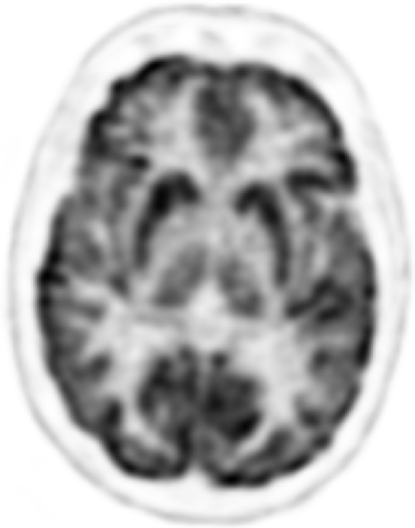


Positron Emission Tomography Image Reconstruction with Time-of-Flight Data Using a SPLAT System Model

Master's thesis in Biomedical engineering and
Engineering mathematics and computational science

Erik Linghammar
Richard Nunstedt

DEPARTMENT OF ELECTRICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

Positron Emission Tomography Image Reconstruction with Time-of-Flight Data Using a SPLAT System Model

ERIK LINGHAMMAR
RICHARD NUNSTEDT

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Computer Vision and Medical Image Analysis group
Chalmers University of Technology
Gothenburg, Sweden 2025

Positron Emission Tomography Image Reconstruction with Time-of-Flight Data Using a SPLAT System Model

ERIK LINGHAMMAR
RICHARD NUNSTEDT

Supervisor:
C. Ross Schmidtlein
Department of Medical Physics
Memorial Sloan Kettering Cancer Center

Supervisor/Examiner:
Ida Häggström
Department of Electrical Engineering
Chalmers University of Technology

Master's Thesis 2025
Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Reconstructed brain image from a PET scan.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Positron Emission Tomography Image Reconstruction with
Time-of-Flight Data Using a SPLAT System Model
Erik Linghammar, Richard Nunstedt
Department of Electrical Engineering
Chalmers University of Technology

Abstract

This thesis presents the development and evaluation of a physics-based system model tailored for image reconstruction in a novel *time-of-flight* (TOF) *positron emission tomography* (PET) brain scanner. The scanner design features a continuous, multi-layered monolithic thin detector, offering improved spatial resolution and sensitivity compared to conventional PET systems. To fully leverage these hardware advances, a custom system model called the SPLAT model was implemented. The SPLAT system model integrates key resolution loss mechanisms, including non-collinearity, detector blur, and *coincidence timing resolution* (CTR), using a Gaussian framework. In addition, positron range and photon attenuation, both implemented in a form viable for 3D, were incorporated to enable more realistic reconstruction. A simulation tool for PET *list-mode* (LM) data was developed based on this model, enabling controlled evaluation of reconstruction performance. Two simulation strategies were implemented to discretize the continuous detector domain. The forward and backprojection operators were implemented with performance optimizations in C and integrated into an iterative *maximum-likelihood expectation-maximization* (ML-EM) reconstruction pipeline. Image quality was quantitatively assessed using normalized RMSE metrics, with analyses explicitly linking reconstruction accuracy to characteristics of the modeled point spread function.

Keywords: Positron Emission Tomography (PET), Time-of-Flight (TOF), Physics-based system model, Image reconstruction, List-mode data, Resolution modeling, Attenuation correction, Positron range.

Acknowledgements

We would like to express our sincere gratitude to our supervisor at Chalmers, Ida Häggström, for her continuous support, kindness, and encouragement throughout the thesis project. Her insightful feedback and thoughtful suggestions have been very valuable to us.

We are also very thankful to our supervisor C. Ross Schmidtlein at Memorial Sloan Kettering Cancer Center, whose work on the SPLAT system model forms the foundation of this thesis. Ross has been incredibly generous with his time and expertise—answering our many questions, inspiring us with his ideas, and providing access to a well-developed codebase that we could adapt and build our code upon. His contributions have been central to the progress of our work.

A warm thank you to our peer reviewer Marcus Hansen for taking the time to read our report and for providing constructive feedback and insightful questions during our thesis defense. We also want to thank Erik Blomqvist for reading and giving valuable feedback on an early draft of our thesis.

We are grateful to Frida Erhard for transforming our sketches into the clear and well-crafted illustrations featured in this report. We also thank Professor Martin Raum for taking time to share useful tips on improving computational performance.

Finally, we'd like to acknowledge the group study rooms — and especially the large whiteboards — in the "Idéläran" building at Chalmers. These spaces have been the setting for countless whiteboard sessions, which have been not only a fun and rewarding part of this thesis but also an important part of our collaboration throughout our years at Chalmers.

Erik Linghammar, Richard Nunstedt, Gothenburg, June 2025

Contents

1
Introduction

Understanding how the human body looks on the inside is a recurring question that medical physicians consider when diagnosing, monitoring and treating patients. Medical imaging is a group of methods that aim to answer this question without cutting into the body or inserting physical devices [1]. Within this field, nuclear medicine imaging images the biodistribution of radiolabeled compounds, providing insight into physiological and biochemical processes. This distinguishes it from anatomical imaging modalities such as *computed tomography* (CT) and magnetic resonance imaging, which primarily depict structural information [1]. Among the modalities in nuclear medicine, *positron emission tomography* (PET) plays a crucial role with established clinical applications in oncology and cardiology, and an expanding role in the imaging of neurological disorders such as Alzheimer's disease and Parkinson's disease [2].

In pursuit of improved sensitivity and spatial resolution to enable the imaging of smaller brain structures, specialized scanners, such as dedicated brain PET systems, have been developed [3]. The performance of these systems is influenced by several factors: physics-related effects (e.g., positron range, photon non-collinearity and attenuation), scanner characteristics (e.g., detector size, ring diameter and timing resolution), and the image reconstruction algorithm [3]. Advances in hardware design, such as thicker detectors with higher stopping power and reduced scanner diameter, have addressed many of these limitations, increasing photon detection efficiency and minimizing spatial resolution loss [3]. However, despite these hardware improvements, the image reconstruction algorithm remains critical to fully exploit the potential of these systems [3]. Modern iterative reconstruction methods require accurate system models capable of accounting for the geometry and physical phenomena that map the activity distribution to the detected coincidence count distribution [4].

Among recent initiatives addressing these challenges is a larger design study aimed at developing a dedicated brain PET camera, with the goal of improving sensitivity, spatial resolution and *time-of-flight* (TOF) performance through advances in hardware design.

A distinctive feature of the proposed dedicated brain PET system is the use of a continuous detector, in contrast to the conventional segmented detector blocks. This continuous detector geometry introduces new opportunities for image reconstruction, particularly in combination with the high-precision TOF anticipated from the

hardware. However, to fully exploit these advantages, a physics-based system model tailored for a continuous detector domain is required.

To address these challenges, this thesis focuses on the development and implementation of a novel physics-based system model, the SPLAT system model, tailored for TOF *list-mode* (LM) data in continuous detector geometries. In the SPLAT system model a spatial representation of the decay distribution in image space is directly computed by associating each detection with a multivariate Gaussian distribution. Due to the similarity with the 3D rendering technique Gaussian splatting, the model is referred to as the SPLAT system model.

1.1 Purpose

The purpose of this thesis is to contribute to the development and implementation of the the novel physics-based SPLAT system model suited for TOF LM data, expected to form a key component in the image reconstruction for the new brain PET camera.

1.1.1 Thesis objectives

The work presented in this master thesis aims to:

- Implement the SPLAT system model using a Gaussian distribution to model resolution loss due to non-collinearity, detector blur and non-perfect timing resolution.
- Find a method and implement a function that incorporates attenuation correction in the SPLAT system model in 2D, that is also viable in 3D.
- Find a method and implement a function that incorporates positron range modeling in the SPLAT system model in 2D, that is also viable in 3D.
- Develop and implement a simulation tool for PET LM data based on the SPLAT system model, suited for the novel brain PET camera.
- Use the system model in an iterative reconstruction algorithm to reconstruct images from LM data generated with the SPLAT simulation tool.
- Evaluate the results of the reconstructed images using difference measures.

Additionally we aim to provide a description of the novel SPLAT system model and why it is needed.

1.1.2 Scope/Demarcation

Although several parts of the model have straight-forward 3D implementations, the implementation of the SPLAT system model and the simulation methods will be limited to 2D due to the time constraints of the project.

Also, we will limit the use of reconstruction algorithms to ML-EM and no further investigation will be done in this area.

Detection uncertainty due to depth of interaction (DOI) will not be modeled due to time constraints. Also, CTR will only be modeled as a constant average value.

2
Theory

This section presents the theoretical background for PET imaging, the novel brain PET system, and the role of image reconstruction and system modeling. This is followed by the theory and modeling choices in the SPLAT system model, including how resolution loss mechanisms are incorporated. Finally, we describe how the SPLAT model can be used as a tool for simulating LM data.

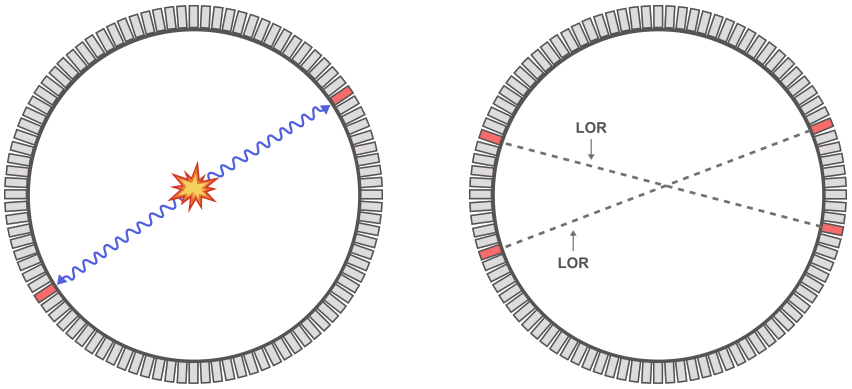
2.1 Introduction to PET

As previously mentioned, nuclear medicine imaging relies on radiolabeled compounds to visualize physiological and biochemical processes. More specifically, it makes use of radioactive substances, called radiotracers, which consist of a chemical compound bound to a positron-emitting radionuclide. These radiotracers are injected, ingested or inhaled by the patient undergoing examination and will spread throughout the body in accordance with the body's natural uptake of the given chemical compound. The radioactive emission of the radionuclide can be detected from outside the body in a detector ring surrounding the object and can be used to reconstruct an image of the activity distribution [1].

Conventional PET systems use a setup of segmented detector blocks in ring formations that surround the object. As the distributed substance undergoes positron (β^+) decay, a positron is emitted and travels a short distance before recombining with an electron which leads to annihilation. From the annihilation a pair of (nearly) back-to-back 511 keV photons are emitted, which may then be detected in the surrounding detector ring. When two photon detections occur within a certain time window one assumes that they originate from the same annihilation event and an *annihilation coincidence detection* (ACD) is recorded, see Figure 2.1a. Information about the event can be stored in several ways e.g. in a sinogram or as LM data. In LM data, information about the two detectors and the time stamps of the detections are stored for each event, illustrated in Table 2.1.

In order to determine the activity distribution in the object, information on where the decay events happened is needed. The basic idea in PET is to make use of the notion of the collinear trajectories of the annihilation photons and the timing difference between the detections.

Each detector pair in a conventional PET camera define a *line of response* (LOR),



(a) (b)

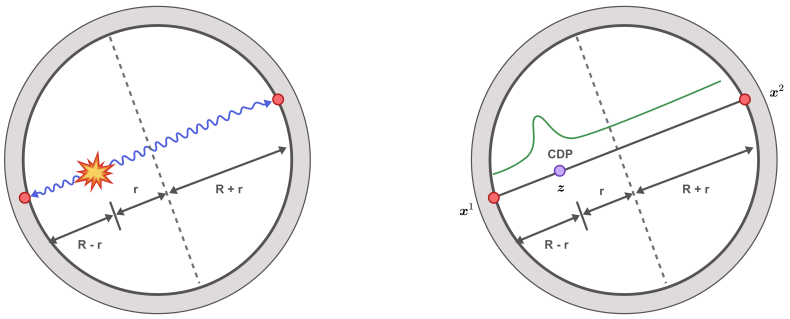
Figure 2.1: The figure shows detector rings with discrete detectors and in **(a)** an *annihilation coincidence detection* (ACD) with the annihilation illustrated in orange and **(b)** examples of *lines of response* (LOR) that connect detector pairs.

Det 1	**Det 2**	t^1	t^2	Δt
$\boldsymbol{x}_1^1$	$\boldsymbol{x}_1^2$	t_1^1	t_1^2	Δt_1
$\boldsymbol{x}_2^1$	$\boldsymbol{x}_2^2$	t_2^1	t_2^2	Δt_2
$\vdots$	$\vdots$	$\vdots$		
$\boldsymbol{x}_n^1$	$\boldsymbol{x}_n^2$	t_n^1	t_n^2	Δt_n

Table 2.1: The table shows an example of list-mode data where each row represents one annihilation coincidence detection (ACD). The detector locations and time stamps for list-mode event k is represented by $\boldsymbol{x}_k^{1,2}$ and $t_k^{1,2}$ respectively with $\Delta t_k = t_k^2 - t_k^1 \geq 0$.

see Figure 2.1b for illustration. Each ACD therefore has an associated LOR along which the event must have taken place. Further, the timing difference, Δt, of the detections of the photon pair, so called TOF, is used in most modern PET scanners to determine where along the LOR the event happened, see figure Figure 2.2. Such a system is simply called TOF-PET. We call the point determined by the LOR and the TOF data the *coincidence detection point* (CDP) as shown in Figure 2.2b (this will be discussed further below).

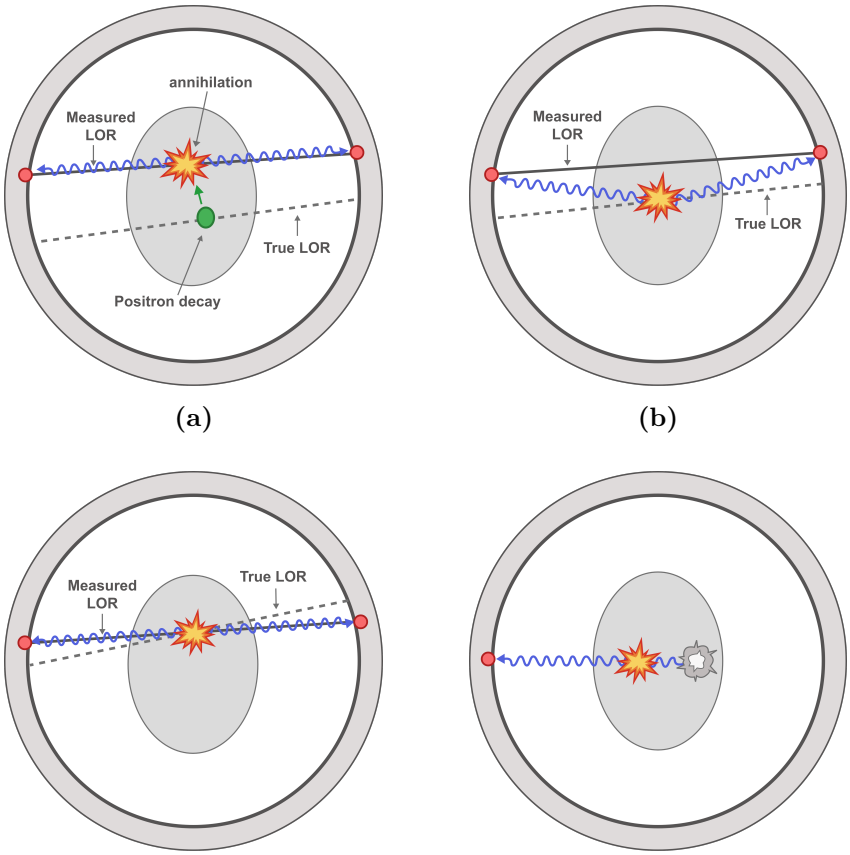
Now, there are several sources of error in measurement that make the location of the decay event uncertain. These uncertainties will blur and degrade the resolution of the reconstructed image. Some of these errors are due to the PET camera itself - its geometry, detector materials and electronics. E.g. the *coincidence timing resolution* (CTR) affect the TOF precision and the detector design affect uncertainty in detection location. Other errors are due to the fact that the detection of an annihilation photon pair is a proxy for what we actually want measure - the positron decay events. Therefore, the positron range (from point of positron decay to annihilation), deviations from collinearity in photon pair trajectory and attenuation of photon flux when passing through matter also degrade the resolution of the reconstructed image. We will refer to these phenomena and uncertainties as resolution loss mechanisms.



(a) **(b)**

Figure 2.2: Illustration of **(a)** an annihilation, shown in orange, and the distance r from the center of the line of response and **(b)** the corresponding *coincidence detection point* (CDP), z, with normal distributed uncertainty due to CTR.

Figure 2.3 illustrates how positron range, non-collinearity, detector uncertainty and attenuation lead to incorrect measurements (or no detection) of LOR.



(a) **(b)** **(c)** **(d)**

Figure 2.3: Illustration of resolution loss mechanisms **(a)** positron range (the distance from decay to annihilation), **(b)** non-collinearity, **(c)** detector blur and **(d)** attenuation.

2.1.1 PET Image Reconstruction

In order to reconstruct PET images, the system model is needed that describe the relationship between the unknown activity distribution image, f, and the measured LM data, g. The model is a mathematical description of how the geometry and physical phenomena (including the resolution loss mechanisms described previously) in the camera maps an activity distribution to LM data. The relation is usually described by a linear system

$$g = \mathcal{A}f + \gamma \tag{2.1}$$

for some system model, $\mathcal{A}$, where the term γ represents random and scattered coincidences. The system model is used in the reconstruction process and will naturally affect the result of the solution. Moving forward, we will not consider γ.

There are two main categories of image reconstruction methods: analytic and iterative [4]. Although analytical methods are still in use, iterative methods have the capability of capturing physical factors that are ignored in analytical algorithms leading to increased spatial resolution and reduced signal-to-noise-ratio.

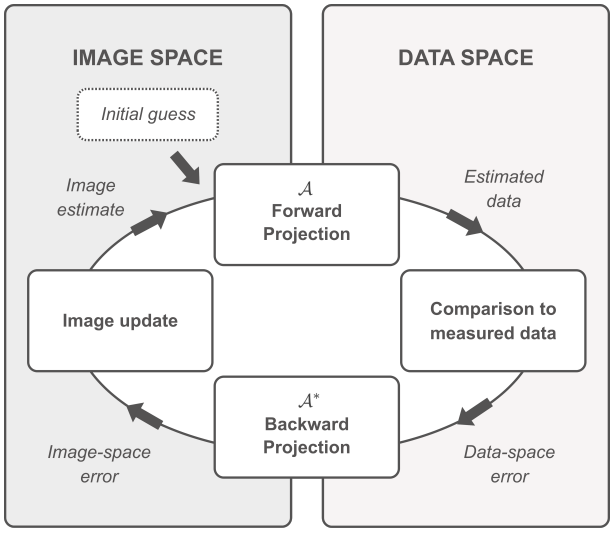


Figure 2.4: Schematic illustration of iterative image reconstruction. The iterative procedure consists of projecting data back and forth between image space and data space. Figure adapted from [4].

The iterative process makes use of the PET system model, $\mathcal{A}$, to project the data back and forth between image space and data space at each iteration, see Figure 2.4. Projecting from image space to the expected LM data is known as forward projection and the reverse operation, projecting from LM data to image space, is known as back projection. Mathematically, we define back projection as $\mathcal{A}^*$, i.e. the adjoint of the system model. The system model therefore decides the result of the reconstructed image.

2.2 The novel brain PET system and LM data

As mentioned above, traditional PET cameras use detector rings of segmented detector blocks defining a finite set of LORs. The novel brain PET camera on the other hand consists of a multilayered monolithic thin-slab detector, that is, several layers of thin overlapping detector sheets in a cylindrical geometry. An important consequence of this design is that the detectors effectively are continuous as opposed to traditional PET where the detector pairs are predefined.

Hence, for the novel system it is no longer appropriate to model the LORs as being defined by each possible detector pair. Rather, the LORs and the detection points can be viewed as coming from a continuous domain, i.e. a continuous cylinder that represents the detector surface with infinitely many possible LORs. Hence, coincidence counts will be defined by two detection points from the continuous domain along with the accompanying TOF data.

The TOF data is used to determine the position of the CDP along its LOR. Figure 2.2 illustrates how the timing difference of the two detections determine the CDP, $\boldsymbol{z}$, via the deviation from the center of the LOR, r. Mathematically, we get that for two detection points on the detector ring, $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$, the CDP is given by

$$\boldsymbol{z} = \left(\frac{1}{2} - \frac{c\Delta t}{2L}\right)\boldsymbol{x}^1 + \left(\frac{1}{2} + \frac{c\Delta t}{2L}\right)\boldsymbol{x}^2 \tag{2.2}$$

where $L = \left\|\boldsymbol{x}^2 - \boldsymbol{x}^1\right\|$, i.e. the length of the LOR, and c denotes the speed of light. By combining $\boldsymbol{z}$ with the angle of the LOR, θ, we get an equivalent representation of the LM data (albeit the absolute time stamps are lost in the conversion).

Moving forward we will use the tuple $(\boldsymbol{z}, \theta)$ to represent the CDP-LOR combination. Here $\theta \in [0, 2\pi)$ is the clockwise angle from the positive y-axis to the LOR. Figure 2.5 illustrates how the CDP-LORs correspond to this representation for n coincidence counts.

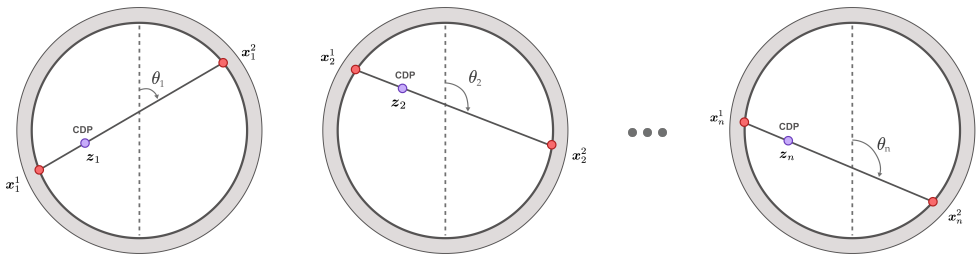


Figure 2.5: Illustration of how each list-mode event k, $k = 1, \ldots, n$, can be represented by the coincidence detection point (CDP), $\boldsymbol{z}_k$, and its line of response (LOR) defined by the angle θ_k and the CDP, $\boldsymbol{z}_k$, that it intersects.

In the SPLAT system model, we refer to the LM data as $g(\boldsymbol{z}, \theta)$. So, the domain of the LM data can be represented by the continuous $(\boldsymbol{z}, \theta)$-space (data space). However, once the recorded LM data has been obtained, with n number of counts, the detection locations may be numbered and each LM event can be represented by

$$\begin{aligned} g_1 &\coloneqq g(\boldsymbol{z}_1, \theta_1) \\ g_2 &\coloneqq g(\boldsymbol{z}_2, \theta_2) \\ &\vdots \\ g_n &\coloneqq g(\boldsymbol{z}_n, \theta_n) \end{aligned}$$

and can be expressed as the vector $\boldsymbol{g} = (g_1, \ldots, g_n)^T$. Specifically, for measured LM data, $\boldsymbol{g}$ will be a vector consisting of n 1's.

2.3 The SPLAT system model

This section presents the theoretical basis and modeling approach for the SPLAT system model, detailing the SPLAT projection operator along with the attenuation correction and positron range operators.

According to Iriarte et al. [4], system models for PET image reconstruction are commonly constructed using one of three methodologies: empirical (measured directly with the PET system), Monte Carlo (highly accurate but computationally expensive), or analytical (fast but often simplistic). A common strategy is to decompose the projection operator into components representing different physical effects, allowing each to be modeled using the most appropriate method. We move on to explain the SPLAT system model.

Recall that the system model is meant to represent the mapping from an activity distribution f to LM data, g. Hence, in order to be accurate, it must account for the resolution loss mechanisms mentioned earlier - positron range, non-collinearity, detector blur, CTR loss and attenuation. We formalize the model below and describe the details in the upcoming section.

The proposed system model in this report may be separated into three main operations - (i) a convolution for positron range modeling, $\mathcal{A}_{\text{pr}}$, (ii) the SPLAT integral operator, $\mathcal{A}_{\text{splat}}$, which models non-collinearity, detector blur and CTR uncertainty and finally (iii) the attenuation correction operator, $\mathcal{A}_{\text{att}}$. We write it as

$$\mathcal{A} = \mathcal{A}_{\text{att}} \underbrace{\mathcal{A}_{\text{CTR}} \mathcal{A}_{\text{det}} \mathcal{A}_{\text{nc}}}_{\mathcal{A}_{\text{splat}}} \mathcal{A}_{\text{pr}}. \tag{2.3}$$

The SPLAT projection operator, $\mathcal{A}_{\text{splat}}$, relies on a multivariate Gaussian distribution. A Gaussian distribution with covariance $S_{(\boldsymbol{z},\theta)}$ is defined in d dimensions as

$$\mathcal{N}(\boldsymbol{x}; S_{(\boldsymbol{z},\theta)}) = \frac{1}{(2\pi)^{d/2} \det\left[S_{(\boldsymbol{z},\theta)}\right]^{1/2}} \exp\left[-\frac{1}{2}\, \boldsymbol{x}^T\, S^{-1}_{(\boldsymbol{z},\theta)}\, \boldsymbol{x}\right] \tag{2.4}$$

with $(\boldsymbol{z}, \theta)$ representing a CDP and its LOR as discussed above. The notations $S_{(\boldsymbol{z},\theta)}$ means to indicate dependence of the covariance on the location of the LOR-CDP as will be discussed below.

For simplicity we here describe the 2D case and assume the 3D extension comes naturally. We consider an activity distribution image $f : \mathcal{I} \subset \mathbb{R}^2 \to \mathbb{R}_+$ and data $g : \mathcal{I} \subset \mathbb{R}^2 \times [0, 2\pi) \to \mathbb{R}_+$ where $\mathbb{R}_+$ denotes the non-negative real numbers.

Then, for $\boldsymbol{z}, \boldsymbol{x} \in \mathcal{I}$ and with μ being the linear attenuation coefficient, the entire system model operator $\mathcal{A}$ is an integral operator and is defined by

$$\Big(\mathcal{A}f\Big)(\boldsymbol{z}, \theta) \coloneqq \underbrace{\exp\left[-\int_{\mathrm{LOR}_{\boldsymbol{z},\theta}} \mu(s)\,\mathrm{d}s\right]}_{\text{attenuation correction}} \underbrace{\int_{\mathbb{R}^2} K(\boldsymbol{x};\, \boldsymbol{z}, \theta)\, f_{\mathrm{pr}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}_{\text{SPLAT projection operator}} \tag{2.5}$$

with the kernel K given by

$$K(\boldsymbol{x};\, \boldsymbol{z}, \theta) = \mathcal{N}(\boldsymbol{x} - \boldsymbol{z}; S_{(\boldsymbol{z},\theta)}). \tag{2.6}$$

Further, f_{pr} denotes the resulting image after positron range convolution with the kernel κ_{pr}, and is given by

$$f_{\mathrm{pr}}(\boldsymbol{x}) = \int_{\mathbb{R}^2} \kappa_{\mathrm{pr}}\Big(\boldsymbol{x} - \boldsymbol{y};\, \hat{\rho}(\boldsymbol{x})\Big)\, f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \tag{2.7}$$

where $\hat{\rho}$ is the tissue density derived from the CT image of the object. The projection operator is hence made up of the positron range convolution with a parameter varying kernel (depending on $\hat{\rho}$), the SPLAT projection operator and the attenuation correction factor and it is the purpose of this work to implement the model.

2.3.1 The SPLAT projection operator

The SPLAT system model is based on the direct computation of an accurate spatial representation of the decay distribution in image space.

In the SPLAT back-projection operator, each SPLAT kernel acts as a *probability density function* (PDF), centered at the CPD and rotated according to the direction of the corresponding LOR, describing the location of the decay. Loosely speaking, an image is created by accumulating the weighted SPLATs for all events, in image space.

The SPLAT forward-projection operator models how each point of the image contribute to the measurement at a certain CDP-LOR $(\boldsymbol{z}, \theta)$ combination. Again loosely speaking, this is done by weighting the image with a SPLAT and integrating the resulting image. The covariance matrix of the SPLAT kernel aims to capture the uncertainty of the measurement due to non-collinearity, detector blur and CTR.

Figure 2.6a illustrates SPLATs positioned at various locations with orientations corresponding to different $(\boldsymbol{z}, \theta)$ combinations. Also, they show varying spreads as

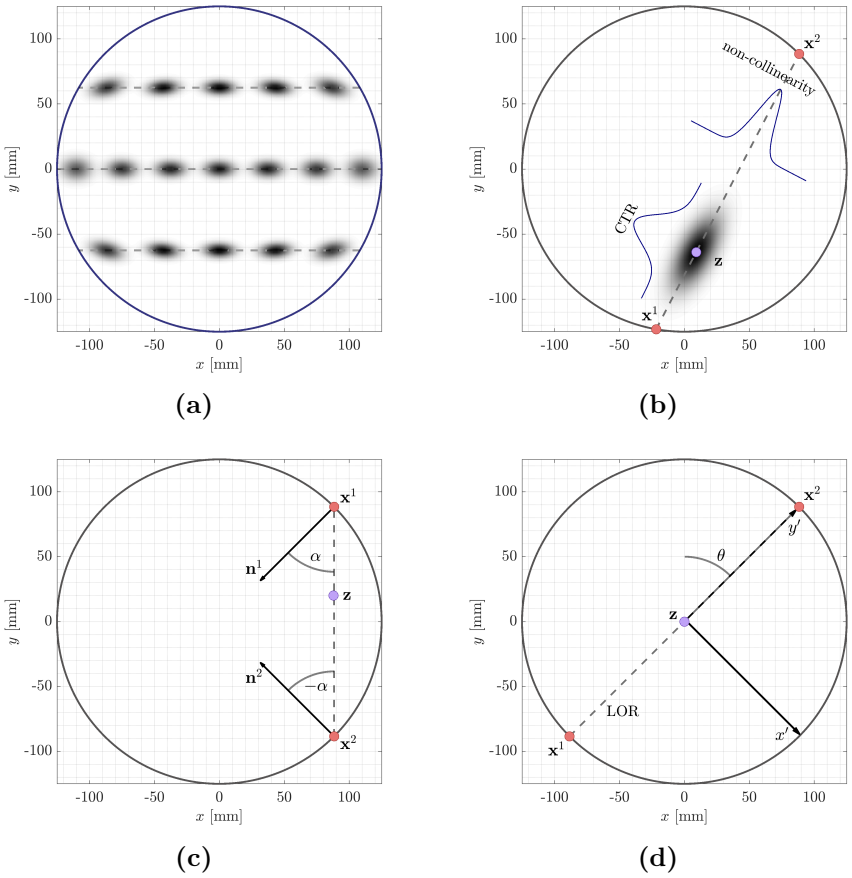


Figure 2.6: Illustration of **(a)** several SPLATs at different positions along different lines of response (LORs) and hence with different sizes and rotation, **(b)** variance due to CTR is parallel to LOR while variance due to non-collinearity is orthogonal to the LOR, **(c)** detection points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ with corresponding normal vectors $\boldsymbol{n}^1$ and $\boldsymbol{n}^2$ with the angle α from normal to LOR defined and finally **(d)** the reference frame of the LOR defined by x' and y' with the angle θ defined from positive y-axis to the LOR.

determined by their respective covariance matrices, $S_{(\boldsymbol{z},\theta)}$.

The three resolution loss mechanisms modeled by the SPLAT happen sequentially and independently from each other. In addition to this, the deviation from collinearity act perpendicular to the LOR, whereas CTR loss act parallel to the LOR, see illustration in Figure 2.6b. This enables us to simply sum the covariance matrices for these two mechanisms.

For detector blur it is a bit more tricky. In the special case where the incident photon is perpendicular to the detector surface, the detector resolution loss acts perpendicular to the LOR, and the SPLAT is hence aligned with the LOR, see the middle LOR in Figure 2.6a. However, when the incident photon is not perpendicular to the detector surface, the covariance matrix of the detector resolution (describing detector resolution perpendicular to the LOR) will have to be rotated, this is illustrated in the upper and lower LORs in figure 2.6a. The rotation is performed with the rotation matrix that aligns the normal of the detector surface with the LOR,

see illustration in figure Figure 2.6c.

Let L be the length of a considered LOR and let $\ell \leq \frac{1}{2}L$ be the distance from detection point 1, $\boldsymbol{x}^1$, to the CDP, $\boldsymbol{z}$. That is, by convention we define detection point 1 to be the one closest to the CDP. Then, with the first dimension perpendicular and second dimension parallel to the LOR, the covariance matrix for the Gaussian SPLATs become

$$S(\ell; L, \boldsymbol{\alpha}) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2_{\text{CTR}} \end{bmatrix} + \begin{bmatrix} \sigma^2_{\text{nc}}(\ell; L) & 0 \\ 0 & 0 \end{bmatrix} + \sum_{i=1,2} R^T(\alpha_i) \begin{bmatrix} \sigma^2_{\text{det}_i}(\ell; L) & 0 \\ 0 & 0 \end{bmatrix} R(\alpha_i)\,, \quad (2.8)$$

where R is the standard counter clockwise rotation matrix defined by

$$R(\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \quad (2.9)$$

that aligns the normal of the detector surface in detection point $\boldsymbol{x}^i$ with the LOR, see an illustration in Figure 2.6c.

Since the covariance matrix in Equation 2.8 is in the reference frame of the LOR the SPLAT needs to be correctly aligned in the image's reference frame. This is done by yet another rotation;

$$S_{(\boldsymbol{z},\theta)} = R^T(\theta)\, S(\ell; L, \boldsymbol{\alpha})\, R(\theta), \quad (2.10)$$

where $S_{(\boldsymbol{z},\theta)}$ is the covariance matrix in the reference frame of the image and $R(\theta)$ is the rotation matrix that would align the image frames vertical axis with the LOR, see Figure 2.6d for an illustration of the geometry.

In the simplest model, the variances would be constant over the whole system- we will refer to this as the *globally invariant* model. But non-collinearity and detector blur can be more accurately modeled if one allows the variances to vary along the LOR. Therefore the covariance matrix in Equation 2.8 varies with ℓ. For illustration, see Figure 2.6 (and Figure 2.8).

The details of this ℓ-dependence are presented in the sections that follow; the resulting formulation is referred to as the *spatially varying* model. Unless stated otherwise, the term the "model" henceforth denotes this spatially varying formulation.

2.3.1.1 Non-collinearity variance

Recall that an annihilation photon pair may have some deviation from the perfect back-to-back trajectory due to residual momentum of the positron. The standard deviation from the collinear 180° trajectories is of an angle of $\varphi = 0.25°$ [5], see illustration in Figure 2.7.

We define a triangle with two vertices at each end of the LOR and the third vertex kept at a constant angle of (180-φ)°. When keeping this angle constant and moving

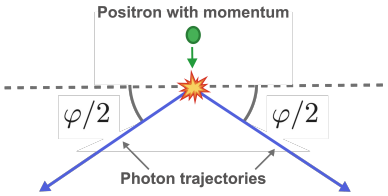


Figure 2.7: Illustration of positron with residual momentum undergoing annihilation leading to non-collinear trajectory of annihilation photons with $\varphi = 0.25°$.

the vertex along the LOR a circular arc is formed (at each side) of the LOR. Figure 2.8 illustrates the triangle with vertices at each end of the LOR and the circular arc in blue. We model the variance due to non-collinearity as the perpendicular distance from the LOR to this circular arc depending in the position along the LOR, ℓ. The expression for the variance is then given by

$$\sigma_{\text{nc}}^2(\ell; L) = \frac{-\sqrt{(b^2 - 4c)} - b}{2} \in \mathbb{R}_+ \tag{2.11}$$

where $b = \ell^2 + (L - \ell)^2 - k^2$, $c = \ell^2(L - \ell)^2$ and $k = \frac{L}{\sin(179.75°)}$.

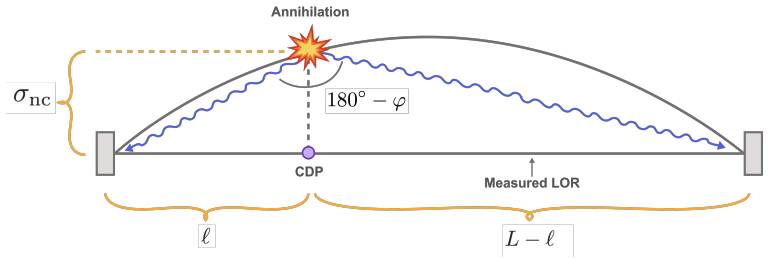


Figure 2.8: Illustration of uncertainty due to non-collinear trajectories. The image shows that the actual annihilation point may deviate from the measured line of response (LOR) between the two detectors and how the standard deviation model depends on the position along the LOR. Here ℓ is the distance from detection point 1 to the coincidence detection point (CDP) and L is the length of the LOR.

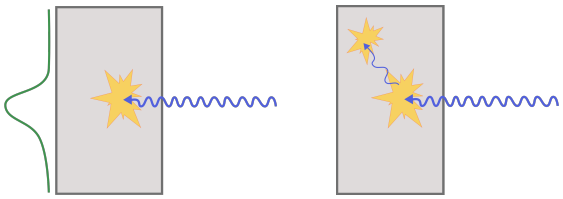
Again, referring back to Figure 2.6a, when comparing the middle SPLATs at each of the three LORs we note that the middle LOR has a wider SPLAT (perpendicular to the LOR) due to non-collinearity variance since it is centered in the scanner.

2.3.1.2 Detector variance

Most PET detectors use scintillator crystals to detect gamma photons [6]. The scintillator crystals emit cascades of visible light through photoelectric effect (PE) and Compton scattering when interacting with the high energy gamma photon [1][6]. The position of the gamma photon is calculated using a center of mass calculation from the light yield [1], see an illustration of the light yield in Figure 2.9a.

There are three major mechanisms that contribute to resolution loss in the detectors [4][5]. There is uncertainty in the center of mass calculation from the photoelectric

light yield [4][7]. There is uncertainty coming from the possibility of Compton scattering and hence light yield on multiple locations in the crystal [4], see illustration in Figure 2.9b. The third mechanism, depth of interaction due to the variation in the depth in the crystal at which the photon interaction happens, will not be modeled in this project.



(a) **(b)**

Figure 2.9: Illustration of detection variance due to **(a)** light yield centroid uncertainty and **(b)** Compton scattering in crystal.

Modeling each detector as independent of the other, the expression for the ℓ-dependent variance for the two detection points in a detector point pair can be written as

$$\begin{aligned} \sigma^2_{\text{det}_1}(\ell; L) &= \sigma^2_{\text{PE,C}} \left(\frac{L-\ell}{L} \right)^2, \text{ and} \\ \sigma^2_{\text{det}_2}(\ell; L) &= \sigma^2_{\text{PE,C}} \left(\frac{\ell}{L} \right)^2 \end{aligned} \tag{2.12}$$

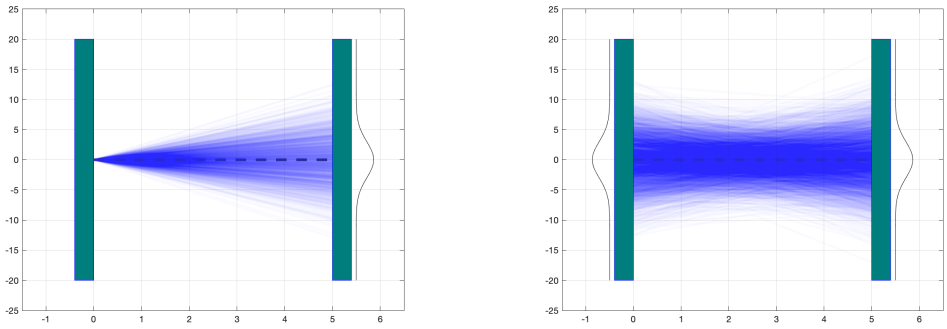
where L is the length of the LOR and $\sigma^2_{\text{PE,C}} \coloneqq 0.7\sigma^2_{\text{C}} + 0.3\sigma^2_{\text{PE}}$, that is, a weighted combination of the two individual variances due to Compton scattering and PE, σ^2_{C} and σ^2_{PE}, respectively.

If we examine e.g. the variance for detector 1 in Equation 2.12 we find that the variance decreases as ℓ gets larger i.e. the further away the CDP is from detector 1. The same relationship holds for detector 2 - the further away the CDP is from detector detector 2, the less variance we get from detector 2. Figure 2.10 illustrates this phenomena in two cases; one case where the detection point is completely certain in one detector but uncertain in the other, one case where both detection points are uncertain.

When summing the variances as in Equation 2.8 and take the special case with parallel detectors we get

$$\sigma^2_{\text{det}_1} + \sigma^2_{\text{det}_2} = \frac{2\sigma^2_{\text{PE,C}}}{L^2} \left(\ell^2 - L\ell + \frac{L^2}{2} \right), \tag{2.13}$$

i.e., a quadratic function in ℓ with minimum in $\ell = \frac{L}{2}$ as illustrated in Figure 2.10b. That is, the minimum total detector variance contribution is when the CDP is in the middle of the LOR. In Figure 2.6c we note that in the middle horizontal LOR the SPLATs get wider perpendicular to the LOR when positioned close to the detectors



(a) **(b)**

Figure 2.10: Illustration of detection uncertainty with green blocks representing detectors and measured LOR in center. Figure **(a)** illustrates a hypothetical case where the left detection points is known and the right detection has $\mathcal{N}$-distributed uncertainty. Figure **(b)** illustrates both detectors having $\mathcal{N}$-distributed uncertainty. The illustration may loosely speaking be interpreted as "LOR density".

due to detector blur. Also, notable is a rotation off the LORs for the top and bottom LOR as the SPLATs get closer to the detectors. This is due to the detector not being perpendicular to the LOR as modeled by the angle α in Equation 2.8.

2.3.1.3 CTR variance

CTR defines the uncertainty of position along the LOR. The main limiting factors of the CTR is the crystal decay time and light output [8]. Within a physical PET camera system, variations in detector quality, electronics and time alignment procedure result in CTR unique for each detector pair, with the PET system CTR often defined as the average CTR for all the pairs [8]. In this system model the CTR will be modeled as constant over the whole system.

2.3.2 Attenuation correction

The attenuation of photon flux through any material depends on the atomic number of the material and the photon energy. Attenuation is characterized by the linear attenuation coefficient, μ, expressed in units of inverse centimeter (cm^{-1}) [9]. Two types of interaction make up the general concept of photon attenuation: absorption from PE and Compton scattering. PE absorbs the entire energy of a photon whereas Compton scattering changes the incident photons' trajectory leaving a lower energy photon to travel a new path. For PET, with its 511 keV photons, the dominating mechanism is Compton scattering which may create trajectories going out of the detector ring leading to reduced counts. In contrast, the dominating mechanism in CT is PE. The term total attenuation is used for the sum of the contributions of the individual attenuation mechanisms.

Since attenuation clearly affects the count of ACD's it needs to be accounted for in the system model. There are several different ways of performing attenuation cor-

rection for PET images. Historically a PET transmission scan was used but it has largely been replaced by CT-based attenuation correction since PET-CT scanners were made available. Magnetic resonance tomography-based correction and deep-learning based methods have also been developed [10].

Attenuation is one of the most important physical effects to correct for in PET data since it may impact both quantitative accuracy and visual quality. Common artifacts in uncorrected images include negative tracer concentration in mediastinal regions and enhanced activity in lungs ("hot lungs") [9].

However, there may also be artifacts arising from an attenuation corrected image e.g. due to differences in patient respiratory movements in a PET and CT image. With knowledge of these artifact one may in clinical practice use a combination of corrected and uncorrected images to get a complete picture [9].

The attenuation correction term, $\mathrm{AC}_{(\boldsymbol{z},\theta)}$, is given by

$$\mathrm{AC}_{(\boldsymbol{z},\theta)} = \exp\left[-\int_{\mathrm{LOR}_{\boldsymbol{z},\theta}} \mu(s)\,\mathrm{d}s\right] \tag{2.14}$$

where the integral of μ is along the LOR defined by $(\boldsymbol{z},\theta)$. The values of μ are derived from the CT image of the object and since they are energy dependent they need to be scaled appropriately to match the energy levels used in PET scan. We will denote the derived values of μ by $\hat{\mu}$.

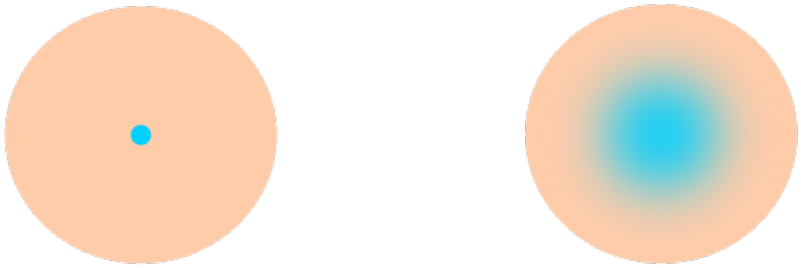
2.3.3 Positron range modeling

The range of the positron for a given decay varies depending on several factors. First of all, there is randomness in its initial energy, resulting in a distribution of ranges. It also depends on the electron density of the surrounding tissue, with higher density leading to a shorter positron range [11]. Also, for a given surrounding tissue different radionuclides have different positron range distributions.

For whole-body scans with the very common radionuclide Fluor-18 (F-18), positron range is of minor concern [12]. For other isotopes with higher positron energies, Gallium-68 (Ga-68) for instance, the positron range becomes more important [12], especially for scanners that strive for optimal spatial resolution.

For a given tissue, consider a point of activity. Since the detector measures annihilation photons and not the decay of the radionuclide, the point of activity will be blurred in the view of the detector due to the positron range (ignoring all other resolution loss mechanisms). See illustration in Figure 2.11.

In forward projection, the positron range operator is applied first, by convolving the image with the positron range kernel [13] as described in Equation 2.3. In back projection, the order is reversed and the positron range convolution is the last operator



(a) **(b)**

Figure 2.11: **(a)** Point of activity in tissue. **(b)** Point of activity in tissue as seen from the view of the detector due to positron range.

applied.

Previous work [12], [14], [15] quantified the resolution loss due to positron range using the density of positron annihilations as a kernel in the convolution. The density of annihilations for different tissues are commonly acquired from Monte Carlo simulations [12]–[17]

Since the positron range depends on tissue, it is desirable to use a kernel that varies with the local density during each step in the convolution [12], [14], [15], [17], as described by Equation 2.7. Unfortunately there is no direct relationship between the values in a CT image and the actual tissue density. However, [12] argues that a bilinear approximation is reasonable. Hence, we let $\hat{\rho}$ denote the approximated density values and define the bilinear model

$$\hat{\rho}\Big(f_{\text{ct}}(\boldsymbol{x})\Big) = \begin{cases} k_1 f_{\text{ct}} + m_1 & 0 \leq f_{\text{ct}}, \\ k_2 f_{\text{ct}} + m_2 & f_{\text{ct}} < 0 \end{cases} \tag{2.15}$$

where f_{ct} is the CT image of the object.

2.4 SPLAT model-based simulation

To evaluate the performance of the proposed system model, we require LM data generated from a known activity distribution. Monte Carlo simulation is considered the gold standard in nuclear medicine imaging due to its detailed modeling of particle interactions and transport [18]. However, its high computational cost makes it impractical for rapid testing or iterative development.

Instead, this thesis proposes a simulation tool based on the SPLAT system model introduced in earlier sections. By sequentially modeling physical resolution loss mechanisms, we aim to produce realistic yet computationally efficient LM data.

Radioactive decay is a stochastic process commonly modeled as a Poisson process. Under the assumption of constant activity during the scan, the expected number

of detected events at a given CDP-LOR, characterized by $(\boldsymbol{z}, \theta)$, is given by the forward projection:

$$\bar{g}(\boldsymbol{z}, \theta) = (\mathcal{A}f)(\boldsymbol{z}, \theta).$$

Each value of $\bar{g}$ corresponds to the mean of a Poisson-distributed random variable:

$$g \sim \text{Poisson}(\bar{g}) = \text{Poisson}(\mathcal{A}f).$$

To simulate physically realistic data, we must account for attenuation. The probability that a decay event reaches the detectors depends not only on its location but also on the attenuation along the LOR. This means we cannot decouple activity and attenuation in the sampling process.

Due to the continuous nature of the detector geometry in the proposed scanner, there is no predefined discrete LORs, making LOR selection a non-trivial problem. Since attenuation must be computed along each LOR to evaluate its contribution to the detection process, we address this by discretizing the space of possible LORs, enabling us to calculate attenuation and assign sampling probabilities.

3
Methods

This section describes methods and implementation for the simulation of LM data using the SPLAT model-based simulation. This is followed by method and implementation of the SPLAT system model, including attenuation correction and positron range modeling. Finally, we describe how images were reconstructed using the *maximum-likelihood expectation-maximization* (ML-EM) image reconstruction algorithm and associated evaluation metrics.

3.1 Implementation of the SPLAT model-based simulation

Two methods of implementation were used to simulate LM data. They differ in how choices of LORs are made and hence how the ground truth object is sampled. They can be thought of as to represent two alternative views on the forward-projection model.

- **Method A:** LORs are defined by selecting detector positions on the detector ring.
- **Method B:** LORs are defined from each pixel center in the image domain using a fixed set of angular directions, producing a grid of pixel-angle combinations.

Both methods were implemented and used interchangeably to generate list-mode data, as they offer complementary strengths. Method A constructs LORs by selecting detector positions, then integrates the activity along each LOR and samples detection positions based on that profile. While this approach is flexible, easily extends to 3D and allows detectors to be placed arbitrarily, it also introduces a somewhat unintuitive sampling step, where spatial information is aggregated and then resampled. Method B, in contrast, operates directly on the image grid, assigning LORs from each pixel center using a fixed set of angles. This method avoids the resampling step, but does not extend as easily to 3D geometries, where evenly spaced angles do not translate to equidistant detector positions on a spherical surface. In practice, both methods were used with fine discretizations to ensure that any differences in the sampling approach did not significantly affect the simulation results.

Before method A or B was used to sample the ground truth, the same upsampling of the ground truth simulation object was made. This was done with MATLAB's built-in functions. Then the positron range operator was applied to the upsampled

simulation object. This was done through convolution of the activity image with a tissue-dependent kernel, as defined in Equation 2.7. The method and implementation of this will be further detailed in subsection 3.2.3.

After sampling the ground truth object using method A and B, the resolution loss mechanisms were simulated. Since the effects are assumed to act independently or in a defined temporal sequence they can be simulated sequentially, following the same structure as the forward projector, see Equation 2.3.

3.1.1 Method A

Method A distributes a desired number of n counts over some choice of LORs (the choice of LORs is an independent step of the algorithm). The method was performed in 2D but could in principle just as well be used in 3D. An integration, F_i, is performed along each LOR_i, $i = \left\{1, \ldots, n_{\mathrm{LORs}}\right\}$ over the ground truth object;

$$F_i = \int_{\mathrm{LOR}_i} f(s)\,\mathrm{d}s. \tag{3.1}$$

Each F_i is then weighted by LOR_i's attenuation correction term, AC_i (defined in Equation 2.14), normalized and scaled by n, to get

$$P_i = \frac{n\,\mathrm{AC}_i\,F_i}{\sum_{j=1}^{n_{\mathrm{LORs}}} \mathrm{AC}_i\,F_i}. \tag{3.2}$$

That is, the desired n number of counts are distributed over the normalized and weighted LORs. The resulting numbers are then considered expected values in a Poisson distribution for each LOR. A random number, N_i, is then drawn for each LOR, that is, $N_i \sim \mathrm{Poisson}(P_i)$.

In order to assign positions of the detections along each LOR a PDF of the simulation object was created for each LOR. This was done by stepping over each LOR with adjustable step size, evaluating $f(s)$ at each point and normalizing. Then, the resulting distributions were sampled N_i times for each LOR and the first simulation step is concluded resulting in simulation points on the predefined LORs.

For the simulations performed with Method A the detectors were placed equidistantly over the detector ring. Figure 3.1 shows an illustration of the resulting LORs for different numbers of detectors. We note in the figure the different patterns for even and odd numbers of detectors; for even numbers e.g. the origin gets heavily sampled while for odd numbers the origin does not get sampled at all unless approaching very high numbers of detectors. The number of LORs defined by n_{det} detectors follow $\binom{n_{\mathrm{det}}}{2}$ which for $n_{\mathrm{det}} = 2000$ result in 1 999 000 LORs. The simulations made for this report were made with an even number of detectors. For simplicity, the LORs could take any angle with respect to the detector normal, e.g. two detectors placed right next to each other also defined a LOR.

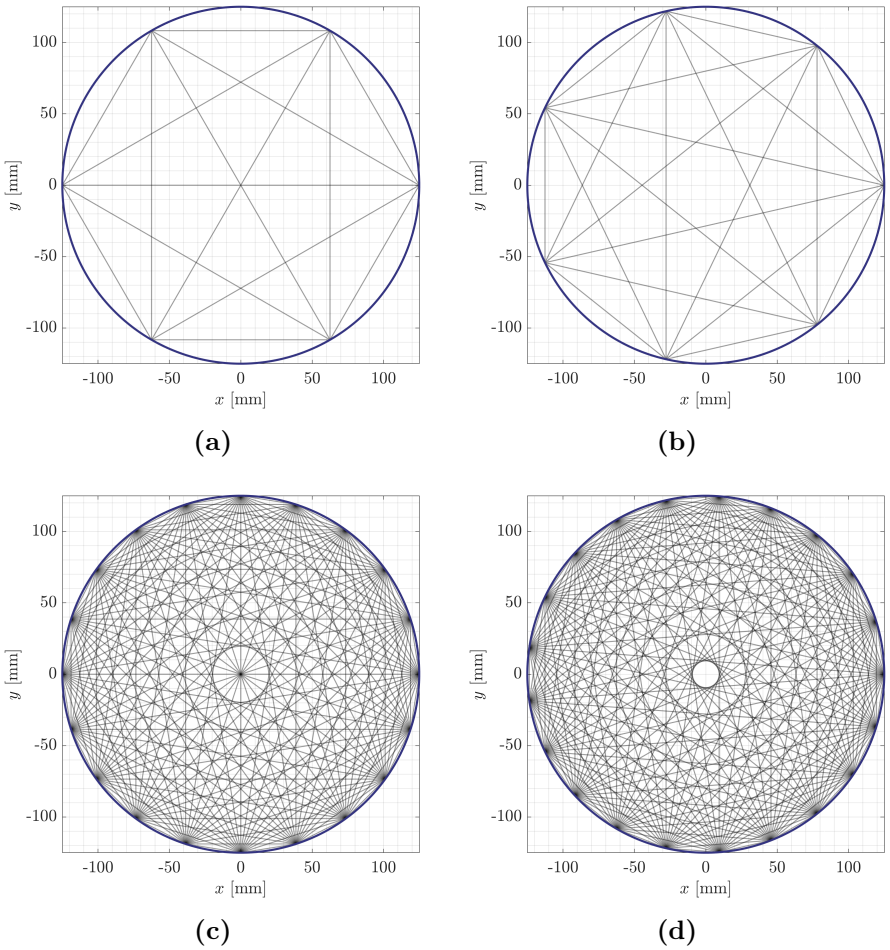


Figure 3.1: The figure shows four different choices of detector discretization with LORs for simulation method A. In **(a)** and **(b)** we have 6 and 7 detectors respectively resulting in 15 and 21 LORs respectively. In **(c)** and **(d)** we have 20 and 21 detectors respectively resulting in 190 and 210 LORs respectively. Note the different patterns for even and odd numbers of detectors.

3.1.2 Method B

In contrast to method A, method B defines candidate LORs from a regular grid of pixel-centered locations combined with a fixed number of angular directions. The desired number of simulated events, n, is distributed across a set of predefined LORs formed from combinations of pixel centers and angles.

To create the sampling space, the simulation object (ground truth image) is up-sampled to an $m_1 \times m_2$ grid, yielding $n_{\text{pixels}} = m_1 \times m_2$ pixels. From each of the n_{pixels} pixel centers, n_θ equally spaced angles, θ_j are defined, resulting in $n_{\text{pixels}} \times n_\theta$ combinations of $(\boldsymbol{z}_i, \theta_j)$, $i = \{1, \ldots, n_{\text{pixels}}\}, j = \{1, \ldots, n_\theta\}$. For each such $(\boldsymbol{z}_i, \theta_j)$, the corresponding attenuation is computed based on the attenuation image. The pixel value is then weighted by the corresponding LORs attenuation correction term $\text{AC}_{\text{LOR}_{(z_i,\theta_j)}}$, scaled by the number of counts and normalized, producing

$$P_{i,j} = \frac{n \, \mathrm{AC}_{\mathrm{LOR}_{(\boldsymbol{z}_i,\theta_j)}} \, f(\boldsymbol{z}_i)}{\sum_{i=1}^{n_{\mathrm{pixels}}} \sum_{j=1}^{n_\theta} \mathrm{AC}_{\mathrm{LOR}_{(\boldsymbol{z}_i,\theta_j)}} \, f(\boldsymbol{z}_i)} \tag{3.3}$$

where $f(\boldsymbol{z}_i)$ is the activity at pixel center $\boldsymbol{z}_i$. As in method A, counts are then drawn from a Poisson distribution, i.e., $N_{i,j} \sim \mathrm{Poisson}(P_{i,j})$, yielding the number of simulated detections for each pixel-angle combination, with no additional sampling for position along the LOR required.

For simulations using this method, the ground truth image was upsampled by a factor of at least 3 (maximum 10), and the number of angular directions was chosen to be at least 3 (maximum 10) times greater than that used for the sensitivity image. Prior to simulation, all values in the ground truth object below $\frac{1}{500}\max(f)$ were set to zero to reduce computational load.

Figure 3.2 shows an illustration of different choices of LORs for simulation method B. Note the different patterns for even and odd numbers of angles for intersecting LORs.

3.1.3 Simulation of resolution loss mechanisms

A brief description of the simulation of resolution loss mechanisms follows.

The resolution loss mechanisms were simulated sequentially in 2D in the order; non-collinearity, detector blur and CTR loss. The non-collinearity simulation consisted of shifting the CDP, $\boldsymbol{z}$, a distance $h \sim \mathcal{N}(0, \sigma^2_{\mathrm{nc}})$ in a direction perpendicular to the LOR and then projecting onto the ring to find the new corresponding detection points.

Resolution loss due to detector blur was modeled by sampling offsets for each detector from $\mathcal{N}(0, \sigma^2_{\mathrm{det}_{1,2}})$ and shifting each detector by this amount along the ring. The CDP was shifted to the new LOR by transferring its original relative position along the LOR to the new LOR.

Again, for CTR we sample from $\mathcal{N}(0, \sigma^2_{\mathrm{CTR}})$ and shift the CDP along the LOR. The detector locations were not changed in this step.

Figure 3.3 shows the ground truth PET and CT images that were used for simulation and image reconstructions unless something else has been stated. The ground truth images are of size 257×257 and the field of view (FOV) is 300 mm. The CT image has been cropped and resampled to fit these settings.

3.2 Implementation of the SPLAT system model

This section describes the implementation of the SPLAT system model, covering the projection operator, attenuation correction, and positron-range modeling.

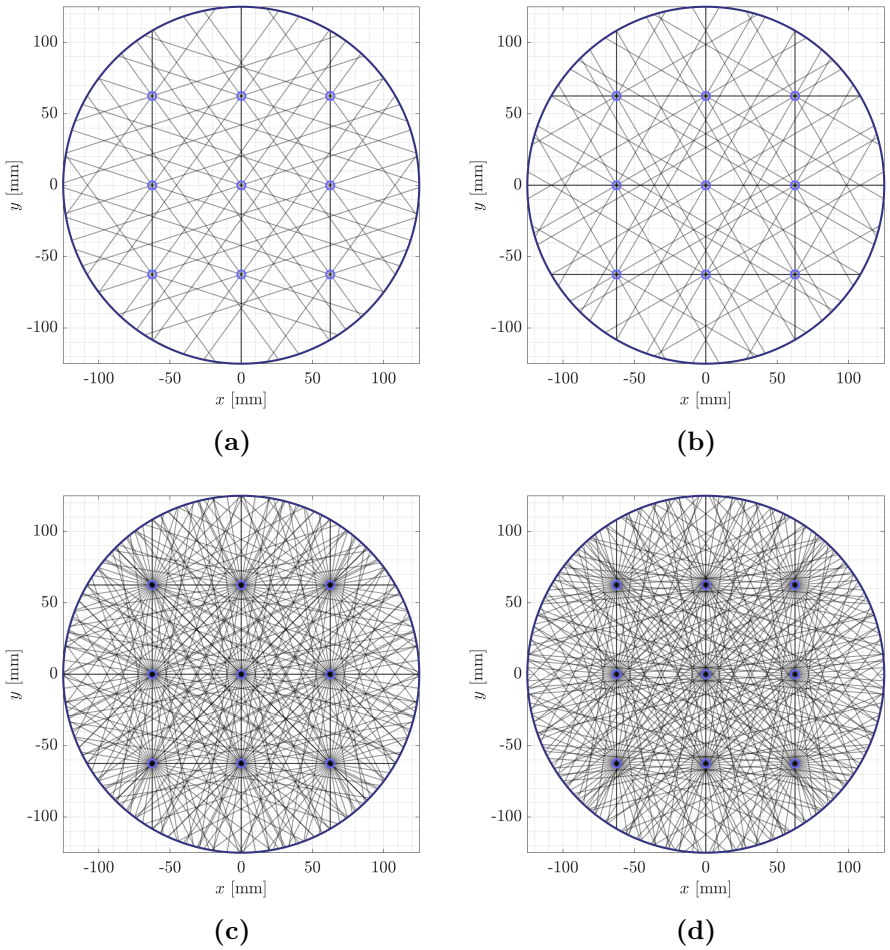
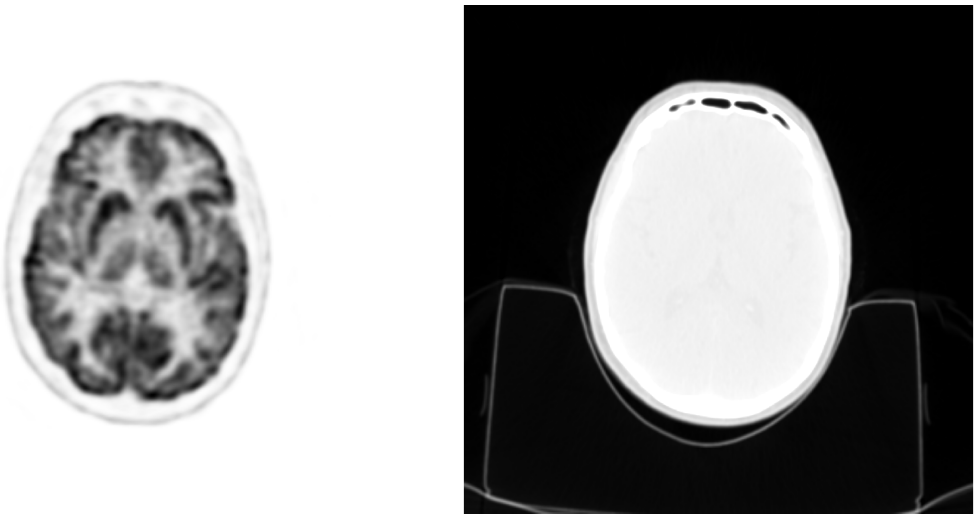


Figure 3.2: Four different choices of LOR discretization for the simulation Method B. For illustration purposes the matrix sizes were set such that only a total of 9 pixels are inside the detector ring. We show in **(a)** 5 angles, **(b)** 6 angles, **(c)**, 20 angles and in **(d)** 21 angles per pixel of intersecting LORs.

3.2.1 Implementation of the SPLAT projection operator

Recall that the SPLAT projection operator, $\mathcal{A}_{\text{splat}}$, takes us from image space to data space and vice versa. While it is an integral operator, we describe here its implemented discrete counterpart. In order to decrease computational load and memory consumption, the SPLATs were evaluated in smaller subgrids centered around the current CDP and then put into its correct position in the image grid. Figure 3.4 shows an illustration of how a subgrid may be used to evaluate a SPLAT at the given subgrid points. The figure shows a contour plot of a SPLAT with center in red and the points of the subgrid in black. The center of the SPLAT may not be on the grid points but no interpolation is needed.

We remark that the SPLAT projection operator used in the image reconstruction algorithm (except in the projection of the sensitivity image) only projects data to and from the measured LM data points, as opposed to the entire data and image spaces. Hence, the operator described here is not just a discretized version of $\mathcal{A}$ but



(a) PET **(b)** CT

Figure 3.3: Ground truth images for list-mode data simulation. We have in **(a)** the PET image and in **(b)** the corresponding CT image. The CT image is displayed in the range [-1000, 100] Hounsfield units.

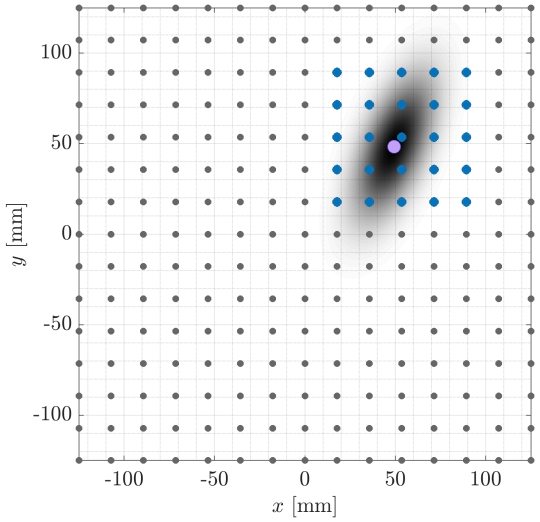


Figure 3.4: Illustration of the subgrid where the SPLAT is evaluated. The center of the SPLAT (the CDP) is shown in purple and the points of subgrid are shown in blue.

also a "reduced" version of $\mathcal{A}$. We denote it A_n to indicate the dependence on the n counts. For the discretized matrix operator this corresponds to removing "unused" rows (rows projecting a zero count) of A.

So, let f be an image of size $n_{\text{pixels}} = m_1 \times m_2$ pixels ordered in column major order. The pixel values are then described by f_j, $j = 1, \ldots, n_{\text{pixels}}$. Each row of the matrix operator A_n corresponds to one $(\boldsymbol{z}_i, \theta_i)$ combination in the LM data and we can express each such row as

$$g_i = \sum_{j=1}^{n_{\text{pixels}}} a_{i,j} f_j \tag{3.4}$$

where $a_{i,j} = \mathcal{N}(\boldsymbol{x}_j - \boldsymbol{z}_i; S_{(\boldsymbol{z}_i,\theta_i)})$ and $\boldsymbol{x}_j$ is the corresponding pixel location to f_j.

The covariance matrices for each event remain the same throughout image reconstruction and hence they need only be calculated once. Since the Gaussian distribution approaches 0 far from its center, values outside the above mentioned subgrids were considered negligible. Because of the use of subgrids $a_{i,j} = 0$ for all j not in the current subgrid in the matrix operator A_n.

The forward- and back-projection algorithms are summarized in Algorithms 1 and 2 respectively.

Algorithm 1 Forward-projection algorithm (A_N)

1: Perform positron range convolution with parameter varying kernel over the entire image
2: **for** each CDP $(\boldsymbol{z}_i, \theta_i)$ **do**
3: Define a subgrid with center at grid point nearest to $\boldsymbol{z}_i$
4: Evaluate $\exp\Big[-\frac{1}{2}\Big((\boldsymbol{x}-\boldsymbol{z}_i)^T \Sigma^{-1}_{\boldsymbol{z}_i,\theta_i}(\boldsymbol{x}-\boldsymbol{z}_i)\Big)\Big]$ at subgrid points and normalize result to sum to 1
5: Perform element-wise multiplication of above result with $f(\boldsymbol{x})$ at subgrid points and sum all terms
6: Add contribution to g_i
7: Multiply g_i by the attenuation factor for LOR_i, i.e set $g_i = \mathrm{AC}_{\mathrm{LOR}_i} g_i$
8: **end for**

Algorithm 2 Back-projection algorithm (A_N^T)

1: **for** each LM event g_i **do**
2: Multiply g_i by the attenuation factor for LOR_i, i.e. set $g_i = \mathrm{AC}_{\mathrm{LOR}_i} g_i$
3: Define a subgrid with center at grid point nearest to $\boldsymbol{z}_i$
4: Evaluate $\exp\Big[-\frac{1}{2}\Big((\boldsymbol{x}-\boldsymbol{z}_i)^T \Sigma^{-1}_{\boldsymbol{z}_i,\theta_i}(\boldsymbol{x}-\boldsymbol{z}_i)\Big)\Big]$ at all grid points in the subgrid
5: Normalize results to sum to 1
6: Scale result by g_i
7: Add results to the corresponding grid indices in image
8: **end for**
9: Perform positron range convolution with parameter varying kernel over the entire image

3.2.1.1 Non-collinearity calculation

Unlike most other computations in the model, non-collinearity variance was evaluated in double precision to avoid truncation errors in subtraction where the number of significant digits in large numbers where very different.

3.2.2 Attenuation correction implementation

Kinahan et. al. [9] discuss different methods for deriving the linear attenuation coefficient, μ, from CT-image Hounsfield units, including the bilinear model used for this report. The model is given (in a similar fashion to the $\hat{\rho}$-map for positron range correction) by

$$\hat{\mu}\Big(f_{\text{ct}}(\boldsymbol{x})\Big) = \begin{cases} k_1 f_{\text{ct}} + \mu_{\text{water}} & f_{\text{ct}} \leq 0 \\ k_2 f_{\text{ct}} + \mu_{\text{water}} & f_{\text{ct}} > 0 \end{cases} \tag{3.5}$$

where

$$k_1 = \frac{\mu_{\text{water,pet}} - \mu_{\text{air,pet}}}{1000}, \quad k_2 = \frac{\mu_{\text{bone,pet}} - \mu_{\text{water,pet}}}{1000} \frac{\mu_{\text{water,ct}} - \mu_{\text{air,ct}}}{\mu_{\text{bone,ct}} - \mu_{\text{water,ct}}} \tag{3.6}$$

with the subscripts representing material and the energy levels of PET and CT, that is, 511 keV and 70 keV effective energy, respectively. From this relation an attenuation image was created from the CT image of the scanned object and the integral in Equation 2.14 was performed for each LOR to obtain the attenuation correction term. The integration along each LOR was implemented in C based on the work by [19]. Such an integral transform is usually called X-ray transform.

3.2.3 Positron range implementation

In this thesis, the positron emitter Gallium-68 (Ga-68) was selected as the radiotracer due to its clinical relevance and relatively long positron range, which makes positron range modeling meaningful. Radial annihilation distributions (aPSFs) derived from Monte Carlo simulations by Carter et al. [16] were used as reference data for positron range. Simulations were performed for three representative tissue types: lung, soft tissue, and cortical bone. See the dotted lines in Figure 3.5 for normalized aPSF as a function of the radius in mm.

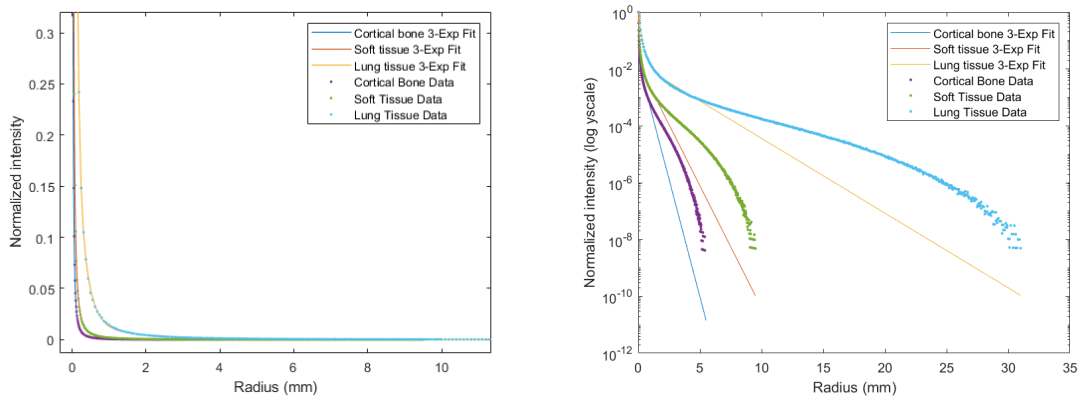
To model the aPSF, a parametric curve was fitted to the Monte Carlo data. Following the approach proposed by Derenzo [20] and adopted in subsequent works [12], [16], [17], the aPSF for each tissue was represented as a weighted sum of exponentials,

$$\text{aPSF}_{\text{model}}(\boldsymbol{x}; \boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{N} a_i \exp\Big[b_i \, \|\boldsymbol{x}\|\Big], \tag{3.7}$$

where $\|\boldsymbol{x}\| = \sqrt{x^2 + y^2}$, i.e. the L^2-norm, and the number of exponential terms was varied ($N \in \{2, 3, 4, 5\}$). The model parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ were optimized with MATLAB's `fit` function from the Curve Fitting Toolbox. Performance was assessed using the coefficient of determination R^2,

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{3.8}$$

where y_i denotes the simulated aPSF values, $\hat{y}_i$ the fitted values, and $\bar{y}$ their sample mean. As reported in Table 3.1, a three-exponential model ($N = 3$) achieved the highest R^2 across all tissues.



(a) **(b)**

Figure 3.5: Parametrized model (lines) fit to simulated positron range data (dots) from Carter et al. [16] for cortical bone, soft tissue and lung tissue with **(a)** linear y-axis and **(b)** logarithmic y-axis.

Tissue Type	2-Exp	3-Exp	4-Exp	5-Exp
Cortical Bone	0.9927	**0.9991**	0.9964	0.9972
Soft Tissue	0.9927	**0.9993**	0.9950	0.9945
Lung Tissue	0.9926	**0.9992**	0.9975	0.9968

Table 3.1: R^2 values for exponential model fits of positron range profiles for the three different tissue types, where the models are sums of 2, 3, 4 or 5 exponentials.

The continuous aPSF model, $\mathrm{aPSF}_{\mathrm{model}}(\boldsymbol{x}; \boldsymbol{a}, \boldsymbol{b})$ was discretized by evaluating it at the physical distances corresponding to the centers of the pixels in a 2D grid surrounding the kernel center. The kernel size was automatically determined by including all pixels with values greater than or equal to a specified threshold, ensuring that the support of the kernel captures the relevant spread of the aPSF.

Three methods were explored for performing convolution with the positron range kernel based on the local tissue density $\hat{\rho}$. The first approach involved fitting model parameters across densities to allow continuous interpolation between tissue types. However, this implementation was found to be computationally too slow for practical use in its naive form. Instead, two faster approaches using linear combinations of precomputed kernels for lung tissue, soft tissue, and cortical bone were considered. In the second of these methods, a unique kernel for each pixel was constructed as a linear interpolation between two of the three tissue-specific kernels, depending on the local value of $\hat{\rho}$. The kernel was defined as

$$\kappa_{\mathrm{pr}}\left(\boldsymbol{x}; \hat{\rho}\right) = \begin{cases} (1 - c_1)\,\mathrm{aPSF}_{\mathrm{soft}}(\boldsymbol{x}) + c_1\,\mathrm{aPSF}_{\mathrm{bone}}(\boldsymbol{x}) & \rho_{\mathrm{soft}} \leq \hat{\rho} \leq \rho_{\mathrm{bone}}, \\ (1 - c_2)\,\mathrm{aPSF}_{\mathrm{lung}}(\boldsymbol{x}) + c_2\,\mathrm{aPSF}_{\mathrm{soft}}(\boldsymbol{x}) & \rho_{\mathrm{lung}} < \hat{\rho} < \rho_{\mathrm{soft}} \end{cases} \tag{3.9}$$

where the interpolation weights c_1 and c_2 are defined such that the contribution from each tissue-specific kernel transitions linearly with respect to the local tissue

density $\hat{\rho}$. Specifically:

$$c_1 = \frac{\hat{\rho} - \rho_{\text{soft}}}{\rho_{\text{bone}} - \rho_{\text{soft}}},$$
$$c_2 = \frac{\hat{\rho} - \rho_{\text{lung}}}{\rho_{\text{soft}} - \rho_{\text{lung}}}.$$

With this formulation, $c_i = 0$ when $\hat{\rho}$ equals the lower bound of the interval, and $c_i = 1$ when it equals the upper bound.

The most efficient method, the triple-image-method, applied convolution to the full image using MATLAB's `conv2` with each of the three fixed kernels, and then combined the resulting images using the same interpolation scheme as in Equation 3.9. With the aPSFs replaced with the convolved images. Although this method does not guarantee that the total image intensity is preserved (unlike pixel-wise normalized kernel combinations), differences were found to be negligible and within numerical precision. Due to its computational efficiency, this method was selected for use in the rest of the work.

3.2.4 Globally invariant vs spatially varying resolution models

Since in the spatially variant model each SPLAT has its own event-specific standard deviations, σ_{CTR} and $\sigma_{\perp}$ change from event to event. To obtain scan-level summaries and be able to compare the globally and spatially invariant models, we replace the individual values with their arithmetic means

$$\bar{\sigma}_{\text{CTR}} = \frac{1}{n}\sum_{k=1}^{n} \sigma_{\text{CTR},k}, \qquad \bar{\sigma}_{\perp} = \frac{1}{n}\sum_{k=1}^{n} \sigma_{\perp,k},$$

where n is the total number of detected coincidences and $\sigma_{\perp,k}$ and $\sigma_{\text{CTR},k}$ is the square root of the diagonal elements in the covariance matrix $S(\ell; L, \boldsymbol{\alpha})$ as defined in Equation 2.8. Note that in the globally invariant model the mean would be the same as the individual values for all k. These two dimensions will also be combined into a single scalar that reflects the overall uncertainty of a SPLAT and hence also the spatial system resolution, we define the average ellipse area

$$A_{\text{ellipse}} \;=\; \pi\, \bar{\sigma}_{\text{CTR}}\, \bar{\sigma}_{\perp}. \tag{3.10}$$

3.3 Image reconstruction

Image reconstruction was performed with the ML-EM algorithm and included both positron range and attenuation correction as part of the system model. The ML-EM algorithm was used to find optimum for the log-likelihood function. The iteration scheme was formulated as

$$f^{(k+1)} = \frac{f^{(k)}}{A^T\mathbf{1}} A_n^T \left(\frac{\mathbf{1}_n}{A_n f^{(k)}} \right) \tag{3.11}$$

where $\mathbf{1}_n = \boldsymbol{g}_n$, i.e. the measured LM data consisting of a vector of only 1's as discussed in section 2.2, and $f^{(k)}$ is the reconstructed image at iteration k and $A^T\mathbf{1}$ is the sensitivity image (discussed below). For the 0:th iteration (the "initial guess") $f^{(0)}$ was defined as a circle with uniform activity across the FOV. We note, as stated before, that the iteration scheme projects data back and forth between image space and data space to converge and that both the discretized operator A and the discretized and reduced operator A_n is used.

All brain images were reconstructed from LM data simulations using Method A or B on the ground truth objects displayed in Figure 3.3 with a system radius of 125 mm.

3.3.1 Construction of the Sensitivity Image

The sensitivity image, $\mathcal{S}$, is defined by the back projection of a unity level activity distribution distributed over the entire FOV. Mathematically, we write it as

$$\mathcal{S} = A^T\mathbf{1} \tag{3.12}$$

where $\mathbf{1}$ is a vector of ones representing unity activity at each position in the entire discretized data space. The discretization was performed over 128 (half the matrix size) angles for each pixel inside the system radius with some margin. Note here that this particular back projection is intended to project the entire data space into image space, unlike the operator A_n which only covers the points where LM events have been recorded.

3.3.2 Convergence and metrics

The convergence criterion for the iterative image reconstruction was defined to be met if

$$\Delta \coloneqq \frac{\left\| f^{(k)} - f^{(k-1)} \right\|}{\left\| f^{(k-1)} \right\|} = \frac{\left(\sum_{j=1}^{n_{\text{pixels}}} \left(f_j^{(k)} - f_j^{(k-1)} \right)^2 \right)^{1/2}}{\left(\sum_{j=1}^{n_{\text{pixels}}} \left(f_j^{(k-1)} \right)^2 \right)^{1/2}} < 10^{-3}, \tag{3.13}$$

that is, if the relative change, taken in the L^2-norm, between successive iterations $< 10^{-3}$. However, in all cases but one, this threshold was not reached within the chosen maximum of 50 iterations.

So instead of relying on a fixed iteration number, the iteration with the minimum *normalized root mean square error* (NRMSE) within the *region of interest* (ROI) was selected for all subsequent analyses. The NRMSE was defined as

$$\text{NRMSE} = \frac{\left(\frac{1}{N} \sum_{j \in \text{ROI}} \left(f_j - \hat{f}_j \right)^2 \right)^{1/2}}{\bar{f}_{\text{ROI}}} \tag{3.14}$$

where f_j and $\hat{f}_j$ are pixel intensities for the ground truth and reconstructed images, respectively, and N is the number of pixels in the ROI. The denominator $\bar{f}_{\text{ROI}}$ is the

Figure 3.6: *Region of interest* (ROI) mask used for *normalized root mean square error* (NRMSE) calculation, covering the soft tissue region of the ground truth object.

mean intensity of the ground truth within the ROI. The ROI used for this evaluation covered the soft tissue region of the ground truth object, as shown in Figure 3.6.

3.4 Code optimization

In the iterative reconstruction, data is projected forward and backward for many iterations [4]. Hence if the operators are too slow they have no clinical use. The projectors were implemented in C via MATLAB's MEX API and was called from MATLAB in the reconstruction loop. The programming made use of vector intrinsic instructions (AVX2/AVX512) in order to optimize computational times. Computational times were also reduced due the subgrids discussed above and (with some exceptions) the use of single precision (32 bit) floating point numbers as opposed to double precision were used to reduce memory consumption.

4
Results

This chapter provides outputs from the simulation steps in the SPLAT model-based simulation followed by results from the SPLAT system model. Finally, we display reconstructed images and provide quantitative similarity measures.

4.1 SPLAT model-based simulation

Figure 4.1 shows the effect of positron range applied to the upsampled ground-truth PET object as the first step in the simulation pipeline. The ground truth PET image in Figure 4.1a was normalized by its max value before the positron range operator was applied. Figure 4.1b displays the result after applying the positron-range operator. Figure 4.1c shows the difference image between Figure 4.1a and Figure 4.1b, included to highlight the subtle changes introduced by the blur. From the grayscale colorbar, we note that the difference is on the order of a few percent. Figure 4.2 shows the different simulation steps of sequential resolution loss mechanisms.

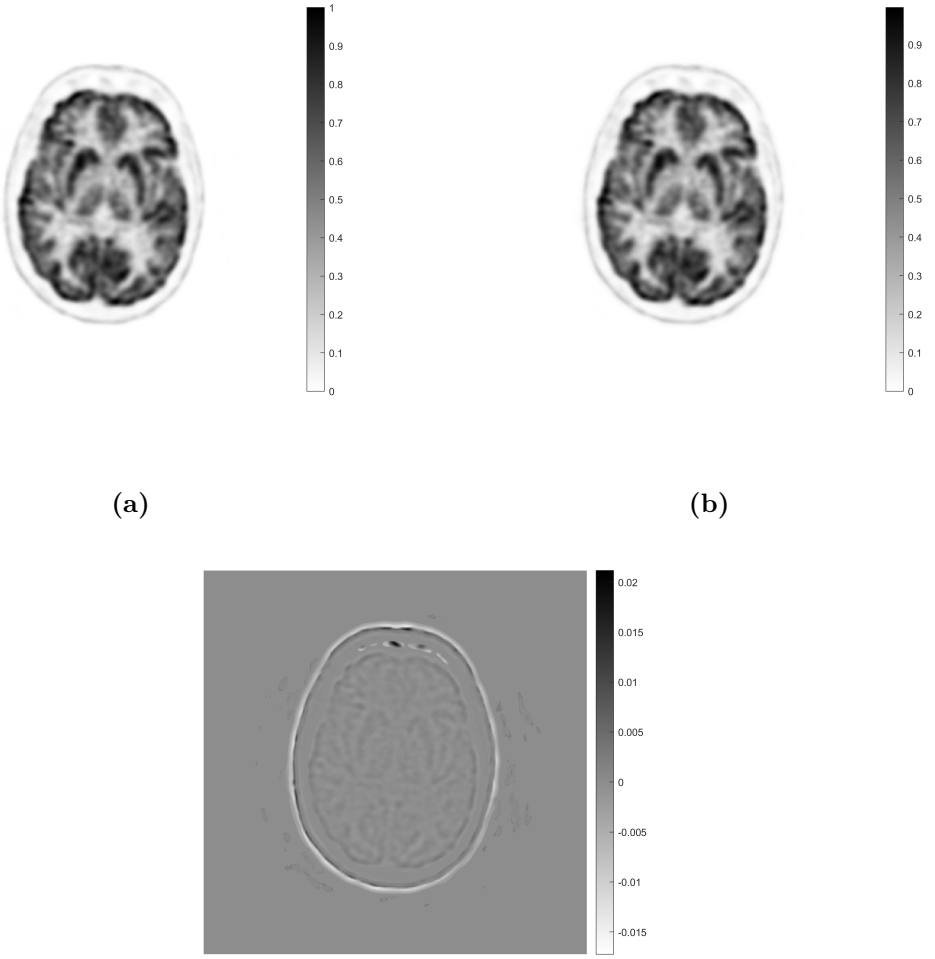
Figure 4.3 shows an example of 10 000 simulation points from the ground-truth object in Figure 3.3a before (blue) and after (brown) resolution loss simulation with 16 detectors (120 LORs) and 100 detectors (4950 LORs), using method A. As expected the blue points are on the original LORs while the brown points, after resolution loss mechanisms, have some deviation from the original LORs.

4.2 The SPLAT system model

This section provides back projections from the SPLAT system model, covering the SPLAT projection operator, attenuation correction, and positron-range modeling. It also includes a comparison between the globally invariant and spatially varying SPLAT model.

4.2.1 The SPLAT projection operator

Figure 4.4 shows the projections from the SPLAT projection operator for 100 sampled counts from a uniform disk, using parameter settings **(a)**-**(h)** listed in Table 4.1. The eight configurations represent six commercial whole-body PET systems and two expected setups for dedicated brain imaging. The images do not include attenuation



(a) **(b)**

(c)

Figure 4.1: Effect of positron-range blur in the simulation pipeline. **(a)** Upsampled version of the ground truth image in Figure 3.3a. **(b)** Image after by applying the positron-range operator to image **(a)**. **(c)** Difference image between **(a)** and **(b)**, highlighting the magnitude and spatial distribution of the changes.

correction and positron range. All images are normalized to the same grayscale scale to allow visual comparison of the spatial distribution and relative intensity across systems. Note how the uncertainty gets smaller as σ's gets smaller, left to right, top to bottom.

4.2.2 Attenuation

Figure 4.5 demonstrates how modeling attenuation changes the back projection. Without attenuation correction **(a)** the SPLATs are not weighted, leading to an over-estimation in the upper half of the FOV. Applying the CT-derived attenuation

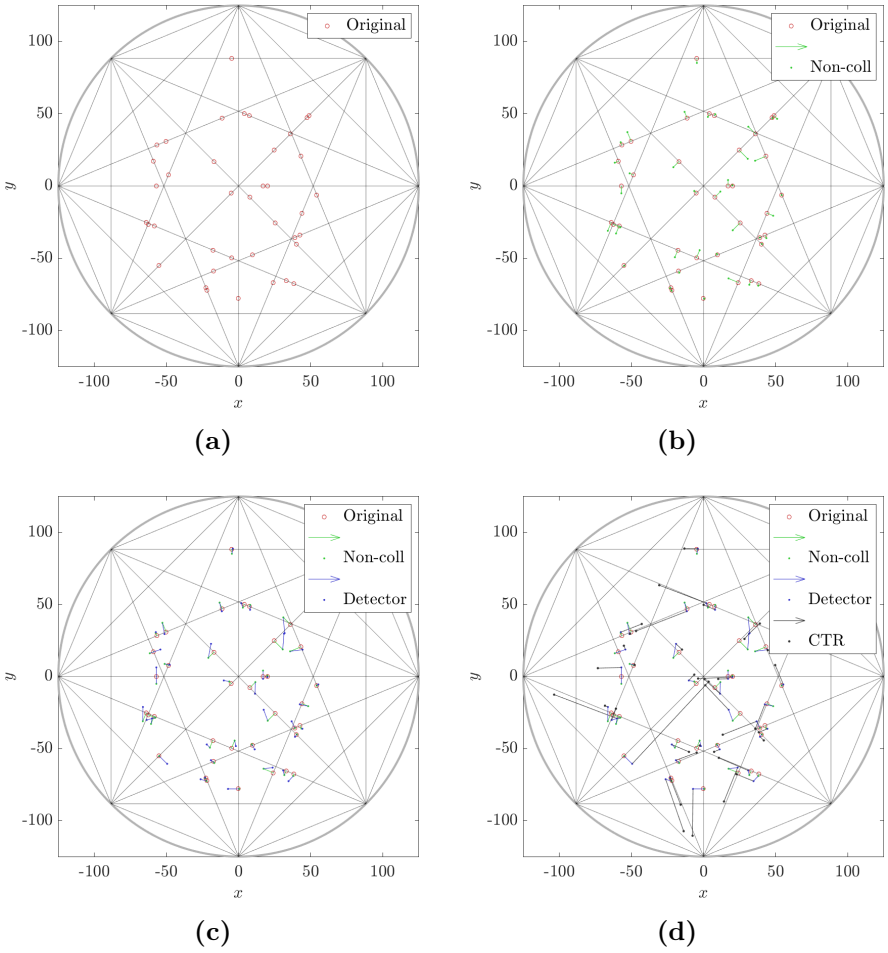


Figure 4.2: Illustration of the simulation steps with sequential resolution loss mechanisms with **(a)** showing the first simulation point at the LOR. The other figures show simulations of **(b)** non-collinearity, **(c)** detector blur and **(d)** timing uncertainty. Simulation parameters were chosen for clarity of illustration, not realism.

map **(b)** weights each SPLAT by its line-integral through attenuating tissue, yielding the corrected back projection shown in **(c)**. The difference between **(a)** and **(c)** highlights the importance of including attenuation in the system model.

4.2.3 Positron range

Figure 4.6 illustrates how positron range blurring alters the SPLATs in Figure 4.5a. When the same blur is combined with attenuation correction Figure 4.6b, both effects act together, producing the result of the full projection operator.

Figure 4.7 shows the result of projecting 10^5 simulated events from the the brain object in Figure 4.3d, using the SPLAT system model operator, including attenuation and positron range. The SPLATs are here displayed in white with black background and the system parameter settings from (h) in Table 4.1.

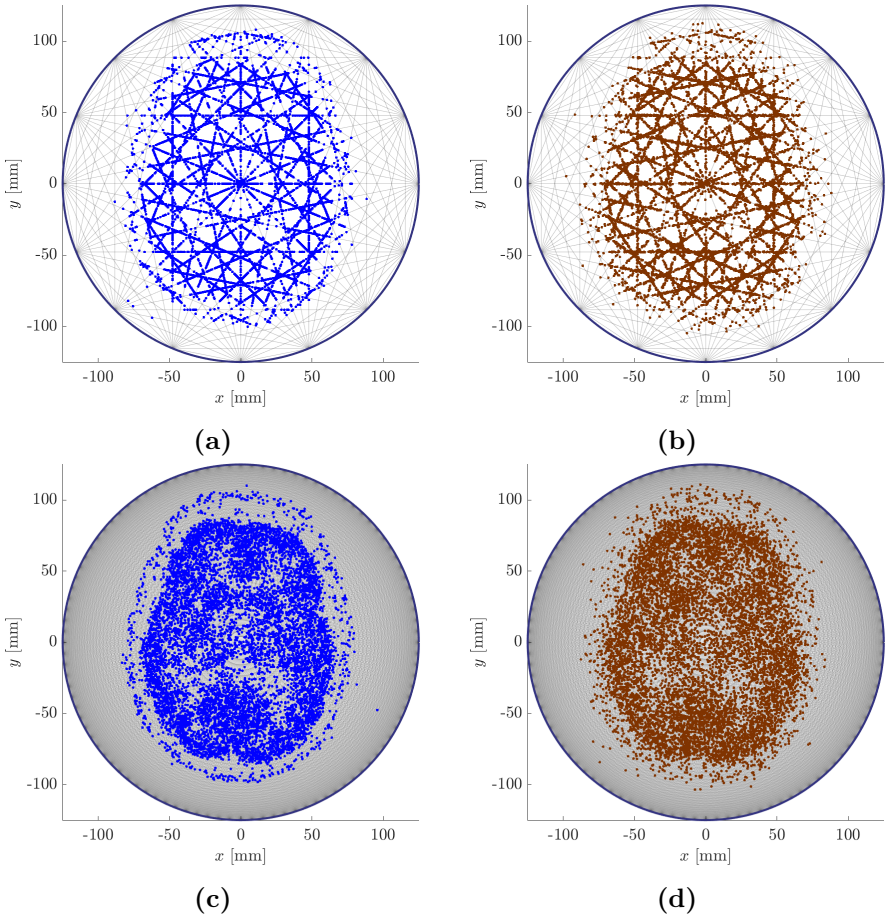


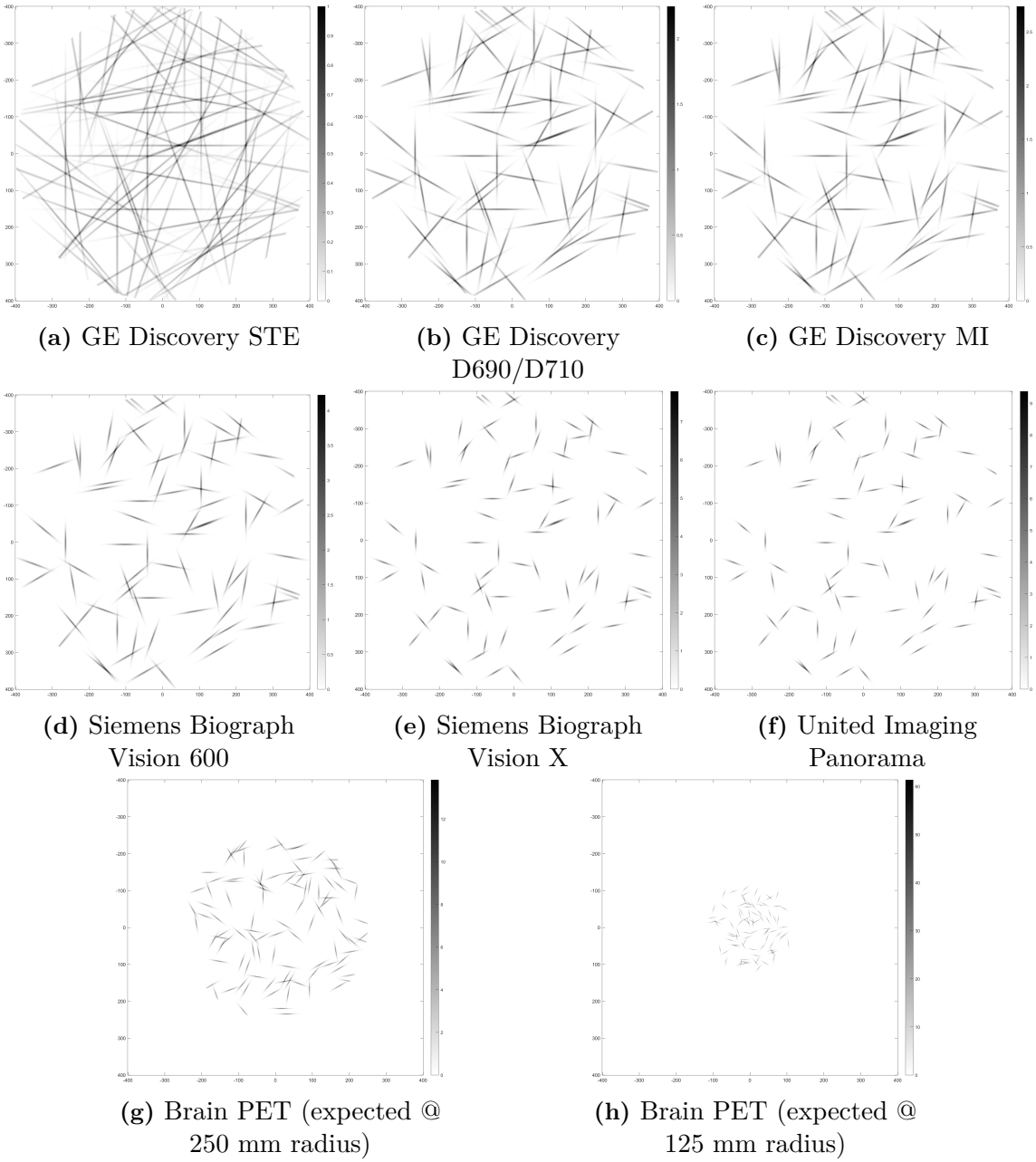
Figure 4.3: Simulation points before (blue) and after (brown) resolution loss simulations for 10 000 counts with Method A. In **(a)** and **(b)** we have 16 detectors/120 LORs and in **(c)** and **(d)** there are 100 detectors/4950 LORs.

4.2.4 Globally invariant vs spatially varying resolution models

Figure 4.8 compares the average standard deviation perpendicular to the LOR, $\bar{\sigma}_\perp$, obtained with the simple globally invariant model (in this model the average standard deviation is the same as the standard deviation) and the spatially varying model. The average was for 10^5 counts for all of the system settings in Table 4.1. Note that the difference in absolute numbers are larger for system with larger standard deviations although the relative difference between the models are quite similar for all system setups.

4.3 Image reconstruction

Figure 4.9 shows high count (10^7) reconstructed images at ML-EM iterations 1, 7, and 20 using data simulated with Method A. The system parameters correspond to label **(h)** in Table 4.1.



(a) GE Discovery STE

(b) GE Discovery D690/D710

(c) GE Discovery MI

(d) Siemens Biograph Vision 600

(e) Siemens Biograph Vision X

(f) United Imaging Panorama

(g) Brain PET (expected @ 250 mm radius)

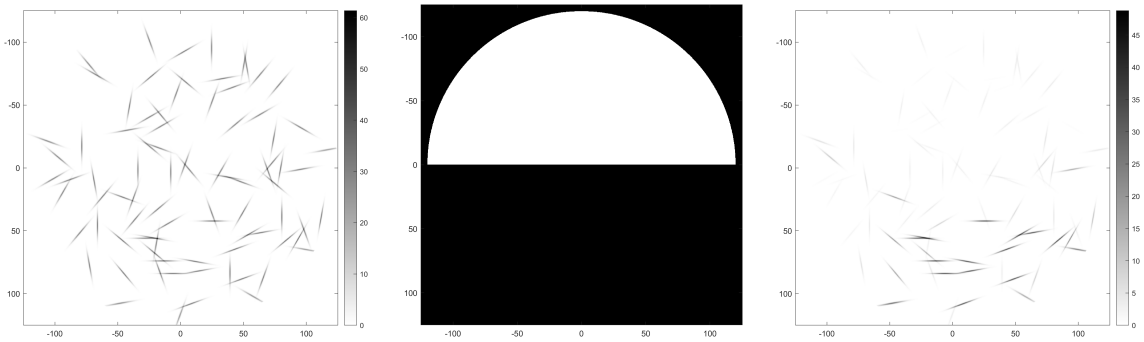
(h) Brain PET (expected @ 125 mm radius)

Figure 4.4: Back projection of 100 events sampled from a uniform activity distribution without attenuation correction and positron range. Using parameter settings from eight PET systems, including six commercial whole-body scanners and two expected configurations for dedicated brain PET. For system names and corresponding parameter values, see Table 4.1. All images share the same grayscale color scale, have the same FOV and is normalized to the maximum value of the first image, enabling direct visual comparison of the spatial concentration and intensity of each back projection.

Figure 4.10 shows NRMSE and relative L^2-norm change across all iterations for the reconstruction corresponding to the images shown at iterations 1, 7 and 20 in Figure 4.9. Note that the minimum NRMSE happens at iteration 7 and that the reconstruction algorithm have not converged at iteration 20.

Label	System	σ_{CTR} [mm]	$\sigma_{\text{PE,C}}$ [mm]	Sys rad [mm]
(a)	GE Discovery STE	-	2.1233	400
(b)	GE Discovery D690/D710	42.97	1.9110	400
(c)	GE Discovery MI	35.01	1.9110	400
(d)	Siemens Biograph Vision 600	24.51	1.6986	400
(e)	Siemens Biograph Vision X	13.69	1.4863	400
(f)	United Imaging Panorama	12.41	1.2740	400
(g)	Brain PET (expected @ 250 mm)	12.73	1.0617	250
(h)	Brain PET (expected @ 125 mm)	6.37	0.4247	125
(i)	Limit of σ_{CTR} for LYSO detectors	3.82	0.4247	125
(j)	Intermediate step	1.91	0.4247	125
(k)	Intermediate step	1.27	0.4247	125
(l)	10 ps CTR challenge [21]	0.64	0.4247	125

Table 4.1: Approximate system dependent resolution parameter settings used in the results. **(a)**-**(h)** being PET systems, including six commercial whole-body scanners and two expected configurations for dedicated brain PET. CTR and detector resolution are expressed as spatial standard deviations in mm.

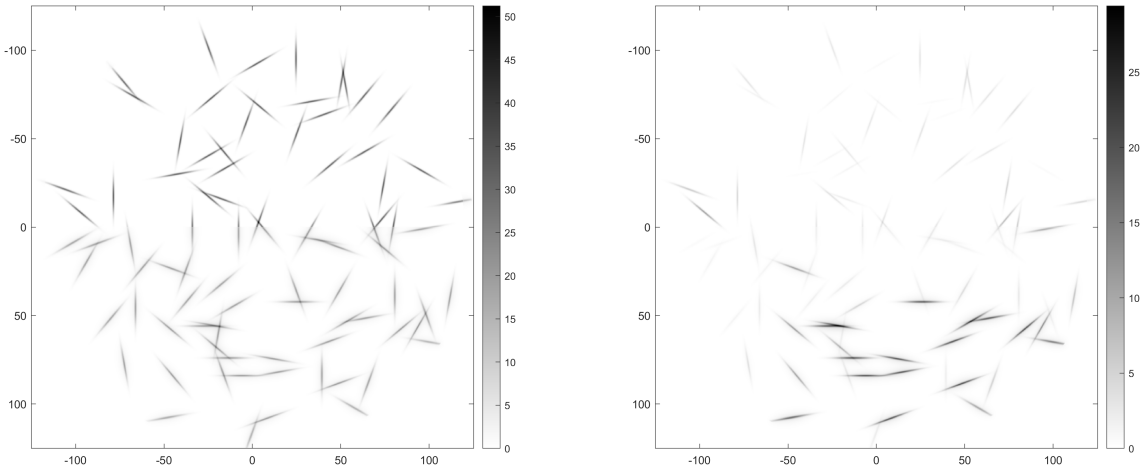


(a) **(b)** **(c)**

Figure 4.5: Impact of attenuation correction in the projector. **(a)** SPLATs from Figure 4.4h, displayed within a reduced field of view (FOV). **(b)** Binary attenuation image derived from the ground-truth CT image in Figure 3.3b. **(c)** Attenuation-weighted back projection of the SPLATs in **(a)** using the image in **(b)**.

To investigate the impact of system resolution parameters and iteration number, Figure 4.11 shows reconstructed images for simulated data of 10^5 counts, using simulation Method B. The columns represent different iterations, 1, best NRMSE and 50. The rows represent different system parameter settings, labels **(a)**, **(e)** and **(h)** in Table 4.1 although all image reconstructions, regardless of resolution parameters, used a system radius of 125 mm.

Figure 4.12 shows the NRMSE and the relative L^2-norm change across all iterations for the reconstructions corresponding to parameter settings **(a)**, **(c)**, **(e)**, **(h)** and **(l)** in Table 4.1. This illustrates how resolution parameters influence the number of



(a) **(b)**

Figure 4.6: Impact of positron-range blurring in the projector. **(a)** Back projection from Figure 4.5a after applying positron-range blurring (no attenuation). **(b)** Same blurring applied to the attenuation-weighted image in Figure 4.5c; both attenuation and positron-range effects are therefore included. The density image $\hat{\rho}$ was derived from the attenuation image in Figure 4.5b. Note that the positron range operator blurs the SPLATs and reduces their peak intensity compared to Figure 4.5a.

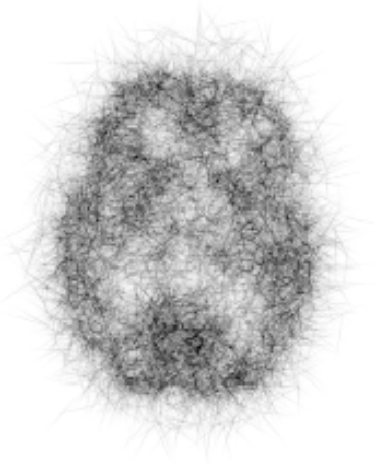


Figure 4.7: Projection of the simulated events from the brain shown in Figure 4.3d, including attenuation and positron range. SPLATs are displayed in white with a black background.

iterations needed for minimum NRMSE and speed of convergence. Note that **(i)** is the only parameter setting that converges before iteration 50.

To summarize the relationship between system resolution and reconstruction accuracy, Figure 4.13 plots the minimum NRMSE against the SPLAT kernel area A_{ellipse} for all settings in Table 4.1.

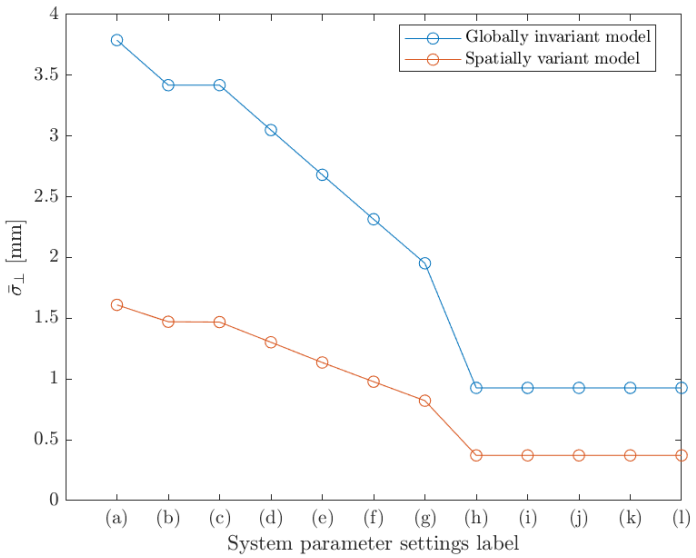
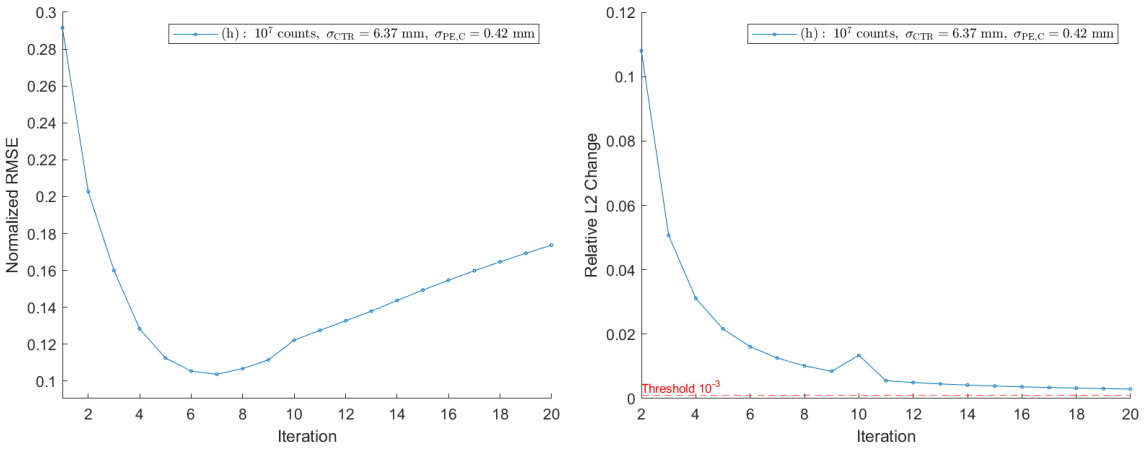


Figure 4.8: Average standard deviations perpendicular to the LOR $\bar{\sigma}_\perp$ for each of the system-parameter settings in Table 4.1. The globally invariant model assuming a single, system-wide variance (i.e. the standard deviation is equal to the average standard deviation in this model), whereas the model adapting the standard deviation locally along the LOR. We note that as expected across all settings, the spatially variant model yields consistently lower $\bar{\sigma}_\perp$ than the globally invariant model, we also note that the difference is substantial.

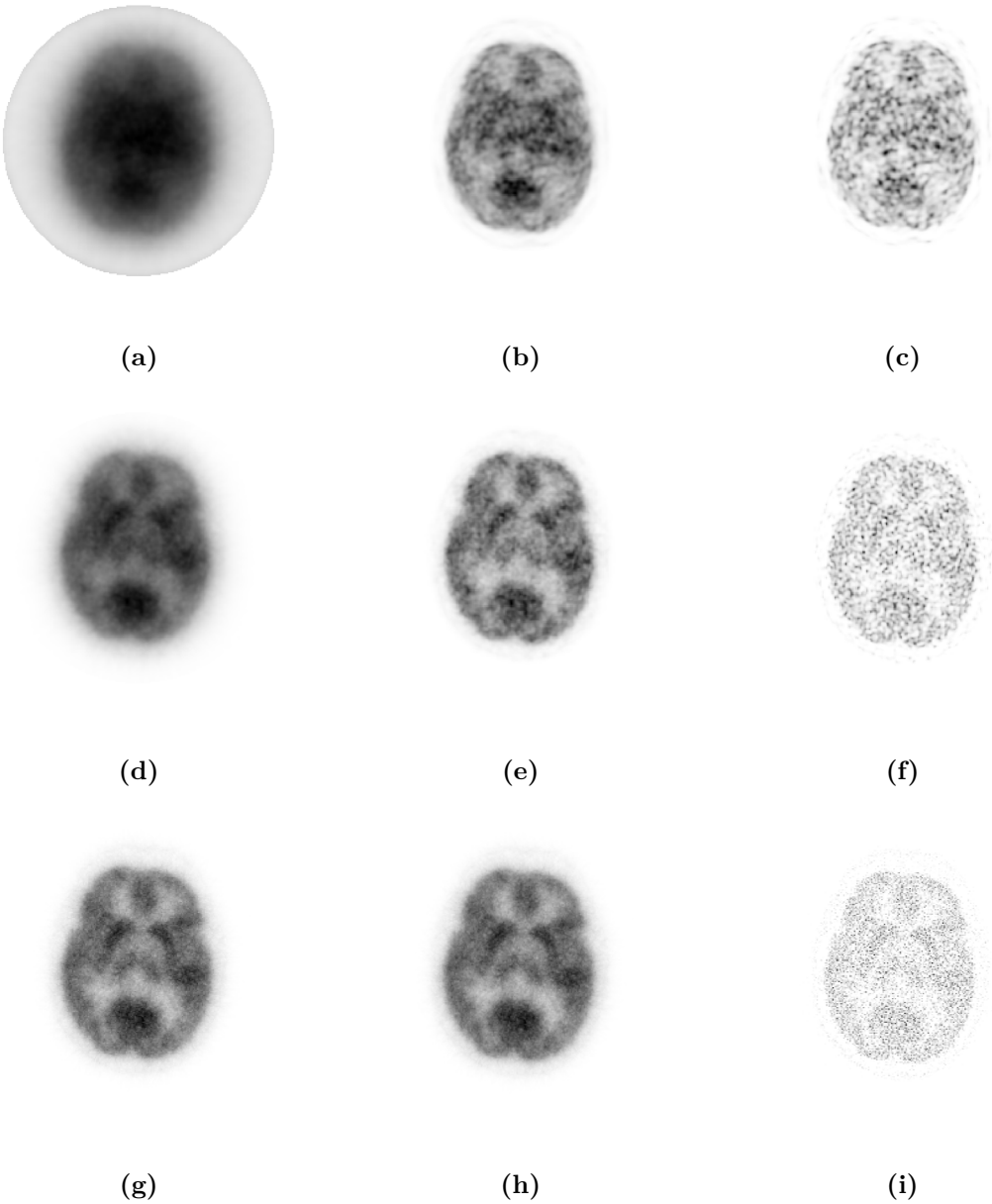
(a) **(b)** **(c)**

Figure 4.9: Reconstructed images for 10^7 counts, at iteration 1, 7 and 20 in **(a)**, **(b)** and **(c)** respectively, using parameter settings corresponding to (h) in table Table 4.1.



(a) **(b)**

Figure 4.10: **(a)** NRMSE and **(b)** relative L^2-norm change plotted across all 20 iterations for the reconstruction process. The corresponding reconstructed images at iterations 1, 7, and 20 are shown in Figure 4.9.

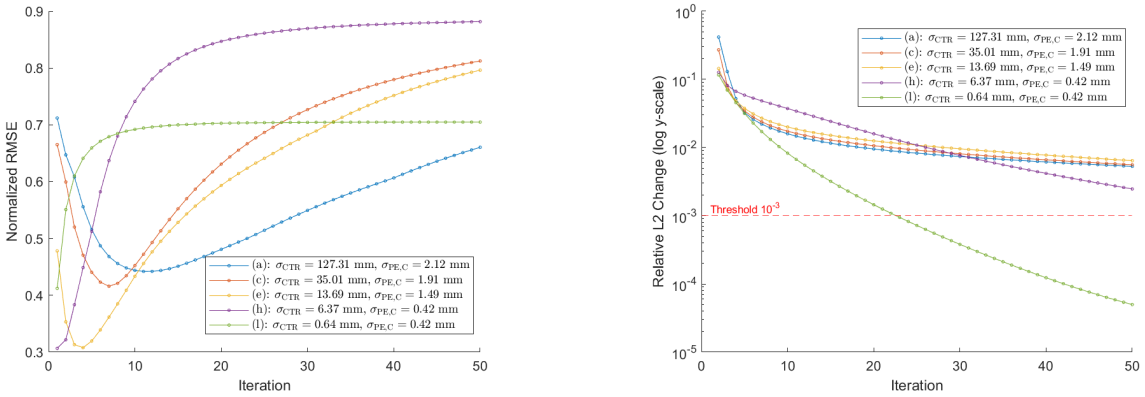


(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g) **(h)** **(i)**

Figure 4.11: Reconstructed images at different iterations and parameter settings. Columns 1 and 3 correspond to iterations 1 and 50, respectively, while column 2 shows the reconstruction with the minimum NRMSE. Row 1 corresponds to label **(a)**, row 2 to label **(e)**, and row 3 to label **(h)** in Table 4.1.



(a) **(b)**

Figure 4.12: **(a)** NRMSE and **(b)** relative L^2-norm change across all reconstruction iterations for parameter settings **(a)**, **(c)**, **(e)**, **(h)**, and **(i)** in Table 4.1. Note that the vertical axis is logarithmic in figure **(b)**.

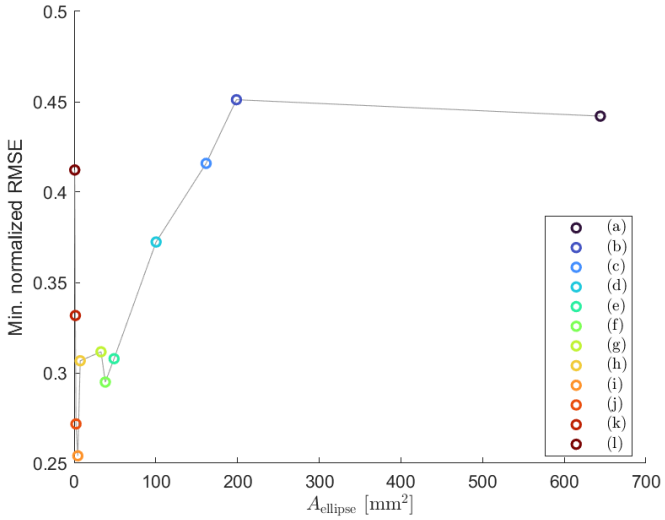


Figure 4.13: Minimum NRMSE as a function of SPLAT area, A_{ellipse}, for the reconstructions corresponding to the parameter setup in Table 4.1.

5
Discussion

This section discusses the modeling, methods, and results presented in the thesis. We begin with the SPLAT model-based simulation, followed by the SPLAT system model and its components. This is followed by a discussion on image reconstruction. Finally, we reflect on potential extensions to 3D and considerations related to code runtime.

5.1 SPLAT model-based simulation

For simplicity of implementation method A used all possible detector pairs even though some LORs would have too much angulation with respect to the detector. That is, two detection points right next to each other had a LOR defined between them. However, the use of masks (setting values to zero outside of object) to prevent counts to far from the ground truth object make it unlikely to have counts in such a LOR. Still, for improved realism and possibly efficiency this issue need to be resolved.

For method B, all pixels considered (pixels with non-zero values) are already in the ground truth object due to masking. The LORs possible are therefore defined by the pixels of the ground truth object and the number of angles n_θ. One advantage of this approach is that it is likely that no unnecessary data points are considered by the algorithm.

The two methods produce different biases in origins of detection points. For method A a bin-size is set by the user which decides the step size along the LOR when creating the LOR PDF. The implication being that each simulation point will be on the border of such step. The question of how to choose the appropriate bin-size in relation to the matrix size remains unanswered in this report. For Method B the simulation points are constrained to the pixel grid and is hence related to how the upsampling of the ground truth object has been done.

With proper binning and adaptation of the data simulated from the methods A and B they have the possibility to be used for conventional PET and even SPECT as well. Another possible use of the simulation methods is to relatively quickly produce large LM data sets for the training of deep-learning models. As in many other areas, the limited amount of training data available for PET is a bottle neck for deep-learning applications.

5.1.1 Consequences of different choices of LOR discretization

The illustrations in Figure 3.1 and Figure 3.2 shows different selections of LORs. Moses [5] points out that (in conventional PET) the number of LORs determine how the activity distribution, f, gets sampled and is hence a fundamental limit of resolution for PET. For the simulations in this project a very large number of LORs have been used in order to mimic the effectively continuous novel brain PET scanner. In addition the simulation of the resolution loss mechanisms add randomness to the measured LOR producing more realistic TOF LM data. For the extension to 3D some further considerations need to be made. Method B used a equidistant angles between LORs at each pixel in 2D, however, there is no analytical analogue for placing equidistantly spaced points on a sphere. This implies that the LOR directions must be decided with some approximation - of which there are several alternatives. For method A the choice of detection points is arbitrary and it would be straight-forward to place detectors with equal distance apart on a cylinder with LORs between all of them.

5.2 The SPLAT system model

5.2.1 Positron range

Our positron range model deliberately ignores edge effects: we convolve the with a spatially invariant, isotropic kernel, implicitly assuming that the positron range is radially the same in all directions. In reality, a positron that leaves a high-density medium for a low-density one (or vice-versa) experiences an asymmetric spread; a truly accurate kernel would therefore have to adapt its shape to the surrounding voxels. Because the triple-image-method (that convolves the full image using MATLAB's `conv2` with each of the three fixed kernels and then combines the resulting images) hinges on spatial invariance, it would not work. The other positron range methods explored could work with a spatially variant kernel, but neither could be slotted into SPLAT without substantial re-engineering. Implementing a unique, non-isotropic, position-specific kernel for every pixel would in the worst case require orders of magnitude more computational burden that lay beyond the scope of this thesis. Nevertheless, this type of positron range correction has recently been demonstrated by Kertész et al. [13].

5.2.2 Globally invariant vs spatially varying resolution models

The results in Figure 4.8 showed that the difference between the globally and spatially invariant resolution models is larger in absolute terms for systems with lower resolution. This means that the improvement gained from switching to the spatially invariant model is especially important for those lower resolution systems. However, even for high-performing systems, like the dedicated brain PET scanner, we observed that the spatially invariant model consistently produced values that were

less than half of those from the globally invariant model. This indicates that the gain is not only present but also substantial even in these more precise systems. Since the spatially varying covariance matrices are precomputed before the iterative reconstruction starts, they do not add a significant runtime cost and are not a bottleneck in the reconstruction process. Taken together, these results make a strong case for always preferring the spatially invariant model over the global one. This is, assuming that the model is accurate and that smaller standard deviation results in better images.

We chose to analyze only the diagonal elements of $S(\ell; L, \boldsymbol{\alpha})$ in Figure 4.8. It is worth noting, however, that because the detector blur covariance is rotated before being added to the other components in Equation 2.8, the resulting covariance matrix $S(\ell; L, \boldsymbol{\alpha})$ is generally not diagonal. A more complete characterization of the spread —e.g., through eigenvalue analysis or rotation to principal axes — might better capture the effective spatial resolution in both dimensions. However, since the non-zero activity in the PET object is located relatively far from the detectors and the detector variance $\sigma_{\mathrm{PE,C}}$ is small, the rotation effects are minimal.

5.3 Image reconstruction

5.3.1 Convergence and NRMSE

Determining how many iterations to run during reconstruction was not straightforward. Our results showed that more iterations do not lead to better-looking images. As described in [22], high iteration numbers can lead to increased spatial variance in the reconstructed images. They suggest using strategies such as different stopping criteria, post-reconstruction filtering or Grenander's method of sieves [23] to address this, particularly in cases where no ground truth is available. In contrast, we relied on access to ground truth and used the image with the lowest normalized RMSE (NRMSE) for each setting, assuming this represented the best result. This served as a practical stopping criterion. A more detailed investigation into stopping criteria would have been necessary in a thesis focused specifically on image reconstruction, but given the scope of this project, we deem this simple approach sufficient.

5.3.2 Normalizing images before NRMSE

All reconstructed images were divided by their maximum pixel value before computing the normalized RMSE. This step was necessary because the ground-truth object was weighted in the simulation to achieve a prescribed total number of counts, making the reconstructed image count-dependent. Whereas the ground truth object reflects an activity in each pixel. In addition to this the values in the images keep growing during the iterative reconstruction because we did not run the algorithm to convergence. For LM data with 10^5 counts, even at iteration 50 the reconstructed images contained substantially smaller values than the ground truth. As a result, the raw difference between a high-iteration image and the ground truth would shrink simply because the values on average become more similar, even if the image ap-

pearance deteriorates due to the increased spatial variance (described in the previous section). Normalizing by the maximum voxel value removes this bias arising from differences in total activity.

An alternative we also considered was to scale the reconstructed images by an *activity-recovery factor* chosen so that the total number of counts in the scaled reconstructed images would match the unweighted ground truth image. Hence reflecting the true activity and not being count dependent. However, because the algorithm was not run to full convergence, the appropriate scaling factor would vary with iteration, inevitably favoring whichever iteration happened to align best with the ground truth activity. Lacking a principled way to select this factor without introducing such bias, we rejected that option and adopted simple max-value normalization instead.

5.3.3 Reconstruction performance and system resolution

In Figure 4.13 where the minimum NRMSE for each of the systems **(a)**–**(l)** in Table 4.1 plotted against their spatial resolution, represented here by the average ellipse area, A_{ellipse} given by Equation 3.10. There is no clear or consistent trend linking better resolution to improved reconstruction performance. In fact, some systems with smaller point-spread functions - i.e., lower average standard deviations - show relatively high NRMSE, while others with worse resolution perform better. This behavior was somewhat unexpected, as one might intuitively expect sharper resolution to lead to better reconstructions, especially when evaluated using NRMSE over a fixed region of interest.

There are several reasons why this relationship may not be so straightforward. First, these reconstructions were done with fewer counts than our main reconstructions, since processing all systems at full statistics would have been too computationally expensive. Second, as discussed above all images were not reconstructed to convergence. Third, we evaluated the NRMSE inside a fixed 125 mm radius across all systems, even though the first six systems have a 400 mm field of view. For those systems, the evaluation region covers only a fraction of the image, which may skew the comparison.

Also, the average ellipse area, A_{ellipse}, representing the spatial resolution used for each system does not either fully capture the effects that impact image quality. Notably, the additional variance from positron range is not reflected in the average standard deviation. Furthermore, the SPLATs are scaled to compensate for attenuation, which is also not reflected in the average ellipse area measure.

In summary, although the average ellipse area, A_{ellipse}, as we have defined it has some limitations. As previously discussed, a refined definition may still serve as a useful indicator of spatial resolution and might have a stronger trend linking reconstruction performance and RMSE. Given limited experience within this, and acknowledging

that many other factors were not controlled or fully accounted for, we refrain from drawing definitive conclusions, as doing so would be beyond the scope of this thesis.

5.4 Numerical error

5.4.1 Discretization error

The SPLATs are evaluated at the grid points in image space and do hence not require any interpolation to find accurate values. The normalization of the SPLATs were done by dividing by the sum of the values in the subgrid which although it does normalize may have substantial error from its continuous counterpart.

5.4.2 Floating point error

In order to keep down memory usage and improve computational speed single precision (32 bit) floating point number where used in MATLAB and with C (MEX) files with a few exceptions. The computation of variance due to non-collinearity in Equation 2.11, which includes arithmetic with numbers having very different orders of magnitude, used double precision in order to avoid truncation that produce obviously incorrect results.

5.4.3 Error due to the use of subgrids

The subgrids used for evaluating the SPLATs where set to have a size proportional to the maximum standard deviation. For typical values this corresponds to values of the SPLATS of order 10^{-13} at the border of the subgrid whereas the largest value in such a SPLAT is approximately 10^{-1}. This implies that the addition of values outside the subgrid would have been truncated in final result due to single precision. Hence, given single precision, the use of subgrids are not likely to affect the computation in any meaningful way, if at all. For double precision on the other hand values of this size can affect the result albeit negligible in this context.

5.5 Extension to 3D

The extension of the positron range operator to 3D could in principle be done by adding a third dimension to the kernel and evaluating its values in an analogous way and using `convn` instead of `conv2`. Similarly for the SPLAT projector, the extension basically is done by adding an index for the third dimension to the code and using a 3D subgrid. As for the attenuation operator, it was implemented to work in 3D and would only require the addition of 3D input to operate in 3D. Of, course working in 3D would come with significantly increased computational times.

With simulation method A relying on the X-ray transform algorithm (both for creation of sinogram and attenuation correction), the first step of the simulation is already functional in 3D. For method B one would require projection of detection points onto the cylinder in 3D (as opposed to the circle in 2D) requiring more code development. Similarly for the resolution loss mechanisms, they would have to be adapted for shifting of points and projection onto a cylinder instead of a circle which is a more complicated task.

5.6 Code runtime

Parts of the code used in the project was optimized in C. Due to the fact that the code ran on the authors personal computers of different types evaluation of the performance was difficult. Also, proper optimization would require sufficient knowledge of the system that the could would run on. The reconstruction of images simulated with 10^7 counts took more than a day and would hence not be clinically usable. However, we expect that significantly reduced computational times would be possible given more computational resources (such as GPU acceleration) and professional level optimization.

6
Conclusion

This thesis set out to implement a physics-based system model for list-mode, time-of-flight PET data (the SPLAT system model) acquired with a continuous, multilayer monolithic detector. By embedding the resolution-loss mechanisms: non-collinearity, detector blur, and coincidence-timing uncertainty, in a spatially varying Gaussian and incorporating both positron-range and photon-attenuation effects, the model addresses the specific challenges posed by the dedicated brain-PET geometry.

Contributions

(i) **Projection operator.** A forward operator $\mathcal{A}_{\text{splat}}$ was implemented whose event-specific covariance matrix captures (a) non-collinearity variance varies with location, (b) detector blur rotated into the LOR frame, and (c) a constant CTR subsection 2.3.1. These covariances were pre-computed and rotated into the local LOR frame, with each Gaussian SPLAT evaluated on a truncated local sub-grid to reduce both memory usage and interpolation errors.

(ii) **Physics add-ons.**
- Attenuation correction ($\mathcal{A}_{\text{att}}$) uses a CT-derived attenuation image transformed via a bilinear Hounsfield units–μ relation and line-integral evaluation in C subsection 3.2.2.
- Positron-range blurring ($\mathcal{A}_{\text{pr}}$) is handled with three exponential kernels (lung, soft tissue, cortical bone) fitted to *GATE* Monte Carlo data, and applied by convolving the image with each tissue-specific kernel followed by interpolation based on the density of the local tissue, subsection 3.2.3.

(iii) **Complete pipeline.** The full SPLAT system model operator $\mathcal{A} = \mathcal{A}_{\text{att}}\mathcal{A}_{\text{splat}}\mathcal{A}_{\text{pr}}$ was embedded in an ML-EM reconstruction, section 3.3.

(iv) **Simulation framework.** Two list-mode data simulators—"Method A" (detector-pair sampling) and "Method B" (pixel-center × angle grid)—were developed to validate the model under controlled conditions, section 3.1.

Limitations

- All results are presently 2D; a full 3D implementation will raise the computational cost, demanding GPU acceleration or further code optimization.
- Positron-range kernels assume isotropy, neglecting boundary effects (e.g. soft-tissue–lung interfaces).

- Depth-of-interaction blur in monolithic crystals is not yet modeled.
- Stop criteria relied on ground truth; clinical scans require alternative metrics.
- Single-precision arithmetic was used for most operations; a 3D pipeline may need mixed-precision or selected double-precision support to avoid numerical instability.

Future work

Several aspects remain to be explored beyond the scope of this thesis. Below are selected directions for future work that we consider relevant.

1. **Full 3D SPLAT.** Extend the sub-grid evaluation, positron-range convolution, and LOR sampling to volumetric data using GPU-accelerated kernels.
2. **Depth-of-interaction modeling.** Introduce a layer-dependent variance term in the SPLAT covariance for monolithic-detector depth-of-interaction uncertainty.
3. **Reconstruction.** While a basic ML–EM reconstruction pipeline was implemented to demonstrate that the system model behaves as expected, a comprehensive evaluation of reconstruction techniques was not within the scope of this thesis. Future work should explore how the SPLAT model can be integrated into more advanced or regularized reconstruction frameworks, possibly using deep learning models.
4. **Clinical validation.** Test SPLAT on other phantoms or patient data to confirm its clinical utility.
5. **Edge-aware positron range.** Adopt anisotropic, boundary-sensitive kernels to model asymmetric range effects across tissue interfaces.

Final remarks

This thesis demonstrated that a spatially variant SPLAT system model, integrating non-collinearity, detector blur, coincidence timing resolution, attenuation, and positron-range effects, can be successfully implemented within iterative PET reconstruction frameworks. While extending the model to full 3D implementations and incorporating additional physical effects remains challenging, the developed framework provides a solid foundation for advancing high-resolution TOF-PET imaging, particularly for specialized continuous-detector systems.

Bibliography

[1] J. L. Prince and J. M. Links, *Medical imaging signals and systems*, eng, 2., [updated] ed. Boston: Pearson, 2015, ISBN: 978-0-13-214518-3.

[2] S. Anand, H. Singh, and A. Dash, "Clinical Applications of PET and PET-CT", en, *Medical Journal Armed Forces India*, vol. 65, no. 4, pp. 353–358, Oct. 2009, ISSN: 03771237. DOI: 10.1016/S0377-1237(09)80099-3. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377123709800993 (visited on 02/04/2025).

[3] C. Catana, "Development of Dedicated Brain PET Imaging Devices: Recent Advances and Future Perspectives", en, *Journal of Nuclear Medicine*, vol. 60, no. 8, pp. 1044–1052, Aug. 2019, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.118.217901. [Online]. Available: http://jnm.snmjournals.org/lookup/doi/10.2967/jnumed.118.217901 (visited on 04/27/2025).

[4] A. Iriarte, R. Marabini, S. Matej, C. Sorzano, and R. Lewitt, "System models for PET statistical iterative reconstruction: A review", en, *Computerized Medical Imaging and Graphics*, vol. 48, pp. 30–48, Mar. 2016, ISSN: 08956111. DOI: 10.1016/j.compmedimag.2015.12.003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0895611115001901 (visited on 04/11/2025).

[5] W. W. Moses, "Fundamental Limits of Spatial Resolution in PET", *Nuclear instruments & methods in physics research. Section A, Accelerators, spectrometers, detectors and associated equipment*, vol. 648 Supplement 1, S236–S240, Aug. 2011, ISSN: 0168-9002. DOI: 10.1016/j.nima.2010.11.092. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144741/ (visited on 02/09/2025).

[6] A. Gonzalez-Montoro, M. N. Ullah, and C. S. Levin, "Advances in Detector Instrumentation for PET", en, *Journal of Nuclear Medicine*, vol. 63, no. 8, pp. 1138–1144, Aug. 2022, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.121.262509. [Online]. Available: http://jnm.snmjournals.org/lookup/doi/10.2967/jnumed.121.262509 (visited on 02/11/2025).

[7] A. Gonzalez-Montoro, A. J. Gonzalez, S. Pourashraf, *et al.*, "Evolution of PET Detectors and Event Positioning Algorithms Using Monolithic Scintillation Crystals", *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 3, pp. 282–305, May 2021, ISSN: 2469-7311, 2469-7303. DOI: 10.1109/TRPMS.2021.3059181. [Online]. Available: https://ieeexplore.ieee.org/document/9353691/ (visited on 04/14/2025).

[8] M. Conti, "State of the art and challenges of time-of-flight PET", en, *Physica Medica*, vol. 25, no. 1, pp. 1–11, Mar. 2009, ISSN: 11201797. DOI: 10.1016/j.ejmp.2008.10.001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1120179708000975 (visited on 04/17/2025).

[9] P. E. Kinahan, B. H. Hasegawa, and T. Beyer, "X-ray-based attenuation correction for positron emission tomography/computed tomography scanners", en, *Seminars in Nuclear Medicine*, vol. 33, no. 3, pp. 166–179, Jul. 2003, ISSN: 00012998. DOI: 10.1053/snuc.2003.127307. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0001299803700038 (visited on 03/11/2025).

[10] J. S. Lee, "A Review of Deep-Learning-Based Approaches for Attenuation Correction in Positron Emission Tomography", *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 2, pp. 160–184, Mar. 2021, ISSN: 2469-7311, 2469-7303. DOI: 10.1109/TRPMS.2020.3009269. [Online]. Available: https://ieeexplore.ieee.org/document/9143173/ (visited on 04/08/2025).

[11] M. D. Harpen, "Positronium: Review of symmetry, conserved quantities and decay for the radiological physicist", en, *Medical Physics*, vol. 31, no. 1, pp. 57–61, Dec. 2003, ISSN: 00942405. DOI: 10.1118/1.1630494. [Online]. Available: http://doi.wiley.com/10.1118/1.1630494 (visited on 02/14/2025).

[12] A. Alessio and L. MacDonald, "Spatially variant positron range modeling derived from CT for PET image reconstruction", in *2008 IEEE Nuclear Science Symposium Conference Record*, Dresden, Germany: IEEE, Oct. 2008, pp. 3637–3640, ISBN: 978-1-4244-2714-7. DOI: 10.1109/NSSMIC.2008.4774106. [Online]. Available: https://ieeexplore.ieee.org/document/4774106/ (visited on 03/11/2025).

[13] H. Kertész, T. Beyer, V. Panin, *et al.*, "Implementation of a Spatially-Variant and Tissue-Dependent Positron Range Correction for PET/CT Imaging", *Frontiers in Physiology*, vol. 13, p. 818 463, Mar. 2022, ISSN: 1664-042X. DOI: 10.3389/fphys.2022.818463. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2022.818463/full (visited on 03/07/2025).

[14] J. Cal-González, J. Herraiz, S. España, *et al.*, "Study of CT-based positron range correction in high resolution 3D PET imaging", en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 648, S172–S175, Aug. 2011, ISSN: 01689002. DOI: 10.1016/j.nima.2010.12.041. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0168900210027828 (visited on 04/08/2025).

[15] J. Cal-Gonzalez, M. Perez-Liva, J. L. Herraiz, J. J. Vaquero, M. Desco, and J. M. Udias, "Tissue-Dependent and Spatially-Variant Positron Range Correction in 3D PET", *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2394–2403, Nov. 2015, ISSN: 0278-0062, 1558-254X. DOI: 10.1109/TMI.2015.2436711. [Online]. Available: http://ieeexplore.ieee.org/document/7112159/ (visited on 03/10/2025).

[16] L. M. Carter, A. L. Kesner, E. C. Pratt, *et al.*, "The Impact of Positron Range on PET Resolution, Evaluated with Phantoms and PHITS Monte Carlo Simulations for Conventional and Non-conventional Radionuclides", en, *Molecular Imaging and Biology*, vol. 22, no. 1, pp. 73–84, Feb. 2020, ISSN: 1536-1632, 1860-2002. DOI: 10.1007/s11307-019-01337-2. [Online]. Available: http://link.springer.com/10.1007/s11307-019-01337-2 (visited on 03/07/2025).

[17] J. Cal-González, J. L. Herraiz, S. España, *et al.*, "Positron range estimations with PeneloPET", *Physics in Medicine and Biology*, vol. 58, no. 15, pp. 5127–5152, Aug. 2013, ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/58/15/5127. [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/58/15/5127 (visited on 03/07/2025).

[18] S. Staelens and I. Buvat, "Monte Carlo Simulations in Nuclear Medicine Imaging", en, in *Advances in Biomedical Engineering*, Elsevier, 2009, pp. 177–209, ISBN: 978-0-444-53075-2. DOI: 10.1016/B978-0-444-53075-2.00005-8. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780444530752000058 (visited on 05/18/2025).

[19] H. Gao, "Fast parallel algorithms for the x-ray transform and its adjoint", en, *Medical Physics*, vol. 39, no. 11, pp. 7110–7120, Nov. 2012, ISSN: 0094-2405, 2473-4209. DOI: 10.1118/1.4761867. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/10.1118/1.4761867 (visited on 04/07/2025).

[20] S. Derenzo and B. (L. B. L. California Univ., *Precision measurement of annihilation point spread distributions for medically important positron emitters*, Apr. 1979.

[21] P. Lecoq, C. Morel, J. O. Prior, *et al.*, "Roadmap toward the 10 ps time-of-flight PET challenge", *Physics in Medicine & Biology*, vol. 65, no. 21, 21RM01, Oct. 2020, ISSN: 1361-6560. DOI: 10.1088/1361-6560/ab9500. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6560/ab9500 (visited on 05/30/2025).

[22] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography", *Physics in Medicine and Biology*, vol. 51, no. 15, R541–R578, Aug. 2006, ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/51/15/R01. [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/51/15/R01 (visited on 05/15/2025).

[23] U. Grenander, "Taxonomic Patterns", in *Regular Structures*, F. John, L. Sirovich, and J. P. LaSalle, Eds., vol. 33, Series Title: Applied Mathematical Sciences, New York, NY: Springer New York, 1981, pp. 423–450, ISBN: 978-0-387-90560-0 978-1-4612-5905-3. DOI: 10.1007/978-1-4612-5905-3_9. [Online]. Available: http://link.springer.com/10.1007/978-1-4612-5905-3_9 (visited on 05/17/2025).

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY