Aus der Klinik und Poliklinik für Nuklearmedizin
Direktor: Herr Prof. Dr. med. Jörg Kotzerke

---

# Development of methods for time efficient scatter correction and improved attenuation correction in time-of-flight PET/MR

Dissertationsschrift

zur Erlangung des akademischen Grades

Doktor der Medizintechnologie

Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt

der Medizinischen Fakultät Carl Gustav Carus

der Technischen Universität Dresden

von

Diplom-Physiker Pavel Nikulin

aus Samara, Russische Föderation

Dresden 2019

1. Gutachter: _____

2. Gutachter: _____

Tag der mündlichen Prüfung:

gez: _____

Vorsitzender der Promotionskommission

# Contents

# Introduction

Since their discovery by Wilhelm Röntgen in 1895, the use of X-rays and, in the second half of the $20^{th}$ century, $\gamma$ radiation for imaging the interior of the human body has developed rapidly and revolutionized medical diagnosis. Especially the development of tomographic techniques (also including nuclear Magnetic Resonance Imaging, MRI) has contributed to the immense success and ever-increasing importance of medical imaging. Positron Emission Tomography (PET) is one of these techniques. Compared to Computed X-ray Tomography (CT) and MRI, PET has been adopted in clinical routine somewhat later and has now been used for about 25 years in this context.

PET allows acquiring three-dimensional images of transport and metabolic processes using short-lived positron ($\beta^+$) emitting radioactive tracers. The unique functional imaging capabilities provided by PET have proven extremely useful in a variety of neurological, cardiological, and oncological applications.

The operational principle of PET is as follows. After administration of the chosen positron emitting radiotracer, the substance distributes in the human body in a tracer-dependent way and according to the patient's physiological/metabolic state. On decay of the radioactive label, the emitted positron undergoes annihilation with an electron in the surrounding matter accompanied by emission of two 511 keV photons in opposite directions. The detection of this photon pair in coincidence, i.e. within an extremely short time interval of a few nanoseconds, defines the so-called Line-Of-Response (LOR) along which the preceding radioactive decay (the "event") has happened. Modern scanners are also capable of time-of-flight (TOF) measurements, i.e. sufficiently precise determination of differences in photon arrival time of both photons contributing to a detected event. TOF-resolved events bear information regarding approximate decay position along the LOR. Collecting this LOR and TOF information for a large number of events provides the basis for tomographic image reconstruction of the tracer concentration distribution within the patient's body. The huge amount of data which have to be processed and the complexity of the physical processes influencing the measurement process makes the reconstruction task algorithmically challenging and computationally demanding.

PET is inherently a quantitative method allowing to assess the regional tracer concentration distribution in absolute terms. This offers unique advantages for diagnostic functional imaging and also the possibility to apply pharmacokinetic approaches (tracer kinetic modeling see, e.g., (van den Hoff, 2005)) provided sufficient quantitative accuracy

of the PET images is guaranteed. However, this is not easy to achieve. The event detection process is influenced by multiple factors (scanner geometry, detector properties, finite positron range prior to annihilation, photon absorption and scatter etc.) which all affect the ultimate detection probability and correct interpretation of the detected events and, consequently, the achievable spatial resolution and local quantitative accuracy of the reconstructed images. It is necessary to adequately correct for all relevant sources of error in order to really utilize the full potential of the PET method.

Among these corrections, attenuation correction is one of the most important ones. The absence of attenuation correction leads to gross errors in reconstructed activity concentrations, making quantitative imaging impossible. Attenuation correction requires determination of the fractional attenuation of 511 keV photon pairs along each LOR. This is equivalent to knowing the spatial distribution of the linear attenuation coefficient, $\mu$, within the imaged object ("$\mu$-map"). Naturally, the $\mu$-map is patient dependent and has, therefore, to be obtained on an individual basis. In the case of standalone PET or combined PET/CT systems, a dedicated attenuation measurement can be performed. For standalone PET machines (no longer produced nowadays), radioactive sources emitting 511 keV (or, rarely, 662 keV) photons were used for transmission measurements. In the case of PET/CT machines, the CT component can be utilized instead to perform the PET attenuation correction.

During the last decade combined PET/MR has been introduced as a further option. One of the first scanners of this kind — Philips Ingenuity TF PET/MR (Philips Healthcare, Best, The Netherlands) — has been in operation at the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) since 2011 and is now installed in the Center for Radiation Research in Oncology (Oncoray) in Dresden, jointly operated by HZDR and the University Hospital of the Technical University Dresden. Improving the quantitative accuracy and quality of the image reconstruction for this system is the principal aim of the present thesis.

Due to the lack of either a CT component or radioactive transmission sources, PET/MR systems no longer allow direct measurement of the photon attenuation. This poses completely new challenges to quantitative PET. Furthermore, the Ingenuity PET/MR is capable of TOF measurements. On the one hand, TOF imaging provides benefits in terms of reduced image noise and improved convergence rate (Conti et al., 2005; Vunckx et al., 2010) compared to non-TOF imaging. On the other hand, the TOF-aware scatter correction, which is required for quantitative PET, is way more complex and computationally demanding than the non-TOF one as will be explained below.

The usual attenuation correction strategy in PET/MR is to use MR-based tissue-type identification to choose a suitable value for the attenuation coefficient of the respective tissue. MR-based attenuation correction methods include segmentation-based (Martinez-Möller et al., 2009; Hu et al., 2009; Schulz et al., 2011), atlas-based (Rota Kops & Herzog, 2007), and CNN-based (Convolutional Neural Networks) (Han, 2017) variants. The last years have seen clear progress regarding generation of accurate head attenuation maps, especially when utilizing dedicated bone-detection MR sequences (Sekine et al., 2016;

Mérida et al., 2017; Rota Kops et al., 2017; Sousa et al., 2018; Wiesinger et al., 2018; Gong et al., 2018). The situation is more unfavorable regarding whole-body attenuation correction, however. In fact, only three- (air, lungs and soft tissue) and four-tissues-classes (air, lungs, soft and adipose tissue) segmentation methods are currently available in clinical PET/MR systems. These approaches do not account for individual variations in lung attenuation coefficients and completely ignore the presence of bones which can cause up to 30% errors in quantification of lung and bone lesions activity, respectively (Keereman et al., 2011; Samarin et al., 2012; Akbarzadeh et al., 2013). Moreover, any kind of MR-based attenuation correction technique is prone to errors in the presence of metal implants in the patient's body since these cause regional signal voids (so-called susceptibility artifacts (Keller et al., 2013)). Even though the adverse effects of metal-induced artifacts can be reduced or corrected (Bezrukov et al., 2013; Ladefoged et al., 2013; Schramm et al., 2014), all MR-based methods remain susceptible to mis-segmentation and systematic errors in the resulting attenuation maps.

Consequently, there is demand for alternative approaches and, indeed, a different strategy to solve the problem of attenuation correction in PET/MR has emerged. It relies on inherent redundancy of the emission data itself which allows, in theory, to derive attenuation map and activity image simultaneously in a joint reconstruction. This algorithm, the Maximum Likelihood Activity and Attenuation estimation (MLAA) method, was first proposed in (Nuyts et al., 1999). It combines two well-established reconstruction algorithms for emission tomography — Maximum Likelihood Expectation Maximization (MLEM) (Shepp & Vardi, 1982) — and transmission tomography — Maximum Likelihood for Transmission tomography (MLTR) (Nuyts et al., 1998).

Originally intended as a method for attenuation correction for standalone PET scanners, it did not find large adoption at the time due to considerable cross-talk between attenuation and activity. However, the emergence of PET/MR as well as developments in PET scanner hardware allowing for TOF measurements caused renewed interest in MLAA. For one, TOF image reconstruction is less sensitive to errors in the attenuation map (Turkington & Wilson, 2009; Conti, 2011). More important in the present context, utilizing the TOF information for activity reconstruction within MLAA reduces localized cross-talk between reconstructed attenuation and activity images and in principle allows correct determination of attenuation coefficients up to a global scaling constant (Salomon et al., 2011; Defrise et al., 2012; Rezaei et al., 2012). On the other hand, it is widely believed that a TOF-aware implementation of the MLTR algorithm would not provide any practical benefits, see e.g. (Rezaei et al., 2016), but no thorough investigations of this topic were published so far.

In the last few years, numerous implementations of TOF-enhanced MLAA and MLAA-like algorithms have appeared, e.g. (Ahn et al., 2012; Rezaei et al., 2014; Mehranian & Zaidi, 2014; Zhu et al., 2016; Salvo & Defrise, 2017; Hwang et al., 2018) as well as non-TOF MLAA variations constrained with MR-based priors (Heußer et al., 2016; Benoit et al., 2016). Also, emission-based metal artifact reduction was proposed in (Fuin et al.,

2017). Joint reconstruction methods have already demonstrated distinct improvements over conventional MR image segmentation-based attenuation correction in whole-body applications (Boellaard et al., 2014; Mehranian & Zaidi, 2015; Ahn et al., 2018; Rezaei et al., 2018). On the other hand, compared to accepted reference — MLEM with CT attenuation correction — tracer uptake errors in focal lesions can still exceed 10% with joint reconstruction methods. A comprehensive review of this topic can be found in (Berker & Li, 2016).

TOF activity image reconstruction, which is crucial for a viable MLAA implementation, has its own challenges. Specifically, the non-uniform TOF distribution of scattered events has to be accounted for. Therefore, conventional scatter correction (SC) methods, notably the Single Scatter Simulation (SSS) approach (Watson, 2000), cannot be used without modification. The accepted way to estimate the scatter time distribution is to include TOF modeling directly into the scatter simulation process which leads to the TOF extension of the SSS algorithm (Werner et al., 2006; Watson, 2007). However, the increased complexity of the TOF-SSS algorithm causes an increase in computation time by an implementation-dependent factor of about 3–7. This can result in significant image reconstruction slow-down for certain practically relevant choices of iteration scheme and reconstruction parameters.

Alternative TOF-SC approaches exist which allow to avoid this substantial computational overhead and to perform TOF scatter estimation only slightly slower than non-TOF algorithms. The common key idea is to use a non-TOF scatter correction algorithm — in particular SSS — to estimate the spatial distribution of scattered events in each LOR and to model scatter time distribution in a separate step that utilizes simplifying assumptions. The first two approaches of this kind were introduced in (Conti et al., 2005) and were intended as a temporary solution since full TOF-SSS was still under development at the time and not yet available. In the so-called simple scaling approach, non-TOF scatter sinograms are produced first and then rescaled to fit the emission data registered outside of the imaged object's boundary individually for each TOF bin. As reported in (Conti et al., 2005), this leads to significant overestimates of scatter in the center of large objects. The second proposed approach (radial distortion and scaling) performs better in this regard but ignores dependency of the scatter TOF profiles on the given activity distribution which is especially problematic under high contrast conditions (e.g. between bladder and surrounding tissue).

A further accelerated TOF-SC method was presented in (Jin et al., 2013) as part of the MOLAR image reconstruction for the Siemens Biograph mCT (Siemens Medical Solutions USA, Inc.). The key idea of this method is based on the assumption that both scatter and true events have similar time distribution which, therefore, can be estimated from the measured data alone. The specifics of the data compression performed by the mCT scanner and a relatively small number of TOF bins (13) allowed to avoid excessive noise in the estimated scatter time distribution. However, this approach is not suitable for the reconstruction of uncompressed list-mode data with a larger number of TOF bins

employed. Moreover, the initial assumption of the method will not hold for LORs in the vicinity of large structures like the bladder exhibiting high activity concentrations. In this case, the scatter TOF distribution will be significantly affected by neighboring activity while the trues TOF distribution is expected to remain relatively flat. Overall, to this day, a satisfactory alternative to TOF-SSS is still missing.

Apart from a working quantitative TOF reconstruction, further requirements have to be fulfilled in order to realize a valid MLAA implementation. Recently, it has been found that MLAA is extremely sensitive to inaccuracies in scatter correction as well as to errors in time offsets and time resolution calibration (Cheng et al., 2016a; Zhu et al., 2017; Nuyts et al., 2018). In the light of this fact, it is relevant that the time resolution of a TOF PET system is to different extent count-rate-dependent (Surti et al., 2007). However, count-rate-dependent time resolution calibration is usually not provided by the vendors.

The present work is concerned with contributions to improved quantitative PET/MR image reconstruction, specifically targeting the Philips Ingenuity PET/MR system operated by HZDR and addressing the demand for a viable MLAA implementation and accelerated TOF scatter correction described previously. It builds on previous developments performed in our group. These include the introduction of improved MR-based attenuation correction (Schramm et al., 2014) and the development of a Tube of response High resolution OSEM Reconstruction (THOR) (Lougovski et al., 2014, 2015). THOR has already demonstrated superior image resolution compared to the standard vendor reconstruction software due to better system modelling and utilization of list-mode data without any compression. Prior to the developments described in this work, THOR did not utilize TOF information at all and only supported MR-based attenuation correction via externally generated $\mu$-maps. Starting from here, the following developments were performed in the present work:

1. THOR was extended to handle TOF data with variable count-rate-dependent time resolution. TOF-SSS was integrated into THOR.

2. A novel Maximum Likelihood Time Resolution Estimation (MLRES) algorithm for time resolution calibration was proposed and evaluated. The algorithm utilizes the joint reconstruction principle to maximize the likelihood by varying activity image and time resolution in an alternating manner. It is assumed that the likelihood reaches its maximum at the true time resolution of the scanner. This approach is related to the one in (Vandenberghe et al., 2007) but MLRES utilizes full TOF image reconstruction and accounts for TOF scatter correction. The latter is required for the algorithm utilization in the real-world applications.

3. A new time efficient TOF-SC method, the Immediate Scatter Approximation (ISA) was developed. It shares the principal idea of separate estimation of scatter spatial distribution (via SSS) and scatter time distribution via a dedicated fast algorithm

with existing approaches. One key difference is that the proposed scatter time distribution algorithm works reliably by design for all practically relevant event numbers and subdivisions into TOF bins. ISA also addresses the activity distribution dependence of the scatter TOF profiles. The proposed approach was evaluated in dedicated phantom measurements providing challenging high activity contrast conditions and compared against several other reconstruction schemes including TOF-SSS. A further evaluation was performed in representative clinical patient data sets.

4. A list-mode MLAA algorithm including corrections for scatter (either TOF-SSS or ISA) and random events was developed, implemented, and evaluated in phantom and patient data. So far, only very few such implementations have been reported (Mollet & Vandenberghe, 2014; Rezaei et al., 2015; Cheng et al., 2016b). Only one of them (Cheng et al., 2016b) includes corrections for scatter and random events and none of them where evaluated in clinical patient data. Both, TOF and non-TOF list-mode MLTR are supported by our implementation which allowed us to perform the comparison of these two variants.

The thesis text is organized as follows. Chapter 1 gives an overview over PET physics, data acquisition, and image reconstruction. Chapter 2 describes the hard- and software as well as algorithms (including the newly developed ones) used in this work. The chapter also comprises a description of the applied data evaluation and validation procedures. Chapter 3 contains the detailed report of the obtained results which then are discussed in Chapter 4.

# Chapter 1

# Fundamentals

## 1.1 Positron Emission Tomography

Positron Emission Tomography (PET) is a nuclear medicine imaging technique which allows to non-invasively image functional processes in the patient body. Like Single-Photon Emission Computed Tomography (SPECT) , PET relies on detecting gamma quanta emitted from within the body by dedicated radioactive *tracers*. A tracer is a substance with specially tailored properties which is administered to the patient in sufficiently small quantities (in order not to disturb the investigated system/process) and allows to assess the physiological or metabolical process in question. Usage of radioactive tracers for medical imaging became possible after the discovery of radioactive labeling of chemical compounds by George de Hevesy in the early $20^{th}$ century for which he was awarded the Nobel Prize in Chemistry in 1943. He demonstrated that molecules with one of the stable atoms replaced by a radioactive isotope have – prior to radioactive decay – the same biological properties as their stable counterparts while being detectable at the level of single molecules due to the emitted radiation upon decay. Since then a variety of dedicated tracers for visualizing of different processes in the body have become available, making PET and SPECT versatile tools with a wide spectrum of applications, notably in oncology, neurology, and cardiology. As the names suggest, PET is restricted to the use of positron ($\beta^+$) emitting compounds which subsequently yield pairs of annihilation photons that are detected by the scanner while SPECT utilizes $\beta^-$ or "pure" $\gamma$ emitters yielding one or more photons during the nuclear decay process. Therefore, PET and SPECT require different radionuclides which in turn leads to tracers with different chemical properties.

### 1.1.1 PET tracers

Positron emitters suitable for application in PET need to have suitable radioactive half-lives and chemical properties. Also, their production in sufficiently high amounts must be practicable. Still, there is a variety of potentially interesting isotopes (Conti & Eriksson, 2016):

- Standard isotopes (with half-lives between one minute and a few hours): $^{11}$C, $^{13}$N, $^{15}$O, $^{18}$F, $^{68}$Ga, $^{82}$Rb;
- Non-standard isotopes (with half-lives of more than twelve hours): $^{64}$Cu, $^{76}$Br, $^{86}$Y, $^{89}$Zr, $^{124}$I.

Today, the radioisotopes of the first group are routinely used in PET while interest in the second group is expected to grow in light of recent developments (Schlyer, 2004; Vallabhajosula et al., 2011).

The short half-life of most PET isotopes poses a major challenge in the production and delivery of radiopharmaceuticals. For example, $^{15}$O and $^{82}$Rb with half-lives of 1 to 2 min have to be administered to the patient nearly immediately after production. Consequently, with the exception of generator-based nuclei (currently only $^{68}$Ga and $^{82}$Rb), a PET facility usually needs to be located in the close vicinity of the isotope production site (usually a cyclotron). Currently, only $^{18}$F-based tracers (half-life 110 min) are partly transported over longer distances. Proper logistics as well as optimization of the radiopharmaceutical production process are, therefore, important aspects regarding feasibility of clinical PET.

Despite the large number of known PET tracers with potential clinical applications, about 90% of all PET investigations worldwide are carried out with a single substance: $^{18}$F-FDG. Due to its high clinical relevance, especially in whole-body applications, $^{18}$F-FDG will be the mainly considered tracer in this work.

### 1.1.2 $^{18}$F-FDG

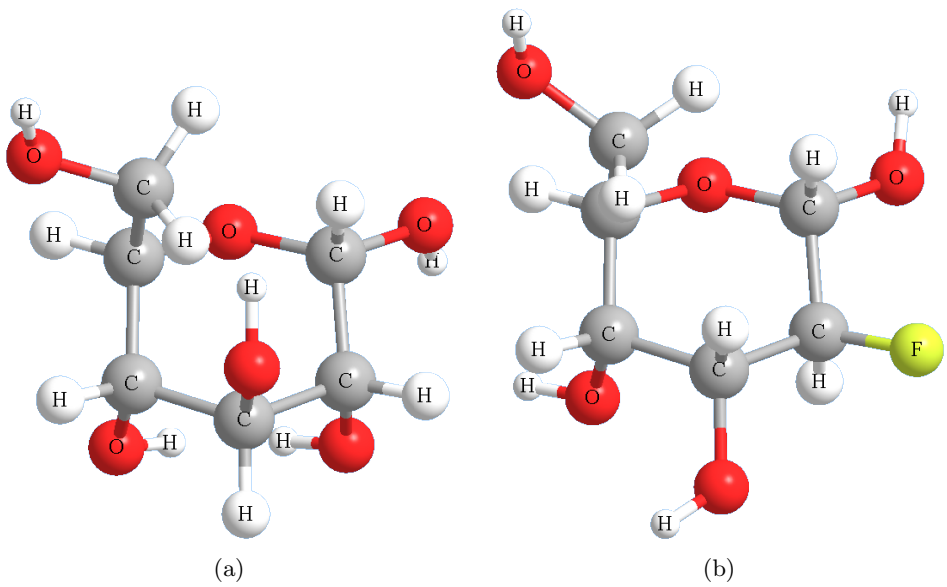


Figure 1.1: *Glucose molecule (a) and FDG molecule (b). (Taken from (Lougovski, 2012))*

2-deoxy-2-($^{18}$F)fluoro-D-glucose ($^{18}$F-FDG) is a glucose molecule analog with the hydroxyl group in C-2 position being replaced with a radioactive fluorine isotope $^{18}$F, see Fig. 1.1.

It has been shown that FDG as well as 2-deoxyglucose, another glucose analog, follows the same metabolical path as normal glucose until a certain point in the metabolic chain (Sols & Crane, 1954; Tewson & Krohn, 1998). It means that human cells consume FDG with the same rate as they would consume regular glucose. Moreover, the glycolysis process is being initiated on FDG molecules with hexokinase phosphorylating them and forming FDG-6-phosphate as a result. Phosphorylation prevents FDG from diffusing back the cellular membrane while the lack of hydroxyl group in C-2 position inhibits the further progression of glycolysis effectively trapping FDG molecules inside of the cell and leading to their accumulation. The accumulation rate is proportional to the tissue glucose consumption rate and tracer blood concentration. Since the administered amount of the radiopharmaceutical is extremely low compared to systemic glucose content the actual metabolic activity of the organism remains unaltered. Elevated FDG uptake is exhibited by the energy-avid brain and heart as well as tumor or inflamed tissues. High FDG concentrations are also observed in the urinary system due to excretion of surplus FDG. Radioactive decay of $^{18}$F in $^{18}$F-FDG molecules transforms fluorine into the stable oxygen isotope $^{18}O^{-}$ which after a combination with $H^{+}$ from surrounding water forms a hydroxyl group. As a result, FDG-6-phosphate becomes a regular glucose-6-phosphate which can be metabolized further.

Highly elevated glucose consumption exhibited by tumor cells makes FDG-PET an effective tool for cancer diagnosis, staging, and restaging especially for whole-body applications when metastases are in question (Kelloff et al., 2005). Besides oncology, $^{18}$F-FDG finds its use in neurology and cardiology as well (Gambhir et al., 2001), making it the most popular PET tracer to date.

## 1.2 Physics in PET

The data acquisition procedure in PET is influenced by different physical processes including radionuclide decay, positron propagation and annihilation, radiation transport, $\gamma$-quanta detection, and event registration. The following chapter provides a short overview of the most important physical processes to be considered in this context.

### 1.2.1 Positron emission and annihilation

As mentioned in Section 1.1.1, the observation of the tracer molecule distribution by a PET scanner becomes possible due to their labeling with radioactive isotopes. Radioactive decay is a stochastic process characterized by the constant decay probability of a single nucleus per unit time as well as independence of decays from each other. It means that the number of nuclear decays $dN$ happening over the infinitely short time interval $dt$ in an ensemble of $N$ radioactive particles is proportional to $N$ and $dt$:

$$dN \propto N\,dt. \tag{1.1}$$

Therefore, the number of radioactive nuclei decreases over time according to the radioactive decay law

$$N(t) = N_0\, e^{-\Lambda t}, \tag{1.2a}$$

$$\Lambda = \frac{\log 2}{T_{1/2}}, \tag{1.2b}$$

where $t$ is the time, $N_0$ is the number of radioactive particles in the system at $t = 0$, $\Lambda$ is the so-called decay constant, and $T_{1/2}$ is the half-life of the radioisotope. Half-life is the time in which half of the radioactive particles in a system decay. It is not to be confused with the *mean lifetime* of the radioactive particle which is equal to reciprocal of $\Lambda$. The number of observed radioactive decays per unit time usually closely follows the Poisson statistics.

Radioactive decay is a common name for a whole class of different nuclear transformation processes. Of concern in the context of the present thesis are

- **$\beta^+$-decay:** the process of emitting a positron ($\beta^+$-particle) and an electron neutrino. It results in formation of an isobar of the initial nuclide with atomic number $Z$ decreased by 1

  $$^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \rightarrow {}^{\mathrm{A}}_{\mathrm{Z-1}}\mathrm{Y} + \mathrm{e}^+ + \nu. \tag{1.3}$$

  The emitted positron subsequently annihilates with an electron in the surrounding medium accompanied by the creation of a pair of 511 keV photons which is the radiation utilized for PET imaging.

- **Electron capture (EC):** the process of capturing an electron from an inner shell of the atom by an atomic nucleus and emitting an electron neutrino. Like $\beta^+$-decay, it results in formation of an isobar of the initial nuclide with atomic number $Z$ decreased by 1

  $$^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} + \mathrm{e}^- \rightarrow {}^{\mathrm{A}}_{\mathrm{Z-1}}\mathrm{Y} + \nu. \tag{1.4}$$

  This process belongs to the $\beta$-decay class of processes being a competitive decay channel for $\beta^+$-decay. The branching ratio between these two depends on the penetration of the electron wave function into the nucleus. For light atoms the penetration is small and, therefore, $\beta^+$-decay is the dominating decay channel in this case. For heavier elements, inner electrons come closer to the nucleus which increases the probability of the process.

Competition between electron capture and $\beta^+$-decay reduces the positron yield. For example, $\beta^+$-decay branching ratio is 99.8% for $^{11}$C, 96.9% for $^{18}$F, and only 88.9% for $^{68}$Ga (Conti & Eriksson, 2016). This effect has to be accounted for during scanner calibration and when quantifying activity concentrations in the reconstructed images.

The positrons created by the $\beta^+$-decay mostly possess substantial (non-thermal) kinetic energy. This excess energy is dissipated by numerous interactions with surrounding

matter until thermal equilibrium is reached. "In flight" annihilation prior to this point has a probability as low as 2% (Harpen, 2004). The distance from the decay location to the point where thermalization is achieved is referred to as *positron range*. The average and maximum positron range depend on energy released in the decay process and on the medium in which propagation takes place. Therefore, it varies among different radioisotopes. For example, mean positron range for $^{18}$F in water is 0.6 mm while for $^{68}$Ga it is 3.5 mm (Conti & Eriksson, 2016). The finite positron range is one of the factors contributing to the principal limit of obtainable spatial resolution in PET.

Annihilation can take in two ways. The positron can either interact directly with an electron of the medium and annihilate immediately into two photons or first form a short-lived bound state with the electron called positronium (Ps). The positronium yield in organic tissues has not been measured directly. Experiments using water as medium suggest a lower bound of 37% (Castellaz et al., 2002; Harpen, 2004). Depending on the relative orientation of the electron and positron spins, Ps can be formed as orthopositronium (o-Ps, parallel spins) or as parapositronium (p-Ps, antiparallel spins). The yield ratio between these two is 3:1, respectively, as has also been demonstrated experimentally (Castellaz et al., 2002). p-Ps annihilates after a mean lifetime of 125 ps into 2 $\gamma$-quanta. This $2\gamma$ decay is forbidden by parity conservation for o-Ps which, therefore, only can decay into 3 $\gamma$-quanta which is much less likely. Consequently, the mean lifetime of o-Ps in vacuum is much larger than that of p-Ps, namely 142 ns (Ley & Werth, 2001). However, the measured o-Ps lifetime in water is considerably shorter (1.8 ns). This discrepancy is explained by the pick-off $2\gamma$ annihilation of the bound positron in Ps with an electron of the surrounding medium. Overall, the $3\gamma$ process, therefore, occurs only in a negligible fraction of the annihilations.

Assuming a center-of-mass reference frame of the system of annihilating positron and electron and accounting for the energy and momentum conservation laws, one can conclude that two $\gamma$-quanta formed in the annihilation process both have an energy of $m_e c^2 = 511$ keV, where $m_e$ is a rest mass of the electron and $c$ is the speed of light and their momentum vectors are antiparallel. Thus, in the center-of-mass system, annihilation photons are propagating in opposite directions along the straight line crossing the annihilation point. However, in the laboratory frame of reference, the sum of positron and electron momenta is usually different from zero even after positron thermalization. Therefore, the photons propagation paths deviate from the aforementioned straight line by a small angle (usually $< 0.5°$). The effect is usually referred to as *photon accolinearity* and is a further factor limiting the principally achievable spatial resolution in PET.

### 1.2.2 Interaction of radiation with matter

Photons can interact with matter via multiple electromagnetic processes (see (Hubbell, 1999) for an overview). The relative probability of undergoing certain interactions depends on the photon energy. In the present context, only the following processes are of relevance:

- **Photoelectric effect.** This denotes the absorption of a photon by an atomic electron and emission of the electron from the atom. This might be written as $\gamma + e^- \rightarrow e^-$. The process plays a significant role in the interaction of low-energy X-ray radiation with matter but is of no real importance for the 511 keV photons used in PET.

- **Compton scattering.** This denotes inelastic photon scattering at quasi-free atomic electrons which might be written as $\gamma + e^- \rightarrow \gamma' + e^-$. This is essentially the exclusive process leading to attenuation and scatter in PET (Hubbell, 1969). The total Compton cross section (Klein & Nishina, 1929) is given by

$$\sigma_c(E) = 2\pi r_e^2 \left\{ \frac{1+\alpha(E)}{\alpha^2(E)} \left[ \frac{2+2\alpha(E)}{1+2\alpha(E)} - \frac{1}{\alpha(E)} \log(1+2\alpha(E)) \right] + \frac{1}{2\alpha(E)} \log(1+2\alpha(E)) - \frac{1+3\alpha(E)}{(1+2\alpha(E))^2} \right\}, \quad (1.5)$$

where $E$ is the energy of the incident photon, $\alpha(E) = E/m_e c^2$, and $r_e = e^2/m_e c^2$ is the so-called classical electron radius ($e$ and $m_e$ are electron charge and mass). The dependence of the scattered photon energy on the scatter angle $\theta$ can be written in the form

$$E' = \frac{E}{1+\alpha(E)\,(1-\cos\theta)}. \quad (1.6)$$

### 1.2.3 Photons detection

There are three main classes of detectors which can be used for detection of high-energy $\gamma$-quanta: proportional gas chambers, semiconductor detectors, and scintillation detectors using different types of scintillators. Only the latter are used in current PET systems. Scintillators absorb high energy photons and emit visible light. All scintillators used in PET are inorganic crystals since these have the highest stopping power for 511 keV radiation due to their high density.

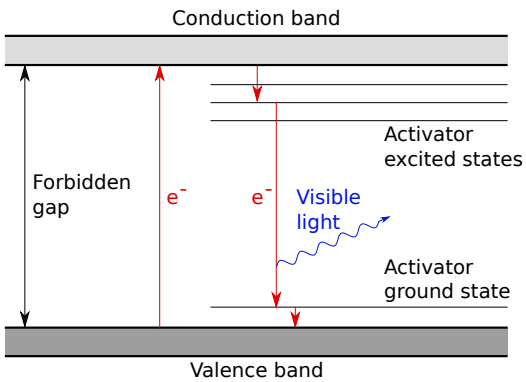


Figure 1.2: *Energy states and electronic transitions in a scintillation crystal.*

Scintillation crystals exhibit a specific electronic structure, see Fig. 1.2. For one, there is a set of energy bands. The lower band is usually referred to as *valence band* and is completely filled. The next available band is called *conduction band* since electrons there are delocalized and can contribute to macroscopic electric current. Both bands are separated by a *forbidden gap* usually containing no allowed energy states. The energy dissipated by an incident photon can translate into excitation of valence electrons and transfer them into the conduction band. The de-excitation process is accompanied by

emission of photons, usually with wavelengths in the UV range. Since UV photons cannot be efficiently detected, the electronic structure of the crystal is modified deliberately by adding certain impurities to the crystal called *activators* in order to increase the wavelength of the scintillation photons. The presence of activators adds further energy levels in the forbidden gap, allowing electrons to de-excite in multiple steps in a process called *luminescence*. Since the difference between activator states is lower than the width of the forbidden gap, the luminescence photons are emitted in the visible spectrum which is desirable for the registration process. The emitted light has an isotropic distribution and its amount is proportional to the energy deposited by the incident $\gamma$-quantum.

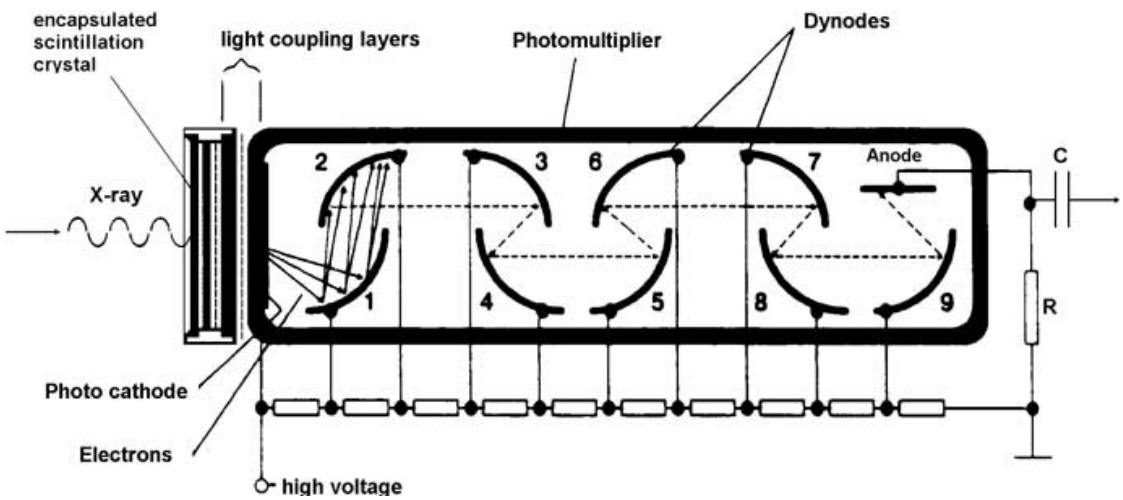


Figure 1.3: *Scintillation crystal attached to a photomultiplier tube. (Taken from Bruker AXS GmbH, Karlruhe, Germany)*

The second component of a scintillation detector converts the scintillation light into an electric signal. The devices used for this purpose need to have a high detection efficiency for the scintillation light and produce sufficiently strong electric signals proportional to the incident amount of scintillation light (and, thus, also proportional to the deposited energy by the incident $\gamma$-quantum). Today, the following options are available:

- **Photomultiplier tubes (PMTs).** PMTs (Fig. 1.3) are a special sort of high voltage vacuum tubes that have been developed more than 80 years ago. They are still in wide use in a large variety of applications. PMTs contain a thin photocathode and a series of amplification stages (dynodes). Scintillation light hitting the photocathode causes emission of electrons via the photoelectric effect. Ejected electrons are being accelerated in the electrostatic field towards the first dynode causing emission of an increased number of secondary electrons on impact. The process repeats multiple times on the subsequent dynodes which results in an avalanche of electrons reaching the anode and causing a strong electric signal. The amplification factor of PMTs can exceed $10^6$ (Lecomte, 2009). The main disadvantages of PMTs are their big size and extreme susceptibility to magnetic fields which can drastically compromise

performance if not thoroughly shielded.

- **Avalanche photodiodes (APDs).** APDs are semiconductor photodetectors with an internal high-voltage signal amplification zone. The electric signal in APDs appears as a result of the interaction of visible or near-UV light with electrons in a semiconductor causing production of electron-hole pairs. Released electrons and holes, driven by applied voltage, participate in electric current conduction. Upon reaching the narrow high-voltage zone, free electrons induce an avalanche of secondary electrons by impact ionization leading to the signal amplification. Depending on the applied voltage, APDs can work in a proportional mode where the signal current is proportional to the number of absorbed photons and in a Geiger mode where any incoming photon can cause an avalanche breakdown generating binary (yes/no) signals. The main problem with APD detectors is the high amount of noise caused by spontaneous electron/hole pair generation. In order to keep noise within acceptable limits, APDs usually are operated at relatively low voltage and, consequently, low amplification of 50-150 leading to vastly inferior signal-to-noise ratio compared to PMTs (Lecomte, 2009). Nevertheless, APDs have several advantages over PMTs such as better photon detection efficiency and basically complete resistance to strong magnetic fields.

- **Silicon photomultipliers (SiPMs).** SiMPs represent arrays of thousands of small ($20 \times 20$ to $100 \times 100$ $\mu$m), tightly packed avalanche photodiodes operating in Geiger mode. Scintillation light hitting the detector cells causes highly amplified breakdown discharge signals. The signals from all cells are summed up in order to form an integral signal of the whole detector element. Operation in the Geiger mode solves the main problem of APDs — low gain (it can reach $10^6$ with SiPMs) — while the pixelization approach allows to reduce noise (Lecomte, 2009). Like APDs, SiPMs are insensitive to magnetic fields and thus especially suitable for combined PET/MR systems. A weak point of SiMPs is their susceptibility to small changes in operating temperature.

Until very recently, PMTs were the only option in positron emission tomography. In fact, most of the currently installed PET systems are still using PMTs. However, PET systems utilizing APDs and, especially, SiPMs have started to emerge throughout the last years. Currently, there are five different systems available in the market (Philips Vereos PET/CT, Siemens Biograph mMR, Siemens Vision PET/CT, GE Discovery MI PET/CT, GE Signa PET/MRI (Delso et al., 2011; Rausch et al., 2018)) and this trend can be expected to continue.

### 1.2.4 Events registration

Detection of a $\beta^+$-decay is based on detection of pairs of annihilation photons created subsequent to the decay. These photons have to be registered in coincidence, i.e. within

a short time interval $T$. The time interval $2T$ is conventionally called *coincidence time window*. The width of the coincidence time window is a parameter which depends on the scanner field-of-view (FOV) size and, to some extent, the precision of measuring the arrival time difference of the photons in the pair (coincidence time resolution). The typical value here is 6–8 ns for a clinical whole-body scanner (for comparison: photons travel 30 cm in 1 ns).

Suppose that the first photon hits detector $A$ and the second photon hits detector $B$ as shown in Fig. 1.4 (top left). A detector pair in coincidence forms a single *detector element*. Taking into account that annihilation photons travel in opposite directions, we can assume that the radioactive decay happened somewhere on the straight line connecting crystals $A$ and $B$ which is usually referred to as Line-Of-Response (LOR). This description is not completely accurate, though. First of all, finite positron range (Section 1.2.1) and accolinearity have to be considered. Intercrystal $\gamma$-quanta scattering (Section 1.2.2) as well as finite detector size (Section 1.2.3), too, compromise accuracy of the event localization. For this reason, the LOR model should rather be considered as idealization of a finite-size Tube-Of-Response (TOR).

Apart from true coincidences, there are further types of coincidences ("events") to be considered, see Fig. 1.4:

- **Scattered events.** One (or both) of the two annihilation photons can undergo (Compton) scatter and change propagation direction and energy. As a consequence, the scattered photons can hit a detector at a position different from the original destination point resulting in misregistration of the event in a different LOR not crossing the emission point. Part of the scattered events can be identified (and discarded) by the scanner hardware with the help of energy discrimination. The remaining, undiscriminated scatter events still account for a substantial fraction of all events and represent a major problem for image reconstruction.

- **Random events.** Two annihilation photons originating from two different decays can accidentally be detected in coincidence while the remaining photons from the two decays escape detection. This causes a spurious event on the respective LOR. If more than two photons are accidentally detected in coincidence (so-called "multiple coincidences"), the scanner hardware is able to identify these events as spurious and can discard them.

Without correction, scattered and random events can severely compromise the image reconstruction process and degrade the resulting image quality. Since scattered and random events together can exceed 50% of the measured coincidences, it is mandatory to handle them properly. Fortunately, there are several techniques available allowing to reduce the amount of registered randoms and scatter at the hardware level as well as during image reconstruction, see Section 1.3.3.

Most of the current PET systems are capable of not only measuring two photon coincidences for the purpose of LOR determination but also can provide information regarding

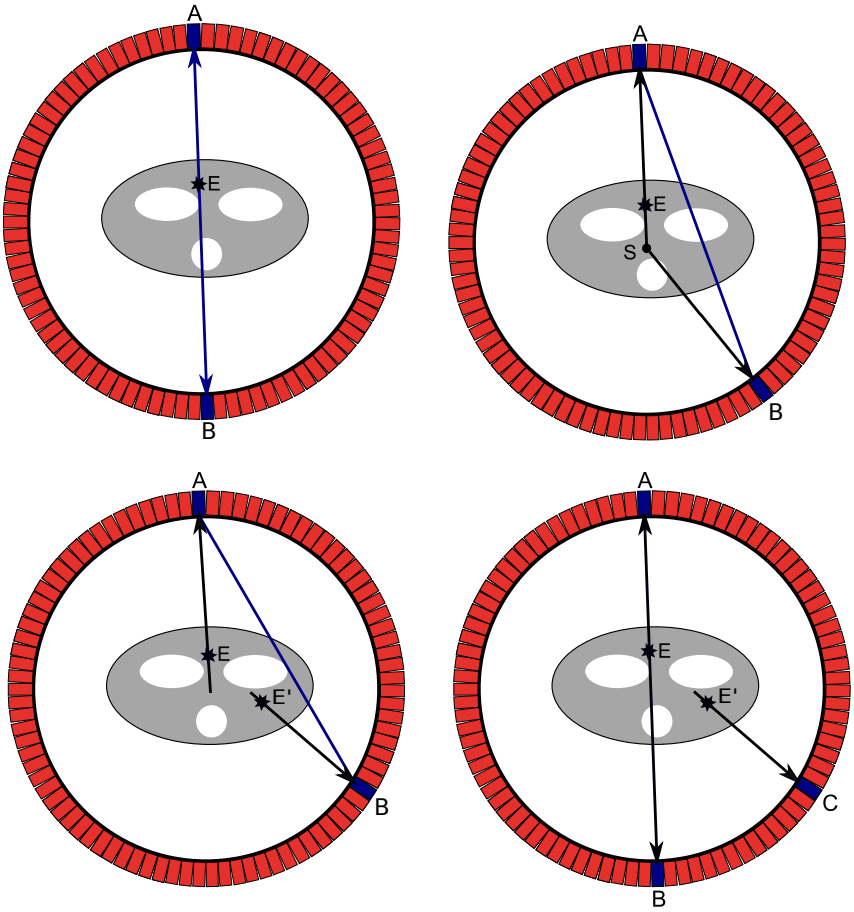


Figure 1.4: *Illustration of different types of events possible in PET: true event (top left), scattered event (top right), random event (bottom left), multiple event (bottom right). $E$ and $E'$ indicate emission points, $A$, $B$, and $C$ are triggered detectors, arrows show the photon paths and the blue line is an LOR corresponding to an event. (Modified after (Lougovski, 2012))*

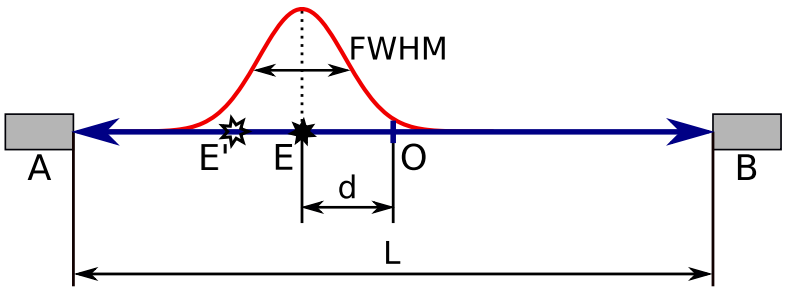


Figure 1.5: *Time-of-flight acquisition principle. $A$ and $B$ are detectors, $O$ is the center of the LOR, $E$ is the true decay location, and $E'$ is the measured decay location. $L$ designates the total length of the LOR. $d$ is the distance between $E$ and the LOR center. Detection probability at position $E'$ is distributed according to the red curve and characterized by the FWHM of this TOF measurement.*

the arrival time difference of both photons at the respective detectors. This capability is usually referred to as Time-Of-Flight (TOF) PET. Consider a pair of photons emitted in opposite directions at point $E$ and detected in LOR $AB$ (Fig. 1.5). Then the arrival time

difference $\delta$ can be written in the form

$$\delta = t_B - t_A = \frac{L/2 + d}{c} - \frac{L/2 - d}{c} = \frac{2d}{c}, \tag{1.7}$$

where $t_A$ and $t_B$ are the arrival times of the photons at detectors $A$ and $B$, respectively, $L$ is the LOR length, $d$ is the (signed) distance between point $E$ and the LOR center, and $c$ is the speed of light. Equation (1.7) demonstrates that exact knowledge of $\delta$ would yield the actual position $d$ of the decay event. However, the actually available time resolution is limited and characterized by the Full Width at Half Maximum (FWHM) of the Gaussian probability distribution of the TOF measurement for a point source. In order to be useful, the time resolution needs to be good enough to allow decay localization at least comparable or preferably distinctly better than the diameter of the imaged object. For example, the first clinical TOF-capable scanner Philips Gemini PET/CT has a mean time resolution of approximately 650 ps which corresponds to event localization uncertainty of 9.75 cm (Surti et al., 2007) while more recent systems provide time resolution down to 310 ps FWHM (Rausch et al., 2018). Utilization of the TOF information for image reconstruction offers various benefits in terms of image quality and convergence speed, see Section 2.3.

The data acquired by a PET scanner can be stored in two principally different ways:

- **Sinogram data.** In this format, the acquired data are stored as a set of special $2D$ histograms (called sinograms). Each element in the sinogram corresponds to an LOR and contains the number of events registered within this LOR over the duration of the measurement. The LORs are encoded using two coordinates: an azimuth angle describing the orientation of the LOR and the radial offset of the LOR (minimum distance to center), see Fig. 1.6. Sinograms are often interpolated in order to provide uniform sampling in both coordinates. There is one sinogram for each combination of detector rings in coincidence conventionally called *plane-of-response* or just plane. The data in sinogram representation are usually compressed in order to reduce disk space requirements and accelerate subsequent image reconstruction. This is achieved by reducing the sampling rate in angular and axial (along the scanner axis) directions, i.e. summing adjacent LORs from neighboring planes and angles. The drawback of the resampling procedure is possibly reduced image resolution compared to uncompressed data. The sinogram approach can be further extended to storing TOF data. In this case, the events are distributed among several TOF-bins (typically 13 or more) according to their measured TOF differences. Then a separate sinogram for each TOF-bin and plane is built, correspondingly increasing the resulting file size compared to the non-TOF case.

- **List-mode data.** List-mode is an uncompressed data format representing a chronological stream of the individual registered events. For each event, different bits of information are stored. The provided information can vary between different PET systems. List-mode files typically take up more disk-space than corresponding

sinogram files. Also, list-mode-based reconstructions are usually slower than their sinogram-based counterparts. Main advantages of utilizing list-mode data are improved image quality and the possibility of utilizing event-based motion-correction.

## 1.3 PET image reconstruction

Mathematically, reconstruction of tomographic images from projection data belongs to the group of so-called ill-posed inverse problems. There are two major classes of reconstruction algorithms generally used in PET: analytical and iterative reconstructions. Analytical algorithms utilize a closed-form solution for the inverse problem in order to derive the activity image. Iterative algorithms are search procedures successively optimizing image estimates by performing a comparison of predicted and measured projection data (from which correction factors to the current image estimate result). A very recent development is the emergence of reconstruction algorithms based on data-driven supervised learning using Convolutional Neural Networks (CNNs) (Zhu et al., 2018). Despite promising initial results, CNN-based reconstructions are still well behind the established PET reconstruction techniques in terms of image quality. Their ultimate usefulness and reliability have yet to be proven and they are not further considered in the present work.

### 1.3.1 Analytical reconstruction

For the sake of simplicity, we will consider the 2D reconstruction problem. Extension to the 3D case is possible but is substantially more complex (Bailey et al., 2005).

Let us introduce the following LOR parametrization: by $\phi \in [0, \pi)$ we denote the angle between LOR's normal vector and $x$-axis of the laboratory frame and with $s \in (-\infty, \infty)$ we designate the signed distance between the LOR and center of the scanner FOV (Fig. 1.6). Then the expected amount of counts in the LOR $(s, \phi)$ can be calculated as a line integral

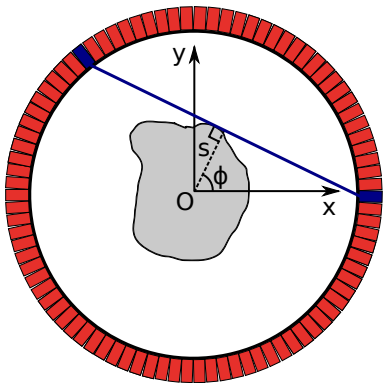


Figure 1.6: *LOR coordinates in PET. $s$ is the (signed) distance between the LOR and center of the scanner FOV, and $\phi$ is the angle between the $x$-axis and the LOR's normal vector.*

$$p(s, \phi) = \int_{-\infty}^{\infty} \lambda(s\cos\phi - t\sin\phi,\ s\sin\phi + t\cos\phi)\, dt, \quad (1.8)$$

where $\lambda(x, y)$ is the tracer concentration in the scanner FOV. The given formula is correct up to a scan-time-dependent scaling factor. Here we ignored the image quality degrading effects described in Section 1.2.4 and approximated the realistic tubes-of-response by infinitely thin lines connecting the centers of the effective detector apertures. Equation (1.8) describes the 2D Radon transform from image $\lambda$ to its projections $p$. The inverse Radon transform derives the image from the acquired projections and solves the tomographic problem.

In order to derive a closed-form inverse Radon transform formula, we assume that projection space $(s, \phi)$ is infinitely sampled and function $p(s, \phi)$ is continuous. Performing Fourier transform $\mathcal{F}$ of $p(s, \phi)$ with respect to $s$, we get a relation between projections and image in frequency space

$$\begin{aligned} P(\nu, \phi) &= (\mathcal{F}p)(\nu, \phi) = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds\, dt\, \lambda(s\cos\phi - t\sin\phi,\ s\sin\phi + t\cos\phi)\ \exp(-2\pi i\, \nu s) = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx\, dy\, \lambda(x, y) \exp[-2\pi i\, (x\,\nu\cos\phi + y\,\nu\sin\phi)] = \\ &= \Lambda(\nu\cos\phi, \nu\sin\phi), \end{aligned} \tag{1.9}$$

where $\Lambda(\nu_x, \nu_y) = (\mathcal{F}_2\, \lambda)(\nu_x, \nu_y)$ and $\mathcal{F}_2$ is the 2D Fourier transform. Equation (1.9) is usually referred to as *central section theorem* (Bailey et al., 2005) and it plays a key role in the reconstruction algorithm derivation. The theorem implies that for any given angle $\phi$ the 1D Fourier transform of parallel projections $p(s, \phi)$ defines $\Lambda(\nu_x, \nu_y)$ on a line in the frequency plain which forms an angle $\phi$ with the $\nu_x$-axis. Therefore, rotation of the projection angle $\phi$ allows recovering the whole $\Lambda(\nu_x, \nu_y)$ distribution. Going back from the frequency domain to Cartesian coordinates, we finally arrive at the desired activity image

$$\begin{aligned} \lambda(x, y) &= (\mathcal{F}_2^{-1}\Lambda)(x, y) = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\nu_x\, d\nu_y\ \Lambda(\nu_x, \nu_y) \exp[2\pi i\, (x\,\nu_x + y\,\nu_y)] = \\ &= \int_{0}^{\pi} d\phi \int_{-\infty}^{\infty} d\nu\ \nu\, \Lambda(\nu\cos\phi, \nu\sin\phi)\ \exp[2\pi i\, (x\,\nu\cos\phi + y\,\nu\sin\phi)] = \\ &= \int_{0}^{\pi} d\phi\ p^F(x\cos\phi + y\sin\phi, \phi)\,, \end{aligned} \tag{1.10}$$

where $\mathcal{F}_2^{-1}$ is inverse 2D Fourier transform and

$$p^F(s, \phi) = \int_{-\infty}^{\infty} d\nu\ \nu\, P(\nu, \phi)\ \exp(2\pi i\, s\nu) \tag{1.11}$$

is a projection $p(s, \phi)$ filtered with a *ramp filter* defined by its kernel $\nu$ in the frequency domain. The integral over $\phi$ in the last part of (1.10) is conventionally called a back-projection operation $B$. It has the geometrical meaning of representing the sum over the contributions from all LORs crossing the point $(x, y)$. This combination of filtering and backprojection explains the name for this reconstruction method: *Filtered Backprojection* algorithm (FBP). Symbolically, the FBP algorithm can be written in the following short

form

$$\lambda(x, y) = B\{\mathcal{F}^{-1}[\,\nu\,(\mathcal{F}p)(\nu, \phi)\,]\}. \tag{1.12}$$

Due to their computational efficiency, analytical reconstructions were used nearly exclusively in the clinical context for a long time. A major drawback of analytical reconstructions is the inability to gracefully handle data inconsistencies (including random errors — noise). As is typical for ill-posed inverse problems, small deviations of the measured data from the "true projections" can cause large errors in the reconstructed image. Additionally, the derivation of the reconstruction algorithm itself requires several simplifications such as ignoring most of the physical effects inherent to the acquisition process and assuming continuous projections data. The acquired data can be precorrected in order to minimize the adverse effects of the aforementioned simplifications but they cannot be eliminated completely.

### 1.3.2 Iterative reconstruction

Unlike the analytical approaches which fundamentally rely on the assumption of continuous data and image sampling in their derivation, iterative algorithms account for the discrete nature of measured data and image representation of the tracer distribution from the very beginning. Also, the acquisition model used in iterative reconstruction allows accurate description of, both, 2D and 3D acquisition in a similar manner without any substantial modifications of the reconstruction algorithm itself. Finally, incorporation of the noise model into the reconstruction process allows iterative algorithms to achieve superior noise characteristics compared to FBP.

Any iterative reconstruction consists of several elements which might be grouped into image model, acquisition model, data model, objective function, and optimization algorithm as shown in Fig. 1.7.

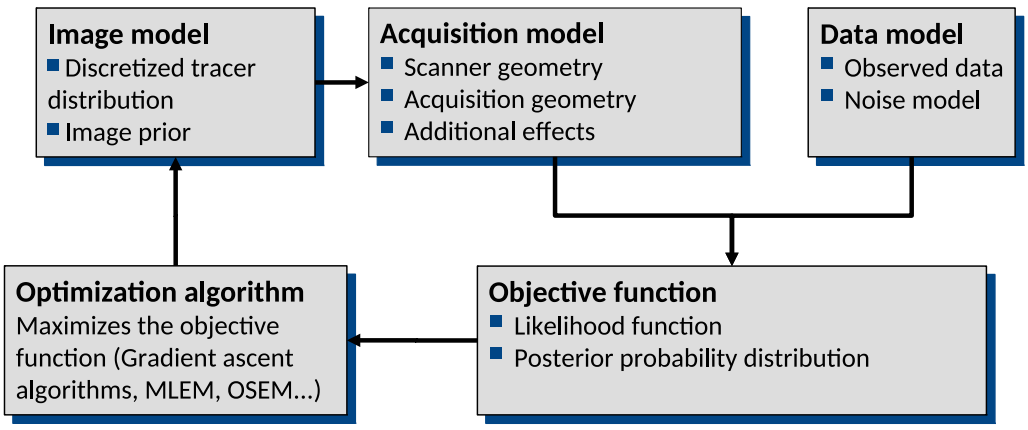


Figure 1.7: *Elements of iterative reconstruction.*

### Image model

The spatial distribution of the tracer (activity distribution) in the imaged object can be described by a continuous function $\lambda(x, y, z)$ defined in $V \subset \mathbb{R}^3$. Then, the number of decays (for a given time interval) in an infinitely small volume around $(x, y, z)$ is a Poisson variable with a mean proportional to $\lambda(x, y, z)$. Since the tracer concentration and the corresponding decay rate are proportional, the difference between them will be ignored in the following. For computational purposes, the continuous activity distribution has to be discretized. In order to do this, we introduce the finite set of $M$ voxels defined by their density functions $\rho_j(x, y, z)$ with $j = 1, \ldots, M$ such that

$$\bigcup_j \operatorname{supp}\left(\rho_j(x, y, z)\right) \supseteq V, \tag{1.13}$$

where supp stands for function support. We also require linear independence of the functions in the sense that neither of $\rho_j(x, y, z)$ can be represented as a linear combination of $\rho_k(x, y, z)$, $k \neq j$. The choice of the basis functions $\rho_j(x, y, z)$ is not unique and it affects the reconstruction results. Cubic voxels are the most popular option, but, e.g, spherical or "blob"-type voxels (Lewitt, 1990; Matej & Lewitt, 1996) can be used instead. With the basis functions defined, one can approximate the activity distribution $\lambda(x, y, z)$ as a linear combination of $\rho_j(x, y, z)$:

$$\lambda(x, y, z) \approx \sum_j \lambda_j \, \rho_j(x, y, z). \tag{1.14}$$

The set of $\lambda_j$ is a discrete representation of $\lambda(x, y, z)$ which will be used further in the reconstruction. In the simple case of rectangular non-overlapping voxels, $\lambda_j$ is an integral of $\lambda(x, y, z)$ over the volume $V_j$ corresponding to voxel $j$

$$\lambda_j = \int_{V_j} \lambda(x, y, z) \, dx \, dy \, dz. \tag{1.15}$$

Together with the image itself, the image model can also include prior knowledge regarding properties of the activity distribution in the human body. A popular choice is a Gaussian prior which enforce image smoothness and penalizes large activity variations between neighboring voxels in order to reduce the noise level. The latest efforts in this field are associated with utilization of CNN based denoising priors trained on clinical data (Kim et al., 2018). Another group of approaches involves exploiting information from other imaging modalities like CT or MR which are usually paired with PET, see e.g. (Ehrhardt et al., 2015). This kind of regularization allows to improve image resolution and decrease noise level under the assumption that the tracer concentration changes slowly within the distinct anatomical regions observable in CT and MR. Violations of this assumption, however, lead to bias in the reconstructed activity values and image artifacts at high regularization strength (Schramm et al., 2018).

### Acquisition model

The acquisition or *forward model* describes the event detection process in the PET scanner and allows to estimate the mean number of events registered by detector element $i$ if tracer distribution $\boldsymbol{\lambda}$ is given. The acquisition model includes

- **Scanner model:** scanner geometry, detector efficiencies and dead-time, and time-of-flight capabilities
- **Acquisition geometry:** position of the imaged object relative to the scanner, possible rigid and non-rigid movement of the tracer distribution during the scan
- **Additional effects:** registration of scattered and random events, photon attenuation, resolution degradation (Section 1.2.4), time dependence of activity concentration due to radioactive decay

The accuracy of the forward model determines the achievable reconstructed image quality. An insufficient acquisition model can cause a large variety of different effects from image blurring or suboptimal image noise characteristics to substantial errors in the reconstructed activity values (when omitting attenuation correction) or even total failure of getting a usable image (if scanner geometry used is wrong). The most accurate modeling of the acquisition process can be achieved with Monte-Carlo simulation but since the simulation process is extremely time-consuming, especially if a low variance in the simulated data is required, it is impossible to use it for clinical reconstruction. Therefore, analytical and semi-analytical methods are commonly used for forward modeling.

Ignoring scattered and random events, the forward model of non-TOF measurements can be derived as follows. Let $n_j$ be the (unobservable) amount of decays occurred in voxel $j$. These values are independent Poisson variables with expectations $E(n_j|\lambda_j) = \lambda_j$. Here we assume the probability of registering decay in voxel $j$ by detector element $i = 1, \ldots, N$ to be known $P(i|j) = c_{ij}$. Values $c_{ij}$ are traditionally called *system matrix* and they are mainly defined by geometry of the system and voxel grid but further physical effects can be taken into account. We will discuss the system matrix calculation details further in Section 2.2.2. Knowledge of the system matrix allows us to estimate the number of counts registered within each LOR $i$ for a given activity distribution $\boldsymbol{\lambda}$:

$$\overline{y}_i = E(n_i|\boldsymbol{\lambda}) = \sum_j c_{ij}\lambda_j. \tag{1.16}$$

This follows from the fact that

$$n_i = \sum_j n_{ij}, \tag{1.17}$$

where $n_{ij}$ is the number of photons originating in voxel $j$ and detected in LOR $i$. Indeed, $n_{ij}$ as well as $n_i$ are independent variables and from the definition of $c_{ij}$ we see that

$$E(n_{ij}|\boldsymbol{\lambda}) = P(i|j)E(n_j|\lambda_j) = c_{ij}\lambda_j \tag{1.18}$$

and, therefore, (1.16) results from (1.17) and linearity of the expectation operation.

Equation (1.16) is conventionally called the *forward projector*. The forward projector is the core of the acquisition model and, in the given form, it allows to describe all relevant effects except scattered and random events. These can be accounted for as additive contributions $s_i$ and $r_i$ to $\overline{y}_i$ that have to be estimated separately.

### Data model

While analytic reconstruction operates on sinogram data, iterative reconstructions are more flexible, supporting, both, sinogram and list-mode data. In both cases, the statistical errors contained in the data can be assumed to be adequately described by the Poisson distribution

$$P(y_i|\langle y_i\rangle) = e^{-\langle y_i\rangle}\,\frac{\langle y_i\rangle^{y_i}}{y_i!} \tag{1.19}$$

yielding the probability of registering $y_i$ events in LOR $i$ if the expectation (mean) value is $\langle y_i\rangle$. If the data are precorrected for scatter and randoms, they do no longer follow a Poisson distribution but can still be described approximately with a shifted Poisson model (Yavuz & Fessler, 1997).

### Objective function

In order to formalize the problem of finding the best estimate of the image $\boldsymbol{\lambda}$ for the measured data, the *objective function* is introduced. The objective function $Q(\boldsymbol{\lambda})$ is defined as the conditional probability of an activity distribution $\boldsymbol{\lambda}$ generating the given (measured) counts $\boldsymbol{y}$. The image $\boldsymbol{\lambda}^*$ which maximizes $Q(\boldsymbol{\lambda})$ is the most probable image

$$\boldsymbol{\lambda}^* = \arg\max Q(\boldsymbol{\lambda}). \tag{1.20}$$

Determination of an optimal estimate of $\boldsymbol{\lambda}^*$ is the goal of the iterative reconstruction process.

In a Bayesian approach, the objective function can be written as

$$Q(\boldsymbol{\lambda}) := P(\boldsymbol{\lambda}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\lambda})\,P(\boldsymbol{\lambda})}{P(\boldsymbol{y})}. \tag{1.21}$$

For convenience, the logarithm of $Q(\boldsymbol{\lambda})$ is commonly used

$$q(\boldsymbol{\lambda}) := \log\,P(\boldsymbol{y}|\boldsymbol{\lambda}) + \log\,P(\boldsymbol{\lambda}) \tag{1.22}$$

as the effective objective function. This is possible since $q$ is a strictly monotonically increasing function of $Q$. The third summand ($\log\,P(\boldsymbol{y})$) could be omitted here since it does not depend on the activity image and, therefore, is irrelevant to the optimization task (1.20).

The probability $P(\boldsymbol{y}|\boldsymbol{\lambda})$ in (1.22) is typically referred to as the *likelihood function* $L(\boldsymbol{\lambda})$.

The likelihood assesses the degree of agreement between measured data and the current image estimate when employing the given acquisition and noise models. Since the numbers of counts registered in different LORs are independent of each other the probability for the whole data set factorizes into a product of probabilities for each LOR

$$L(\boldsymbol{\lambda}) := P(\boldsymbol{y}|\boldsymbol{\lambda}) = \prod_i P(y_i|\,\overline{y}_i), \tag{1.23}$$

where $\overline{y}_i = \overline{y}_i(\boldsymbol{\lambda})$ is the forward projector defined in (1.16). If Poisson noise (1.19) is assumed then the likelihood function takes the form

$$L(\boldsymbol{\lambda}) = \prod_i e^{-\overline{y}_i} \frac{\overline{y}_i{}^{y_i}}{y_i!}. \tag{1.24}$$

Transition to the log-likelihood $l(\boldsymbol{\lambda})$, which is required in (1.22), allows replacing the product over LORs with a sum

$$l(\boldsymbol{\lambda}) := \log\, L(\boldsymbol{\lambda}) = \sum_i (y_i \,\log \overline{y}_i - \overline{y}_i) + C, \tag{1.25}$$

where $C$ is a constant independent of $\boldsymbol{\lambda}$ and, thus, can be omitted.

The summand $\log P(\boldsymbol{\lambda})$ in (1.22) penalizes convergence to the images which are *a priori* unlikely according to the utilized image model. If no prior knowledge is considered then the cost function coincides with the log-likelihood $q(\boldsymbol{\lambda}) = l(\boldsymbol{\lambda})$. Otherwise, following Bayesian terminology, $Q(\boldsymbol{\lambda})$ is a *posterior probability distribution* and $q(\boldsymbol{\lambda})$ is in this case frequently referred to as *penalized likelihood* in the literature.

### Optimization algorithm

The optimization algorithm solves equation (1.20) iteratively providing a sequence of images $\boldsymbol{\lambda}^{(p)}$, $p = 0, 1, \ldots$, which converge to the true solution $\boldsymbol{\lambda}^*$

$$\lim_{p\to\infty} \boldsymbol{\lambda}^{(p)} = \boldsymbol{\lambda}^*. \tag{1.26}$$

For practical reasons, it is required that the sequence $\{\boldsymbol{\lambda}^{(p)}\}$ increases the cost function monotonically: $q(\boldsymbol{\lambda}^{(p+1)}) \geq q(\boldsymbol{\lambda}^{(p)})$. Further desirable properties of the optimization algorithm are computational efficiency and fast convergence independent of the initial image estimate $\boldsymbol{\lambda}^{(0)}$.

There is a large number of numerical optimization algorithms available (Bailey et al., 2005) among which the most popular for PET applications are MLEM (Maximum Likelihood Estimation Maximization) and OSEM (Ordered Subsets Estimation Maximization) for non-penalized image reconstruction and their analogs for *maximum a posteriori* (MAP) reconstruction. MLEM and OSEM algorithms will be further discussed in Section 2.2.1.

### 1.3.3 Data corrections

As mentioned in Section 1.3, it is necessary to include certain corrections into the reconstruction process, specifically corrections regarding detector normalization, radioactive decay, photon attenuation and scatter as well as random coincidences. Normalization and decay corrections will be discussed in Section 2.2.2 and the others are discussed below.

#### Attenuation correction

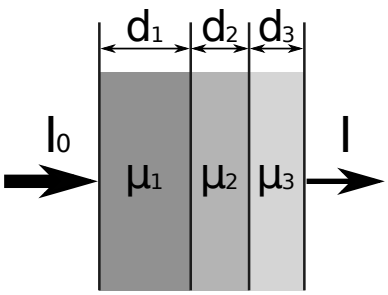


Figure 1.8: *Lambert-Beer law.*

The attenuation of a given photon flux $I_0$ resulting from traversal through a medium is described by the Lambert-Beer law

$$I = I_0 \exp\left[-\int_0^d \mu(x)\,dx\right], \tag{1.27}$$

where $I_0$ and $I$ are initial and final intensities, respectively, $\mu(x)$ is the linear attenuation coefficient of the medium and $d$ is traversed path length. In the example shown in Fig. 1.8 $\mu$ is a stepwise constant function assuming a value $\mu_i$ for layer $i$ with thickness $d_i$ in which case formula (1.27) can be rewritten as $I = I_0 \exp\Big(-\sum_i \mu_i\, d_i\Big)$.

Considering the passage of a single photon through the medium, (1.27) can be interpreted as the probability of the photon to survive the passage up to path length $d$ without being eliminated (by absorption or scatter) from the incident beam. For a pair of annihilation photons emerging from a point $x'$ on LOR $i$ with length $L$ we can calculate the respective "survival probabilities " according to

$$a_{i,1} = \exp\left[-\int_{-L/2}^{x'} \mu_i(x)\,dx\right], \tag{1.28a}$$

$$a_{i,2} = \exp\left[-\int_{x'}^{L/2} \mu_i(x)\,dx\right], \tag{1.28b}$$

where $\mu_i(x)$ is the attenuation coefficient along LOR $i$. Consequently, the coincidence detection probability (or, equivalently, photon flux) for LOR $i$ is reduced by the so-called attenuation factor

$$a_i = a_{i,1} \times a_{i,2} = \exp\left[-\int_{-L/2}^{x'} \mu_i(x)\,dx - \int_{x'}^{L/2} \mu_i(x)\,dx\right] = \exp\left[-\int_{-L/2}^{L/2} \mu_i(x)\,dx\right]. \tag{1.29}$$

As can be seen, the attenuation factor $a_i$ (and thus the change in coincidence detection probability) is independent of the point of origin of the photon pair along the LOR. This is a very advantageous property of the coincidence detection utilized in PET in comparison

to SPECT where the depth-dependent single photon attenuation is much more difficult to correct. It especially enables to determine the attenuation factor by a transmission measurement with a radioactive or X-ray source.

The spatial distribution of the linear attenuation coefficient can be discretized in the same way as the activity image as a set of $\mu_j$ values corresponding to a voxel $j$. We will call this representation the *attenuation map* or *$\mu$-map*. In this discrete representation, the integral in (1.29) is replaced by the sum

$$a_i = \exp\left[-\sum_{j\in\text{voxels}} l_{ij}\,\mu_j\right], \tag{1.30}$$

where $l_{ij}$ are suitable weighting factors accounting for the contribution of voxel $j$ to the attenuation factor for LOR $i$.

As stated in Section 1.2.2, attenuation of 511 keV $\gamma$-radiation is essentially only due to Compton scattering (and this remains true also for somewhat lower photon energies of a few hundred keV). The scatter probability of a photon of incident energy $E$ by a single electron is expressed in terms of the total cross section $\sigma_\mathrm{c}(E)$. Multiplication of the cross section by the electron density $n$ yields the linear attenuation coefficient

$$\mu(E) = n\,\sigma_c(E)\,. \tag{1.31}$$

Formula (1.31) together with (1.5) allows the recalculation of the attenuation coefficients

$$\mu(E) = \mu(E_0)\frac{\sigma_c(E)}{\sigma_c(E_0)}\,, \tag{1.32}$$

if $\mu(E_0)$ is known and $E$ does not become too small (in which case Compton scatter would no longer be the only relevant process). Equation (1.32) provides the means to assess the attenuation of the scattered photons the energy of which is bellow $E_0 = 511\,\mathrm{keV}$.

Depending on the considered PET scanner design, different methods are utilized to determine the $\mu$-map.

- **Standalone PET.** Early clinical PET systems used 511 keV emitting radioactive sources — usually $^{68}$Ge — rotating around the patient for transmission measurements in order to determine the attenuation factors directly. The corresponding attenuation map can be subsequently reconstructed from the acquired transmission data. This is principally the best approach since the correct $\gamma$ energy can be used for the transmission measurement but it suffers substantially from limited statistical accuracy and long acquisition times. In current PET/CT and PET/MR systems this method is no longer provided.

- **PET/CT.** For PET/CT systems, the CT component is used for attenuation correction by performing a scaling procedure that translates the attenuation coefficients determined at the typical X-ray energies to 511 keV as described in e.g. (Kinahan

et al., 2003). Proper handling of both photoelectric effect and Compton scattering (Section 1.2.2) as sources of attenuation in CT is achieved by using bilinear scaling function. Today, this procedure is by far the most widely used approach.

- **PET/MR.** Attenuation correction in PET/MR is less straightforward than it is in PET/CT since photon attenuation cannot be measured directly at all by these systems. There are two possible solutions to this problem

  - *MR-based attenuation correction.* The principal idea is to perform a tissue type classification of the MR images and then to insert the principally known attenuation coefficients of the different tissues to set up the $\mu$-map. There are different ways how to perform the tissue type classification such as direct MR-image segmentation (Martinez-Möller et al., 2009; Schulz et al., 2011), atlas-based methods (Rota Kops & Herzog, 2007), and the use of convolution neural networks (Han, 2017). All these methods are prone to artifacts which are caused by segmentation errors, presence of endoprostheses, etc. The principal problem is the total lack of direct information regarding the actual occurring photon attenuation.

  - *Emission-based attenuation correction.* Principally, TOF PET data contain enough information about both activity and attenuation coefficient distributions up to a global scaling factor (Defrise et al., 2012). This fact is exploited by the Maximum Likelihood Activity and Attenuation estimation algorithm (Rezaei et al., 2012) and its various modifications. The algorithm maximizes the likelihood function by updating $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ alternately until convergence. More details will be given in Section 2.5.

### Randoms correction

The contribution of random coincidences to the acquired coincidence data can be estimated by two different methods. Usually, only one of both methods is supported by any given PET system.

- **Using the singles rates.** The randoms rate in an LOR $AB$ can be computed as $2T\,S_A\,S_B$ if the singles rates $S_A$ and $S_B$ for detectors $A$ and $B$, respectively, are provided by the scanner ($2T$ is the coincidence time window).

- **Using a delayed coincidence channel.** In this approach, the number of random coincidences is measured directly by introducing a delay into one of the singles channels feeding the coincidence processor. The delay is selected to be much larger than the width of the coincidence time window in order to completely suppress the true coincidences. Since the number of detected randoms is unaffected by this delay, all coincidences detected in the delayed channel are random.

### Scatter correction

A substantial fraction of the detected coincidences involves scattered photons. This is the case despite the fact that coincidences are accepted by the event-processing electronics only in a certain energy window which already suppresses a notable fraction of the actually occurring scatter. Unfortunately, the limited energy resolution of the available PET detectors does not allow to suppress scatter more efficiently in this way. The remaining scatter in the measured coincidence data has, therefore, to be estimated and subtracted from the measured data (or, alternatively, included into the forward model of the iterative reconstruction).

Existing scatter estimation methods might be categorized as follow (Markiewicz et al., 2007):

1. **Tail fitting.** For each projection angle, a simple model (e.g. a Gaussian) is fitted to the distribution of registered coincidences in LORs not crossing the object (the so-called "scatter tails" region) and then extrapolated to the remaining LORs (those crossing the object under the considered projection angle). The method is prone to errors for asymmetrical and inhomogeneous objects.

2. **Multiple energy window techniques.** Data are acquired in two or more energy windows from which the fraction of scatter and true events is estimated based on their expected detection probability ratios within the different windows. The method suffers from the violation of underlying assumptions in the case of big and inhomogeneous objects or from insufficient statistical accuracy in the additional energy windows.

3. **Integral transformation in projection space.** The scatter event distribution is estimated as convolution of the true activity distribution projections with a suitable scatter kernel. Construction of a suitable scatter kernel which adequately takes into account the given activity distribution and scatter medium geometry is problematic, however.

4. **Integral transformation from image to projection space.** In this approach, scatter computation is performed analytically, based on a physical model of the actual scattering process where the given scatter medium as well as activity distribution are taken into account. The Single Scatter Simulation described in Section 2.2.3 follows this strategy.

5. **Monte Carlo simulation.** An event-by-event simulation of the full radiation transport and measurement process is performed. The method is very time-consuming and unsuitable for routine use in PET image reconstruction.

# Chapter 2

# Materials and Methods

## 2.1 Philips Ingenuity PET/MR scanner

Patient and phantom scans were performed with an Ingenuity TF PET/MR device (Philips Healthcare, Best, The Netherlands). This is a hybrid imaging system which combines the Philips Gemini TF time-of-flight PET with the Achieva 3T X-series MRI machine. Significant magnetic shielding was added to the PET hardware in order to avoid interference of the MRI's high magnetic field with the PMTs of the PET system. Furthermore, the PMT gains are calibrated in the presence of the MRI system in order to account for any residual magnetic fields possibly penetrating the shielding.

Figure 2.1: *Philips Ingenuity TF PET/MR scanner. MR part is on the left-hand side of the image and PET part on the right-hand side. (Taken from Philips Healthcare, Best, The Netherlands)*

The PET scanner consists of 28 detector modules forming a cylinder with 90.3 cm diameter. Each detector module consists of a $23 \times 44$ array of LYSO scintillation crystals, $4 \times 4 \times 22$ mm$^3$ each, resulting in a total of 28336 detector crystals grouped in 44 detector rings, see Fig. 2.2. Signal readout is performed by 420 PMTs coupled to the crystals forming a pixelated Anger-logic detector. Taking into account the housing, the patient bore diameter is equal to 70.7 cm. The given detector configuration results in 18 cm axial field-of-view (FOV) and a transaxial field-of-view of up to 67.6 cm which might be restricted to reduce storage requirements and accelerate image reconstruction. Two configurations are used in clinical practice: 57.6 cm FOV for whole-body scans and 25.6 cm for brain investigations. LORs not crossing the selected transaxial FOV are ignored by the scanner hardware.

The energy resolution of the system is specified as better than 13% and time resolution as better than 550 ps which is in agreement with measurements performed in (Zaidi et al., 2011) where the authors observed energy resolution of approximately 11.6% and time

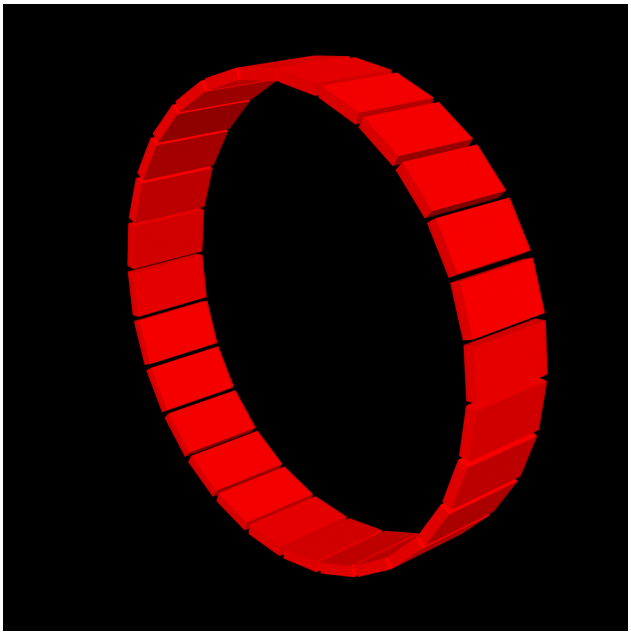

Figure 2.2: *Philips Ingenuity TF PET/MR PET detector array geometry.*

resolution of about 525 ps. However, these values were obtained with count rates close to zero, and both figures are known to deteriorate with increasing count rate for this scanner design (Surti et al., 2007).

The system uses an energy window of 460–665 keV and a coincidence time window of 6 ns. Delayed coincidence window acquisition is implemented to enable randoms correction. Data are stored in proprietary Philips uncompressed list-mode format with a list-mode TOF bin width of 25 ps. Further information about the Ingenuity TF PET/MR system can be found in (Zaidi et al., 2011).

### 2.1.1 Philips Ingenuity PET/MR scanner reconstruction software

Two image reconstruction procedures are provided by the vendor. A sinogram based 3D Row-Action Maximum Likelihood Algorithm (3D-RAMLA) is used for brain investigations while a list-mode ordered-subsets expectation maximization technique, featuring blob basis functions and incorporating time-of-flight information (BLOB-OS-TF) is used for whole-body investigations. Note that 3D-RAMLA does not allow to utilize time-of-flight information. Both reconstruction algorithms include normalization, attenuation, dead-time, randoms, and scatter corrections.

Due to the inability to perform attenuation measurements, either with 511 keV transmission sources such as $^{68}$Ge/$^{68}$Ga or with X-rays/CT, attenuation correction in PET/MR is not straightforward (see also Section 1.3.3). Philips addresses the issue by performing MR image segmentation and subsequent tissue type classification (Hu et al., 2009; Schulz et al., 2011). Altogether, three classes are used: air, lung, and soft tissue whose linear attenuation coefficients are set to 0, 0.022, and 0.096 cm$^{-1}$ respectively. The reconstruction software also allows using external $\mu$-maps generated with more elaborate methods.

For example, 4-class MR image segmentation using the Dixon MR sequence for adipose tissues classification as proposed in (Martinez-Möller et al., 2009) was implemented by our group. Despite its principal advantages, this method is not used in clinical routine since it requires an additional time-consuming MR acquisition.

The vendor software also provides an attenuation map truncation compensation protocol in order to address the problem of significantly smaller field-of-view of the MR part of the scanner compared to PET one. According to our observations, the protocol serves its purpose reasonably well in general but can fail under certain circumstances (Schramm et al., 2013). Moreover, the possible presence of susceptibility artifacts in the MR images caused by metal implants in the patient body disrupts the standard segmentation procedure and causes generation of erroneous attenuation maps. To reduce the adverse effects of metal-implant-induced artifacts, an in-house MR-based attenuation correction tool was developed in our group (Schramm et al., 2014). It utilizes the PET emission data in order to determine the body contour, identifies regions affected by artifacts and corrects MR image segmentation and truncation errors.

For randoms correction, the delayed coincidence channel approach is used while scatter correction is based on the well known Single Scatter Simulation (SSS) algorithm. The overall scale of the generated scatter distribution is determined by matching the distribution to the measured "scatter tails" (Polycarpou et al., 2011).

The reconstruction software allows adjusting several parameters which influence reconstruction speed/accuracy and the degree of image smoothing. Unfortunately, no information is available on how exactly these parameters modify the iteration scheme and post-filtering smoothing kernel.

## 2.2 THOR reconstruction software

The reconstruction software used in this work is called THOR (Tube of Response High Resolution OSEM Reconstruction). The name refers to the used method for system matrix calculation (see Section 2.2.2). This is an in-house developed reconstruction research toolkit based on the ordered subset accelerated version of Maximum Likelihood Expectation Maximization (MLEM) algorithm by L. A. Shepp and Y. Vardi (Shepp & Vardi, 1982; Hudson & Larkin, 1994). THOR was first presented in (Lougovski et al., 2014) and further investigated in (Lougovski et al., 2015). For reference, we provide an overview of the concepts underlying THOR and describe the version of THOR prior to implementation of the modifications proposed and described within this work.

### 2.2.1 Reconstruction algorithm

The derivation of the MLEM algorithm and the proof of its convergence are given in the Appendix. Here we provide the canonical MLEM iteration formula without any data

corrections applied

$$\lambda_j^{(p+1)} = \lambda_j^{(p)} \frac{1}{\sum\limits_i c_{ij}} \sum_i c_{ij} \frac{y_i}{\sum\limits_{j'} c_{ij'}\lambda_{j'}}, \tag{2.1}$$

where $\lambda_j$ is the activity concentration estimate in voxel $j$ in the iteration $p$, $y_i$ is the number of counts registered in LOR $i$, and $c_{ij}$ is the system matrix element for LOR $i$ and voxel $j$ (Section 1.3.2). The sum over all voxels in the denominator is the forward projector (1.16). It represents the projection of the activity distribution onto a given detector, i.e the (estimated) count integral in the respective LOR. The measured-to-estimated LOR counts ratio is projected back into the image by the outer sum over $i$ and, therefore, this operation is usually referred to as *back projection*. The final division by the sensitivity term $S_j = \sum_i c_{ij}$ accounts for unequal LOR coverage of the voxels.

Equation (2.1) guarantees non-negativity of the image estimate $\boldsymbol{\lambda}$ provided a non-negative initial estimate $\boldsymbol{\lambda}^{(0)}$ (we use $\lambda_j^{(0)} = 1$, $j = 1, \ldots, J$). Usually, a pre-defined, protocol-specific number of iterations is used for reconstruction.

The iteration formula (2.1) can be easily extended to list-mode (LM) acquisitions (LM-MLEM). The complete derivation of LM-MLEM is given in (Parra & Barrett, 1998). Since for list-mode acquisition the individual events are stored, the sum over all LORs $i$ is replaced by a sum over all events $e$:

$$\lambda_j^{(p+1)} = \lambda_j^{(p)} \frac{1}{\sum\limits_i c_{ij}} \sum_e c_{i_e j} \frac{1}{\sum\limits_{j'} c_{i_e j'}\lambda_{j'}}, \tag{2.2}$$

where $i_e$ denotes an LOR corresponding to coincidence event $e$. After grouping all the events which belong to the same LOR and taking into account that there are $y_i$ events registered within the LOR, the list-mode and canonical MLEM formulas become the same if the system matrix elements $c_{i_e j}$ do not differ between different events in the LOR. We will be using this reconstruction method in this work.

MLEM reconstruction is computational very intensive so there is demand for acceleration. A straightforward way to achieve this is called Ordered Subsets Estimation Maximization (OSEM) (Hudson & Larkin, 1994). For OSEM reconstruction, the input data (LOR projections or list-mode) are separated into $M$ subsets

$$D = D_1 \cup D_2 \cup D_3 \cup ... \cup D_M,$$

with

$$D_i \cap D_j = \begin{cases} D_i & , i = j \\ \varnothing & , i \neq j \end{cases} \tag{2.3}$$

(i.e. the subsets don't overlap). There are different possibilities of how to subdivide the data. For list-mode, temporal or geometrical criteria can be used. In our implementation of LM-OSEM, we group the events by LOR. Subset $D_i$ is formed from the events which

belong to LOR subset $L_i$, $i = 1 \dots M$. The subsets $L_i$ are chosen in a way that ensures homogeneous LORs distribution among planes, views, and radial elements, so the subsets can be processed in any order.

For LM-OSEM the iteration formula can be represented as (Levkovitz et al., 2001)

$$\lambda_j^{(p+1)} = \lambda_j^{(p)} \frac{1}{\sum\limits_{i \in L_k} c_{ij}} \sum_{e \in D_k} c_{i_e j} \frac{1}{\sum\limits_{j\prime} c_{i_e j\prime} \lambda_{j\prime}}, \quad k = 1 + p \bmod M, \tag{2.4}$$

where mod is the modulo operation. The algorithm cycles through the subsets using a new portion of data for each subsequent activity image update. We call an *OSEM iteration* one full cycle over all subsets $1 \dots M$. Note that an OSEM iteration takes approximately the same time as an MLEM iteration, but OSEM requires $M$ times fewer iterations to converge (thus accelerating the reconstruction by this factor).

### 2.2.2 System Matrix

The system matrix is a key component of iterative PET reconstruction. Since system matrix element $c_{ij}$ is defined as the probability of registering a photon pair originating from a decay in voxel $j$ in LOR $i$ the system matrix represents the effective model used for description of the annihilation photons generation, propagation, and detection process. The system matrix usually only describes detection of the unscattered correlated photon pairs, i.e. true events. The system matrix includes the geometrical probability of registering an event within a given LOR as well as relevant physical effects influencing the detection process. These can conceptually be divided into two groups:

1. **Resolution degrading**
   - Positron range
   - Photon accolinearity
   - Photon intercrystal scattering
2. **Signal diminishing**
   - Photon attenuation
   - Limited crystal efficiency
   - Activity concentration reduction due to radioactive decay

For more details, see Section 1.2.

There are different possibilities of how to account for these effects (or to neglect them completely) before or during reconstruction. These include different data precorrection techniques and image and sinogram blurring on the one hand, and incorporation into the system matrix on the other hand. We have chosen the latter approach. The mentioned resolution degrading effects were accounted for by modifying the geometrical probability calculation in a semi-phenomenological manner. Effects of the second group enter the

system matrix as multiplicative factors. In the following, we describe the system matrix computation in more detail.

### Geometrical detection probability

By geometrical detection probability we denote the probability of detecting a photon pair originating from voxel $j$ with detector pair $i$ in the absence of any scatter and assuming ideal detectors (detection guaranteed if the photons pass through the respective detector's effective "aperture"). In the following, we use the convention that $c_{ij}$ denotes this geometrical probability. All other contributions to the system matrix will be expressed explicitly by separate factors.

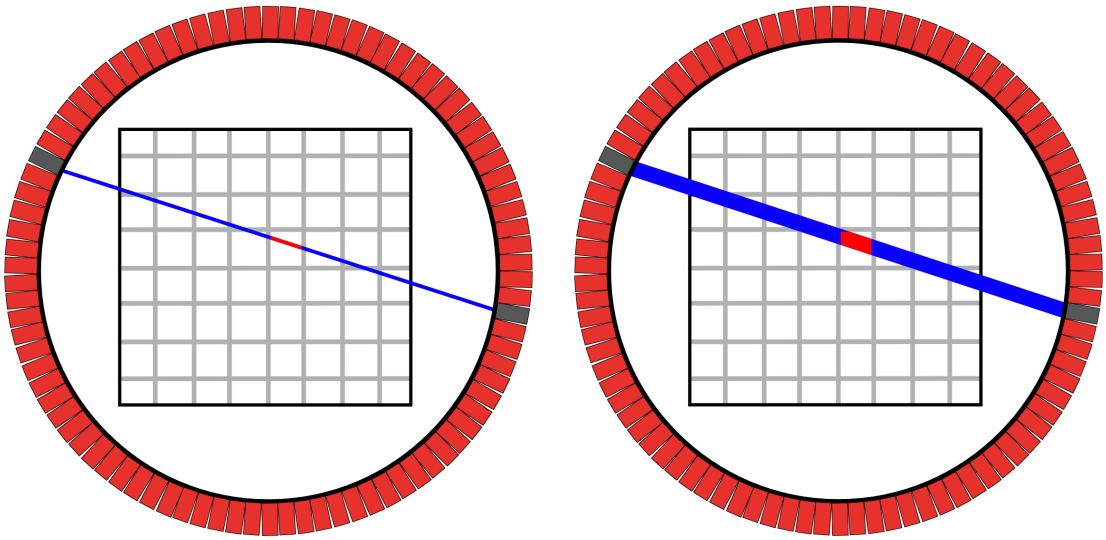

Figure 2.3: *Length (left) and volume (right) as a measure of intersection. (Taken from (Lougovski, 2012))*

THOR utilizes a *volume-of-intersection* approach instead of the widely used *length-of-intersection* as a measure of detection probability (Lougovski et al., 2014). The difference between both approaches is illustrated in Fig. 2.3. Basically, the difference is to use a *Tubes-Of-Response* (TOR) connecting the effective detector apertures instead of a *Lines-Of-Response* (LOR) connecting the centers of these apertures. The system matrix elements are then given by

$$c_{ij} = \frac{V_{\text{int}}(i,j)}{V_{\text{tube}}(i)}, \tag{2.5}$$

where $V_{\text{int}}(i,j)$ is the volume-of-intersection of TOR $i$ with voxel $j$ and $V_{\text{tube}}(i)$ is the total volume of TOR $i$.

The TOR approach offers a flexible way to model the event detection with finite detectors of possibly spatial variant detection efficiency as well as positron range, photon acollinearity, etc. Variation of the TOR radius (and possibly its density) are a means to naturally describe the different resolution degrading effects mentioned above. The

relation between the TOR-based approach and certain LOR-based approaches to image reconstruction was further investigated in (Lougovski et al., 2015).

In order to allow *on-the-fly* $V_{\text{int}}(i,j)$ volume computation, approximate symmetries were employed. First of all, voxels were described as spheres instead of cubes arranged in the same rectangular grid. Choosing a sphere radius of

$$R_{\text{vox}} = \sqrt[3]{\frac{3\,V_{\text{vox}}}{4\,\pi}} \tag{2.6}$$

makes the sphere volume equal to that of the original cubic voxel, $V_{\text{vox}}$. Second, a constant density cylinder was chosen as the TOR model. The TOR radius is a tunable parameter which accounts for the assorted resolution degrading effects and the finite detector size. A reasonable choice for the tube radius is

$$R^0_{\text{tube}} = \sqrt{\frac{S_{\text{det}}}{\pi}}, \tag{2.7}$$

where $S_{det}$ is the effective detector aperture. It was shown in (Lougovski et al., 2014), that this choice of voxel and tube size already allows achieving sub-5mm reconstructed resolution without causing Gibbs (or ringing) artifacts at object boundaries. After further testing, we adjusted tube radius to $R_{\text{tube}} = 1.2\,R^0_{\text{tube}}$ in the present study which was found to yield the best overall results for $^{18}$F-FDG studies.

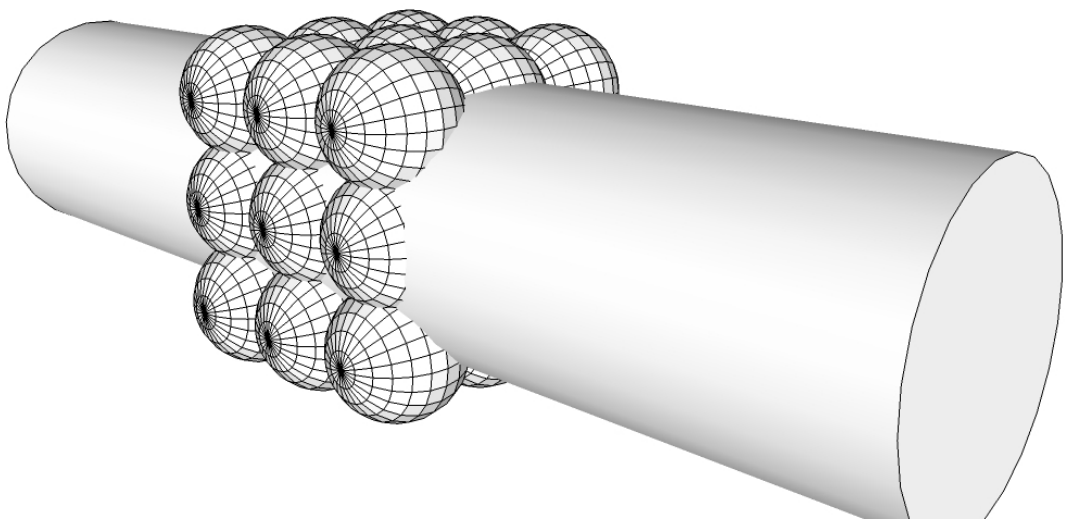

Figure 2.4: *Cylindrical LOR intersects voxel grid with spherical voxels. (Taken from (Lougovski, 2012))*

The utilized model of spherical voxels and cylindrical TORs is illustrated in Fig. 2.4. The significant computational advantage of this model stems from the fact that the intersection volume between TOR $i$ and voxel $j$, $V_{\text{int}}(i,j) = V(d_{ij})$, is only a function of the "impact parameter" $d_{ij}$, the distance between the TOR axis and voxel center. The impact parameter is fast to compute during forward projection while a lookup table for

$V(d)$ can be precomputed prior to reconstruction using elliptical integrals (Lamarche & Leroy, June 1990) and is easily stored in memory.

### Normalization

The system response to activity in the FOV is not uniform across different LORs. The variations are related to differences between detector crystals and signal processing electronics as well as to purely geometrical effects. The process of correcting for unequal LOR sensitivities is known as *normalization*. Normalization factors for individual LORs are measured during scanner calibration which is part of routine maintenance. Since direct measurement of the normalization factors is a time-consuming process the component based approach is utilized by the manufacturer (Wang et al., 2007). The approach assumes that LOR normalization factors can be represented as

$$n_i = n_i^{Geom}\, \varepsilon_A\, \varepsilon_B, \tag{2.8}$$

where $n_i^{Geom}$ represents a geometry factor for LOR $i$ and $\varepsilon_A$ and $\varepsilon_B$ are intrinsic efficiencies of the detectors defining the LOR. The detector efficiencies are determined using a dedicated cylinder phantom and the geometry factor is derived using a plane source as suggested in (Wang et al., 2007).

### Attenuation correction

The basic principles of attenuation correction in PET were explained in Section 1.3.3. Applying these to the TOR system matrix calculation approach used in THOR, we compute the coefficients $l_{ij}$ needed in (1.30) as

$$l_{ij} = L\, c_{ij}, \tag{2.9}$$

where $L$ is the total length of the considered TOR. Definition (2.9) turns the sum over all voxels in (1.30) into a product of the length of the TOR and the mean attenuation coefficient $\langle \mu \rangle_i$ within the TOR,

$$a_i = \exp\left[ -L \sum_{j \in \text{voxels}} \frac{V_{\text{int}}(i,j)}{V_{\text{tube}}}\, \mu_j \right] = \exp\big( -L\, \langle \mu \rangle_i \big)\,, \tag{2.10}$$

which is in correspondence with (1.29).

The required attenuation maps were obtained either externally (by MR- or transmission-scan-based methods) or internally by the MLAA approach discussed in Section 2.5.

### Decay correction

Most PET tracers have very short radioactive half-lives which leads to non-negligible decay of the tracer during the course of the measurement. For example, the popular PET isotope

$^{18}$F has a half-life of about 110 min while the acquisition can last up to more than one hour.

According to the radioactive decay law (1.2), the remaining activity $A(t)$ at time $t$ is related to initial activity at the chosen reference time $A(t_0) = A_0$ by

$$A(t) = A_0 \, 2^{-(t-t_0)/T_{1/2}} \,,$$

where $T_{1/2}$ is the radioactive half-life. Defining the decay correction factor $d_e$

$$d_e = 2^{(t_e-t_0)/T_{1/2}} \,, \tag{2.11}$$

where $t_e$ is the detection time for event $e$ and applying the radioactive decay law to the forward projector (1.16) we get the decay corrected count rate expectation in LOR $i$

$$\overline{y}_i = \frac{1}{d_e} \sum_j c_{ij} \lambda_j$$

which replaces the corresponding expression in Eq. (2.1).

### 2.2.3 Scatter correction

We are using the Single Scatter Simulation algorithm in order to estimate the amount of scattered events $s_i$ in each LOR $i$. SSS was chosen among the available alternatives listed in Section 1.3.3 since it provides a combination of still acceptable efficiency and good accuracy. The SSS formula is an analytical description of the photon single scattering process

$$s_i^{\mathrm{SSS}} = \int\limits_{V_S} dV_S \left( \frac{\sigma_{AS}\, \sigma_{BS}}{4\pi\, R_{AS}^2 R_{BS}^2} \right) \frac{\mu_S}{\sigma_c} \frac{d\sigma_c}{d\Omega} \left[ I^A + I^B \right] , \tag{2.12}$$

where $V_S$ is the scatter medium volume and $S$ denotes the scatter point inside volume element $dV_S$. $A$ and $B$ denote the detectors defining the LOR $i$. $R_{AS}$ and $R_{BS}$ denote the distances between point $S$ and detectors $A$ and $B$, respectively. $\sigma_{AS}$ and $\sigma_{BS}$ are the surface areas of the detectors $A$ and $B$, respectively, as they appear to an observer at the point $S$. $\mu_S$ here is the linear attenuation coefficient of the scattering medium at $S$. $\sigma_c$ is the total Compton cross section (1.5) and $d\sigma_c/d\Omega$ the differential Compton cross section at 511 keV (Klein & Nishina, 1929):

$$\frac{d\sigma_c}{d\Omega} = r_e^2 \, (2 - \cos\theta)^{-2} \left[ (2 - \cos\theta)^{-1} + 2 - \cos\theta - \sin^2\theta \right] / 2 \,, \tag{2.13}$$

where $r_e = e^2/m_e c^2$ is the so-called classical electron radius ($e$ and $m_e$: electron charge and mass) and $\theta$ is the scatter angle, see Fig. 2.5.

$I^A$ and $I^B$ account for variable photon emission intensity, photon attenuation, and photon detection probability assuming that the positron annihilation happened somewhere

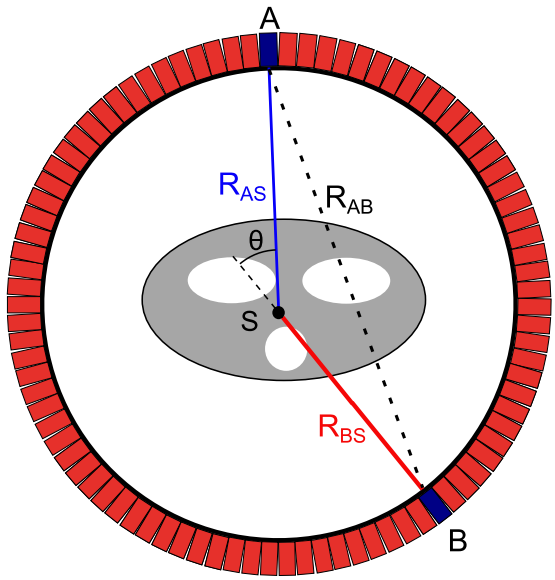


Figure 2.5: *Scattering of $\gamma$-quanta in the body. (Modified after (Lougovski, 2012))*

along the lines $AS$ and $BS$, respectively

$$I^A = \epsilon_{AS}\,\epsilon'_{BS}\,\exp\left[-\int_S^A \mu(s)\,ds - \int_S^B \mu'(s)\,ds\right]\int_S^A \lambda(s)\,ds\,, \tag{2.14a}$$

$$I^B = \epsilon'_{AS}\,\epsilon_{BS}\,\exp\left[-\int_S^A \mu'(s)\,ds - \int_S^B \mu(s)\,ds\right]\int_S^B \lambda(s)\,ds\,, \tag{2.14b}$$

where $\lambda(s)$ and $\mu(s)$ are activity and attenuation image samples taken along the lines $AS$ and $BS$, and $\epsilon_{AS}$ and $\epsilon_{BS}$ are the efficiencies of detectors $A$ and $B$, respectively, for 511 keV $\gamma$-quanta. Primed quantities represent the respective values calculated for the scattered photons of energy

$$E' = \frac{m_e c^2}{2 - \cos\theta}. \tag{2.15}$$

Attenuation coefficients $\mu'$ can be rescaled to the target energy using (1.32). For the detector efficiencies a model similar to the one described in (Accorsi et al., 2004) is used

$$\begin{aligned}\epsilon_{AS}(E) &= \frac{\varepsilon_A}{\sqrt{2\pi}\sigma_E}\int_{E_{LLD}}^{E_{ULD}} \exp\left[-\frac{(E'-E)^2}{2\,\sigma_E^2}\right]dE' = \\ &= \frac{\varepsilon_A}{2}\left\{\mathrm{erf}\left[\frac{2\sqrt{\log 2}\,(E_{ULD}-E)}{E_{FWHM}}\right] - \mathrm{erf}\left[\frac{2\sqrt{\log 2}\,(E_{LLD}-E)}{E_{FWHM}}\right]\right\}\end{aligned} \tag{2.16}$$

($E_{LLD}$ and $E_{ULD}$: lower and upper energies thresholds, $\sigma_E$ and $E_{FWHM}$: standard deviation and Full Width at Half Maximum (FWHM) of the Gaussian used for describing the energy resolution, $\varepsilon_A$: intrinsic crystal efficiency defined through the scanner calibration procedure). The relevant values of all parameters for the Ingenuity PET/MR scanner are

listed in Section 2.1.

For numerical integration of (2.12) , we use a sparse set of scatter points $S$ as suggested in (Watson, 2000). The points are distributed evenly in the field-of-view with a mean spacing of 22.5 mm. Points with $\mu_S < 0.04\text{cm}^{-1}$ are excluded since their contribution to the integral is relatively small. The SSS algorithm iterates over the scatter points and determines the contribution of each point to the total scatter in LOR $AB$ by computing

1. Scatter probability $\frac{\mu_S}{\sigma_c}\frac{d\sigma_c}{d\Omega}$
2. Geometrical detection probability $\frac{\sigma_{AS}\,\sigma_{BS}}{4\pi\,R_{AS}^2 R_{BS}^2}$
3. Number of decays along the lines connecting scatter point $S$ and detectors $A$ and $B$: $\int\limits_S^A \lambda(s)\,ds$ and $\int\limits_S^B \lambda(s)\,ds$, respectively
4. Photon attenuation $\exp\left[-\int\limits_S^A \mu(s)\,ds - \int\limits_S^B \mu'(s)\,ds\right]$ and energy-dependent detection probability $\epsilon_{AS}\,\epsilon'_{BS}$ for a photon pair emitted somewhere along the line $AS$ (and similar quantities for a photon pair emitted along $BS$).

A naive straightforward implementation of SSS results in intolerably long computation times. The algorithm can be accelerated as follows. First of all, the scatter distribution is very smooth, i.e it varies only very slowly across different LORs. Consequently, it is permissible to estimate the scatter contribution only for a small subset of the available LORs and to interpolate this distribution for the remaining LORs. This optimization alone accelerates SSS by two orders of the magnitude or even more. A further improvement arises from the structure of the contributions $I^A$ and $I^B$ in (2.14) which are derived from line integrals through the activity and attenuation maps. These integrals can be precalculated for all scatter point/detector pairs and then reused multiple times for different LORs.

The scatter distribution estimated via (2.12) can not be used directly for scatter correction of the measured data. For one, the contribution of multiple scatter has to be considered. Second, only the contribution of scatter originating from decays in the effective total FOV covered by the given scan is calculated. In most cases, however, not the whole body of the patient is imaged and scattered photons coming from outside of the FOV remain unaccounted for. The standard method to address both problems is a scatter scaling procedure. More details on this topic and on our implementation of the scatter scaling are given in Section 2.7.1.

### 2.2.4 Randoms correction

The Philips Ingenuity PET/MR uses delayed channel coincidences for randoms estimation. However, the measurements are typically too noisy for direct use in list-mode reconstruction. Variance reduction techniques are thus required. We are using the single-plane

variance reduction method due to Casey (SP-C) (Casey & Hoffman, 1986) which was also evaluated for 3D-PET (Badawi et al., 1999). This algorithm does not introduce any bias into the estimated randoms and allows to balance between computational efficiency and the achieved variance reduction. We assume that number of expected delayed and, therefore, random events can be calculated as

$$r_{AB} = k\, S_A S_B, \tag{2.17}$$

where $S_A$ and $S_B$ are the singles rates of detectors $A$ and $B$, respectively, and $k$ is a coefficient depending on scan duration and proportional to coincidence time window $2T$. We can sum both sides of (2.17) over a certain detector group $G_A$ and then get

$$\sum_{D \in G_A} r_{AD} = k\, S_A \sum_{D \in G_A} S_D, \tag{2.18}$$

$$M\, \langle r \rangle_{A\,G_A} = k\, M\, S_A\, \langle S \rangle_{G_A}, \tag{2.19}$$

where $M$ is the number of detectors in group $G_A$, $\langle S \rangle_{G_A}$ is the mean value of singles rates of the detectors in the group $G_A$, and $\langle r \rangle_{A\,G_A}$ is the mean value of delayed events registered in the LOR group $AG_A = \{AC\}_{C \in G_A}$. From (2.19) we can obtain the singles-rate estimate for detector $A$

$$S_A = \frac{1}{k} \frac{\langle r \rangle_{A\,G_A}}{\langle S \rangle_{G_A}}. \tag{2.20}$$

Combining (2.17) and (2.20) we finally get the denoised estimate of random event count in the LOR $AB$

$$r_{AB} = \frac{\langle r \rangle_{A\,G_A} \langle r \rangle_{B\,G_B}}{k \langle S \rangle_{G_A} \langle S \rangle_{G_B}} = \frac{\langle r \rangle_{A\,G_A} \langle r \rangle_{B\,G_B}}{\langle r \rangle_{G_A G_B}}, \tag{2.21}$$

where $\langle r \rangle_{G_A G_B}$ is the average of delayed events measured in the LOR group $G_A G_B$ which is connecting detector groups $G_A$ and $G_B$. In our reconstruction, we use 8 non-overlapping groups of adjacent detectors. $G_A$ and $G_B$ are defined as the groups opposite to detectors $A$ and $B$, respectively.

### 2.2.5 Complete MLEM algorithm

Including the aforementioned corrections into (2.2) the fully corrected MLEM iteration formula finally reads

$$\lambda_j^{(p+1)} = \lambda_j^{(p)} \frac{1}{\sum\limits_i a_i\, n_i\, c_{ij}} \sum_e a_{i_e} n_{i_e} c_{i_e j} \frac{d_e}{\sum\limits_k a_{i_e} n_{i_e} c_{i_e k}\, \lambda_k^{(p)} + s_{i_e} + r_{i_e}}, \tag{2.22}$$

where $\lambda_j^{(p)}$ denotes estimated activity concentration in voxel $j$ after the $p^{th}$ iteration, $i$ enumerates the different LORs, $a_i$ is the attenuation and $n_i$ the normalization factor, and $c_{ik}$ is the system matrix element for LOR $i$ and voxel $k$. $s_i$ and $r_i$ are estimated additive contributions of scattered and random events in LOR $i$.

Note that attenuation correction and normalization are incorporated into the sensitivity

$$S_j = \sum_i a_i \, n_i \, c_{ij}, \tag{2.23}$$

as well as into the forward projector (together with scatter and random corrections)

$$\overline{y}_i = \sum_k a_i \, n_i \, c_{ik} \lambda_k + s_i + r_i. \tag{2.24}$$

This makes the algorithm (2.22) a list-mode analog of the ordinary Poisson OSEM (Politte & Snyder, 1991; Conti et al., 2005) which does properly handle Poisson statistics of the data.

### 2.2.6 Reconstruction of multiple bed positions

The axial FOV of clinical PET scanners is limited (18 cm for Ingenuity TF PET/MR). In order to obtain a whole-body image, the acquisition is split into a number of scans of sequential bed positions. A slight overlap between subsequent bed position allows compensating for the reduction of the scanner sensitivity at the axial FOV edges.

THOR combines the information from multiple bed positions in LOR space and simulates a "virtual total body scanner" covering the whole patient which has several advantages. The whole-body image is reconstructed as a whole contrary to conventional bed-by-bed reconstruction. This allows better handling of information from the bed overlap regions since no stitching of multiple separate images is required. Moreover, this approach is beneficial in terms of scatter correction quality since the whole continuous scatter distribution can be generated at once with out-of-FOV (ooFOV) scatter simulation being seamlessly integrated.

The benefits of our virtual total body scanner approach come at the price of considerably increased RAM requirements to store the data for all the bed positions simultaneously. Additionally, the scatter estimation process for a virtual total body scanner demands more computational resources due to the greatly increased axial FOV and, therefore, increased number of possible scatter points to consider.

### 2.2.7 Distributed computation

List-mode PET image reconstruction is a computationally demanding task which poses a great challenge to the computer hardware in use. Reconstruction of a typical clinical list-mode data set obtained with a modern PET scanner on a single CPU core requires days of calculations and is, therefore, practically unusable. Efficient parallel implementation of the reconstruction software is necessary in order to make reconstruction times compatible with clinical needs.

To this end, THOR was split into a *server* and *client* applications. The client is a small program which initializes the reconstruction process, connects to instances of the

server program and delegates the reconstruction workload to them. It also interacts with the server instances during the reconstruction to ensure proper synchronization between them.

The server application is responsible for the actual computations. There is one server instance running on each of the computers participating in the reconstruction, allowing for distribution of the workload over the available machines. Moreover, the computations are parallelized in the server application to take advantage of multi-core CPUs. Parallelization is facilitated by the structure of the MLEM formula (2.25) where each term of the sum is associated with a single LOR. Therefore the computation can be easily split into multiple threads each processing a distinct subset of LORs. The same applies to other reconstruction algorithms implemented in THOR. The results of the computations are transferred back to the supervising client as light-weight voxel data arrays which do not cause relevant network load.

Currently, most of the code is already optimized for parallel computation, but certain tasks are still performed in single-threaded mode either because they are computationally inexpensive or because the parallelization possibilities are limited. In detail, the current state of affairs is as follows:

- Client
  - Single-threaded
    1. Input/Output operations
    2. Image-space operations
    3. Aggregating the data from the servers
  - Multi-threaded
    1. Interaction with servers
    2. Network data transfer
- Server
  - Single-threaded
    1. Input/Output operations
    2. Image-space operations
    3. SSS scatter points generation
    4. Aggregating the data from the threads
  - Multi-threaded
    1. Sensitivity calculation
    2. Iterative algorithms (MLEM, MLRES, MLAA)
    3. Single Scatter Simulation (TOF and non-TOF)
    4. ISA algorithm

5. Scatter scaling
6. Scatter interpolation

Procedures whose parallel implementation was achieved within the present work have been underlined.

Despite the parallel implementation of the most computationally intensive tasks, reconstruction speed does not scale linearly with the number of servers and processing cores. This is explained by limited storage and network performance as well as the remaining fraction of single-threaded operations and the requirement for server synchronization during the reconstruction process.

For the present work, THOR reconstruction was performed on a ten servers setup (8 × 28 cores Intel Xeon E5-2690 v4 + 1 × 24 cores Intel Xeon E5-2690 v3 + 1 × 24 cores Intel Xeon E5-2697 v2, 128 GB RAM each). Calculations were distributed among the servers proportional to their individual performance.

## 2.3 TOF extension of the reconstruction algorithm

Utilization of TOF information (Section 1.2.4) provides clear benefits for iterative PET image reconstruction in terms of reduced image noise and improved convergence rate (Conti et al., 2005; Vunckx et al., 2010). Moreover, it has been shown that TOF image reconstruction is more robust in the presence of data inconsistencies such as erroneous attenuation map or detector normalization (Turkington & Wilson, 2009; Conti, 2011; ter Voert et al., 2017), and provides better lesion detectability (Surti et al., 2006; Kadrmas et al., 2009; El Fakhri et al., 2011; Mühlematter et al., 2018). Finally, utilizing the TOF information allows reducing the cross-talk between reconstructed activity and attenuation maps in joint image reconstruction (Defrise et al., 2012).

As mentioned in Section 2.1, the Philips Ingenuity PET/MR scanner supports TOF measurements. Naturally, proper treatment of the additional TOF information within the reconstruction software is necessary. In this section, we introduce the TOF extension of the list-mode MLEM algorithm (2.22). The difficulties of quantitative TOF reconstruction as well as possible solutions are further discussed in Section 2.4 and Section 2.7.2.

### 2.3.1 TOF MLEM

The list-mode version of the time-of-flight MLEM algorithm can be represented as follows

$$\lambda_j^{(p+1)} = \lambda_j^{(p)} \frac{1}{\sum\limits_i a_i\, n_i\, c_{ij}} \sum_e a_{i_e} n_{i_e} c_{i_e j,\delta_e}(\sigma_{\mathrm{TOF,e}}) \times \\ \times \frac{d_e}{\sum\limits_k a_{i_e} n_{i_e} c_{i_e k,\delta_e}(\sigma_{\mathrm{TOF,e}})\, \lambda_k^{(p)} + s_{i_e,\delta_e}(\sigma_{\mathrm{TOF,e}}) + r_{i_e,\delta_e}}\,, \quad (2.25)$$

where $c_{ik,\delta}(\sigma_{\mathrm{TOF}})$ is a system matrix element for LOR $i$ and voxel $k$. The further index $\delta$ explicitly denotes the dependency on the difference of photon arrival times at both detectors. $\sigma_{\mathrm{TOF}}$ is the standard deviation of the Gaussian used to describe the finite time resolution (response characteristic) of the coincidence measurement (which can be count-rate-dependent (Surti et al., 2007)). Subscript $e$ in $i_e$, $\delta_e$ and $\sigma_{\mathrm{TOF,e}}$ indicates that the respective quantity refers to a specific single event $e$. Accordingly, $s_{i,\delta}(\sigma_{\mathrm{TOF}})$ and $r_{i,\delta}$ are TOF dependent additive contributions of scattered and random events in LOR $i$. In the following, we use the convention that omission of index $\delta$ designates the non-TOF value of the respective quantity. Otherwise, the notation follows (2.22).

The TOF forward projector for TOF difference $\delta$ is then given by

$$\overline{y}_{i,\delta} = a_i\, n_i \sum_j c_{ij,\delta}(\sigma_{\mathrm{TOF}})\, \lambda_j + s_{i,\delta}(\sigma_{\mathrm{TOF}}) + r_{i,\delta}. \tag{2.26}$$

The TOF system matrix elements can be obtained from the corresponding non-TOF system matrix elements by multiplication with the relative probability of registering a photon pair emitted from voxel $j$ in LOR $i$ considering the given TOF difference

$$c_{ij,\delta}(\sigma_{\mathrm{TOF}}) = c_{ij}\, w_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}}), \tag{2.27}$$

where $\delta_{ij}$ is a signed TOF-difference of the center of the voxel $j$ along LOR $i$. Considering $\delta$ and $\sigma_{\mathrm{TOF}}$ as parameters, the probability is normalized to fulfill

$$\frac{1}{2\,T} \int_{-T}^{T} w_i(\delta'; \delta, \sigma_{\mathrm{TOF}})\, d\delta' = 1, \qquad \delta \in [-T,\, T],\ \sigma_{\mathrm{TOF}} > 0\,, \tag{2.28}$$

where $2T$ is the coincidence time window.

For illustration, we present the TOF-weights function $w_i$ for the limiting cases $\sigma_{\mathrm{TOF}} \to 0$ and $\sigma_{\mathrm{TOF}} \to \infty$, respectively.

For $\sigma_{\mathrm{TOF}} \to 0$ (infinite time resolution) the TOF-weight $w_i^0(\delta_{ij}; \delta)$ can be represented as a boxcar function, corresponding to zero detection probability if the measured annihilation location lies outside of voxel $j$:

$$w_i^0(\delta_{ij}; \delta) \equiv w_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}} \to 0) = \begin{cases} \dfrac{2\,T}{\Delta T_i}, & \delta_{ij} - \delta \in \left[-\dfrac{\Delta T_i}{2}, \dfrac{\Delta T_i}{2}\right], \\ 0, & \text{otherwise}, \end{cases} \tag{2.29}$$

where $\delta_j$ is arrival time difference for a photon pair emitted from the center of the voxel $j$, and $\Delta T_i$ is the effective length of the voxel along the LOR $i$ (using units such that $c = 1$).

For $\sigma_{\mathrm{TOF}} \to \infty$ (infinitely low time resolution) we simply get

$$w_i^\infty(\delta_{ij}; \delta) \equiv w_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}} \to \infty) = \begin{cases} 1, & \delta_{ij}, \delta \in [-T,\, T], \\ 0, & \text{otherwise}, \end{cases} \tag{2.30}$$

since all outcomes within the given coincidence time window are equally probable in this case (so that (2.27) approaches the non-TOF limit, $c_{ij,\delta}(\sigma_{\mathrm{TOF}} \to \infty) \to c_{ij}$, as it should).

In the general case of finite time resolution, the response characteristic of the coincidence measurement is adequately described by a Gaussian (truncated at the borders of the used coincidence time window) with standard deviation $\sigma_{\mathrm{TOF}}$. By integrating this "TOF-kernel" over the considered voxel $j$ we obtain the probability density of detecting the photon pair emitted from this voxel with TOF-difference $\delta$. Dividing this probability by the probability density function of the uniform (non-TOF) distribution we get

$$w_i(\delta_{ij};\delta,\sigma_{\mathrm{TOF}}) = \begin{cases} A(\delta,\sigma_{\mathrm{TOF}}) \displaystyle\int_{\delta_{ij}-\Delta T_i/2}^{\delta_{ij}+\Delta T_i/2} \exp\left[-\frac{(\delta'-\delta)^2}{2\,{\sigma_{\mathrm{TOF}}}^2}\right] d\delta', & \delta_{ij},\delta \in [-T,\,T], \\ 0, & \text{otherwise}, \end{cases} \tag{2.31}$$

where $A(\delta,\sigma_{\mathrm{TOF}})$ is a scaling factor that ensures that the normalization (2.28) is fulfilled. Since for real-world PET systems $\sigma_{\mathrm{TOF}}$ is about 1-2 orders of magnitude larger than the typical voxel size, variations in the integrand are in fact negligible and (2.31) approaches

$$\widetilde{w}_i(\delta_{ij};\delta,\sigma_{\mathrm{TOF}}) = \begin{cases} \widetilde{A}(\delta,\sigma_{\mathrm{TOF}}) \exp\left[-\frac{(\delta_{ij}-\delta)^2}{2\,{\sigma_{\mathrm{TOF}}}^2}\right], & \delta_{ij},\delta \in [-T,\,T], \\ 0, & \text{otherwise}, \end{cases} \tag{2.32a}$$

$$\widetilde{A}(\delta,\sigma_{\mathrm{TOF}}) = 2\,T \left\{ \int_{-T}^{T} \exp\left[-\frac{(\delta'-\delta)^2}{2\,{\sigma_{\mathrm{TOF}}}^2}\right] d\delta' \right\}^{-1}. \tag{2.32b}$$

Note that the presented formalism relies on the assumption that the TOF information is discretized using sufficiently small bins so that discretization errors are negligible and the use of continuous functions is permissible. Considering specifically the Philips Ingenuity PET/MR, this assumption is fulfilled: this system uses a TOF bin size of 25 ps which is more than 20 times smaller than actual time resolution (see Section 2.1).

## 2.4 TOF scatter correction

While the random coincidences contribution in (2.25) remains unaltered in comparison to the non-TOF case $r_{i,\delta} \equiv r_i$ (since the randoms exhibit a uniform time distribution (Conti et al., 2005)), this is not the case for scatter events. In order to account for the non-uniform time distribution of the scatter we express the TOF scatter estimate $s_{i,\delta}(\sigma_{\mathrm{TOF}})$ for LOR $i$ as a product of the non-TOF scatter value $s_i$ and TOF weight factor $W_i(\delta,\sigma_{\mathrm{TOF}})$

$$s_{i,\delta}(\sigma_{\mathrm{TOF}}) = s_i\, W_i(\delta,\sigma_{\mathrm{TOF}}). \tag{2.33}$$

$W_i(\delta, \sigma_{\mathrm{TOF}})$ thus represents the integral effect of TOF-weighting on the whole scatter estimate in the LOR. For calculation of $W_i(\delta, \sigma_{\mathrm{TOF}})$ we use a two-step procedure. First, we calculate "true" scatter time distributions which correspond to an "ideal" scanner with infinite time resolution. For practical reasons, we store a discrete approximation $m_{i\tau}$ of this distribution using a sufficiently high number $N_\tau$ of TOF bins and normalize it as follows

$$\sum_{\tau=1}^{N_\tau} m_{i\tau} = 1. \tag{2.34}$$

The required $W_i(\delta, \sigma_{\mathrm{TOF}})$ coefficients are then computed for each event individually during the MLEM update as a scaled and weighted sum of the $m_{i\tau}$

$$W_i(\delta, \sigma_{\mathrm{TOF}}) = N_\tau \, B(\delta, \sigma_{\mathrm{TOF}}) \sum_{\tau=1}^{N_\tau} m_{i\tau} \int\limits_{\Delta_\tau} \exp\left[-\frac{(\delta' - \delta)^2}{2\,{\sigma_{\mathrm{TOF}}}^2}\right] d\delta' \,, \tag{2.35a}$$

$$B(\delta, \sigma_{\mathrm{TOF}}) = \left\{ \sum_{\tau=1}^{N_\tau} \int\limits_{\Delta_\tau} \exp\left[-\frac{(\delta' - \delta)^2}{2\,{\sigma_{\mathrm{TOF}}}^2}\right] d\delta' \right\}^{-1} , \tag{2.35b}$$

where $B(\delta, \sigma_{\mathrm{TOF}})$ is a further normalization constant. The integrals in (2.35a) and (2.35b) extend over the time intervals $\Delta_\tau$ corresponding to the individual time bins $\tau$. The sum of these integrals in (2.35b) is thus equal to the integral over all possible TOF time differences along the whole LOR. Equation (2.35a) serves as our model to describe the influence of finite time resolution being a convolution of a truncated Gaussian kernel and the discretized scatter TOF-distribution $m_{i\tau}$ with a proper normalization enforced by (2.33) and (2.34). Alternatively, the equation (2.35a) can be interpreted as a "forward projection" of the scatter time distribution defined by $m_{i\tau}$ with weights (2.31) computed for TOF-bins instead of voxels. Obviously, in the non-TOF limit $W_i(\delta, \sigma_{\mathrm{TOF}} \to \infty) \equiv 1$ whereas for infinitely increasing time resolution $W_i(\delta, \sigma_{\mathrm{TOF}} \to 0)$ becomes asymptotically identical to the product of the respective $m_{i\tau}$ and $N_\tau$.

The crucial difference between the different TOF-SC methods considered in this work is the way they estimate the scatter TOF distribution $m_{i\tau}$. The most naive approach is to assume a uniform distribution of the scatter over all time bins which enables the use of the non-TOF SSS estimate within the TOF reconstruction. TOF-SSS estimates $s_i$ and $m_{i\tau}$ simultaneously and serves as our reference method. The newly proposed ISA method combines a non-TOF SSS–based computation of the scatter distribution $s_i$ with a dedicated fast algorithm to estimate the scatter time distribution $m_{i\tau}$. We now describe these different approaches in more detail.

### 2.4.1 Non-TOF SSS

Formally, SSS yields scatter within the considered LOR after integration over all photon arrival time differences, i.e. it does not provide the probability distribution of the scatter over the different TOF bins. Nevertheless, one might consider using SSS in a TOF re-

construction by assuming a uniform distribution of the scatter over all TOF bins. This is equivalent to assuming that all TOF weights $W_i(\delta, \sigma_{\mathrm{TOF}})$ are equal to 1. We have included this hybrid approach in our investigation to evaluate the importance of adequately modeling the time structure of the scatter.

### 2.4.2 TOF-SSS

Unlike the original SSS, the time-of-flight extension of the algorithm discriminates between simulated scatter events according to their point of original photon pair emission $E$. This allows to calculate the corresponding TOF difference (we are using units such that the speed of light $c = 1$) $\delta = R_{EA} - (R_{ES} + R_{SB})$ at which they would be registered ($R_{EA}$, $R_{ES}$, and $R_{SB}$ are the lengths of the lines $EA$, $ES$, and $SB$, respectively, see Fig. 2.6). In order to achieve this goal, a new TOF detection efficiency function $\epsilon_\tau(\delta)$ is introduced which yields the probability of detecting an event with a TOF difference $\delta$ within a TOF bin $\tau$. Using this function, one can represent the amount of scatter registered within TOF bin $\tau$ as

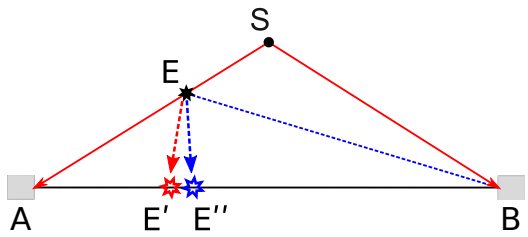


Figure 2.6: *Illustration of ISA and TOF-SSS. A, B: detectors, S: scatter point, E: emission point. E′ is the corresponding apparent source position as resulting from TOF-SSS while E″ is the position resulting from ISA. Note that the difference between E′ and E″ is small despite considering an LOR with a very large radial offset from E and scatter at a rather distant point.*

$$s_{i,\tau}^{\mathrm{TOF-SSS}} = \int_{V_S} dV_S \left( \frac{\sigma_{AS}\,\sigma_{BS}}{4\pi\, R_{AS}^2 R_{BS}^2} \right) \frac{\mu_S}{\sigma_c} \frac{d\sigma_c}{d\Omega} \left[ I_\tau^A + I_\tau^B \right], \tag{2.36a}$$

$$I_\tau^A = \epsilon_{AS}\,\epsilon'_{BS} \exp\left[ -\int_S^A \mu(s)\,ds - \int_S^B \mu'(s)\,ds \right] \times \int_S^A \epsilon_\tau[R_{AS} - R_{BS} - 2s]\,\lambda(s)\,ds\,, \tag{2.36b}$$

$$I_\tau^B = \epsilon'_{AS}\,\epsilon_{BS} \exp\left[ -\int_S^A \mu'(s)\,ds - \int_S^B \mu(s)\,ds \right] \times \int_S^B \epsilon_\tau[R_{AS} - R_{BS} + 2s]\,\lambda(s)\,ds\,. \tag{2.36c}$$

The notation follows (2.12) and (2.14). Note that $I_\tau^A$ and $I_\tau^B$ are no longer scalars but arrays representing two halves of the time distribution of scatter events originating somewhere on the lines $AS$ and $BS$, respectively. For non-TOF SSS, significant acceleration was achieved through precalculating and storing line integrals of attenuation and activity for each pair of scatter point and detector. In TOF-SSS, however, activity integrals are functions of TOF bin $\tau$ and therefore have to be stored accordingly.

Another complication arises from the fact that TOF detection efficiency depends on $(R_{AS} - R_{BS})$ which is the TOF difference of the point $S$ considering LOR $AB$. This translates into an LOR-dependent shift of the activity TOF integrals. This shift cannot

be precalculated since the same cached activity distributions are used for estimating scatter in multiple LORs. Therefore, the unshifted integrals

$$\int_S^A \epsilon_\tau(2s)\,\lambda(s)\,ds \tag{2.37}$$

are precomputed and the required TOF shifts are applied later during the combination of the precomputed distributions for each scatter point and LOR.

Since in our implementation the finite time resolution of the coincidence measurement is handled separately in (2.35a), we use perfect TOF bin separation here:

$$\epsilon_\tau(\delta) = \begin{cases} 1, & \delta \in \Delta_\tau, \\ 0, & \text{otherwise.} \end{cases} \tag{2.38}$$

The spatial and temporal scatter distributions $s_i^{\text{SSS}}$ and $m_{i,\tau}$ can then be separated and are given by

$$m_{i,\tau} = \frac{s_{i,\tau}^{\text{TOF-SSS}}}{s_i^{\text{SSS}}}, \tag{2.39}$$

$$s_i^{\text{SSS}} = \sum_\tau s_{i,\tau}^{\text{TOF-SSS}}. \tag{2.40}$$

Altogether, the TOF-SSS workflow can be described as follows

1. Generate a set of scatter points $S$
2. Compute attenuation line integrals and binned activity distributions for each pair of detector $A$ and scatter point $S$
3. For each LOR $AB$ and scatter point $S$:
   (a) Scale precomputed activity TOF distributions along the lines $AS$ and $BS$ according to (2.36b) and (2.36c)
   (b) Combine them and impose offset $(R_{AS} - R_{BS})$
   (c) Multiply the resulting TOF distribution by scatter point weighting factors according to (2.36a) and merge the result with the current scatter TOF distribution estimate for the LOR $AB$
4. Interpolate the obtained scatter distribution $s_{i,\tau}^{\text{TOF-SSS}}$ as described in Section 2.2.3.
5. Separate scatter spatial distribution $s_i$ and temporal distribution $m_{i,\tau}$
6. Scale $s_i$ to fit the scatter tails as described in Section 2.7.1

The LOR undersampling factor and scatter point locations were the same as those for SSS. Note that steps 2 and 3 involve manipulations on length $N_\tau$ arrays which results in a considerable slowdown of the scatter correction compared to SSS.

### 2.4.3 Immediate Scatter Approximation (ISA)

Scatter correction via the TOF-SSS approach described in the preceding section is very time-consuming and slows down the image reconstruction considerably. Therefore, as part of the present work a time-efficient approximation of TOF-SSS has been developed. In doing so, the objective was to model with sufficient accuracy the dependency of scatter TOF profiles on activity distribution (while neglecting the additional influence of spatial variations of the attenuation coefficient). The relevant geometry is shown in Fig. 2.6. The method utilizes the fact that photon pairs emitted in point $E$ and detected in LOR $AB$ after a single scatter at point $S$ will, in most cases, exhibit an arrival time difference close to $(R_{EA} - R_{EB})$. For illustration, Figures 2.7 and 2.8 demonstrate this behavior for a 2D acquisition of a point source in the center of a cylindrical water phantom. In this example, the TOF difference should thus be close to zero for most events. The simulation underlying these plots scans a sufficiently fine grid of scatter locations and scatter angles, computing for each grid point the resulting LOR and TOF difference as well the relative scatter and detection probability in the given energy window. As can be seen in Fig. 2.7, the probability of detecting a scattered event rapidly decreases with increasing LOR distance as well as TOF difference. This behavior arises due to a combination of two factors. First, the actual arrival time difference is $(R_{EA} - (R_{ES} + R_{SB}))$ and for LORs with a sufficiently small orthogonal distance from $E$ (corresponding to small angle forward scattering) we have $R_{ES} + R_{SB} \approx R_{EB}$ even for relatively distant scatter points $S$ (for which $R_{ES}$ is not small compared to $R_{EB}$). Second, although for more distant LORs the dependency of measured TOF information on the distance $R_{ES}$ becomes more prominent, the relative contribution of the considered emission point to an LOR rapidly decreases with increasing distance of the LOR from $E$. The reason for this is lower scatter probability at large angles in combination with lower detection/acceptance probability of the scattered photon by $B$. The latter is a consequence of the reduced energy of the scattered photon which reduces overlap of the detector signal with the given energy window of 460–665 keV for the considered scanner. Therefore, for the overwhelming number of events contributing to a given LOR the TOF difference of the registered scatter events essentially only depends on the position of the emission point relative to the LOR while the dependence on the actual scatter point position might be neglected.

To a lesser extent, approximate independence of scatter point location is also valid for the scatter angle: for most events the actual scatter angle is not much larger than for a scatter occurring immediately after the emission at $E$. So to a rough approximation one can compute the scatter angle for that limiting case only and neglect here, too, the dependence on actual scatter position.

Altogether, we surmise that it should be sufficient to estimate scatter profiles $m_{i\tau}$ based on the assumption that for all scatter events only one of the photons scatters and does so immediately after emission. This is the key assumption of our proposed *Immediate Scatter Approximation* (ISA) method. Our ISA implementation models the object as a

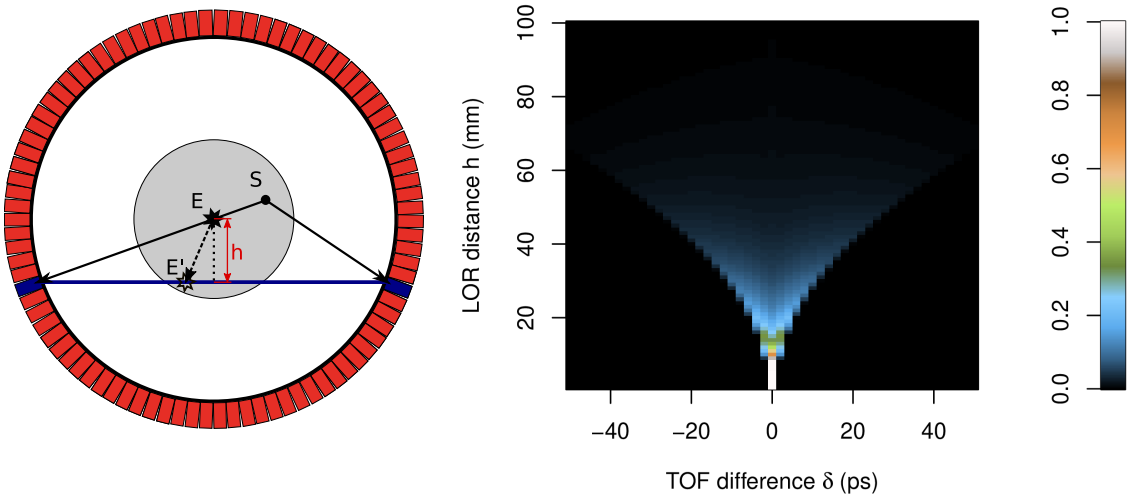


Figure 2.7: *Illustration of the TOF difference vs. the LOR distance distribution of the scattered events produced by a point source located at the center of a cylindrical water phantom of 40 cm diameter centered in the field of view. The chosen scanner geometry and the energy acquisition window are that of the Philips Ingenuity PET/MR. The 2D case is considered. The influence of energy-dependent attenuation (different before and after scatter), anisotropic scatter probability, energy-dependent detection probability, and position-dependent solid angles are included while multiple scatter was neglected. The plot shows the relative probability of detecting a scattered event at TOF difference $\delta$ in an LOR with radial offset $h$. The distribution rapidly approaches zero with an increasing LOR distance and TOF difference.*

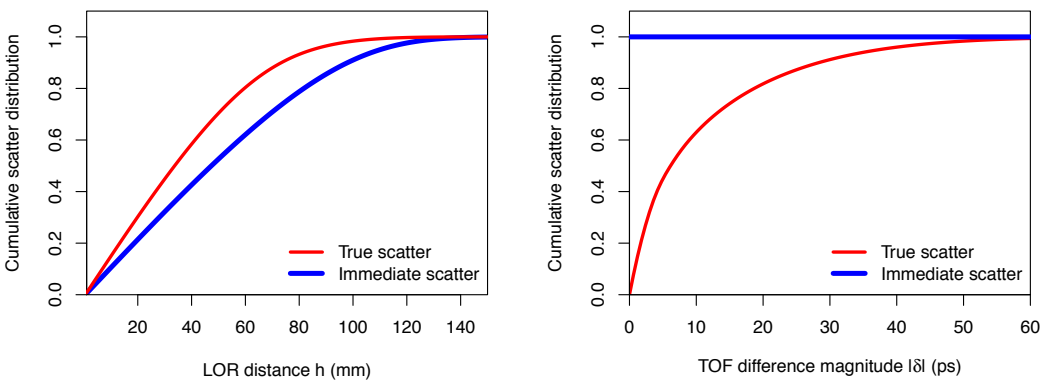


Figure 2.8: *Cumulative distribution functions (red curves) of LOR distance (left) and TOF difference (right) obtained by integrating the probability density in Fig. 2.7 along the corresponding axis. Shown in blue are the predictions resulting from ISA for this configuration (zero TOF difference for all scattered events, modest overemphasis of larger LOR distances). Note the small range of actually occurring TOF differences (60 ps corresponding to 9 mm shift in apparent event position along the LOR relative to ISA prediction).*

relatively small set of emission points $E$ with activities $I_E$ taken from the current image estimate. The locations of the emission points are chosen to coincide with those of the

scatter points used for SSS according to (Watson, 2000).

For each LOR $AB$ and emission point $E$, the arrival time difference of any event originating from that point is estimated as $\delta_{i,E} = R_{EA} - R_{EB}$ (neglecting the residual scatter point dependency as explained above). By repeating this procedure for all emission points one arrives at the scatter time distribution

$$m_{i\tau} = N_i \sum_{\substack{E,\\ \delta_{i,E}\in\Delta_\tau}} I_E \, \frac{d\sigma_c}{d\Omega}(\theta)\, \epsilon(\theta)\, \Delta\Omega_{EA}\, \Delta\Omega_{EB}\,, \tag{2.41}$$

where $\theta$ is scattering angle (for scatter happening directly at $E$), $[d\sigma_c/d\Omega](\theta)$ is the differential Compton cross section at 511 keV, and $\epsilon(\theta)$ is the energy-dependent detection probability for the scattered photon. $\Delta\Omega_{EA}$ and $\Delta\Omega_{EB}$ are the solid angles extended by both detectors at $E$. $N_i$ is a normalization coefficient ensuring satisfaction of (2.34). Note that the quantities $[d\sigma_c/d\Omega](\theta)$, $\epsilon(\theta)$, $\Delta\Omega_{EA}$, $\Delta\Omega_{EB}$ in (2.41) are already required for SSS and can, therefore, be recycled by ISA if the emission points are chosen to coincide with the SSS scatter points.

## 2.5 Emission-based attenuation correction

As explained in Section 1.3.3, attenuation correction is not straightforward in PET/MR. Furthermore, only a few of the available options are suitable for whole-body investigations. The problem arises from large anatomical inter-subject variability which poses a major challenge for algorithms relying on MR-based tissue classification. This problem might be avoided by solely utilizing the PET information for attenuation correction. Such an emission-based attenuation correction algorithm, called Maximum Likelihood reconstruction of Attenuation and Activity (MLAA), was first presented in (Nuyts et al., 1999) for non-TOF PET and, subsequently, an improved version of it appeared in (Salomon et al., 2011) and (Rezaei et al., 2012). Recent implementations utilize the available TOF information in order to reduce the cross-talk between reconstructed attenuation and activity (Defrise et al., 2012) which was a major problem in the initial implementation.

The key idea of the method is to maximize the likelihood function with respect to, both, activity and attenuation image. The maximization is performed by changing activity and attenuation maps alternately with separate maximum likelihood algorithms for emission and transmission tomography. Common choices here are TOF-MLEM, described in detail in Section 2.3.1, and Maximum Likelihood for Transmission tomography (MLTR) designed initially for CT reconstructions (Nuyts et al., 1998). The original algorithm design assumes sinogram-based implementation and, therefore, is not directly applicable for list-mode reconstruction. List-mode versions of MLTR were presented in (Mollet & Vandenberghe, 2014) and (Rezaei et al., 2015) but the additive contributions, i.e. scatter and randoms, were ignored in both cases. List-mode MLAA with all necessary corrections was used in (Cheng et al., 2016b) but no details of the algorithm were given. We, therefore, found

it necessary to derive a fully corrected list-mode MLTR algorithm from scratch in order to incorporate it into our LM-MLAA implementation. This derivation is presented in the following section.

### 2.5.1 Maximum likelihood for transmission tomography

In our derivation, we are following the approach described in (Van Slambrouck & Nuyts, 2014) and (Rezaei et al., 2016). The proof of convergence of the algorithm will be given under the assumption that scatter and random contributions can be neglected.

In analogy to (1.25), it can be shown that the log-likelihood function of TOF list-mode emission data is given by

$$l(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_e \log \overline{y}_{i_e,\delta_e} - \sum_i \overline{y}_i \,, \tag{2.42}$$

where non-TOF and TOF forward projectors $\overline{y}_i$ and $\overline{y}_{i,\delta}$ are defined in (2.24) and (2.26), respectively. For the sake of convenience, we omit an explicit time resolution specification and introduce a separate notation for "pure" forward projectors $\psi$ without scatter and random corrections

$$\psi_i = a_i \, n_i \sum_j c_{ij} \, \lambda_j, \tag{2.43a}$$

$$\psi_{i,\delta} = a_i \, n_i \sum_j c_{ij,\delta} \, \lambda_j, \tag{2.43b}$$

where all notations are shared with (2.24) and (2.26). The fully corrected forward projectors can then be rewritten in the new notation as

$$\overline{y}_i = \psi_i + s_i + r_i, \tag{2.44a}$$

$$\overline{y}_{i,\delta} = \psi_{i,\delta} + s_{i,\delta} + r_{i,\delta}. \tag{2.44b}$$

MLTR assumes the likelihood to be only a function of $\boldsymbol{\mu}$. The optimization problem of finding the $\boldsymbol{\mu}$ value which maximizes the likelihood is solved by employing the strategy of maximizing a suitable separable surrogate function instead of the likelihood itself. Using the quadratic approximation of the Taylor expansion, we get

$$l(\boldsymbol{\lambda}, \boldsymbol{\mu} + \delta\boldsymbol{\mu}) \simeq \tilde{l}(\boldsymbol{\lambda}, \boldsymbol{\mu} + \delta\boldsymbol{\mu}) = \\ l(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_j \frac{\partial \, l}{\partial \mu_j}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \, \delta\mu_j + \frac{1}{2} \sum_{j,\,k} \frac{\partial^2 \, l}{\partial \mu_j \partial \mu_k}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \, \delta\mu_j \, \delta\mu_k, \tag{2.45}$$

where $\delta\boldsymbol{\mu}$ is some small increment of $\boldsymbol{\mu}$. Taking into account that the second derivatives of the likelihood are all negative and

$$2 \, \delta\mu_j \, \delta\mu_k \leq \delta\mu_j^2 + \delta\mu_k^2 \tag{2.46}$$

we obtain the surrogate $S(\boldsymbol{\lambda}, \boldsymbol{\mu}, \delta\boldsymbol{\mu})$

$$\tilde{l}(\boldsymbol{\lambda}, \boldsymbol{\mu} + \delta\boldsymbol{\mu}) \geq S(\boldsymbol{\lambda}, \boldsymbol{\mu}, \delta\boldsymbol{\mu}) = l(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_j \frac{\partial\, l}{\partial \mu_j}(\boldsymbol{\lambda}, \boldsymbol{\mu})\, \delta\mu_j + \frac{1}{4}\sum_{j,k} \frac{\partial^2\, l}{\partial \mu_j\, \partial \mu_k}(\boldsymbol{\lambda}, \boldsymbol{\mu})\, \delta\mu_j^2 + \frac{1}{4}\sum_{j,k} \frac{\partial^2\, l}{\partial \mu_j\, \partial \mu_k}(\boldsymbol{\lambda}, \boldsymbol{\mu})\, \delta\mu_k^2. \quad (2.47)$$

Taking into account the symmetry of the second derivative, one can notice that the last two sums in (2.47) are identical and, therefore,

$$S(\boldsymbol{\lambda}, \boldsymbol{\mu}, \delta\boldsymbol{\mu}) \;=\; l(\boldsymbol{\lambda}, \boldsymbol{\mu}) \;+\; \sum_j \frac{\partial\, l}{\partial \mu_j}(\boldsymbol{\lambda}, \boldsymbol{\mu})\, \delta\mu_j \;+\; \frac{1}{2}\sum_j \left(\sum_k \frac{\partial^2\, l}{\partial \mu_j\, \partial \mu_k}(\boldsymbol{\lambda}, \boldsymbol{\mu})\right) \delta\mu_j^2. \quad (2.48)$$

The MLTR update can then be defined as a step of $S(\boldsymbol{\lambda}, \boldsymbol{\mu}, \delta\boldsymbol{\mu})$ maximization

$$\delta\boldsymbol{\mu}^{(p)} = \arg\max_{\delta\boldsymbol{\mu}} S(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)}, \delta\boldsymbol{\mu}), \quad (2.49a)$$

$$\boldsymbol{\mu}^{(p+1)} = \boldsymbol{\mu}^{(p)} + \delta\boldsymbol{\mu}^{(p)}. \quad (2.49b)$$

It is easy to see that the sequence $\boldsymbol{\mu}^0$, $\boldsymbol{\mu}^1$, ... increases $S(\boldsymbol{\lambda}, \boldsymbol{\mu}, 0)$ monotonically, i.e. $S(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p+1)}, 0) \geq S(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)}, 0)$, for all $p$, since

$$S(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p+1)}, 0) = l(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p+1)}) \simeq \tilde{l}(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)} + \delta\boldsymbol{\mu}^{(p)}) \geq S(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)}, \delta\boldsymbol{\mu}^{(p)}) \geq S(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)}, 0), \quad (2.50)$$

as follows from (2.47), (2.49a) and (2.49b). Monotonicity and boundedness of the sequence $\{S(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)}, 0)\}$ implies its convergence to some $\overline{S} = S(\boldsymbol{\lambda}, \overline{\boldsymbol{\mu}}, 0)$ with $\{\boldsymbol{\mu}^{(p)}\}$ converging to some $\overline{\boldsymbol{\mu}}$. Substituting (2.48) to (2.49a) we obtain the following expression for $\delta\boldsymbol{\mu}^{(p)}$

$$\delta\mu_j^{(p)} = -\frac{\dfrac{\partial\, l}{\partial \mu_j}(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)})}{\displaystyle\sum_k \frac{\partial^2\, l}{\partial \mu_k\, \partial \mu_j}(\boldsymbol{\lambda}, \boldsymbol{\mu}^{(p)})}. \quad (2.51)$$

Convergence of $\{\boldsymbol{\mu}^{(p)}\}$ demands convergence of $\{\delta\boldsymbol{\mu}^{(p)}\}$ to zero. Therefore, it follows from (2.51) that

$$\frac{\partial\, l}{\partial \mu_j}(\boldsymbol{\lambda}, \overline{\boldsymbol{\mu}}) = 0 \quad (2.52)$$

for all $j$ at the convergence point $\overline{\boldsymbol{\mu}}$. Thus, the likelihood as a function of $\boldsymbol{\mu}$ converges to its maximum with sequence $\{\delta\boldsymbol{\mu}^{(p)}\}$ and the concavity of the likelihood guarantees that the maximum is global.

The likelihood derivatives are straightforward to calculate by substituting the exact

form of attenuation factors (1.30) into (2.42) and performing direct differentiation

$$\frac{\partial\, a_i}{\partial \mu_j} = \frac{\partial}{\partial \mu_j} \exp\left(-\sum_k l_{ik}\, \mu_k\right) = -\, l_{ij}\, a_i, \tag{2.53a}$$

$$\frac{\partial\, l}{\partial \mu_j} = \sum_i l_{ij}\, \psi_i - \sum_e l_{i_e j}\, \frac{\psi_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}}, \tag{2.53b}$$

$$\frac{\partial^2\, l}{\partial \mu_k\, \partial \mu_j} = \sum_e l_{i_e k}\, l_{i_e j} \left[\frac{\psi_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}} - \left(\frac{\psi_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}}\right)^2\right] - \sum_i l_{ik} l_{ij}\, \psi_i, \tag{2.53c}$$

where all notations are the same as in Section 2.2.2 and Section 2.3.1. Combining (2.49b), (2.51), and (2.53), we finally get an expression for the TOF LM-MLTR update

$$\mu_j^{(p+1)} = \mu_j^{(p)} + \frac{\sum\limits_e l_{i_e j}\, \dfrac{\psi_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}} - \sum\limits_i l_{ij}\, \psi_i}{\sum\limits_e l_{i_e j} \left[\dfrac{\psi_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}} - \left(\dfrac{\psi_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}}\right)^2\right] \overline{L}_{i_e} - \sum\limits_i l_{ij}\, \psi_i\, \overline{L}_i}\,, \tag{2.54}$$

where $\overline{L}_i = \sum_j l_{ij}$ is the length of intersection of LOR $i$ with the image. The ratio $\psi_{i_e,\delta_e}/\overline{y}_{i_e,\delta_e}$ in this formula can be interpreted as an estimated probability for the event $e$ registered in LOR $i_e$ with TOF difference $\delta_e$ to be a true event.

Formula (2.51) shows that MLTR is a variation of the (scaled) gradient ascent optimization algorithm. The first derivative of the likelihood defines the algorithm's step "direction" and the second derivative (always negative) only modifies the step "size" which affects the convergence speed. In gradient ascent algorithms, in general, the step size is allowed to be selected rather arbitrarily as long as it does not prevent convergence and, therefore, one can multiply $\delta\mu_j^{(p)}$ by some factor $\alpha^{(p)} > 0$ to improve the algorithm performance. Furthermore, the convergence properties can be modified even more drastically as described in (Van Slambrouck & Nuyts, 2014), leading either to a "patchwork" MLTR which modifies only some subset of the voxels per iteration or to the so-called convex algorithm ensuring faster convergence in high-attenuating regions (Lange & Fessler, 1995).

We have started our derivation with the likelihood of TOF data (2.42). Therefore, algorithms (2.25) and (2.54) optimize the same likelihood function by design. However, it is widely believed (Rezaei et al., 2016) that the available TOF information can be ignored in the attenuation estimation step of the joint reconstruction and the non-TOF MLTR algorithm can be used in MLAA instead. Non-TOF MLTR can be derived directly from the non-TOF likelihood (1.24) or it can be obtained from (2.54) by assuming that the ratios $\psi_{i,\delta}/\overline{y}_{i,\delta}$ are independent of TOF difference $\delta$ (Rezaei et al., 2016)

$$\frac{\psi_{i,\delta}}{\overline{y}_{i,\delta}} \approx \frac{\psi_i}{\overline{y}_i}\,. \tag{2.55}$$

By substituting (2.55) into (2.54) we get the formula for a non-TOF LM-MLTR update

$$\mu_j^{(p+1)} = \mu_j^{(p)} + \frac{\sum\limits_e l_{i_e j} \, \dfrac{\psi_{i_e}}{\overline{y}_{i_e}} - \sum\limits_i l_{ij} \, \psi_i}{\sum\limits_e l_{i_e j} \left[ \dfrac{\psi_{i_e}}{\overline{y}_{i_e}} - \left( \dfrac{\psi_{i_e}}{\overline{y}_{i_e}} \right)^2 \right] \overline{L}_{i_e} - \sum\limits_i l_{ij} \, \psi_i \, \overline{L}_i} \, . \tag{2.56}$$

Since TOF information is ignored in (2.56), it is easy to group all the events by LOR and take advantage of computing the factors $\psi_i/\overline{y}_i$ only once per LOR. The resulting formula is a direct analogue of (2.56), suitable for a sinogram-based implementation

$$\mu_j^{(p+1)} = \mu_j^{(p)} + \frac{\sum\limits_i l_{ij} \, \psi_i \left( \dfrac{y_i}{\overline{y}_i} - 1 \right)}{\sum\limits_i l_{ij} \, \psi_i \left( \dfrac{y_i}{\overline{y}_i} - 1 - \dfrac{y_i \, \psi_i}{\overline{y}_i^2} \right) \overline{L}_i} \, . \tag{2.57}$$

Usually, formula (2.57) is further simplified by assuming that $y_i/\overline{y_i} \approx 1$ in the denominator, allowing to exclude the measured data from the denominator completely (Rezaei et al., 2016). As discussed before, changes in the denominator do not affect the convergence point $\overline{\boldsymbol{\mu}}$ of the algorithm but only alter the convergence speed. We, therefore, do not perform the aforementioned MLTR modification due to the negligible impact on algorithm performance and results. Assumption (2.55), on the other hand, introduces major changes in the numerator of the MLTR formula which in theory can alter the solution towards which the algorithm converges. However, so far the differences between TOF and non-TOF MLTR in the context of MLAA reconstruction have never been thoroughly investigated. For this reason, we have implemented both versions of MLTR, (2.54) and (2.56), and integrated them in THOR for further analysis.

LM-MLTR can be accelerated with an ordered subsets approach by splitting the sums over all events in (2.54) and (2.56) into sums over event subsets $D_i$, $i = 1 \ldots M$, as explained in Section 2.2.1. Accordingly, the sum over all LORs has to be reduced to a sum over LOR subset $L_i$ corresponding to event subset $D_i$. We used the same subsets definition for MLEM and MLTR in our implementation.

### 2.5.2 Maximum likelihood reconstruction of attenuation and activity

As explained above, the MLAA algorithm maximizes the likelihood function by alternating updates of activity and attenuation images via MLEM and MLTR algorithms, respectively. The general MLAA workflow is illustrated in Fig. 2.9. The concrete iteration scheme does vary in different implementations. There are two main possibilities:

1. Interleaving updates

2. Interleaving reconstructions

The first option implies alternating updates of the activity and attenuation images during

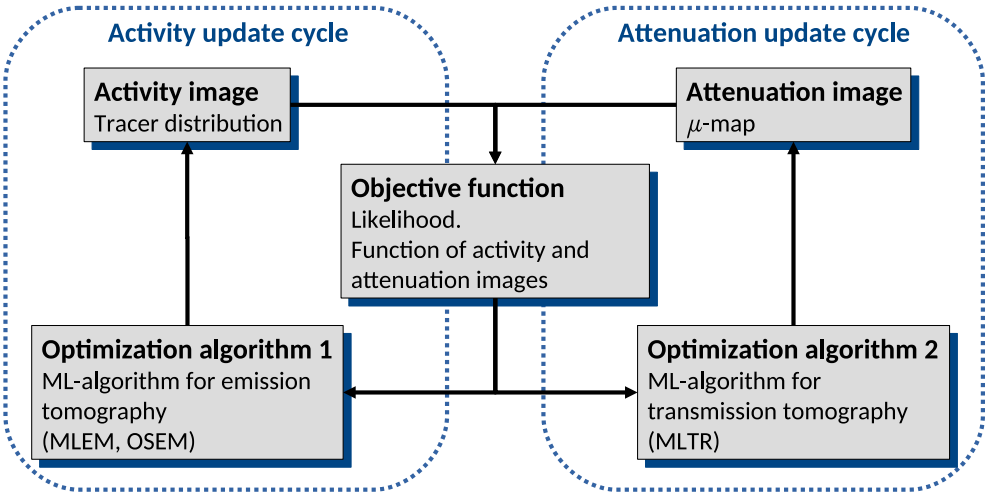


Figure 2.9: *General MLAA algorithm workflow.*

a single reconstruction where one MLEM update is immediately followed by several MLTR updates (MLTR converges slower than MLEM), see Fig. 2.10. Variations of this approach were analyzed and discussed in detail in (Presotto et al., 2015), concluding that at least four MLTR updates should be applied per single MLEM update.

The other iteration scheme uses multiple consecutive MLEM reconstructions of the same dataset separated by several MLTR updates, see Fig. 2.11. In this case, each MLEM reconstruction is performed with a single attenuation map which is only modified between the reconstructions.

In general, the first approach is preferable for clinical applications since it requires performing MLEM reconstruction only once which results in a severalfold speed advantage. On the other hand, the second approach, being impractical in a clinical routine, is better suited for investigation of MLAA capabilities and limitations. There are several reasons for this. First of all, MLTR updates, in this case, use a fully corrected and converged activity image as well as accurate scatter estimate (scatter is updated 2–3 times during the MLEM reconstruction). Moreover, scatter is being recalculated in every MLEM cycle based on the latest available attenuation map. Another advantage of the interleaved reconstruction

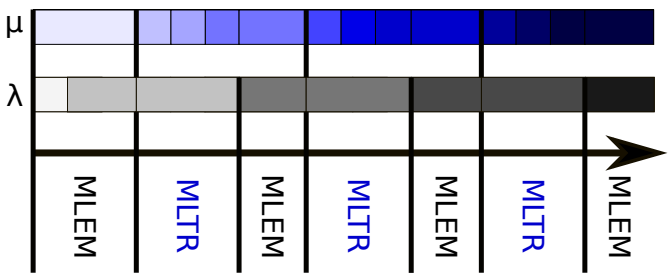


Figure 2.10: *MLAA algorithm with interleaved updates. Horizontal bars represent μ-map (top) and activity image (bottom) estimated during the reconstruction. Different colors represent changes in the estimated images and darker colors correspond approach to convergence.*

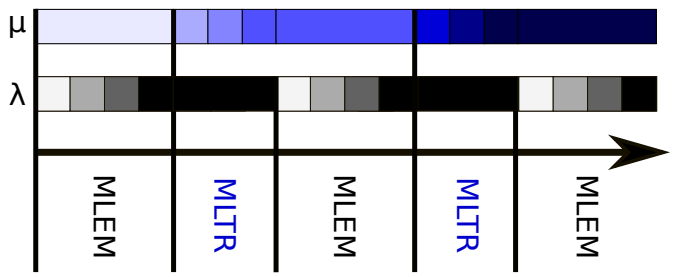


Figure 2.11: *MLAA algorithm with interleaved reconstructions. Horizontal bars represent $\mu$-map (top) and activity image (bottom) estimated during the reconstruction. Different colors represent changes in the estimated images and darker colors correspond to an image closer to convergence.*

MLAA scheme is a higher degree of freedom in adjusting MLEM and MLTR individual iteration schemes. It is possible to use the existing optimized reconstruction parameters for MLEM reconstruction and then just select an appropriate number of MLEM cycle repetitions and MLTR iterations in order to ensure overall algorithm convergence. For these reasons, we have implemented MLAA using the interleaved reconstructions scheme.

There are several practical complications one has to handle in order to avoid errors in the reconstructed images and achieve good MLAA performance:

1. **Image scale**

   Even though the utilization of TOF information in MLAA reconstruction allows to reduce local cross-talk between attenuation and activity, the global scale of activity and attenuation images remains undefined. Indeed, if there are two sets of images $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ and $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ such that $\lambda_j = C\,\lambda'_j$ and $a_i = a'_i\,/C$ for some $C > 0$ and all $i$ and $j$, then they will yield the same likelihood values $l(\boldsymbol{\lambda}, \boldsymbol{\mu}) = l(\boldsymbol{\lambda}', \boldsymbol{\mu}')$, see (2.42), and if $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ maximizes the likelihood, then $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ maximizes it too. Therefore, it is crucial to fix the scale of the reconstructed images in order to prevent quantitative errors.

   Within this study, we fixed the attenuation map scale as follows. We assume that soft tissue is a dominant tissue class. In order to find the soft tissue voxels, the denoised attenuation map (using a 12 mm FWHM Gaussian filter) was histogrammed with a bin width of 0.015 cm$^{-1}$ and the most populated $\mu$-range was selected. Our assumption implies that voxels with $\mu$-values within this range should correspond to soft tissue. Therefore, their average attenuation coefficient $\mu_{\text{avg}}$ should be equal to the respective $\mu$-value, $\mu_{\text{soft}} = 0.096$ cm$^{-1}$. This condition was enforced by the global attenuation map rescaling

$$\boldsymbol{\mu}'^{(p)} = \boldsymbol{\mu}^{(p)}\,\frac{\mu_{\text{soft}}}{\mu_{\text{avg}}}, \tag{2.58}$$

   where $\boldsymbol{\mu}^{(p)}$ and $\boldsymbol{\mu}'^{(p)}$ are attenuation maps before and after rescaling, respectively. The procedure is performed after every MLTR update to prevent a drift of the $\mu$-map global scale.

2. **Algorithm initialization**

   Generally speaking, MLAA does not guarantee the convergence to the global maximum of the likelihood as a function of $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ since the existence of local maxima is not excluded (Rezaei et al., 2012). Therefore, it is recommended to use a good attenuation map estimate for algorithm initialization. The common choice here is a homogeneous $\mu$-map generated by determining the body outline from non-attenuation-corrected activity image and assigning the attenuation coefficient of water to the interior (Cheng et al., 2016b). In our implementation, we take advantage of the available MR scan and use a (truncation compensated) MR-based attenuation map as an initial $\mu$-map guess $\boldsymbol{\mu}^{(0)}$.

3. **Missing attenuation data**

   LORs which do not cross the activity support $A$ need to be handled with special care. For these LORs, $\psi_i \approx 0$ and all of the registered events have to be either scatter or randoms and thus contain no actual information about the attenuation factors $a_i$. Therefore, voxels outside of the activity support are intersected by a large number of uninformative LORs resulting in increased noise in the background region and in an increased chance of convergence to some local likelihood maximum instead of the global one (Nuyts et al., 1999). Restricting MLTR to update the $\mu$-map only inside of the activity support is a viable solution to the problem. However, if the activity support was delineated wrongly or if there are unaccounted attenuating objects in the FOV (e.g. a flexible MR coil or headphones), then MLAA will perform suboptimally. Without the possibility to update the restricted attenuation map regions, the algorithm will erroneously predict more attenuating material inside the patient body causing undesirable artifacts.

   Another way of reducing the background noise while still allowing modification of the $\mu$-map outside of the patient support is suggesting to the algorithm the likeliest $\mu$-value instead of enforcing it. *A priori* knowledge regarding the attenuation in the background can be introduced into the algorithm by modifying the likelihood function in a way that prefers solutions with a predefined set of $\mu$-values outside of the activity support (Nuyts et al., 2013; Heußer et al., 2017). Another method was proposed in (Nuyts et al., 1999). It reduces background noise by adding information regarding attenuation in the LORs which do not contain any activity. We will ignore the patient bed and MR hardware for now and will assume that for these LORs the attenuation is close to zero. This assumption enters into the formula (2.57) through modified forward projector and events count

$$y'_i = y_i, \qquad \psi'_i = \psi_i, \qquad \text{if } i \cap A \neq 0, \tag{2.59a}$$

$$y'_i = B + s_i + r_i, \qquad \psi'_i = a_i\, n_i\, B, \qquad \text{otherwise}, \tag{2.59b}$$

   where $B > 0$ is an arbitrary constant specifying the regularization strength and

primed values are the new values for event count and forward projector. We used $B = 1$ in the present work. The proposed modifications are not directly compatible with the list-mode algorithms (2.54) and (2.56). Therefore, we suggest computing the contribution of the "empty" LORs in the list-mode MLTR in the same way as in (2.57).

4. **Erroneous structures in the background**

   It is still possible that even after regularization (2.59) MLAA generates some attenuating structures in the background which do not correspond to any real attenuating objects in the FOV. These attenuation map artifacts do not affect the attenuation correction performance but can interfere with scatter estimation procedures and should be removed. Normally, the $\mu$-values of spurious structures grow slower than the ones corresponding to actual attenuating medium. This observation can, thus, be used for a "cleanup" of the attenuation map: we remove too small ($\mu < h$) $\mu$-values outside of the activity support every second MLTR update, thus suppressing slowly growing attenuating structures in the background

$$\mu_j'^{(p)} = 0, \qquad \text{if } j \notin A,\ \mu_j^{(p)} < h,\ p \bmod 2 = 0, \tag{2.60a}$$

$$\mu_j'^{(p)} = \mu_j^{(p)}, \qquad \text{otherwise}, \tag{2.60b}$$

   where primed quantities designate the values after correction. We used $h = 0.005\ \mathrm{cm}^{-1}$ in the present study.

5. **Negative attenuation values**

   Gradient-ascent algorithms do not guarantee non-negativeness of estimated $\mu$-values and this constraint has to be enforced explicitly. We, therefore, remove all negative $\mu$-values from the attenuation map after every MLTR update

$$\mu_j'^{(p)} = \max\left(\mu_j^{(p)},\ 0\right), \tag{2.61}$$

   where primed quantities designate the values after correction.

6. **High attenuation map noise**

   Attenuation maps estimated from emission data exhibit substantially higher noise levels than those obtained with CT- or MR-based methods. Noise in the attenuation map translates into increased image noise level and should be minimized. This can be achieved through $\mu$-map regularization (Ahn et al., 2012) or by simple smoothing of the $\mu$-map increment image $\delta\boldsymbol{\mu}^{(p)}$. Utilizing the latter approach, we perform a Gaussian filtering (FWHM = 4 mm) of the $\delta\boldsymbol{\mu}^{(p)}$ for every $p$.

7. **Altering of patient bed and MR hardware attenuation**

   The $\mu$-map modifications performed during the MLTR update do also affect the pre-built well-known attenuation maps of the patient bed and MR coils. To revert

the changes made by the algorithm, we restore the bed and coils template in the attenuation map after every MLTR update.

## 2.6 Data acquisition and reconstruction

All data were acquired in list-mode with the Philips Ingenuity PET/MR scanner at the University Hospital Carl Gustav Carus, Dresden. The applied activity measurements were performed with the clinical dose calibrator ISOMED 1000 (Nuklear-Medizintechnik, Dresden, Germany). Dose calibrator and scanner are cross-calibrated quarterly as described in (Maus et al., 2014). The calibration accuracy is count-rate-dependent. It is better than 20% for singles rates $< 40$ Mcps and better than 10% for singles rates $< 10$ Mcps.

### 2.6.1 Phantom data

PET phantoms are dedicated vessels containing suitably shaped solid or hollow fillable inserts to simulate relevant tissues and activity distributions. Since material composition and geometry are known, the attenuation map of the phantom can easily be determined. Activity concentration in each compartment, too, can be accurately measured prior to (and independent of) the PET acquisition. Phantoms thus allow conducting studies with a known ground truth typically unavailable for patient scans. Therefore, phantoms are used in scanner and reconstruction calibration, performance measurements, and quality assurance. The list of phantoms and their configurations as used in the present work is given below.

#### SUV phantom

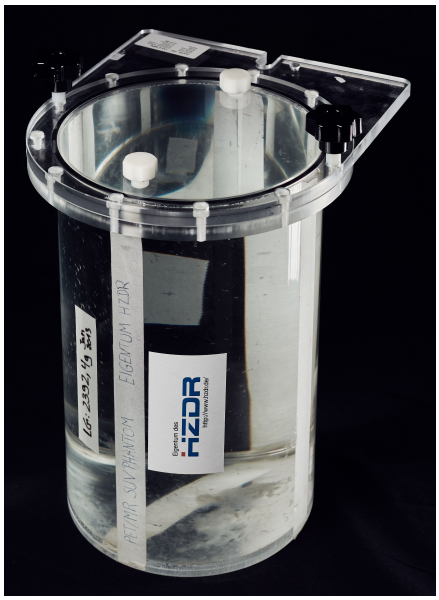


Figure 2.12: *SUV calibration phantom*

The standard phantom supplied with the Ingenuity PET/MR for SUV calibration and validation procedures is shown in Fig. 2.12. The phantom is a hollow acrylic glass cylinder with dimensions

- inner radius: 19.5 cm,
- outer radius: 20 cm,
- height: 31.3 cm,
- volume: 9345 mL.

The phantom is filled with water to which the desired amount of radioactive tracer is added. Our measurements were performed using an initial activity of 180 MBq $^{18}$F-FDG and lasted 6 hours (about 3.3 half-lives of $^{18}$F) which covers a high dynamic count-rate range (count rates change by about a factor of ten during the measurement).

### Cylinder phantom with spherical inserts

The phantom is part of phantom set *L981602* (PTW-Freiburg, Freiburg, Germany) and consists of a hollow acrylic glass cylinder with interchangeable cover that holds six thin-walled glass spheres of different size. The phantom is shown in Fig. 2.13. Its dimensions are

- inner diameter: 19.4 cm,
- outer diameter: 20 cm,
- height: 19 cm,
- volume: 5616 mL.

The sphere dimensions are summarized in Table 2.1.

Figure 2.13: *Cylinder phantom with spherical inserts.*

| Sphere No. | Outer diameter (mm) | Inner diameter (mm) | Volume (mL) |
|---|---|---|---|
| 1 | 12.0 | 9.7 | 0.47 |
| 2 | 14.5 | 12.9 | 1.12 |
| 3 | 18.4 | 16.8 | 2.48 |
| 4 | 23.5 | 21.3 | 5.03 |
| 5 | 29.7 | 27.2 | 10.59 |
| 6 | 39.2 | 37.0 | 26.50 |

Table 2.1: *Dimensions of the six spherical inserts*

The spheres and the background can be filled independently with different activity concentrations to achieve the desired activity levels and image contrasts. For the present work, all spheres were filled with a common activity concentration. Three different configurations were used representing high, medium and low contrasts, respectively, see Table 2.2.

| Measurement | Spheres activity concentration (kBq/mL) | Background activity concentration (kBq/mL) | Contrast | Duration (s) |
|---|---|---|---|---|
| 1 | 69.42 | 3.68 | 18.86 | 600 |
| 2 | 55.60 | 5.39 | 10.31 | 840 |
| 3 | 45.38 | 8.86 | 5.12 | 1080 |

Table 2.2: *Cylinder phantom study specifications*

### Whole-body phantom

This phantom, too, is part of phantom set *L981602* (PTW-Freiburg, Freiburg, Germany) and made of acrylic glass. The geometry is shown in Fig. 2.14. The phantom has a volume of 9650 mL and possesses an interchangeable top cover that enables attachment of different inserts. Three different configurations were used in the present work.

**1. No inserts (homogeneous)**

A nine-hour-long scan was performed with 140 MBq $^{18}$F-FDG activity mixed in the water-filled phantom to investigate the scanner performance over a very large dynamic count-rate range.

**2. Cylindrical inserts**

The phantom in this configuration hosts three cylinders of 5 cm outer diameter, see Fig. 2.15. Two of the cylinders contain air and the last one is made of polytetrafluoroethylene (PTFE, e.g. Teflon™) whose linear attenuation coefficient $\mu = 0.182$ cm$^{-1}$ is close to that of cortical bone ($\mu \approx 0.172$ cm$^{-1}$). The background was filled with water solution of $^{18}$F-FDG with activity of 40 MBq and scanned for ten minutes. This setup was used to evaluate the performance of the attenuation correction algorithms.

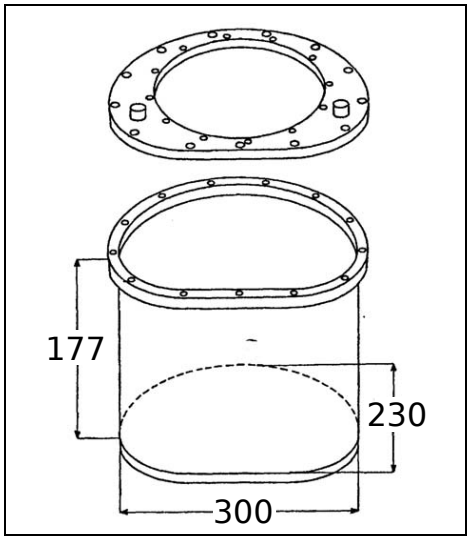


Figure 2.14: *Whole-body phantom schematic view. (Modified after the phantom user manual, PTW-Freiburg)*

**3. Bladder and lesion inserts (pelvis region phantom)**

This configuration models the extreme high contrast conditions typical for the pelvic region. It will be referred to as "pelvis region phantom" in the following. This configuration was used to evaluate the performance of scatter correction in a worst-case scenario. Two spherical inserts are used in this case. The bigger one ($R = 36$ mm) is located in the center and represents the bladder. The second one ($R = 13.8$ mm) is located 5 cm away from the "bladder" to describe a tumor lesion near the bladder, see Fig. 2.15. We chose a bladder:lesion:background activity concentration ratio of 40:5:1 which is comparable to concentration ratios frequently observed in the pelvic region in clinical $^{18}$F-FDG PET scans.

### 2.6.2 Clinical data

The clinical dataset contains three representative patient studies (two whole-body acquisitions and one brain scan). Informed consent was obtained from all subjects. All patient data were anonymized.

**Patient A (whole-body)**

The patient had a cervical lymph node metastases of squamous-cell carcinoma. He received 333 MBq dose of $^{18}$F-FDG and underwent 10-bed-positions whole-body PET scan starting 65 minutes post-injection with 2 minutes scan time per bed position.

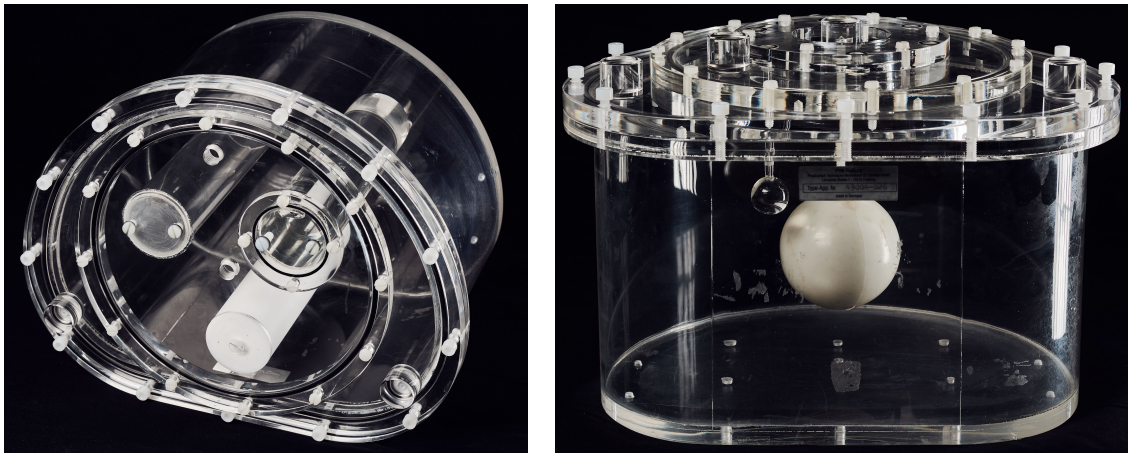

Figure 2.15: *Whole-body phantom. Left: cylindrical air and PTFE inserts for attenuation correction assessment. Right: spherical water-filled inserts representing bladder and tumor lesion for scatter-correction assessment.*

#### Patient B (whole-body)

The first patient had five adenocarcinoma metastasis in the liver. He received 309 MBq $^{18}$F-FDG injection and underwent 10-bed-positions whole-body PET scan 59 min post-injection. The patient had stent implants in the descending and abdominal aorta, and in the vessels of the mesentery and the kidneys. These implants are prone to cause metal-induced artifacts in MR images. The vicinity of the potentially affected areas to the lungs can affect lungs delineation from the MR image and, therefore, poses a great challenge to any MR-based attenuation correction technique.

#### Patient C (brain)

The patient exhibited a gray matter heterotopia in the left mesial temporal lobe. A brain scan of 20 minutes was performed starting 30 minutes after the injection of 210 MBq of $^{18}$F-FDG.

### 2.6.3 ROIs definition

Regions-Of-Interest (ROI) based analysis was used in the present thesis in order to evaluate the quantitative performance of the proposed methods. Description of the ROIs is given in the following.

#### Pelvis region phantom

The pelvis region phantom is dedicated to scatter correction quality evaluation. Therefore, of special interest in this context is the concentric neighborhood of the bladder since it is most strongly affected by inaccuracies of the scatter correction. A frequently observed problem in this region is a severe scatter overcorrection (known as photopenic or halo artifact) which thus provides a critical test of SC accuracy. The activity in the background was determined from a set of 14 cylindrical regions-of-interest (ROI) with a total volume

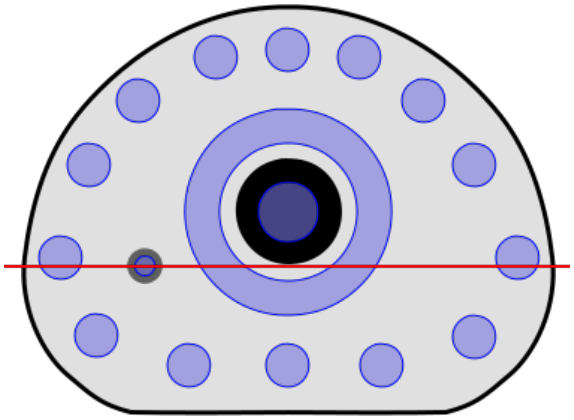

Figure 2.16: *Used pelvis region phantom geometry and ROI positions (light gray: background, dark gray: lesion, black: bladder, light blue: ROIs). The line profiles through the reconstructed images shown in Fig. 3.8 as well as the TOF profiles in Fig. 3.9 are along the horizontal red line through the center of the lesion and tangential to the bladder.*

of 805 $cm^3$ equally spaced along the rim approximately 3 cm inward from the edge of the phantom. Activities in the bladder and the lesion were measured in spherical ROIs of 59 $cm^3$ and 2 $cm^3$, respectively, centered within the respective region (Fig. 2.16).

#### Whole-body phantom with cylindrical inserts

The whole-body phantom with cylindrical inserts is divided into four compartments with different attenuation. Estimated linear attenuation coefficients in these compartments were taken from cylindrical ROIs which were placed inside the inserts and in the background. The ROIs in the inserts had a common size and volume (110 $cm^3$). The background $\mu$-value was measured in a set of two ROIs with a total volume of 1643 $cm^3$. The ROIs radii were selected small enough to exclude the influence of partial volume effects.

#### Patient 2

The noise in the reconstructed activity images and attenuation maps was determined as fractional standard deviation of the voxel values in the spherical ROI of 33 $cm^3$ placed in the liver.

### 2.6.4 Reconstruction parameters

The image reconstruction parameters for THOR are listed in Table 2.3. The parameters were the same for both TOF and non-TOF reconstructions. No post-smoothing was applied. The scatter TOF bin number for each reconstruction protocol was determined with the optimization procedure described in Section 2.7.3. 25 TOF bins yield a bin width of 240 ps and 50 TOF bins yield a bin width of 120 ps. Note, that MLRES and MLAA reconstruction algorithms utilize different iteration schemes which are specified in the respective sections.

| Study | Iterations × Subsets | Voxels size (mm) | Grid size (voxels) | TOF-bins | Sc. updates at (%) |
|---|---|---|---|---|---|
| Phantom | 5 × 12 | 4 | 144 × 144 × 45 | 50 | 20/40/60 |
| Clinical (WB) | 2 × 12 | 4 | 144 × 144 × 234 | 25 | 30/60 |
| Clinical (WB single bed) | 2 × 12 | 4 | 144 × 144 × 45 | 25 | 30/60 |
| Clinical (Brain) | 3 × 12 | 2 | 256 × 256 × 90 | 50 | 20/40/60 |

Table 2.3: *Iteration schemes and SC parameters for image reconstruction with THOR.*

For TOF reconstruction, initial scatter estimation was performed in TOF mode with respective algorithm and the scatter time distributions were derived (i.e. the time demanding procedure of scatter time distribution calculation was performed only once per reconstruction). Non-TOF SSS was used for subsequent scatter updates during the iterative reconstruction process. A simplified scheme of the reconstruction process is shown in Fig. 2.17. Two scatter updates were done for the whole-body scans while three scatter updates were done for the brain and phantom scans since these were reconstructed with a protocol using an increased number of iterations.

If the different was not explicitly stated, attenuation maps were generated from MR image segmentation into 3 classes (air, lungs, soft tissue) with the in-house toolkit (Section 2.1.1). Truncation compensation was performed when necessary.

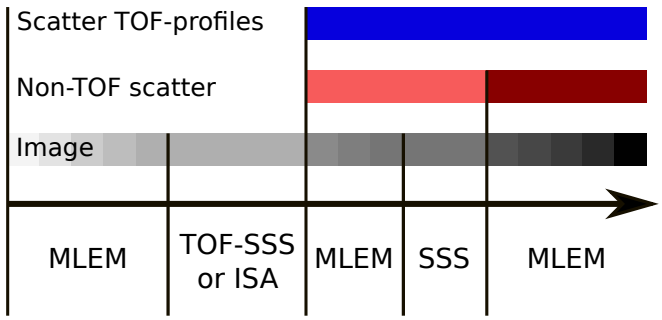


Figure 2.17: *Simplified scheme of the reconstruction process. Horizontal bars represent the quantities $m_{i\tau}$, $s_i$, $\lambda_j$ estimated during the reconstruction. Different color shades indicate changes in the value of the respective quantity (the case of two scatter updates and 12 updates of the initial image estimate is shown).*

## 2.7 Optimization of reconstruction parameters

The THOR reconstruction tool has several parameters which can be adjusted in order to achieve optimal reconstruction accuracy and performance. Among these are scatter scaling parameters, the number of used scatter TOF bins, and scanner time resolution.

### 2.7.1 Scatter scaling

As mentioned in Section 2.2.3, the Single Scatter Simulation does not account for multiple scatter events and out-of-FOV scatter. Therefore, scatter scaling procedure is required,

see e.g. (Polycarpou et al., 2011). The scatter scaling approach is based on the assumption that the total scatter including single, multiple, and ooFOV scatter has the same shape as the single scatter in each plane. It is then possible to scale the SSS scatter distribution to match the overall scatter amount. The accuracy of this procedure depends on the chosen scaling method. Mostly variants of scatter tail fitting are used but other methods have been proposed as well, such as maximum likelihood scatter scaling (Rezaei et al., 2017a) or Monte Carlo simulation based scaling (Jinghan Ye et al., 2014).

Within this work, we are using the scatter tail fitting method. It exploits the fact that for LORs which are not crossing the activity support all measured events are either randoms or scatter. Since the randoms contribution can be reliably estimated (Section 2.2.4), the scatter tails can be used to determine the overall scaling factor for the whole scatter distribution in the plane-of-response (Section 1.2.4). For this, the least squares method is normally used. The aim here is to find a scaling factor $b > 0$ which minimizes the sum

$$\sum_{i \in \text{tails}} k_i \, (y_i - r_i - b \, s_i^{\text{SSS}})^2, \tag{2.62}$$

where $k_i$ is the LOR weight. The resulting scatter estimate is defined then by

$$s_i = b \, s_i^{\text{SSS}} \tag{2.63}$$

The resulting fit quality strongly depends on the choice of the weighting coefficients $k_i$. The most straightforward approach here is to use an unweighted fitting procedure corresponding to the choice $k_i = 1$ for all $i$. Another weighting method is based on the assumption that close-to-object LORs contain the most reliable information about scatter. Hence, only these LORs should be used during the fitting. We investigated several implementations of this weighting scheme where $k_i$ was set to 1 for the first 3, 5, and 7 LORs, respectively, closest to the activity support for each projection angle and to 0 otherwise.

The last considered weighting method is inspired by the structure of the MLEM iteration formula (2.22) itself. As it can be seen from it, scatter errors are amplified by a factor of $1/a_i$ in the scatter tails region since the forward projector is close to zero. This amplification factor can be larger than 1 outside of the activity support due to the presence of the patient bed. This suggests using $k_i = (1/a_i)^2$ as a weighting factor during the least squares tail fitting process. Other powers of this factor $1/a_i$, $(1/a_i)^3$, $(1/a_i)^4$, $(1/a_i)^5$, $(1/a_i)^6$ were also investigated.

We determine the activity support by suitable thresholding of the attenuation map excluding the patient bed. A safety margin of 10 mm was added to the derived body outline in order to avoid any contamination of the LORs used for tail fitting by residual activity contributions due to finite resolution or residual patient motion. Scaling coefficients were calculated and applied to the scatter in each plane independently in order to allow for variations in a scatter scale due to the variable fraction of the ooFOV scatter along the

scanner axis.

The aforementioned methods were incorporated into the THOR scatter correction routine and evaluated on the pelvis region phantom data (Section 2.6.1). The data were reconstructed in the non-TOF mode. The reconstructed activities in the ROIs defined in Section 2.6.3 were analyzed and the best performing method was selected for further use in the reconstruction.

### 2.7.2 Time resolution calibration

Time resolution is an important characteristic of TOF-capable PET system. Higher time resolution means more reliable measurements of the photon arrival time differences allowing to achieve higher benefits from TOF-imaging. It is important to use a correct time resolution during image reconstruction in order to avoid artifacts in MLEM (Daube-Witherspoon et al., 2006) and MLAA (Cheng et al., 2016a). Time resolution depends on the scintillation crystals response time as well as on the performance of the signal processing electronics. In this context, it is relevant to recognize that the time resolution of a time-of-flight PET system is count-rate-dependent (Surti et al., 2007). However, count-rate-dependent time resolution calibration was not provided by the vendor for our scanner which is a common practice. We, therefore, developed such a procedure which is compatible with clinical routine and is also applicable retrospectively to existing data.

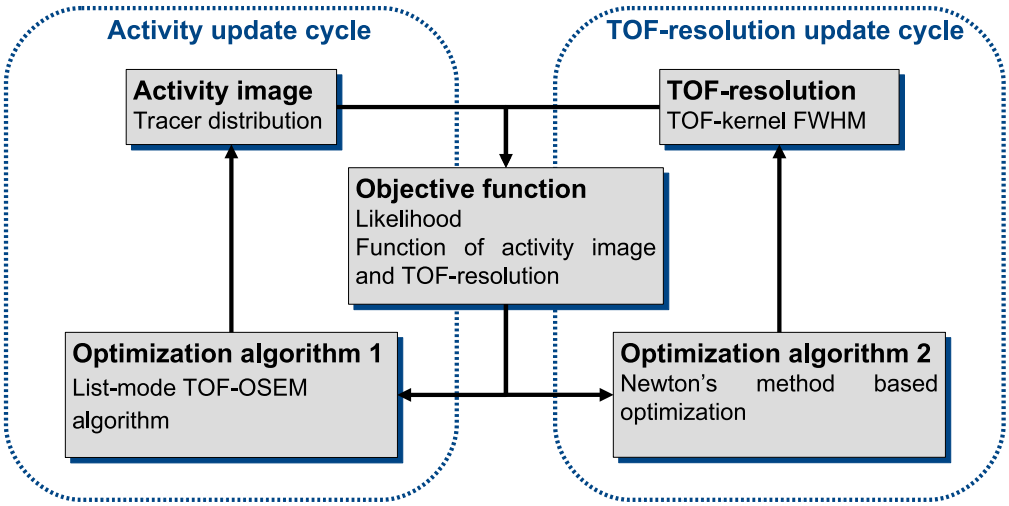


Figure 2.18: *MLRES algorithm general scheme.*

We propose a novel Maximum Likelihood Time Resolution Estimation (MLRES) algorithm that maximizes the likelihood by updating activity image and TOF-kernel width alternately. TOF-MLEM is used to update the activity image and Newton's-method-based maximization of the likelihood is performed to update the time resolution. The idea of this method is similar to that of MLAA. It is assumed that the likelihood (2.42) reaches its maximum when both true activity distribution and true time resolution were used (Vandenberghe et al., 2007). The simplified scheme of the algorithm is shown in Fig. 2.18.

Newton's optimization method used for time resolution estimation iteratively maximizes the second-order Taylor expansion of the likelihood assuming that the global likelihood maximum is located close enough to initial $\sigma_{\mathrm{TOF}}^{(0)}$ guess

$$\sigma_{\mathrm{TOF}}^{(p+1)} = \sigma_{\mathrm{TOF}}^{(p)} - \frac{\dot{l}\left(\boldsymbol{\lambda}^{(p)}, \sigma_{\mathrm{TOF}}^{(p)}\right)}{\ddot{l}\left(\boldsymbol{\lambda}^{(p)}, \sigma_{\mathrm{TOF}}^{(p)}\right)}, \tag{2.64}$$

where index $p$ is used to enumerate both activity and time resolution updates. Single and double dots above the letters designate, respectively, the first and the second derivative over $\sigma_{\mathrm{TOF}}$

$$\dot{l}\left(\boldsymbol{\lambda}, \sigma_{\mathrm{TOF}}\right) = \frac{\partial\, l}{\partial \sigma_{\mathrm{TOF}}}\left(\boldsymbol{\lambda}, \sigma_{\mathrm{TOF}}\right), \tag{2.65a}$$

$$\ddot{l}\left(\boldsymbol{\lambda}, \sigma_{\mathrm{TOF}}\right) = \frac{\partial^2\, l}{\partial {\sigma_{\mathrm{TOF}}}^2}\left(\boldsymbol{\lambda}, \sigma_{\mathrm{TOF}}\right). \tag{2.65b}$$

Direct calculation of the derivatives of the likelihood (2.42) in (2.64) gives us the time resolution update

$$\sigma_{\mathrm{TOF}}^{(p+1)} = \sigma_{\mathrm{TOF}}^{(p)} - \left.\frac{\displaystyle\sum_e \frac{\dot{\overline{y}}_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}}}{\displaystyle\sum_e \left[\frac{\ddot{\overline{y}}_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}} - \left(\frac{\dot{\overline{y}}_{i_e,\delta_e}}{\overline{y}_{i_e,\delta_e}}\right)^2\right]}\right|_{\sigma_{\mathrm{TOF}}=\sigma_{\mathrm{TOF}}^{(p)}}, \tag{2.66}$$

where $\overline{y}_{i,\delta}$ is defined in (2.26) and

$$\dot{\overline{y}}_{i,\delta} = a_i\, n_i \sum_j c_{ij}\, \dot{\widetilde{w}}_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}})\, \lambda_j + s_i\, \dot{W}_i(\delta, \sigma_{\mathrm{TOF}}), \tag{2.67a}$$

$$\ddot{\overline{y}}_{i,\delta} = a_i\, n_i \sum_j c_{ij}\, \ddot{\widetilde{w}}_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}})\, \lambda_j + s_i\, \ddot{W}_i(\delta, \sigma_{\mathrm{TOF}}), \tag{2.67b}$$

as follows from (2.26), (2.27) and (2.33) with TOF-weighting factors $\widetilde{w}_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}})$ and $W_i(\delta, \sigma_{\mathrm{TOF}})$ given by (2.32a) and (2.35a).

We further assume that the number of scatter TOF bins used is reasonably large (say, $N_\tau \geq 50$) and that the effects of the TOF-kernel truncation can be ignored for most of the events. In this approximation normalization factors $\widetilde{A}(\delta, \sigma_{\mathrm{TOF}})$ and $B(\delta, \sigma_{\mathrm{TOF}})$ are proportional to the standard Gaussian normalization factor. The derivatives of the TOF weighting factors in (2.67) are then given by

$$\dot{\widetilde{w}}_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}}) = \widetilde{w}_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}}) \left[\frac{(\delta_{ij} - \delta)^2}{\sigma_{\mathrm{TOF}}^3} - \frac{1}{\sigma_{\mathrm{TOF}}}\right], \tag{2.68a}$$

$$\ddot{\widetilde{w}}_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}}) = \widetilde{w}_i(\delta_{ij}; \delta, \sigma_{\mathrm{TOF}}) \frac{1}{\sigma_{\mathrm{TOF}}^2} \left[2 - 5\,\frac{(\delta_{ij} - \delta)^2}{\sigma_{\mathrm{TOF}}^2} + \frac{(\delta_{ij} - \delta)^4}{\sigma_{\mathrm{TOF}}^4}\right], \tag{2.68b}$$

$$\dot{W}_i(\delta, \sigma_{\mathrm{TOF}}) = N_\tau\, B(\delta, \sigma_{\mathrm{TOF}}) \times \\ \times \sum_{\tau=1}^{N_\tau} m_{i\tau} \left[ \frac{(\delta_\tau - \delta)^2}{\sigma_{\mathrm{TOF}}^3} - \frac{1}{\sigma_{\mathrm{TOF}}} \right] \int\limits_{\Delta_\tau} \exp\left[ -\frac{(\delta' - \delta)^2}{2\,{\sigma_{\mathrm{TOF}}}^2} \right] d\delta', \tag{2.68c}$$

$$\ddot{W}_i(\delta, \sigma_{\mathrm{TOF}}) = N_\tau\, B(\delta, \sigma_{\mathrm{TOF}}) \times \\ \times \sum_{\tau=1}^{N_\tau} \frac{m_{i\tau}}{\sigma_{\mathrm{TOF}}^2} \left[ 2 - 5\,\frac{(\delta_\tau - \delta)^2}{\sigma_{\mathrm{TOF}}^2} + \frac{(\delta_\tau - \delta)^4}{\sigma_{\mathrm{TOF}}^4} \right] \int\limits_{\Delta_\tau} \exp\left[ -\frac{(\delta' - \delta)^2}{2\,{\sigma_{\mathrm{TOF}}}^2} \right] d\delta', \tag{2.68d}$$

where $\delta_\tau$ is a time difference corresponding to the center of TOF bin $\tau$ and all other notations were introduced before. As can be seen, the proposed algorithm requires three forward projection operations per event with different TOF weighting. It is not necessary, however, to recalculate the system matrix elements multiple times per projection operation since values $c_{ij}$ can be computed only once per LOR and then reused. MLRES can be further accelerated with the ordered subsets like MLEM and MLTR, see Section 2.2.1 and Section 2.5.1.

A closer look at (2.66) shows that outside of the activity support, only scatter data is used to determine time resolution. Moreover, the contribution of such LORs can be relatively high especially for small objects. Since the algorithm compares the measured scattered data against the simulated one, the possible errors in the estimated scatter can translate into the errors in the estimated time resolution. In order to avoid this risk, we restrict the events set used in (2.66) and ignore all the events from the LORs which do not cross the activity support.

The scanner time resolution calibration was performed using SUV phantom (Section 2.6.1). The wide range of singles rates (about 5 – 30 Mcps) the phantom study was performed at allows deriving the calibration curve which covers the clinically relevant count-rate interval and even extends beyond it. 32 short time frames (150 – 600 s) were extracted from the six-hours-long study and reconstructed separately in order to prevent large variations of the singles rate exceeding 0.3 Mcps. The iteration scheme of five MLEM iterations with 12 subsets was chosen (Table 2.3) while five time resolution updates per single MLEM update were performed in order to achieve convergence of the algorithm. 100 TOF bins were used to avoid sampling errors. Iterations were initialized with a time resolution of 700 ps in all cases. TOF-SSS algorithm was used for scatter correction.

### 2.7.3 TOF bins number

The number of scatter TOF bins, $N_\tau$, determines the sampling of the stored scatter time profiles $m_{i\tau}$. Increasingly finer sampling reduces discretization errors but rapidly leads to prohibitive RAM requirements. Therefore, it is important to determine a suitable value for $N_\tau$ that yields sufficient scatter correction accuracy while not exceeding available memory.

For this analysis, we employed the same strategy as in Section 2.7.1 and used the pelvis region phantom for precise scatter correction quality evaluation. The scatter correction

was performed with the reference TOF-SSS.

The TOF bins number was varied in the range [1-100] and the dependency of the reconstructed ROI activities (Section 2.6.3) on $N_\tau$ was determined. Suitable values of $N_\tau$ were determined for the whole-body and brain reconstruction protocols, respectively, while taking into account the existing requirements regarding desirable numerical accuracy and available RAM.

## 2.8 Validation

### 2.8.1 MLRES validation

The MLRES method was validated by applying it to a set of phantom and patient studies different from the SUV phantom dataset used for the initial calibration. This set includes the phantom datasets listed in Section 2.6.1 and the clinical datasets listed in Section 2.6.2. Long phantom studies were split into short time frames to reduce variations in the singles rate to below 0.3 Mcps. MR-based attenuation maps were used for the reconstruction. The same iteration scheme and time resolution start value were used for both calibration and validation purposes. Results were compared with the initial time resolution calibration and the mean absolute and relative deviations from it were calculated.

Since global convergence of MLRES is not guaranteed, we assessed the robustness of the algorithm to initialization. For this, we performed MLRES reconstructions of the same SUV phantom study with varying initial values of the TOF-kernel width in the range of $[500 - 1000]$ ps. The obtained results were compared and the deviations from the mean were computed.

Performance of the proposed method for time resolution prediction was additionally evaluated with MLAA reconstruction. It is known that the MLAA algorithm is sensitive to errors in the scanner calibration in general (Nuyts et al., 2018) and to errors in the time resolution calibration in particular (Cheng et al., 2016a). Thus, the whole-body phantom with three cylindrical inserts (Section 2.6.1) was reconstructed with LM-MLAA algorithm (non-TOF MLTR version) using time resolution of 550 ps provided by the vendor and 685 ps estimated with MLRES. The transmission-scan-based attenuation map of the phantom was used as a reference. The $\mu$-map was acquired with ECAT EXACT HR+ (Siemens, Knoxville, Tennessee) PET scanner. The corresponding reference activity image was reconstructed with TOF-MLEM using time resolution of 685 ps.

The TOF-SSS algorithm was used for scatter correction for all reconstructions.

### 2.8.2 ISA accuracy and performance

The different TOF-SC algorithms were evaluated on a set of phantom and patient scans. The phantom modeled extreme high contrast conditions in the pelvic region in order to evaluate scatter correction in a worst-case scenario. The phantom setup is described in Section 2.6.1. The accuracy of the SC algorithms was quantitatively assessed by comparing

image-derived activity ratios of relevant regions with their true values. Performance of the different TOF-SC algorithms was additionally assessed in scans of the patients A and C (Section 2.6.2) representing typical whole-body and brain investigations, receptively.

The data were reconstructed with THOR in TOF and non-TOF mode. Reconstruction settings are listed in Table 2.3.

Altogether, 5 different reconstruction/scatter correction schemes were evaluated:

1. **Vendor**: vendor provided reconstruction (phantom/whole-body: TOF BLOB-OS-TF, brain: non-TOF 3D-RAMLA)
2. **Non-TOF recon**: non-TOF THOR + SSS
3. **SSS**: TOF-THOR + SSS (distributing scatter equally over TOF bins)
4. **ISA**: TOF-THOR + ISA
5. **TOF-SSS**: TOF-THOR + TOF-SSS

The vendor-provided reconstruction was included for comparison against our work and used with its "highest quality" setting.

### 2.8.3 Quality of MLAA attenuation correction

Two versions of the MLAA algorithm featuring non-TOF (2.56) and TOF (2.54) LM-MLTR as a $\mu$-map estimation tool were evaluated on phantom and patient data. MLAA reconstructed activity and attenuation images were compared to those obtained with the in-house developed MR-based $\mu$-map generation tool described in Section 2.1.1 in combination with TOF-MLEM. In the further text, we will use the abbreviation MRTR to refer to our custom $\mu$-map generator. The in-house developed method was used in the present investigation due to its superior performance compared to the vendor-provided toolkit. All reconstructions were performed with both ISA and TOF-SSS in order to assess the influence of the selected SC method on the joint reconstruction procedure.

Altogether, six different reconstruction regimes were utilized:

1. **MRTR + ISA**: TOF-MLEM with ISA scatter correction and MR-based $\mu$-map
2. **MRTR + TOF-SSS**: TOF-MLEM with TOF-SSS scatter correction and MR-based $\mu$-map
3. **MLTR + ISA**: MLAA (TOF-MLEM with non-TOF MLTR) with ISA scatter correction
4. **MLTR + TOF-SSS**: MLAA (TOF-MLEM with non-TOF MLTR) with TOF-SSS scatter correction
5. **TOF-MLTR + ISA**: MLAA (TOF-MLEM with TOF-MLTR) with ISA scatter correction

6. **TOF-MLTR + TOF-SSS**: MLAA (TOF-MLEM with TOF-MLTR) with TOF-SSS scatter correction

The accuracy of the estimated linear attenuation coefficients was evaluated with the whole-body phantom with three cylindrical inserts (Section 2.6.1). The reconstructed $\mu$-values were determined based on the ROI analysis and compared with the known attenuation coefficients of air, water, and PTFE accordingly to the analyzed region.

The pelvis region phantom was employed for the algorithm performance analysis under the high activity contrast and extensive scatter conditions. Reconstructed activity and estimated attenuation values were sampled from the ROIs described in Section 2.6.3 and were compared with the reference values. Since the phantom was filled with water, linear attenuation coefficients of 0.096 $cm^{-1}$ were expected in all ROIs. Any deviations from this value can be attributed to either sensitivity of the selected reconstruction method to errors in the estimated scatter or to the cross-talk between attenuation and activity.

MLAA capabilities in a clinical context were demonstrated using the examples of whole-body and brain patient studies. Datasets of the patients B and C were utilized for this purpose, respectively.

The iteration schemes and reconstruction parameters which were used for the activity reconstruction are listed in Table 2.3. The attenuation map reconstruction with MLAA was performed according to the interleaved reconstruction scheme (Fig. 2.11). Two MLTR iterations with 12 subsets were applied in-between the MLEM reconstruction cycles.

Convergence properties of the different MLAA implementations were evaluated using the whole-body phantom with cylindrical inserts. 10 MLTR reconstructions were applied and the estimated $\mu$-values accuracy was assessed. Both MRTR and homogeneous water-filled attenuation maps were used for the algorithm initialization. TOF-SSS was used for the scatter correction in this case. Consequent phantom and clinical data reconstructions were performed with the determined optimal number of MLTR reconstructions of 7 resulting in a total of $7 \times 2 \times 12 = 168$ MLTR updates.

# Chapter 3

# Results

## 3.1 Optimization of reconstruction parameters

### 3.1.1 Scatter scaling

Different variants of scatter scaling (see Section 2.7.1) were investigated using the pelvis region phantom. The reconstructed activity concentrations were compared to the true activity concentrations prepared using a cross-calibrated dose calibrator.

For a representative transaxial slice, the reconstructed images resulting from the different variants of scatter tail weighting are presented in Fig. 3.1. The ROI based quantification of these images yields the results in Table 3.1.

As is to be expected, the immediate vicinity of the "bladder" is the most sensitive region regarding differences between the tail fitting methods. Without scatter tail weighting, the activity concentration within this "halo region" did deviate by 33% from the true value. The situation improved somewhat when only considering the 3 LORs closest to the object: the deviation from ground truth was reduced to 26.0% in this case. Including 5 or 7 LORs yielded essentially the same result (with a tendency to increase the error slightly).

Attenuation-weighted tail fitting is distinctly superior in comparison. $(1/a_i)^2$ weighting, which is suggested by the fundamental MLEM properties, already reduced the error to 22.3% in the halo region. A further reduction of the halo artifact strength occurred when increasing the attenuation dependence of the weighting by raising the weighting factor to higher powers $n$. $n = 3$ reduced the underestimate in the halo region to 12.4% and $n = 4$ to 8.4%. The higher powers up to $n = 6$ did not lead to further significant changes. This is also obvious in Fig. 3.1 where images for weighting schemes $(1/a_i)^4$, $(1/a_i)^5$, and $(1/a_i)^6$ look virtually identical. Altogether, weighting with $(1/a_i)^4$ yielded satisfactory quantitative agreement (see Table 3.1) in all ROIs (Section 2.6.3) and was, therefore, used for all further reconstructions.

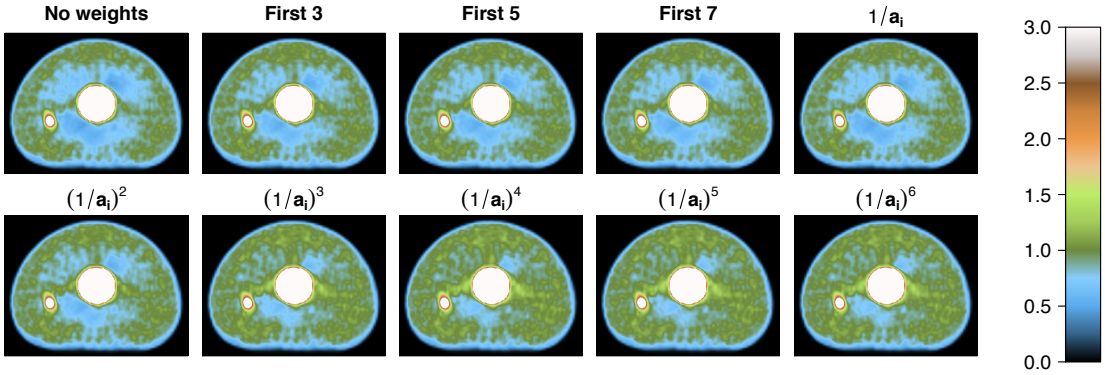


Figure 3.1: *Comparison of different scatter tail weighting methods. A representative transaxial slice from a measurement with the pelvis region phantom is shown. True activity concentration in the background is equal to one in the chosen units. Values below one thus indicate scatter overcorrection in the affected region. Images are thresholded as indicated at the colorbar to facilitate visual assessment of the relevant dynamic range (true contrast in the data is 40:1). "No weights" denotes unweighted scatter tail fitting. "First N" utilizes only the first N LORs not intersecting the object. "$(1/a_i)^n$" utilizes the reciprocal of the specified power of the attenuation factors as LOR weights in the fitting procedure.*

| Weighting method | Reconstructed activity deviation (%) | | | |
|---|---|---|---|---|
| | Δ Lesion | Δ Bladder | Δ Background | Δ Halo |
| No weighting | −12.9 | −2.7 | −7.2 | −33.3 |
| First 3 | −12.1 | −2.4 | −5.7 | −26.0 |
| First 5 | −12.1 | −2.5 | −5.9 | −26.7 |
| First 7 | −12.2 | −2.5 | −6.0 | −27.2 |
| $1/a_i$ | −12.6 | −2.6 | −6.6 | −30.5 |
| $(1/a_i)^2$ | −11.6 | −2.3 | −5.0 | −22.3 |
| $(1/a_i)^3$ | −10.4 | −1.9 | −3.0 | −12.4 |
| $(1/a_i)^4$ | −10.0 | −1.7 | −2.2 | −8.4 |
| $(1/a_i)^5$ | −10.0 | −1.7 | −2.2 | −8.2 |
| $(1/a_i)^6$ | −10.1 | −1.8 | −2.5 | −9.3 |

Table 3.1: *Quantitative comparison of scatter scaling methods. Shown are the deviations from true target region activities. "No weights" denotes unweighted scatter tail fitting. "First N" utilizes only the first N LORs not intersecting the object. "$(1/a_i)^n$" utilizes the reciprocal of the specified power of the attenuation factors as LOR weights in the fitting procedure.*

### 3.1.2 Time resolution calibration

Time resolution of the Ingenuity PET/MR scanner was estimated with the MLRES algorithm using the SUV calibration phantom (see Section 2.7.2). Our evaluation yields a perfect linear dependency ($R^2 = 1$) of time resolution on count rate, as demonstrated in Fig. 3.2. According to these results, the resolution approaches 551 ps as the count rate approaches a value of zero while the resolution degrades by 145 ps when increasing the singles count rate by ten million per second. Compared to the zero count-rate limit, time resolution is thus reduced distinctly for clinically relevant count rates and drops to about 840 ps for a singles rate of 20 Mcps ($= 20 \times 10^6$ counts per second).

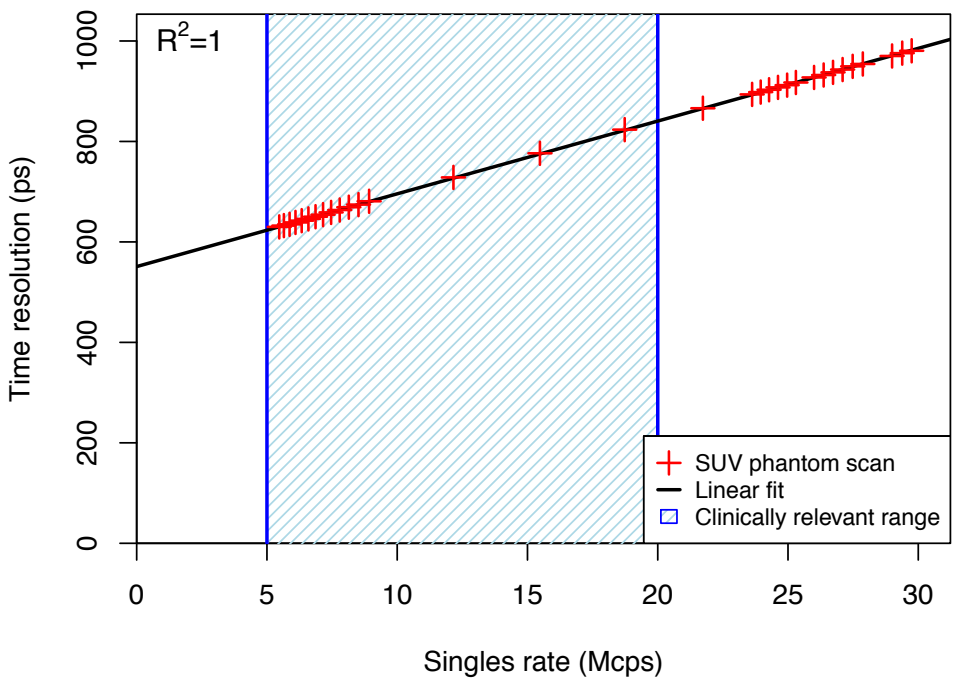


Figure 3.2: *Time resolution of Philips Ingenuity PET/MR as a function of singles rate estimated with MLRES. The range of clinically relevant singles rates is shown in blue.*

### 3.1.3 TOF bins number

The impact of the TOF bins number on accuracy of the TOF-aware scatter correction was evaluated with the pelvis region phantom, as described in Section 2.7.3. The dependency of reconstructed mean activity concentration in different regions on the number of TOF bins used to store the scatter time distributions is presented in Fig. 3.3. Asymptotically, the chosen number of TOF bins has no longer any influence on the reconstructed images which can be understood as being a consequence of negligible TOF discretization errors in this regime. But before this limit is reached, the chosen number of TOF bins has a pronounced effect and oscillating changes of the reconstructed values as a function of TOF bin number are apparent (discriminating between odd and even numbers of TOF bins). The oscillations vanish for $N_\tau \geq 16$, but only for $N_\tau \geq 50$ the asymptotic limit is approached.

According to these findings, we chose to use 50 TOF bins where feasible and used a reduced number of 25 TOF bins for those study types where available computer memory did not allow to use a larger number. As is obvious from Fig. 3.3, even the lower number suffices to reduce residual deviations from the asymptotic limit to essentially negligible levels.

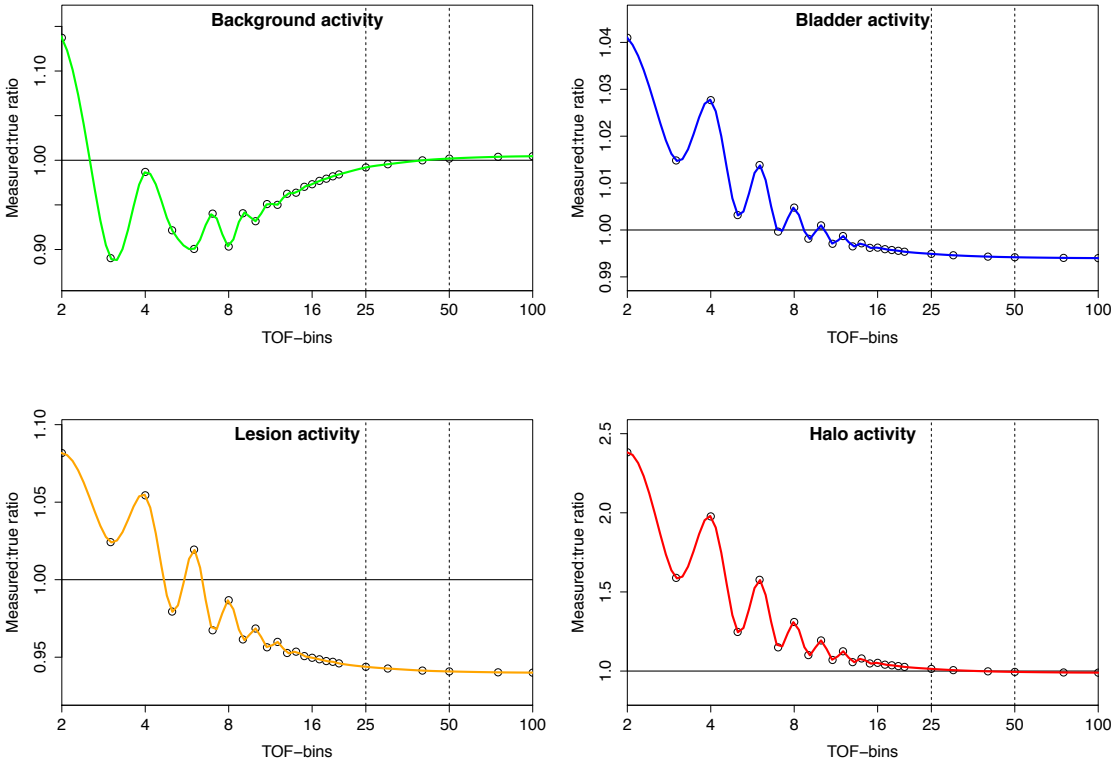


Figure 3.3: *Dependency of reconstructed activity values in the different ROIs on the amount of scatter TOF bins $N_\tau$. The ratios of the reconstructed activities to the ground truth are shown.*

## 3.2 Validation

### 3.2.1 MLRES validation

Consistency of the MLRES algorithm was validated in independently acquired phantom and clinical data as described in Section 2.8.1. The results are shown in Fig. 3.4. As can be seen, the validation data exhibit a small systematic deviation from the prediction of the performed calibration measurements. Mean deviation from the time resolution calibration is 21.6 ps (2.83%) over the investigated dynamic range. The relative error is increasing with increasing singles rate but remains below 7% over the whole range of singles rates. Overall, MLRES yields consistent results for clinical and phantom data. Interestingly, the highest deviations from predicted values were observed in the homogeneous whole-body phantom measurements. The mean absolute error for the remaining dataset was only 6.6 ps.

The dependency of the MLRES result on the start value for the time resolution is demonstrated in Fig. 3.5. MLRES results are deviating from the mean by not more than approximately 4 ps ($\approx$ 0.5%) for start values between 500 and 1000 ps. These results show that the algorithm is robust against the initialization and confirm that the chosen iteration scheme is sufficient to achieve adequate (if not perfect) convergence.

The importance of using the correct time resolution for MLAA reconstruction is demon-

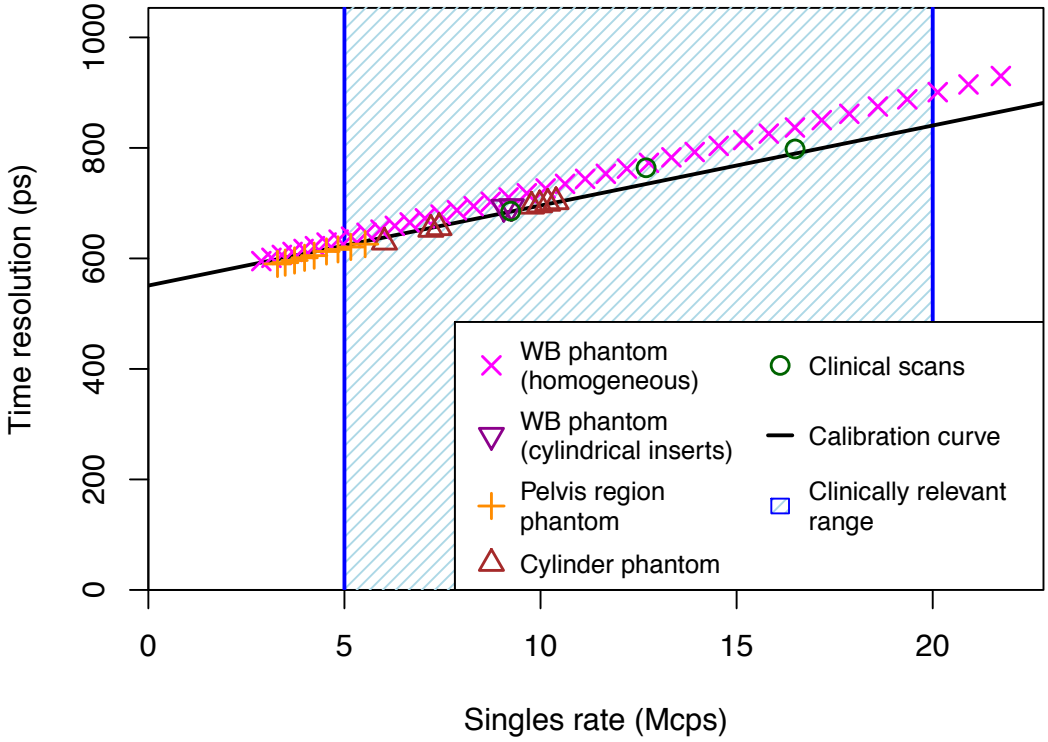


Figure 3.4: *Cross-validation of time resolution calibration. The range of clinically relevant singles rate is shown in blue.*

strated in Fig. 3.6. Distinct activity and attenuation artifacts at the rim of the phantom are present if the vendor-provided value of 550 ps is used. The artifacts disappear when the MLRES estimate (=685 ps in this case) is used and concordance with the transmission-based attenuation map and TOF-MLEM reconstruction is restored.

### 3.2.2 ISA accuracy and performance

Different variants of TOF scatter correction were investigated using the pelvis region phantom and clinical studies, see Section 2.8.2. Selected slices from the reconstructed image volumes are shown in Fig. 3.7. A quantitative evaluation is presented in Table 3.2. Line profiles intersecting both the lesion and the bladder are shown in Fig. 3.8.

The vendor (BLOB-OS-TF) reconstruction cannot handle this extreme high contrast case. Scatter is heavily overcorrected near the "bladder", resulting in a massive halo artifact with a spurious 65.8% signal drop. Activity in the background remote from the central sphere approaches its true value (= 1 in the units used in Fig. 3.7) only at the very edge of the phantom.

In comparison, non-TOF THOR utilizing SSS exhibits less severe but non-negligible

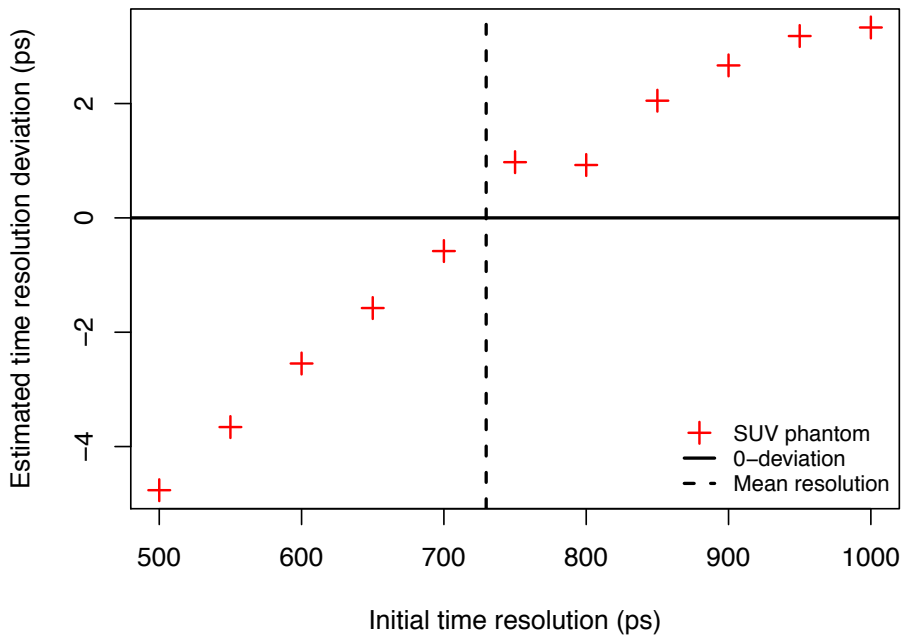


Figure 3.5: *Influence of chosen start value for the MLRES algorithm on a fit result. The start value has only a minimal effect on the final result (uncertainty range of about −4 to 3 ps).*

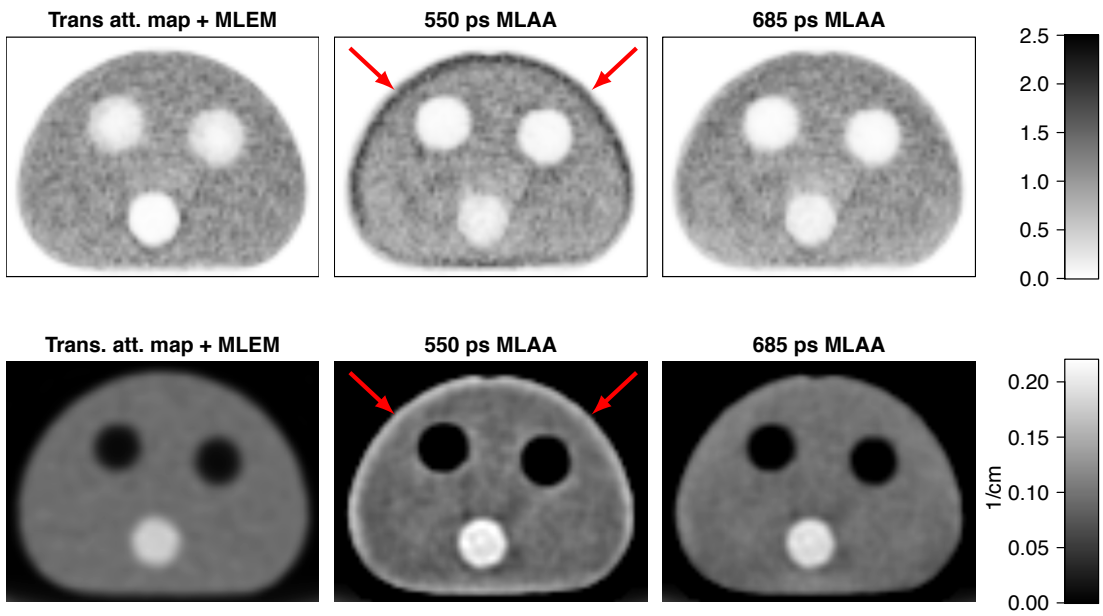


Figure 3.6: *Activity (top) and attenuation (bottom) images of the whole body phantom with cylindrical inserts. Left: transmission scan plus TOF-MLEM reconstruction. Middle: MLAA reconstruction assuming a time resolution of 550 ps (vendor-provided value). Note the indicated artifacts at the rim of the phantom. Right: MLAA reconstruction assuming a time resolution of 685 ps (MLRES estimate). True activity concentration in the background is equal to one in the chosen units.*

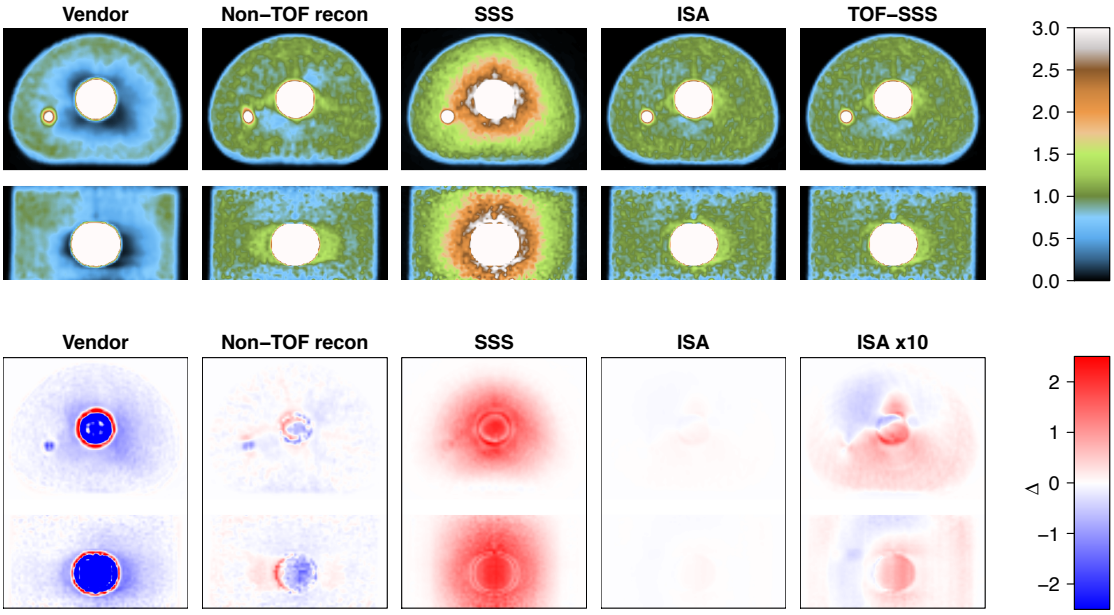


Figure 3.7: *Comparison of different scatter correction and reconstruction methods in the pelvis phantom study. The top two rows are activity images and the bottom two rows are absolute differences between selected reconstruction method and reference TOF-SSS. Top and bottom rows in each group show transaxial and coronal view, respectively. Image captions are explained in Section 2.8.2. True activity concentration in the background is equal to one in the chosen units. Images are thresholded as indicated at the colorbar to facilitate visual assessment of the relevant dynamic range (true contrast in the data is 40:1).*

scatter correction artifacts. For instance, one observes 8.4% activity underestimate (over-correction of scatter) in the halo region. Moreover, activity distribution around the central sphere is visibly nonuniform and introduces spurious structures in the reconstructed images.

Switching from non-TOF THOR to TOF-THOR while keeping simple SSS is not a viable approach as can be seen in the 3$^{\text{rd}}$ column of Fig. 3.7: here, scatter is heavily underestimated resulting in 138.3% higher activity around the "bladder" compared to the ground truth.

Utilizing TOF reconstruction together with TOF-aware scatter correction improves the situation distinctly as can be seen in the last two columns of Fig. 3.7. In fact, ISA and TOF-SSS produce near identical results. Both are able to suppress the halo artifact nearly completely and adequately reproduce the correct activity ratios between the different regions. The remaining small differences between ISA and TOF-SSS, as shown in the last two columns in the bottom half of Fig. 3.7, result in a small overestimate of mean background by 1.5%. All other measures are virtually identical, see Table 3.2.

The difference in scatter time profiles estimated with ISA and TOF-SSS is demonstrated in Fig. 3.9. The left plot shows the scatter TOF distribution obtained with TOF-SSS and its ISA approximation for the horizontal LOR shown in Fig. 2.16. The right plot shows the profiles — the actually applied TOF weights $w_{i,\delta}$ according to (2.35a) —

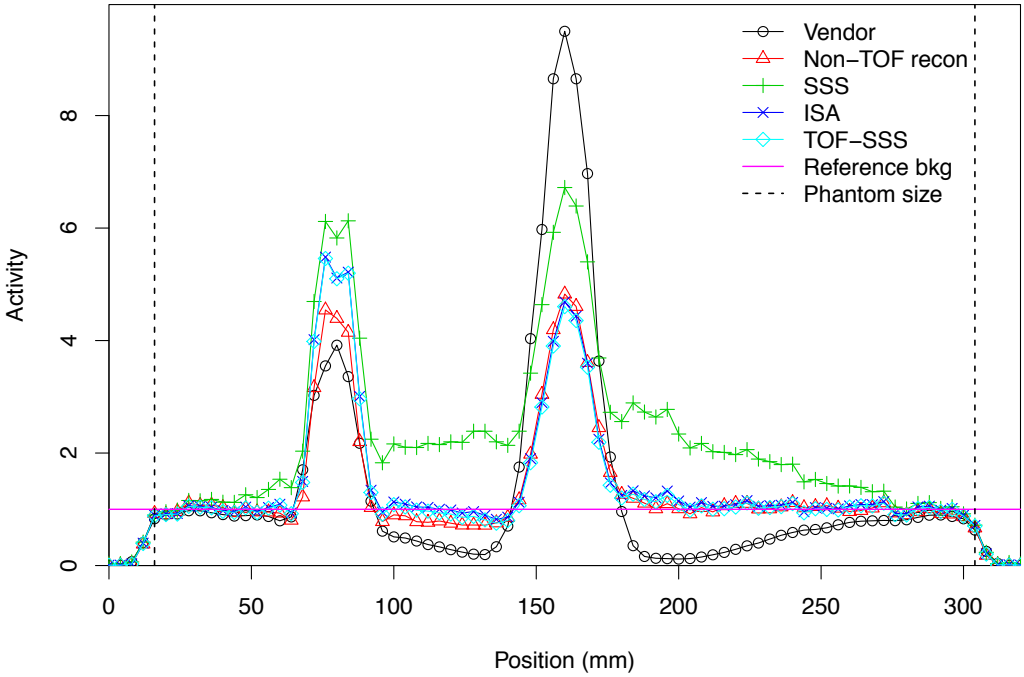


Figure 3.8: *Line profiles through the reconstructed images in Fig. 3.7 along the line indicated in Fig. 2.16. The left peak represents the intersection with the lesion, the right peak spill-out of signal from the bladder. True activity concentration in the lesion is equal to five in the chosen units.*

| SC & reconstruction method | Reconstructed activity deviation (%) | | | |
|---|---|---|---|---|
| | Δ Lesion | Δ Bladder | Δ Background | Δ Halo |
| Vendor recon | −16.3 | −13.6 | −12.1 | −65.8 |
| THOR: non-TOF | −10.0 | −1.7 | −2.2 | −8.4 |
| THOR: SSS | 8.2 | 4.1 | 13.6 | 138.3 |
| THOR: ISA | −5.7 | −0.4 | 1.5 | 2.5 |
| THOR: TOF-SSS | −5.9 | −0.6 | 0.2 | −0.5 |

Table 3.2: *Quantitative comparison of scatter correction and reconstruction methods. Shown are the deviations from true target region activities.*

when taking into account the finite time resolution of the PET scanner. It is apparent, that the deviations of ISA from TOF-SSS are mostly very small and larger deviations are sharply localized (notably at the edges of the central peak). The smoothing effect of the finite time resolution eliminates these differences essentially completely as is obvious on the right-hand side of Fig. 3.9.

The different reconstruction and scatter correction methods were also evaluated in representative clinical PET scans (patients A and C, Section 2.6.2). A whole-body investigation is shown in Fig. 3.10 and a long duration/high statistics brain investigation in Fig. 3.11 (as mentioned in Section 2.1.1, in the latter case the vendor uses a different

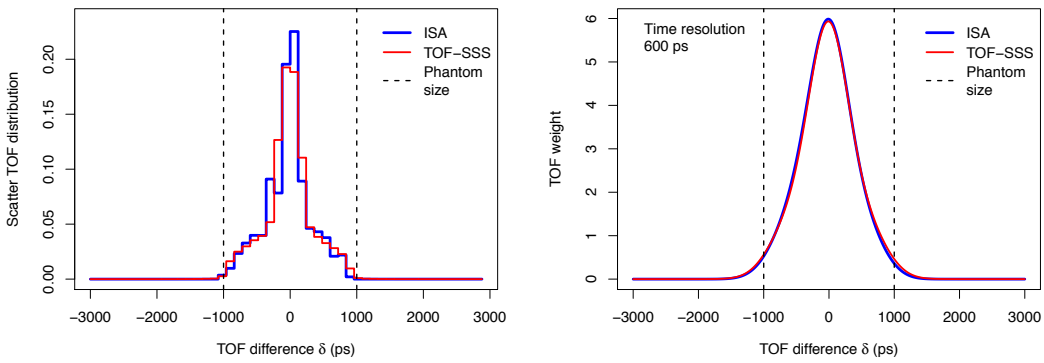


Figure 3.9: *Representative TOF profile $m_{i\tau}$ and its ISA approximation (left) and the corresponding TOF weights $W_i(\delta, \sigma_{\mathrm{TOF}})$ (right) as a function of TOF difference $\delta$ at $\sigma_{\mathrm{TOF}} = 600$ ps. The respective distributions were obtained in the pelvis region phantom for an LOR along the red line in Fig. 2.16.*

reconstruction than for whole-body studies). The results are qualitatively in agreement with those from the phantom study. Due to the less extreme contrast conditions deviations from the TOF-SSS results are visually less obvious, but quantitatively relevant (except for ISA which again yields virtually the same results as TOF-SSS) as can be appreciated from the difference images in the bottom half of both figures.

| SC & reconstruction method | Whole body | | Whole body (single bed) | | Brain | |
|---|---|---|---|---|---|---|
| | Time (min) | Time (%) | Time (min) | Time (%) | Time (min) | Time (%) |
| THOR: SSS | 5.0 | 100.0 | 0.3 | 100.0 | 0.4 | 100.0 |
| THOR: ISA | 5.8 | 115.7 | 0.3 | 124.6 | 0.5 | 124.1 |
| THOR: TOF-SSS | 23.8 | 477.2 | 1.5 | 581.5 | 1.1 | 273.6 |

Table 3.3: *Performance comparison of different scatter estimation methods. Scatter calculation time compared to SSS is shown.*

| SC & reconstruction method | Whole body | | Whole body (single bed) | | Brain | |
|---|---|---|---|---|---|---|
| | Time (min) | Time (%) | Time (min) | Time (%) | Time (min) | Time (%) |
| THOR: SSS | 21.5 | 100.0 | 3.0 | 100.0 | 16.3 | 100.0 |
| THOR: non-TOF | 20.1 | 93.5 | 2.9 | 96.4 | 13.4 | 82.5 |
| THOR: ISA | 22.4 | 103.9 | 3.1 | 101.6 | 16.4 | 100.5 |
| THOR: TOF-SSS | 41.6 | 193.3 | 4.2 | 141.0 | 16.9 | 104.1 |

Table 3.4: *Performance comparison of different scatter correction and reconstruction methods. Reconstruction time compared to THOR: SSS is shown.*

The runtime performance of the implemented approaches for these clinical data is summarized in Table 3.3 and Table 3.4. As can be seen, ISA outperforms TOF-SSS by a factor of 2–5 and is only 16–24% slower than non-TOF SSS. The performance gain of ISA over TOF-SSS translates into nearly a factor of two acceleration of the whole-body

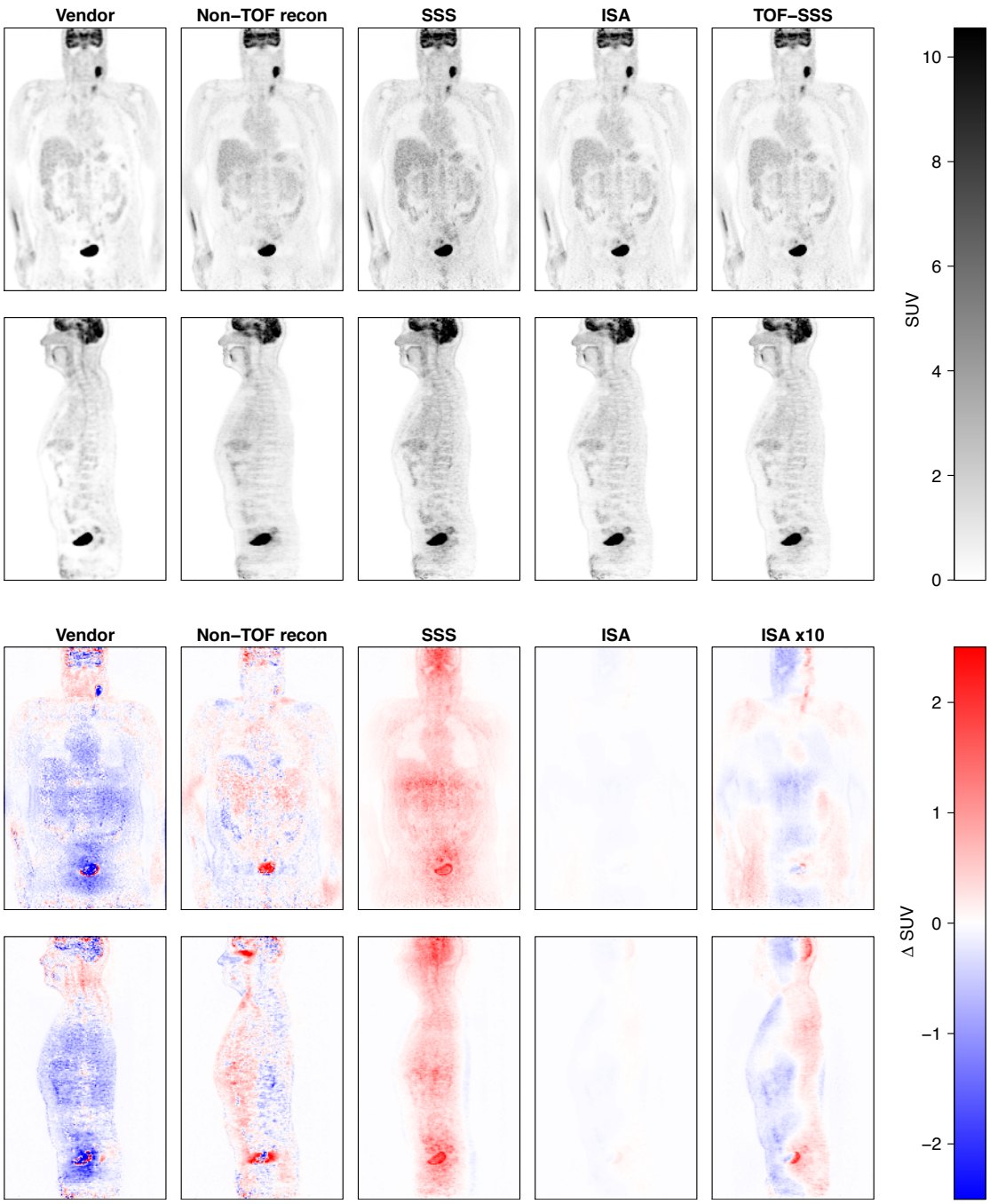


Figure 3.10: *Comparison of scatter correction and reconstruction methods in a clinical whole-body study. Image captions are explained in Section 2.8.2. The top two rows are activity images reconstructed with the different algorithms, and the bottom two rows are absolute differences between the indicated respective reconstruction method and TOF-SSS. Top and bottom rows in each group show coronal and sagittal views, respectively. In the different variants of TOF reconstruction red regions reflect undercorrection of scatter and blue ones overcorrection relative to TOF-SSS.*

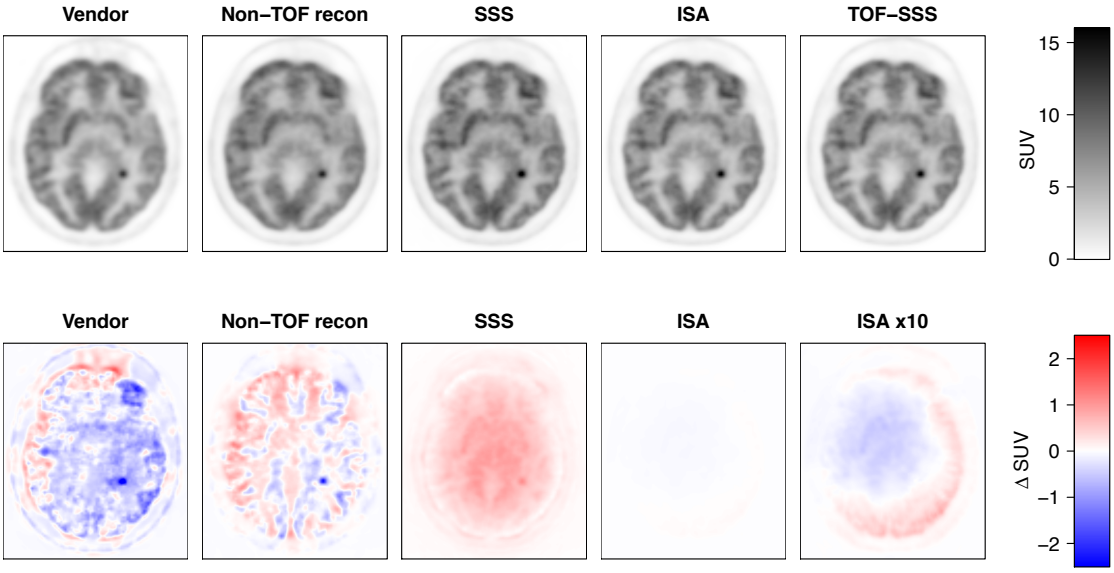


Figure 3.11: *Comparison of scatter correction and reconstruction methods in a clinical brain study. Image captions are explained in Section 2.8.2. The top row shows a transaxial view of the activity images reconstructed with the different algorithms, and the bottom row the absolute differences between the indicated respective reconstruction method and TOF-SSS. In TOF reconstructions red regions reflect undercorrection of scatter and blue ones overcorrection relative to TOF-SSS.*

reconstruction, resulting in a reconstruction time comparable to the one required when using ordinary SSS which indicates that scatter correction is responsible for a substantial part of total reconstruction time. Reconstructing a single bed position of the same study we observe acceleration of the reconstruction by a factor of 1.4 in comparison to TOF-SSS when using ISA and basically identical performance compared to SSS. For the long-duration (large number of events) brain investigation the overall time gain achieved with ISA vs. TOF-SSS is only minimal ($< 4\%$).

### 3.2.3 Quality of MLAA attenuation correction

The convergence properties of different MLAA variants were investigated using the whole-body phantom with cylindrical inserts. The attenuation correction algorithms performance was also evaluated with selected phantom and clinical datasets (see Section 2.8.3).

#### Convergence properties

The convergence of MLAA-reconstructed linear attenuation coefficients with increasing number of MLTR updates is shown in Fig. 3.12 for the case when using a homogeneous $\mu$-map to initialize the iterations. The convergence rates of non-TOF and TOF MLTR as a part of MLAA are similar although the TOF variant converges slightly slower in the air-filled regions. Both algorithms converge much faster in the bone-like region than in the air region. With the used iteration scheme, full convergence is achieved after 150–200 updates or 7–8 MLTR reconstruction cycles. TOF-MLTR reproduces the true attenuation

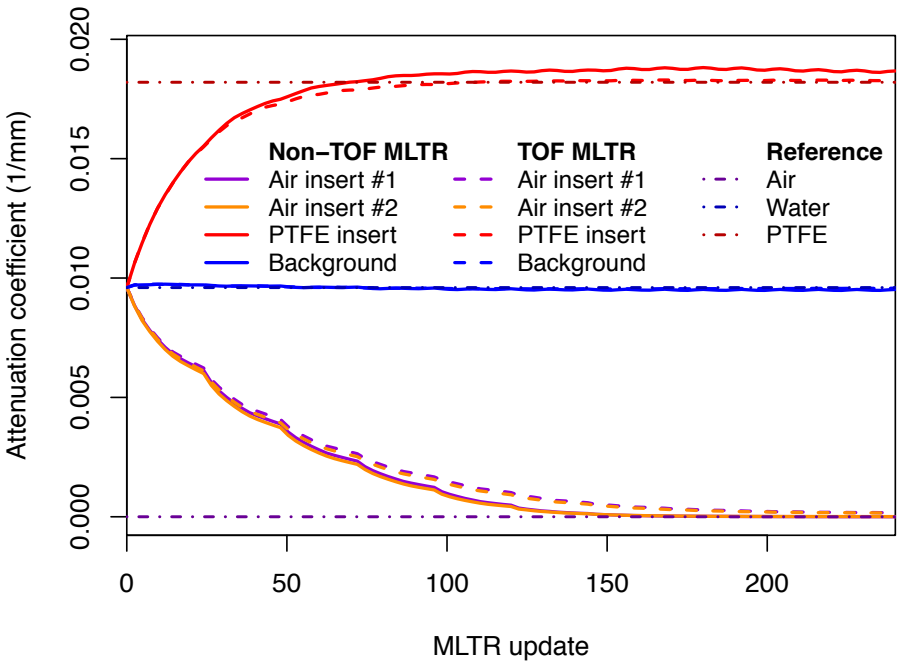


Figure 3.12: *Comparison of MLAA implementations with non-TOF and TOF MLTR. A homogeneous $\mu$-map was used for algorithm initialization.*

coefficient in the PTFE insert nearly exactly while non-TOF variant overestimates it by about 3%.

Initialization of the algorithms with the MRTR-derived $\mu$-map yields the results shown in Fig. 3.13. This $\mu$-map provides a better initial estimate of the actual $\mu$-value in the air-filled cavities than the homogeneous map used in the previous run. Consequently, faster convergence of the MLAA algorithm is observed in these regions: the true mean $\mu$-value in the air-filled cylinders is reached after about 50 updates.

### Image quality: phantom studies

Reconstructed activity and attenuation images obtained with different combinations of $\mu$-map estimation and TOF-SC methods are shown in Fig. 3.14. $\mu$-values obtained with the different methods in different regions of the phantom are summarized in Table 3.5. For comparison, a T1-weighted MR image is shown as well in Fig. 3.14. As can be seen, no differentiation between air and PTFE inserts would be possible with MR data alone, making derivation of a valid MR-based attenuation map unfeasible. The MRTR algorithm, according to its internal logic, assigned the attenuation value of lungs to the two upper cylinders and that of water to the bottom cylinder, which is incorrect.

The MLAA reconstructions, on the other hand, are able to correctly identify PTFE and air. The true $\mu$-value of PTFE ($\mu = 0.182$ cm$^{-1}$) was estimated with different

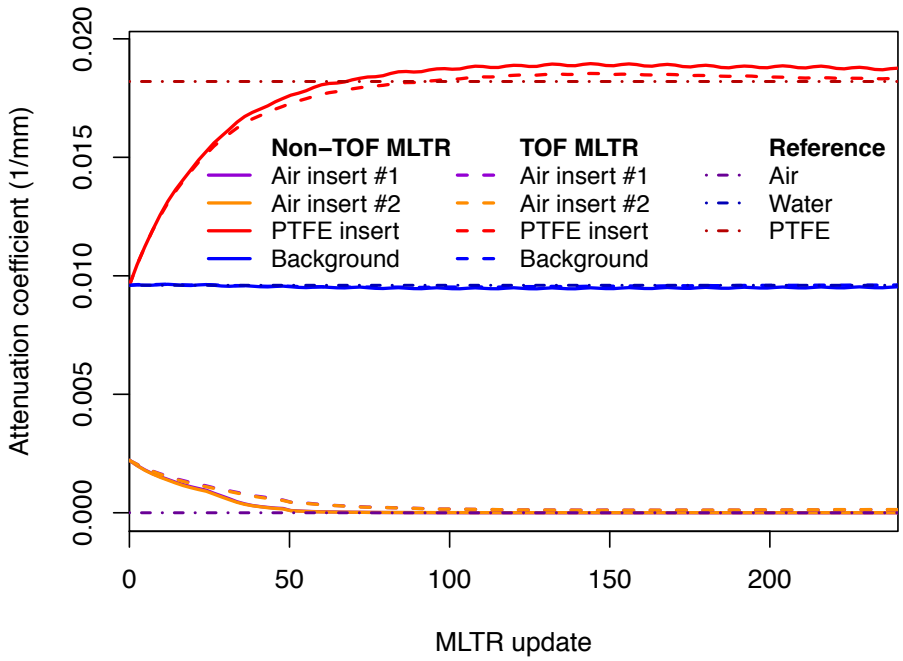


Figure 3.13: *Comparison of MLAA implementations with non-TOF and TOF MLTR. The MR-based $\mu$-map (MRTR), was used for algorithm initialization (Fig. 3.14).*

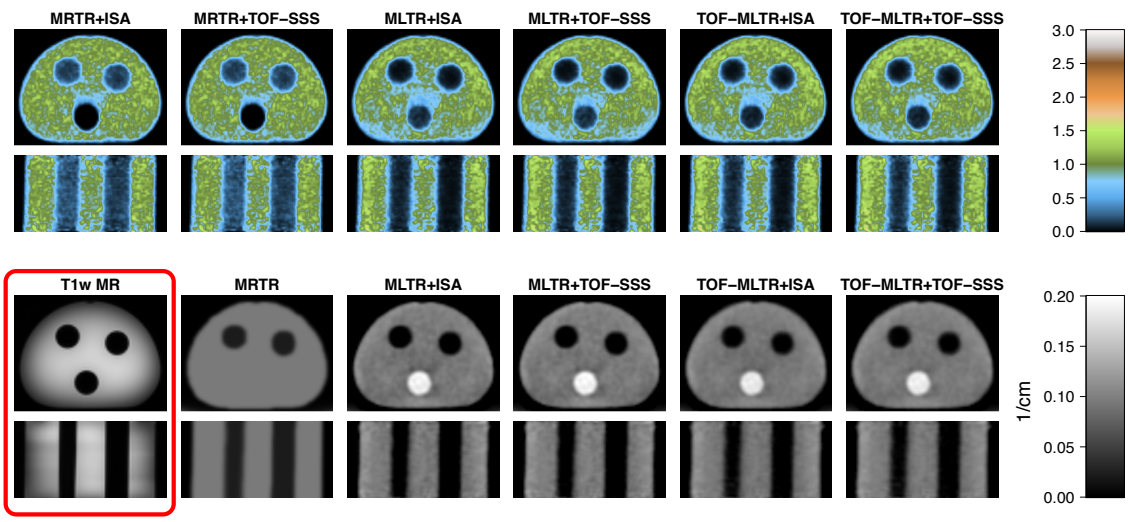


Figure 3.14: *Transaxial (rows 1, 3) and coronal (rows 2, 4) views of activity images (rows 1, 2) and attenuation maps (rows 3, 4) of the whole-body phantom with three cylindrical inserts. Each column represents a different combination of $\mu$-map estimation and TOF-SC methods. Two inserts are air-filled, the third consists of PTFE. Image captions are explained in Section 2.8.3. A T1-weighted MR image (T1w MR) is shown for comparison.*

accuracy by different MLAA implementations. TOF-MLTR-based reconstructions yielded the most accurate results (0.1819 cm$^{-1}$ with ISA and 0.1849 cm$^{-1}$ with TOF-SSS scatter

| $\mu$-map estimation method | SC method | Reconstructed attenuation (1/cm) | | | |
|---|---|---|---|---|---|
| | | Background | Air insert #1 | Air insert #2 | PTFE insert |
| True values | — | 0.0960 | 0.0000 | 0.0000 | 0.1820 |
| MRTR | — | 0.0960 | 0.0221 | 0.0221 | 0.0960 |
| MLTR | ISA | 0.0948 | 0.0000 | 0.0000 | 0.1875 |
| MLTR | TOF-SSS | 0.0951 | 0.0001 | 0.0001 | 0.1899 |
| TOF-MLTR | ISA | 0.0951 | 0.0007 | 0.0007 | 0.1819 |
| TOF-MLTR | TOF-SSS | 0.0958 | 0.0011 | 0.0012 | 0.1849 |

Table 3.5: *Comparison of attenuation coefficients in selected regions of the whole-body phantom with three cylindrical inserts derived with different combinations of $\mu$-map estimation and TOF-SC methods ("—" indicates that scatter correction is not required). Abbreviations are explained in Section 2.8.3. The true $\mu$-values in the corresponding ROIs are given for comparison.*

corrections). MLTR-based reconstructions yielded 0.1875 cm$^{-1}$ and 0.1899 cm$^{-1}$ with ISA and TOF-SSS SC, respectively, i.e. approximately a 3% overestimate. All MLAA-estimated air attenuation coefficients were close to zero ([0.0000–0.0012] cm$^{-1}$). MLTR-based MLAA reconstructions yielded air attenuation values closest to the true value, presumably due to faster convergence. Mean background $\mu$-values were in the range of [0.0948–0.0958] cm$^{-1}$ for all the MLAA implementations, slightly underestimating the attenuation coefficient of water. The true value was reproduced exactly (at the specified level of accuracy) by the $\mu$-map scaling procedure.

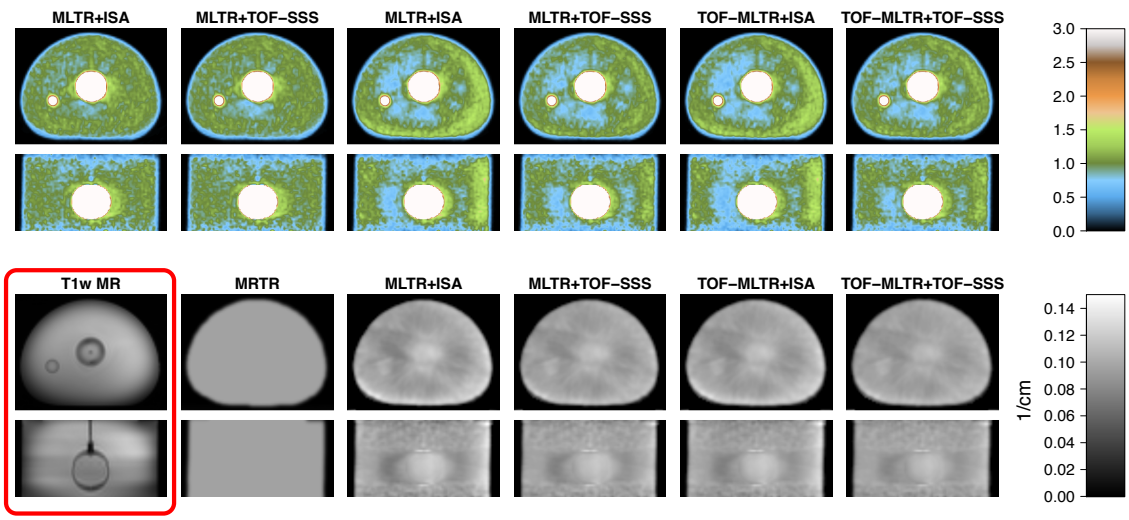


Figure 3.15: *Transaxial (rows 1, 3) and coronal (rows 2, 4) views of activity images (rows 1, 2) and attenuation maps (rows 3, 4) of the pelvis region phantom. Each column represents a different combination of $\mu$-map estimation and TOF-SC methods. Image captions are explained in Section 2.8.3. A T1-weighted MR image (T1w MR) is shown for comparison.*

Results for the pelvis region phantom are shown in Fig. 3.15. This extreme case of very high activity concentration contrasts and nearly perfectly homogeneous attenuation map (neglecting the very minor influence of structural materials like supporting rods and insert walls) is not handled completely satisfactory by MLAA (as is also apparent in Table 3.6): only in the halo region the attenuation coefficient is estimated correctly.

Elsewhere, $\mu$ is overestimated to a different extent. The largest overestimation is observed in the bladder insert with the least serious overestimation obtained by a combination of TOF-MLTR and TOF-SSS (yielding $\mu = 0.1075$ cm$^{-1}$). In general, one can conclude that joint reconstruction with TOF-MLTR and TOF-SSS produces the least biased $\mu$-map among the considered MLAA implementations but still does exhibit substantial artifacts which in turn lead to errors in the reconstructed activity distribution.

| $\mu$-map estimation method | SC method | Reconstructed attenuation (1/cm) | | | |
|---|---|---|---|---|---|
| | | Lesion | Bladder | Background | Halo |
| True values | — | 0.0960 | 0.0960 | 0.0960 | 0.0960 |
| MRTR | — | 0.0960 | 0.0960 | 0.0959 | 0.0960 |
| MLTR | ISA | 0.1016 | 0.1138 | 0.1054 | 0.0961 |
| MLTR | TOF-SSS | 0.1026 | 0.1100 | 0.1010 | 0.0966 |
| TOF-MLTR | ISA | 0.1019 | 0.1111 | 0.1046 | 0.0965 |
| TOF-MLTR | TOF-SSS | 0.1030 | 0.1075 | 0.1007 | 0.0963 |

Table 3.6: *Comparison of attenuation coefficients in selected regions of the pelvis region phantom derived with different combinations of $\mu$-map estimation and TOF-SC methods ("—" indicates that scatter correction is not required). Abbreviations are explained in Section 2.8.3. True $\mu$-values in the corresponding ROIs are given for comparison.*

The propagation of the attenuation map artifacts to the activity image is clearly visible in Fig. 3.15. The MLAA-reconstructed images demonstrate discernible spurious background inhomogeneities and left-to-right asymmetry. The latter effect is more noticeable for ISA-based reconstructions. The ROI analysis results in Table 3.7 underline that non-regularized joint reconstruction is very sensitive to minor uncertainties in the estimated scatter (which are mostly irrelevant for standard MLEM). For example, using ISA for scatter correction in MLAA leads to 15.2–16.7% activity overestimation in the bladder and 13.0–15.0% overestimation in the background. Utilizing TOF-SSS instead reduces the respective values to 8.8–9.7% and 4.7–6.1%. The difference between TOF and non-TOF MLTR, on the other hand, is only minimal and does not exceed 1.5% for most of the ROIs.

| $\mu$-map estimation method | SC method | Reconstructed activity deviation (%) | | | |
|---|---|---|---|---|---|
| | | $\Delta$ Lesion | $\Delta$ Bladder | $\Delta$ Background | $\Delta$ Halo |
| MRTR | ISA | −5.7 | −0.4 | 1.5 | 2.5 |
| MRTR | TOF-SSS | −5.9 | −0.6 | 0.2 | −0.5 |
| MLTR | ISA | −2.1 | 16.7 | 15.0 | −2.6 |
| MLTR | TOF-SSS | −2.8 | 9.7 | 4.7 | −5.8 |
| TOF-MLTR | ISA | −1.8 | 15.2 | 13.0 | −6.5 |
| TOF-MLTR | TOF-SSS | −1.4 | 8.8 | 6.1 | −5.4 |

Table 3.7: *Comparison of reconstructed activities in selected regions of the pelvis region phantom derived with different combinations of $\mu$-map estimation and TOF-SC methods. Abbreviations are explained in Section 2.8.3. Shown are the deviations from true target region activities.*

### Image quality: clinical studies

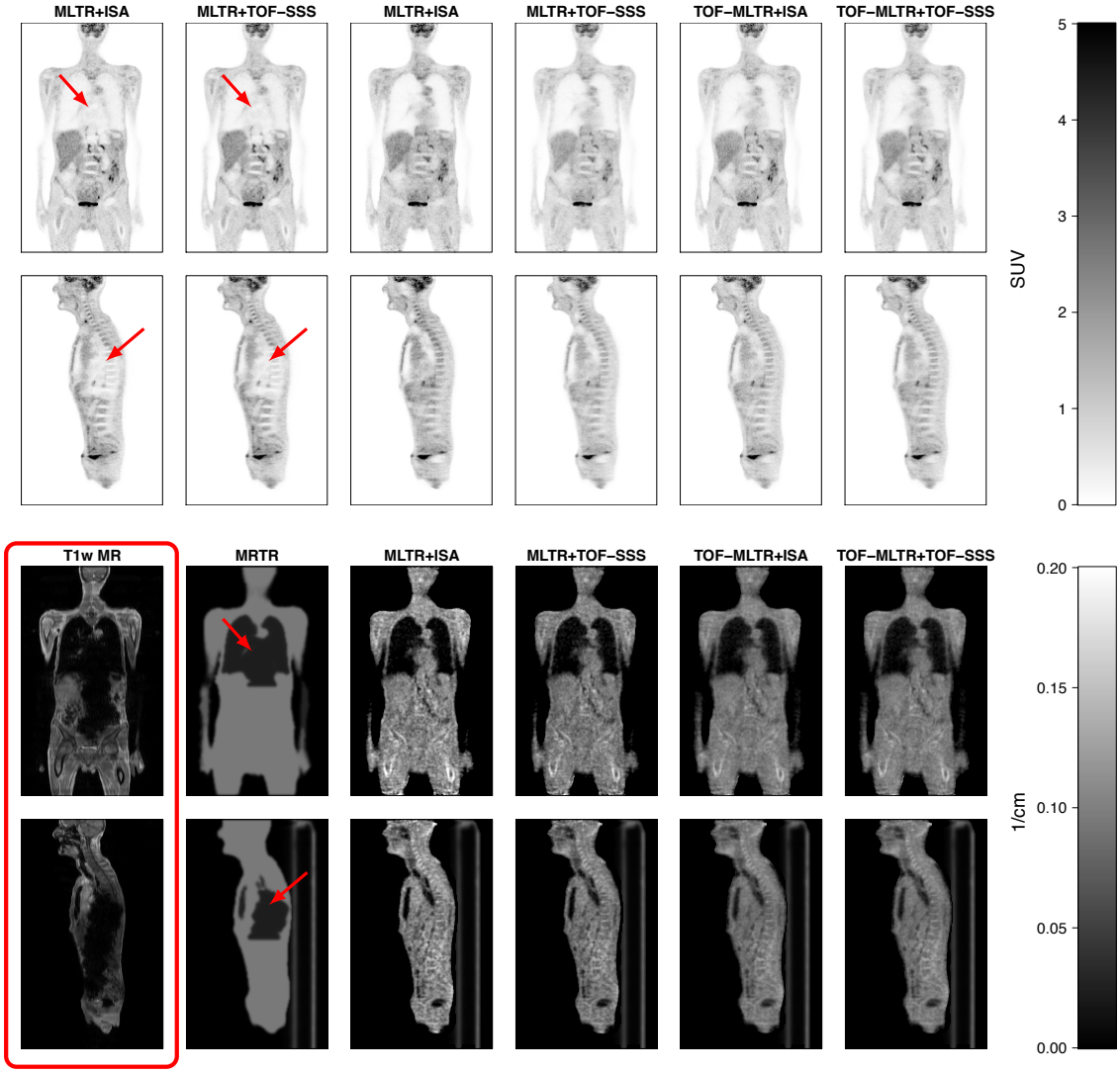


Figure 3.16: *Comparison of activity images (top rows) and attenuation maps (bottom rows) of the whole-body patient study obtained using different $\mu$-map estimation methods in conjunction with different TOF-SC methods. Top and bottom rows in each group show coronal and sagittal views, respectively. Image captions are explained in Section 2.8.3. T1-weighted MR image (framed red) is given in the bottom row for comparison. Artifacts in automatically generated MRTR attenuation map and corresponding activity images are shown by arrows.*

Reconstructed activity and attenuation images of a clinical whole-body investigation (patient B, Section 2.6.2) obtained with different combinations of $\mu$-map estimation and TOF-SC methods are shown in Fig. 3.16. The MR image exhibits a severe metal artifact in the abdomen and thorax area overlapping partly with the lungs. The MRTR generation tool automatically identifies the lungs as a region with very low MR-signal and suitable size and position. As such, the attenuation coefficient was correctly assigned to the lungs, however, a large part of the metal artifact was also incorrectly identified as part of the lung. This leads to a massive activity drop in the thorax in the reconstructed images.

Application of the joint reconstruction techniques allows to determine the actual lungs boundaries and correctly estimate the attenuation coefficients in the surrounding regions. Additionally, MLAA recovered the correct position of air in the whole respiratory tract as well as the air cavity behind the bladder which is also visible in the MR image. Finally, the large osseous structures such as the pelvic bone, femur, and spine are identifiable in the reconstructed $\mu$-maps.

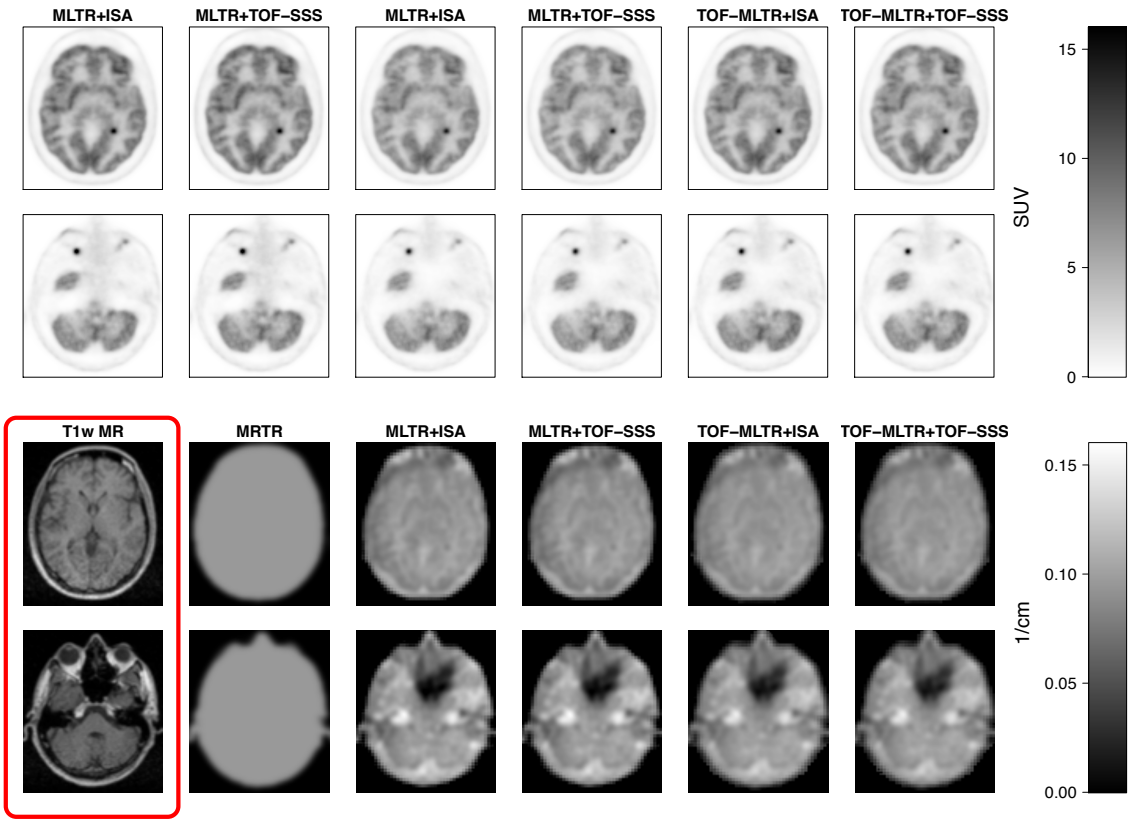


Figure 3.17: *Comparison of activity images (top rows) and attenuation maps (bottom rows) of the brain patient study obtained using different $\mu$-map estimation methods in conjunction with different TOF-SC methods. Top and bottom rows in each group show different coronal slices of the same image. Image captions are explained in Section 2.8.3. T1-weighted MR image (framed red) is given in the bottom row for comparison.*

Reconstructed activity and attenuation images of a clinical brain investigation (patient B, Section 2.6.2) reconstructed using different implementations of MLEM and MLAA are shown in Fig. 3.17. The MRTR $\mu$-map generation tool does not account for osseous structures at all but simply fills the delineated head contour with the attenuation coefficient of water. MLAA algorithms, on the other hand, are able to identify these structures as well as the air in the nasal cavity, paranasal sinuses, and inner ear. All recovered structures with low attenuation correspond to low signal intensity regions in attenuation MR image. However, only larger bones were properly recovered by MLAA due to partial volume effects. Moreover, a difference in $\mu$-values between gray and white matter is visible in the upper row of the attenuation images. This can be attributed to cross-talk between attenu-

ation and activity as there is a pronounced difference in glucose metabolism between gray and white matter while both tissue types have essentially the same attenuation coefficient.

In these clinical examples, differences between ISA and TOF-SSS based reconstructions are very small. The apparent differences in images, like the one seen in ISA + MLTR versus TOF-SSS + MLTR in Fig. 3.16, are presumably due to imperfections of the activity and attenuation scaling procedure used in the present investigation. The differences between non-TOF and TOF MLTR are distinctly more pronounced: non-TOF MLTR leads to a higher air-to-tissue contrast in the $\mu$-maps, presumably due to faster convergence in low attenuation regions. On the other hand, lower $\mu$-map noise was observed with TOF-MLTR when applied to the relatively low-statistics whole-body investigation. The values of 5.5% versus 7.9% (fractional standard deviations) were obtained in a liver ROI with the TOF and the non-TOF MLTR, respectively. The reduced amount of noise in the $\mu$-maps achieved with TOF-MLTR translates into minor (only about 1%) noise reduction in the corresponding activity images in the given example.

# Chapter 4

# Discussion

The present work addresses two persistent issues of image reconstruction for time-of-flight (TOF) PET: improvement of emission-based attenuation correction and time efficiency of TOF scatter correction. Due to the missing capability to measure the attenuation directly, improving attenuation correction using the MLAA technique is of special relevance for PET/MR while accelerating TOF scatter correction is of equal importance for TOF-capable PET/CT systems as well.

The following discussion is subdivided into three main parts, mostly adhering to the same logical structure as Sections 2.7-2.8 and Chapter 3.

## 4.1 Optimization of reconstruction parameters

A crucial requirement for accurate scatter correction is reliable scaling of the derived scatter distribution. We chose a weighted scatter tail fitting technique for this purpose. The performed measurements (Fig. 3.1) and their analysis (Table 3.1) revealed that rather strong weighting of the LORs in the scatter tails with $(1/a_i)^4$ ($a_i$ is the LOR-specific attenuation factor) led to the best results. This weighting strongly favors LORs crossing the patient bed (i.e. those not crossing the phantom or patient but still exhibiting notable attenuation). Other tail fitting approaches, such as exclusive usage of closest-to-object LORs or unweighted fitting, demonstrated inferior quality in our evaluation. One possible explanation for this behavior could be that the SSS algorithm potentially predicts scatter distributions with higher accuracy for LORs intersecting attenuating structures devoid of activity (like the patient bed) compared to "empty" LORs. Further investigation of this conjecture would, however, require a dedicated simulation study which could not be included in the present investigation. In the present context, the empirical finding that the applied weighting scheme and resulting scaling of the scatter distribution leads to satisfactory quantitative performance of the scatter correction is sufficient. Very recently, a seemingly promising new alternative solution to the scatter scaling problem has been proposed (maximum likelihood scatter scaling (Rezaei et al., 2017a)). It would be interesting, in the future, to compare performance of this new approach with the one developed in the

present work.

Knowledge of the actual time resolution operational in the considered PET scan is mandatory for a viable MLAA implementation. Since vendor-provided figures regarding the time resolution are not necessarily reliable and do not cover count-rate-dependent effects at all, a new algorithm was developed and implemented to determine the time resolution as a function of count rate. This algorithm (MLRES) is based on the maximum likelihood principle. Using phantom data covering the full dynamic range of relevant count rates, we observed a perfectly linear dependency of time resolution on singles rate (Fig. 3.2). This observation is in concordance with previously published results for the Philips Gemini TF PET/CT scanner (Philips Healthcare, Best, The Netherlands) (Surti et al., 2007) which uses a PET design very similar to that of the Ingenuity PET/MR. Quantitatively, however, there are differences regarding the count-rate dependency of the time resolution between both systems. Our results demonstrate that at count rates close to zero the Ingenuity PET/MR possesses approximately 30 ps better time resolution than the Gemini PET/CT. On the other hand, the time resolution of the Ingenuity PET/MR degrades about two times faster than that of the Gemini PET/CT.

Stability and consistency of the MLRES procedure were investigated by also evaluating various additional phantom and clinical data sets (Fig. 3.4). We could demonstrate very good concordance of results for most of the data sets which implies that the procedure is not notably affected by changes in the regional activity distribution. For one of the performed measurements this assessment is not strictly correct, however. This measurement was performed with a homogeneously filled whole-body phantom. At higher count rates, systematic deviations from the time resolution expected according to the preceding calibration measurement can be seen in Fig. 3.4. Since the deviations are much smaller than the changes in time resolution over the practically relevant range of count rates, the performed time resolution calibration is not invalidated although a small negative bias would be the consequence for this special geometry/phantom. Considering possible explanations for the discrepant behavior of the homogeneously filled whole-body phantom, two things come to mind: this configuration exhibits a high scatter fraction and, more relevant, is devoid of any internal structures which could provide additional information to the algorithm. This conjecture is supported by the observation that adding cylindrical inserts devoid of activity (and thus visible structure) to the phantom restores complete concordance with the calibration measurement, see Fig. 3.4. In any case, as already stated above, the object dependence is a minor effect that manifests itself only in extreme corner cases without invalidating use of MLRES for determination of a sufficiently accurate value of the effective time resolution.

As with some iterative algorithms (like MLAA), convergence of MLRES to the desired solution (here: global maximum of the likelihood) is not automatically guaranteed. Our investigation of the convergence properties revealed, however, that the algorithm always unambiguously converges to the same solution, independent of the chosen start value

(initial estimate of the time resolution) as Fig. 3.5 demonstrates.

Overall, MLRES thus presents a viable means to reliably determine the count-rate-dependent time resolution of a given PET system. As mentioned before, this information is especially important for MLAA reconstruction: as Fig. 3.6 demonstrates, underestimating the actual TOF kernel width leads to clear activity and attenuation artifacts at the rim of the imaged object. This observation also is in concordance with the literature (Zhu et al., 2017).

In the present context, it is relevant to mention that the effective time resolution (the one accounting for all sources of errors) depends on the accuracy of the time offsets calibration (Clementel et al., 2013; Werner & Karp, 2013). Time offsets are additive corrections which have to be applied to the photon registration time differences in order to compensate for variations in detector response speed. These corrections are determined for each LOR during the routine scanner maintenance. However, the accuracy at which these correction factors are calculated was found to be insufficient for joint reconstruction (Nuyts et al., 2018) and more accurate data-driven calibration methods have been proposed (Rezaei et al., 2017b; Defrise et al., 2018). Published simulation results demonstrate that in the context of MLAA 10% error in the TOF offsets is equivalent to 20% error in TOF kernel width used during reconstruction (Zhu et al., 2017). Therefore, we assume that implementation of the improved time offsets calibration scheme would substantially reduce effective time resolution (as estimated with MLRES) and, consequently, would allow using narrower TOF kernels without appearance of rim MLAA artifacts. This assumption has to be, however, verified in the future studies.

A further parameter that needs to be adjusted is the used scatter TOF bin width or, equivalently, the total number of TOF bins used for scatter time distributions storage (not to be confused with the TOF bins used in the coincidence list-mode data). As usual, it is desirable to minimize the number of bins — and, therefore, memory demand — as far as possible without causing notable discretization errors. Due to the absence of the fine details in the scatter time profiles, the scatter TOF bin width can be chosen to be rather large without introducing notable errors. In the present investigation, between 25 and 50 TOF bins turned out to be a reasonable choice where the lower number was selected for whole-body studies to comply with the available memory resources. In this case, the bin width equals 240 ps. Although this is only 2–3 times smaller than the time resolution of the scanner ($>$ 600 ps in clinical operation), no adverse effects on the accuracy of the scatter correction are observed, see Fig. 3.3. This can be qualitatively understood from Fig. 3.9 which demonstrates that the TOF profiles for patient-sized objects are much broader than the time resolution of the considered system and that the stated sampling width actually suffices to capture the relevant information contained in the profiles (this is especially true for the profiles after convolution with the given time resolution of the scanner, see the right-hand side in Fig. 3.9).

## 4.2 ISA

One of the main results of the present work is that ISA-based scatter correction yields reconstructed images that are virtually identical to those obtained using full TOF-SSS while accelerating the computation of the required TOF scatter by a factor of 2–5. Our phantom results demonstrate that ISA, like TOF-SSS, is capable of minimizing scatter related image artifacts and deviations from true contrast between different regions even under extreme high contrast conditions. According to these results, ISA could be used as a drop-in replacement for TOF-SSS in clinical PET image reconstruction without compromising quantitative accuracy.

The very similar quantitative performance of ISA and TOF-SSS can be traced back to the fact that the resulting scatter TOF weights are almost identical, see Fig. 3.9. The TOF-weights are computed from an estimate of the scatter time distributions (corresponding to infinite time resolution) by convolution with a scanner-specific TOF kernel. The scatter in the LOR considered for illustration has a non-trivial time profile shape (superposition of a broad and a sharply peaked distribution) caused by immediate vicinity of the central sphere of the phantom. Despite these challenging conditions, the scatter time distribution estimated by TOF-SSS is remarkably similar to its ISA approximation. Remaining differences are negligible in view of the scanner's limited time resolution. The scatter TOF weights in Fig. 3.9 were calculated for the modest time resolution of the Philips Ingenuity PET/MR scanner used in the present study, but the same reasoning will remain valid for any realistic time resolution (say $\geq 200\,\mathrm{ps}$). Altogether, it is the good approximation of the scatter time distributions which justifies the ISA approach.

In comparison to ISA and TOF-SSS, the other reconstruction and scatter correction schemes have to be considered unsatisfactory to a varying extent. The discrepancies are most pronounced in the phantom measurements shown in Fig. 3.7 due to the chosen very high contrast but are also visible in the clinical data sets in Figures 3.10 and 3.11. Regarding the vendor provided reconstruction, no implementation details are available and we are not in a position to explain the reason for the observed overcorrection of scatter in most areas. The effect becomes severe at high contrasts and generates notable artifacts not only in the phantom but in the whole-body scan in Figures 3.10, too.

Compared to non-TOF THOR, ISA and TOF-SSS reconstructions clearly show the benefits of fully utilizing available TOF information in order to improve quantitative accuracy and avoidance of image artifacts. Regarding specific reasons for inferior performance of non-TOF THOR, we verified that insufficient convergence of non-TOF vs. TOF reconstruction is not the cause (increasing the iteration count did not lead to relevant changes). Rather, the differences seemingly have to be attributed to principally inferior performance of non-TOF reconstruction and scatter correction. Regarding the clinical data sets, they might also be partly explained by the known sensitivity of non-TOF reconstruction to possible inconsistencies/errors in the MR-derived attenuation maps (which are much better tolerated by TOF reconstruction).

Compared to the non-TOF reconstruction, the attempt to combine TOF reconstruction with non-TOF scatter correction (by distributing the scatter equally over all TOF bins) leads to distinctly more severe image artifacts and increased quantitative errors resulting from a rather pronounced under-correction of the scatter (and thus a positive bias in the resulting images). Explanation of this observation is rather straightforward if one considers the scatter TOF profiles in Fig. 3.9 which reveal that the scatter is by no means uniformly distributed over the TOF bins but rather localized within the imaged object boundary. Disregarding this behavior results, in our case, in scatter underestimation by a factor of up to 6.

The speed advantage of ISA over TOF-SSS arises from the fact that it requires neither TOF forward projections in the SSS caching step nor heavy cached arrays manipulations and merging for each scatter point and LOR in the combination step, effectively replacing them by very few (about 5) floating point operations. The reduction in overall image reconstruction time achieved when replacing full TOF-SSS by ISA depends of course on the considered specific implementation of iterative image reconstruction as well as on the input data.

In this study, we have used THOR, our tube-of-response list-mode OSEM reconstruction. This type of reconstruction, notably the resolution modeling achieved by the tube-of-response approach, is computationally quite intensive and THOR thus is not as fast as highly optimized scanner specific sinogram based reconstructions. Nevertheless, we observe nearly a factor of two reduction in overall computation time of whole-body studies when performing the initial scatter time distribution computation with ISA rather than with TOF-SSS. This demonstrates that in this case the scatter estimation procedure is responsible for a large fraction of the computation time. The acceleration factor reduces to 1.4 if a single bed position of a whole-body study is being reconstructed whereas the scatter estimation alone is accelerated by a factor of about five, demonstrating the reduced contribution of scatter correction to total reconstruction time. For the long-duration/high-statistics brain scan, the overall reconstruction time — which in this case is completely dominated by the actual image reconstruction process — is only reduced by about 4%.

To better understand the rather pronounced differences between the different scenarios (whole-body, low statistics/large object single-bed, high statistics/small object single-bed scan) it first should be noted that the computation times for MLEM update and scatter simulation, respectively, scale differently with scan statistics, number of bed positions, and dimensions of the reconstruction grid. Second, the number of emission/scatter points over which the SC algorithm iterates is much higher for scans of thorax or abdomen than in brain investigations due to distinctly larger transaxial object extension. The number of emission/scatter points is further increased in multi-bed studies by utilizing adjacent bed positions to also account for out-of-FOV scatter (Section 2.2.6) which explains the discrepancy between one-bed and multi-bed whole-body reconstruction acceleration factors.

Overall, it is thus clear that the benefits of switching from TOF-SSS to ISA are most

important for whole-body studies but also are notable for single-bed scans of the thorax or abdomen. In view of the fact that oncological whole-body studies are currently by far the most frequent application of clinical PET, an achievable factor of up to two reduction in overall reconstruction time for this type of study when using TOF-aware scatter correction seems highly relevant.

The influence of small variations/errors in the underlying attenuation map (e.g., differences between attenuation maps derived in PET/MR and PET/CT, respectively) has been found to be negligible for non-TOF SSS in brain PET (Burgos et al., 2014; Teuho et al., 2017) and the situation should be similar for whole-body PET. ISA does not utilize attenuation information beyond the non-TOF SSS calculation effectively used for correct scaling of the TOF profiles. It might thus be concluded that ISA is not affected any more by uncertainties of the attenuation map than non-TOF SSS.

ISA also has some potential shortcomings. Unlike TOF-SSS, ISA does not account for the influence of spatially variant attenuation along $ES$ (Fig. 2.6) on the resulting TOF scatter profiles. However, in our data, we were so far not able to see any adverse consequences of this fact, e.g. in the lungs in Fig. 3.10. Nevertheless, a more comprehensive investigation of this question might be worthwhile.

Another potential shortcoming of ISA is that it does not explicitly account for the influence of the actual scatter medium geometry on the contributions from different emission points. This could be modeled in (2.41) by an additional emission point position and scatter angle dependent factor accounting for the difference between ISA and TOF-SSS predictions such as those shown in Fig. 2.8. However, the presented results and preliminary testing of tentative further modifications to ISA do not indicate that this effect is a practically relevant issue, presumably, because the averaging of the contributions from a huge number of emission points in (2.41) leads to canceling of the residual differences present for the individual emission points.

## 4.3 MLAA

The implemented TOF-extension of the THOR reconstruction tool together with development of a time resolution calibration method provided the basis for a viable implementation of the MLAA algorithm. Our implementation has some unique traits.

First of all, there are only very few list-mode MLAA (LM-MLAA) implementations with scatter and randoms corrections included (Cheng et al., 2016b). Only by including these corrections, LM-MLAA can be sensibly applied to clinical data sets and its usefulness can be evaluated beyond simulations and phantom studies.

Another unique feature of our LM-MLAA implementation is the possibility to use either non-TOF or TOF-aware LM-MLTR for the attenuation map updates and to compare both approaches directly within the framework of otherwise identical image reconstruction. Our corresponding results show that TOF-MLTR is more accurate than the non-TOF version, see Fig. 3.12 and Fig. 3.13. It also exhibits lower noise levels in the reconstructed

attenuation maps ($\mu$-maps) and activity images. This is especially noticeable in whole-body investigations (Fig. 3.16). On the other hand, TOF-MLTR converges slightly slower in regions with low attenuation coefficients such as air cavities but the effect is only minor and of no practical relevance. It is nevertheless interesting to understand this behavior. On first glance, it seems somewhat counter-intuitive since TOF-MLEM not only exhibits reduced noise but also converges faster than non-TOF MLEM and the same might be expected for TOF-MLTR.

The explanation for this discrepant behavior of MLEM and MLTR is related to the fundamentally different impact of TOF information on activity and attenuation reconstruction, respectively. Since attenuation for a coincidence event in the given LOR does not depend on the location of the photon pair creation, TOF information bears no additional value for transmission reconstruction if only true coincidences are considered. Utilization of the TOF information only affects the ratio between estimated additive contributions (randoms, scatter) and estimated trues. The improved noise characteristics and accuracy of TOF-MLTR can, thus, be understood as a consequence of better consistency of randoms and scatter correction in TOF-MLTR and TOF-MLEM as well as the fact that these two algorithms optimize the same TOF-likelihood function (2.42) while regular MLTR optimizes the non-TOF likelihood (1.25). On the other hand, convergence speed differences of two MLTR implementations is, presumably, the consequence of the bias in the converged estimate of $\mu$-values with non-TOF MLTR.

MLAA is very sensitive to data inconsistencies caused by inaccuracies in scatter correction and scanner calibration (Zhu et al., 2017; Nuyts et al., 2018). For this reason, it was important to evaluate the consequences of using the newly proposed ISA scatter correction in the context of MLAA. The direct comparison of ISA and TOF-SSS (with TOF and non-TOF MLTR) shows that in most cases the difference between using either ISA or TOF-SSS for MLAA is negligible and much lower then the difference between both MLAA implementations (incorporating either non-TOF or TOF-MLTR) when using the same scatter correction approach, see Figures 3.14, 3.16, and 3.17. Only for the extreme high activity contrast realized in the pelvis region phantom, a difference between ISA and TOF-SSS becomes apparent. In fact, neither of both methods led to satisfactory MLAA results in this challenging case. For instance, up to 20% variations of $\mu$-values were observed in the reconstructed attenuation maps of this completely water-filled phantom (ignoring the very minor distortions caused by support rods and insert walls), see Table 3.6. Furthermore, a clear left-to-right asymmetry is apparent which is somewhat more pronounced for ISA. Altogether, the investigated MLAA variants are not able to deal with such extreme cases with a combination of very high activity contrasts and very low attenuation contrast.

Regarding performance for more usual clinical conditions, our comparison between MR-based and emission-based attenuation correction demonstrates that the implemented MLAA algorithms are capable of recovering air cavities and osseous structures in the $\mu$-maps. These structures are missing in the MR-based attenuation maps (which also

were used for initialization of the MLAA reconstruction). The obtained $\mu$-values are in reasonable agreement with the expected population-based values. Unfortunately, no CT-based $\mu$-maps were available which could have served as a reference for evaluation of absolute quantitative accuracy with MLAA *in vivo* in the patient studies. At least for the phantom study shown in Fig. 3.14, the true attenuation coefficients of the different regions are reproduced with reasonable accuracy, demonstrating the principle usefulness and potential of MLAA.

One especially interesting aspect of MLAA for PET/MR concerns the ability of MLAA to reduce metal-implants-induced artifacts which can spoil MR-based $\mu$-maps. Despite some progress regarding automated metal artifacts correction (Schramm et al., 2014), the fundamental problem of tissue misclassification in regions affected by MR susceptibility artifacts is not solved as demonstrated by Fig. 3.16. The Figure shows a patient with stent implant causing a massive MR signal void in the chest and abdomen. In this example, this spurious signal void intersects the lungs and prevents proper lung segmentation of the MR image and thus leads to an erroneous attenuation map. A dedicated correction tool developed in our group was used in this work and allowed supression of artifact propagation into the abdomen but still failed for the chest region. MLAA, on the other hand, can handle the presence of metal implants quite well as was confirmed in the present investigation. This finding is also in accord with the literature (Fuin et al., 2017).

Despite these promising results, the implemented MLAA algorithm still have severe shortcomings which prevent immediate clinical application. First, the implemented algorithm for global $\mu$-map scaling is not sufficiently reliable: while working well in our phantom studies, it sometimes fails for clinical data sets see Fig. 3.16. This is a well known problem and several solutions have already been proposed in the literature (Cheng et al., 2016a; Feng et al., 2018; Rezaei et al., 2018). Combining the available MR information with MLAA is an obvious possibility. Notably, MR-based MLAA regularization seems to be an appealing approach. Since MR excels in imaging of soft tissues, the information provided by this modality might be utilized to drive the MLAA estimate of the attenuation coefficients for soft-tissue voxels — as determined from MR data — to the correct value (near 0.096 $cm^{-1}$) (Ahn et al., 2012; Mehranian & Zaidi, 2014). Feasibility of this approach has been reported (Mehranian & Zaidi, 2015). More recently, even the possibility to apply MR-enhanced MLAA for non-TOF joint reconstruction tasks has been reported (Heußer et al., 2016; Benoit et al., 2016).

MR-based regularization seems to be the solution for another problem, also observed in our MLAA implementation (Ahn et al., 2018). This is the non-negligible local cross-talk between attenuation and activity. Despite the theoretically demonstrated resolution of the cross-talk problem by utilization of TOF information (Defrise et al., 2012), we were not able to confirm a complete cross-talk artifact elimination in practice, see Fig. 3.17. However, this might be a consequence of the relatively low time resolution of our PET/MR system together with possibly inaccurate TOF-offsets calibration (Nuyts et al., 2018). Finally, utilization of the MR-derived priors allows to extend MLAA applications for non-FDG

tracers with more specific uptake in target regions (Ahn et al., 2018). This would be relevant, e.g., for $^{18}$F-Fluoride studies (Grant et al., 2007) which are targeting osseous structures that are generally affected by erroneous MR-based attenuation correction.

Overall, more work on MLAA will be necessary to unambiguously demonstrate its suitability for clinical routine use. This includes implementing of MR-based MLAA regularization which is going to be our next step. To the extent that some of the remaining problems are related to small residual errors of the available scatter correction procedures, the development of advanced SC methods beyond single-scatter-approximation-based approaches seems to be worthwhile. After these steps, an improved PET image reconstruction with THOR utilizing MLAA for attenuation correction could be introduced into clinical routine to provide physicians more quantitatively accurate images for cancer diagnostics and therapy assessment.

# Summary

The present work addresses two persistent issues of image reconstruction for time-of-flight (TOF) PET: acceleration of TOF scatter correction and improvement of emission-based attenuation correction. Due to the missing capability to measure photon attenuation directly, improving attenuation correction by joint reconstruction of the activity and attenuation coefficient distribution using the MLAA technique is of special relevance for PET/MR while accelerating TOF scatter correction is of equal importance for TOF-capable PET/CT systems as well.

To achieve the stated goals, in a first step the high-resolution PET image reconstruction THOR, previously developed in our group, was adapted to take advantage of the TOF information delivered by state-of-the-art PET systems. TOF-aware image reconstruction reduces image noise and improves convergence rate both of which is highly desirable.

Based on these adaptations, this thesis describes new developments for improvement of TOF scatter correction and MLAA reconstruction and reports results obtained with the new algorithms on the Philips Ingenuity PET/MR jointly operated by the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) and the University Hospital.

A crucial requirement for quantitative TOF image reconstruction is TOF-aware scatter correction. The currently accepted reference method — the TOF extension of the single scatter simulation approach (TOF-SSS) — was implemented as part of the TOF-related modifications of THOR. The major drawback of TOF-SSS is a 3–7 fold increase in computation time required for the scatter estimation, compared to regular SSS, which in turn does lead to a considerable image reconstruction slowdown. This problem was addressed by development and implementation of a novel accelerated TOF scatter correction algorithm called ISA. This new algorithm proved to be a viable alternative to TOF-SSS and speeds up scatter correction by a factor of up to five in comparison to TOF-SSS. Images reconstructed using ISA are in excellent quantitative agreement with those obtained when using TOF-SSS while overall reconstruction time is reduced by a factor of two in whole-body investigations. This can be considered a major achievement especially with regard to the use of advanced image reconstruction in a clinical context.

The second major topic of this thesis is contribution to improved attenuation correction in PET/MR by utilization of MLAA reconstruction. First of all, knowledge of the actual time resolution operational in the considered PET scan is mandatory for a viable MLAA implementation. Since vendor-provided figures regarding the time resolution are

not necessarily reliable and do not cover count-rate dependent effects at all, a new algorithm was developed and implemented to determine the time resolution as a function of count rate. This algorithm (MLRES) is based on the maximum likelihood principle and allows to determine the functional dependency of the time resolution of the Philips Ingenuity PET/MR on the given count rate and to integrate this information into THOR. Notably, the present work proves that the time resolution of the Ingenuity PET/MR can degrade by more than 250 ps for the clinically relevant range of count rates in comparison to the vendor-provided figure of 550 ps which is only realized in the limit of extremely low count rates.

Based on the previously described developments, MLAA could be integrated into THOR. The performed list-mode MLAA implementation is capable of deriving realistic, patient-specific attenuation maps. Especially, correct identification of osseous structures and air cavities could be demonstrated which is very difficult or even impossible with MR-based approaches to attenuation correction. Moreover, we have confirmed that MLAA is capable of reducing metal-induced artifacts which are otherwise present in MR-based attenuation maps. However, the detailed analysis of the obtained MLAA results revealed remaining problems regarding stability of global scaling as well as local cross-talk between activity and attenuation estimates. Therefore, further work beyond the scope of the present work will be necessary to address these remaining issues.

# Zusammenfassung

In der vorliegenden Dissertation wurden zwei fortdauernde Probleme der Bildrekonstruktion in der *time-of-flight* (TOF) PET bearbeitet: Beschleunigung der TOF-Streukorrektur sowie Verbesserung der emissionsbasierten Schwächungskorrektur. Aufgrund der fehlenden Möglichkeit, die Photonenabschwächung direkt zu messen, ist eine Verbesserung der Schwächungskorrektur durch eine gemeinsame Rekonstruktion der Aktivitäts- und Schwächungskoeffizienten-Verteilung mittels der MLAA-Methode von besonderer Bedeutung für die PET/MRT, während eine Beschleunigung der TOF-Streukorrektur gleichermaßen auch für TOF-fähige PET/CT-Systeme relevant ist.

Für das Erreichen dieser Ziele wurde in einem ersten Schritt die hochauflösende PET-Bildrekonstruktion THOR, die bereits zuvor in unserer Gruppe entwickelt wurde, angepasst, um die TOF-Information nutzen zu können, welche von allen modernen PET-Systemen zur Verfügung gestellt wird. Die Nutzung der TOF-Information in der Bildrekonstruktion führt zu reduziertem Bildrauschen und zu einer verbesserten Konvergenzgeschwindigkeit.

Basierend auf diesen Anpassungen werden in der vorliegenden Arbeit neue Entwicklungen für eine Verbesserung der TOF-Streukorrektur und der MLAA-Rekonstruktion beschrieben. Es werden sodann Ergebnisse vorgestellt, welche mit den neuen Algorithmen am Philips Ingenuity PET/MRT-Gerät erzielt wurden, das gemeinsam vom Helmholtz-Zentrum Dresden-Rossendorf (HZDR) und dem Universitätsklinikum betrieben wird.

Eine wesentliche Voraussetzung für eine quantitative TOF-Bildrekonstruktionen ist eine Streukorrektur, welche die TOF-Information mit einbezieht. Die derzeit übliche Referenzmethode hierfür ist eine TOF-Erweiterung des *single scatter simulation* Ansatzes (TOF-SSS). Diese Methode wurde im Rahmen der TOF-Erweiterung von THOR implementiert. Der größte Nachteil der TOF-SSS ist eine 3–7-fach erhöhte Rechenzeit für die Berechnung der Streuschätzung im Vergleich zur non-TOF-SSS, wodurch die Bildrekonstruktionsdauer deutlich erhöht wird. Um dieses Problem zu beheben, wurde eine neue, schnellere TOF-Streukorrektur (ISA) entwickelt und implementiert. Es konnte gezeigt werden, dass dieser neue Algorithmus eine brauchbare Alternative zur TOF-SSS darstellt, welche die Rechenzeit auf ein Fünftel reduziert, wobei mithilfe von ISA und TOF-SSS rekonstruierte Schnittbilder quantitativ ausgezeichnet übereinstimmen. Die Gesamtrekonstruktionszeit konnte mithilfe ISA bei Ganzkörperuntersuchungen insgesamt um den Faktor Zwei reduziert werden. Dies kann als maßgeblicher Fortschritt betrachtet wer-

den, speziell im Hinblick auf die Nutzung fortgeschrittener Bildrekonstruktionsverfahren im klinischen Umfeld.

Das zweite große Thema dieser Arbeit ist ein Beitrag zur verbesserten Schwächungskorrektur in der PET/MRT mittels MLAA-Rekonstruktion. Hierfür ist zunächst eine genaue Kenntnis der tatsächlichen Zeitauflösung in der betrachten PET-Aufnahme zwingend notwendig. Da die vom Hersteller zur Verfügung gestellten Zahlen nicht immer verlässlich sind und zudem die Zählratenabhängigkeit nicht berücksichtigen, wurde ein neuer Algorithmus entwickelt und implementiert, um die Zeitauflösung in Abhängigkeit von der Zählrate zu bestimmen. Dieser Algorithmus (MLRES) basiert auf dem *maximum likelihood* Prinzip und erlaubt es, die funktionale Abhängigkeit der Zeitauflösung des Philips Ingenuity PET/MRT von der Zählrate zu bestimmen. In der vorliegenden Arbeit konnte insbesondere gezeigt werden, dass sich die Zeitauflösung des Ingenuity PET/MRT im klinisch relevanten Zählratenbereich um mehr als 250 ps gegenüber der vom Hersteller genannten Auflösung von 550 ps verschlechtern kann, welche tatsächlich nur bei extrem niedrigen Zählraten erreicht wird.

Basierend auf den oben beschrieben Entwicklungen konnte MLAA in THOR integriert werden. Die MLAA-Implementierung erlaubt die Generierung realistischer patientenspezifischer Schwächungsbilder. Es konnte insbesondere gezeigt werden, dass auch Knochen und Hohlräume korrekt identifiziert werden, was mittels MRT-basierter Schwächungskorrektur sehr schwierig oder sogar unmöglich ist. Zudem konnten wir bestätigen, dass es mit MLAA möglich ist, metallbedingte Artefakte zu reduzieren, die ansonsten in den MRT-basierten Schwächungsbildern immer zu finden sind. Eine detaillierte Analyse der Ergebnisse zeigte allerdings verbleibende Probleme bezüglich der globalen Skalierung und des lokalen Übersprechens zwischen Aktivitäts- und Schwächungsschätzung auf. Daher werden zusätzliche Entwicklungen erforderlich sein, um auch diese Defizite zu beheben.

# List of Abbreviations

**3D-RAMLA** 3D Row-Action Maximum Likelihood Algorithm

**$^{18}$F-FDG** 2-deoxy-2-($^{18}$F)fluoro-D-glucose

**APD** Avalanche Photodiode

**CNN** Convolutional Neural Networks

**CPU** Central Processing Unit

**CT** Computed Tomography

**FBP** Filtered Backprojection

**FOV** Field-Of-View

**FWHM** Full Width at Half Maximum

**HZDR** Helmholtz-Zentrum Dresden-Rossendorf

**ISA** Immediate Scatter Approximation

**LM** List-Mode

**LOR** Line-Of-Response

**LYSO** Lutetium-yttrium oxyorthosilicate

**MAP** Maximum A Posteriori

**MLAA** Maximum Likelihood Activity and Attenuation

**MLEM** Maximum Likelihood Expectation Maximization

**MLRES** Maximum Likelihood Time Resolution Estimation

**MLTR** Maximum Likelihood for Transmission tomography

**MRI/MR** Magnetic Resonance Imaging

**MRTR** In-house MR-based $\mu$-map generator

**ooFOV** out-of-FOV

**OSEM** Ordered Subsets Estimation Maximization

**PET** Positron Emission Tomography

**PMT** Photomultiplier tube

**PTFE** Polytetrafluoroethylene

**RAM** Random Access Memory

**ROI** Region-of-interest

**SC** Scatter correction

**SiPM** Silicon photomultiplier

**SPECT** Single Photon Emission Computed Tomography

**SSS** Single Scatter Simulation

**SUV** Standardized uptake value

**THOR** Tube of response High resolution OSEM Reconstruction

**TOF** Time-Of-Flight

**TOR** Tube-Of-Response

**UV** Ultraviolet

# List of Figures

# List of Tables

# Bibliography

Accorsi R, Adam LE, Werner ME, Karp JS. 2004. Optimization of a fully 3D single scatter simulation algorithm for 3D PET. Physics in medicine and biology 49:2577–2598. 42

Ahn S, Cheng L, Shanbhag DD, Qian H, Kaushik SS, Jansen FP, Wiesinger F. 2018. Joint estimation of activity and attenuation for PET using pragmatic MR-based prior: application to clinical TOF PET/MR whole-body data for FDG and non-FDG tracers. Physics in Medicine & Biology 63:045006. 8, 102, 103

Ahn S, Qian H, Manjeshwar RM. 2012. Convergent iterative algorithms for joint reconstruction of activity and attenuation from time-of-flight PET data. In: 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC). IEEE, 3695–3700. 7, 63, 102

Akbarzadeh A, Ay MR, Ahmadian A, Riahi Alam N, Zaidi H. 2013. MRI-guided attenuation correction in whole-body PET/MR: Assessment of the effect of bone attenuation. Annals of Nuclear Medicine 27:152–162. 7

Badawi RD, Miller MP, Bailey DL, Marsden PK. 1999. Randoms variance reduction in 3D PET. Physics in Medicine and Biology 44:941–954. 44

Bailey DL, Townsend DW, Valk PE, Maisey MN (eds) . 2005. Positron Emission Tomography. Springer-Verlag, London. 22, 23, 28

Benoit D, Ladefoged CN, Rezaei A, Keller SH, Andersen FL, Højgaard L, Hansen AE, Holm S, Nuyts J. 2016. Optimized MLAA for quantitative non-TOF PET/MR of the brain. Physics in Medicine and Biology 61:8854–8874. 7, 102

Berker Y, Li Y. 2016. Attenuation correction in emission tomography using the emission data — A review. Medical Physics 43:807–832. 8

Bezrukov I, Schmidt H, Mantlik F, Schwenzer N, Brendle C, Scholkopf B, Pichler BJ. 2013. MR-Based Attenuation Correction Methods for Improved PET Quantification in Lesions Within Bone and Susceptibility Artifact Regions. Journal of Nuclear Medicine :1–7. 7

Boellaard R, Hofman MBM, Hoekstra OS, Lammertsma AA. 2014. Accurate PET/MR Quantification Using Time of Flight MLAA Image Reconstruction. Molecular Imaging and Biology . 8

Burgos N, Thielemans K, Cardoso MJ, Markiewicz P, Jiao J, Dickson J, Duncan JS, Atkinson D, Arridge SR, Hutton BF, Ourselin S. 2014. Effect of scatter correction when comparing attenuation maps: Application to brain PET/MR. In: 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–5. 100

Casey ME, Hoffman EJ. 1986. Quantitation in Positron Emission Computed Tomography: 7. A Technique to Reduce Noise in Accidental Coincidence Measurements and Coincidence Efficiency Calibration. Journal of Computer Assisted Tomography 10:845–850. 44

Castellaz P, Siegle A, Stoll H. 2002. Positron Age-Momentum-Correlation (AMOC) Measurements on Organic Liquids. Journal of Nuclear and Radiochemical Sciences 3:R1–R7. 15

Cheng JCK, Salomon A, Yaqub M, Boellaard R. 2016a. Impact of mis-matched TOF kernel in TOF-OSEM and TOF-MLAA reconstructions. Journal of Nuclear Medicine 57:203. 9, 71, 74, 102

Cheng JCK, Salomon A, Yaqub M, Boellaard R. 2016b. Investigation of practical initial attenuation image estimates in TOF-MLAA reconstruction for PET/MR. Medical Physics 43:4163–4173. 10, 55, 62, 100

Clementel E, Mollet P, Vandenberghe S. 2013. Effect of local TOF kernel miscalibrations on contrast-noise in TOF PET. IEEE Transactions on Nuclear Science 60:1578–1588. 97

Conti M. 2011. Why is TOF PET reconstruction a more robust method in the presence of inconsistent data? Physics in medicine and biology 56:155–68. 7, 47

Conti M, Bendriem B, Casey M, Chen M, Kehren F, Michel C, Panin V. 2005. First experimental results of time-of-flight reconstruction on an LSO PET scanner. Physics in medicine and biology 50:4507–4526. 6, 8, 45, 47, 49

Conti M, Eriksson L. 2016. Physics of pure and non-pure positron emitters for PET: A review and a discussion. EJNMMI Physics 3. 11, 14, 15

Daube-Witherspoon ME, Surti S, Matej S, Werner M, Jayanthi S, Karp JS. 2006. Influence of Time-of-Flight Kernel Accuracy in TOF-PET Reconstruction. In: 2006 IEEE Nuclear Science Symposium Conference Record, vol. 3. IEEE, 1723–1727. 71

Defrise M, Rezaei A, Nuyts J. 2012. Time-of-flight PET data determine the attenuation sinogram up to a constant. Physics in Medicine and Biology 57:885–899. 7, 31, 47, 55, 102

Defrise M, Rezaei A, Nuyts J. 2018. Time-of-flight PET time calibration using data consistency. Physics in Medicine & Biology 63:105006. 97

Delso G, Furst S, Jakoby B, Ladebeck R, Ganter C, Nekolla SG, Schwaiger M, Ziegler SI. 2011. Performance Measurements of the Siemens mMR Integrated Whole-Body PET/MR Scanner. Journal of Nuclear Medicine 52:1914–1922. 18

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B (Methodological) 39:1–38. 133

Ehrhardt MJ, Thielemans K, Pizarro L, Atkinson D, Ourselin S, Hutton BF, Arridge SR. 2015. Joint reconstruction of PET-MRI by exploiting structural similarity. Inverse Problems 31. 25

El Fakhri G, Surti S, Trott CM, Scheuermann J, Karp JS. 2011. Improvement in Lesion Detection with Whole-Body Oncologic Time-of-Flight PET. Journal of Nuclear Medicine 52:347–353. 47

Feng T, Wang J, Li H. 2018. Joint activity and attenuation estimation for PET with TOF data and single events. Physics in Medicine & Biology 63:245017. 102

Fuin N, Pedemonte S, Catalano OA, Izquierdo-Garcia D, Soricelli A, Salvatore M, Heberlein K, Hooker JM, Van Leemput K, Catana C. 2017. PET/MRI in the Presence of Metal Implants: Completion of the Attenuation Map from PET Emission Data. Journal of Nuclear Medicine 58:840–845. 7, 102

Gambhir SS, Czernin J, Schwimmer J, Silverman DH, Coleman RE, Phelps ME. 2001. A Tabulated Summary of the FDG PET Literature. Journal of nuclear medicine : official publication, Society of Nuclear Medicine 42:1S–93S. 13

Gong K, Yang J, Kim K, El Fakhri G, Seo Y, Li Q. 2018. Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images. Physics in Medicine & Biology 63:125011. 7

Grant FD, Fahey FH, Packard AB, Davis RT, Alavi A, Treves ST. 2007. Skeletal PET with 18F-Fluoride: Applying New Technology to an Old Tracer. Journal of Nuclear Medicine 49:68–78. 103

Han X. 2017. MR-based synthetic CT generation using a deep convolutional neural network method. Medical Physics 44:1408–1419. 6, 31

Harpen MD. 2004. Positronium: review of symmetry, conserved quantities and decay for the radiological physicist. Medical physics 31:57–61. 15

Heußer T, Rank CM, Berker Y, Freitag MT, Kachelrieß M. 2017. MLAA-based attenuation correction of flexible hardware components in hybrid PET/MR imaging. EJNMMI Physics 4. 62

Heußer T, Rank CM, Freitag MT, Dimitrakopoulou-Strauss A, Schlemmer HP, Beyer T, Kachelriess M. 2016. MR-Consistent Simultaneous Reconstruction of Attenuation and Activity for Non-TOF PET/MR. IEEE Transactions on Nuclear Science 63:2443–2451. 7, 102

Hu Z, Ojha N, Renisch S, Schulz V, Torres I, Buhl A, Pal D, Muswick G, Penatzer J, Guo T, Bönert P, Tung C, Kaste J, Morich M, Havens T, Maniawski P, Schäfer W, Günther RW, Krombach GA, Shao L. 2009. MR-based Attenuation Correction for a Whole-body Sequential PET/MR System. In: IEEE Nuclear Science Symposium Conference Record. IEEE Nuclear Science Symposium, 3508–3512. 6, 34

Hubbell JH. 1969. Photon cross sections, attenuation coefficients, and energy absorption coefficients from 10 keV to 100 GeV. Natl Standards Reference Data System Natl Bureau of Standards 29 :87. 16

Hubbell JH. 1999. Review of photon interaction cross section data in the medical and biological context. Physics in Medicine and Biology 44. 15

Hudson HM, Larkin RS. 1994. Ordered Subsets of Projection Data. IEEE Trans. Med. Imaging 13:601–609. 35, 36

Hwang D, Kim KY, Kang SK, Seo S, Paeng JC, Lee DS, Lee JS. 2018. Improving the Accuracy of Simultaneously Reconstructed Activity and Attenuation Maps Using Deep Learning. Journal of Nuclear Medicine 59:1624–1629. 7

Jin X, Chan C, Mulnix T, Panin V, Casey ME, Liu C, Carson RE. 2013. List-mode reconstruction for the Biograph mCT with physics modeling and event-by-event motion correction. Physics in Medicine and Biology 58:5567–5591. 8

Jinghan Ye, Xiyun Song, Zhiqiang Hu. 2014. Scatter correction with combined single-scatter simulation and Monte Carlo simulation for 3D PET. In: 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–3. 70

Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW. 2009. Impact of Time-of-Flight on PET Tumor Detection. Journal of Nuclear Medicine 50:1315–1323. 47

Keereman V, Holen RV, Mollet P, Vandenberghe S. 2011. The effect of errors in segmented attenuation maps on PET quantification. Medical physics 38:6010–6019. 7

Keller SH, Holm S, Hansen AE, Sattler B, Andersen F, Klausen TL, Højgaard L, Kjær A, Beyer T. 2013. Image artifacts from MR-based attenuation correction in clinical, whole-body PET/MRI. Magnetic Resonance Materials in Physics, Biology and Medicine 26:173–181. 7

Kelloff GJ, Hoffman JM, Johnson B, Scher HI, Siegel BA, Cheng EY, Cheson BD, O'Shaughnessy J, Guyton KZ, Mankoff DA, Shankar L, Larson SM, Sigman CC, Schilsky RL, Sullivan DC. 2005. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. Clinical cancer research : an official journal of the American Association for Cancer Research 11:2785–808. 13

Kim K, Wu D, Gong K, Dutta J, Kim JH, Son YD, Kim HK, El Fakhri G, Li Q. 2018. Penalized PET Reconstruction Using Deep Learning Prior and Local Linear Fitting. IEEE Transactions on Medical Imaging 37:1478–1487. 25

Kinahan PE, Hasegawa BH, Beyer T. 2003. X-ray-based attenuation correction for positron emission tomography/computed tomography scanners. Seminars in nuclear medicine 33:166–79. 30

Klein O, Nishina Y. 1929. Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. Zeitschrift für Physik 52:853–868. 16, 41

Ladefoged CN, Andersen FL, Keller SH, Löfgren J, Hansen AE, Holm S, Højgaard L, Beyer T. 2013. PET/MR imaging of the pelvis in the presence of endoprostheses: reducing image artifacts and increasing accuracy through inpainting. European Journal of Nuclear Medicine and Molecular Imaging 40:594–601. 7

Lamarche F, Leroy C. June 1990. Evaluation of the volume of intersection of a sphere with a cylinder by elliptic integrals. Computer Physics Communications 59 Issue 2:359–369. 40

Lange K, Fessler J. 1995. Globally convergent algorithms for maximum a posteriori transmission tomography. IEEE Transactions on Image Processing 4:1430–1438. 58

Lecomte R. 2009. Novel detector technology for clinical PET. European journal of nuclear medicine and molecular imaging 36 Suppl 1:S69–85. 17, 18

Levkovitz R, Falikman D, Zibulevsky M, Ben-Tal A, Nemirovski A. 2001. The design and implementation of COSEM, an iterative algorithm for fully 3-D listmode data. IEEE transactions on medical imaging 20:633–42. 37

Lewitt RM. 1990. Multidimensional digital image representation using generalired Kaiser-Bessel window functions. Journal of the Optical Society of America A 7:1834–46. 25

Ley R, Werth G. 2001. The Hydrogen Atom: Precision Physics of Simple Atomic Systems, chap. Positronium: Theory Versus Experiment. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 407–418. 15

Lougovski A. 2012. High Resolution Fully 3D List-mode PET Reconstruction with Integrated Event-based Motion Compensation. Ph.D. thesis, Technische Universität, Medizinische Fakultät Dresden. 12, 20, 38, 39, 42, 111

Lougovski A, Hofheinz F, Maus J, Schramm G, van den Hoff J. 2015. On the relation between Kaiser-Bessel blob and tube of response based modelling of the system matrix in iterative PET image reconstruction. Physics in Medicine and Biology 60:4209–4224. 9, 35, 39

Lougovski A, Hofheinz F, Maus J, Schramm G, Will E, van den Hoff J. 2014. A volume of intersection approach for on-the-fly system matrix calculation in 3D PET image reconstruction. Physics in Medicine and Biology 59:561–577. 9, 35, 38, 39

Markiewicz PJ, Tamal M, Julyan PJ, Hastings DL, Reader aJ. 2007. High accuracy multiple scatter modelling for 3D whole body PET. Physics in medicine and biology 52:829–47. 32

Martinez-Möller A, Souvatzoglou M, Delso G, Bundschuh Ra, Chefd'hotel C, Ziegler SI, Navab N, Schwaiger M, Nekolla SG. 2009. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. Journal of Nuclear Medicine 50:520–6. 6, 31, 35

Matej S, Lewitt R. 1996. Practical Considerations for 3-D Image Reconstruction Using Spherically Symmetric volume Elements. IEEE Transactions on Medical Imaging 15:68–78. 25

Maus J, Schramm G, Christ U. 2014. A Technical Report on Philips PET SUV Calibration/Validation procedures. Technical report, Helmholtz-Zentrum Dresden-Rossendorf. 64

Mehranian A, Zaidi H. 2014. Joint estimation of activity and attenuation in PET/MR using MR-constrained Gaussian priors. In: 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–4. 7, 102

Mehranian A, Zaidi H. 2015. Clinical Assessment of Emission- and Segmentation-Based MR-Guided Attenuation Correction in Whole-Body Time-of-Flight PET/MR Imaging. Journal of Nuclear Medicine 56:877–883. 8, 102

Mérida I, Reilhac A, Redouté J, Heckemann RA, Costes N, Hammers A. 2017. Multi-atlas attenuation correction supports full quantification of static and dynamic brain PET data in PET-MR. Physics in Medicine and Biology 62:2834–2858. 7

Mollet P, Vandenberghe S. 2014. Comparison of transmission- and emission-based attenuation correction for TOF-PET/MRI. In: 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–5. 10, 55

Mühlematter UJ, Nagel HW, Becker A, Mueller J, Vokinger KN, de Galiza Barbosa F, ter Voert EEGT, Veit-Haibach P, Burger IA. 2018. Impact of time-of-flight PET on quantification accuracy and lesion detection in simultaneous $^{18}$F-choline PET/MRI for prostate cancer. EJNMMI Research 8:41. 47

Nuyts J, Bal G, Kehren F, Fenchel M, Michel C, Watson C. 2013. Completion of a truncated attenuation image from the attenuated PET emission data. IEEE Transactions on Medical Imaging 32:237–246. 62

Nuyts J, Dupont P, Stroobants S, Benninck R, Mortelmans L, Suetens P. 1999. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. IEEE Transactions on Medical Imaging 18:393–403. 7, 55, 62

Nuyts J, Man BD, Dupont P, Defrise M, Suetens P, Mortelmans L. 1998. Iterative reconstruction for helical CT: a simulation study. Physics in Medicine and Biology 43:729–737. 7, 55

Nuyts J, Rezaei A, Defrise M. 2018. The Validation Problem of Joint Emission/Transmission Reconstruction From TOF-PET Projections. IEEE Transactions on Radiation and Plasma Medical Sciences 2:273–278. 9, 74, 97, 101, 102

Parra L, Barrett HH. 1998. List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET. IEEE transactions on medical imaging 17:228–235. 36

Politte DG, Snyder DL. 1991. Corrections for Accidental Coincidences and Attenuation in Maximum-Likelihood Image Reconstruction for Positron-Emission Tomography. IEEE Trans. Med. Imaging 10:82–89. 45

Polycarpou I, Thielemans K, Manjeshwar R, Aguiar P, Marsden PK, Tsoumpas C. 2011. Comparative evaluation of scatter correction in 3D PET using different scatter-level approximations. Annals of Nuclear Medicine 25:643–649. 35, 70

Presotto L, Bettinardi V, Gianolli L, Perani D. 2015. Alternating strategies and ordered subset acceleration schemes for maximum likelihood activity and attenuation reconstruction in time-of-flight PET. In: 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–4. 60

Rausch I, Ruiz A, Valverde-Pascual I, Cal-Gonzalez J, Beyer T, Carrio I. 2018. Performance evaluation of the Philips Vereos PET/CT system according to the NEMA NU2-2012 standard. Journal of Nuclear Medicine [in press] DOI: 10.2967/jnumed.118.215541. 18, 21

Rezaei A, Bickell M, Fulton R, Nuyts J. 2015. Joint activity and attenuation reconstruction of listmode TOF-PET data. In: 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–3. 10, 55

Rezaei A, Defrise M, Bal G, Michel C, Conti M, Watson C, Nuyts J. 2012. Simultaneous reconstruction of activity and attenuation in time-of-flight PET. IEEE transactions on Medical Imaging 31:2224–33. 7, 31, 55, 62

Rezaei A, Defrise M, Nuyts J. 2014. ML-reconstruction for TOF-PET with Simultaneous Estimation of the Attenuation Factors. IEEE transactions on medical imaging 33:1563–1572. 7

Rezaei A, Deroose CM, Vahle T, Boada F, Nuyts J. 2018. Joint Reconstruction of Activity and Attenuation in Time-of-Flight PET: A Quantitative Analysis. Journal of Nuclear Medicine 59:1630–1635. 8, 102

Rezaei A, Michel C, Casey ME, Nuyts J. 2016. Simultaneous reconstruction of the activity image and registration of the CT image in TOF-PET. Physics in Medicine and Biology 61:1852–1874. 7, 56, 58, 59

Rezaei A, Salvo K, Vahle T, Panin V, Casey M, Boada F, Defrise M, Nuyts J. 2017a. Plane-dependent ML scatter scaling: 3D extension of the 2D simulated single scatter (SSS) estimate. Physics in Medicine & Biology 62:6515–6531. 70, 95

Rezaei A, Schramm G, Van Laere K, Nuyts J. 2017b. Data driven time alignment for TOF-PET. In: 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–3. 97

Rota Kops E, Hautzel H, Herzog H, Lerche C, Shah NJ. 2017. Attenuation Correction of Cerebellum in PET/MR Data. In: 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–2. 7

Rota Kops E, Herzog H. 2007. Alternative methods for attenuation correction for PET images in MR-PET scanners. In: 2007 IEEE Nuclear Science Symposium Conference Record, vol. 6. IEEE, 4327–4330. 6, 31

Salomon A, Goedicke A, Schweizer B, Aach T, Schulz V. 2011. Simultaneous reconstruction of activity and attenuation for PET/MR. IEEE Transactions on Medical Imaging 30:804–13. 7, 55

Salvo K, Defrise M. 2017. sMLACF: a generalized expectation-maximization algorithm for TOF-PET to reconstruct the activity and attenuation simultaneously. Physics in Medicine & Biology 62:8283–8313. 7

Samarin A, Burger C, Wollenweber SD, Crook DW, Burger Ia, Schmid DT, von Schulthess GK, Kuhn FP. 2012. PET/MR imaging of bone lesions — implications for PET quantification from imperfect attenuation correction. European Journal of Nuclear Medicine and Molecular Imaging 39:1154–1160. 7

Schlyer DJ. 2004. PET tracers and radiochemistry. Annals of the Academy of Medicine, Singapore 33:146–54. 12

Schramm G, Holler M, Rezaei A, Vunckx K, Knoll F, Bredies K, Boada F, Nuyts J. 2018. Evaluation of Parallel Level Sets and Bowsher's Method as Segmentation-Free Anatomical

Priors for Time-of-Flight PET Reconstruction. IEEE Transactions on Medical Imaging 37:590–603. 25

Schramm G, Langner J, Hofheinz F, Petr J, Lougovski A, Beuthien-Baumann B, Platzek I, van den Hoff J. 2013. Influence and Compensation of Truncation Artifacts in MR-Based Attenuation Correction in PET/MR. IEEE Transactions on Medical Imaging 32:2056–2063. 35

Schramm G, Maus J, Hofheinz F, Petr J, Lougovski A, Beuthien-Baumann B, Platzek I, van den Hoff J. 2014. Evaluation and automatic correction of metal-implant-induced artifacts in MR-based attenuation correction in whole-body PET/MR imaging. Physics in Medicine and Biology 59:2713–2726. 7, 9, 35, 102

Schulz V, Torres-Espallardo I, Renisch S, Hu Z, Ojha N, Börnert P, Perkuhn M, Niendorf T, Schäfer WM, Brockmann H, Krohn T, Buhl A, Günther RW, Mottaghy FM, Krombach GA. 2011. Automatic, three-segment, MR-based attenuation correction for whole-body PET/MR data. European Journal of Nuclear Medicine and Molecular Imaging 38:138–152. 6, 31, 34

Sekine T, Buck A, Delso G, ter Voert EEGW, Huellner M, Veit-Haibach P, Warnock G. 2016. Evaluation of Atlas-Based Attenuation Correction for Integrated PET/MR in Human Brain: Application of a Head Atlas and Comparison to True CT-Based Attenuation Correction. Journal of Nuclear Medicine 57:215–220. 6

Shepp LA, Vardi Y. 1982. Maximum likelihood reconstruction for emission tomography. IEEE Transactions on Medical Imaging 1:113–22. 7, 35, 133

Sols A, Crane RK. 1954. Substrate specificity of brain hexokinase. The Journal of biological chemistry 210:581–595. 13

Sousa JM, Appel L, Engström M, Papadimitriou S, Nyholm D, Larsson EM, Ahlström H, Lubberink M. 2018. Evaluation of zero-echo-time attenuation correction for integrated PET/MR brain imaging — comparison to head atlas and $^{68}$Ge-transmission-based attenuation correction. EJNMMI Physics 5:20. 7

Surti S, Karp JS, Popescu LM, Daube-Witherspoon ME, Werner M. 2006. Investigation of time-of-flight benefit for fully 3-D PET. IEEE Transactions on Medical Imaging 25:529–538. 47

Surti S, Kuhn A, Werner ME, Perkins AE, Kolthammer J, Karp JS. 2007. Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. Journal of Nuclear Medicine 48:471–480. 9, 21, 34, 48, 71, 96

ter Voert EE, Veit-Haibach P, Ahn S, Wiesinger F, Khalighi MM, Levin CS, Iagaru AH, Zaharchuk G, Huellner M, Delso G. 2017. Clinical evaluation of TOF versus non-TOF on

PET artifacts in simultaneous PET/MR: a dual centre experience. European Journal of Nuclear Medicine and Molecular Imaging 44:1223–1233. 47

Teuho J, Saunavaara V, Tolvanen T, Tuokkola T, Karlsson A, Tuisku J, Teräs M. 2017. Quantitative Evaluation of two Scatter Correction Techniques for $^{18}$F-FDG Brain PET/MRI in Regard to MR-Based Attenuation Correction. Journal of Nuclear Medicine 58:1691–1698. 100

Tewson TJ, Krohn Ka. 1998. PET radiopharmaceuticals: state-of-the-art and future prospects. Seminars in nuclear medicine 28:221–34. 13

Turkington T, Wilson J. 2009. Attenuation artifacts and time-of-flight PET. In: 2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC). IEEE, 2997–2999. 7, 47

Vallabhajosula S, Solnes L, Vallabhajosula B. 2011. A Broad Overview of Positron Emission Tomography Radiopharmaceuticals and Clinical Applications: What Is New? Seminars in Nuclear Medicine 41:246–264. 12

van den Hoff J. 2005. Principles of quantitative positron emission tomography. Amino Acids 29:341–353. 5

Van Slambrouck K, Nuyts J. 2014. Reconstruction Scheme for Accelerated Maximum Likelihood Reconstruction: The Patchwork Structure. IEEE Transactions on Nuclear Science 61:173–181. 56, 58

Vandenberghe S, Matej S, Daube-Witherspoon M, Guerchaft M, Verhaeghe J, Bol A, van Elmbt L, Lemahieu I, Karp J. 2007. Determining timing resolution from TOF-PET emission data. In: 2007 IEEE Nuclear Science Symposium Conference Record, vol. 4. IEEE, 2727–2731. 9, 71

Vunckx K, Zhou L, Matej S, Defrise M, Nuyts J. 2010. Fisher information-based evaluation of image quality for time-of-flight PET. IEEE transactions on medical imaging 29:311–21. 6, 47

Wang W, Hu Z, Gagnon D. 2007. A New Component Approach to Efficiency Normalization for 3D PET. IEEE Transactions on Nuclear Science 54:92–99. 40

Watson C. 2000. New, Faster, Image-Based Scatter Correction for 3D PET. IEEE Transations on Nuclear Science 47:1587–1594. 8, 43, 55

Watson CC. 2007. Extension of Single Scatter Simulation to Scatter Correction of Time-of-Flight PET. IEEE Transations on Nuclear Science 54:1679–1686. 8

Werner ME, Karp JS. 2013. TOF PET offset calibration from clinical data. Physics in Medicine and Biology 58:4031–4046. 97

Werner ME, Surti S, Karp JS. 2006. Implementation and Evaluation of a 3D PET Single Scatter Simulation with TOF Modeling. In: 2006 IEEE Nuclear Science Symposium Conference Record. IEEE, 1768–1773. 8

Wiesinger F, Bylund M, Yang J, Kaushik S, Shanbhag D, Ahn S, Jonsson JH, Lundman JA, Hope T, Nyholm T, Larson P, Cozzini C. 2018. Zero TE-based pseudo-CT image conversion in the head and its application in PET/MR attenuation correction and MR-guided radiation therapy planning. Magnetic Resonance in Medicine 80:1440–1451. 7

Yavuz M, Fessler JA. 1997. New statistical models for randoms-precorrected pet scans. In: Duncan J, Gindi G (eds) Information Processing in Medical Imaging. Springer Berlin Heidelberg, Berlin, Heidelberg, 190–203. 27

Zaidi H, Ojha N, Morich M, Griesmer J, Hu Z, Maniawski P, Ratib O, Izquierdo-Garcia D, Fayad Za, Shao L. 2011. Design and performance evaluation of a whole-body Ingenuity TF PET-MRI system. Physics in medicine and biology 56:3091–106. 33, 34

Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. 2018. Image reconstruction by domain-transform manifold learning. Nature 555:487–492. 22

Zhu W, Feng T, Chen M, Dong Y, Bao J, Li H. 2016. Ultra efficient and robust estimation of the attenuation map in PET imaging. In: 2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD), vol. 3. IEEE, 1–3. 7

Zhu W, Feng T, Dong Y, Bao J, Li H. 2017. A Systematic Study on Factors Influencing the Accuracy of MLAA. In: 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 1–3. 9, 97, 101

# Appendix: derivation of the MLEM algorithm

Here we provide a brief justification and derivation of the MLEM algorithm. We will be following the definitions and assumptions introduced in Section 1.3.2. Ignoring scatter and random events for now, we presume the forward model (1.16), no prior knowledge, and pure Poisson noise (1.19). Therefore, the cost function is given by the Poisson log-likelihood (1.25) and the acquisition process is fully described by the system matrix elements $c_{ij}$. From the probabilistic definition of $c_{ij}$, it is clear that the following condition is fulfilled:

$$S_j = \sum_i c_{ij} \leq 1, \tag{4.1}$$

where $S_j$ is the scanner *sensitivity* to events originating from voxel $j$. It represents the fractional detection probability (ratio of detected to total number of decay events in voxel $j$). $S_j = 1$ would represent the ideal case of a scanner with full ($4\pi$) solid angle coverage and no photon losses, i.e. 100% detection probability. From (4.1) and (1.18) it follows that

$$\sum_i E(n_{ij}|\boldsymbol{\lambda}) = S_j \, E(n_j|\boldsymbol{\lambda}) \, . \tag{4.2}$$

Before we continue let us demonstrate some useful relations which follow from the properties of the Poisson processes we are dealing with. It is well known that for two Poisson variables $X$ and $Y$ with means $\lambda_X$ and $\lambda_Y$, respectively, the conditional expectation of $X$ given $X + Y$ is

$$E(X|X+Y) = (X+Y)\frac{\lambda_X}{\lambda_X + \lambda_Y}. \tag{4.3}$$

Consequently, considering the number of all the detected events originating in voxel $j$, $n_j^d = \sum_i n_{ij}$, and the number of all the photon pairs emitted from $j$, $n_j$, one can conclude that

$$E(n_j^d|n_j) = \frac{n_j}{E(n_j|\boldsymbol{\lambda})} \sum_i E(n_{ij}|\boldsymbol{\lambda}) = n_j S_j. \tag{4.4}$$

We can then calculate the expected number of detected events originating in voxel $j$ when

$\boldsymbol{y}$ and $\boldsymbol{\lambda}$ are given. Directly using the definition of expectation operation we get

$$E(n_j^d|\boldsymbol{\lambda},\boldsymbol{y}) = \sum_{n_j^d=0}^{\infty} n_j^d\, P(n_j^d|\boldsymbol{\lambda},\boldsymbol{y}) = \sum_{n_j^d=0}^{\infty}\sum_{n_j=0}^{\infty} n_j^d\, P(n_j^d|n_j)\, P(n_j|\boldsymbol{\lambda},\boldsymbol{y}) =$$
$$= \sum_{n_j=0}^{\infty} E(n_j^d|n_j)\, P(n_j|\boldsymbol{\lambda},\boldsymbol{y}) = S_j \sum_{n_j=0}^{\infty} n_j\, P(n_j|\boldsymbol{\lambda},\boldsymbol{y})$$

In this way we get a useful relation between conditional expectations of detected and emitted photon pairs originating from a certain voxel which is similar to (4.2) but has a different conditioning:

$$\sum_i E(n_{ij}|\boldsymbol{\lambda},\boldsymbol{y}) = S_j\, E(n_j|\boldsymbol{\lambda},\boldsymbol{y}). \tag{4.5}$$

Considering (1.16), (4.3), and the natural condition $\sum_j n_{ij} = y_i$ we see that

$$E(n_{ij}|\boldsymbol{\lambda},\boldsymbol{y}) = y_i\, \frac{c_{ij}\lambda_j}{\sum\limits_{j\prime} c_{ij\prime}\lambda_{j\prime}}. \tag{4.6}$$

The optimization problem only makes sense if the objective function has one and only one global maximum. The Log-likelihood function satisfies this condition since it is concave which can be seen as follows. By direct differentiation of (1.25) we get

$$\frac{\partial\, l(\boldsymbol{\lambda})}{\partial \lambda_j} = \sum_i \left[ c_{ij}\, \frac{y_i}{\sum\limits_{j\prime} c_{ij\prime}\lambda_{j\prime}} - c_{ij} \right]. \tag{4.7}$$

Using (1.16), (4.6), (1.18), and (4.5) this can be rewritten as

$$\begin{aligned}\frac{\partial\, l(\boldsymbol{\lambda})}{\partial \lambda_j} =& \frac{1}{\lambda_j} \sum_i \left[ y_i\, \frac{c_{ij}\lambda_j}{\sum\limits_{j\prime} c_{ij\prime}\lambda_{j\prime}} - c_{ij}\lambda_j \right] = \\ =& \frac{1}{\lambda_j} \sum_i \left[E(n_{ij}|\boldsymbol{\lambda},\boldsymbol{y}) - E(n_{ij}|\boldsymbol{\lambda})\right] = \\ =& \frac{S_j}{\lambda_j} \left[E(n_j|\boldsymbol{\lambda},\boldsymbol{y}) - E(n_j|\boldsymbol{\lambda})\right].\end{aligned} \tag{4.8}$$

Equation (4.8) demonstrates that if $l(\boldsymbol{\lambda})$ reaches its maximum at a certain $\boldsymbol{\lambda}$ then conditional and unconditional expectations of $n_j$ given $\boldsymbol{\lambda}$ are equal to each other for all voxels $j$. Further differentiation of (4.7) leads us to

$$\frac{\partial^2\, l(\boldsymbol{\lambda})}{\partial \lambda_j \partial \lambda_k} = -\sum_i c_{ij} c_{ik}\, \frac{y_i}{\overline{y}_i{}^2}. \tag{4.9}$$

The matrix elements in (4.9) form the Hessian matrix $H(\boldsymbol{\lambda})$ of the log-likelihood function.

For any $\boldsymbol{x} = (x_1, \ldots, x_J)^T$, the quadratic form $\boldsymbol{x}^T H(\boldsymbol{\lambda})\boldsymbol{x}$ can be represented as

$$\boldsymbol{x}^T\, H(\boldsymbol{\lambda})\, \boldsymbol{x} = -\sum_i \frac{y_i}{\overline{y}_i{}^2} \sum_j c_{ij} x_j \sum_k c_{ik} x_k = -\sum_i \frac{y_i}{\overline{y}_i{}^2} \left( \sum_j c_{ij} x_j \right)^2 \tag{4.10}$$

and since all the summands in (4.10) are non-negative, the quadratic form (and thus the Hessian) is negative semidefinite which proofs that the log-likelihood $l(\boldsymbol{\lambda})$ is concave.

Following (Shepp & Vardi, 1982) we search for the image which maximizes the likelihood by applying the Estimation Maximization algorithm (Dempster et al., 1977). The algorithm consists of two steps:

1. Estimation (*E-step*): Estimate the number of decays in each voxel $j$ from current activity estimate $\boldsymbol{\lambda}^{(p)}$ and measured data $\boldsymbol{y}$

$$n_j^{(p)} = E(n_j | \boldsymbol{\lambda}^{(p)}, \boldsymbol{y}) \tag{4.11}$$

2. Maximization (*M-step*): Determine new image estimate $\boldsymbol{\lambda}^{(p+1)}$ as a solution of

$$E(n_j | \boldsymbol{\lambda}^{(p+1)}) = n_j^{(p)} \tag{4.12}$$

The value $p$ is the iteration counter. From the equations above it follows that if the algorithm converges to a certain image $\boldsymbol{\lambda}^{(\infty)}$ then $\boldsymbol{\lambda}^{(p+1)} = \boldsymbol{\lambda}^{(p)} = \boldsymbol{\lambda}^{(\infty)}$ and, therefore, $E(n_j|\boldsymbol{\lambda}^{(\infty)}) = E(n_j|\boldsymbol{\lambda}^{(\infty)}, \boldsymbol{y})$. Using these expectations in (4.8), we see that $\boldsymbol{\lambda}^{(\infty)}$ actually maximizes the likelihood. Convergence is secured by special properties of Estimation Maximization algorithms: it follows from (Dempster et al., 1977), Theorem 1 that $l(\boldsymbol{\lambda}^{(p+1)}) \geq l(\boldsymbol{\lambda}^{(p)})$ for any $\boldsymbol{\lambda}^{(p+1)}$ satisfying (4.11) and (4.12). Equality holds only if $\boldsymbol{\lambda}^{(p+1)} = \boldsymbol{\lambda}^{(p)}$ which implies convergence. Monotonic increase of the sequence ( $l(\boldsymbol{\lambda}^{(0)})$, $l(\boldsymbol{\lambda}^{(1)}), \ldots$ ) together with continuity and boundedness of the likelihood function $l(\boldsymbol{\lambda})$ following from its concavity implies the convergence of the algorithm and, thus, the existence of the likelihood maximizing solution $\boldsymbol{\lambda}^{(\infty)}$.

The MLEM algorithm can be brought into a more convenient form. Equation in (4.12) is trivial and the right-hand side of (4.11) can be easily obtained from (4.5) and (4.6). Thus, substituting $n_j^{(p)}$ in (4.12) by (4.11) we, finally, get (2.1).

# Acknowledgments

The Ph.D. study is now over and the dissertation is finally finished. The last three years were not only exciting but also difficult for me: new country, new culture, new people, and the whole new scientific field. It would have been impossible for me to reach this point and overcome all difficulties without help from my colleges and friends.

First of all, I would like to express my gratitude and appreciation to my supervisor, Prof. Dr. Jörg van den Hoff who provided me an opportunity to work in his excellent group. I am thankful for his wise guidance, precious advice, meaningful discussions, and, of course, for his patience. I would like to give a special thanks to Dr. Alexandr Lougovski for his kind help in the first few months of my stay in Germany as well as for development and implementation of THOR — the reconstruction tool without which my work would have been impossible. I would also like to thank all my colleges in HZDR and University Hospital Carl Gustav Carus I had the pleasure to work with, especially Dr. Jens Maus, Dr. Frank Hofheinz, Dr. Jan Petr, Dr. Liane Oehme, Prof. Dr. Jörg Kotzerke, Dr. Irina Korovina, Dr. Nadia Licciardello, and Dr. Massimo Sgarzi. Their knowledge, expertise, and assistance were invaluable for me on my way to understanding the new topic of medical physics, writing the thesis and preparing for the defense procedure.

The time I spent working in the Institute of Radiopharmaceutical Cancer Research, HZDR, was truly great! For exceptional working conditions, I would like to express my gratitude to Prof. Dr. Jörg Steinbach and Prof. Dr. Michael Bachmann.

Finally, I sincerely thank my friends and family for the support they gave and continue giving to me when I need it the most.

**Anlage 1**

**Technische Universität Dresden**
**Medizinische Fakultät Carl Gustav Carus**
**Promotionsordnung vom 24. Juli 2011**

**Erklärungen zur Eröffnung des Promotionsverfahrens**

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

2. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten:
Prof. Dr. Jörg van den Hoff, Dr. Jens Maus, Dr. Frank Hofheinz, Dr. Jan Petr, Dr. Nadia Licciardello, Dr. Massimo Sgarzi, Dr. Irina Korovina

3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

5. Die Inhalte dieser Dissertation wurden in folgender Form veröffentlicht:

Publikationen:

i) Nikulin, P., Maus, J., Hofheinz, F., Lougovski, A., van den Hoff, J.
Time efficient scatter correction for time-of-flight PET: the immediate scatter approximation.
Physics in Medicine and Biology [accepted for publication].

Vorträge:

i) Nikulin, P., Lougovski, A., Hofheinz, F., Maus, J., van den Hoff, J.
Time efficient scatter correction in Time-Of-Flight PET image reconstruction
PSMR 2016 - $5^{th}$ Conference on PET/MR and SPECT/MR, Cologne 2016

ii) Nikulin, P., Lougovski, A., Hofheinz, F., Maus, J., van den Hoff, J.
Scatter correction in TOF and non-TOF PET image reconstruction in THOR
Seminar on PET image reconstruction, Leuven 2016

iii) Nikulin, P., Lougovski, A., Hofheinz, F., Maus, J., van den Hoff, J.
Time efficient scatter correction in Time-Of-Flight PET image reconstruction
Jahrestagung der Deutschen Gesellschaft für Nuklearmedizin, Dresden 2017

iv) Nikulin, P., Lougovski, A., Hofheinz, F., Maus, J., van den Hoff, J.
Comparison of MLAA-derived attenuation maps with and without utilisation of time-of-flight information in the attenuation estimation step
PSMR 2017 - $6^{th}$ Conference on PET/MR and SPECT/MR, Lisbon 2017

v) Nikulin, P., Lougovski, A., Hofheinz, F., Maus, J., van den Hoff, J.
A Maximum-Likelihood Timing Resolution Estimation algorithm for TOF-PET
Jahrestagung der Deutschen Gesellschaft für Nuklearmedizin, Bremen 2018

vi) Nikulin, P., Lougovski, A., Hofheinz, F., Maus, J., van den Hoff, J.
An Algorithm for Maximum-Likelihood Estimation of the Timing Resolution in TOF-PET
PSMR 2018 - $7^{th}$ Conference on PET/MR and SPECT/MR, Isola d'Elba 2018

6. Ich bestätige, dass es keine zurückliegenden erfolglosen Promotionsverfahren gab.

7. Ich bestätige, dass ich die Promotionsordnung der Medizinischen Fakultät der Technischen Universität Dresden anerkenne.

8. Ich habe die Zitierrichtlinien für Dissertationen an der Medizinischen Fakultät der Technischen Universität Dresden zur Kenntnis genommen und befolgt.

Ort, Datum

Unterschrift des Doktoranden

**Anlage 2**

**Hiermit bestätige ich die Einhaltung der folgenden aktuellen gesetzlichen Vorgaben im Rahmen meiner Dissertation**

- das zustimmende Votum der Ethikkommission bei Klinischen Studien, epidemiologischen Untersuchungen mit Personenbezug oder Sachverhalten, die das Medizinproduktegesetz betreffen
  *nicht zutreffend*
- die Einhaltung der Bestimmungen des Tierschutzgesetzes
  *nicht zutreffend*
- die Einhaltung des Gentechnikgesetzes
  *nicht zutreffend*
- die Einhaltung von Datenschutzbestimmungen der Medizinischen Fakultät und des Universitätsklinikums Carl Gustav Carus.

Ort, Datum

Unterschrift des Doktoranden