Geometry Optimization and Evaluation of PET inserts for Simultaneous PET/MR Neuroimaging

by

Mohammadreza Teimoorisichani

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Physics and Astronomy

University of Manitoba

Winnipeg

Abstract

A positron emission tomography (PET) insert with a compact ring diameter is currently under design by our group. The PET insert is a brain-dedicated scanner which is designed to retrofit into the Siemens Magnetom 7T brain Magnetic Resonance (MR) scanner. A dual-layer offset (DLO) design is considered for the detectors of the PET insert to provide depth-of-interaction information and reduce resolution degradation due to parallax error. A wide range of detector geometries was evaluated through Monte Carlo simulations to characterize the performance of each detector and study the optimum detector geometry for the PET insert. More than 200 DLO detectors with a scintillation material of LYSO and detector thicknesses between 10 to 30 mm, different layer thickness ratios and crystal sizes of 0.5 to 4 mm were evaluated in this study. The effects of each detector geometry on radial mispositioning of the incident photons, coincidence response functions, scanner sensitivity and scanner resolution were studied. The effects of layer-misassignment in DLO detectors due to inter-crystal scatter were also studied. The count rate performance of several scanners with different detector sizes and thicknesses were evaluated to provide a comprehensive model for the count rate performance of each scanner with a wide range of deadtime properties. A scalable dual-GPU list-mode image reconstruction algorithm has also been developed, which was used in the evaluation of different brain PET scanners. Based on the results of this work, a single-ended readout DLO detector with an area of between 1.5 to 9.6 cm^2, a total/front/back layer thickness of 15/6/9 or 20/8/12 mm, and a crystal size of 2 to 3 mm is recommended.

Preface

Except for certain sections in Chapter 8, all of the work presented henceforth was conducted by the author, Mohammadreza Teimoorisichani. All required permissions to reuse the published material have been granted and are indicated at the beginning of related chapters.

The original idea of the chosen GPU implementation for iterative list-mode image reconstruction was proposed by:

Cui, J.Y., Pratx, G., Prevrhal, S. and Levin, C.S., 2011. "Fully 3D list-mode time-of-flight PET image reconstruction on GPUs using CUDA", Medical Physics, 38(12), pp.6775-86.

The idea of intraoperative prostate imaging as explained in Section 8.1 was proposed by Dr. Jonathan Thiessen and his collaborators at the Lawson Imaging Centre in London, Ontario. The small animal PET insert (NuPET™) described in Section 8.2 and the indexed histogram data format used in this scanner were developed by Cubresa, Winnipeg, Manitoba (a Canadian medical imaging instrumentation vendor).

A version of the detector geometry optimization work presented in Chapter 4 has been previously published in:

Teimoorisichani, M. and Goertzen, A. L., 2016, October. "Geometry optimization of dual-layer offset detectors for compact ring diameter PET systems," In 2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD) IEEE.

and

Teimoorisichani, M. and Goertzen, A. L., 2019. “Geometry Optimization of a Dual-layer Offset Detector for Use in Simultaneous PET/MR Neuroimaging” IEEE Transactions on Radiation and Plasma Medical Sciences, 3(3), pp. 275–284.

A version of the count rate study explained in Chapter 5 has been submitted to Physics in Medicine and Biology for publication. A manuscript summarizing the method and results of Chapter 6 has been approved for publication in Physics and Medicine and Biology under the title:

Teimoorisichani, M. and Goertzen, A. L., 2019. “A study of inter-crystal scatter in dual-layer offset scintillator arrays for brain-dedicated PET scanners” Physics in Medicine and Biology.

A separate manuscript on the proposed algorithms and reported results in Chapter 7 and Chapter 8 will be prepared for publication.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
APD	Avalanche photodiode
API	Application programming interface
BGO	Bismuth germanate
CRF	Coincidence response function
CT	Computed tomography
CTW	Coincidence timing window
CPU	Central processing unit
CUDA	Compute Unified Device Architecture
DER	Dual-ended readout
DG	Distributed gradient
DLO	Dual-layer offset
DLR	Discrete-layer readout
DML-EM	Distributed ML-EM
DOI	Depth-of-interaction
EC	Electron capture
FBP	Filtered backprojection
FOV	Field-of-view
FWHM	Full width at half maximum
FWTM	Full width at tenth maximum
GATE	Geant4 application for tomographic emission
GPU	Graphics processing unit
GSO	Gadolinium orthosilicate
HRRT	High-resolution research tomograph
LAC	Linear attenuation coefficient
LFS	Lutetium fine silicate

LGSO	Lutetium-gadolinium oxyorthosilicate
LLD	Lower level discriminator
LOR	Line-of-response
LSO	Lutetium oxyorthosilicate
LuAP	Lutetium aluminum perovskite
LUT	lookup table
LuYAP	Lutetium-yttrium aluminum perovskite
LYSO	Lutetium-yttrium oxyorthosilicate
MAP	Maximum a posteriori
MC	Monte Carlo
ML-EM	Maximum likelihood expectation maximization
MPPC	Multi-pixel photon counter
MRF	Markov random field
MRI	Magnetic resonance imaging
MSRB	Multi-slice rebinning
NaI	Sodium iodide
NECR	Noise equivalent count rate
OS-EM	Ordered-subsets expectation maximization
OSMAPOSL	Ordered Subset version of Green's MAP One-Step-Late algorithm
OSL	One-step-late
P2P	Peer-to-peer
PCIe	Peripheral component interconnect express
PET	Positron emission tomography
PM	Photomultiplier
PS-PM	Position-sensitive photomultiplier
PSF	Point spread function
PSMA	Prostate specific membrane antigen
SD	Standard deviation
SER	Single-ended readout
SI	(French) Système international d'unités (International system of units)

SiPM	Silicon photomultiplier
SNR	Signal-to-noise ratio
SoA	Structure of arrays
SPECT	Single photon emission computed tomography
STL	Standard template library
ToF	Time-of-flight
TOR	Tube of response
ULD	Upper level discriminator
UTE	Ultrashort echo time

Chapter 1 : Introduction

Positron emission tomography (PET) is a *nuclear medicine* imaging technique that is widely used for functional imaging in human and animal studies. As images obtained from a PET scanner include minimal anatomical information, integrating PET scanners with anatomical imaging modalities has been an active area of research. We are currently designing a PET scanner that can be used in simultaneous anato-functional imaging of the human brain. The brain PET scanner is to be designed such that it can function both as a stand-alone PET scanner and as a *magnetic resonance imaging* (MRI) compatible PET-insert. This dissertation begins with an overview of the necessary physics behind nuclear medicine. In Sections 2.2 to 2.6 of Chapter 2, basic concepts and definitions in nuclear medicine are explained. As a large portion of this dissertation involves studying the performance of PET scanners, the focus of the introductory sections is on the impact of detector geometrical parameters on the performance of a PET scanner. In Sections 2.7 to 2.9, the data generation and processing pipeline in a PET scanner are also discussed, followed by an overview of the existing limitations of the current generation PET scanners in Section 2.10.

Numerical simulations were chosen in this work to perform a detailed evaluation of the performance of PET detectors suitable for our brain PET scanner. Since, a large portion of the presented data in this dissertation was obtained from simulations Section 2.11 of Chapter 2 reviews the basis of *Monte Carlo* (MC) simulations with a focus on nuclear medicine imaging applications.

An important goal of this project was to design an optimum detector for the brain-dedicated PET scanner. Since the geometry optimization project required quantitative evaluation of the images of several potential PET scanners, a review on the basis of image reconstruction is done in

Sections 3.1 and 3.2 of Chapter 3. In particular, *iterative reconstruction algorithms* were used in this project and therefore a thorough explanation of iterative image reconstruction algorithms in nuclear medicine is given. Due to the large number of image reconstructions required for the geometry optimization project, we chose to design a highly optimized reconstruction framework using *graphics processing units* (GPUs). In Section 3.3, basic concepts of GPU structure and programming related to the developed image reconstruction algorithm are explained along with a literature review on the available GPU-based image reconstruction algorithms.

Chapter 4 marks the beginning of the detector geometry optimization process in this work. In particular, the geometry optimization explained in Chapter 4 was designed given a fixed detector area for the brain PET scanner and a detector architecture of *dual-layer offset* (DLO). Given the geometry of the chosen brain MR scanner, a set of geometrical properties are discussed as geometrical assumptions and variables that are the target of the optimization process in this chapter. The details of the geometrical assumptions and variables to be optimized are explained in Section 4.3 followed by an explanation of the details of detector and scanner modeling in Sections 4.4 to 4.6. Since the geometry optimization experiment was conducted through MC simulations, an investigation on the validation of the performed MC simulations is done in Section 4.7 prior to the results of the geometry optimization experiment in Section 4.8.

The geometry optimization algorithm discussed in Chapter 4 evaluates the performance of various DLO detectors from the standpoint of *scanner spatial resolution* and *scanner sensitivity*. Therefore, several unrelated detector performance metrics such as *deadtime* are not included in the experiments of Chapter 4. To complete the geometry optimization step, the *count rate performance* of the brain PET scanners should also be examined. Studying the count rate performance of PET detectors suitable for a brain PET scanner also provides a theoretical model for the count rate

performance of each potential scanner. Therefore, a comprehensive study on the count rate performance of tens of possible PET scanners was conducted as explained in Chapter 5. The importance of detector area on the count rate performance of a PET scanner is discussed in Section 5.2. A wide range of geometrical properties and performance metrics were considered in this count rate evaluation study which are reviewed in Sections 5.3 and 5.4. We performed the count rate study with single layer detectors as the DLO design does not affect the count rate performance of a PET scanner. The results of the count rate study are discussed in Sections 5.5 and 5.6. The results of the geometry optimization project (Chapter 4) and detector count rate study (Chapter 5) are complementary to each other.

The geometry optimization project in this dissertation was aimed to determine the optimal detector geometry for a PET insert that can be retrofit into a selected MR scanner. Nonetheless, the optimization algorithm explained in Chapter 4 and Chapter 5 provides a generic optimization scheme that can be extended to different DLO detectors for various target scanners (e.g. a small-animal or whole-body PET scanner).

A DLO detector, as explained in Chapter 4, may misposition events to an incorrect layer when photons scatter between the crystal elements of the detector. Although the findings of Chapter 4 and Chapter 5 conclude the detector geometry optimization project, we investigated a new DLO design in Chapter 6 that can potentially reduce event mispositioning in a DLO detector. The so-called *discrete layer readout* (DLR) detector design was assessed for DLO detectors and compared to the previously used DLO detectors. The theoretical impact of inter-crystal scatter in DLO detectors and the potential benefits of a DLR-DLO design are fully explored in Sections 6.3 and 6.4, respectively. Benefiting from the results of the geometry optimization study in Chapter 4, several DLO detectors with various thicknesses were chosen for this study that are explained in

Section 6.5. The performance of the selected DLO detectors in terms of *event positioning* and *response profile* was thoroughly studied in Sections 6.8 and 6.9. As it was important to determine whether layer mispositioning in DLO detectors can deteriorate the quality of the PET images, an imaging study was also conducted that is discussed in Section 6.10.

To complete the geometry optimization task in this project, it was necessary to reconstruct and compare images of several candidate brain PET scanners. Therefore, a scalable GPU-based image reconstruction algorithm was developed to evaluate images of the brain PET scanners. Chapter 7 explains the steps involved in the development of the mentioned image reconstruction algorithm for brain PET scanners. Details of the hardware used and memory management in GPUs are discussed in Sections 7.2 and 7.3 followed by further technical implementation details in Sections 7.4 to 7.8. The performance of the developed image reconstruction algorithm was evaluated in Section 7.9 and sample reconstructed images are shown in Section 7.10.

An important advantage of the developed GPU-based image reconstruction algorithm for brain PET scanners is the high efficiency of the developed algorithm. As our collaborators at the Lawson Imaging Centre in London, Ontario required a fast and reliable reconstruction engine for their small-animal PET scanner, *NuPET*™ (Cubresa, Winnipeg, Manitoba), a similar GPU-based reconstruction algorithm was also developed for NuPET. Chapter 8 discusses the developed GPU-based image reconstruction algorithm for NuPET in detail. The project for which the NuPET reconstruction algorithm was developed is briefly reviewed in Section 8.1. The properties of the NuPET scanner and its data format are discussed in Sections 8.2 and 8.3, respectively. Since the developed reconstruction algorithm for NuPET was developed for a known scanner with fixed geometrical properties, an accurate *component-based normalization* method was also developed

for this scanner that is explained in Section 8.4. Technical details of the developed image reconstruction algorithm can be found in Sections 8.5 and 8.6 followed by a performance analysis of the reconstruction algorithm and sample reconstructed images obtained from MC simulation and acquired data in Sections 8.7 to 8.9.

Lastly, a summary of the findings of this dissertation is given in Chapter 9. In Section 9.1, the results of the detector geometry optimization project are summarized followed by a summary of the developed GPU-based image reconstruction algorithms in Section 9.2. In each section, we also discuss possible future directions related to each area.

Chapter 2 : Introduction to PET and PET Instrumentation

In this chapter, a review of underlying PET physics, instrumentation and clinical applications is provided.

2.1 Basics of Radioactivity Relevant to PET Imaging

Originating from a Greek word meaning *indivisible*, the smallest unit of each material that conserves the chemical properties of the material is called an *atom*. The discovery of subatomic particles, electron, proton and neutrons and *radioactivity* in the late 1800s and early 1900s, by J. J. Thompson, E. Rutherford, H. Becquerel and M. Curie, paved the path toward the formation of a new branch of medicine called *nuclear medicine*. A model for the atom was proposed by E. Rutherford and N. Bohr in early 1900s in which the atom was considered as a miniature planetary system with the electrons orbiting around a relatively heavy positively charged nucleus [1]. The understanding of nuclear mass, energy level and angular momentum provided a basis for explaining nuclear processes, in which a nucleus can change its state and even identity. Elements of such nuclei are referred to as *radioisotopes* and the term *radioactive* was used with the discovery of radioisotopes [2].

The number of protons in the nucleus of an atom is called *atomic number* (*Z*) and the total number of protons and neutrons is known as *mass number* (*A*) or *nucleon* number. The number of neutrons *N* in a nucleus is therefore *A* – *Z*. An element with a chemical symbol *X* is shown by ${}^{A}X$, ${}_{Z}^{A}X$ or ${}_{Z}^{A}X_{N}$. Naturally existing elements have atomic numbers of $1 \le Z \le 92$ and neutron numbers of $0 \le N \le 146$. Electrons and nucleons in atoms both have quantized energy levels, however, the energy level spacing in nuclei is many times larger than that in atoms. While the *electrostatic* attraction force between the protons in the nucleus and electrons holds electrons in

their atomic orbits, a strong nuclear force holds the nucleus. The nuclear force is known as the *strong force* and, unlike the electrostatic force, is charge independent. The theories and models proposed for the nuclear force are beyond the scope of this dissertation. Interested readers are encouraged to review Chapter 4 of [3] and Chapter 16 of [2].

There are currently more than 3000 known nuclides out of which only 266 have a stable ground state. The rest are known as radioactive nuclides which eventually undergo radioactive *decay*. While most stable light nuclei (roughly $A < 40$) have an equal number of protons and neutrons, heavier stable nuclei have an excessive number of neutrons to supply the additional binding energy needed for the nucleus stability. A plot of neutron number N versus proton number Z for all known nuclides is shown in Figure 2.1. The gray-shaded area represents the known radioactive nuclides with a half-life (explained later) of greater than 1 millisecond and the blue dots represent the 266 known stable nuclides [1, 3].

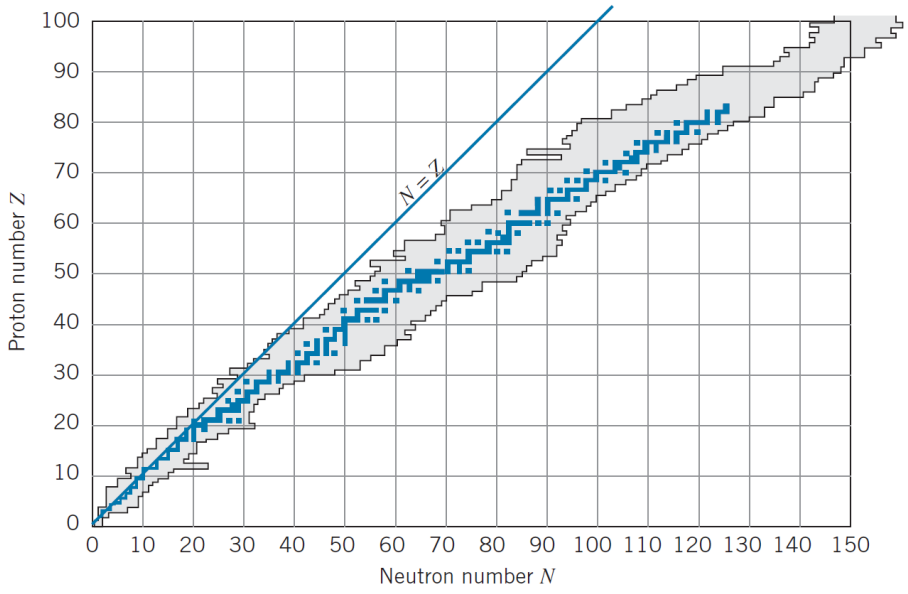

Figure 2.1. A graph of known stable (blue) and radioactive (gray) nuclei (Figure 12.11 of [1]. Image reproduced with permission of the rights holder, John Wiley & Sons, Inc.).

Usually, for each atomic number *A*, there are one or two stable nuclides and the rest are unstable and radioactive. The radioactive nuclides have excessive energy and decay into other nuclides by emitting radiation to reduce their energy level. Since the SI unit of energy (*Joule*) is a large unit for atomic and nuclear physics, energy is often expressed using a smaller unit called the *electron volt* (eV) such that

$$1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J}\ . \tag{2.1}$$

Various radioisotopes have different rates of decay, however, generally the radioactive disintegration follows Poisson statistics. The *half-life* ($T_{1/2}$) of a radioisotope is defined as the time by which the number of radioactive nuclei has decreased to half its initial value. Some commonly used radioisotopes in PET are carbon-11 (^{11}C) with a half-life of 20.3 minutes, fluorine-18 (^{18}F) with a half-life of about 109.7 minutes and oxygen-15 (^{15}O) with a half-life of 122.2 seconds [4]. The decay rate of a radioisotope source, $\left.\frac{dN}{dt}\right|_{decay}$, defines its *activity* and is given by the fundamental law of radioactive decay

$$\left.\frac{dN}{dt}\right|_{decay} = -\lambda N \tag{2.2}$$

in which N is the number of radioactive nuclei and λ is the *decay constant* defined as the probability per unit time of the decay of any given nucleus. The relationship between the decay constant and half-life of a radioisotope is given by

$$T_{1/2} = \frac{\ln(2)}{\lambda}. \tag{2.3}$$

Integrating equation (2.2) over a time t results in

$$N(t) = N_0 e^{-\lambda t} \tag{2.4}$$

which gives the number of remaining radionuclides N after time t from an initial number of radionuclides N_0 for a radioisotope with a decay constant of λ. The SI unit for radioactivity is *becquerel* (Bq) such that 1 Bq is defined as 1 decay/s. Since Bq is a small unit, the activity of a radioisotope sample is often reported in kilo-becquerel (kBq), mega-becquerel (MBq) or giga-becquerel (GBq). Each radionuclide may decay by one or multiple modes of radioactive decay and emission type. Some of the relevant decay modes are briefly explained in the following [4]:

Beta (β) *decay* occurs when a neutron in the nucleus changes to a proton or vice versa. A neutron, in radionuclides with an excessive number of neutrons, may decay in the nucleus to a proton, an electron (β^-)[1] and a particle called the *antineutrino* ($\bar{\nu}$). Antineutrino, the antiparticle of a *neutrino* (ν) ("little neutral one" in Italian), is a particle with a very small mass and no electrical charge[2]. Therefore, a β^- decay of a radioisotope X results in

$$^{A}_{Z}X_{N} \longrightarrow {}^{A}_{Z+1}X'_{N-1} + \beta^- + \bar{\nu} \tag{2.5}$$

where X' is a chemically different isotope which may or may not be radioactive. Another form of β decay occurs in radionuclides with an excessive number of protons, such that a proton decays into a neutron (n), a positive electron (*positron* shown by β^+) and a neutrino (ν). A positron is the antiparticle of an electron. A radioisotope X undergoing a β^+ decay results in

[1] Electrons emitted from the nucleus by a nuclear decay (and not from an electron shell) are indicated by β.
[2] While the existence of neutrinos was first proposed in 1930 to maintain the conservation laws in a beta decay, it was only in 1956 that an experiment proved the existence of neutrinos in a nuclear reactor. Details of this experiment can be found under the section 'Observation of the Neutrino' in Chapter 17 of [2]. While it was believed that the neutrino is a massless particle, it is now believed that the neutrino's mass is not exactly zero and its mass is possibly of order 10^{-5} times the electron mass [5].

$$ {}^{A}_{Z}X_{N} \longrightarrow {}^{A}_{Z-1}X'_{N+1} + \beta^{+} + \nu \ . \tag{2.6} $$

Positron emitting radioisotopes often incorporate a competing nuclear decay process called the *electron capture* (EC). In electron capture, a proton in the nucleus captures a shell electron from inner orbital electrons of the atom and creates the following process

$$ {}^{A}_{Z}X_{N} + \mathrm{e}^{-} \longrightarrow {}^{A}_{Z-1}X'_{N+1} + \nu \ . \tag{2.7} $$

Alpha decay is another type of radioactive decay in which a nucleus disintegrates into a lighter nucleus and an alpha particle (α) (the nucleus of ^{4}He) such that

$$ {}^{A}_{Z}X_{N} \longrightarrow {}^{A-4}_{Z-2}X'_{N-2} + {}^{4}_{2}\alpha_{2} \ . \tag{2.8} $$

The nucleus of some radioisotopes may remain in an excited state following an α, β^{-} or β^{+} decay. These radioisotopes are called *metastable* and are particularly useful in nuclear medicine. Metastable radionuclides will reach their ground state after a deexcitation of the excessive energy by emitting one or more *photons* (electromagnetic radiation). These photons are known as nuclear *gamma rays* (γ rays) and are often more energetic than the *characteristic X-rays*[1]. An alternative to gamma-ray emission is *internal conversion*. In an internal conversion, the excessive energy of an excited nuclei is internally absorbed by an orbital electron (usually from an inner shell) of the atom which causes the electron to be ejected. The so-called *conversion electron* can only be ejected if the excitation energy is sufficient to overcome the binding energy of the electron. Technetium-99m (^{99m}Tc), with a half-life of 6 hours, is an example of a metastable radionuclide that is by far the most commonly used radionuclide in nuclear medicine imaging studies [4].

[1] Characteristic X-rays are photons emitted when an electron moves from an outer atomic shell to an inner shell. The energy of such characteristic X-rays is equal to the difference in the binding energy of the electron between the two shells.

A *decay scheme* is often used to present the possible radioactive decay modes of a radioisotope. Figure 2.2 shows the decay scheme of two commonly used radioisotopes in nuclear medicine imaging, ^{99m}Tc and ^{18}F [6].

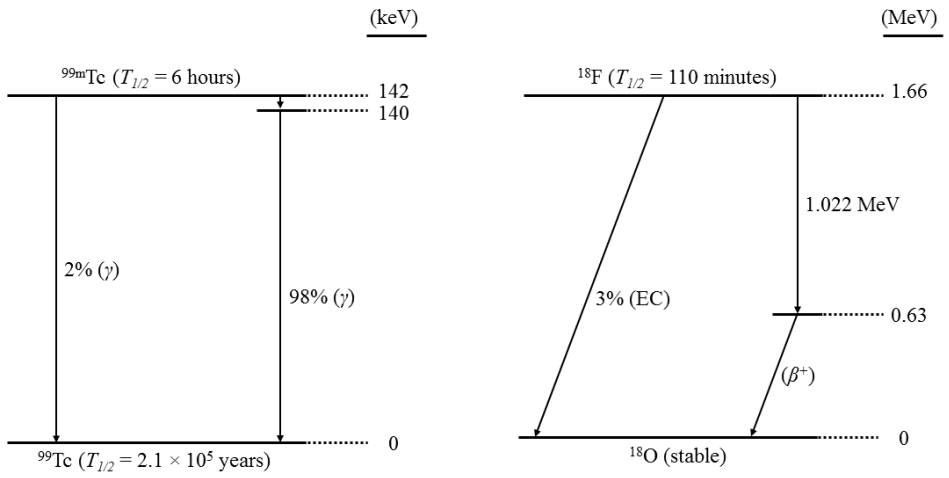

Figure 2.2. The decay scheme of ^{99m}Tc (left) and ^{18}F (right). Different energy levels indicated in keV and MeV show the amount of excessive nuclei energy and the percentages reflect the abundance of each decay mode.

2.2 Interaction of Radiation with Matter

Charged particles (such as α and β), uncharged particles (such as neutrons), and electromagnetic radiation (photons) transfer their energy to matter through different mechanisms that will be discussed in this section.

2.2.1 **Interaction of Charged Particles**

Charged particles lose their energy by *collisions* with atoms and molecules. Although the term *collision* is used, it should be clarified that these interactions are in fact as a result of electrical attraction or repulsion forces and not physical collisions. If an orbital electron is ejected due to the collision, an *ionization* interaction occurs. In an ionization interaction, the charged particle loses some of its energy and a *secondary electron* is generated with a kinetic energy equal to the gained energy from the collision minus the binding energy of the electron. The secondary electron, called

a *delta ray* (δ), may be sufficiently energetic to cause secondary ionization. A charged particle may also cause an *excitation* during which an orbital electron is raised to the next energy level (*excited state*). The last important type of interaction occurs when a travelling charged particle interacts with the nucleus of an atom. The charged particle often deflects and decelerates by the strong electrical force exerted on it by the nucleus. The particle will then lose some of its energy due to the deceleration. The lost energy is released as a photon called *bremsstrahlung* ("braking radiation" in German) [7]. While a high energy α particle (e.g. 10 MeV) may travel up to a few hundred μm in soft tissue, a high energy electron may travel up to a few cm in soft tissue [4].

The last relevant interaction type is the positron-electron *annihilation*. Positrons generated from a β^+ decay, generally travel up to a few mm in a medium, such as tissue, and then combine with a normal electron to form a very unstable particle known as *positronium* which can live up to a few nanoseconds. The positronium then undergoes an annihilation that generates two [almost] oppositely directed 511 keV photons. Because the two particles forming the positronium are moving, the annihilation photons may be emitted in directions slightly off from exactly 180º. The generated back-to-back photons are known as annihilation photons and are the basis of PET imaging.

2.2.2 **Interaction of Uncharged Particles**

Unlike charged particles, uncharged particles and photons cannot interact with matter using the Coulomb force[1] and cause ionization directly. However, through various mechanisms, indirect ionization can still occur. Neutrons interact with the nucleus of the target material. Often interactions of neutrons with matter are studied for slow (energy $\lesssim$ 0.5 eV) and fast neutrons

[1] The Coulomb force is the attraction or repulsion force between the charged particles and is defined according to Coulomb's law.

separately. Such interactions will not be discussed here as they are not relevant to the work done in this dissertation.

Photons are electromagnetic radiation that carry energy proportional to their frequency. The relationship between a photon's energy (E), its radiation frequency (ν) and wavelength (λ) is given by

$$E = h\nu = \frac{h\mathrm{c}}{\lambda} \tag{2.9}$$

where c is the speed of light in vacuum and h is *Planck's constant* given by

$$h = 6.626\times10^{-34}\ \text{J.s}\ = 4.135\times10^{-15}\ \text{eV.s}\ . \tag{2.10}$$

Photons interact with matter through various means. The most important ones that are relevant to nuclear medicine are the *photoelectric effect, Compton scattering, pair production* and *coherent (Rayleigh) scattering*.

In a photoelectric interaction, the kinetic energy of an incident photon is completely absorbed by an atom and the photon vanishes. The atom ejects an orbital electron (often from a tightly bound atomic shell such as the K shell) that is known as the *photoelectron*. In a photoelectric absorption for a photon with an initial energy of $E_\gamma = h\nu$, the photoelectron will receive an energy E_{e^-} given by

$$E_{e^-} = E_\gamma - E_b \tag{2.11}$$

where E_b is the binding energy of the photoelectron in its original shell. The ionized atom then absorbs a close by electron, by capturing either an electron from another shell or a free electron, to fill its vacancy. The photoelectric process is more likely for photons of relatively low energy

and materials of high atomic number *Z*. The probability of photoelectric absorption per atom (τ) for a photon with an energy of E_γ and a target material with an atomic number *Z* is proportional to

$$\tau \propto \frac{Z^n}{E_\gamma^{3.5}} \tag{2.12}$$

where the exponent *n* varies between 4 and 5 depending on the photon energy [8]. The photoelectric probability may spike for photon energies equal to the binding energy of the electrons in the target atom at different atomic shells such as the K-shell (*K-edge*).

Compton scattering is the most probable interaction type for typical photon energies emitted by commonly used radioisotope sources in nuclear medicine. In Compton scattering, often an incident photon interacts with an unbound electron which is initially at rest and transfers a part of its energy to the electron. As a result, the scattered photon deflects at an angle θ and the initially steady electron, now called the *recoil electron*, starts moving at angle ϕ with respect to the initial incident photon's direction as shown in Figure 2.3.

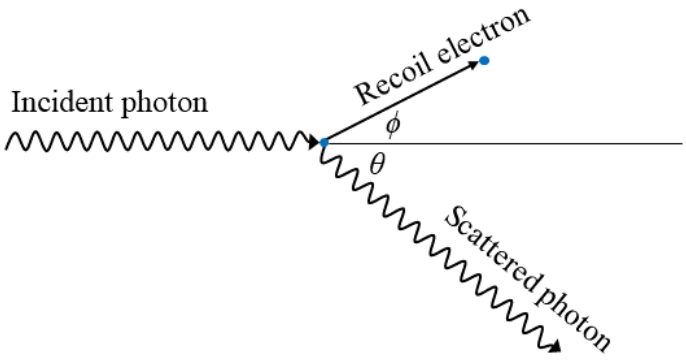

Figure 2.3. A schematic drawing of Compton scattering.

The frequency of the incident photon reduces from ν to ν' and, therefore, it experiences a partial energy loss. The relation between the initial energy of the photon $h\nu$ and the scattered photon energy $h\nu'$ is given by

$$h\nu' = \frac{h\nu}{1+\frac{h\nu}{m_0 c^2}(1-\cos\theta)} \tag{2.13}$$

where $m_0 c^2$ is the rest-mass energy of the electron (511 keV). The transferred energy to the recoil electron can be as low as nearly 0 eV. The maximum energy transferred to the recoil electron, on the other hand, corresponds to 180-degree *backscattering* events. Using equation (2.13) for θ = 180°, the backscattered photon energy ($h\nu'_{\min}$) is given by

$$h\nu'_{\min} = \frac{h\nu}{1+\frac{2h\nu}{m_0 c^2}} \ . \tag{2.14}$$

The probability of Compton scattering per atom (σ) increases linearly with the atomic number of the target. The angular distribution of scattered photons is predicted by the *Klein-Nishina* formula for the differential scattering cross-section $d\sigma/d\Omega$

$$\frac{d\sigma}{d\Omega} = Zr_0^2\left(\frac{1}{1+\alpha(1-\cos\theta)}\right)^2\left(\frac{1+\cos^2\theta}{2}\right)\left(1+\frac{\alpha^2(1-\cos\theta)^2}{(1+\cos^2\theta)(1+\alpha(1-\cos\theta))}\right) \tag{2.15}$$

where *Z* is the atomic number of the target, r_0 is a constant known as the classical electron radius, θ is the scattering angle of the photon and α is defined as $h\nu/m_0 c^2$ [8]. Roughly speaking, at relatively low energies of about 10 – 100 keV, the scattered photon is more likely to be directed in either the forward or backward direction and less likely to be at right angles (90°). At high energies

of greater than ~500 keV, scattering becomes mainly forward directed. The relative probability of Compton scattering per unit of solid angle against scattering angle θ is plotted in Figure 2.4 for various photon energies.

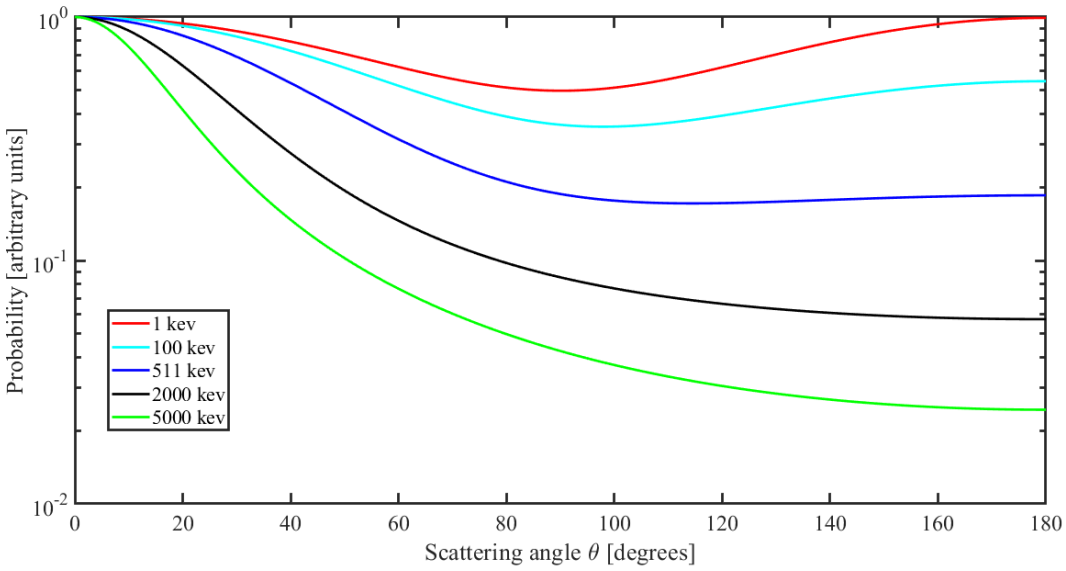

Figure 2.4. Angular probability of Compton scattering for photons of various energies.

When the photon energy exceeds twice the rest-mass energy of an electron (i.e. 1.02 MeV), the photon may disappear and get replaced by an electron-positron pair in the Coulomb field of a nucleus. The probability of this process (κ), known as *pair production*, cannot be described in a simple expression. Nevertheless, pair production is more likely for energetic photons. The rest of the 1.02 MeV energy of the photon will be split into the kinetic energy of the electron and the positron. As explained in Section 2.2.1, the positron will eventually annihilate after depositing all or most of its kinetic energy in the medium. Figure 2.5 shows an illustration of the photoelectric absorption, Compton scattering and pair production.

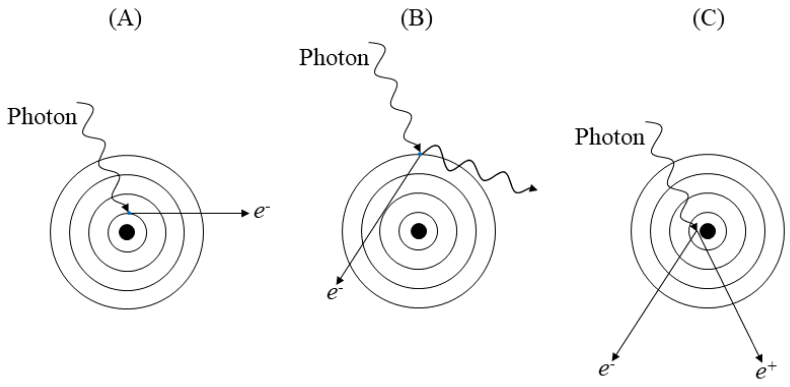

Figure 2.5. A drawing of the photoelectric absorption (A), Compton scattering (B) and pair production (C). The nucleus of the atom is shown as a black-filled circle at the centre with electron orbitals around it.

The last relevant photon interaction type is coherent (Rayleigh) scattering. Unlike Compton scattering, in coherent scattering, the photon interacts coherently with all the electrons of the target atom. However, the photon retains its energy after the scattering event and does not ionize or excite the atom. Therefore, the photon energy remains intact and only a change in the direction of the scattered photon is observed. Coherent scattering has a high probability (σ_τ) for low energy photons (< 50 keV) and absorbers with a high atomic number. While coherent scattering has an important contribution in X-ray imaging modalities, it is of little importance in nuclear medicine [4].

The mentioned photon interaction probabilities can be divided by the density of the medium (ρ) and expressed as *mass attenuation coefficients* to remain independent of the density of the absorber. Figure 2.6 presents the photon interaction probability for different types of interaction with lutetium oxyorthosilicate (LSO) (a material used in radiation detectors) as a function of photon energy.

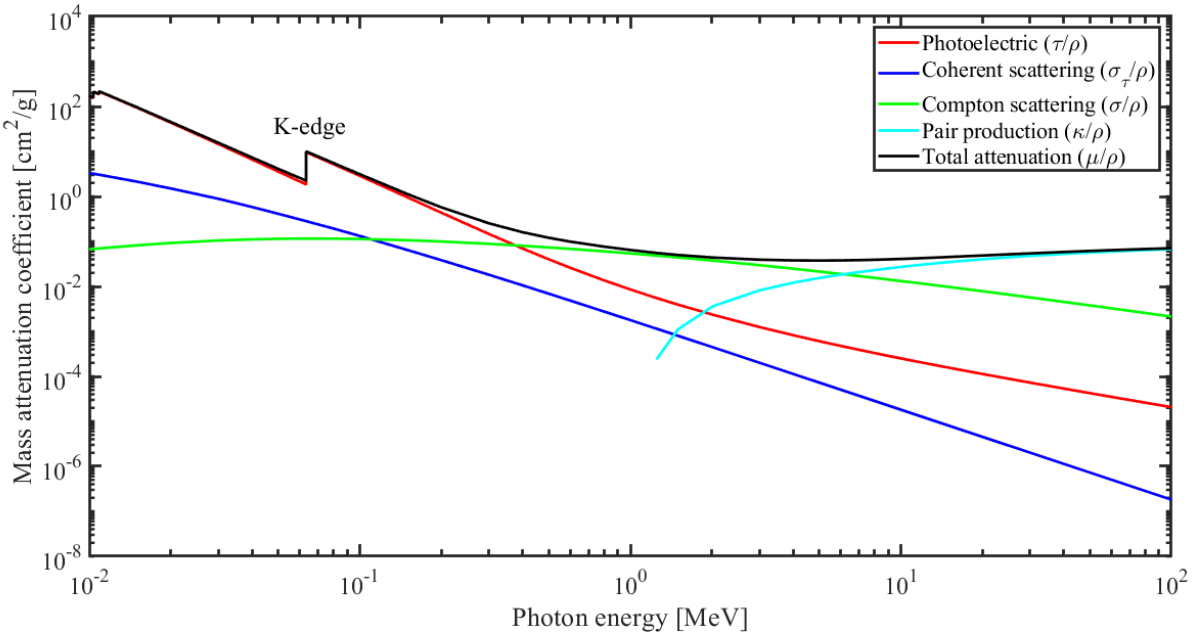

Figure 2.6. Mass attenuation coefficient for various photon interaction types versus photon energy in LSO. Data obtained from [9].

The spike labeled as the K-edge in Figure 2.6 corresponds to the photoelectric events for the K-shell electrons and the black curve shows the sum of all mass attenuation coefficients components, known as the *mass attenuation coefficient* (μ/ρ). The dominancy of each interaction with different materials for a range of photon energies is shown in Figure 2.7.

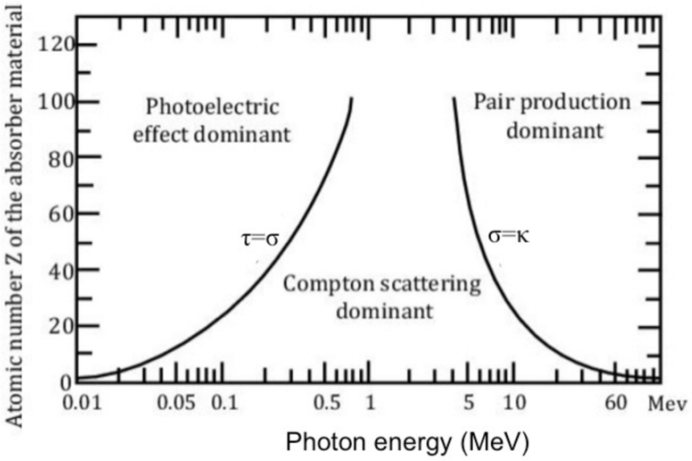

Figure 2.7. The most probable interaction for absorbers with different atomic numbers versus photon energy (Figure 2.20 of [8]. Image reproduced with permission of the rights holder, John Wiley & Sons, Inc.).

A narrow beam of photons gets *attenuated* by an absorber either by absorption or by scattering. The amount of attenuation in the absorber can be quantified if the photon energy and the attenuation coefficient of the absorber are known. The *linear attenuation coefficient* (LAC) μ for an absorber is given by

$$\mu = \tau \text{ (photoelectric)} + \sigma \text{ (Compton)} + \kappa \text{ (pair production)} \ . \tag{2.16}$$

The intensity of a narrow beam of monoenergetic photons with a known energy through an absorber with a thickness of *t* can be calculated by

$$\frac{I}{I_0} = e^{-\mu t} \tag{2.17}$$

where I_0 and I represent the initial and final intensity of the photon beam, respectively. Photon attenuation in a medium can also be characterized by its *mean free path* λ. The mean free path is defined as the average distance a photon travels in the absorber before an interaction takes place and can be calculated by [8]

$$\lambda = \frac{\int_0^\infty x e^{-\mu x} dx}{\int_0^\infty e^{-\mu x} dx} = \frac{1}{\mu} \ . \tag{2.18}$$

2.3 Overview of Nuclear Medicine

The first attempts to map the spatial distribution of an injected radioactive material in the human body dates back to the 1940s. The first 2-dimensional (2D) images were built through mapping the distribution of an injected radioactive material using a gas-filled radiation detector. By the mid-1950's, several English hospitals used templates, fitted over the neck, to mark the radioactivity distribution on a grid of 1 cm squares to create an "image" of the distribution. In

1951, the first attempt to map the activity distribution in the brain through the use of back-to-back annihilation photons was published. The concept was later expanded and the first area sensitive positron imaging device (the first PET scanner known as PC-1) was designed in 1968 and tested in 1970. Improvements in the generation of radiopharmaceuticals, and specially introducing a generator for ^{99m}Tc in 1957, accelerated nuclear medicine imaging developments. The idea of creating 3-dimensional (3D) tomographic images was first developed in 1963 and the first prototype of a *single photon emission computed tomography* (SPECT) camera was built in 1977. Although PET was developed before SPECT, PET scanners to this date are less common as they are more expensive and require a medical cyclotron to produce positron-emitting radionuclides, such as ^{18}F. More details on the history of nuclear medicine and the photographs of the first scanners and their obtained images can be found in [10-14].

Nowadays, nuclear medicine imaging is a widely used diagnostic technique to collect functional information from the human body. A short-lived radioisotope, emitting gamma rays often in the range of about 100 to 500 keV, is either inhaled, swallowed or injected into the body. The radioisotope may be *labeled* to chemical compounds, such as glucose, to enable targeted imaging. The distribution of the radioactive compound (the *radiotracer*) in the body is imaged using radiation detectors around the patient to provide an image of the distribution of the radiotracer in the body. Modern nuclear medicine imaging includes *planar* and *tomographic imaging*. In planar imaging, a 2D superimposed image of the distribution of the radiotracer in the body is acquired from a certain point-of-view. In tomographic imaging however, a 3D image of the distribution of the radioactivity is created through collecting data from various angles and using a *computed tomographic* (CT) *image reconstruction* technique. The following is a brief overview

of the basis of the two common nuclear medicine tomographic imaging techniques, SPECT and PET.

2.3.1 **SPECT**

In SPECT, the patient is injected with a gamma-emitting radiotracer, such as ^{99m}Tc compounds. A *gamma camera* is then used to acquire 2D *projection* images from several equally spaced angular intervals around the patient. A gamma camera is the most widely used imaging device in nuclear medicine that has three essential parts. The first component is the *collimator* that is made of small holes surrounded by a high atomic number material (e.g., lead) to absorb non-parallel gamma rays. The second part is a layer of a so-called *scintillator material*. Scintillators are usually high atomic number materials that convert the energy deposited in them by a gamma-ray to *optical photons* or visible light. Scintillator materials are explained in Section 2.5.1. The last part of a gamma camera is the *photodetectors* or conventionally *photomultiplier tubes* (PM tubes) that convert the optical photons to an electronic signal (there may also be a layer of a light guide between the scintillator and PM tube that optical photons use to reach to PM tubes). Further details on conventional and modern photodetectors are explained in Section 2.5.2.

The projections are then used in a tomographic image *reconstruction technique* to reconstruct an image of the distribution of the radiotracer in the body. The basis of different tomographic reconstruction techniques is described in detail in Chapter 3. A schematic illustration of a SPECT camera is shown in Figure 2.8.

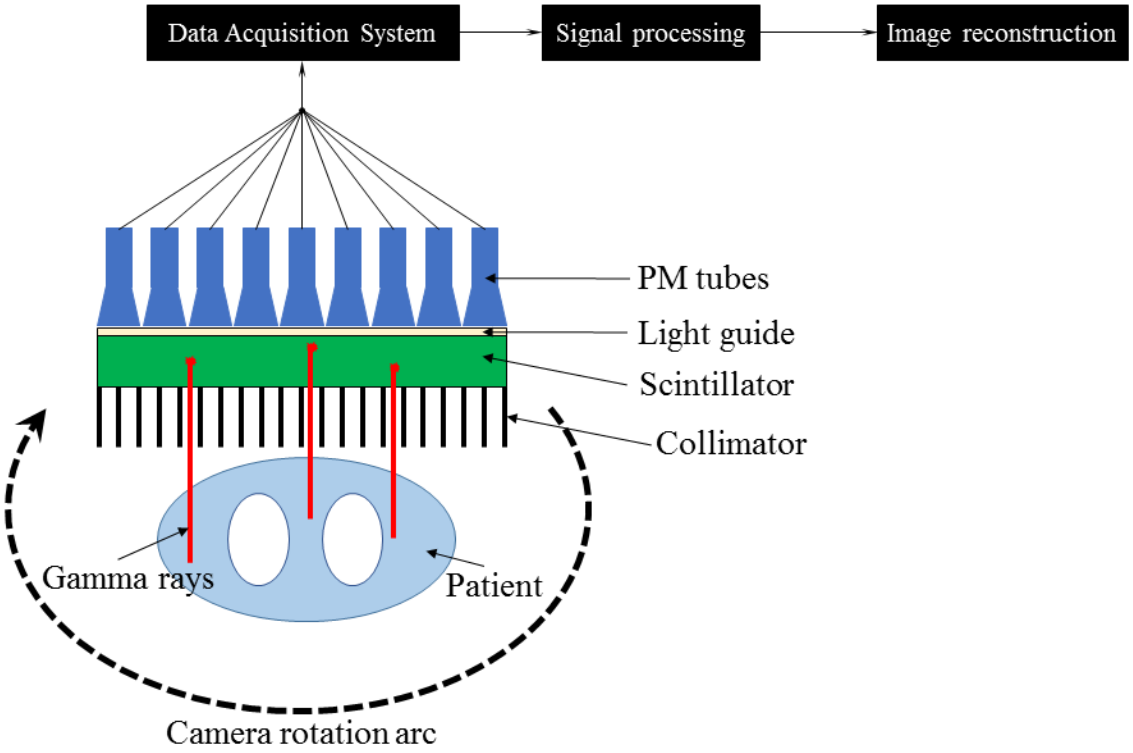

Figure 2.8. Basic components of a SPECT system.

2.3.2 **PET**

PET is a widely used functional tomographic imaging modality. In a PET scan, short-lived positron-emitting radionuclides, such as ^{18}F, are used to acquire images. Similar to SPECT, the radionuclide may be bound to biological molecules to create a radiotracer. As the radionuclides decay in the body, positrons are being emitted. Each positron absorbs a nearby electron and undergoes a positron-electron annihilation and, subsequently, an emission of a pair of back-to-back 511 keV photons. The photons are then detected by a series of photon detectors around the patient. The detectors used in PET are similar to the detectors used in a gamma camera. However, unlike in SPECT, detectors are usually arranged as a ring of detectors around the patient to facilitate the detection of back-to-back annihilation photons. The most common form of PET detectors comprises pixelated elements of an *inorganic scintillator* arranged into detector blocks

as illustrated in Figure 2.9. More details on the detectors used in modern PET scanners can be found in Section 2.5.

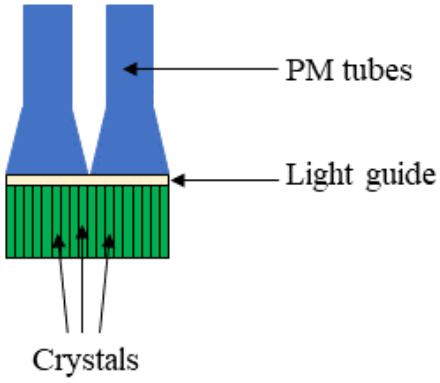

Figure 2.9. Basic components of a PET detector block.

The ring of detectors around the patient detects annihilation photons. A *single* refers to the signal generated as a result of photon detection in a PET detector. When two singles arrive within a predefined timing window, a *coincidence* is generated. The mentioned timing window is known as the *coincidence timing window* (CTW). The line connecting the two detection points of the annihilation photons (two singles) is called a *line-of-response* (LOR). Detected coincidences in PET may be categorized into 3 major types of coincidences as shown in Figure 2.10. When the back-to-back annihilation photons are detected without undergoing a scattering interaction in the object, a *true coincidence* is detected. If one or both of the annihilation photons undergo Compton scattering, a *scatter coincidence* is generated. Finally, if two annihilation photons, originating from different annihilation events, are detected at the same time, a *random coincidence* is generated. While in a true coincidence the annihilation position lies along the detected LOR, in scatter and random coincidences, a false LOR is detected that may be distant from the annihilation position. Such LORs contribute to noise in the image and degrade the quality of the reconstructed images.

The spatial and temporal information of the detected LORs provide a basis for reconstructing an image of the radiotracer distribution within the body.

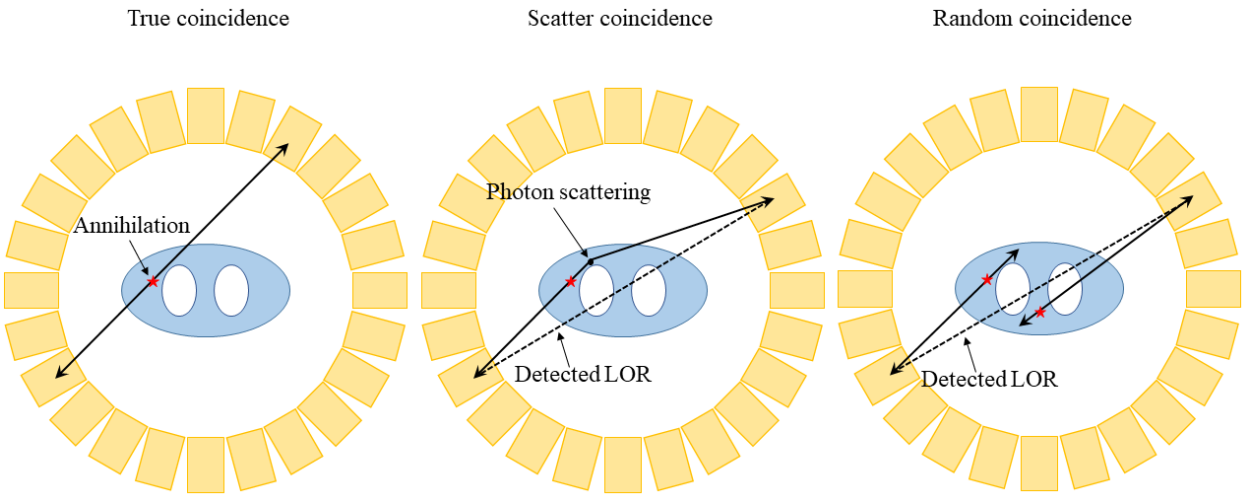

Figure 2.10. Different coincidence types in PET.

PET imaging has played a significant role in oncology, cardiovascular studies and neurology. The most commonly performed PET scan in the world is done with ^{18}F attached to fluorodeoxyglucose (FDG) to study glucose metabolism in the body. In neurology, PET has been used to diagnose neurodegenerative diseases, such as Alzheimer's disease or Parkinson's disease, epilepsy, neurodevelopmental disorders and psychiatric disorders [15]. The resurgent interest in different neurological studies with PET has been so significant that many groups have started investigating brain-dedicated PET scanners [16-21]. As this dissertation focuses on brain-dedicated PET scanners, the advantage of a brain-dedicated scanner and a review on the existing scanners are discussed in Section 2.6.

2.4 Multimodality Imaging

Images obtained from a nuclear medicine modality provide information on the functionality of each organ. Nonetheless, only minimal anatomical information is provided from

the nuclear medicine images. Therefore, often SPECT or PET images are acquired along with an anatomical imaging modality, such as CT or MRI. The obtained coregistered anatomical images from these so-called *hybrid scanners* not only provide clinically valuable complementary information, but also allow for a compensation of image degrading artifacts that occur during the SPECT or PET acquisition (these image degrading artifacts are reviewed in Section 2.8). In this section, common multimodality approaches for PET and an overview of their clinical applications are discussed.

2.4.1 **PET/CT**

At present, most PET scanners are manufactured and sold as PET/CT scanners, such as [22-27]. The PET and CT scanners are merged into a single gantry and configured in a *tandem* arrangement with an axial distance of about 60 to 120 cm between the axial centre of the two components [4]. Hybrid scanners are comprised of two separate modalities built into the same gantry. Therefore, the CT and PET data are acquired sequentially while the patient translates between the modalities. The known relative positions of these modalities enable an accurate registration of the images. The co-registered PET and CT images not only have several clinical diagnostic values, but also allow for a quantitative correction of the PET images (see Section 2.8). An example of a modern hybrid PET/CT scanner with reconstructed images of a patient with papillary carcinoma of the thyroid shown in Figure 2.11.

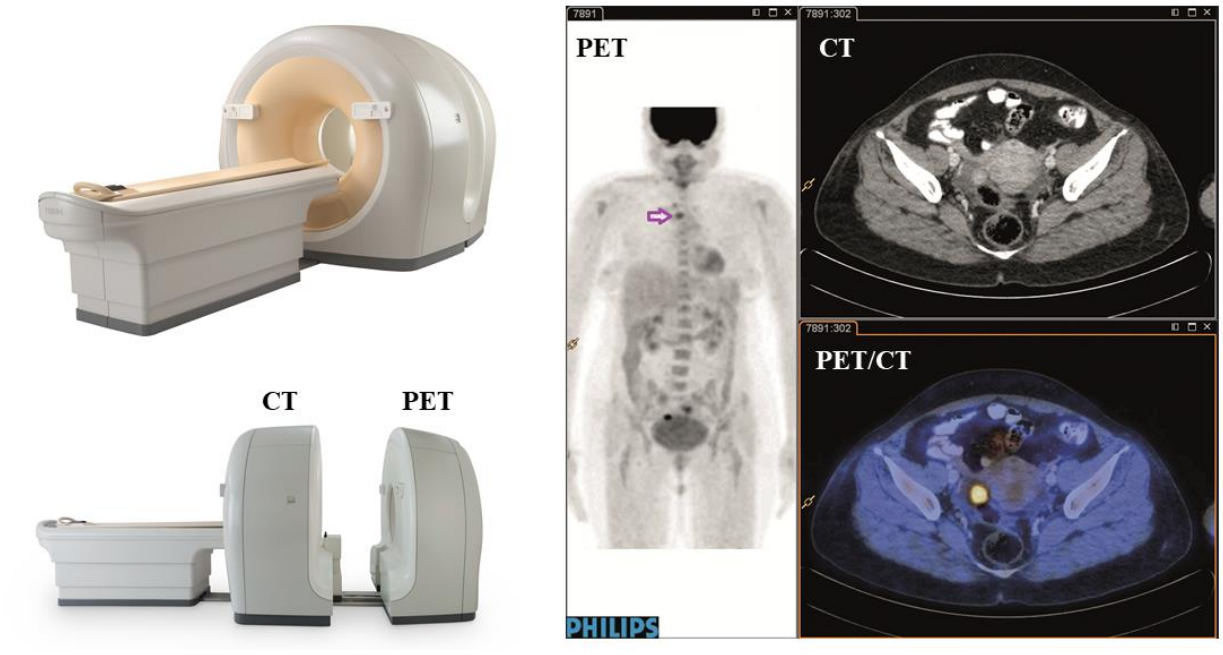

Figure 2.11. The Philips Ingenuity PET/CT scanner with sample patient images (courtesy of Philips Healthcare, Andover, MA [28]).

2.4.2 **PET/MR**

The integration of PET and MRI has been of significant interest and an active research area within the past few decades due to several reasons. First, MR images have excellent soft tissue contrast and in many cases are preferred over anatomical images obtained by CT, particularly for brain imaging. Also, an MRI scanner can provide images with complementary functional and physiological information. Another significant advantage of a PET/MR scanner is that it potentially allows for real simultaneous acquisition of PET and MR as a full integration of the scanners is feasible [4]. Several researches have proposed different motion correction approaches based on the simultaneously acquired MR information [29-31]. Also, as shown by several studies, a strong magnetic field reduces the positron range [32, 33]. The reduction in positron range decreases image blurring in PET images due to the positron range and improves image in-plane (*transaxial*) resolution [34, 35].

The first prototype hybrid PET/MR scanner was built in 1997 [36] and the first commercially manufactured PET/MR scanner was introduced in 2005 [37]. A non-trivial challenge in building a PET/MR scanner was to protect the conventional PM tubes from the strong magnetic field generated by the MR scanner. As explained in Section 2.5.2.1, PM tubes accelerate electrons by applying a high voltage. In the strong magnetic field of an MR scanner, however, the Lorentz force is applied on the moving electrons which interferes with the efficiency of the PM tube. Therefore, all modern integrated PET/MR scanners are manufactured with *solid-state photodetectors* that are explained in Section 2.5.2.2. Currently, there are three variants of PET/MR scanners as depicted in Figure 2.12 that are discussed in more detail in the following sections. For a more comprehensive review of current PET/MR designs, clinical applications, clinical potentials and challenges, references [36, 38-42] are suggested.

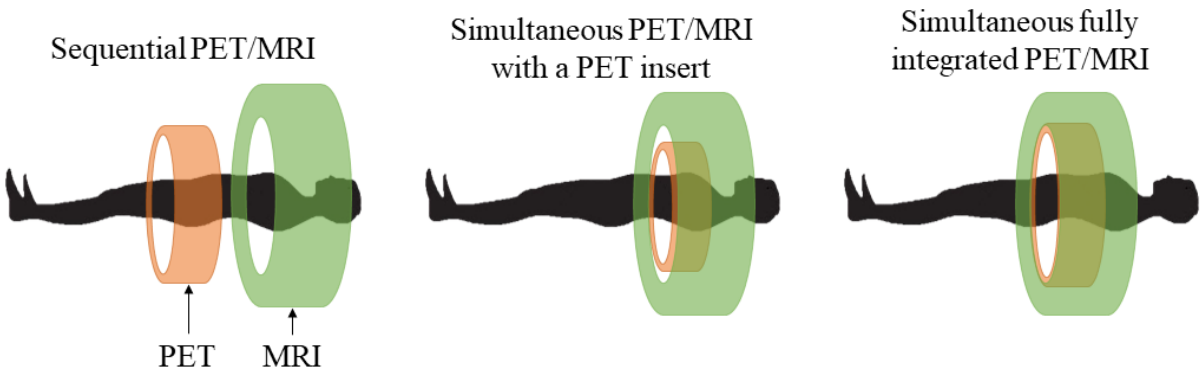

Figure 2.12. Different configurations of PET/MRI.

2.4.3 **Sequential PET/MR**

In a sequential PET/MR scanner, such as [43, 44], co-registered PET and MR images are acquired sequentially, similar to a PET/CT scanner. The sequential acquisition of PET and MR images may extend the overall scan time and is prone to misalignments due to patient motion. The co-registered MR images are used to improve clinical diagnosis and apply quantitative corrections

to the PET images. The Philips Ingenuity time-of-flight PET/MR system [44] is an example of a sequential PET/MR system that is based on lutetium-yttrium oxyorthosilicate (LYSO) scintillators and PM tubes.

2.4.4 **Fully Integrated PET/MR**

The aim of a fully integrated PET/MR scanner is to integrate both systems with no reduction in the performance of the PET and MR components. A fully integrated PET/MR scanner reduces the overall acquisition time of an equivalent sequential PET/MR scanner and enables novel simultaneous imaging protocols. Development of solid-state photodetectors that are compact and MR-compatible led to the development of such scanners. The Siemens mMR is an example of a fully integrated whole-body PET/MR scanner [25]. The Siemens mMR is comprised a 3 T (tesla) superconductor magnet with a length of 163 cm and a bore diameter of 60 cm and a PET scanner with LSO detectors coupled to *avalanche photodiodes* (APDs) (see Section 2.5.2.2) as the photomultiplier. Sample co-registered PET/MR images of a whole-body ^{18}F-FDG scan of a patient with the Siemens mMR PET/MR and the Biograph PET/CT (for comparison) are shown in Figure 2.13-(A) and Figure 2.13-(B), respectively. Figure 2.13-(C) and Figure 2.13-(D) show coronal T2-weighted MR images of a healthy individual obtained with the Siemens mMR (C) and Siemens Verio MR scanner (D).

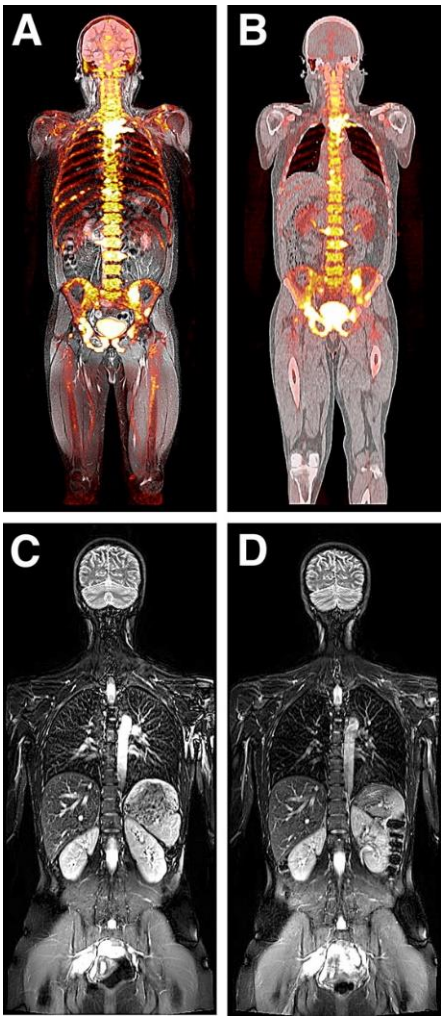

Figure 2.13. Sample PET/MR (A), PET/CT (B), MR with the Siemens mMR (C) and MR with the Siemens Verio (D) images (figure originally published in [25]. © SNMMI).

2.4.5 **MR-compatible PET Inserts**

A PET insert is built such that it can be inserted into, or removed from an MR scanner. A PET insert must be portable and fully MR-compatible. When inserted into an MR scanner, a PET insert provides virtually all the benefits of a fully integrated PET/MR scanner. A significant motivation behind the idea of building PET inserts is the high cost of a fully integrated PET/MR scanner. The Siemens brainPET[1] insert [17] is an experimental brain-dedicated PET insert that has

[1] In Chapter 7, we also use the term BrainPET to refer to a subset of scanners studied in this dissertation. The BrainPET scanners used in the current work are different than the Siemens brainPET insert.

been manufactured to fit into the bore a Siemens 3 T MRI system. The Siemens brainPET insert had APD-based detectors coupled with LSO crystals. Recently, with the development of solid-state photodetectors, there has been a growing interest in the development of PET inserts, specially for preclinical and brain-dedicated MR scanners [17, 45-49].

2.4.6 **Clinical Applications**

Hybrid modalities, such as PET/CT and PET/MR, have been a valuable diagnostic tool. While PET/CT has evolved to be one of the most powerful tools in various fields and particularly oncology [50], PET/MR has shown the potential to broaden the horizons in the field of molecular imaging due to the reasons discussed in Section 2.4.2. An integrated PET/MR system provides real-time co-registered anato-functional images of the human body with complementary information. Images obtained from MRI can reveal information on morphological features, water motion in tissue (through diffusion-weighted imaging) [51], vascular anatomy [52], oxygen consumption [53] and several more biological properties. On the other hand, PET images can demonstrate other complementary features such as the blood flow [54], metabolism [55], receptors [56] and transmitters [57].

In brain PET/MR studies, the hippocampus and thalamus have been the most attractive candidates due to their lack of substructural contrast and resolution in MRI [58]. The hippocampus is involved in human behaviors such as memory and spatial navigation and the thalamus, consisting of many subnuclei, has several diverse functions [59, 60]. It was with a high resolution brain PET/MR system that, for the first time, functional activity in the substructures of the hippocampus and thalamus was modeled [61, 62]. Simultaneous PET/MR has also improved measuring glucose metabolism and serotonin transporter activity in the raphe nuclei which contain serotonergic neurons. The raphe nuclei that remained identifiable with conventional neuroimaging

approaches, mediate several functions, including sleep, appetite and mood [63]. Other applications, potential benefits and challenges of PET/MR in neurology are thoroughly discussed in [39, 41, 46, 58, 64].

2.5 PET Detectors

The most common form of detectors used in PET scanners are based on *inorganic scintillators*. In Section 2.5.1, a review of the basis of photon detection using *organic* and inorganic scintillators is reported, followed by an overview of conventional and modern photodetectors in Section 2.5.2. Lastly, in Section 2.5.3, methods on *depth-of-interaction* (DOI) encoding detectors (conveniently called *DOI detectors*) are reviewed.

2.5.1 **Scintillation Detectors**

The basis of a *scintillation detector* is to detect ionizing radiation by the scintillation light produced in certain materials known as *scintillators*. The process of emitting visible light from a substance following its excitation is known as *fluorescence*. An ideal scintillation detector converts the kinetic energy of the ionizing radiation into detectable light with high efficiency (*scintillation efficiency*) and linearly. Scintillation materials can be categorized into organic and inorganic scintillators that are explained in Section 2.5.1.1. Also, scintillation detectors used in modern PET scanners may use pixelated or non-pixelated scintillators that are reviewed in Section 2.5.1.2.

2.5.1.1 Organic and inorganic scintillators

The most commonly used scintillators in nuclear medicine imaging devices include inorganic crystals. The scintillation process is due to the characteristics of the crystal structure of the scintillators, meaning that they only scintillate in crystalline form. While some inorganic crystals are scintillators in their pure state, most are *impurely activated*, meaning that the crystals

have small amounts of impurity atoms of other elements. Impurity atoms cause disturbance in the normal structure of the crystal matrix. For instance, LYSO(Ce) indicates a cerium-doped LYSO scintillator or NaI(Tl) indicates a thallium-doped sodium iodide.

An ideal scintillator for a PET scanner should have certain properties. It should have a high density and a high effective atomic number to be a good absorber for 511 keV photons. A good scintillator should also yield a high number of optical photons and be transparent to its own scintillation emission to prevent self-absorption of the generated optical photons. It is also preferred that the scintillator has a short decay time of the induced luminescence and has a desirable *index of refraction.* For instance, the glass entrance window of a PM tube usually has an index of refraction of about 1.5. Therefore, it is most desirable to have a scintillator with a similar index of refraction when using such PM tubes [4, 8]. Properties of some common scintillators used in SPECT and PET scanners are shown in Table 2.1.

Table 2.1. Properties of selected scintillators used in nuclear medicine (data taken from [4, 8, 65]). For LYSO, a mixing fraction of 0.105 YSO/LSO was assumed [66].

	NaI(Tl)	BGO	LSO(Ce)	YSO(Ce)	LuAP(Ce)	LYSO(Ce)	Plastic
Density (g/cm^3)	3.67	7.13	7.40	4.54	8.34	7.13	1.03
Effective atomic number	50	73	66	39	65	65	12
Decay time (ns)	230	300	40	70	18	40	2
Photon yield per keV	38	8	20-30	24	12	20-30	10
Index of refraction	1.85	2.15	1.82	1.80	1.97	1.82	1.58
Peak emission (nm)	415	480	420	420	365	420	Various

NaI(Tl): Thallium-doped Sodium iodide, BGO: Bismuth germanate, LSO(Ce): Cerium-doped lutetium oxyorthosilicate, YSO(Ce): Cerium-doped yttrium oxyorthosilicate, LuAP(Ce): Cerium-doped lutetium aluminum perovskite, LYSO(Ce): Cerium-doped lutetium yttrium oxyorthosilicate.

Most modern systems use lutetium-based scintillators (e.g. LSO or LYSO) due to their excellent combination of high density and stopping power, high luminosity and short decay time. The crystals used in a PET detector are sealed with a reflector material (e.g. Teflon, aluminum or stainless-steel jacket) with a transparent glass or optical window at one end that permit the exit of

the optical photons from the crystals to the photomultiplier. A layer of *photoemissive* substance is coated at the entrance window of a PM tube that is called the *photocathode*. Once optical photons hit the photocathode, electrons eject from the photocathode and move towards the *dynodes* of the PM tube (see Section 2.5.2.1). Figure 2.14 shows the typical assembly of a PET detector block with PM tubes (a single pixel is shown for greater clarity). The scintillator and the photomultiplier should be located in a light-tight surrounding to prevent the interference of outside optical photons.

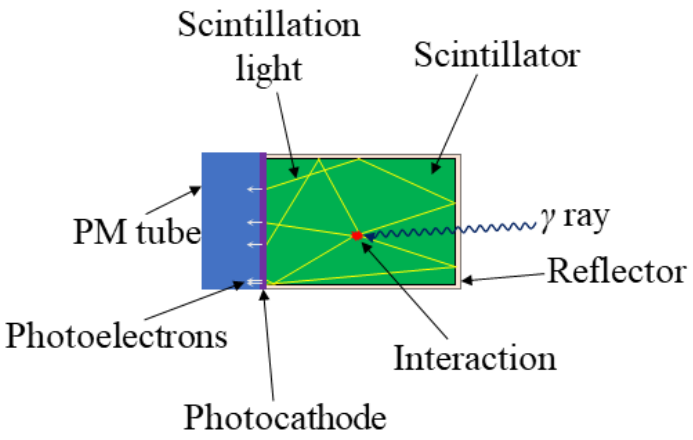

Figure 2.14. Assembly of a typical scintillation detector and an example of optical photon collection.

The fluorescence process in an organic scintillator is an inherent molecular property. Optical photons in an organic scintillator arise from transitions in the energy level structure of a single molecule and therefore can be observed from a given molecular species independent of its physical state. Therefore, these scintillators preserve their scintillation property whether they are in solid, liquid or gaseous form. The most common form of such scintillators that have been evaluated for a PET scanner is the *plastic scintillator* that can be produced by dissolving an organic scintillator in a solvent that can be then polymerized [8]. Because these materials are relatively inexpensive some groups have recently investigated the use of such detectors for PET scanners.

Details of such plastic detectors can be found in [67-71] and are not discussed here. Nonetheless, typical properties for plastic scintillators are stated in Table 2.1 for comparison purposes.

2.5.1.2 Pixelated and non-pixelated detectors

A conventional PET detector consists of an array of inorganic scintillators in each detector block. Therefore, a discrete event positioning scheme should be used such that each event is assigned to a crystal with a known position in the scanner. However, several groups have proposed *monolithic scintillation* detectors (a slab of continuous scintillator) in which the scintillator in each detector block is not pixelated. Because of the large crystal volume and the absence of inter-crystal boundary, a monolithic detector can potentially improve light transport within the detector which enhances the timing resolution of the detector. For example, Borghi *et al*. [72] have shown a timing resolution of 147 ps for a monolithic LYSO crystal with 22 mm thickness. Monolithic detectors also can provide a continuous 3D event positioning within the detector. However, monolithic crystals have shown a large nonuniformity in the spatial resolution across the detector due to the truncation of light transport at the edges of the detector block. Detectors with monolithic crystals are further discussed in Section 2.5.3.2.

2.5.2 **Photodetectors**

Optical photons generated in a scintillator are detected by photodetectors at one (or sometimes two) side(s) of the scintillation material where no optical reflector exists. *Photodetectors* are used to detect these optical photons and generate an electronic pulse. In this section, the basis of conventional PM tubes and modern solid-state photodetectors is discussed.

2.5.2.1 Photomultiplier tubes

Basic components of a typical PM tube are shown in Figure 2.15.

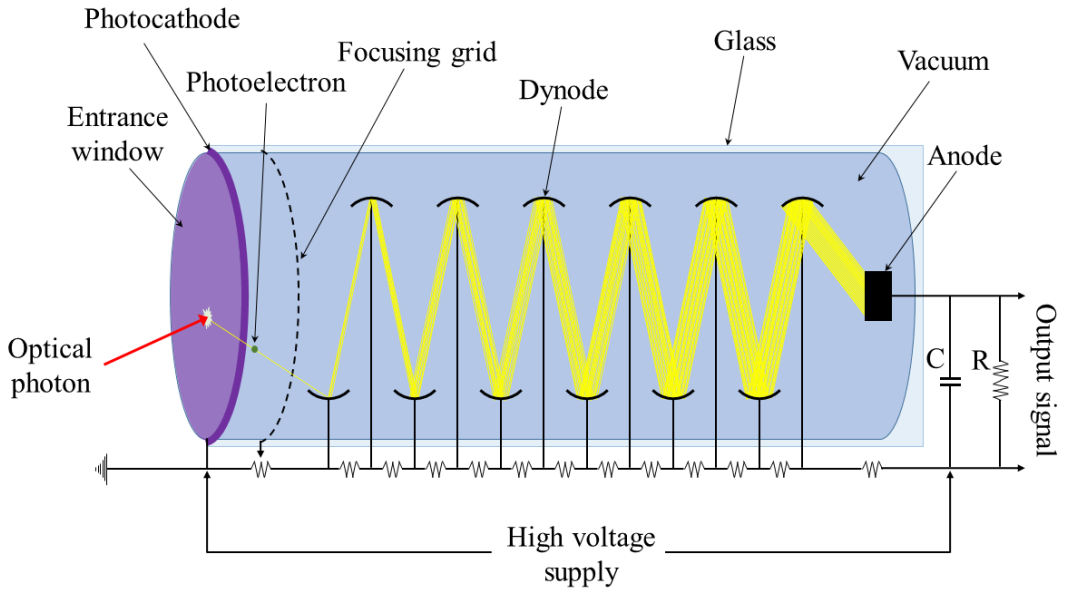

Figure 2.15. Basic components of a PM tube.

The photocathode, typically made of cesium antimony or other bialkali compounds, ejects photoelectrons once struck by photons of visible light. The photoelectrons then travel toward the dynode through a focusing grid. The dynodes are coated with a material that has a relatively high secondary emission characteristics which ejects several secondary electrons when hit by an electron. The multiplication factor per dynode is about 3 to 6. The dynodes are maintained at a potential difference of about 50 to 150 V (volt) such that the multiplication of the electrons repeats several times (about 9 to 12 times such that a gain of 10^7 to 10^8 is achieved). Eventually, a stream of electrons is collected at the anode to produce a relatively high amplitude pulse. To provide the potential difference required for electron multiplication in a PM tube a high voltage (larger than 1000 V) is required. It is also important that the PM tube is sealed and evacuated. Position-sensitive and multichannel PM tubes, that determine the location of incident light, have also been used in several nuclear medicine cameras [4].

2.5.2.2 Solid-state photodetectors

Solid-state photodetectors are an alternative to conventional PM tubes for the detection of optical photons. Solid-state photodetectors or photodiodes are light-sensitive semiconductor detectors that detect the optical (visible) photons. Such photodetectors can be made of different materials, most commonly *silicon* (Si). The most important type of photodiode used in PET detectors is silicon-based photodiodes. The optical photons generated from within the scintillator ionize silicon such that the produced charge is proportional to the number of scintillation light photons. The silicon-based APD is an example of such photodetectors that can produce 10^2 to 10^3 electrons per each electron produced from the optical photons. The magnitude of the pulse generated from an APD is proportional to the energy deposited by the photoelectrons [4, 8].

Avalanche photodiodes can also operate in so-called *Geiger* mode at which a higher bias voltage is applied such that the pulse generated by the APD becomes independent of the energy of the incident photon (~10^6 amplification). In practice, a large number of very small Geiger-mode APDs, each about 20 to 50 μm, are integrated into a single unit. Each one of the APDs is known as a *cell* and the overall magnitude of the generated output pulse is proportional to the number of discharged cells. Since the number of discharged cells depends on the total deposited photon energy, the signal amplitude reflects the deposited energy. Such devices are known as *multi-pixel photon counters* (MPPCs) or *silicon photomultipliers* (SiPMs). Silicon-based photodetectors, such as SiPMs, are the photodetector of choice for modern PET detectors due to several reasons. For instance, these photodetectors can be made in compact sizes with a thickness of as small as a few mm. The small size of such photodetectors allows for a direct coupling of each pixel of a pixelated scintillator to a photomultiplier. Also, unlike conventional PM tubes, APDs and SiPMs are

insensitive to magnetic fields which make them highly suitable for hybrid PET/MR scanners. Furthermore, modern SiPMs have high photon detection efficiency and low timing jitter [15].

2.5.3 **Depth of Interaction Capable Detectors**

In a scintillation detector, when a 511 keV photon enters the detector, the depth at which the photon can be absorbed varies. The varying depth of photon interaction in the detector introduces an uncertainty in the positioning of the LOR which in turn introduces a blurring effect in the image. For example, the radial uncertainty in the positioning of two LORs is shown in Figure 2.16. The blurring caused by such phenomena is called *parallax error* in PET. Parallax error in PET introduces resolution degradation at large radial offset (compare w_2 with w_1 in Figure 2.16) which could be particularly troubling for small-diameter scanners such as a brain-dedicated or small-animal scanner or scanners with a long axial field-of-view (FOV).

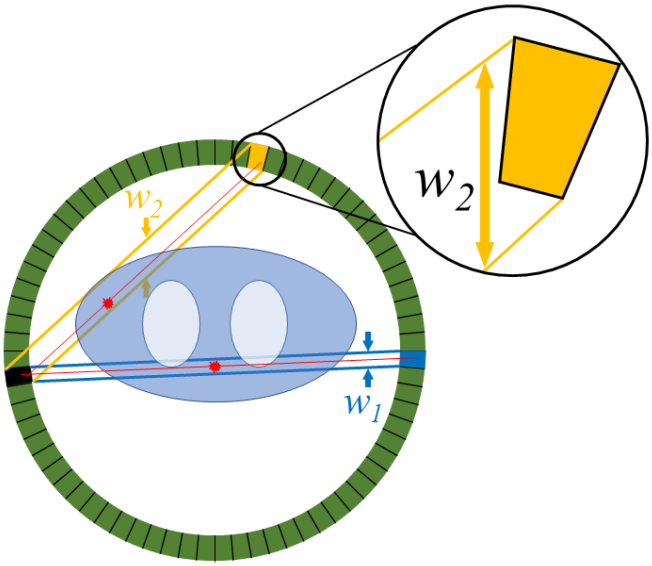

Figure 2.16. Schematic illustration of the basis of parallax error in PET.

A frequent technique to reduce parallax error in PET is to use DOI detectors. As the name implies, a DOI detector provides information on the depth at which the photon is detected and

therefore can reduce the parallax error in PET. Figure 2.17 shows an example of how a DOI detector can reduce parallax error in a PET scanner (compare w_2 in Figure 2.16 and Figure 2.17).

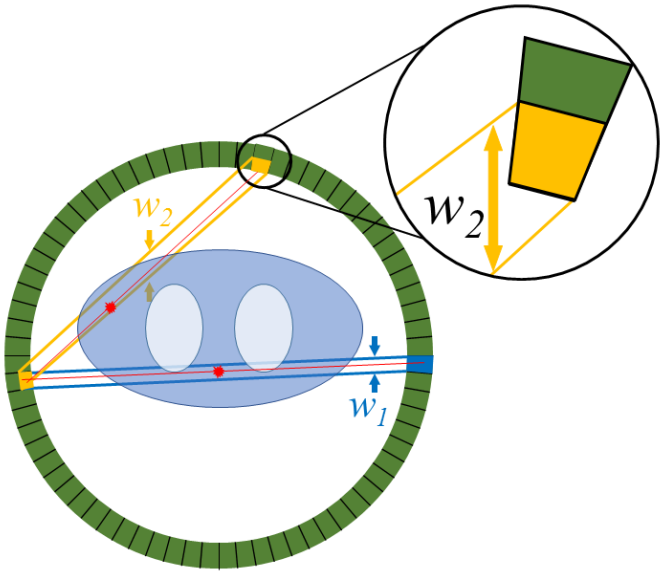

Figure 2.17. The use of a DOI detector in a PET scanner to reduce parallax error.

Various approaches that have been investigated for DOI-capable detectors are shown in Figure 2.18 and discussed in the following sections.

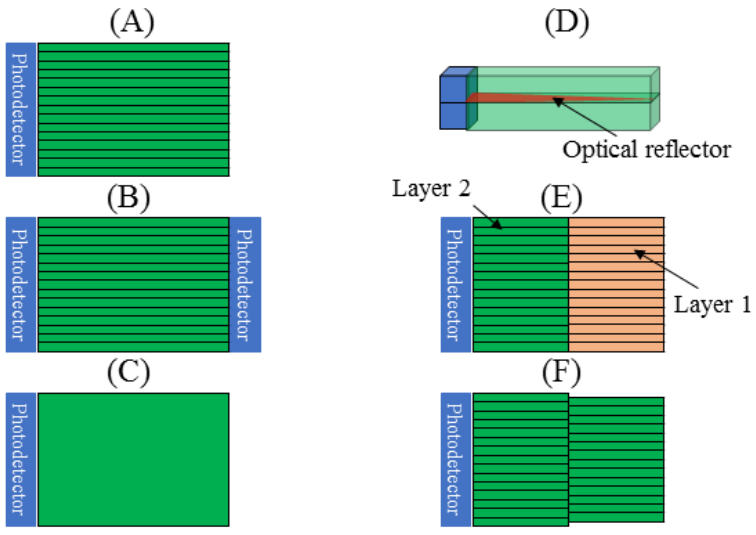

Figure 2.18. Overview of a (A) non-DOI detector, (B) dual-ended readout detector, (C) monolithic scintillator, (D) single-ended readout with continuous DOI (E) phoswich design and (F) dual-layer offset detector.

2.5.3.1 Dual-ended readout continuous depth of interaction

Optical photons generated in the scintillator of a detector can be collected from both ends of the detector, as shown in Figure 2.18-(B), which is called *dual-ended readout* or DER (as opposed to a conventional *single-ended readout* or SER detector). The ratio of the signal amplitudes from the two ends of the detector corresponds to the depth at which the optical photons were generated. Such a detector design has been proposed and evaluated in studies such as [73-77]. Compared with other DOI techniques, this design has the advantage of providing continuous DOI information. However, a DER detector is more expensive as it requires twice the number of photodetectors in comparison with other designs. Such designs also require compact front end electronics to reduce the gamma-ray attenuation by the photodetector and electronics [78].

2.5.3.2 Single-ended readout continuous depth of interaction

Several groups have investigated methods to create SER detectors that can provide DOI information. Such DOI detector designs can be categorized into two groups: 1) Detectors with a monolithic crystal block in which a continuous crystal slab is used as a detector as shown in Figure 2.18-(C) and 2) detectors with small crystal arrays with certain optical reflector arrangements as illustrated in Figure 2.18-(D).

A monolithic crystal can 3-dimensionally position events in the detector. As light dispersion depends on the distance between the generation point of the optical photons and the photodetector, DOI in a monolithic detector can be determined by measuring the extent of light dispersion on the photodetector. Also, the image of the centroid of light dispersion on photodetector shows the other 2 coordinates of the interaction. Event positioning in a monolithic detector is often distorted around the edges of the detector block due to the restricted light spread at the crystal edges [79]. Several groups have investigated such detectors [80-84] and tried to

improve the event positioning scheme through statistical-based positioning algorithms [81, 85, 86].

To acquire continuous DOI information from small crystal elements suitable for a small-animal PET scanner, some researchers have studied the use of novel optical reflector arrangements between crystal arrays. An example of a triangular optical reflector between crystals, suggested in [87, 88], is shown in Figure 2.18-(D). Triangular optical reflectors between the crystals facilitate a light sharing between the crystals such that the light sharing ratio depends on the DOI position. A detector design similar to Figure 2.18-(D) requires a one-to-one coupling of individual crystals with a compact photodetector such that the exact light sharing ratio between neighboring crystals can be measured. Lewellen *et al.* in [87] achieved a DOI resolution of 3 to 4 mm using a similar approach for lutetium fine silicate (LFS) detectors with a thickness of 30 mm.

2.5.3.3 Multi-layer discrete depth of interaction

Several detector designs have been proposed in the literature to obtain discrete DOI information through stacking layers of scintillators and a SER scheme. A traditional design, known as *phoswich* (phosphor sandwich), includes two layers of scintillators with different decay time constants (Figure 2.18-(E)). For instance, a phoswich detector may consist of LSO/LuYAP (lutetium-yttrium aluminum perovskite) layers as suggested by [89]. Due to a difference in the decay time constants of the two layers, a different pulse shape from each layer is obtained which is used to differentiate the layer of interaction [90]. A disadvantage of multi-layer detectors is the light collection loss that frequently happens in multi-layer detectors. The accuracy of DOI in these detectors depends on the thickness of each layer [78].

Other groups have studied multi-layer detectors with different optical reflector arrangement in each layer to obtain DOI information. Due to the different reflector arrangements

in each layer, the centroid of each crystal can be distinctly positioned on the detector flood map and therefore the crystal and the layer of interaction can be distinguished. Examples of such detectors with 2, 3, 4 and 8 layers are given in [91-94]. A common challenge in all these studies is that the crystal centroids at the edge are moved to the centre because of the limited light dispersion at the edge of the detector block. Therefore, crystal misidentification is likely for these detectors [78].

A *dual-layer offset* or DLO detector, shown in Figure 2.18-(F), is also another DOI detector that has been studied in [45, 48, 95-102]. The crystals in each layer of a DLO detector have an offset of one-half a crystal pitch so that the crystals in each layer can be identified in the flood map. Similar to other multi-layer detectors, the DOI accuracy in a DLO detector depends on the chosen layer thicknesses. A DLO detector can be constructed with the same scintillation material in both layers. Disadvantages of the DLO design include difficult crystal segmentation specially for corner crystals [103] and providing only two levels of DOI. In 2014, a PET insert was built in our lab that provides DOI information through its DLO detectors [45] and, as will be discussed later, DLO detectors are extensively studied in this dissertation. Similar to the DLO design, a four-layer DOI detector with relative offset was investigated through numerical simulations in [104, 105]. No prototype detector or scanner has been built based on this design.

2.5.3.4 Discrete depth of interaction with direct measurements

Detectors in this category directly collect optical photons from several depths within the detector. In other words, photodetectors are located between crystal layers so that the discrete DOI information is directly obtained by photodetectors. For example, Levin in [106] constructed a detector by stacking 8 layers of small LYSO crystal arrays mounted on a layer of APD arrays to

directly readout individual crystals. There are also other similar designs with SiPM arrays in which several layers of crystals with SiPM arrays are used to acquire direct DOI information [107, 108].

Detectors capable of direct DOI measurements are difficult to build and expensive. Also, this technique requires very thin circuit boards and photodetectors to minimize photon attenuation. However, because of the direct readout of each small crystal, these detectors can distinguish inter-crystal scattered photons in the detector which increases the positioning accuracy of the events and, subsequently, the image quality. Such detectors also have nearly complete light collection (>90%) which in turn improves the energy resolution [78].

2.6 Brain-dedicated PET Scanners

Several groups have investigated brain-dedicated PET scanners either as a stand-alone PET scanner or a PET insert. Most brain-dedicated PET scanners have a block detector design with a reduced FOV due to their smaller gantry size (compared to a whole-body scanner). A substantial amount of current research in brain-dedicated scanners is being done with DOI detectors. Currently, a wide spectrum of detector and geometry designs exist for brain-dedicated scanners. The following is a review of the details of some of the proposed brain-dedicated PET scanners in the literature with a focus on detector geometry and scanner performance metrics.

In an early attempt in 2002, Watanabe *et al.* [16] reported the performance metrics of a brain-dedicated stand-alone PET scanner based on detector blocks with an array of bismuth germanate (BGO) crystals. The scanner had retractable *septa* in between slices such that both 2D and fully 3D acquisitions were feasible. Detectors in the proposed PET scanner had 8 × 4 crystals elements with a size of 2.8 × 6.55 × 30 mm^3 with position-sensitive PM tubes (PS-PM tubes). The detectors were arranged such that a transaxial/ axial FOV of 330/ 163 mm is achieved. The scanner

was reported to have a *spatial resolution* (see Section 2.9.1) of 2.9 mm at the centre of the FOV with a sensitivity (see Section 2.9.2) of 6.4 kcps/kBq/ml in 2D mode when slice septa were in place. No sensitivity in 3D mode was reported for this scanner.

The high-resolution research tomograph (HRRT) is another brain-dedicated stand-alone PET scanner that has different generations. The first prototype HRRT was installed in 1999 [109, 110]. The first generation dual-layer HRRT was built with eight panel detector heads arranged in an octagon such that a transaxial/ axial FOV of 312/ 252 mm was provided [111]. Each detector head of the first generation HRRT consisted of 117 double layer gadolinium orthosilicate (GSO)/LSO blocks such that DOI information can be obtained by a difference in the light decay in the layers. Crystal faces of the HRRT had a size of 2.1×2.1 mm^2 with a thickness of 7.5 mm in each layer (a total of 15 mm thickness) [111]. Detectors of the first generation HRRT then upgraded to phoswich detectors in which an LSO/LYSO detector replaced the previous detectors of the first generation. Both generations used PM tubes as their photosensors and had similar geometrical properties for their gantries. A spatial resolution of 2.5 mm (isotropic) was reported for the first and second HRRT generation [20, 111-113].

Sheikhzadeh *et al*. [114] in 2017 proposed a brain-dedicated scanner with a ring diameter of 357 mm, an axial length of 204 mm and 4 rings of detectors. The detectors of this so-called *NeuroPET* scanner has two layers of cerium-doped LYSO crystal arrays stacked on top of each other in the form of a DLO detector. The DLO detector used in [114] had 22×22 crystals in the outer layer and 21×21 crystals in the inner layer with a size of $2.3 \times 2.3 \times 10$ mm^3 in both layers. Detector analysis of NeuroPET was done via numerical simulations and no performance metrics for the scanner have been reported so far.

Highly portable brain-dedicated PET scanners have also been proposed to allow for more versatile brain functional studies. The *PET-hat*, for example, was designed such that a wearable single ring PET scanner can acquire brain PET images while the patient is sat, as shown in Figure 2.19 [115, 116].

Figure 2.19. The PET-hat system (left) and the PET-hat with a subject (middle and right) © 2011 IEEE [115].

Detectors of the PET-hat consisted of two layers of GSO scintillators with crystals of size 4.9 × 5.9 × 7 mm^3 in the inner layer and 4.9 × 5.9 × 8 mm^3 in the outer layer. Different concentrations of cerium were used in the two layers to differ the decay time and create the DOI detector. The scanner achieved a transaxial/ axial spatial resolution of 4.0/ 3.5 mm at the centre of the FOV with a peak sensitivity of about 0.7% (see Section 2.9.2).

Another example of a brain-dedicated PET scanner is the jPET-D4 [117-119]. The jPET-D4 was built with four-layer DOI detectors with GSO as the choice of scintillator and PS-PM tubes as the choice of photomultiplier. Each layer of the jPET-D4 detectors had 16 × 16 crystals with a size of 2.9 × 2.9 × 7.5 mm^3 and the detectors were arranged such that a transaxial/ axial FOV of 300/ 260 mm is provided. The method for DOI discrimination was based on pulse shape discrimination and *anger type calculations* (see Section 2.7.1) [93]. The jPET-D4 showed a spatial resolution of 3 mm almost uniformly across the FOV. The use of semiconductor detectors in brain-

dedicated PET scanners have also been proposed in [120-123] that will not be explained here. Also, several other non-cylindrical PET scanners (e.g. a helmet-like PET scanner), suitable for brain studies, have been investigated in the literature that are not discussed here. Examples of such scanners can be seen in [21, 124-129].

The first clinical brain-dedicated PET insert was developed by Siemens in 2007 for the Siemens Trio 3 T MR scanner [37, 130]. The developed PET insert consisted of 23,040 LSO scintillator elements grouped into detectors of 12 × 12 crystals with an individual crystal size of $2.5 \times 2.5 \times 20$ mm^3. The detectors were arranged such that an inner ring diameter of 355 mm and an axial FOV of about 190 mm were provided. An array of APDs was used to readout detectors with no DOI information. A similar spatial resolution to the HRRT (about 2.5 mm at the centre of the FOV) was reported for this scanner. More details on the performance evaluation metrics of this scanner can be found in [17]. Similarly, Jung *et al.* in 2015 [18] developed a brain PET insert for the Siemens Trio 3 T MR scanner. Detector blocks of the PET insert were composed of an array of 16 × 16 LYSO crystal elements, with a size of $3 \times 3 \times 20$ mm^3 each, coupled with an array of silicon photomultipliers or SiPMs. A spatial resolution of about 3 mm at the centre of the FOV and a peak sensitivity of about 0.8% for this scanner was reported. Due to the absence of any DOI information, the radial spatial resolution of the PET scanner was degraded by about 56% at 10 cm off-centre of the FOV.

A detector for a brain-dedicated PET insert suitable for the Siemens Trio 3 T MR scanner was developed in [131] that collects DOI information through a multi-layer detector. The proposed detector in [131] had 4 layers of lutetium-gadolinium oxyorthosilicate (LGSO) scintillators with different optical reflector arrangements in each layer [94, 132, 133] (see Section 2.5.3.3). Crystals with an individual size of $2.9 \times 2.9 \times 5.0$ mm^3 in each layer, were grouped in detector blocks of

12 × 4 × 4 crystals and readout by an array of MPPCs. This detector design was later updated in [134] such that each detector has 6 × 6 × 4 LYSO crystals with a size of 1.45 × 1.45 × 4.5 mm^3 so that a uniform radial spatial resolution of 1.45 mm can be achievable.

The MINDView brain PET is another brain-dedicated PET insert, suitable for the Siemens mMR 3 T MR scanner, that has been described in [135-138]. Although staggered pixelated crystal arrays had been investigated for this scanner in [137], monolithic LYSO blocks of 50 × 50 × 20 mm^3 were used in the built prototype that were readout by an array of SiPMs. This design provides a continuous DOI as described in Section 2.5.3.2. An inner diameter of 330 mm and an axial length of 152 mm were considered for this scanner. System evaluations in [135] suggested that the scanner can achieve a spatial resolution of 1.6 mm over the entire FOV and a peak sensitivity of 2.7%.

2.7 Data Acquisition in PET

Radiation detectors in a PET scanner are aimed to detect photons originated from positron-electron annihilations. As mentioned in Section 2.3.2, the detection of two singles within a pre-defined CTW is referred to as a *coincidence*. A list or a histogram of the recorded coincidences are then used to reconstruct an image of the radioactivity distribution within the object. In the following sections, sorting coincidences and different methods of recording data in PET are discussed.

2.7.1 **Coincidence Sorting**

When 511 keV photons enter a PET detector, they may undergo different interactions with the scintillation material. In a one-to-one coupled detector, in which the optical photons from each crystal are collected separately, the number of optical photons corresponds to the energy deposited

in the crystal. In detectors where a collection of crystal arrays is readout with a photodetector, the number of optical photons corresponds to the energy deposited in all crystal elements. In such detectors, the position of the detected photon is determined through comparing the signal strengths from various sides (or corners) of the photodetector. The method of positioning events in a scintillator detector by using the signal magnitude from different sides of a scintillator is known as *Anger logic* [139-141]. The total magnitude of the output signal corresponds to the deposited energy of one or multiple photon interactions in the detector.

Following the terminology used in *Geant4*[1] (**Ge**ometry **an**d **t**racking) and *GATE* (Geant4 application for tomographic emission) [142-144], each interaction is referred to as a *hit* and the output signal of the detector is referred to as a *single* in this dissertation. A single may comprise one or multiple hits within a detector. Modern PET detectors record the time of each single with a precision of about hundreds of picoseconds to a few nanoseconds that is known as the *timing resolution* of the detector [145-148]. In a PET scanner, singles may be stored for further processing or discarded depending on whether they meet certain criteria. Included in these criteria is that the energy of the recorded singles should lie within a predefined energy window. The lower threshold of this energy window, known as the *lower level discriminator* (LLD), is aimed to reject scatter coincidences (see Figure 2.10), while the upper threshold, known as the *upper level discriminator* (ULD), rejects pileup (see Section 2.8.4).

As mentioned in Section 2.3.2, coincidences are created if two singles are detected within a certain CTW (e.g. 10 ns). Two singles detected within a predefined CTW can only create a coincidence if a minimum angular difference (minimum block difference) between the two

[1] Geant4 is a Monte Carlo simulation package as explained in Section 2.11.

involved detectors is met. If more than two singles are detected within CTW, a *multiple coincidence* is created which can be handled differently in different PET scanners. The explained concepts and terms in sorting coincidences are illustrated in Figure 2.20.

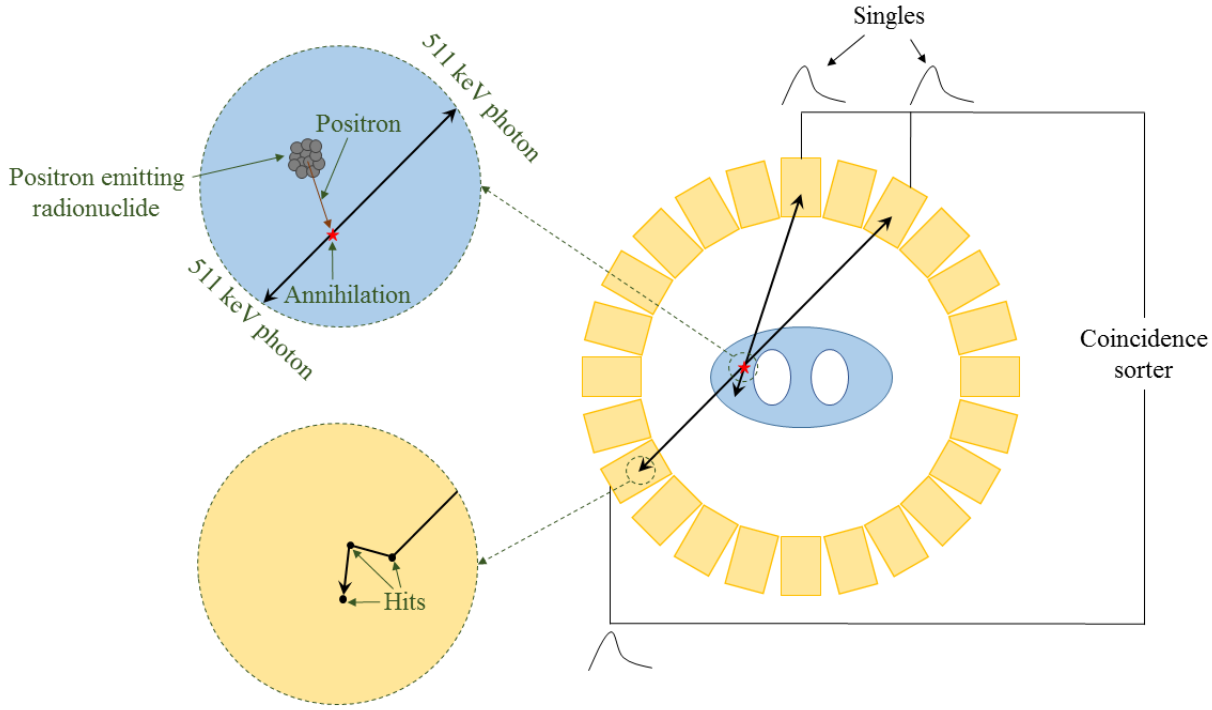

Figure 2.20. Hits, singles and coincidences in PET. If the 3 singles shown in this figure are detected within a CTW, a multiple coincidence is recorded.

Each true coincidence creates an LOR that [ideally] crosses the positron-electron annihilation location. The minimum block difference in a PET scanner determines the largest radial distance for an LOR. In other words, the transaxial FOV of a PET scanner is in fact determined by ring diameter and the minimum block difference of the PET scanner as shown by Figure 2.21.

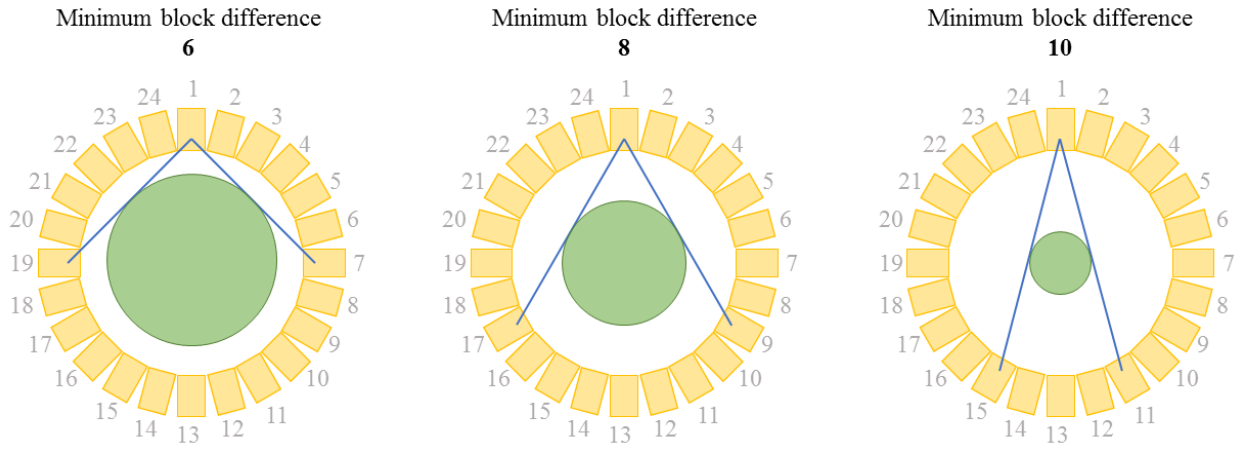

Figure 2.21. Transaxial FOV of a PET scanner (green circle) with various minimum block differences. Detectors are numbered as shown next to each detector.

2.7.2 **List-mode Data Acquisition**

In a PET scan, accepted coincidences may be stored in a list of events where for each coincidence, crystal numbers (crystal IDs) of the involved crystals in the coincidence are recorded. If the scanner is capable of recording the time for each single with a precision of about less than a nanosecond, a time stamp for each coincidence is also stored to be used in the image reconstruction process. Such scanners are known as *time-of-flight* (ToF) scanners. The ToF information can be used to reduce the length of each LOR to Δd which is calculated by

$$\Delta d = \frac{\Delta t \times c}{2} \tag{2.19}$$

in which Δt is the difference in the arrival times of the photons and c is the speed of light (3×10^{10} cm/s).

List-mode data are specifically valuable in cases where preserving original LOR sampling in the scanner is preferred [149]. Reconstructing list-mode data requires the use of a statistical list-mode image reconstruction technique that will be discussed in Section 3.2. In a statistical list-mode image reconstruction algorithm, a more accurate modeling of the physical effects is feasible than

analytical *sinogram*-based reconstructions which can improve the quality of PET images (Section 2.7.3). List-mode data can also incorporate ToF information for each coincidence that can be used to improve image *signal-to-noise ratio* (SNR) [150, 151]. Furthermore, a list-mode reconstruction algorithm has been shown to perform much better than sinogram-based reconstructions when recorded PET data suffer from low statistics [152-154]. Therefore, list-mode acquisition and reconstruction techniques are becoming more popular and have been used in this dissertation as well. Nonetheless, an overview of sinogram formation in PET is given in the next section.

2.7.3 **Sinogram Formation**

A *sinogram* is a histogram of the number of detected coincidences per each LOR. Depending on the LOR's geometrical orientation, each LOR is assigned into a sinogram *bin*. In reality, the total number of sinogram bins may be less than the possible number of LORs in the scanner as certain LORs are grouped into a bin. In 2D PET, the bin corresponding to an LOR depends on 2 distinct geometrical properties of the LOR: s, the radial distance of the LOR from the origin, and φ, the orientation angle of the LOR. In 3D PET, however, the bin corresponding to an LOR depends on 4 distinct geometrical properties of the LOR: s and φ as defined, z, the axial position of the LOR, and θ, the axial slope of the LOR as shown in Figure 2.22.

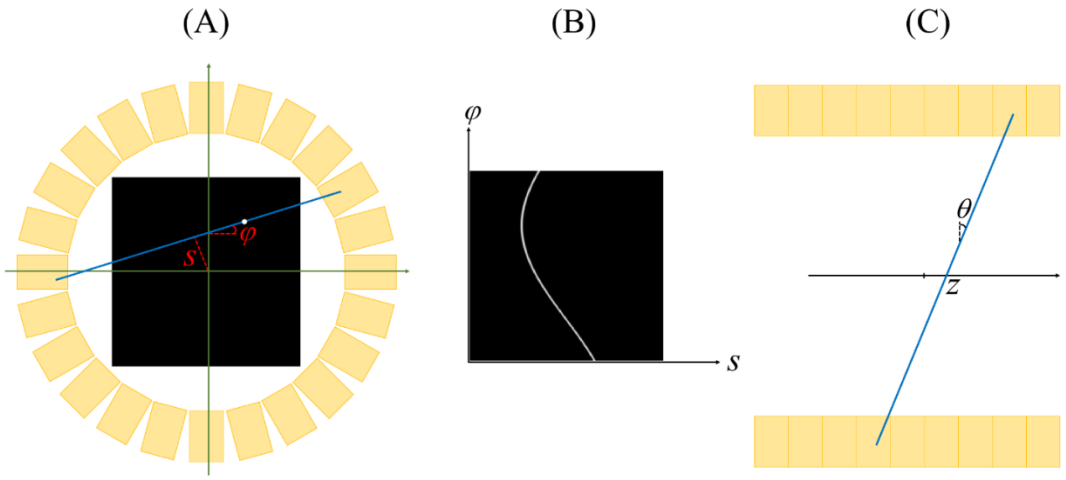

Figure 2.22. (A) A point source (the white dot) in PET with a sample LOR, (B) sinogram corresponding to the point source shown in (A) and (C) a sample LOR with an axial distance z and an axial slope θ in 3D PET.

Each row of the sinogram is a 1D distribution of the total number of recorded coincidences at a certain angle, also known as a *projection*. Therefore, a projection can be mathematically modeled as a set of line integrals at a certain view. A sinogram is composed of projections from all angles φ and, as shown in Figure 2.22-(A) and Figure 2.22-(B), discrete data at all angles can be collected on the transverse plane (except when there are gaps between the detector blocks, a case ignored in this discussion). However, as shown in Figure 2.22-(C), θ is limited due to the limited axial extent of the scanner. The missing projections cause incomplete datasets which is a problem in analytical reconstruction algorithms. This problem can be overcome by 3D *rebinning algorithms* which sort the 3D data into a set of ordinary 2D sinograms [155, 156]. The generated set of sinograms are then reconstructed using a reconstruction technique.

2.8 Quantitative PET Imaging

PET is known as a powerful *quantitative imaging* technique, meaning that the radiotracer concentration in an image voxel can be accurately measured. A number of corrections must be applied to the images in order to obtain quantitatively accurate images that are discussed in the following sections.

2.8.1 **Attenuation Correction**

In a PET scan, photons emitted inside the body are attenuated as a result of the photoelectric interaction or photon scattering. The probability of a back-to-back 511 keV photon pair travelling a distance *T* within the body reaches the detectors with no interactions within the body is given by

$$P = e^{-\mu T} \tag{2.20}$$

in which μ is the LAC at 511 keV. For instance, the LAC of 511 keV photons for soft tissue and bone is about 0.095 cm^{-1} and 0.13 cm^{-1}, respectively [4]. While attenuation correction is the most important consideration for quantitative PET in clinical practice, it is relatively easy to implement if co-registered CT images are available (for instance, through a hybrid PET/CT scanner). Attenuation correction in PET requires a quantitatively accurate and aligned *attenuation map*. An attenuation map is the map of LACs of the object which conventionally was obtained through a bilinear curve that converts CT numbers in Hounsfield units[1] to LACs at 511 keV [157].

Attenuation correction in PET/MR, however, is more challenging as MR images do not represent electron density. In fact, the strength of an acquired MR signal is a function of proton density (the amount of water) which complicates the generation of attenuation maps for PET. A quantitative inaccuracy of up to 45% is reported when segmentation errors (a 'segmentation error' refers to a misidentification of a tissue type) occur in the derivation of attenuation maps from MR images [158]. Currently, there are three methods for generating the attenuation map in PET/MR.

The first group of methods are *template-based* approaches. The methods in this category are based on a library of coregistered attenuation maps and MR templates. For each patient, a

[1] CT numbers or Hounsfield units (named after the inventor of medical CT scanners) have a linear relationship with respect to the LAC and is defined as CT number = $1000 \times (\mu - \mu_{water}) / \mu_{water}$.

nonlinear registration transformation is found that adapts the MR template to the patient MR image. The same transformation is then applied to the attenuation map template to generate the attenuation map of the patient [159].

There are also *segmentation* approaches that aim to generate patient specific pseudo-CT images by segmenting the acquired MR images into various tissue classes. Then, the patient attenuation map is generated by assigning the corresponding LAC to each tissue class. The MR images need to be acquired with a previously defined MR sequence. In one of the studies using this approach for brain PET/MR, MR voxels within the skull were assigned three different LAC values based on a classification of each voxel to either skin, brain or bone [160].

The last category of PET/MR attenuation correction techniques is known as *sequence-based* approaches. Typically, bone structures cannot be seen in MR images obtained by pulse sequences routinely done in clinical studies and so the differentiation of bone from air is a difficult task. Ultrashort echo time (UTE) sequences were specifically introduced to visualize tissues with a short T2 and so can enable the visualization of the bones [161]. Therefore, UTE images contain enough attenuation-related information that the attenuation map generation can be done solely based on them. Similar to direct segmentation-based approaches, UTE images will then be segmented.

2.8.2 **Scatter Correction**

As shown in Figure 2.10, photons scattered in the patient can be assigned to an apparent LOR far from the true annihilation point. Scatter coincidences in PET can largely contribute to image noise and degrade quantitative accuracy of PET images. As the predominant mode of 511 keV photon interaction in scintillators is Compton scattering, many 511 keV photons also scatter

within the scintillator and only deposit part of their energy. Currently, there are two main groups of scatter correction techniques in PET.

The first approach is to estimate the contribution of scatter coincidences to any given LOR (or projection) by modeling Compton scattering for the calculated attenuation map discussed in the previous section. The estimated scatter contribution can then be subtracted from the projection profiles or, more accurately, modeled in the image reconstruction algorithm. The second approach is to estimate the contribution of the scattered photons based on an evaluation of the projection profiles outside the object. Since coincidences detected outside the object boundary are scatter coincidences (assuming that random coincidences are already subtracted), scatter contribution within the object is extrapolated by measuring the scatter component outside the object. In this method, the estimated scatter contribution is subtracted from the projections before image reconstruction [4, 162].

2.8.3 **Random Correction**

The distribution of random coincidences is relatively uniform across the FOV. Similar to scatter coincidences, random coincidences also affect the quantitative accuracy of the images and reduce image contrast. Random coincidences are estimated and subtracted from the projections prior to image reconstruction or modeled within the reconstruction algorithm. The rate of random coincidences can vary largely with the amount of activity in or outside the FOV. A common approach to estimate the rate of random events is by using the singles rate of each detector, known as the *singles-based estimation* method. In this approach, the rate of random coincidences for an LOR between connecting detectors *i* and *j* in a scanner (R_{ij}) is described by

$$R_{ij} = 2\tau S_i S_j \tag{2.21}$$

where τ is the scanner's CTW and S_i and S_j are the singles rate of detectors i and j, respectively [4]. The other common approach to estimate the random contribution is known as the *delayed window technique*. The delayed window technique is based on an assumption that the mean number of random coincidences detected along any LOR in a typical CTW is equal to the mean number of random coincidences measured in a delayed coincidence window. For instance, the CTW (which is typically between 4 to 12 ns) can be delayed by 64 ns such that coincidences are created from a pair of singles that have a timing difference of 64 to 76 ns (when CTW for prompt events is 12 ns). Detected coincidences in the delayed coincidence window are definitely random coincidences as no true or scatter coincidence can be detected in this CTW. The rate of detected random coincidences using the delayed timing window is considered similar to that using the un-delayed (i.e. prompt) window [163].

2.8.4 **Deadtime Correction**

As described before, optical photons are produced when annihilation photons interact with the scintillation material of a PET detector. The optical photons are then collected and amplified by a photomultiplier to generate an electronic signal. If a second photon interacts with the scintillation detector before the previous event is processed and the signal is generated, the detector is unable to generate a new signal. The minimum time required to complete processing an event is defined as the detector *deadtime*. The event loss due to detector deadtime is specifically a problem at high count rates. Another relevant problem in high count rates is pulse *pileup*, which occurs when a signal from an event piles up on the signal from the previous event. In pileup, the total recorded energy by the detector is simply the sum of all photon energies.

Deadtime can be modeled as *paralyzable* or *non-paralyzable*. A non-paralyzable detector does not accept any new events for processing when it is still dealing with an earlier event.

Therefore, in a detector with a non-paralyzable deadtime model, any new events will be lost during the processing time of the earlier event. A paralyzable detector, on the other hand, processes all events and extends the deadtime when a new pulse is generated. The two models for deadtime are shown in Figure 2.23.

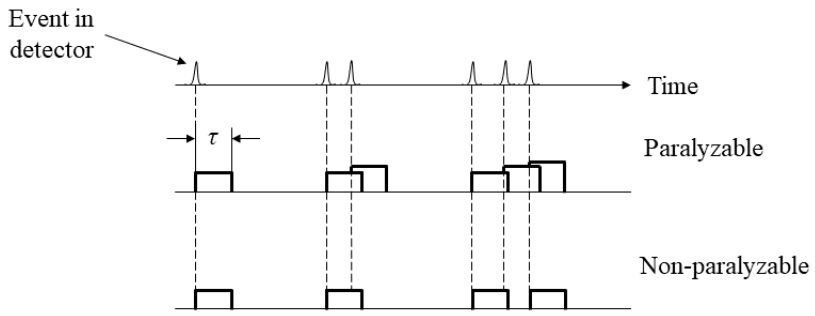

Figure 2.23. Paralyzable and non-paralyzable deadtime models with a deadtime of τ.

Note how in the example shown in Figure 2.23 for the paralyzable model, three counts are recorded in total with an added amplitude to the last two signals, whereas four counts are recorded for the non-paralyzable deadtime model. The time τ is also known as the *deadtime constant*.

In a detector with non-paralyzable deadtime, the relationship between the observed count rate m with true count rate n and system deadtime τ is given by

$$n = \frac{m}{1 - m\tau} \ . \tag{2.22}$$

For a paralyzable deadtime model, this relationship is given by [8]

$$m = ne^{-n\tau} \ . \tag{2.23}$$

A graph of the observed count rate against the true count rate, using equations (2.22) and (2.23), is shown in Figure 2.24 for a deadtime constant of τ = 750 ns.

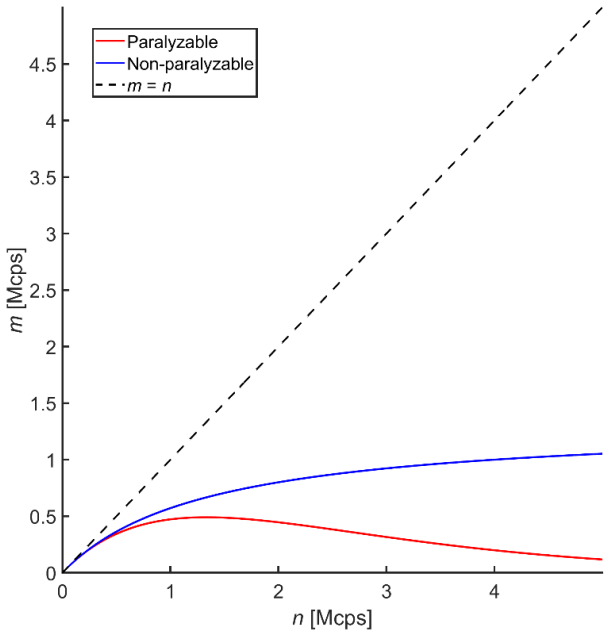

Figure 2.24. A graph of the observed rate *m* as a function of the true rate *n* for detectors with paralyzable and non-paralyzable deadtime with a deadtime constant of 750 ns.

Deadtime correction can be done by applying an adjusted deadtime correction factor for each detector or through applying a known deadtime model on the scanner globally.

2.8.5 **Normalization**

In a PET scanner, there are variations in the detection efficiency of various LORs. The efficiency variation is due to gain differences in the photomultipliers and the properties of each scintillation crystal. *Normalization* in PET refers to the process of correcting for such variations. The most straight-forward approach to calculate normalization coefficients for each LOR is by exposing all LORs to some known activity which can be accomplished by irradiating all detector pairs to a 511 keV photon source with a known geometry. This method is known as *direct normalization*. The normalization coefficient for each LOR between crystals *i* and *j*, in the absence of attenuation, scatter and random, is then calculated as

$$n_{ij} = \frac{N_{ij}}{A_{ij}} \tag{2.24}$$

in which N_{ij} is the recorded number of coincidences between crystals i and j and A_{ij} is the total activity along the LOR.

Direct normalization requires recording a large number of LORs to achieve a reasonable statistical accuracy. Since the normalization source should also be a low activity source with minimal attenuation, a long acquisition time is commonly required for direct normalization. The long acquisition time can be inconvenient, especially considering that PET normalization is usually done on a weekly or monthly basis. Also, as more modern scanners have more crystal elements and LORs (up to hundreds of millions), direct normalization is not the preferred method.

The other normalization method, known as the *component-based* method, employs a reduction method to resolve the issues of direct normalization. Normalization coefficients obtained from the component-based approach are calculated as the product of intrinsic crystal efficiencies and geometrical factors. Intrinsic crystal efficiencies account for the variation in crystal efficiencies and are calculated from the average sum of coincidence efficiencies between each crystal and all opposite crystals that are in coincidence. The geometrical factors account for photon incidence angle and are completely dependent on the geometry of the scanner. Therefore, the geometrical factors need to be calculated once only by either analytical calculations, numerical simulations or measurements [164, 165]. Further details and implementation of the component-based method are discussed in Section 8.4.

2.9 Performance Characteristics in PET

The performance of a PET scanner is evaluated through several measurements. A guideline to perform standard evaluations of small-animal and clinical PET scanners is given in NEMA NU 4 [166] and NEMA NU 2 [167], respectively. Some of the most common performance measurement metrics that are used in this dissertation are discussed below.

2.9.1 **Spatial Resolution**

The *spatial resolution* of a PET scanner is a measure of the ability of the scanner to distinguish between two points after image reconstruction. To measure the spatial resolution of a PET scanner, images of a point source at various radial offsets are reconstructed and the full width at half maximum (FWHM) and full width at tenth maximum (FWTM) of the reconstructed point source are measured. The spatial resolution FWHM of a PET scanner Γ at the centre of the scanner can be roughly estimated by

$$\Gamma = a.\sqrt{g^2 + (0.0022D)^2 + s^2 + b^2} \tag{2.25}$$

where a is a factor affected by the choice of image reconstruction algorithm, g is the geometrical spatial resolution (determined by the geometrical properties of the PET detectors), D is the diameter of the detector ring (the term 0.0022D is due to the back-to-back photon acollinearity), s is the source size including positron range and b is a general term accounting for other blurring effects [168].

2.9.2 **Sensitivity**

The sensitivity of a PET scanner is defined as the rate of true coincidences for a given source strength. As sensitivity varies significantly throughout the FOV, often a sensitivity profile along the central axis of the PET scanner is shown for the scanner. The maximum sensitivity (*peak*

sensitivity) of a cylindrical PET scanner is at the geometrical centre of the scanner. The sensitivity of a PET scanner is measured with a low activity source to minimize the rate of random coincidences and deadtime losses. Higher sensitivity scanners are preferred as they reduce noise in the images and allow for PET scans with a lower injected activity.

2.9.3 **Noise equivalent count rate**

A common performance parameter that reflects the count rate performance of a PET scanner is the *noise equivalent count rate* (NECR). The NECR is the equivalent count rate that gives rise to the same statistical noise level as the observed count rate after random and scatter coincidences are corrected [4]. The count rate measured by a PET scanner (prompt count rate or R_{prompt}) comprises true (R_{true}), scatter ($R_{scatter}$) and random coincidences (R_{random}) such that

$$R_{prompt} = R_{true} + R_{scatter} + R_{random} \ . \tag{2.26}$$

The NECR is then defined as

$$\text{NECR} = \frac{R_{true}^2}{R_{true} + R_{scatter} + \alpha R_{random}} \tag{2.27}$$

where α is a constant depending on the method chosen for the estimation of random events (in this dissertation $\alpha = 1$). The NECR relates to the SNR of the reconstructed images and is therefore of significant importance [169]. Often, the NECR of a scanner is plotted as a function of activity in a uniform cylinder along with plots for true, random and scatter coincidence rates.

2.10 PET Limitations

Despite the impressive gains in the performance of PET scanners, PET encounters several limitations that will be reviewed in this section.

2.10.1 **Image Signal-to-noise Ratio**

Images obtained from a PET scan suffer from a low SNR due to the poor photon statistics. Low SNR reduces the ability to detect small lesions and increases image noise which is particularly troubling for low count or dynamic studies. Theoretically, low SNR can be mitigated by increasing scan time or the injected dose of the radiotracer. However, the duration of a diagnostic PET scan is already about 10 – 20 minutes which is much longer than many other imaging modalities. Also, according to the current imaging protocols, a typical PET scan delivers an effective radiation dose of about 5 – 8 mSv[1] to the patient which increases as the injected radiotracer dose increases (to compare, a chest X-ray delivers an effective dose of about 0.1 mSv) [4, 15]. In neurological studies with PET, a long scan time is often highly desired to increase the image SNR which increases the likelihood of patient motion during the scan [170].

2.10.2 **Spatial Resolution**

The spatial resolution of a PET scanner is limited by several factors. Most clinical scanners have scintillation crystals with a face size of about 3 to 4 mm and a depth of about 15 to 25 mm. The limited size of each crystal introduces an intrinsic resolution of about half a crystal width at the centre of the gantry. Also, positrons emitted from certain radionuclides can travel up to several millimeters before they undergo positron-electron annihilation. For example, the effective range of positrons from ^{18}F and rubidium-82 (^{82}Rb) are about 0.5 mm and 3 mm in water, respectively. Photon acollinearity is another limiting factor on the theoretical resolution of a scanner. In reality, the spatial resolution of a scanner is also limited by event localization uncertainty in the detector and the image reconstruction technique [15].

[1] Sievert (Sv) is the SI unit of effective dose.

2.10.3 **Acquisition Time and Cost**

As mentioned in Section 2.10.1, a PET scan is a relatively long process. The long acquisition time of a PET scan reduces patient throughput and makes the PET scan more prone to motion artifacts as any movement during the scan can blur the images. Although total-body PET scanners with an extended axial FOV (up to 2 meters) have been suggested as a solution to reduce patient dose and acquisition time, these scanners are very expensive [171].

Scintillation materials used in a PET scanner are the most expensive part of the scanner. Therefore, while using thicker detectors and extending the scanner with more detectors increase the sensitivity of the scanner and reduce the required acquisition time for a subset of brain imaging tasks, they make the scanner significantly more expensive.

2.11 Monte Carlo Simulation in PET Imaging Application

Monte Carlo (MC) simulations in PET imaging refer to a numerical approach based on random sampling to simulate a PET scanner. One of the most commonly used MC-based software toolkits that simulates the passage of particles through matter and has applications well beyond medical imaging is *Geant4* that has been developed by CERN [172-174]. The MC simulation package GATE [142-144], was developed by the international OpenGATE collaboration to employ a dedicated scripting mechanism to extend the native command interpreter of Geant4 to medical imaging and radiation therapy applications. GATE has been well validated against real world measurements and used extensively in detector development and optimization studies. Also, some open source image reconstruction packages (such as STIR [175] or CASToR [176]) now accept data generated from GATE simulations. In this dissertation, the majority of the presented data have been obtained through MC simulations with GATE.

Chapter 3 : Tomographic Imaging Reconstruction in PET

In this chapter, the process of reconstructing an image from the recorded sinograms or list of coincidences is explained. Analytical reconstruction approaches are briefly explained in Section 3.1. Section 3.2 reviews the basis of iterative reconstruction algorithms in PET. Finally, in Section 3.3, different implementations for GPU-based image reconstruction algorithms are studied.

3.1 Analytical Reconstruction Techniques

The *Projections* of an image can be defined by a mathematical operator called the *Radon transform* [177, 178]. The Radon transform (projection) of an image $f(x, y)$ along an angle φ and at a distance s from the origin (see Figure 2.22) is fundamentally the summations of all image values along that direction that is given by

$$g(s,\varphi) = \int_{-\infty}^{+\infty} f(s\cos\varphi - u\sin\varphi, s\sin\varphi + u\cos\varphi)du \ . \tag{3.1}$$

A sinogram $g(s,\varphi)$ is created from all recorded coincidences in a PET scanner at all possible s and φ values. The most common analytical approaches for reconstructing an image from a sinogram are *backprojection* and *filtered backprojection* (FBP) that are reviewed below.

3.1.1 **Simple Backprojection**

The *backprojection* operator accumulates the contribution of all LORs passing through any point (x,y) in the 2D image and is mathematically given by

$$b(x,y) = \int_{0}^{\pi} g(s,\varphi)d\varphi \ . \tag{3.2}$$

In PET, the number of angles at which projections are acquired is limited and a simple backprojection of the acquired projections can lead to a blurry image as shown in Figure 3.1. While increasing the number of angles φ can reduce the blurring effect in the image, images reconstructed with simple backprojection always suffer from this artifact. The reason is that the backprojection operator produces images that are convolved with a $1/s$ function. The blur in simple backprojection is commonly known as $1/s$ (or more commonly $1/r$) blurring.

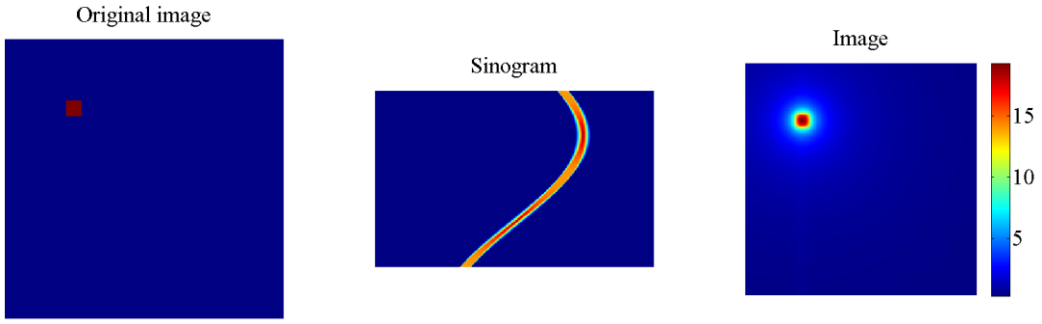

Figure 3.1. Sinogram of a sample image and its reconstructed image using backprojection.

As a result of the $1/s$ blurring, the relationship between the final reconstructed image $b'(x,y)$ and $b(x,y)$ can be expressed by

$$b'(x,y) = b(x,y) * (1/s) \tag{3.3}$$

in which * represents the convolution operator [4].

3.1.2 **Filtered Backprojection**

To reduce the effects of $1/s$ blurring in the reconstructed image, the backprojected image can be deconvolved with the $1/s$ function. It has been proven that deconvolving the image by the $1/s$ function is equivalent to filtering each projection in the frequency domain with the *ramp filter* (or a slightly modified version of the ramp filter) and then backproject the filtered projections. The explained method is known as *filtered backprojection* (FBP) and can be mathematically shown as

$$b(x,y) = \int_0^{\pi} g_f^*(s,\varphi)\, d\varphi \tag{3.4}$$

in which $g_f^*(s,\varphi)$ is the Radon transform of the image that has been filtered by the ramp filter in the frequency domain. Several filters may be used in FBP, such as the Ramp, Hamming, Hann or Butterworth filter. As the ramp filter amplifies high frequency noise, a windowed ramp filter is used in FBP reconstructions such that frequency components above a cut-off threshold are set to zero [163]. A sample image reconstructed with FBP, using the Hann filter, is shown in Figure 3.2.

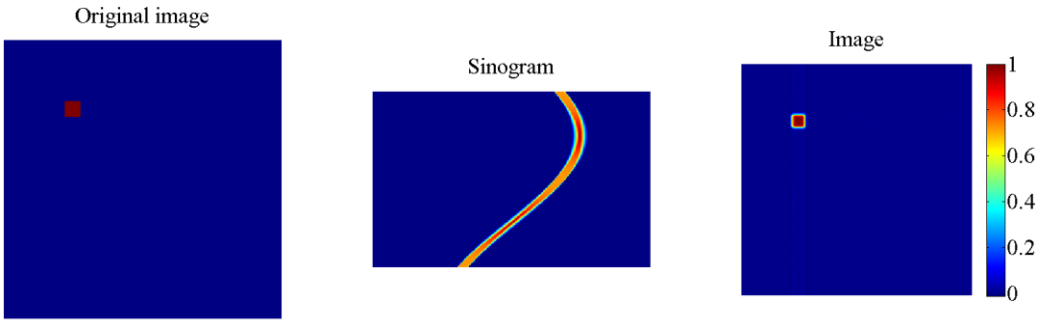

Figure 3.2. Sinogram of a sample image and its reconstructed image using filtered backprojection.

Further details of FBP are not discussed in this section as FBP has not been used in this dissertation. The example shown in Figure 3.2 is a simple 2D image reconstruction in which the axial slope or tan(θ) (θ is defined in Figure 2.22) is assumed to be 0. In a fully 3D PET scanner however, data are collected from other non-zero axial slopes where tan(θ) is limited due to the truncated axial FOV of the scanner. As FBP assumes measured data are complete from all possible angles, an extension of 2D FBP to 3D requires a complete dataset along all oblique (non-zero tan(θ)) sinogram planes. As pointed out in Section 2.7.3, rebinning algorithms such as Fourier rebinning or 3D reprojection [155] are commonly used in an analytical 3D PET reconstruction technique to reconstruct images from incomplete datasets. The details of these techniques can be found in [179].

Reconstruction methods such as FBP are fast and are not computationally heavy. Nevertheless, images obtained from an analytical image reconstruction algorithm suffer from a low resolution as such algorithms cannot model the physical phenomena involved in a PET scanner.

3.2 Iterative Reconstruction Techniques

In this dissertation, iterative reconstruction algorithms have been used to reconstruct images obtained from simulated and real PET scanners. In principle, mathematical models based on statistical theory can be built to incorporate every physical and statistical phenomenon involved in the PET acquisition in the reconstruction algorithm. Assuming a simple 2D reconstruction example, an image $f(x,y)$ and its corresponding sinogram $g(s,\varphi)$ are matrices that can be considered as vectors by linearizing the matrices (stacking the rows of the matrices). This way, g_i and f_j are created where i and j are indices pointing to a sinogram bin and an image pixel, respectively. The value g_i of any sinogram bin is a weighted sum of the m pixel values in the image which can be written as

$$g_i = a_{i1} f_1 + a_{i2} f_2 + ... + a_{im} f_m = \sum_{j=1}^{m} a_{ij} f_j \tag{3.5}$$

where a_{ij} are elements of a matrix A known as the *system matrix*. An element a_{ij} is a weighting factor representing the contribution of pixel j to the number of counts detected in sinogram bin i. The product shown in equation (3.5) can be concisely written as $g = Af$ indicating that the matrix A can generate a stack of projections g from an image f. The process of generating projections from an image f is known as *[forward] projection*. Any element a_{ij} is a weighting factor representing the contribution of pixel j to the number of counts detected in bin i. In a simple example, matrix A can be a binary matrix with elements 0 or 1, meaning that a given pixel either

does or does not contribute to the number of counts detected in any sinogram bin. To make this concept clear, the system matrix of a 2D PET scanner with limited LOR arrangements is described next. Inspired by [180], an illustration of orthogonal projections for a 3 × 3 image is shown in Figure 3.3.

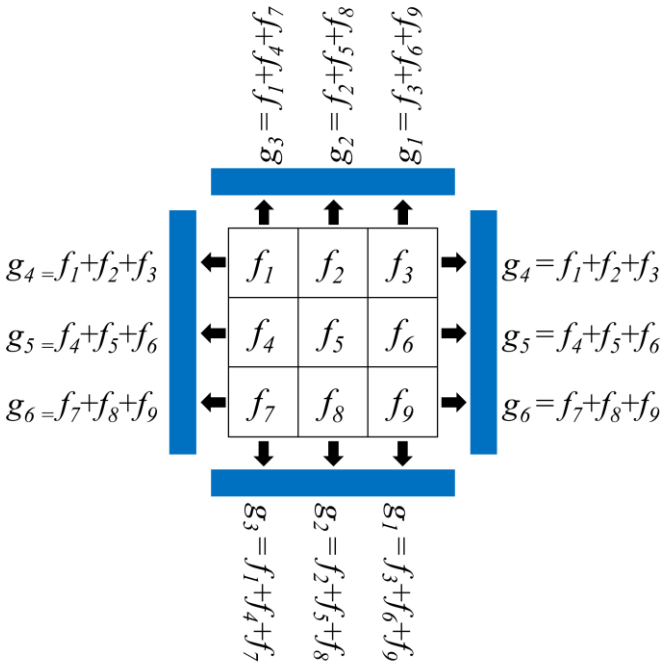

Figure 3.3. The projections for a 3 × 3 image assuming only 6 LORs are allowed.

The relationship between sinogram bins g_i and image pixels f_j for this example is

$$\begin{cases} g_1 = f_3 + f_6 + f_9 \\ g_2 = f_2 + f_5 + f_8 \\ g_3 = f_1 + f_4 + f_7 \\ g_4 = f_1 + f_2 + f_3 \\ g_5 = f_4 + f_5 + f_6 \\ g_6 = f_7 + f_8 + f_9. \end{cases} \tag{3.6}$$

A system matrix representing the set of relationships shown in equation (3.6) can then be shown as

$$
\begin{array}{cc}
 & \begin{matrix} f \Rightarrow 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \quad g \Downarrow \\
A = & \begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix}
\end{array}
\tag{3.7}
$$

in which each row of the system matrix corresponds to a sinogram bin and each column corresponds to an image pixel. Therefore, $g = Af$ generates the projections of any image f. In more realistic cases, the values of a_{ij} may be any real value between 0 and 1. For example, in an accurately modelled system matrix, the probability of detecting a coincidence that travels a long distance within the PET detectors should be different from that of a coincidence that travels a short distance in the detectors. The reconstruction algorithms explained in the next sections use the system matrix to perform the projection operator. Therefore, an accurate determination of the elements of the system matrix (a_{ij}) is crucial for a precise reconstruction. For now, an assumption is made that the system matrix is carefully calculated and that the effects of attenuation, normalization and detector response are accounted for in the system matrix. Theoretically, the image f can be directly calculated by the equation $f = A^{-1}g$, where A^{-1} is the inverse matrix of A. In practice, this direct inversion is not practical due to one or more of the following problems [180]

1. Solving the equation $f = A^{-1}g$, even for a very low-resolution image, is computationally intensive,
2. A^{-1} may not exist,
3. A^{-1} may not be unique, and

4. A^{-1} may be ill-conditioned, meaning that small changes in the sinogram data *g* may produce large differences in the results *f*.

Various optimization methods have been proposed to find the best estimate for *f* given *g*. *Maximum likelihood expectation maximization* (ML-EM) is an algorithm that aims to maximize the likelihood of the reconstructed image in an iterative fashion as described below.

3.2.1 **Maximum Likelihood Expectation Maximization**

The goal of an iterative statistical reconstruction technique is to find the mean number of disintegrations $\bar{f}$ in the image that can produce the sinogram *g* with the highest likelihood. Assume $\bar{f}_j$ is the mean number of disintegrations in pixel *j* in an image with *m* pixels in total. The mean number of events $\bar{g}_i$ detected by sinogram bin *i* is the sum of the mean number of decays emitted from each pixel $\bar{f}_j$, which is

$$\bar{g}_i = \sum_{j=1}^{m} a_{ij} \bar{f}_j. \tag{3.8}$$

Measurements in PET are subject to variations due to the Poisson nature of the radioactive disintegrations. So, the number of photons emitted from the *m* pixels and detected by bin *i* is a Poisson variable. The probability of detecting g_i photons is then given by

$$P(g_i) = \frac{e^{-\bar{g}_i} \bar{g}_i^{g_i}}{g_i!} \tag{3.9}$$

where ! denotes the factorial operator. The *i* Poisson variables are independent. The conditional probability $P(g \mid \bar{f})$ of observing the vector *g* (projection) when the emission map is $\bar{f}$ is the

product of the individual probabilities $P(g_i)$. To define this conditional probability, a likelihood function $L(\bar{f})$ is defined by

$$\begin{aligned} L(\bar{f}) &= P(g \mid \bar{f}) \\ &= P(g_1)P(g_2)...P(g_n) \\ &= \prod_{i=1}^{n} P(g_i) \\ &= \prod_{i=1}^{n} \frac{e^{-\bar{g}_i}\bar{g}_i^{g_i}}{g_i!}. \end{aligned} \tag{3.10}$$

The highest value for the likelihood $L(\bar{f})$ (also maximum likelihood estimate of $\bar{f}$) can be found by computing the derivative of $L(\bar{f})$. However, it is easier to work with the log of the likelihood function. Considering $l(f) = \ln(L(f))$, where ln is the natural log, equation (3.10) can be written as

$$l(\bar{f}) = \sum_{i=1}^{n}(-\bar{g}_i + g_i \ln(\bar{g}_i) - \ln(g_i!)). \tag{3.11}$$

Substituting $\bar{g}_i$ in equation (3.11) with equation (3.8) yields

$$l(\bar{f}) = \sum_{i=1}^{n}\left(-\sum_{j=1}^{m} a_{ij}\bar{f}_j + g_i \ln\left(\sum_{j=1}^{m} a_{ij}\bar{f}_j \right) - \ln(g_i!) \right). \tag{3.12}$$

Equation (3.12) is the basis of many statistical reconstruction techniques, such as ML-EM. This equation calculates the probability of observing a projection dataset for any mean image $\bar{f}$. The aim in ML-EM is to find an image that has the highest probability to yield g. So, the vector $\bar{f}$ for which $l(\bar{f})$ is maximal is considered as the best estimate for the solution. It has been proved

[180] that $l(\bar{f})$ has 1 and only 1 maximum which can be found by setting the derivative of $l(\bar{f})$ to zero

$$\frac{\partial l(\bar{f})}{\partial \bar{f}_j} = -\sum_{i=1}^{n} a_{ij} + \sum_{i=1}^{n} \frac{g_i}{\sum_{j'=1}^{m} a_{ij'} \bar{f}_{j'}} a_{ij} = 0 \tag{3.13}$$

which can also be written as

$$\bar{f}_j \frac{\partial l(\bar{f})}{\partial \bar{f}_j} = -\bar{f}_j \sum_{i=1}^{n} a_{ij} + \bar{f}_j \sum_{i=1}^{n} \frac{g_i}{\sum_{j'=1}^{m} a_{ij'} \bar{f}_{j'}} a_{ij} = 0, \tag{3.14}$$

and therefore

$$\bar{f}_j = \frac{\bar{f}_j}{\sum_{i=1}^{n} a_{ij}} \sum_{i=1}^{n} \frac{g_i}{\sum_{j'=1}^{m} a_{ij'} \bar{f}_{j'}} a_{ij}. \tag{3.15}$$

Equation (3.15) leads to the iterative scheme of the ML-EM algorithm as proposed by [181]

$$\bar{f}_j^{(k+1)} = \frac{\bar{f}_j^{(k)}}{\sum_{i=1}^{n} a_{ij}} \sum_{i=1}^{n} \frac{g_i}{\sum_{j'=1}^{m} a_{ij'} \bar{f}_{j'}^{(k)}} a_{ij} \tag{3.16}$$

in which *k* is the iteration number.

The term $\sum_{j'=1}^{m} a_{ij'} \bar{f}_{j'}^{(k)}$ is the current estimate of the mean number of counts in sinogram bin *i* which represents the projection operator. The ratio of g_i to the forward projected value for sinogram bin *n* is the ratio of the measured number of counts to the current estimate of the mean number of counts for that sinogram bin. The term $\sum_{i=1}^{n} (g_i / \sum_{j'=1}^{m} a_{ij'} \bar{f}_{j'}^{(k)}) a_{ij}$ represents the

backprojection operator for pixel *j*. Therefore, as suggested by [180], equation (3.16) for the entire image can conceptually be seen as

$$\text{Image}^{(k+1)} = \text{Image}^{(k)} \times \text{Normalized backprojection of} \left(\frac{\text{Measured projections}}{\text{Projections of image}^{(k)}} \right). \quad (3.17)$$

When dealing with a list of the recorded coincidence (*list-mode*), equation (3.17) can be reformulated as

$$\overline{f}_j^{(k+1)} = \frac{\overline{f}_j^{(k)}}{\sum_{i=1}^{n} a_{ij}} \sum_{i=1}^{N} \frac{1}{\sum_{j'=1}^{m} a_{ij'} \overline{f}_{j'}^{(k)}} a_{ij} \quad (3.18)$$

in which *N* is the total number of recorded events. Equation (3.18) is the basis of list-mode ML-EM.

3.2.2 **Ordered-subset Expectation Maximization**

Convergence in the ML-EM algorithm may require a large number of iterations which can make the reconstruction algorithm slow. Hudson and Larkin [182] proposed the use of *subsets* (or *blocks*) in ML-EM to accelerate the reconstruction algorithm by introducing *ordered-subsets expectation maximization* (OS-EM). To accelerate convergence in OS-EM, the acquired projections are distributed and the ML-EM algorithm is applied to each subset. The OS-EM algorithm is widely used in clinical practice nowadays due to its convenience and high image quality. This algorithm is mathematically defined as

$$\overline{f}_j^{k,(l+1)} = \frac{\overline{f}_j^{k,l}}{\sum_{i=1}^{n} a_{ij}} \sum_{i \in T^l} \frac{g_i}{\sum_{j'=1}^{m} a_{ij'} \overline{f}_{j'}^{k,l}} a_{ij} \quad (3.19)$$

where k and l represent the iteration and subset number, respectively, and T^l is the lth subset of the sinogram bins that are used to update the image in subset l. Similar to the list-mode ML-EM formulation in equation (3.18), a list-mode OS-EM is given by

$$\overline{f}_j^{k,(l+1)} = \frac{\overline{f}_j^{k,l}}{\sum_{i=1}^{n} a_{ij}} \sum_{i \in S^l} \frac{1}{\sum_{j'=1}^{m} a_{ij'} \overline{f}_{j'}^{k,l}} a_{ij} \tag{3.20}$$

in which S^l represents the list of coincidences to be used in subset l.

3.2.3 **Maximum a Posteriori**

Many models of physical effects occurring during image acquisition can be incorporated into an iterative reconstruction algorithm such as ML-EM. As one can expect, the images obtained with ML-EM are more appealing than the images obtained with a conventional analytical reconstruction. Nonetheless, the reconstructed images tend to become noisy as the number of iterations increases. Thus, various algorithms were proposed that while maintaining the estimated projections as close as possible to measured projections, minimize the image noise. The introduction of prior knowledge as a constraint that may favor convergence of the EM process is called *regularization*. For convenience purposes, $\overline{f}$ and $\overline{g}$ are replaced with f and g in the following equations, such that f and g refer to the image and sinogram domain, respectively.

According to Bayesian theorem [183], the prior knowledge about the reconstructed image f can be introduced in the EM algorithm such that

$$P(f \mid g) = \frac{P(g \mid f)P(f)}{P(g)} \tag{3.21}$$

where $P(g \mid f)$ is the likelihood function defined in equation (3.10), $P(f)$ is the prior function that contains the a priori knowledge of the image, $P(g)$ is the *a priori* probability distribution of the measurements and $P(f \mid g)$ is the *a posteriori* probability distribution of the image. In Bayesian image reconstruction, the *maximum a posteriori* (MAP) estimate of the image is found by maximizing the posterior probability density function

$$f = \arg\max_{f \geq 0} \left\{ \ln P(g \mid f) + \ln P(f) - \ln P(g) \right\}. \tag{3.22}$$

It should be noted that $P(g)$ is the a priori probability distribution of the measurements, independent of f and essentially constant once the data are acquired. Therefore, the term $\ln P(g)$ can be withdrawn from equation (3.22), which gives

$$f = \arg\max_{f \geq 0} \left\{ \ln P(g \mid f) + \ln P(f) \right\}. \tag{3.23}$$

Equation (3.23) can be interpreted as maximizing a penalized likelihood function where the additive penalty is $\ln P(f)$, hence the term *penalized ML*. As stated by [184], there are differences in MAP and penalized ML from a statistical point of view, however MAP and penalized ML are considered equivalent in many PET image reconstruction contexts. Various approaches have been proposed to form the prior distribution for the image $P(f)$. The most common framework is done by using the *Markov random field* (MRF) prior [185]. The MRF is a probabilistic image model with properties that allow one to readily specify the joint distribution of images (within a scale factor) in terms of local *potentials* that describe the interactions between groups of voxels that are mutual neighbors [184]. MRF models are practically used via a theorem stating the equivalence between MRF's and *Gibbs distribution* [186]. The MRF-Gibbs equivalence theorem states that the joint density of an MRF is a Gibbs distribution. The Gibbs distribution allows us to specify a prior in terms of local interactions since all MRFs have the property that the

conditional probability of any voxel in the image depends only on the values of the voxels in a local neighborhood of that voxel. The joint density in the form of a Gibbs distribution is

$$P(f) = \frac{1}{Z} e^{-\beta U(f)} \tag{3.24}$$

where Z is a normalization constant, β is a parameter that determines the degree of smoothness of the image, and $U(f)$ is the *Gibbs energy function* which is formed as the sum

$$U(f) = \sum_{c \in \mathbb{C}} V_c(f) \tag{3.25}$$

in which $V_c(f)$ is a set of potential functions, each defined on a subset of voxels called a *clique*. A clique $c \in \mathbb{C}$ ('clique' comes from a basic concept in graph theory - In this context, cliques are composed of a series of voxels that are common neighbors) consists of one or more voxels, all of which are mutual neighbors of each other where $\mathbb{C}$ is the set of all possible cliques.

Replacing equation (3.24) into equation (3.23) and dropping the normalization factor Z gives

$$f = \arg\max_{f \geq 0} \left\{ \ln P(g \mid f) - \beta U(f) \right\} \tag{3.26}$$

where the parameter β is known as the regularization parameter and the Gibbs energy function $U(f)$ is commonly known as a *regularizer* or *prior*.

A maximization process similar to the derivation of equation (3.16) leads to an iterative scheme, known as the *one-step-late* (OSL) algorithm, originally described by [187]

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{\sum_{i=1}^{n} a_{ij} + \beta \frac{\partial}{\partial f_j} U\left(f_j^{(k)}\right)\Big|_{f=f^{(k)}}} \sum_{i=1}^{n} \frac{g_i}{\sum_{j'=1}^{m} a_{ij'} f_{j'}^{(k)}} a_{ij} \ . \tag{3.27}$$

The algorithm described in equation (3.27) is known as the MAP-EM reconstruction algorithm using the OSL approach.

One of the most commonly used simple prior terms that has been introduced is the quadratic prior [180, 185] in which the derivative of the energy function $U(f)$ is defined as

$$\frac{\partial}{\partial f_j^{(k)}} U\left(f_j^{(k)}\right) = \sum_{c \in \mathbb{C}_j} w_{jc}\left(f_j^{(k)} - f_c^{(k)}\right) \tag{3.28}$$

where $\mathbb{C}_j$ is a clique of voxels in the neighborhood of pixel j and w_{jc} is a weight indicating the strength of neighborliness between pixels j and c such that w_{ic} is 1 if j and c are orthogonally nearest neighbors, $1/\sqrt{2}$ for diagonal nearest neighbors and 0 otherwise.

3.3 A Review of GPU-based Reconstruction Algorithms in PET

The use of GPUs in image reconstruction, and specifically list-mode PET image reconstruction, is becoming more popular. A brief technical overview of a programmable GPU is given in Sections 3.3.1 and 3.3.2 followed by an explanation on multi-GPU programming in Section 3.3.3. The explanation in this chapter is focused on the use of GPUs in image reconstruction. More general theoretical and practical programming guidelines can be found in several references cited below and specifically reference [188]. A relevant literature review on previous effort to implement GPU-compatible PET list-mode image reconstruction techniques is done in Sections 3.3.4 and 3.3.5.

3.3.1 **GPU Architecture Model**

Programmable GPUs developed by *Nvidia* are categorized into different generations according to their *microarchitecture* and *compute capabilities*. The GPUs employed in the image

reconstruction step of this dissertation were two *Nvidia GeForce GTX Titan Black* cards. A Titan Black graphic card belongs to the *Kepler* class microarchitecture and has a compute capability of *3.5*. As the programming model used to implement the image reconstruction algorithm was *Compute Unified Device Architecture* (CUDA), an overview of the related CUDA properties of the used graphic cards are shown in Table 3.1.

Table 3.1. Selected properties of Nvidia GeForce GTX Titan Black. Properties obtained from the CUDA *Device Query* tool provided by Nvidia.

Total amount of global memory	6084 MBytes
Number of CUDA cores	2880
Number of multiprocessors	15
CUDA cores per multiprocessor	192
GPU clock rate	889 – 980 MHz
Maximum texture dimension (1D)	65536
Total amount of constant memory	65536 bytes
Total amount of shared memory per block	49152 bytes
Warp size	32
Maximum number of threads per block	1024

In order to develop an efficient GPU-based image reconstruction, understanding the CUDA execution model is essential. The CUDA execution model is vastly different from parallel execution through multithreading of a central processing unit (CPU). The type of parallelism GPUs offer can be best described by the term *single instruction, multiple thread* (SIMT). A few basic definitions are required to explain how SIMT works.

The terms *host* and *device* are used to refer to the CPU and the GPU and their associated memories, respectively. A CUDA function is often referred to as a *kernel*. When the host launches a kernel, the kernel execution takes place on a number of *threads*. A *streaming multiprocessor*

(SM) is a scalable array of processors designed to support concurrent kernel execution of several threads. There exists a two-level thread hierarchy to organize the threads. The threads are grouped into *blocks of threads* and the thread blocks are grouped into *grids of blocks* with scalable dimensions as shown in Figure 3.4.

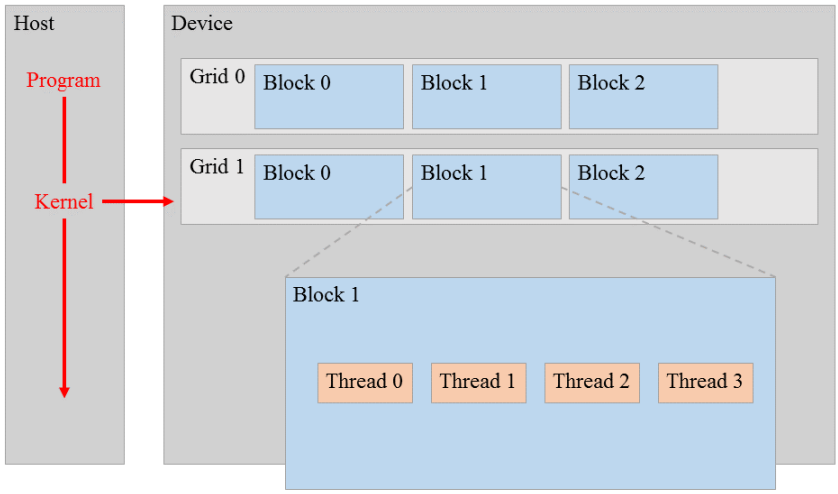

Figure 3.4. Categorization of threads into thread blocks and blocks into grids (inspired by Figure 2-5 of [189]).

When a grid of thread blocks is launched, the thread blocks are distributed among different SMs. Threads assigned to an SM are further grouped into *warps*, each consisting of 32 consecutive threads. The *warp scheduler* is in charge of distributing thread blocks into warps and the *instruction dispatch unit* sends instructions to the threads grouped into a warp. All threads in a warp are essentially executed following the SIMT model. According to the SIMT execution model, all threads within a warp must execute the same instructions, otherwise *thread divergence* occurs. When threads of a warp diverge, serial execution of threads (as opposed to parallel execution) occurs.

As shown in Table 3.1, each SM in a Kepler GPU has 192 CUDA cores. An SM in Kepler devices also has 4 warp schedulers, 8 instruction dispatch units, 32 special function units (SFU),

32 load/store units (LD/ST), 65536 registers and dedicated shared and texture memories as shown in Figure 3.5.

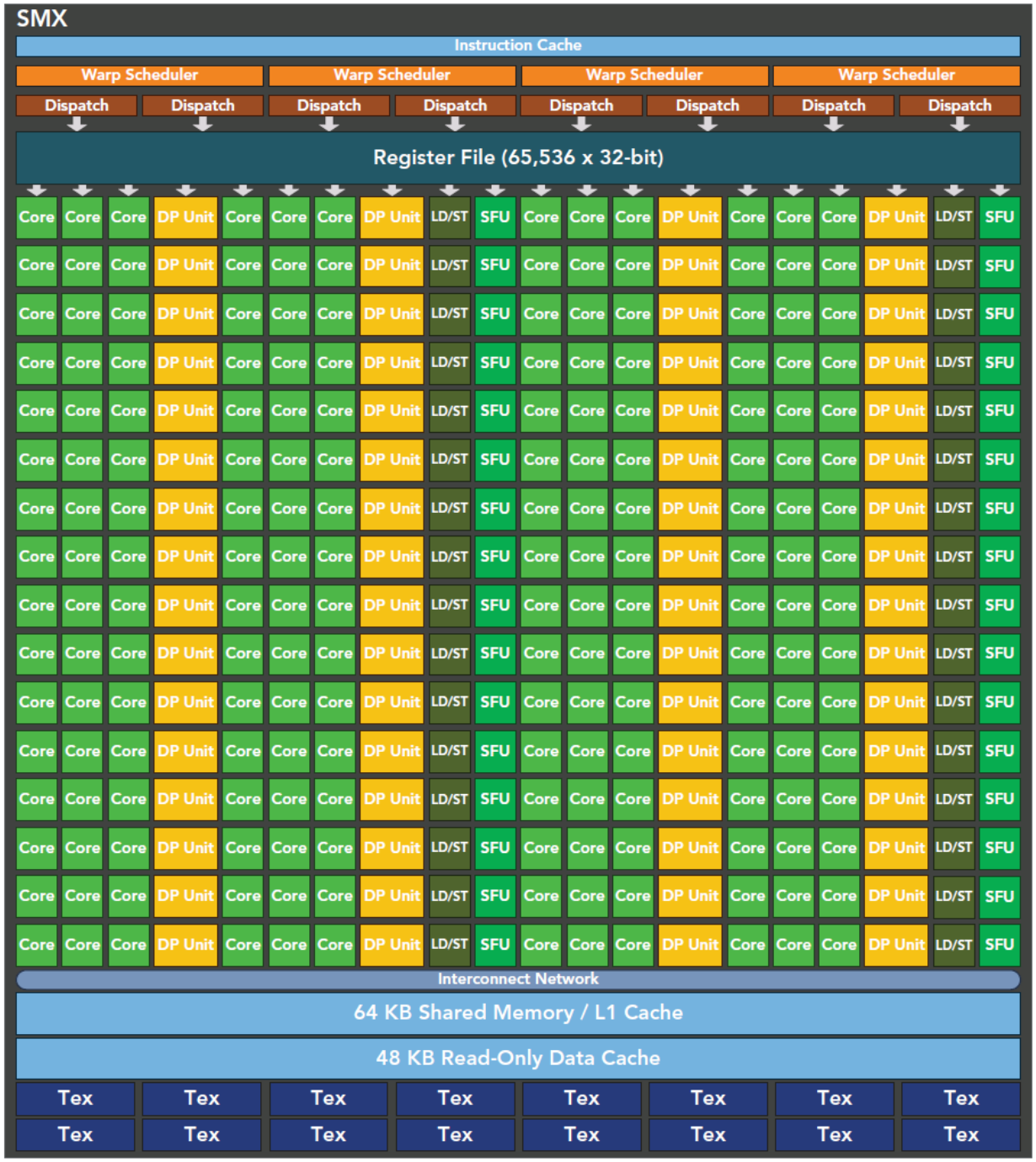

Figure 3.5. Structure of a SM in a Kepler GPU (Figure 3-7 of [189]. Image reproduced with permission of the right holder, John Wiley & Sons, Inc.).

3.3.2 **GPU Memory Layout**

The efficient use of memory hierarchy in a GPU is crucial to maintain the performance of a CUDA program. Therefore, the available memories in a Titan Black are reviewed in this section. It should be noted that, unlike the CPU memory hierarchy in which cache memories are not programmable, several programmable memory models are offered by the CUDA memory model. As a general rule, smaller memories have faster access from the threads.

Registers (see Figure 3.5) are the smallest and fastest memory on a GPU. Each thread has private access to registers and a kernel holds frequently accessed variables in a register. There is a hardware limit of 255 registers per thread in a Titan Black. Variables that cannot fit into the available registers (due to the lack of space) spill into *local memories*. Local memories are not shown in Figure 3.5 as they do not physically exist. Local memories are in fact a part of the global memory that is employed by a thread to act as its own dedicated memory. *Shared memory* is an on-chip memory with extremely low latency and high bandwidth. As all threads within the same block have access to the same shared memory address, shared memory can be employed as an efficient way of inter-thread communications. Shared memory has read/write ability, however, this memory is limited to 48 KB. *Constant memory* is a limited read-only memory with a cap of 64 KB (different than the shown read-only data cache in Figure 3.5). Variables located in constant memory are accessible by all threads and in fact all kernels at all time. *Texture memory* is another dedicated read-only memory that can be accessed by all threads. Texture memory uses a previously defined texture and is optimized for spatial locality. It also has the unique ability of interpolation that allows for a more accurate on-the-fly calculation of texture objects. Lastly, *global memory* is the largest memory available on a GPU that also has the highest latency. Similar to constant memory, variables defined on global memory last during the lifetime of the application [189].

3.3.3 **Multi-GPU Programming**

Multiple GPUs can be used to increase the throughput of a CUDA program. Since in this dissertation multiple GPUs were connected over the same *peripheral component interconnect express* (PCIe) bus in a single node, this method of employing multiple GPUs is discussed in this section.

GPUs connected to the same PCIe bus can have access to the data located on the global memory of other GPUs. Inter-GPU (*inter-node*) communications can be managed such that the task and/or data are split between the devices. The CUDA *peer-to-peer* (P2P) application programming interface (API) enables direct inter-node communications. More specifically, the CUDA P2P API allows for two methods of communication: P2P access and P2P transfer. Through P2P access, different GPUs can load and store data across the GPUs within a CUDA kernel while through a P2P transfer, data can be copied between the GPUs.

3.3.4 **GPUs in PET Image Reconstruction**

The most popular GPU computing API in PET image reconstruction is CUDA as it provides a convenient C/C++ extension. While some GPU-based image reconstruction algorithms were developed with APIs other than CUDA (such as OpenGL), the majority of current GPU-based image reconstruction algorithms benefit from CUDA libraries. System matrix elements are usually found on-the-fly for each sinogram bin or LOR assuming a pre-defined shift-invariant response function for the PET scanner. Although in studies such as [190] and [191], shift-variant functions have been studied to model the detector response function, the use of shift-invariant functions are still preferred due to its high computability with CUDA.

Forward and backprojection operators in an iterative image reconstruction algorithm can be performed in two different manners: *voxel-driven*, in which the contribution of all voxels to

each line (LOR) is found, or *line-driven*, in which the contribution of each line (LOR) to all voxels is found (note that this is only a matter of implementation and there are no differences in reconstructed images) [192]. An approach was suggested in [193] for ToF PET list-mode image reconstruction that is the basis of the developed list-mode image reconstruction algorithm in this work as discussed in Chapter 7 and Chapter 8. The proposed list-mode image reconstruction algorithm in [193] employed a line-driven approach for both the forward and backprojection operators such that no further CPU-GPU communication is required once the reconstruction begins.

3.3.5 **Distributed ML-EM**

Dividing the workload of an iterative image reconstruction algorithm to multiple-GPUs can potentially lead to yet another level of parallelism. The theoretical gain that can be achieved by using multiple GPUs is in fact limited by the amount of required inter-node communications according to *Amdahl's law* [194]. List-mode ML-EM, as expressed by equation (3.18), was implemented on multiple GPUs to quantify the gain one can achieve by employing multiple GPUs in Cui *et al.* [195]. It was seen that a simple implementation of standard list-mode ML-EM, which was referred to as the *distributed gradient* (DG) algorithm in [195], is only optimum when up to two GPUs are used and the use of more GPUs may require a reformulation of the list-mode ML-EM. The reformulated algorithm, called *distributed ML-EM* (DML-EM), demonstrated a large speed-up gain for up to about 10 GPUs. The details of the reformulation in DML-EM are not discussed in this dissertation as the algorithm is not used. Nevertheless, Figure 3.6 shows the graph obtained by Cui *et al*. [195] comparing the performance of DML-EM against DG.

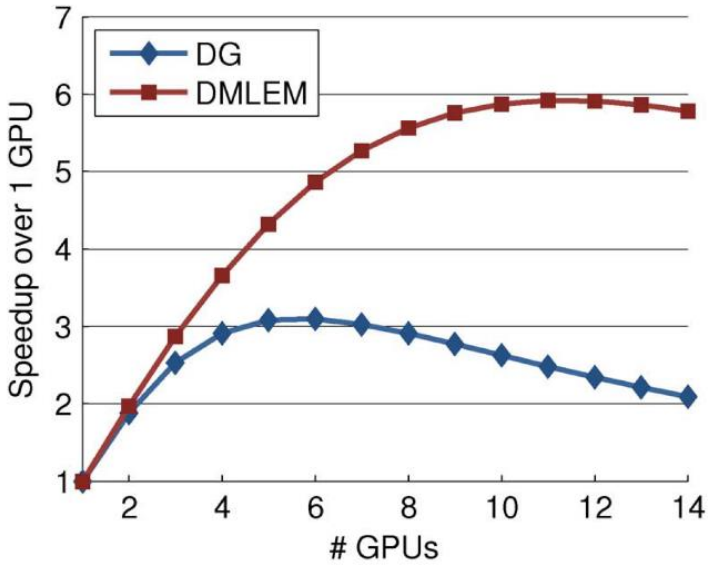

Figure 3.6. Performance comparison of multi-GPU DML-EM and DG with respect to a single GPU (© 2013 IEEE [195]).

Based on the results of this study, the DG algorithm was chosen in the dual-GPU implementation of list-mode ML-EM as explained in Chapter 7 and Chapter 8.

Chapter 4 : Detector Geometry Optimization for a Brain PET System with Fixed Detector Area

Reuse Permission

4.1 Introduction

The current chapter is the first part of 3 chapters that investigate design and performance optimization of detectors for a brain-dedicated PET insert. The geometry optimization project in this chapter has been conducted assuming a fixed detector area for DLO detectors suitable for a brain-dedicated PET insert. Later in Chapter 5, the effects of detector area on the performance of PET detectors and in Chapter 6, the effects of layer misassignment due to inter-crystal scatter in DLO detectors will be studied in detail.

The objective of the presented work in this chapter was to design an *optimum* DLO detector for a brain-dedicated PET insert given a predefined detector area. Optimization of a detector can be interpreted differently depending on what criteria are considered as *optimum*. The chosen MR scanner for the PET insert is discussed in Section 4.2 while Section 4.3 includes an overview of the fixed and variable geometrical properties considered for the detectors in this optimization study. Monte Carlo simulations were used to evaluate the performance of various detectors. The details of the MC simulation platform are explained in Section 4.4 followed by a description of the algorithms to generate *singles* and *coincidences* in Sections 4.5 and 4.6, respectively. The method of processing data from MC simulations was compared against measurements in Section 4.7 to validate the simulations and singles and coincidence generation algorithms. Finally,

the details of the chosen detector geometry optimization method are fully discussed in Sections 4.8 and 4.9.

4.2 Geometrical Properties of the Chosen MR Scanner

The chosen MR scanner is the *Siemens Magnetom 7 T brain MR scanner* located at the Lawson Imaging Centre in London, Ontario. This scanner is a state-of-the-art research-based ultra-high field MR scanner that can be used for human brain or animal imaging. The inner bore diameter of the MR scanner is 360 mm [197]. The bore diameter of the MR scanner and room required for an MR coil set a limit on the size of the PET insert.

4.3 Geometrical Parameters of PET Detectors

A series of geometrical assumptions and constraints were considered for the PET scanner that are discussed in the following sections. It should be noted that the fixed geometrical assumptions explained in Section 4.3.1 are specific to the current chapter and are further varied and studied later in Chapter 5.

4.3.1 **Fixed Geometrical Properties**

The detector block size assumes that each detector uses a 4 × 4 group of SensL ArrayC SiPMs (SensL Inc., Cork, Ireland). The size of each SiPM is 7.1 × 7.1 mm^2 which provides a detector area of about 8 cm^2. Considering 7 mm room for the SiPM board thickness and detector casing, the PET insert accommodates 38 detector blocks per ring. Assuming 6 rings of detectors, the axial FOV will be 170.4 mm, allowing imaging of the entire human brain in a single FOV. The choice of crystal material is LYSO ($Lu_{1.79}Y_{0.21}SiO_5$) with a mixing fraction of 0.105 yttrium to lutetium and single-ended DLO detector architecture, as shown in Figure 2.18-(F), is to be used in the detector design.

4.3.2 **Geometrical Variables**

A schematic drawing of a DLO detector block as used in this study is shown in Figure 4.1.

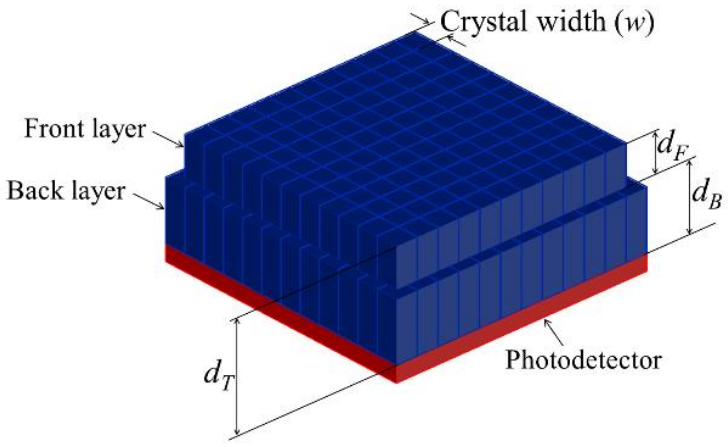

Figure 4.1. Drawing of a DLO detector block with total thickness d_T, front/back layer thickness d_F/d_B and crystal width w (© 2018 IEEE [99]).

Different detector arrangements were considered by varying the geometrical parameters of the PET detector including the total detector thickness d_T, the front/back layer thickness d_F/d_B and crystal with w. As explained in the sections below, a total of 5 (number of d_T) × 5 (number of w) × 9 (number of layer ratios for DLO detectors) = 225 possible DLO detector geometries were studied. The selected numerical values for each geometrical parameter and their impact on the PET scanner are mentioned in the following sections.

4.3.2.1 Detector total thickness

Detectors with d_T values of 10, 15, 20, 25 and 30 mm were considered in this section. A minimum thickness of 10 mm was chosen assuming that a detector thinner than 10 mm does not have sufficient detection efficiency required for a clinical human PET scanner. The maximum thickness was selected to be 30 mm since detectors thicker than this reduce the inner diameter of the system to less than 280 mm given the fixed outer diameter of the PET insert. It should also be

noted that thicker detectors require a higher amount of scintillation material which increases the cost of the scanner.

4.3.2.2 Layer thickness ratio

Layer thickness ratio determines the DOI accuracy and detection efficiency in each layer. For each total thickness d_T, the thickness of the front layer d_F was varied in increments of $0.1 \times d_T$ in the range of $0.1 \times d_T$ to $1.0 \times d_T$. In each case, the thickness of the back layer d_B is given by $d_T - d_F$. A front layer thickness of $1.0 \times d_T$ represents a typical single layer detector which does not provide any DOI information. This case was also considered for comparison purposes between the performances of DLO detectors with their equivalent single layer detector.

4.3.2.3 Crystal pitch

Block detectors with ($n \times n$) crystals in the front layer and (n+1) × (n+1) crystals in the back layer were considered in this work. The studied number of crystals in this work were (6×6) + (7×7), (8×8) + (9×9), (11×11) + (12×12), (23×23) + (24×24) and (39×39) + (40×40). These designs correspond to a *nominal crystal width* (w) of 4, 3, 2, 1 and 0.5 mm, respectively. A gap of 80 μm was assumed to exist between adjacent crystals due to the reflector surrounding around each crystal. The term *nominal crystal width* is used in this work, since for each geometry, depending on the number of crystals n, total detector thickness and the thickness ratio between the layers, the actual crystal pitch was calculated such that the gap between adjacent detector blocks is zero (i.e. adjacent detector blocks just touch). In this work, the terms *crystal width*, *crystal pitch* and *nominal crystal pitch* are used interchangeably. Table 4.1 shows a summary of the exact crystal widths used for each detector geometry.

Table 4.1. Minimum, maximum, average and standard deviation (SD) of the exact crystal widths used for various detector geometries.

Number of crystals $(n\times n) + (n+1)\times(n+1)$	Nominal crystal width (w)	Minimum/maximum of exact crystal widths	Average ± SD of exact crystal widths
(6×6) + (7×7)	4 mm	3.418 mm/4.057 mm	3.839 ± 0.160 mm
(8×8) + (9×9)	3 mm	2.659 mm/3.156 mm	2.973 ± 0.121 mm
(11×11) + (12×12)	2 mm	1.994 mm/2.367 mm	2.213 ± 0.091 mm
(23×23) + (24×24)	1 mm	0.997 mm/1.163 mm	1.084 ± 0.049 mm
(39×39) + (40×40)	0.5 mm	0.598 mm/0.686 mm	0.642 ± 0.030 mm

A minimum of 0.5 mm crystal width was chosen since in practice this is at the limit of being reliably constructed and read out and because the *packing fraction* (the ratio of the scintillator volume in a detector block to entire volume of the detector) becomes less than 0.75 for smaller crystals. Figure 4.2 shows the packing fraction of detectors with various total thicknesses and crystal arrangements. Since packing fraction is not a function of layer thickness ratio, a d_B/d_T ratio of 0.5 was used in this figure.

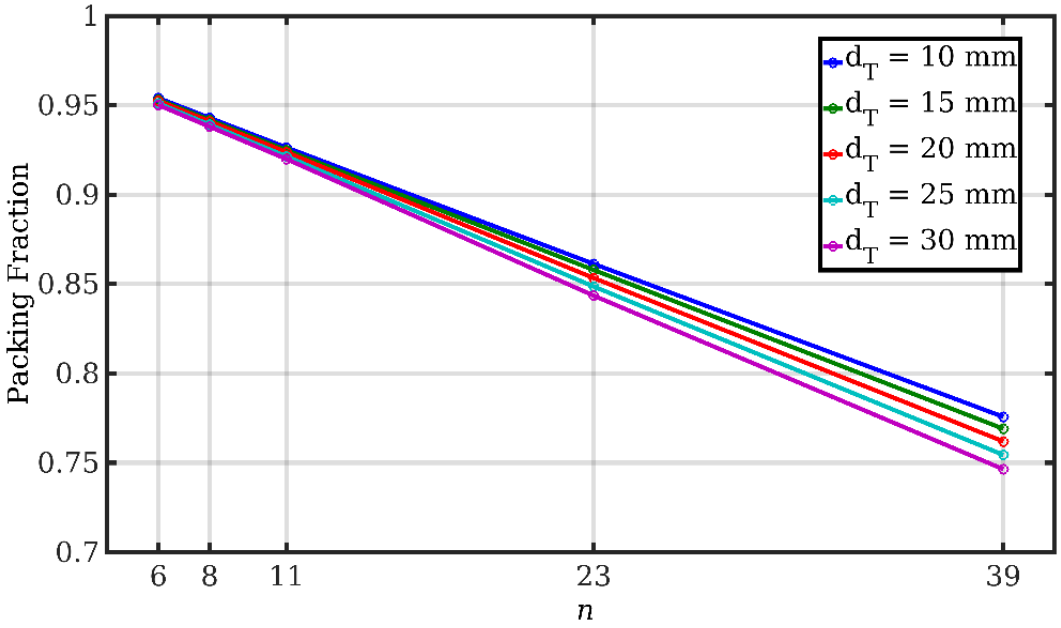

Figure 4.2. Packing fraction for DLO detectors with various total thicknesses, a d_B/d_T ratio of 0.5 and $(n+1) \times (n+1) + n \times n$ crystal elements per block.

4.4 Monte Carlo Simulation Platform

The performance of detectors was evaluated through MC studies using GATE v7.1 [142, 143]. The ratio of lutetium to yttrium in the default LYSO in the GATE material library was updated such that its properties follow that of the LYSO considered in this work as explained in Section 4.3.1. Optical photons and electronic signal processing effects, such as dead-time, were not modeled in the work explained in this chapter. A list of all generated *hits* within the detectors was extracted (with no blurring on energy or timing information) from all simulations and transferred to an in-house algorithm to generate *singles* (*events*) and *coincidences*. Note that the same terminology for hits, singles (events) and coincidences has been used in Chapter 5 and Chapter 6 as well.

4.5 Singles Generation Algorithm

An *event* was defined as the energy weighted centroid of corresponding photon hits on the XZ plane (see Figure 4.3 for axis definition). Events were generated from the list of recorded hits and blur the deposited energy by assuming an energy resolution of 12.5% at 511 keV [45]. If the blurred deposited energy of the event was within an energy window of 350 to 650 keV, the event was saved in a list mode format for further processing. For each accepted event, XYZ coordinates were defined based on the centroid of the event's energy deposited using a crystal lookup table (LUT) in a manner similar to defining a crystal of interaction using a detector flood image map. It is important to note that due to the interleaved pattern of the crystal layers in flood images for a DLO detector, it is possible that the event is mispositioned into the wrong layer based on the energy-weighted centroid which will be further discussed in Chapter 6. Figure 4.3 shows an example of a single generated from two hits in a DLO detector.

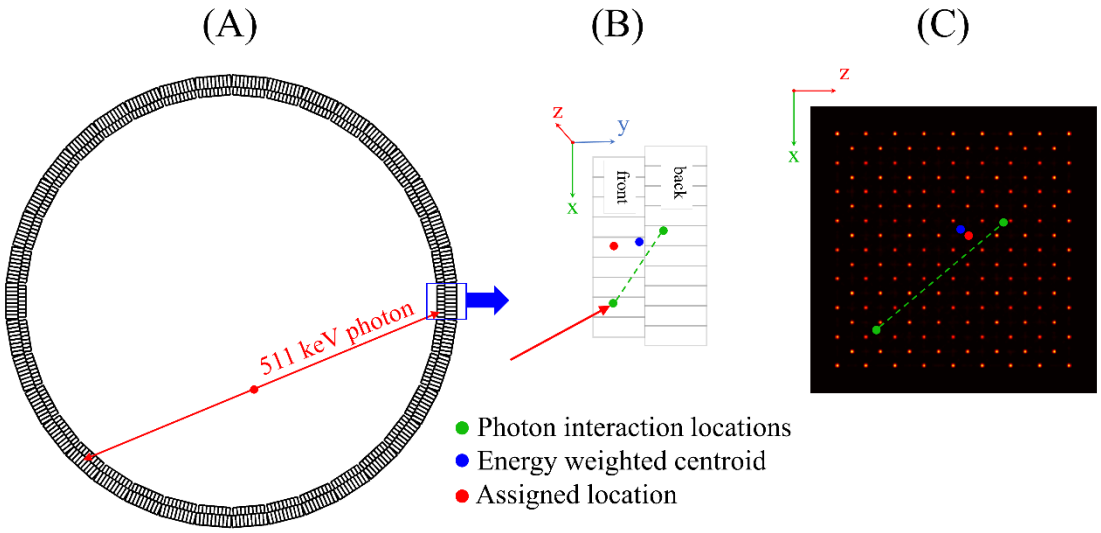

Figure 4.3. (A) Drawing of a sample PET scanner and (B) one of its detectors with a single generated from two hits. Image (C) shows the locations of the hits and assigned location on the flood map.

4.6 Coincidence Generation Algorithm

For studies in which coincidences were analyzed, the previously generated list of singles was used to extract coincidences assuming a CTW of 10 ns. Also, the exact timing information of the singles was blurred assuming a timing resolution of 2.5 ns. In this work, a constant value for the energy and timing resolutions of all studied detectors was assumed. Also, no threshold for the required block separation was assumed (i.e. no minimum block difference). In all studies, event rates were kept sufficiently low to avoid any effects due to pileup or dead-time. Multiple coincidences were discarded such that a policy '*KillAll*' (per GATE's terminology) is implemented.

4.7 Simulation Validation

In order to validate the MC simulations, *coincidence response functions* (CRFs) of a small animal PET insert with DLO detectors available in our group were used. The available small animal PET scanner is built of one ring of detectors with 16 detector blocks per ring. The scintillation material is LYSO:Ce and each detector block is comprised of (21×9)/(22×10) crystals

with a crystal pitch of 1.27 mm and 4/6 mm layer thicknesses in the front/back layer. The inner diameter of the small animal PET scanner is 65.8 mm. Further details of this scanner can be found in [45, 48].

An experiment was conducted in which a spherical ^{22}Na source with a diameter of 0.25 mm embedded in a $10 \times 10 \times 10$ mm^3 plastic cube was stepped in increments of 1/8th of the crystal pitch (0.15875 mm). Data were acquired for 60 seconds per source step with an energy window of $300 - 800$ keV. CRFs of the front and back layer crystals of two detector pairs in the two layers were separately measured as shown in Figure 4.4 and compared with the results obtained from a MC simulation study.

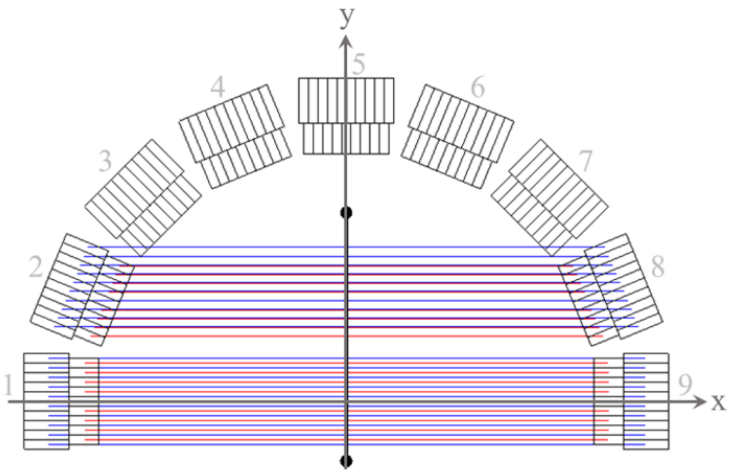

Figure 4.4. Schematics of the DLO PET scanner used in MC simulation validation experiment. A point source was stepped between the two bold points on the y-axis and the number of coincidences in each pair of opposite crystals of the front (red lines) and back layer (blue lines) was recorded at each step (© 2018 IEEE [99]).

The method of finding CRFs from the simulated data is explained in detail in Section 4.8.2.1. Figure 4.5-(A) illustrates the response functions obtained from GATE simulations between block pairs 1-9 and 2-8 (see Figure 4.4 for block numbers) while Figure 4.5-(B) shows the measured response functions.

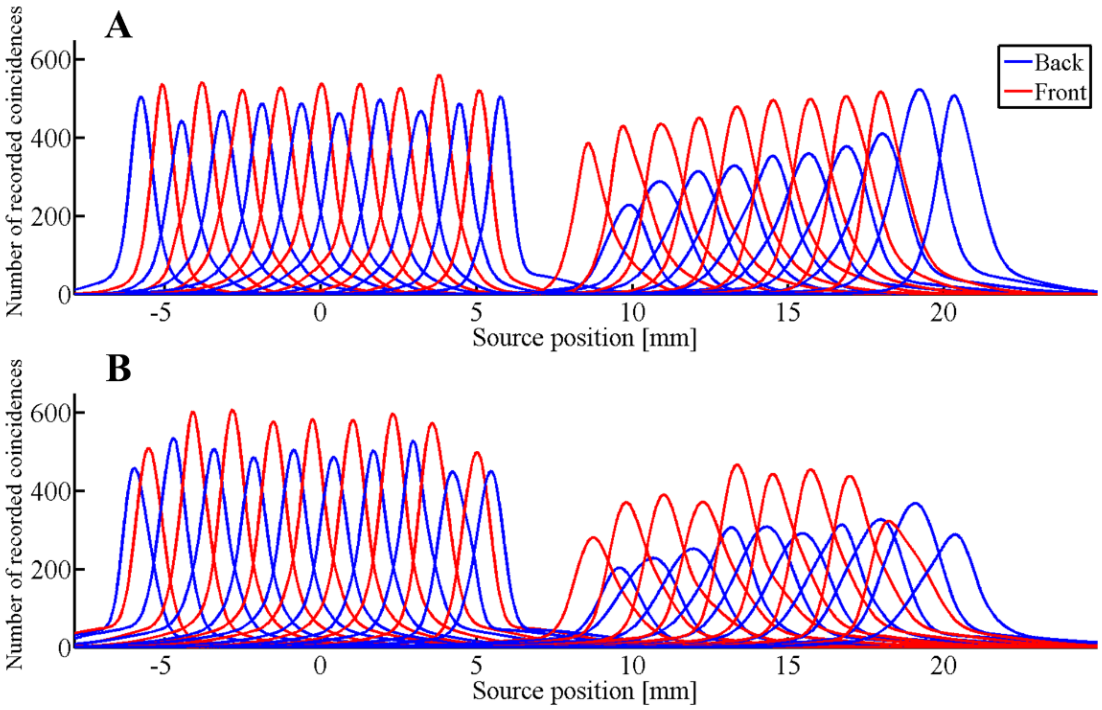

Figure 4.5. CRFs obtained from (A) MC simulations and (B) measurements. In these plots, the red and blue curves show the fit CRFs obtained from the front and back layer crystals, respectively (© 2018 IEEE [99]).

The two sets of profiles show reasonable visual agreement in shape and total counts, with differences mainly in the second pair of detector blocks likely due to uncertainty in exact alignment of the detectors in the physical system. The average and standard deviation of the CRF FWHM and FWTM for each pair of detector blocks are presented in Table 4.2.

Table 4.2. Average CRF FWHM and FWTM for CRFs of the small animal DLO scanner used in MC simulation validation experiment obtained experimentally and from simulations (© 2018 IEEE [99]).

Block pair		Front layer FWHM [mm]	Front layer FWTM [mm]	Back layer FWHM [mm]	Back layer FWTM [mm]
1 – 9	Experiment	1.05 ± 0.03	2.74 ± 0.24	1.08 ± 0.07	2.83 ± 0.33
	MC simulations	1.00 ± 0.08	2.65 ± 0.47	1.04 ± 0.11	2.79 ± 0.50
2 – 8	Experiment	1.72 ± 0.18	4.12 ± 0.27	2.10 ± 0.21	4.69 ± 0.25
	MC simulations	1.44 ± 0.10	3.54 ± 0.27	1.85 ± 0.17	4.00 ± 0.45

4.8 Geometry Optimization Scheme

An algorithm was developed to model and compare the performance of each detector from several aspects. Three sets of separate simulation studies were conducted to compare various performance metrics for each detector. The optimization process is shown in Figure 4.6 and is discussed in detail in the following sections.

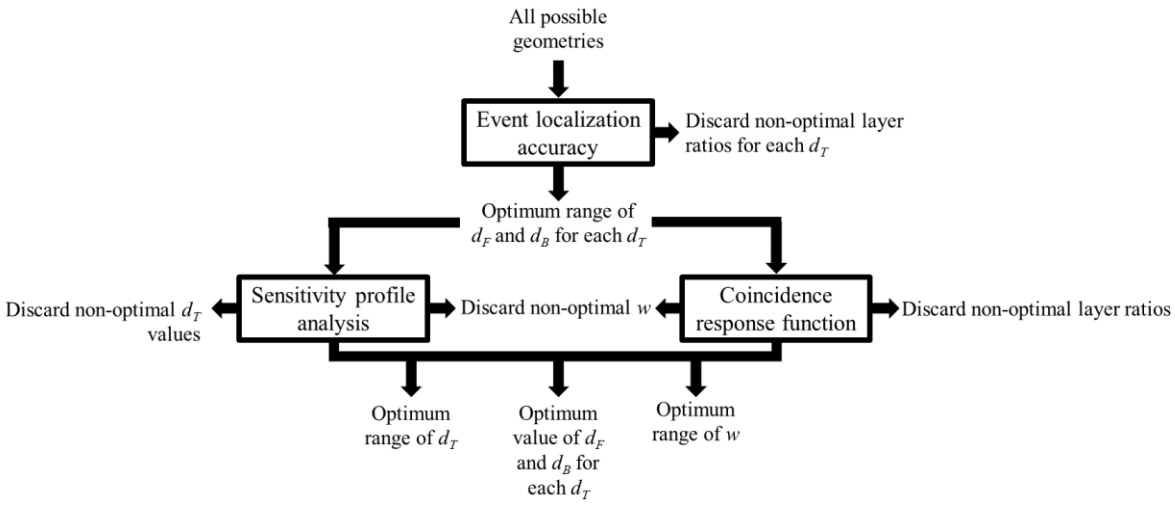

Figure 4.6. Flowchart of the optimization algorithm (© 2018 IEEE [99]).

4.8.1 **Evaluation of Event Localization Accuracy**

The purpose of this experiment was to find a range of optimum layer thickness ratios for each total thickness d_T through measuring the DOI accuracy provided by each detector.

4.8.1.1 Method

Similar to the technique used in [89] and [98], a single detector block with various geometries was simulated and the corner front layer crystal of the detector was irradiated with a uniform 511 keV photon beam at various incidence angles (α) as shown in Figure 4.7-(A). In this study, incidence angles of 0º, 14º (α_1), 27º (α_2), 37º (α_3) and 45º (α_4) were studied, mimicking values of increasing radial offsets in increments of 25% of the inner radius of the PET scanner as

shown in Figure 4.7-(B). In an example illustrated in Figure 4.7-(A), an incident photon is shown with beam angle α that undergoes Compton scatter at point 2 followed by photoelectric interaction at point 3. The *radial displacement R* between the assigned crystal's centroid (point 0) and the true LOR (the line connecting point 1 to 2) is shown in gray and labeled as *R*. Vector **v** shows the direction of *R* (normal to the line connecting point 1 to 2) and **s** is the distance between points 0 and 2. Point 4, shown in blue, represents the energy weighted location of the centroids corresponding to points 2 and 3. Note that point 4 does not lie along the line connecting points 2 and 3 in this example because the locations used for calculating the energy weighted centroid are the crystal centroids rather than the exact interaction locations.

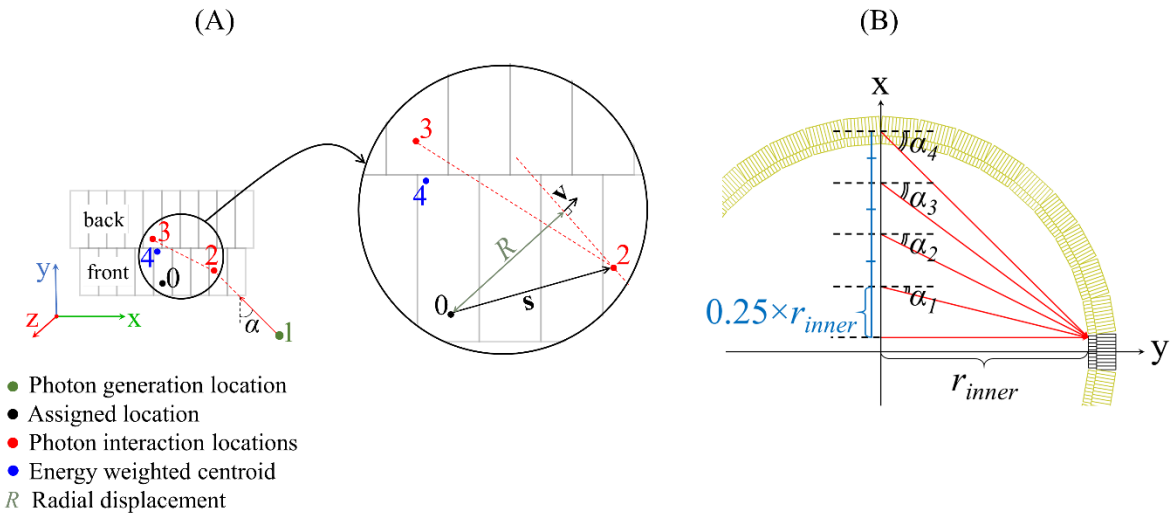

Figure 4.7. An illustration of (A) the radial displacement for an example incident photon and (B) different incidence angles α corresponding to different radial displacements (© 2018 IEEE [99]).

The resolution and resolution uniformity of a PET scanner is directly related to *R*. The radial displacement *R* is mathematically defined as

$$R = \frac{\mathbf{v}.\mathbf{s}}{|\mathbf{v}|} = \hat{\mathbf{v}}.\mathbf{s} = \frac{(x_2 - x_1)(y_1 - y_0) - (x_1 - x_0)(y_2 - y_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} \tag{4.1}$$

where x_i and y_i refer to the x and y coordinates of point i and vectors **v** and **s** are as defined in Figure 4.7-(A). For each detector irradiated with a beam of 511 keV photons at any given incidence angle, the radial displacements R was calculated for each incident photon.

An ideal detector is one that can maintain the radial displacement values as low as possible. To reduce the number of detectors in this section, the dependency of the radial displacement to crystal pitch w was investigated in an experiment in which a DLO detector block with a d_T of 30 mm, w values of 0.5, 1, 2 and 4 mm, and various layer thickness ratios was irradiated at various incidence angles α and the *average radial displacement* (R_{ave}) was measured in each set-up. As will be shown in Section 4.8.1.2, the optimum thickness ratio was seen to be independent of w and therefore the rest of the simulations for this section were performed only for detectors with a w of 2 mm. For each d_T, the average absolute radial displacement of all recorded events (R_{ave}) was calculated. A subset of detector geometries that minimizes R_{ave} for various incidence angles was selected as the optimum thickness split between the two layers. The chosen subset for each d_T included the two or three-layer ratios that yielded the smallest values of R_{ave}. The optimum layer ratios were then used in other experiments to further limit the number of optimum detector geometries.

4.8.1.2 Results

Sample histograms of radial displacements R for a detector block with a d_T of 20 mm, d_F of 2, 8, 14 and 20 mm, d_B of 18, 12, 6 and 0 mm and w of 2 mm and α values of 14º and 45º are shown in Figure 4.8-(A) and Figure 4.8-(B), respectively.

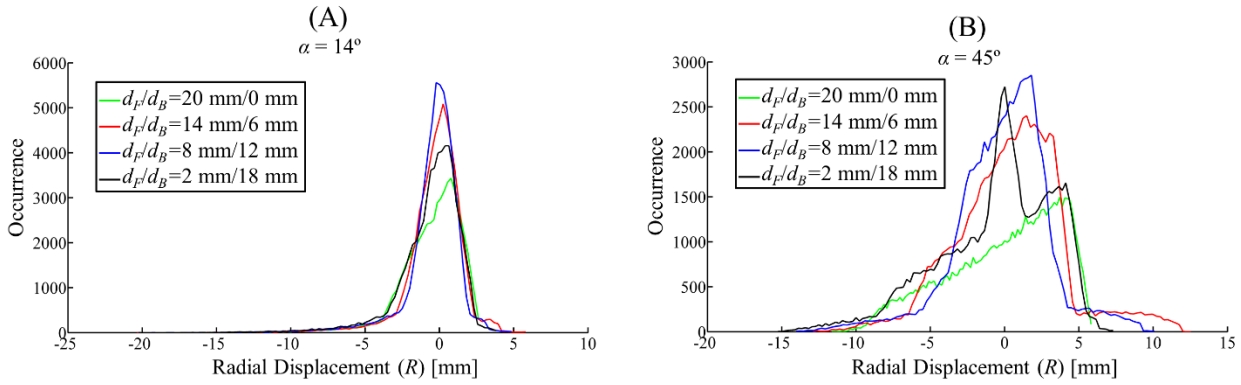

Figure 4.8. Histograms of radial displacement R between the first point of interaction and the assigned crystal centroid for a DLO detector block with a d_T of 20 mm, various layer thickness ratios, w of 2 mm and beam incidence angles α of 14° (A) and 45° (B). Each color corresponds to a certain d_F and d_B value as shown in the legends (© 2018 IEEE [196]).

Figure 4.9 shows the results of the average displacement analysis for a total detector thickness of 30 mm when w was set to 0.5, 1, 2 and 4 mm. It can be seen that the optimum d_B value (at which the displacement is minimum for various incidence angles) is the same for various crystal sizes, although the absolute displacement changes with various w values.

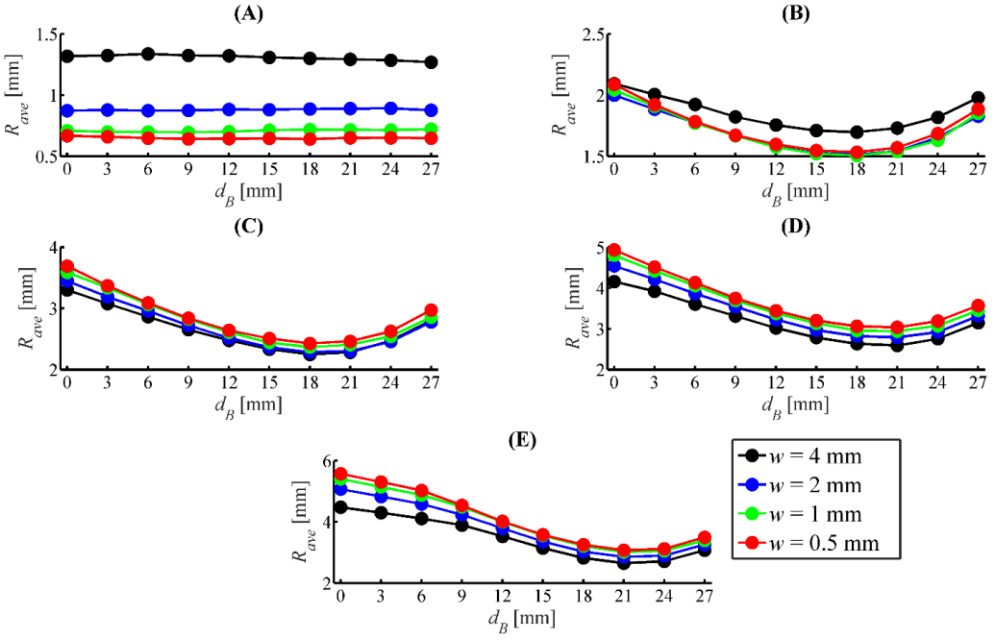

Figure 4.9. Average radial displacement R_{ave} vs. back layer thickness d_B for detectors with a d_T of 30 mm and various w values as described by the legend. Figures (A) to (E) correspond to incidence angles of α = 0°, 14°, 27°, 37° and 45°, respectively (© 2018 IEEE [99]).

The average radial displacement R_{ave} for each detector geometry was then calculated for a crystal pitch of w = 2 mm. Using the calculated average radial displacements, graphs of R_{ave} against back layer thickness d_B were plotted for detectors with total thicknesses of 10, 15, 20, 25 and 30 mm and are shown in Figure 4.10. For a beam with an incidence angle of 45º, the event positioning error is minimized at d_F/d_B of 5/5, 6/9, 8/12, 7.5/17.5 and 9/21 for d_T of 10, 15, 20, 25 and 30 mm, respectively. At these thickness splits, for α of 45º, R_{ave} is reduced by 31.3%, 33.4%, 36.8%, 39.9% and 43.7% as compared with a single layer detector block (i.e. d_B = 0 mm) for d_T of 10, 15, 20, 25 and 30 mm, respectively. For each d_T, minima of the displacement curves corresponding to various incidence angles (other than α = 0º) are extracted. The layer thicknesses that minimize the radial displacement for each d_T at all incidence angles (optimum d_F and optimum d_B) are shown in Table 4.3.

Table 4.3. Range of optimum d_F and d_B for DLO detectors with a total thickness of 10 to 30 mm and the percentage of R_{ave} reduction in comparison with a single layer detector at α = 45º (© 2018 IEEE [99]).

d_T [mm]	Optimum d_F [mm]	Optimum d_B [mm]	% reduction in R_{ave} for α = 45º
10	4	6	30.0
	5	5	31.3
	6	4	30.0
15	6	9	33.4
	7.5	7.5	33.0
20	8	12	36.8
	10	10	35.1
25	7.5	17.5	39.9
	10	15	39.3
30	6	24	42.9
	9	21	43.7
	12	18	40.3

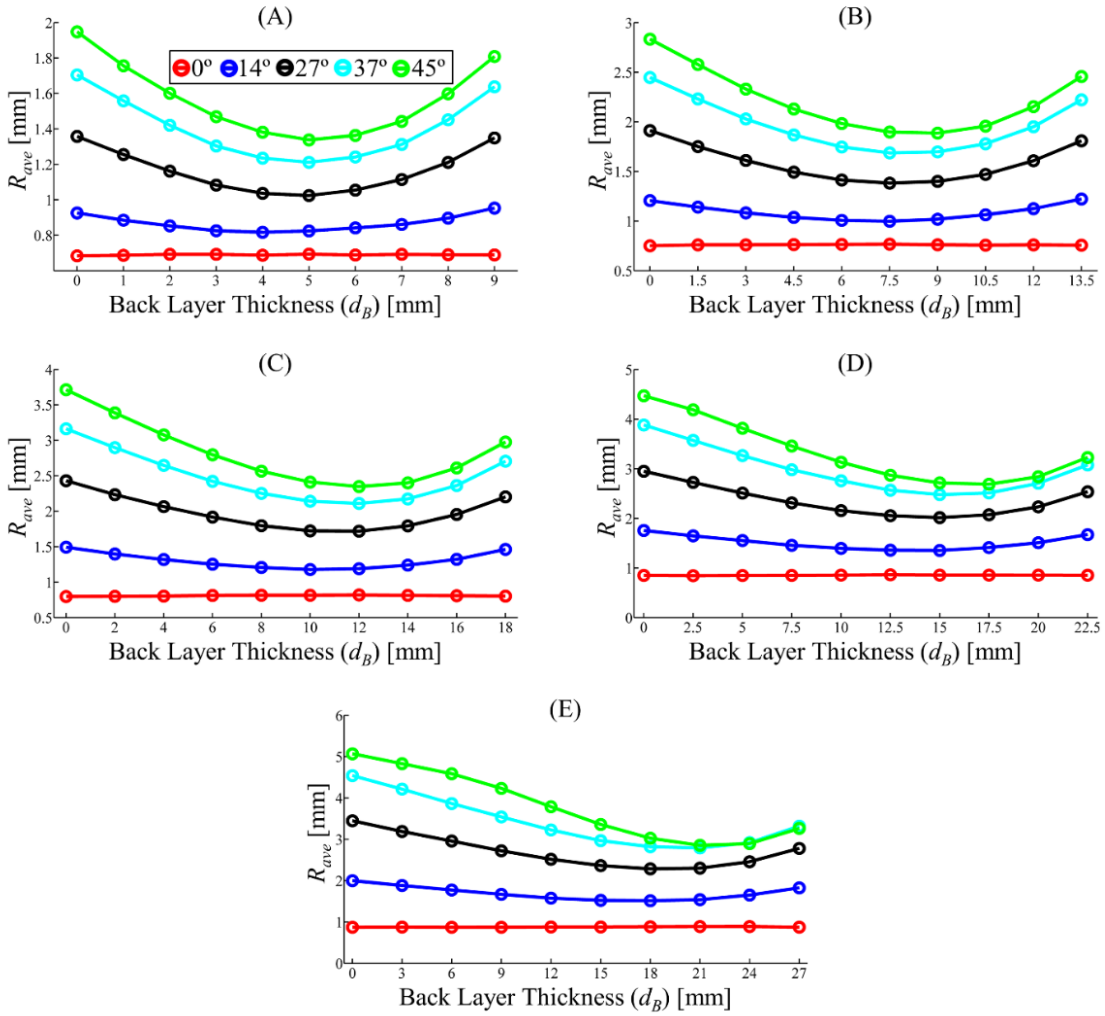

Figure 4.10. Graphs of R_{ave} vs. d_B for detector blocks with a w of 2 mm and total thicknesses of 10, 15, 20, 25 and 30 mm are shown in (A), (B), (C), (D) and (E), respectively (© 2018 IEEE [196]).

4.8.2 **Coincidence Response Function Analysis**

The aim of CRF analysis was to further limit the optimum thickness ratio for each d_T and evaluate the effects of various crystal widths on the performance of each detector.

4.8.2.1 Method

The detector geometries that passed the previous optimization step successfully were considered for a simulation of a half-ring PET scanner (only a half-ring PET scanner was simulated for each detector geometry to speed up the simulations). Coincidence response functions of

opposite crystal elements were measured by counting the number of coincidences between them when stepping a point source radially. A spherical point source (radius: 0.225 mm) embedded in a 10 × 10 × 10 mm^3 water-filled phantom was used in measuring the CRFs of each selected geometry. The dimensions of the source were chosen such that the source mimics the ^{22}Na point used in the NEMA NU4 spatial resolution measurement [166].

The source activity was set to 100 kBq with a long half-life so that the decay rate is constant during data acquisition. The length of each source step was set to *w*/8. For each detector geometry, a constant acquisition time per source step was used. The acquisition time was chosen such that for normal coincidences between detector number 1 and 20 (see Figure 4.11 for detector numbers), the CRF peak reached at least 300 counts. The CRFs for the front and back layers of detector pairs 1 – 20 to 6 – 15 were plotted individually. The average and maximum radial offset of each opposite pair of detector block is shown in Figure 4.11.

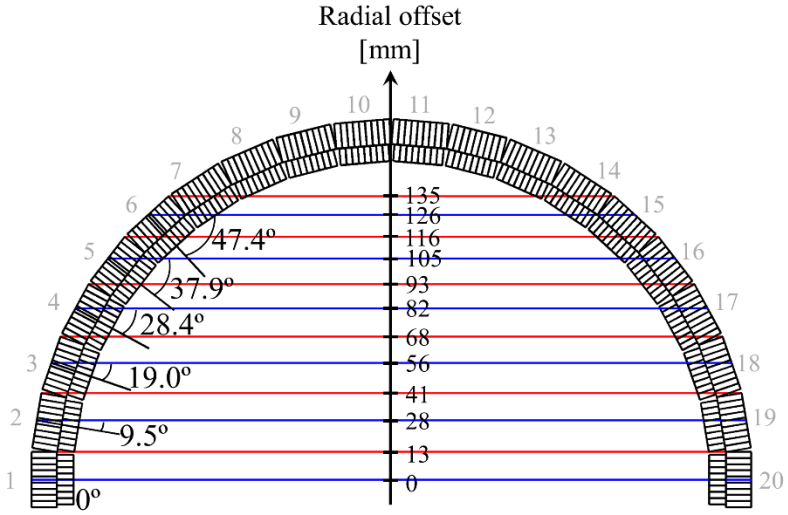

Figure 4.11. Schematics of a half-ring PET scanner used in CRF analysis (© 2018 IEEE [99]).

Each profile was fit with a sum of three Gaussian functions using least-squares fitting to reduce abrupt changes in the profiles and model any deviations from a typical Gaussian function

(e.g. some CRFs may have a considerable skewness due to radial elongation or blurring). The average and SD of the FWHM and FWTM of the CRF of each pair of opposite crystals (averaged over all crystals of each layer of the detector block pair) were then calculated for all scanners to facilitate a quantitative comparison.

As explained by equation (2.25), the spatial resolution of a PET scanner is a direct function of crystal width. Considering only geometrical factors, for two detectors with a width of w in coincidence, the response function is a triangular shape with FWHM equal to $w/2$ (i.e. FWHM/w of 0.5) if a point source is equidistant from the two detectors. The response function FWHM approaches w as the point source becomes closer to either of the detectors [4]. Therefore, in this work, the experimental ratio of the response function FWHM to crystal width (FWHM/w) for detectors with various crystal widths was also investigated.

The acquired CRFs model the elongation of the radial resolution with increasing radial offset and can offer a reliable estimate of the best achievable radial resolution for each scanner at various radial offsets [101]. Since the response function FWHM is a function of incidence angle, a single metric by which the average resolution of various scanners can be compared was defined as the *effective resolution*. Effective resolution of each scanner was defined to be the weighted average of the response function FWHM vs. radial offset. The weight for each pair of detector blocks was set to the maximum radial offset of the detector blocks pair (i.e. 13, 41, 68, 93, 116 and 135 mm for CRFs of detector pairs 1 – 20, 2 – 19, …, 6 – 15, respectively, as shown in Figure 4.11). The weights were chosen to monotonically increase with radial offset to account for the increase in the circumference of the circle corresponding to each radial offset. The maximum radial offset was selected over mean radial offset for weighting since the mean radial offset of detector pair 1 – 20 is zero. By measuring a radially-weighted resolution, the absolute radial resolution of the

scanners was compared. The effective resolution, denoted as ξ, was calculated for each studied scanner based on the following equation

$$\xi = \frac{\sum_i w_i r_i}{\sum_i w_i} \tag{4.2}$$

in which w_i is the maximum radial offset for each detector block i and its opposite detector block (corresponding to the red lines in Figure 4.11) and r_i is the mean FWHM of the CRFs for the pair of detector block. It should be noted that the effective resolution *does not* represent an estimate of the scanner resolution. Instead, it only provides a metric for comparing the radial resolution of different scanners in a relative manner. For comparison purposes, response functions of single layer PET scanners with total thicknesses of 10, 20 and 30 mm were also investigated.

For each group of detectors with identical total thicknesses and comparable CRF FWHM values, preference was given to a layer ratio that provided a closer coincidence detection ratio between the two layers. Detector blocks that have a similar coincidence detection rate in both layers were preferred since in practice having similar layer detection efficiencies reduces the impact of layer mispositioning errors, makes the flood map segmentation easier, and simplifies normalization since all crystals have approximately the same efficiency. The ratio of detected coincidences between the two layers was calculated by integrating over the obtained CRFs of each crystal pair followed by an averaging over the crystals of each layer in each opposite detector pair.

4.8.2.2 Results

Detector geometries that were outside the range of optimum d_F/d_B for each d_T (as described in Table 4.3) were excluded from the response function analysis. Therefore, a total of 60 (12 as defined in Table 4.3 and 5 crystal widths for each d_T) DLO detectors were simulated for CRF

evaluation. Also, for comparison purposes, the performance of single layer PET scanners having similar crystal pitch and thicknesses were investigated. Figure 4.12 shows two sample plots of response functions corresponding to a DLO PET scanner with a total detector thickness of 20 mm, front layer thickness of 8 mm and crystal width of 4 mm (Figure 4.12-(A)) and 2 mm (Figure 4.12-(B)).

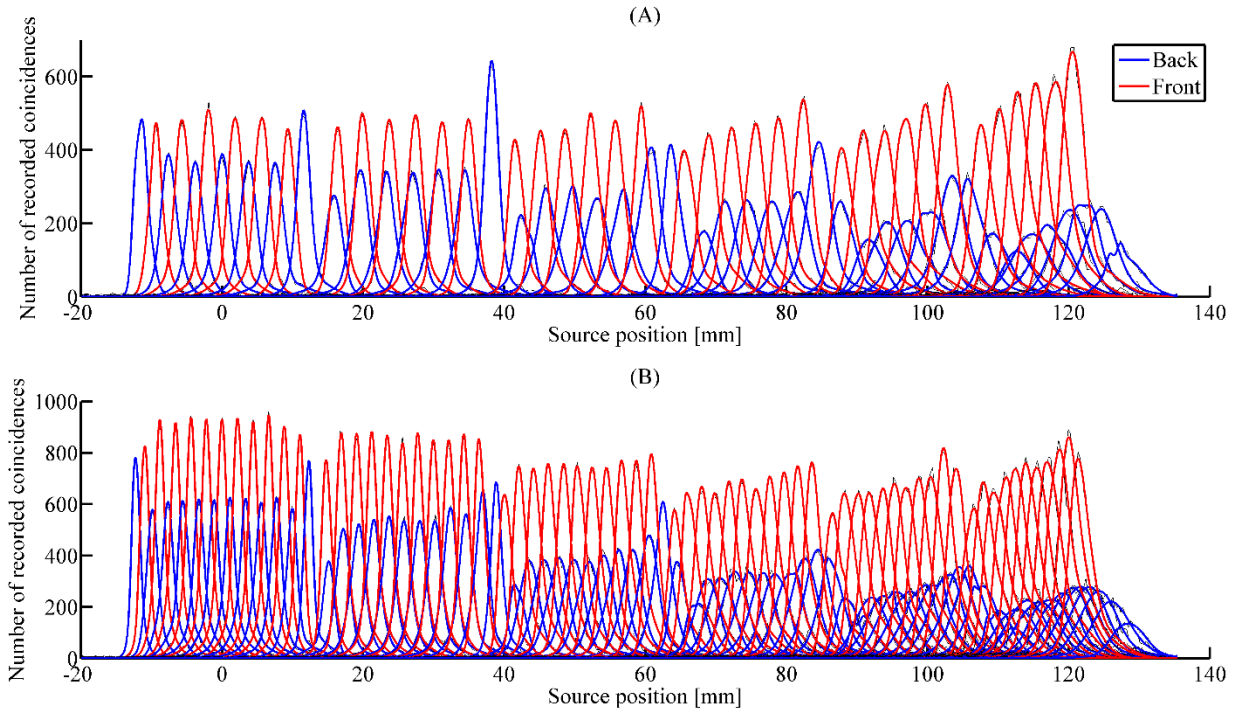

Figure 4.12. Sample CRFs of DLO PET scanners with a total detector thickness of 20 mm and a front layer thickness of 8 mm. The CRFs shown in (A) correspond to a crystal width of 4 mm while CRFs shown in (B) correspond to a crystal width of 2 mm. The underlying black curves are profiles drawn using the data before fitting (© 2018 IEEE [99]).

From the obtained CRFs for each detector geometry, the FWHM of the response function at various radial offsets was calculated. A surface plot of the FWHM of the response functions (averaged over crystals of each layer of each block) as a function of radial offset and crystal pitch w is shown in Figure 4.13, Figure 4.14, Figure 4.15, Figure 4.16 and Figure 4.17 for DLO detectors with d_T values of 10, 15, 20, 25 and 30 mm, respectively.

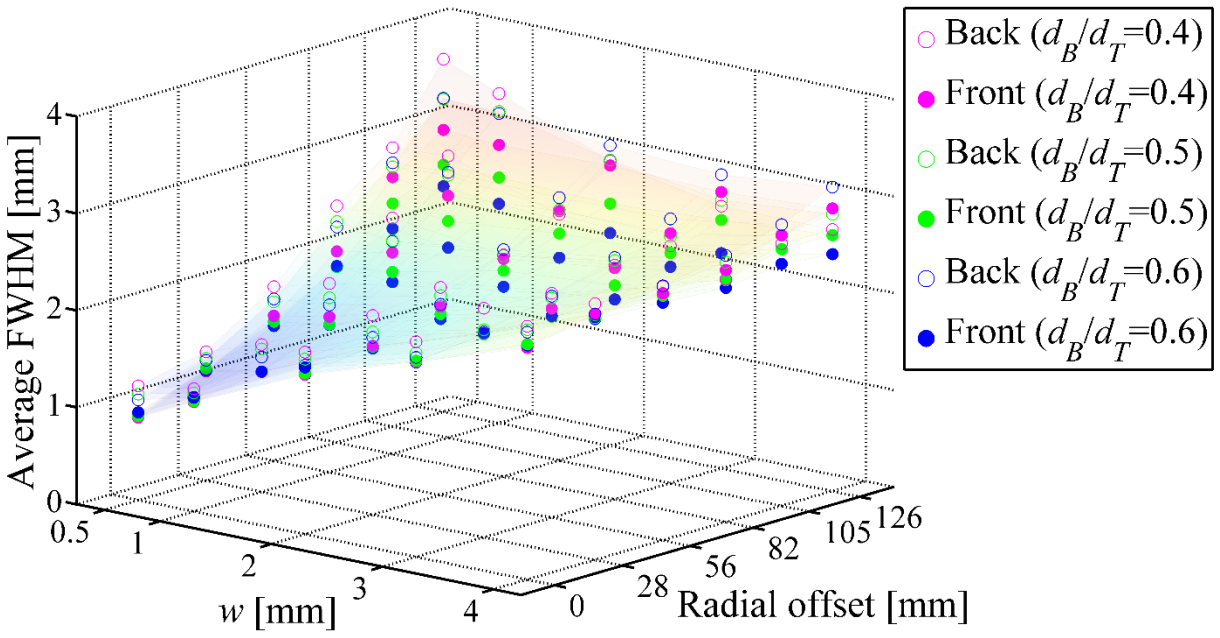

Figure 4.13. Surface plot of CRF FWHM against crystal pitch w and radial offset for DLO detectors with a d_T of 10 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively.

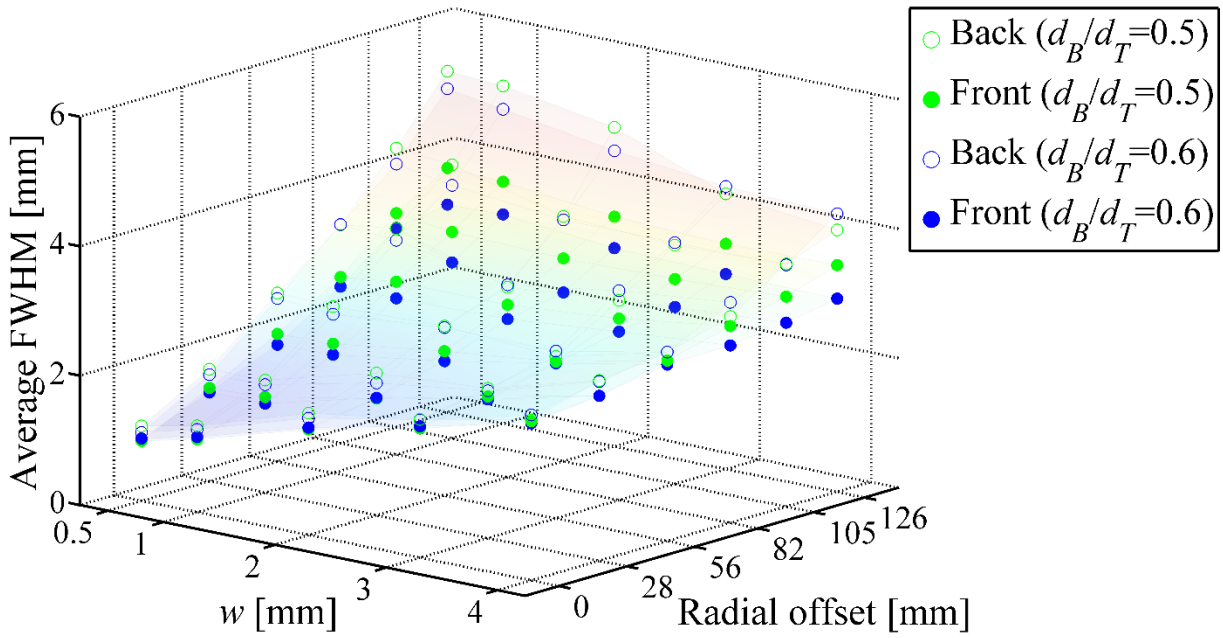

Figure 4.14. Surface plot of CRF FWHM against crystal pitch w and radial offset for DLO detectors with a d_T of 15 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively.

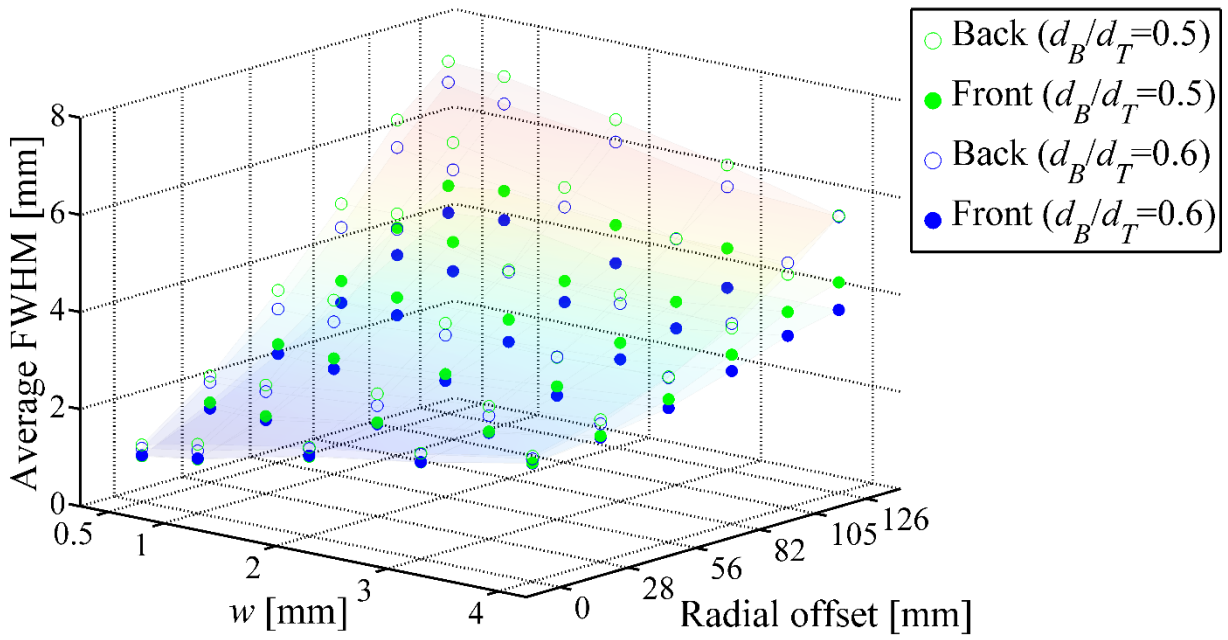

Figure 4.15. Surface plot of CRF FWHM against crystal pitch *w* and radial offset for DLO detectors with a d_T of 20 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively (© 2018 IEEE [99]).

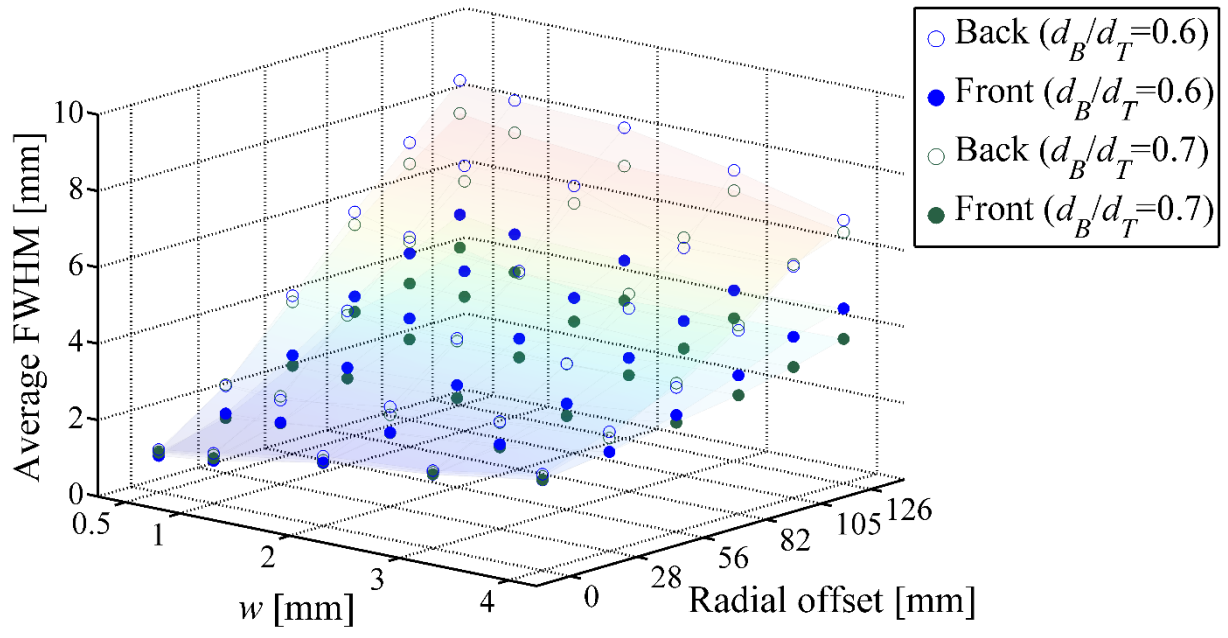

Figure 4.16. Surface plot of CRF FWHM against crystal pitch *w* and radial offset for DLO detectors with a d_T of 25 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively.

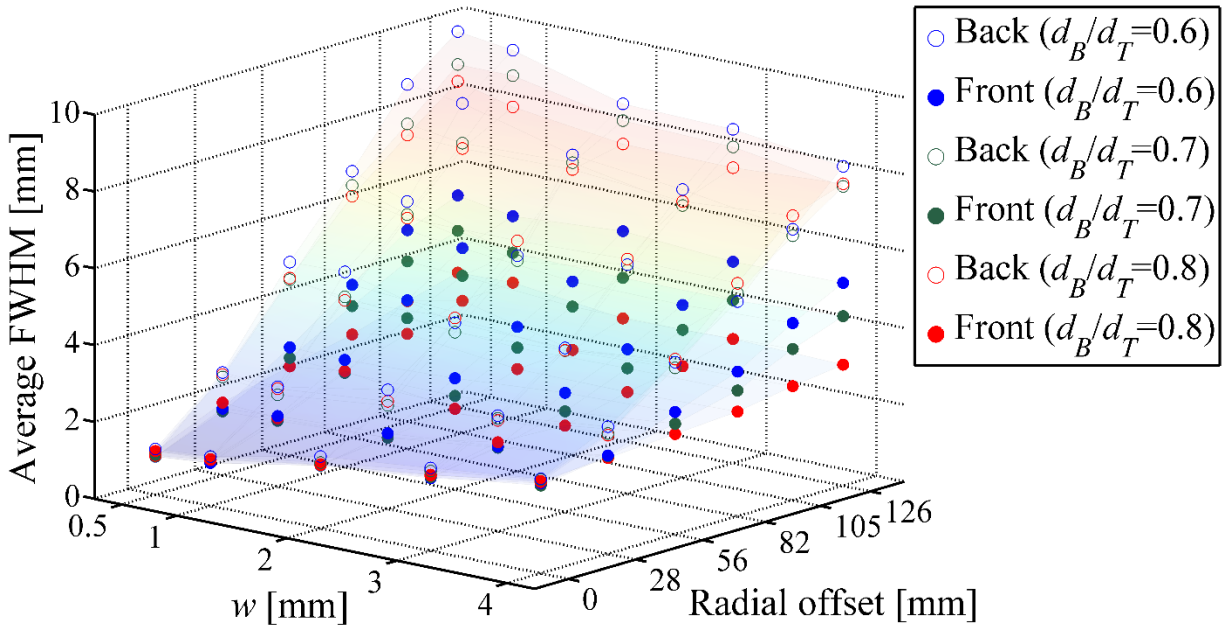

Figure 4.17. Surface plot of CRF FWHM against crystal pitch w and radial offset for DLO detectors with a d_T of 30 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively.

In the CRF plots, performance of each layer is demonstrated separately for comparison purposes. Overall, response functions for bottom layer crystals were worse than for front layer crystals in all DLO scanners. This was expected as intra-detector scattered photons are more likely to be absorbed in the back layer due to the forward-directed distribution of scatter at 511 keV. Similarly, surface plots of the CRF FWHM/w as a function of radial offset and crystal pitch for DLO detectors with d_T values of 10, 15, 20, 25 and 30 mm are shown in Figure 4.18, Figure 4.19, Figure 4.20, Figure 4.21 and Figure 4.22, respectively.

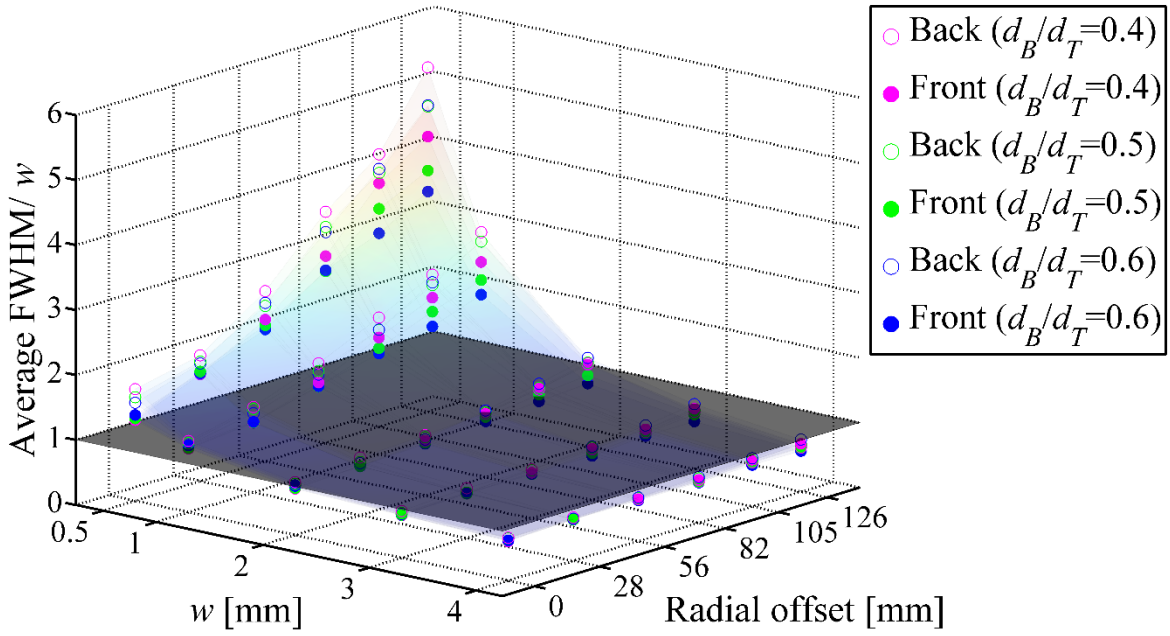

Figure 4.18. Surface plot of CRF FWHM/w against crystal pitch w and radial offset for DLO detectors with a d_T of 10 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively. The solid black surface represents the CRF FWHM/w of one.

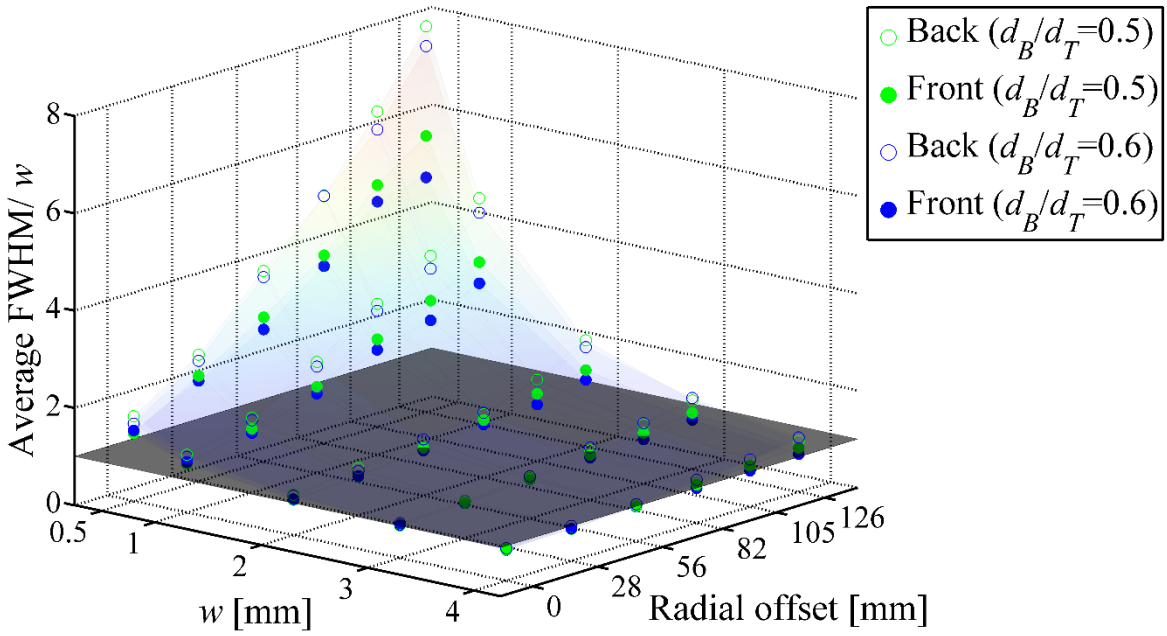

Figure 4.19. Surface plot of CRF FWHM/w against crystal pitch w and radial offset for DLO detectors with a d_T of 15 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively. The solid black surface represents the CRF FWHM/w of one.

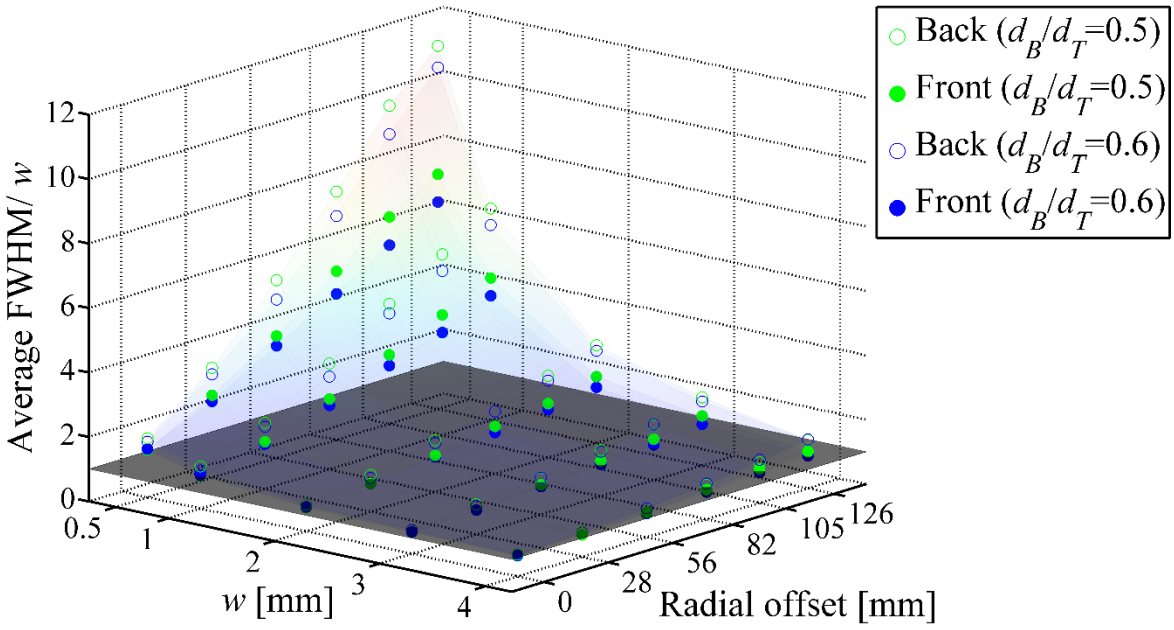

Figure 4.20. Surface plot of CRF FWHM/w against crystal pitch w and radial offset for DLO detectors with a d_T of 20 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively. The solid black surface represents the CRF FWHM/w of one.

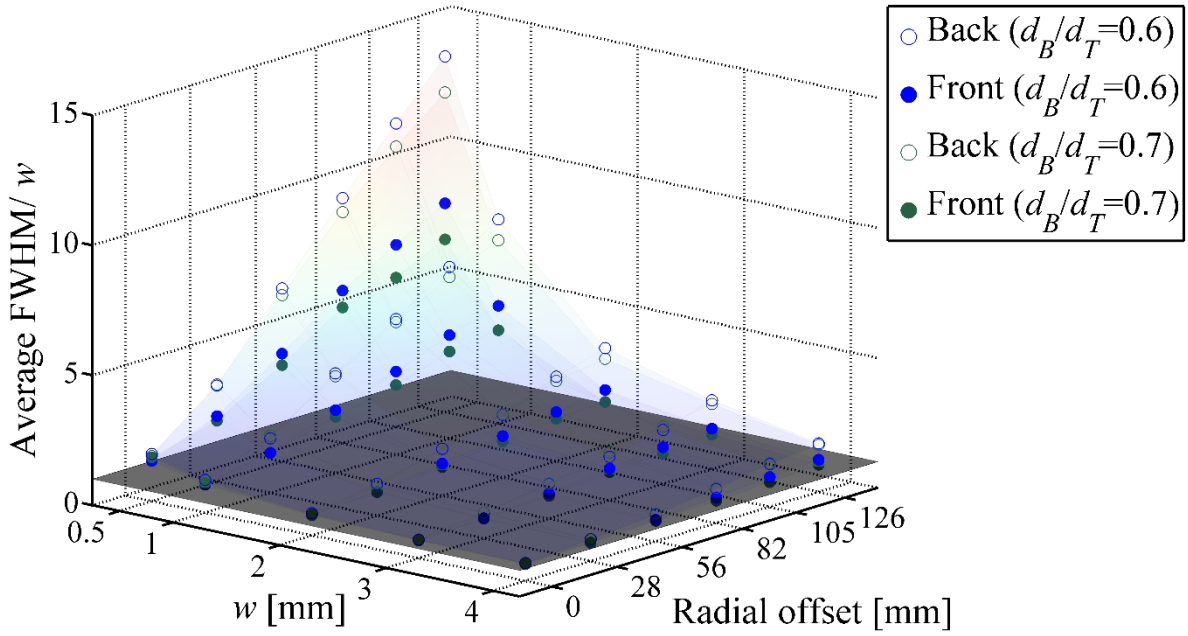

Figure 4.21. Surface plot of CRF FWHM/w against crystal pitch w and radial offset for DLO detectors with a d_T of 25 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively. The solid black surface represents the CRF FWHM/w of one.

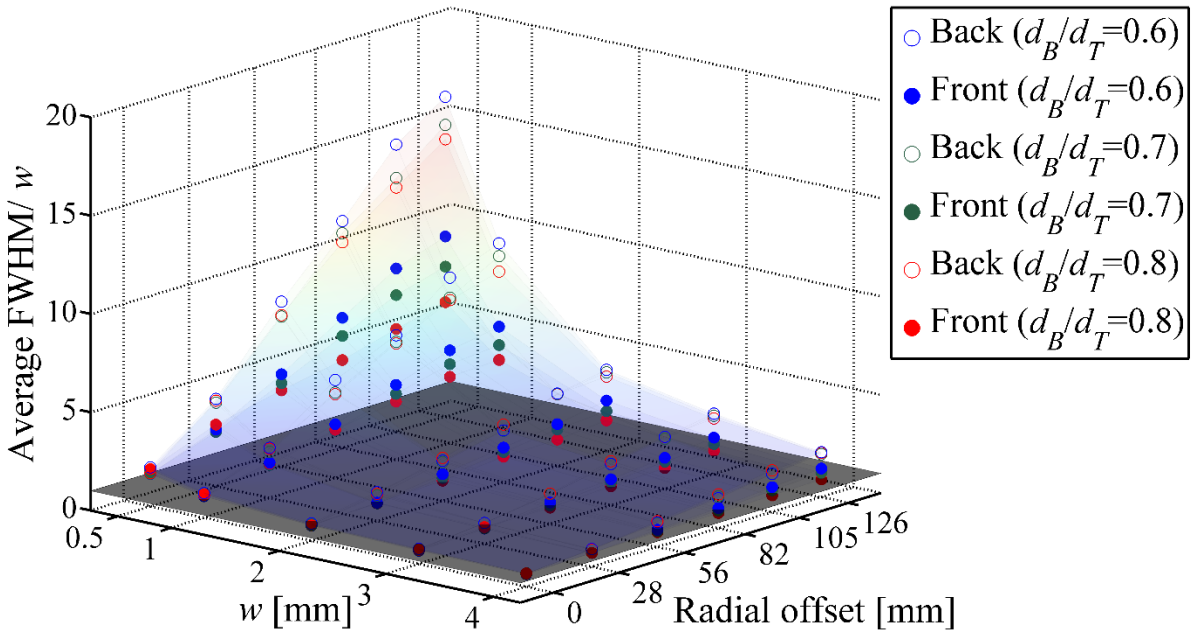

Figure 4.22. Surface plot of CRF FWHM/w against crystal pitch w and radial offset for DLO detectors with a d_T of 30 mm. Each color represents a layer thickness ratio as defined by the legend and filled and unfilled markers refer to the front and back layer, respectively. The solid black surface represents the CRF FWHM/w of one.

Similarly, surface plots of the CRF FWHM as a function of radial offset and crystal pitch w for single layer scanners with d_T values of 10, 20 and 30 mm are shown in Figure 4.23, Figure 4.24 and Figure 4.25, respectively. Also, surface plots of the CRF FWHM/w for the same order of detector thicknesses are shown in Figure 4.26, Figure 4.27 and Figure 4.28, respectively.

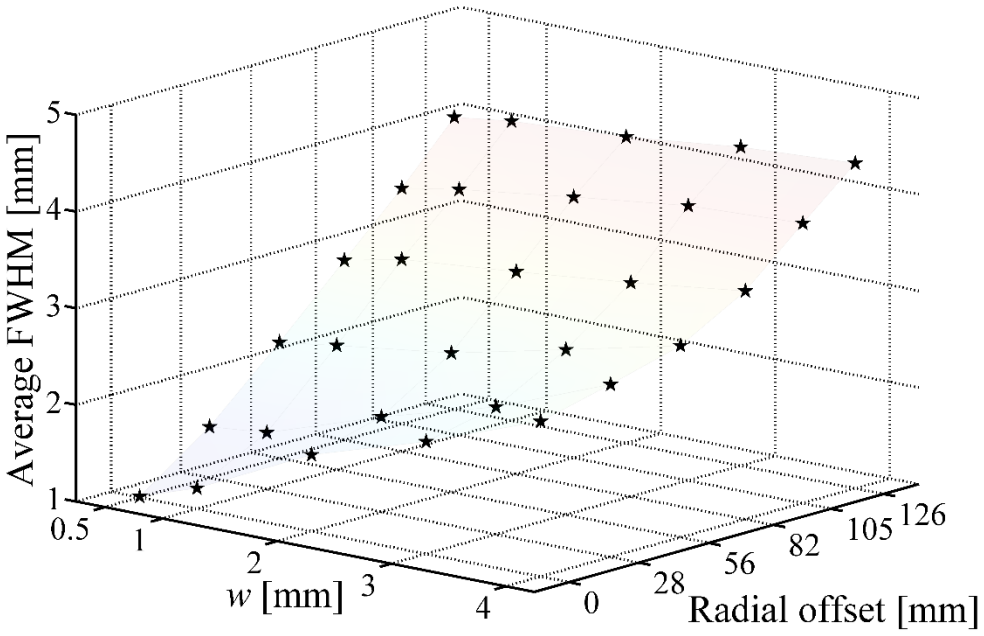

Figure 4.23. Surface plot of CRF FWHM against crystal pitch *w* and radial offset for single layer detectors with a d_T of 10 mm.

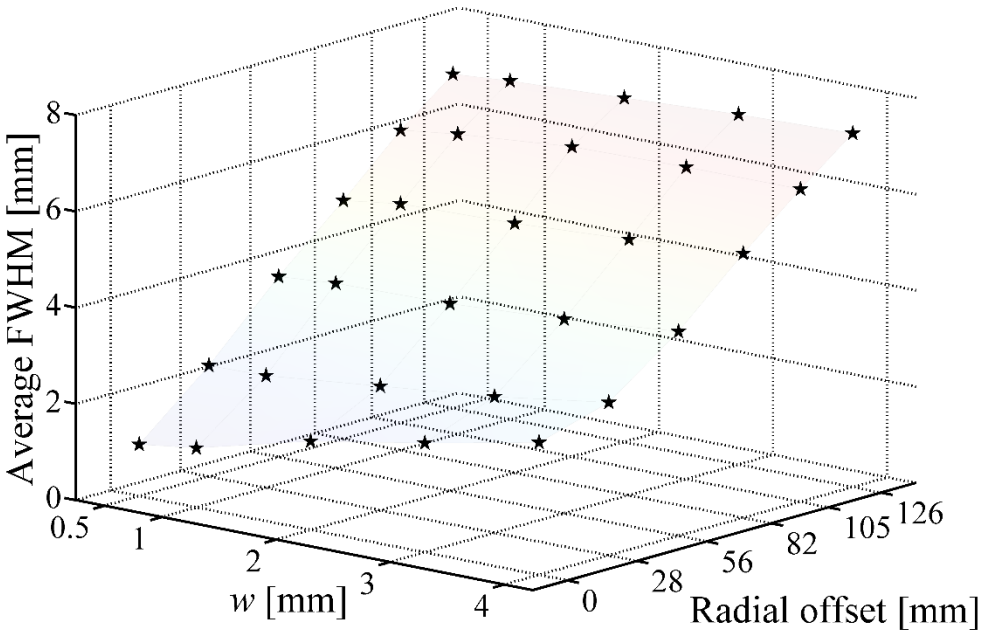

Figure 4.24. Surface plot of CRF FWHM against crystal pitch *w* and radial offset for single layer detectors with a d_T of 20 mm.

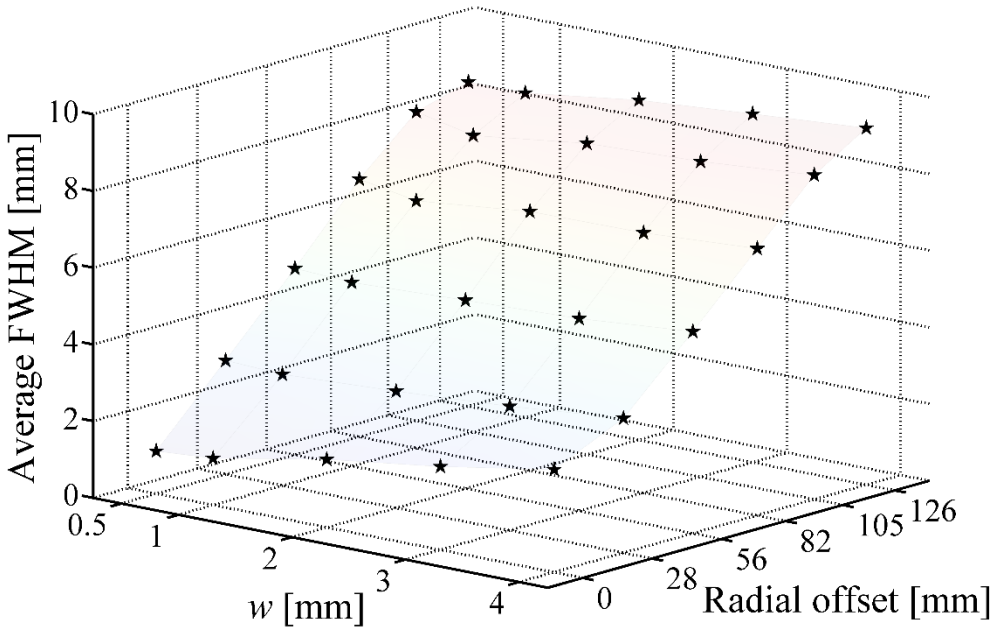

Figure 4.25. Surface plot of CRF FWHM against crystal pitch *w* and radial offset for single layer detectors with a d_T of 30 mm.

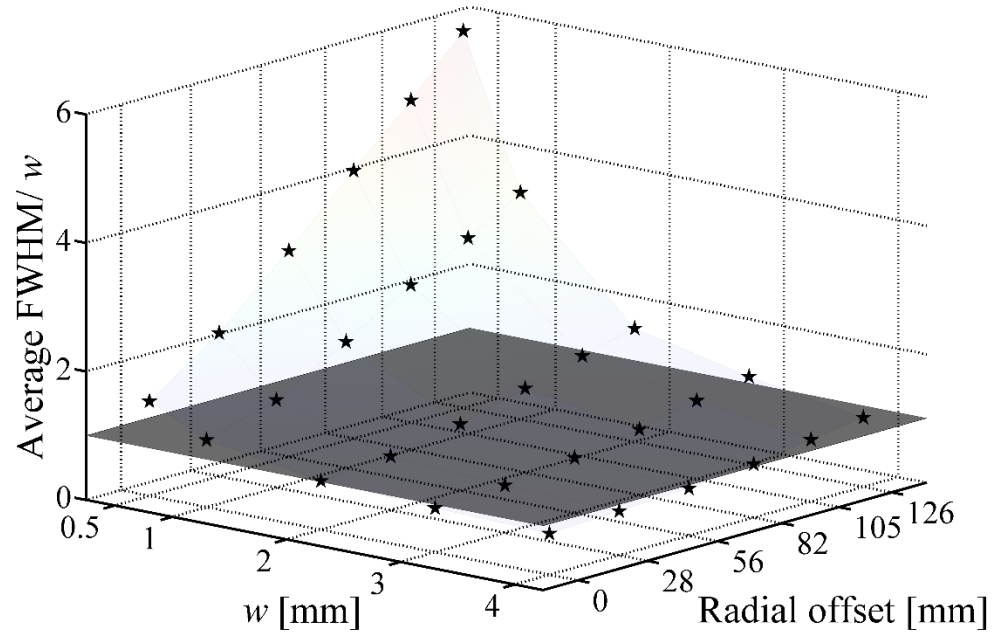

Figure 4.26. Surface plot of CRF FWHM/*w* against crystal pitch *w* and radial offset for single layer detectors with a d_T of 10 mm.

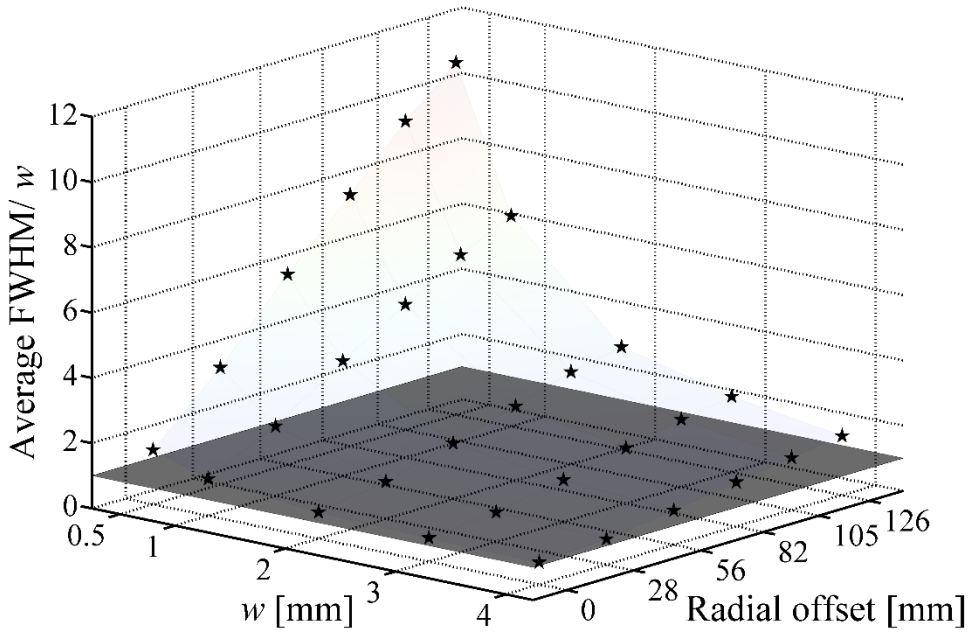

Figure 4.27. Surface plot of CRF FWHM/*w* against crystal pitch *w* and radial offset for single layer detectors with a d_T of 20 mm.

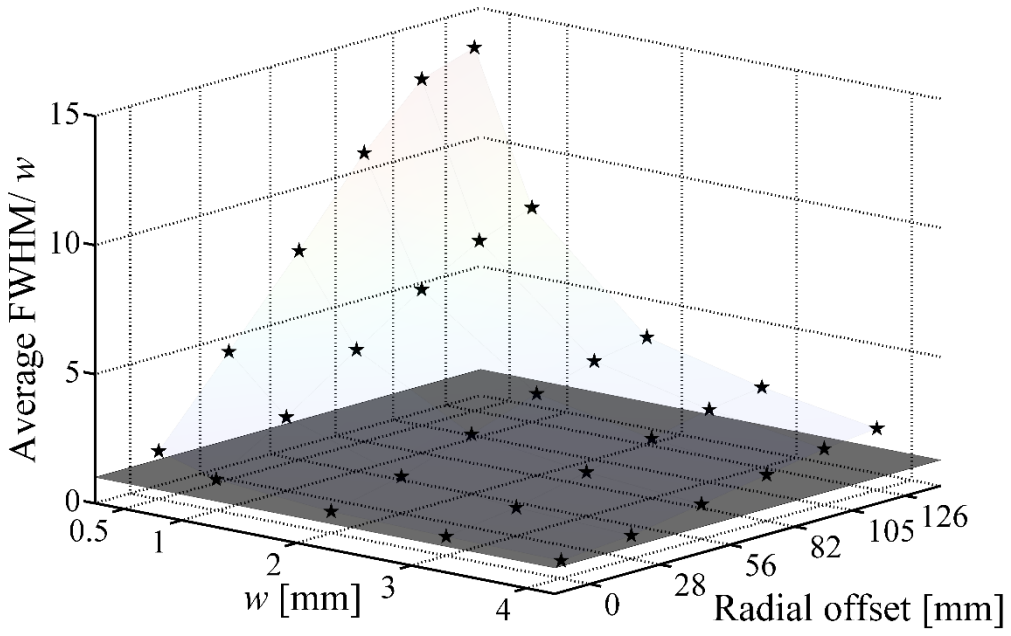

Figure 4.28. Surface plot of CRF FWHM/*w* against crystal pitch *w* and radial offset for single layer detectors with a d_T of 30 mm.

The calculated response function FWHM for each scanner was then used in equation (4.2) to facilitate calculation of the effective resolution ξ for each scanner. A plot of the calculated effective resolution for all scanners is shown in Figure 4.29. In Figure 4.29, each data point represents the effective resolution of a separate scanner, where colors and shapes of the markers define the crystal width and the layer thickness ratio, respectively. Single layer scanners with a d_T of 10, 20 and 30 mm were also considered for comparison purposes.

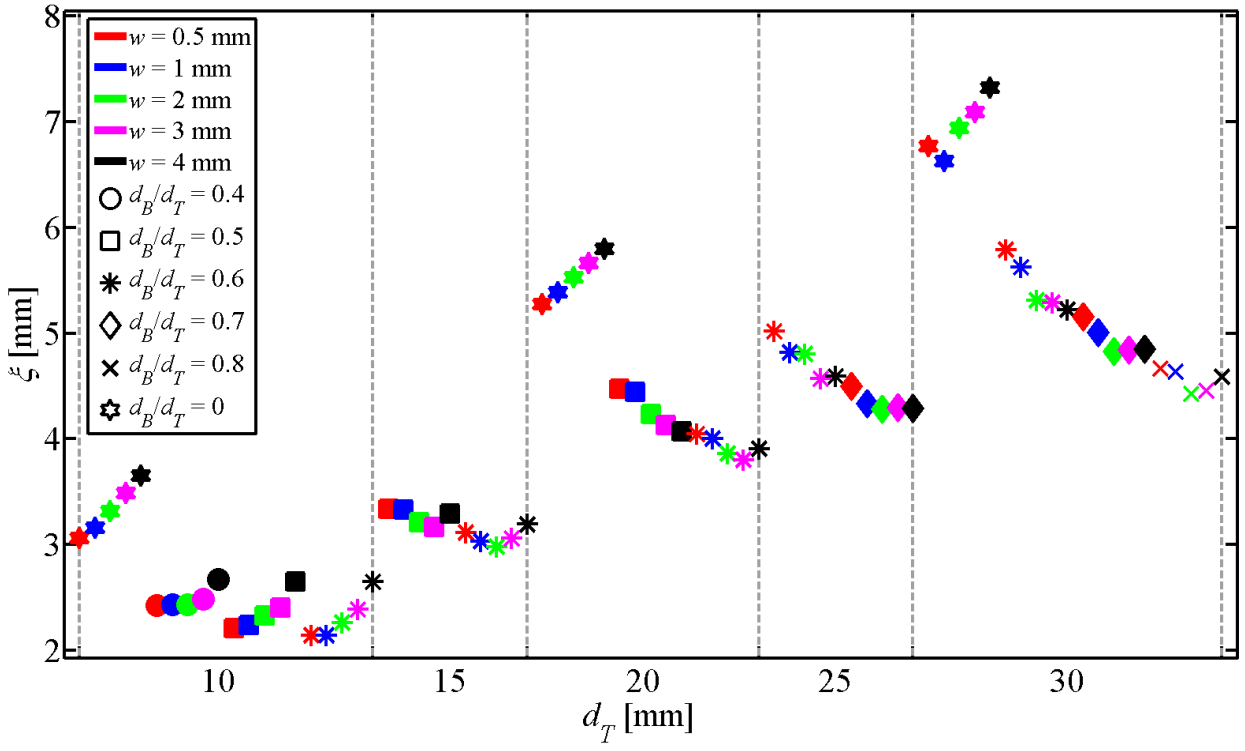

Figure 4.29. A plot of the effective resolution ξ for all DLO and selected single layer scanners considered in the CRF experiment. In this plot, each data point presents the performance of one scanner geometry. Colors red, blue, green, magenta and black represent a w of 0.5, 1, 2, 3 and 4 mm, respectively. Circles, squares, asterisks, diamonds, crosses and hexagrams represent a d_B/d_T of 0.4, 0.5, 0.6, 0.7, 0.8 and 0 (single layer), respectively (© 2018 IEEE [99]).

From the obtained CRFs, we also measured the ratio of the number of detected coincidences in the front to back layer. The notation ϵ represents the average number of detected coincidences in crystals of each layer of a detector block. Since this ratio was seen to be largely dependent on the thickness ratio of the two layers rather than the crystal width, here we only

present the results for DLO detectors with a w of 2 mm. A plot of average $\epsilon_{front}/\epsilon_{back}$ (the ratio of the average number of detected coincidences in the front to back layer) for DLO detectors with a d_T of 10 to 30 mm is shown in Figure 4.30. Note that these plots are based on the coincidences that were used in CRF derivation and do not include all recorded coincidences at all angles.

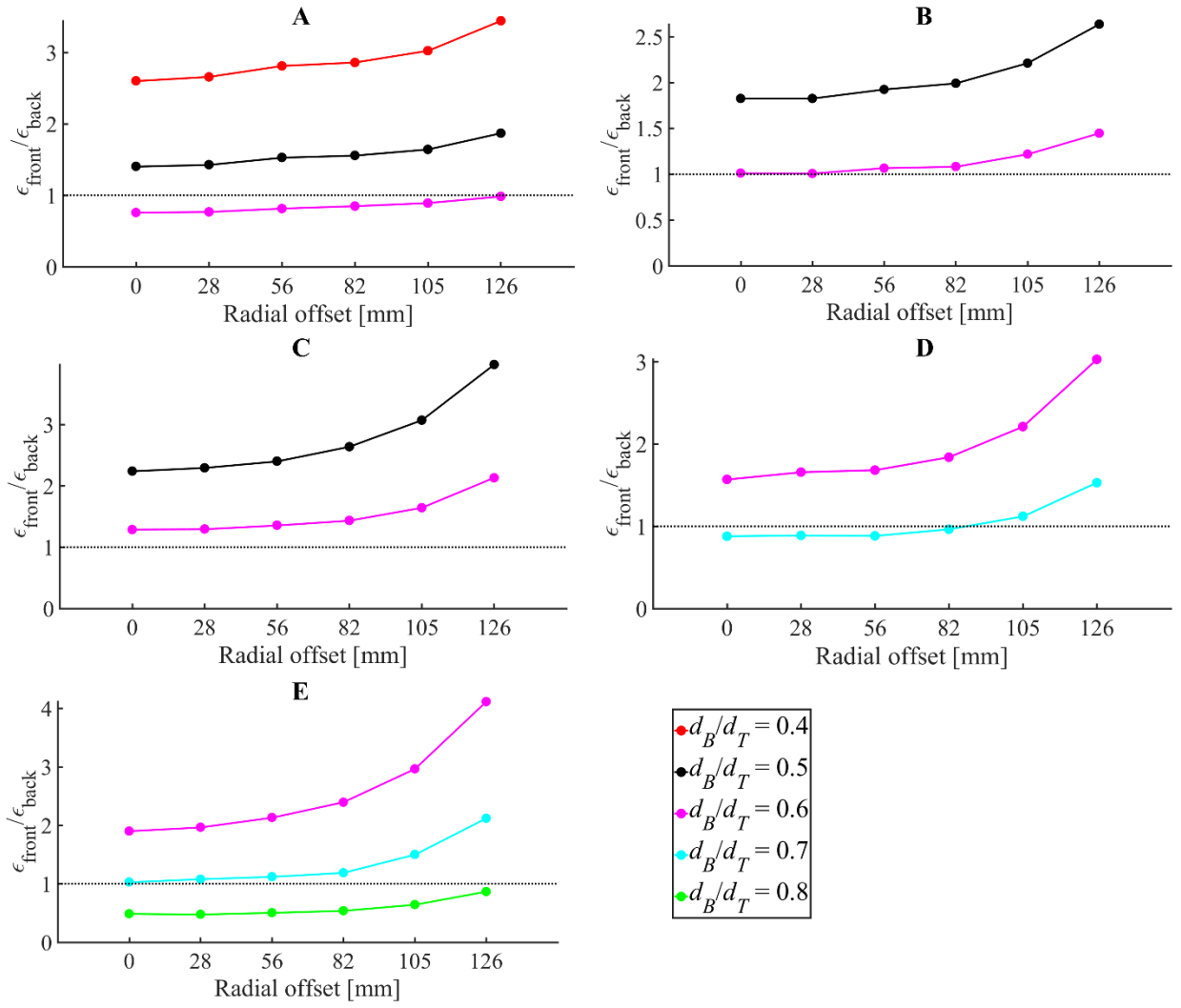

Figure 4.30. Average coincidence detection ratio of the front to back layer (ϵ_{front} / ϵ_{back}) for DLO detectors with a crystal width of 2 mm and total thicknesses of 10, 15, 20, 25 and 30 mm are shown in (A) to (E), respectively. The red, black, magenta, cyan and green curves represent a d_B/d_T layer ratio of 0.4, 0.5, 0.6, 0.7 and 0.8, respectively. The dashed line in all plots represents a ϵ_{front} / ϵ_{back} value of one (© 2018 IEEE [99]).

4.8.3 **Sensitivity Analysis**

In the two previous optimization steps, the sensitivity of the scanners was ignored. In a brain dedicated PET scanner however, sensitivity is an important performance metric. All scanners studied in the CRF experiment were studied for sensitivity analysis as explained below.

4.8.3.1 Method

Sensitivity profiles were seen to be largely dependent only on the total detector thickness. Thus, the results of only one selected layer ratio for each d_T are presented here. Scanners with a DLO detector and d_T values of 10, 15, 20, 25 and 30 mm were assumed to have d_F/d_B layer thicknesses of 4/6, 6/9, 10/10, 10/15 and 12/18 mm, respectively. The chosen thickness ratios were selected arbitrarily. An axial sensitivity profile of each scanner was measured by simulating a rod phantom with 4 mm diameter and 165 mm length located at the transaxial center of the scanner. The simulations were done without any attenuating material in the phantom and direct back-to-back photon pairs were simulated. Activity in the phantom was assumed to remain constant at 1 MBq for the duration of the acquisition (120 seconds). Similar to NEMA guidelines [167], single slice rebinning was used to rebin the oblique LORs using a slice thickness of 5 mm. The *scanner sensitivity* was then calculated by averaging sensitivity over the 5 central slices (i.e. central 25 mm).

4.8.3.2 Results

The results of only 25 selected DLO scanners are shown here (5 total thickness with a selected layer ratio × 5 crystal width). A similar sensitivity analysis was also done for equivalent single layer scanners, however, similar to layer thickness ratio, single layer scanners were seen to yield a similar sensitivity value as an equivalent DLO scanner with the same total thickness and

crystal width. For the selected geometries, scanner sensitivity was measured in kcps/MBq. A graph of the sensitivity for these scanners is shown in Figure 4.31.

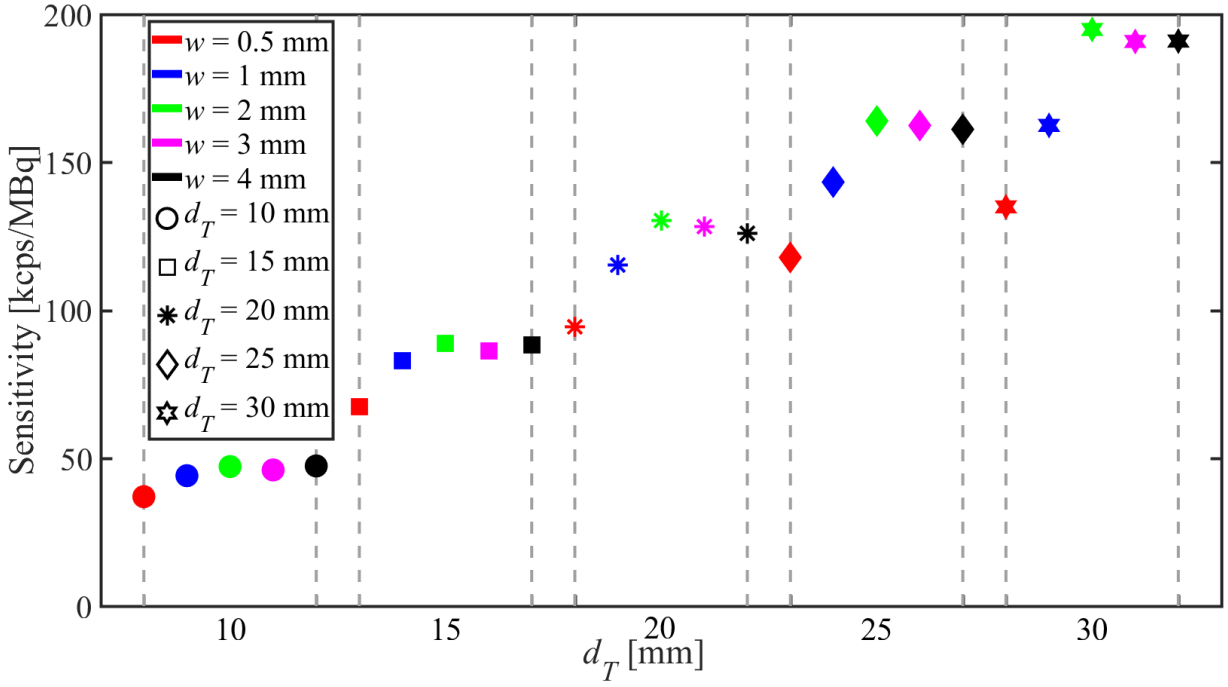

Figure 4.31. A plot of scanner sensitivity for various DLO scanner geometries. Only one layer thickness ratio per total thickness was studied. In this plot, circles, squares, asterisks, diamonds and hexagrams represent DLO scanners with a d_T of 10, 15, 20, 25 and 30 mm, respectively and colors red, blue, green, magenta and black represent a w of 0.5, 1, 2, 3 and 4 mm, respectively (© 2018 IEEE [99]).

4.9 Discussion and Conclusion

The radial displacement between the first point of interaction in the detector and the assigned position in the target crystal (R) is a dominant factor in defining the spatial resolution of scanner. For events normal to the surface of the detector, the distribution of the radial displacement is mainly a factor of crystal width. However, for large incidence angles, the absolute value of R and its mean value R_{ave} vary with other geometrical factors such as detector total thickness and thickness ratio between the two layers as well as crystal width. As demonstrated in Figure 4.9, optimum thickness ratio is a function of incidence angle and can be considered independent of crystal width. Depending on the incidence angle, a range of optimum values for the layer thickness

ratio can be determined. It should be noted that in a PET scan, coincidences are collected over a wide range of incidence angles, with a maximum defined by the geometrical properties of the scanner and minimum acceptable block and ring separation for a valid coincidence. The plots of R_{ave} for various layer ratios shown in Figure 4.9 suggest that while characterization of R_{ave} for normal and near-normal events largely depends on crystal width, it mainly is independent of crystal width for large incidence angles. At these large angles, the resolution is primarily a function of DOI accuracy (i.e. layer ratio).

For DLO detectors with a d_T of 10 mm, it was found that a thicker front layer yielded a slightly better average front layer response function FWHM at small radial offsets. However, at large radial offsets, a better front layer response function FWHM was obtained when a thinner front layer was used. The effective resolution of these scanners was slightly lower for a thicker back layer (d_B/d_T of 0.6) for all crystal widths as shown in Figure 4.29. Also, our simulations showed that at this layer split, the number of detected coincidences in the two layers is closer to each other (see Figure 4.30-(A)). Therefore, a front/back layer split of 4/6 mm can be considered superior to the other layer ratios for DLO detectors with a d_T of 10 mm. For detectors with total thicknesses of 15, 20 and 25 mm, d_B/d_T ratios of 0.6, 0.6 and 0.7, respectively, led to a better effective resolution as well as a more uniform coincidence detection rate between the two layers. Therefore, d_F/d_B ratios of 6/9, 8/12 and 7.5/17.5 mm can be considered most optimal for total thicknesses of 15, 20 and 25 mm, respectively. For DLO detectors with a d_T of 30 mm, a d_B/d_T ratio of 0.8 showed superior effective resolution in comparison with other layer splits for all crystal widths. However, a d_B/d_T of 0.7 led to a more uniform coincidence detection efficiency between the two layers. So, d_F/d_B values of 6/24 and 9/21 mm were preferred over 12/18 mm. Although in some studies, such as [114], researchers have simply considered equal layer thicknesses for DLO

(or other multi-layer) scintillation detectors, the results summarized in Sections 4.8.1 and 4.8.2 suggest that often a thinner front layer leads to a more optimum detector design and that a careful layer thickness optimization is required in the design of a multi-layer detector.

The surface plots of the CRF FWHM against layer thickness ratio and crystal width suggest that the response function FWHM generally decreases for a smaller *w*. However, the relationship between *w* and response function FWHM is not linear. The data shown in Figure 4.29 (effective resolution) and the graphs of response function FWHM/*w* suggest that all DLO and single layer detector geometries using crystal pitches of equal to or smaller than 1 mm lead to a response function FWHM/*w* ratio of greater than one. The loss in response function FWHM/*w* is an indicator that the potential gain one can obtain from using smaller crystal widths decreases sharply when using detectors with a crystal width of smaller than 2 mm. In particular, the benefits of a smaller crystal pitch in DLO detectors could be seen for the thinnest evaluated detectors (d_T = 10 mm) while for thicker detectors, crystal widths of 0.5 and 1 mm resulted in a greater effective resolution compared with a similar detector with a larger crystal width. The superiority of 2 to 3 mm crystal widths to a smaller value was an important finding that had not been a subject of investigation in the literature and could be of significant importance to other researchers in the field. As will be further discussed in this dissertation, the high ratio of response function FWHM/*w* and poor effective resolution for small crystal widths are due to a higher amount of inter-crystal scattering and the fact that a larger d_T enhances parallax error at greater radial offsets such that the effects of this error dominate over the intrinsic spatial resolution of the detector. The impact of inter-crystal scatter and consequent layer misassignment on DLO detectors will be studied in detail in Chapter 6. Furthermore, although using a smaller crystal width potentially helps to improve image resolution, it may increase the image noise too as it yields poorer statistics for each recorded LOR.

Comparing the CRFs of DLO detectors with their equivalent single layer detectors, it was seen that the response function of the front layer crystals is always better than their equivalent single layer detector, whereas the back layer crystals may show very close response functions to those of a single layer detector. Nonetheless, the effective resolution of the studied single layer detectors was higher than their equivalent DLO detectors regardless of their layer split.

The trend of sensitivity in scanners (Figure 4.31) showed a sensitivity improvement of up to a factor of ~4 when increasing total thickness from 10 to 30 mm. Note that the increase in the sensitivity of a scanner is the result of two factors: first, increased likelihood of photon interaction within the detectors and second, increasing the solid angle coverage for coincidences (all scanners had the same outer diameter, therefore the inner diameter of a scanner decreases as detector thickness increases). For any arbitrary total thickness, the sensitivity varies with w such that it peaks at a w of 2 mm and it oscillates slightly thereafter. The low sensitivity of detectors with a small w is a direct consequence of the low packing fraction (see Figure 4.2). The fluctuation in the sensitivity of scanners with crystals of larger than 1 mm is due to geometrical considerations. It should be noted that due to the half-crystal-pitch offset between the front and back layer crystals, the gap between the front layers of adjacent detectors increases as the crystal size exceeds 2 mm (e.g. less scintillation material volume is carried by detectors with a w of 3 mm than detectors with a w of 2 mm). This gap affects the detection efficiency rate of the scanner and is likely the cause of the sensitivity reduction. Optimization of w and d_T may require a compromise for the most important performance metrics of the scanner. It was seen that for DLO detectors with a d_T of 10 mm, a small crystal size provides better resolution but lower sensitivity. For total thicknesses of 15, 20 and 25 mm, the sensitivity of the scanners with a crystal pitch of 2 mm or larger was higher than other crystal sizes while yielding a lower effective resolution. For a d_T of 30 mm, the highest

achievable sensitivity occurred for a w of 2 mm at 194.9 kcps/MBq (for w = 3 or 4 mm sensitivity is ~191 kcps/MBq). The data showed that increasing d_T could consistently increase the sensitivity of the scanner and if geometrical variables were chosen wisely, prevents a rapid loss of resolution. Table 4.4 summarizes the findings for various geometrical parameters of a DLO detector.

Table 4.4. Summary of the optimum geometrical values (© 2018 IEEE [99]).

d_T [mm]	Optimum d_F/d_B [mm]	Optimum w [mm]	Approximate sensitivity [kcps/MBq]	Range of ξ [mm]
10	4 / 6	1, 2 and 3	47	2.15 to 2.39
15	6 / 9	2 and 3	88	2.98 to 3.06
20	8 / 12	2, 3 and 4	129	3.80 to 3.91
25	7.5 / 17.5	2, 3 and 4	163	4.28 to 4.29
30	6 / 24[a]	2, 3 and 4	192	4.43 to 4.59
	9 /21[b]	2, 3 and 4	192	4.83 to 4.85

[a] better ξ

[b] ϵ_{front} / ϵ_{back} closer to 1 for radial offsets less than 105 mm

In conclusion, it was shown that a DLO detector could reduce the radial blurring in PET by providing DOI information. In this chapter, an attempt was made to optimize the geometry of a DLO detector for use in a dedicated brain PET insert given a fixed detector area. The effects of three geometrical variables, total detector thickness, layer thickness ratio, and crystal width were studied through a series of MC simulations. A range of optimum layer thickness ratios was found that minimizes the average radial displacement between the first point of interaction and the assigned crystal's centroid (R_{ave}) for various incidence angle beams. Studying the CRFs of the scanners along with their sensitivity profiles, optimum front/back layer thickness values were found to be 4/6, 6/9, 8/12, 7.5/17.5, and 6/24 or 9/21 mm for total thicknesses of 10, 15, 20, 25 and 30 mm, respectively. Coincidence detection efficiency between the layers was considered when other performance metrics were not significantly different. It was found that a crystal size of 0.5

mm does not improve the performance metrics of the scanner and that a thicker detector may lead to a scanner with a considerably lower spatial resolution if its geometrical parameters are not carefully chosen.

The geometry optimization algorithm presented in Figure 4.6 was an important proposal in this work which can be used by other detector development research groups for different scanners, detector types and scintillation materials. Nonetheless, a number of factors were not considered in the suggested optimization algorithm. Among the most important ones are the cost of detectors and the difficulty in flood map segmentation for detectors with extremely small crystal widths. It should be noted that the placement of the MR gradient coil in the PET insert has not been considered in this study. Therefore, even though a total scintillator thickness of up to 30 mm was assumed in this optimization study, in reality detectors with a total thickness of larger than 20 mm may limit the required room for a brain-dedicated PET insert. Furthermore, the performance of each PET system was not evaluated using metrics such as NECR or NEMA phantom imaging in this chapter. In the next chapter however, a thorough investigation on the count rate performance and NECR of several PET detectors will be performed.

Chapter 5 : Study of Brain PET Count Rate Performance as a Function of Detector Area and Deadtime

5.1 Introduction

In Chapter 4, an optimum set of geometrical parameters for the detectors of the brain PET insert was suggested. Detector optimization in Chapter 4 was done assuming that the detectors of the PET insert have a predetermined detector area of 28.4 × 28.4 mm^2 and a scintillator array structure of DLO to provide DOI information. The chosen detector area was based on the assumption of using a 4 × 4 array of SensL ArrayC SiPMs. It should be noted that the measured performance metrics discussed in Chapter 4 (namely, event localization accuracy, CRF and scanner sensitivity) were independent of the count rate performance of the scanner. The count rate performance of a PET scanner, however, is an important performance metric that should be carefully modelled and considered when designing a PET detector. The current chapter investigates the impact of detector size (detector area) and deadtime on the count rate performance of a brain-dedicated PET insert. As the evaluated performance metrics in Chapter 4 were count rate independent, the findings of Chapter 4 and Chapter 5 are considered complementary. Furthermore, it should be mentioned that the count rate study in this chapter provides a comprehensive reference for the count rate performance of brain-dedicated PET inserts that can be of benefit to researchers within the field.

5.2 Importance of Detector Area

A number of detectors with a small detector area, suitable for a small-animal or brain-dedicated PET scanner, have been studied in the past. A small detector with a few crystal elements can be constructed by a one-to-one coupling of the scintillator to one SiPM or MPPC, facilitating

an individual crystal readout. As an example, a one-to-one coupling of individual crystals with MPPC arrays has been studied in [198-203]. Studies on one-to-one coupled detectors, such as [203, 204], demonstrated that such detectors can potentially lead to a sub-millimeter spatial resolution in a prototype small-animal PET/CT scanner. Detectors with an individual crystal readout scheme, however, require processing a large number of signal channels. Also, while these detectors can offer a high degree of inter-crystal scatter rejection and a low deadtime at the block level, they have a lower detection efficiency than a larger detector [200]. Detectors with a large area, on the other hand, have an improved detection sensitivity as they have a lower event loss rate due to inter-crystal scatter. Also, large detectors require handling a lower number of signal channels. Nonetheless, the count rate performance of large detectors is more easily deteriorated by detector deadtime.

5.3 Geometrical Parameters of Detectors

Similar to the experiments performed in the previous chapters, MC simulations were chosen to study the count rate performance of scanners with a wide range of detector geometrical parameters and deadtime performance metrics. The count rate performance of brain-dedicated PET scanners with a range of detector geometries was evaluated through Monte Carlo simulations using GATE v6.2 [143] as described below.

5.3.1 **Fixed Geometrical Parameters**

All PET scanners were simulated as cylindrical PET scanners with a fixed inner diameter of 32 cm and an axial length of 20 cm, giving a total detector area of approximately 2,000 cm^2, suitable for a PET insert for the chosen brain-dedicated MR scanner. Even though our brain-dedicated PET insert is to be constructed with DOI detectors through a DLO design, only single layer detectors were studied in this work as DOI does not relate to the count rate performance

metrics studied in this work. The detectors were simulated as pixelated scintillator arrays with varying detector area such that detector blocks of 1 to 2500 crystal elements were examined. According to the findings of Chapter 4, a crystal pitch of ~2 mm was selected for all detectors. Minimum block separation for an acceptable coincidence was set for each scanner geometry such that a minimum transaxial FOV of 25 cm is provided. Two sample detector blocks and transaxial views of their corresponding scanners are shown in Figure 5.1. While in the previous studies, LYSO was selected as the scintillator, the chosen scintillating material for the detectors in this chapter was LSO[1] with an assumed energy resolution of 12% at 511 keV and an ideal timing resolution. While this work considers pixelated scintillator arrays, the results are readily applicable to monolithic designs also since pileup and deadtime are considered at the block level, as described below.

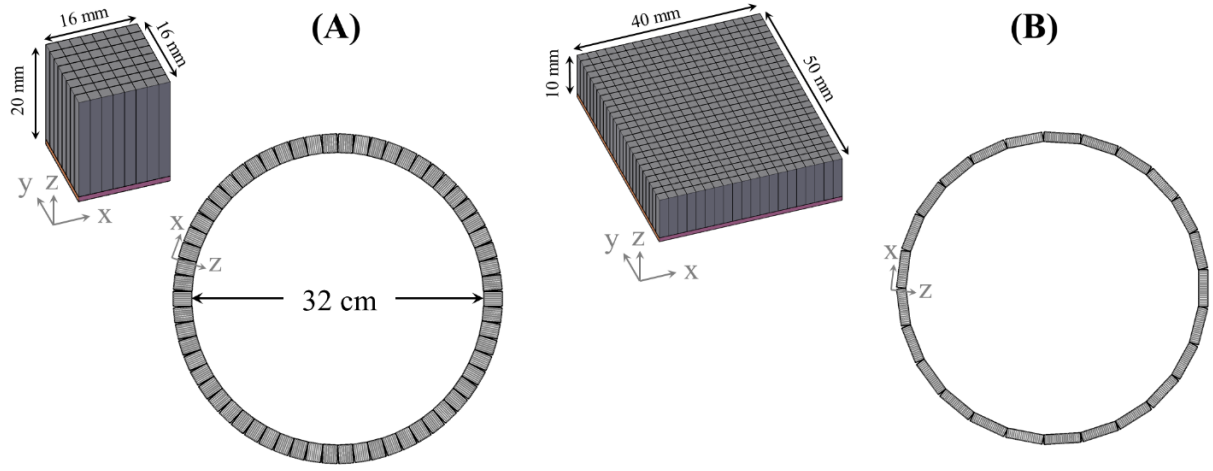

Figure 5.1. Schematic drawings of detector blocks and the transaxial views of two sample PET scanners. Scanners have an inner ring diameter of 32 cm with (A) 62 detector blocks per ring, 8 crystals per block transaxially, 8 crystals per block axially and a scintillator thickness of 20 mm and (B) 25 detector blocks per ring, 20 crystals per block transaxially, 25 crystals per block axially and a scintillator thickness of 10 mm.

[1] As LYSO has different variations, LSO was chosen in this chapter so that the results are applicable to other related researches.

5.3.2 **Detector Performance and Geometrical Variables**

Several geometrical variables and performance metrics were chosen for the studied detectors as explained in detail in the following sections.

5.3.2.1 Detector area

A total of 36 different detector areas between 0.04 cm^2 (1 crystal per detector block) to 101.37 cm^2 (2500 crystals per detector block) were studied. Crystals in each detector block were arranged such that no gap between the adjacent crystals exists. The absence of any gaps between the crystals of each detector causes a sensitivity increase which in turn may intensify the effects of deadtime on the count rate performance of the studied detectors. The exact crystal pitch was adjusted for each geometry such that adjacent detector blocks on the transverse and axial planes just touch. The mean crystal pitch along the transverse direction was 2.03 ± 0.03 mm (range: 2.00 to 2.12 mm) and along the axial direction was 2.02 ± 0.03 mm (range: 2.00 to 2.08 mm). The number of crystals per detector block in this chapter is shown as 'number of transaxial crystals × number of axial crystals'. Table 5.1 shows the geometrical properties of the chosen detector areas and their corresponding PET scanners.

5.3.2.2 Detector thickness

Detectors with scintillator thicknesses of 10, 15 and 20 mm were considered in this work. Unlike the design scheme used in Chapter 4, all scanners had a fixed inner diameter of 32 cm and therefore scanners with different detector thicknesses had different outer diameters.

5.3.2.3 Deadtime and pileup

In this work, deadtime only at the detector level was assumed. An ideal data acquisition system was considered such that no further deadtime contribution down the signal processing stream exists. However, to account for all components of the effective system deadtime and other

contributing elements to the scanner deadtime caused by the downstream data acquisition system in reality, a wide range of deadtime constants and types were considered. In fact, deadtime in this work was considered to be the effective deadtime of the scanner which includes the effects of both detector deadtime and pileup as pileup was not explicitly modeled. It should be noted that, in fact, detectors with paralyzable deadtime incorporate the effects of pileup to a large degree. Nonetheless, the occurrence of low-energy events piling up and contributing to the detected singles was ignored in this study. Detectors with paralyzable and non-paralyzable deadtime models (in separate experiments) with deadtime constants of 0, 250, 500, 750, 1000, 1500, 2000, 3000 and 4000 ns were considered in this study [205, 206]. The deadtime model and constant in each case was assumed to represent the cumulative deadtime properties of all components of the effective system deadtime.

5.3.2.4 Coincidence timing window

The coincidence timing window or CTW has a direct effect on the scanner's coincidence detection rate and, in particular, the rate of random coincidences. Often a limit on the CTW is the timing resolution of the detectors. In this work, the NECR of the scanners with CTW values of 4 to 12 ns was considered. To study the effects of the CTW on the NECR of the scanners with various detector sizes, the head phantom (see Section 5.4) at a given activity of 75 MBq and a fixed LLD of 350 keV was used. Detectors with thicknesses of 10 to 20 mm and areas of 0.04 cm^2 (1 × 1 crystal) to 101.37 cm^2 (25 × 100 crystals) were considered for this study. Since the relative effects of CTW on NECR were seen to be independent on the deadtime model and constant, a predefined deadtime type and constant of paralyzable and 750 ns, respectively, were assumed for all detectors in this section.

Table 5.1. Geometrical properties of the chosen detectors and scanners.

No. of crystals per block (transverse)	No. of crystals per block (axial)	No. of detector rings in scanner	No. of detector blocks per ring	Detector area [cm^2]
1	1	100	502	0.04
1	2	50	502	0.08
2	2	50	251	0.16
3	3	33	167	0.36
4	4	25	125	0.64
5	5	20	100	1.01
6	6	16	83	1.51
7	7	14	71	2.02
8	8	12	62	2.70
9	9	11	55	3.33
10	10	10	50	4.03
11	11	9	45	4.97
12	12	8	41	6.14
14	14	7	35	8.23
14	16	6	35	9.60
15	16	6	33	10.19
18	16	6	28	12.02
16	20	5	31	13.02
20	20	5	25	16.17
22	20	5	22	18.40
20	25	4	25	20.21
23	25	4	22	23.00
25	25	4	20	25.34
28	25	4	18	28.21
22	33	3	22	30.67
25	33	3	20	33.79
28	33	3	18	37.62
30	33	3	16	42.43
22	50	2	22	46.01
25	50	2	20	50.68
28	50	2	18	56.42
30	50	2	16	63.65
18	100	1	28	72.11
20	100	1	25	80.85
22	100	1	22	92.02
25	100	1	20	101.37

5.3.2.5 Energy thresholds

In each simulation, the number of singles per detector and coincidences along with their energy information were recorded. For each simulation, the singles rate per detector was calculated with LLD values of 0, 150, 200, 250, 300, 350, 400, 425 and 450 keV and a fixed ULD of 650

keV. Using the known history of each recorded coincidence (provided by GATE), the rates of true, scatter and random coincidences were also measured for LLD values of 0, 250, 300, 350, 400, 425 and 450 keV and a fixed ULD of 650 keV.

5.4 Phantoms

Two phantoms representing the human head and whole-body were considered in this study as explained below. In all simulations, the radioisotope was assumed to emit direct back-to-back 511 keV photons with a half-life much longer than the acquisition time.

5.4.1 **Head Phantom**

The *head phantom* was a water-filled cylindrical phantom with a diameter of 20 cm and an axial length of 17.5 cm. The head phantom was imaged with activities of 12.5, 25, 37.5, 50, 62.5, 75, 87.5, 100, 112.5, 125, 150, 175, 200, 225 and 250 MBq. In studies where the head phantom was used, the head phantom was centered at the axial centre of the scanner as shown in Figure 5.2-(A).

5.4.2 **Whole-body Phantom**

The *whole-body phantom* was a water-filled cylindrical phantom with a diameter of 20 cm and axial length of 70 cm. It was imaged with the same activity concentrations as the head phantom (i.e. activities of 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 and 1000 MBq). While the head phantom was centered at the axial centre of the scanner, the whole-body phantom was placed such that the first 17.5 cm part of it was at an identical position as the head phantom as shown in Figure 5.2-(B).

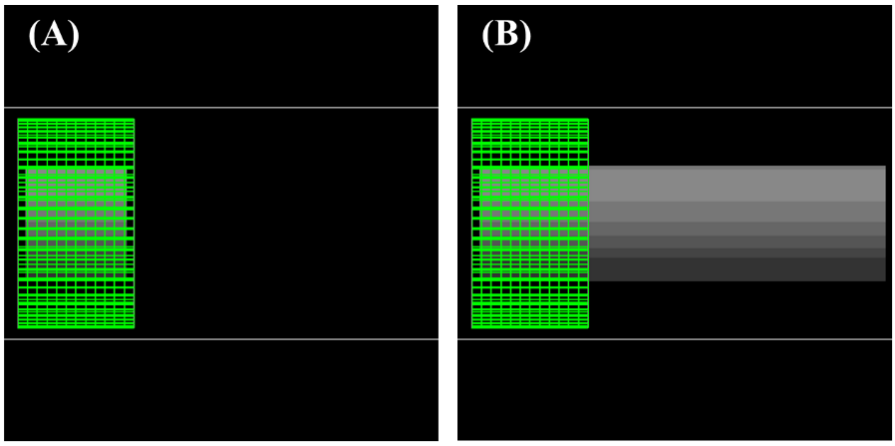

Figure 5.2. An image of the head (A) and whole-body (B) phantoms positioned in a scanner. The phantoms and scanners are shown in gray and green, respectively.

5.5 System Characterization

The work described in this chapter incorporated 174,960 distinct simulations (requiring about 5,120 hours of CPU time in total on a single core of an Intel Xeon X5647 CPU) to cover all detector and phantom parameters considered. Each simulation time was adjusted such that 10^6 decays (and not events) in the FOV (and not phantom) of the scanner are simulated. The coincidence policy '*takeAllGoods*' (per GATE's terminology) was implemented in all simulations, according to which, all possible lines-of-response are accepted in case of multiple coincidence. From the results of the obtained simulations, the following performance metrics were studied.

5.5.1 **Singles Rate**

Figure 5.3 and Figure 5.4 show the system singles rate for scanners with a detector thickness of 10 mm and deadtime constants of 0 to 4000 ns for non-paralyzable and paralyzable deadtime models, respectively, when the head phantom is imaged with an LLD of 350 keV. Similar plots for a detector thickness of 15 mm are presented in Figure 5.5 and Figure 5.6. Also, plots for a detector thickness of 20 mm are shown in Figure 5.7 and Figure 5.8.

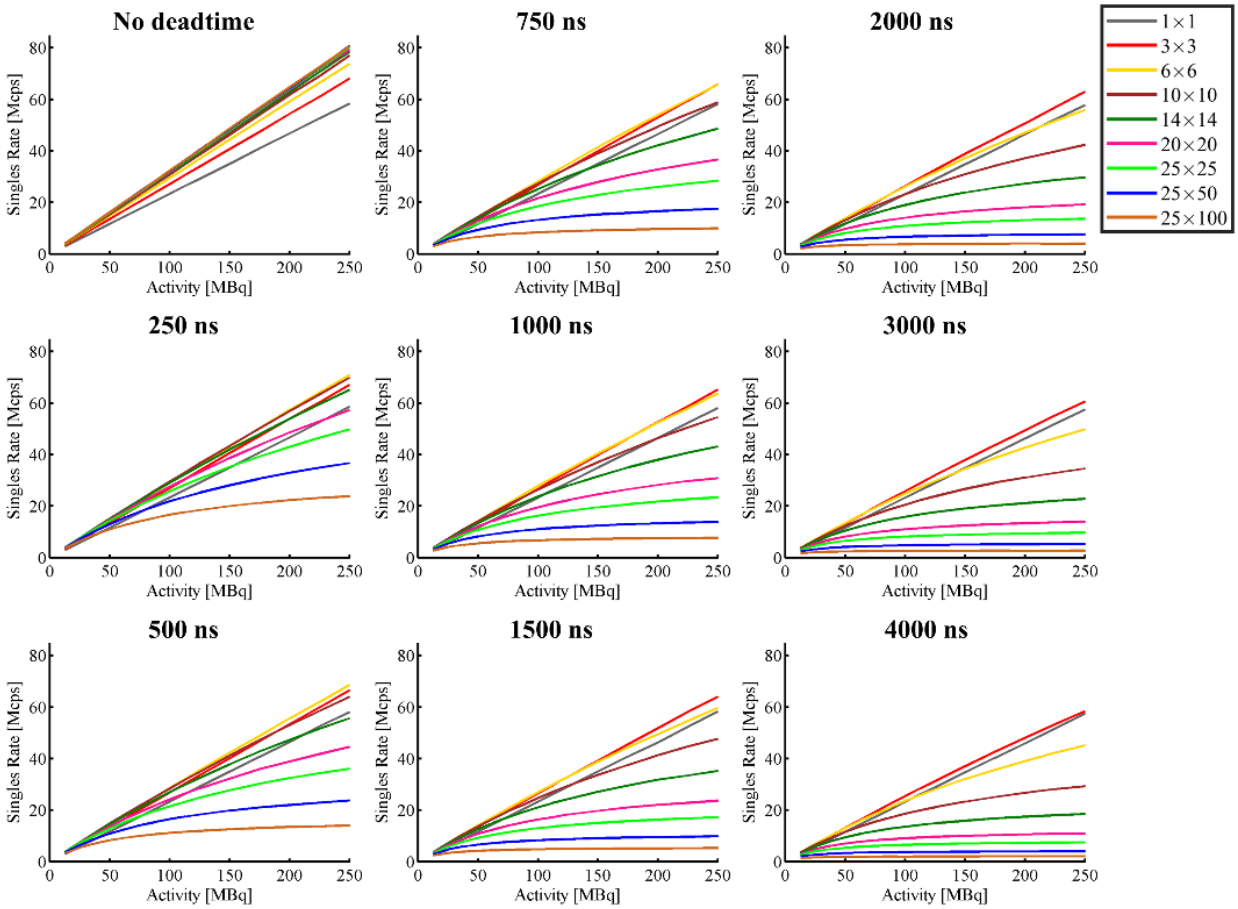

Figure 5.3. Plots of the scanner singles rate vs activity in the head phantom for brain-dedicated scanners with a detector thickness of 10 mm and various detector sizes as defined by the legend. A non-paralyzable deadtime model with various deadtime constants, as shown on the title of each plot, was considered.

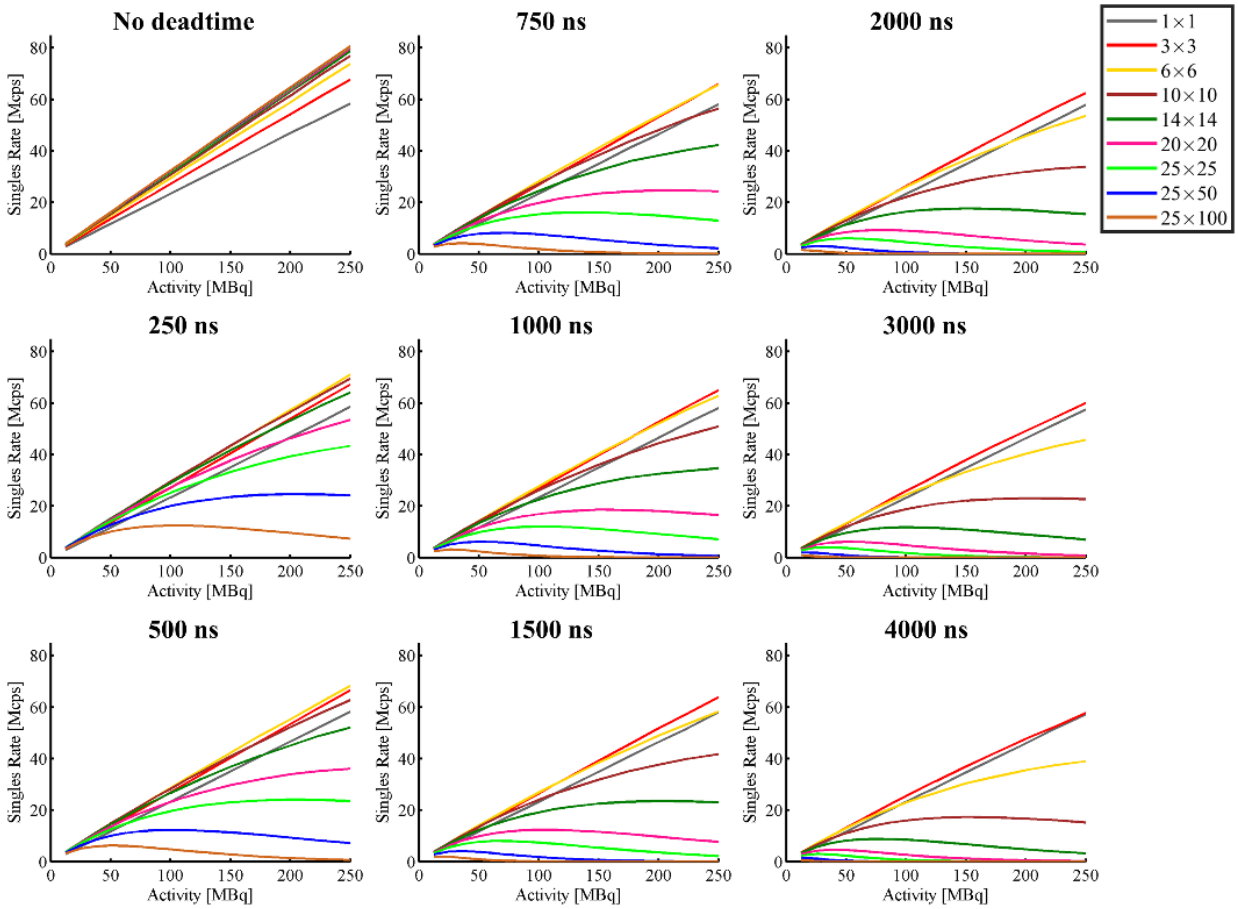

Figure 5.4. Plots of the scanner singles rate vs activity in the head phantom for brain-dedicated scanners with a detector thickness of 10 mm and various detector sizes as defined by the legend. A paralyzable deadtime model with various deadtime constants, as shown on the title of each plot, was considered.

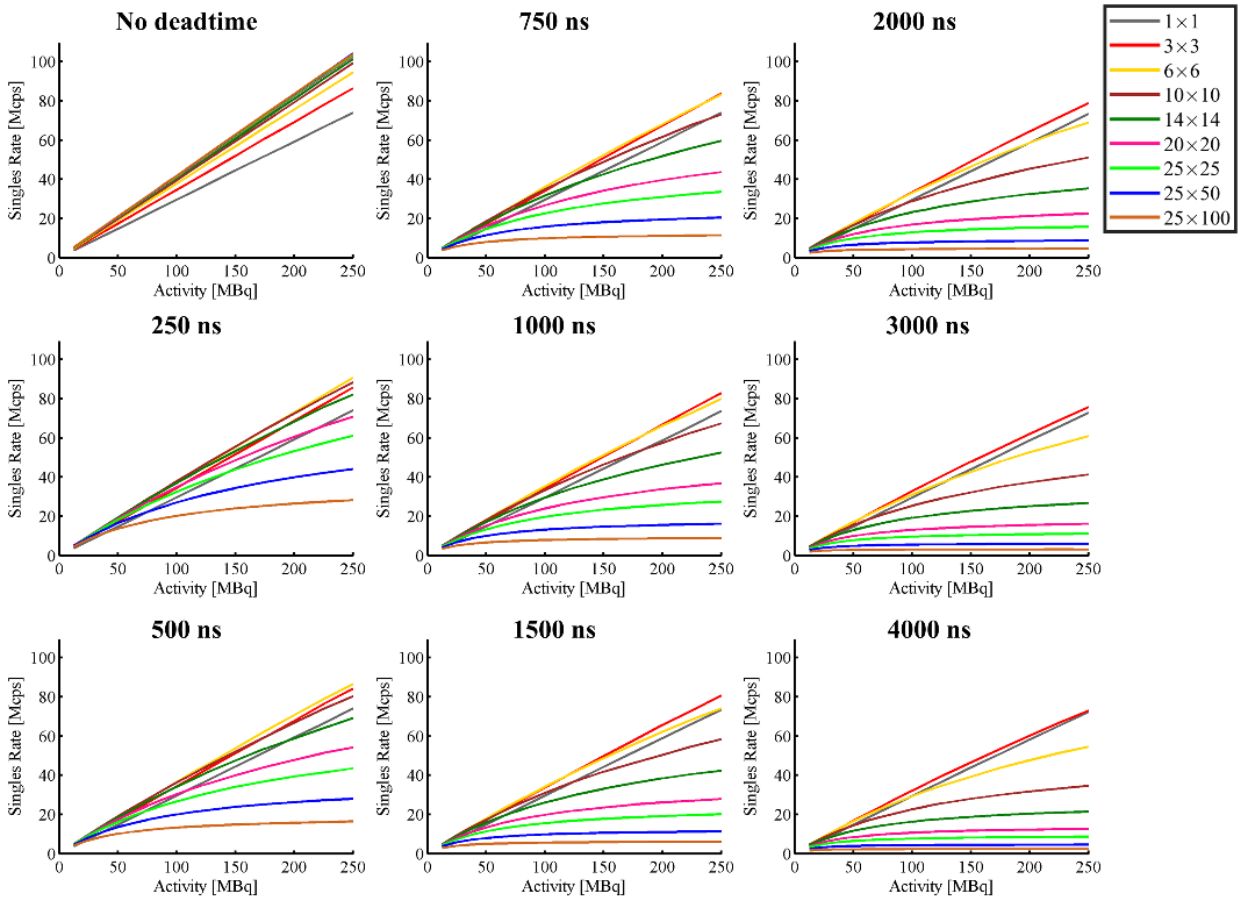

Figure 5.5. Plots of the scanner singles rate vs activity in the head phantom for brain-dedicated scanners with a detector thickness of 15 mm and various detector sizes as defined by the legend. A non-paralyzable deadtime model with various deadtime constants, as shown on the title of each plot, was considered.

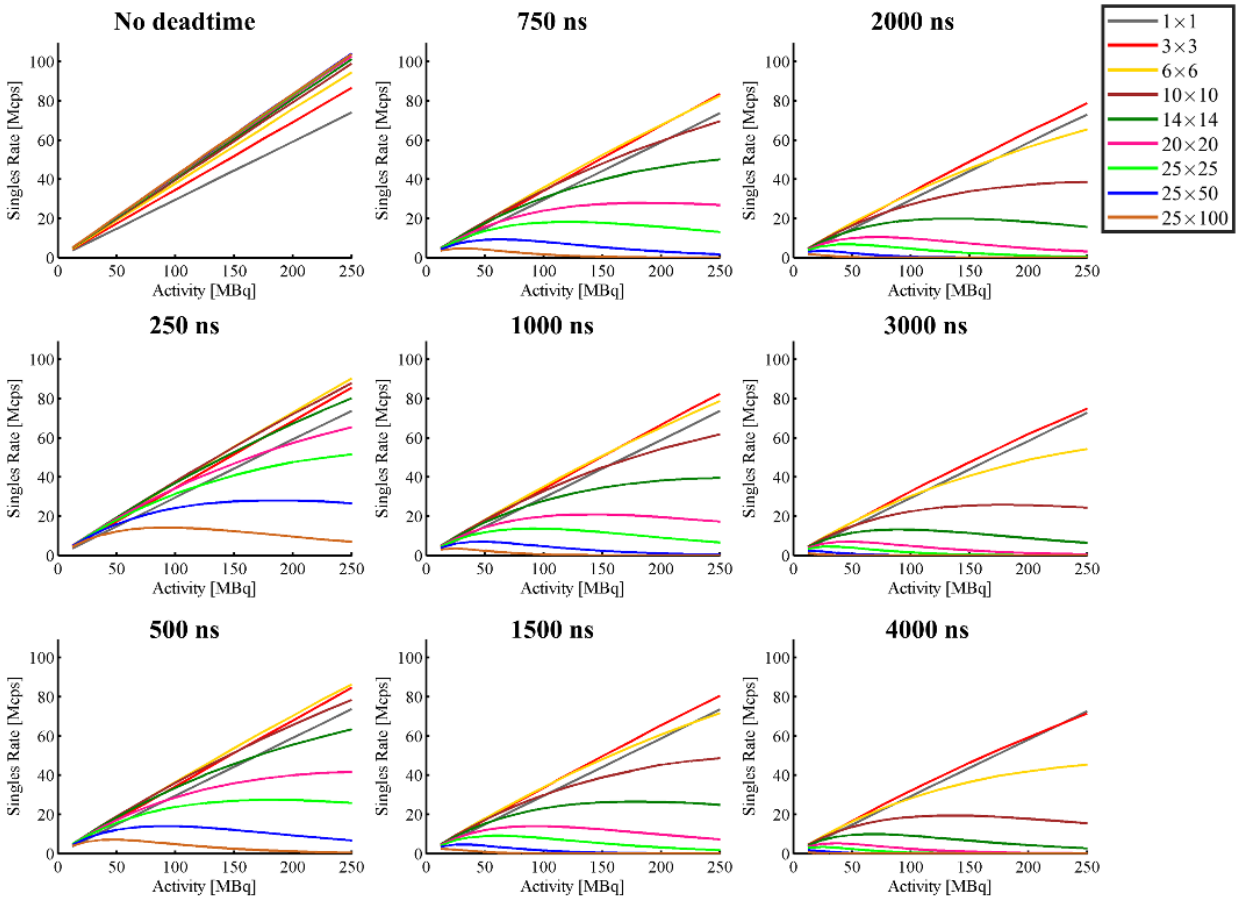

Figure 5.6. Plots of the scanner singles rate vs activity in the head phantom for brain-dedicated scanners with a detector thickness of 15 mm and various detector sizes as defined by the legend. A paralyzable deadtime model with various deadtime constants, as shown on the title of each plot, was considered.

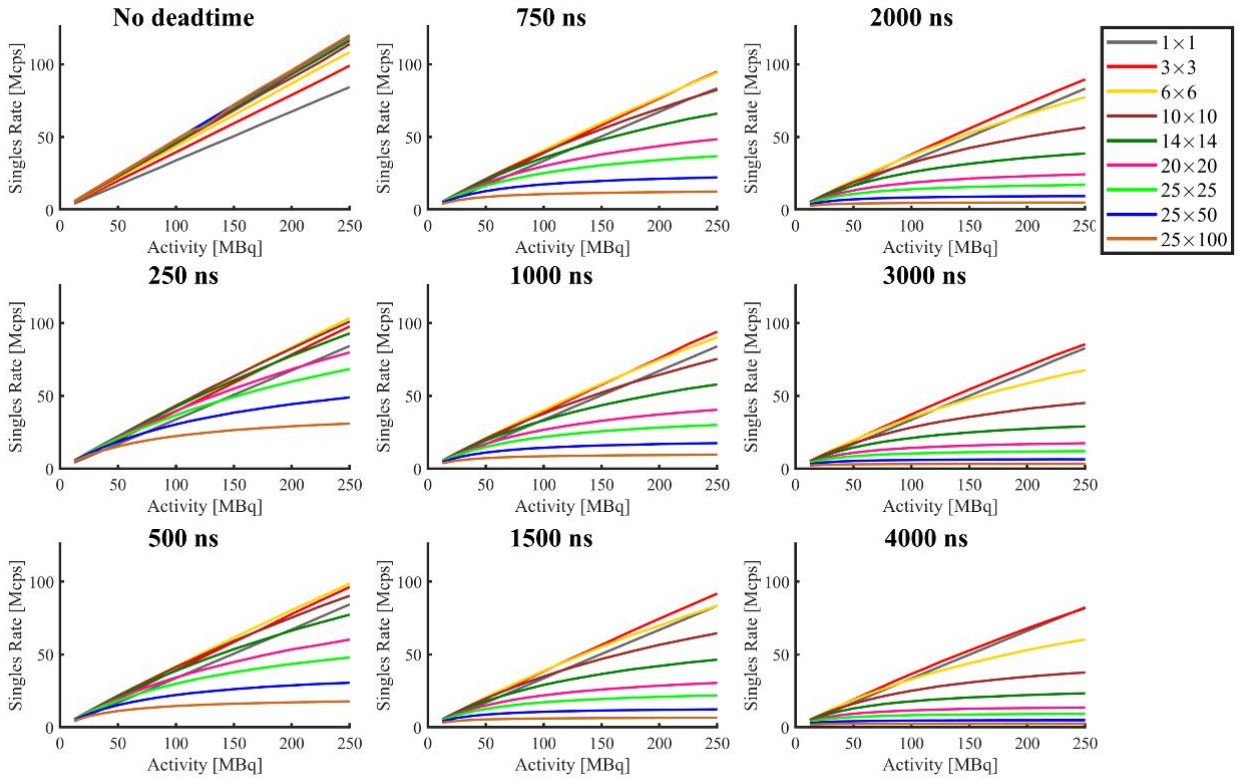

Figure 5.7. Plots of the scanner singles rate vs activity in the head phantom for brain-dedicated scanners with a detector thickness of 20 mm and various detector sizes as defined by the legend. A non-paralyzable deadtime model with various deadtime constants, as shown on the title of each plot, was considered.

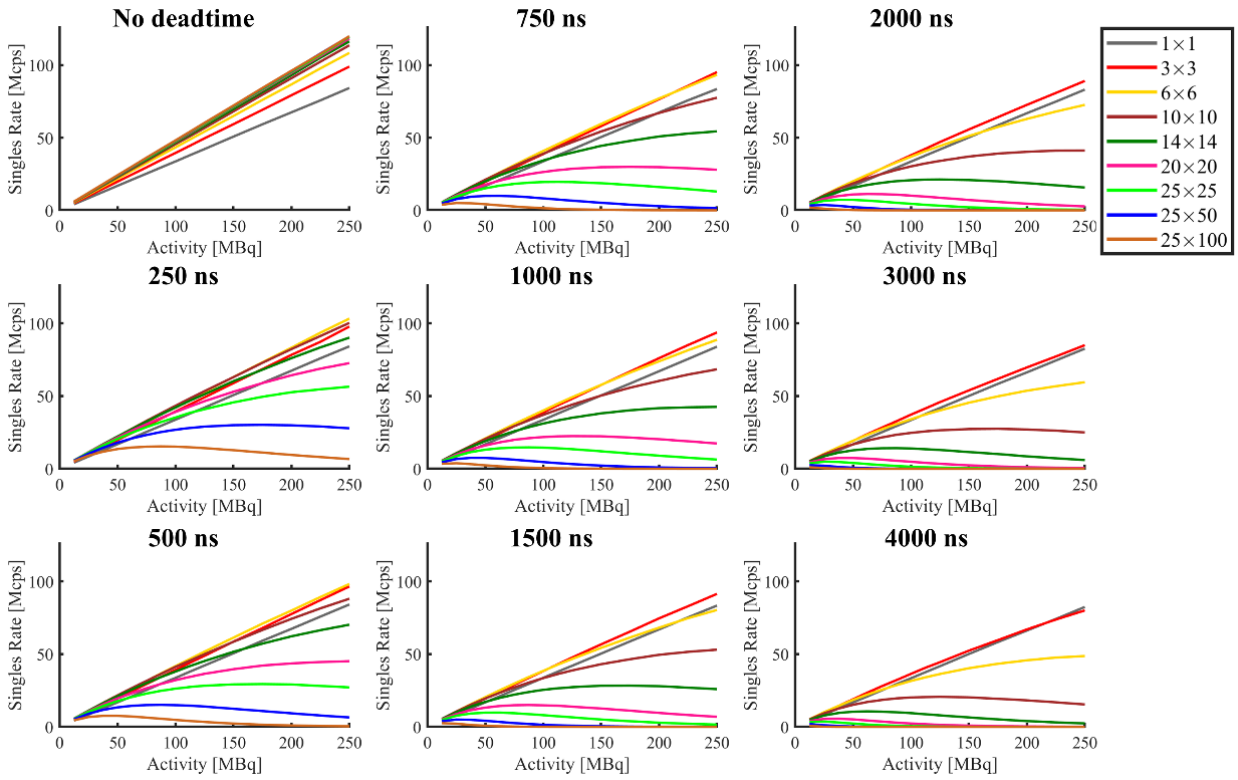

Figure 5.8. Plots of the scanner singles rate vs activity in the head phantom for brain-dedicated scanners with a detector thickness of 20 mm and various detector sizes as defined by the legend. A paralyzable deadtime model with various deadtime constants, as shown on the title of each plot, was considered.

Since the total volume of the scintillator material is constant for scanners with any given detector thickness, one can expect a linear trend between the singles rate and activity when no deadtime is considered. However, as shown in the 'no deadtime' plot of the figures above, the singles rate for systems having detectors with a small number of crystals is lower than that of scanners with a larger detector when an LLD of 350 keV is used. The reason that the singles rate drops for very small detectors is because of the loss of the Compton scattered photons between the detectors (i.e. less of the scatter is intra-detector). In the presence of deadtime, as expected, large detectors demonstrated a loss of count rate, more for paralyzable deadtime models.

To evaluate the effects of LLD on the singles rate, the singles rate was measured for a selection of scanners and an LLD of 0 to 450 keV. The scanners had detector areas of 0.04 to

101.37 cm^2, detector thicknesses of 10 to 20 mm and deadtime constants of 0, 750 and 4000 ns. Figure 5.9 shows plots of scanner singles rate vs. detector area for the case of imaging the head phantom with an activity of 75 MBq (activity concentration of 13.6 kBq/cc).

When no deadtime exists, it can be seen that for any given detector thickness and LLD lower than 250 keV, a smaller detector yields a higher singles rate as each photon may contribute to more than one valid event. For LLD values equal to or greater than 250 keV, however, a larger detector increases the scanner singles rate when no deadtime is present. Introducing deadtime, on the other hand, can sharply reduce the singles rate of the detectors as the detector size increases. For example, a significant drop in the singles rate of the scanner occurs when the detector area exceeds ~9.6 cm^2 (14 × 16 crystals) in the case of a 750 ns paralyzable deadtime or ~2.7 cm^2 (8 × 8 crystals) in case of a 4000 ns paralyzable deadtime.

It can be useful to estimate detector singles rates on a basis of events per unit area per unit activity in the phantom in order to quickly estimate the expected event rate for an arbitrary detector size. In this work, a scanner with a detector array of 15 × 16 crystals (~10 cm^2) and 10/ 15/ 20 mm thickness with no deadtime and LLD of 0 keV resulted in event rates (singles rate) of 394/ 451/ 483 cps/cm^2/MBq when the head phantom source was simulated. When an LLD of 350 keV is set, the event rate drops to 157/ 203/ 233 cps/cm^2/MBq. It should be noted that the reported events rates per unit area were calculated assuming no gaps between the adjacent crystal elements and therefore can be considered as the singles rate in a monolithic detector.

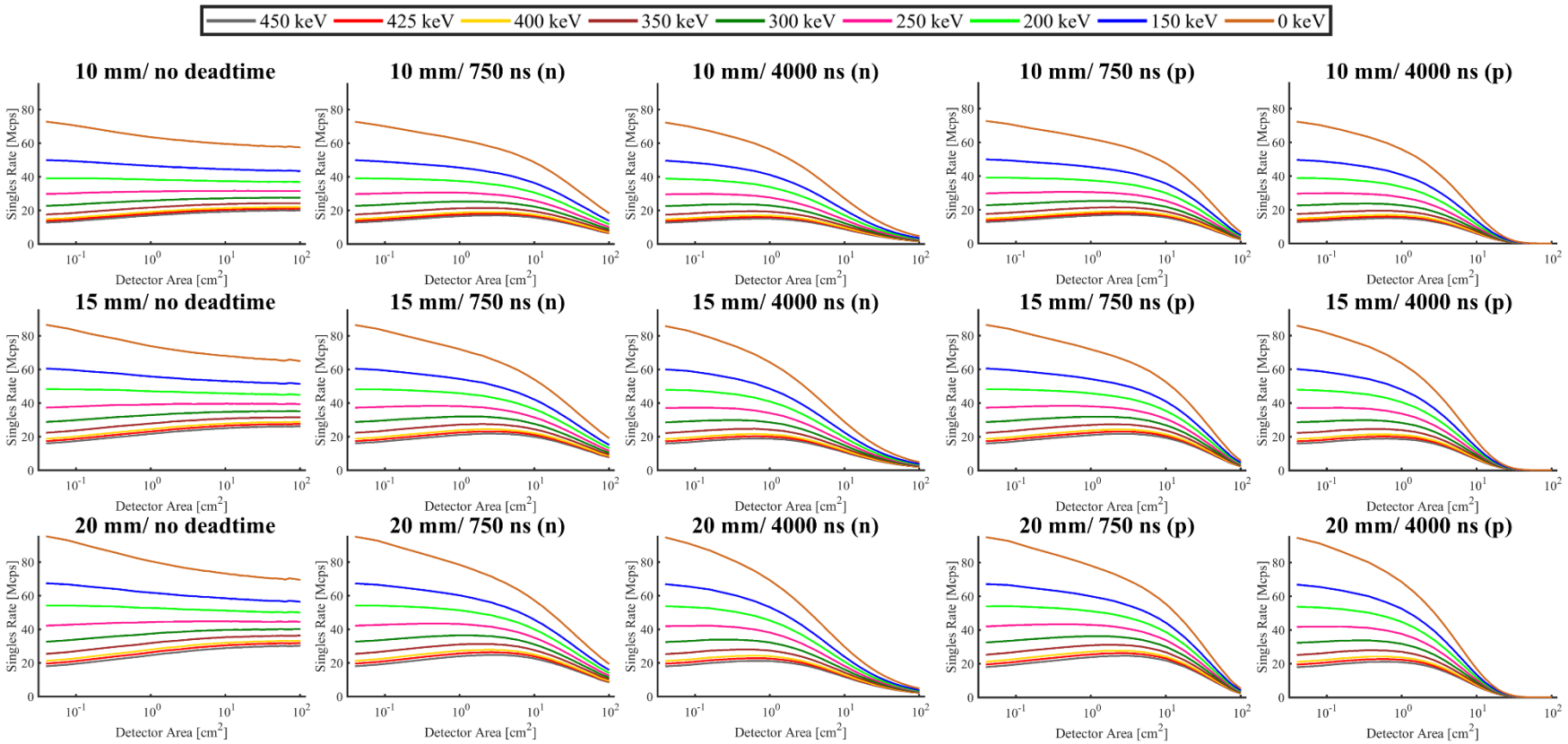

Figure 5.9. Plots of scanner singles rate vs detector area on a logarithmic scale for brain-dedicated scanners with detectors of various thicknesses and deadtime constants. The singles rates were obtained using the head phantom with 75 MBq of activity. Each color represents a different LLD as defined by the legend. The title of each plot indicates the detector thickness, deadtime constant and (if applicable) deadtime model (p for paralyzable and n for non-paralyzable) of the corresponding scanner, respectively.

5.5.2 **Scanner Sensitivity**

The results of the Singles Rate study discussed in Section 5.5.1 suggested that the loss of inter-crystal scattered photons in small detectors can lead to a reduction in detection sensitivity when an LLD of 250 keV or larger is used. On the other hand, larger detectors may also cause a loss in the scanner sensitivity due to a rather significant influence of deadtime, even though sensitivity is measured with a low activity source.

In this experiment, the head phantom with 12.5 MBq of activity was used to measure scanner sensitivity at a fixed LLD of 350 keV and a CTW of 8 ns. Figure 5.10 shows the sensitivity plots of scanners for all detector thicknesses, detector areas and deadtime models and constants considered in this work. It can be seen that, for cases where no deadtime is present (black triangles), sensitivity increases with detector area as more inter-crystal scattered photons are accepted as valid singles and stabilizes once detectors get larger than ~10 cm^2 (15 × 16 crystals). Introducing deadtime noticeably reduces the sensitivity for scanners with detectors larger than ~4 cm^2 (10 × 10 crystals), with larger detectors showing an increasing event loss.

It was observed that scanners with a paralyzable deadtime model have slightly lower sensitivity than an equivalent scanner with a non-paralyzable deadtime model. It should also be noted that, as expected, scanners with a thicker detector have a higher sensitivity. For instance, a 10/ 15/ 20 mm thick detector with 10 × 10 crystal array having 1000 ns non-paralyzable deadtime gives a sensitivity of 1.1%/ 1.9%/ 2.5%.

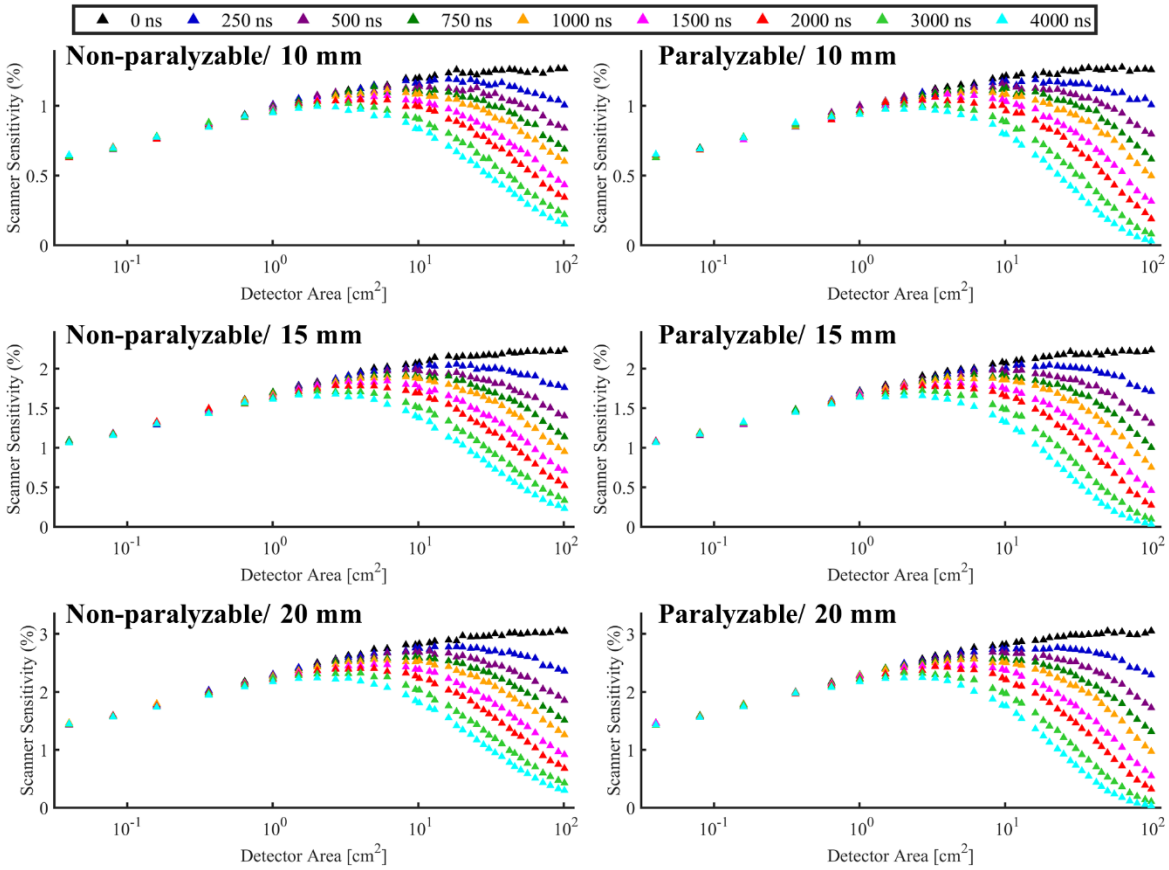

Figure 5.10. Sensitivity plots of scanners vs detector area on a logarithmic scale for various detector thicknesses and deadtime models. The title of each plot indicates the deadtime model and detector thickness of the plot and different colors indicate different deadtime constants as defined by the legend. An activity of 12.5 MBq in the head phantom with an LLD of 350 keV and a CTW of 8 ns was used.

5.5.3 **Noise Equivalent Count Rate**

Noise equivalent count rates for scanners with a detector thickness of 10 mm and crystal array sizes ranging from 1 × 1 to 25 × 100 crystals are shown in Figure 5.11 and Figure 5.12 for non-paralyzable and paralyzable deadtime models, respectively. In these NECR plots, a CTW of 8 ns, an LLD of 350 keV and the head phantom were used. Similar NECR graphs are shown in Figure 5.13 (non-paralyzable) and Figure 5.14 (paralyzable) for a detector thickness of 15 mm and in Figure 5.15 (non-paralyzable) and Figure 5.16 (paralyzable) for a detector thickness of 20 mm.

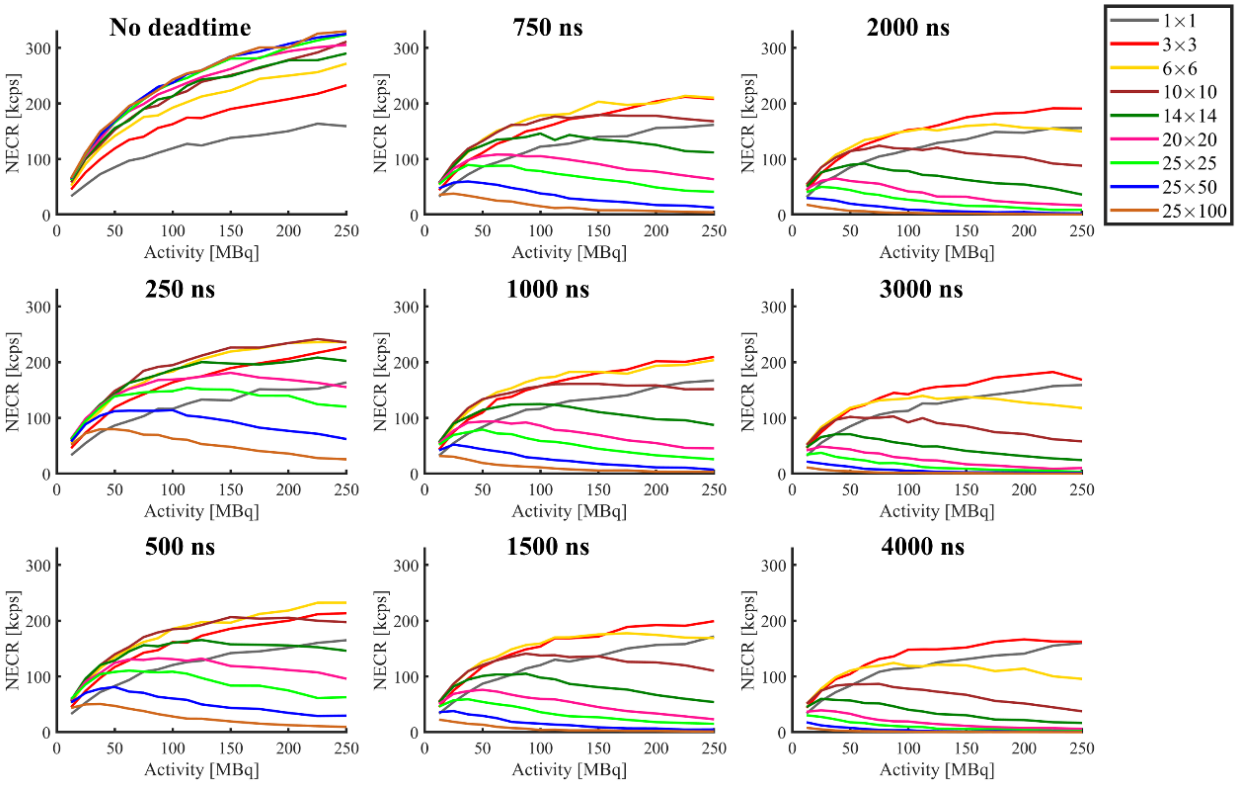

Figure 5.11. Plots of the scanner NECR vs activity in the head phantom for brain-dedicated scanners with detector blocks of thickness 10 mm and selected detector sizes as defined by the legend. A non-paralyzable deadtime model with deadtime constants shown on the title of each plot was considered.

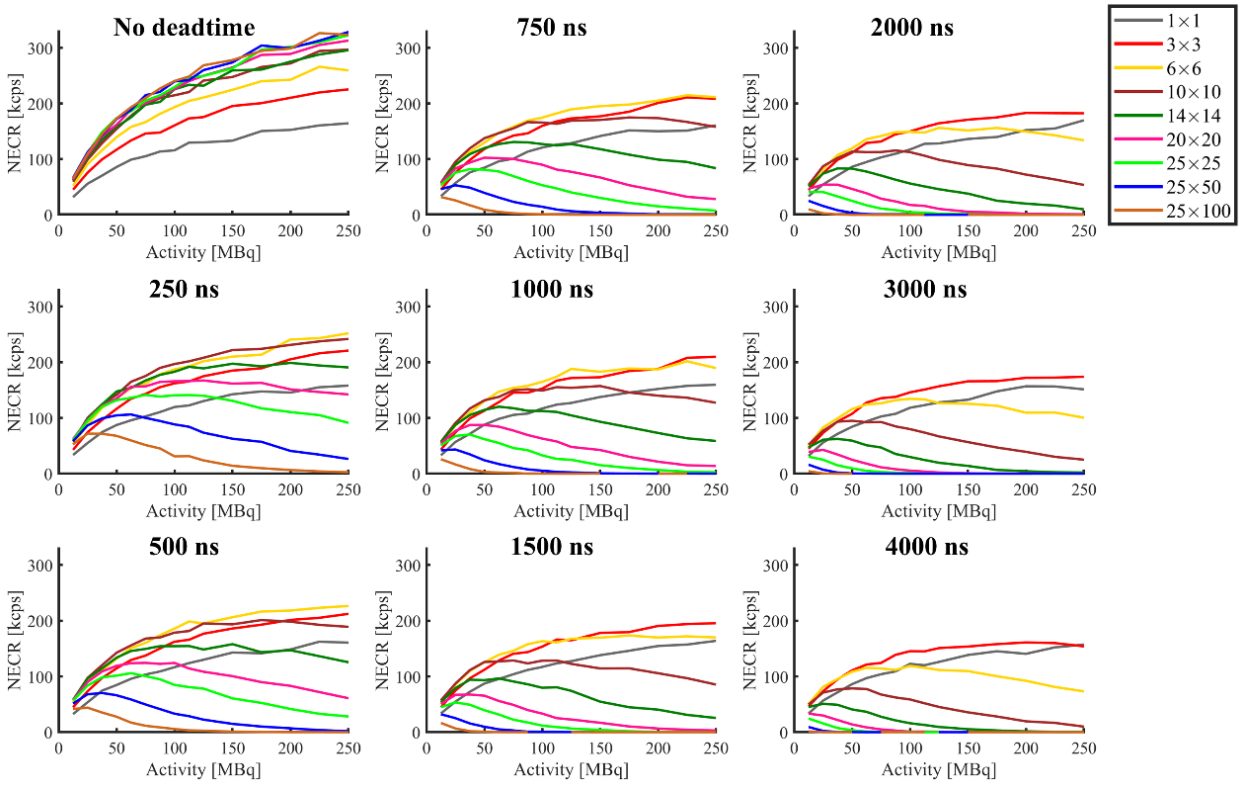

Figure 5.12. Plots of the scanner NECR vs activity in the head phantom for brain-dedicated scanners with detector blocks of thickness 10 mm and selected detector sizes as defined by the legend. A paralyzable deadtime model with deadtime constants shown on the title of each plot was considered.

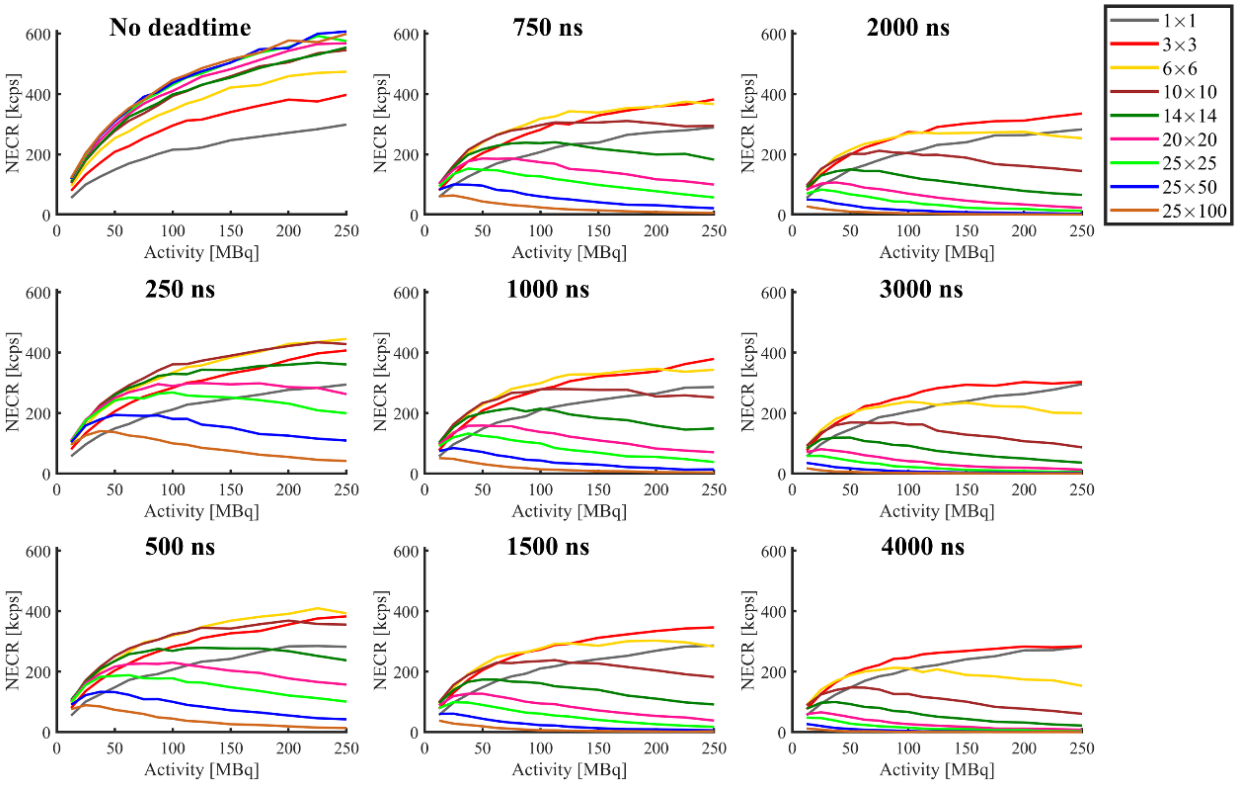

Figure 5.13. Plots of the scanner NECR vs activity in the head phantom for brain-dedicated scanners with detector blocks of thickness 15 mm and selected detector sizes as defined by the legend. A non-paralyzable deadtime model with deadtime constants shown on the title of each plot was considered.

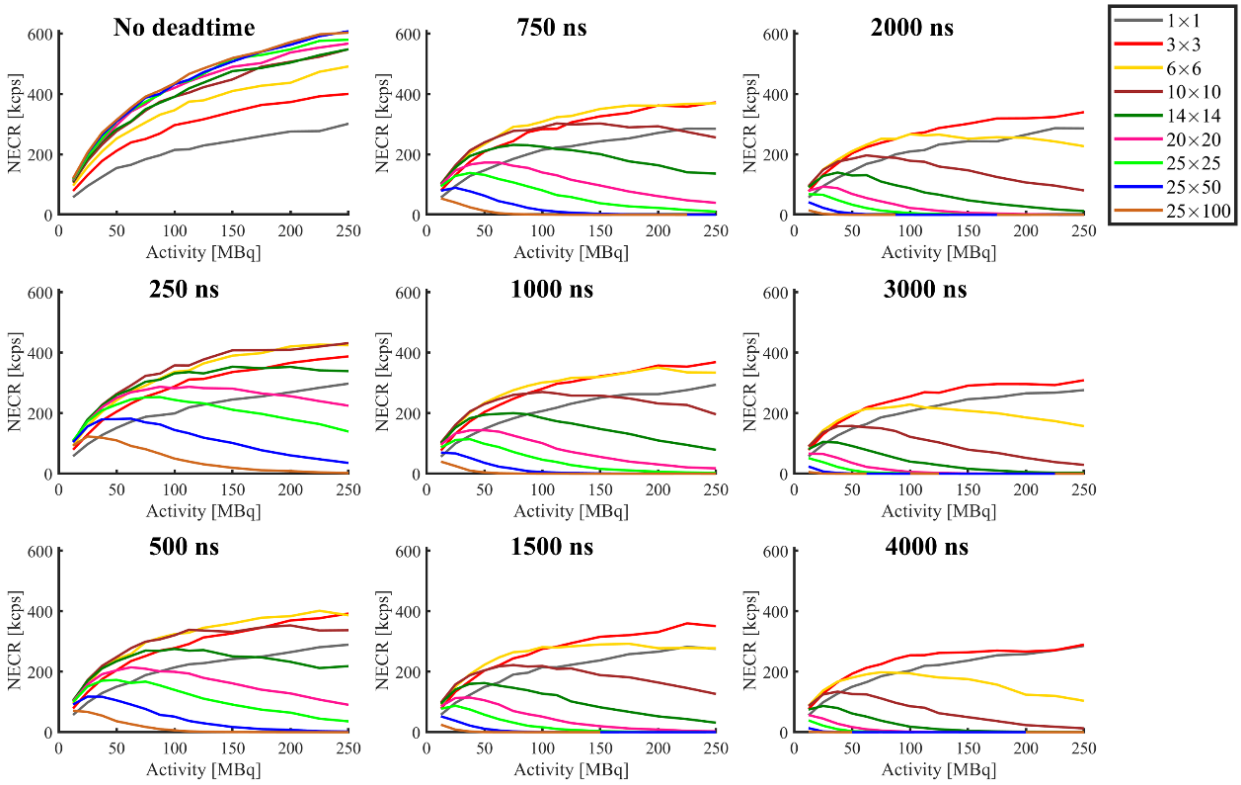

Figure 5.14. Plots of the scanner NECR vs activity in the head phantom for brain-dedicated scanners with detector blocks of thickness 15 mm and selected detector sizes as defined by the legend. A paralyzable deadtime model with deadtime constants shown on the title of each plot was considered.

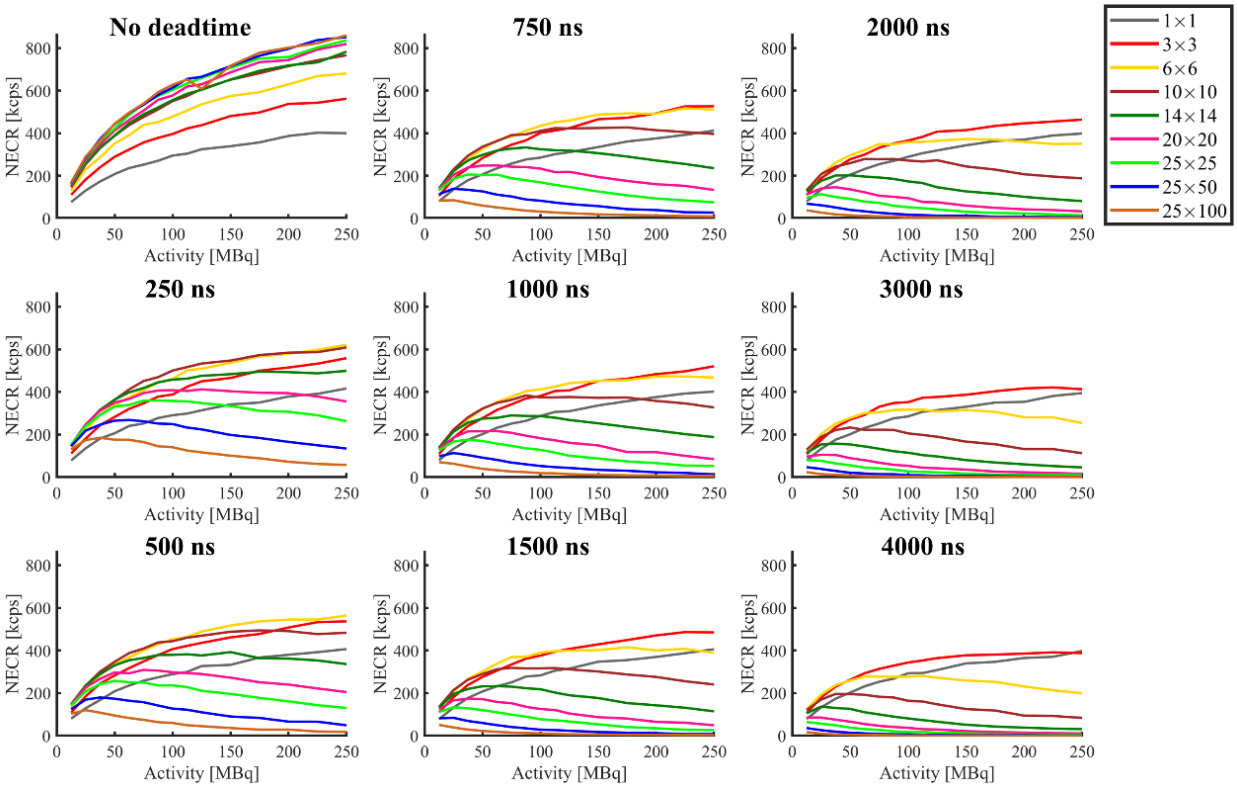

Figure 5.15. Plots of the scanner NECR vs activity in the head phantom for brain-dedicated scanners with detector blocks of thickness 20 mm and selected detector sizes as defined by the legend. A non-paralyzable deadtime model with deadtime constants shown on the title of each plot was considered.

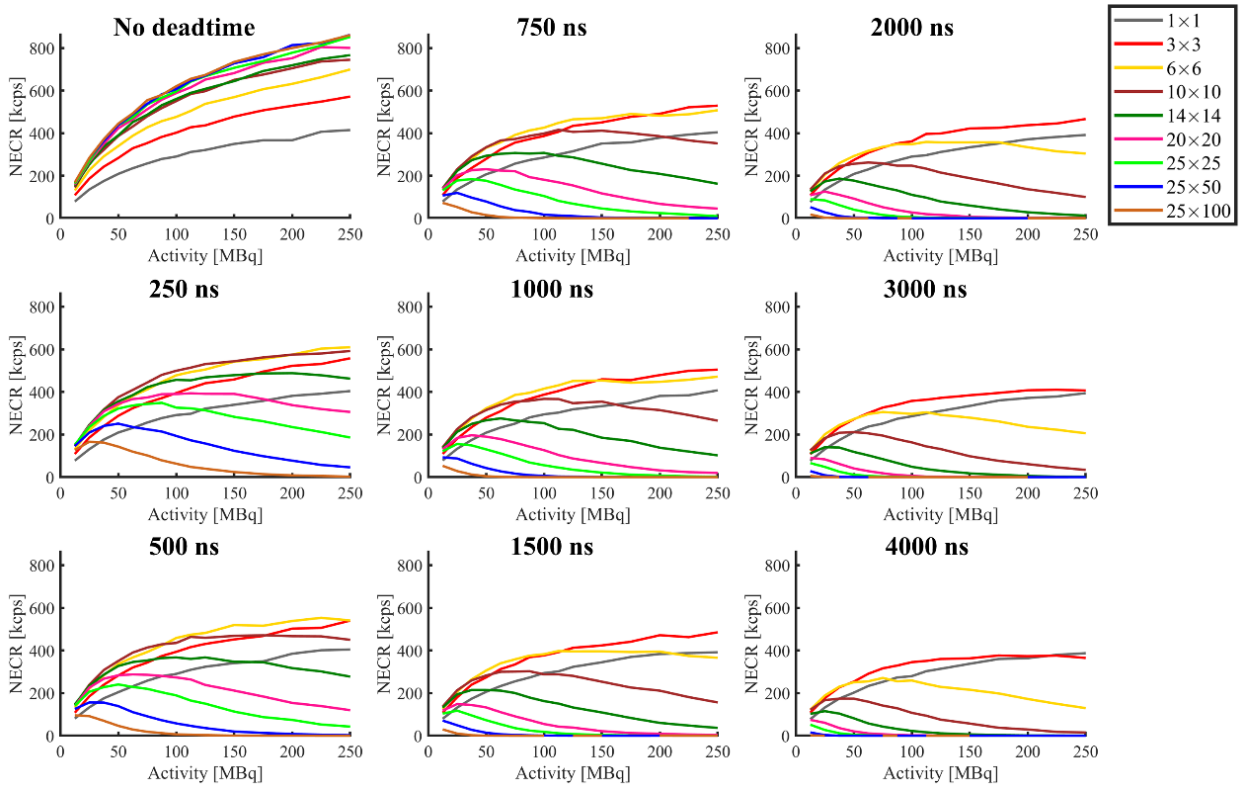

Figure 5.16. Plots of the scanner NECR vs activity in the head phantom for brain-dedicated scanners with detector blocks of thickness 20 mm and selected detector sizes as defined by the legend. A paralyzable deadtime model with deadtime constants shown on the title of each plot was considered.

When no deadtime is present, NECR simply increases with activity and ultimately is limited by the randoms rate, so the use of a large detector is justified as it improves the NECR. However, deadtime rapidly effects the NECR curve as the deadtime constant increases, especially for large detectors. Peak NECR, peak NECR activity and peak NECR activity concentration of brain-dedicated scanners with detector thicknesses of 10 to 20 mm and selected detector sizes and deadtime constants were extracted from the simulation data and are summarized in Table 5.2 for non-paralyzable deadtimes and in Table 5.3 for paralyzable deadtimes. In these tables, detectors with crystal arrays of 6 × 6, 10 × 10, 14 × 14, 20 × 20 and 25 × 25 and deadtime constants of 500, 1000, 2000 and 3000 ns were included.

Table 5.2. Peak NECR and peak NECR activity for selected scanners with various detector thicknesses, detector sizes, deadtime constants and a non-paralyzable deadtime model. The first number shown under the 'Peak NECR activity (act. conc.)' column shows the peak NECR activity of the phantom in MBq, while the second number in parentheses shows the peak NECR activity concentration in kBq/cc.

		Thickness 10 mm		Thickness 15 mm		Thickness 20 mm	
Crystals per block	Deadtime constant [ns]	Peak NECR [kcps]	Peak NECR activity (act. conc.) [MBq] (kBq/cc)	Peak NECR [kcps]	Peak NECR activity (act. conc.) [MBq] (kBq/cc)	Peak NECR [kcps]	Peak NECR activity (act. conc.) [MBq] (kBq/cc)
6 × 6	500	232.38	250 (45.47)	409.90	225 (40.93)	564.15	250 (45.47)
	1000	203.87	250 (45.47)	346.67	200 (36.38)	472.68	200 (36.38)
	2000	162.77	175 (31.83)	274.80	200 (36.38)	372.69	150 (27.28)
	3000	139.76	112.5 (20.46)	237.96	100 (18.19)	316.66	100 (18.19)
10 × 10	500	206.64	150 (27.28)	369.00	200 (36.38)	494.06	175 (31.83)
	1000	161.15	150 (27.28)	280.19	112.5 (20.46)	383.64	87.5 (15.92)
	2000	124.30	75 (13.64)	212.25	75 (13.64)	278.59	87.5 (15.92)
	3000	102.62	87.5 (15.92)	169.59	87.5 (15.92)	232.87	50 (9.09)
14 × 14	500	165.47	125 (22.74)	279.07	125 (22.74)	392.11	150 (27.28)
	1000	124.83	100 (18.19)	216.26	75 (13.64)	289.84	75 (13.64)
	2000	92.01	62.5 (11.37)	150.47	50 (9.09)	202.00	50 (9.09)
	3000	70.85	50 (9.09)	119.28	50 (9.09)	157.37	37.5 (6.82)
20 × 20	500	132.83	87.5 (15.92)	230.19	100 (18.19)	308.72	75 (13.64)
	1000	93.70	62.5 (11.37)	159.11	50 (9.09)	216.60	62.5 (11.37)
	2000	65.38	37.5 (6.82)	107.11	37.5 (6.82)	145.88	37.5 (6.82)
	3000	48.55	25 (4.55)	80.85	25 (4.55)	105.17	25 (4.55)
25 × 25	500	110.57	62.5 (11.37)	188.62	62.5 (11.37)	257.40	50 (9.09)
	1000	79.14	50 (9.09)	132.69	37.5 (6.82)	175.97	37.5 (6.82)
	2000	50.20	25 (4.55)	83.43	25 (4.55)	113.20	25 (4.55)
	3000	37.49	25 (4.55)	58.79	12.5 (2.27)	80.21	12.5 (2.27)

Table 5.3. Peak NECR and peak NECR activity for selected scanners with various detector thicknesses, detector sizes, deadtime constants and a paralyzable deadtime model. The first number shown under the 'Peak NECR activity (act. conc.)' column shows the peak NECR activity of the phantom in MBq, while the second number in parentheses shows the peak NECR activity concentration in kBq/cc.

		Thickness 10 mm		Thickness 15 mm		Thickness 20 mm	
Crystals per block	Deadtime constant [ns]	Peak NECR [kcps]	Peak NECR activity (act. conc.) [MBq] (kBq/cc)	Peak NECR [kcps]	Peak NECR activity (act. conc.) [MBq] (kBq/cc)	Peak NECR [kcps]	Peak NECR activity (act. conc.) [MBq] (kBq/cc)
6 × 6	500	226.61	250 (45.47)	401.61	225 (40.93)	553.15	225 (40.93)
	1000	201.62	225 (40.93)	350.96	200 (36.38)	471.72	250 (45.47)
	2000	156.26	125 (22.74)	266.96	100 (18.19)	358.80	112.5 (20.46)
	3000	134.39	100 (18.19)	228.80	100 (18.19)	305.66	75 (13.64)
10 × 10	500	201.31	175 (31.83)	353.19	200 (36.38)	471.29	175 (31.83)
	1000	157.38	150 (27.28)	270.16	100 (18.19)	367.40	100 (18.19)
	2000	115.67	87.5 (15.92)	196.98	62.5 (11.37)	262.41	62.5 (11.37)
	3000	94.87	50 (9.09)	157.94	50 (9.09)	211.26	50 (9.09)
14 × 14	500	157.99	150 (27.28)	275.52	100 (18.19)	367.35	125 (22.74)
	1000	120.13	62.5 (11.37)	200.17	75 (13.64)	275.65	62.5 (11.37)
	2000	83.37	37.5 (6.82)	139.93	37.5 (6.82)	186.60	37.5 (6.82)
	3000	62.45	37.5 (6.82)	105.18	25 (4.55)	140.64	25 (4.55)
20 × 20	500	124.50	75 (13.64)	214.73	62.5 (11.37)	287.73	62.5 (11.37)
	1000	87.40	50 (9.09)	143.69	50 (9.09)	196.63	37.5 (6.82)
	2000	54.19	25 (4.55)	93.24	25 (4.55)	125.50	25 (4.55)
	3000	42.82	25 (4.55)	66.18	12.5 (2.27)	89.91	12.5 (2.27)
25 × 25	500	105.93	62.5 (11.37)	172.84	50 (9.09)	240.45	50 (9.09)
	1000	70.25	37.5 (6.82)	114.21	37.5 (6.82)	155.85	25 (4.55)
	2000	40.96	25 (4.55)	68.37	12.5 (2.27)	89.27	12.5 (2.27)
	3000	30.43	12.5 (2.27)	51.26	12.5 (2.27)	66.01	12.5 (2.27)

5.5.4 **Coincidence Timing Window**

Graphs of scanner NECR at 75 MBq in the head phantom against detector area with CTW values of 4, 8 and 12 ns are shown in Figure 5.17. An LLD of 350 keV and a paralyzable deadtime of 750 ns were assumed in obtaining these data. Since simulations were done with no blurring on the timing data (i.e. ideal timing resolution), it is expected that a smaller CTW improves the NECR. It was seen that reducing CTW from 12 ns to 8 ns can increase the NECR by an average (averaged over all detector areas) of (24.8 ± 6.8)%, (23.1 ± 7.8)% and (21.9 ± 6.1)% for detectors with thicknesses of 10, 15 and 20 mm, respectively. Furthermore, reducing the CTW from 8 ns to 4 ns can increase the NECR by an average of (37.5 ± 13.0)%, (37.9 ± 15.4)% and (34.8 ± 5.9)% for detectors with thicknesses of 10, 15 and 20 mm, respectively.

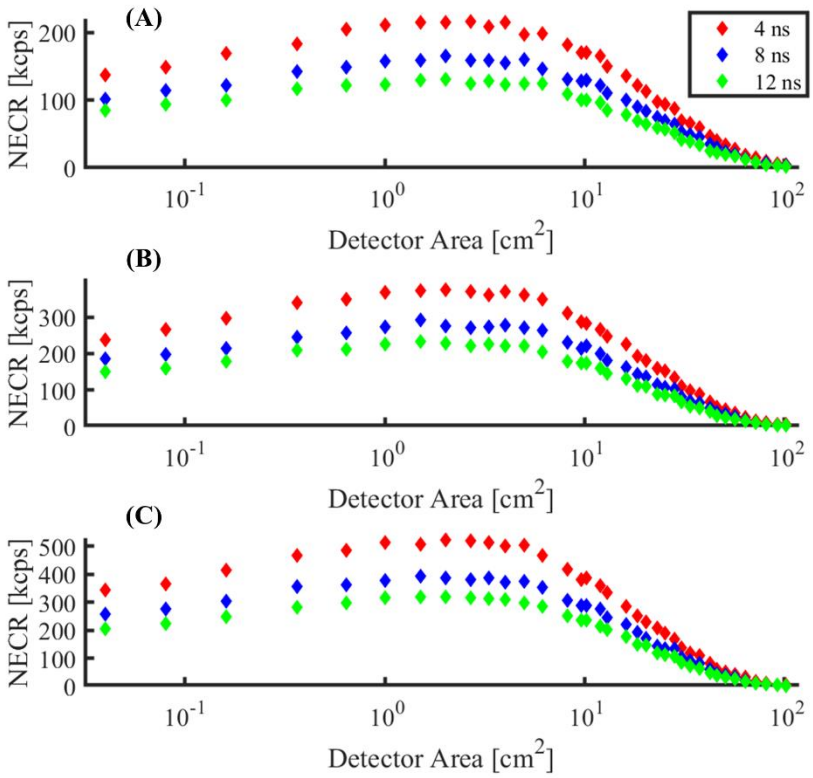

Figure 5.17. Plots of NECR at 75 MBq in the head phantom vs detector area on a logarithmic scale for different CTW values. The graphs correspond to scanners with detector thicknesses of (A) 10, (B) 15 and (C) 20 mm. An LLD of 350 keV with a paralyzable deadtime of 750 ns were considered. Each marker represents a different CTW as defined by the legend.

5.5.5 **Lower Level Discriminator**

To demonstrate the effects of LLD on the NECR and scanner sensitivity, plots of NECR at 75 MBq in the head phantom vs scanner sensitivity for LLDs of 0, 250, 300, 350, 400, 425 and 450 keV are shown in Figure 5.18 for scanners with detector thicknesses of 10, 15 and 20 mm and paralyzable and non-paralyzable deadtime models. In this experiment, only scanners with a detector area of 4.03 cm^2 (10 × 10 crystals), a deadtime constant of 750 ns and a CTW of 8 ns were included. The ULD was kept fixed at 650 keV.

It can be seen that reducing the LLD from 450 keV to 300 keV increases scanner sensitivity while rapidly degrading the NECR. Decreasing the LLD to less than 300 keV however, increases the NECR again while improving scanner sensitivity. To better understand the reason behind the

NECR rise for low LLDs, the count rate of each coincidence type was investigated for one selected experiment. The rate of true, scatter, random and prompt coincidences as well as the NECR for one of the mentioned experiments corresponding to a detector thickness of 20 mm and a deadtime type of non-paralyzable are shown in Table 5.4. The data suggest that reducing the LLD from 450 keV to 300 keV keeps the contribution of random coincidences virtually stable while introducing a decrease in the ratio of true coincidences and an increase in the ratio of scatter coincidences. The lower ratio of true coincidences and higher ratio of scatter and random coincidences degrade the NECR. However, further reducing the LLD from 300 keV towards no energy cut severely reduces the ratio of true coincidences, increases the ratio of random coincidences and maintains the ratio of scatter coincidences almost identical at ~25%. The change in behavior for LLD values lower than ~300 keV is due to the fact that scattering at 511 keV is mainly forward directed and less probable for low energy back-scattered photons. Also, it should be noted that the single scattered photons in the phantom have a minimum energy of 170.3 keV. In other words, the NECR increase for a low LLD is because the scatter contribution remains virtually steady and only the random coincidences rate increases once the LLD reduces to below 300 keV.

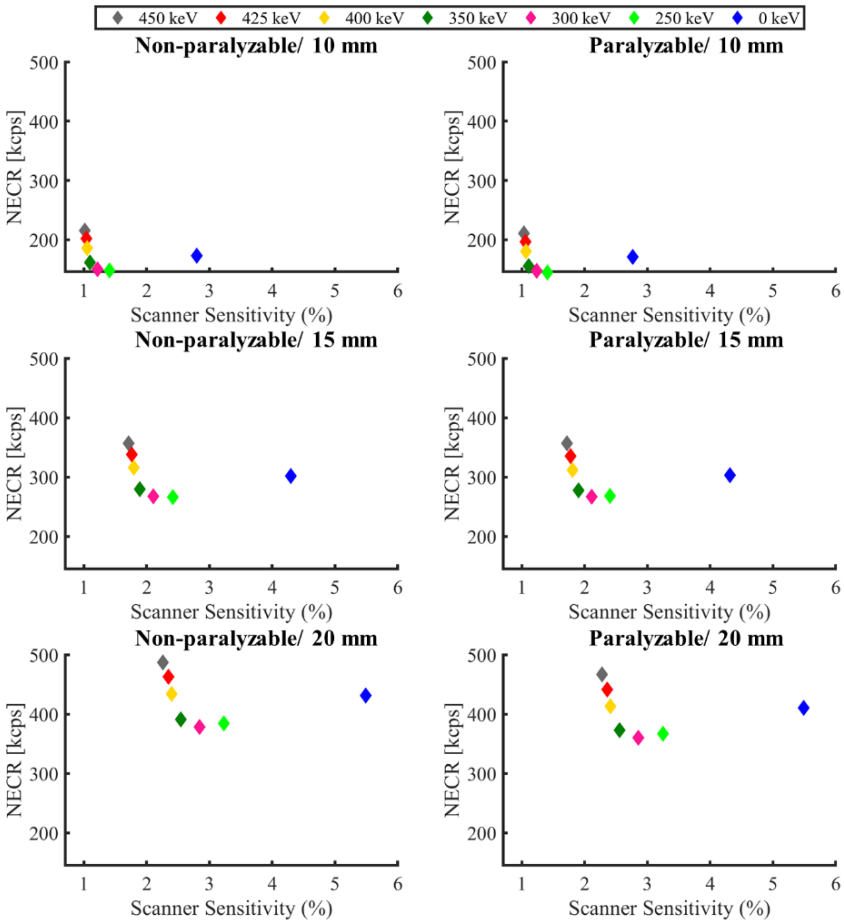

Figure 5.18. Plots of NECR at 75 MBq in the head phantom (activity concentration of 13.6 kBq/cc) vs sensitivity for scanners with various deadtime models and detector thicknesses as shown in each plot's title (detector thickness is written in parentheses). Selected brain-dedicated PET scanners with a detector of 10 × 10 crystal arrays, deadtime constant of 750 ns, CTW of 8 ns and various LLDs, as defined by the legend, were chosen.

Table 5.4. The count rate of different coincidence types in kcps with their ratios (shown in parentheses) and NECR for a selected scanner (detector thickness of 20 mm, 10 × 10 crystals per detector, non-paralyzable deadtime of 750 ns) obtained with the head phantom with an activity of 75 MBq. The numbers shown on the top of the table indicate the LLD values.

	0 keV	250 keV	300 keV	350 keV	400 keV	425 keV	450 keV
True	3384.98 (12.75%)	1993.80 (19.28%)	1746.45 (21.65%)	1568.40 (24.91%)	1480.95 (29.29%)	1450.20 (31.89%)	1398.98 (34.82%)
Random	16327.88 (61.50%)	5726.40 (55.39%)	4370.55 (54.18%)	3369.60 (53.52%)	2700.38 (53.42%)	2438.03 (53.62%)	2166.75 (53.94%)
Scatter	6837.30 (25.75%)	2618.85 (25.33%)	1949.33 (24.17%)	1358.55 (21.58%)	873.75 (17.29%)	659.03 (14.49%)	451.50 (11.24%)
Prompt	26550.15 (100%)	10339.05 (100%)	8066.33 (100%)	6296.55 (100%)	5054.85 (100%)	4547.25 (100%)	4017.23 (100%)
NECR	431.56	384.49	378.13	390.67	433.88	462.46	487.18

5.5.6 **Outside FOV Activity**

Figure 5.19 shows the scanner singles rate vs detector area for scanners with detector thicknesses of 10, 15 and 20 mm when the head and whole-body phantoms had an activity concentration of 13.6 kBq/cc (i.e. 75 MBq in the head phantom and 300 MBq in the whole-body phantom). Paralyzable and non-paralyzable deadtime models with a deadtime constant of 750 ns, a CTW of 8 ns and an LLD of 350 keV were considered to acquire the singles data. The singles rate plots show that the outside activity significantly increases the scanner singles rate for detectors of up to ~25 cm^2. The count rate performance of larger detectors is severely dominated by the effects of deadtime, resulting in deadtime losses such that the body phantom singles rate is less than the head phantom singles rate.

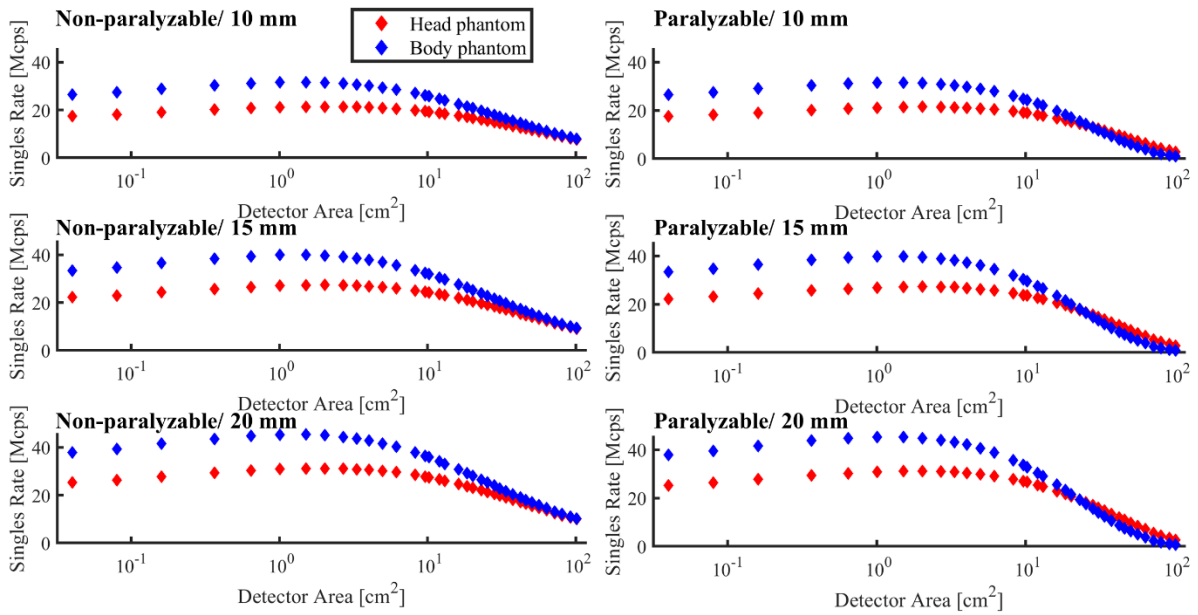

Figure 5.19. Plots of scanner singles rate vs detector area on a logarithmic scale for the head and body phantom (as defined by the legend) at an activity concentration of 13.6 kBq/cc in the phantoms. The plots correspond to scanners with a paralyzable (right) and non-paralyzable (left) deadtime with a deadtime constant of 750 ns, a CTW of 8 ns, an LLD of 350 keV and different detector thicknesses as shown in the title of each plot.

The increase in the singles rate contributes to the rate of random coincidences and affects the NECR of the scanner. The rate of true, scatter and random coincidences as well as the NECR

at 13.6 kBq/cc for PET scanners with various detector geometries are shown in Figure 5.20 for the head and whole-body phantoms. Similarly, a CTW of 8 ns and an LLD of 350 keV were applied in this section. Only the results for detectors with a paralyzable deadtime of 750 ns are shown in Figure 5.20.

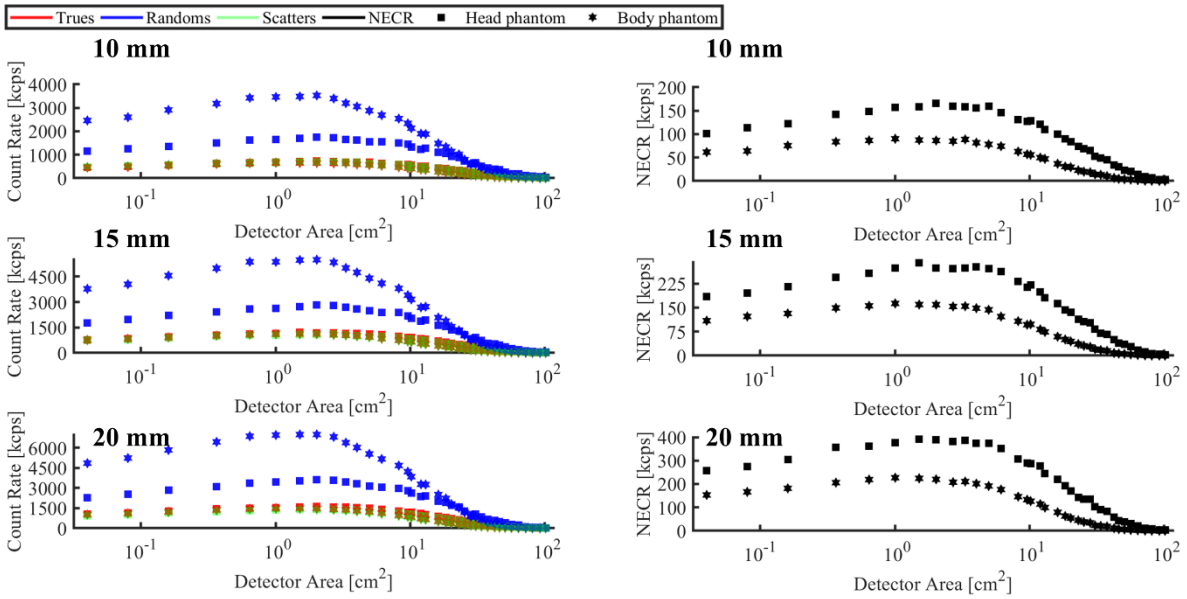

Figure 5.20. Plots of the true, random and scatter coincidence rate (left) with NECR (right) calculated at an activity concentration of 13.6 kBq/cc for scanners with detector designs of various areas on a logarithmic scale and different phantoms. An LLD of 350 keV, a CTW of 8 ns and a paralyzable deadtime of 750 ns were used for all plots. The graphs from top to bottom correspond to scanners with detector thicknesses of 10, 15 and 20 mm, respectively, as indicated on the top of each plot. Each marker and line type represent a different type of count rate and phantom as defined by the legend.

Generally, for any given detector thickness, enlarging the detector area led to a higher contribution of random coincidences which in turn makes the NECR more degraded by the outside activity. For instance, 10 mm detectors with 3 × 3 (area = 0.36 cm^2), 10 × 10 (area = 4.03 cm^2) and 20 × 20 (area = 16.17 cm^2) crystal arrays experienced 41.2%, 47.3% and 84.3% reductions in their NECR, respectively, when the head phantom was substituted with the body phantom. A similar trend was seen for detectors with thicker scintillators, although a thicker detector generally yielded a higher absolute NECR. Detectors with a thickness of 15 mm and 3 × 3 (area = 0.36 cm^2), 10 ×

10 (area = 4.03 cm^2) and 20 × 20 (area = 16.17 cm^2) crystal arrays experienced 39.0%, 46.6% and 63.4% reductions in their NECR, respectively, when activity outside the FOV was introduced. Lastly, the reduction rates for the same order of 20 mm thick detectors were 42.2%, 46.0% and 64.1% due to the introduction of activity outside the FOV.

5.6 Discussion and Conclusion

In Chapter 4, an investigation was reported on the effects of detector geometrical factors on the performance of a brain-dedicated PET insert from the point of view of the scanner resolution and resolution uniformity given a fixed detector area. In this chapter, the study has been extended to evaluate the effects of the detector size and deadtime on the count rate performance of brain-dedicated PET inserts. The count rate of singles and true, random, scatter and prompt coincidences were recorded for brain-dedicated scanners with detector thicknesses of 10 to 20 mm and areas of 0.04 (1 crystal per detector block) to 101.37 cm^2 (2500 crystals pet detector block). For each detector geometry, a scanner was designed such that the scanner can be retrofit into the Siemens Magnetom 7T brain-dedicated MR scanner.

The singles rate analysis showed that a scanner with detectors of very few crystals (e.g. 1 × 1) may have a lower singles rate than larger detectors because of the lack of the intra-detector scattered photons contribution. The lower singles rate in small detectors in turn causes a reduction in scanner sensitivity as shown in Figure 5.10. On the other hand, singles rate behavior of large detectors is primarily determined by the thickness of the scintillator and their deadtime properties. For large detectors and a given LLD, the singles rate can rapidly level off for non-paralyzable deadtimes, or fall for paralyzablde deadtimes, as the activity increases. As shown in the plots of Figure 5.9, a sudden drop in the singles rate occurs when the detector size increases beyond a threshold. This threshold, for each activity level, is primarily a function of the deadtime constant.

The sensitivity plots shown in Figure 5.10 also suggest that, depending on the deadtime properties of the scanner, large detectors can result in a reduction in scanner sensitivity. It should be mentioned that detectors with a few crystals, such as [203, 204], can theoretically improve the spatial resolution of the scanner, however spatial resolution was not investigated in this chapter.

In several reported experiments in this chapter, selected LLD (often 350 keV) and CTW (often 8 ns) values were used while studying the effects of other variables on the count rate performance of scanners. While the selected default values were not intended to necessarily represent the commonly chosen values in a modern PET scanner, the effects of these variables on the target performance metrics were studied in a relative fashion.

The NECR graphs presented in this chapter showed some noticeable statistical fluctuations. The 'no deadtime' plots of Figure 5.11 and Figure 5.12 (and similarly Figure 5.13 and Figure 5.14 and also Figure 5.15 and Figure 5.16) were in fact obtained from two identical simulations with different random number generator seeds. Therefore, the two plots should be the same when no statistical uncertainties are present and are expected to be smooth. However, the plots are not identical and show variations. Our investigations suggested that these fluctuations originated from somewhat large statistical uncertainties in the number of recorded coincidences, even though each simulation contained a million decays in the FOV. To investigate this issue, NECR at 3 activity levels of 62.5, 75 and 87.5 MBq in the head phantom was measured 5 additional times for a selected scanner. The selected scanner had a detector thickness of 15 mm, detectors with 10 × 10 crystals (area = 4.03 cm^2), a CTW of 8 ns, an LLD of 350 keV and a 2000 ns paralyzable deadtime. Figure 5.21 shows the NECR curve for this scanner with repeated values. While the relative NECR uncertainty (standard deviation over the mean of all 6 measurements) at 62.5 MBq was measured to be 0.5%, the relative NECR uncertainties at 75 and 87.5 MBq were 2.4% and 2.2%, respectively.

This is a quantitative example of the statistical uncertainty that existed in this count rate study. Even though increasing the number of decays could have reduced the uncertainty, only a million decays in the FOV were generated to maintain a reasonable execution time for all simulations.

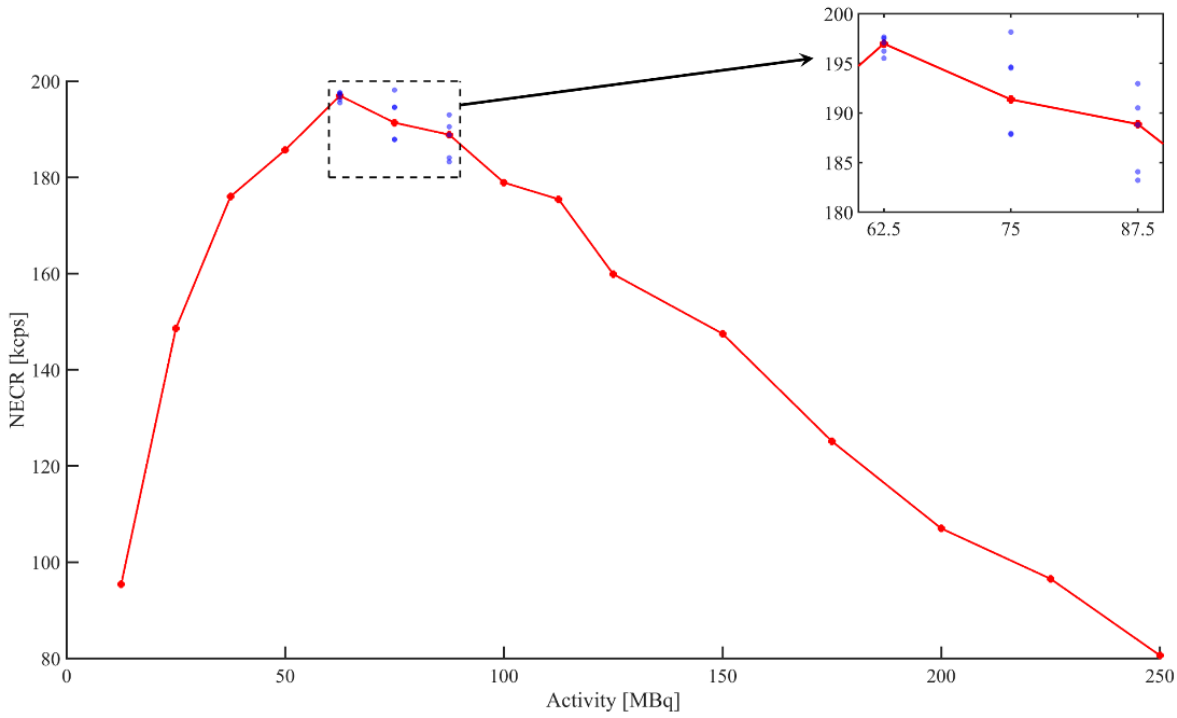

Figure 5.21. A plot of the NECR curve and repeated NECR values (five times) for 3 selected activity values in the head phantom using a chosen scanner. The red curve shows the NECR used in the analysis of this chapter and the blue points indicate the calculated NECR values for each repeated simulation.

The NECR plots of scanners with selected detectors provide a quantitative basis on how large detectors can result in a deterioration in the count rate performance of the scanner. Reducing the detector size, on the other hand, increases the number of output channels, the cost of the scanner and the complexity of processing output signals. The results of this chapter showed that, interestingly enough, the NECR performance of detectors with a very small area (less than ~0.64 cm^2 or 4 × 4 crystals per detector) was inferior to larger detectors due to the loss of singles and a reduced scanner sensitivity. The presented count rate performance of detectors with a small area can be of significant interest to other researchers seeking to develop one-to-one coupled detectors.

The data presented in this work quantifies the balance between the count rate performance of a scanner and its detector size for a wide range of detector deadtime models and constants. For example, a scanner with 20 mm thick detectors with 6 × 6 crystals (area = 1.51 cm^2) and 1000 ns paralyzable deadtime has a peak NECR of 471.71 kcps at 45.47 kBq/cc. Increasing the number of crystals per detector to 10 × 10, 14 × 14, 20 × 20 or 25 × 25 (area = 4.03, 8.23, 16.17, 25.34 cm^2, respectively), changes the peak NECR to 383.64 kcps at 15.92 kBq/cc, 289.84 kcps at 13.64 kBq/cc. 216.60 kcps at 11.37 kBq/cc and 175.97 kcps at 6.82 kBq/cc, respectively. The data presented in this work suggest that when detector deadtime is in the range of 750 to 1000 ns (typical values for SiPM based detectors [205, 206]) the optimal detector size is in the range of 1.5 cm^2 to 9.6 cm^2 for the geometry of a brain-dedicated PET insert.

As expected, increasing detector thickness enhanced the peak NECR for the studied scanners. For the same scanner mentioned earlier with detectors of 6 × 6 crystals (area = 1.51 cm^2) and a 1000 ns paralyzable deadtime, the peak NECR increases from 203.87 kcps at 45.47 kBq/cc to 346.67 kcps at 36.38 kBq/cc and 472.68 kcps at 36.38 kBq/cc when detector thickness increases from 10 to 15 and 20 mm, respectively.

The benefits of a narrower CTW are evident for NECR plots shown in Figure 5.17 for a count rate experiment at 75 MBq. Studies such as [207] and [208], investigated a variable CTW for coincidence processing as a function of ring difference and showed that a variable CTW can potentially improve the count rate performance of a scanner. Such techniques were not studied in this work as all studied brain-dedicated scanners have a relatively (relative to a whole-body scanner) short axial length of 20 cm.

A steep reduction in the NECR of the scanner observed when the LLD was reduced from 450 to 300 keV (see Figure 5.18) emphasizes the importance of LLD on the count rate performance

of a scanner. The reason for an improvement in the NECR when the LLD is reduced to less than 300 keV is discussed in detail in Section 5.5.5. However, a more thorough investigation of the effects of using a low LLD on image quality is recommended as a large number of random coincidences may degrade contrast recovery of the images and introduce image noise. Also, as pileup was not explicitly modeled in this work, the effects of piled up low energy events contributing to valid singles were not studied. Nevertheless, the occurrence rate of piled up events at the studied activity level (75 MBq in the head phantom) for this experiment is expected to be sufficiently low. For instance, a 14 × 16 crystal elements detector with an area of 9.6 cm^2 recorded a detector singles rate of about 230 kcps (when no deadtime or LLD is applied) which is a modest count rate with minimal pileup effects for a PET detector with modern data acquisition system.

The greatest contribution of the activity outside the FOV of the scanner was in the rate of random coincidences which in turn degraded the NECR performance of the scanner. As activity outside the FOV can be particularly troubling for brain-dedicated scanners, a shield on the caudal end of the scanner has been proposed in the past. For example, the NeuroShield, an external lead shield that was designed to reduce events from the activity outside the FOV, was suggested in [209]. This device successfully reduced the contribution of random coincidences originating from outside of the FOV activity. As reported in [210], such shields can improve the NECR of a scanner in brain studies by a considerable margin.

Although a large number of detector geometries and count rate experiments were considered in this study, certain limitations also existed. As discussed earlier, one of the main limitations of this work was the encountered statistical limitations to maintain the practicality aspect of the work. Also, the lack of a separate pulse pileup model was one the main limitations of this work. The absence of a study of the image SNR, resolution and resolution uniformity for

the studied scanners is another limitation of this work. Nonetheless, the effects of detector geometrical variables on several image quality related factors have been separately studied in Chapter 4 which complements the results of this study. A complete collection of optical photons was assumed in this work with no further deadtime down the data processing pipe line. In reality though, a partial light collection associated with larger detector blocks, electronic noise and additional deadtime may cause the performance of the scanner to deviate from the performance metrics reported in this study. Also, as the studied scanners are being evaluated for a brain-dedicated PET insert with limited gantry size, no time-of-flight investigation was considered in this work.

In conclusion, a comprehensive investigation of the count rate performance of several scanners suitable for a brain-dedicated PET insert was performed in this chapter. All geometrical assumptions and variables were defined such that the PET insert can be retrofit into the Siemens Magnetom 7T brain-dedicated MR scanner. Geometrical and deadtime characteristics of the detectors were varied to study the performance of a large number of LSO detectors with thicknesses between 10 to 20 mm, areas between 0.04 to 101.37 cm^2, deadtime constants of 0 to 4000 ns with models of paralyzable and non-paralyzable deadtime, LLD values of 0 to 450 keV, CTW values of 4 to 12 ns and the head and whole-body phantoms. A model for the scanner singles rate and NECR were shown to demonstrate how the geometrical variables and performance metrics affect the count rate performance of the scanner. It was shown that detectors with a few crystal elements have a reduced sensitivity while large detectors are highly likely to show count rate losses when deadtime presents. The importance of a short CTW was shown by studying the effects of CTW on NECR for a given activity and a model for the change in NECR and scanner sensitivity when LLD varies was presented. The effects of activity outside the FOV on the count rate

performance of scanners with various detector thickness and sizes were also shown. Analyzing the results of the count rate study in this chapter and assuming a deadtime of about 750 to 1000 ns, we recommend a detector size of about 1.5 to 9.6 cm^2 for our brain-dedicated PET insert.

Chapter 6 : Event Mispositioning due to Layer Misassignment in DLO Detectors

Reuse Permission

6.1 Introduction

As explained in Section 2.5.3.3 and Chapter 4, a DLO detector enables DOI through light sharing between two layers of scintillation arrays with a *single-ended readout* (SER) scheme. However, as briefly pointed out in Section 4.9, the SER scheme in DLO detectors may lead to a layer misassignment when inter-crystal scattering occurs. The aim of this work is to study inter-crystal scattering and evaluate the effects of layer misidentification in DLO detectors on the performance of selected brain-dedicated PET scanners discussed in the previous chapter. Similar to the methodology of Chapter 4, GATE v7.1 [142, 143] was used to simulate DLO detectors in the current chapter. The inter-crystal scatter investigation reported in the current chapter does not concern the count rate performance of the scanner. Therefore, the findings of this chapter should be considered independent of those of Chapter 5.

6.2 Inter-crystal Scatter in Pixelated Scintillators

Various studies have investigated the effects of inter-crystal scatter on the performance of PET detectors. Through MC simulations, various crystal positioning schemes were investigated in Shao *et al*. [212] to study the effects of inter-crystal scatter on the CRFs of LSO and BGO detectors at various incidence angles. It was shown that the weighted energy scheme (which is essentially what happens in a conventional block detector) leads to the worst position detection accuracy.

Also, the results of [212] suggested that inter-crystal scatter has negligible effects on the FWHM of the CRFs and it only adds a slight increase to the tail of the obtained CRFs for the studied single layer detectors. Consistent with the findings of [212], Levin *et al*. [213] also showed that reducing inter-crystal scatter in a PET detector only leads to an improvement in the FWTM of the CRFs and does not affect the FWHM of the CRFs. The effects of inter-crystal scatter on various discrete and continuous DOI detectors (i.e. phoswich, DLO and dual-ended readout) were studied in [214] by simulating the CRFs of the detectors with and without inter-crystal scatter. The authors suggested that while all DOI-enabled detectors reduce parallax error, inter-crystal scatter is not an important component of spatial resolution loss, especially at large radial offsets. Investigations in [212-214] did not include the relation of variation in CRFs and image quality. The effects of inter-crystal scatter on images were however presented by Park *et al*. [215]. In [215], various detector geometries with crystal widths between 0.5 to 2 mm, detector thicknesses between 10 to 30 mm and scintillation materials of BGO, LSO and NaI were considered. For each detector, simulations were conducted to evaluate various positioning algorithms assuming an individual crystal readout. Park *et al*. [215] demonstrated that while inter-crystal scatter does not have an effect on the FWHM of CRFs, it reduces the image contrast. Also, it was suggested that crystal elements with a larger pitch and lower thickness can minimize the degrading effects of inter-crystal scatter.

It has been shown that a one-to-one coupling of crystal elements to an array of photodetectors, such as multi-pixel photon counters, allows for an individual scintillation element readout and a highly efficient light collection for all crystals [106, 216]. Ota *et al*. [200] suggested a technique to create an interaction-order scheme based on Compton kinematics to improve the spatial resolution. The proposed technique in [200] only applies to single inter-crystal scatter events. Consistent with several other publications, the results of [200] showed that the FWTM of

the CRFs has the greatest improvement when individual crystal readout is used. The authors suggested evaluating the effects of various readout schemes through reconstructing images in a future work.

6.3 Inter-crystal Scatter in DLO Detectors

Similar to the method in Section 4.3.1, the selected scintillation material for the proposed brain-dedicated PET insert is LYSO ($Lu_{1.79}Y_{0.21}SiO_5$) with a mixing fraction of 0.105 yttrium to lutetium. For the selected LYSO detector, the ratio of Compton scattered photons to photons with other interactions (i.e. photoelectric and coherent scattering) is 0.64 [9]. The high probability of Compton scattering causes a large amount of inter-crystal scatter in the detector, especially for small crystal elements.

In a DLO detector, inter-crystal scatter can lead to both crystal and layer misidentification. In this chapter, the effects of inter-crystal scatter and layer misidentification in DLO detectors are studied within a range of detectors with various total thicknesses between 15 to 25 mm.

6.4 Discrete-layer Readout DLO Detector

One of the advantages of a DLO detector (Figure 6.1-(A)) is that DOI information can be obtained through SER as shown in Figure 6.1-(B). However, to investigate the effects of layer misidentification as a result of inter-crystal scatter in DLO detectors, we introduce a *discrete-layer readout* (DLR) DLO scheme as shown in Figure 6.1-(C).

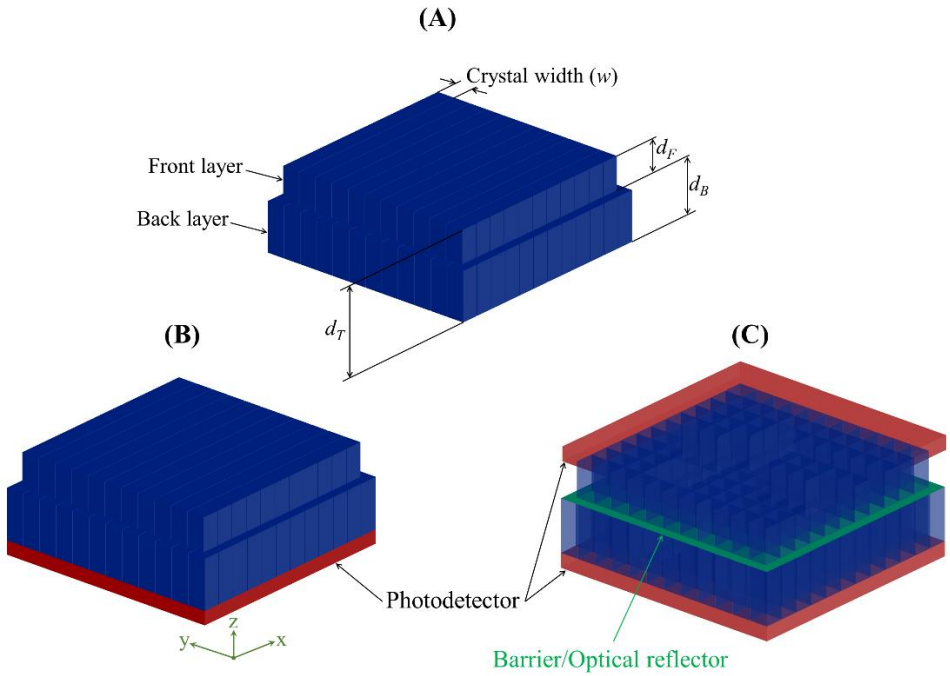

Figure 6.1. (A) Crystal arrangement of a DLO detector. The bottom images show a drawing of a SER-DLO (B) and a DLR-DLO (C) block detector (© 2019 [211]).

In a DLR-DLO detector, each layer is read out independently and, unlike a typical SER-DLO detector, light sharing between the two layers is prohibited as an optical reflector between the layers exists. Essentially, different crystal identification schemes (flood map segmentation techniques) were employed for SER- and DLR-DLO detectors. A SER-DLO detector yields a single detector flood map with an interleaved pattern for crystals in different layers. A DLR-DLO detector, however, is read out through two independent photodetectors collecting optical photons at the top and bottom of the detector. Sample flood maps of a SER- and DLR-DLO detector are shown in Figure 6.2-(A) and Figure 6.2-(B), respectively. Also, an example of how SER may assign an event to an incorrect layer in a DLO detector is illustrated in Figure 6.2-(C). Figure 6.2-(D) shows how the same event is assigned to a crystal in the layer in which all photon interactions occurred for the DLR-DLO readout case.

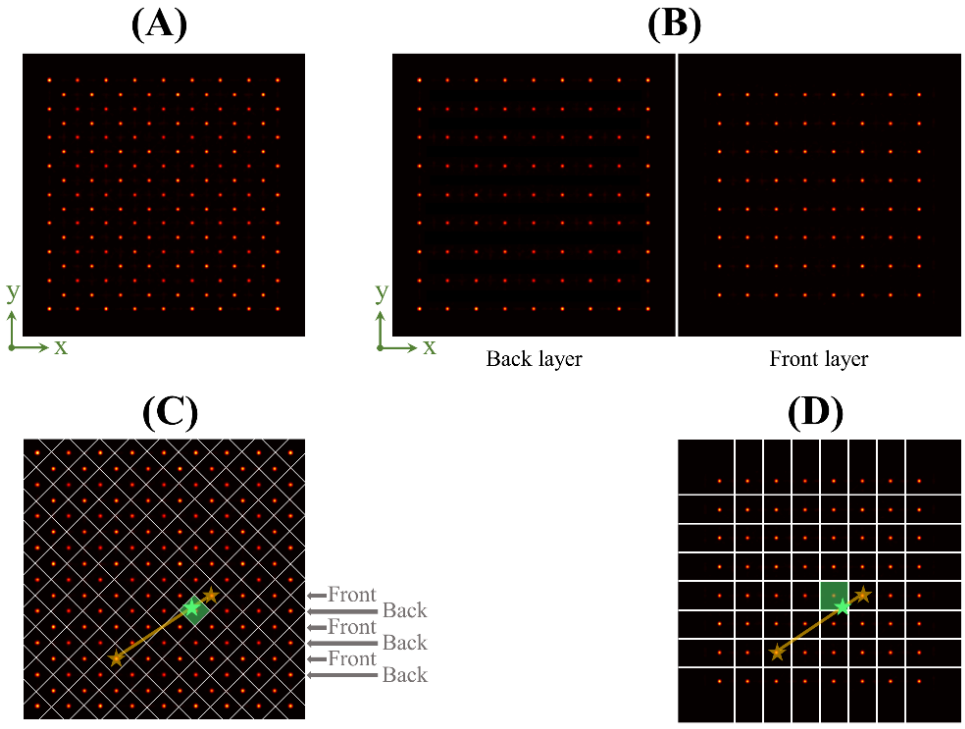

Figure 6.2. Sample flood maps of a SER-DLO (A) and DLR-DLO (B) detector. An example trajectory of a photon undergoing a Compton scatter followed by a photoelectric interaction in the front layer of a DLO detector is shown in yellow in the bottom images. The green stars represent the location of the energy weighted centroid of the incident photons and the green highlighted segments indicate the assigned crystals. As shown in (C), while the photon only deposited its energy in the front layer crystals, the energy weighted centroid lays within the boundary of a bottom layer crystal when SER is used which leads to a layer misidentification (© 2019 [211]).

6.5 Geometrical Parameters of Detectors

Similar to the description of the detectors and scanners used in Chapter 4, fixed and variable geometrical parameters considered for the detectors in this chapter are discussed in the following sections.

6.5.1 **Fixed Geometrical Parameters**

Since layer misassignments in a segmented SER-DLO detector are not related to detector area, detectors were chosen in accordance with the descriptions in Section 4.3.1. Specifically, all scanners were assumed to be constructed of 6 rings of detectors and 38 block detectors per ring such that the outermost diameter of the scanner is 346 mm and an axial coverage of 170 mm is

provided. Each detector used a 4 × 4 group of SensL ArrayC SiPMs where the size of each SiPM is 7.1 × 7.1 mm^2. Among the most important advantages of LYSO is that it has a relatively high effective atomic number (Z) of 64 and a high linear attenuation coefficient of 0.86 cm^{-1} for 511 keV photons. It is also a fast scintillation material with a decay time of 41 ns [163].

6.5.2 **Geometrical Variables**

Three different DLO detectors were considered in this work by varying the total detector thickness (d_T) between 15 and 25 mm. Each block detector was assumed to be constructed of 8 × 8 crystals in the front layer and 9 × 9 crystals in the back layer. The crystal width was assumed to have a nominal value of 3 mm such that the adjacent block detectors in each detector ring just touch. Also, the front/back layer thickness (d_F/d_B) for each total thickness d_T was chosen according to the findings of Chapter 4 summarized in Table 4.4. Table 6.1 summarizes the geometrical properties of the DLO detectors considered in this chapter.

Table 6.1. Geometrical variables of the 3 DLO detectors used in this study. The detectors were assumed to have (8×8) + (9×9) crystal arrays. For each case, crystal width *w* was adjusted such that the adjacent blocks in each ring just touch.

d_T [mm]	d_F / d_B [mm]	*w* [mm]
15	6 / 9	2.99
20	8 / 12	2.93
25	7.5 / 17.5	2.83

6.6 Singles Generation Algorithm

Similar to the approach explained in Sections 4.5 and 4.6, an in-house code was written to generate *singles* and *coincidences* from the list of all recorded *hits*. A timing resolution of 2.5 ns and an energy resolution of 12.5% at 511 keV were assumed for all scanners. As pointed out previously, while light sharing occurs between the layers of a SER-DLO detector, no optical sharing between the layers of DLR-DLO detectors was assumed. In the case of SER-DLO, from

the list of recorded hits, an energy weighted centroid of each set of hits in a block was calculated. If the total blurred deposited energy in the block was within an energy window of 350 to 650 keV, the energy weighted centroid coordinates were then mapped onto the flood map of the detector to assign it to a crystal and form a single (Figure 6.2-(C)). For DLR-DLO detectors, hits in each layer were processed independently to create *singles*. Due to the chosen energy window of 350 – 650 keV, only interactions either in the front or back layer of each detector could contribute to a valid single. Then, all energy weighted centroids were assigned to a crystal using a flood map look-up table for each scanner.

6.7 Coincidence Generation Algorithm

The process of generating coincidences from singles was identical to that described in Section 4.6. From the generated singles, a list of coincidences was created assuming a coincidence timing window of 10 ns. The process of creating coincidences from singles was the same for both the SER- and DLR-DLO detectors. No scatter or randoms corrections were employed, however, the activity level was kept sufficiently low to maintain random coincidences at a negligible rate in all simulations.

6.8 Analysis of Inter-crystal Scattering Pattern

Since the aim of this chapter was to examine the effects of inter-crystal scatter on DLO detectors, the scattering pattern in the selected DLO detectors was studied in this section.

6.8.1 **Method**

Individual DLO detector blocks with total thicknesses of 15, 20 and 25 mm were simulated and irradiated by a uniform beam of 511 keV photons. All hits (photon interactions) within each detector were traced to study the pattern and effects of inter-crystal scatter in DLO detectors with

different total thicknesses. Singles were generated according to the technique detailed in Section 6.6. The beam was defined to have an area equal to the crystal face size and was perpendicularly pointed toward the central crystal of the front and back layer of each DLO detector in two separate experiments as shown in Figure 6.3-(A) and Figure 6.3-(B) (i.e. beams offset by ½ crystal pitch in both directions for two experiments). These experiments are referred to as the *front-layer hits experiment* and *back-layer hits experiment*, respectively. The simulation continued until 10^5 singles in the layer of interest were generated. The number of singles assigned to each crystal of each layer in each experiment was then determined.

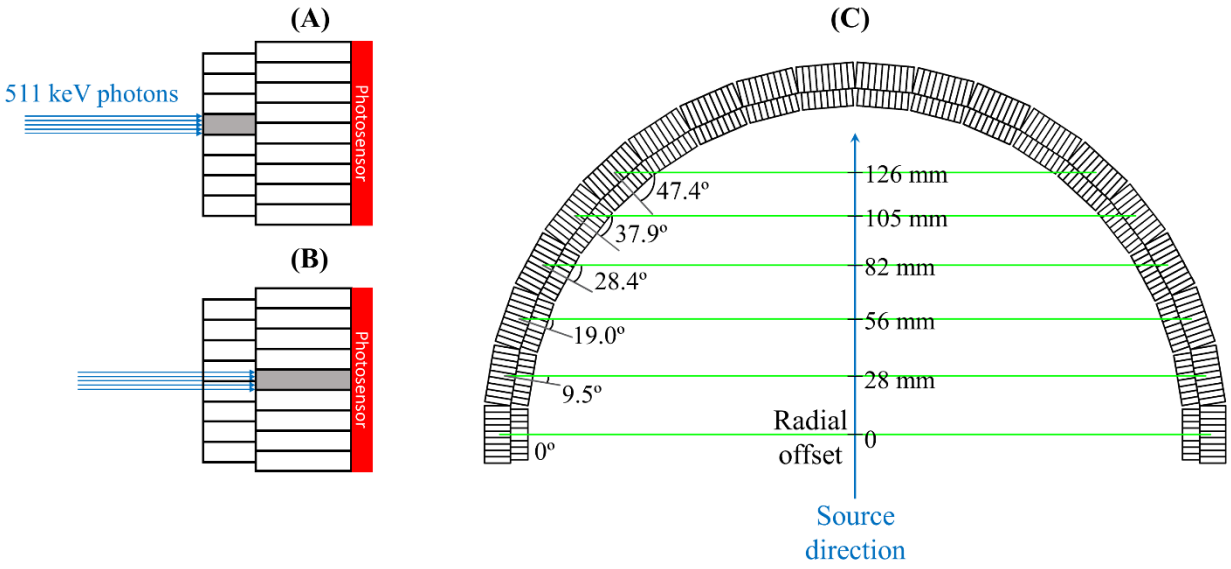

Figure 6.3. An illustration of the front-layer hits experiment (A) and back-layer hits experiment (B) with the target crystal shaded in grey. Image (C) shows the average radial offset and its corresponding incidence angle for each opposing block detector pair in the coincidence response function experiment (© 2019 [211]).

6.8.2 **Result**

A map of 10^4 hits when a central crystal element of each DLO detector (front- and back-layer hits experiments) is irradiated is shown in Figure 6.4. It should be emphasized that the presented hit maps in this figure are independent of the readout technique.

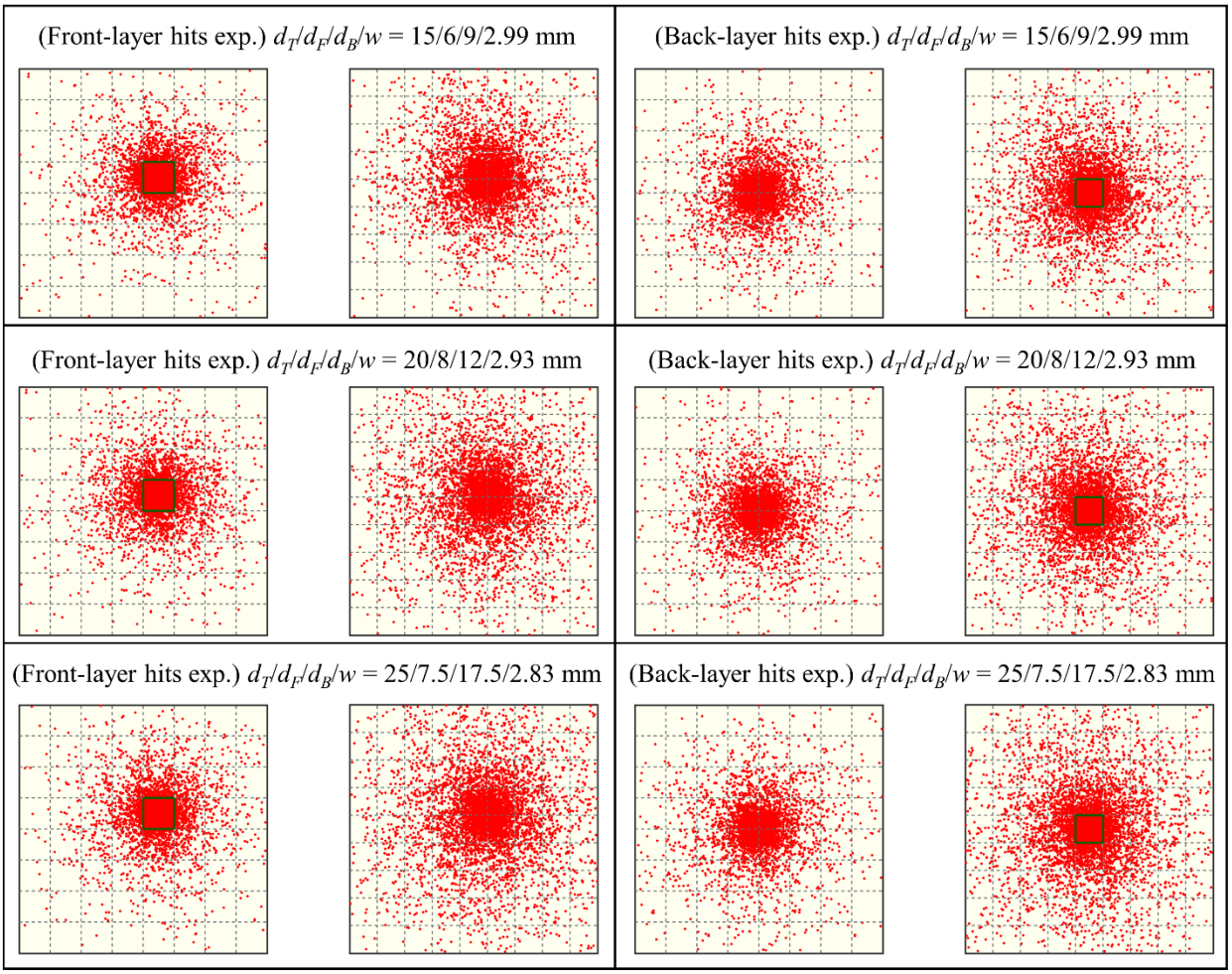

Figure 6.4. A map of 10^4 *hit* locations in each layer of DLO detectors with (8×8) + (9×9) crystal arrays with a nominal crystal width of 3 mm and various total/front/back thicknesses. Each pair of plots correspond to the hits map of the front (left) and back (right) layer crystals in one experiment in which a middle crystal element in either the front or the back layer was the beam target. The target crystal element in each simulation study is indicated by a green square (© 2019 [211]).

In the plots of Figure 6.4, the exact coordinates of all interactions are used regardless of the amount of their deposited energy. The crystal which was entirely covered by the beam width is indicated by a green square. The blurred energy weighted centroid of each set of hits, corresponding to a single, was then mapped onto a crystal look-up table to be assigned to a crystal. Figure 6.5 shows the ratio of the assigned singles to each crystal (relative to the total number of recorded singles in the detector) assuming SER-DLO readout scheme.

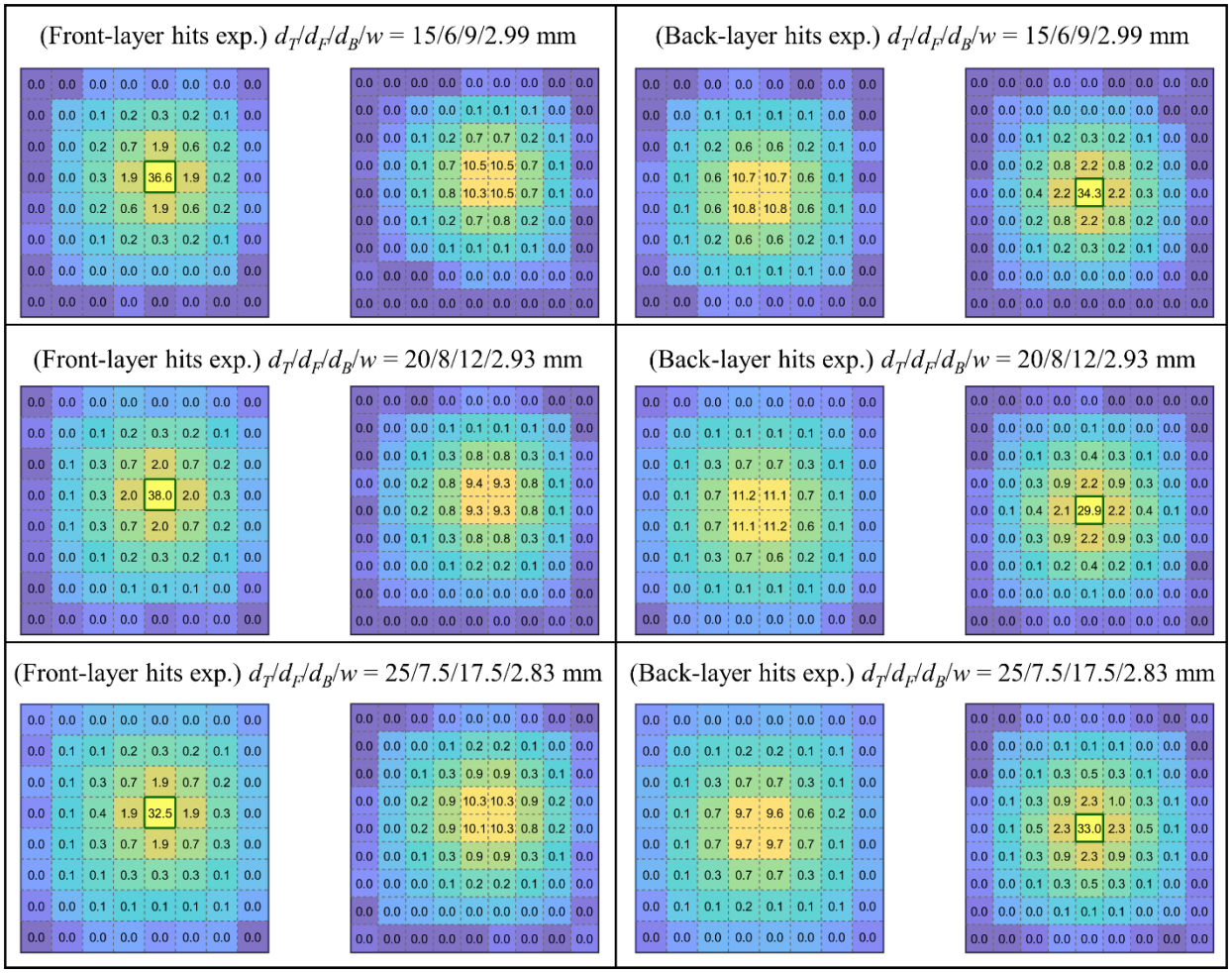

Figure 6.5. A map of the percentage of the assigned singles to each crystal element of DLO detectors with (8×8) + (9×9) crystal arrays with a nominal crystal width of 3 mm and various total/front/back thicknesses. Each pair of plots correspond to the singles map of the front (left) and back (right) layer crystals in one experiment in which a middle crystal element in either the front or the back layer was the beam target. The target crystal element in each simulation study is indicated by a green square. Colors correspond to the number of the assigned singles to crystals on a logarithmic scale (© 2019 [211]).

The number of recorded and correctly positioned singles in the front- and back-layer hits experiments are reported in Table 6.2. Note that in this table, it was assumed that a *correctly positioned* single is a single that is assigned to the same directly irradiated crystal. However, since the overlying or underlying four crystals in the other layer of the target crystal may also represent a correctly positioned event, the number of singles in these crystals were also measured.

Table 6.2. Results of front- and back-layer hits experiments for DLO detectors with a total thickness of d_T.

d_T [mm]	Target layer	Number of singles in front layer	Number of singles in back layer	Number of singles assigned to the target crystal	Number of singles assigned to crystals other than the target crystal	Number (percentage) of singles assigned to the target and its four over/underlying crystals	Number (percentage) of singles assigned to crystals other than the target and its four over/underlying crystals
15	Front†	100,000	99,936	73,220	126,716	156,749 (78.40%)	43,187 (21.60%)
15	Back††	99,818	100,000	68,541	131,277	154,288 (77.21%)	45,530 (22.79%)
20	Front	100,000	88,586	71,762	116,824	142,214 (75.41%)	46,372 (24.59%)
20	Back	112,441	100,000	63,525	148,916	158,238 (74.49%)	54,203 (25.51%)
25	Front	100,000	109,094	67,909	141,185	153,633 (73.48%)	55,461 (26.52%)
25	Back	91,691	100,000	63,308	128,383	137,510 (71.74%)	54,181 (28.26%)

† Front-layer hits experiment

†† Back-layer hits experiment

6.9 Coincidence Response Function Analysis

The first step in comparing the performance of SER- and DLR-DLO detectors was to measure the CRFs of the scanners as explained below.

6.9.1 **Method**

For this experiment, a single half-ring PET scanner, similar to the image shown in Figure 6.3-(C), was only simulated in each case. Similar to method used in Section 4.8.2.1, the bottom half of the scanner was not included as only hits data in the top half were used. A spherical source with a radius of 0.225 mm embedded in a $10 \times 10 \times 10$ mm^3 water cube was stepped along the radial axis of the scanner in increments of $w/8$ (w is the crystal width). From the list of recorded hits for each source step, a list of coincidences was created and the number of coincidences between the opposite crystals of the front and back layer were recorded. The source was assumed to emit back-to-back photons with an acollinearity with a FWHM of 0.5º. Acquisition time per source step was chosen such that the relative uncertainty for each CRF is below 0.05 (i.e. $1 / n <$

0.05^2 where *n* is the peak value of CRFs for normal coincidences). The time per source step varied for different detector geometries but remained constant for all steps of each detector geometry. Once the number of recorded coincidences for each source position was measured, each CRF was fit with the sum of 3 Gaussian functions to facilitate the measurement of the FWHM and FWTM (collectively called FWXM) of the CRFs. For each detector geometry, CRFs of the first 6 pairs of opposite detector blocks were measured. The average radial offset and incidence angle of each opposite pair of detector blocks are shown in Figure 6.3-(C). The generated CRFs were also compared with those of an equivalent single layer scanner to evaluate the benefits of DLO detectors.

6.9.2 **Result**

Coincidence response profiles of the first 3 pairs of opposing detector blocks for a DLO scanner with a total thickness of 20 mm and various detector readout techniques are shown in Figure 6.6.

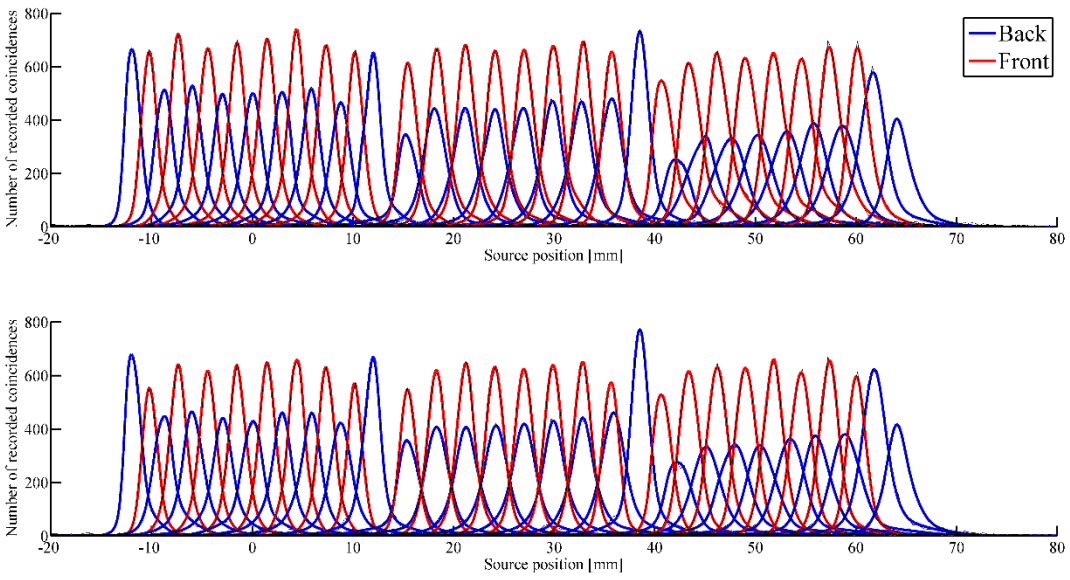

Figure 6.6. Coincidence response functions of the first 3 opposing detector block pairs for SER-DLO (top) and DLR-DLO (bottom) scanners with a total detector thickness of 20 mm (© 2019 [211]).

The average FWXM of the CRFs for each considered DLO and single layer detector block pair was calculated and plotted in Figure 6.7.

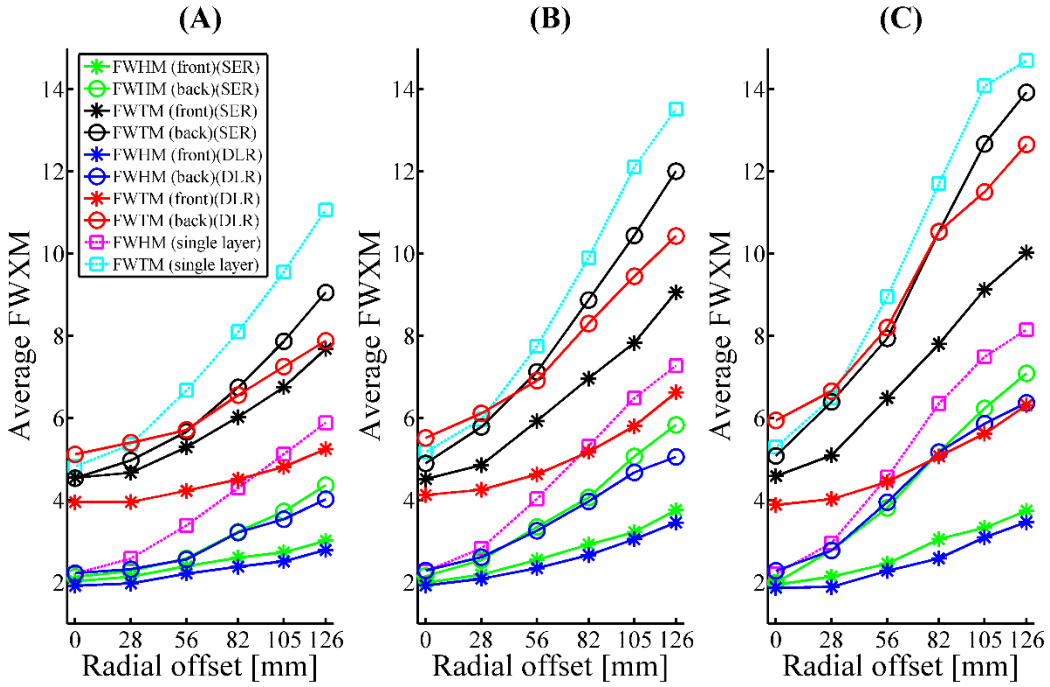

Figure 6.7. Average FWXM of the CRFs for each detector block pair with total thicknesses of 15 (A), 20 (B) and 25 (C) mm at various radial offsets. Various colors and shapes of the markers are defined in the legend (© 2019 [211]).

When comparing SER- with DLR-DLO detectors, a mild improvement in the CRF FWHM of the front layer can be seen when DLR-DLO is used (compare the green and blue stars in Figure 6.7). The FWHM of the back layer crystals show a consistent trend of superior FWHM for SER-DLO for small radial offsets, and superior FWHM for DLR-DLO at larger radial offsets (see the green and blue circles in Figure 6.7). The FWTM of the front layer CRFs is where the largest differences can be seen for DLR vs. SER readout in DLO detectors (compare the black and red stars in Figure 6.7). For all DLO detectors, the front layer FWTM improved considerably when SER was replaced with DLR. The back layer FWTM of DLO detectors showed a similar trend as the back layer FWHM, with a better/worse CRF FWTM observed for the SER-DLO detectors at low/high radial offsets (compare the black and red circles in Figure 6.7). The FWHM and FWTM

of the CRFs of the studied single layer detector without DOI are shown with purple and cyan squares, respectively. The benefit of a DLO detector in maintaining resolution uniformity is evident by comparing the FWHM of the CRFs of single layer scanners with that of equivalent DLO detectors. The FWTM of the single layer detectors, however, was close to the FWTM of the back layer of DLO detectors, especially for thicker detectors.

6.10 Phantom Imaging Study

To evaluate the effects of layer misidentification on image quality, images of a Derenzo-like phantom were also reconstructed for all DLO and their equivalent single layer PET scanners as explained below.

6.10.1 **Method**

To better understand the effects of layer misidentification on image quality, simulations of a Derenzo-like phantom with a diameter of 130 mm and a length of 37 mm was done for all studied DLO and their equivalent single layer PET scanners. The diameters of the hot rods in the phantom were 3.0, 4.0, 6.0, 8.2, 10.2 and 12.2 mm and the spacing between the centers of the neighbor rods was equal to twice the diameter of the rods. The phantom was filled with water mixed with a positron emitting radionuclide (half-life much greater than the acquisition time) with an activity concentration of 125 Bq/cm^3 in the background and 1 kBq/cm^3 in the rods (hot-to-background activity concentration ratio = 8). The phantom was positioned at a radial offset of [-3 -3] cm along the transaxial plane at the axial centre of the scanners. Acquisitions were carried out until sufficient hits data for 10^8 coincidences were collected. A list of 10^8 coincidences was then generated from the obtained data in a list-mode format. Since the purpose of this study was to compare the images with different detector readout modes (and not to measure the absolute spatial resolution), images were reconstructed using list-mode ML-EM [182] with 35 iterations. Image reconstruction was

done through a GPU-based list-mode reconstruction technique similar to [193] using an in-house reconstruction engine. The reconstruction package was developed such that it can accommodate list-mode data from any single layer or DLO brain-dedicated PET scanner as long as the geometrical assumptions explained in Section 6.5.1 are met. Details of the developed GPU-based image reconstruction algorithm are not discussed in the current chapter and can be found in Chapter 7.

The images were reconstructed in a grid of 330 × 330 × 110 voxels with an isotropic voxel size of 0.425 mm. A spatially-invariant response function in the form of a Gaussian function with a FWHM equal to *w* for each scanner was modeled in both the forward and backprojection of the developed reconstruction method. To generate a sensitivity image for each scanner accounting for phantom attenuation, a backprojection of 10^8 randomly generated coincidences was done for each scanner according to the method explained in [217]. Since the phantom was located at a radial offset of [-3 -3] cm, the reconstruction volume was also transaxially shifted accordingly to keep the phantom at the centre of the reconstruction volume. Figure 6.8-(A) shows a transaxial view of the phantom for a selected DLO scanner with the transaxial boundaries of the reconstruction volume. While images were compensated for attenuation effects, no scatter and randoms corrections were applied. However, the activity level was kept sufficiently low to minimize random events and possible effects due to detector pileup and deadtime. No post-filtering was applied on the reconstructed images.

The central slice of the images was used for quantitative evaluations. As shown in Figure 6.8-(B), a set of profiles along the outermost rods of the phantom was plotted and the average and standard deviation of the peak-to-valley ratios of each sets of profiles (corresponding to a group of rods with the same diameter) were calculated.

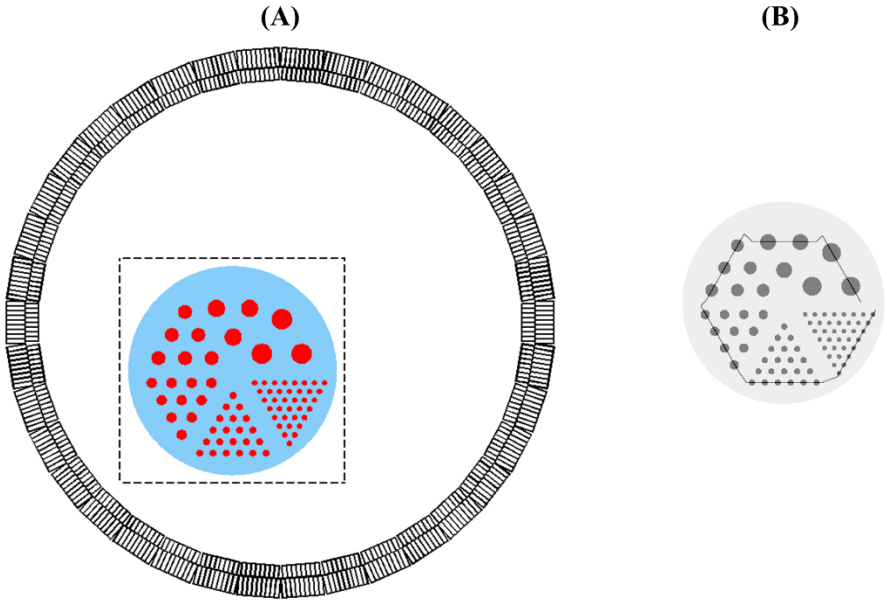

Figure 6.8. Schematic drawing of the Derenzo-like phantom used in the image evaluation study (A). The phantom background, shown in blue, was filled with radioactive water as background. The hot rods, shown in red, had a diameter of 3.0, 4.0, 6.0, 8.2, 10.2 and 12.2 mm and were filled with an activity with a hot-to-background activity concentration ratio of 8:1. Transaxial reconstruction area is shown by the dashed lines. The scanner shown here had DLO detectors with a total/front/back layer thickness of 20/8/12 mm. Image (B) shows the profile used to evaluate the quality of each image (© 2019 [211]).

6.10.2 **Result**

The central slice of the reconstructed Derenzo-like phantom used in the imaging study for the studied DLO and single layer PET scanners is shown in Figure 6.9. The drawn profiles (as defined in Figure 6.8-(B)) over these images are shown in Figure 6.10.

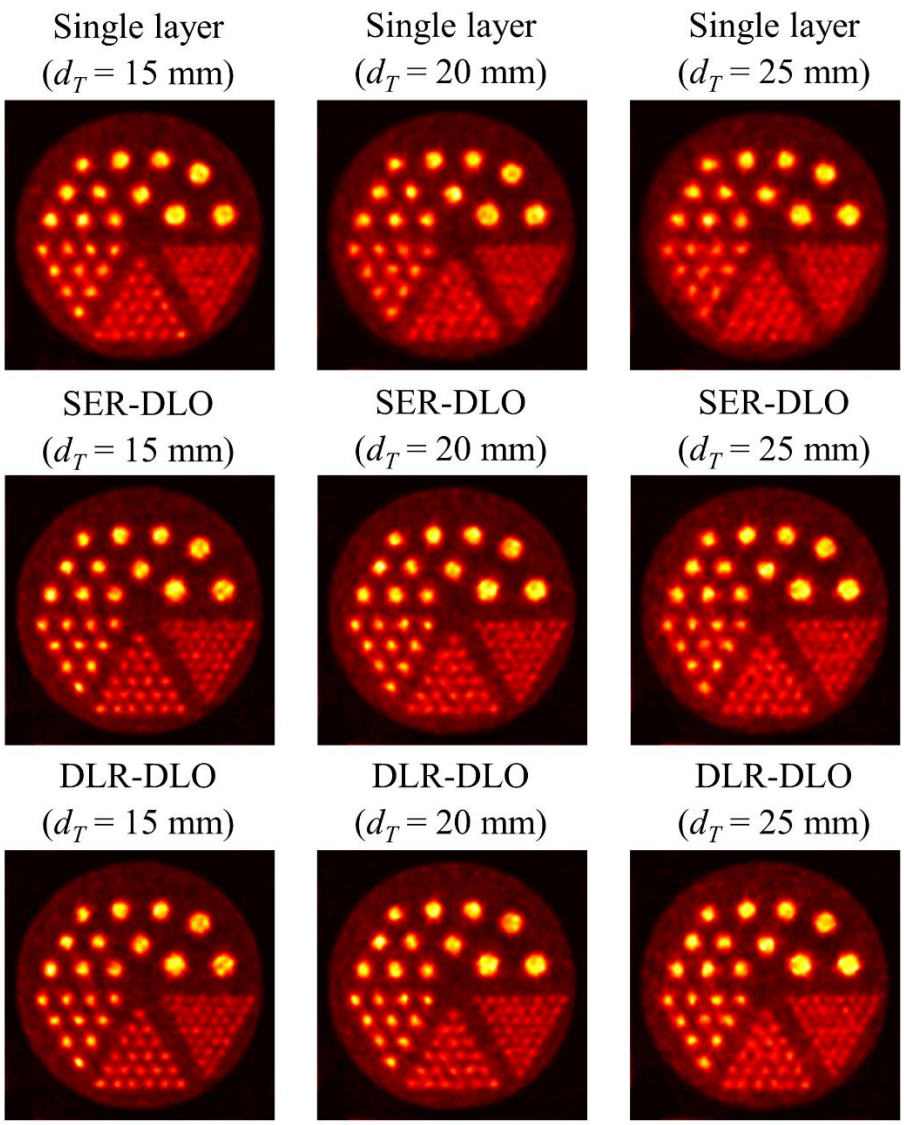

Figure 6.9. The central slice of the reconstructed images of a Derenzo-like phantom with single layer (top row), SER-DLO (middle row) and DLR-DLO (bottom row) PET scanners. From left to right, each column of images corresponds to a scanner with detectors of total thickness of 15, 20 and 25 mm, respectively. Image intensity is normalized to the maximum pixel value in each slice (© 2019 [211]).

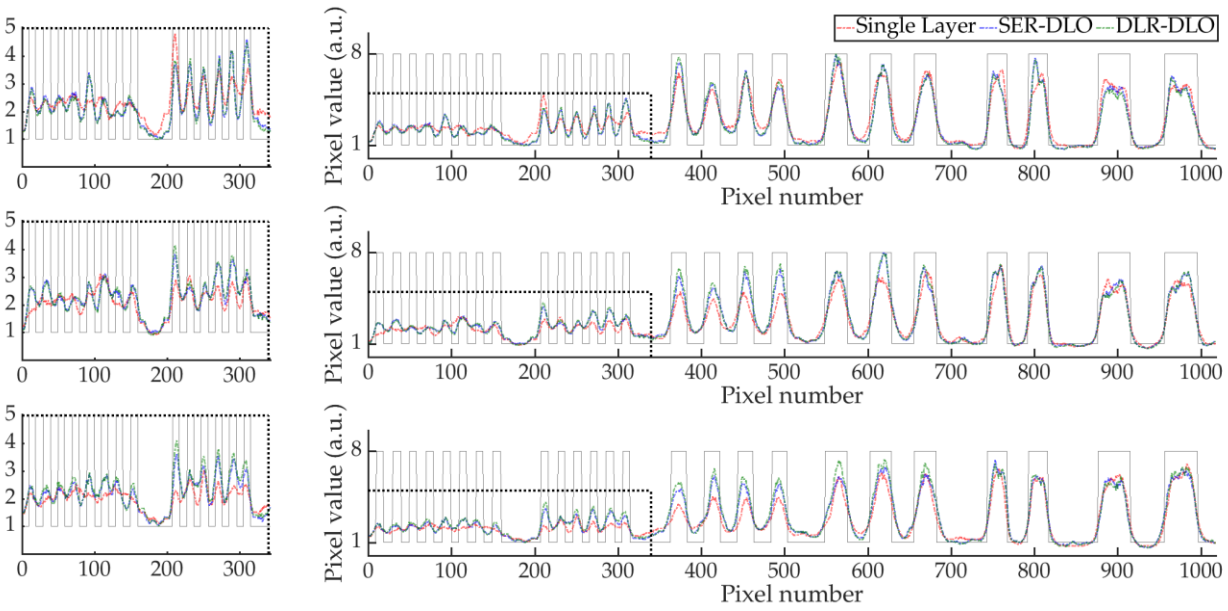

Figure 6.10. Profiles of the reconstructed images for scanners with a total detector thickness of 15 (top), 20 (middle) and 25 (bottom) mm are shown on the right. These profiles are normalized to the maximum pixel intensity in the slice and the solid black line corresponds to the reference profile. The profiles of the first two sets of rods (diameters: 3 and 4 mm) are enlarged and shown in the plots on the left (© 2019 [211]).

As mentioned before, the average peak-to-valley ratio for each group of rods with the same diameter was used to compare the images. Graphs of average peak-to-valley ratio for profiles of various diameter rods with single layer and DLO scanners with different readout types are plotted in Figure 6.11. A DLO detector led to a higher mean peak-to-valley ratio than its equivalent single-layer scanner for all rod diameters. The peak-to-valley ratios obtained from DLO scanners with different detector readout modes were similar, with the DLR-DLO having a very small advantage in peak-to-valley ratio as compared with the SER-DLO design.

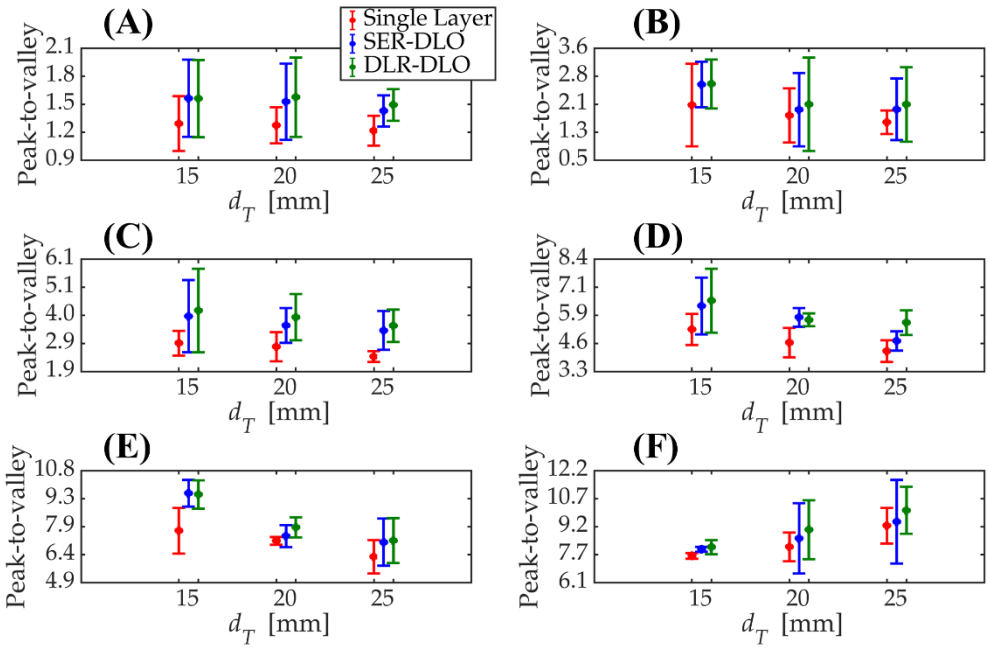

Figure 6.11. Graphs of the average peak-to-valley for phantom image profiles on rods with a diameter of 3 (A), 4 (B), 6 (C), 8.2 (D), 10.2 (E) and 12.2 (F) mm (© 2019 [211]).

6.11 Discussion and Conclusion

At 511 keV, the angular distribution of Compton scattered photons in a detector is forward-directed. Therefore, it is expected that intra-detector scattered photons have a greater effect on the back layer. Studying the map of all hits from a uniform photon beam in different detector blocks showed that the back layer of a DLO detector is more effected by scattered photons. As shown in Table 6.2, a greater influence of scattered photons in the back layer can be seen for thicker detectors.

Examining the number of singles assigned to each crystal (as presented in Figure 6.5 and Table 6.2), it can be seen that the majority of the incident photons are assigned to crystals other than the target crystal. Assigning an event with an incidence angle of 90° to a crystal other than the target crystal in the target layer or the four overlaying back or front crystals in the other layer leads to a radial mispositioning of the event. It was shown that a thicker detector leads to a lower crystal

positioning accuracy. It should be noted that since the beam was irradiating the detector at a normal angle, any mispositoning of a single along the XY plane (see Figure 6.1 for the coordinate system) is equal to the radial displacement between the first hit and the assigned single centroid. The mentioned radial displacement is of particular interest as it directly determines the amount of parallax error in a PET scanner.

The CRF experiment in this study was performed to evaluate the effects of the mispositioned singles on CRFs. At the centre of the detector ring, the CRF FWHM was within 0.41 mm for all considered detector designs. However, as expected, at greater radial offsets a thinner detector maintained the CRF FWHM more uniform as the magnitude of parallax error increases with the total detector thickness. The more pronounced increase in the FWHM of back layer crystals (in comparison with the front layer) is due to various factors. Firstly, the fact that the layer splits were chosen such that the back layer thickness is always greater than that of the front layer. Secondly, the difference between the thicknesses of the two layers increases as the total detector thickness increases, leading to a more dominant influence of the scattered photons on the back layer. Finally, the front layer acts as a scattering medium in front of the back layer arrays and the apparent width of the front layer increases as the incidence angle increases which leads to a higher degree of scattering. Although an even layer split could potentially lead to a lower CRF FWHM for the back layer crystals, the stated layer thicknesses were chosen because such layer splits minimize the parallax error at large radial offsets and yield a relatively uniform coincidence detection efficiency in both layers as discussed in Chapter 4.

Discrete layer readout was seen to reduce the degradation of the FWTM of the CRFs vs. radial offset for DLO detectors, especially for the front layer CRFs. The behavior of the front layer

CRFs is indeed similar to what is shown in other studies such as [212, 213, 215]. However, the CRF analysis suggests that the behavior of the back layer CRFs is different than front layer CRFs.

An important part of the study explained in this chapter was the imaging experiment discussed in Section 6.10. The aim of this imaging study was to translate the changes in the FWXM of the CRFs to image resolution as the effects of the changes in the CRF FWXM on image quality have not been explored for DLO detectors in the literature. The same number of coincidences and reconstruction parameters were chosen for all reconstructions to maintain a fair comparison between the images. The large hot rods in the reconstructed images shown in Figure 6.9, especially for scanners with a d_T of 15 mm, suffer from edge artifacts due to the use of a shift-invariant response function (point spread function) system model in the projection domain of our image reconstruction algorithm. It has been shown that edge artifacts appear if an incorrect system resolution model is used and are observed at the point of discontinuity which can lead to an overshoot in small objects [218]. Since the response function used in the image reconstruction algorithm of this study was a shift-invariant Gaussian function, edge artifacts were inevitable and, therefore, we chose to reconstruct the images with a point spread function that is a reasonable representation of the average response functions found in Figure 6.7. Although the appearance of artifacts suggests that the resolution of the small rods may have been over-estimated, we believe that the results of the imaging study are valid for the purpose of conducting a relative comparison between image profiles obtained from scanners with the same detector geometry.

Visual inspections of the reconstructed images shown in Figure 6.9 suggest that while DLO detectors provide images with a higher resolvability in comparison with equivalent single layer detectors, no significant differences between the images of SER- and DLR-DLO scanners were noticed. Consistently, studying the peak-to-valley ratio for different rod sizes in different images

showed that images obtained from scanners with DLO detectors provide a higher average peak-to-valley ratio in comparison with those obtained from single layer scanners. However, the peak-to-valley ratios of profiles for SER- and DLR-DLO detectors were close to each other with mild improvements for DLR-DLO detectors. Therefore, even though the FWTM of the front layer CRFs of DLR-DLO detectors showed apparent improvements relative to SER-DLO, the image quality showed negligible differences.

In conclusion, the results suggest that the FWTM of the CRFs of the front layer is most affected by layer misidentification. The imaging study revealed that the improvements in the FWTM of the CRFs for the DLR-DLO design translate into only minor gains in the peak-to-valley ratios measured with high contrast hot-rod phantoms, and thus will not necessarily result in noticeable improvements in the quality of the reconstructed brain PET images. As a result, for our brain PET application we propose to use a SER-DLO detector design.

Chapter 7 : Development of a Dual-GPU Image Reconstruction for BrainPET

In the performance evaluation process of various detectors, a scalable image reconstruction algorithm was needed to reconstruct images from the simulation of brain PET scanners with various detector geometries. In particular, an image reconstruction package compatible with different detector geometries was required to perform the imaging study explained in Section 6.10. Also, as will be mentioned in Chapter 9, a number of other reconstruction attempts were made that are not discussed in this dissertation. Due to the large number of candidate geometries, it was desired to reconstruct the images in a timely manner. To accomplish this goal, a dual-GPU list-mode image reconstruction algorithm was developed. This chapter reviews the basis of the developed image reconstruction algorithm from an implementation point-of-view. The theoretical formulation of the list-mode reconstruction algorithm can be found in Section 3.2. The forward and backprojection steps in the developed image reconstruction algorithm were inspired by [193] and [195]. Readers interested in the implementation details of the algorithm are encouraged to review [193, 195]. Although the original implementation of the algorithm benefits from *Thrust* libraries[1], the reconstruction algorithm explained in this chapter (and similarly in Chapter 8) does not use these libraries.

While the developed reconstruction algorithm for the candidate brain PET scanners accounts for geometrical sensitivity variations and attenuation effects, it does not have a scatter correction algorithm implemented. We refer to the developed dual-GPU reconstruction package

[1] Thrust is a Standard Template Library (STL) based C++ library that can be used to implement high performance CUDA applications through the use of a high-level interface.

in this chapter as '*DGPUBrainPET*'. The manual for DGPUBrainPET can be found in Appendix A.

7.1 Geometrical Description of BrainPET

BrainPET here is an umbrella term referring to a collection of all brain-dedicated scanners that follow a set of known geometrical properties. To be specific, the reconstruction algorithm described in this chapter was developed such that it is compatible with any arbitrary scanner that follows the geometrical properties defined in Section 4.3. BrainPET scanners had single layer or DLO detectors with a detector area of 28.4 × 28.4 mm^2 with 38 detector blocks per ring (n_{DPR}) and 6 rings of detectors (n_{RD}). Detector total thickness, layer thickness ratio and crystal pitch were treated as variables.

All DLO BrainPET scanners had DLO detector blocks with an array of (n × n) + (n+1) × (n+1) crystals. The number of possible LORs (n_{LOR}) in such a BrainPET is given by

$$n_{LOR} = \left(n_{CRYS} \times n_{BIC} \times \left((n+1)\times(n+1) + n \times n\right)\right) \Big/ 2 \tag{7.1}$$

in which n_{CRYS} is the total number of crystal elements in the scanner which is given by

$$n_{CRYS} = n_{RD} \times n_{DPR} \times \left((n+1)\times(n+1) + n \times n\right) \tag{7.2}$$

and n_{BIC} is the number of detector blocks that are in coincidence with any given detector block in the scanner which is calculated as

$$n_{BIC} = n_{RD} \times \left(n_{DPR} - 2 \times \left(\min_{BD} - 1\right) - 1\right) \tag{7.3}$$

where min_{BD} is the minimum block separation required for a coincidence. A value of 8 was considered for min_{BD} in all imaging studies included in this dissertation. Although BrainPET scanners had a relatively short axial extent, they can have a large number of LORs due to the DLO design. As an example, a scanner with DLO detector blocks with (6×6) + (7×7) crystals per block (i.e. $n = 6$) has 113,663,700 LORs and a scanner with detectors of (39×39) + (40×40) crystals per block has over 150 billion LORs (153,239,764,212 LORs)[1]. Due to the large number of existing LORs for some BrainPET scanners, a very sparse event density in LORs can be expected for such scanners. The varying geometry of the BrainPET scanners and the large number of available LORs for some scanners make *list-mode reconstruction* suitable for the evaluation of BrainPET scanners.

7.2 CPU and GPU Hardware

A system with two Nvidia GPUs on the same PCIe root node was used. Both GPUs were GeForce Titan Black (Nvidia, Santa Clara, CA, USA, 2014) with a *compute capability* of 3.5. An overview of the GeForce Titan Black properties and its architecture can be found in Section 3.3.1. Direct P2P access between the two devices was enabled through the CUDA P2P API. The proposed dual-GPU image reconstruction technique in this work was compared to a CPU program executing on a single thread of an Intel Xeon E2630 2.3 GHz CPU (Intel, Santa Clara, CA, USA, 2012). Also, the developed reconstruction algorithm was evaluated on a single Nvidia GeForce GTX 980 and a single Nvidia GeForce GTX 1070 Ti in separate experiments to compare the performance of these graphic cards with a single Nvidia GeForce GTX Titan Black.

The GTX 980 belongs to the *Maxwell* microarchitecture (a successor to *Kepler*) and the GTX 1070 Ti belongs to the *Pascal* microarchitecture (a successor to *Maxwell*). The GTX 980 has

[1] As a reference, the 2-meter long total-body EXPLORER scanner with more than 400,000 crystal elements has over 50 billion LORs which is almost 100 times more than the number of LORs in a current clinical scanner [219].

2048 CUDA cores with a GPU clock rate of 1126 – 1216 MHz while the GTX 1070 Ti has 2432 CUDA cores with a GPU clock rate of 1607 – 1683 MHz. A complete list of these graphic cards' properties and specifications can be found in [188]. The developed CUDA programs in this dissertation were compiled with NVCC using CUDA Toolkit 9.2 and 10 and G++ compiler using C++11, C++14 and C++17 libraries with option -O3 in a Linux environment.

7.3 Dual-GPU Memory Layout

In the current implementation of the dual-GPU list-mode ML-EM, the image is segmented into one or several 3D *cubes* such that a slice of the cube along each axis of the coordinate system (*tile*) can be fit into the shared memory. A *tile* is defined as a 2D image section with $T_d \times T_d$ pixels ($T_d \leq 110$) such that each tile requires less than 48 KB (size of the shared memory per SM). In the developed image reconstruction technique, threads within each thread block process a tile separately. This technique, commonly known as *tiling* in CUDA programming, reduces the global memory access by caching each tile into the fast shared memory [220]. The developed reconstruction algorithm for BrainPET requires the number of voxels along each dimension to be a multiple of T_d and only involves the second GPU if the image can be split into more than one cube. For example, given a T_d of 100, an image with 100 × 100 × 100 voxels is reconstructed using only one GPU and an image with 200 × 200 × 100 voxels is reconstructed using two GPUs.

Recorded coincidences are grouped according to their predominant direction (X, Y, or Z as will be explained in Section 7.4) and stored in a *structure of arrays* (SoA). Each coincidence is represented with two integers indicating the indices of the involved crystal elements in the scanner. The created SoA, including a list of all recorded coincidences, is then transferred to the global memory of both devices to decrease the occurrence of P2P interactions. For each recorded LOR, additional memory for a forward projected value and an attenuation coefficient is also required.

To reduce the size required to store the spatial information of each coincidence, a crystal LUT is defined for each BrainPET candidate on the global memory of both GPUs to host the Cartesian coordinates of the centroids of all crystal elements. The reconstructed image, attenuation map (attenuation image) and sensitivity image are transferred to the global memory of one of the devices. A 1D *texture* is defined in the texture memory of each GPU to model a *shift-invariant point spread function* (PSF) as a Gaussian function with a FWHM equal to the crystal pitch of the scanner with 2048 elements.

Each thread in a thread block processes the contribution of each pixel in a limited segment of the tile as will be discussed in Sections 7.5 and 7.6. Figure 7.1 illustrates the memory layout and the fundamental basis of the proposed dual-GPU reconstruction algorithm.

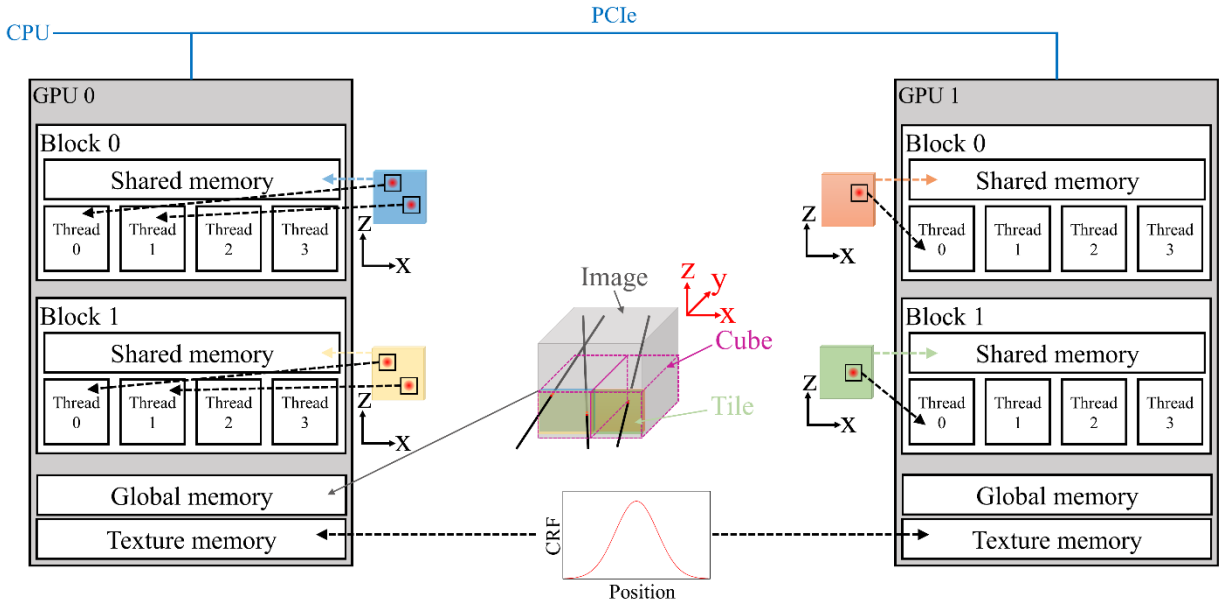

Figure 7.1. Memory layout and caching mechanism in the developed dual-GPU reconstruction algorithm. A predefined PSF is defined as 1D texture in the texture memory of both devices while the image is loaded into the global memory of GPU 0. From each cube of the image, a tile is extracted and loaded into the shared memory of each thread block. Each thread finds the contribution of all pixels around the line-tile intersection as shown by the black square in each tile.

7.4 Line Partitioning

As mentioned earlier, LORs are grouped according to their predominant direction. For instance, for an LOR with unit direction $\mathbf{u}=\left[d_X, d_Y, d_Z\right]$, the predominant direction is *X* if $\left|d_X\right| \geq \left|d_Y\right|$ and $\left|d_X\right| \geq \left|d_Z\right|$ or *Y* if $\left|d_Y\right| \geq \left|d_X\right|$ and $\left|d_Y\right| \geq \left|d_Z\right|$. Partitioning LORs according to their predominant direction allows for slice-by-slice processing of the coincidences and creating a *tube-of-response* (TOR) for each coincidence with a limited TOR-tile intersection width. A window around each line-tile intersection is shown in Figure 7.1 to illustrate the limitation of the TOR-tile intersection width. Cui *et al*. [193] suggested that the described line partitioning maintains the width of the TOR-tile intersection less than a threshold for all LORs. Therefore, a fixed TOR-tile intersection window width can be used for all LORs (i.e. all threads) which consequently prevents thread divergence. As found by [193], the maximum TOR-tile intersection width ($W_{\max}$) for TORs with a predefined width of T_W can be

$$W_{\max} = \sqrt{2}T_W + S \tag{7.4}$$

where *S* is the slice thickness along the predominant line direction (e.g. if the predominant direction of an LOR is *X*, *S* is the image pixel size along the *X*-axis).

7.5 Forward Projection

A line-driven forward projection operator has been used in the developed image reconstruction algorithm. To find the contribution of each tile to each TOR, the position of the LOR-tile intersection is found in the first step. Then, a 2D patch with a size of $W_{\max} \times W_{\max}$ is considered around the LOR-tile intersection (see the black square around the LOR-tile intersections in Figure 7.1). For each pixel in this window with a distance *r* from the LOR-tile

intersection, a weight $f(r)$ is fetched from the texture memory and used to accumulate the contribution of all pixels. Note that this method does not account for spatial changes in the response function across the FOV (i.e. a shift-invariant PSF is assumed). The projection is a cumulative step in which the total contribution of all image pixel values weighted by the PSF model in all slices is calculated for each coincidence. The described process is repeated such that the contribution of all tiles in all cubes in each GPU to each coincidence is cumulated. If the image is split into more than one cube, each GPU holds a partial contribution of the image and therefore a merging step is required prior to the beginning of the backprojection step. Analysis of the developed reconstruction algorithm showed that merging forward projection values from the two GPUs through P2P access is not a significant bottleneck of the program. Although, as discussed in Section 3.3.5, the required inter-node communications will be a bottleneck when more than two GPUs are used. Once the forward projection values are merged, the values are written back to the global memory of both devices to be used in the backprojection step. The complete forward projection algorithm is summarized in Figure 7.2.

7.6 Backprojection

Line-driven and voxel-driven backprojection GPU implementations have been proposed and evaluated in [193] and [221], respectively. The chosen backprojection in this work was line-driven such that for each cube, a similar scheme to the forward projection is employed to write to a backprojection matrix. Because of the large spatial sparsity existing between the interactions of TORs and image voxels in list-mode data, more than one line may attempt to write to the same pixel in the tile which leads to a *race condition*. *Atomic functions* (see CUDA Toolkit Documentation [188] for technical details) should then be used in the backprojection kernel to ensure quantitative correctness of the reconstructed image. Hardware atomic operations for single-

precision floating points are available in the Kepler architecture and are used in the backprojection kernel as shown in Figure 7.3.

Kernel 1: Forward projection kernel

```
for each tile do
  Load the tile into the shared memory
  for each line assigned to a thread do
    Find a patch around the line-tile intersection
    Accumulate tile contribution to the line
    Write projection values to global memory
  end for
end for
```

Kernel 2: Inter-GPU P2P communication

```
for each line assigned to a thread do
  Add projection values from devices
  Write the projection values back to the global memory of both GPUs
end for
```

Figure 7.2. Flowchart of the forward projection kernel for each cube to be executed on each thread (Kernel 1). If the image has more than 1 cube, the forward projection values from the two devices will be merged through a P2P inter-node communication (Kernel 2).

Kernel 3: Backprojection kernel

```
for each tile do
  Load the tile into the shared memory
  for each line assigned to a thread do
    Find a patch around the line-tile intersection
    Atomically write the contribution of the line to the tile in the shared memory
  end for
end for
```

Figure 7.3. Flowchart of the backprojection kernel for each cube to be executed on each thread.

7.7 Incorporating Priors

The backprojection matrix created from the accumulation of all backprojected TORs is then used in a multiplicative update step to complete the ML-EM algorithm for one iteration (or subset). The multiplicative update step can be done very efficiently on a GPU such that each thread

calculates the updated image value for a certain number of voxels and write them back to the global memory. Unlike the forward and backprojection steps, no memory caching is required in this step.

In the current implementation of the dual-GPU reconstruction algorithm, cubes are updated separately and then stitched together to form the whole image. While this method does not cause any inhomogeneity in the cube boundaries in an ML-EM algorithm, it cannot be used if a prior is introduced in the multiplicative image update step. The unavailability of a complete backprojection matrix causes the derivate of the potential function for cliques to be incomplete at the boundaries of the cubes which in turn introduces artifacts in the updated image. The quadratic prior as defined in equation (3.28) is implemented in the image reconstruction algorithm for images with 1 cube (i.e. single GPU reconstruction). The regularization parameter β, as defined in equation (3.27), defines the influence of the regularization term and therefore the degree of smoothness.

7.8 Further Technical Details

Several further optimization techniques have been proposed for GPU-based image reconstruction algorithms in the literature. Often the target of such techniques is the timing profile of the reconstruction algorithm and not the quality of the reconstructed images (as the image quality is expected to remain intact). Therefore, such implementation and optimization details are not discussed in this dissertation. Nonetheless, a technical discussion of possible optimization techniques can be found in references such as [193] and [192]. Briefly, in the current implementation of the dual-GPU list-mode reconstruction algorithm for BrainPET, *fast math* libraries (`-use_fast_math`), loop unrolling (`#pragma unroll`) and CUDA *streams* were investigated and implemented where performance improvements were seen.

The outermost loop in the forward and backprojection kernels (Kernel 1 in Figure 7.2 and Kernel 3 in Figure 7.3) is over the tiles along the predominant direction ('*slices*' along the predominant direction). To avoid thread divergence, all concurrent threads should loop over the same number of tiles. However, as lines are randomly arranged in list-mode data, the number of *involved tiles* (tiles intersecting the line) may vary largely for various lines as shown in Figure 7.4.

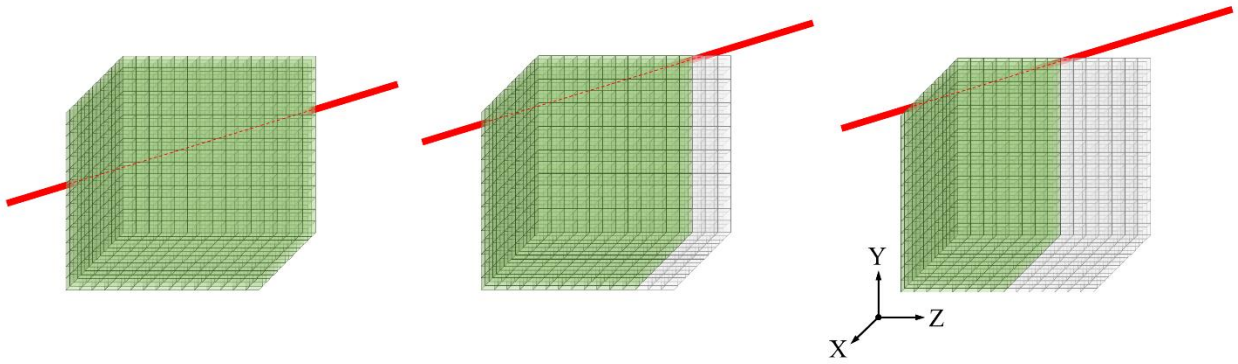

Figure 7.4. Sample LORs with a predominant direction of *Z* that intersect with a varying number of slices in each example. The *involved slices* are shown in green while the slices with no intersection are shown in grey.

In the current implementation of the forward and backprojection kernels the outermost loop runs for the maximum possible number of tiles as defined by the size of each cube. A predefined maximum loop limit avoids thread divergence at the cost of severe calculation redundancy which is an imperfection of the current implementation.

7.9 Performance analysis

7.9.1 Execution timing profiles

The patch size, W_{max}, is determined according to the physical dimensions of the image voxels, crystal pitch of the detectors and the width of the Gaussian function defined as the shift-invariant PSF. As each thread loops over the pixels in a window with a width W_{max}, the execution time of the developed image reconstruction algorithm varies with W_{max}. Processing time for 1

million randomly generated coincidences for a complete iteration (forward projection, backprojection and multiplicative image update) was measured using single- and dual-GPU implementations of the developed algorithm and are shown in Table 7.1 for various patch sizes.

Table 7.1. Processing time of 1 million randomly generated coincidences using single- and dual-GPU implementations.

Image voxel size	Number of threads per block	Number of blocks	Patch pixel size (W_{max})	Single GPU time (ms)	Dual GPU time (ms)
40 × 40 × 80	1024	40	1	125	78
40 × 40 × 80	1024	40	3	166	88
40 × 40 × 80	1024	40	5	245	127
60 × 60 × 120	1024	60	1	302	170
60 × 60 × 120	1024	60	3	305	171
60 × 60 × 120	1024	60	5	392	200
80 × 80 × 160	1024	80	1	676	341
80 × 80 × 160	1024	80	3	687	343
80 × 80 × 160	1024	80	5	747	397
100 × 100 × 200	1024	100	1	990	520
100 × 100 × 200	1024	100	3	982	535
100 × 100 × 200	1024	100	5	1050	556
200 × 200 × 200	1024	200	1	3880	2100
200 × 200 × 200	1024	200	3	3974	2118
200 × 200 × 200	1024	200	5	4058	2218

7.9.2 **Thread Block and Grid Size**

The number of threads per block (block size) and blocks per grid (grid size) in a CUDA program can influence the efficiency of the program. To provide a quantitative model for the performance of the developed GPU-based image reconstruction algorithm (single-GPU only) as a function of block and grid sizes, about 10^8 randomly generated lines were processed for one complete iteration (including forward projection, backprojection and multiplicative image update) and the processing time was measured for each instance. An image with 100 × 100 × 100 voxels with an isotropic voxel size of 1.5 mm was used such that a W_{max} of 5 pixels were considered for all events. In this experiment, a single GeForce GTX 980 and a single GeForce GTX 1070 Ti were

also used to compare the performance of a GeForce GTX Titan Black with these two graphic cards. Figure 7.5 shows the performance of all 3 graphic cards when block sizes of 256, 512 and 1024 and grid sizes of 1 to 100 were used.

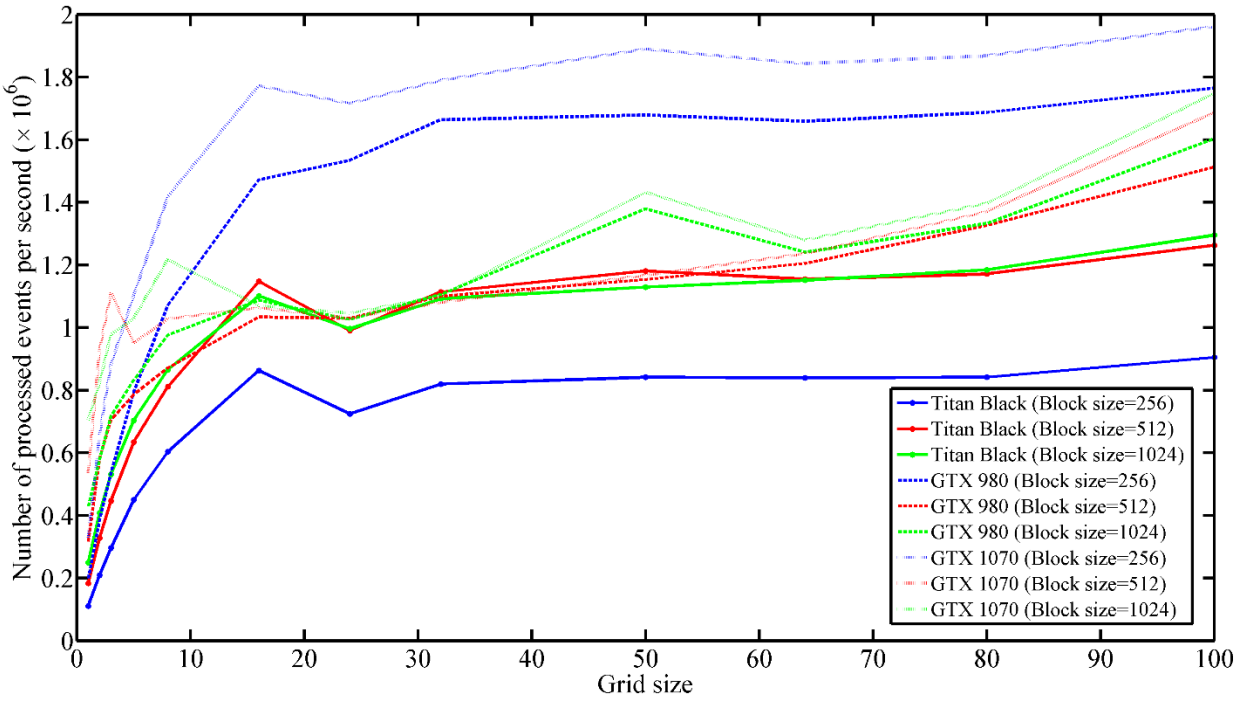

Figure 7.5. A graph of the number of processed lines per second vs grid size for various devices with different CUDA execution parameters.

A shown by Figure 7.5, a low number for the grid size deteriorates the performance of the reconstruction algorithm. The highest performance for each device was achievable when a grid size of 100 is used (i.e. even distribution of one slice per block). Interestingly enough, a block size of 256 (i.e. 256 threads per block) outperformed larger block sizes for the GTX 980 and the GTX 1070 Ti while a block size of 1024 led to the fastest performance for the GTX Titan Black. Due to the unavailability of these devices in our laboratory (other than GTX Titan Black), complementary experiments and performance analyses were not feasible. Nonetheless, the performance profiles shown in Figure 7.5 provides a valuable guideline on how to choose the CUDA execution parameters for the developed GPU-based image reconstruction algorithm. It should again be

emphasized that changing the CUDA execution parameters (block and grid size) is only a matter of performance and does not affect the quality of the reconstructed images.

7.9.3 **Point Spread Function Modeling**

The developed image reconstruction algorithm described in this chapter was employed to reconstruct the images shown in Figure 6.9 for various single layer and DLO brain-dedicated scanners. As discussed in Section 6.11, the images shown in Figure 6.9 suffered from edge artifacts which were due to the use of a shift-invariant PSF model in the reconstruction process. Images shown in Figure 6.9 were obtained using 35 iterations and 1 subset and the edge artifacts were seen to become more visible as the number of iterations increased. For instance, Figure 7.6 shows the same set of images as in Figure 6.9 but reconstructed with 50 iterations and 1 subset. The shown images in Figure 7.6 correspond to BrainPET single layer and SRE- and DLR-DLO scanners with detector thicknesses of 15, 20 and 25 mm and a nominal crystal pitch of 3 mm. It can be seen that using a shift-invariant PSF in the projection domain of the image reconstruction algorithm introduces edge artifacts to the images because of the inevitable mismatch between the actual response function and a shift-invariant PSF modeled in the developed reconstruction algorithm.

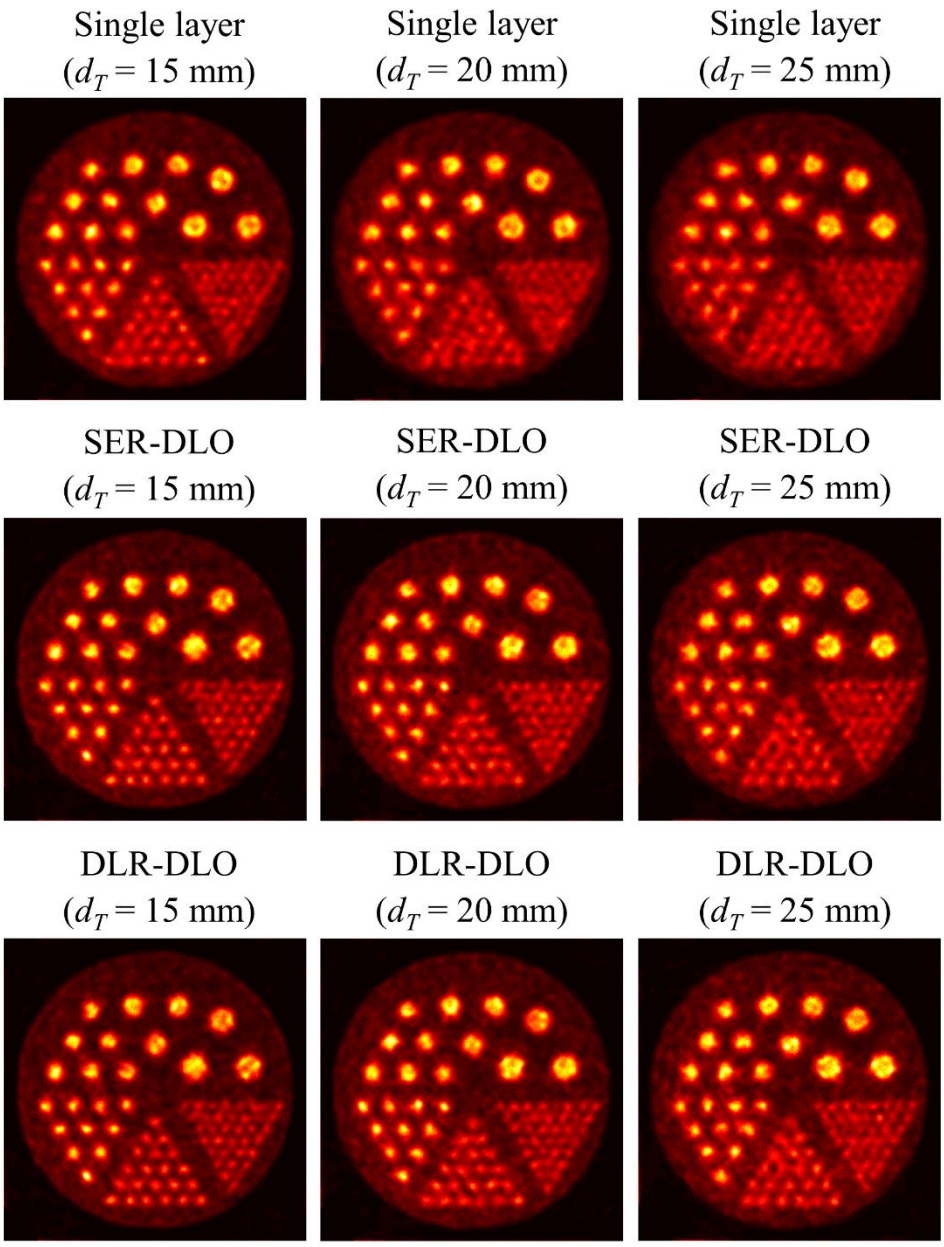

Figure 7.6. The central slice of the reconstructed images of a Derenzo-like phantom with single layer (top row), SER-DLO (middle row) and DLR-DLO (bottom row) brain-dedicated PET scanners. From left to right, each column of images corresponds to a scanner with detectors of total thickness of 15, 20 and 25 mm, respectively. Image intensity is normalized to the maximum pixel value in each slice. The shown images are equivalent to images shown in Figure 6.9 but reconstructed with 50 iterations.

The system response models obtained from the CRF analysis (discussed in Section 4.8.2) confirms that the true PSF is largely a function of radial displacement. As various studies have suggested that an accurate [218, 222] or an under-estimated response function [223] should be

used in the PSF model, under-estimated, more realistic and over-estimated shift-invariant PSFs were evaluated to study the impact of edge artifacts on the images. The data obtained from simulating the same Derenzo-like phantom used in Figure 7.6 for single layer and SER-DLO scanners with a total detector thickness of 20 mm using 100 million events were used to further study the effects of shift-invariant PSF modeling on the image quality. Unlike the images reconstructed in Figure 7.6 and Section 6.10, the images in this study were reconstructed with an isotropic voxel size of 1.25 mm in a grid of 110 × 110 × 110 voxels. The reconstructed images are shown in Figure 7.7. In Figure 7.7, '*R*' represents the ratio of the PSF FWHM to the crystal pitch (*R* = 1 has been used in this dissertation) and a non-PSF list-mode reconstruction ($W_{\max} = 1$) was also performed for comparison purposes. The images shown in Figure 7.7 were reconstructed with 30 iterations and 3 subsets to enhance edge artifacts.

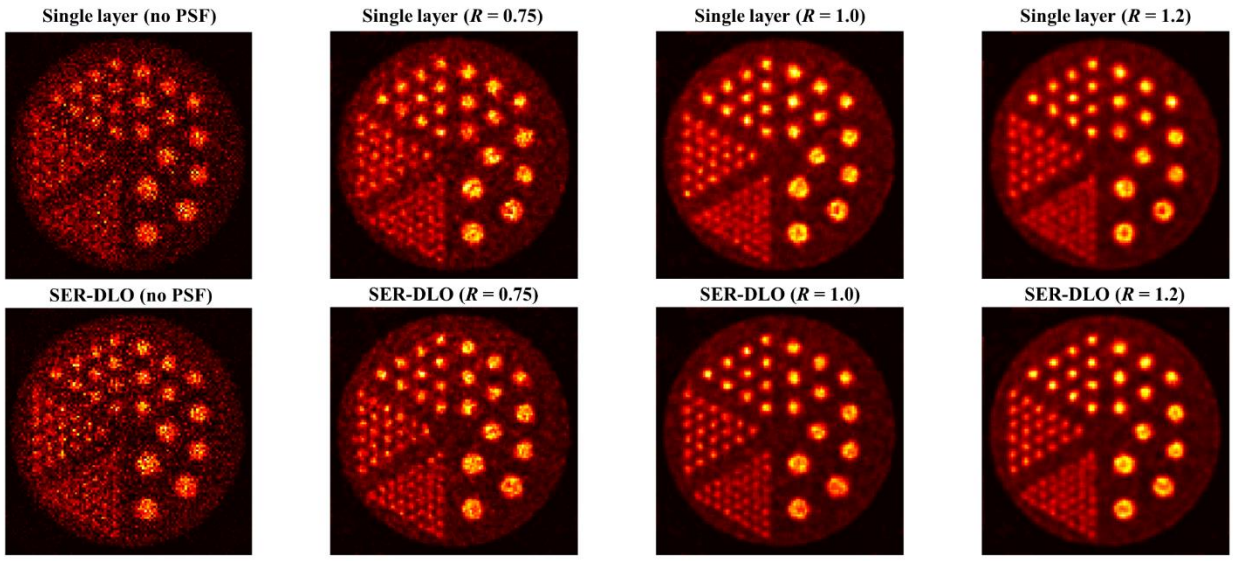

Figure 7.7. Central slice of the reconstructed images of a Derenzo-like phantom with different PSF kernels for single layer (top) and SER-DLO (bottom) scanners with a total detector thickness of 20 mm. The PSF was modeled in the projection domain with a FWHM of '*R*' times crystal pitch. Image intensity is normalized to the maximum pixel value in each slice.

As shown by the images shown in Figure 7.7, a non-PSF reconstruction has led to images suffering from a high noise level as the voxel size was much smaller than the crystal pitch of the

scanner. In all other reconstructed images with PSF modeling, different degrees of edge artifacts can be seen. While remedies as simple as post reconstruction filtering with a low-pass filter or modifying the PSF kernel have been suggested in the literature, such techniques may not be appropriate for the quantitative comparison studies done in this dissertation due to their comparative nature.

7.10 CPU and GPU Reconstructed Images

Due to the possibility of accumulating single precision floating point errors when performing arithmetic calculations and possible race conditions in a CUDA program, it is common to validate the reconstructed images on a GPU against similarly reconstructed images on a CPU. MC simulated data of a hot rod phantom (phantom diameter/length: 10/3.7 cm) with rod diameters between 0.9 to 4.7 mm were reconstructed using the developed single-GPU and CPU image reconstruction algorithms to ensure consistency between the reconstructed images. The simulated scanner was an arbitrary BrainPET scanner in which each detector block comprised 8 × 8 crystals and each crystal element had a size of $3 \times 3 \times 20$ mm^3. The image reconstructions were done using 4.0×10^6 coincidences with 15 iterations and 1 subset. No attenuation or scatter correction was applied. The images were reconstructed in a grid of 100 × 100 × 100 voxels with an isotropic voxel size of 1.5 mm. Figure 7.8 shows the reconstructed images of the rod phantom using the developed GPU and its equivalent CPU algorithms as well as a vertical profile used to compare the image profiles. The reconstructed images shown in Figure 7.8 reflected a high degree of consistency between the results of the GPU and CPU implementations.

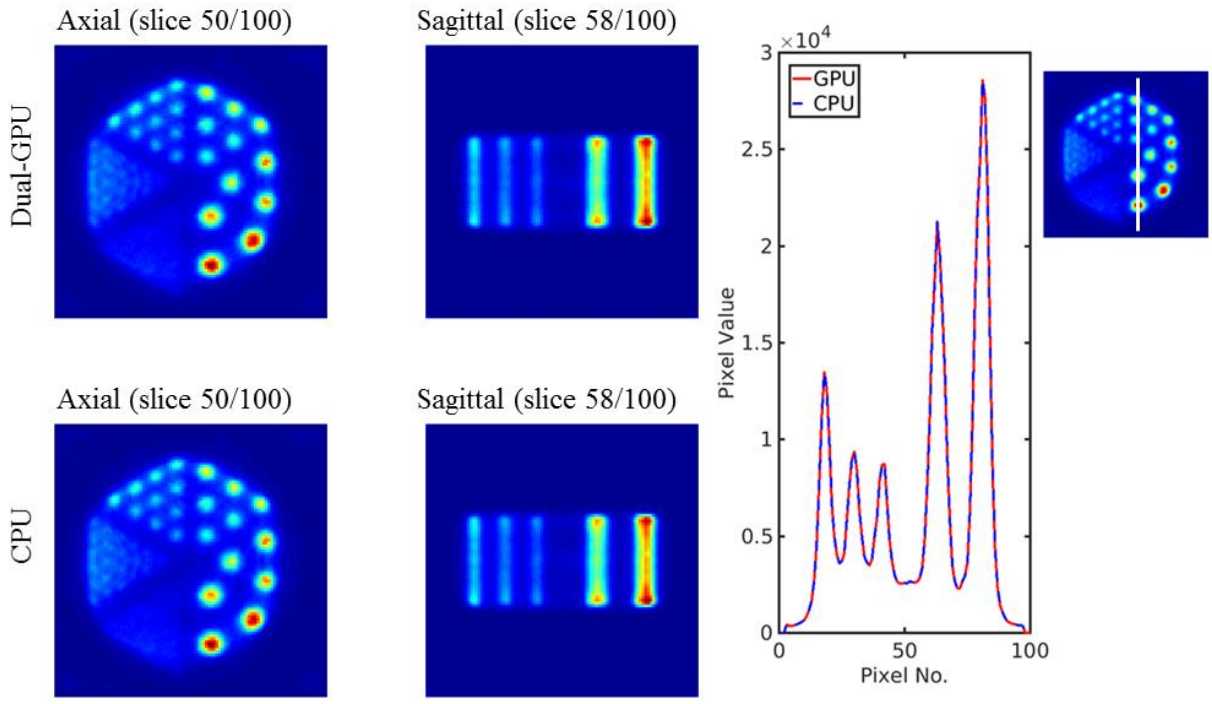

Figure 7.8. Sample reconstructed images and profiles of the hot rod phantom used in the reconstruction validation experiment. The image on the left shows the profile location. Image intensity is normalized to the maximum pixel value in each slice.

7.11 Summary

A framework for a scalable dual-GPU-based list-mode ML-EM was explained in this chapter that follows the standard ML-EM formulation with fully consistent line-driven forward and backprojection operators. The proposed reconstruction algorithm was developed such that it remains compatible with all possible single layer and DLO BrainPET scanners to facilitate a fast and reliable evaluation of each candidate scanner. While different CPU-based scalable reconstruction programs (e.g. STIR and CASToR) have been developed in the past and are widely used by different researchers, no equivalent GPU-based reconstruction algorithm is available (to the best of our knowledge). The developed dual-GPU list-mode reconstruction algorithm benefited from a novel memory layout that is suitable for fast high resolution reconstructions. Incorporation of priors in a dual-GPU-based image reconstruction algorithm is also another novel part of the

developed image reconstruction algorithm. We employed two GPUs to speed up the reconstruction even further and develop a suitable reconstruction engine for our geometry optimization project. It was seen that the proposed dual-GPU approach decreased the reconstruction time by 38% to 50% of the time required using a single GPU. The gain obtained from the dual-GPU approach in comparison with the CPU approach considerably varied when changing W_{max}. For an image with 2.0×10^6 voxels, the dual-GPU approach was measured to be 46/ 275/ 686 times faster than an equivalent CPU approach when W_{max} was set to 1/ 3/ 5. A high degree of consistency between the results of the developed GPU and CPU implementations was observed as shown by Figure 7.8.

In conclusion, the analysis of the developed algorithm showed that the developed GPU approach could improve the reconstruction time by up to a factor of two without reformulating list-mode ML-EM when running on two GPUs (as compared to a single GPU execution). We did not include any ToF information, however, investigations in [193] have shown that incorporating ToF can increase the efficiency of the reconstruction algorithm by up to 55%. Similar to the results obtained by [195], it is expected that the timing performance does not scale up when using more than a few GPUs due to severe inter-node communication.

Chapter 8 : Development of a Dual-GPU Image Reconstruction for NuPET™

Throughout the development of a dual-GPU reconstruction algorithm for BrainPET scanners, we learned that our collaborators at the Lawson Imaging Centre (London, Ontario) require a fast reconstruction software for an ongoing clinical project. The target PET scanner in the project was a small-animal PET insert called *NuPET*™. The NuPET scanner is a commercially available small-animal MR-compatible PET insert that is manufactured by *Cubresa*, Winnipeg, Manitoba. A similar GPU-based list-mode ML-EM reconstruction algorithm as explained in Chapter 7 was developed for NuPET that is explained in detail in this chapter.

The developed reconstruction algorithm for NuPET is fundamentally similar to the previously explained algorithm for BrainPET scanners. However, it has been modified to benefit from both GPUs at all times. Since the geometry and data format of this scanner are different than that of the BrainPET scanners, the developed reconstruction algorithm was modified to account for the NuPET list-mode data format. It should be emphasized that the considered modifications in the image reconstruction algorithm in the current chapter were *not* aimed to improve the efficiency or the accuracy of the reconstruction technique. Instead, these modifications were applied to better accommodate the data format of NuPET and be compatible with the standard reconstruction settings as defined by Cubresa. The following sections include an overview of the clinical project (Section 8.1), the geometry (Section 8.2) and the data format (Section 8.3) of NuPET and explain the details of the developed normalization (Section 8.4) and image reconstruction algorithms (Section 8.6) for NuPET. Performance analysis and sample

reconstructed images obtained from the simulation of NuPET and acquired data for various phantoms are presented as well in Sections 8.7 and 8.8.

We refer to the developed dual-GPU reconstruction package in this chapter as '*DGPUNuPET*'. The manual for DGPUNuPET can be found in Appendix B. Also, the manual for the developed component-based normalization algorithm for NuPET (referred to as '*NormNuPET*') can be found in Appendix C.

8.1 Intraoperative Prostate Imaging

Radical prostatectomy is a surgical process which involves the removal of the prostate gland and the seminal vesicles for patients with prostate complications such as prostate cancer. The most common technique of radical prostatectomy is *open surgery radical prostatectomy* (ORP) [224]. In a typical ORP, once the prostate is removed, it is evaluated by pathologists to determine the *Gleason grade*, tumour extent, and surgical margins. The presence of tumor cells within the surgical margins increases the risk of prostate cancer recurrence by 2 to3-fold [225]. Imaging the *prostate specific membrane antigen* (PSMA) with PET is a high sensitivity approach while high resolution MRI has shown a high capability in assessing tumor margins. The high specificity of PSMA based radiotracers has been a remarkable development in imaging prostate cancer in nuclear medicine in recent years. PSMA is expressed in the vast majority of prostate cancer tissue specimens and its degree of expression correlates with a number of important metrics of prostate cancer tumor aggressiveness [226, 227]. Often PSMA is labeled with gallium-68 (^{68}Ga) for PET imaging. Our collaborators at the Lawson Imaging Centre have recently initiated a project to employ high resolution PSMA-PET/MRI imaging to improve open surgery radial prostectomy. It has been proposed to image the patient with PSMA-PET/MRI pre-surgery and image the resected prostate after ORP to evaluate the complete removal of the cancerous cells. An ex-vivo

imaging of the prostate will be done immediately after resection using the available small-animal NuPET scanner on-site and the surgeon will receive a PSMA-PET margin report to help guide additional tissue resection if necessary. Due to the intraoperative imaging protocol, a fast (ideally less than 5 minutes) high resolution image reconstruction is demanded. An optical imaging approach with a similar goal has recently been suggested in [228] in which Cerenkov luminescence imaging is used to evaluate the tumor margins of resected breast tumors.

8.2 Description of NuPET

The NuPET scanner is a small-animal PET insert that has geometrical properties very similar to the small-animal PET insert explained in Section 4.7. A major difference is that NuPET has two rings of detectors whereas the scanner explained in Section 4.7 had only one ring of detectors. More specifically, NuPET has an inner bore diameter of 79.5 mm and an effective axial length (the axial length of the FOV) of 67.8 mm. The NuPET scanner has 2 rings of detector blocks with 16 detector blocks per ring. Each detector has a DLO architecture with 11 × 25 crystals in the front layer and 12 × 26 crystals in the back layer. The detectors are made of LYSO with a crystal pitch of 1.28 mm and thickness of 4/6 mm in the front/back layer.

In the MC simulations of NuPET, an energy resolution of 12.5% at 511 keV, energy window of 300 to 700 keV, timing resolution of 2.5 ns and CTW of 10 ns were applied. A schematic drawing of the NuPET scanner is shown in Figure 8.1.

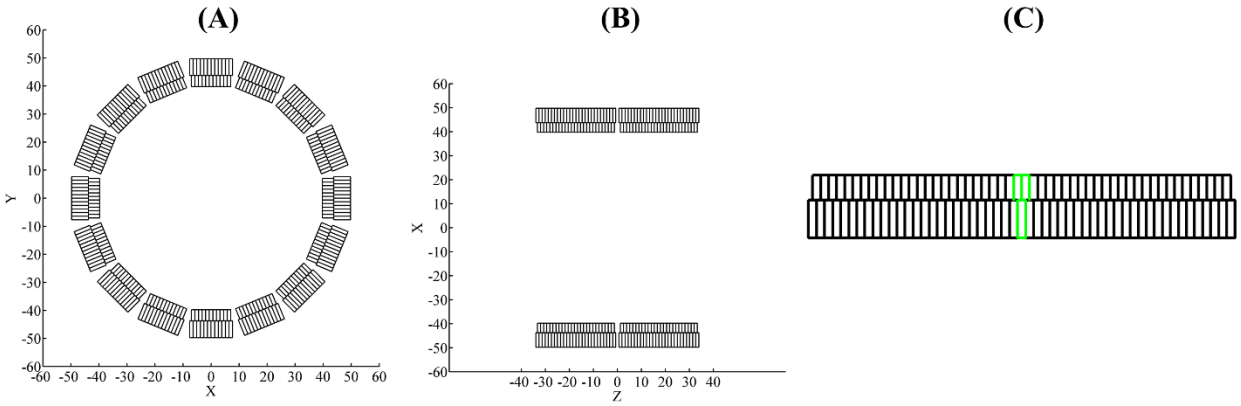

Figure 8.1. A schematic drawing of the NuPET scanner from the transverse (A) and axial (B) views. The added crystal elements, shown in green in image (C), demonstrate the arbitrary crystals assumed in creating the indexed histogram data.

Per Cubresa's terminology, images from the NuPET scanner can be reconstructed using two different *zoom* values (grid and voxel sizes). The most common reconstruction setting is to reconstruct images into a 3D grid of 92 × 92 × 105 voxels with an isotropic voxel size of 0.64 mm (*zoom 1*). Images can also be reconstructed in a grid of 184 × 184 × 210 voxels with an isotropic voxel size of 0.32 mm (*zoom 2*). The current implementation of the iterative image reconstruction algorithm for NuPET (created by Cubresa) benefits from a shift-varying pre-stored system matrix the elements of which have been previously calculated through analytical calculations in an approach similar to [229].

8.3 Indexed Histogram Data

The list-mode data generated by NuPET is in a format known as an *indexed histogram*. In practice, the indexed histogram data is a histogram of the number of recorded LORs along all possible LORs. This format of data is aimed to maintain the original spatial sampling of the recorded events with no rebinning. To assign a histogram index to each LOR, arbitrary crystals were assumed in between the two rings of detectors as if the scanner was built of one ring of

detectors where each detector has 11 × 52 crystals in the front layer and 12 × 53 crystals in the back layer. The arbitrary crystals are shown in green in Figure 8.1-(C).

While the indexed histogram data assume a total of 81,718,784 possible LORs for NuPET, this scanner has 77,183,456 physically existing LORs (the extra LORs are due to assuming arbitrary crystal elements). Over the course of this project, a number of indexed histogram files were received from Cubresa and the Lawson Imaging Centre. The obtained data for each scan included two indexed histogram files corresponding to the true and [an estimation of the] random events. Also, the indexed histogram files obtained from the scan of an annular phantom was provided with each dataset that were used in a normalization step as will be explained in Section 8.4. Each histogram bin in the indexed histogram files is a 4-byte float number representing the number of recorded events along a certain LOR after a correction for detector deadtime. All provided data had previously been corrected for deadtime.

8.4 Component Normalization for NuPET

Since the BrainPET reconstruction algorithm (Chapter 7) was developed for simulated PET scanners with various geometries, no normalization was considered for the BrainPET scanners. Instead, the sensitivity image was obtained through a backprojection of all or a large number of randomly selected LORs assuming the same weight (geometrical probability) for all LORs in a method similar to [217]. In a real scanner though, the efficiency of each LOR varies due to several factors which can be broadly categorized into *geometrical factors* and *crystal intrinsic efficiency* factors. The current section explains the method used for the calculation of these factors for NuPET.

The purpose of a normalization method in PET is to account for variations in the detection probability of different LORs. Therefore, the most straightforward approach to calculate a normalization coefficient for each LOR is by exposing all LORs to a known activity. As mentioned in Section 2.8.5, such a method is known as *direct normalization* and requires a source with minimal attenuation and a long acquisition time to achieve reasonable statistical accuracy. Using the direct normalization method, a normalization coefficient n_{ij} for each LOR connecting crystals i and j can be calculated using equation (2.24)

$$n_{ij} = \frac{N_{ij}}{A_{ij}} \tag{2.24}$$

in which N_{ij} is the number of recorded LORs using the normalization source and A_{ij} is the activity along the crystals i and j. Sample reconstructed images that are shown in this chapter and are provided by the Lawson Imaging Centre have been corrected using the direct normalization approach.

An accurate *component-based normalization* was developed for NuPET to increase the quantitative accuracy of the reconstructed images [66, 164, 230, 231]. Because component normalization does not require a large number of events, it can be done more frequently while providing a more accurate normalization coefficient for each LOR. The normalization coefficient n_{ij} can be considered as a product of *geometrical factors* g_{ij} and *crystal intrinsic efficiency factors* for crystals i (ε_i) and j (ε_j) such that

$$n_{ij} = \varepsilon_i \varepsilon_j g_{ij} \ . \tag{8.1}$$

The geometrical factors (g) account for sensitivity variations between different LORs. As the name implies, such variations are due to different crystal arrangements and orientations. For example, in Figure 8.2, three different LORs are shown such that two LORs share the same geometrical factor while the other one has a different value. While the purpose of component normalization is to reduce the number of independent normalization coefficients, the geometrical factor for each LOR g_{ij} must be calculated independently as it cannot be separated into further components for each individual crystal i and j. Nonetheless, the calculation of geometrical factors can be done only once for each scanner as the geometry of the scanner does not change. Approaches similar to [190] have been proposed in the literature to calculate the geometrical factors through analytical calculations. Such techniques have not been investigated in this dissertation. Instead, MC simulations were used to find the geometrical factors for each LOR.

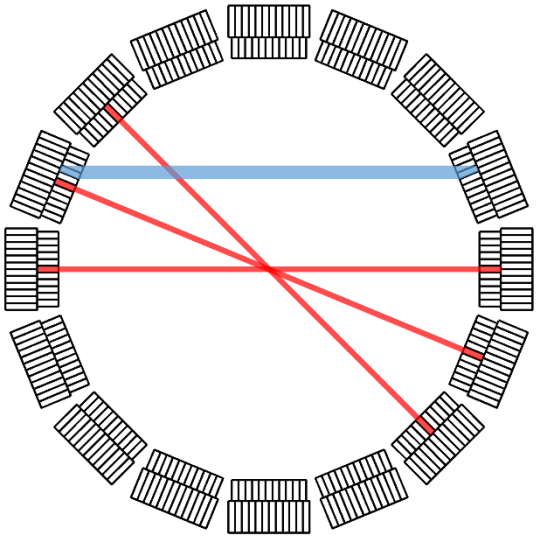

Figure 8.2. Four sample LORs with two different geometrical factors. The LORs shown in red share the same geometrical factors while the blue LOR has a different value.

To increase the accuracy of the calculated geometrical factors, LORs sharing the same geometrical factors (*degenerate LORs*) were grouped together to find an average coefficient. To find a group of degenerate LORs, each LOR was rotated 16 times (or 8 times for certain LORs) to

complete a 2π rotation, then mirrored vertically (about the Y-axis as defined by Figure 8.1) and rotated 16 times (or 8 times for certain LORs) again. For instance, LORs connecting the central crystal elements of the inner layer crystals of two opposite detector blocks in the transverse plane (such as the red LORs in Figure 8.2) require 8 rotations. Finally, the initial LOR was mirrored axially and the mentioned 32 (16 + 16) or 16 (8 + 8) rotations were repeated. An example of a group of degenerate LORs can be seen in Figure 8.3. All degenerate LORs are essentially located at the same radial distance, but not all LORs with the same radial distance can necessarily be grouped into the same degenerate LORs. In the NuPET scanner, most LORs have a degeneracy factor of 64 while some have a degeneracy factor of 32 or 16.

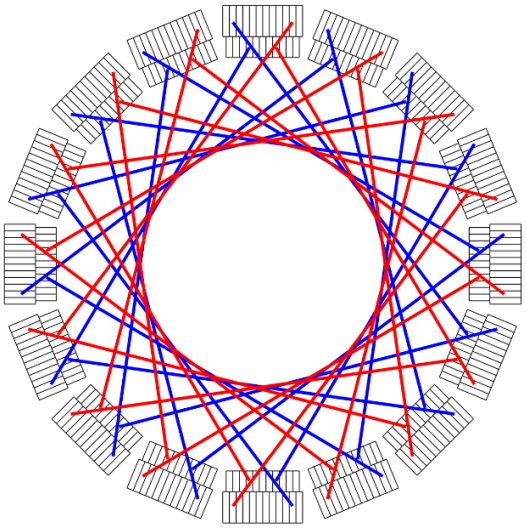

Figure 8.3. An example of degenerate LORs sharing the same geometrical factors along the transverse plane in NuPET. Each blue LOR is obtained through a series of 2π/16 degree rotation and each red LOR is obtained by mirroring a blue LOR followed by 16 rotations.

Categorizing degenerate LORs into the same group reduces the number of required geometrical calculations for 81,718,784 LORs to only 1,280,480 *unique LORs* for the NuPET scanner.

To calculate the geometrical normalization coefficients for each group of degenerate LORs, a large attenuation-free cylindrical phantom (diameter/length: 78.0/70.0 mm) was simulated in GATE with an activity of 100 Bq and no half-life. A very low activity was used in this phantom to ensure a negligible random contribution. The simulation was executed for a very long time until nearly one billion coincidences were recorded. The geometrical factor for a group of degenerate LORs (in group *S*) was then calculated as

$$g_S = \frac{\sum_{ij \in S} N_{ij}}{A_S \text{size}(S)} \tag{8.2}$$

where A_S is the amount of activity exposing crystals *i* and *j* for any pair of crystals *i*-*j* in group *S* and size(*S*) is the degeneracy factor of group *S*. Obviously, the amount of activity exposing all crystal pairs *i*-*j* in a degenerate group *S* must be identical. The choice of a uniform cylindrical phantom satisfies this condition for all groups of degenerate LORs. Figure 8.4 shows a histogram of the geometrical factors for all possible LORs in the NuPET scanner.

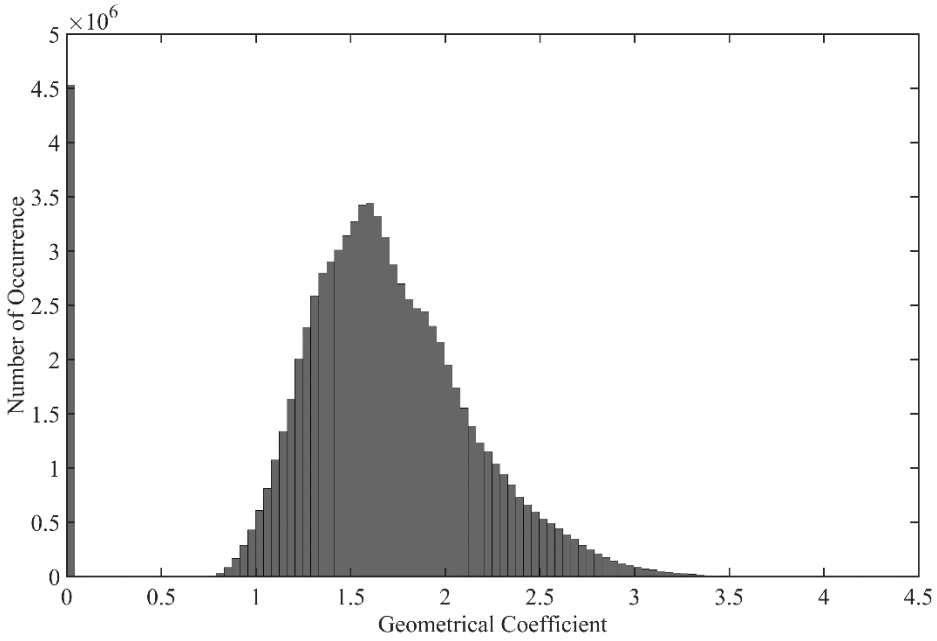

Figure 8.4. A histogram of the geometrical coefficients for all LORs in the NuPET scanner.

As mentioned before, the geometrical factors are constant and need to be found only once. The crystal intrinsic efficiency factors (ε), however, should be calculated periodically as crystals efficiency factors drift over time. The extended version of the component-based normalization due to Defrise to 3D data [66, 155, 231] was used in this dissertation to find the crystal intrinsic efficiency factors ε. The normalization phantom designed by Cubresa is an annular phantom with an inner diameter of 50 mm, an outer diameter of 58 mm and an axial length of ~134 mm. The phantom, uniformly filled with radioactivity, is placed at the centre of the gantry and then scanned to enable the calculation of normalization coefficients for each LOR.

The component-based normalization method groups the crystals of the PET scanner into G groups of M crystal elements. In the developed component-based normalization method for NuPET, each group G incorporates $(12 \times 53) + (11 \times 52) = 1208$ (i.e. $M = 1208$) crystals out of which 1174 crystals physically exist, and the other 34 crystals are arbitrary crystal elements assumed between the two rings of detectors (crystals shown in green in Figure 8.1-(C)). For each

LOR connecting crystals *i*-*j*, a measurement of the product of the intrinsic coefficients for crystals *i* and *j* can be found as

$$\eta_{ij} = \frac{N_{ij}}{g_{ij} A_{ij}} = \varepsilon_i \varepsilon_j \ . \tag{8.3}$$

Note that an A_{ij} of 0 for an LOR can lead to a miscalculation, indicating that the normalization phantom must be designed such that *all* LORs that will be considered in the reconstruction are exposed to activity. A mean detector efficiency $\langle \varepsilon \rangle_A$ for all crystals in group *A* is defined as

$$\langle \varepsilon \rangle_A = \frac{1}{M} \sum_{i \in A} \varepsilon_i \ . \tag{8.4}$$

The average LOR efficiency between crystal *i* in group *A* and all crystals of another group *B* is

$$\langle \eta \rangle_{i,B} = \frac{1}{M} \sum_{j \in B} \eta_{ij} \tag{8.5}$$

and therefore, the average LOR efficiency of all LORs between group two groups *A* and *B*, $\langle \eta \rangle_{A,B}$ is

$$\langle \eta \rangle_{A,B} = \frac{1}{M^2} \sum_{i \in A} \sum_{j \in B} \eta_{ij} \ . \tag{8.6}$$

Using equations (2.24), (8.1), (8.5) and (8.6), $\langle \eta \rangle_{A,B}$ and $\langle \eta \rangle_{i,B}$ can be written as

$$\langle \eta \rangle_{A,B} = \langle \varepsilon \rangle_A \langle \varepsilon \rangle_B \tag{8.7}$$

$$\langle \eta \rangle_{i,B} = \varepsilon_i \langle \varepsilon \rangle_B \ . \tag{8.8}$$

Defrise *et al.* [231] suggested a calculation method for each ε_i as follows. Each detector should be in coincidence with 2*K*+1 other detectors and all LORs should be uniformly illuminated. A group *A* with detectors *i* is assumed such that $AM \le i \le AM + M$. Defrise *et al*. [165] showed that if the group size *M* is chosen such that group *A* is in coincidence with 3 opposing groups, using the previously measured η_{ij}, the average efficiency of group *A* ($\langle \varepsilon \rangle_A$) can be calculated as

$$\langle \varepsilon \rangle_A = \left(\langle \eta \rangle_{A,A+G/2} \prod_{k=0}^{G/2-1} \frac{\langle \eta \rangle_{A+k,A+k+G/2}}{\langle \eta \rangle_{A+k+1,A+k+G/2}} \right)^{1/2} \tag{8.9}$$

and by using equations (8.7) and (8.8), an equation for each crystal intrinsic efficiency can be found as

$$\varepsilon_i = \frac{1}{3} \sum_{q=-1}^{1} \frac{\langle \eta \rangle_{i,(A+G/2+q)}}{\langle \varepsilon \rangle_{A+G/2+q}} \ . \tag{8.10}$$

Badawi *et al*. [164] extended equation (8.9) to cases where group *A* is in coincidence with *T* opposing groups, where *T* is an odd number, and found a more general form of equation (8.9) as

$$\langle \varepsilon \rangle_A = \frac{1}{T-1} \sum_{P=-S}^{S-1} \left(\langle \eta \rangle_{A,A+G/2} \prod_{k=0}^{G/2-1} \frac{\langle \eta \rangle_{A+k,A+k+G/2-p}}{\langle \eta \rangle_{A+k+1,A+k+G/2-p}} \right)^{1/2} \tag{8.11}$$

where $S = (T-1)/2$. Also, a more general form of equation (8.10) was found to be

$$\varepsilon_i = \frac{1}{T} \sum_{q=-S}^{S} \frac{\langle \eta \rangle_{i,(A+G/2+q)}}{\langle \varepsilon \rangle_{A+G/2+q}} \ . \tag{8.12}$$

The NuPET scanner is designed such that each detector is in coincidence with 7 opposing detector blocks (i.e. a minimum block difference of 5). The currently designed annular normalization phantom with an inner/outer diameter of 50/58 mm misses certain LORs as shown in Figure 8.5.

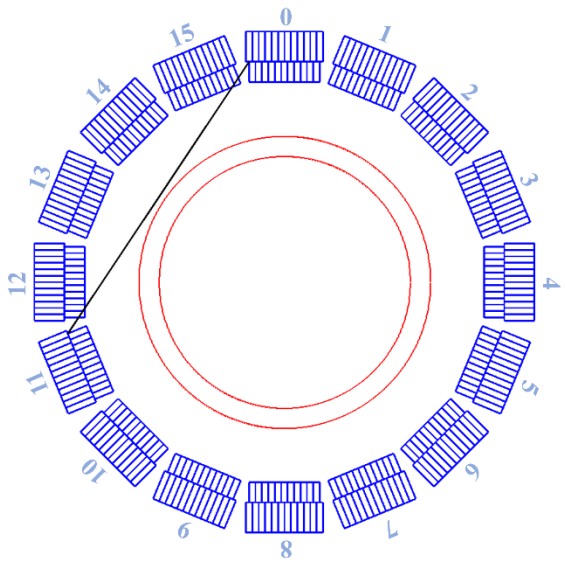

Figure 8.5. A drawing of the designed annular normalization phantom and an example of an LOR that cannot be recorded using the annular normalization phantom.

The lack of complete data in turn introduces quantitative instabilities in the corner crystals of the back layer if a T of 7 is chosen. To demonstrate the problem, Figure 8.6 shows a map of the crystal efficiency factors (ε_i) for all crystals i of the simulated NuPET scanner when T was set to 7. Note that because these data were obtained from GATE with no perturbation in the crystal efficiencies, a uniform crystal efficiency map for each detector is expected.

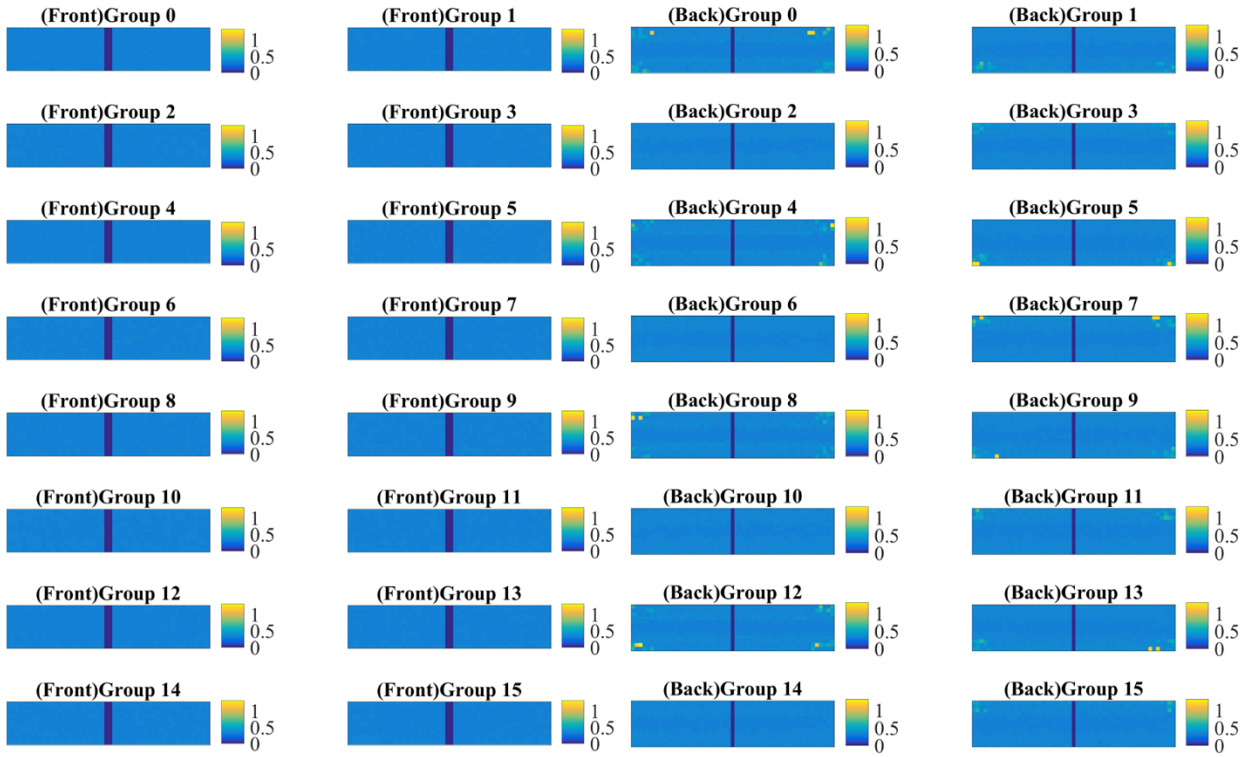

Figure 8.6. A map of the crystal efficiency ε_i for all crystals *i* of the simulated NuPET scanner when using the described component normalization method with $T = 7$. Note how the edge and corner crystals in the back layer show a spike in their efficiency factors. Image intensity for all images is globally normalized.

While a uniform crystal efficiency map for all crystals was expected, some crystals of the back layer (around the corner of each detector) showed random instabilities. Such incorrect crystal efficiency factors in turn introduce artifacts and affect the quantitative accuracy of the reconstructed images. Figure 8.7 shows a map of crystal efficiency factors for the simulated NuPET scanner when *T* was set to 5.

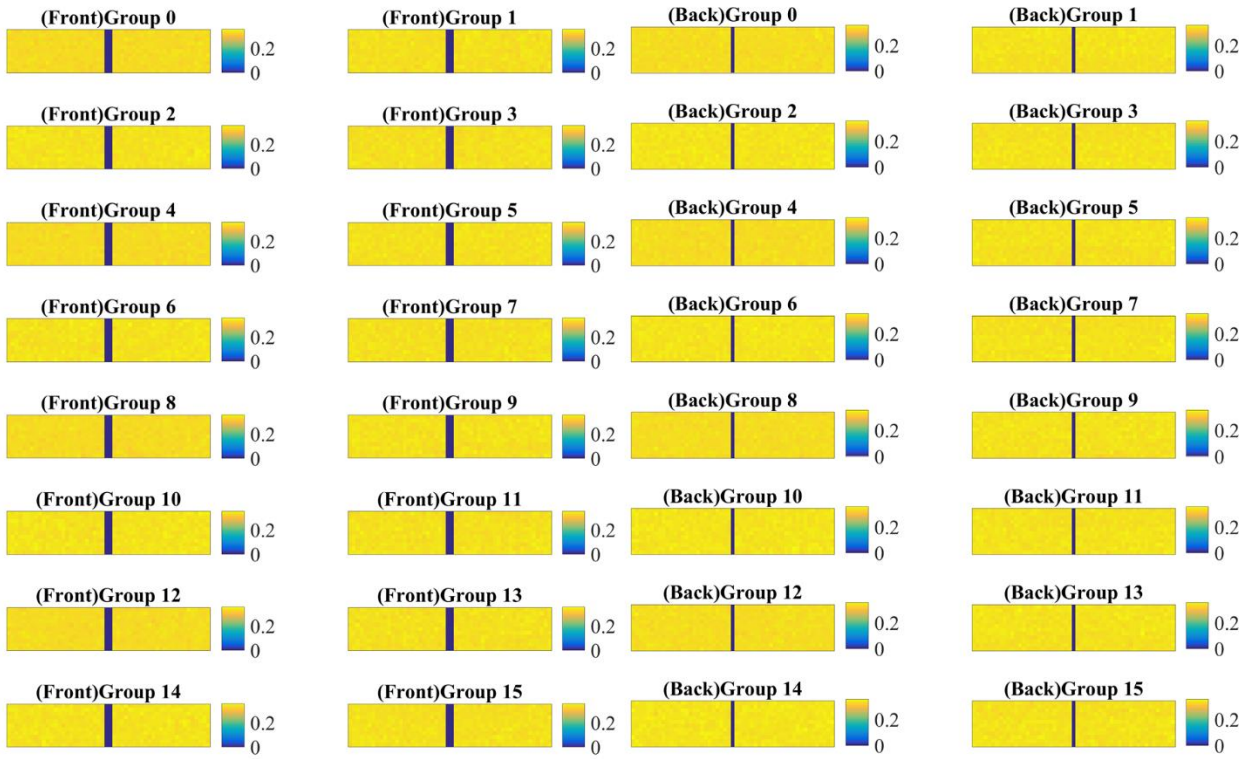

Figure 8.7. A map of the crystal efficiency ε_i for all crystals *i* of the simulated NuPET scanner when using the described component normalization method with $T = 5$. Image intensity for all images is globally normalized.

A significantly more uniform crystal efficiency map for the crystals of each detector was observed when a T of 5 was used. Therefore, for all further imaging studies T was assumed to be 5. The reconstructed images of the phantoms shown in this chapter are corrected for normalization using the developed component-based normalization for NuPET. However, some normalization data may date back to several months[1] before the data acquisition of the phantom. Due to the high likelihood of a drift in the crystal intrinsic efficiency factors over time, normalization artifacts are expected when outdated normalization data are used (to the best of our knowledge, no published study has recommended how often the normalization factors should be recalculated). The amount of activity used in each scan for the data shown in this chapter was also unknown to us.

[1] The exact date and condition of all data acquisitions (not simulations) were unknown.

Normalization artifacts are also likely to occur when a significantly different amount of activity is used in the normalization phantom than the activity used in the phantom.

8.5 GPU Hardware

Two Nvidia GeForce GTX Titan Black graphic cards were employed to develop a dual-GPU list-mode ML-EM image reconstruction algorithm for NuPET. The properties of these graphic cards have been previously explained in Table 3.1. To maintain consistency between the developed dual-GPU image reconstruction and the NuPET's integrated image reconstruction algorithm, images were reconstructed using the same voxel size (0.64/0.32 mm isotopically for zoom 1/2). Since it was desirable to engage both GPUs for all reconstructions, the developed reconstruction algorithm for BrainPET scanners was modified as will be explained in the next section. The developed CUDA program for NuPET was compiled with NVCC using CUDA Toolkit 10 and G++ compiler using C++17 libraries with option -O3 in a Linux environment.

8.6 GPU Memory Layout in NuPET

As explained in Section 7.3, the developed BrainPET image reconstruction algorithm benefits from the second GPU only when the image can be split into more than 1 cube. To split the reconstruction task between the two GPUs at all times, a different memory layout was implemented for NuPET, such that the list of coincidences is divided between the devices (as opposed to the BrainPET implementation in which the image is divided between the devices). The image is broken into 3D cubes from which tiles of $T_d \times T_d$ pixels are extracted and loaded into the shared memory of each block. Given the reconstruction grid size of 92 × 92 × 105 voxels for zoom 1, the tile size (T_d) was set to 105 in all reconstructions. Initial evaluation showed that a uniform grid size of 105/210 for zoom 1/2 results in a faster reconstruction than a grid size of 92 × 92 × 105 for zoom 1 and 184 × 184 × 210 voxels for zoom 2. Therefore, images were reconstructed in

a grid of 105 × 105 × 105 for zoom 1 (i.e. 1 cube) and 210 × 210 × 210 for zoom 2 (i.e 8 cubes) and then truncated into 92 × 92 × 105 (zoom 1) or 184 × 184 × 210 (zoom 2) for presentation.

Similar to the approach discussed in Section 7.4, the recorded coincidences were partitioned according to their predominant direction and stored in a SoA. Because NuPET data are arranged into an indexed histogram format, all zero bins (LORs with zero record) were discarded and the list of coincidences is shuffled if more than 1 subset is used (recall that unlike the BrainPET data where the LOR positions in the list-mode data were very sparse, the indexed histogram format sorts the LORs according to their coordinates). Unlike the memory layout for BrainPET, the entire list [histogram] of coincidences are then split into the global memory of the two devices. The global memory of both devices also hosts a crystal LUT for NuPET. Similar to the BrainPET reconstruction algorithm, a 1D texture is defined in the texture memory of each GPU to model a shift-invariant PSF as a Gaussian function with a user-defined FWHM (the default is equal to the crystal pitch of the scanner or 1.28 mm) with 2048 elements.

The memory layout for NuPET is schematically shown in Figure 8.8. For comparison, this memory layout can be compared to Figure 7.1 in which the memory layout for the BrainPET scanners was shown.

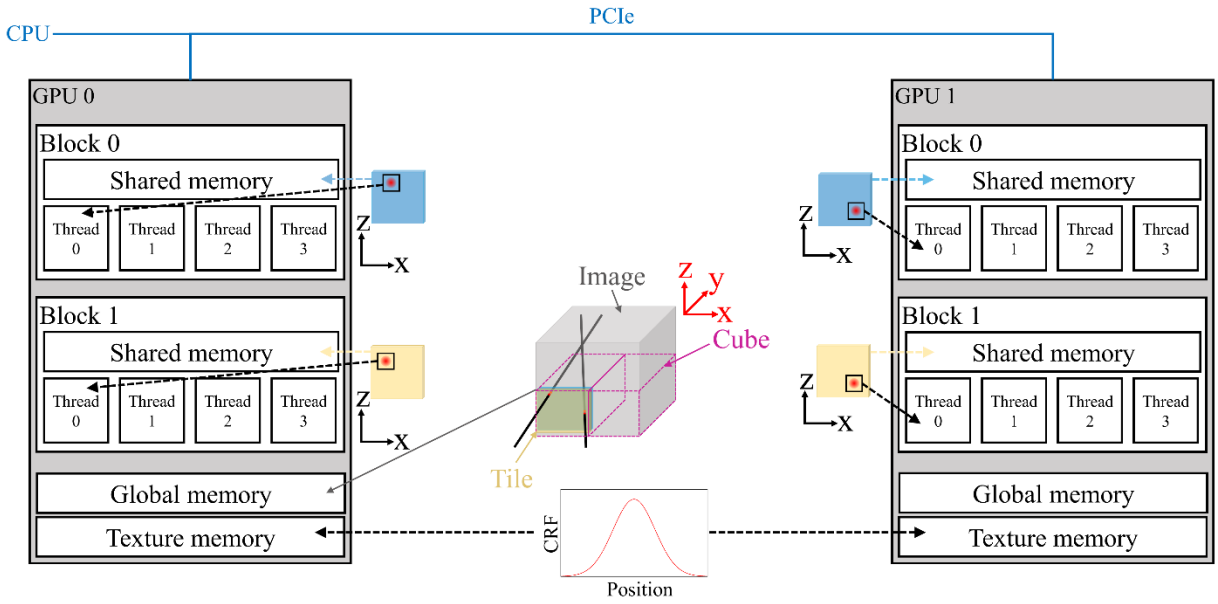

Figure 8.8. Memory layout in the developed dual-GPU reconstruction algorithm for NuPET.

8.7 Image Reconstruction and Performance Analysis

The developed forward and backprojection kernels are essentially similar to the algorithms discussed for the BrainPET scanners in Sections 7.5 and 7.6. Unlike the developed BrainPET algorithm, the backprojection matrices on the two GPUs are merged into a single backprojection matrix prior to the multiplicative image update step in the NuPET implementation. The existence of a single backprojection matrix allows for the incorporation of a prior term regardless of the number of voxels in the image (therefore, MAP reconstruction can be safely used regardless of the image voxel size). To compare the performance of the developed dual-GPU image reconstruction for NuPET against that of the BrainPET scanners on a single and dual GPU, varying numbers of randomly generated LORs were reconstructed (forward projection, backprojection and image update for 1 iteration) using each technique. Ten different sets of LORs, between 1.0×10^5 to 8.2×10^7 LORs, were generated and the processing time was measured for each case. Three different grid sizes with a large number of voxels were chosen for this comparison study. Namely, grid sizes of $210 \times 210 \times 210$, $315 \times 315 \times 315$ and $420 \times 420 \times 420$ voxels were selected. All reconstructions

were done given a patch size ($T_w \times T_w$) of 5 × 5 pixels to generate a TOR. Also, in all reconstructions 1024 threads per block with 512 thread blocks were assumed. The measured elapsed times for grid sizes of 210 × 210 × 210, 315 × 315 × 315 and 420 × 420 × 420 voxels are shown in Figure 8.9.

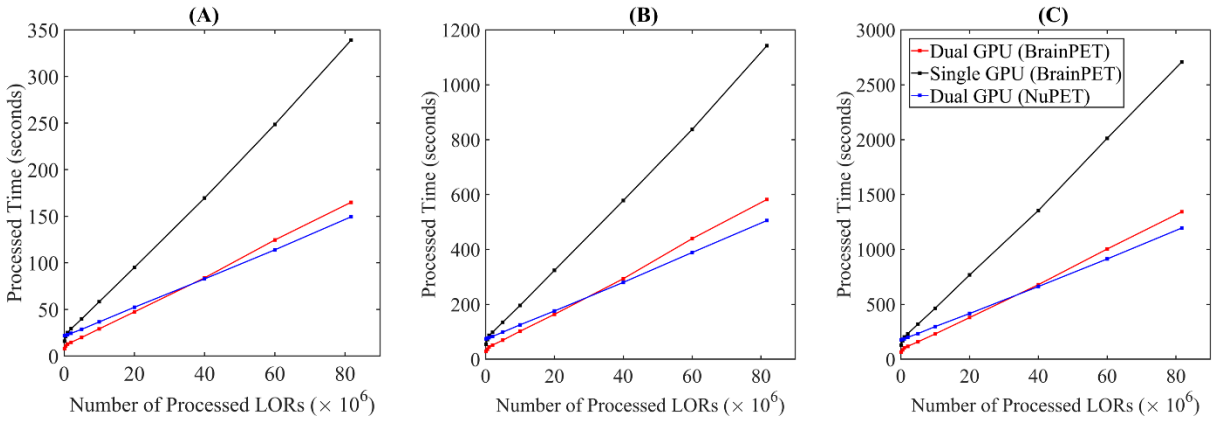

Figure 8.9. Processing time to complete one iteration of list-mode ML-EM for different numbers of events using single and different dual-GPU implementations. The list of coincidences in each case was reconstructed in a grid of (A) 210 × 210 × 210, (B) 315 × 315 × 315 and (C) 420 × 420 × 420 voxels with a fixed T_w width of 5 voxels.

The plots in Figure 8.9 show that the BrainPET implementation has a faster processing time than the NuPET implementation for less than ~35 million events. However, the NuPET implementation was more optimum when more than ~35 million events are processed. The explained trend seems to be independent of the number of image voxels. The plots shown in Figure 8.9 can be used to estimate the reconstruction time for various reconstruction settings. Lastly, it should be noted that while a grid size of 210 × 210 × 210 is actually used as zoom 2 in the NuPET image reconstruction algorithm, the other grid sizes (i.e. 315 × 315 × 315 and 420 × 420 × 420) are only used for comparison purposes in this section.

Currently, only list-mode ML-EM and MAP reconstruction with the quadratic prior (equation (3.28)) has been implemented in the developed reconstruction algorithm for NuPET. Considering that the data processing pipeline of the NuPET scanner calculates an estimated

contribution of the random events to each coincidence bin, the list-mode ML-EM algorithm, as described by equation (3.20), was reformulated as [152, 232]

$$\overline{f}_j^{k,(l+1)} = \frac{\overline{f}_j^{k,l}}{\sum_{i=1}^{n} a_{ij}} \sum_{i \in S^l} \frac{1}{\sum_{j'=1}^{m} a_{ij'} \overline{f}_{j'}^{k,l} + \overline{r}_i} a_{ij} \tag{8.13}$$

in which $\overline{r}_i$ is an estimated contribution of the random events to LOR *i* and all other variables are as defined previously in equation (3.20).

8.8 NEMA NU-4 Image Quality Phantom Images

During the course of this project, several datasets of NuPET acquisitions were provided to us. In the following sections, reconstructed images using simulated and acquired data from the NuPET scanner are shown. While all these images were corrected for random events, no attenuation correction was applied due to the absence of a co-registered attenuation map.

8.8.1 **Monte Carlo Simulations**

The performance analysis of the developed image reconstruction algorithm for NuPET was begun by evaluating the reconstruction algorithm on simulated data. A GATE simulation of the NEMA NU-4 phantom [166] was done with a very low activity of 370 Bq in the phantom with an infinite half-life. The chosen low activity virtually eliminates the contribution of random events. The simulation continued until nearly 1.7×10^8 events were recorded. Coincidences were then binned into an indexed histogram format and passed into the developed dual-GPU image reconstruction algorithm. The PSF was modeled as a Gaussian function with a FWHM of 0.768 mm (0.6 × the crystal pitch of the scanner) in these reconstructions. Images were reconstructed without applying a prior term using 50 iterations and 1 subset with both zoom 1 and zoom 2 in two

different reconstructions. Sample slices of the reconstructed images are shown in Figure 8.10 for zoom 1 and Figure 8.11 for zoom 2.

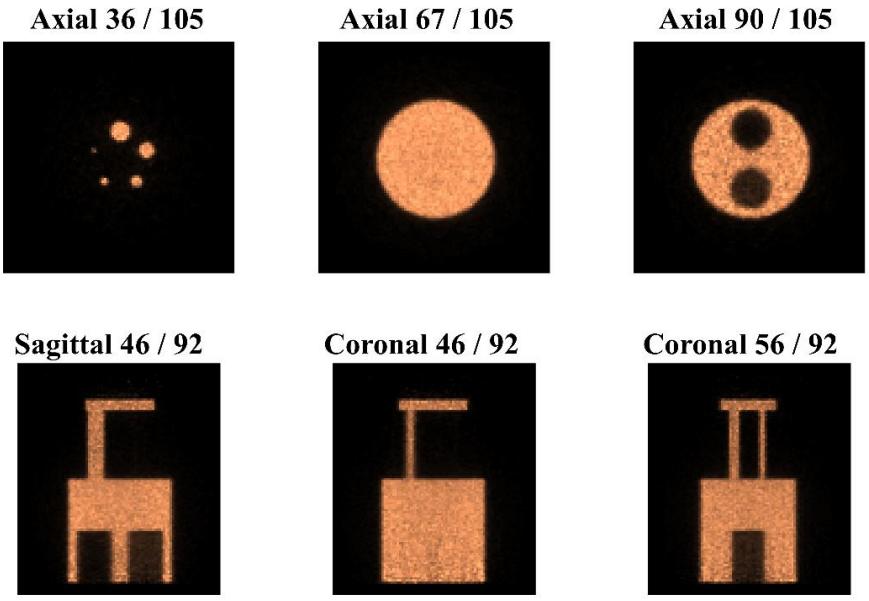

Figure 8.10. Sample reconstructed slices of a NEMA NU-4 image quality phantom. A list of ~170 million events, obtained from a GATE simulation, was used to reconstruct these images using zoom 1 with 50 iterations. The intensity of each image is normalized to its maximum value in the slice.

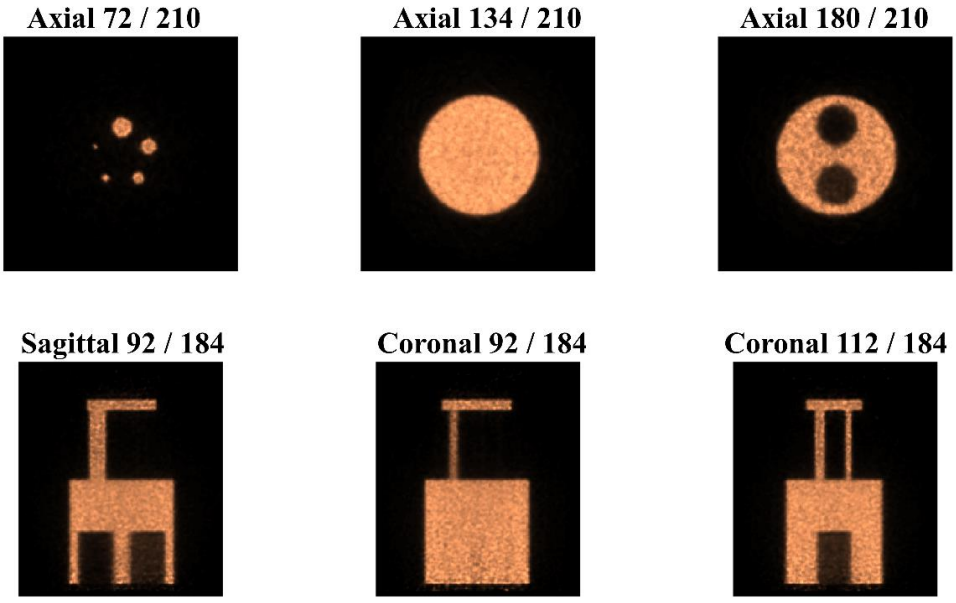

Figure 8.11. Sample reconstructed slices of a NEMA NU-4 image quality phantom. A list of ~170 million events, obtained from a GATE simulation, was used to reconstruct these images using zoom 2 with 50 iterations. The intensity of each image is normalized to its maximum value in the slice.

The reconstructed images were visually inspected with no quantitative analysis. The images of both zoom values showed no obvious artifacts. Images reconstructed with zoom 1 were reconstructed in 11.5 seconds per iteration (about 9.5 minutes for 50 iterations) and it took 109.4 seconds per iteration (about 89.5 minutes for 50 iterations) for images of zoom 2. It might be interesting to note that in this relatively long acquisition, only about ~39 million of the possible 81.7 million histogram bins (to be exact, 48.1% of the indexed histogram bins) were filled and the rest remained empty.

8.8.2 **Acquired Data**

Because the data provided to us were quantitatively corrected for deadtime and the NuPET acquisition software stores the acquisition parameters in a separate Dicom file, the exact number of recorded coincidences was unknown for the data in this section. Figure 8.12 shows an example set of the reconstructed images using zoom 1 with 50 iterations. For this set of recorded data (dataset 1), a set of normalization coefficients (incorporating all normalization components) was provided by Cubresa in lieu of the normalization phantom scan data. The normalization coefficients were obtained using the direct normalization method. The provided normalization coefficients were directly employed in the developed reconstruction algorithm. In these images, a shift-invariant PSF with a FWHM of 1.28 mm (i.e. equal to the crystal pitch) was modeled in both the forward and backprojection kernels. As the phantom was placed with a different orientation for each scan, it was not feasible to show matching slices for different scans. Therefore, selected slices showing cross sections across various parts of the NEMA NU-4 phantom are shown. About 63% of the indexed histogram bins were filled and a reconstruction time of ~14 seconds per iteration was recorded for this reconstruction.

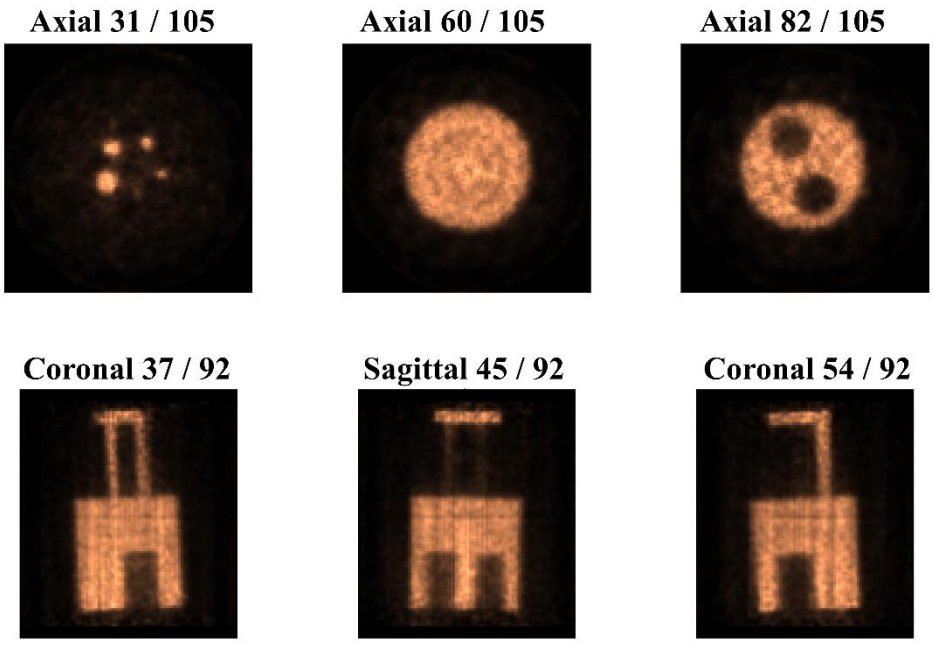

Figure 8.12. Sample reconstructed slices of a NEMA NU-4 image quality phantom. The data were acquired by Cubresa (dataset 1) and the images were reconstructed using zoom 1 with 50 iterations. The intensity of each image is normalized to its maximum value in the slice.

The random hot and cold strips along the sagittal and coronal images indicate artifacts likely due to a mismatch between the normalization data and the acquired data (since a drift in the normalization factors may occur over time, the normalization scan should ideally be acquired with a short time gap from the actual data acquisition). Since the raw normalization data (the scan of the annular phantom) were not available for this dataset, further investigations were not feasible.

The developed image reconstruction algorithm was also evaluated on a second set of data (dataset 2) for the NEMA NU-4 phantom. Since the raw data of the normalization phantom were also available, the developed component-based normalization was used to find the normalization coefficients using the method described in Section 8.4. The indexed histogram data suggested that about 40% of the possible LORs had recorded events and the rest were empty. The exact number of recorded LORs was not known. Images of this dataset were also reconstructed using 50

iterations and 1 subset with zoom 1 and are shown in Figure 8.13. A shift-invariant PSF with a FWHM of 1.28 mm was modeled in the image domain of the reconstruction algorithm.

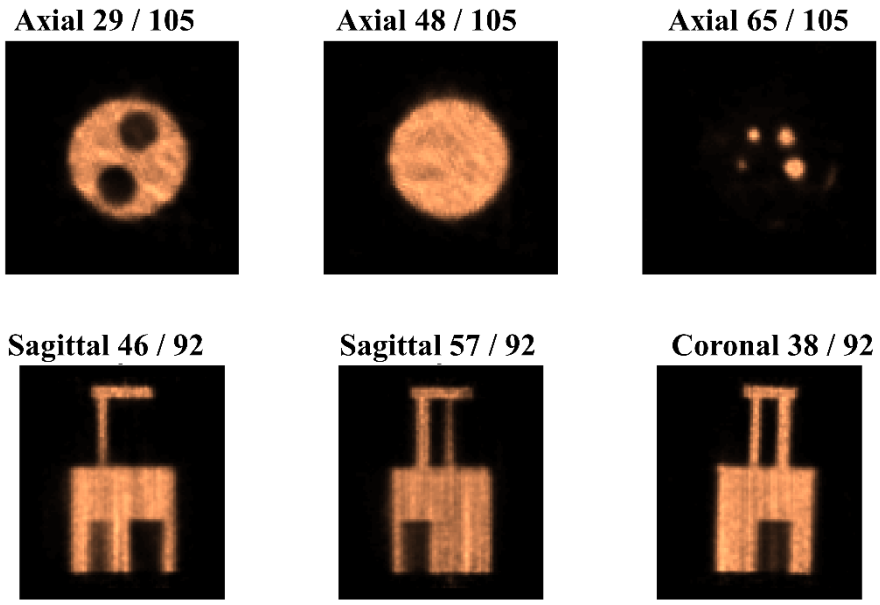

Figure 8.13. Sample reconstructed slices of a NEMA NU-4 image quality phantom. The data were acquired by Cubresa (dataset 2) and the images were reconstructed using zoom 1 with 50 iterations. The intensity of each image is normalized to its maximum value in the slice.

A reconstruction time of 10.8 seconds per iteration was recorded for the images of dataset 2. The images shown in Figure 8.13 demonstrated striking artifacts which could be due to a mismatch between the normalization and phantom acquisition data (e.g. a long gap between the two acquisition dates) or also partially due to an inevitable mismatch between the true PSF and the chosen shift-invariant PSF.

A third dataset (dataset 3) was also received from Cubresa and used to evaluate the developed reconstructed algorithm. Since the raw data for the scan of the annular normalization phantom were available, the developed component-based normalization method was employed to find the normalization coefficients for each LOR. Inspecting the provided data showed that 73.8% of the indexed histogram bins were filled (about 60 million out of 81.7 million). Images were

reconstructed using zoom 1 with 50 iterations and 1 subset. The average reconstruction time was 17.8 seconds per iteration. Similar to the previous two datasets, PSF was modeled as a shift-invariant Gaussian function with a FWHM of 1.28 mm. The reconstructed images of dataset 3 are shown in Figure 8.14.

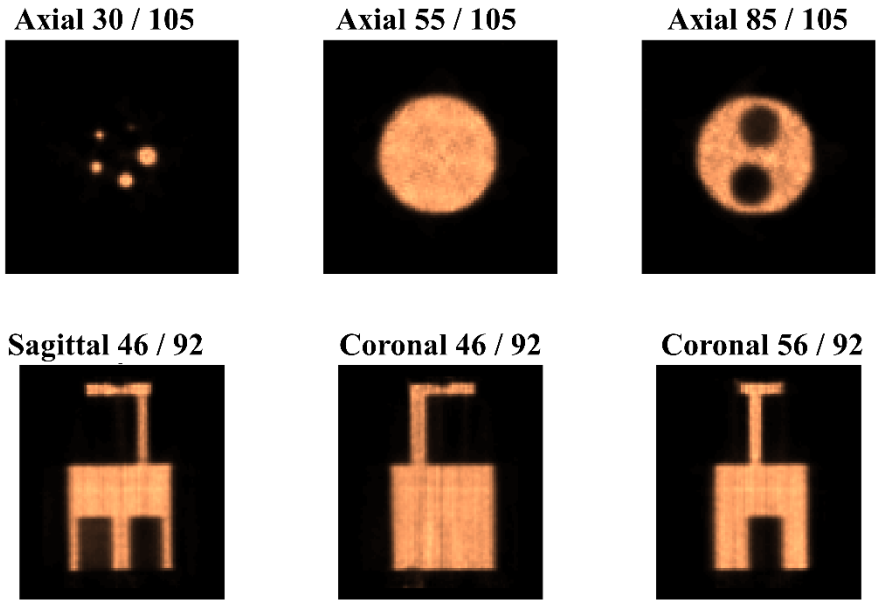

Figure 8.14. Sample reconstructed slices of a NEMA NU-4 image quality phantom. The data were acquired by Cubresa (dataset 3) and the images were reconstructed using zoom 1 with 50 iterations. The intensity of each image is normalized to its maximum value in the slice.

Visual inspection of the images suggests that the reconstructed images for dataset 3 have a relatively better image quality with mild artifacts consistently present along the axial direction.

Images of dataset 3 were also reconstructed using the developed dual-GPU MAP reconstruction algorithm with a quadratic penalty ($\beta = 300$). The images were reconstructed using 50 iterations and 1 subset with zoom 1. The same chosen slices shown in Figure 8.14 are shown in Figure 8.15.

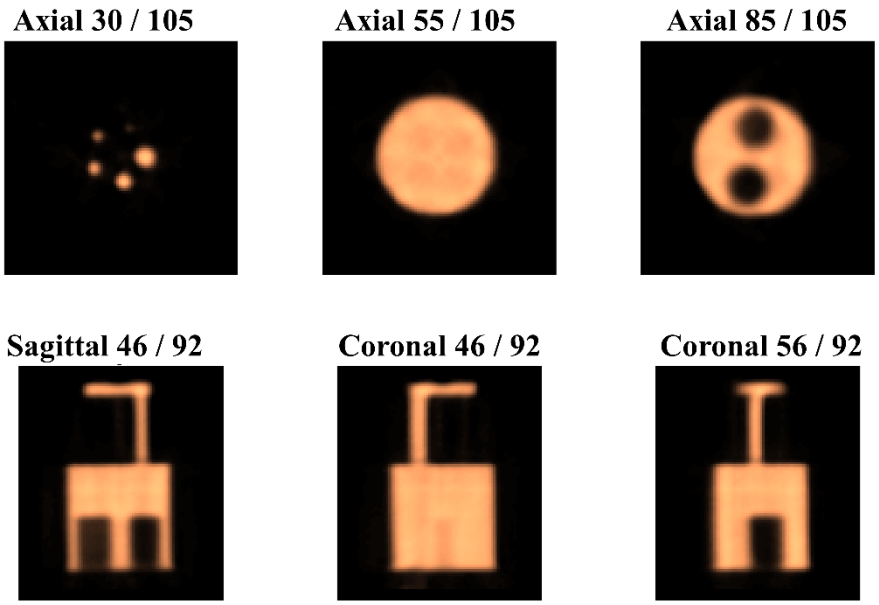

Figure 8.15. Sample reconstructed slices of a NEMA NU-4 image quality phantom. The data were acquired by Cubresa (dataset 3) and the images were reconstructed using the developed MAP reconstruction algorithm with β 300 and zoom 1 with 50 iterations. The intensity of each image is normalized to its maximum value in the slice.

No further quantitative evaluation was performed on the images as multiple datasets were used in this section. Incorporating a quadratic penalty term in the reconstructed images shown in Figure 8.15 maintained the uniformity of the images better than the list-mode ML-EM images shown in Figure 8.14. The existence of a penalty term avoids noisy images even when a larger number of iterations is used.

8.9 Uniform Phantom Images

Recently, data for a scan of a uniform phantom were provided to us by the Lawson Imaging Centre that are discussed in this section. The phantom is a uniform cylindrical phantom with a diameter of ~35 mm and a long axial length such that it covers the entire FOV along the axial direction. About 50.7 million out of 81.7 million (~62%) of the indexed histogram bins were filled in the available data. Images were reconstructed with 25 iterations and 4 subsets using zoom 1

with an average reconstruction time of 6.4 seconds per subset (25.6 seconds per iteration). Axial images of slice number 57 are shown in Figure 8.16.

We were also provided with some reconstructed images by the default NuPET reconstruction software using 2D FBP and 3D reprojection FBP analytical approaches [155], a list-mode ML-EM with a shift-varying PSF model [229] and MAP reconstructed images using STIR [175] with the so-called Ordered Subset Version of Green's MAP One-step-late algorithm [187] (or *OSMAPOSL* as referred to by the STIR community). These reconstructed images were also shown for comparison purposes in Figure 8.16. Note that images reconstructed with the FBP approach (Figure 8.16-(A) and Figure 8.16-(D)) contain negative pixels. The images shown in Figure 8.16 that are reconstructed with an iterative approach demonstrated superior image quality in comparison with analytical approaches. The use of a quadratic prior term in Figure 8.16-(F) (β was set to 750), prevented noisy images. While the reconstructed images with the NuPET reconstruction software incorporated direct normalization, a component-based normalization was used in the case of the dual-GPU reconstruction.

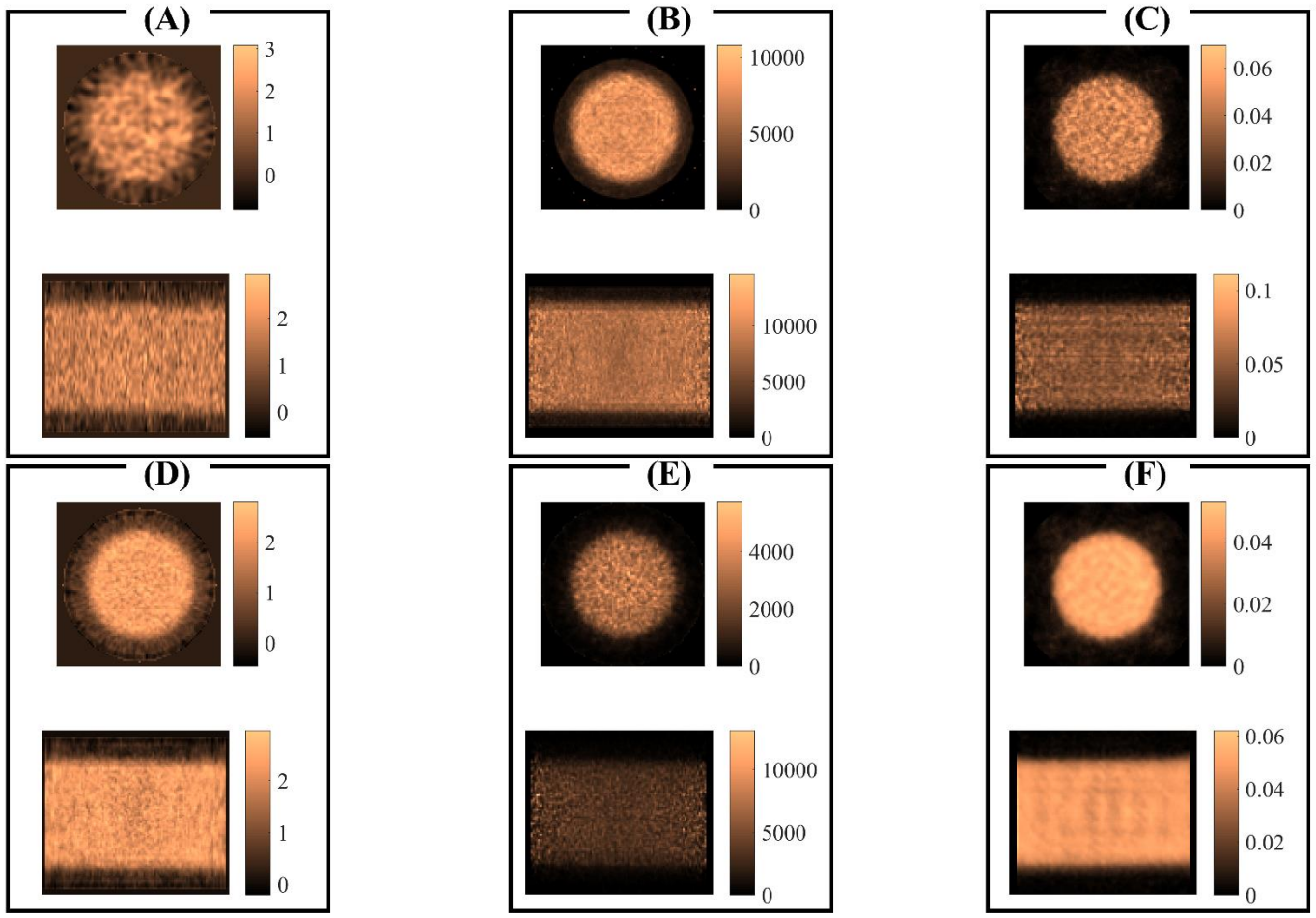

Figure 8.16. Sample reconstructed images of a uniform phantom with different reconstruction algorithms: (A) 2D FBP, (B) OSMAPOSL, (C) the developed dual-GPU list-mode ML-EM algorithm, (D) 3D reprojection FBP, (E) NuPET's integrated list-mode ML-EM with a shift-varying PSF model and (F) the developed dual-GPU list-mode MAP reconstruction with a quadratic penalty (β = 750). In each group, the top image shows axial slice #57/110 and the bottom image shows coronal slice #45/92. All images were reconstructed with an iterative approach were reconstructed using 25 iterations and 4 subsets. Image values are not normalized to the same maximum pixel value and the intensity of each image is normalized to its maximum value (images (A), (B), (D) and (E): image courtesy of the Lawson Imaging Centre, London, ON).

To better demonstrate statistical fluctuations in the images, a horizontal profile (across the middle row of the axial slices in Figure 8.16) was drawn for each image that is shown in Figure 8.17. The plots of Figure 8.17 is an example of how the developed MAP reconstruction can prevent noise in the reconstructed images.

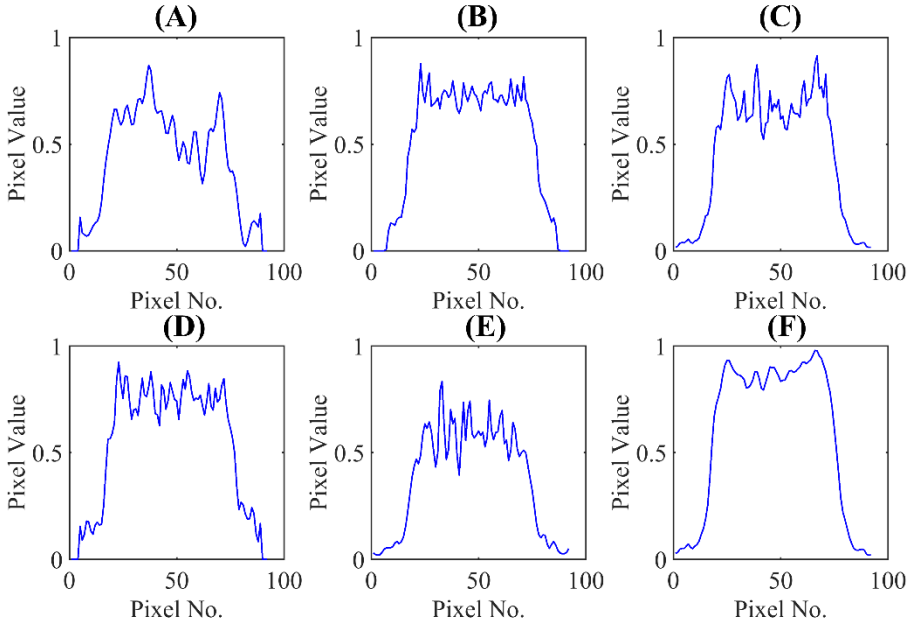

Figure 8.17. Normalized horizontal profiles across the centre of the shown axial images in Figure 8.16. Profiles (A) to (E) correspond to the top (axial) images (A) to (E) shown in Figure 8.16, respectively, such that: (A) 2D FBP, (B) OSMAPOSL, (C) the developed dual-GPU list-mode ML-EM algorithm, (D) 3D reprojection FBP, (E) NuPET's integrated list-mode ML-EM with a shift-varying PSF model and (F) the developed dual-GPU list-mode MAP reconstruction with a quadratic penalty (β = 750). Each profile is normalized to the maximum pixel intensity in the corresponding image.

8.10 Summary

The previously developed dual-GPU reconstruction algorithm for BrainPET scanners was revised to activate both GPUs at all times when reconstructing NuPET data and be compatible with the indexed histogram data format. A component-based normalization method was developed to calculate accurate normalization coefficients for each acquired dataset. With respect to the processing time, it was shown that the developed image reconstruction algorithm for NuPET performs worse/better than the BrainPET algorithm for less/more than ~35 million [unique] LORs

when running on two GPUs. Nonetheless, the dual-GPU implementation for NuPET benefits from both GPUs for the determined image voxel sizes (zoom 1 and 2).

The images of the acquired data from MC simulations with GATE were reconstructed using the developed reconstruction algorithm and showed no obvious artifacts. The images of the NEMA NU-4 image quality phantom reconstructed from acquired data with the NuPET scanner however, showed varying degrees of nonuniformities. Since no systematic artifacts were identifiable across different sets of reconstructed images, the source of the artifacts could be a mismatch between the normalization and acquired NEMA NU-4 image quality data (e.g. long gap between the acquisition dates of the normalization and NEMA NU-4 image quality phantoms). The amount of activity in the phantoms and system count rate for each acquisition were not provided to us. The observed artifacts may also be due to a large activity difference (and consequently system count rate difference) in the NEMA NU-4 image quality and the annular normalization phantom.

A quadratic penalty term was also implemented and the images reconstructed using the MAP algorithm with the quadratic prior were shown in this chapter as well. Our initial evaluation of the developed MAP algorithm showed that the implemented GPU-based MAP algorithm with the quadratic prior can better suppress noise in the images when comparing to the equivalent list-mode ML-EM images. A more thorough evaluation of the developed list-mode ML-EM and MAP reconstruction algorithms for NuPET with new datasets is recommended.

The developed dual-GPU image reconstruction algorithm for the NuPET scanner has a reconstruction time of maximum 18 seconds per iteration (zoom 1). The overall reconstruction time can further be reduced by splitting list-mode data into subsets. The achieved timing performance is sufficient for the mentioned intraoperative prostate imaging project explained in

Section 8.1, especially when considering that intra-operative prostate PET scans usually have a much lower number of recorded events (due to the low activity in prostate and short acquisition time). Our collaborators at the Lawson Imaging Centre have developed a prostate phantom for the imaging study. However, no prostate data (either from the phantom or a patient) have yet been acquired. Therefore, we could not evaluate the developed algorithm on clinical prostate data.

Chapter 9 : Summary

In this chapter, the findings of this dissertation are summarized. Based on the results obtained in the conducted experiments, a series of recommendations are also made and the areas for future work are discussed. The summary of this dissertation is broken into two sections concerning the geometry optimization project and the developed GPU-based image reconstruction algorithms which are discussed in Sections 9.1 and 9.2, respectively.

9.1 BrainPET Detector Geometry Optimization

The limited bore diameter of the chosen brain-dedicated MR scanner (360 mm) encourages a careful design for the detectors of the PET insert. Since parallax error potentially degrades the spatial resolution of the PET insert at large radial offsets, we chose DLO as our choice of detector design to obtain DOI information. More than 200 possible DLO detectors were considered for careful evaluations through several experiments. DLO detectors with total thicknesses of 10, 15, 20, 25 and 30 mm with 9 possible layer split ratios (in increments of 10% of the total thickness) were studied. Also, for each detector, nominal crystal pitches of 0.5, 1, 2, 3 and 4 mm were considered. All experiments were conducted through MC simulations with GATE and were validated against measurement using the CRFs of a small-animal PET scanner available in our laboratory.

We began the geometry optimization process by assuming a predefined detector area of 28.4×28.4 mm^2. The assumed detector area was selected based on the size of a SensL ArrayC SiPM array. A geometry optimization algorithm for the detectors of the brain PET insert was developed that is described by Figure 4.6. The developed geometry optimization algorithm in Chapter 4 was dedicated to performance metrics that are largely independent of the detector

area. Namely, the event positioning accuracy, coincidence response profiles and sensitivity of each detector and its corresponding PET scanner were carefully modeled in the experiments discussed in Chapter 4.

The event localization accuracy experiment was designed to find a range of optimum layer thickness ratios for each total thickness. As a result of this experiment, we limited the range of optimum layer thicknesses to 2 or 3 layer ratios for each total thickness as presented in Table 4.3. Using the results of the event localization accuracy experiment, we limited the number of possible geometries and simulated the CRFs of all remaining detector geometries. It was seen that a crystal size of less than 2 mm often leads to a large CRF FWHM/w ratio (except for a total thickness of 10 mm). Also, it was concluded that a d_F/d_B of 4/6, 6/9, 8/12, 7.5/17.5 and 6/24 or 9/21 mm can be considered as the optimum layer thickness ratio for detectors with a total thickness of 10, 15, 20, 25 and 30 mm, respectively. Furthermore, effective resolution (ξ) as a novel concept was introduced and used in this dissertation as a novel metric to measure a radially weighted average resolution of a PET scanner as defined by equation (4.2).

The sensitivity profiles of all scanners used in the CRF experiment were also estimated through another set of MC simulations to provide an estimation of the sensitivity performance of each scanner. It was seen that detectors with a small crystal size (i.e. 0.5 and 1 mm) might lead to scanners with a lower sensitivity than that of scanners with larger crystal sizes due to their lower packing fraction and higher inter-crystal scatter rate. Also, it was shown that smaller crystal sizes can lead to a non-optimal effective resolution, especially for detectors with a thickness of larger than 10 mm. The findings of the proposed geometry optimization algorithm for DLO detectors are summarized in Table 4.4.

As mentioned earlier, the proposed detector geometry optimization algorithm in Chapter 4 did not consider the effects of detector area on the performance of the PET scanner. In Chapter 5 however, the effects of detector area on the performance of the brain-dedicated PET insert were carefully examined to complement the findings of Chapter 4. We included a wide range of possible deadtime model and constants (paralyzable and non-paralyzable models with deadtime constants of 0 to 4000 ns) and thoroughly examined the count rate performance of each scanner. Since the DLO design and crystal pitch have no impact on the count rate performance of the PET scanners, single layer detectors with a crystal pitch of 2 mm were included in this study. We included detector thicknesses of 10, 15 and 20 mm in this count rate performance evaluation study. Detector thicknesses of 25 and 30 mm were not included as the results of Chapter 4 showed a relatively large effective resolution ($\xi > 4$ mm) for such detectors. It should also be noted that even though a thicker detector can improve the sensitivity of the scanner (see Figure 4.31), a thinner detector may be preferred due to its better resolution (see Figure 4.29), geometrical consideration (larger bore size for the PET insert) and cost considerations (less scintillation material – recall that scintillation material is one of the most expensive components of a PET scanner).

The count rate performance of brain PET scanners with 36 different detector areas and several deadtime characteristics were carefully examined through studying the singles rate, scanner sensitivity and the NECR of each scanner. We also evaluated the effects of CTW, LLD and outside FOV activity on the count rate performance of all considered detectors. The reported count rate study in Chapter 5 also provides a detailed model for the count rate performance of PET detectors with several geometrical and deadtime properties which can be used as a reference for other researchers working on detector development. The count rate performance data presented in this dissertation assumed a single effective deadtime model for each case with no explicit modeling

of pileup. Therefore, the count rate performance values presented should be used with cautious as pileup contribution may invalidate some of the reported performance metrics at high activities.

It should be clarified that the final selection criteria for the geometrical properties of the detectors of the brain-dedicated PET insert (the *optimum* geometry) depend on the acceptable trade-off between scanner sensitivity, spatial resolution and cost considerations. Therefore, the performance of various brain-dedicated scanners was modelled in this dissertation from different aspects without an explicit recommendation of a detector geometry as the optimum geometry. In other words, this dissertation can be served as a thorough optimization and modeling study that presents a comprehensive model for the performance of a brain-dedicated PET insert with different detector geometries. Given the findings of the geometry optimization investigations in this work, we propose a SER-DLO detector with a scintillation material of LYSO, total/front/back layer thickness of 15/6/9 to 20/8/12 mm and crystal width of 2 to 3 mm. In the presence of a realistic effective deadtime of about 750 to 1000 ns for our brain PET scanner, we also recommend a detector area of 1.5 cm^2 to 9.6 cm^2. Note that the presented geometry optimization work in Chapter 4 assumed a fixed detector area of 8.07 cm^2 (2.84 × 2.84 cm^2). Nonetheless, the experiments and findings of Chapter 4 are mainly independent of the detector area.

The limitations that existed in this geometry optimization project have been discussed in Sections 4.9 and 5.6. In particular, a thorough imaging study is recommended to compare the image SNR for scanners of various DLO detector geometries. Also, an imaging study can examine the effects of LLD on image quality, which is particularly interesting as the count rate study in Chapter 5 showed that a lower LLD might increase the scanner NECR. Furthermore, a study regarding the potential effects of pileup on the found performance metrics at high count rates is recommended.

While the presented model for the count rate performance of PET detectors in Chapter 5 includes several detector performance metrics, it is also recommended to incorporate timing blurring in a separate set of the simulations in a future work. In particular, the effects of CTW on the scanner NECR could be of interest when a non-ideal timing resolution is considered for the detectors.

As DLO was our choice of detector design to acquire DOI information, a thorough study on the behavior of DLO detectors was performed to quantify the frequency and impact of layer misassignments in DLO detectors. This investigation, described in Chapter 6, was done through an introduction and evaluation of the DLR-DLO design. An in-depth analysis of the mispositioned events followed by a CRF study, comparing the CRFs of SER- and DLR-DLO detectors, showed that the layer misassignment in SER-DLO detectors primarily increases the FWTM of the CRFs (Figure 6.7). The imaging study showed that no significant improvements in the image quality could be achieved when SER-DLO detectors were replaced by DLR-DLO detectors. Therefore, a SER scheme was selected for the DLO detectors of the brain PET scanner.

Although the focus of the geometry optimization scheme in this dissertation was on detector development for a specific brain-dedicated PET insert, many of the reported results are readily applicable to detectors being designed for other scanners. For instance, the geometry optimization algorithm explained in Figure 4.6 may be used in other multi-layer detector development studies for various PET scanners. Also, the optimum layer values obtained from the event localization accuracy experiment (Section 4.8.1) are valid for any dual-layer LYSO detector regardless of the scanner's bore diameter.

A number of experiments were also conducted as a part of the geometry optimization project that were unsuccessful or remained incomplete and therefore were not discussed in this

dissertation. In the hope that mentioning these attempts can guide other researchers, they will be briefly mentioned here:

1. In some initial studies during the geometry optimization project, images of several DLO brain PET scanners were analytically reconstructed for a point source phantom with no background activity. The methodology and results of this imaging study were not included in this thesis as our final geometry optimization algorithm (Figure 4.6) did not incorporate an imaging study. In the developed image reconstruction algorithm, the 3D data were rebinned into a set of 2D sinograms which were then reconstructed using 2D FBP. The chosen method of rebinning was *multi-slice rebinning* (MSRB) as described by [233]. As reported in the original publication describing this method, images reconstructed with MSRB require an axial deblurring step post-reconstruction. It was found that parameterizing the axial deblurring step, such that it remains compatible for all single layer and DLO detectors, is a complex and to a certain degree cumbersome problem. This project was however completed and a scalable sinogram generation algorithm was developed that is compatible with all brain PET scanners. Nonetheless, the axial blurring step was found to be unreliable and require manual modifications based on the scanner geometry.
2. In the execution of the count rate study as explained in Chapter 5, a caudal lead shield covering one side of the scanner was also included in a separate chain of simulations. The aim of this experiment was to evaluate whether such shielding can reduce the effects of outside FOV activity on the NECR. The shield was designed as an annular lead shield with an inner diameter of 320 mm and a thickness of either 10 or 20 mm (in separate experiments) such that it covers one side of all detectors in the caudal end of the scanner. Through analyzing the results of the simulations, it became obvious that such shielding cannot

meaningfully improve (or really change) the NECR due to two reasons: A) although the shielding protected the first ring of detector blocks against outside FOV activity, the inner diameter of the annular shield was smaller than what was required to protect the other detectors in other detector rings, and B) due to the high statistical uncertainty in the NECR data (see Figure 5.21), it was difficult to perform any meaningful statistical comparison between NECR plots that were very close to each other. As briefly mentioned in Section 5.6, it is recommended to repeat these simulations with a more carefully designed shield.

9.2 Dual-GPU Image Reconstruction Algorithm

The developed dual-GPU list-mode image reconstruction algorithm was based on a previously developed algorithm originally described by [193, 195]. The developed reconstruction algorithm explained in Chapter 7 was developed as part of the geometry optimization project for the imaging study explain in Section 6.10. We incorporated a shift-invariant Gaussian-like PSF in the forward and backprojection kernels of the developed reconstruction algorithms for brain PET scanners and NuPET. An overview of the dual-GPU implementation and evaluation of the developed reconstruction for brain-dedicated PET scanners and NuPET were discussed in Chapter 6 and Chapter 7, respectively. The efficiency of the developed reconstruction algorithms as well as sample reconstructed images were also presented.

Since the brain PET image reconstruction algorithm was primarily designed for high resolution reconstructions, it was developed such that the second GPU is only involved if the image can be split into more than 1 cube. The developed image reconstruction algorithm for NuPET however, uses both GPUs regardless of the number of cubes in the image. It was shown that a reconstruction time of fewer than 18 seconds per iterations is achievable and sample reconstructed images from GATE and real scanners were shown as well.

A component-based normalization method was also developed for NuPET such that a normalization coefficient is calculated for each LOR by processing the data obtained from scanning an annular normalization phantom. We did not incorporate a normalization algorithm for brain PET scanners due to the large number of candidate detector geometries. Instead, a uniform detection probability was assumed for all recorded events as the calculation of geometrical probabilities for each scanner requires a very long acquisition time. As a result, caution should be used when working with the developed image reconstruction algorithm for brain PET scanners as it may lead to large normalization artifacts, especially if there is a substantial difference between the coincidence detection probabilities in the front and back layer crystals of the scanner.

Although the images reconstructed from simulating the NuPET scanner showed no clear artifacts, the images reconstructed from acquired data showed non-uniformity artifacts as discussed in Section 8.10. While possible reasons for these artifacts were discussed in Chapter 8, a more thorough investigation of the origin of the artifacts is recommended. It is also highly recommended to group the LORs with the same number of *involved slices* when invoking the forward and backprojection kernels to further optimize the developed dual-GPU image reconstruction algorithm (see Section 7.8). Such optimizations could improve the efficiency of the developed reconstruction algorithm significantly. Also, similar to the so-called *one-pass list-mode ML-EM* [153], the developed dual-GPU reconstruction package can be modified such that it allows for a virtually online reconstruction algorithm by reconstructing subsets of data as the data are being fed into the reconstruction engine.

In the end, we share some of the many attempts that either failed or remained incomplete in the image reconstruction project and make some recommendations for future work:

1. An analytical method for an online calculation of the CRFs in a PET scanner was suggested in an excellent paper by Pratx *et al*. [190]. During the development of the image reconstruction algorithm for brain PET scanners, we extended the proposed method from [190] to DLO detectors and obtained initial results. The method and results were not discussed in this dissertation as a shift-invariant Gaussian PSF was selected for the final implementation. Though we never fully evaluated and confirmed the results of our DLO extension of [190], the preliminary results showed some promising agreements between the analytically calculated CRFs and the CRFs obtained from MC simulations (Section 4.8.2). The developed method based on [190], is highly compatible with GPUs and may calculate the CRF for any pair of crystals on-the-fly using a GPU. In theory, incorporating an accurate and fast analytical approach to find CRFs eliminates the need for a pre-calculation of geometrical components and allows for modeling of a shift-varying PSF into the system matrix. Integrating a method similar to [190] for the calculation of a shift-variant PSF into the developed GPU-based image reconstruction algorithms is highly recommended.
2. Other than the proposed method by Pratx *et al*. [190], other analytical methods to find an accurate model for the PSF were also attempted. In particular, a method developed by Zhang *et al*. [229] was evaluated and thoroughly tested for various single layer and DLO brain PET scanners. The proposed method in [229] attempts to find the system matrix elements analytically by generating a large number of uniformly sampled LORs from each image voxel and tracking the photon paths within crystal elements. The proposed method in [229] uses *Siddon's ray tracing* [234] to find the path length of each coincidence in the 'target crystal elements' and other crystal elements ('attenuation crystal elements') and finds a geometrical probability for each coincidence based on the measured path lengths. This

method was also extended for DLO detectors and implemented for various single layer and DLO brain PET scanners. The developed method was thoroughly evaluated but the results were not discussed in this dissertation. After careful evaluation of the developed method and through personal correspondence with the authors of [229], we learned that the analytically calculated response functions using this method cannot produce an accurate model for the PSF of a DLO scanner. The main reason that this method lacked accuracy was its oversimplification of the PSF model and the fact that the proposed algorithm in [229] ignores several physical phenomena occurring in a detector such as inter-crystal scattering. The second reason that this method was not further pursued was that this method (and generally methods involving Siddon's ray tracing) was not an ideal method for a GPU implementation.

3. The current version of the developed dual-GPU reconstruction algorithms (for brain PET scanners and NuPET) extracts 3D segments of the image (cube) and processes each cube using a tiling technique. The extraction and stitching cubes were introduced purely for simplicity and easier programming. In fact, tiles could be extracted directly from and written back directly to the image which may increase the efficiency of the developed image reconstruction algorithm even further. Therefore, it is recommended to reformulate the developed reconstruction algorithms such that the tiles are directly loaded into the shared memory of each thread block from the reconstructed image.
4. Incorporation of other prior terms (other than the already implemented quadratic prior term) is also recommended. In particular, an incorporation of an anatomically guided prior term (such as [184, 235]) could be of interest. Nonetheless, the limited memory and the low

number of available registers per thread may make the implementation a non-trivial challenge.

In conclusion, a quantitative model for the performance metrics of brain-dedicated PET inserts with different detector geometries was provided in this work. The performance of various DLO detectors was evaluated with a focus on mitigating parallax error, improving spatial resolution and improving the count rate performance of the PET insert. A scalable GPU-based image reconstruction algorithm was also developed that was used in the geometry optimization project. Also, a dual-GPU list-mode image reconstruction algorithm was developed for NuPET to facilitate fast reconstruction of indexed histogram data during radical prostatectomy.

Appendix A : User Guide for BrainPET Image Reconstruction Algorithm (DGPUBrainPET)

The source codes of the developed package as well as sample data files are under the name 'DGPUBrainPET'. Note that the algorithm implemented in DGPUBrainPET strictly requires two GPUs and executing this program on one GPU cannot be done without major modifications of the source codes. Nonetheless, a single-GPU version of DGPUBrainPET has also been developed (under the name '*SGPUBrainPET*').

As mentioned in Chapter 7, DGPUBrainPET was developed to be scalable such that it can easily reconstruct images for any single layer or DLO BrainPET scanner. The computer on which the DGPUBrainPET has been developed has had 3 mounted graphics cards. While the two Nvidia GeForce GTX Titan Black graphics cards were used for CUDA programming, the 3rd one is used for visualization. Therefore, the DGPUBrainPET package is hard coded to verify the availability of two GTX Titan Black graphics cards in every execution to ensure that the kernels are executed on the target graphic cards. If the Titan Back graphics cards are replaced by another pair of graphics cards supporting CUDA, function `check_cuda_devices()` (in file `device_src.cu`) should be modified followed by a recompilation.

The DGPUBrainPET program is designed to be executed on a 64-bit UNIX environment and as of April 2019, a *Makefil*e associated with the program is available to compile the program. No environment variables are available and therefore, to compile the program, the user can simply run the command `make`. If the compilation of all components (C++ and CUDA scripts) are successful, an executable file called `<generic_em_multi_gpu>` is generated. Optionally, one can add the executable file to the environment by doing

`export PATH=$PATH:/address/to/parent/folder` (bash)

assuming that `<generic_em_multi_gpu>` is located at

`/address/to/parent/folder`.

The source code files of the DGPUBrainPET package are briefly described in Table A.1.

Table A.1. A list of source files in the DGPUBrainPET package.

File name	Description
`Makefile`	Compile all source/header files
`image_recon_em.h`	C++ header file
`image_recon_em_gpu.h`	CUDA header file
`image_recon_em.cu`	The `main()` function
`host_src.cpp`	Host source codes, except normalization-related functions
`host_normalization.cpp`	Host source codes for normalization-related functions
`device_src.cu`	Parent functions calling CUDA kernels
`attenuation_coef_src.cu`	CUDA kernels related to attenuation correction calculation
`forward_proj_src.cu`	CUDA kernel for forward projection operator
`backward_proj_src.cu`	CUDA kernel for backprojection operator
`conv_map_src.cu`	CUDA kernels related to MAP reconstruction
`misc_both_gpu_src.cu`	Other CUDA kernels

Executing the program can be done through

```
./generic_em_multi_gpu  INPIUT_ARGUMENTS
```

When the program is executed with no input arguments, a detailed guide on how to use the program will be shown. Alternatively, the guide can be seen by executing

```
./generic_em_multi_gpu  --help
```

Before continuing to explain further usage details of the developed program, two important data classes must be defined as

`CRYSLUT_TYPE`: 4-byte wide single-precision float
`CRYSNUM_TYPE`: 4-byte wide unsigned integer

BrainPET scanners, as discussed in Section 4.3, have 38 detector blocks per ring (n_{DPR}) with 6 rings of detectors (n_{RD}) as shown in Figure A.1.

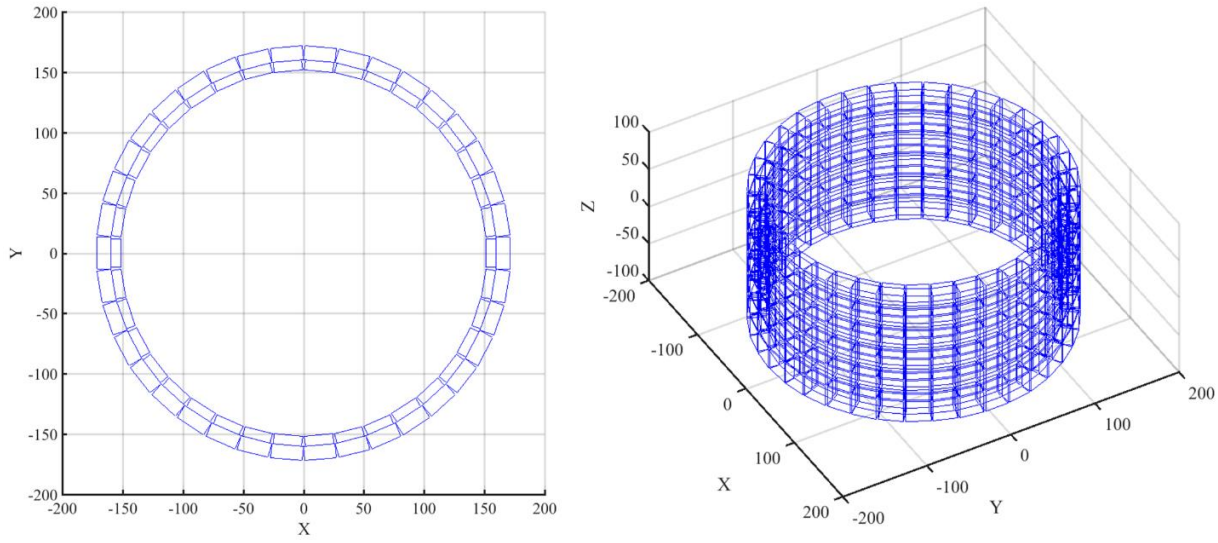

Figure A.1. 2D (left) and 3D (right) illustrations of an example BrainPET scanner with coordinate systems.

The DGPUBrainPET program reads data files (e.g. crystal LUT files or coincidence files) directly through an address provided by the user and not through a header file. For each scanner, a crystal LUT is required. A binary crystal LUT file should be created such that it includes the position of the centroid of each crystal element i [C_{iX} C_{iY} C_{iZ}] (given the coordinate system as shown in Figure A.1) with type `CRYSLUT_TYPE` in a format similar to what is shown in Figure A.2 (note that little-endian format is assumed). Note that the relationship between crystal numbers and their centroid Cartesian coordinates is arbitrary, meaning that users can define their own method of crystal numbering. It is, however, critical that a consistent numbering approach is used in both the crystal LUT and the given list of events. In this thesis, the approach used for crystal numbering was similar to that of CASToR [176] as explained below.

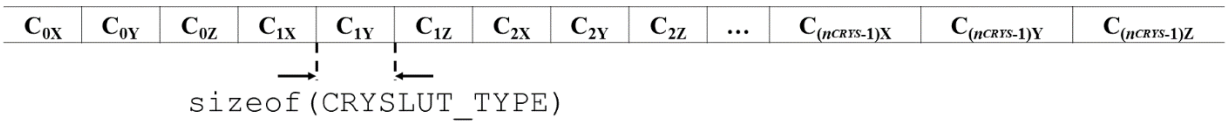

Figure A.2. A block of memory showing the format of the crystal LUT for a BrainPET scanner with n_{CRYS} crystal elements.

Crystal elements are numbered between 0 to $n_{CRYS} - 1$ where n_{CRYS} is the number of crystals in the PET scanner which is defined by equation (7.2) for DLO BrainPET scanners. For any single layer BrainPET scanner with detector blocks of $n \times n$ crystals, n_{CRYS} is simply given by

$$n_{CRYS} = n_{RD} \times n_{DPR} \times n \times n \ .$$

In order to number the crystals of a BrainPET scanner, each crystal is identified based on the following four geometrical properties: `blockID`, `crystalID`, `ringID` and `layerID` that are defined by Figure A.3.

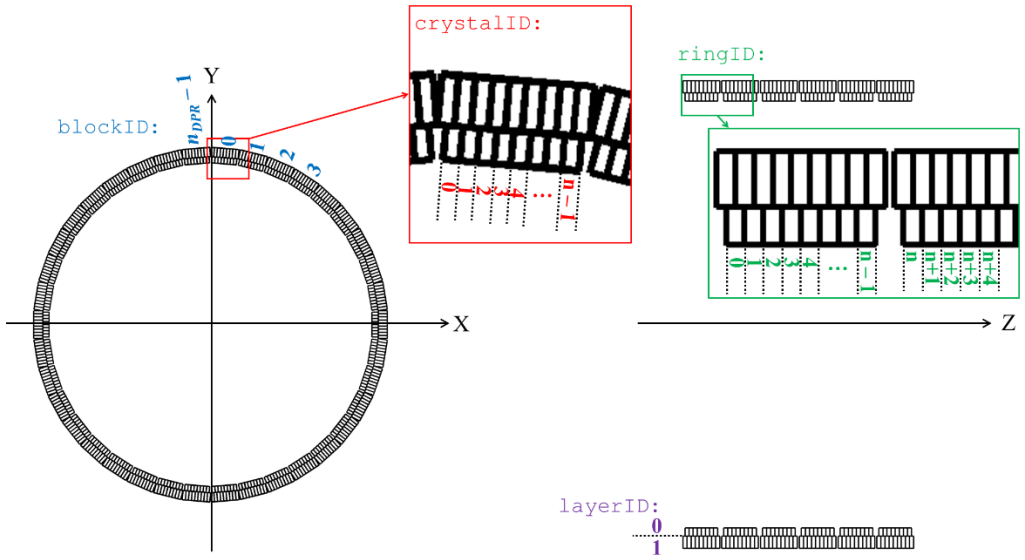

Figure A.3. Schematic illustration of how `blockID`, `crystalID`, `ringID` and `layerID` are defined.

The relationship between the crystal number (`crysNo`) of an arbitrary crystal in the scanner and its `blockID`, `crystalID`, `ringID` and `layerID` is given by

$$\texttt{crysNo} = \texttt{crystalID} + \texttt{blockID} \times m + \texttt{ringID} \times n_{DPR} \times m + \texttt{layerID} \times n_{RD} \times n_{DPR} \times n \times n$$

in which m is either n if `layerID` is 0 or $n + 1$ if `layerID` is 1.

The data generated from GATE simulations should be recorded in a binary file where each coincidence is indicated with two integers with class `CRYSNUM_TYPE`. The binary file including the list of coincidences is structured similar to what is shown in Figure A.4.

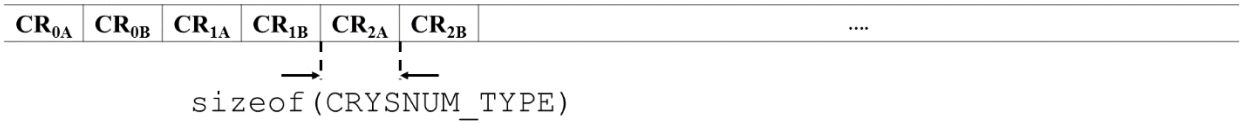

Figure A.4. A block of memory including the list of recorded coincidences where CR_{iA} and CR_{iB} represent the crystal numbers of the i^{th} recorded coincidence.

To execute the reconstruction program, it should be called with some required and optional arguments. For instance,

```
./generic_em_multi_gpu --no 9 --dy 20 --dooverdy 0.6
--nvoxels 110 110 110 --dimvoxels 1.8 1.8 2.5 --addtocryslut
/address/to/crysLUT.lut --addtooutputfolder
/address/to/a/folder --outputfilename fileName --nitr 10
--nsub 5 --addtocoinfile /address/to/coincidenceFile.bin
```

will execute a reconstruction for a DLO scanner with detectors of total thickness 20 mm, back-to-total layer ratio of 0.6 (i.e. back/front layer = 12/8 mm) and detector blocks of (9 × 9) + (8 × 8) crystals (i.e. argument after option `--no` is in fact $n + 1$). The crystal LUT for this scanner is given at `/address/to/crysLUT.lut` and the binary coincidence file is located at `/address/to/coincidenceFile.bin`. The reconstruction will be performed in a grid of

110 × 110 × 110 voxels with a voxel size of 1.8 × 1.8 × 2.5 mm^3. The reconstruction will use 10 iterations of list-mode ML-EM with 5 subsets and the reconstructed images will be saved as a binary file with name `fileName_i.bin` (for iteration `i`) at address `/address/to/a/folder`.

Note that the crystal pitch of the scanner is not an input. Instead, based on the given number of crystals per detector, the program finds the crystal pitch of the scanner such that the adjacent detector blocks in the transverse plane just touch (see Figure A.1 for example).

A list of complete command line options are as follows:

`--nO` [required]: Should be followed by an integer that indicates the number of crystals in the back layer along one direction. Each detector block has `(no×no) + (no-1)×(no-1)` crystal elements if the scanner has DLO detectors. If the scanner is a single layer scanner, each detector block has `(no-1) × (no-1)` crystal elements (i.e. back layer does not exist). The relationship between `nO` and the previously defined *n* in this dissertation is simply $n = \mathtt{no} - 1$,

`--dy` [required]: The trailing number indicates the total detector thickness in each detector block in mm,

`--dooverdy` [required]: Should be followed by a number indicating the ratio of the back layer thickness to the total detector thickness. This number can be in the range of [0 1) such that 0 indicates a single layer scanner,

`--nvoxels` [required]: Should be followed by number of voxels in the image along the [X Y Z] axes in this format `'`$\mathtt{N_x\ N_y\ N_z}$`'` (without quotation marks). Note that $\mathtt{N_x}$, $\mathtt{N_y}$ and $\mathtt{N_z}$ must be a multiple of the tile dimension. The tile dimension along each axis is hard coded as 110.

`--dimvoxels` [required]: Should be followed by voxel dimension along the [X Y Z] axes in mm in this format 'd_x d_y d_z' (without quotation marks),

`--addtocryslut` [required]: Indicates the address to the binary crystal LUT file as explained above,

`--addtooutputfolder` [required]: Indicates the address to a folder where the reconstructed images (and their headers) will be saved,

`--outputfilename` [required]: Specifies the body file name for the output image file. Note that for each iteration *i*, a binary image and a text header file `fileName_i.bin` and `fileName_i.hdr` will be stored,

`--nitr` [required if `--sens_gen_mode` is *not* enabled]: Number of iterations to be used for the reconstruction,

`--nsub` [required if `--sens_gen_mode` is *not* enabled]: Number of subsets to be used for the reconstruction,

`--addtocoinfile` [required if `--sens_gen_mode` is *not* enabled]: Address to a binary coincidence file including the list of recorded events,

`--set_w_to_one` [optional]: If this option is on, a TOR with a width of *one pixel* only is considered. In other words, instead of a TOR for each coincidence, an LOR with a width of 1 pixel is assumed,

`--sens_gen_mode` [optional]: If this option is on, the program will generate a sensitivity image for the target BrainPET scanner through backprojecting 50×10^6 random events (see option `--nsamples` too),

`--nsamples` [optional]: By default, the program backprojects 50×10^6 random events to generate a sensitivity image. Using `--nsamples N` instead backprojects `N` randomly generated events (large numbers are highly recommended),

`--qp` [optional]: Enables MAP reconstruction using a quadratic prior,

`--beta` [optional]: If MAP reconstruction is requested, a default β value of 0.0015 will be used. The argument after this option allows for an introduction of a different β value,

`--addtosensim` [optional]: This option should be followed with an address to a binary file (`single-precision float`) of the sensitivity image. The size of the sensitivity image should be the same as the reconstructed image. If no sensitivity image is provided the program assumes a uniform sensitivity image,

`--addtoattnim` [optional]: This option should be followed with an address to a binary attenuation file (`single-precision float`). The size of the attenuation map should be the same as the reconstructed image. If no attenuation image is provided, the program assumes a zero attenuation map. The unit of the given attenuation map should be 1/mm,

`--gshift` [optional]: Applies a global shift to centre of the reconstructed image. The coordinate shift is expressed as 'S_x S_y S_z' (without quotation marks) in mm,

`--nstreams` [optional]: The argument after this option determines of the number of CUDA streams. The number can be in the range of [1 32] (default: 8). Changing this value affects the efficiency of the forward projection kernel,

`--crffwhmtodx` [optional]: The argument after this option determines the ratio between CRF FWHM/*w* (*w*: crystal pitch). For instance, if a scanner has a crystal pitch of 2 mm, `--crffwhmtodx 0.5` enforces a Gaussian shift-invariant PSF with a FWHM of 0.5 × 2 mm = 1 mm in the forward and backprojection kernels. The default number is 1,

`--maskout` [optional]: The argument after this option determines the ratio by which all transaxial slices of the image will be truncated. By default, no truncation will be done on the images. The given number should be in the range of [0 1],

`--nsliceout` [optional]: The argument after this option determines the number of slices on each end of the image to be excluded from the image (default: 0),

`--verbose` [optional]: Verbosity level which can be 0, 1 or 2 (default: 0),

`--help` [optional]: Shows the help information and exit.

Appendix B : User Guide for NuPET Image Reconstruction Algorithm (DGPUNuPET)

The source codes of the developed package as well as sample data files are under the name 'DGPUNuPET'. Note that the algorithm implemented in DGPUNuPET strictly requires two GPUs and executing this program on one GPU cannot be done without major modifications of the source codes. No single-GPU implementation of DGPUNuPET has been developed in this work.

The DGPUNuPET program, as described in Chapter 8, is designed to be specifically compatible with the NuPET indexed histogram data. Due to the reasons discussed in Appendix A, the program verifies the presence of two Nvidia GeForce GTX Titan Black graphic cards during each execution. Similar to the DGPUBrainPET package, if the Titan Back graphic cards are replaced by another pair of graphic cards supporting CUDA, function `check_cuda_devices()` (in file `device_src.cu`) should be modified followed by a recompilation.

The technical explanation here in this appendix applies to the developed DGPUNuPET package as of April 2019. A Makefile is available to handle the code compilation process on a 64-bit UNIX environment. The command `make` compiles the program and generates an executable file called `<nupet_gpu_recon>`. The following command adds the generated executable file to the environment (optional)

`export PATH=$PATH:/address/to/parent/folder` (bash)

assuming that `<nupet_gpu_recon>` is located at

`/address/to/parent/folder.`

The source code files of the DGPUNuPET package are briefly described in Table B.1.

Table B.1. A list of source files in the DGPUNuPET package.

File name	Description
`Makefile`	Compile all source/header files
`image_recon_em.h`	C++ header file
`image_recon_em_gpu.h`	CUDA header file
`image_recon_em.cu`	The `main()` function
`host_src.cpp`	Host source codes, except normalization-related functions
`device_src.cu`	Parent functions calling CUDA kernels
`attenuation_coef_src.cu`	CUDA kernels related to attenuation correction calculation
`forward_proj_src.cu`	CUDA kernel for forward projection operator
`backward_proj_src.cu`	CUDA kernel for backprojection operator
`map_gpu_src.cu`	CUDA kernels related to MAP reconstruction
`misc_both_gpu_src.cu`	Other CUDA kernels

Executing the program can be done through

```
./nupet_gpu_recon  INPIUT_ARGUMENTS
```

To see a list of the required and optional input arguments one can simply run

`./nupet_gpu_recon  --help [-h]` or

```
./nupet_gpu_recon
```

The indexed histogram data are an array of 81,719,040 elements (recall that the NuPET scanner has 81,718,784 LORs assuming extra arbitrary crystals shown in Figure 8.1). The first 256 elements are reserved for header bytes and the rest include indexed histogram information. Each element is a 4-byte single-precision float number. The histogram data are arranged such that coincidences from different crystal layer combinations (e.g. front-front, back-front, etc.) are sorted in a row. Following the 1024 bytes (256 × 4 bytes) reserved for header information, a total of

18,322,304 elements are reserved for LORs connecting front-front crystals, 22,651,776 elements for back-back LORs, 20,372,352 for front-back LORs and finally another 20,372,352 for back-front LORs. Therefore, all indexed histogram data files have 81,719,040 elements with a size of 326.9 MB. Note that the mentioned file size is independent of the number of recorded events. For each PET reconstruction, two indexed histogram files are required representing [an estimation of] the number of true and random events per LOR in separate files.

Since the code was developed for a scanner with known geometrical properties, no crystal LUT is required as an input. To execute the reconstruction program, it should be called with some required and optional arguments similar to the developed BrainPET reconstruction algorithm. Command line options have a long '`--`' and a short '`-`' option format as described in the following:

`--zoom [-z]` [required]: Indicates the zoom number such that `-z 1` reconstructs an image with 105 × 105 × 105 voxels and a uniform voxel size of 0.64 mm and `-z 2` reconstructs an image with 120 × 210 × 210 voxels and a uniform voxel size of 0.32 mm,

`--addtooutputfolder [-a]` [required]: Indicates the address to a folder where the reconstructed images (and their headers) will be saved,

`--outputfilename [-o]` [required]: Specifies the body file name for the output image file. Note that for each iteration `i`, a binary image and a text header file `fileName_i.bin` and `fileName_i.hdr` will be stored,

`--nitr [-i]` [required if `--sens_gen_mode` or `--fwd_proj` are *not* enabled]: Number of iterations to be used for the reconstruction,

`--nsub` [required if `--sens_gen_mode` or `--fwd_proj` are *not* enabled]: Number of subsets to be used for the reconstruction,

`--addtotruesfile [-t]` [required if `--sens_gen_mode` or `--fwd_proj` are *not* enabled]: Address to an indexed histogram file containing the list of recorded true coincidences,

`--addtorandfile [-r]` [required if `--sens_gen_mode` or `--fwd_proj` are *not* enabled]: Address to an indexed histogram file containing the list of recorded random coincidences,

`--set_w_to_one [-w]` [optional]: If this option is on, a TOR with a width of *one pixel* only is considered. In other words, instead of a TOR for each coincidence, an LOR with a width of 1 pixel is assumed,

`--sens_gen_mode [-sg]` [optional]: If this option is on, the program will generate a sensitivity image through a complete backprojection of all LORs,

`--fwd_proj [-fp]` [optional]: If this option is on, the program will forward project a uniform or a user-input image (see `--addtoinitialim`) and write the forward projected values in an indexed histogram file to disk (no image will be generated),

`--addtoinitialim [-ai]` [optional]: Points to an address to a binary file of an image to be used as an initial image. The image must have the same size as the reconstructed image and all pixels should have a positive value (only in `--fwd_proj` mode, pixel values of 0 are allowed for an initial image),

`--addtogeomprobfile [-ag]` [optional]: The following argument should point to an indexed histogram binary file of the geometrical coefficients,

`--addtonormfile [-an]` [optional]: The following argument should point to an indexed histogram binary file of the normalization coefficients. The normalization coefficients can be calculated using the NormNuPET package (see Appendix C),

`--addtosensim [-as]` [optional]: Points to an address to a binary file of the sensitivity image. The size of the sensitivity image should be the same as the reconstructed image. If no sensitivity image is provided the program assumes a uniform sensitivity image,

`--addtoattnim [-at]` [optional]: Points to a binary attenuation image. The size of the attenuation map should be the same as the reconstructed image. If no attenuation image is provided, the program assumes a zero attenuation map. The unit of the given attenuation map should be 1/mm,

`--qp [-qp]` [optional]: Enables MAP reconstruction using a quadratic prior,

`--beta [-b]` [optional]: If MAP reconstruction is requested, a default β value of 0.05 will be used. The argument after this option allows for an introduction of a different β value,

`--nstreams [-n]` [optional]: The argument after this option determines the number of CUDA streams. The number can be in the range of [1 32] (default: 8). Changing this value affects the efficiency of the forward projection kernel,

`--crffwhmtodx [-c]` [optional]: The argument after this option determines the ratio of CRF FWHM/w (w: 1.28 mm). For instance, `--crffwhmtodx 0.5` enforces a Gaussian shift-

invariant PSF with a FWHM of 0.5 × 1.28 mm = 0.64 mm in the forward and backprojection kernels. The default number is 1,

`--maskout [-m]` [optional]: The argument after this option determines the ratio by which all transaxial slices of the image will be truncated. By default, no truncation will be done on the images. The given number should be in the range of [0 1],

`--nsliceout [-ns]` [optional]: The argument after this option determines the number of slices on each end of the image to be excluded from the image (default: 0),

`--verbose [-v]` [optional]: Verbosity level which can be 0, 1 or 2 (default: 0),

`--help [-h]` [optional]: Shows the help information and exit.

Appendix C : User Guide for NuPET Component Normalization Algorithm (NormNuPET)

The developed package for component normalization for the NuPET scanner as well as sample data files are under the name 'NormNuPET'. The NormNuPET is a C++ program that runs on a CPU with no GPU involvements. The developed algorithm is in a single `cpp` file named `main.cpp`.

The component-based normalization algorithm for NuPET was implemented in a C++ program as explained in Section 8.4. The program can calculate the normalization coefficients for all indexed histogram bins (LORs) in less than 3 seconds. In the developed algorithm, *T* in equation (8.12) was hard coded to be 5. This program runs on a single CPU core in a UNIX environment. Due to the absence of a Makefile for this program, compilation should be manually done manually, such as

```
g++ -O3 main.cpp -o <ProgramFileName>
```

which generates a binary executable file called `ProgramFileName`. To execute the program, the program requires the following arguments:

`-f`: Address to the forward projected values of a uniform image in an indexed histogram format. This file can be generated using option `--fwd_proj [-fp]` of `<nupet_gpu_recon>`,

`-t`: Address to the indexed histogram file of the true events obtained from the scan of a normalization phantom,

`-r`: Address to the indexed histogram file of the random events obtained from the scan of a normalization phantom,

`-a` : Address to a folder where the generated normalization coefficient file will be written to,

`-o` : Filename of the generated normalization file without extension. e.g. `-o someFileName` generates an indexed histogram file `someFileName.ih`.

References

1. Krane, K.S., *Modern Physics*. 2nd Edition ed. Vol. Wiley-VCH. 1995. 608.
2. Taylor, J.R., M.A. Dubson, and C.D. Zafiratos, *Modern physics for scientists and engineers*. 2004: Prentice-Hall.
3. Tipler, P.A. and R. Llewellyn, *Modern physics*. 2008: Macmillan.
4. Cherry, S.R., J.A. Sorenson, and M.E. Phelps, *Physics in nuclear medicine e-Book*. 2012: Elsevier Health Sciences.
5. Abe, K., et al., *Upper bound on neutrino mass based on T2K neutrino timing measurements.* Physical Review D, 2016. **93**(1).
6. Magill, J., et al., *Karlsruhe nuclide chart*. 2015, Technical report, Nucleonica GmbH.
7. Kamal, A., *Nuclear physics*. 2014: Springer.
8. Knoll, G.F., *Radiation detection and measurement*. 2010: John Wiley & Sons.
9. Berger, M. and S. Seltzer, *XCOM Photon Cross Sections.* Version 3.1, NISTIR, 1999.
10. Wagner, H.N., *A brief history of positron emission tomography (PET).* Seminars in Nuclear Medicine, 1998. **28**(3): p. 213-220.
11. Rich, D.A., *A brief history of positron emission tomography.* Journal of nuclear medicine technology, 1997. **25**(1): p. 4-11.
12. Graham, L.S., et al., *Nuclear medicine from Becquerel to the present.* Radiographics, 1989. **9**(6): p. 1189-202.
13. Carlson, S., *A Glance At The History Of Nuclear Medicine.* Acta Oncologica, 1995. **34**(8): p. 1095-1102.
14. Brownell, G.L., *A history of positron imaging.* Physics Research Laboratory, Massachusetts General Hospital, MIT, 1999: p. 1.
15. Berg, E. and S.R. Cherry, *Innovations in Instrumentation for Positron Emission Tomography.* Semin Nucl Med, 2018. **48**(4): p. 311-331.
16. Watanabe, M., et al., *A new high-resolution PET scanner dedicated to brain research.* IEEE Transactions on Nuclear Science, 2002. **49**(3): p. 634-639.
17. Kolb, A., et al., *Technical performance evaluation of a human brain PET/MRI system.* Eur Radiol, 2012. **22**(8): p. 1776-88.
18. Jung, J.H., et al., *Development of PET/MRI with insertable PET for simultaneous PET and MR imaging of human brain.* Med Phys, 2015. **42**(5): p. 2354-63.
19. Jung, J., et al., *Development of PET insert for simultaneous PET/MR imaging of human brain.* EJNMMI Phys, 2014. **1**(Suppl 1): p. A8.
20. de Jong, H.W., et al., *Performance evaluation of the ECAT HRRT: an LSO-LYSO double layer high resolution, high sensitivity scanner.* Phys Med Biol, 2007. **52**(5): p. 1505-26.
21. Ahmed, A.M., et al., *Simulation study comparing the helmet-chin PET with a cylindrical PET of the same number of detectors.* Phys Med Biol, 2017. **62**(11): p. 4541-4550.
22. Vaidya, M., et al., *Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer.* Radiother Oncol, 2012. **102**(2): p. 239-45.
23. Szanda, I., et al., *National Electrical Manufacturers Association NU-4 performance evaluation of the PET component of the NanoPET/CT preclinical PET/CT scanner.* J Nucl Med, 2011. **52**(11): p. 1741-7.
24. Jakoby, B.W., et al., *Physical and clinical performance of the mCT time-of-flight PET/CT scanner.* Phys Med Biol, 2011. **56**(8): p. 2375-89.

25. Delso, G., et al., *Performance measurements of the Siemens mMR integrated whole-body PET/MR scanner.* J Nucl Med, 2011. **52**(12): p. 1914-22.
26. De Ponti, E., et al., *Performance measurements for the PET/CT Discovery-600 using NEMA NU 2-2007 standards.* Med Phys, 2011. **38**(2): p. 968-74.
27. Bettinardi, V., et al., *Physical performance of the new hybrid PETCT Discovery-690.* Med Phys, 2011. **38**(10): p. 5394-411.
28. *Philips healthcare*. Available from: https://www.philips.ca/healthcare/product/HC882456/ingenuity-tf-petct-system/overview.
29. Chun, S.Y., et al., *MRI-based nonrigid motion correction in simultaneous PET/MRI.* J Nucl Med, 2012. **53**(8): p. 1284-91.
30. Ouyang, J., Q. Li, and G. El Fakhri, *Magnetic resonance-based motion correction for positron emission tomography imaging.* Semin Nucl Med, 2013. **43**(1): p. 60-7.
31. Petibon, Y., et al., *Cardiac motion compensation and resolution modeling in simultaneous PET-MR: a cardiac lesion detection study.* Phys Med Biol, 2013. **58**(7): p. 2085-102.
32. Raylman, R.R., B.E. Hammer, and N.L. Christensen, *Combined MRI-PET scanner: a Monte Carlo evaluation of the improvements in PET resolution due to the effects of a static homogeneous magnetic field.* IEEE Transactions on Nuclear Science, 1996. **43**(4): p. 2406-2412.
33. Wirrwar, A., et al., *4.5 tesla magnetic field reduces range of high-energy positrons-potential implications for positron emission tomography.* IEEE Transactions on Nuclear Science, 1997. **44**(2): p. 184-189.
34. Kraus, R., G. Delso, and S.I. Ziegler, *Simulation Study of Tissue-Specific Positron Range Correction for the New Biograph mMR Whole-Body PET/MR System.* IEEE Transactions on Nuclear Science, 2012. **59**(5): p. 1900-1909.
35. Shah, N.J., et al., *Effects of magnetic fields of up to 9.4 T on resolution and contrast of PET images as measured with an MR-BrainPET.* PLoS One, 2014. **9**(4): p. e95250.
36. Shao, Y., et al., *Simultaneous PET and MR imaging.* Physics in Medicine and Biology, 1997. **42**(10): p. 1965-1970.
37. Cherry, S.R., A.Y. Louie, and R.E. Jacobs, *The Integration of Positron Emission Tomography With Magnetic Resonance Imaging.* Proceedings of the IEEE, 2008. **96**(3): p. 416-438.
38. Musafargani, S., et al., *PET/MRI: a frontier in era of complementary hybrid imaging.* Eur J Hybrid Imaging, 2018. **2**(1): p. 12.
39. Pichler, B.J., et al., *PET/MRI: paving the way for the next generation of clinical multimodality imaging applications.* J Nucl Med, 2010. **51**(3): p. 333-6.
40. Vandenberghe, S. and P.K. Marsden, *PET-MRI: a review of challenges and solutions in the development of integrated multimodality imaging.* Phys Med Biol, 2015. **60**(4): p. R115-54.
41. Werner, P., et al., *Current status and future role of brain PET/MRI in clinical and research settings.* Eur J Nucl Med Mol Imaging, 2015. **42**(3): p. 512-26.
42. Mannheim, J.G., et al., *PET/MRI Hybrid Systems.* Semin Nucl Med, 2018. **48**(4): p. 332-347.
43. Nagy, K., et al., *Performance evaluation of the small-animal nanoScan PET/MRI system.* J Nucl Med, 2013. **54**(10): p. 1825-32.

44. Zaidi, H., et al., *Design and performance evaluation of a whole-body Ingenuity TF PET-MRI system.* Phys Med Biol, 2011. **56**(10): p. 3091-106.
45. Goertzen, A.L., et al., *First Results From a High-Resolution Small Animal SiPM PET Insert for PET/MR Imaging at 7T.* IEEE Transactions on Nuclear Science, 2016. **63**(5): p. 2424-2433.
46. Heiss, W.D., *The potential of PET/MR for brain imaging.* Eur J Nucl Med Mol Imaging, 2009. **36 Suppl 1**: p. S105-12.
47. Schlemmer, H.P., et al., *Simultaneous MR/PET imaging of the human brain: feasibility study.* Radiology, 2008. **248**(3): p. 1028-35.
48. Stortz, G., et al., *Performance of a PET Insert for High-Resolution Small-Animal PET/MRI at 7 Tesla.* J Nucl Med, 2018. **59**(3): p. 536-542.
49. Thompson, C.J., et al., *Development of a PET scanner for simultaneously imaging small animals with MRI and PET.* Sensors (Basel), 2014. **14**(8): p. 14654-71.
50. Townsend, D.W., *Dual-modality imaging: combining anatomy and function.* J Nucl Med, 2008. **49**(6): p. 938-55.
51. Chang, S.-C., et al., *Diffusion-weighted MRI features of brain abscess and cystic or necrotic brain tumors.* Clinical Imaging, 2002. **26**(4): p. 227-236.
52. Felmlee, J.P. and R.L. Ehman, *Spatial presaturation: a method for suppressing flow artifacts and improving depiction of vascular anatomy in MR imaging.* Radiology, 1987. **164**(2): p. 559-64.
53. Pekar, J., et al., *In Vivo measurement of cerebral oxygen consumption and blood flow using17O magnetic resonance imaging.* Magnetic Resonance in Medicine, 1991. **21**(2): p. 313-319.
54. Lortie, M., et al., *Quantification of myocardial blood flow with 82Rb dynamic PET imaging.* Eur J Nucl Med Mol Imaging, 2007. **34**(11): p. 1765-74.
55. Mosconi, L., et al., *FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease.* Eur J Nucl Med Mol Imaging, 2009. **36**(5): p. 811-22.
56. Steeves, T.D., et al., *Increased striatal dopamine release in Parkinsonian patients with pathological gambling: a [11C] raclopride PET study.* Brain, 2009. **132**(Pt 5): p. 1376-85.
57. Moore, R.Y., et al., *Monoamine neuron innervation of the normal human brain: an 18F-DOPA PET study.* Brain Research, 2003. **982**(2): p. 137-145.
58. Cho, Z.H., et al., *In-vivo human brain molecular imaging with a brain-dedicated PET/MRI system.* MAGMA, 2013. **26**(1): p. 71-9.
59. Blum, K.I. and L.F. Abbott, *A Model of Spatial Map Formation in the Hippocampus of the Rat.* Neural Computation, 1996. **8**(1): p. 85-93.
60. Burgess, N., E.A. Maguire, and J. O'Keefe, *The Human Hippocampus and Spatial and Episodic Memory.* Neuron, 2002. **35**(4): p. 625-641.
61. Cho, Z.H., et al., *Observation of glucose metabolism in the thalamic nuclei by fusion PET/MRI.* J Nucl Med, 2011. **52**(3): p. 401-4.
62. Cho, Z.H., et al., *Substructural hippocampal glucose metabolism observed on PET/MRI.* J Nucl Med, 2010. **51**(10): p. 1545-8.
63. Son, Y.D., et al., *Glucose metabolism of the midline nuclei raphe in the brainstem observed by PET-MRI fusion imaging.* Neuroimage, 2012. **59**(2): p. 1094-7.
64. Zaidi, H. and M.-L. Montandon, *The new challenges of brain PET imaging technology.* Current Medical Imaging Reviews, 2006. **2**(1): p. 3-13.

65. Kimble, T., M. Chou, and B.H.T. Chai, *Scintillation properties of LYSO crystals*, in *2002 IEEE Nuclear Science Symposium Conference Record*. 2003. p. 1434-1437.
66. Stortz, G., *Development of a small animal MR compatible PET insert*. 2016, University of British Columbia.
67. Kaminska, D., et al., *A feasibility study of ortho-positronium decays measurement with the J-PET scanner based on plastic scintillators.* Eur Phys J C Part Fields, 2016. **76**: p. 445.
68. Moskal, P., et al., *J-PET: A Novel TOF-PET Detector based on Plastic Scintillators*, in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*. 2016. p. 1-3.
69. Moskal, P., et al., *Test of a single module of the J-PET scanner based on plastic scintillators.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2014. **764**: p. 317-321.
70. Moskal, P., et al., *Time resolution of the plastic scintillator strips with matrix photomultiplier readout for J-PET tomograph.* Phys Med Biol, 2016. **61**(5): p. 2025-47.
71. Raczyński, L., et al., *Compressive sensing of signals generated in plastic scintillators in a novel J-PET instrument.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2015. **786**: p. 105-112.
72. Borghi, G., et al., *A 32 mm x 32 mm x 22 mm monolithic LYSO:Ce detector with dual-sided digital photon counter readout for ultrahigh-performance TOF-PET and TOF-PET/MRI.* Phys Med Biol, 2016. **61**(13): p. 4929-49.
73. Abreu, M.C., et al., *Design and evaluation of the clear-PEM scanner for positron emission mammography.* IEEE Transactions on Nuclear Science, 2006. **53**(1): p. 71-77.
74. Du, H., Y. Yang, and S.R. Cherry, *Measurements of wavelength shifting (WLS) fibre readout for a highly multiplexed, depth-encoding PET detector.* Phys Med Biol, 2007. **52**(9): p. 2499-514.
75. Shao, Y., et al., *Design studies of a high resolution PET detector using APD arrays.* IEEE Transactions on Nuclear Science, 2000. **47**(3): p. 1051-1057.
76. Yang, Y., et al., *Depth of interaction resolution measurements for a high resolution PET detector using position sensitive avalanche photodiodes.* Phys Med Biol, 2006. **51**(9): p. 2131-42.
77. Yang, Y., et al., *A prototype PET scanner with DOI-encoding detectors.* J Nucl Med, 2008. **49**(7): p. 1132-40.
78. Ito, M., S.J. Hong, and J.S. Lee, *Positron emission tomography (PET) detectors with depth-of- interaction (DOI) capability.* Biomedical Engineering Letters, 2011. **1**(2): p. 70-81.
79. Joung, J., R.S. Miyaoka, and T.K. Lewellen, *cMiCE: a high resolution animal PET using continuous LSO with a statistics based positioning scheme.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2002. **489**(1-3): p. 584-598.
80. Bruyndonckx, P., et al., *Towards a continuous crystal APD-based PET detector design.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2007. **571**(1-2): p. 182-186.
81. Bruyndonckx, P., et al., *Study of Spatial Resolution and Depth of Interaction of APD-Based PET Detector Modules Using Light Sharing Schemes.* IEEE Transactions on Nuclear Science, 2003. **50**(5): p. 1415-1419.
82. Hyun Chung, Y., et al., *Preliminary experimental results of a quasi-monolithic detector with DOI capability for a small animal PET.* Nuclear Instruments and Methods in Physics

Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2010. **621**(1-3): p. 590-594.
83. Lee, S.-J., et al., *A cross-stack quasi-monolithic detector with DOI capability for a small animal PET*. 2008: p. 3805-3809.
84. Lewellen, T.K., *Recent developments in PET detector technology*. Phys Med Biol, 2008. **53**(17): p. R287-317.
85. Ling, T., T.K. Lewellen, and R.S. Miyaoka, *Depth of interaction decoding of a continuous crystal detector module.* Phys Med Biol, 2007. **52**(8): p. 2213-28.
86. van der Laan, D.J., et al., *Spatial Resolution in Position-Sensitive Monolithic Scintillation Detectors*. 2006: p. 2506-2510.
87. Lewellen, T.K., M. Janes, and R.S. Miyaoka, *DMice - a depth-of-interaction detector design for PET scanners.* 2004. **4**: p. 2388-2392.
88. Miyaoka, R.S., et al., *Design of a depth of interaction (DOI) PET detector module.* IEEE Transactions on Nuclear Science, 1998. **45**(3): p. 1069-1073.
89. Yong Hyun, C., et al., *Optimization of dual layer phoswich detector consisting of LSO and LuYAP for small animal PET.* 2004: p. 2257-2261.
90. Chandrikamohan, P. and T.A. DeVol, *Comparison of Pulse Shape Discrimination Methods for Phoswich and CsI:Tl Detectors.* IEEE Transactions on Nuclear Science, 2007. **54**(2): p. 398-403.
91. Inadama, N., et al., *8-Layer DOI Encoding of 3-Dimensional Crystal Array.* IEEE Transactions on Nuclear Science, 2006. **53**(5): p. 2523-2528.
92. Inadama, N., et al., *DOI PET detectors with scintillation crystals cut as triangular prisms.* 2008: p. 3942-3945.
93. Murayama, H., et al., *Depth encoding multicrystal detectors for PET.* IEEE Transactions on Nuclear Science, 1998. **45**(3): p. 1152-1157.
94. Tsuda, T., et al., *A four-Layer depth of interaction detector block for small animal PET.* IEEE Transactions on Nuclear Science, 2004. **51**(5): p. 2537-2542.
95. Liu, H., et al., *Development of a depth of interaction detector for γ-rays.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2001. **459**(1-2): p. 182-190.
96. Nan, Z., et al., *Anode position and last dynode timing circuits for dual-layer BGO scintillator with PS-PMT based modular PET detectors.* IEEE Transactions on Nuclear Science, 2002. **49**(5): p. 2203-2207.
97. Parl, C., et al., *Dual layer doI detector modules for a dedicated mouse brain PET/MRI.* Phys Med Biol, 2019. **64**(5): p. 055004.
98. Stortz, G., et al. *Simulation guided optimization of dual layer offset detector design for use in small animal PET*. in *2011 IEEE Nuclear Science Symposium Conference Record*. 2011. IEEE.
99. Teimoorisichani, M. and A.L. Goertzen, *Geometry Optimization of a Dual-layer Offset Detector for Use in Simultaneous PET/MR Neuroimaging.* IEEE Transactions on Radiation and Plasma Medical Sciences, 2018: p. 1-1.
100. Thompson, C., et al., *Comparison of single and dual layer detector blocks for pre-clinical MRI–PET.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2013. **702**: p. 56-58.

101. Thompson, C.J., et al., *Evaluation of High Density Pixellated Crystal Blocks With SiPM Readout as Candidates for PET/MR Detectors in a Small Animal PET Insert.* IEEE Transactions on Nuclear Science, 2012. **59**(5): p. 1791-1797.
102. Grogg, K.S., et al., *National Electrical Manufacturers Association and Clinical Evaluation of a Novel Brain PET/CT Scanner.* J Nucl Med, 2016. **57**(4): p. 646-52.
103. Schellenberg, G., G. Stortz, and A.L. Goertzen, *An algorithm for automatic crystal identification in pixelated scintillation detectors using thin plate splines and Gaussian mixture models.* Phys Med Biol, 2016. **61**(3): p. N90-N101.
104. Chung, Y.H., et al., *Monte Carlo simulation of a four-layer DOI detector with relative offset in animal PET.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2011. **626-627**: p. 43-50.
105. Ito, M., et al., *A Four-Layer DOI Detector With a Relative Offset for Use in an Animal PET System.* IEEE Transactions on Nuclear Science, 2010. **57**(3): p. 976-981.
106. Levin, C.S., *Design of a high-resolution and high-sensitivity scintillation crystal array for PET with nearly complete light collection.* IEEE Transactions on Nuclear Science, 2002. **49**(5): p. 2236-2243.
107. Beltrame, P., et al., *Construction and tests of demonstrator modules for a 3-D axial PET system for brain or small animal imaging.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2011. **636**(1): p. S226-S230.
108. Braem, A., et al., *High precision axial coordinate readout for an axial 3-D PET detector module using a wave length shifter strip matrix.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2007. **580**(3): p. 1513-1521.
109. Knöß, C., *Evaluation and optimization of the high resolution research tomograph (HRRT).* 2004.
110. Schmand, M., et al., *Performance results of a new DOI detector block for a high resolution PET-LSO research tomograph HRRT.* IEEE Transactions on Nuclear Science, 1998. **45**(6): p. 3000-3006.
111. Wienhard, K., et al., *The ECAT HRRT: performance and first clinical application of the new high resolution research tomograph.* IEEE Transactions on Nuclear Science, 2002. **49**(1): p. 104-110.
112. Sossi, V., et al. *The second generation HRRT-a multi-centre scanner performance investigation*. in *IEEE Nuclear Science Symposium Conference Record, 2005*. 2005. IEEE.
113. Boellaard, R., et al., *Characterization of a single LSO crystal layer High Resolution Research Tomograph.* Physics in Medicine and Biology, 2003. **48**(4): p. 429-448.
114. Sheikhzadeh, P., et al., *Development and validation of an accurate GATE model for NeuroPET scanner.* Phys Med, 2017. **40**: p. 59-65.
115. Yamamoto, S., et al., *Development of a Brain PET System, PET-Hat: A Wearable PET System for Brain Research.* IEEE Transactions on Nuclear Science, 2011. **58**(3): p. 668-673.
116. Yamamoto, S., et al., *Development of PET-Hat: Wearable PET system for brain research.* Journal of Nuclear Medicine, 2009. **50**(supplement 2): p. 1532-1532.

117. Yamaya, T., et al., *Preliminary resolution performance of the prototype system for a 4-Layer DOI-PET scanner: jPET-D4.* IEEE Transactions on Nuclear Science, 2006. **53**(3): p. 1123-1128.
118. Yamaya, T., et al., *First Human Brain Imaging by the jPET-D4 Prototype With a Pre-Computed System Matrix.* IEEE Transactions on Nuclear Science, 2008. **55**(5): p. 2482-2492.
119. Yoshida, E., et al. *The jPET-D4: performance evaluation of four-layer DOI-PET scanner using the NEMA NU2-2001 standard*. in *2006 IEEE Nuclear Science Symposium Conference Record*. 2006. IEEE.
120. Katoh, N., et al., *A new brain positron emission tomography scanner with semiconductor detectors for target volume delineation and radiotherapy treatment planning in patients with nasopharyngeal carcinoma.* Int J Radiat Oncol Biol Phys, 2012. **82**(4): p. e671-6.
121. Mikhaylova, E., et al., *Simulation of the expected performance of a seamless scanner for brain PET based on highly pixelated CdTe detectors.* IEEE Trans Med Imaging, 2014. **33**(2): p. 332-9.
122. Morimoto, Y., et al., *Development of a 3D Brain PET Scanner Using CdTe Semiconductor Detectors and Its First Clinical Application.* IEEE Transactions on Nuclear Science, 2011. **58**(5): p. 2181-2189.
123. Morimoto, Y., et al., *Performance of a prototype brain PET scanner based on semiconductor detectors.* Journal of Nuclear Medicine, 2008. **49**(supplement 1): p. 122P-122P.
124. Ahmed, A.M., et al., *Investigation of the optimal detector arrangement for the helmet-chin PET – A simulation study.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2017. **858**: p. 96-100.
125. Athanasiades, A., et al., *A new PET scanner for functional brain imaging based on 2-mm straw detectors*, in *2008 IEEE Nuclear Science Symposium Conference Record*. 2008. p. 5445-5451.
126. Gong, K., et al., *Designing a compact high performance brain PET scanner-simulation study.* Phys Med Biol, 2016. **61**(10): p. 3681-97.
127. Tashima, H. and T. Yamaya, *Proposed helmet PET geometries with add-on detectors for high sensitivity brain imaging.* Phys Med Biol, 2016. **61**(19): p. 7205-7220.
128. Tashima, H., et al., *Development of the helmet-chin PET prototype*, in *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. 2015. p. 1-3.
129. Yamaya, T., et al., *First prototype of a compact helmet-chin PET for high-sensitivity brain imaging.* Journal of Nuclear Medicine, 2015. **56**(supplement 3): p. 317-317.
130. Schlemmer, H., et al. *Combined clinical MR/PET: feasibility of simultaneous imaging*. in *Proceedings of the 15th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*. 2007.
131. Nishikido, F., et al., *Feasibility of a brain-dedicated PET-MRI system using four-layer DOI detectors integrated with an RF head coil.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2014. **756**: p. 6-13.
132. Tsuda, T., et al., *Measurement of 32 ~ 8 ~ 4 LYSO Crystal Responses of DOI Detector for jPET-RD*, in *IEEE Nuclear Science Symposium Conference Record, 2005*. 2005. p. 2881-2884.

133. Tsuda, T., et al., *Performance evaluation of a subset of a four-layer LSO detector for a small animal DOI PET scanner: jPET-RD.* IEEE Transactions on Nuclear Science, 2006. **53**(1): p. 35-39.
134. Nishikido, F., et al., *Development of 1.45-mm resolution four-layer DOI–PET detector for simultaneous measurement in 3T MRI.* Radiological Physics and Technology, 2015. **8**(1): p. 111-119.
135. Benlloch, J.M., et al., *The MINDVIEW project: First results.* Eur Psychiatry, 2018. **50**: p. 21-27.
136. González, A.J., et al., *Progress report on the MindView brain PET detector module based on large area SiPMs arrays.* EJNMMI Physics, 2014. **1**(1): p. A66.
137. González, A.J., et al., *The MINDView brain PET detector, feasibility study based on SiPM arrays.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2016. **818**: p. 82-90.
138. Gonzalez, A.J., et al., *A novel brain PET insert for the MINDView project*, in *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. 2014. p. 1-4.
139. Anger, H.O., *Scintillation Camera.* Review of Scientific Instruments, 1958. **29**(1): p. 27-33.
140. Peterson, T.E. and L.R. Furenlid, *SPECT detectors: the Anger Camera and beyond.* Phys Med Biol, 2011. **56**(17): p. R145-82.
141. Tai, Y.C. and R. Laforest, *Instrumentation aspects of animal PET.* Annu Rev Biomed Eng, 2005. **7**: p. 255-85.
142. Jan, S., et al., *GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy.* Phys Med Biol, 2011. **56**(4): p. 881-901.
143. Jan, S., et al., *GATE: a simulation toolkit for PET and SPECT.* Physics in Medicine and Biology, 2004. **49**(19): p. 4543-4561.
144. Strulab, D., et al., *GATE (geant4 application for tomographic emission): a PET/SPECT general-purpose simulation platform.* Nuclear Physics B - Proceedings Supplements, 2003. **125**: p. 75-79.
145. Moses, W.W., C. Granja, and C. Leroy, *Recent Advances and Future Advances in Time-of-Flight PET.* 2010: p. 119-125.
146. Moszynski, M., et al., *New Prospects for Time-of-Flight PET With LSO Scintillators.* IEEE Transactions on Nuclear Science, 2006. **53**(5): p. 2484-2488.
147. Schaart, D.R., et al., *LaBr(3):Ce and SiPMs for time-of-flight PET: achieving 100 ps coincidence resolving time.* Phys Med Biol, 2010. **55**(7): p. N179-89.
148. van Dam, H.T., et al., *Sub-200 ps CRT in monolithic scintillator PET detectors using digital SiPM arrays and maximum likelihood interaction time estimation.* Phys Med Biol, 2013. **58**(10): p. 3243-57.
149. Parra, L. and H.H. Barrett, *List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET.* IEEE Trans Med Imaging, 1998. **17**(2): p. 228-35.
150. Karp, J.S., et al., *Benefit of time-of-flight in PET: experimental and clinical results.* J Nucl Med, 2008. **49**(3): p. 462-70.
151. Moses, W.W., *Time of flight in pet revisited.* IEEE Transactions on Nuclear Science, 2003. **50**(5): p. 1325-1330.
152. Rahmim, A., et al., *Statistical list-mode image reconstruction for the high resolution research tomograph.* Physics in Medicine and Biology, 2004. **49**(18): p. 4239-4258.

153. Reader, A.J., et al., *One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays.* IEEE Transactions on Nuclear Science, 2002. **49**(3): p. 693-699.
154. Reader, A.J. and H. Zaidi, *Advances in PET Image Reconstruction.* PET Clin, 2007. **2**(2): p. 173-90.
155. Defrise, M., et al., *Exact and approximate rebinning algorithms for 3-D PET data.* IEEE Trans Med Imaging, 1997. **16**(2): p. 145-58.
156. Liu, X., et al., *Exact rebinning methods for three-dimensional PET.* IEEE Trans Med Imaging, 1999. **18**(8): p. 657-64.
157. Chuanyong, B., et al., *A generalized model for the conversion from ct numbers to linear attenuation coefficients.* IEEE Transactions on Nuclear Science, 2003. **50**(5): p. 1510-1515.
158. Keereman, V., et al., *The effect of errors in segmented attenuation maps on PET quantification.* Med Phys, 2011. **38**(11): p. 6010-9.
159. Kops, E.R. and H. Herzog, *Template based attenuation correction for PET in MR-PET scanners.* 2008: p. 3786-3789.
160. Hofmann, M., et al., *Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques.* Eur J Nucl Med Mol Imaging, 2009. **36 Suppl 1**: p. S93-104.
161. Keereman, V., et al., *MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences.* J Nucl Med, 2010. **51**(5): p. 812-8.
162. Zaidi, H., *Scatter modelling and correction strategies in fully 3-D PET.* Nucl Med Commun, 2001. **22**(11): p. 1181-4.
163. Khalil, M.M., *Basic science of PET imaging*. 2017: Springer.
164. Badawi, R.D., M.A. Lodge, and P.K. Marsden, *Algorithms for calculating detector efficiency normalization coefficients for true coincidences in 3D PET.* Physics in Medicine and Biology, 1998. **43**(1): p. 189-205.
165. Saha, G.B., *Basics of PET imaging: physics, chemistry, and regulations*. 2015: Springer.
166. NEMA, N.E.M.A., *Standard Publication NU 4-2008, Performance Measurements of Small Animal Positron Emission Tomographs (NEMA, Rosslyn, VA, 2008).* 2008.
167. NEMA, N.E.M.A., *Standards Publication NU 2-2018, Performance Measurements of Positron Emission Tomographs (NEMA, Rosslyn, VA, 2018).* 2018.
168. Moses, W.W., *Fundamental Limits of Spatial Resolution in PET.* Nucl Instrum Methods Phys Res A, 2011. **648 Supplement 1**: p. S236-S240.
169. Strother, S.C., M.E. Casey, and E.J. Hoffman, *Measuring PET scanner sensitivity: relating countrates to image signal-to-noise ratios using noise equivalents counts.* IEEE Transactions on Nuclear Science, 1990. **37**(2): p. 783-788.
170. Bloomfield, P.M., et al., *The design and implementation of a motion correction scheme for neurological PET.* Phys Med Biol, 2003. **48**(8): p. 959-78.
171. Cherry, S.R., et al., *Total-body imaging: Transforming the role of positron emission tomography.* Sci Transl Med, 2017. **9**(381).
172. Agostinelli, S., et al., *Geant4—a simulation toolkit.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2003. **506**(3): p. 250-303.
173. Allison, J., et al., *Geant4 developments and applications.* IEEE Transactions on Nuclear Science, 2006. **53**(1): p. 270-278.

174. Allison, J., et al., *Recent developments in G eant 4.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2016. **835**: p. 186-225.
175. Thielemans, K., et al., *STIR: software for tomographic image reconstruction release 2.* Phys Med Biol, 2012. **57**(4): p. 867-83.
176. Merlin, T., et al., *CASToR: a generic data organization and processing code framework for multi-modal and multi-dimensional tomographic reconstruction.* Phys Med Biol, 2018. **63**(18): p. 185005.
177. RADON, J., *Uber die Bestimmung von Funktionen durch ihre Integralwerte langs gewisser Mannigfatigkeiten.* Berichte Sachsische Akademie der Wissenschaften (Leipzig), 1971. **69**: p. 262-277.
178. Radon, J., *On the Determination of Functions from Their Integral Values along Certain Manifolds.* IEEE Trans Med Imaging, 1986. **5**(4): p. 170-6.
179. Bendriem, B. and D.W. Townsend, *The theory and practice of 3D PET*. Vol. 32. 2013: Springer Science & Business Media.
180. Bruyant, P.P., *Analytic and iterative reconstruction algorithms in SPECT.* J Nucl Med, 2002. **43**(10): p. 1343-58.
181. Lange, K. and R. Carson, *EM reconstruction algorithms for emission and transmission tomography.* J Comput Assist Tomogr, 1984. **8**(2): p. 306-16.
182. Hudson, H.M. and R.S. Larkin, *Accelerated image reconstruction using ordered subsets of projection data.* IEEE Trans Med Imaging, 1994. **13**(4): p. 601-9.
183. Kendall, M.G., *Advanced Theory Of Statistics Vol-I*. 1943: Charles Griffin: London.
184. Bai, B., Q. Li, and R.M. Leahy, *Magnetic resonance-guided positron emission tomography image reconstruction.* Semin Nucl Med, 2013. **43**(1): p. 30-44.
185. Geman, S. and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images.* IEEE Trans Pattern Anal Mach Intell, 1984. **6**(6): p. 721-41.
186. Besag, J., *Spatial Interaction and the Statistical Analysis of Lattice Systems.* Journal of the Royal Statistical Society: Series B (Methodological), 1974. **36**(2): p. 192-225.
187. Green, P.J., *Bayesian reconstructions from emission tomography data using a modified EM algorithm.* IEEE Trans Med Imaging, 1990. **9**(1): p. 84-93.
188. NVIDIA. *Programming Guide:: CUDA Toolkit Documentation*. 2019 02/2019]; Available from: https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html.
189. Cheng, J., M. Grossman, and T. McKercher, *Professional Cuda C Programming*. 2014: John Wiley & Sons.
190. Pratx, G. and C. Levin, *Online detector response calculations for high-resolution PET image reconstruction.* Phys Med Biol, 2011. **56**(13): p. 4023-40.
191. Pratx, G., et al., *Fast, accurate and shift-varying line projections for iterative reconstruction using the GPU.* IEEE Trans Med Imaging, 2009. **28**(3): p. 435-45.
192. Pratx, G. and L. Xing, *GPU computing in medical physics: a review.* Med Phys, 2011. **38**(5): p. 2685-97.
193. Cui, J.Y., et al., *Fully 3D list-mode time-of-flight PET image reconstruction on GPUs using CUDA.* Med Phys, 2011. **38**(12): p. 6775-86.
194. Rodgers, D.P., *Improvements in multiprocessor system design.* ACM SIGARCH Computer Architecture News, 1985. **13**(3): p. 225-231.
195. Cui, J., et al., *Distributed MLEM: an iterative tomographic image reconstruction algorithm for distributed memory architectures.* IEEE Trans Med Imaging, 2013. **32**(5): p. 957-67.

196. Teimoorisichani, M. and A.L. Goertzen, *Geometry optimization of dual-layer offset detectors for compact ring diameter PET systems.* 2016: p. 1-3.
197. Connell, I.R., et al., *Design of a parallel transmit head coil at 7T with magnetic wall distributed filters.* IEEE Trans Med Imaging, 2015. **34**(4): p. 836-45.
198. Yamaya, T., et al., *A SiPM-based isotropic-3D PET detector X'tal cube with a three-dimensional array of 1 mm(3) crystals.* Phys Med Biol, 2011. **56**(21): p. 6793-807.
199. Vandenbroucke, A., et al., *Performance characterization of a new high resolution PET scintillation detector.* Phys Med Biol, 2010. **55**(19): p. 5895-911.
200. Ota, R., et al., *Evaluation of a Sub-Millimeter Resolution PET Detector With a 1.2 mm Pitch TSV-MPPC Array One-to-One Coupled to LFS Scintillator Crystals and Inter-Crystal Scatter Studies With Individual Signal Readout.* IEEE Transactions on Radiation and Plasma Medical Sciences, 2017. **1**(1): p. 15-22.
201. Levin, C.S. and H. Zaidi, *Current Trends in Preclinical PET System Design.* PET Clin, 2007. **2**(2): p. 125-60.
202. Inadama, N., et al., *A DOI PET detector having extended X'tal cube structure*, in *IEEE Nuclear Science Symposuim & Medical Imaging Conference*. 2010. p. 3346-3348.
203. Bergeron, M., et al., *LabPET II, an APD-based Detector Module with PET and Counting CT Imaging Capabilities.* IEEE Transactions on Nuclear Science, 2015. **62**(3): p. 756-765.
204. Bergeron, M., et al., *LabPET II, an APD-based PET detector module with counting CT imaging capability*, in *2011 IEEE Nuclear Science Symposium Conference Record*. 2011. p. 3543-3547.
205. Ferramacho, L., *Measurements from different SiPMs with TOFPET v1 ASIC*, in *3rd FAST WG3/4 meeting*. 2017, PETsys Electronics.
206. Sportelli, G., et al., *The TRIMAGE PET Data Acquisition System: Initial Results.* IEEE Transactions on Radiation and Plasma Medical Sciences, 2017. **1**(2): p. 168-177.
207. Poon, J.K., et al., *Optimal whole-body PET scanner configurations for different volumes of LSO scintillator: a simulation study.* Phys Med Biol, 2012. **57**(13): p. 4077-94.
208. Conti, M., *Tailoring PET Time Coincidence Window Using CT Morphological Information.* IEEE Transactions on Nuclear Science, 2007. **54**(5): p. 1599-1605.
209. Thompson, C.J., S. Kecani, and S. Boelen, *Evaluation of a neck-shield for use during neurological studies with a whole-body PET scanner*, in *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No.00CH37149)*. 2000. p. 17/31-17/36.
210. Thompson, C.J. and J.J. Moreno-Cantu, *Measurement of the change in noise-effective count rate during PET brain studies with additional shielding.* IEEE Transactions on Nuclear Science, 2002. **49**(5): p. 2057-2061.
211. Teimoorisichani, M. and A.L. Goertzen, *A study of inter-crystal scatter in dual-layer offset scintillator arrays for brain-dedicated PET scanners.* Phys Med Biol, 2019.
212. Shao, Y., et al., *A study of inter-crystal scatter in small scintillator arrays designed for high resolution PET imaging.* IEEE Transactions on Nuclear Science, 1996. **43**(3): p. 1938-1944.
213. Levin, C.S., et al., *Compton scatter and X-ray crosstalk and the use of very thin intercrystal septa in high-resolution PET detectors.* IEEE Transactions on Nuclear Science, 1997. **44**(2): p. 218-224.
214. Degenhardt, C., et al., *Impact of Intercrystal Crosstalk on Depth-of-Interaction Information in PET Detectors.* IEEE Transactions on Nuclear Science, 2007. **54**(3): p. 427-432.

215. Park, S.J., W.L. Rogers, and N.H. Clinthorne, *Effect of intercrystal Compton scatter on efficiency and image noise in small animal PET module.* 2004: p. 2272-2277.
216. Stickel, J.R. and S.R. Cherry, *High-resolution PET detector design: modelling components of intrinsic spatial resolution.* Physics in Medicine and Biology, 2005. **50**(2): p. 179-195.
217. Qi, J., *Calculation of the Sensitivity Image in List-Mode Reconstruction for PET.* IEEE Transactions on Nuclear Science, 2006. **53**(5): p. 2746-2751.
218. Rahmim, A., J. Qi, and V. Sossi, *Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls.* Med Phys, 2013. **40**(6): p. 064301.
219. Zhang, X., et al., *Quantitative image reconstruction for total-body PET imaging using the 2-meter long EXPLORER scanner.* Phys Med Biol, 2017. **62**(6): p. 2465-2485.
220. Ruetsch, G. and P. Micikevicius, *Optimizing matrix transpose in CUDA.* Nvidia CUDA SDK Application Note, 2009. **18**.
221. Markiewicz, P.J., et al., *High throughput CUDA implementation of accurate geometric modelling for iterative reconstruction of PET data.* 2014: p. 1-4.
222. Reader, A.J., et al., *EM Algorithm System Modeling by Image-Space Techniques for PET Reconstruction.* IEEE Transactions on Nuclear Science, 2003. **50**(5): p. 1392-1397.
223. Tong, S., et al., *Properties and Mitigation of Edge Artifacts in PSF-Based PET Reconstruction.* IEEE Transactions on Nuclear Science, 2011. **58**(5): p. 2264-2275.
224. D'Amico, A.V., *Biochemical Outcome After Radical Prostatectomy, External Beam Radiation Therapy, or Interstitial Radiation Therapy for Clinically Localized Prostate Cancer.* Jama, 1998. **280**(11).
225. Iczkowski, K.A. and M.S. Lucia, *Frequency of positive surgical margin at prostatectomy and its effect on patient outcome.* Prostate Cancer, 2011. **2011**: p. 673021.
226. Rowe, S.P., et al., *PET imaging of prostate-specific membrane antigen in prostate cancer: current state of the art and future challenges.* Prostate Cancer Prostatic Dis, 2016. **19**(3): p. 223-30.
227. Sasikumar, A., *Specificity of (68)Ga-PSMA PET/CT for Prostate Cancer - Myths and Reality.* Indian J Nucl Med, 2017. **32**(1): p. 11-12.
228. Grootendorst, M.R., et al., *Intraoperative Assessment of Tumor Resection Margins in Breast-Conserving Surgery Using (18)F-FDG Cerenkov Luminescence Imaging: A First-in-Human Feasibility Study.* J Nucl Med, 2017. **58**(6): p. 891-898.
229. Zhang, X., et al., *Development and evaluation of a LOR-based image reconstruction with 3D system response modeling for a PET insert with dual-layer offset crystal design.* Phys Med Biol, 2013. **58**(23): p. 8379-99.
230. Badawi, R.D. and P.K. Marsden, *Developments in component-based normalization for 3D PET.* Physics in Medicine and Biology, 1999. **44**(2): p. 571-594.
231. Defrise, M., et al., *A normalization technique for 3D PET data.* Physics in Medicine and Biology, 1991. **36**(7): p. 939-952.
232. Rahmim, A., et al., *Statistical dynamic image reconstruction in state-of-the-art high-resolution PET.* Phys Med Biol, 2005. **50**(20): p. 4887-912.
233. Lewitt, R.M., G. Muehllehner, and J.S. Karp, *Three-dimensional image reconstruction for PET by multi-slice rebinning and axial image filtering.* Physics in Medicine and Biology, 1994. **39**(3): p. 321-339.
234. Siddon, R.L., *Fast calculation of the exact radiological path for a three-dimensional CT array.* Med Phys, 1985. **12**(2): p. 252-5.

235. Tang, J. and A. Rahmim, *Anatomy assisted PET image reconstruction incorporating multi-resolution joint entropy.* Phys Med Biol, 2015. **60**(1): p. 31-48.