xi) = -\int_0^\lambda \mu(\boldsymbol{r}_n + \lambda'\boldsymbol{d}_n, E_n)d\lambda',
\tag{2.84}
$$

with ξ being a uniform random number. In case of a homogeneous distribution of the attenuation coefficient $\mu(\boldsymbol{r}, E) = \mu(E)$, the integral term equals $\lambda\mu(E_n)$. Thus, there is an analytic inverse that is given as:

$$
\lambda = -\frac{1}{\mu(E_n)}\ln(\xi).
\tag{2.85}
$$

Otherwise, i.e. if $\mu(\boldsymbol{r}, E) \neq \mu(E)$, equation (2.84) has to be solved for λ numerically. Since this inversion might be computationally expensive, a so-called Woodcock sampling is often used instead [81]. This approach determines the path length in an iterative manner according to the following update scheme:

$$
\lambda_{m+1} = \lambda_m - \frac{1}{\mu_{\max}(E_n)}\ln(\xi_m),
\tag{2.86}
$$

and

$$
\lambda_0 = -\frac{1}{\mu_{\max}(E_n)}\ln(\xi_m).
\tag{2.87}
$$

Here, the path length of the mth iteration is accepted with a probability equal to $\mu(\boldsymbol{r}_n + \lambda_m + \boldsymbol{d}_n, E_n)/\mu_{\max}(E_n)$, with $\boldsymbol{r}_n$ being the starting point of the x-ray photon and $\boldsymbol{d}_n$ being its flight direction. Subsequently, the accepted path length is used to calculate the next interaction point $\boldsymbol{r}_{n+1}$ as:

$$
\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + \lambda_m\boldsymbol{d}_n.
\tag{2.88}
$$

iii.) Sampling the Interaction Effect

Most x-ray imaging techniques use x-ray energies below 1022 keV. Therefore, the relevant x-ray matter interactions are absorption through the photoelectric effect (PE), Compton scattering (CS) and Rayleigh scattering (RS)(see section 2.1.2). Given the energy E_n of an x-ray photon, the conditional probabilities of the three effects are given as:

$$p(\text{PE}|E_n) = \frac{\mu_{\text{PE}}(E_n)}{\mu_{\text{PE}}(E_n) + \mu_{\text{CS}}(E_n) + \mu_{\text{RS}}(E_n)} \tag{2.89}$$

$$p(\text{CS}|E_n) = \frac{\mu_{\text{CS}}(E_n)}{\mu_{\text{PE}}(E_n) + \mu_{\text{CS}}(E_n) + \mu_{\text{RS}}(E_n)} \tag{2.90}$$

$$p(\text{RS}|E_n) = \frac{\mu_{\text{RS}}(E_n)}{\mu_{\text{PE}}(E_n) + \mu_{\text{CS}}(E_n) + \mu_{\text{RS}}(E_n)}, \tag{2.91}$$

where μ_i denotes the attenuation coefficient due to the interaction effect $i \in \{\text{PE, CS, RS}\}$. Given a uniform random number ξ, the interaction effect can be sampled using the inverse transform method as:

$$\text{Photoelectric effect} \quad \text{if} \quad \xi \leq p(\text{PE}|E_n) \tag{2.92}$$

$$\text{Compton scattering if} \quad p(\text{PE}|E_n) < \xi \leq p(\text{PE}|E_n) + p(\text{CS}|E_n) \tag{2.93}$$

$$\text{Rayleigh scattering if} \quad p(\text{PE}|E_n) + p(\text{CS}|E_n) < \xi \leq 1 \tag{2.94}$$

iv.) Updating the Flight Direction

The flight direction is updated according to the actual energy E_n of the x-ray photon as well as the interaction effect that has been sampled. In case of an interaction via the photoelectric effect the x-ray photon is absorbed, and thus the flight direction is not updated but the simulation is terminated at this point. Considering Compton and Rayleigh scattering, the scatter angles are distributed according to their differential cross sections $\frac{d\sigma}{d\Omega}(\varphi, \vartheta, E)$(see section 2.1.2). In both cases, the differential cross section does not depend on the azimuth angle φ, i.e. the angle within the plane perpendicular to the x-ray photon's flight direction $\boldsymbol{d}_n$. Thus, it can be sampled similar to equation (2.79) as $\phi = 2\pi\xi$, with ξ being a uniform random number. Given the energy E_n, the conditional PDFs $p_i(\vartheta|E_n)$ for the polar angle can be calculated as:

$$p_i(\vartheta|E_n) = \frac{2\pi \sin(\vartheta)\frac{d\sigma_i}{d\Omega}(\varphi, \vartheta, E)}{\int_0^{2\pi} d\varphi \int_0^\pi \sin(\vartheta)d\vartheta \frac{d\sigma_i}{d\Omega}(\varphi, \vartheta, E)}, \tag{2.95}$$

with $i \in \{\text{CS, RS}\}$ denoting the interaction effect. Typically, MC codes do not use analytic expressions of the differential cross section but rather rely on tabulated values of a certain database such as the EPDL for instance. Therefore, the polar angle can be derived using the discrete equivalent of equation (2.95) and an appropriate discrete sampling approach (e.g. Walker's method). With $\boldsymbol{d}_{n,\perp}$ being a vector perpendicular to the initial flight direction $\boldsymbol{d}_n$, the new flight direction $\boldsymbol{d}_{n+1}$ can be calculated as:

$$\boldsymbol{d}_{n+1} = R_{\boldsymbol{d}_n}(\varphi)R_{\boldsymbol{d}_{n,\perp}}(\vartheta)\boldsymbol{d}_n, \tag{2.96}$$

where $R_{\boldsymbol{n}}(\alpha)$ denotes a rotation matrix that rotates a given vector by the angle α around the axis $\boldsymbol{n}$.

v.) Updating the Energy

The energy is updated according to the interaction effect and the polar scatter angle ϑ. In case of a photoelectric interaction, the photon is absorbed. Thus, the energy E_{n+1} equals zero. In case of Compton scattering the energy is updated according to equation (2.14) while Rayleigh scattering is an elastic process that leaves the energy unchanged. Thus, the energy updates are given as follows:

$$\text{Potoelectric effect:} \quad E_{n+1} = 0, \tag{2.97}$$

$$\text{Compton scattering:} \quad E_{n+1} = \frac{E_n}{1 + \frac{E_n}{m_0 c^2}(1 - \cos\vartheta)}, \tag{2.98}$$

$$\text{Rayleigh scattering:} \quad E_{n+1} = E_n. \tag{2.99}$$

2.2.4 Variance Reduction

According to equation (2.59) the variance of any MC estimate can be reduced by increasing the number of samples or, the number of simulated x-ray photons, respectively. However, as the computation time increases linearly with the number of samples, the term variance reduction rather refers to the application of more sophisticated sampling techniques which lead to a decrease of the variance without affecting the computation time significantly. The basic idea can be motivated as follows. Suppose to be interested in solving the integral

$$I = \int_{-3}^{3} g(x)dx = \int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \approx \frac{1}{N} \sum_n \frac{g(X_n)}{p(X_n)}, \tag{2.100}$$

where the latter term represents the MC estimate according to equation (2.58), and X_n is a random variable sampled from $p(x)$. Here, $p(x)$ can be any function that can be interpreted as a PDF. Figure 2.9 shows the evaluation of the integral using three different PDFs, namely a uniform PDF, a triangular PDF and an exponential PDF. Obviously, the variance is smaller the more similar the PDF is to the function to be evaluated. Thus, the variance can be reduced by adjusting the sampling such that regions with a high contribution to the integral are sampled more often.

Unfortunately, the application of this strategy to x-ray imaging is more complicated as the PDF of the quantity to be evaluated is (usually) unknown. As described in section 2.2.3, samples are drawn by simulating single x-ray photon tracks according to conditional PDFs. However, this procedure might not sample the unknown PDF optimally. A demonstrative example is given in figure 2.10. Using a conventional path length sampling according to equation (2.85), points that (almost) do not contribute to the scatter distribution at all are sampled with the same probability as points with a high contribution. Thus, using a conventional sampling might simulate several unnecessary photon tracks. Furthermore, it can be seen that the contribution highly depends on

the imaging setup. Therefore, an optimal sampling of the unknown PDF often requires dedicated prior knowledge. A brief description of the most common strategies is given in the following. For a more comprehensive discussion of variance reduction approaches the reader is referred to Bielajew and Rogers [82] or Mainegra-Hing and Kawrakow [83].
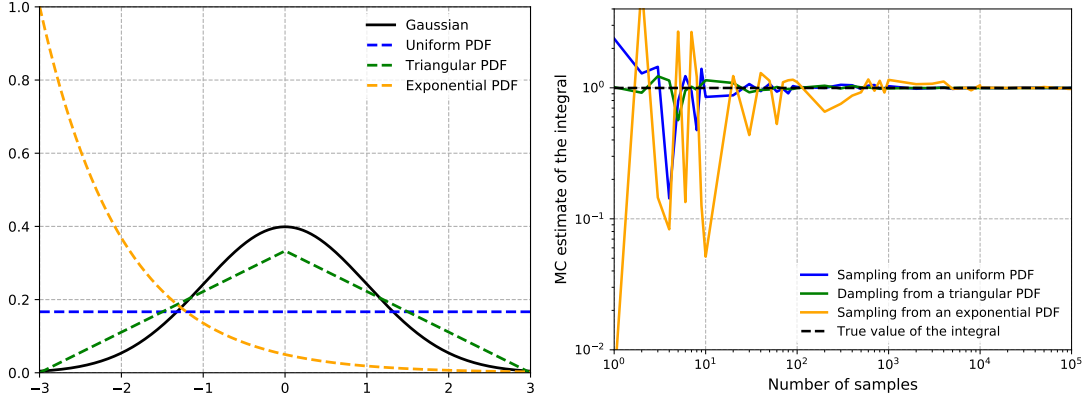
Figure 2.9: MC estimate of the integral from -3 to 3 over a Gaussian with zero mean and unit variance using different sample numbers (right) and different PDFs (left). The lowest variance can be achieved using a triangular PDF which is most similar to the Gaussian.

i.) Basic Principle

The basic principle of most variance reduction approaches is to sample points with a high contribution to the quantity to be evaluated more frequently. Thus, samples have to be drawn according to biased (conditional) PDFs $p_{\text{bias}}(x)$ that have the desired properties but, consequently, do not represent the underlying physics correctly. To keep the MC simulation unbiased, the improper sampling is accounted by introducing a weight such that

$$w_{\text{bias}} \cdot p_{\text{bias}}(x) = p_{\text{real}}(x). \tag{2.101}$$

Given a sequence of random samples $\{X_{\text{bias}}\}_N$ and the corresponding weights $\{w_{\text{bias}}\}_N$, the MC estimate of the expected value of X can be calculated as:

$$E(X) \approx \frac{1}{N} \sum_n X_{n,\text{real}} = \frac{1}{N} \sum_n w_{n,\text{bias}} \cdot X_{n,\text{bias}}. \tag{2.102}$$

ii.) Stratified Sampling

Considering MC as an estimator of the integral given in equation (2.71), stratified sampling makes use of the fundamental property of the Riemann integral:

$$\text{E}(f) = \int_I f(x)p(x)dx = \sum_{m=1}^{M} \int_{I_m} f(x)p(x)dx, \tag{2.103}$$
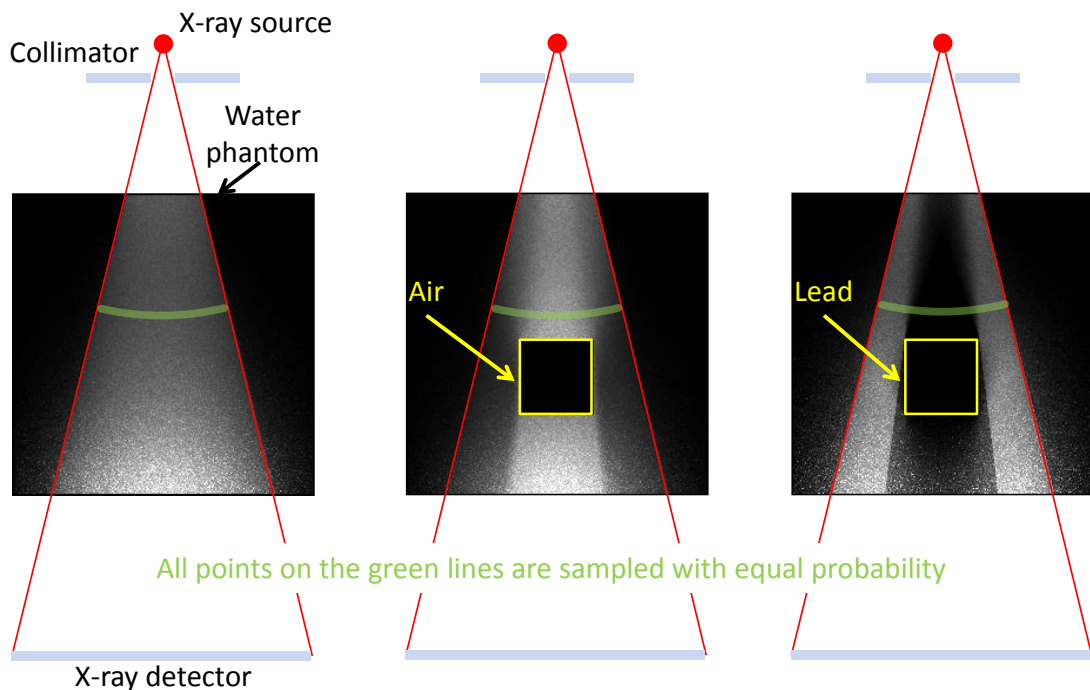
Figure 2.10: MC scatter simulation of a water phantom with three different setups. The grayscale image indicates the contribution of every pixel to the simulated scatter distribution (bright = high contribution). Using a straightforward sampling approach, all points of the green curve are sampled with the same probability although their contributions to the scatter distribution differ significantly.

where $\{I_m\}$ are M mutually exclusive sub-intervals of I. With $w_m = \int_{I_m} p(x)dx$ being the probability of drawing a random sample from $p(x)$ within the interval I_m, the latter term of equation (2.103) can be rewritten as:

$$\sum_{m=1}^{M} \int_{I_m} f(x)p(x)dx = \sum_{m=1}^{M} \int_{I_m} f(x)w_m p(x|x \in I_m)dx, \tag{2.104}$$

where $p(x|x \in I_m)$ denotes the conditional PDF of $p(x)$ given that $X \in I_m$. Thus, an unbiased MC estimator of the integral in equation (2.103) is given by:

$$\mathrm{E}(f)|_{\mathrm{strat}} \approx \sum_{m=1}^{M} \frac{w_m}{N_m} \sum_{i=1}^{N_m} f(X_{im}), \tag{2.105}$$

where N_m denotes the number of samples of the sub-interval m and X_{im} the i-th random number distributed according to $p(x|x \in I_m)$. With $\mu_m = \int_{I_m} f(x)p(x|x \in I_m)dx$ and $\sigma_m^2 = \int_{I_m} (f(x) - \mu_m)^2 p(x|x \in I_m)dx$ being the mean and the variance of the m-th sub-interval, the variance of the stratified sampling estimate (equation (2.105)) can be

calculated as:

$$\text{Var}(f)|_{\text{strat}} = \sum_{m=1}^{M} w_m^2 \frac{\sigma_m^2}{N_m}. \tag{2.106}$$

Obviously, the variance can be reduced by distributing a given number of samples N smartly among the sub-intervals. However, typically the variance σ_m^2 of the sub-interval is unknown. Therefore, it needs to be estimated by the sample variance or by an educated guess. Considering, for instance, the setup depicted in figure 2.10. An x-ray photon will not contribute to the scatter signal if it hits the collimator. Thus, it is beneficial to divide the solid angle to be sampled into two intervals, one of them corresponding to the acceptance angle of the collimator. Sampling this interval more often reduces the variance of the scatter estimate. It has to be noted that stratified sampling is similar to a sampling scheme that uses a biased PDF such that the important regions are sampled more often.

iii.)   Forced Interaction

Forced interaction refers to modifying the PDF in such a way that interactions within a certain region of interest or via a desirable interaction effect are promoted. Suppose, for instance, to be interested in x-ray scattering. In that case, an interaction via the photoelectric effect is not desirable as it absorbs the x-ray photon, and thus, terminates the simulation of the track at this point. Consequently, Compton and Rayleigh scattering can be forced by not sampling the interaction effect according to the PDFs given equations (2.89), (2.90), and (2.91), but by neglecting the photoelectric effect. The biased sampling is accounted for by decreasing the weight of the x-ray photon by the probability of a Compton or Rayleigh interaction, i.e. by $w_{\text{bias}} = 1 - \mu_{\text{PE}}(E)/\mu(E)$. Similarly, it is beneficial to force interactions with the x-ray detector. Otherwise it may happen that simulated x-ray photons leave the volume of interest without contributing to the signal at all.

iv.)   Splitting and Russian Roulette

Splitting refers to transforming a particle that is described by its weight w_0 and its state S_0 into N split particles with state $S_{\text{split}} = S_0$ and weight $w_{\text{split}} = w_0/N$. This technique is particularly useful if the interest of the simulation is focused on a certain spatial region. Here, splitting should be applied when the particle approaches the region of interest. Russian roulette can be seen as the reverse process. It kills a particle with a probability of K when the particle moves away from the region of interest. To keep the simulation unbiased, the particle's weight is increased by a factor of $1/(1 - K)$ if it survived the Russian roulette. Here, the efficiency of these methods highly relies on the parameters S and K as well as on the strategy that is used to decide when a particle is splitted and when a particle is subject to a Russian roulette.

2.3 Artificial Neural Networks and Deep Learning

Artificial neural networks (ANNs) are computational frameworks that are inspired by the way information is processed in the human brain. Similar to biological neural networks, they are based on an ensemble of artificial neurons. Here, the key advantage over other machine learning approaches is that such systems are able to learn complex tasks from observational data without the need for explicit feature engineering. For example, an ANN can learn to recognize cats by just considering a set of labeled images, but without having any prior knowledge on cats. In that process, deep learning refers to techniques that make the ANN learn a certain task.

As a deep learning based approach to estimate x-ray scattering in real-time is presented in this thesis, the corresponding basics are briefly reviewed in the following. For a more comprehensive introduction to deep learning, the reader is referred to reference [84].

2.3.1 Artifical Neurons

The basic computational unit of an ANN is the artificial neuron that is designed in analogy to its biological counterpart (see figure 2.11). In the biological case, each neuron receives input signals from its dendrites. If the sum of all input signals exceeds a certain threshold, the neuron generates an action potential that is transmitted along its axons, which are connected via synapses to the dendrites of other neurons.

A similar operation is performed by the artificial neuron as depicted in figure 2.11. Here, the artificial neuron receives a given number of inputs $\{a_i\}$ as well as a bias b. Each of the inputs is assigned a weight w_i ,which might be interpreted as synaptic strength or relative importance of the associated input. To generate the output, the neuron applies an activation function σ to the weighted sum of the inputs and the bias. Here, the purpose of the activation function is to introduce a non-linearity in the output, which is necessary as most problems to be solved are non-linear. Practically, there is a variety of different activation functions that are used in artificial neurons or ANNs, respectively (see figure 2.12). While the first artificial models used simple threshold
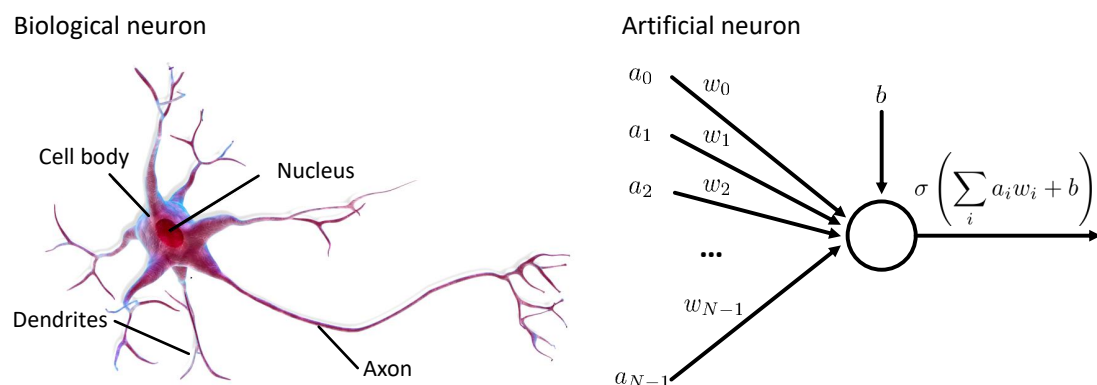
Figure 2.11: Structure of biological and artificial neurons.
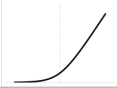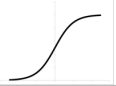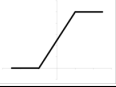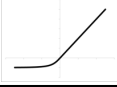
Function	Equation	Plot	Function	Equation	Plot
Identity	$f(x) = x$		Softplus	$f(x) = \log(1 + \exp x)$	
Sigmoid	$f(x) = \dfrac{1}{1 + e^{-x}}$		ReLU	$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	
Hard sigmoid	$f(x) = \begin{cases} 0 & \text{for } x < -\alpha \\ \frac{\alpha + x}{2\alpha} & \text{for } -\alpha \leq x < \alpha \\ 1 & \text{for } x \geq \alpha \end{cases}$		Leaky ReLU	$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	
Tanh	$f(x) = \dfrac{2}{1 + e^{-2x}} - 1$		ELU	$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	
Softsign	$f(x) = \dfrac{x}{1 + |x|}$		Inverse square root LU	$f(x) = \begin{cases} \frac{x}{\sqrt{1 + \alpha x^2}} & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	

Figure 2.12: Common activation functions of artificial neurons. Note that the identity is usually only used in the output layer.

functions, equivalent to the Heaviside step function, more continuous functions such as sigmoids became popular in the 1980s. However, recent works mostly rely on rectified linear unit (ReLU) activation functions or variants of it (see figure 2.12) [85].

2.3.2 Feedforward Artificial Neural Networks

Feedforward ANNs, also referred to as multilayer perceptrons (MLPs), were one of the first and most simple types of ANNs. Therefore, they are briefly discussed in the following to explain some of the basic properties of neural networks.

i.) Architecture and Nomenclature

The basic structure of a feedforward ANN as well as the corresponding nomenclature is depicted in figure 2.13. It consists of a set of artificial neurons (see section 2.3.1) that are arranged in layers. In contrast to other types of ANNs, these layers are connected in such a way that information is propagated only in one direction and only between neighboring layers.

The first layer, receiving the external input to the ANN, is referred to as input layer. It does not perform any computation but only passes the input to the next layer. The output of the ANN can be accessed via its last layer, the so-called output layer. Any layer in between, i.e. layers that do not interact with the external environment, are referred to as hidden layers. While feedforward ANNs only have one input and one output layer, they may have multiple hidden layers.

Most commonly, the layers of feedforward ANNs are designed to be fully connected, i.e. any neuron of a certain layer receives an input from all neurons of the previous layer. As described in section 2.3.1, the output of the receiving neuron is calculated as the
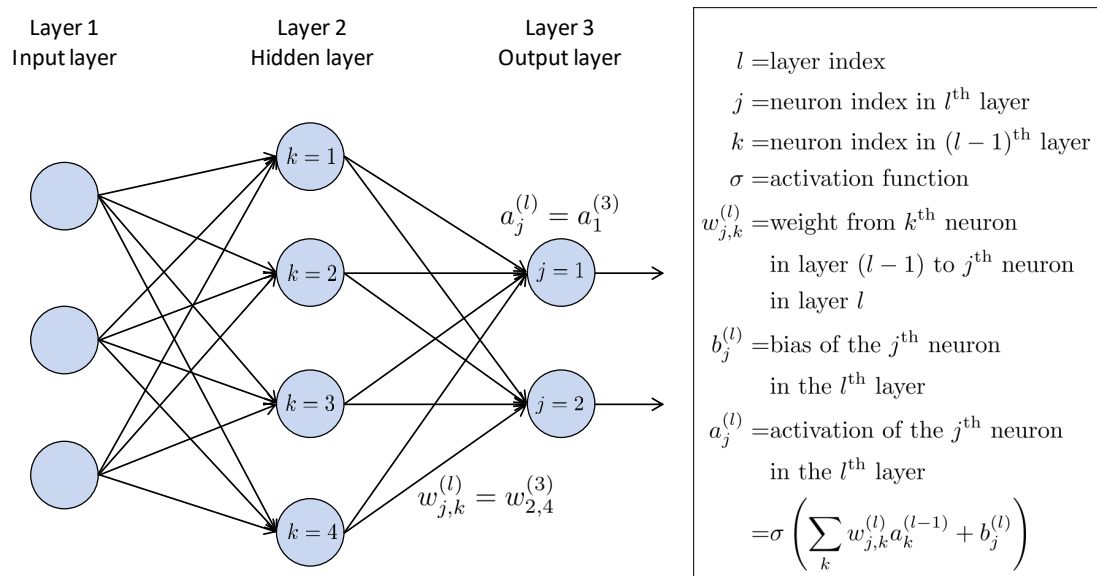
Figure 2.13: Basic architecture and nomenclature of feedforward ANNs. It has to be noted that these networks are not restricted to a single hidden layer but may have multiple hidden layers.

value of its activation function applied to the weighted sum of all inputs and the bias. Here, all weights and all biases are trainable parameters that are optimized such that the network performs a predefined task. The corresponding training procedure will be discussed in more detail in section 2.3.3.

ii.) Feedforward Computation

Given an input vector $\boldsymbol{x}$, the output of a feedforward ANN can be calculated by sequentially applying the mapping of each layer. Considering the example given in figure 2.13, the network might perform the mapping $f : \mathbb{R}^3 \to \mathbb{R}^2$, which is given as:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{a}^{(3)}(\boldsymbol{a}^{(2)}(\boldsymbol{a}^{(1)}(\boldsymbol{x}))), \tag{2.107}$$

where $\boldsymbol{a}^{(i)}$ denotes the output activation of the i-th layer. As the first layer is the input layer, its activation equals the identity mapping, i.e. $\boldsymbol{a}^{(1)}(\boldsymbol{x}) = \boldsymbol{x}$. Denoting the weight of the kth neuron in the first layer to the jth neuron in the second layer as $w_{j,k}^{(2)}$ and the corresponding bias of the jth neuron as $b_j^{(2)}$, the activation of the jth neuron is given as:

$$a_j^{(2)} = \sigma \left(\sum_k w_{j,k}^{(2)} a_k^{(1)} + b_j^{(2)} \right). \tag{2.108}$$

Arranging the weights in a matrix $\boldsymbol{W}^{(2)}$ and the biases in a vector $\boldsymbol{b}^{(2)}$, the calculation of the activation vector $\boldsymbol{a}^{(2)}$ is given as:

$$
\begin{aligned}
\boldsymbol{a}^{(2)} &= \boldsymbol{\sigma}\left(W^{(2)}\boldsymbol{a}^{(1)} + \boldsymbol{b}^{(2)}\right) \\
&= \boldsymbol{\sigma}\left(
\begin{bmatrix}
w_{1,1}^{(2)} & w_{1,2}^{(2)} & w_{1,3}^{(2)} \\
w_{2,1}^{(2)} & w_{2,2}^{(2)} & w_{2,3}^{(2)} \\
w_{3,1}^{(2)} & w_{3,2}^{(2)} & w_{3,3}^{(2)} \\
w_{4,1}^{(2)} & w_{4,2}^{(2)} & w_{4,3}^{(2)}
\end{bmatrix}
\cdot
\begin{bmatrix}
a_1^{(1)} \\
a_2^{(1)} \\
a_3^{(1)}
\end{bmatrix}
+
\begin{bmatrix}
b_1^{(2)} \\
b_2^{(2)} \\
b_3^{(2)} \\
b_4^{(2)}
\end{bmatrix}
\right),
\end{aligned}
\tag{2.109}
$$

where, $\boldsymbol{\sigma}(\boldsymbol{x})$ applies the activation to each element of the vector. Similarly, the activation of the third layer, or the output respectively, can be calculated as:

$$
\begin{aligned}
\boldsymbol{a}^{(3)} &= \boldsymbol{\sigma}\left(W^{(3)}\boldsymbol{a}^{(2)} + \boldsymbol{b}^{(3)}\right) \\
&= \boldsymbol{\sigma}\left(
\begin{bmatrix}
w_{1,1}^{(3)} & w_{1,2}^{(3)} & w_{1,3}^{(3)} & w_{1,4}^{(3)} \\
w_{2,1}^{(3)} & w_{2,2}^{(3)} & w_{2,3}^{(3)} & w_{2,4}^{(3)}
\end{bmatrix}
\cdot
\begin{bmatrix}
a_1^{(2)} \\
a_2^{(2)} \\
a_3^{(2)} \\
a_4^{(2)}
\end{bmatrix}
+
\begin{bmatrix}
b_1^{(3)} \\
b_2^{(3)}
\end{bmatrix}
\right)
\end{aligned}
\tag{2.110}
$$

Therefore, the mapping performed by a feedforward network can be implemented as repeated matrix multiplications interwoven with activation functions. It has to be noted that a set of input vectors can be efficiently processed in parallel by rearranging them into an input matrix $\boldsymbol{X}$ where each column represents one input vector.

iii.) Representational Power

ANNs can be viewed as mathematical models to approximate some unknown function $\tilde{f} : X \rightarrow Y$. This leads to the question of the representational power of ANNs, or more specifically, whether there is any function that cannot be modeled using an ANN. The universal approximation theorem answers this question [86]. It states that a feedforward ANN with at least one hidden layer, a suitable non-linear activation function (e.g. a sigmoid activation), and a linear output layer can approximate any Borel measurable function with any desired accuracy.

A rather intuitive explanation of the universal approximation theorem can be derived as described in reference [87]. It considers the network depicted in figure 2.14 that consists of one input neuron, two hidden neurons with a sigmoid activation function $\sigma(z) = (1 + e^{-z})^{-1}$ and one output neuron with a linear activation function. Given an input x, the activation of the first neuron in the hidden layer can be calculated as:

$$
a_1^{(2)}(x, w_{1,1}^{(2)}, b_1^{(2)}) = \frac{1}{1 + e^{-w_{1,1}^{(2)}x + b_1^{(2)}}}.
\tag{2.111}
$$

If the weight is set to a value $w_{1,1}^{(2)} \gg 1$, the activation approximates the Heaviside step function Θ:

$$
a_1^{(2)}(x, w_{1,1}^{(2)}, b_1^{(2)}) \overset{w_{1,1}^{(2)} \gg 1}{\approx} \Theta(x + \frac{b_1^{(2)}}{w_{1,1}^{(2)}}).
\tag{2.112}
$$

41
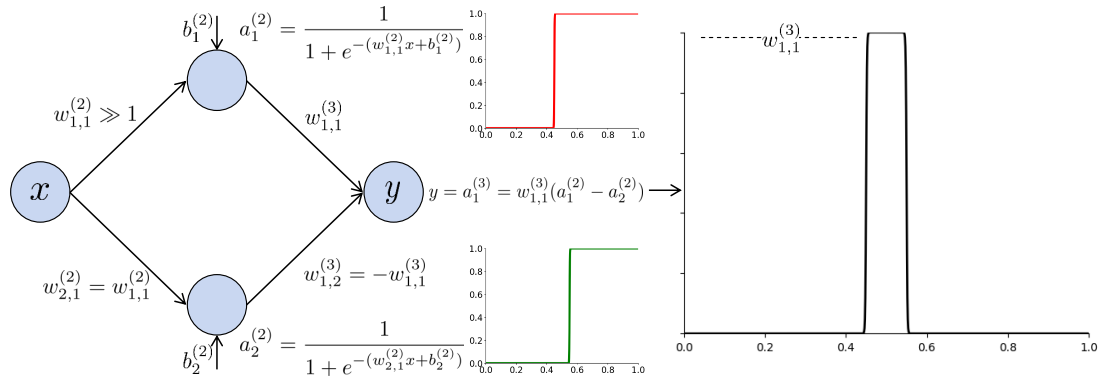
Figure 2.14: Approximation of a rectangular function to using an ANN with two hidden neurons and a sigmoid activation function.

Similarly, the activation of the second neuron $(w_{2,1}^{(2)} = w_{1,1}^{(2)})$ in the hidden layer can be approximated as:

$$a_2^{(2)}(x, w_{2,1}^{(2)}, b_2^{(2)}) \overset{w_{2,1}^{(2)} \gg 1}{\approx} \Theta(x + \frac{b_2^{(2)}}{w_{1,1}^{(2)}}). \tag{2.113}$$

Setting $b_1^{(2)} > b_2^{(2)}$, and the weights of the output layer to $w_{1,1}^{(3)} = -w_{1,2}^{(3)}$, the activation of the linear output layer approximates a boxcar function:

$$y = a_1^{(3)} = w_{1,1}^{(3)}(a_1^{(2)} - a_2^{(2)}) \approx w_{1,1}^{(3)} \, \mathrm{II}\left(\frac{x - \frac{1}{2}\left(\frac{b_1^{(2)} - b_2^{(2)}}{w_{1,1}^{(2)}} \right)}{\frac{b_1^{(2)} - b_2^{(2)}}{w_{1,1}^{(2)}}} \right), \tag{2.114}$$

where $\mathrm{II}(x)$ is the rectangular function:

$$\mathrm{II}(x) = \begin{cases} 0, & \text{if } |x| > \frac{1}{2} \\ 1, & \text{if } |x| \leq \frac{1}{2}. \end{cases} \tag{2.115}$$

Additional, arbitrarily shaped boxcar functions can be constructed in the same way by adding pairs of neurons to the hidden layer. Thus, by a suitable choice of the weights and biases, any function $f(x)$ can be approximated as the sum of these boxcar functions. Similarly, this explanation can be extended to higher dimensional functions.

However, it has to be noted that the universal approximation theorem is practically a rather weak statement. Considering the 1D example given above, the sum of boxcar functions could be directly used as an universal approximator, but no one would suggest to use them as the basis for a machine learning algorithm. Furthermore, empirical observations show that deeper networks perform better than single-hidden-layer networks, despite their representational power is equal.

2.3.3 Training of Artificial Neural Networks

As mentioned in the previous section, ANNs aim to approximate an unknown function $\tilde{f} : X \to Y$, e.g. a classifier that maps a given input $\boldsymbol{x}$ to a certain class y. To do so, a neural network defines the mapping $\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\theta})$, with $\boldsymbol{x} \in X$, and optimizes the parameter vector $\boldsymbol{\theta}$ (usually the weights and the biases) such that $f(\boldsymbol{x}, \boldsymbol{\theta})$ is a suitable approximation to $\tilde{f}$. Here, the process of finding (or learning) the optimal parameter set is referred to as training. The corresponding basics are briefly discussed in the following.

i.) Training Data and Testing Data

ANNs are designed to learn the function $\tilde{f}$ using a set of representative examples which is denoted as training data set. Here, it is distinguished between supervised and unsupervised learning. The former requires the training data set to be labeled, i.e. every training example to be a pair $(\boldsymbol{x}^{(i)}, \boldsymbol{y}_{\text{true}}^{(i)})$ of an input $\boldsymbol{x}^{(i)} \in X$ and the corresponding ground truth $\boldsymbol{y}_{\text{true}}^{(i)} = \tilde{f}(\boldsymbol{x}^{(i)})$. Unsupervised learning, in contrast, refers to learning in absence of a ground truth. In that case the training data set only consists of examples $\boldsymbol{x}^{(i)} \in X$. Thus, there are infinitely many possible solutions which need to be constrained in a some way to reasonable outputs. This can, for instance, be achieved by an appropriate choice of the loss function (see next section).

Since the training data set $X_{\text{train}} = \{\boldsymbol{x}^{(i)}\}$ is only a proper subset of the input space X, the training rather leads to an approximation of the function $h : X_{\text{train}} \to Y$. Therefore, the performance of an ANN needs to be evaluated for a testing data set $X_{\text{test}} = \{\boldsymbol{x}^{(j)}\}$ with $\boldsymbol{x}^{(j)} \in X$ and $\boldsymbol{x}^{(j)} \notin X_{\text{train}}$. Although this does not proof that the ANN applies in a similar way to any other unknown input, a good performance on the testing data gives at least some confidence that a suitable mapping has been learned.

ii.) Loss Function

Similar to other optimization problems, the parameter vector $\boldsymbol{\theta}$ of a neural network is determined by minimizing (or maximizing) an objective function, also referred to as loss function L. Here, the loss function provides a qualitative measure, rating the discrepancy between the mapping $f(\boldsymbol{x}, \boldsymbol{\theta})$ defined by the neural network and the unknown function $\tilde{f}$ to be approximated.

Considering supervised learning, the loss function usually compares the output of the ANN and corresponding ground truth (or their statistics) according to a similarity metric. The appropriate choice of the loss function depends on the problem to be solved. In case of a regression problem, for instance, the mean squared error (MSE) loss is a common choice.

$$L_{\text{MSE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=0}^{N-1} L_{\text{MSE}}^{(i)}(\boldsymbol{x}^{(i)}, \boldsymbol{y}_{\text{true}}^{(i)}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i=0}^{N-1} \| f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}) - \boldsymbol{y}_{\text{true}}^{(i)} \|_2^2, \qquad (2.116)$$

with $L^{(i)}$ being the per-example loss. Classification problems, in contrast, rather rely

on probability based loss functions such as the cross entropy (CE) loss:

$$L_{\mathrm{CE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=0}^{N-1} L_{\mathrm{CE}}^{(i)}(\boldsymbol{x}^{(i)}, \boldsymbol{y}_{\mathrm{true}}^{(i)}, \boldsymbol{\theta}) = \sum_{i=0}^{N-1} \sum_{j=0}^{J-1} \boldsymbol{y}_{j,\mathrm{true}}^{(i)} \ln f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta})_j. \qquad (2.117)$$

It has to be noted that the activation of the output layer needs to fit the requirements of the loss function. e.g. using the CE loss function requires the output to be interpreted as probability distribution, i.e. $\sum_j f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta})_j \overset{!}{=} 1$.

Considering unsupervised learning, there is no ground truth $\boldsymbol{y}_{\mathrm{true}}^{(i)}$ available. Thus, the loss function is rather based on some sort of prior knowledge on the target distribution Y. For instance Y may be the distribution of smooth CT images, while the input distribution X may be the distribution of noisy CT images. Thus, the loss function might look as follows:

$$L(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=0}^{N-1} L^{(i)}(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}) = \sum_{i=0}^{N-1} \|f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}) - \boldsymbol{x}^{(i)}\|^2 + \lambda \|\nabla f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta})\|^2, \quad (2.118)$$

where the latter term enforces the smoothness, while first term restricts the solution to outputs that are similar to the input, i.e. for that particular example to outputs that look like CT images. Therefore, ANNs that are based on this principle are referred to as autoencoders. It has to be noted there are several other approaches to unsupervised learning. A more detailed discussion can be found in reference [84].

iii.) Gradient-Based Learning

Most commonly, the parameters of an ANN are learned using a gradient decent approach [88] or a variant of it. It can be motivated by the following relation that holds true for any differentiable function $f(\boldsymbol{a})$ and a small but positive constant η:

$$f(\boldsymbol{a} - \eta \nabla f(\boldsymbol{a})) \leq f(\boldsymbol{a}). \qquad (2.119)$$

I.e. moving the argument of a differentiable function along the direction of its negative gradient will decrease its value. Thus, starting with an initial guess $\boldsymbol{a}_0$, the following sequence

$$\boldsymbol{a}_{n+1} = \boldsymbol{a}_n - \eta \nabla_{\boldsymbol{a}} f(\boldsymbol{a}_n) \qquad (2.120)$$

converges to a local minimum of $f(\boldsymbol{a})$, at least for particular choices of η. As η controls the speed of convergence, it is also referred to as the learning rate.

Given a training data set $\{(\boldsymbol{x}_i, y_{\mathrm{true}}^{(i)})\}$, an ANN defining the mapping $\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\theta})$, and a differentiable loss function $L(\boldsymbol{\theta})$, the gradient descent approach yields the following update scheme to find the parameter vector that locally minimizes the loss:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \eta \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_n). \qquad (2.121)$$

It has to be noted that due to the nonlinearities of the ANN, the loss function is nonconvex with respect to the parameters $\boldsymbol{\theta}$. Consequently, the gradient descent will

most like converge to a local minimum. Therefore, it is important to use a proper initialization. Furthermore, increasing the size of the training data set potentially helps converging to a local minimum closer to the global optimum.

Stochastic Gradient Descent and Mini-Batch Gradient Descent   Calculating a parameter update according to equation (2.121) requires to evaluate the gradient for any example $\boldsymbol{x}^{(i)} \in X_{\text{train}}$ of the training data set:

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \frac{1}{N} \sum_i^N \nabla_{\boldsymbol{\theta}} L^{(i)}(\boldsymbol{x}^{(i)}, \boldsymbol{y}_{\text{true}}^{(i)}, \boldsymbol{\theta}). \tag{2.122}$$

Consequently, the calculation of the update step is computationally expensive, especially in case of large training data sets. Therefore, the gradient $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$ is typically approximated by evaluating only a subset $X_{B,\text{train}} = \{\boldsymbol{x}^{(b)}\}$ with $b \leq B < N$ of the training data set:

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) \approx \frac{1}{B} \sum_b^B \nabla_{\boldsymbol{\theta}} L^{(b)}(\boldsymbol{x}^{(b)}, \boldsymbol{y}_{\text{true}}^{(b)}, \boldsymbol{\theta}). \tag{2.123}$$

If the subset contains only a single element i.e. $B = 1$, this variant is referred to as stochastic gradient descent while for any other $B < N$ it is referred to as mini-batch gradient descent and B is referred to as batch size.

Momentum   Momentum is a concept that has been introduced to accelerate learning in case of small but consistent gradients, high curvature, or noisy gradients [89]. It accumulates the past gradients by an exponentially decaying average and moves within the direction of that average. Therefore, it defines the velocity $\boldsymbol{v}$ that is updated as:

$$\boldsymbol{v}_{n+1} = \alpha \boldsymbol{v}_n + \eta \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_n). \tag{2.124}$$

with α being the decay constant. The corresponding update of the parameter vector including the momentum concept is given as:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \boldsymbol{v}_{n+1}. \tag{2.125}$$

Algorithms with Adaptive Learning Rates   The learning rate η is one of the most important parameters to set in ANNs as it strongly affects the performance of the model. Setting the learning rate too small leads to slow convergence while setting the learning rate too high may lead to no convergence at all. Furthermore, it is beneficial to adapt the learning rate of each parameter of the network individually. At first, this issue was addressed by the AdaGrad algorithm [90]. It accumulates the the squared values of the gradient in every update step:

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_n) \odot \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_n), \tag{2.126}$$

with $\odot$ being the Hadamard product, and scales the learning rate according to the inverse square root. Thus, the parameter update is given as:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \frac{\eta}{\sqrt{\boldsymbol{r}_{n+1}} + \epsilon} \odot \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_n), \tag{2.127}$$

with ϵ being a small constant to prevent a division by 0. It has to be noted that the square root and the division are applied element-wise. In that way, the learning rate of parameters with a high partial derivative decreases more rapidly, pushing the update towards more gently sloped directions of the parameter space.

Other variants of adaptive learning rate algorithms such as RMSProp [91], or Adam [92] rely on a similar principle, i.e. changing the learning rate individually according to some function of the accumulated gradient.

iv.) Backpropagation

Gradient-based learning requires to evaluate the gradient of the loss function with respect to the network's parameters. A straight forward way to compute the gradient or the partial derivative of the ith parameter θ_i, respectively, would be to evaluate the difference quotient:

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i} \approx \frac{L(\boldsymbol{\theta} + \epsilon \cdot \hat{\boldsymbol{e}}_i) - L(\boldsymbol{\theta})}{\epsilon}, \tag{2.128}$$

where $\hat{\boldsymbol{e}}_i$ is the unit vector corresponding to the entry θ_i. This, however, requires to calculate $N + 1$ passes through the network to calculate the partial derivatives of N parameters. Since ANNs typically have up to several million parameters, such an approach is rather impractical.

An efficient way to calculate the gradients is provided by the so-called backpropagation algorithm that has been proposed in the 1970s, but has not emerged in the field of ANNs until the famous publication of Rumelhart et al. in 1986 [93].

Defining the the error $\delta_j^{(l)}$ of jth neuron in the lth layer with respect to the loss function L as:

$$\delta_j^{(l)} \equiv \frac{\partial L}{\partial z_j^{(l)}}, \tag{2.129}$$

with $z_j^{(l)}$ being the output of the corresponding neuron prior to the activation $a_j^{(l)} = \sigma(z_j^{(l)})$:

$$z_j^{(l)} = \sum_k w_{j,k}^{(l)} a_k^{(l-1)} + b_j^{(l)} = \sum_k w_{j,k}^{(l)} \sigma(z_k^{(l-1)}) + b_j^{(l)}, \tag{2.130}$$

the backpropagation algorithm calculates the gradients for a given example $\boldsymbol{x}^{(i)} \in X_{\text{train}}$ using the following steps.

1. **Input:** The input activation $\boldsymbol{a}^{(1)}$ is set to $\boldsymbol{x}^{(i)}$.

2. **Feedforward:** Calculation of $z_j^{(l)}$ for every neuron j in every layer $l \in \{2, \ldots, L\}$ by a forward pass through the network (see section 2.3.2):

$$\boldsymbol{z}^{(l)} = \boldsymbol{W}^{(l)} \boldsymbol{a}^{(l-1)} + \boldsymbol{b}^{(l)} \text{ with } \boldsymbol{a}^{(l)} = \sigma(\boldsymbol{z}^{(l)}). \tag{2.131}$$

3. **Output error:** Calculation of the error of the output layer (note that $\nabla_a L$ refers to the gradient of the loss function and must not be confused with the index of the output layer):

$$\delta_j^{(L)} = \frac{\partial L}{\partial z_j^{(L)}} = \frac{\partial L}{\partial a_j^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} = \frac{\partial L}{\partial a_j^{(L)}} \sigma'(z_j^{(L)}) \tag{2.132}$$

$$\Leftrightarrow \boldsymbol{\delta}^{(L)} = \nabla_a L \odot \sigma'(\boldsymbol{z}^{(L)}) \tag{2.133}$$

4. **Backpropagation of the error:** The error of any previos layer l is calculated recursively by:

$$\delta_j^{(l)} = \frac{\partial L}{\partial z_j^{(l)}} = \sum_k \frac{\partial L}{\partial z_k^{(l+1)}} \frac{\partial z_k^{(l+1)}}{\partial z_j^{(l)}} = \sum_k \frac{\partial z_k^{(l+1)}}{\partial z_j^{(l)}} \delta_k^{(l+1)} \tag{2.134}$$

According to equation (2.130), the partial derivative can be evaluated as:

$$\frac{\partial z_k^{(l+1)}}{\partial z_j^{(l)}} = \frac{\partial}{\partial z_j^{(l)}} \sum_j w_{k,j}^{(l+1)} \sigma(z_j^{(l)}) + b_k^{(l+1)} = w_{k,j}^{(l+1)} \sigma'(z_j^{(l)}) \tag{2.135}$$

Substituting equation (2.135) back into equation (2.134) yields:

$$\delta_j^{(l)} = \sum_k w_{k,j}^{(l+1)} \delta_k^{(l+1)} \sigma'(z_j^{(l)}) \tag{2.136}$$

$$\Leftrightarrow \boldsymbol{\delta}^{(l)} = ((\boldsymbol{W}^{(l+1)})^T \boldsymbol{\delta}^{(l+1)}) \odot \sigma'(\boldsymbol{z}^{(l)}). \tag{2.137}$$

5. **Output:** The gradients with respect to the weights and the biases are given as:

$$\frac{\partial L}{\partial w_{j,k}^l} = \frac{\partial L}{\partial z_k^{(l)}} \frac{\partial z_k^{(l)}}{\partial w_{j,k}^{(l)}} = a_k^{(l-1)} \delta_j^{(l)} \tag{2.138}$$

$$\frac{\partial L}{\partial b_j^l} = \frac{\partial L}{\partial z_k^{(l)}} \frac{\partial z_k^{(l)}}{\partial b_j^{(l)}} = \delta_j^{(l)} \tag{2.139}$$

Thus, the backpropagation approach allows to calculate all gradients by a one forward and one backward pass through the network.

v.)   Hyperparameters and Validation Data

Any open parameter of an ANN that is not optimized during the training process is referred to as hyperparameter. Hyperparameters include the parameters of the optimization algorithm, such as the learning rate η, but may also be used to define the layout of the ANN, e.g. the number of layers or the number of artificial neurons per layer. Since the particular choice of the hyperparameters strongly affects the performance of an ANN, several effort is made to tune the hyper parameters (see e.g. [84]). The concept of hyperparameter tuning typically involves the use of a so-called validation data set. It consists of a set of unknown examples that is not used to learn the weights and biases, but differs from the testing data set. Now, the optimal realization of the hyperparameters is considered to be the one that minimizes the loss on the validation data. To have an unbiased estimator of the ANN's performance, it still needs to be evaluated for the testing data.

vi.)   Generalization, Underfitting and Overfitting

The major challenge of deep learning is to find a suitable approximation of an unknown function that applies not only to examples of the training data set, but also to unknown examples. This property is referred to as generalization and corresponds to a small loss on the training data set as well as on the testing data set. In this context it is distinguished between underfitting and overfitting as depicted in figure 2.15. Here, underfitting occurs if the representational power of the model is too small. In that case, a small loss cannot even be achieved on the training data set. Overfitting, in contrast, corresponds to having a small training loss but considerably higher errors on the testing data set. This might happen if the model does not fit the task to be learned. One may think of it as fitting a higher order polynomial to data that have been sampled a according to a second order polynomial, for instance. Another common cause of overfitting is an inappropriate composition of the training data set, e.g. if the function to be approximated is defined on the interval $[0, 1]$ but the training data set only represents the interval $[0, 0.5]$.
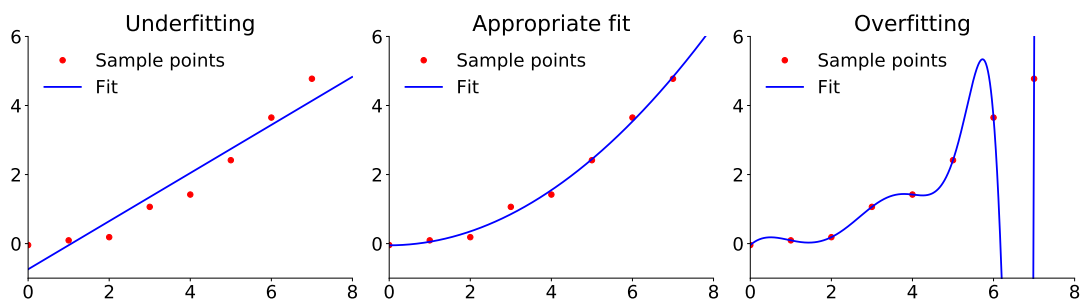
Figure 2.15: Example of underfitting, apropriate fitting and overfitting. Here, the red points correspond to the training examples while the blue curve corresponds to the function learned during the training phase.

While the problem of underfitting can be tackled more easily by increasing the complexity of the ANN (i.e. increasing the number of layers or the number of artificial neurons per layer), it is more difficult to prevent the ANN from overfitting. Some of the most common strategies to do so are the following.

Training Data and Data Augmentation Most obviously, the problem of overfitting can be tackled by increasing the size of the training data set, and thereby, reducing the number of possible solutions leading to a small loss. However, practically it is often difficult to acquire a huge amount of training data. Therefore, the existing training data are often altered artificially, e.g. by applying some sort of transformation, to generate additional data. This strategy is referred to as data augmentation.

Penalizing Large Weights Penalizing large weights is another strategy to prevent overfitting. This is motivated by the fact that a large weight make the output strongly depend on the input corresponding to that weight. Practically, this is usually realized by adding a penalty such as the l1-norm or the l2-norm of the weight matrix to the loss function.

Early Stopping When training ANNs, it is often observed that the training loss steadily decreases for each parameter update. Evaluating the loss after each update for unknown data, however, shows that it starts to increase at some point. Thus, a simple approach to prevent overfitting is to use a validation data set and to stop the training once the validation loss starts increasing.

Dropout Dropout provides a computational inexpensive approach to regularize a broad class of ANN models [94]. It is inspired by so-called bagging approaches that average (in the most simple case) the output of multiple ANNs that have been trained to learn the same task. Here, the averaging process helps reduce the error of individual ANNs, and therefore, to reduce overfitting. Since training several ANNs is computationally expensive, the basic ideal of dropout is to drop individual neurons out of the network in every epoch of the training phase. For instance, this can be realized by setting the activation the network's neurons to zero with a probability $1 - P$. Thus, dropout can be seen as training an ensemble of subnetworks at the same time. Once the network is trained, their average is approximated by using the full network with each neuron's output weighted by a factor of P.

2.3.4 Convolutional Neural Networks

Convolutional neural networks represent a class of ANNs that are designed to process data with a grid-like topology such as 2D images or 3D volumes for instance [95]. The architecture of convolutional neural networks is inspired by the organization of neurons in the animal visual cortex [96]. Here, cortical neurons only respond to stimuli that are received in a particular region (receptive field) of the visual field. In that process two different types of cells were found to be involved. Simple cells that have a maximum

activation when patterns with a particular orientation are present in their receptive field, and complex cells that respond similar to simple cells but have a larger receptive field and a lower sensitivity to the exact position of the pattern [97]. Convolutional neural networks realize a similar behavior by their two basic building blocks depicted in figure 2.16: convolution layers and pooling layers.

i.) Convolution Layers

In principle, convolution layers are very similar to the ones introduced in section 2.3.2: they are made up of artificial neurons with learnable weights and biases. Each neuron calculates the dot product of its inputs and its weights, adds the bias, and applies an activation function. Thus, a convolution layer also represents a differentiable function that can be trained, as described in section 2.3.3, i.e. by minimizing a loss function using a gradient-based approach and backpropagation.

However, in contrast to fully connected layers, convolution layers arrange their neurons in a grid-like structure and restrict the receptive field of a single neuron to only a small patch of the previous layer (see figure 2.16). Here, each layer uses shared weights, i.e. any neuron within a particular layer uses the same weights to calculate its output D. Interpreting the weights to be entries of a kernel K with a size equal to the receptive field of the neuron, the operation applied to a given 2D input S can be expressed in terms of a convolution:

$$D_{i,j} = \sigma\left((S * K)_{i,j} + b\right) = \sigma\left(\sum_{i',j'} S_{i-i',j-j'} K_{i',j'} + b\right), \tag{2.140}$$

where i and j denote the pixel indices, and σ and b refer to the activation function and the bias, respectively. It has to be noted that convolutional layers are not restricted to 2D data but process 1D or higher dimensional data with a grid-like structure in a similar same way.

The output D is also referred to as feature map as it indicates whether and where a certain feature, encoded by the kernel K, is present in the input. Rather than extracting just a single feature, a convolution layer typically uses a certain number of G kernels to extract G different features simultaneously. This adds an additional feature dimension to every layer $D_{i,j} \rightarrow D_{i,j,g}$. Similarly, the kernel receives a third dimension once a an input with f features is processed $K_{i,j} \rightarrow K_{i,j,f}$. In that case the output is given as:

$$D_{i,j,g} = \sigma\left(\sum_f (S_f * K_f^{(g)})_{i,j} + b^{(g)}\right) = \sigma\left(\sum_f \sum_{i',j'} S_{i-i',j-j',f} K_{i,j,f}^{(g)} + b^{(g)}\right). \tag{2.141}$$

Here, the summation over f is required to keep the dimension of the feature map. Otherwise, every convolution layer would add an addition dimension to the feature map.

Stacking several convolution layers in that way allows the network to extract high-level features from a given input. Since the weights of the kernels are learnable parameters, a convolutional neural network does not rely on handcrafted features but is trained to find meaningful features itself during training phase.
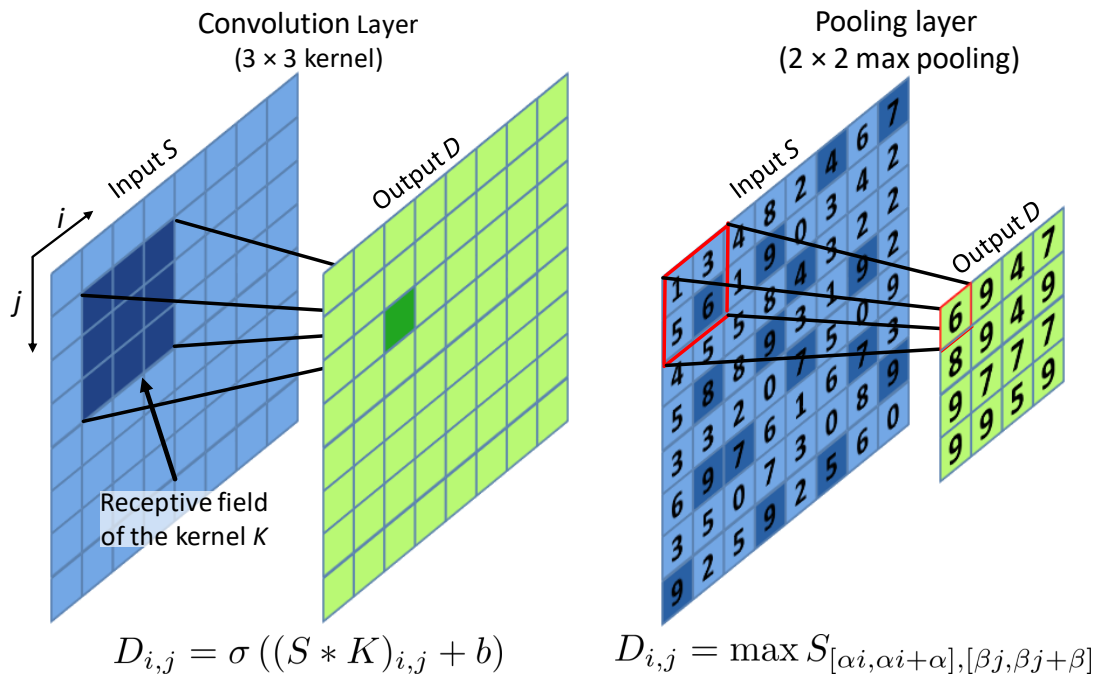
Figure 2.16: Basic building blocks of convolutional neural networks.

ii.) Pooling Layers

Pooling layers are typically used in convolutional neural networks in between successive convolution layers. Here, the objective of the pooling layer is to reduce the spatial dimensions of the input while extracting the dominant features. Therefore, every unit of the pooling layer replaces the data within its receptive field by some sort of summary statistic. Most commonly, this is done by applying a 2×2 maximum operation (max pooling, see figure 2.16) or by averaging over non-overlapping 2×2 patches (average pooling). Here, the pooling layer operates independently on every feature map of the input. E.g. a 2×2 pooling of an $I \times J \times F$ input (width $\times$ height $\times$ # features) yields an $I/2 \times J/2 \times F$ output. Since max pooling and average pooling are non-trainable operations some authors prefer to use strided convolutions instead. Therefore, the convolution operation described in the previous section is applied with a stride of s to achieve an $s \times s$ pooling. Thus, it is potentially possible to learn the most suitable summary statistic during training phase rather than using a predefined operation.

3 | Simulation of the CT Data Acquisition

Precise models describing the CT data acquisition process are required in several areas of CT imaging such as iterative reconstruction [21], [22], [98], [99], protocol optimization [100] or artifact correction [24], [101], [102]. The latter is also the focus of this thesis. Here, the most common CT artifacts caused by x-ray scattering, beam hardening, partial volume effects and off-focal radiation shall be corrected using a simulation-based correction approach. Most accurately, these effects could be modeled using MC methods. However, for the sake of performance it is advantageous to use analytic expressions whenever it is possible. This sections sets up the corresponding physical models and validates them by a comparison to MC simulations or reference measurements.

3.1 Material and Methods

3.1.1 Geometry and Object Specification

The basic setup of an industrial cone-beam CT (CBCT) system is depicted in figure 3.1. The geometry of a certain view angle α is completely described by the four vectors $\boldsymbol{s}(\alpha), \boldsymbol{o}(\alpha), \boldsymbol{u}(\alpha)$ and $\boldsymbol{v}(\alpha)$. Here, the vector $\boldsymbol{s}$ points from the isocenter, that is defined to be at $(0, 0, 0)$, to the position of the x-ray source. The vector $\boldsymbol{o}$ defines the origin of the x-ray detector and points to the center of the first detector pixel. The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ have a magnitude equal to the pixel spacing and span the detector plane. Thus, the position $\boldsymbol{d}$ of the detector pixel (n_u, n_v) is given as:

$$\boldsymbol{d} = \boldsymbol{o} + n_u \cdot \boldsymbol{u} + n_v \cdot \boldsymbol{v}. \tag{3.1}$$

The object to be simulated is either given as a voxel volume or is defined by its surface in terms of a triangular mesh. The voxel volume is defined on a regular grid with origin $\boldsymbol{o} = (x_0, y_0, z_0)^T$, spacing $\boldsymbol{x} = (dx, 0, 0)^T$, $\boldsymbol{y} = (0, dy, 0)^T$, $\boldsymbol{z} = (0, 0, dz)^T$ and dimensions (N_x, N_y, N_z). Similar to equation (3.1), every voxel (n_x, n_y, n_z) can be related to a physical coordinate $\boldsymbol{r}$ by the following transform:

$$\boldsymbol{r} = \boldsymbol{o} + n_x \cdot \boldsymbol{x} + n_y \cdot \boldsymbol{y} + n_z \cdot \boldsymbol{z}. \tag{3.2}$$

A voxel-based representation has the advantage that every voxel can be assigned a different property e.g. a different density or different attenuation, respectively. A

surface mesh, in contrast, can only be used to describe homogeneous objects but has the advantage of lower memory requirements and higher computational performance, especially when it comes to line integral calculations.
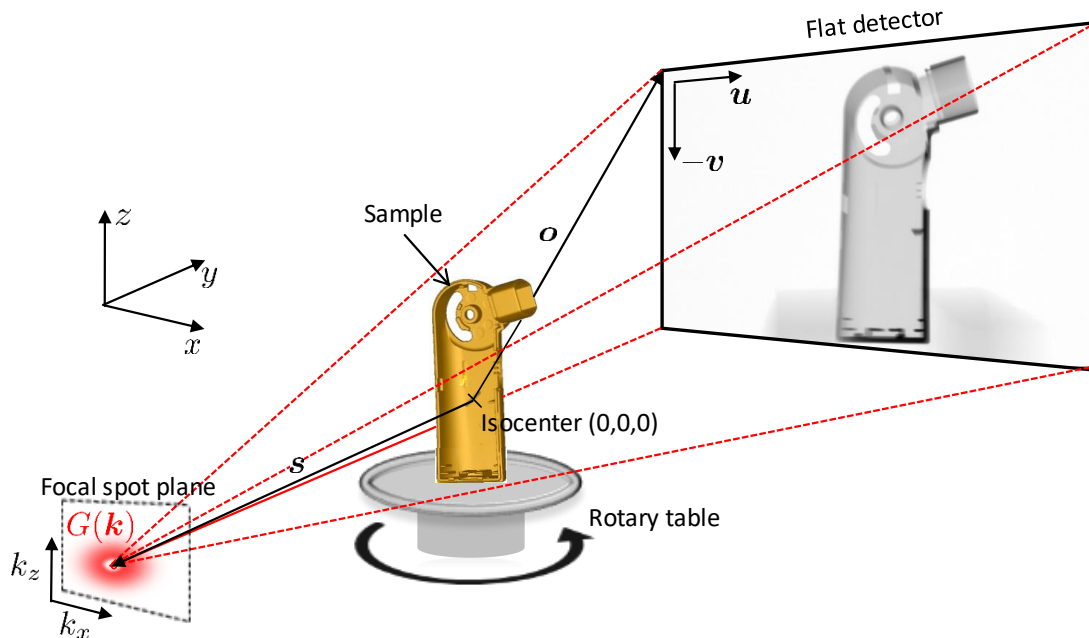
Figure 3.1: Geometry and basic setup of a CBCT system.

3.1.2   X-Ray Spectra of Transmission X-Ray Tubes

The simulation of x-ray emission spectra of transmission x-ray tubes is based on a slight modification of the method proposed by Tucker et al. [103]. In their paper the x-ray spectrum was estimated using semiempirical models approximating the generation of bremsstrahlung and characteristic x-rays. The corresponding derivation of these models is briefly reviewed in the following.

i.)   Bremsstrahlung Model

Considering an electron with kinetic energy T, deflected in the field of a nucleus with charge Ze, the differential cross section for the emission of a photon with an energy between E and $E + dE$ is modeled according to reference [104]:

$$d\sigma_{\text{rad}} = \alpha r_e^2 \, Z^2 \, B \, \frac{T + m_0 c^2}{T} \frac{dE}{E}, \tag{3.3}$$

where α denotes the fine structure constant, r_e the classical electron radius, and B a slowly varying function of Z and T that is determined as described in iii.). For an x-ray target with density ρ and atomic mass A, there are $N = \rho/A$ atoms per unit volume.

Accordingly, the average bremsstrahlung energy, emitted by an electron traversing the distance dx along its incident direction, can be calculated as:

$$dT_{\text{rad}}(E) = \frac{\rho}{A} E \, d\sigma_{\text{rad}} \, dx, \tag{3.4}$$

Penetrating a distance dx, the electron loses the energy dT. The fraction dT_{rad}/dT corresponds to the radiative energy loss. Thus, the bremsstrahlung intensity generated by an electron with an initial energy T_0 can be calculated as:

$$EN(E)dE = \int_E^{T_0} \left(\frac{dT_{\text{rad}}}{dT} \right) dT, \tag{3.5}$$

where $N(E)dE$ refers to the number of photons having an energy between E and dE. Substituting equation (3.4) and (3.3) yields:

$$N(E)dE = \frac{\alpha r_e^2 Z^2}{A} \frac{dE}{E} \int_E^{T_0} B \frac{T + m_0 c^2}{T} \left(\frac{1}{\rho} \frac{dT}{dx} \right)^{-1} dT, \tag{3.6}$$

where $\frac{1}{\rho} \frac{dT}{dx}$ is the mass stopping power that can be derived by an interpolation of tabulated data. Finally, target attenuation $F(x, E)$ of x-rays generated at depth x needs to be considered. Considering a perpendicular incidence of electrons onto a transmission target of thickness D, it is given according to the Beer-Lambert law (see equation (2.6)):

$$F(x, E) = e^{-\mu(E)(D-x)}. \tag{3.7}$$

To use this expression in equation (3.6), the position x needs to be related to the kinetic energy of the electron T. This is done using the Thomson-Whiddington relation [105]:

$$\rho c x = T_0^2 - T^2, \tag{3.8}$$

where c is the empirical Thomson-Whiddington constant. Including the target attenuation, the bremsstrahlung spectrum is given as:

$$N(E)dE = \frac{\alpha r_e^2 Z^2}{A} \frac{dE}{E} \int_E^{T_0} e^{-\mu(E)(D - \frac{T_0^2 - T^2}{\rho c})} B \frac{T + m_0 c^2}{T} \left(\frac{1}{\rho} \frac{dT}{dx} \right)^{-1} dT. \tag{3.9}$$

ii.) Characteristic X-Ray Model

Theoretical and experimental findings indicate that the intensity I_K of characteristic x-rays can be approximated as:

$$I_K = A_K (T_0 - E_K)^{n_K}, \tag{3.10}$$

where, A_K is a proportionality constant, E_K is the k-shell binding energy, and T_0 is the initial electron energy. The exponent n_K typically takes values in the range of 1.6 to 1.7 [106].

Since electrons loose energy while penetrating deeper into the target, there is a decreasing probability for the generation of characteristic x-rays with increasing depth. Tucker et al. modeled this behavior using a parabolic distribution function:

$$P(x) = \frac{3}{2}\left(1 - \frac{x^2}{R^2}\right), \tag{3.11}$$

where R refers to the distance at which the average kinetic energy of electrons equals E_K. Including the target attenuation in a similar way as described in the previous section, the total number of characteristic x-rays can be modeled as:

$$N(E_i) = A_K \left(\frac{T_0}{E_K} - 1\right)^{n_K} f(E_i) \int_0^R P(x) e^{-\mu(E_i)(D-x)} dx, \tag{3.12}$$

Here, the fractional yield $f(E_i)$ of the of the E_i characteristic x-ray is taken from tabulated data [107]. A_K and n_K are model parameters that are determined as described in iii.).

iii.)   Determination of the Open Parameters

In the paper of Tucker et al., the function B in equation (3.9) is parametrized as:

$$B = (A_0 + A_1 T_0)\left(1 + B_1\left(\frac{E}{T}\right) + B_2\left(\frac{E}{T}\right)^2 + B_3\left(\frac{E}{T}\right)^3 + B_4\left(\frac{E}{T}\right)^4\right). \tag{3.13}$$

The open parameters $A_0, A_1, B_1, B_2, B_3, B_4$ as well as the parameters A_K and n_K of the characteristic x-ray model were determined such that the model reproduces measured spectra of a reflection x-ray tube.

In this thesis, this model was supposed to be used to simulate transmission x-ray spectra. Therefore, the open parameters had to be recalibrated to fit this purpose. To do so, Monte Carlo (MC) simulations of transmission x-ray tubes were performed using the Geant4 (v. 10.4) MC code. Simulations were performed for tungsten targets assuming different target thicknesses and different tube voltages. Here, the target thickness was set in steps of 2 µm between 4 µm and 24 µm. The tube voltage was set in steps of 20 kV between 80 kV and 300 kV. Subsequently, the open parameters of the Tucker model were optimized by minimizing the mean squared error between its prediction and the MC spectra.

3.1.3   X-Ray Object Interactions

Interactions of x-ray photons with the object to be measured can be divided into primary interactions and x-ray scattering interactions. While primary interactions can be simulated using analytic expressions, there is no closed-form solution to describe x-ray scattering. Therefore, primary interactions and x-ray scattering are modeled separately. The corresponding models are described in the following.

i.)   Primary Interactions

Primary interactions are described by the Beer-Lambert law according to equation (2.6). Given the x-ray spectrum $w(E)$, the prefilter attenuation $\kappa(E)$, the detector efficiency $\eta(E)$ as well as the distribution of the object's attenuation coefficient $\mu(\boldsymbol{r})$, the normalized primary intensity can be calculated as:

$$\frac{I_p}{I_0}(\boldsymbol{d}) = \frac{\int w(E)\kappa(E)\eta(E)e^{-\int_0^1 \mu(\boldsymbol{s}+\lambda\cdot(\boldsymbol{d}-\boldsymbol{s}),E)d\lambda}dE}{\int w(E)\kappa(E)\eta(E)dE}, \tag{3.14}$$

Here, the x-ray spectrum $w(E)$ and the detector efficiency $\eta(E)$ are modeled as described in section 3.1.2 and section 3.1.5, respectively. The prefilter attenuation is modeled as $\kappa(E) = e^{-\mu_p(E)\cdot l_p}$, where $\mu_p(E)$ is the prefilter material's attenuation coefficient and l_p is the corresponding intersection length. To simulate primary intensities, equation (3.14) is evaluated for the center of every detector pixel with $\boldsymbol{d}$ being calculated according to equation (3.1). To account for partial volume effects, a given number of subsamples may be calculated and averaged in intensity domain for every detector pixel.

ii.)   X-Ray Scattering

Analytic approaches to calculate the distribution of scattered x-rays are typically restricted to single scattering [108]. In order to derive accurate scatter estimates, x-ray scattering is modeled here using MC methods. For the sake of computational performance, the MC scatter simulation is implemented from scratch rather than using an existing MC code such as Geant4 [109] for instance. Since the MC code differs at several points from the straight forward implementation given in section 2.2.3, the basic structure is briefly described in the following.

Input to the MC Simulation   The simulation of x-ray scattering for a given setup requires to specify the acquisition geometry, the object geometry, its composition as well as the underlying physics (see figure 3.2).

Acquisition geometry   Here, the acquisition geometry is specified in terms of the vector $\boldsymbol{s}$ pointing to the position of the x-ray source and a triangle mesh representing the x-ray detector. In case of a typical flat detector with quadratic pixels, each pixel is composed of two triangles or their three vertices, respectively. However, due to the flexibility of triangle meshes the simulation is not restricted to simple detector layouts, but allows to model arbitrarily shaped detectors. Optionally, a preweighting mesh and an anti-scatter grid can be provided. The preweighting mesh allows to model prefilters or shaped filters that are commonly used in clinical CT. Therefore, every triangle is assigned a spectral weight. Once an x-ray photon with energy E_γ intersects with a triangle of the preweighting grid, it is weighted accordingly. Similarly anti-scatter grids, which are commonly used to suppress the contribution of scattered x-ray photons, can be defined.

Object specification   The MC simulation is designed to operate on voxel volumes that are defined according to section 3.1.1. The composition of the object is specified by a material voxel volume that assigns every voxel a certain material. Similarly, a density voxel volume is provided to specify the density of every voxel.

Physics   The simulation of random photon tracks requires to know their initial spectral distribution as well as their material and energy dependent interaction probabilities. The former is passed to the simulation in terms of the (normalized) x-ray spectrum that is evaluated at a given number of discrete points. The interaction probabilities for every material, specified by the material voxel volume, are given in terms of the materials' attenuation coefficients as well as their differential scattering cross sections. Both of them are derived by interpolation of tabulated values given in the EPDL and are passed to the simulation as discrete arrays.
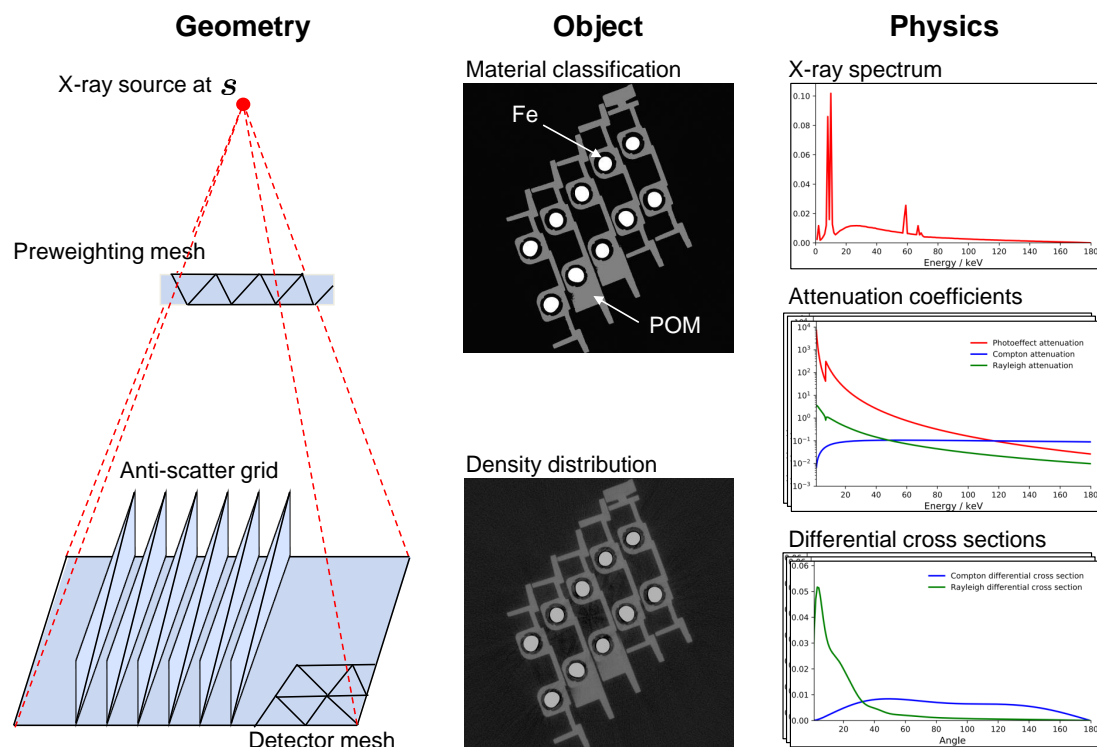
Figure 3.2: Input to the MC simulation. The acquisition geometry is specified in terms of a source vector $\boldsymbol{s}$ and a detector triangle mesh such that arbitrary shapes can be realized. Optionally, a preweighting mesh and an anti-scatter grid can be specified as a triangle mesh. The object information is provided in terms of a voxel volumes that assign ever voxel a certain material as well as a density. The physics are given by the spectral distribution of incident x-rays in terms of an x-ray spectrum as well as by the attenuation coefficients and the differential scatter cross sections for every material.

Simulation of Random Photon Tracks   The general procedure to simulate random photon tracks is similar as described in section 2.2.3. However, a biased sampling scheme is applied here at several points to increase the performance of the simulation. Therefore, a weight w is assigned to each photon. Initially, the weight w_0 is set equal to 1 and is updated every time an interaction is not sampled according to the physically "correct" PDF (see section 2.2.4). The main steps of the simulation are as follows.

Sampling the initial energy   The initial energy distribution of x-ray photons is given by the normalized x-ray spectrum that is provided as input to the simulation. As the spectrum is evaluated on a discrete grid, the energy E_0 is sampled using Walker's method as described in section 2.2.2. As the sampling is not biased, the weight remains unchanged.

Sampling the first interaction point   Here, the first interaction point is determined by sampling a random voxel from a uniform distribution. Since interactions in air occur with a very low probability, the corresponding voxels are not considered in the sampling process. Having N_{start} voxels representing a material different than air, the probability of drawing the voxel n_s with $0 \leq n_s < N_{\text{start}}$ is given as:

$$P_{\text{1st, bias}}(n_s) = \frac{1}{N_{\text{start}}}. \tag{3.15}$$

Assuming an isotropic emission of the x-ray source, the physically correct probability density to have the first interaction at position $\boldsymbol{r}$ can be calculated as described in section 2.2.3. According to equations (2.77), (2.78), and (2.82) it is given as:

$$p_{\text{1st, real}}(\boldsymbol{r}) = \frac{1}{2\pi} \frac{\sin \vartheta(\boldsymbol{r})}{2} \mu(\boldsymbol{r}) e^{- \int_0^1 \mu(\boldsymbol{s} + \lambda(\boldsymbol{r} - \boldsymbol{s}), E_0) d\lambda}. \tag{3.16}$$

The probability to have the first interaction within the voxel corresponding to $\boldsymbol{r}$ requires to integrate equation (3.16) over the volume V of that voxel. Here, the integration is approximated as follows:

$$\begin{aligned} P_{\text{1st, real}}(n_x, n_y, n_z) &= \int_V p_{\text{1st, real}}(\boldsymbol{r}) dV \\ &\approx \frac{\Delta V}{4\pi \boldsymbol{r}^2} \mu(\boldsymbol{r}) e^{- \int_0^1 \mu(\boldsymbol{s} + \lambda(\boldsymbol{r} - \boldsymbol{s}), E_0) d\lambda}, \end{aligned} \tag{3.17}$$

where ΔV denotes the volume of the voxel. To account for the biased sampling, the photon's weight is updated according to the fraction of the real and the biased sampling probability:

$$w_{n+1} = w_n \cdot \frac{P_{\text{1st, real}}}{P_{\text{1st, bias}}} = \frac{\Delta V}{4\pi \boldsymbol{r}^2} \mu(\boldsymbol{r}) e^{- \int_0^\lambda \mu(\boldsymbol{r}_n + \lambda' \boldsymbol{d}_n, E_0) d\lambda'} \cdot N_{\text{start}} \tag{3.18}$$

It has to be noted that the advantage of this biased sampling scheme is twofold: first, to ensures that any simulated x-ray interacts with the object and second, to promote interactions more close to the detector as they have a higher impact on the scatter distribution.

Forced detection and photon splitting   To avoid simulating photon tracks that do not contribute to the scatter distribution, i.e. photons that leave the volume of interest without hitting the detector, a forced detection approach is implemented. The basic principle of that approach is depicted in figure 3.3. Instead of scoring the contribution to the scatter distribution only when the photon hits the detector, its contribution is scored at every interaction point. Here, this is realized by the introduction of so-called split photons. Once an interaction point is sampled, N_s split photons, that share all properties except for the weight with their master photon, are generated. To keep the simulation unbiased, the weight of the split photons is reduced according to their number:

$$w_{0,\text{split}} = \frac{w_n}{N_s}. \tag{3.19}$$

For each split photon an interaction effect as well as the corresponding flight direction is sampled(see below). However, in contrast to the master photons, no further interactions are simulated for split photons. Thus, whenever their flight direction intersects with a detector triangle, their contribution is scored (forced detection). Otherwise, they are removed from the simulation. To account for the negligence of split photon interactions, the split photon weight has to be decreased accordingly. Considering the forced detection approach, the probability $P_{\text{det, bias}}$ of the photon not being absorbed (i.e. reaching the detector) equals 1:

$$P_{\text{det, bias}} = 1. \tag{3.20}$$

The physically correct probability, however, is given by the Beer-Lambert law as:

$$P_{\text{det, real}} = e^{-\int_0^1 \mu(\boldsymbol{r}+\lambda(\boldsymbol{r}-\boldsymbol{q}),E)d\lambda}, \tag{3.21}$$

where $\boldsymbol{r}$ is the interaction point of the master photon, or the starting point of the split photon, respectively, and $\boldsymbol{q}$ is the intersection point with the detector triangle. Thus, the split photon weight is given as:

$$w_{1,\text{split}} = w_0 \cdot \frac{P_{\text{det, real}}}{P_{\text{det, bias}}} = w_0 \cdot e^{-\int_0^1 \mu(\boldsymbol{r}+\lambda(\boldsymbol{r}-\boldsymbol{q}),E)d\lambda}. \tag{3.22}$$

Here, interactions in the detector are not included in the MC simulation but are considered in terms of the detector efficiency $\eta(E, d, \Phi)$ that can be calculated e.g. as described in section 3.1.5. Thus, any split photon hitting a certain detector triangle m of thickness d with an incident angle Φ contributes to the scatter distribution $S(m, E)$ as follows:

$$S(m, E) \mathrel{+}= w_{1,\text{split}} \cdot \eta(E, d, \Phi). \tag{3.23}$$

Once all split photons are processed, the simulation continues with the simulation of the random track of the master photon.
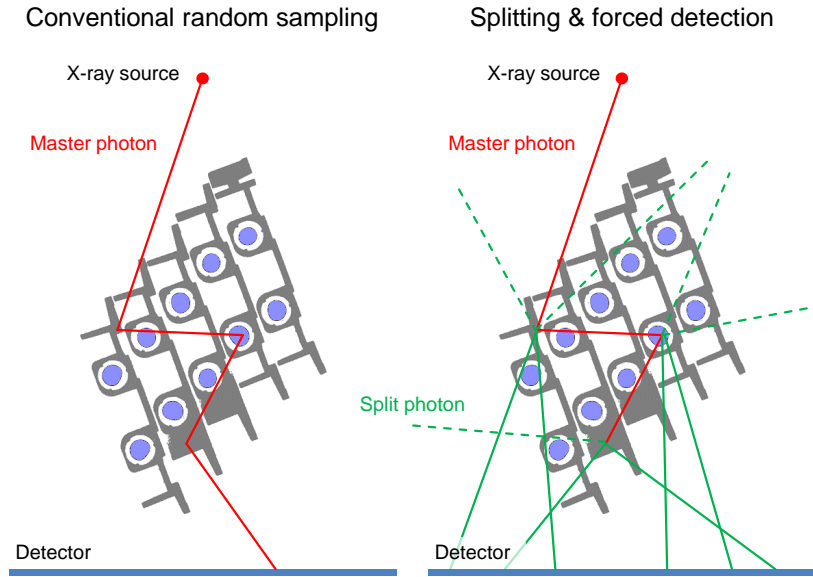
Figure 3.3: Conventional random sampling (left) and particle slitting and forced detection (right). The conventional sampling approach simulates a random photon track until the photon leaves the volume of interest or hits the detector. Particle splitting, in contrast, generates a given number of split photons at any interaction point and scores their contribution to the detector signal immediately when they point towards the detector.

Sampling the interaction effect   Considering energies below 1022 keV, x-ray photons can interact via the photoelectric effect (PE), Compton scattering (CS) and Rayleigh scattering (RS). However, as a photoelectric effect would terminate the simulation without contributing to the scatter distribution, it is not considered explicitly. Thus, only Compton and Rayleigh scattering are sampled with the following probabilities:

$$P(CS) = \frac{\mu_{\mathrm{CS}}(E)}{\mu_{\mathrm{CS}}(E) + \mu_{\mathrm{RS}}(E)} \tag{3.24}$$

$$P(RS) = \frac{\mu_{\mathrm{RS}}(E)}{\mu_{\mathrm{CS}}(E) + \mu_{\mathrm{RS}}(E)}. \tag{3.25}$$

The probability that the photon interacted via the photoelectric effect is equal to $(1 - \mu_{\mathrm{PE}}(E)/\mu(E))$. It's negligence has to be accounted by the decreasing the photon weight accordingly:

$$w_{n+1} = w_n \cdot \left(1 - \frac{\mu_{\mathrm{PE}}(E)}{\mu(E)}\right). \tag{3.26}$$

Updating the energy of the x-ray photon   The energy of the simulated x-ray photon has to be updated according to the interaction effect. This is done as described in section 2.2.3.

Updating the flight direction of the x-ray photon The flight direction is updated according to the differential cross sections that are given as input to the simulation. Here, the update does not differ from the straight forward implementation given in section 2.2.3.

Sampling the path length Once the energy and the flight direction are updated, the path length is sampled to determine the next interaction point of the master photon. Given the attenuation of every material, the path length sampling is performed using a Woodcock tracking as described in section 2.2.3.

Termination of the simulation The simulation of a random photon track is terminated once the master photon leaves the volume. It has to be noted that the master photon does not contribute to the scatter distribution directly, but only via its split photons.

3.1.4 Focal Spot Distribution

Ideally, all x-ray photons are emitted from a single point at position $\boldsymbol{s}$ (see figure 3.1). However, practically the area x-ray photons are generated has a certain extension and is rather described by a focal spot distribution. This distribution is determined by several factors such as the power of the x-ray tube, the electromagnetic fields to focus the electron beam, the geometry of the x-ray tube as well as its physical properties. Typically, it is composed of a narrow distribution that corresponds to the area electrons are focused on and flat but broad distribution, referred to as off-focal radiation, that is associated with electron backscattering.

Given the projection $G(\boldsymbol{k})$ of the focal spot distribution into the focal spot plane (k_x-k_z-plane parallel to the detector plane), the normalized intensities can be calculated as:

$$\frac{I_{p,\text{off}}}{I_{0,\text{off}}}(\boldsymbol{d}) = \frac{\int G(\boldsymbol{k}) \int \tilde{w}(E) e^{-\int_0^1 \mu(\boldsymbol{s}+\boldsymbol{k}+\lambda\cdot(\boldsymbol{d}-(\boldsymbol{s}+\boldsymbol{k})),E)d\lambda} dE\, d^2k}{\int G(\boldsymbol{k}) \int \tilde{w}(E) dE\, d^2k}, \tag{3.27}$$

with $\tilde{w}(E) = w(E)\kappa(E)\eta(E)$ being product of the source spectrum, the prefilter attenuation and the detector efficiency.

i.) Simulation of Projection Data

Given the distribution function $G(\boldsymbol{k})$, projection data can be simulated by evaluating equation (3.27) for a given number of discrete grid points (k_x, k_z) within the focal spot plane. Subsequently, the normalized intensity is given as their weighted sum with the corresponding weight $G(\boldsymbol{k})$. However, as the computation time increases linearly with the number of grid points, this approach is not the method of choice if performance is an issue. Therefore, a convolution-based approach as derived in the following is used instead.

Considering the projection of a thin sample of thickness dl located at position y_0, the line integral over the attenuation coefficient given in equation (3.27) can be approximated as:

$$\int_0^1 \mu(\boldsymbol{s} + \boldsymbol{k} + \lambda \cdot (\boldsymbol{d} - (\boldsymbol{s} + \boldsymbol{k})), E)d\lambda \approx \mu(\boldsymbol{s} + \boldsymbol{k} + \lambda_0 \cdot (\boldsymbol{d} - (\boldsymbol{s} + \boldsymbol{k})), E)dl, \quad (3.28)$$

where $\lambda_0 = \frac{y_0 - s_y}{u_y - s_y}$ is given according to the intersection point with the x-z-plane at $y = y_0$. Inserting this approximation into equation (3.27) yields:

$$
\begin{aligned}
\frac{I_{p,\text{off}}}{I_{0,\text{off}}}(\boldsymbol{d}) &= \frac{\int G(\boldsymbol{k}) \int \tilde{w}(E) e^{-\int_0^1 \mu(\boldsymbol{s}+\boldsymbol{k}+\lambda\cdot(\boldsymbol{d}-(\boldsymbol{s}+\boldsymbol{k})),E)d\lambda} dE \, d^2k}{\int G(\boldsymbol{k}) \int \tilde{w}(E) dE \, d^2k} \\
&\approx \frac{\int G(\boldsymbol{k}) \int \tilde{w}(E) e^{-\mu(\boldsymbol{s}+\boldsymbol{k}+\lambda_0\cdot(\boldsymbol{d}-(\boldsymbol{s}+\boldsymbol{k})),E)dl} dE \, d^2k}{\int G(\boldsymbol{k}) \int \tilde{w}(E) dE \, d^2k} \\
&= \frac{\int G(\boldsymbol{k}) \int \tilde{w}(E) e^{-\mu(\lambda_0(\boldsymbol{d}+(\boldsymbol{s}+\boldsymbol{k})(\frac{1-\lambda_0}{\lambda_0}))),E)dl} dE \, d^2k}{\int G(\boldsymbol{k}) \int \tilde{w}(E) dE \, d^2k} \\
&= \frac{\int \left(\frac{\lambda_0}{1-\lambda_0}\right)^2 G(-\frac{\lambda_0}{1-\lambda_0}\boldsymbol{\kappa}) \int \tilde{w}(E) e^{-\mu(\lambda_0(\boldsymbol{d}-\boldsymbol{\kappa}),E)dl} dE \, d^2\kappa}{\int G(\boldsymbol{k}) \int \tilde{w}(E) dE \, d^2k} \\
&= \tilde{G} * \frac{I_p}{I_0}(\boldsymbol{d}), \quad (3.29)
\end{aligned}
$$

where $\tilde{G}(\boldsymbol{d}) = \frac{\left(\frac{\lambda_0}{1-\lambda_0}\right)^2 G(-\frac{\lambda_0}{1-\lambda_0}\boldsymbol{d})}{\int G(\boldsymbol{k}) d^2k}$ represents the mirrored projection of the focal spot distribution into the detector plane and $\frac{I_p}{I_0}(\boldsymbol{d})$ denotes the ideal polychromatic forward projection, i.e. the polychromatic forward projection for a point-like focal spot at position $\boldsymbol{s}$. Thus, the effect of the non-ideal focal spot distribution can be modeled approximately by a single forward projection followed by a convolution operation. For the sake of performance, this model is also applied to thick samples.

ii.) Determination of the Distribution Function $G(\boldsymbol{k})$

Edge Measurement The spatial distribution $G(\boldsymbol{k})$ of the emitted x-rays can be reconstructed directly using a transmission measurement of an absorbing edge as depicted in figure 3.4. Suppose to have a measurement of a straight line within the x-z-plane at y_0. According to equation (3.29) this so called line-spread function (LSF) is given as:

$$\text{LSF}(\vartheta, \xi) = \int \tilde{G}(\boldsymbol{d}) \delta(\boldsymbol{d} \cdot \hat{\boldsymbol{n}}_\vartheta - \xi) d^2u. \quad (3.30)$$

Obviously, the LSF equals the Radon transform of the function $\tilde{G}$. Thus, the reconstruction problem is similar to the CT reconstruction problem described in section 2.1.4. Consequently, the function $\tilde{G}$ can be reconstructed using an FBP approach:

$$\tilde{G}(\boldsymbol{d}) = \mathcal{R}^{-1}\text{LSF}(\vartheta, \xi). \quad (3.31)$$

However, due to the lack of appropriate line samples it is more convenient to measure the edge-spread function (ESF) as depicted in figure 3.4. Given the ESF, the LSF can be calculated as its derivative. Assuming a rotational symmetry of the focal spot distribution, $\tilde{G}$ can be determined from a single transmission measurement.
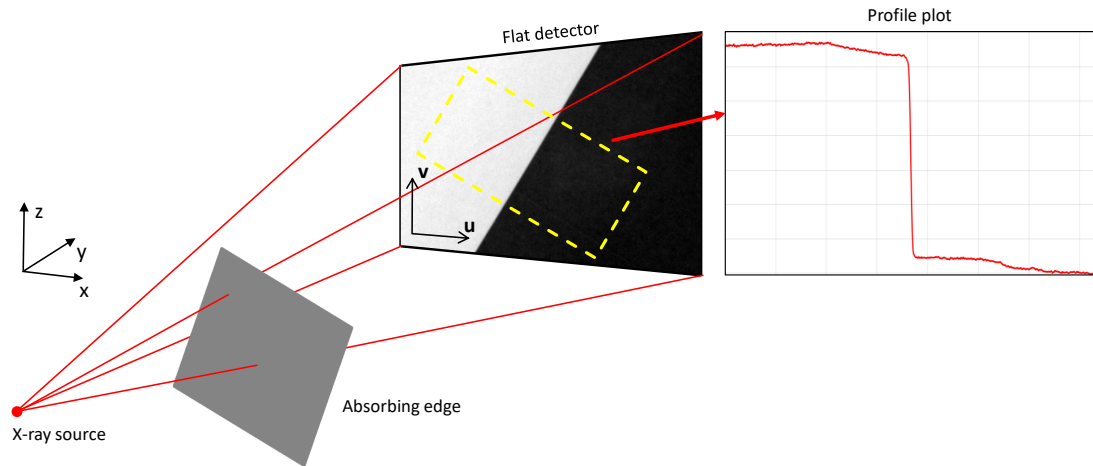
Figure 3.4: Setup to determine the focal spot distribution using an absorbing edge.

Calibration Sphere Measurement Using an edge measurement is potentially the most accurate way to determine the distribution function $G(\boldsymbol{k})$. However, it is rather complicated to integrate an edge measurement in the workflow of a CT scan. Therefore, it is proposed to determine the function $G(\boldsymbol{k})$ or its projection $\tilde{G}(\boldsymbol{d})$, respectively, from the measurement of a calibration sphere. These high-precision crystalline spheres have well-known dimensions and are typically measured prior to any CT acquisition to calibrate the geometry of the system as well as the geometrical distortion of the detector. Using this approach, the determination of the distribution function is formulated as an optimization problem. Therefore, it is modeled using a a weighted sum of a set of basis functions. Here, estimates derived from an edge measurement, as depicted in figure 3.4, suggest that the basis functions can be chosen as Gaussian functions and step functions. Thus, the following parameterization is proposed:

$$\tilde{G}_{\text{est}}(r, \vartheta, \boldsymbol{\theta}) = a_0 + \Theta(b_0 - |r|) + \sum_{i=1}^{i=4} c_i e^{-a_i \cdot (r-b_i)^2}, \tag{3.32}$$

with $\boldsymbol{\theta} = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ denoting the parameter vector and Θ denoting the Heaviside function. It has to be noted that a rotational symmetry of $\tilde{G}$ is assumed here, and therefore, there is no dependence on ϑ.

Given the calibration sphere measurement I_{meas}, the open parameters are determined by minimizing the following cost function with respect to $\boldsymbol{\theta}$.

$$C_{\text{off}}(\boldsymbol{\theta}) = \|\tilde{G}_{\text{est}}(\boldsymbol{\theta}) * \frac{I_p}{I_0} - I_{\text{meas}}\|^2. \tag{3.33}$$

It has to be noted that the convolution is evaluated on a Cartesian grid. Therefore, the distribution function given in equation (3.32) is resampled onto a Cartesian grid.

3.1.5 X-Ray Detection

The x-ray detection process is typically characterized by the detection efficiency $\eta(E_\gamma)$, i.e. the fraction of energy that is absorbed per incident x-ray photon with energy E_γ:

$$\eta(E_\gamma) = \frac{E_{\text{dep}}(E_\gamma)}{E_\gamma}, \tag{3.34}$$

where E_{dep} refers to the energy deposition in the detector. Given a detector of thickness d and attenuation $\mu(E)$, a common approximation of the detector efficiency is given by:

$$\eta(E_\gamma) \approx 1 - e^{-\mu(E_\gamma) \cdot d}, \tag{3.35}$$

which equals the probability of the x-ray photon undergoing an interaction within the detector. Thus, it is assumed that any interaction leads to the deposition of the complete energy of the x-ray photon. Since this assumption does not hold true, especially for Compton and Rayleigh scattering, the detector efficiency is typically overestimated by this model. Therefore, the attenuation coefficient is often replaced by the so-called energy transfer coefficient μ_{tr} that corresponds to the mean fraction of energy transferred to kinetic energy of electrons per unit path length [110]. This assumption, however, leads to an underestimation of the detector efficiency since higher order interactions are neglected. Therefore, a more accurate model is derived as follows.

In general, the interactions depicted in figure 3.5 can occur when x-ray photons hit the sensitive area of the detector. Here, the first interaction via the photoelectric effect, Compton scattering or Rayleigh scattering is referred to as first-order interaction while single or multiple interactions of fluorescence or scattered x-rays are referred to as higher-order interactions. Calculating the detector efficiency according to equation (3.34) requires to sum up the energy deposition of these interactions. Thus, the energy deposition of an x-ray photon with energy E_γ within a detector of thickness d can be calculated as:

$$E_{\text{dep}}(E_\gamma, d) = \sum_i E_{\text{dep},i} = \sum_i \int_0^d p_{\text{int},i}(r|E_\gamma)\bar{E}_i(r, E_\gamma)dr, \tag{3.36}$$

where $p_{\text{int},i}(r|E_\gamma)$ is the PDF for an interaction at position r, and $\bar{E}_i(r, E_\gamma)$ is the mean energy transfer to the medium given a certain interaction i. To estimate the energy deposition according to equation (3.36), the following models and approximations are applied:

i.) *Probability of first-order interactions*

According to equation (2.82), the PDF $p_{\text{int}}(r|E_\gamma)$ for an interaction at position r within a homogeneous detector material with attenuation $\mu(E_\gamma)$ is given as:

$$p_{\text{int}}(r|E_\gamma) = \mu(E_\gamma)e^{-\mu(E_\gamma) \cdot r}. \tag{3.37}$$
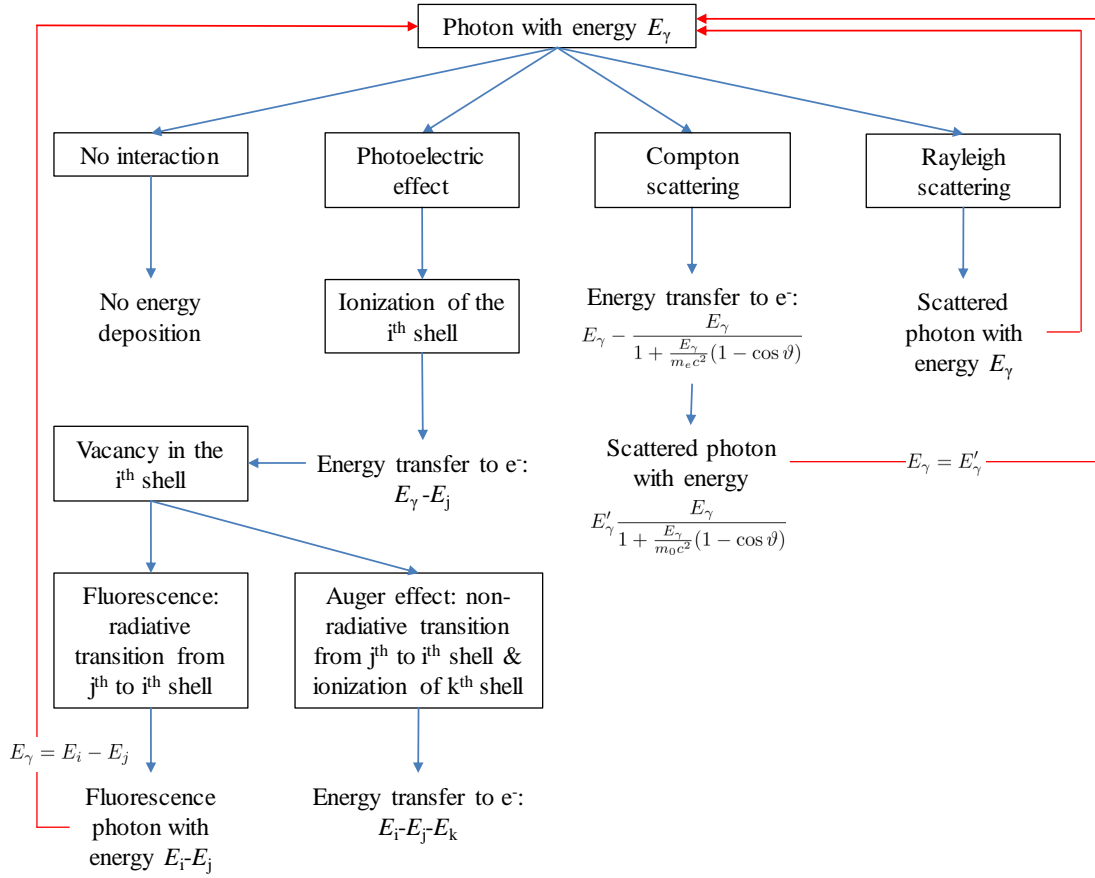
Figure 3.5: Sketch of possible interactions in the x-ray detector.

The probability of the interaction being a photoelectric interaction (PE), Compton scattering (CS) or Rayleigh scattering (RS), equals $P_i(E) = \frac{\mu_i(E)}{\mu(E)}$ with $i \in \{\text{PE, CS, RS}\}$ and $\mu_k(E)$ being the interaction effect specific attenuation coefficient. Thus, the complete PDF is given as:

$$p_{\text{int},i}(r|E_\gamma) = \mu(E_\gamma) \, e^{-\mu(E_\gamma) \cdot r} \, \frac{\mu_i(E_\gamma)}{\mu(E_\gamma)}, \tag{3.38}$$

ii.)   *Energy deposition of first-order interactions via the photoelectric effect*

As described in section 2.1.2, a photoelectric interaction leads to the absorption of the incident x-ray photon leaving the absorbing atom in an excited state. With a certain probability the atom deexcites emitting a fluorescence photon that might escape the detector. Therefore, it is assumed that the energy deposition of a photoelectric interaction equals the energy E_γ of the incident x-ray photon minus the mean energy that is transferred to fluorescence photons. For the sake of simplicity, the energy transfer is restricted to K_α and K_β fluorescence photons

as they occur most likely and have sufficient energy to escape the detector [111]. Given the probability P_{K_α} and P_{K_β} for a K_α and a K_β fluorescence, the mean energy transfer is given as:

$$\bar{E}_{PE}(r, E_\gamma) = (E_\gamma - P_{K_\alpha}E_{K_\alpha} - P_{K_\beta}E_{K_\beta}) \tag{3.39}$$

where E_{K_α} and E_{K_β} are the energies of the K_α and a K_β fluorescence photons. The probability $P_{K_{\alpha/\beta}}$ can be calculated as the product of the k-shell ionization probability $P_{PE,\,K}$, the k-shell fluorescence yield Φ, i.e. the fraction of k-shell vacancies that are filled by a radiative transition, and the probability $P_{L\rightarrow K/M\rightarrow K}$ that a K-shell vacancy is filled by an electron of the L-shell or the M-shell, respectively. Here, $P_{PE,\,K}$ is approximated as:

$$P_{PE,\,K} \approx \begin{cases} \frac{\mu(E_K+\epsilon)-\mu(E_K-\epsilon)}{\mu(E_K+\epsilon)}, & \text{if } E > E_K \\ 0, & \text{else} \end{cases} \tag{3.40}$$

where E_K is the k-edge energy. Φ and $P_{L\rightarrow K/M\rightarrow K}$, in contrast, are taken from data tables given in the evaluated atomic data library (EADL) [107].

Substituting equation (3.38) and equation (3.39) into equation (3.36) yields the energy deposition of first-order photoelectric interactions::

$$\begin{aligned} E_{dep,\,PE}(E_\gamma, d) &= \int_0^d p_{int,\,PE}(r|E_\gamma)\bar{E}_{PE}(r, E_\gamma)dr \\ &= \int_0^d \mu(E_\gamma)e^{-\mu(E_\gamma)\cdot r}\frac{\mu_{PE}(E)}{\mu(E)}(E_\gamma - P_{K_\alpha}E_{K_\alpha} - P_{K_\beta}E_{K_\beta})dr \\ &= \frac{\mu_{PE}(E_\gamma)}{\mu(E_\gamma)}(1 - e^{-\mu(E_\gamma)\cdot d})\cdot(E_\gamma - P_{K_\alpha}E_{K_\alpha} - P_{K_\beta}E_{K_\beta}) \end{aligned} \tag{3.41}$$

iii.) *Energy deposition of first-order interactions via Compton scattering*

A Compton scattered x-ray photon transfers a fraction of it's energy to atomic electrons. According to equation (2.14), the energy transfer can be calculated as:

$$E_e'(\vartheta, E_\gamma) = E_\gamma - E_\gamma'(\vartheta) = E_\gamma\left(1 - \frac{1}{1 + \frac{E_\gamma}{m_ec^2}(1 - \cos(\vartheta))}\right), \tag{3.42}$$

with E_γ being the energy of the incident x-ray photon and ϑ being the scatter angle. Given the angular distribution of Compton scattered x-ray photons in terms of the differential Compton scatter cross section $\frac{d\sigma_{CS}}{d\vartheta}$, the mean energy transfer can be calculated as the weighted average from 0 to π:

$$\bar{E}_{CS}(r, E_\gamma) = \int_0^\pi \frac{d\sigma_{CS}}{d\vartheta}E_e'(\vartheta, E_\gamma)d\vartheta. \tag{3.43}$$

Thus, the energy deposition of first-order Compton scatter interactions is given as:

$$
\begin{aligned}
E_{\text{dep, CS}}(E_\gamma, d) &= \int_0^d p_{\text{int, CS}}(r|E_\gamma)\bar{E}_{\text{CS}}(r, E_\gamma)dr \\
&= \int_0^d \mu(E_\gamma)e^{-\mu(E_\gamma)\cdot r}\frac{\mu_{\text{CS}}(E_\gamma)}{\mu(E_\gamma)}\int_0^\pi \frac{d\sigma_{\text{CS}}}{d\vartheta}E'_{\text{e}}(\vartheta, E_\gamma)d\vartheta dr \\
&= \frac{\mu_{\text{CS}}(E_\gamma)}{\mu(E_\gamma)}(1 - e^{-\mu(E_\gamma)\cdot d})\int_0^\pi \frac{d\sigma_{\text{CS}}}{d\vartheta}E'_{\text{e}}(\vartheta, E_\gamma)d\vartheta. \quad (3.44)
\end{aligned}
$$

iv.) Energy deposition of first-order interactions via Rayleigh scattering

Since Rayleigh scattering is an elastic process, there is no energy transfer to electrons and therefore no energy deposition:

$$
E_{\text{dep, RS}}(E_\gamma, d) = 0. \quad (3.45)
$$

v.) Higher-order interactions

It is assumed that any interaction of fluorescence or scattered x-ray photons leads to the deposition of their complete energy. Thus, all calculations here are restricted to secondary interactions. Given a primary interaction at position r, the fluorescence or the scattered photon interacts with the detector material with a probability equal to $1 - e^{-\mu(E)\cdot l(r,\vartheta)}$, where ϑ is the polar angle of the scattered photon and $l(r,\vartheta)$ is the intersection length with the detector. Thus, the mean energy transfer can be calculated as:

$$
\bar{E}_{\text{2nd},i}(r, E_\gamma) = \int_0^\pi \frac{d\sigma_i}{d\vartheta}E_{\text{2nd}}(\vartheta, E_\gamma)(1 - e^{-\mu(E_{\text{2nd}}(\vartheta,E_\gamma))\cdot l(r,\vartheta)})d\vartheta, \quad (3.46)
$$

where $\frac{d\sigma_i}{d\vartheta}$ is the differential cross section for the considered interaction and E_{2nd} is the energy of the fluorescence or scattered x-ray photons.

It has to be noted that equation (3.46) assumes a rotational symmetry with respect to the azimuth angle φ (otherwise, the intersection length also depends on φ). Therefore, equation (3.46) only holds true for flat detectors and x-ray photons with an incident angle perpendicular to the detector's surface. The treatment of skew incident angles will be discussed later.

vi.) Energy deposition of fluorescence photons

It is assumed that fluorescence photons are emitted isotropically ($\frac{d\sigma}{d\vartheta} = \frac{\sin\vartheta}{2}$). K_α fluorescence photons that are emitted with the probability P_{K_α} have the energy K_{K_α} while K_β fluorescence photons (probability P_{K_β}) have the energy E_{K_β}. Thus, according to equation (3.46) and equation (3.36), their energy deposition is given as:

$$
\begin{aligned}
E_{\text{dep, 2nd PE}}(E_\gamma, d) = \int_0^d \mu_{\text{PE}}(r, E_\gamma)e^{-\mu(r,E_\gamma)\cdot r}\int_0^\pi \frac{\sin\vartheta}{2}P_{K_\alpha}E_{K_\alpha} \\
+ (1 - e^{-\mu(E_{K_\alpha})\cdot l(r,\vartheta)}) + P_{K_\beta}E_{K_\beta}(1 - e^{-\mu(E_{K_\beta})\cdot l(r,\vartheta)})d\vartheta dr,
\end{aligned}
$$

$$(3.47)$$

vii.) Energy deposition of Compton scattered x-ray photons

The angular distribution of Compton scattered x-ray photons is given by the differential Compton scatter cross section $\frac{d\sigma_{\text{CS}}}{d\vartheta}$. The remaining energy E'_γ of the scattered photon is given by equation (2.14). Substitution into equation (3.36) yields:

$$E_{\text{dep, 2nd CS}}(E_\gamma, d) = \int_0^d \mu_{\text{CS}}(r, E_\gamma) e^{-\mu(r, E_\gamma) \cdot r} \int_0^\pi \frac{d\sigma_{\text{CS}}}{d\vartheta} (1 - e^{-\mu(E'_\gamma(\vartheta)) \cdot l(r, \vartheta)}) d\vartheta dr$$

$$(3.48)$$

viii.) Energy deposition of Rayleigh scattered x-ray photons

The angular distribution of Rayleigh scattered x-ray photons is given by the differential Rayleigh scatter cross section $\frac{d\sigma_{\text{RS}}}{d\vartheta}$. The energy of the scattered photons equals the energy E_γ of the incident photon. Thus, the energy deposition can be calculated as:

$$E_{\text{dep, 2nd RS}}(E_\gamma, d) = \int_0^d \mu_{\text{RS}}(r, E_\gamma) e^{-\mu(r, E_\gamma) \cdot r} \int_0^\pi \frac{d\sigma_{\text{RS}}}{d\vartheta} (1 - e^{-\mu(E_\gamma) \cdot l(r, \vartheta)}) d\vartheta dr$$

$$(3.49)$$

ix.) Treatment of skew incident angles

The calculations above assume the incident angle of the x-ray photon to be perpendicular to the detector's surface. In principle, they could be adjusted such that they also hold true for skew incident angles. However, for the sake of performance the incident angle Φ is accounted by applying a multiplicative correction term $c(E_\gamma, \Phi)$ such that:

$$\eta(E_\gamma, d, \Phi) = \eta(E_\gamma, d) \cdot c(E_\gamma, d, \Phi). \tag{3.50}$$

Here, the correction term is calculated according to the simple detector efficiency model given by equation (3.35):

$$c(E_\gamma, d, \Phi) = \frac{1 - e^{-\mu(E_\gamma)\frac{d}{\cos \Phi}}}{1 - e^{-\mu(E_\gamma) \cdot d}}, \tag{3.51}$$

where $\frac{d}{\cos \Phi}$ accounts for the increased intersection length with the detector.

Substituting the energy deposition of first- and higher-order interactions into equation (3.34) yields an expression for detector efficiency which can be evaluated by numerical integration over the detector thickness d and the polar angle ϑ. Here, the use of a step size of $\Delta d = 5$ µm and $\Delta \vartheta = 1°$ seem to be a good compromise between speed and accuracy.

3.2 Results and Validation

3.2.1 X-Ray Emission Spectra

The simulation of transmission x-ray spectra is based on a modification of the model of Tucker et al. As the original model has been developed for reflection x-ray tubes, its parameters were recalibrated here to fit MC simulations of transmission tubes. The corresponding parameters are summarized in table 3.1. It has to be noted that all other parameters were kept unchanged. Exemplary spectra including a 1 mm aluminum prefilter are shown in figure 3.6. While there are some deviations in the low energy range of the spectrum, higher energies are reproduced quite accurately. Evaluating the mean absolute percentage error (MAPE) between the predicted spectra and the MC reference for all tube voltages and all target thicknesses, there is an error of 9.7 %.
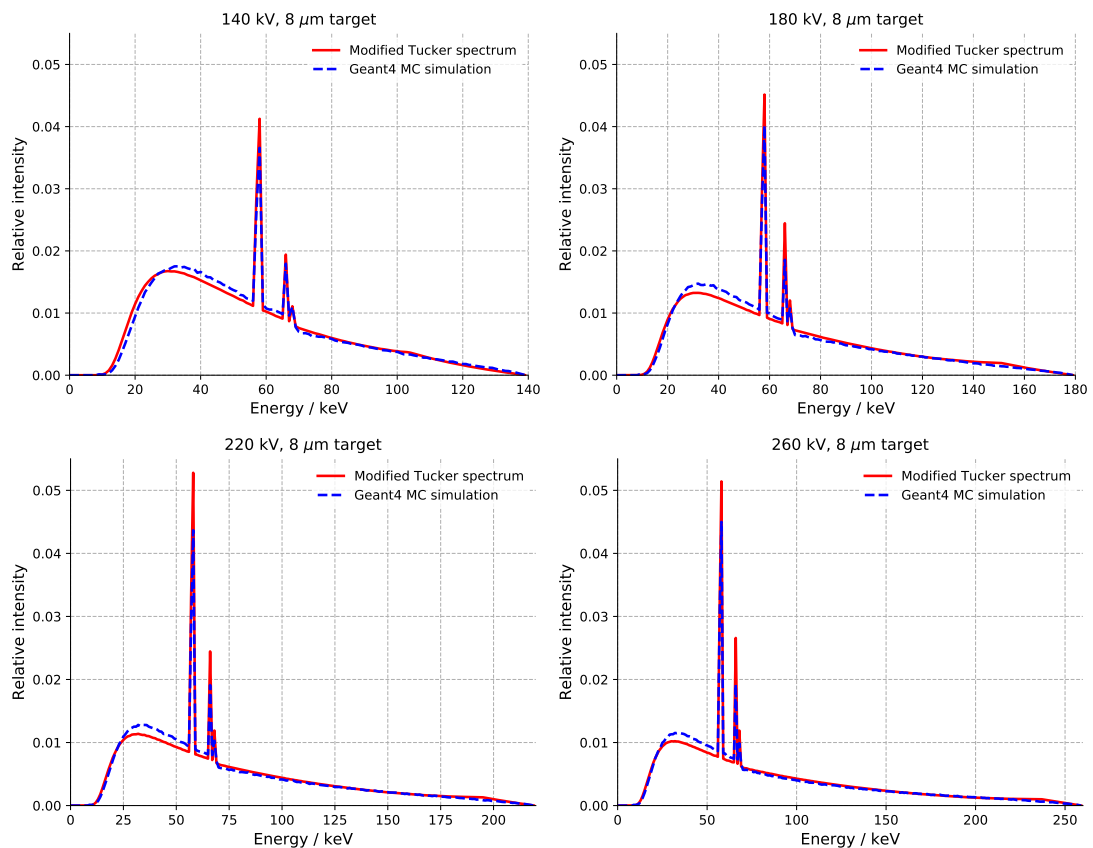
Figure 3.6: Exemplary results of the x-ray spectra derived with the modified Tucker model. All spectra are prefiltered with 1 mm of aluminum.

Table 3.1: Parameter of the spectrum model.

	B_1	B_2	B_3	B_4	A_K	n_K
Tucker et al.	-5.049	10.847	- 10.516	3.842	1.349×10^{-3}	1.648
Recalibration	0.088	1.565	-6.398	4.302	3.233×10^{-3}	1.652

3.2.2 X-Ray Scattering

To validate the scatter predictions of the MC simulation described in section 3.1.3 (referred to as DKFZ MC code), they are compared against scatter distributions calculated with the well established Geant4 (v. 10.4) MC code [109]. Since the Geant4 MC simulation is only used for reference, it is not optimized for performance but relies on a straight forward implementation as described in section 2.2.3. For both MC codes, simulations were performed using a CBCT setup with an ideal (100 % efficiency) 30×30 cm detector, a source-to-isocenter distance of 300 mm and an isocenter-to-detector distance of 300 mm (see figure 3.7). The object to be simulated consists of different sized spheres (10 mm and 25 mm radius) of different materials (water, aluminum and iron). Furthermore, simulations were performed using different x-ray spectra with a maximal tube voltage of 50 kV, 100 kV, 200 kV and 400 kV. A qualitative comparison of the scatter distributions for 10^{10} x-ray photons is shown in figure 3.8. A quantitative comparison in terms of the mean absolute percentage error (MAPE) between the DKFZ MC simulation and the Geant4 simulation yields deviations of 2.1 % (50 kV spectrum), 1.7 % (100 kV spectrum), 1.4 % (200 kV spectrum), and 1.2 % (400 kV spectrum). Thus, there is a good accordance between both MC codes, while the remaining deviations may result from the discretization that is required for the DKFZ simulation or from slightly different cross section tables that are used for the simulation.
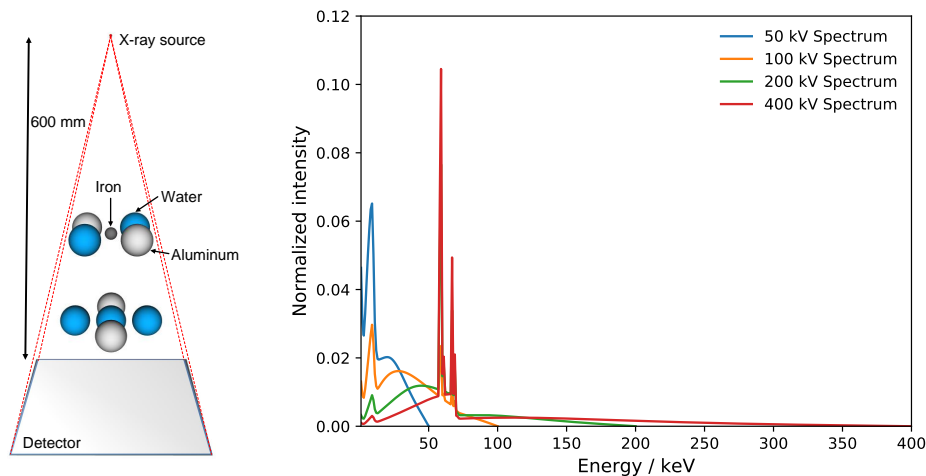
Figure 3.7: Setup for the validation of the MC scatter simulation (left) and x-ray tube spectra that were used for the simulation (right).
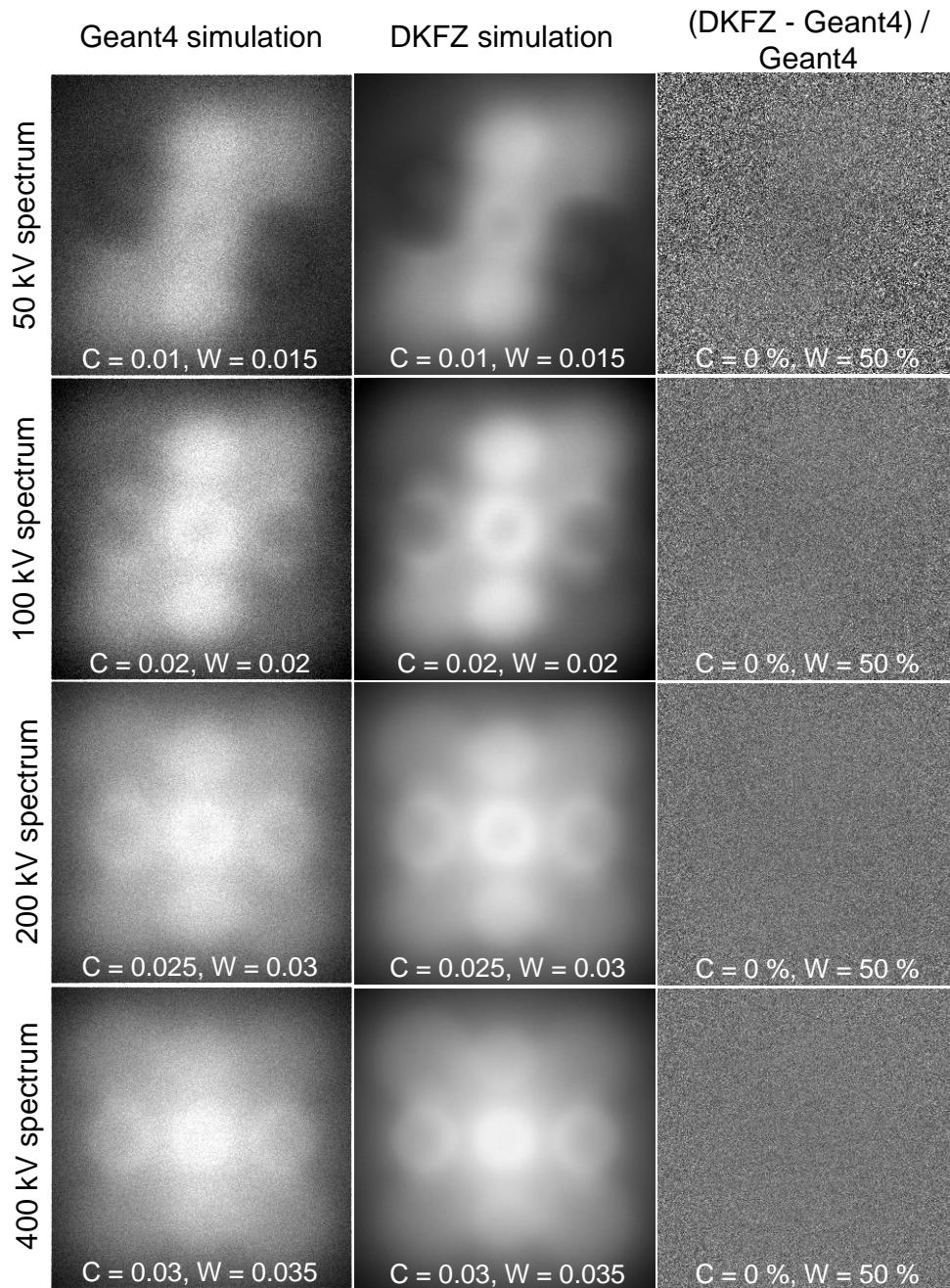
Figure 3.8: Scatter simulations of 10^{10} x-ray photons for the sphere phantom shown in figure 3.7 at four different tube voltages between 50 kV and 400 kV (top row to bottom row). The left column shows the scatter distribution calculated using the Geant4 MC code, the middle column shows the results for the MC code developed in this thesis, while the relative difference is shown in the right column.

3.2.3 Focal Spot Distribution

To evaluate the limits of the convolution-based approach to model the focal spot distribution, simulations using the line pair phantom depicted in figure 3.9 were performed. The phantom consists of 50 line pairs that were assigned increasing densities such that there are projection values between 0 and 10. For the sake of simplicity, the simulations were restricted to a 2D setup as depicted in figure 3.9. The focal spot distribution was modeled as a Gaussian function with varying standard deviations, namely 0.05 mm, 0.1 mm, 0.5 mm, 1.0 mm and 2.0 mm. For each simulation 50000 x-positions of the x-ray source were sampled randomly from the corresponding Gaussian distribution. Subsequently, the respective forward projections were averaged in intensity domain. The proposed method, in contrast, applied a convolution operation to a single forward projection assuming a point-like focal spot. As described in section 3.1.4, this method is theoretically equivalent to a real simulation in case of (infinitely) thin objects. Therefore, the line pairs were simulated with different heights of 1 mm, 10 mm, 50 mm, 100 mm, and 200 mm. In any case, the line pairs were focused on the x-ray source (see figure 3.9). Furthermore, different magnifications (1.25, 2.0, and 5.0) were investigated by shifting the line pair phantom relative to the x-ray source while keeping the source-detector distance fixed at 1000 mm. To have always the same detector signal, the spacing of the line pairs was changed according to the magnification. Thus, there is a spacing of 3.125 lp/cm for the low magnification, a spacing of 5.0 lp/cm for the medium magnification and 12.5 lp/cm for the high magnification.
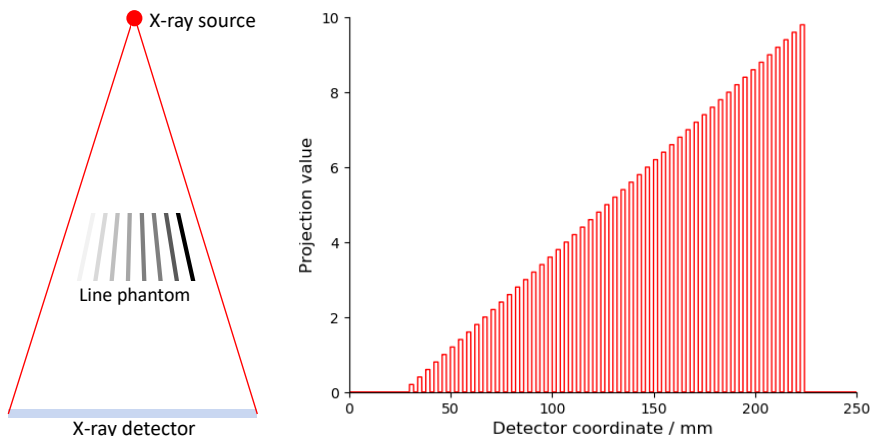
Figure 3.9: Sketch of the line pair phantom (left) and the corresponding projection values for a point-like focal spot (right). The line pairs are simulated with increasing density, resulting in a decreasing intensity.

Qualitative results of the real simulation and the proposed model are shown in figure 3.10 for a setup with a magnification of 1.25. Here, it can be observed that the convolution model is able to reproduce the physically correct simulation with only minor deviations if either the width of the focal spot distribution is small or the extension of the object in source-detector direction is small. Otherwise (see figure 3.10, bottom right)
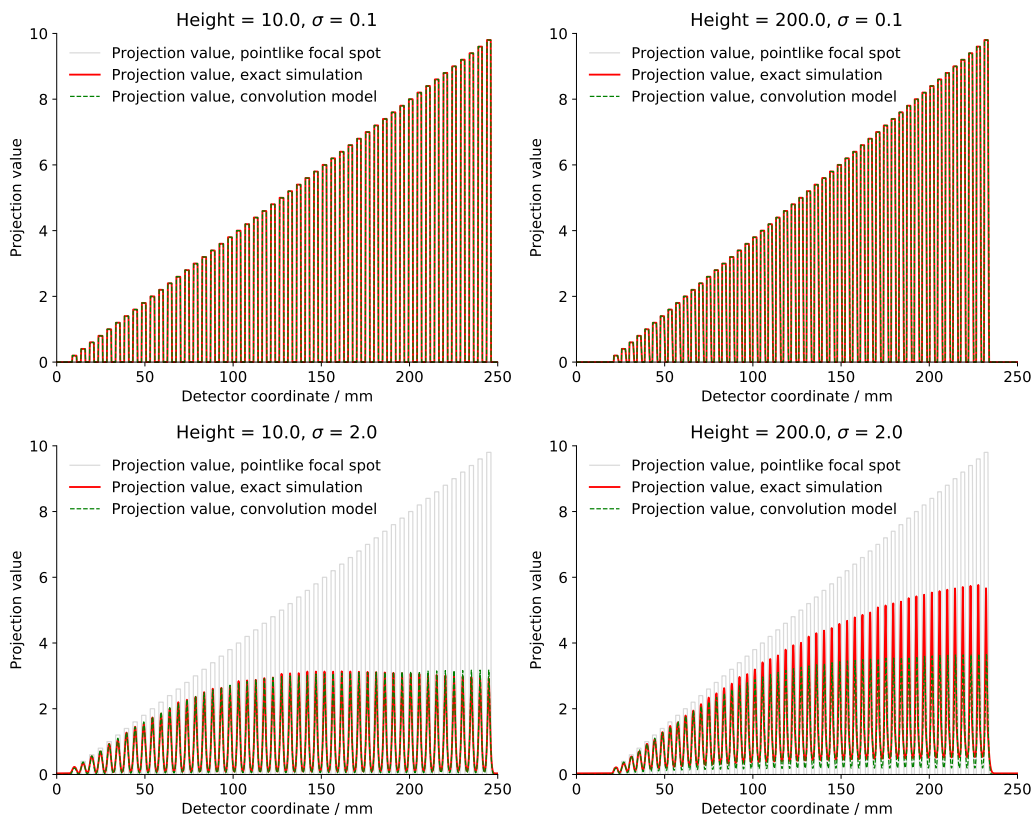
Figure 3.10: Comparison of the proposed convolution-based model to a physically correct simulation using a setup with a magnification of 1.25.

higher deviations may arise. A quantitative evaluation in terms of the mean absolute percentage error (MAPE) between the convolution model and the reference simulation is provided in table 3.2. Here, similar trends can be observed. The accuracy decreases by increasing the width of the focal spot distribution or the height of the line pairs. Furthermore, simulations that were performed with a small magnification show slightly lower deviations. This might be explained by the fact that effective width of the focal spot distribution, i.e. the focal spot distribution projected to the isocenter, decreases for decreasing magnifications. However, for most practical cases the convolution model seems to be a valid approximation to a physically correct simulation.

A practical example of the application of the convolution-based model is depicted in figure 3.11. Here, the projection of the focal spot distribution $\tilde{G}(\boldsymbol{d})$ was estimated using a calibration sphere measurement as described in section 3.1.4. To investigate whether this distribution also applies to other measurements, a luster terminal was measured subsequently. As shown in the second row of figure 3.11, the application of the convolution-based model reduced the discrepancies between the simulation and the measurement significantly which indicates the applicability of this approach.

Table 3.2: Mean absolute percentage error of the convolution-based focal spot simulation for the line pair phantom.

	Standard deviation of the Gaussian				
	0.05 mm	0.1 mm	0.5 mm	1.0 mm	2.0 mm
Height of the lines					
Magnification 1.25					
1.0 mm	0.6	0.5	0.5	0.6	0.7
10.0 mm	0.5	0.6	1.1	1.6	3.9
50.0 mm	0.5	0.8	3.2	3.9	9.1
100.0 mm	0.7	1.1	5.7	11.1	13.2
200.0 mm	0.8	1.5	8.7	17.2	20.2
Magnification 2.0					
1.0 mm	0.6	0.7	0.7	0.8	0.9
10.0 mm	0.5	0.6	3.0	5.2	6.5
50.0 mm	0.8	1.4	3.9	7.6	9.3
100.0 mm	1.2	2.4	6.9	12.4	15.6
200.0 mm	1.9	4.1	13.1	18.3	24.2
Magnification 5.0					
1.0 mm	0.5	0.5	0.6	0.7	0.9
10.0 mm	0.9	1.1	1.6	4.2	11.9
50.0 mm	3.6	5.7	10.2	21.3	33.0
100.0 mm	6.0	10.0	16.9	26.3	32.9
200.0 mm	3.6	6.9	12.7	17.5	21.2
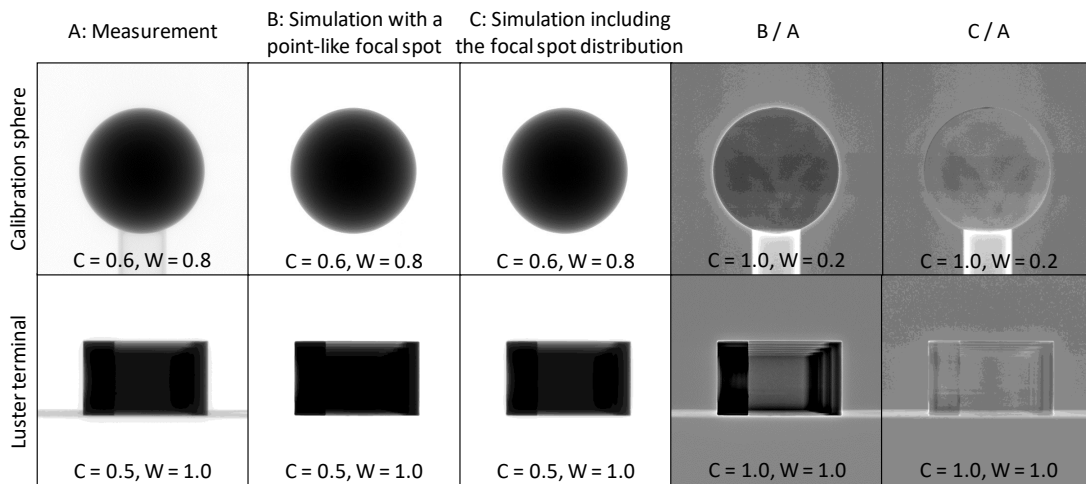
Figure 3.11: Simulation with and without the focal spot model. The kernel representing the projection of the focal spot distribution was determined using the calibration sphere measurement in the top row. The same kernel also applies to the measurement of the luster terminal.

3.2.4   X-Ray Detection

To evaluate the accuracy of the detector model derived in section 3.1.5, it was compared against an • MC simulation which is considered as ground truth. The reference MC simulation was implemented using the well established Geant4 MC framework (v. 10.4)[109]. Simulations were performed for cesium iodine (CsI), cadmium telluride (CdTe) and gadolinium oxysulfide (Gd_2O_2S) which are three commonly used detector materials in CT. For each material 300 energies distributed equally between 1 keV and 300 keV were simulated using different detector thicknesses (0.25 mm, 0.5 mm, 1.0 mm) and different incident angles (0°, 25°). Additionally, the detector efficiency was calculated according to the simple model given by equation (3.35). The corresponding results are shown in figure 3.12 (CsI), figure 3.13 (CdTe), and figure 3.14 (Gd_2O_2S). For all configurations the proposed model is very close to the ground truth MC simulation except for some small deviations in the range of the k-edge energy. More simple models, in contrast, lead to significantly higher deviations, especially for lower energies. A qualitative evaluation in terms of the mean absolute percentage error (MAPE) with respect to the ground truth is given in table 3.3. The lowest accuracy is achieved modeling the detector efficiency as $\eta(E_\gamma) = 1 - e^{-\mu(E_\gamma) \cdot d}$. This leads to an overestimation of the detector efficiency with deviations from the ground truth between 18 % and 44 % on average. More accurate results can be achieved by replacing the attenuation coefficient by the mean energy transfer coefficient. In that case, however, the detector efficiency is slightly underestimated leading to a MAPE between 7 % and 14 %. The proposed model clearly outperforms the two simple models and deviates from the ground truth MC simulation by only 1 % to 4 % while being about 1000 times faster.

Table 3.3: Mean absolute percentage error with respect to the MC ground truth of the three detector efficiency models for energies between 1 keV and 300 keV.

| Configuration | | | $1 - e^{-\mu(E_\gamma) \cdot d}$ | $1 - e^{-\mu_{tr}(E_\gamma) \cdot d}$ | Proposed model |
Material	d / mm	Φ			
CsI	0.25	0° / 25°	43.7 / 43.0 %	7.5 / 7.1 %	4.3 / 4.4 %
CsI	0.5	0° / 25°	35.5 / 32.2 %	8.3 / 8.2 %	3.5 / 3.4 %
CsI	1.0	0° / 25°	25.2 / 25.0 %	10.5 / 10.4 %	2.2 / 2.1 %
CdTe	0.25	0° / 25°	33.5 / 33.3 %	10.8 / 10.5 %	2.7 / 2.6 %
CdTe	0.5	0° / 25°	26.3 / 26.2 %	12.9 / 12.8 %	2.0 / 1.9 %
CdTe	1.0	0° / 25°	21.1 / 20.9 %	13.9 / 13.7 %	1.9 / 2.0 %
Gd_2O_2S	0.25	0° / 25°	35.7 / 35.4 %	8.8 / 8.4 %	3.9 / 3.9 %
Gd_2O_2S	0.5	0° / 25°	25.8 / 25.7 %	11.2 / 10.8 %	2.4 / 2.4 %
Gd_2O_2S	1.0	0° / 25°	18.6 / 18.3 %	12.2 / 12.0 %	1.6 / 1.4 %
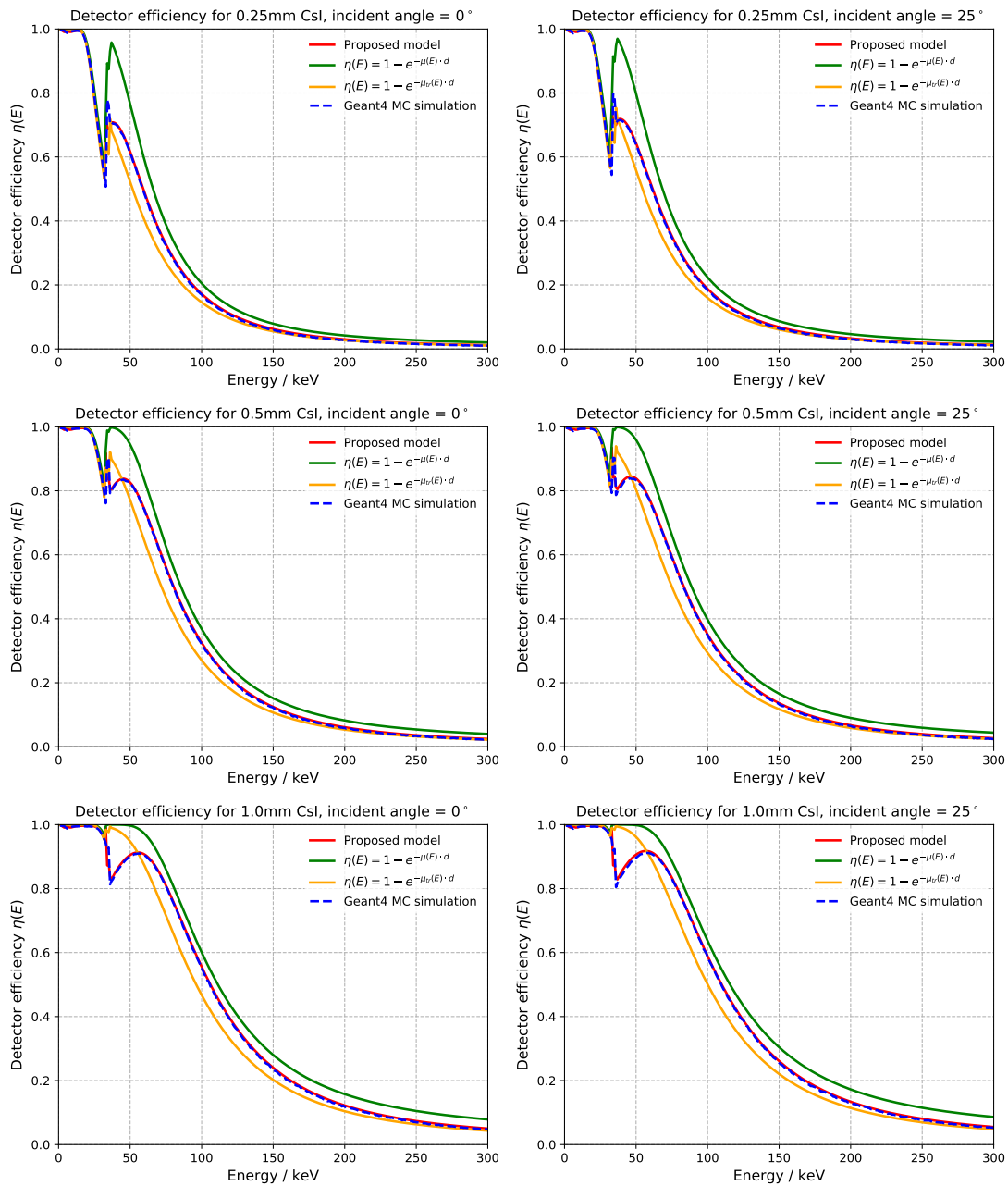
Figure 3.12: Comparison of different models to calculate the detector efficiency of CsI. A Geant4 MC simulation (dashed blue curve) which is considered as ground truth serves as reference. Simulations were performed for detector thicknesses of 0.25 mm (top row), 0.5 mm (middle row), and 1.0 mm (bottom row). The two columns show different incident angles (left: 0°, right: 25°).
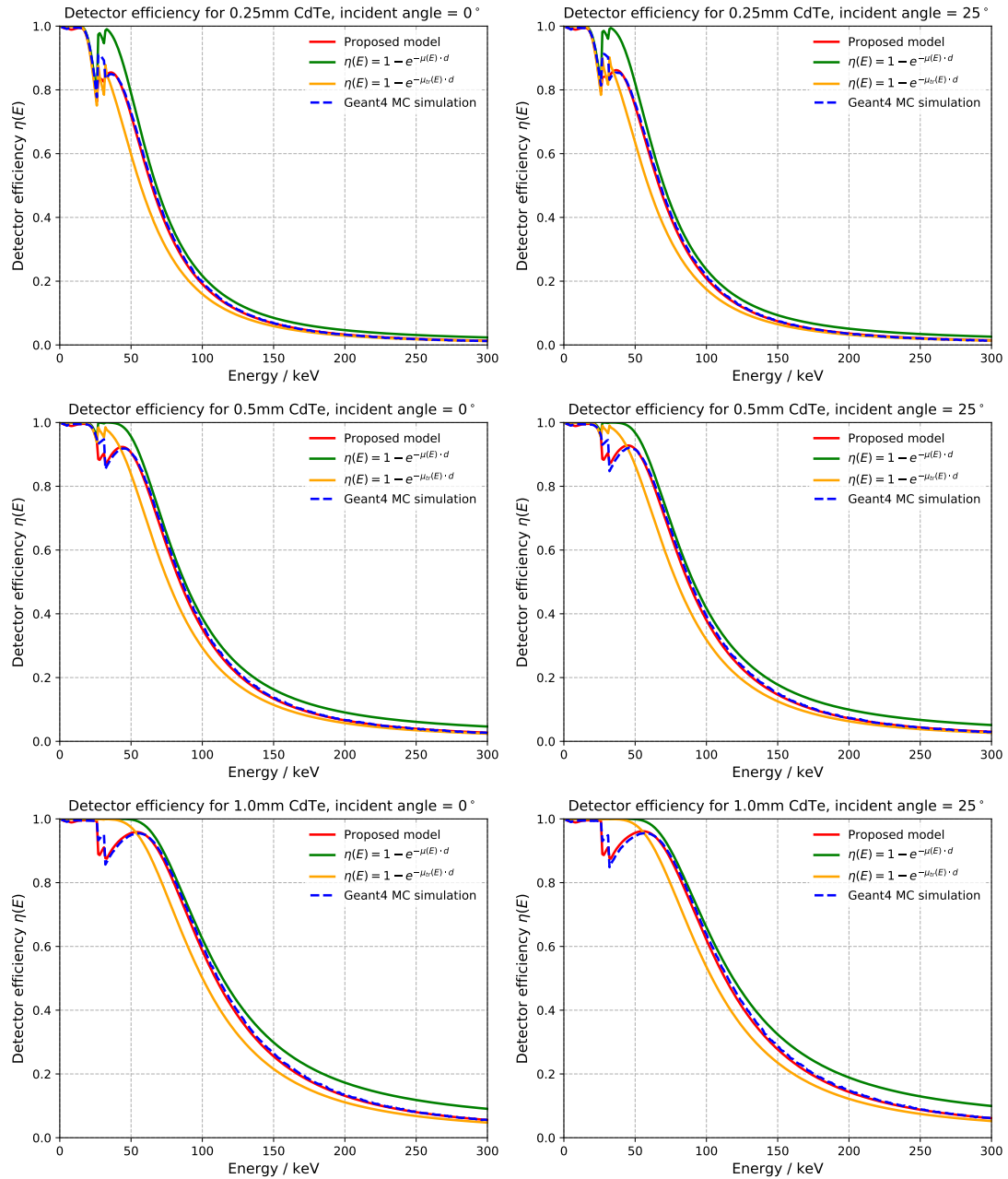
Figure 3.13: Comparison of different models to calculate the detector efficiency of CdTe. A Geant4 MC simulation (dashed blue curve) which is considered as ground truth serves as reference. Simulations were performed for detector thicknesses of 0.25 mm (top row), 0.5 mm (middle row), and 1.0 mm (bottom row). The two columns show different incident angles (left: 0°, right: 25°).
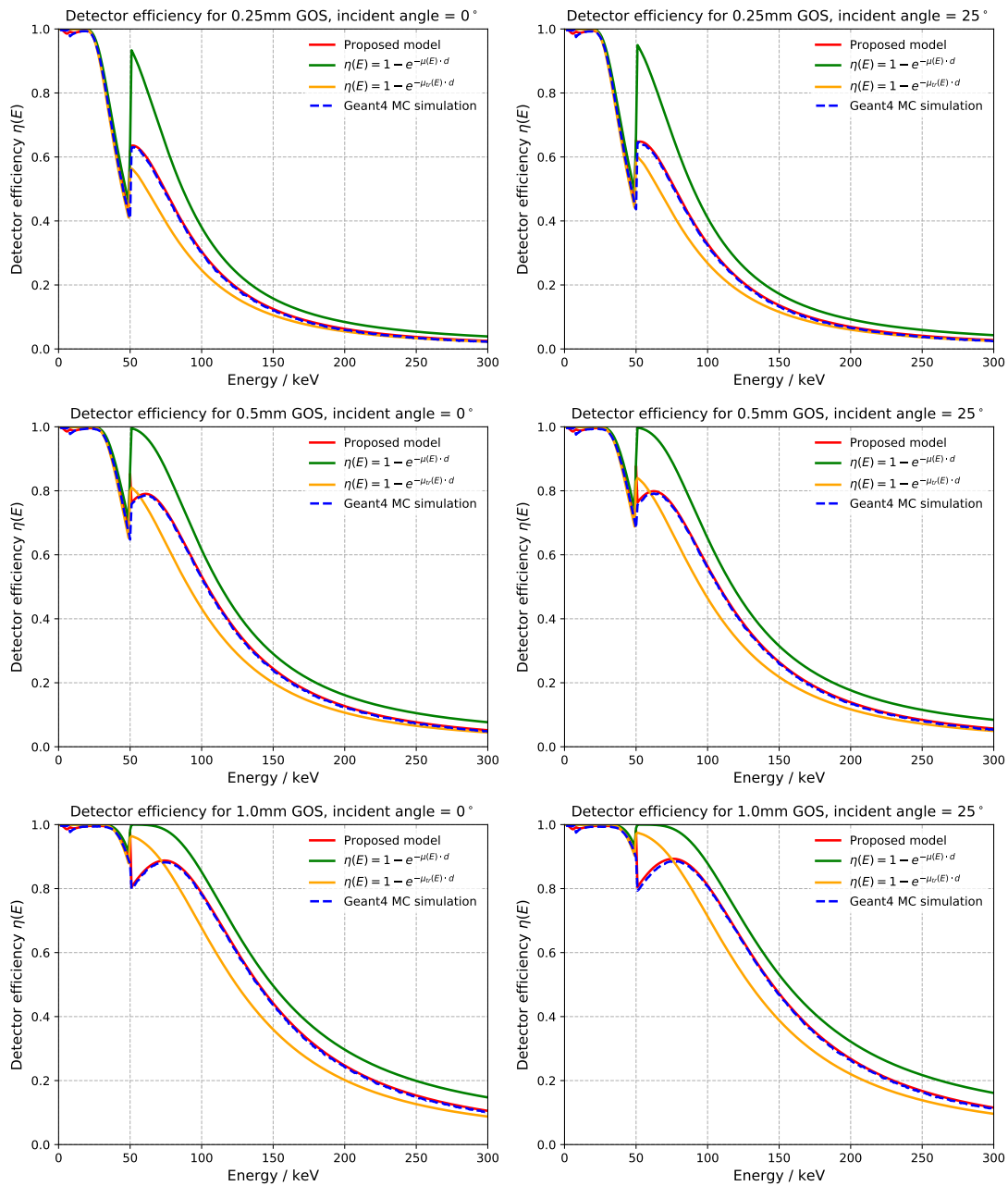
Figure 3.14: Comparison of different models to calculate the detector efficiency of Gd_2O_2S. A Geant4 MC simulation (dashed blue curve) which is considered as ground truth serves as reference. Simulations were performed for detector thicknesses of 0.25 mm (top row), 0.5 mm (middle row), and 1.0 mm (bottom row). The two columns show different incident angles (left: 0°, right: 25°).

3.3 Discussion

This chapter presents different models to describe the CT data acquisition process. Besides accuracy, the focus was set on computational performance. Therefore, analytical approximations were derived whenever possible. The particular models are briefly discussed in the following.

X-Ray Spectra The simulation of x-ray spectra is based on a model proposed by Tucker et al. to describe spectra of reflection x-ray tubes [103]. It could be shown that recalibrating the open parameters allows to apply this model to simulate spectra of transmission x-ray tubes, as they are used in this thesis. Here, the mean absolute percentage error with respect to reference MC simulations is about 10 %. These deviations are probably a result of the simplified treatment of the electron energy distribution. Here, electrons contribute to the bremsstrahlung spectrum according to their average kinetic energy as given by the Thomson-Whiddington law. However, the kinetic energy of of 200 keV electrons penetrating a tungsten target is, for instance, about 140 keV after 10 µm. Thus, applying the Tucker model to thin transmission targets leads to the complete disregard of lower energy electrons. To increase the accuracy of the predicted spectra, it seems to be necessary to include the depth-dependent electron energy distribution within the model. This has been done in a slightly different approach by Poludniowski et al. using look-up tables of precalculated MC electron energy distributions [112], [113]. While being potentially more accurate, the requirement of MC simulations makes it rather impractical. Alternatively, the x-ray spectra could be estimated directly from calibration measurements as proposed e.g. by Leinweber et al. [114]. However, the results presented in chapter 4 indicate that the proposed approach provides a sufficient accuracy to fit the purpose of artifact correction.

Focal Spot Distribution The effect of non-ideal focal spot distributions is modeled here using a convolution-based approach. It was shown that this approach is equivalent to a physically exact simulation for infinitely thin objects. Increasing the object thickness leads to a reduced accuracy, especially in case of large focal-spot sizes. Due to the perspective geometry, the effective focal spot distribution changes in source-detector direction. Therefore, a better accuracy might be achieved by subdividing the object along this axis and to process every subobject with a specific convolution kernel. However, for most practical cases the proposed approach seems to provide a sufficient accuracy while being computationally efficient. Compared to a physically more exact simulation that required, let's say, at least 5×5 samples of the focal spot, the runtime is about 25 times faster.

Besides the computational model, the focal spot distribution itself needs to be determined. Here, a pragmatic approach was introduced that derives the focal spot distribution from a calibration sphere measurement. It was shown that the distribution derived from such a measurement also applies to subsequent CT acquisitions. Since the calibration sphere is typically measured prior to the actual CT measurement, the

proposed approach can be easily integrated in the workflow of a CT scan without the need for additional calibration procedures.

X-Ray Scattering X-ray scattering is the only effect that was not modeled using analytic approximations, but rather MC methods. To achieve a reasonable performance, several variance reduction techniques were implemented. This reduces the typical runtimes to about 20 s for a single projection. Several applications, however, would greatly benefit from faster scatter estimation approaches. To meet this demand, the so-called deep scatter estimation (DSE) is introduced in this thesis. Therefore, the reader is referred to chapter 5 for a more comprehensive discussion on x-ray scatter estimation.

Detector Efficiency The detector efficiency model relies on a precise simulation of first order interactions and an approximate treatment of higher order interactions. It could be demonstrated that the proposed model applies to different detector materials and different detector dimension. Here, it yields estimates of the detector efficiency that differ by less than 4 % from MC simulations. Considering the computational performance, the model yields the detector response for a 300 kV spectrum in about 0.9 s on a single core central processing unit (CPU). Thus, the proposed model provides a fast alternative to MC methods that clearly outperforms commonly used approximations of the detector efficiency.

4 | Simulation-Based Artifact Correction

This chapter describes a novel approach to correct for CT artifacts using simulations of the CT measurement process. After a brief review of prior work, the potential of the simulation-based artifact correction (SBAC) is demonstrated for simulated data as well as measurements of single- and multi-material components and compared to state of the art artifact correction approaches. It has to be noted that parts of this chapter have been published in reference [102].

4.1 Background and Prior Work

As highlighted in chapter 1, CT reconstructions of highly attenuating or multi-material components are often corrupted by artifacts which impair the metrological assessment. Existing artifact correction approaches that have been proposed to address this issue are briefly reviewed in the following. Since medical CT has to deal with similar artifacts, especially in case of patients with metallic implants, most of them have been initially proposed for medical applications.

In principal, artifact correction approaches can be divided into iterative approaches, post- or precorrection approaches. Iterative approaches, which became very popular in clinical CT in recent years, try to solve the reconstruction problem in an iterative manner. Therefore, they usually set up a forward model that predicts projection data given an estimate of the CT image. This estimate is refined in every iteration step by calculating an update based on the discrepancy between the prediction and the measured projection data. This procedure is repeated until convergence is reached, i.e. the prediction fits the measurement. Depending on the design of the forward model, iterative approaches are able to account for the most common CT artifacts [22]–[24], [115], [116]. Furthermore, they allow for the incorporation of prior knowledge such as the shape of the measured component or image sparsity for instance [17], [21], [117]–[119]. However, the need for at least a few forward and backprojections makes iterative approaches computationally expensive. Especially in metrological CT that has to deal with comparably large data sets this turns out to be a major limitation.

Postcorrection approaches are often used as a faster alternative. These approaches apply correction terms, which are motivated by empirical or physical considerations, to

an analytic reconstruction. Therefore, they typically set up models with a given number of open parameters that can potentially account for a certain artifact. Subsequently, the open parameters are adjusted such that an appropriate metric, sensitive to artifacts in image domain, is minimized [16], [19], [25], [120]. Thus, no dedicated prior knowledge such as the x-ray spectrum for instance is required. In addition to computational performance this can be seen as further advantage compared to iterative methods. More recently, the use of DCNNs has been proposed as another realization of postcorrection approaches [34]–[36].

Besides being applied in image domain, correction terms can also be applied in projection space prior to the reconstruction. There are several flavors of these precorrection approaches. Considering multi-material components containing metal, so-called metal artifact reduction (MAR) algorithms are frequently applied. Initially MAR approaches have been proposed for the correction of artifacts caused by metalic implants in medical CT [121]. Their basic principle relies on the identification of the metal trace within the acquired projection data and its subsequent replacement with some sort of interpolated data [20], [122]–[124]. The corrected projections are then reconstructed and the metal is reinserted to the CT image. While this strategy is potentially useful for multi-material components with a small amount of metal [125], interpolation errors may degrade the correction result for higher metal fractions.

The correction of single-material components, in contrast, often relies on approaches similar to water precorrection in clinical CT [126]. These approaches aim to invert the relationship between the measured projection data and the intersection length through the object being investigated. Since, the projection values strictly increases with the intersection length (see equation (2.8)), there is a unique solution of that inversion. Typically, it is implemented as an analytic function or a look-up table that can subsequently be used to map a given measurement to ideal data [18], [127]. Here, the correction function is either derived from theoretical considerations, i.e. by numerical inversion of a certain physical model describing the data acquisition or by performing calibration measurements of a known component. However, strictly speaking only beam hardening artifacts can be corrected using this approach. In case of other artifacts such as x-ray scattering or off-focal radiation, there is no unique relationship between projection values and intersection lengths. Consequently, these effects have to be considered a priori. While there is only little prior work on off-focal correction, scatter correction is reviewed more detailed in chapter 5.

4.2 Material and Methods

4.2.1 Basic Principle

Analytic reconstruction approaches assume the measured projection data p to be the monochromatic x-ray transform (denoted as X) of the measured component:

$$p = \mathsf{X}f, \tag{4.1}$$

with f being the CT image, i.e. the spatial distribution of the component's attenuation coefficient. Similar to the 2D case described in section 2.1.4, the function f can be reconstructed from a set of projection images by applying the inverse x-ray transform which is typically implemented as FBP.

However, conventional CT systems usually acquire projection data q which deviate from the ideal case given by equation (4.1). There are several effects that may be responsible for these deviations. Here, beam hardening, off-focal radiation, x-ray scattering and partial volume effects have been identified to be the most severe ones. Considering these effects, q can be expressed as described in chapter 3:

$$q(\boldsymbol{d}) = -\ln\left[\frac{\int G(\boldsymbol{k})\int \tilde{w}(E)e^{-\int_0^1 \mu(\boldsymbol{s}+\boldsymbol{k}+\lambda\cdot(\boldsymbol{d}-(\boldsymbol{s}+\boldsymbol{k})),E)d\lambda}dE\,d^2k}{\int G(\boldsymbol{k})\int \tilde{w}(E)dE\,d^2k} + S\right], \qquad (4.2)$$

with S denoting the scatter distribution, $G(\boldsymbol{k})$ being the spatial distribution of x-rays within the focal spot plane, $\tilde{w}(E)$ being the detected x-ray spectrum, $\int_0^1 \mu(\boldsymbol{s}+\boldsymbol{k}+\lambda\cdot(\boldsymbol{d}-(\boldsymbol{s}+\boldsymbol{k})),E)d\lambda$ being the line integral over the attenuation coefficient and $\boldsymbol{s}$ and $\boldsymbol{d}$ denoting the position of the source and detector element, respectively.

A comparison of equation (4.1) and equation (4.2) shows that $p \neq q$. Thus, an analytic reconstruction, i.e. the application of the inverse x-ray transform operator X^{-1}, does not yield f but rather an image g that is corrupted by artifacts a:

$$\begin{aligned}
g = \mathsf{X}^{-1}q &= \mathsf{X}^{-1}p + \mathsf{X}^{-1}(q-p) \\
&= f + \mathsf{X}^{-1}(q-p) \\
&= f + a.
\end{aligned} \qquad (4.3)$$

Accordingly, the simulation-based artifact correction (SBAC) aims to derive an estimate $\tilde{a}$ of the artifact term such that a corrected image can be calculate as:

$$f_{\text{SBAC}} = g - \tilde{a} = g - \mathsf{X}^{-1}(\tilde{q}-\tilde{p}), \qquad (4.4)$$

with $\tilde{q}$ and $\tilde{p}$ being estimates of real and ideal projection data, respectively. The basic workflow of the SBAC is depicted in figure 4.1. It relies on precise simulations of the CT measurement process to derive the estimates $\tilde{q}$ and $\tilde{p}$ which are then used to calculate $\tilde{a}$. Here, the simulations are based on a prior model f_{prior} of the component, e.g a computer-aided design (CAD) model. However, in most cases f_{prior} can also be generated directly from the measurement by a segmentation of the uncorrected reconstruction g:

$$f_{\text{prior}} = Tg, \qquad (4.5)$$

where T represents the corresponding segmentation operator. If not stated otherwise, the operator T is implemented here as a marching cubes algorithm that yields a surface mesh of the component [128].
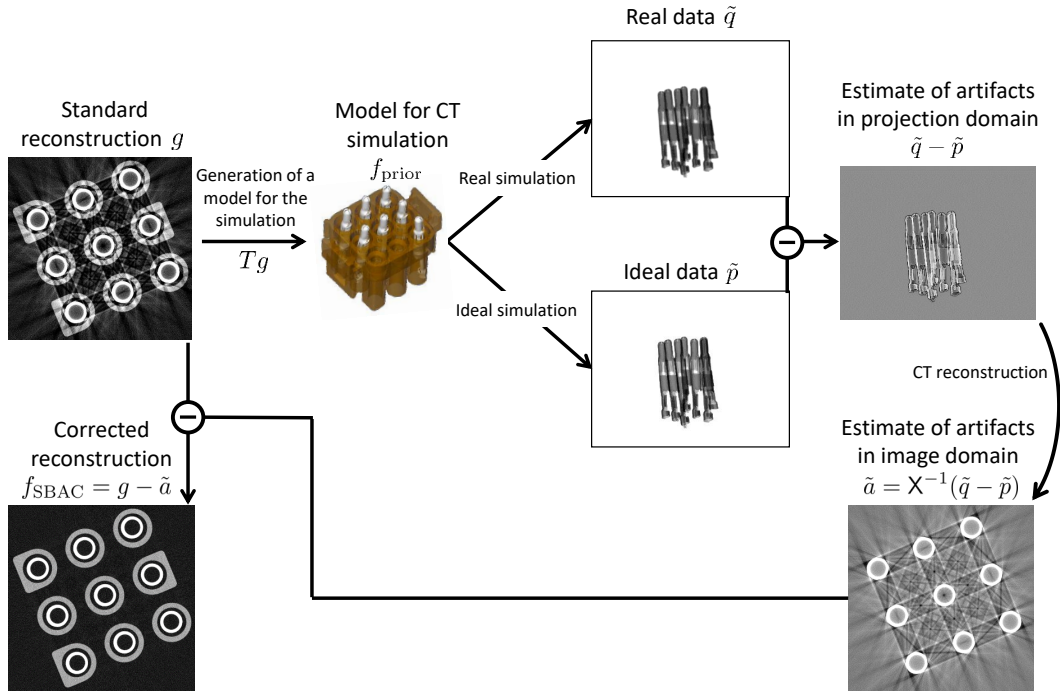
Figure 4.1: Schematic of the SBAC workflow. An initial reconstruction g is used to generate a prior model. Based on that model, a real and an ideal simulation is performed. Their difference represents the artifacts within the measurement and can be used to calculate a correction term for the initial reconstruction.

4.2.2 Simulation of Ideal and Real Projection Data

To simulate ideal projection data p as well as real projection data q, the SBAC applies the physical models derived in chapter 3 as follows.

i.) Simulation of Ideal Data

Ideal data are simulated as a CT measurement with a monochromatic x-ray source, a point-like focal spot, an ideal detector and no x-ray scattering. Thus, an ideal measurement can be described by equation 4.2 with $G(\boldsymbol{k}) = \delta(0)$, $\tilde{w}(E) = \delta(E_0)$ and $S = 0$:

$$\tilde{p}(\boldsymbol{d}) = -\ln\left[\frac{\int \delta(0) \int \delta(E_0) e^{-\int_0^1 \mu(\boldsymbol{s}+\boldsymbol{k}+\lambda\cdot(\boldsymbol{d}-(\boldsymbol{s}+\boldsymbol{k})),E)d\lambda} dE\, d^2k}{\int \delta(0) \int \delta(E_0) dE\, d^2k}\right] \tag{4.6}$$

$$= -\ln\left[e^{-\int_0^1 \mu(\boldsymbol{s}+\lambda\cdot(\boldsymbol{d}-\boldsymbol{s}),E_0)}\right] \tag{4.7}$$

$$= \int_0^1 \mu(\boldsymbol{s} + \lambda \cdot (\boldsymbol{d} - \boldsymbol{s}), E_0). \tag{4.8}$$

Here, E_0 represents a reference energy that can, in principle, be set to an arbitrary non-negative value. However, to have similar CT values as the artifact image, it is set equal to the center of mass of the spectrum:

$$E_0 = \frac{\int \tilde{w} E \, dE}{\int \tilde{w}(E) dE}. \tag{4.9}$$

Furthermore, it is assumed that any material of the object to be investigated has a homogeneous density distribution. In that case the line integral over the attenuation coefficient given in equation (4.8) can be calculated as:

$$\int_0^1 \mu(\boldsymbol{s} + \lambda \cdot (\boldsymbol{d} - \boldsymbol{s}), E_0) d\lambda = \sum_i \mu_i(E_0) \cdot l_i(\boldsymbol{s}, \boldsymbol{d}), \tag{4.10}$$

with $l_i(\boldsymbol{s}, \boldsymbol{d})$ being the intersection length through the object for a straight line from $\boldsymbol{s}$ to $\boldsymbol{d}$. Thus, it can be calculated by determining the intersection points of a given ray with the surface mesh that is used as prior model f_{prior}.

ii.) Beam Hardening and Partial Volume Effects

Beam hardening is simulated according to equation (3.14) using the x-ray spectrum model described in section 3.1.2 and the detector model described in section 3.1.5. Similar to the simulation of ideal data, a homogeneous density distribution is assumed such that the line integral over the attenuation coefficient can be calculated as $\sum_i \mu_i(E) \cdot l_i(\boldsymbol{s}, \boldsymbol{d})$.

To account for partial volume effects, 4×4 subsamples are calculated and averaged in intensity domain for every detector pixel.

iii.) Simulation of X-Ray Scattering

X-ray scattering is simulated based on the prior model f_{prior} using the MC simulation described in section 3.1.3. To increase the performance of the MC simulation, the size of the prior model is decreased by resampling onto a lower resolution voxel grid. Since x-ray scatter distributions are known to be low frequent, the corresponding loss of resolution has only a minor effect on the accuracy of the scatter prediction but speeds it up significantly. Similarly, the detector is downsampled prior to the MC simulation. Finally, the scatter prediction is upsampled again to the original detector size.

iv.) Simulation of Off-Focal Radiation

The focal spot distribution including off-focal radiation is simulated using the convolution-based model derived in section 3.1.4. The corresponding convolution kernel representing the projection of the focal spot distribution into the detector plane was determined according to section 3.1.4 using a calibration sphere measurement.

v.)   Simulation of Cone-Beam Artifacts

The use of circular scan trajectories in combination with a cone-beam setup leads to the introduction of so-called cone-beam artifacts as this trajectory only allows for an exact CT reconstruction within the mid-plane. If $\tilde{q}$ and $\tilde{p}$ are simulated in cone-beam geometry, both of them show similar cone-beam artifacts as the measurement. Thus a correction is not possible. Therefore, the simulation has to be performed using a geometry that allows for an exact reconstruction. Here, a parallel beam geometry is used for that purpose. In that case, however, the reconstruction operator X^{-1} is different for $\tilde{q}$ and $\tilde{p}$. Consequently, the subtraction has to be calculated post to reconstruction:

$$\tilde{f} = g - \tilde{a} = g - (\mathsf{X}^{-1}\tilde{q} - \mathsf{X}_{\parallel}^{-1}\tilde{p}), \tag{4.11}$$

and one additional reconstruction is required.

4.2.3   Simulation Study

A quantitative evaluation of the SBAC was performed using simulated projection data. As prior for the simulation a CAD model, representing a multi-material component composed of poly methyl methacrylate (PMMA) and copper, was designed. Projections were simulating according to equation (4.2) assuming a cone-beam setup with a 1944×1536 flat detector and a pixel size of 150×150 µm. The detected x-ray spectrum was simulated with a tube voltage of 225 kV using the models described in section 3.1.2 and section 3.1.5. X-ray scattering was simulated using the MC simulation described in section 3.1.3 while the effect the non-ideal focal spot distribution was modeled by a convolution with a Gaussian function in intensity domain. The ground truth is given by the reconstruction of an ideal simulation as described in section 4.2.2.

4.2.4   Measured Data

In order to demonstrate the potential of the SBAC for real data, measurements were conducted on a commercial industrial CT system (Werth TomoScope® 200) that is equipped with a 225 kV micro-focus x-ray tube and a 3888×3072 flat detector with a pixel size of 74.8×74.8 µm. For any measurement, the detector was operated in the 2×2 binning mode leading the an effective pixel size of 149.6×149.6 µm. However, it has to be noted that the pixel size in the isocenter, which is the important quantity to characterize the spatial resolution, differs for all measurements as they are acquired with different magnifications. The single- and multi-material components that have been measured for this study are shown in figure 4.2. The measurement parameters as well as the maximum and the mean intersection lengths are summarized in table 4.1. All measurements were reconstructed analytically using the FBP-type Feldkamp-David-Kress algorithm [8].
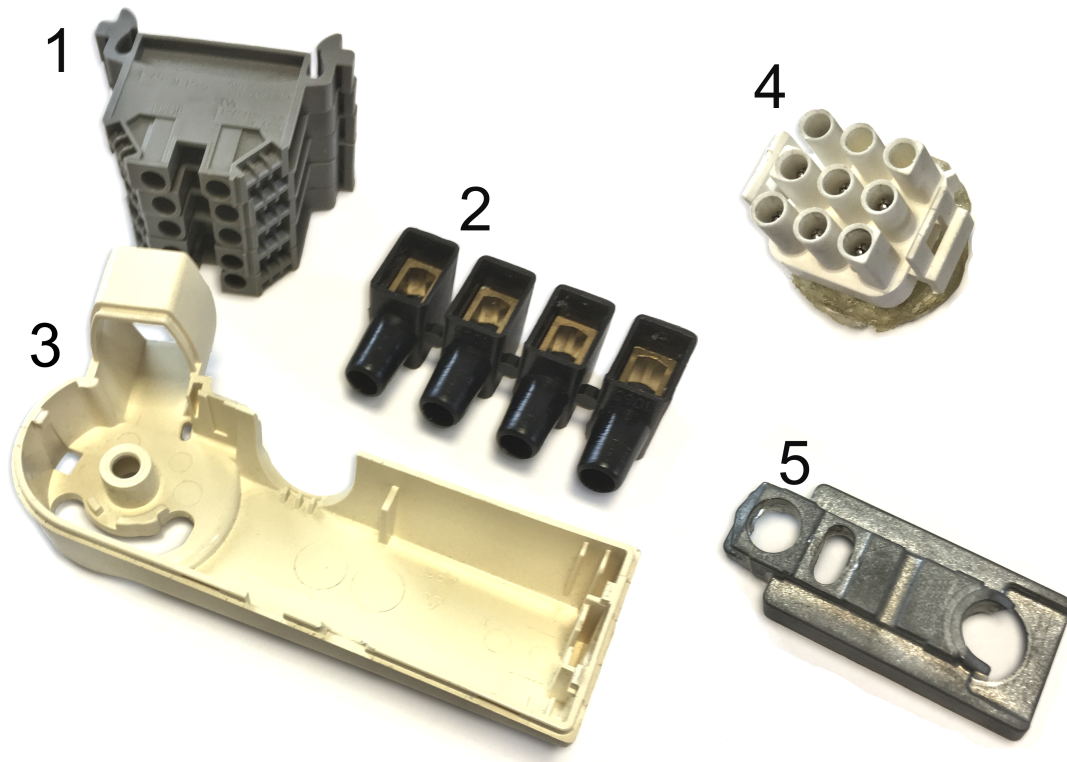
Figure 4.2: Photograph of the components to test the SBAC. 1. Electrical plug 1, 2. Luster terminal, 3. Plastic Inhalator, 4. Electrical plug 2, 5. Die-cast zinc hinge.

Table 4.1: Acquisition parameters of single-material and multi-material measurements as well as the mean and the maximum intersection length (L). It has to be noted that these measures refer only to the metal intersection length in case of multi-material components.

Sample	Voltage	Current	Prefilter	Projections per 360°	Pixel size at isocenter	Mean L / Max. L
Plug 1	225 kV	170 µA	1.2 mm Sn	1200	41 µm	3.0 mm / 15.9 mm
Luster terminal	225 kV	170 µA	1.2 mm Sn	1200	41 µm	4.9 mm / 29.8 mm
Inhalator	160 kV	90 µA	0.5 mm Al	800	80 µm	4.3 mm / 42.6 mm
Plug 2	225 kV	170 µA	1.2 mm Sn	1200	25 µm	1.1 mm / 6.8 mm
Zinc hinge	215 kV	180 µA	1.0 mm Sn	800	43 µm	3.8 mm / 20.5 mm

4.2.5   Evaluation of Single-Material Measurements

The focus of the single- material measurements was set on dimensional accuracy. A dimensional evaluation of a given component is usually performed by comparing the surface mesh calculated from the CT reconstruction to a certain reference such as a CAD model for instance. The presence of CT artifacts, however, may cause significant deviations of that surface mesh from the real surface. Here, a quantitative evaluation of these deviations before and after applying the SBAC is performed by a comparison to a reference measurement that is considered as ground truth. In case of the die-cast zinc hinge, a tactile measurement with a commercial coordinate measurement machine (Werth Touch Probe TP 200) serves as reference. Therefore, a given number of predefined points on the surface of the die-cast zinc hinge is sampled using a tactile probe (maximum permissible probing error = 2 µm). Subsequently, the surface mesh extracted from the CT reconstruction was compared to that point cloud.

The correction of cone-beam artifacts using the SBAC, was tested for a plastic inhalator that was measured with small and a large cone-angle. While the the acquisition with the large cone-angle shows severe cone-beam artifacts, the other acquisition is almost free of artifacts and was therefore used as a reference.

4.2.6   Evaluation of Multi-Material Measurements

In general, multi-material components cannot be assessed using a tactile measurement without disassembling them. Therefore, a quantitative evaluation in terms of a comparison to an ideal reference was performed for simulated data which were generated as described in section 4.2.3. To evaluate the performance of the SBAC for measured data, corrected images were compared to two state of the art artifact correction approaches: the normalized metal artifact reduction (NMAR) [20] and, and an iterative reconstruction with total variation (TV) regularization [17], [129]. The NMAR belongs to the inpainting-based correction approaches that identify the metal trace within the acquired projection data and replace it by interpolated data. Additionally, NMAR makes use of prior knowledge that is incorporated to improve the accuracy of the interpolation. The iterative reconstruction, in contrast, derives a corrected image by minimizing the following cost function C in an iterative manner for f:

$$C = ||\mathsf{X}f - q|| + \lambda \cdot \mathrm{TV}(f), \tag{4.12}$$

where λ is a weighting factor and $\mathrm{TV}(f)$ refers to the total variation operator that is defined a:

$$\mathrm{TV}(f) = \int |\nabla f| dV. \tag{4.13}$$

4.3 Results

4.3.1 Simulation Study

The projection data were simulated according to section 4.2.3, and were reconstructed using the analytic Feldkamp–David–Kress algorithm [8]. This initial reconstruction was used as a prior model for the SBAC. For that purpose, it was segmented into air, plastic and metal. As the metal artifacts are mainly propagated into the plastic part, the corresponding plastic segmentation is degraded similarly. The metal part can usually be segmented more accurately. Therefore, the simulation of beam hardening, off-focal radiation and partial volume effects was based on the metal prior only. It is assumed that this strategy has only minor impact on the accuracy of the correction as these effects are caused predominantly by the metal part. The simulation of x-ray scattering, in contrast, was based on both the plastic prior as well as the metal prior. Since scatter distributions are usually smooth, inaccuracies of the segmentation can be neglected for that purpose.

CT reconstructions showing the correction result of the SBAC as well as the two reference approaches introduced in section 4.2.6 are displayed in figure 4.3. Here, the NMAR as well as the iterative reconstruction cannot remove the streak artifacts efficiently. The SBAC, in contrast, leads to images that are almost free of artifacts. Considering the difference to the ground truth, there are only small deviations mainly related to the CT value of the metal part.
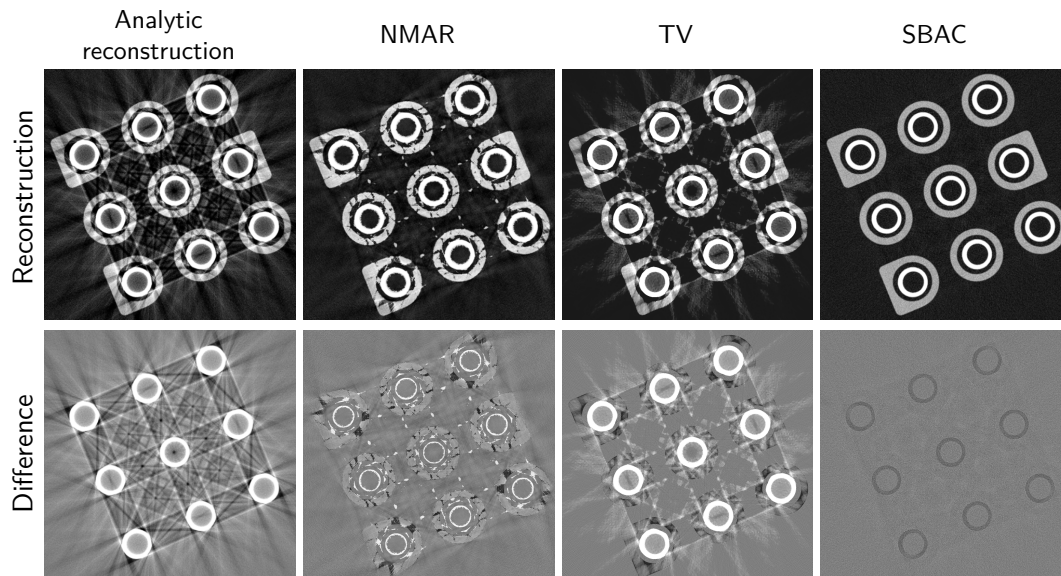
Figure 4.3: Top row: CT reconstructions of simulated data using different correction approaches (the gray scale is centered to the attenuation of plastic: $C = 0.012$ mm^{-1}, $W = 0.030$ mm^{-1}). Bottom row: Difference to the ground truth (the gray scale is centered to zero: $C = 0.00$ mm^{-1} / $W = 0.05$ mm^{-1}).

4.3.2 Measured Data

i.) Single-Material Components

A qualitative evaluation of the single-material measurements is shown in figure 4.4. Here, the zinc hinge measurement shows CT artifacts that are caused by x-ray scattering, beam hardening as well as off-focal radiation. The reconstruction of the inhalator measurement, in contrast, shows mainly cone-beam artifacts. In any case, the application of the SBAC yields CT images that are almost free of artifacts. However, since the visual impression does not necessarily correlate with the dimensional accuracy of the CT measurement, a dimensional evaluation was performed in addition. To do so, a surface mesh was calculated from every CT reconstruction and compared against a reference that is not corrupted by artifacts. The corresponding results are shown in figure 4.5. Considering the die-cast zinc hinge, there are large deviations from the tactile reference measurement that exceed the tolerances specified by the manufacturer by up to 400 %. These deviations can be reduced significantly by applying the proposed SBAC. Here, the dimensional assessment shows a good agreement between the CT and the tactile probe (note that the deviations along the edges of the component result from missing sample points of the tactile measurement).

The inhalator measurement demonstrates that similar results can be achieved for CT reconstructions that are corrupted by cone-beam artifacts. Here, the surface meshes were compared against a surface mesh calculated from a CT measurement with a narrow cone-angle. Without correction there are deviations, especially in the periphery of the field of measurement where the cone-angle is large, that exceed the tolerances of the manufacturer more than 200 %. Applying the SBAC reduces these deviations well below values of 40 %.

ii.) Multi-Material Components

The correction of multi-material components using the SBAC was evaluated for three typical components with different metal fractions or metal intersection lengths, respectively (see table 4.1). Since multi-material components cannot be assessed entirely using a tactile probe, two commonly used artifact correction algorithms, namely NMAR and an iterative reconstruction with TV regularization, were implemented as a reference. The corresponding CT reconstructions as well as an analytic reconstruction and the SBAC result are shown in figure 4.6. As expected, the presence of metal leads to severe artifacts appearing as dark steaks if no correction is applied. In contrast to clinical CT where NMAR usually leads to a considerable improvement of image quality, it fails to improve the multi-material measurements here. Similarly, the iterative reconstruction approach does not remove streak artifacts but only leads to a small improvement. Compared to the analytic reconstruction and the reference approaches, the SBAC is able to correct for almost all artifacts and yields CT volumes that allow for a clear discrimination between plastic and metal.
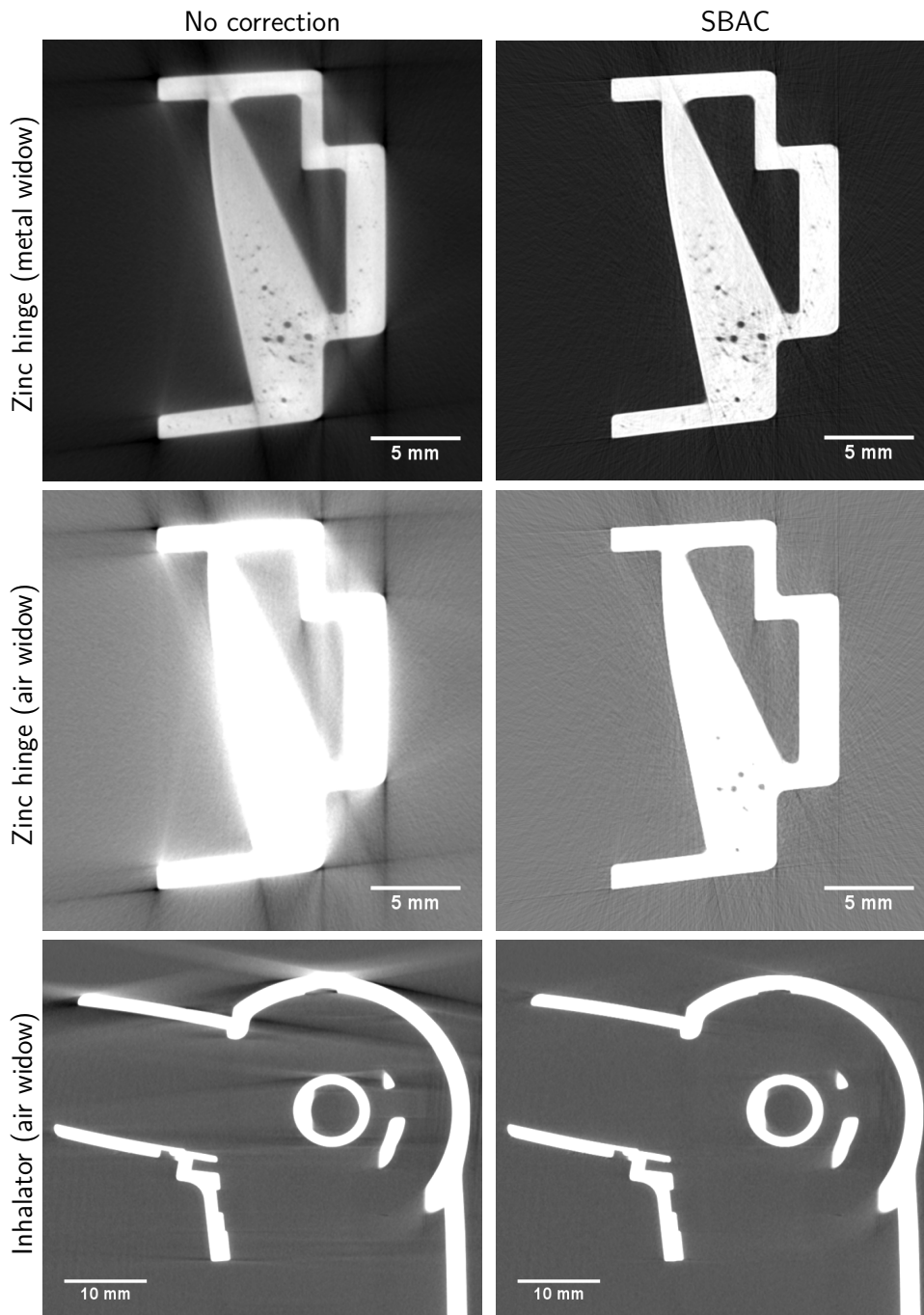
Figure 4.4: Reconstruction of the measured die-cast zinc hinge (top) and the inhalator (bottom) without and with simulation-based artifact correction (left and right). The zinc hinge is displayed at two different window levels centered to metal and centered to air (top and middle).
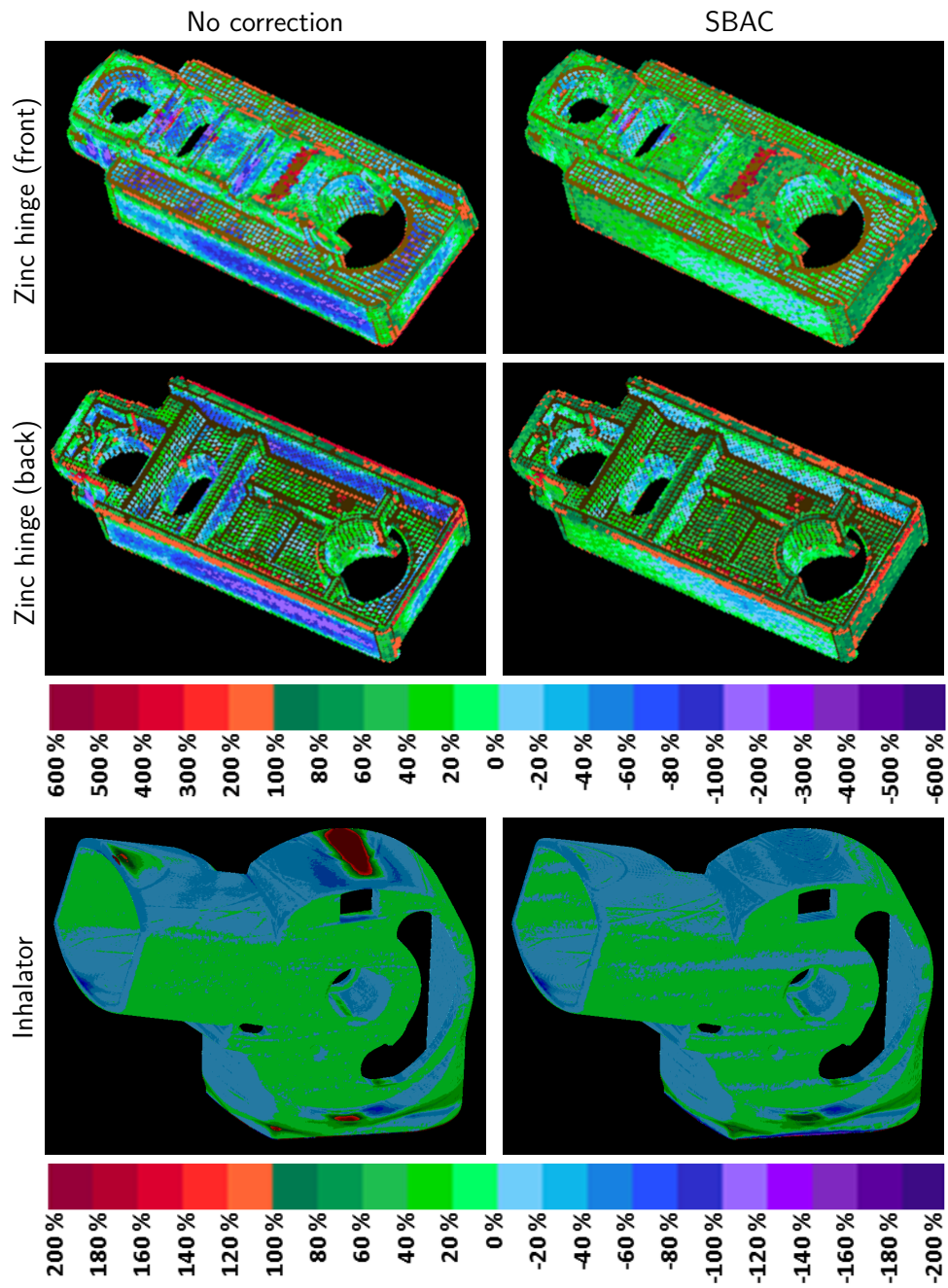
Figure 4.5: Dimensional evaluation of the CT measurement of the die-cast zinc hinge (top) and the inhalator (bottom) with and without simulation-based artifact correction. The color scale refers to the tolerances specified by the manufacturer. Here, 100 % and -100 % are the maximum acceptable deviations of the CT measurement from the reference measurement.
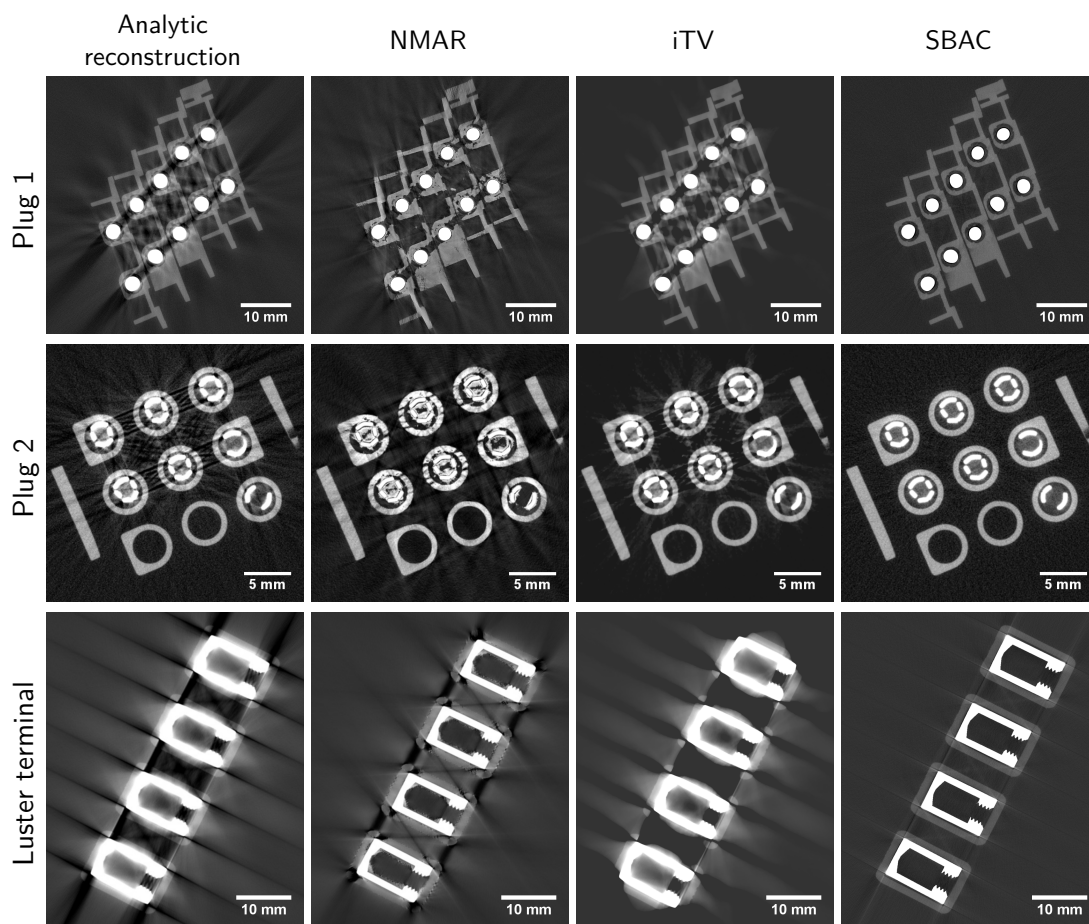
Figure 4.6: CT reconstructions of two different multi-material plugs (top row C = 0.012 mm^{-1} / W = 0.030 mm^{-1}, middle row C = 0.02 mm^{-1} / W = 0.06 mm^{-1}) and a luster terminal (bottom row C = 0.03 mm^{-1} / W = 0.10 mm^{-1}).

4.4 Discussion

This chapter presents the simulation-based artifact correction (SBAC), a novel approach to correct for CT artifacts using precise simulations of the CT measurement process. While the basic principle of the SBAC allows to correct for any artifact that is modeled appropriately within the simulation, the focus of this study was set on the correction of beam hardening, x-ray scattering, off-focal radiation, partial volume effects and cone-beam artifacts which are considered to be the most prominent artifacts in CBCT. In order to evaluate the potential of the SBAC, it was applied to simulated and measured data of highly attenuating as well as multi-material components that currently pose a major challenge to metrological CT. A first observation of these experiments concerns the model of the component that is required as prior for the simulation. In principle, the CAD model that comes with most components can be used for that purpose. However, for several reasons it is beneficial to be independent of additional CAD data. First and most obvious, there are cases where no CAD model is available or where the CAD model only describes parts of the component to be investigated. Second, the use of the CAD model requires to align it with the measured component. Thus, an additional registration routine is required that needs to be robust against CT artifacts. Third, there are certain applications that prohibit the use of CAD data within an artifact correction framework as this can potentially bias the corrected measurement towards the CAD model. Therefore, the prior model was calculated in this study directly from an initial CT reconstruction by segmentation. It turns out that accurate correction results can be achieved even if that segmentation deviates to some extend from a true representation of the component. In case of single material components these deviations, which correspond to several voxels, can be seen in the left column of figure 4.5. As they do not change the appearance of most CT artifacts the SBAC correction term also applies to images that differ by a few voxels. Considering multi-material components, also the complete negligence of the plastic part in the prior has only a minor influence on the accuracy of the correction (see figure 4.3). This is because the contribution to CT artifacts, except for x-ray scattering, is dominated by the metal part of the object. Therefore, the difference between the ideal and the real simulation of the metal part seems to be a good estimate for the present artifacts. Since this does not hold true for x-ray scattering, it is the only artifact that requires the consideration of the plastic part of the component.

Furthermore, it could be shown that the SBAC clearly outperforms other commonly used artifact correction approaches, namely the NMAR and an iterative reconstruction with TV regularization. Both of these approaches are adapted from medical CT and show convincing correction results in that field. Here, however, they are not able to remove the artifacts in a similar way. This is a result of the high metal fraction of the components investigated in this study. As the NMAR relies on the replacement of measured data with interpolated data, it works best if the metal fraction is small. Similarly, the iterative reconstruction only leads to a minor improvement of image quality whereas the SBAC yields images that are almost free of artifacts.

Besides the visual impression of image quality, the dimensional accuracy of a certain

artifact correction approach is an important issue for several applications. Since these quantities are not necessarily correlated, a dimensional evaluation was performed for the single-material components by a comparison to an artifact-free reference. Here, the results demonstrate that the SBAC is able to reduce these deviations significantly compared to a standard analytic reconstruction. Therefore, the SBAC increases the validity of the metrological assessment and makes CT applicable even to highly attenuating components or multi-material components, respectively.

5 | Real-Time Scatter Estimation using a Deep Convolutional Neural Network

X-ray scattering is a major source of artifacts in most CT applications. Especially in CBCT, which has to deal with high scatter-to-primary ratios, the CT reconstructions suffer from severe streak and cupping artifacts. As the presence of scatter artifacts has a strong impact on the quality of any metrological assessment, accurate scatter correction is necessary. However, considering existing scatter correction approaches, there is always a trade-off between accuracy and computational performance. Therefore, a novel approach that is able to overcome this drawback using a deep convolutional neural network (DCNN) is presented in the following. After a brief review of prior work on scatter correction, the proposed approach is introduced and evaluated for simulated and measured data. It has to be noted that parts of the results have also been published in reference [130] and reference [131].

5.1 Background and Prior Work

In general, there are two typical strategies to reduce the impact of scattered x-rays on CT image quality: scatter suppression and scatter estimation. The former approach is based on the use of additional hardware, such as anti-scatter grids or collimators, which are designed to reduce the number of scattered x-rays reaching the detector [26]. Scatter estimation approaches, in contrast, aim to remove scattered x-rays that are present in the measured data. Therefore, they derive an estimate of the scatter distribution that is subtracted post to data acquisition [27]. One option to derive this estimate is to use dedicated hardware, e.g. primary modulation grids or beam blockers, that allows to distinguish between primary and scattered x-rays [132]–[139]. Other approaches use software-based solutions that set up empirical, physical or consistency-based models that predict or approximate x-ray scattering [19], [140]–[153].

The gold standard among these methods is Monte Carlo (MC) simulation which is able to model the entire physics of the CT data acquisition process, and therefore,

yields highly accurate scatter estimates [27]. On the downside, MC simulations are very time-consuming and cannot be applied in real-time using conventional hardware. Furthermore, they need prior information such as the material distribution and the density distribution that have to be estimated a priori [154].

So-called kernel-based scatter estimation (KSE) approaches are therefore used as a fast alternative to MC. Basically, there are two flavors of kernel-based scatter estimation (KSE) approaches. The first one estimates scattered x-rays as an integral transform of a scatter source term and a scatter propagation kernel [140], [141], [155]–[157]. In that process, the scatter source term, which is typically derived from a simplified theoretical model (e.g. only single scattering in forward direction is considered), represents the fraction of x-rays that are scattered along a straight line starting at the position of the x-ray source. The scatter propagation kernel reflects the spatial spreading of scattered x-rays and is usually calibrated to fit reference measurements or MC simulations. Since the multiplication of these two quantities represents the scatter distribution for a single ray, the total scatter distribution can be estimated by integrating the contribution of any ray between the x-ray source and a detector element.

Instead of using analytic models to approximate scattering along a certain line, other KSE approaches perform needle-beam MC simulations of primitive geometries, e.g. ellipsoids or cuboids, with varying dimensions [143], [145]–[147]. The corresponding scatter distributions are then stored as a look-up table. To estimate x-ray scattering of a given measurement each detector pixel is assigned one of the precalculated needle-beam scatter distributions according to a similarity metric. Similar to model-based KSE approaches, all contributions are integrated to calculate the total scatter distribution. To account for the different shape of the actual object and the primitive geometry that has been used to generate the look-up table, additional correction terms are often included within the correction framework [147].

However, being real-time capable, KSE approaches are by far less accurate than MC simulations. Furthermore, it needs several efforts to calibrate the open parameters of these models to fit different acquisition parameters as well as different components. Therefore, hybrid approaches have been proposed that incorporate MC information within the KSE framework [149], however, at the cost of computational performance. To achieve real-time performance without losing accuracy, this chapter introduces the deep scatter estimation (DSE) and demonstrates the potential for different industrial applications and compares it against state of the art KSE approaches.

5.2   Material and Methods

5.2.1   Kernel-Based Scatter Estimation

Kernel-based scatter estimation (KSE) approaches estimate scatter S_{est} by an integral transform of a so-called scatter source term $T(p)$ and a scatter propagation kernel G [26]:

$$S_{\text{est}}(\boldsymbol{d}) = \int T(p)(\boldsymbol{d}')G(\boldsymbol{d}, \boldsymbol{d}', \boldsymbol{c})d^3d', \tag{5.1}$$

with $\boldsymbol{d}$ denoting the detector coordinate, p denoting the logarithm of the primary intensity $p = -\ln(I/I_0)$ and $\boldsymbol{c}$ denoting a parameter vector that is typically calibrated such that the model reproduces reference measurements or MC simulations. In recent year several models to implement equation (5.1) have been proposed [140], [141], [143], [147], [149], [155]. Here a slight modification of the model of Ohnesorge et al. that has been proposed by Baer et al. is used as a reference [149]. This model defines the scatter source term as:

$$T(p) = K \cdot p \cdot e^{-p}, \tag{5.2}$$

which has the physical meaning of a forward scatter intensity and can be derived by the following consideration. Assume an x-ray photon is emitted at position $(0, 0, 0)$ and heads along $\boldsymbol{t}$ through an object with attenuation $\mu(\boldsymbol{r})$. According to equation (2.82), the probability of a scatter interaction between $\lambda \boldsymbol{t}$ and $(\lambda + d\lambda)\boldsymbol{t}$ is proportional to $\mu(\lambda \boldsymbol{t})e^{-\int_0^\lambda \mu(\lambda' \boldsymbol{t})d\lambda'}$. The probability of the scattered x-ray photon reaching the detector at position $\boldsymbol{d} = \lambda_{\boldsymbol{d}}\boldsymbol{t}$ is given as $e^{-\int_\lambda^{\lambda_{\boldsymbol{d}}} \mu(\lambda' \boldsymbol{t})d\lambda'}$. Accounting for all interactions on the line from $(0, 0, 0)$ to $\boldsymbol{d}$ requires to integrate the product of these probabilities over λ: $\int_0^{\lambda_{\boldsymbol{d}}} \mu(\lambda \boldsymbol{t})e^{-\int_0^\lambda \mu(\lambda' \boldsymbol{t})d\lambda'}e^{-\int_\lambda^{\lambda_{\boldsymbol{d}}} \mu(\lambda' \boldsymbol{t})d\lambda'}d\lambda$ which can be evaluated as $p \cdot e^{-p}$ with $p = \int_0^{\lambda_{\boldsymbol{d}}} \mu(\lambda \boldsymbol{t})$. The parameter K in equation (5.2) is an open parameter of the model and can be interpreted as a proportionality factor that accounts for the differential cross section for forward scattering that has been neglected in the derivation above.

The scatter kernel $G(\boldsymbol{d}, \boldsymbol{d}', \boldsymbol{c})$ is defined as:

$$G(\boldsymbol{d}, \boldsymbol{d}', \boldsymbol{c}) = \sum_{\pm} e^{-c_1((\boldsymbol{d}-\boldsymbol{d}')\hat{\boldsymbol{e}}_1 \pm c_2)^2} \cdot \sum_{\pm} e^{-c_3((\boldsymbol{d}-\boldsymbol{d}')\hat{\boldsymbol{e}}_2 \pm c_4)^2}. \tag{5.3}$$

This parametrization of G as a shift invariant function further simplifies the calculation of equation (5.1) as it can be expressed as a convolution operation.

Here, the open parameters $\{K, \boldsymbol{c}\}$ of the model are determined by minimizing the squared difference between the model's scatter prediction and an MC simulation:

$$\{K, \boldsymbol{c}\} = \underset{K, \boldsymbol{c}}{\operatorname{argmin}} \sum_{n, \boldsymbol{d}} \|S_{\text{est}}(n, \boldsymbol{d}, K, \boldsymbol{c}) - S_{\text{MC}}(n, \boldsymbol{d})\|_2^2, \tag{5.4}$$

where, n represents the sample index and S_{MC} denotes the MC scatter prediction.

Here it is assumed that these parameters, which are only calibrated once, apply to any measurement and any view angle subsequently. However, practically this assumption leads to a decrease of performance. Therefore, Baer et al. proposed to use a distinct set of parameters $\{K, \boldsymbol{c}\}_n$ for any projection to be corrected using a similar minimization scheme as equation (5.4):

$$\{K, \boldsymbol{c}\}_n = \underset{K, \boldsymbol{c}}{\operatorname{argmin}} \sum_{\boldsymbol{d}} \|S_{\text{est}}(n, \boldsymbol{d}, K, \boldsymbol{c}) - \tilde{S}_{\text{MC}}(n, \boldsymbol{d})\|_2^2. \tag{5.5}$$

In contrast to the KSE approach, this so-called hybrid scatter estimation (HSE) uses only a very coarse MC simulation $\tilde{S}_{\text{MC}}$ to achieve a reasonable performance. Thus, it can be seen as physics-based regularizer to the noisy MC scatter prediction. In this study the HSE was used as a second reference approach.

5.2.2 Deep Scatter Estimation (DSE)

Conventional KSE approaches come with the drawback of being restricted to a predefined model that is based on simplified assumptions and approximations on x-ray scattering. Therefore, they might not generalize suitably to arbitrary cases. However, given only the projection data as input, it is challenging to develop more comprehensive models and to set their parameters appropriately. Potentially, deep learning based approaches are able to overcome these drawbacks as, in contrast to conventional approaches, they do not use handcrafted models or features to solve a certain problem but learn the most suitable model intrinsically from a large set of training data. Therefore, the DSE uses a deep convolutional neural network (DCNN) that is trained to predict x-ray scattering. The proposed DSE network, which is similar to the U-net proposed by Ronneberger et al. [39], is shown in figure 5.1. It is an encoder-decoder network with concatenated skip connections between the encoding path and the decoding path that help to restore high resolution features and to improve the training convergence. The encoding path, that extracts a hierarchy of features of the input images, consists of seven stages. Each applies three convolution layers with a 3×3 kernel size and ReLU activation. While the spatial dimensions are reduced by applying a convolution with a stride equal to two, the features are doubled starting with 16 features in the first layer. The decoding path is designed similarly but applies a 2×2 nearest neighbor upsampling instead of the strided convolutions to restore the spatial dimensions.

The DSE network was implemented using the open source framework Keras (v. 2.15) in combination with the Tensorflow backend (v. 1.7). Given only a function of the acquired projection data as input, DSE was trained by optimizing the network's weights $\boldsymbol{w}$ and biases $\boldsymbol{b}$ to minimize the mean absolute percentage error (MAPE) between an MC scatter prediction S_{MC} and the network's output:

$$\{\boldsymbol{w}, \boldsymbol{b}\} = \operatorname*{argmin}_{\boldsymbol{w}, \boldsymbol{b}} \frac{100}{M} \sum_{n, \boldsymbol{d}} \left| \frac{\mathrm{DSE}(n, \boldsymbol{d}, \boldsymbol{w}, \boldsymbol{b}) - S_{\mathrm{MC}}(n, \boldsymbol{d})}{S_{\mathrm{MC}}(n, \boldsymbol{d})} \right|, \tag{5.6}$$

with n denoting the sample index, $\boldsymbol{d}$ denoting the detector coordinate, and M the product of the number of training samples and detector pixels. The weights were initialized with a Glorot uniform distribution [158]. The biases were initialized with zeros. The training was performed for 80 epochs on an NVIDIA Quadro P6000 using an Adam optimizer and a batch size of 16.

As x-ray scattering is known to be low frequent, the DSE network does not take the full size projection data as input, but downsamples them to have with 256×256 pixels. Thus, the DSE scatter estimate needs to be upsampled to the original projection size prior to scatter correction.

Feature dimensions:
256×256 128×128 64×64 32×32 16×16 8×8 4×4 8×8 16×16 32×32 64×64 128×128 256×256

Number of features of the convolutional layer:
16   32   64   128   256   512   1024   512   256   128   64   32   16 / 1

▷ 3 × 3 Convolution (stride = 1), ReLU   ▷ 3 × 3 Convolution (stride = 2), ReLU   ▷ 1 × 1 Convolution (stride = 1), ReLU   ◆ 2 × 2 Upsampling

○ Depth concatenate
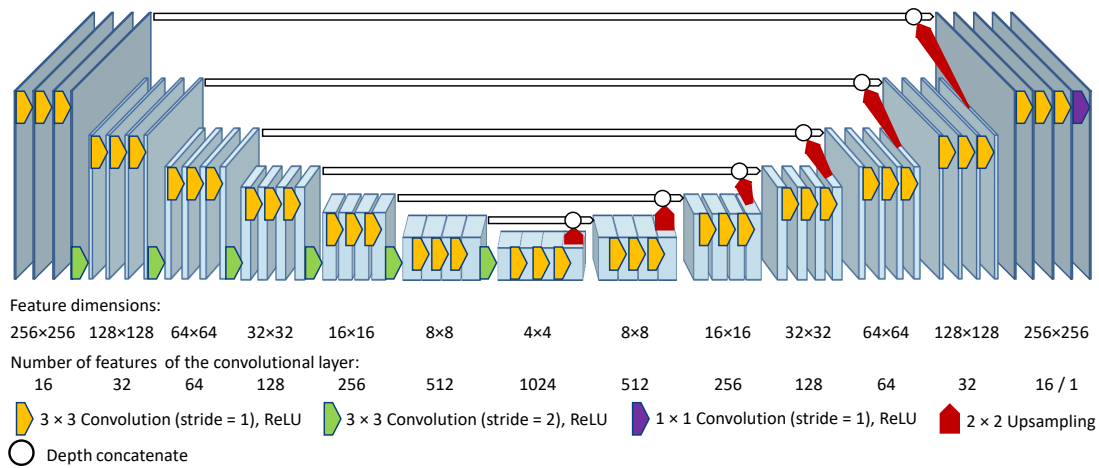
Figure 5.1: Architecture of the DSE deep convolutional neural network. Note that the network does not take the full size projection data as input, but a downsampled 256 × 256 version. Thus, the DSE scatter estimate needs to be upsampled prior to scatter correction.

5.2.3 Simulation Study

Practically, it is advantageous to have a scatter estimation approach that can be applied to different components and different acquisition conditions without major adjustments. The potential of DSE as well as the two reference approaches to do so, was investigated using simulated data. The scatter estimation approaches were evaluated with respect to:

i.) Generalization to different tube voltages

ii.) Generalization to different noise levels

iii.) Generalization to different materials

iv.) Generalization to different magnifications

v.) Generalization to different components

All simulations were performed using CAD prior models from a public database (www.grabcad.com). Here, the experiments i.) - iv.) are based on the compressor wheel models shown in figure 5.2, while experiment v.) was based on the six models shown in figure 5.3. Given the prior model, artificial projections of an industrial CT system were simulated as:

$$p_{\text{sim}} = -\ln\left(\frac{I + N_p + S_{\text{MC}}}{I_0}\right), \tag{5.7}$$

with I being the polychromatic primary intensity according to equation (3.14), N_p being Poisson distributed noise, S_{MC} being the distribution of scattered x-rays that

was calculated using the MC simulation described in section 3.1.3, and I_0 being the flat field image, i.e. an image without the component inside the field of measurement (FOM). For each component and each parameter set 36 views, distributed equally over a scan range of 360°, were simulated. Additionally, DSE was investigated with respect to the mapping to be learned by the network. To do so, DSE was trained to for the following mappings:

- $M_{\mathrm{ep}}: \quad e^{-p_{\mathrm{sim}}} \qquad \longrightarrow \quad S_{\mathrm{MC}}$
- $M_{\mathrm{p}}: \quad p_{\mathrm{sim}} \qquad\qquad \longrightarrow \quad S_{\mathrm{MC}}$
- $M_{\mathrm{pep}}: p_{\mathrm{sim}} \cdot e^{-p_{\mathrm{sim}}} \quad \longrightarrow \quad S_{\mathrm{MC}}$

For any mapping the data were divided into two sets. A training data set that is used to optimize the network's parameters and testing data set to investigate the performance for unknown data. A detailed description of the simulation parameters for each experiment can be found in the following sections. The training of the DSE network was performed as described section 5.2.2. The open parameters of the KSE approach were determined according to equation (5.4) using the same training data as DSE. As the HSE approach is based on a recalibration of the open parameters for every projection, it is applied directly to the testing data set using a coarse MC simulation. To quantify the performance of the three scatter estimation approaches, the mean absolute percentage error (MAPE) with respect to ground truth MC scatter distribution S_{MC} was used as a performance measure (see equation (5.6)).

To evaluate the impact of the quality of the scatter estimates on CT images, CT reconstructions of scatter-corrected projections were performed. Therefore, a 360° circle scan with 720 projections was simulated according to equation (5.7) using the six components shown in figure 5.3 as prior. For any projection scatter was estimated using DSE, KSE, and HSE. This estimate was subtracted in intensity domain to derive a corrected data set. The scatter-corrected data were precorrected for beam-hardening using a similar approach to water-precorrection in clinical CT [126], (i.e. by numerical inversion of the relationship between polychromatic projection values and intersection lengths). Subsequently, the projections were reconstructed on a $1024 \times 1024 \times 1024$ voxel grid with 0.2 mm voxel size using the analytic Feldkamp algorithm [8].
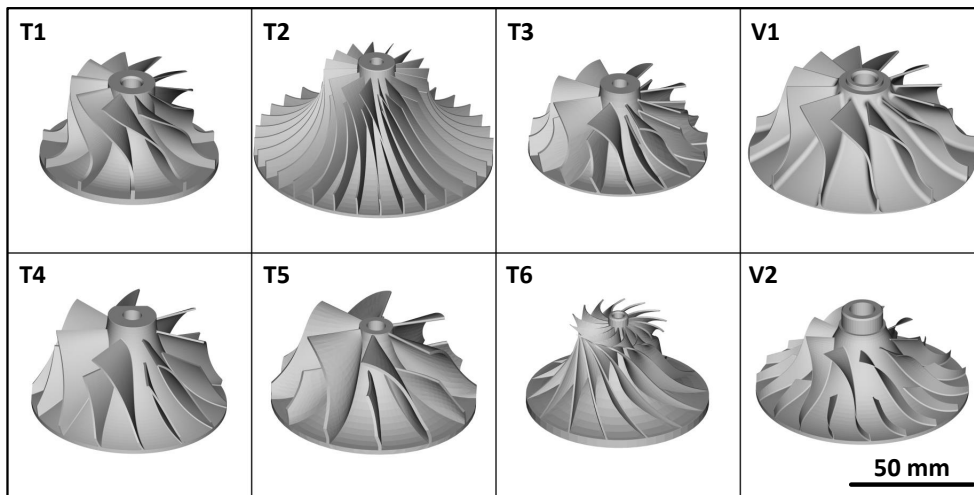
Figure 5.2: Compressor wheel models that have been used to generate simulated data. Simulations of model T1-T6 have been used for training while simulations of model V1 and V2 have been used for testing.
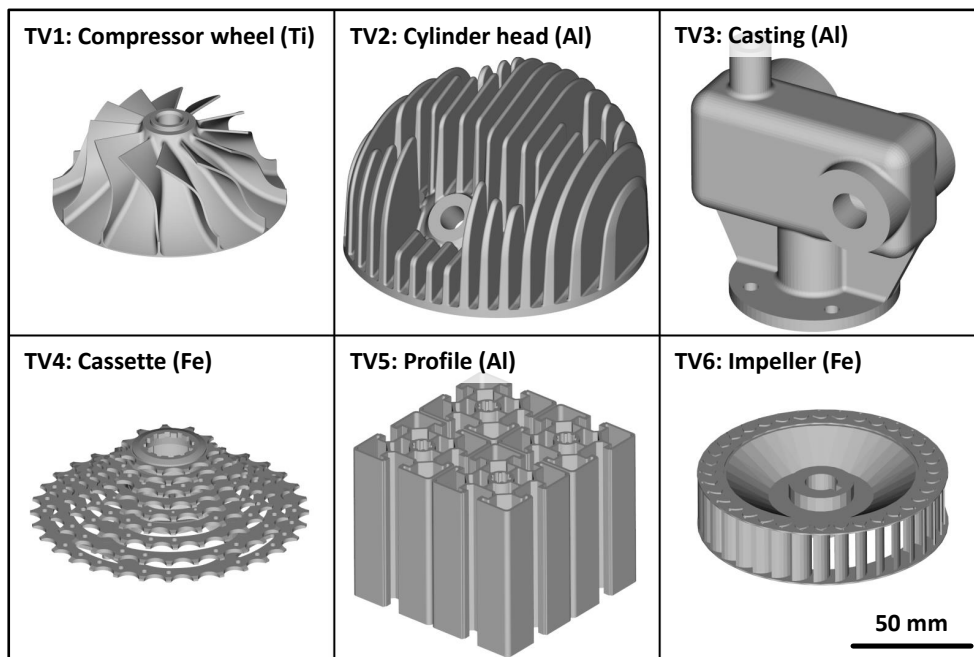
Figure 5.3: Models of the compressor wheel (TV1), the cylinder head (TV2), the casting (TV3), the cassette (TV4), the profile (TV5), and the impeller (TV5) that have been used to generate simulated data to investigate the generalization to different components.

i.) Generalization to Different Tube Voltages

In contrast to clinical CT that uses a few predefined tube voltages, typically between 70 kV and 150 kV, industrial CT uses a wider range of tube voltages. Therefore, it is desirable that a scatter estimation approach does not need to be optimized for every tube voltage separately. This issue was investigated using simulations of compressor wheels (see figure 5.2) at different tube voltages, namely 100 kV, 150 kV, 200 kV, 250 kV, 300 kV, 350 kV, and 400 kV. The training data were generated according to equation 5.7 using the models T1–T6 while the scatter estimation was tested on data of the models V1 and V2. For each prior model and each tube voltage, simulations were performed using different scaling, different displacements and different tilt angles of the model. The corresponding simulation parameters are summarized in table 5.1. For each of the three mappings (M_{ep}, M_p, M_{pep}) given above, DSE was trained nine times: for any of the seven tube voltages separately, for 100 kV and 400 kV data, and for all data together. To evaluate the generalization, any of the nine DSE networks was applied to the test data of each tube voltage. The KSE approach was optimized and tested in a similar way. As stated above, HSE was optimized and applied for the test data only.

Table 5.1: Parameters of the simulation study on the generalization of the scatter estimation approaches with respect to different tube voltages.

Sample	Training	Testing
Models (see figure 5.2)	T1–T6	V1, V2
Source-to-isocenter distance	375 mm	375 mm
Source-to-detector distance	1000 mm	1000 mm
View angle	$0° - 360°$, $\Delta\alpha = 10°$	$0° - 360°$, $\Delta\alpha = 10°$
Detector elements	1024×1024	1024×1024
Detector pixel size	0.4 mm $\times$ 0.4 mm	0.4 mm $\times$ 0.4 mm
Tube voltage	100 kV – 400 kV, $\Delta V = 50$ kV	100 kV – 400 kV, $\Delta V = 50$ kV
Prefilter	1 mm Sn	1 mm Sn
Detector material	1 mm CsI	1 mm CsI
Photons per pixel	10k	10k
Object scaling	0.9, 1.0, 1.1	0.9, 1.0, 1.1
Object z-displacement	-20 mm, 0 mm, 20 mm	-20 mm, 0 mm, 20 mm
Object tilt angle	$0°$, $18°$, $72°$	$36°$, $54°$
Object material	Al	Al
Object density	2.7 g/cm^3	2.7 g/cm^3
Data augmentation	Horizontal flipping	Horizontal flipping
Samples	$6\times36\times3\times3\times2 = 11664$ for every tube voltage	$2\times36\times3\times3\times2\times2 = 2592$ for every tube voltage

ii.)   **Generalization to Different Noise Levels**

Depending on the requirements on image statistics, resolution and scan time, CT acquisitions are conducted with different tube currents or different integration times, respectively. The number of photons being detected scales linearly with these two quantities. Thus, different tube currents or integration times lead to different noise levels with the standard deviation being proportional to their square root. To investigate the generalization of DSE and the two reference approaches to different noise levels, simulations of the compressor wheels shown in figure 5.2 were performed. Similar to i.) different models and different simulation parameters were used for training and testing. However, in contrast to the previous study, the tube voltage was set to a fixed value of 250 kV but the noise level was varied by simulating different numbers of photons per detector pixel (0.5k, 1k, 2k, 5k, 10k, 20k, 50k). A summary of all simulation parameters can be found in table 5.2. Again DSE and KSE were optimized once for each noise level separately, once for data that were simulated with 500 and 50000 photons per pixel, and once using all noise levels together. Subsequently, each DSE network and each KSE parameter set was applied to any noise level's testing data set.

Table 5.2: Parameters of the simulation study on the generalization of the scatter estimation approaches with respect to different noise levels.

Sample	Training	Testing
Models (see figure 5.2)	T1–T6	V1, V2
Source-to-isocenter distance	375 mm	375 mm
Source-to-detector distance	1000 mm	1000 mm
View angle	$0° - 360°, \Delta\alpha = 10°$	$0° - 360°, \Delta\alpha = 10°$
Detector elements	1024×1024	1024×1024
Detector pixel size	0.4 mm $\times$ 0.4 mm	0.4 mm $\times$ 0.4 mm
Tube voltage	250 kV	250 kV
Prefilter	1 mm Sn	1 mm Sn
Detector material	1 mm CsI	1 mm CsI
Photons per pixel	0.5k, 1k, 2k, 5k, 10k, 20k, 50k	0.5k, 1k, 2k, 5k, 10k, 20k, 50k
Object scaling	0.9, 1.0, 1.1	0.9, 1.0, 1.1
Object z-displacement	-20 mm, 0 mm, 20 mm	-20 mm, 0 mm, 20 mm
Object tilt angle	$0°, 18°, 72°$	$36°, 54°$
Object material	Al	Al
Object density	2.7 g/cm^3	2.7 g/cm^3
Data augmentation	Horizontal flipping	Horizontal flipping
Samples	$6\times36\times3\times3\times3\times2 = 11664$ for every noise level	$2\times36\times3\times3\times2\times2 = 2592$ for every noise level

iii.) Generalization to Different Materials

Typically, industrial CT systems are designed to allow for the measurement of various objects differing in size, shape and material. Therefore, scatter estimation approaches should be able to handle different materials without the need for a material specific calibration. This is especially important as the exact material composition is not always known to the operator of the CT system (e.g. measurement service providers). To investigate this issue, simulations of the compressor wheels shown in figure 5.2 were performed assuming different object materials and different densities, namely: water ($\rho_1 = 1$ g/cm^3, $\rho_2 = 1.5$ g/cm^3,), aluminum ($\rho_1 = 2.7$ g/cm^3, $\rho_2 = 4.0$ g/cm^3,), titan ($\rho_1 = 4.5$ g/cm^3, $\rho_2 = 6.7$ g/cm^3,) and iron ($\rho_1 = 7.8$ g/cm^3). A list of all simulation parameters is given in table 5.3. Similar to the previous sections DSE and KSE were optimized once using the training data set of each material separately, once using training data of H$_2$O ($\rho_1 = 1$ g/cm^3) and iron ($\rho_1 = 7.8$ g/cm^3), and once using all training data together.

Table 5.3: Parameters of the simulation study on the generalization of the scatter estimation approaches with respect to different object materials.

Sample	Training	Testing
Models (see figure 5.2)	T1–T6	V1, V2
Source-to-isocenter distance	375 mm	375 mm
Source-to-detector distance	1000 mm	1000 mm
View angle	$0° - 360°$, $\Delta\alpha = 10°$	$0° - 360°$, $\Delta\alpha = 10°$
Detector elements	1024×1024	1024×1024
Detector pixel size	0.4 mm $\times$ 0.4 mm	0.4 mm $\times$ 0.4 mm
Tube voltage	250 kV	250 kV
Prefilter	1 mm Sn	1 mm Sn
Detector material	1 mm CsI	1 mm CsI
Photons per pixel	10k	10k
Object scaling	0.9, 1.0, 1.1	0.9, 1.0, 1.1
Object z-displacement	-20 mm, 0 mm, 20 mm	-20 mm, 0 mm, 20 mm
Object tilt angle	$0°$, $18°$, $72°$	$36°$, $54°$
Object material	H$_2$O, Al, Ti, Fe	H$_2$O, Al, Ti, Fe
Object density	H$_2$O: 1.0 g/cm^3, 1.5 g/cm^3,	H$_2$O: 1.0 g/cm^3, 1.5 g/cm^3,
	Al: 2.7 g/cm^3, 4.0 g/cm^3,	2.7 g/cm^3, 4.0 g/cm^3,
	Ti: 4.5 g/cm^3, 6.7 g/cm^3,	4.5 g/cm^3, 6.7 g/cm^3,
	Fe: 7.8 g/cm^3	Fe: 7.8 g/cm^3
Data augmentation	Horizontal flipping	Horizontal flipping
Samples	$6\times36\times3\times3\times3\times2 = 11664$	$2\times36\times3\times3\times2\times2 = 2592$
	for every material	for every material

iv.) Generalization to Different Magnifications

As depicted in figure 3.1, most industrial CT systems are table-top CBCT systems that place the sample on a rotary table and keep source and detector fixed. Several vendors mount the rotary table on a linear axis which can be moved in source-detector direction. By doing so, the magnification m, that is defined as

$$m = \frac{\text{source-to-detector distance}}{\text{source-to-isocenter distance}}, \tag{5.8}$$

can be changed. Accordingly, the change of the magnification also leads to a change of the FOM and the effective pixel size in the isocenter, or the resolution, respectively.

To investigate the performance of DSE as well as the reference approaches to estimate scatter for varying magnifications, simulations of the compressor wheels (figure 5.2) were performed. Different magnifications were realized using a fixed source-to-detector distance of 1000 mm but different source-to-isocenter distances of 125 mm, 250 mm, 375 mm 500 mm, and 625 mm. Since practical applications usually measure components with the maximum spatial resolution i.e. such that their x-ray projection covers (almost) the complete detector, the size of the compressor wheel is scaled to meet this criterion. The corresponding scaling factor γ is calculated as:

$$\gamma = \frac{\text{source-to-isocenter distance}}{375 \text{ mm}}, \tag{5.9}$$

with 375 mm being the reference source-isocenter distance as it is used by the other experiments.

Since that scaling leads to an increase of the intersection lengths, there are higher projection values for simulations that have been performed with lower magnification. However, this effect can be reverted by scaling the object's density ρ by the reciprocal of γ (recall that the attenuation of an x-ray of energy E traversing the distance l through an object with mass attenuation coefficient $\frac{\mu(E)}{\rho_0}$ is given by $\exp(-\frac{\mu(E)}{\rho_0} \cdot \rho \cdot l)$). In that case, the primary intensity is distributed equally for any magnification. Thus, projections only differ by their respective scatter distributions which increase with decreasing isocenter-detector distances. To investigate whether the scatter estimation approaches can handle this (rather theoretical) scenario, a second simulation study was performed that scales the object as well as the density of the object. The corresponding simulation parameters for both experiments are summarized in table 5.4. To evaluate the generalization of DSE and KSE, they were optimized once using the training data of every setting separately, once using the training data that have been simulated with a source-isocenter distance of 125 mm and 625 mm, and once using all data together.

Table 5.4: Parameters of the simulation study on the generalization of the scatter estimation approaches with respect to different magnifications. Note that the scaling factor γ is chosen according to equation (5.9).

Sample	Training	Testing
Models (see figure 5.2)	T1–T6	V1, V2
Source-to-isocenter distance	125 mm, 250 mm, 375 mm, 500 mm, 625 mm	125 mm, 250 mm, 375 mm, 500 mm, 625 mm
Source-to-detector distance	1000 mm	1000 mm
View angle	$0° - 360°$, $\Delta\alpha = 10°$	$0° - 360°$, $\Delta\alpha = 10°$
Detector elements	1024×1024	1024×1024
Detector pixel size	0.4 mm $\times$ 0.4 mm	0.4 mm $\times$ 0.4 mm
Tube voltage	250 kV	250 kV
Prefilter	1 mm Sn	1 mm Sn
Detector material	1 mm CsI	1 mm CsI
Photons per pixel	10k	10k
Object scaling	$\gamma \cdot (0.9, 1.0, 1.1)$	$\gamma \cdot (0.9, 1.0, 1.1)$
Object z-displacement	-20 mm, 0 mm, 20 mm	-20 mm, 0 mm, 20 mm
Object tilt angle	$0°, 18°, 72°$	$36°, 54°$
Object material	Al	Al
Object density	$\gamma^{-1} \cdot 2.7$ g/cm^3	$\gamma^{-1} \cdot 2.7$ g/cm^3
Data augmentation	Horizontal flipping	Horizontal flipping
Samples	$6 \times 36 \times 3 \times 3 \times 3 \times 2 = 11664$ for every magnification	$2 \times 36 \times 3 \times 3 \times 2 \times 2 = 2592$ for every magnification

v.)   Generalization to Different Components

There is broad spectrum of components that can potentially be tested using x-ray computed tomography. Therefore, it is desirable to have a scatter estimation approach that shows similar performance for different components without the need for dedicated components-specific optimization. To investigate the ability of DSE, KSE and HSE to generalize to different components, simulations of six components (see figure 5.3) that differ in size, shape and material have been performed. For each component a training data set and a testing data set was simulated. To make sure that the training data does not resemble the testing data, it was generated using different tube voltages, different tilt angles of the component and different scaling factors. A complete list of all simulation parameters can be found in table 5.5. The optimization and testing of DSE and KSE is similar to what is described in the previous sections. For each component a network is trained or a parameter set is calculated, respectively, by using only the training data of that particular component. Additionally, one optimization is performed using the training data of the cassette (TV4) and the profile (TV5) and one that uses all available training data.

Table 5.5: Parameters of the simulation study on the generalization of the scatter estimation approaches with respect to different components. Note that the scaling factor γ is chosen according to equation (5.9).

Sample	Training	Testing
Models (see figure 5.3)	TV1–TV6	TV1–TV6
Source-to-isocenter distance	250 mm, 375 mm, 500 mm	250 mm, 375 mm, 500 mm
Source-to-detector distance	1000 mm	1000 mm
View angle	$0° - 360°$, $\Delta\alpha = 10°$	$0° - 360°$, $\Delta\alpha = 10°$
Detector elements	1024×1024	1024×1024
Detector pixel size	0.4 mm $\times$ 0.4 mm	0.4 mm $\times$ 0.4 mm
Tube voltage	150 kV, 200 kV, 300 kV	250 kV, 350 kV
Prefilter	1 mm Sn	1 mm Sn
Detector material	1 mm CsI	1 mm CsI
Photons per pixel	10k	10k
Object scaling	$\gamma \cdot \{ 0.9, 1.1 \}$	$\gamma \cdot 1.0$
Object z-displacement	-20 mm, 0 mm, 20 mm	-20 mm, 0 mm, 20 mm
Object tilt angle	$0°$, $18°$, $72°$	$36°$, $54°$
Object material	Al (TV2, TV3, TV5),	Al (TV2, TV3, TV5),
	Ti (TV1),	Ti (TV1),
	Fe (TV4, TV6)	Fe (TV4, TV6)
Object density	2.7 g/cm^3 (TV2,TV3,TV5)	2.7 g/cm^3 (TV2,TV3,TV5)
	4.5 g/cm^3 (TV1)	4.5 g/cm^3 (TV1)
	7.8 g/cm^3 (TV4, TV6)	7.8 g/cm^3 (TV4, TV6)
Data augmentation	Horizontal flipping	Horizontal flipping
Samples	$3\times36\times3\times2\times3\times6 = 11664$	$3\times36\times2\times3\times2\times2 = 2592$
	for every component	for every component

5.2.4 Measured Data

To test the application of DSE to real data, measurements of an aluminum profile were performed at our DKFZ table-top CT system which is equipped with a Varian 4030 flat detector and a 110 kV Hamamatsu micro-focus x-ray source (see figure 5.4). To train the DSE network, representative training data needs to be generated. Basically, there are different strategies that can be used to do so. Probably the most accurate way is to derive the training data from reference measurements, e.g. measurements with and without anti-scatter grid or with and without collimation. However, practically it is difficult to acquire a huge amount of data following this strategy. Therefore, training data were generated using CT simulations based on the models derived in chapter 3. Here, the simulations were designed to resemble the measurement data in terms of the acquisition geometry and the acquisition conditions. Primary radiation was simulated analytically including the source and detector models described in section 3.1.2 and section 3.1.5, respectively. The effect of the focal spot distribution including off-focal
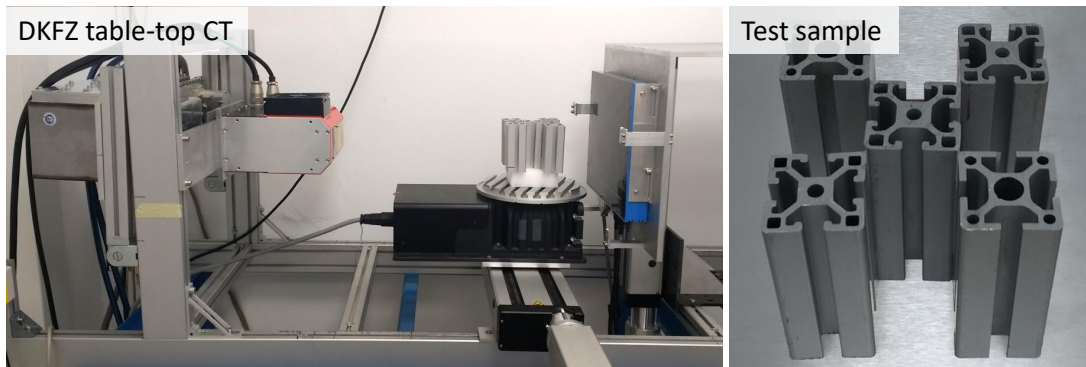
Figure 5.4: DKFZ table-top setup to test the application of DSE to measured data (left) and the corresponding test sample (right).

radiation was modeled using the convolution-based approach introduced in section 3.1.4, while the distribution itself was derived from an edge measurement. X-ray scattering was simulated using the MC simulation described in section 3.1.3.

All simulations were based on the prior models that have been used for the simulation study (see figure 5.3). However, since our x-ray tube can only be operated at a maximum voltage of 110 kV, it is not able to penetrate titanium or iron appropriately. Therefore, the material of all prior models was set to aluminum. A complete list of all simulation and measurement parameters is given in table 5.6.

Table 5.6: Parameters for the simulated training data set and the measurement.

	Training	Measurement
Models (see figure 5.3)	TV1–TV6	-
Source-to-isocenter distance	100 mm, 110 mm, 120 mm	110 mm
Source-to-detector distance	580 mm	580 mm
View angle	$0° - 360°$, $\Delta\alpha = 10°$	$0° - 360°$, $\Delta\alpha = 0.5°$
Detector elements	768×768	768×768
Detector pixel size	0.388 mm×0.388 mm	0.388 mm×0.388 mm
Tube voltage	100 kV, 110 kV, 120 kV	110 kV
Prefilter	1.0 mm, 2.0 mm Cu	2.0 mm Cu
Detector material	1 mm CsI	1 mm CsI
Photons per pixel	10k	-
Object scaling	1.0	-
Object z-displacement	0 mm	-
Object tilt angle	$0°, 30°, 60°, 90°$	$0°$
Object material	Al	Al
Object density	2.7 g/cm^3	2.7 g/cm^3
Samples	15552	720

Given the primary and the scatter simulation, training samples were calculated according to equation (5.7), followed by a 3×3 downsampling to as size of 256×256. Subsequently, KSE and DSE were optimized as described in section 5.2.1 and 5.2.2, by minimizing the error between the scatter prediction and the MC simulation. The parameters of HSE, were optimized using a coarse MC simulation based on an initial reconstruction. Since there is no scatter ground truth for measured data, a slit scan acquisition with a narrow collimation was performed as a reference.

5.3 Results

5.3.1 Simulation Study

i.) Generalization to Different Tube Voltages

The generalization of the DSE as well as the two reference approaches with respect to varying tube voltages was evaluated as described in section 5.2.3. Qualitative results of the corresponding scatter estimates for an exemplary projection are shown in figure 5.5. Quantitative results on the generalization of the three scatter estimation approaches including the three mappings M_{ep}, M_p, and M_{pep} to be learned by the DSE network are summarized in table 5.7.

In general, it can can be observed that the accuracy of any scatter estimation approach (except for a few KSE cases) is reduced if it is applied to estimate scatter for a tube voltage that was not included in the training data set. However, compared to the variation of other parameters (see following sections) the reduction is rather small. Here, the accuracy decreases by 1.6 % / 100 kV (KSE), 4.4 % / 100 kV (DSE, M_{ep}), 0.9 % / 100 kV (DSE, M_p) and 1.1 % / 100 kV (DSE, M_{pep}) on average (note that these are absolute measures in terms of percentage points). However, table 5.7 shows that there is no need to train for each tube voltage separately as training for all tube voltages does not lead to a major decrease of accuracy but applies to all tube voltages in a similar way. Even the training on only two tube voltages (100 kV and 400 kV) leads to a high accuracy for any of the seven tube voltages.

Comparing the accuracy among the different scatter estimation approaches, KSE shows the poorest performance, as expected. In any case, the MAPE between the ground truth and the scatter estimate is larger than 12 %. Using a distinct parameter set for every projection to be processed, like HSE does, improves the accuracy to a MAPE between 5.0 % (400 kV) and 6.0 % (100 kV). DSE clearly outperforms the reference approaches if it is trained to learn a mapping based on projection data (M_p) or "pep" data (M_{pep}). Here, a training on all available training data yields a MAPE between 1.2 % (400 kV) and 1.5 % (100 kV) for the M_p mapping. Similar accuracy is achieved for the M_{pep} mapping with a MAPE between 1.5 % (400 kV) and 1.6 % (100 kV). In contrast, training DSE to learn M_{ep} mapping leads to scatter estimates with a significantly lower accuracy. Here, the MAPE is between 7.5 % (250 kV) and 8.0 % (400 kV).
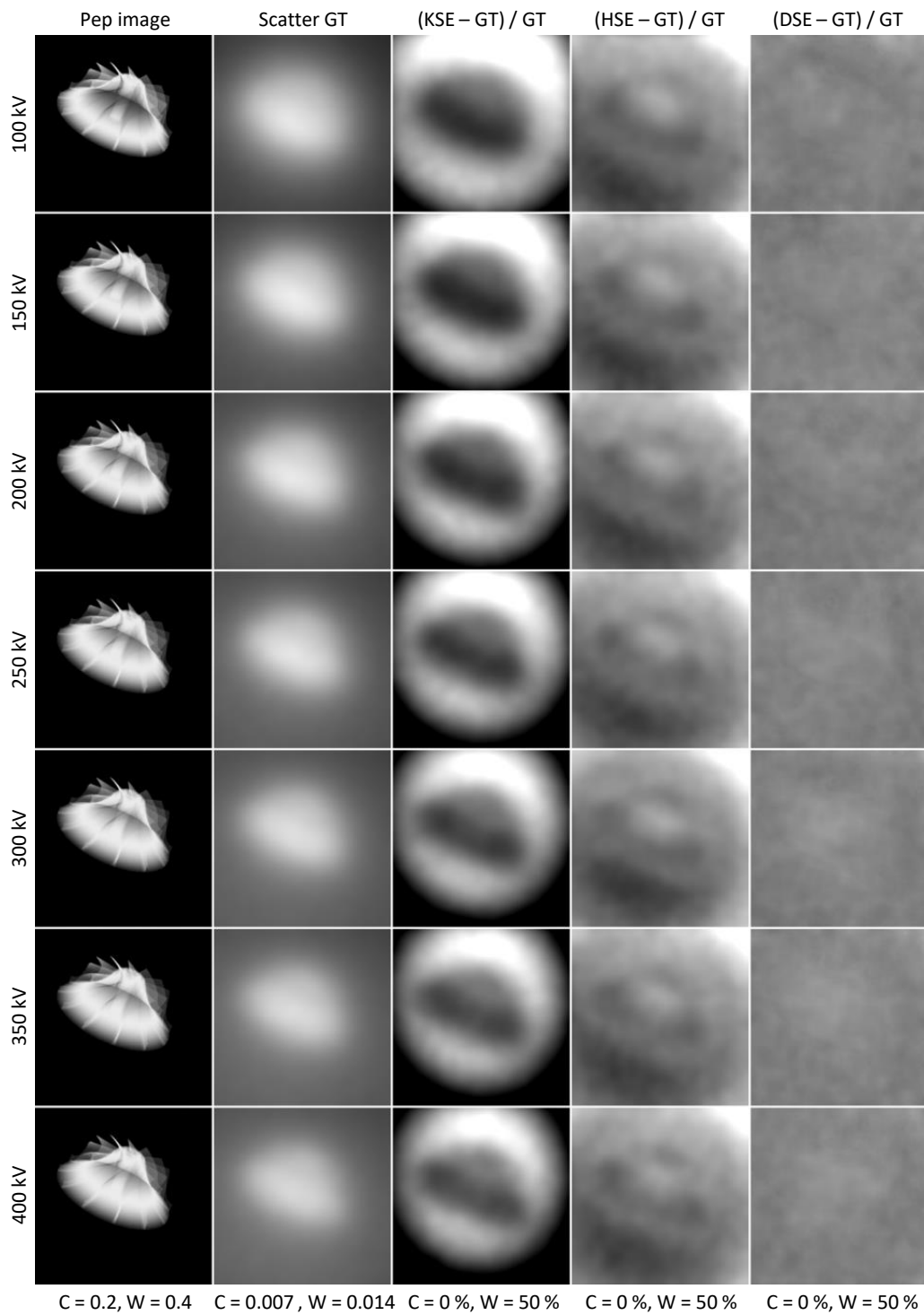
Figure 5.5: Percentage error of KSE, HSE and DSE (mapping M_{pep}) scatter estimates with respect to the MC ground truth (GT) for different tube voltages. KSE and DSE were optimized using all available training data (100 kV to 400 kV). HSE was optimized directly for the test data.

Table 5.7: Mean absolute percentage error of KSE, HSE and DSE scatter estimates for different tube voltages.

| | Testing | | | | | | |
	100 kV	150 kV	200 kV	250 kV	300 kV	350 kV	400 kV
Training							
KSE							
100 kV	12.9	16.4	18.1	19.3	20.2	20.9	21.6
150 kV	13.0	15.4	16.8	17.8	18.4	19.0	19.5
200 kV	13.4	14.8	15.7	16.2	16.6	17.0	17.3
250 kV	14.7	14.9	15.1	15.3	15.5	15.7	15.9
300 kV	16.2	15.3	15.0	14.9	14.8	14.8	14.9
350 kV	16.7	15.5	15.0	14.7	14.6	14.5	14.6
400 kV	19.5	16.9	15.5	14.7	14.2	13.9	13.7
100 kV & 400 kV	13.4	14.7	15.6	16.1	16.5	16.8	17.2
All tube voltages	13.9	14.8	15.3	15.7	16.0	16.3	16.6
HSE							
100 kV	6.0	5.5	5.3	5.2	5.2	5.1	5.0
DSE, $M_{\mathrm{ep}} : e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$							
100 kV	7.5	9.2	11.6	13.7	15.4	16.7	18.0
150 kV	12.2	7.0	7.7	9.3	10.8	12.0	13.3
200 kV	16.4	8.6	6.7	7.1	8.0	8.9	10.0
250 kV	21.4	10.9	6.8	5.8	6.2	7.0	8.0
300 kV	23.3	12.9	8.3	6.3	5.9	6.1	6.7
350 kV	27.1	15.2	10.0	7.5	6.6	6.5	6.8
400 kV	31.5	18.7	12.7	9.4	7.6	6.7	6.2
100 kV & 400 kV	7.1	9.8	10.3	10.0	9.6	9.3	9.0
All tube voltages	7.8	7.7	7.6	7.5	7.6	7.8	8.0
DSE, $M_{\mathrm{p}} : p_{\mathrm{sim}} \longrightarrow S_{\mathrm{MC}}$							
100 kV	1.6	2.0	2.4	2.7	3.2	3.6	4.2
150 kV	2.2	1.3	1.7	2.3	2.9	3.3	3.7
200 kV	2.6	1.7	1.2	1.5	1.9	2.3	2.7
250 kV	4.3	2.9	1.7	1.2	1.5	1.9	2.4
300 kV	4.0	3.0	2.2	1.7	1.4	1.5	1.8
350 kV	4.5	3.6	2.6	1.8	1.4	1.2	1.4
400 kV	3.8	3.3	2.6	2.0	1.6	1.4	1.3
100 kV & 400 kV	1.9	2.9	3.6	2.9	2.1	1.6	1.4
All tube voltages	1.5	1.4	1.4	1.3	1.3	1.3	1.2
DSE, $M_{\mathrm{pep}} : p_{\mathrm{sim}} \cdot e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$							
100 kV	1.6	2.5	3.0	3.5	3.9	4.4	4.8
150 kV	2.4	1.3	1.8	2.5	3.0	3.4	3.8
200 kV	2.6	1.9	1.5	1.6	1.8	2.2	2.5
250 kV	4.0	2.8	1.7	1.1	1.4	1.8	2.2
300 kV	4.1	3.2	2.2	1.4	1.2	1.4	1.7
350 kV	5.0	3.8	2.7	1.9	1.4	1.2	1.3
400 kV	4.9	3.9	3.0	2.3	1.7	1.4	1.2
100 kV & 400 kV	1.6	2.1	2.4	2.0	1.6	1.4	1.3
All tube voltages	1.6	1.5	1.5	1.5	1.5	1.5	1.5

ii.)   Generalization to Different Noise Levels

The generalization to different noise levels was investigated as described in section 5.2.3 by simulating different photon numbers per detector pixel (0.5k to 50k). The corresponding results in terms of the MAPE between the scatter estimates of the testing data set and the MC ground truth are summarized in table 5.8 (note that qualitative results are not shown here as they are almost equal to the fourth row of figure 5.5).

It can be observed, that the accuracy of KSE with a MAPE of 15.3 % and the accuracy of HSE with a MAPE of 5.2 % is independent of the noise level. This is probably a result of large receptive field of the convolution kernel that is used to estimate scattering.

Considering DSE the behavior for varying noise levels highly depends on the mapping to be learned by the neural network. The M_{ep} mapping yields a MAPE of 6.0 %$\pm$0.1 %, irrespective of the noise level it has been trained on. Training DSE to learn the M_{p} mapping or the M_{pep} mapping, in contrast, shows a strong dependence on the noise level. While networks that have been trained on a high noise level still show a considerable performance when they are applied to low noise data, the MAPE increases significantly if networks trained on low noise data are applied to high noise data. An exemplary case that demonstrates this behavior is shown in figure 5.6. While the result of the M_{ep} mapping is not affected by the high noise level, both of the mappings M_{p} and M_{pep} lead to a noise-related texture outside the object as well as an overestimation of scatter inside the object.

However, DSE generalizes well to different noise levels once low and high noise levels are included within the training data set. Training on data that have been simulated with 0.5k photons per pixel and 50k photons per pixel, yields scatter estimates with MAPE lower than 3.2 % (M_{p}) or 3.3 % (M_{pep}), respectively, for any other noise level. Similar to the variation of tube voltage in the previous section, training on data of all noise levels can further improve the accuracy.

| Scatter GT | DSE estimate, M_{ep} | DSE estimate, M_{p} | DSE estimate, M_{pep} |
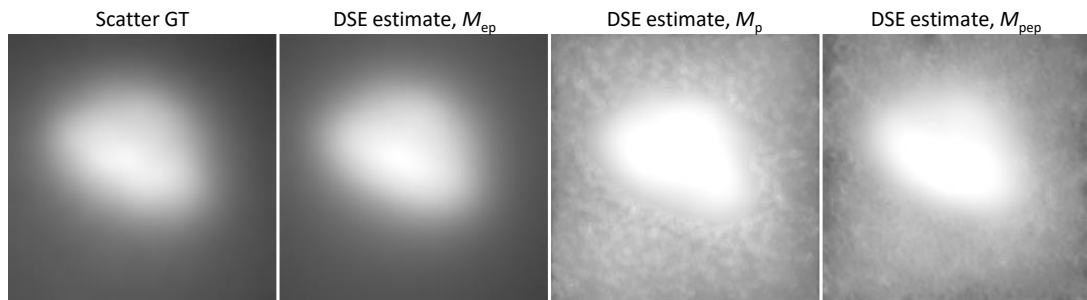
Figure 5.6: Ground truth (GT) and DSE scatter estimates for high noise data (0.5k photons per pixel) using a network that has been trained for low noise data (50k photons per pixel). The three last images show the results for the three different DSE mappings that have been investigated.

Table 5.8: Mean absolute percentage error of KSE, HSE and DSE scatter estimates for different noise levels or number of photons per pixel, respectively.

| | Testing | | | | | | |
	0.5k	1k	2k	5k	10k	20k	50k
Training							
KSE							
0.5k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
1k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
2k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
5k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
10k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
20k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
50k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
0.5k & 50k	15.3	15.3	15.3	15.3	15.3	15.3	15.3
All noise levels	15.3	15.3	15.3	15.3	15.3	15.3	15.3
HSE							
0.5k	5.2	5.2	5.2	5.2	5.2	5.2	5.2
DSE, $M_{\mathrm{ep}} : e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$							
0.5k	6.0	6.0	6.0	6.0	6.0	6.0	6.0
1k	6.0	6.0	6.0	6.0	6.0	6.0	6.0
2k	6.1	6.1	6.1	6.1	6.1	6.1	6.1
5k	6.0	6.0	6.0	6.0	6.0	6.0	6.0
10k	6.0	6.0	6.0	6.0	6.0	6.0	6.0
20k	6.1	6.1	6.1	6.1	6.1	6.1	6.1
50k	6.0	6.0	6.0	6.0	6.0	6.0	6.0
0.5k & 50k	6.1	6.1	6.1	6.1	6.1	6.1	6.1
All noise levels	5.9	5.9	5.9	5.9	5.9	5.9	5.9
DSE, $M_{\mathrm{p}} : p_{\mathrm{sim}} \longrightarrow S_{\mathrm{MC}}$							
0.5k	2.8	3.2	3.7	4.0	4.1	4.1	4.1
1k	4.2	2.3	2.6	3.1	3.4	3.5	3.6
2k	6.1	6.1	6.1	6.0	6.0	6.0	6.0
5k	38.5	13.0	2.6	1.5	1.5	1.6	1.7
10k	52.5	26.5	5.4	1.5	1.2	1.2	1.2
20k	72.1	38.1	12.1	1.8	1.2	1.0	1.0
50k	87.2	55.6	29.0	5.7	1.5	1.0	0.9
0.5k & 50k	2.8	3.1	3.2	2.2	1.6	1.3	1.1
All noise levels	3.0	2.4	2.0	1.7	1.5	1.4	1.3
DSE, $M_{\mathrm{pep}} : p_{\mathrm{sim}} \cdot e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$							
0.5k	2.8	3.8	4.8	5.4	5.7	5.8	5.8
1k	4.1	2.3	2.5	3.1	3.4	3.5	3.6
2k	7.5	3.0	1.9	2.0	2.2	2.3	2.4
5k	15.6	6.0	2.4	1.4	1.4	1.5	1.6
10k	23.3	10.1	3.7	1.5	1.2	1.2	1.2
20k	35.5	17.0	6.9	1.9	1.2	1.0	0.9
50k	59.1	27.7	12.0	3.5	1.4	1.0	0.8
0.5k & 50k	2.8	3.1	3.3	2.4	1.5	1.2	1.1
All noise levels	2.8	2.4	2.0	1.6	1.4	1.3	1.2

iii.) Generalization to Different Materials

The generalization to different materials was investigated for simulated projections of compressor wheels that were assigned different materials (H_2O, Al, Ti, Fe) and different densities. Qualitative results showing the deviation of the scatter estimates from the MC ground truth are shown in figure 5.7 for an exemplary projection. Quantitative results in terms of the MAPE between the scatter estimates of the testing data set and the ground truth are summarized in table 5.9.

As to be expected, any scatter estimation approach shows a decreased performance if it is applied to a material that was not present in the training data set. Besides, it can be observed that, compared to a variation of the object material, the scatter estimation approaches are more robust against a variation of the density. In this study H_2O, Al, and Ti were simulated with two different densities that differ by a factor 1.5. Scatter estimation approaches that were trained on only one of them, also show a good performance when they are applied to test data of the other density. Considering figure 5.7, this might be explained by the fact the scatter distributions for the two different densities but the same material are very similar. Furthermore, KSE, HSE and DSE perform worse the more absorbing the material becomes. In that case the scatter distribution is less homogeneous which might be more difficult to learn.

Comparing the different scatter estimation approaches in terms of accuracy, KSE again shows the poorest performance. Using a parameter set that has been derived by an optimization that is based on all available training data, the MAPE is between 15.8 % (Al, $\rho = 2.7$ g/cm^3) and 29.6 % (Fe, $\rho = 7.8$ g/cm^3). HSE improves the accuracy ans yields scatter estimates with a MAPE between 3.6 % (H_2O, $\rho = 1.0$ g/cm^3) and 9.5 % (Fe, $\rho = 7.8$ g/cm^3). DSE trained on all data to learn the mapping M_{ep} shows a similar performance with a MAPE between 3.9 % (H_2O, $\rho = 1.0$ g/cm^3) and 9.5 % (Al, $\rho = 4.0$ g/cm^3). The highest accuracy can be achieved by training the DSE network to learn the mapping M_p or the mapping M_{pep}. The former yields a MAPE on the test data set between 1.2 % (H_2O, $\rho = 1.0$ g/cm^3) and 1.9 % (Fe, $\rho = 7.8$ g/cm^3), while the latter yields a MAPE between 1.2 % (H_2O, $\rho = 1.0$ g/cm^3) and 2.0 % (Fe, $\rho = 7.8$ g/cm^3).

iv.) Generalization to Different Magnifications

The generalization of KSE, HSE, and DSE was investigated as described in section 5.2.3 by simulating compressor wheels with different source-to-isocenter distances (125 mm to 625 mm). Keeping the source-to-detector distance fixed, leads to different magnifications, and therefore, also a different FOM. Since metrological CT measurements are usually performed such that the object covers almost the complete FOM, the size of the compressor wheels was changed according to equation (5.9). A second experiment was performed that not only changes the size of the object but also its density with the inverse of equation (5.9). By doing so, projections simulated with different magnifications show the same primary intensity distribution but differ only by their scatter distribution. Qualitative results for an exemplary projection are shown in figure 5.8 and figure 5.9,

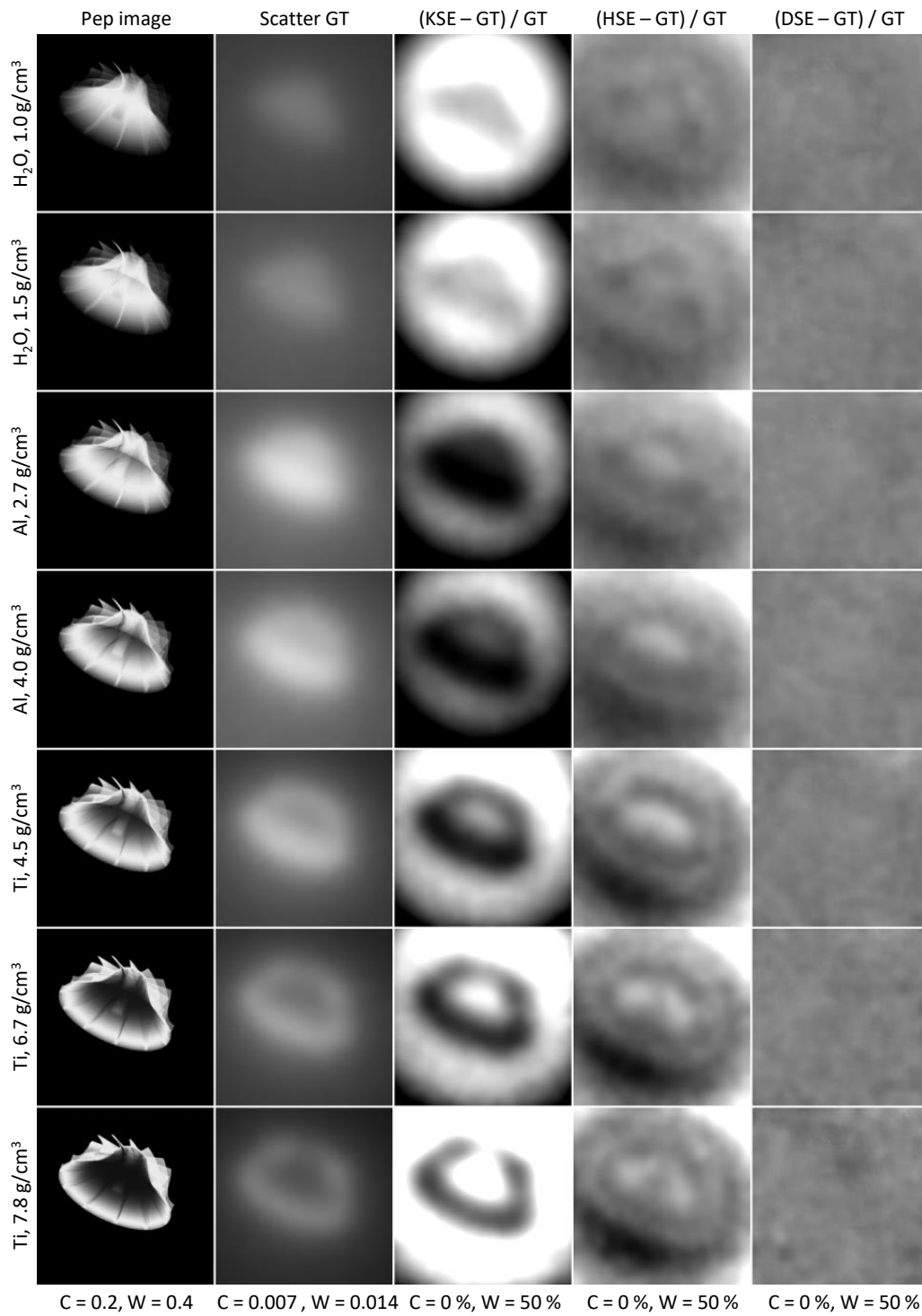|	Pep image	Scatter GT	(KSE – GT) / GT	(HSE – GT) / GT	(DSE – GT) / GT
H_2O, 1.0 g/cm^3					
H_2O, 1.5 g/cm^3					
Al, 2.7 g/cm^3					
Al, 4.0 g/cm^3					
Ti, 4.5 g/cm^3					
Ti, 6.7 g/cm^3					
Ti, 7.8 g/cm^3					
	C = 0.2, W = 0.4	C = 0.007, W = 0.014	C = 0 %, W = 50 %	C = 0 %, W = 50 %	C = 0 %, W = 50 %

Figure 5.7: Percentage error of KSE, HSE and DSE (mapping M_{pep}) scatter estimates with respect to the MC ground truth (GT) for different materials. KSE and DSE were optimized using a training data set containing all materials. HSE was optimized directly for the test data.

119

Table 5.9: Mean absolute percentage error of KSE, HSE and DSE scatter estimates for different materials.

| | Testing | | | | | | |
	H_20, ρ_1	H_20, ρ_2	Al, ρ_1	Al, ρ_2	Ti, ρ_1	Ti, ρ_2	Fe, ρ_1
Training							
KSE							
H_2O, $\rho_1 = 1.0\,\mathrm{g/cm^3}$	9.3	9.8	18.5	20.9	18.7	18.6	30.0
H_2O, $\rho_2 = 1.5\,\mathrm{g/cm^3}$	10.4	9.1	16.1	18.3	21.0	21.0	34.6
Al, $\rho_1 = 2.7\,\mathrm{g/cm^3}$	32.8	28.3	15.3	15.3	19.7	22.0	40.1
Al, $\rho_2 = 4.0\,\mathrm{g/cm^3}$	33.5	28.5	15.4	14.4	23.6	25.8	45.7
Ti, $\rho_1 = 4.5\,\mathrm{g/cm^3}$	31.4	29.0	19.4	20.8	15.3	16.5	27.0
Ti, $\rho_2 = 6.7\,\mathrm{g/cm^3}$	30.1	27.4	20.0	21.5	14.9	16.2	26.9
Fe, $\rho_1 = 7.8\,\mathrm{g/cm^3}$	26.8	26.5	27.9	29.9	19.3	18.5	17.6
H_2O & Fe, ρ_1	18.1	18.2	24.9	27.2	16.6	16.0	17.9
All materials	23.4	20.8	15.8	17.5	15.5	16.4	29.6
HSE							
H_20, $\rho = 1.0\,\mathrm{g/cm^3}$	3.6	3.9	5.2	6.4	7.6	8.6	9.5
DSE, $M_\mathrm{ep} : e^{-p_\mathrm{sim}} \longrightarrow S_\mathrm{MC}$							
H_20, $\rho_1 = 1.0\,\mathrm{g/cm^3}$	4.2	6.4	16.6	19.4	23.1	25.1	45.2
H_20, $\rho_2 = 1.5\,\mathrm{g/cm^3}$	4.3	4.6	14.0	16.3	39.8	57.4	107.4
Al, $\rho_1 = 2.7\,\mathrm{g/cm^3}$	17.6	16.3	6.2	11.0	45.9	71.2	127.0
Al, $\rho_2 = 4.0\,\mathrm{g/cm^3}$	19.4	15.7	6.9	6.8	36.0	59.1	110.1
Ti, $\rho_1 = 4.5\,\mathrm{g/cm^3}$	25.6	27.0	22.6	19.8	9.0	16.2	43.4
Ti, $\rho_2 = 6.7\,\mathrm{g/cm^3}$	28.9	30.5	28.1	25.0	10.4	9.2	24.2
Fe, $\rho_1 = 7.8\,\mathrm{g/cm^3}$	41.9	40.9	40.2	36.6	21.8	17.8	8.4
H_2O & Fe, ρ_1	4.9	9.3	20.2	26.0	16.4	14.4	8.7
All materials	3.9	7.0	5.8	9.5	8.3	8.4	8.8
DSE, $M_\mathrm{p} : p_\mathrm{sim} \longrightarrow S_\mathrm{MC}$							
H_20, $\rho_1 = 1.0\,\mathrm{g/cm^3}$	1.5	2.2	10.2	10.1	10.4	9.7	11.9
H_20, $\rho_2 = 1.5\,\mathrm{g/cm^3}$	1.1	1.1	8.7	8.6	11.9	12.6	19.5
Al, $\rho_1 = 2.7\,\mathrm{g/cm^3}$	11.3	9.2	1.2	1.9	3.4	4.9	8.3
Al, $\rho_2 = 4.0\,\mathrm{g/cm^3}$	11.2	9.1	1.2	1.2	2.9	3.0	5.0
Ti, $\rho_1 = 4.5\,\mathrm{g/cm^3}$	12.4	10.8	2.0	2.2	1.6	2.7	4.9
Ti, $\rho_2 = 6.7\,\mathrm{g/cm^3}$	7.0	7.1	3.1	2.8	2.2	1.8	2.6
Fe, $\rho_1 = 7.8\,\mathrm{g/cm^3}$	8.1	7.0	5.5	3.7	2.8	2.2	2.1
H_2O & Fe, ρ_1	1.2	1.3	5.0	2.5	2.1	2.0	2.0
All materials	1.2	1.3	1.1	1.3	1.4	1.6	1.9
DSE, $M_\mathrm{pep} : p_\mathrm{sim} \cdot e^{-p_\mathrm{sim}} \longrightarrow S_\mathrm{MC}$							
H_20, $\rho_1 = 1.0\,\mathrm{g/cm^3}$	1.4	2.0	10.7	11.5	12.0	11.3	12.0
H_20, $\rho_2 = 1.5\,\mathrm{g/cm^3}$	1.4	1.3	8.4	8.5	10.7	10.4	11.7
Al, $\rho_1 = 2.7\,\mathrm{g/cm^3}$	12.2	9.1	1.2	1.6	2.7	3.8	7.5
Al, $\rho_2 = 4.0\,\mathrm{g/cm^3}$	10.9	8.9	1.2	1.2	3.1	3.7	6.0
Ti, $\rho_1 = 4.5\,\mathrm{g/cm^3}$	10.6	9.7	1.8	2.1	1.6	2.1	3.3
Ti, $\rho_2 = 6.7\,\mathrm{g/cm^3}$	11.5	9.4	2.2	2.1	1.9	1.7	2.5
Fe, $\rho_1 = 7.8\,\mathrm{g/cm^3}$	9.6	8.4	3.6	3.1	2.3	1.8	1.9
H_2O & Fe, ρ_1	1.5	1.6	5.2	3.6	2.4	2.3	2.2
All materials	1.2	1.3	1.2	1.3	1.5	1.6	2.0

respectively. Quantitative results summarizing the MAPE with respect to the MC ground truth are given in table 5.10 and table 5.11, respectively.

Compared to the variation of the tube voltage, the noise level or the material, the variation of the magnification has the strongest impact on the amplitude of the scatter distribution. However, especially for the second experiment, in which the density is scaled, the input to the scatter estimation approaches is very similar (see e.g. "pep" image in figure 5.9). Therefore, it is difficult to distinguish between two data sets that have been acquired with different magnifications. As a result, there are high deviations once a scatter estimation approach is applied to a magnification that is not included in the training data set. Here, the performance is worst if a scatter estimation approach trained on data with a source-to-isocenter distance of 625 mm is applied to testing data that were simulated with a source-to-isocenter distance of 125 mm. In that case KSE yields scatter estimates with deviations up to 284.4 %, DSE, M_{ep} shows deviations up to 204.4 %, DSE, M_{p} shows deviations up to 63.0 %, and DSE, M_{pep} shows deviations up to 84.9 %. Since HSE is optimized for the testing data set only, the accuracy is similar to the other experiments with a MAPE between 3.2 % (fixed density, 125 mm) and 9.6 % (fixed density, 625 mm).

Furthermore, it can be observed that KSE seems to have too few degrees of freedom to account for different magnifications appropriately. Even if it is trained on all available data, there are high deviations ranging from 16 % (fixed density, 500 mm) to 118.8 % (fixed density, 125 mm). DSE trained to learn the M_{p} or the M_{pep} mapping, in contrast, already leads to reasonable results if it is trained on data that were simulated with a source-isocenter distance of 125 mm and 625 mm. Here, the M_{pep} mapping shows the smallest deviations with a MAPE less than 6.1 %. Similar to the other experiments of the simulation study, a further improvement can be achieved by incorporating data of all magnifications in the training data set. In that case the M_{p} and the M_{pep} mapping lead to scatter estimates that deviate less than 1.6 % from the ground truth.

Interestingly, a comparison of the two experiments shows only minor differences regarding the accuracy of the scatter estimates. This suggests that the DSE neural network is able to extract meaningful features even if there is an equal distribution of the primary intensity but only a slightly different scatter distribution.
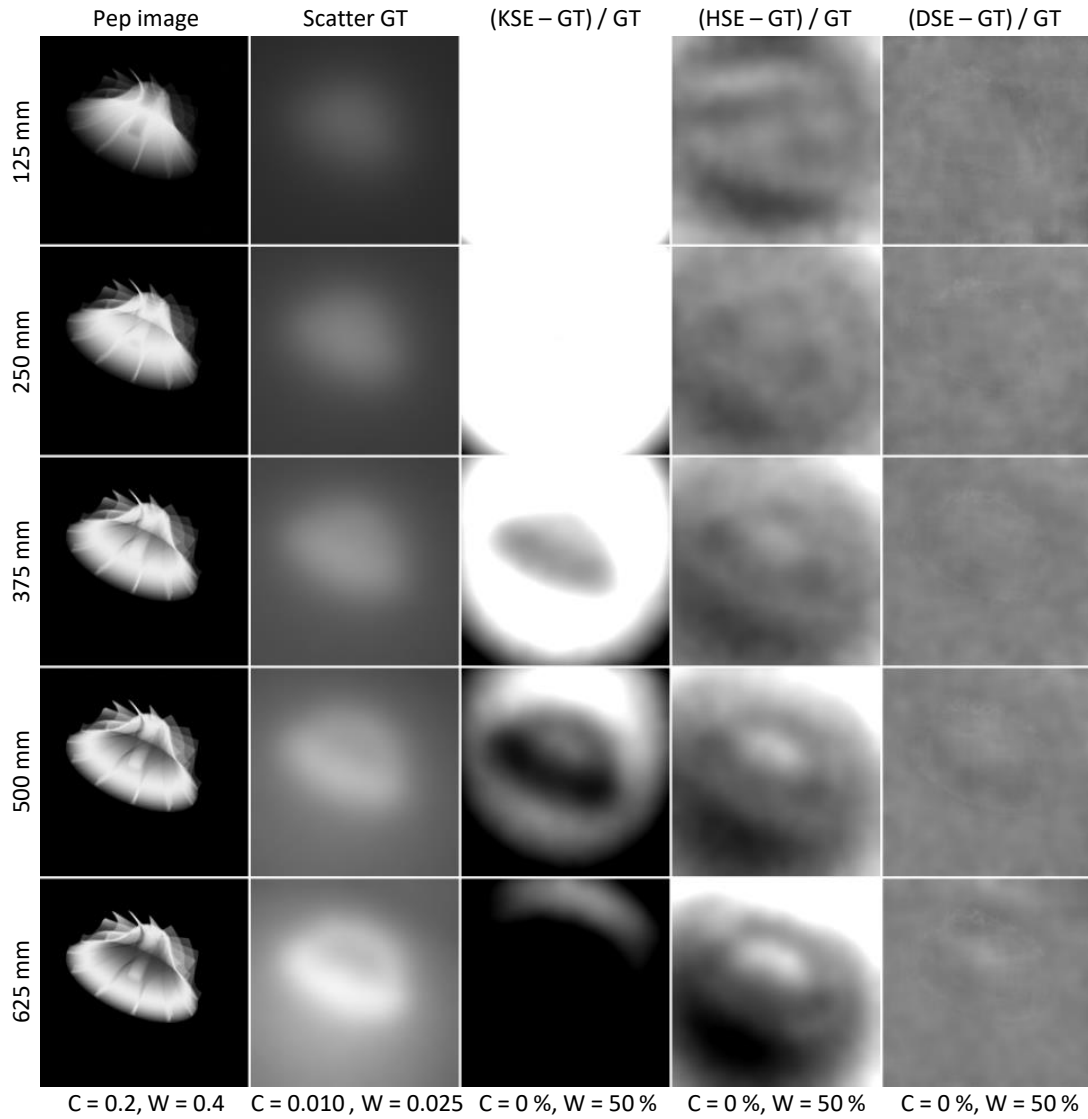
Figure 5.8: Percentage error of KSE, HSE and DSE (mapping M_{pep}) scatter estimates for different magnifications. Here the size of the compressor wheel was changed according to equation (5.9) while the density was set to a fixed value of 2.7 g/cm^3. KSE and DSE were optimized using all available training data (125 mm to 625 mm). HSE was optimized directly for the test data.

Table 5.10: Mean absolute percentage error of KSE, HSE and DSE scatter estimates for different magnifications but fixed densities. The labels refer to the source-to-isocenter distance while the source-to-detector distance is set to a fixed value of 1000 mm. Given these measures, the magnification is calculated according to equation (5.8).

| | Testing | | | | |
	125 mm	250 mm	375 mm	500 mm	625 mm
Training					
KSE					
125 mm	16.6	33.1	49.5	64.1	76.1
250 mm	39.0	18.6	36.0	53.6	69.9
375 mm	72.7	34.2	15.3	30.9	53.0
500 mm	145.0	86.6	41.2	14.4	34.7
625 mm	284.4	185.1	113.7	52.7	14.8
125 mm & 625 mm	189.3	127.9	71.3	26.1	22.5
All magnifications	118.8	70.5	31.5	16.0	40.1
HSE					
125 mm	3.2	3.8	5.2	7.1	9.6
DSE, $M_{\mathrm{ep}}: e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$					
125 mm	3.3	23.8	40.6	55.2	68.0
250 mm	24.3	4.1	19.3	36.8	54.0
375 mm	57.8	25.3	6.1	21.1	42.0
500 mm	106.0	62.9	31.1	8.3	27.1
625 mm	195.1	122.1	83.0	41.8	10.8
125 mm & 625 mm	2.9	94.5	85.5	42.8	11.2
All magnifications	3.6	7.6	16.4	22.0	17.0
DSE, $M_{\mathrm{p}}: p_{\mathrm{sim}} \longrightarrow S_{\mathrm{MC}}$					
125 mm	2.2	10.4	20.4	33.5	48.1
250 mm	6.6	1.4	3.3	8.4	19.2
375 mm	14.4	2.9	1.2	4.4	16.1
500 mm	34.1	6.3	2.6	1.0	4.3
625 mm	63.0	13.6	5.8	2.9	1.0
125 mm & 625 mm	2.2	17.7	10.7	4.4	1.2
All magnifications	1.5	1.3	1.2	1.1	1.1
DSE, $M_{\mathrm{pep}}: p_{\mathrm{sim}} \cdot e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$					
125 mm	3.1	21.8	37.1	51.7	63.7
250 mm	4.9	1.2	3.4	7.0	12.0
375 mm	12.9	2.8	1.3	3.7	12.6
500 mm	24.7	6.1	3.0	1.0	5.0
625 mm	84.9	18.4	5.8	2.4	1.0
125 mm & 625 mm	2.4	6.1	5.0	2.1	1.0
All magnifications	1.6	1.3	1.1	1.0	1.0

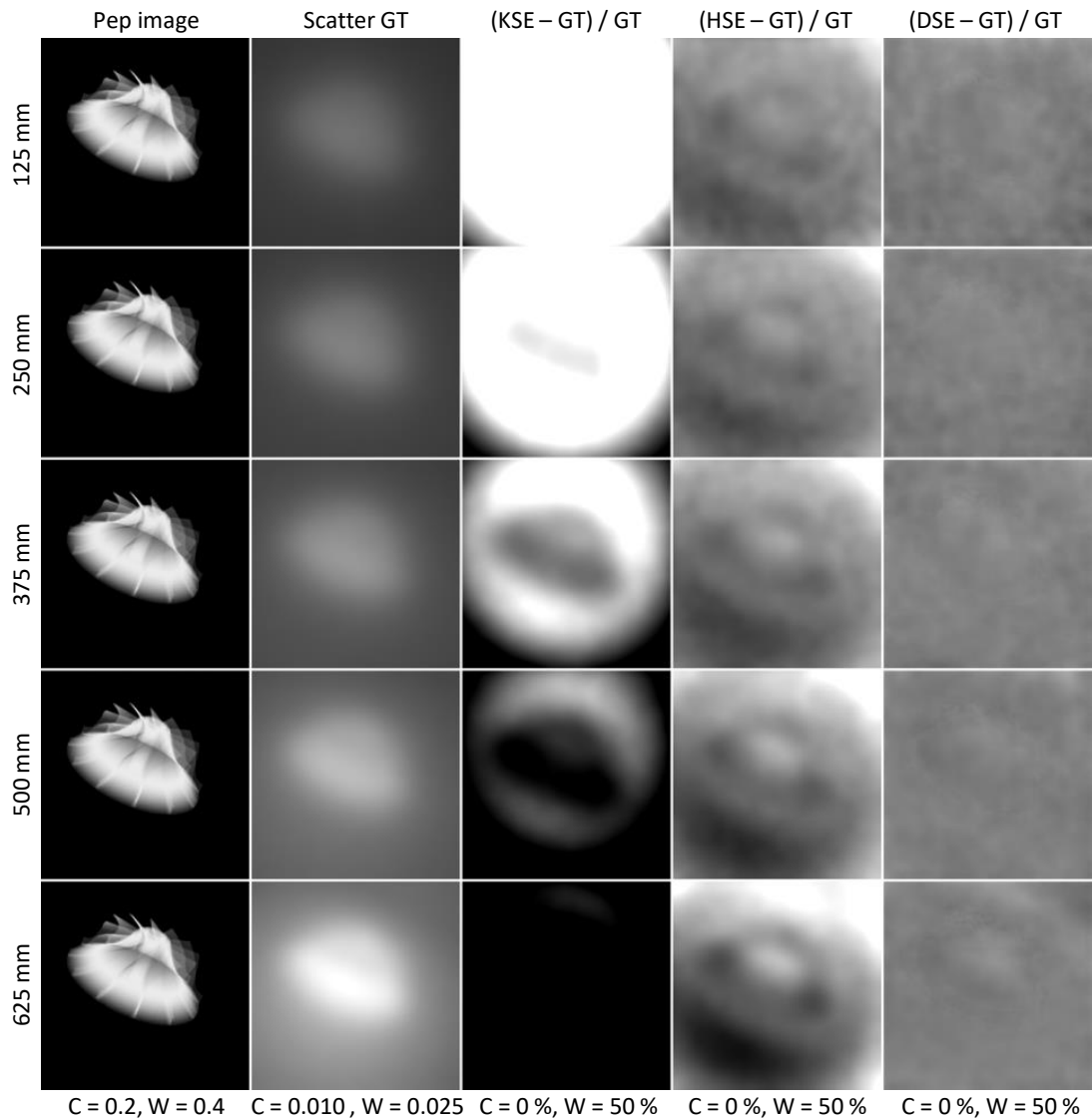Figure 5.9: Percentage error of KSE, HSE and DSE (mapping M_{pep}) scatter estimates for different magnifications. Here the size of the compressor wheel was changed for every magnification according to equation (5.9) while the density was changed according to the inverse of (5.9). KSE and DSE were optimized using all available training data (125 mm - 625 mm). HSE was optimized directly for the test data.

Table 5.11: Mean absolute percentage error of KSE, HSE and DSE scatter estimates for different magnifications. The object density is scaled according to the inverse of equation (5.9). The labels refer to the source-isocenter distance while the source-detector distance is set to a fixed value of 1000 mm. Given these measures, the magnification is calculated according to equation (5.8).

| | Testing | | | | |
	125 mm	250 mm	375 mm	500 mm	625 mm
Training					
KSE					
125 mm	14.8	22.0	38.4	53.2	72.6
250 mm	24.8	15.3	24.8	44.0	65.9
375 mm	54.2	29.8	15.3	29.1	51.1
500 mm	110.7	73.5	37.6	14.8	35.0
625 mm	211.7	157.1	100.0	48.6	14.8
125 mm & 625 mm	92.4	58.2	28.6	16.9	38.5
All magnifications	73.2	43.7	19.2	21.1	44.7
HSE					
125 mm	4.3	4.6	5.2	6.2	7.6
DSE, $M_{\mathrm{ep}} : e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$					
125 mm	5.1	16.6	34.1	50.9	66.1
250 mm	22.7	5.5	20.5	40.7	59.0
375 mm	55.5	27.9	6.0	24.9	48.1
500 mm	108.7	71.6	35.2	6.8	30.4
625 mm	204.4	150.2	97.0	46.5	7.7
125 mm & 625 mm	4.7	17.6	32.8	30.9	6.0
All magnifications	5.2	3.9	3.7	3.8	5.1
DSE, $M_{\mathrm{p}} : p_{\mathrm{sim}} \longrightarrow S_{\mathrm{MC}}$					
125 mm	1.5	3.3	6.7	13.8	28.4
250 mm	3.9	1.4	3.2	6.8	16.8
375 mm	6.1	3.3	1.2	3.6	12.2
500 mm	12.8	7.4	3.7	1.1	4.3
625 mm	19.9	11.9	6.8	3.5	1.0
125 mm & 625 mm	1.5	3.8	6.3	3.7	1.0
All magnifications	1.5	1.3	1.2	1.1	1.1
DSE, $M_{\mathrm{pep}} : p_{\mathrm{sim}} \cdot e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$					
125 mm	1.5	4.2	7.5	13.6	28.3
250 mm	7.7	1.8	5.1	8.6	15.5
375 mm	6.2	3.2	1.3	4.5	16.7
500 mm	9.7	5.7	3.0	1.0	4.4
625 mm	24.7	12.4	6.4	3.2	1.0
125 mm & 625 mm	1.4	3.1	3.5	2.9	1.0
All magnifications	1.4	1.3	1.2	1.1	1.0

v.)   Generalization to Different Components

The generalization of KSE, HSE and DSE to different components was investigated by performing simulations based on the six CAD models shown in figure 5.3 and the parameters given in table 5.5. Exemplarily, the accuracy of the three scatter estimation approaches is shown for a single projection in figure 5.10. The results of the quantitative evaluation are given in table 5.13.

Similar to the other experiments, it can be observed that a certain scatter estimation approach trained on data of a single component only, shows a decreased performance when it is applied to data of any other component that was not included within the training data set. However, there is a difference between the different components that seems to be related to their material. A scatter estimation approach that has been trained on one of the three aluminum components (cylinder head, casting, profile) also yields reasonable accuracy when it is applied to the other two components without being trained on them. In that case the average MAPE decreases from 14.5 % to 15.2 % (KSE), 10.0 % to 20.0 % (DSE, M_{ep}), 1.0 % to 5.4 % (DSE, M_p), and 1.0 % to 3.1 % (DSE, M_{pep}). In contrast, the performance is reduced significantly if it is applied to a component with higher attenuation, i.e. the compressor wheel (titanium), the bicycle cassette (iron) or the impeller (iron). Here, the average MAPE can be evaluated as 56.8 % (KSE), 87.4 % (DSE, M_{ep}), 55.6 % (DSE, M_p), and 28.2 % (DSE, M_{pep}). Interestingly, the deviation is much lower if training and testing are performed the other way round, i.e. training on one of the high attenuation components and testing on one of the aluminum components. This might be explained by the fact that the high attenuation components have a wider range of possible projection values. Thus, a network trained on aluminum components only might fail as it has not seen these projection values during training.

Considering a training on data of all components, there are similar trends for KSE. While testing on the aluminum components yields almost the same accuracy compared to training and testing on only one of them. Here the deviations are between 14.8 % (profile) and 15.5 % (cylinder head). Higher deviations up to 51.8 % (impeller) can be observed for the three components with high attenuation. This trend is also visible for HSE. However, as its parameters are recalibrated for every projection there is a lower MAPE which is between 2.0 % (profile) and 8.1 % (cassette). A similar MAPE is achieved using DSE when it is trained to learn the mapping M_{ep}. In that case it is between 4.3 % (profile) and 8.1 % (cassette). However, compared to training and testing on a single component only, the MAPE improves here considerably. This is possibly due to the higher number of samples that is used for the training when projections of all components are used. Furthermore, the higher variability of the samples might help to prevent the network from overfitting. Just as for the other experiments, DSE shows the best performance if it is trained to learn the mapping M_p or the mapping M_{pep}. For both cases there is an almost equal performance with a MAPE ranging from 0.8 % (casting) to 1.4 % (cassette).

To demonstrate the impact of scatter correction on CT images, CT reconstructions were performed for the six components shown in figure 5.3. The simulation parameters

were chosen equal to the parameters of the testing data (see table 5.5). The tube voltage was set to 250 kV and the source-to-isocenter distance as well as the isocenter-to-detector distance was set to 500 mm. For every CT simulation the scatter correction was performed as described in section 5.2.3 by subtracting KSE, HSE, and DSE scatter estimates in intensity domain. The corresponding CT reconstructions are shown in figure 5.11. A qualitative evaluation of the CT value accuracy in terms of the MAPE with respect to a noiseless, scatter free reconstruction is given in table 5.12. It can be observed that all three scatter estimation approaches lead to a significant improvement of the CT values. Similar to the evaluation in projection domain, the highest accuracy can be achieved using DSE followed by HSE and KSE. However, considering the MAPE only, all approaches seem to have similar performance. This is because the MAPE is rather insensitive to artifacts that might be introduced by an inappropriate scatter correction. This can clearly be seen for the KSE correction results (figure 5.11, third column). Here, the overestimation of scatter, especially in regions with high scatter-to-primary ratio, leads to an overestimation of the attenuation, and therefore, to bright streaks in the CT images. Being recalibrated for every projection, HSE further improves image quality but cannot completely account for all scatter artifacts. DSE, in contrast, leads to CT images that are almost equal to the scatter free ground truth.

Table 5.12: Mean absolute percentage error of CT reconstructions of the testing data set with respect to a noiseless scatter free reference. Scatter correction has been performed as described in section 5.2.3 using KSE, HSE and DSE scatter estimates. It has to be noted that errors of the scatter free reconstruction are only a result of the noise.

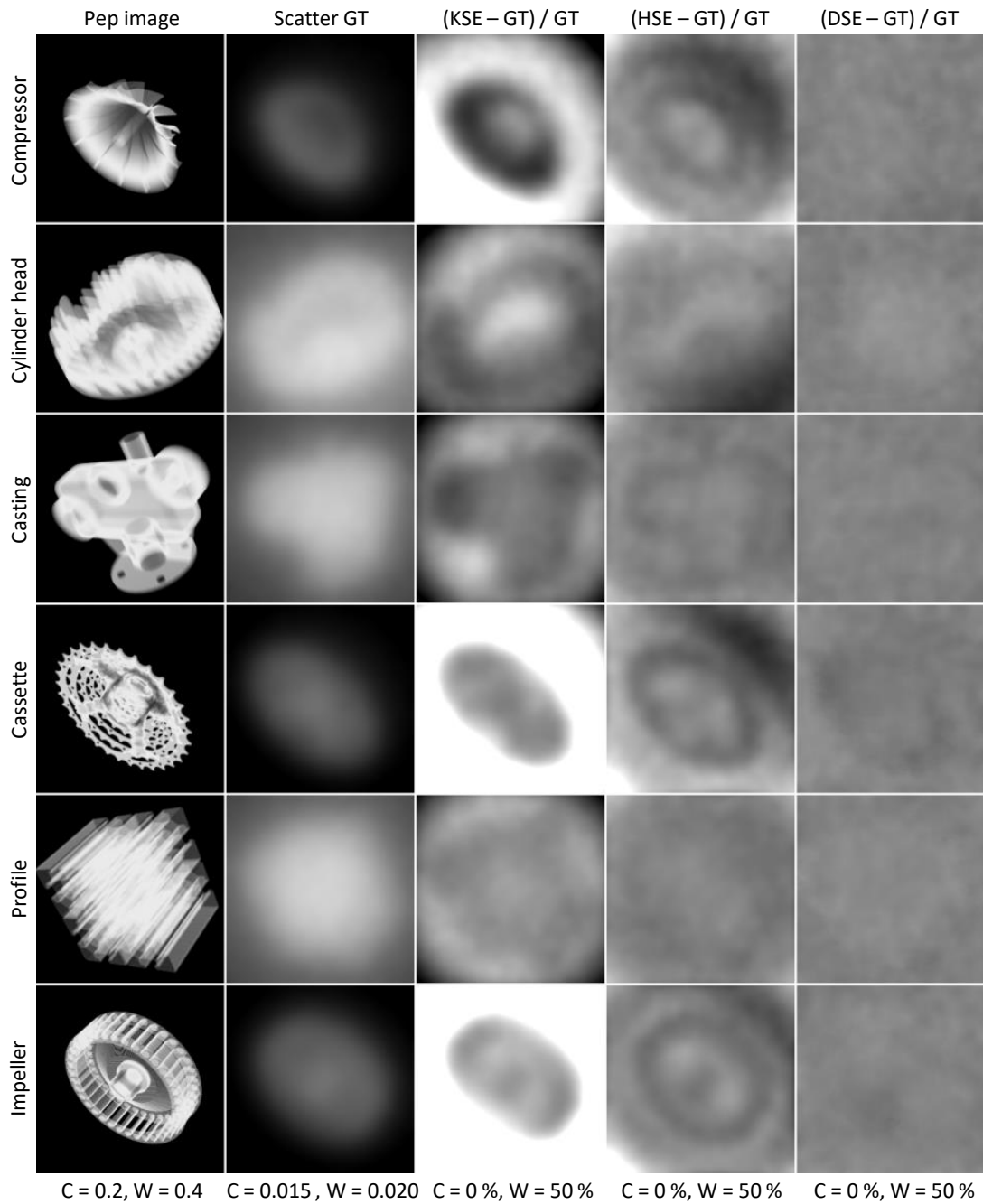|	Scatter free	No correction	KSE	HSE	DSE (M_{pep})
Compressor	6.0	18.3	10.2	7.5	6.2
Cylinder Head	5.7	23.2	11.1	7.3	5.8
Casting	3.9	14.0	8.9	6.0	4.2
Cassette	6.1	19.8	9.3	7.0	6.3
Profile	3.7	11.6	6.4	4.9	4.0
Impeller	5.1	20.9	6.3	5.9	5.3

Figure 5.10: Percentage error of KSE, HSE and DSE (mapping M_{pep}) scatter estimates for different components. KSE and DSE were optimized using a training data set containing all components. HSE was optimized directly for the test data.

Table 5.13: Mean absolute percentage error of KSE, HSE and DSE scatter estimates for different components.

|	Testing Compressor	Cylinder Head	Casting	Cassette	Profile	Impeller
Training						
KSE						
Compressor	15.4	21.6	20.8	24.3	18.9	22.8
Cylinder Head	53.0	14.6	16.3	79.5	15.1	73.7
Casting	34.2	15.3	15.2	60.6	16.7	63.1
Cassette	17.5	29.9	29.1	17.3	26.9	14.9
Profile	31.5	13.9	14.1	56.7	13.7	59.3
Impeller	18.1	31.2	30.3	17.2	28.2	14.6
Profile & Cassette	26.4	15.7	15.4	47.2	13.9	49.4
All Parts	27.8	15.5	15.4	49.5	14.8	51.8
HSE						
Compressor	7.7	5.6	3.5	8.1	2.0	5.6
DSE, $M_{\mathrm{ep}} : e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$						
Compressor	10.5	27.5	51.0	23.6	40.1	29.2
Cylinder Head	77.7	10.0	17.4	86.3	22.8	54.1
Casting	45.7	17.0	9.5	85.1	14.3	40.3
Cassette	17.2	23.9	37.0	9.4	27.6	26.4
Profile	132.8	29.4	19.0	170.8	10.5	94.2
Impeller	22.8	23.9	31.4	20.3	39.0	12.6
Profile & Cassette	21.7	16.4	39.3	12.5	11.4	21.7
All Parts	6.5	5.4	4.3	8.1	4.3	6.2
DSE, $M_{\mathrm{p}} : p_{\mathrm{sim}} \longrightarrow S_{\mathrm{MC}}$						
Compressor	1.2	8.3	4.4	4.8	6.9	4.4
Cylinder Head	36.5	1.0	3.3	49.9	5.0	28.2
Casting	20.8	4.6	1.0	24.6	6.0	12.2
Cassette	3.4	6.3	3.9	1.5	3.8	2.1
Profile	158.4	5.7	7.9	128.7	1.0	40.9
Impeller	4.5	8.9	5.9	7.2	6.4	1.2
Profile & Cassette	2.2	3.3	1.8	1.4	0.9	1.8
All Parts	1.2	1.0	0.8	1.4	0.9	1.0
DSE, $M_{\mathrm{pep}} : p_{\mathrm{sim}} \cdot e^{-p_{\mathrm{sim}}} \longrightarrow S_{\mathrm{MC}}$						
Compressor	1.3	6.1	4.0	4.3	5.3	3.8
Cylinder Head	28.1	1.0	3.8	38.1	4.9	23.4
Casting	16.3	2.3	0.9	24.9	2.1	14.3
Cassette	3.0	5.2	3.3	1.4	3.5	1.5
Profile	35.6	3.5	1.9	48.6	1.3	24.9
Impeller	5.3	8.7	5.3	4.8	6.2	1.1
Profile & Cassette	2.1	3.2	1.8	1.5	1.8	1.7
All Parts	1.2	0.9	0.8	1.4	0.9	1.1
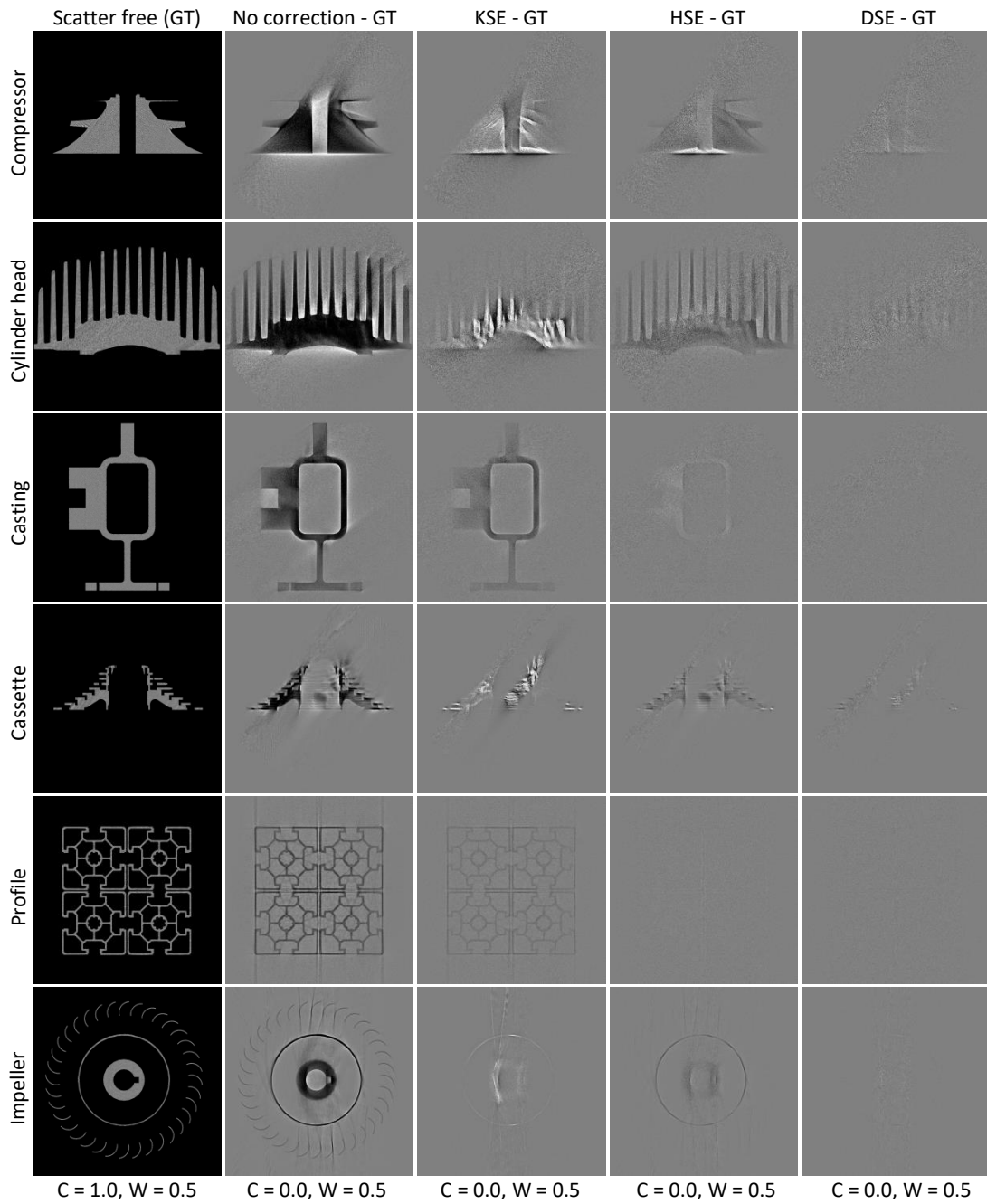
Figure 5.11: Scatter corrected CT reconstructions of simulated data. The simulations were performed assuming a tube voltage of 250 kV. Scatter was estimated using KSE, HSE and DSE (mapping M_{pep}) and subtracted in intensity domain to get a corrected data set. Prior to reconstruction the projections were precorrected such that they represent intersection lengths. Therefore, an ideal reconstruction has a CT value equal to 1.

5.3.2 Measured Data

To evaluate the potential of DSE as well as the reference approaches to estimate scatter for real data, measurements were performed at our DKFZ table-top CT using the settings given in table 5.6. Due to the absence of a ground truth scatter distribution, a slit scan measurement was performed as reference using the same parameters except for the collimation that was set to about 16 detector rows. Similar to the simulation study, the scatter estimates were subtracted in intensity domain to derive a corrected data set. Additionally, a scatter correction using an MC scatter estimate was performed. Prior to reconstruction an aluminum precorrection was applied that converts projection values to aluminum-equivalent intersection lengths. The slit scan reconstruction, the scatter-corrected reconstructions as well as difference images are shown in figure 5.12.

Visually, all scatter estimation approaches are able to reduce the scatter-related artifacts. Compared to the uncorrected reconstruction, all these images are more similar to the slit scan reconstruction. However, especially KSE and HSE lead to the introduction of bright streaks to the CT images. This is caused by an overestimation of scatter which has also been observed for the simulated data. Compared to KSE and HSE, DSE shows a better performance and yields CT images that are similar to the reconstructions that were corrected using an MC scatter estimate. However, considering the difference images (figure 5.12, bottom), both, the DSE and the MC corrected reconstructions, show some streaks which indicate an overestimation of scatter. Potentially, these streaks might also be caused by noise. Since the scatter estimates only represent the (smooth) expectation of the scatter distribution, the scatter related noise remains in the projection data after correction. Thus, compared to a slit scan that does not measure the scatter at all, there is a higher noise level. As this is especially the case for high intersection lengths, steaks may arise along the corresponding rays.

A quantitative evaluation in terms of the MAPE with respect to the slit scans yields similar trends. Here the following errors can be measured: 30.0 % (no correction), 15.9 % (KSE), 13.2 % (HSE), 6.0 % (DSE), and 5.8 % (MC). In comparison to the simulation study the performance of the scatter estimation approaches seems to be slightly decreased. However, it has to be noted that the slit scan is not completely free of scatter. Furthermore, there might be other effects such as detector backscatter, for instance, that have not been modeled in the MC simulation that is used as basis for KSE, HSE and DSE optimization.
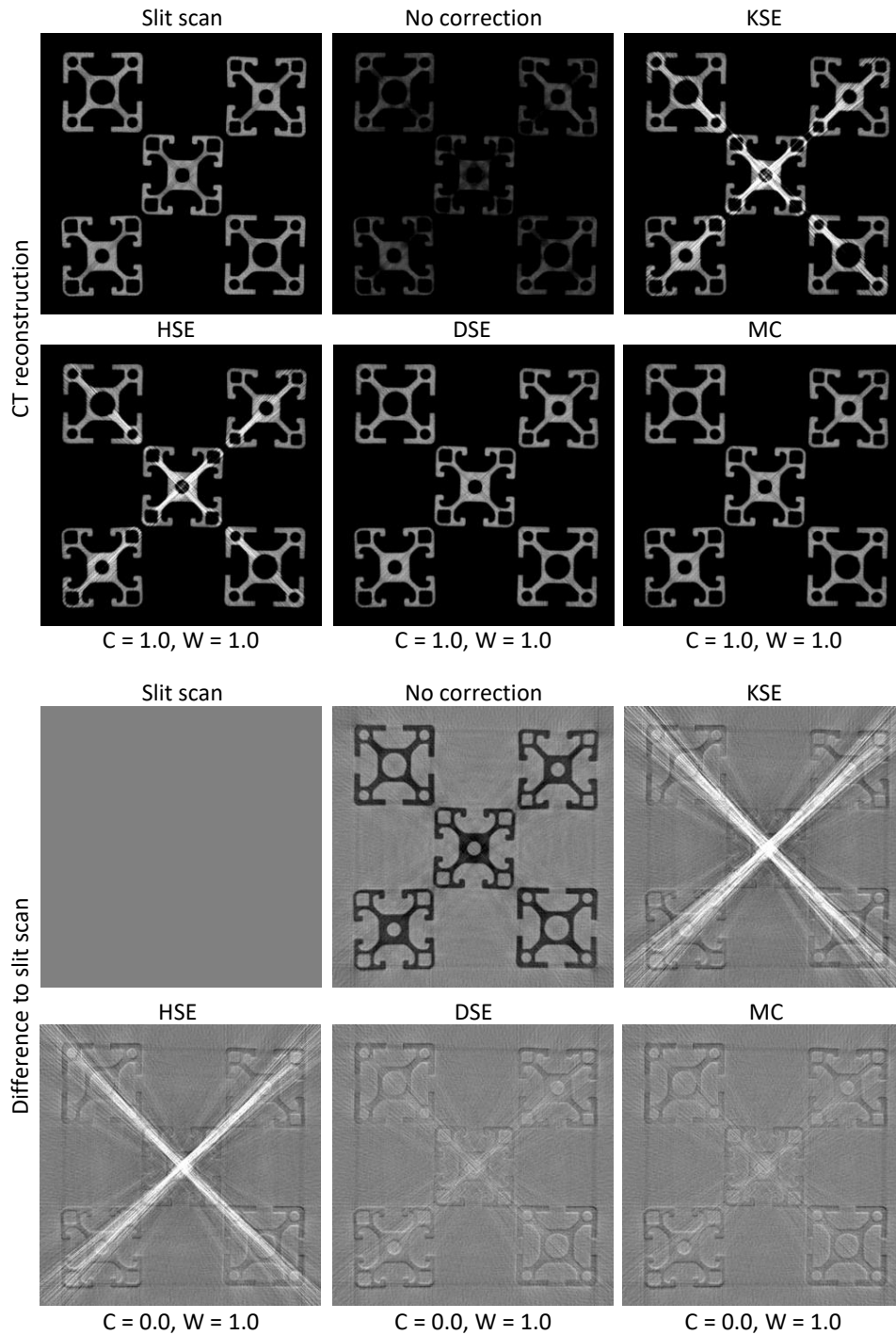
Figure 5.12: Scatter corrected CT reconstructions of measured data. Scatter was estimated using KSE, HSE, DSE (mapping M_{pep}) as well as a MC simulation. Prior to reconstruction the projections were precorrected such that they represent intersection lengths. Therefore, an ideal reconstruction has a CT value equal to 1.

5.4 Discussion

This study demonstrates the potential of using a DCNN for x-ray scatter estimation in industrial CT. Compared to existing scatter estimation approaches such as the KSE or HSE, the proposed neural network based method does not rely on a predefined theoretical model but learns a suitable model itself given labeled training data. Therefore, it can potentially derive more complex models based on local features of the acquired projection images, which are used as input to the network. Once DSE is trained, it provides scatter estimates in about 10 ms, and thus, enables a real-time scatter correction.

Considering a practical application of a certain scatter estimation approach, it has to be robust against variations of the acquisition parameters and the acquisition geometry. Therefore, a simulation study was set up to investigate the generalization of DSE with respect to different tube voltages, different noise levels, different materials, different magnifications, and different components.

Here, training DSE on samples of a single parameter realization only, e.g. on one noise level only, may lead to high deviations when it is tested on a different parameter set. This holds especially true for high noise levels, different magnifications and different components. In contrast, a DSE network trained on one tube voltage only, leads to accurate scatter estimates when it is applied to other tube voltages that were not represented in the training data set. This is possibly a result of the variability of the input and the target distributions. These are very similar for different tube voltages, but differ significantly for e.g. varying magnifications.

In any case, the use of a training data set containing samples with any parameter realization led to scatter estimates that differ by less than 2 % from the MC ground truth. Furthermore, this training does not lead to a major loss of accuracy compared to training and testing on the same parameter realization. Thus, there is no advantage in having a separately trained network for different scatter estimation tasks, but rather a single DSE network trained on representative data can be used for that purpose. Even the training of DSE on samples of two parameter realizations only, led to a good generalization to any other tested acquisition parameter. In that case, the MAPE between the ground truth and the DSE estimate did not exceed 6.1 % (at least for the mapping M_{pep}). Practically, this suggests that DSE does not need to be trained for any possible acquisition parameter, but the acquisition parameter space can be sampled on a coarse grid to generate representative training data.

Considering the accuracy of DSE, the mapping to be learned by the network turned out to be of particular importance. While training DSE to learn a mapping based on logarithmic data (M_{p}) or "pep" data (M_{pep}) yields scatter estimates that are almost equal to the MC ground truth, the intensity based mapping (M_{ep}) performs up to a magnitude worse. It has to be noted that this discrepancy is not a matter of convergence on the training data. While the training loss converges for all mappings to a similar MAPE between 0.5 % and 1.0 %, there is only a difference in terms of the testing loss. This might be explained by the similarity between the input and the scatter distribution. In case of intensity data there is a lower correlation between the scatter to be estimated and the input distribution than for projection and "pep" data. Thus, it might be more

difficult to learn meaningful features that allow for a good generalization on the test data set.

Compared to DSE, the conventional scatter estimation approaches, KSE and HSE, lead to significantly higher errors. Considering all experiments, there is an average MAPE of 27.3 % for KSE and an average MAPE of 5.6 % for HSE. Here, the poor performance of KSE can be explained by the limits of the underlying model to reproduce MC simulations as well as to generalize to different acquisition conditions. As demonstrated by HSE, more sophisticated approaches can improve the performance considerably, however, at the cost of a higher processing times ($\approx$ 7 s / projection). To the author's best knowledge, there is no conventional approach that is able to combine both, accuracy and computational performance.

Similar trends have been observed for measured data. While KSE and HSE led to an overestimation of x-ray scattering, and therefore, introduced streak artifacts to the CT reconstructions, DSE yields CT images similar to the slit scan reference. Compared to the simulation study, there seem to be slightly higher deviations. The reason for that might be twofold: firstly, the slit scan acquisition is not perfectly free of scatter and secondly, there are potentially other sources of scatter. Since only scatter from the object was modeled to simulate the training data set, DSE does not account for scattering from parts of the experimental setup or the backside of the detector, for instance. However, there is no conceptual restriction of DSE to a training on simulated data. To account for effects that can hardly be modeled, it can also be trained to reproduce scatter estimates that have been derived from measurements e.g. using a slit scan, a primary modulation grid or a beam blocker.

Finally, it has to be noted that DSE is not restricted to industrial CT in particular, but can be used with any other x-ray imaging modality, even if they are not of tomographic type.

6 | Summary and Conclusions

Today, x-ray computed tomography (CT) finds a variety of applications in industrial metrology ranging from dimensional inspection and flaw detection to reverse engineering. However, the measurement of highly attenuating or multi-material components remains a major challenge as the corresponding CT reconstructions are often corrupted by CT artifacts. To overcome this limitation, two novel approaches were presented here: the simulation-based artifact correction (SBAC) and the deep scatter estimation (DSE).

Simulation-Based Artifact Correction The SBAC represents a very general approach to correct for various CT artifacts. Given a prior model of the component to be corrected, it performs an accurate simulation of the CT measurement process, modeling all physical effects causing artifacts. The difference between this simulation and an ideal one yields an estimate of the present artifacts, and can be used to correct the corresponding CT measurement. Thus, the SBAC can, in principle, account for any artifact that is considered in the simulation. Here, the simulation was restricted to beam hardening, x-ray scattering, off-focal radiation, partial volume effects and cone-beam artifacts. Therefore, accurate and efficient models describing the generation of x-rays, the interaction of x-rays with the measured object, and the detection of x-rays were developed or existing models were refined, respectively. It could be demonstrated that the proposed models differ by less than 10 % from reference Monte Carlo (MC) simulations, while being orders of magnitude faster.

The potential of the SBAC was evaluated for metrological CT measurements of single- and multi-material components. Based on these measurements, it was shown for the first time that the most common CT artifacts can be corrected with a single correction term derived from CT simulations. Here, the SBAC provided CT reconstructions that showed almost no artifacts and whose quality was clearly superior to state-of-the-art reference approaches. Furthermore, it could be shown that the SBAC not only increases the visual impression of the CT reconstructions, but also their dimensional accuracy. The comparison of surface meshes extracted from SBAC-corrected CT reconstructions revealed only minor deviations from tactile measurements, the gold standard of dimensional metrology. This is important to meet the high demands on the quality of the surface meshes, especially when investigating components with small manufacturing tolerances in the range of a few tens of microns.

Thus, the SBAC helps to extend the applicability of CT in the field of industrial

metrology and to establish CT as an alternative to tactile measurements even for highly attenuating components.

Deep Scatter Estimation   In the context of scatter artifact correction, the DSE was developed to solve the problem of long processing times of accurate scatter estimation approaches. Therefore, DSE trained a deep convolutional neural network (DCNN) to reproduce MC scatter simulations given only a function of the acquired projection data as input. Once, the network is trained, it can be applied to unknown data in real-time (≈ 10 ms / projection). In that, process DSE does not rely on a certain theoretically motivated scattering model, but learns the most suitable model itself from observational data. This is especially an advantage if a certain scatter estimation approach needs to be adapted to novel data, e.g. data that have been acquired with different acquisition parameters or at a different system. While conventional approaches might require to refine the underlying theoretical model, DSE can be adapted by simply exchanging or extending the training data set.

To demonstrate the practical applicability of DSE, different simulation studies and measurements were carried out. Here, the simulation study demonstrated that DSE generalizes well to different tube voltages, different noise levels, different materials, different magnifications, as well as different components. In particular, it could be shown that DSE is clearly superior to conventional reference methods, and thereby, provides scatter distributions that deviate on average by less than 2 % from MC simulations. Furthermore, these experiments suggest that a single DSE network, trained on representative data, can be used universally for various scatter estimation tasks.

A similarly good performance could be observed for measured data of an experimental CBCT system. Here, DSE provided scatter-corrected CT reconstructions whose quality was almost equal to slit scan measurements. Furthermore, this study demonstrated that a DSE network, trained on simulated data, also applies to measured data. This is of particular importance as simulation is an easy way to generate an arbitrary number of training examples. However, it has to be noted that several efforts have been made here to tune the simulations to reproduce measurements of the experimental CBCT system. To which extent the simulation must match the measured data is the subject of further investigations.

Obviously, if a sufficiently accurate simulation is practically not possible, DSE can also be trained using measured data, e.g. of a slit scan or a beam blocker measurement. Conceptually, there are no restrictions for the generation of training data.

It can be concluded that DSE makes an important contribution to improve the accuracy of x-ray scatter correction. Therefore, DSE helps to further improve the quality of metrological CT measurements, especially in case of time-critical appliactions such as inline-CT measurements, for instance, that require real-time capable correction approaches.

Bibliography

[1] G. N. Hounsfield, "A method of and apparatus for examination of a body by radiation such as X or gamma radiation", *The Patent Office, London. Patent Specification 1283915*, 1972.

[2] A. M. Cormack, "Representation of a function by its line integrals, with some radiological applications", *Journal of Applied Physics*, volume 34, number 9, pages 2722–2727, 1963.

[3] A. M. Cormack, "Representation of a function by its line integrals, with some radiological applications. II", *Journal of Applied Physics*, volume 35, number 10, pages 2908–2913, 1964.

[4] A. M. Cormack, "Reconstruction of densities from their projections, with applications in radiological physics", *Physics in Medicine and Biology*, volume 18, number 2, pages 195–207. 1973.

[5] J. W. Kress and L. A. Feldkamp, "X-Ray tomography applied to NDE of ceramics", in *Volume 5: Ceramics; Structures and Dynamics; Controls, Diagnostics and Instrumentation; Education; Process Industries*, 1983, pages 182–187.

[6] W. Gilboy, "X-and γ-ray tomography in NDE applications", *Nuclear Instruments and Methods in Physics Research*, volume 221, number 1, pages 193–200, 1984.

[7] P. Reimers, W. Gilboy, and J. Goebbels, "Recent developments in the industrial application of computerized tomography with ionizing radiation", *NDT International*, volume 17, number 4, pages 197–207, 1984.

[8] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm", *Journal of the Optical Society of America A*, volume 1, number 6, pages 612–619, 1984.

[9] G. Georgeson and R. Bossi, "X-ray CT for quantitative casting material evaluation", *NDT & E International*, volume 27, number 2, page 101, 1994.

[10] S. T. Neel, D. S. Eliasen, and R. N. Yancey, "Dimensional measurement of internal features in complex castings", in *Review of Progress in Quantitative Nondestructive Evaluation*, Boston, MA: Springer US, 1995, pages 689–694.

[11]   S. Trent Neel, R. Gibson, C. R. Daniels, and E. L. Klosterman, "Dimensional accuracy in X-ray computed tomography imaging", in *Review of Progress in Quantitative Nondestructive Evaluation*, Boston, MA: Springer US, 1998, pages 411–412.

[12]   L. De Chiffre, S. Carmignato, J.-P. Kruth, R. Schmitt, and A. Weckenmann, "Industrial applications of computed tomography", *CIRP Annals - Manufacturing Technology*, volume 63, number 2, pages 655–677, 2014.

[13]   J. J. Lifton, A. A. Malcolm, and J. W. McBride, "An experimental study on the influence of scatter and beam hardening in x-ray CT for dimensional metrology", *Measurement Science and Technology*, volume 27, number 1, pages 192–199, 2016.

[14]   T. Buzug, *Computed tomography*. Springer Berlin Heidelberg, 2008.

[15]   J. F. Barrett and N. Keat, "Artifacts in CT: Recognition and avoidance.", *Radiographics*, volume 24, number 6, pages 1679–91, 2004.

[16]   M. Kachelrieß, K. Sourbelle, and W. A. Kalender, "Empirical cupping correction: A first-order raw data precorrection for cone-beam computed tomography", *Medical Physics*, volume 33, number 5, pages 1269–1274, 2006.

[17]   E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization.", *Physics in Medicine and Biology*, volume 53, number 17, pages 4777–4807, 2008.

[18]   M. Krumm, S. Kasperl, and M. Franz, "Reducing non-linear artifacts of multi-material objects in industrial 3D computed tomography", *NDT & E International*, volume 41, number 4, pages 242–251, 2008.

[19]   E. Meyer, C. Maaß, M. Baer, R. Raupach, B. Schmidt, and M. Kachelrieß, "Empirical scatter correction (ESC): A new CT scatter correction method and its application to metal artifact reduction", in *IEEE Medical Imaging Conference*, 2010, pages 2036–2041.

[20]   E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (NMAR) in computed tomography", *Medical Physics*, volume 37, number 2010, pages 5482–5493, 2012.

[21]   J. W. Stayman, Y. Otake, J. L. Prince, A. J. Khanna, and J. H. Siewerdsen, "Model-based tomographic reconstruction of objects containing known components", *IEEE Transactions on Medical Imaging*, volume 31, number 10, pages 1837–1848, 2012.

[22]   J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography.", *Physics in Medicine and Biology*, volume 58, number 12, R63–96, 2013.

[23] L. Brabant, M. Dierick, E. Pauwels, M. N. Boone, and L. Van Hoorebeke, "EDART, a discrete algebraic reconstructing technique for experimental data obtained with high resolution computed tomography.", *Journal of X-ray Science and Technology*, volume 22, number 1, pages 47–61, 2014.

[24] Y. Zhao and M. Li, "Iterative beam hardening correction for multi-material objects", *PLOS ONE*, volume 10, number 12, L. Zeng, Ed., e0144607, 2015.

[25] H. S. Park, D. Hwang, and J. K. Seo, "Metal artifact reduction for polychromatic x-ray CT based on a beam-hardening corrector", *IEEE Transactions on Medical Imaging*, volume 35, number 2, pages 480–487, 2016.

[26] 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*, volume 38, number 7, pages 4296–4311, 2011.

[27] 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*, volume 38, number 9, pages 5186–5199, 2011.

[28] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction", *Medical Physics*, volume 44, number 10, e360–e375, 2017.

[29] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network", *IEEE Transactions on Medical Imaging*, volume 36, number 12, pages 2524–2535, 2017.

[30] X. Zheng, Z. Lu, S. Ravishankar, Y. Long, and J. A. Fessler, "Low dose CT image reconstruction with learned sparsifying transform", in *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2016, pages 1–5.

[31] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network.", *Biomedical Optics Express*, volume 8, number 2, pages 679–694, 2017.

[32] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT", *IEEE Transactions on Medical Imaging*, volume 36, number 12, pages 2536–2545, 2017.

[33] V. Kearney, S. Haaf, A. Sudhyadhom, G. Valdes, and T. D. Solberg, "An unsupervised convolutional neural network-based algorithm for deformable image registration", *Physics in Medicine and Biology*, volume 63, number 18, page 185 017, 2018.

[34] Y. Zhang and H. Yu, "Convolutional neural network based metal artifact reduction in X-ray computed tomography", *IEEE Transactions on Medical Imaging*, volume 37, number 6, pages 1370–1381, 2018.

[35] S. Xu and H. Dang, "Deep residual learning enabled metal artifact reduction in CT", in *Proceedings of the SPIE Medical Imaging Conference*, 2018.

[36] H. Yu, Y. Zhang, and Y. Chu, "Reduction of metal artifacts in x-ray CT images using a convolutional neural network", in *SPIE Optical Engineering and Applications*, 2017.

[37] J. D. Dormer, B. Fei, M. Halicek, L. Ma, C. M. Reilly, and E. Schreibmann, "Heart chamber segmentation from CT using convolutional neural networks", in *Proceedings of the SPIE Medical Imaging Conference*, 2018.

[38] Q. Zheng, H. Delingette, N. Duchateau, and N. Ayache, "3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation", *IEEE Transactions on Medical Imaging*, volume 37, number 9, pages 2137–2148, 2018.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation", *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[40] S. Chen, J. Endres, S. Dorn, J. Maier, M. Lell, M. Kachelrieß, and A. Maier, "A feasibility study of automatic multi-organ segmentation using probabilistic atlas", in *Bildverarbeitung für die Medizin*, 2017.

[41] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: application to sparse-view CT", *IEEE Transactions on Medical Imaging*, volume 37, number 6, pages 1418–1429, 2018.

[42] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier, "Deep learning computed tomography: learning projection-domain weights from image domain in limited angle problems", *IEEE Transactions on Medical Imaging*, volume 37, number 6, pages 1454–1463, 2018.

[43] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution", *IEEE Transactions on Medical Imaging*, volume 37, number 6, pages 1407–1417, 2018.

[44] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "LEARn: learned experts' assessment-based reconstruction network for sparse-data CT", *IEEE Transactions on Medical Imaging*, volume 37, number 6, pages 1333–1347, 2018.

[45] D. Meschede, *Gerthsen Physik*, 24th edition. Springer Berlin Heidelberg, 2010.

[46] H. Bethe, "Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie", *Annalen der Physik*, volume 397, number 3, pages 325–400, 1930.

[47] H. Bethe, "Bremsformel für Elektronen relativistischer Geschwindigkeit", *Zeitschrift für Physik*, volume 76, number 5-6, pages 293–299, 1932.

[48] F. Sauter, "Über die Bremsstrahlung schneller Elektronen", *Annalen der Physik*, volume 412, number 4, pages 404–412, 1934.

[49] H. J. Bhabha, "On the penetrating component of cosmic radiation", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, volume 164, number 917, pages 257–294, 1938.

[50] U. Fano, "Gamma ray attenuation part I basic processes", *Nucleonics*, volume 11, 1953.

[51] A. Beer, "Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten", *Annalen der Physik und Chemie*, volume 162, number 5, pages 78–88, 1852.

[52] J. H. Lambert, *Photometria, sive de mensura et gradibus luminis, colorum et umbrae*. Eberhard Klett, 1760.

[53] P. Bouguer, *Essai d'Optique sur la Gradation de la Lumière*. Paris: Claude Jombert, 1729.

[54] R. H. Pratt, A. Ron, and H. K. Tseng, "Atomic photoelectric effect above 10 keV", *Reviews of Modern Physics*, volume 45, pages 273–325, 1973.

[55] M. Stobbe, "Zur Quantenmechanik photoelektrischer Prozesse", *Annalen der Physik*, volume 399, number 6, pages 661–715, 1930.

[56] D. E. Cullen, J. H. Hubbell, and L. Kissel, "EPDL97: The evaluated photon data library, '97 version", Tech. Rep., 1997.

[57] J. H. Hubbell and H. A. Gimm, "Pair, triplet, and total atomic cross sections (and mass attenuation coefficients) for 1 MeV–100 GeV photons in elements Z =1 to 100", *Journal of Physical and Chemical Reference Data*, volume 9, number 4, pages 1023–1148, 1980.

[58] J. Scofield, "Theoretical photoionization cross sections from 1 to 1500 keV.", Divison of Technical Information Extension, U.S. Atomic Energy Commission, Tech. Rep., 1973.

[59] A. H. Compton, "A quantum theory of the scattering of X-rays by light elements", *Physical Review*, volume 21, number 5, pages 483–502, 1923.

[60] O. Klein and Y. Nishina, "Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac", *Zeitschrift für Physik*, volume 52, number 11-12, pages 853–868, 1929.

[61] J. H. Hubbell, W. J. Veigele, E. A. Briggs, R. T. Brown, D. T. Cromer, and R. J. Howerton, "Atomic form factors, incoherent scattering functions, and photon scattering cross sections", *Journal of Physical and Chemical Reference Data*, volume 4, number 3, pages 471–538, 1975.

[62] J. Hubbell, "Summary of existing information on the incoherent scattering of photons, particularly on the validity of the use of the incoherent scattering function", *Radiation Physics and Chemistry*, volume 50, number 1, pages 113–124, 1997.

[63] M. Born, *Atomic physics*, 8th edition. London: Blackie & Son Ltd, 1969.

[64] G. Wendin, "Anomalous X-ray scattering", *Physica Scripta*, volume 21, number 3-4, pages 535–542, 1980.

[65] D. Cullen, "Program SCATMAN: A code designed to calculate photon coherent scattering anomalous scattering factors and cross sections", *Lawrence Livermore National Laboratory, UCRL-ID-103422*, 1989.

[66] M. Nikl, "Scintillation detectors for x-rays", *Measurement Science and Technology*, volume 17, number 4, R37–R54, 2006.

[67] P. a. Rodnyi, P. Dorenbos, and C. W. E. van Eijk, "Energy loss in inorganic scintillators", *Physica Status Solidi (B)*, volume 187, number 1, pages 15–29, 1995.

[68] C. Greskovich and S. Duclos, "Ceramic scintillators", *Annual Review of Materials Science*, volume 27, number 1, pages 69–88, 1997.

[69] G. Knoll, *Radiation detection and measurement*, Third Edit. New York: John Wiley & Sons, 2000, page 816.

[70] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. Description of system", *The British Journal of Radiology*, volume 46, number 552, pages 1016–1022, 1973.

[71] J. Radon, "Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten", *Akademie der Wissenschaften zu Leipzig*, volume 69, pages 262–277, 1917.

[72] C. Heinzl, "Making X-ray nano-tomography accessible with a stand-alone device", in *AAAS Annual Meeting*, 2015.

[73] G.-L. L. Comte de Buffon, "Essai d'arithmétique morale",

[74] F. James, "A review of pseudorandom number generators", *Computer Physics Communications*, volume 60, number 3, pages 329–344, 1990.

[75] J. Von Neumann, "Various techniques used in connection with random digits", *National Bureau of Standards Applied Mathematics*, volume 3, number x, pages 36–38, 1951.

[76] D. Lehmer, "Mathematical models in large scale computing unit in proceedings of the second symposium on large scale digital computing machinery", *Annals of the Computation Laboratory of Harvard University*, volume XXVI, pages 141–146, 1951.

[77] W. H. Press and S. A. Teukolsky, "Portable random number generators", *Computers in Physics*, volume 6, number 5, page 522, 1992.

[78] G. Marsaglia, "Xorshift RNGs", *Journal of Statistical Software*, volume 8, number 14, 2003.

[79] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: The art of scientific computing*, 3rd edition. New York, NY, USA: Cambridge University Press, 2007.

[80] J. von Neumann, "Various techniques used in connection with random digits", in *Monte Carlo Method*, A. S. Householder, G. E. Forsythe, and H. H. Germond, Eds., Washington, D.C.: U.S. Government Printing Office: National Bureau of Standards Applied Mathematics Series, 12, 1951, pages 36–38.

[81] E. Woodcock, T. Murphy, P. Hemmings, and S. Longworth, "Techniques used in the GEM code for Monte Carlo neutronics calculations in reactors and other systems of complex geometry", *Proc. Conf. on Applications of Computing Methods to Reactor Problems Argonne National Laboratories Report ANL-7050*, number 3, pages 557–577, 1965.

[82] A. F. Bielajew and D. W. O. Rogers, "Variance-reduction techniques", in *Monte Carlo Transport of Electrons and Photons*, Boston, MA: Springer US, 1988, pages 407–419.

[83] E. Mainegra-Hing and I. Kawrakow, "Variance reduction techniques for fast Monte Carlo CBCT scatter correction calculations.", *Physics in Medicine and Biology*, volume 55, number 16, pages 4495–4507, 2010.

[84] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.

[85] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudík, Eds., series Proceedings of Machine Learning Research, volume 15, Fort Lauderdale, FL, USA, 2011, pages 315–323.

[86] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Mathematics of Control, Signals, and Systems*, volume 2, number 4, pages 303–314, 1989.

[87] M. A. Nielsen, *Neural networks and deep learning*. Determination Press, 2015.

[88] A.-L. Cauchy, "Methode generale pour la resolution des systemes d'equations simultanees", *Compte Rendu des Seances de L'Academie des Sciences*, volume 25, number 2, pages 536–538, 1847.

[89] B. Polyak, "Some methods of speeding up the convergence of iteration methods", *USSR Computational Mathematics and Mathematical Physics*, volume 4, number 5, pages 1–17, 1964.

[90] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization", *Journal of Machine Learning Research*, volume 12, pages 2121–2159, 2011.

[91] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups", *IEEE Signal Processing Magazine*, volume 29, number 6, pages 82–97, 2012.

[92] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization", in *International Conference on Learning Representations*, 2015.

[93] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", *Nature*, volume 323, number 6088, pages 533–536, 1986.

[94] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting", *Journal of Machine Learning Research*, volume 15, pages 1929–1958, 2014.

[95] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition", *Neural Computation*, volume 1, number 4, pages 541–551, 1989.

[96] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex", *The Journal of Physiology*, volume 148, number 3, pages 574–591, 1959.

[97] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex", *The Journal of Physiology*, volume 195, number 1, pages 215–243, 1968.

[98] H. Dang, J. W. Stayman, A. Sisniega, J. Xu, W. Zbijewski, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and J. H. Siewerdsen, "Statistical reconstruction for cone-beam CT with a post-artifact-correction noise model: application to high-quality head imaging.", *Physics in Medicine and Biology*, volume 60, number 16, pages 6153–75, 2015.

[99] S. T. II, J. H. Siewerdsen, and J. W. Stayman, "Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise", *Physics in Medicine and Biology*, volume 61, number 1, page 296, 2016.

[100] R. Christoph, H. Weise, C. Leinweber, and M. Kachelrieß, "Examination of a method for the optimization of scan parameters for measuring with computed tomography", in *8th Conference on Industrial Computed Tomography*, 2018.

[101] M. Kachelrieß and W. Kalender, "Improving PET/CT attenuation correction with iterative beam hardening correction", in *IEEE Nuclear Science Symposium Conference Record, 2005*, volume 4, 2005, pages 1905–1909.

[102] J. Maier, C. Leinweber, S. Sawall, H. Stoschus, F. Ballach, T. Müller, M. Hammer, R. Christoph, and M. Kachelrieß, "Simulation-based artifact correction (SBAC) for metrological computed tomography", *Measurement Science and Technology*, volume 28, number 6, page 065 011, 2017.

[103] D. M. Tucker, G. T. Barnes, and D. P. Chakraborty, "Semiempirical model for generating tungsten target x-ray spectra", *Medical Physics*, volume 18, number 2, pages 211–218, 1991.

[104] R. D. Evans, *The atomic nucleus*, 3. 1956, volume 2, 11S–12S.

[105] R. Whiddington, "The transmission of cathode rays through matter", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, volume 86, number 588, pages 360–370, 1912.

[106] N. Dyson, *X-Rays in atomic and nuclear physics*. Cambridge: Cambridge University Press, 1990.

[107] S. T. Perkins, D. E. Cullen, M. H. Chen, J. H. Hubbell, J. Rathkopf, and J. Scofield, "EADL: Evaluated atomic data library of the Lawrence Livermore National Laboratory, USA Summary documentation", International Atomic Energy Agency (IAEA), Tech. Rep., 1994, page 3.

[108] Y. Kyriakou, T. Riedel, and W. A. Kalender, "Combining deterministic and Monte Carlo calculations for fast estimation of scatter intensities in CT.", *Physics in medicine and biology*, volume 51, number 18, pages 4567–4586, 2006.

[109] S. Agostinelli, J. Allison, K. Amako, *et al.*, "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 506, number 3, pages 250–303, 2003.

[110] H. O. Wyckoff, A. Allisy, H. Fränz, W. A. Jennings, A. M. Kellerer, K. Lidén, and H. H. Rossi, "Report 33", *Journal of the International Commission on Radiation Units and Measurements*, volume os17, number 2, NP–NP, 1980.

[111] S. Perkins, D. Cullen, and S. Seltzer, "Tables and graphs of electron-interaction cross sections from 10 eV to 100 GeV derived from the LLNL evaluated electron data library (EEDL), Z = 1–100", Lawrence Livermore National Laboratory (LLNL), Livermore, CA, Tech. Rep., 1991.

[112] G. G. Poludniowski, "Calculation of x-ray spectra emerging from an x-ray tube. Part II. X-ray production and filtration in x-ray targets", *Medical Physics*, volume 34, number 6Part1, pages 2175–2186, 2007.

[113] G. G. Poludniowski and P. M. Evans, "Calculation of x-ray spectra emerging from an x-ray tube. Part I. Electron penetration characteristics in x-ray targets", *Medical Physics*, volume 34, number 6, pages 2164–2174, 2007.

[114] C. Leinweber, J. Maier, and M. Kachelrieß, "X-ray spectrum estimation for accurate attenuation simulation", *Medical Physics*, volume 44, number 12, pages 6183–6194, 2017.

[115] L. Brabant, E. Pauwels, M. Dierick, D. Van Loo, M. Boone, and L. Van Hoorebeke, "A novel beam hardening correction method requiring no prior knowledge, incorporated in an iterative reconstruction algorithm", *NDT & E International*, volume 51, pages 68–73, 2012.

[116] K. Dremel and T. Fuchs, "Scatter simulation and correction in computed tomography: A reconstruction-integrated approach modelling the forward projection", *NDT & E International*, volume 86, pages 132–139, 2017.

[117]   J. W. Stayman, H. Dang, Y. Ding, and J. H. Siewerdsen, "PIRPLE: a penalized-likelihood framework for incorporation of prior images in CT reconstruction", *Physics in Medicine and Biology*, volume 58, number 21, page 7563, 2013.

[118]   M. Schrapp, T. Scharrer, M. Goldammer, S. J. Rupitsch, A. Sutor, H. Ermert, and R. Lerch, "Artifact reduction in non-destructive testing by means of complementary data fusion of X-ray computed tomography and ultrasonic pulse-echo testing", *Measurement Science and Technology*, volume 24, number 12, page 125 403, 2013.

[119]   Y. Liu, A. Beyer, P. Schuetz, J. Hofmann, A. Flisch, and U. Sennhauser, "Cooperative data fusion of transmission and surface scan for improving limited-angle computed tomography reconstruction", *NDT & E International*, volume 83, pages 24–31, 2016.

[120]   Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, "Empirical beam hardening correction (EBHC) for CT.", *Medical Physics*, volume 37, number 10, pages 5179–5187, 2010.

[121]   W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants.", *Radiology*, volume 164, number 2, pages 576–577, 1987.

[122]   A. H. Mahnken, R. Raupach, J. E. Wildberger, B. Jung, N. Heussen, T. G. Flohr, R. W. Günther, and S. Schaller, "A new algorithm for metal artifact reduction in computed tomography", *Investigative Radiology*, volume 38, number 12, pages 769–775, 2003.

[123]   E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Frequency split metal artifact reduction (FSMAR) in computed tomography", *Medical Physics*, volume 39, number 4, page 1904, 2012.

[124]   J. Wang, S. Wang, Y. Chen, J. Wu, J.-L. Coatrieux, and L. Luo, "Metal artifact reduction in CT using fusion based prior image.", *Medical Physics*, volume 40, number 8, page 081 903, 2013.

[125]   A. Amirkhanov, C. Heinzl, M. Reiter, J. Kastner, and E. Groller, "Projection-based metal-artifact reduction for industrial 3D X-ray computed tomography", *IEEE Transactions on Visualization and Computer Graphics*, volume 17, number 12, pages 2193–2202, 2011.

[126]   G. T. Herman, "Correction for beam hardening in computed tomography.", *Physics in Medicine and Biology*, volume 24, number 1, pages 81–106, 1979.

[127]   P. Hammersberg and M. Mångård, "Correction for beam hardening artefacts in computerised tomography.", *Journal of X-Ray Science and Technology*, volume 8, number 1, pages 75–93, 1998.

[128]   W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm", *SIGGRAPH Comput. Graph.*, volume 21, number 4, pages 163–169, 1987.

[129] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based CT image reconstruction applied to clinical data.", *Physics in Medicine and Biology*, volume 56, number 6, pages 1545–1561, 2011.

[130] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): Accurate real-time scatter estimation for X-ray CT using a deep convolutional neural network", *Journal of Nondestructive Evaluation*, volume 37, number 3, page 57, 2018.

[131] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation.", *Medical Physics*, volume 46, number 1, pages 238–249, 2019.

[132] R. Ning, X. Tang, and D. Conover, "X-ray scatter correction algorithm for cone-beam CT imaging", *Medical Physics*, volume 31, number 5, pages 1195–1202, 2004.

[133] L. Zhu, N. Strobel, and R. Fahrig, "X-ray scatter correction for cone-beam CT using moving blocker array", in *Proceedings of the SPIE Medical Imaging Conference*, M. J. Flynn, Ed., volume 5745, 2005, page 251.

[134] L. Zhu, N. R. Bennett, and R. Fahrig, "Scatter correction method for X-ray ct using primary modulation: Theory and preliminary results", *IEEE Transactions on Medical Imaging*, volume 25, number 12, pages 1573–1587, 2006.

[135] K. Schorner, M. Goldammer, K. Stierstorfer, J. Stephan, and P. Boni, "Scatter correction method by temporal primary modulation in X-ray CT", *IEEE Transactions on Nuclear Science*, volume 59, number 6, pages 3278–3285, 2012.

[136] R. Grimmer, R. Fahrig, W. Hinshaw, H. Gao, and M. Kachelrieß, "Empirical cupping correction for ct scanners with primary modulation (ECCP)", *Medical Physics*, volume 39, number 2, pages 825–831, 2012.

[137] L. Ritschl, R. Fahrig, M. Knaup, J. Maier, and M. Kachelrieß, "Robust primary modulation-based scatter estimation for cone-beam CT", *Medical Physics*, volume 42, number 1, pages 469–478, 2015.

[138] S. Hsieh, "Estimating scatter in cone-beam CT with striped ratio grids: A preliminary investigation", *Medical Physics*, volume 43, number 9, pages 5084–5092, 2016.

[139] B. Bier, M. Berger, A. Maier, M. Kachelrieß, L. Ritschl, K. Müller, J.-h. Choi, and R. Fahrig, "Scatter correction using a primary modulator on a clinical angiography C-arm CT system", *Medical Physics*, volume 44, number 9, e125–e137, 2017.

[140] W. Swindell and P. M. Evans, "Scattered radiation in portal images: A Monte Carlo simulation and a simple physical model", *Medical Physics*, volume 23, number 1, pages 63–73, 1996.

[141] B. Ohnesorge, T. Flohr, and K. Klingenbeck-Regn, "Efficient object scatter correction algorithm for third and fourth generation CT scanners", *European Radiology*, volume 9, number 3, pages 563–569, 1999.

[142] W. Zbijewski and F. Beekman, "Fast scatter estimation for cone-beam X-ray CT by combined Monte Carlo tracking and Richardson-Lucy fitting", *IEEE Medical Imaging Conference*, volume 5, number C, pages 2774–2777, 2004.

[143] 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*, volume 53, number 23, pages 6729–6748, 2008.

[144] G. Poludniowski, P. M. Evans, V. N. Hansen, and S. Webb, "An efficient Monte Carlo-based algorithm for scatter correction in keV cone-beam CT", *Physics in Medicine and Biology*, volume 54, number 12, pages 3847–3864, 2009.

[145] J. Star-Lack, M. Sun, A. Kaestner, R. Hassanein, G. Virshup, T. Berkus, and M. Oelhafen, "Efficient scatter correction using asymmetric kernels", in *Proceedings of the SPIE Medical Imaging Conference*, E. Samei and J. Hsieh, Eds., volume 7258, 2009, 72581Z.

[146] M. Meyer, W. A. Kalender, and Y. Kyriakou, "A fast and pragmatic approach for scatter correction in flat-detector CT using elliptic modeling and iterative optimization", *Physics in Medicine and Biology*, volume 55, number 1, pages 99–120, 2010.

[147] M. Sun and J. M. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition", *Physics in Medicine and Biology*, volume 55, number 22, pages 6695–6720, 2010.

[148] W. Yao and K. W. Leszczynski, "An analytical approach to estimating the first order scatter in heterogeneous medium. II. A practical application", *Medical Physics*, volume 36, number 7, pages 3157–3167, 2009.

[149] M. Baer and M. Kachelrieß, "Hybrid scatter correction for CT imaging", *Physics in Medicine and Biology*, volume 57, number 21, pages 6849–6867, 2012.

[150] A. Sossin, V. Rebuffel, J. Tabary, J. M. Létang, N. Freud, and L. Verger, "A novel scatter separation method for multi-energy x-ray imaging", *Physics in Medicine and Biology*, volume 61, number 12, pages 4711–4728, 2016.

[151] A. Maslowski, A. Wang, M. Sun, T. Wareing, I. Davis, and J. Star-Lack, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - part I: Core algorithms and validation", *Medical Physics*, volume 45, number 5, pages 1899–1913, 2018.

[152] A. Wang, A. Maslowski, P. Messmer, M. Lehmann, A. Strzelecki, E. Yu, P. Paysan, M. Brehm, P. Munro, J. Star-Lack, and D. Seghers, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - part II: System modeling, scatter correction, and optimization.", *Medical Physics*, volume 45, number 5, pages 1914–1925, 2018.

[153] M. Hoffmann, W. Tobias, N. Maaß, F. Dennerlein, and A. Maier, "Empirical scatter correction using the epipolar consistency condition", in *5th International Conference on Image Formation in X-Ray Computed Tomography*, 2018.

[154] N. Waltrich, S. Sawall, J. Maier, J. Kuntz, K. Stannigel, K. Lindenberg, and M. Kachelrieß, "Effect of detruncation on the accuracy of Monte Carlo-based scatter estimation in truncated CBCT", *Medical Physics*, volume 45, number 8, pages 3574–3590, 2018.

[155] L. A. Love and R. A. Kruger, "Scatter estimation for a digital radiographic system using convolution filtering", *Medical Physics*, volume 14, number 2, pages 178–185, 1987.

[156] L. Spies, M. Ebert, B. a. Groh, B. M. Hesse, and T. Bortfeld, "Correction of scatter in megavoltage cone-beam CT", *Physics in Medicine and Biology*, volume 46, number 3, pages 821–833, 2001.

[157] J. Rinkel, L. Gerfault, F. Estève, and J.-M. Dinten, "A new method for x-ray scatter correction: first assessment on a cone-beam CT experimental setup", *Physics in Medicine and Biology*, volume 52, number 15, pages 4633–4652, 2007.

[158] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feed-forward neural networks", *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 9, pages 249–256, 2010.

List of publications

Parts of this thesis have been published in the following journal articles and conference contributions:

Journal articles

[1] **J. Maier**, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation.", *Medical Physics*, volume 46, number 1, pages 238–249, 2019.

[2] S. Sawall, A. Hahn, **J. Maier**, J. Kuntz, and M. Kachelrieß, "Technical note: intrinsic raw data-based CT misalignment correction without redundant data", *Medical Physics*, volume 46, number 1, pages 173–179, 2019.

[3] S. Dorn, S. Chen, S. Sawall, **J. Maier**, M. Knaup, M. Uhrig, H.-P. Schlemmer, A. Maier, M. Lell, and M. Kachelrieß, "Towards context-sensitive CT imaging — organ-specific image formation for single (SECT) and dual energy computed tomography (DECT)", *Medical Physics*, volume 45, number 10, pages 4541–4557, 2018.

[4] **J. Maier**, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): Accurate real-time scatter estimation for x-ray CT using a deep convolutional neural network", *Journal of Nondestructive Evaluation*, volume 37, number 3, page 57, 2018.

[5] N. Waltrich, S. Sawall, **J. Maier**, J. Kuntz, K. Stannigel, K. Lindenberg, and M. Kachelrieß, "Effect of detruncation on the accuracy of Monte Carlo-based scatter estimation in truncated CBCT", *Medical Physics*, volume 45, number 8, pages 3574–3590, 2018.

[6] **J. Maier**, C. Leinweber, S. Sawall, H. Stoschus, F. Ballach, T. Müller, M. Hammer, R. Christoph, and M. Kachelrieß, "Simulation-based artifact correction (SBAC) for metrological computed tomography", *Measurement Science and Technology*, volume 28, number 6, page 065 011, 2017.

[7]   C. Leinweber, **J. Maier**, and M. Kachelrieß, "X-ray spectrum estimation for accurate attenuation simulation", *Medical Physics*, volume 44, number 12, pages 6183–6194, 2017.

[8]   S. Sawall, D. Franke, A. Kirchherr, J. Beckendorf, J. Kuntz, **J. Maier**, A. Kraupner, J. Backs, A. Briel, and M. Kachelrieß, "In vivo quantification of myocardial infarction in mice using micro-CT and a novel blood pool agent", *Contrast Media and Molecular Imaging*, volume 2017, 2017.

[9]   S. Sawall, **J. Maier**, C. Leinweber, C. Funck, J. Kuntz, and M. Kachelrieß, "Model-based sphere localization (MBSL) in X-ray projections", *Physics in Medicine and Biology*, volume 62, number 16, pages 6486–6496, 2017.

[10] S. Faby, **J. Maier**, S. Sawall, D. Simons, H.-P. Schlemmer, M. Lell, and M. Kachelrieß, "An efficient computational approach to model statistical correlations in photon counting x-ray detectors", *Medical Physics*, volume 43, number 7, pages 3945–3960, 2016.

[11] M. Brehm, S. Sawall, **J. Maier**, S. Sauppe, and M. Kachelrieß, "Cardiorespiratory motion-compensated micro-CT image reconstruction using an artifact model-based motion estimation", *Medical Physics*, volume 42, number 4, pages 1948–1958, 2015.

[12] L. Ritschl, R. Fahrig, M. Knaup, **J. Maier**, and M. Kachelrieß, "Robust primary modulation-based scatter estimation for cone-beam CT", *Medical Physics*, volume 42, number 1, pages 469–478, 2015.

[13] **J. Maier**, S. Sawall, and M. Kachelrieß, "Assessment of dedicated low-dose cardiac micro-CT reconstruction algorithms using the left ventricular volume of small rodents as a performance measure", *Medical Physics*, volume 41, number 5, page 051 908, 2014.

[14] F. Pfanner, **J. Maier**, T. Allmendinger, T. Flohr, and M. Kachelrieß, "Monitoring internal organ motion with continuous wave radar in CT", *Medical Physics*, volume 40, number 9, page 091 915, 2013.

Conference contributions

[1]   S. Dorn, **J. Maier**, S. Sawall, and M. Kachelrieß, "Iodine quantification accuracy in dual-source dual-energy CT using default parameters and patient-specific calibrations", in *European Congress of Radiology*, 2019.

[2]   S. Dorn, S. Sawall, **J. Maier**, C. Polster, S. Faby, M. Uhrig, H.-P. Schlemmer, and M. Kachelrieß, "Patient size-dependent ultra low dose data completion scan in a whole body photon-counting CT scanner", in *European Congress of Radiology*, 2019.

[3]   E. Eulig, **J. Maier**, M. Knaup, T. Koenig, K. Hörndler, and M. Kachelrieß, "Learned digital subtraction angiography (Deep DSA): Method and application to lower extremities", in *Proceedings of the 16th International Meeting on Fully 3D Image Reconstruction*, 2019.

[4]   **J. Maier**, E. Eulig, S. Dorn, S. Sawall, and M. Kachelrieß, "Real-time patient-specific CT dose estimation for single- and dual-source CT using a deep convolutional neural network", in *European Congress of Radiology*, 2019.

[5]   J. Kuntz, **J. Maier**, M. Kachelrieß, and S. Sawall, "Focal spot deconvolution using convolutional neural networks", in *Proceedings of the SPIE Medical Imaging Conference*, 2019.

[6]   S. Sawall, S. Dorn, **J. Maier**, S. Faby, M. Uhrig, H.-P. Schlemmer, and M. Kachelrieß, "Does iodine CNR improve when switching from today's energy integrating CT to tomorrow's photon-counting CT", in *European Congress of Radiology*, 2019.

[7]   S. Sawall, S. Dorn, **J. Maier**, S. Faby, M. Uhrig, H.-P. Schlemmer, and M. Kachelrieß, "Optimal iodine CNR in a whole body photon-counting CT scanner", in *European Congress of Radiology*, 2019.

[8]   S. Sawall, S. Dorn, L. Klein, **J. Maier**, S. Faby, M. Uhrig, H.-P. Schlemmer, and M. Kachelrieß, "Patienten-adaptierte Schwellwerteinstellungen für optimalen Jodkontrast in einem photonenzählenden Ganzkörper-CT", in *Deutscher Röntgenkongress*, 2019.

[9]   T. Vöth, **J. Maier**, and M. Kachelrieß, "Forward and cross-scatter estimation in dual source CT using a deep convolutional neural network", in *Proceedings of the SPIE Medical Imaging Conference*, 2019.

[10] Y. Berker, **J. Maier**, and M. Kachelrieß, "Deep scatter estimation in PET: Fast scatter correction using a convolutional neural network", in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2018.

[11] J. Dickmann, **J. Maier**, S. Sawall, C. Brönnimann, and M. Kachelrieß, "A count rate-dependent method for spectral distortion correction in photon counting CT", in *Proceedings of the SPIE Medical Imaging Conference*, 2018.

[12] J. Dickmann, **J. Maier**, S. Sawall, C. Brönnimann, and M. Kachelrieß, "Automated threshold selection for photon-counting CT", in *European Congress of Radiology*, 2018.

[13] S. Dorn, S. Chen, **J. Maier**, S. Sawall, M. Knaup, A. Maier, M. Lell, and M. Kachelrieß, "Prior-based multi material decomposition (PBMMD) for dual energy CT", in *5th International Conference on Image Formation in X-Ray Computed Tomography*, 2018.

[14] S. Dorn, S. Chen, F. Pisana, M. Özdemir, **J. Maier**, S. Sawall, M. Knaup, A. Maier, M. Lell, and M. Kachelrieß, "Context-sensitive organ-specific evaluation and analysis of dual-energy computed tomography (DECT)", in *European Congress of Radiology*, 2018.

[15] S. Dorn, S. Chen, S. Sawall, D. Simons, M. May, **J. Maier**, M. Knaup, H.-P. Schlemmer, A. Maier, M. Lell, and M. Kachelrieß, "Organ-specific context-sensitive CT image reconstruction and display", in *Proceedings of the SPIE Medical Imaging Conference*, 2018.

[16] E. Eulig, **J. Maier**, A. Hahn, and M. Kachelrieß, "Deep inpainting for photon-counting cone-beam CT", in *Scientific Assembly and Annual Meeting of the Radiological Society of North America*, 2018.

[17] **J. Maier**, Y. Berker, S. Sawall, and M. Kachelrieß, "Deep scatter estimation (DSE): Feasibility of using a deep convolutional neural network for real-time x-ray scatter prediction in cone-beam CT", in *Proceedings of the SPIE Medical Imaging Conference*, 2018.

[18] **J. Maier**, E. Eulig, S. Dorn, S. Sawall, and M. Kachelrieß, "Real-time patient-specific CT dose estimation using a deep convolutional neural network", in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2018.

[19] **J. Maier**, E. Eulig, S. Sawall, and M. Kachelrieß, "Deep Scatter estimation (DSE) for truncated cone-beam CT (CBCT)", in *Scientific Assembly and Annual Meeting of the Radiological Society of North America*, 2018.

[20] **J. Maier**, S. Sawall, and M. Kachelrieß, "Deep scatter estimation (DSE): Accurate real-time scatter estimation for x-ray CT using a deep convolutional neural network", in *8th Conference on Industrial Computed Tomography*, 2018.

[21] **J. Maier**, S. Sawall, and M. Kachelrieß, "Real-time x-ray scatter estimation for CT and CBCT using a deep convolutional neural network", in *European Congress of Radiology*, volume 9, 2018, B–0736.

[22] J. Kuntz, C. Funk, **J. Maier**, M. Kachelriess, and S. Sawall, "Hochauflösende CT Angiographie in der Kleintiermedizin", in *Sechstes CT-User Meeting der AG-CT der Deutschen Veterinärmedizinischen Gesellschaft (DGK-DVG)*, 2018.

[23] J. Kuntz, C. Funk, **J. Maier**, M. Kachelrieß, and S. Sawall, "Design and evaluation of a prototype high–throughput micro–CT system for in–vivo small animal imaging", in *5th International Conference on Image Formation in X-Ray Computed Tomography*, 2018.

[24] J. Kuntz, C. Funk, **J. Maier**, M. Kachelrieß, and S. Sawall, "In-vivo Coronary Micro–CT of Small Animals for Preclinical Research", in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2018.

[25] N. Waltrich, S. Sawall, **J. Maier**, J. Kuntz, K. Stanigel, K. Lindenberg, and M. Kachelrieß, "Influence of data completion on scatter artifact correction for truncated cone-beam CT data", in *Proceedings of the SPIE Medical Imaging Conference*, 2018.

[26] E. Braig, C. Broennimann, M. Duda, S. Ehn, S. Gkoumas, A. Guiller, M. Habl, J. Herzen, M. Kachelrieß, J. **J. Maier**, K. Mechlem, D. Muenzel, P. Noel, F. Pfeiffer, M. Rissi, E. Rummeny, S. Sawall, T. Sellerer, and P. Zambon, "Hybrid photon counting X-ray detectors for preclinical CT and spectral imaging", in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2017.

[27] S. Chen, J. Endres, S. Dorn, **J. Maier**, M. Lell, M. Kachelrieß, and A. Maier, "A feasibility study of automatic multi-organ segmentation using probabilistic atlas", in *Bildverarbeitung für die Medizin*, 2017.

[28] C. Funck, D. Prox, **J. Maier**, M. Kachelrieß, J. Kuntz, and S. Sawall, "Task-based optimization of in-vivo micro-CT scan protocols using energy integrating and photon counting detectors", in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2017.

[29] C. Leinweber, **J. Maier**, and M. Kachelrieß, "Accurate reconstruction of x-ray spectra in CT from simple transmission measurements", in *European Congress of Radiology*, 2017.

[30] S. Sawall, **J. Maier**, C. Leinweber, D. Prox, C. Funck, J. Kuntz, and M. Kachelrieß, "A hybrid surface-voxel approach for the reconstruction of surfaces from CT projections of non-homogeneous objects", in *Proceedings of the 14th International Meeting on Fully 3D Image Reconstruction*, 2017.

[31] N. Waltrich, S. Aktas, S. Sawall, **J. Maier**, K. Stannigel, K. Lindenberg, and M. Kachelrieß, "Importance of prior information for accurate scatter correction of truncated cone-beam CT (CBCT) data", in *Scientific Assembly and Annual Meeting of the Radiological Society of North America*, 2017.

[32] **J. Maier**, C. Leinweber, S. Sawall, H. Stoschus, and F. Ballach, "Simulation-based metal artifact reduction for computed tomography of multi-material components", in *6th Conference on Industrial Computed Tomography*, 2016.

[33] **J. Maier**, C. Leinweber, S. Sawall, H. Stoschus, F. Ballach, T. Müller, M. Hammer, R. Christoph, and M. Kachelrieß, "Simulation-based artifact correction for computed tomography in metrology", in *4th International Conference on Image Formation in X-Ray Computed Tomography*, 2016.

[34] C. Leinweber, **J. Maier**, and M. Kachelrieß, "CT X-ray spectrum reconstruction with high frequency components", in *Scientific Assembly and Annual Meeting of the Radiological Society of North America*, 2016.

[35] C. Leinweber, **J. Maier**, S. Sawall, H. Stoschus, F. Ballach, T. Müller, M. Hammer, R. Christoph, and M. Kachelrieß, "Attenuation-based reconstruction of low and high frequency components of detected X-ray spectra", in *4th International Conference on Image Formation in X-Ray Computed Tomography*, 2016.

[36] S. Sawall, A. Kraupner, D. Franke, J. Kuntz, KirchherrAnne, **J. Maier**, B. Andreas, and M. Kachelrieß, "CT imaging of myocardial infarction in mice using a novel iodine-based contrast agent", in *World Molecular Imaging Congress*, 2016.

[37] S. Faby, **J. Maier**, D. Simons, H.-P. Schlemmer, M. Lell, and M. Kachelrieß, "A photon counting detector model based on increment matrices to simulate statistically correct detector signals", in *Proceedings of the SPIE Medical Imaging Conference*, volume 9412, 2015.

[38] **J. Maier**, S. Sawall, and M. Kachelrieß, "Estimation of Bow Tie Filter Attenuation from Reconstructed CT Images", in *Scientific Assembly and Annual Meeting of the Radiological Society of North America*, 2015.

[39] S. Sawall, J. Kuntz, **J. Maier**, B. Flach, S. Schüller, and M. Kachelrieß, "CT Reconstruction of surfaces from binary objects", in *3rd International Conference on Image Formation in X-Ray Computed Tomography*, 2014.

[40] **J. Maier**, S. Sawall, and M. Kachelrieß, "Dose-dependent estimation of the left ventricular volume using cone-beam micro-CT", in *Jahrestagung des Interdisziplinären Netzwerks Molekulare Bildgebung*, 2013.

Patents

[1] **J. Maier**, R. Christoph, H. Stoschus, M. Hammer, B. Frederick, M. Kachelrieß, and C. Leinweber, "Verfahren zur Bestimmung des Spektrums einer Vorrichtung zur Erzeugung und Abgabe von Röntgenstrahlung", Deutsches Patent- und Markenamt, DE 102016113864 A1, 2017.

[2] **J. Maier**, R. Christoph, M. Kachelrieß, M. Hammer, and S. Sawall, "Verfahren zur Korrektur von zweidimensionalen Durchstrahlungsbildern, insbesondere für die dimensionelle Messung mit einer Computertomografiesensorik", Deutsches Patent- und Markenamt, DE 102015102334 A1, 2015.

Acknoledgements