Vrije Universiteit Brussel

FACULTY OF SCIENCE
Department of Physics

Use Of Machine Learning Algorithms For Gamma Detection In Positron Emission Tomography

Cedric Lemaître

Use of machine learning algorithms for gamma detection in positron emission tomography

Vrije Universiteit Brussel

IIHE
Interuniversity Institute for High Energies

Cedric Lemaître

Promotoren:
Prof. Dr. P. Bruyndonckx
Prof. Dr. S.P.K. Tavernier

Proefschrift ingediend met het oog op het behalen van de graad van
Doctor in de wetenschappen

Doctoral examination commission
Peter Bruyndonckx (Promotor, VUB ZAP)
Stefaan Tavernier (Co-Promotor en secretaris, VUB ZAP)
Michel Defrise (Chairperson, VUB ZAP)
Rudi Deklerck (VUB, Dr.ir)
Karl Ziemons (Forschungszentrum Julich, Dr.) (external)
Yves D'asseler (Ugent, ZAP) (external)
Bernard Manderick (VUB ZAP)

Special Acknowledgment:

This PhD thesis was funded by

“Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen)” under Grants SB-031062/SB-33062.

Give a man a fish and you feed him for a day.
Teach a man to fish and you feed him for a lifetime.

Chinese Proverb

Acknowledgments

Finally!

It is hard to believe, but you are holding my PhD thesis in your hands. The writing was a difficult mission, certainly in combination with a new job and two kids. However, as long as there are enough people around you who support you, it is possible.

Writing a thesis is perhaps a lonely occupation but a PhD is not something you do on your own. That is why, I would like to express a special word of gratitude to the following people for their support and friendship I encountered during my PhD.

Firstly, I want to thank Stefaan Tavernier for accepting me as PhD student within the PET group. I am even more grateful to Peter Bruyndonckx. He is the person who taught me how “to fish” in this PET world. Thanks a lot for your never ending patience, fruitful discussions and the proof-reading of this thesis. I highly appreciate your way of working; the autonomy and responsibility you gave me. I also want to thank Christian Wastiels. Together we solved a lot of electronic and mechanic scanner problems, often with a lot of fun. I enjoyed this a lot.

A big thanks is also in order to Marleen Goeman for her efficient administrative management. I also wish to thank all members of the PET-team. It was very pleasant to work with you: Jun Dang, Jan Debruyne, Olivier Devroede, Magalie Krieguer, Mateusz Wedrowski & Li Zhi.

In addition, there are of course several other ‘academic’ friends who deserve some special attention; Daan, Eileen, Kevin, Severin, Sofie, ... many thanks for the pleasant lunches at the VUB cafeteria. Jan, I do not know if you have to congratulate me or I have to congratulate you for the four years office’s sharing, but it was definitely an enrichment.

I would like to thank my parents and brothers, relatives and friends. Thank you for your encouragements and the good time I spent with you. Besides, here is the answer on your question:

“Wanneer ga je je thesis afwerken?” - ‘t Is af!’

Last but not least, Sarah, no words can express my gratitude. Your support and encouragements were in the end what made this PhD possible, you brought me all I needed, give me the time and love which were indispensable. Together with Kobe and Wout you give me a lot of happiness during these challenging years. I am also happy we can finally live a normal life, all together and continue with the renovation of our house.

Thanks to all of you!

Table of contents

Introduction

Positron emission tomography (PET) is a nuclear medicine imaging technique based on the detection of gamma rays emitted by positron-emitting short lived isotopes. It is one of the noninvasive technologies that can routinely and quantitatively measure metabolic, biochemical, and functional activity in living tissue. It assesses changes in the function, circulation, and metabolism of body organs. PET is an important modality because PET images can demonstrate pathologic changes in the human body even before they are seen on the other imaging modalities. As the methods and technology of PET imaging have advanced, they have been applied not only to human clinical imaging, but to small research animals as well using dedicated apparatus. Small animal PET allows the study of pharmacological effects of certain drugs over a long period of time in a limited number of animals. It also allows the validation of newly developed radio tracers and it plays a growing role in research on diseases and genome. However, imaging small objects pushes the spatial resolution and sensitivity limits of current PET technology.

This thesis consists in a study at the level of PET detector blocks based on monolithic crystals in combination with a machine learning positioning algorithm. These detector blocks, when compared with pixilated ones, feature a simpler design, lower cost, larger sensitive volume, better energy resolution and comparable or better spatial resolution.
The use of machine learning algorithms allows a positioning estimation that does not suffer much from parallax errors. The study is part of one of the Crystal Clear Projects. The aim of this collaboration is to build new generation PET scanners with high resolution and high sensitivity. The block approach in combination with a Neural network positioning algorithm will be used on a research prototype for human brain PET/MRI imaging.

The first chapter of this thesis depicts the role of PET within the spectrum of medical imaging. It also outlines the different projects within the Crystal Clear Collaboration. In the second chapter, the process flow from tracer admission till image reconstruction is considered. It describes the fundamentals of PET, PET's detection principles and the parameters that determine the quality of PET images.
The spatial resolution obtained with a monolithic LSO block in combination with one of the three analyzed machine learning positioning algorithms is evaluated in chapter 3.

Levenberg-Marquardt and algebraic neural networks as well as support vector machines were considered. The data needed for the evaluation was acquired on a bench set-up with three different block geometries, a rectangular 20x10x10mm3, a bigger 20x10x20mm3 rectangular block and a trapezoidal 20x15.4(11.5)x20mm3 block.

Based on the results of chapter 3, the best -block geometry/positioning algorithm-combination was used on the PET prototype demonstrator. In chapter 4, this prototype demonstrator and an automated method of training data acquisition are reported. In addition the influence of the acquisition parameters on the neural network performance was investigated. In the last chapter some two-dimensional reconstructed images of point sources with different diameters and a home made Derenzo phantom were analyzed. Finally, the main conclusions from this thesis are summarized in a last section.

Chapter 1

The role of PET within the spectrum of medical imaging

1.1 Biomedical imaging modalities

Medical imaging refers to the techniques and processes used to create images of the human body for clinical purposes (medical procedures seeking to reveal diagnose or examine a disease) or medical science (including the study of anatomy and functionality). Medical imaging techniques can be classified according to a number of criteria. A particular classification scheme could use appearance, e.g. tomographic versus non-tomographic images and would group computed tomography (CT) and magnetic resonance imaging (MRI) because of the similarity in image presentation. Another would classify the techniques according to the underlying physics. This is the classification scheme which is used here. Its basis will be the origin and nature of the information carrier e.g. sound waves in ultrasound, X-rays in CT and gamma rays in nuclear medicine [1]. The scheme has two branches relating to the source of information carriers: external or internal (Figure 1.1).

1.1.1 External sources

When the source is external, the body structures modulate the information through interactions with the carriers. In X-ray radiography or CT, an external point source of X-rays is used. The X-rays are partially absorbed when the rays pass through the body. The rays that are neither absorbed nor scattered move in straight lines between the point source and the detector, thus creating a shadow image of the bodily structures.

In ultrasound, an external source of pulsed sound waves is used. Interfaces between different tissues will partially reflect the sound waves. By measuring the time span between the outgoing and incoming sound pulse, images can be reconstructed.

In endoscopy, an external light source illuminates internal organs through a glass fiber. An ocular or small camera is used to observe the reflected light and hence the organ.

1.1.2 Internal sources

The body naturally and continuously radiates heat: its source is internal. In order to image the information carriers, some optics for infrared radiation is needed: a thermographic camera.

Electroencephalography is the neurophysiologic measurement of the electrical activity of the brain. The resulting traces are known as an electroencephalogram (EEG) and represent an electrical signal from a large number of neurons. An electrocardiogram (ECG) is a graphic produced by an electrocardiograph, which records the electrical activity of the heart over time. In both techniques, EEC and ECG, the information arises from an internal change of electrical activity.

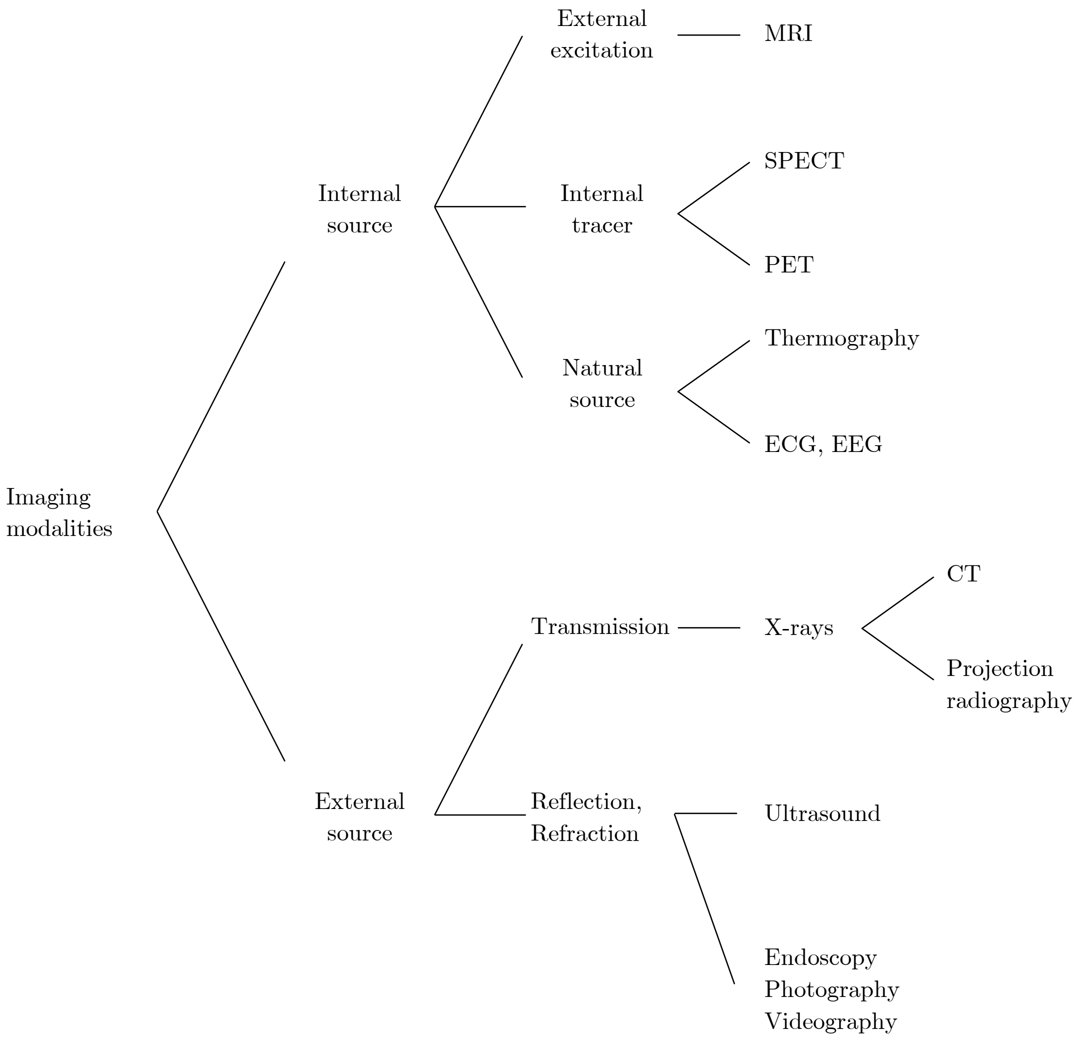


Figure 1.1: A classification of the medical imaging modalities according to their underlying physics

In MRI, the information carriers are radio waves emitted by hydrogen nuclei in the body. Although the body has plenty of hydrogen nuclei, e.g. in water molecules, the nuclei do not naturally emit radio waves. In order for them to do so, they have to be first put in a magnetic field and then excited by means of well-chosen radio wave pulses at specific frequencies. The nuclei then ‘answer’ on their turn by emitting radio waves of similar frequencies. In MRI, the internal sources are always present but they only emit information when excited to do so.

In nuclear imaging, PET and SPECT, the internal information carriers are gamma rays emitted by injected radioisotopes bound to molecules with known biological properties. They are chosen in such a way that their radioactive decay allows for external detection and that their space/time distribution reflects clinical information.

The techniques which rely on ionizing radiation such as X-rays, SPECT and PET subject the patient to a radiation dose that carries a small but finite potential detriment. This places a limit on their usage, particularly in the cases of pregnant women and children. Ultrasound and MRI do not carry these risks. Nevertheless, nuclear medicine techniques like SPECT and PET, yield metabolic, physiologic and pathologic information of the body. They are dependent upon the development of a radiopharmaceutical which offers specific uptake into the organ of interest. By measuring the sites of uptake and the rates of uptake/disappearance, information about organ function can be obtained.

Since the other techniques yield images of physical structural parameters, they are largely concerned with the anatomy of the patient. More specialist techniques such as Doppler ultrasound and functional magnetic resonance imaging (fRMI) have begun to extend their range into physiology also. Although the signal-to-noise ratio (SNR) is weak and quantitative data remains difficult to remain. To clearly make the distinction with fMRI that can now provide some functional information but not at the molecular level, PET and SPECT should be classified as molecular imaging techniques instead of functional techniques (Figure 1.2). However, MRI/ultrasound versus PET/SPECT are rather of complementary nature, instead of being in competition.

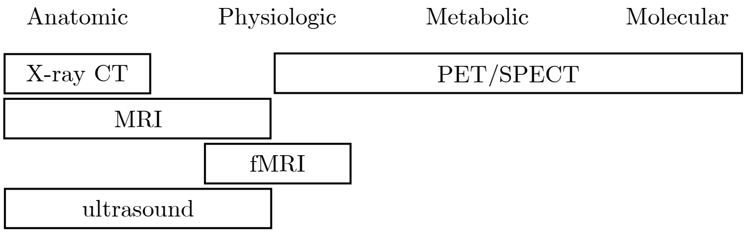


Figure 1.2: This overview illustrates the most common used medical imaging techniques. It covers the entities that can be detected which range from structure, physiology, metabolism/biochemistry, molecular pathway and molecular targets/receptor/binding sites

1.2 The Crystal Clear Collaboration

The Crystal Clear Collaboration (CCC) is an international collaboration grouping several research institutes and universities, working together to develop new generations of scanners for positron emission tomography (Table 1.1)

Table 1.1 Formal members of the Crystal Clear Collaboration

Institute-Company	Location
CERN	Geneva (Switzerland)
CERMEP	Lyon (France)
CIEMAT	Madrid (Spain)
Forschungszentrum Jülich (FZJ)	Jülich (Germany)
Institute for Physical Research	Ashtarak (Armenia)
Institute of Nuclear Problems	Minsk (Belarus)
LIP	Lisbon (Portugal)
Samsung Medical Center	Seoul (Korea)
Université Claude Bernard	Lyon (France)
Université de Lausanne,	Lausanne (Switserland)
Universiteit Gent	Ghent (Belgium)
Vrije Universiteit Brussel, IIHE	Brussels (Belgium)

This collaboration, created in 1990, was at first mainly interested in scintillators for high-energy physics which would be suitable for use at the LHC collider. Later, the interest of the collaboration moved to nuclear medicine and more particular to PET instrumentation development. It was felt that the experience available in the collaboration on scintillation materials and photo detectors would allow a useful contribution to this field.

1.2.1 Small animal PET – The ClearPET-project

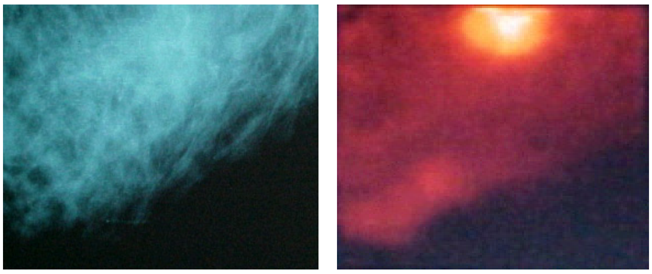
Small animal (rats and mice) PET imaging is being used increasingly as a basic measurement tool in modern biomedical research. Applications of this technology occur in development and evaluation of therapeutic efficacy in small animal models of human disease, visualization and quantification of the site and amount of gene expression, and in many other settings, e.g. visualization and quantification of the movement of various cell types within the body [2]. Newly developed drugs are tested in animal first, both for their biological utility and for their toxicity. Due to the non invasive character of PET, it is possible to study the progress of a disease and its treatment efficiency over time, by performing repeated scans in the same animal. Despite these successes, at least two important technical issues remained to be addressed. First, the intrinsic spatial resolution of contemporary clinical PET scanners is not sufficient to allow definitive visualization of organ substructures in the mouse. Second, in addition to improved spatial resolution, the sensitivity of small animal PET scanners must also be increased if pharmacologic effects are to be avoided. Improved sensitivity has also been identified as an important goal for the image reconstruction. Thus, sensitivity, as well as resolution, should be maximized.

End 90's, these specific needs related to the small animal PET area, motivated the ClearPET project in the development of small animal PET devices. Together with a number of guest laboratories, the institutes provide expertise in different domains of physics instrumentation, biology and medicine. Their research activities have led to the design and construction of five prototypes of a new generation of PET scanners for small animals, which provide depth-of-interaction (DOI) information [3, 4]. The detector heads are based on an 8 x 8 matrix of scintillation crystal elements, read out by a multi-anode photomultiplier tube. Each element consists of a phosphor sandwich, or phoswich, made up of two layers of crystals with different decay times. One layer is formed from cerium-doped lutetium yttrium orthosilicate (LYSO) scintillator material; the other contains cerium-doped lutetium yttrium aluminate perovskite (LuYAP) scintillator, specially developed by the Crystal Clear collaboration and now commercially available from several companies.

Two versions of the scanner are being commercially produced, differing only in the mechanics of the gantry. The ClearPET Neuro is optimized for small primates and features a gantry that can be tilted to allow the animal to be imaged in a sitting position, while the ClearPET Rodent is optimized for rats and mice. These machines are now commercialized by the German company Raytest GmbH under the name ClearPET$^{\text{TM}}$ [5].

1.2.2 Positron Emission Mammography – The ClearPEM-project

The early detection of breast cancer is becoming a priority in healthcare policy of an increasing number of countries. A very large number of women (about one woman in eight) will develop a breast cancer, which is the second leading cause of cancer death. On the other hand, early detection leads to very high cure rate [6]. By the end of the 1970s, the introduction of breast cancer screening by mammography was an important advancement in medical imaging. Permitting early detection of this illness, several randomized, controlled screening studies have shown an overall decrease in breast cancer mortality of up to 30%. Unfortunately, the specificity of conventional X-ray mammography is rather low. In particular for women with dense breast (about 40% of the cases), until 50% of the tumors are not detected by conventional X-ray mammography. Specifying the nature (malign or benign) of tumors using this method is also problematic and this results in a high number of false positive results.



Digital X-ray Mammogram FDG-PET Mammogram

Figure 1.3: Digital X-ray and PET (10 mCi FDG, 4 min acquisition time) mammograms. (*Courtesy of L.P. Adler, Cancer Centre, Philadelphia.*)

A large number of unnecessary biopsies or even auxiliary dissections are therefore performed which have a high cost for the society, not considering the psychological aspect on the women. On the other hand, metabolic techniques using ^{18}F–fluoro-deoxy-glucose (FDG) have demonstrated an excellent sensitivity to malignant tissues (nearly 100%) due to the much higher glucose consumption of cancerous cells as compared to normal tissue.

The cancer detection capability of PET is clearly illustrated in Figure 1.3 comparing X-ray and PET images of the same breast. To respond to the demand for a highly specific device the Crystal Clear Collaboration started in 2002 the design and construction of a dedicated positron emission mammogram, the ClearPEM project [7]. The ClearPEM scanner is developed with three main guidelines: low random background; high sensitivity; and spatial resolution smaller than 2mm. The first requirement arises from the fact that the scanner must cope with a large single photon rate due to the close presence of the heart. In order to increase the sensitivity the ClearPEM imaging system exploits Compton interactions in the detector, in which case the reconstruction of the scattering topology is required. Finally, in order to deliver the required spatial resolution allover the field-of-view without compromising the sensitivity by restricting the angle of the accepted lines-of-response, the detector is able to measure the DOI of the incoming photons. The basic detector module is composed by a matrix of 32 2x2x20 mm^3 LYSO:Ce pixels, readout at both ends by Hamamatsu S8550 APD arrays for DOI capability. Twenty-four of these basic detector modules are mechanically fixed and electrically connected to front and back electronics PCBs forming a super module. Four super modules are mounted in a detector head. In total the ClearPEM scanner has 192 detector modules and 6144 crystals.

1.2.3 A functional PET system for the human brain – The BrainPET-project

During the last two decades, functional brain imaging using PET has advanced constantly, and steadily gained importance in the clinical and research arenas [8]. Nevertheless, emerging clinical and research applications of functional brain imaging promise even greater levels of accuracy and precision and therefore, impose more constraints with respect to the intrinsic performance of the PET tomograph. Continuous efforts to integrate recent research findings for the design of different geometries and various detector read out technologies of PET scanners have become the goal of both the academic community and nuclear medicine industry. Also the software- and hardware-based correlation between anatomical (X-ray CT, MRI) and physiological (PET) information is a promising research field and now offers unique capabilities for the medical imaging community and biomedical researchers. One of the main advantages of dual-modality PET/CT imaging is that PET data are intrinsically aligned to anatomical information from the X-ray CT without the use of external markers or internal landmarks. However the combination of a PET scanner with an MRI scanner has a number of advantages compared to a PET/CT: the anatomy of soft tissue can be visualized better in MRI images and MRI scans do no impose an additional radiation dose to the patient. On the other hand, combining PET with MRI technology is scientifically more challenging owing to the strong magnetic fields.

Within the CCC BrainPET-project, prototype PET systems for the study of the human brain will be developed. Compared with the existing scanners on the market, a device with a better resolution and higher sensitivity will be provided. The result is a significant improvement of the lesions detection and an enhancement of the functional characteristics research of this critical organ. CIEMAT (Madrid, Spain) has already acquired the necessary funding for the construction of a complete Brain-PET prototype based on the detection of photons using monolithic scintillator blocks and localization using computer learning algorithms developed by our VUB-group. These machines will not only be basic research prototypes but fully functional systems. After completion the Spanish prototype will be used for clinical studies in the Hospital Puerta de Hierro (Madrid, Spain).

The use of these block-systems in combination with several positioning algorithms is exactly the subject of this thesis. The prototype of CIEMAT will have 4 rings (12 cm axial field of view). The scanner will consist of rings made up of ~52 detector modules. This gives a scanner diameter of about 40 cm. The anticipated specifications are:

- a detector module consists of 2 layers of trapezium shaped ~20x20x10mm^3 LSO blocks
- each layer is read out by 2 APDs, one on each side
- a ring is made of 52 of these detector modules which correspond with a ring diameters of 40cm
- a spatial resolution of ~1.2mm FWHM which remains rather uniform over the field of view
- MRI insensitive technology. Although the CIEMAT system will not be integrated in an MRI scanner, the technology used will already be made MRI compatible.

Chapter 2

From positron emitter to PET image

PET imaging relies on the nature of the positron and positron decay. The theoretical physicist, P. Dirac, postulated the existence of positive electrons on the basis of the -Dirac equations-[9] and Einstein's theory of relativity. In 1932, experimental physicist, C.D. Anderson, proved the correctness of Dirac's prediction by observing experimentally cosmic rays. He discovered particles with the mass of electrons but their path in a strong magnetic field indicated a positive charge. He called these particles positrons, or positive electrons [10].

Functional imaging with positron-emitting isotopes was first proposed in the early 1950s as an imaging technique that could offer greater sensitivity than conventional nuclear medicine techniques with single photon-emitting isotopes. The SPECT collimator is eliminated and replaced by electronic collimation, a simultaneous detection of both back-to-back annihilation photons.

2.1 From positrons to annihilation photons, the tracer

A PET study involves injecting a compound, which is labeled with a positron-emitting radionuclide, into the patient. The radiolabeled compound is called a radiopharmaceutical, or more commonly, a tracer. The choice of tracer specifies the parameters which are examined. Examples are gene expression, glucose metabolism, protein syntheses, receptor affinity, … (Table 2.1).

Table 2.1 Examples of positron emitting radionuclides and there applications

Radionuclide	Tracer	Application	Half-life
^{18}F	FDG	Glucose metabolism	110min
	Fluoro-DOPA	Receptor affinity	
^{15}O	H_2O	Hemodynamic appl.	123sec
	O_2	Substrate metabolism	
^{11}C	^{11}C-cocaine	Drugs study	20.4min
	11C-ephedrine	Neurotransmitter appl.	
^{82}Rb	^{82}Rb	Myocardial studies	76sec
^{22}Na	e^+-emitter used for detector testing		2.602y

In radioactive decay by positron emission, a proton p in the nucleus is transformed into a neutron n and a positively charged electron e^+ (Equation (2.1)). The positively charged electron -or positron- and an electron neutrino v_e are ejected from the nucleus. Schematically, the process is

$$p \xrightarrow{\beta^+} n + e^+ + v_e \text{ (+energy)} \tag{2.1}$$

The radionuclides that decay via positron emission are proton-rich and move closer to their stable state while giving off a positive charge. Positrons are emitted with a kinetic energy greater than zero, and are slowed down through multiple Coulomb interactions in biological tissue [11]. Energy loss continues until the positron reaches thermal equilibrium with the surrounding medium. After coming to rest, the electron annihilates with an electron. More accurately, the positron and an electron momentarily form an atom called a positronium, which has the positron as it nucleus and a lifetime of about 10^{-10}sec [12]. The positron then combines with the negative electron in an annihilation reaction, in which their masses are converted into energy (Figure 2.1). The mass-energy equivalent of each particle is 511keV. This energy appears in the form of two 511keV annihilation photons, which leave the site of the annihilation in opposite directions. The back-to-back emission of annihilation photons is required for conservation of momentum for a stationary electron-positron pair. That's why positron emitters are useful in nuclear medicine because two photons are generated per nuclear decay event. The precise directional relationship between the annihilation photons permits the use of coincidence-counting techniques. The combined detection of two annihilation photons is called a coincidence, and the line joining the event locations is called a line-of-response (LOR). The activities measured along these LORs are close approximations to line integrals, which adequately sample the activity distribution. Several mathematical algorithms exist to reconstruct the tracer distribution from these line integrals. (see paragraph 2.5)

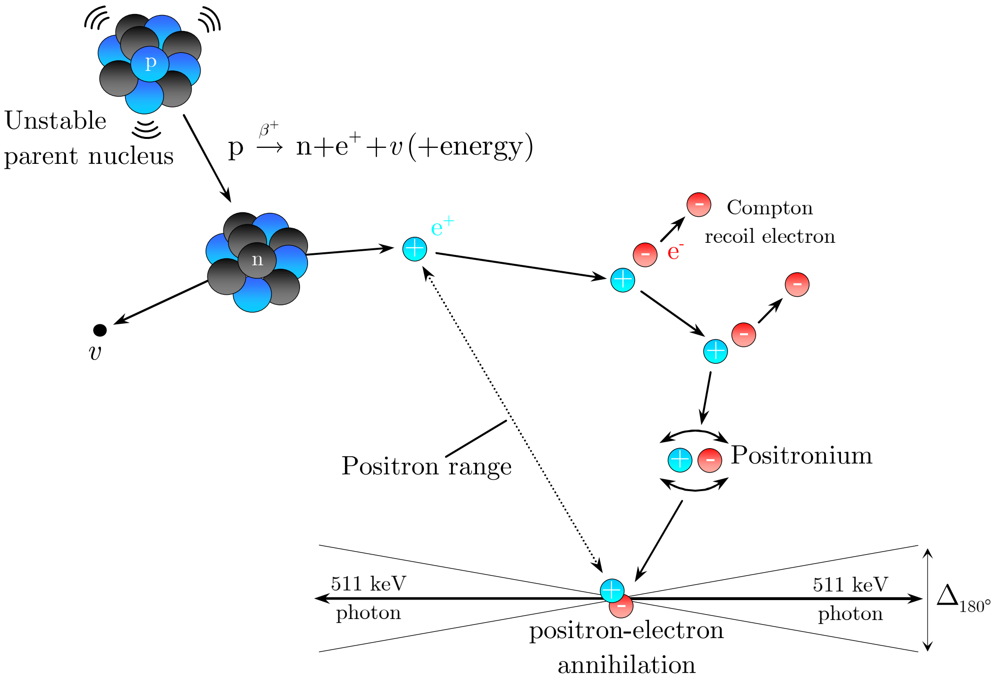


Figure 2.1: Schematic representation of positron emission and the physical aspects of positron-electron annihilation, i.e. positron range and the non-colinearity of the 511 keV annihilation photons (see paragraph 2.1.1)

2.1.1 The fundamental physical resolution limits of PET

Two factors relating to the basic physics of positron emission and annihilation process degrade the intrinsic spatial resolution of a PET system. An ideal tomograph should accurately measure the activity distribution in the body, i.e. the positron emission points. Actually, a tomograph can only detect the annihilation point. The distance the positron travels after being emitted from the nucleus and before annihilation is called the positron range. This range effect degrades the spatial resolution, introducing a blurring in the image. The positron range depends on the electron density of the medium (in denser material, the positron travels a shorter distance, reducing the uncertainty). Also the energy of the emitted positron, which depends on the particular nuclide, influences the positron range. The maximum energies E_{MAX} of the radionuclides used for nuclear medicine are in the range of 0.5 to 5MeV. However, positrons are emitted with a spectrum of energies. Only a small fraction has the full amount of energy available from the decay. The range distributions obtained per radionuclide are centered on the positron point source with long exponential tails rather than Gaussian shaped, thus making conventional full width at half maximum (FWHM) measures uncertain (Figure 2.2).

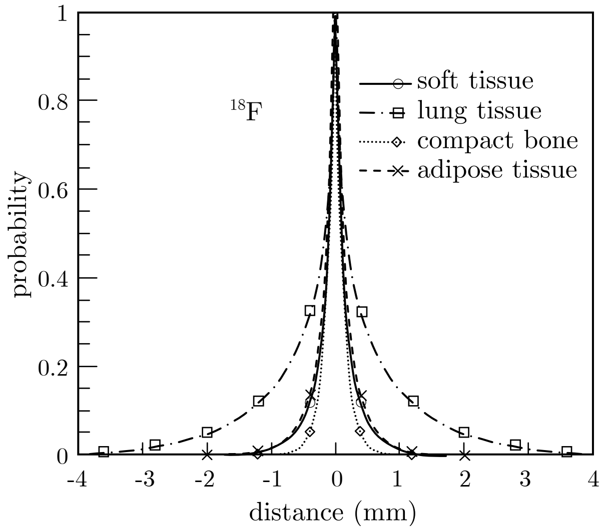


Figure 2.2: Projected annihilation point spread probability distributions normalized to 1.0 at zero distance. [13]

The full width at 20% of the maximum amplitude (FW20M) of the annihilation distributions yields more appropriate values for root mean square addition of spatial resolution loss components [13]. The contribution of this effect to the total spatial resolution is reported in Table 2.2 for several nuclides. A second factor involved in the degradation of PET's resolution is that the annihilation photons almost never are emitted at exactly 180° directions from each other. This effect, which is due to small residual momentum of the positron when it reaches the end of its range, is known as non-colinearity. In general, the annihilation occurs when the positron has reached thermal equilibrium with an electron not at rest. In fact, electrons constitute a statistical system of particles (Fermi gas) with an isotropic distribution of momentum in space.

Table 2.2 Spatial resolution loss due to the tissue- and radionuclide-dependence [13]

Radionuclide	β^+-E_{MAX}	Compact bone FW20M	Soft tissue FW20M	Adipose tissue FW20M	Lung tissue FW20M
^{18}F	0.635MeV	0.42mm	0.54mm	0.58mm	1.52mm
^{15}O	1.72MeV	1.08mm	1.87mm	2.31mm	5.30mm
^{11}C	0.960MeV	0.62mm	0.96mm	1.15mm	2.69mm
^{13}N	1.19MeV	0.72mm	1.26mm	1.55mm	3.50mm
^{82}Rb	3.35MeV	2.68mm	4.10mm	4.30mm	10.50mm

An approximated calculation of the deviation from the colinearity of the emitted photons can be made by just considering the thermal motion of the particles and the conservation of the Fermi momentum. The distribution of the angular deviation, projected on a plane, was measured to be Gaussian with FWHM ~ 0.5° [11]. The effect on the spatial distribution, expressed in terms of FWHM in the centre of a detector ring of diameter D can be parameterized as:

$$\Delta_{180^\circ} \approx \Delta\theta \,\mathrm{x}\, \frac{D}{4} = 0.0022 \,\mathrm{x}\, D \tag{2.2}$$

with both Δ_{180° and D measured in meters. By setting $\Delta\theta$ to a value of 0.5° (8.7mrad), this means a contribution to the spatial resolution of 2.2 mm FWHM per meter of detector separation. The range effect and the non-colinearity, as described above, are fundamental physical resolution limits in PET. Their effects degrade the spatial resolution, causing a blurring in the reconstructed image. This is especially evident for isotopes with a high E_{MAX} (i.e., a long positron range) and for whole-body PET systems, where the detector separation is greater than on dedicated brain or animal scanners.

2.2 Detection of annihilation photons by scintillators

2.2.1 Event types in annihilation coincidence detection

The colinearity of the two annihilation photons allows the so-called electronic collimation of the photons by coincidence detection. An annihilation event is assumed to have occurred when a pair of photons is recorded within a specified coincidence time window, which typically is around 10 nanoseconds. Unfortunately, three different types of coincidences detection can occur: true, random and scattered coincidences.

If the two photons detected within the time window arise from the same annihilation event and the volume connecting both detectors encloses the annihilation point, a true coincidence is measured (Figure 2.3.a). Only these true coincidences allow the measurement of the tracer distribution. On the other hand, the detection of random and scattered coincidences results in inaccurate LOR counting, generating noise in the reconstructed image.

Random coincidences occur when annihilation photons from two unrelated positron annihilation events are detected, within the coincidence time window and are recorded as a single coincidence event (Figure 2.3.b). This will yield wrong information about the position of the annihilation. The random coincidence counting rate in a detector pair is given by:

$$N_{\text{random}} = \Delta_{\text{window}} \text{ x } N_{\text{single,1}} \text{ x } N_{\text{single,2}} \qquad (2.3)$$

where $N_{\text{single,1}}$ and $N_{\text{single,2}}$ are the single event rates detected in each detector individually and Δ_{window} is the length of the coincidence time window. Since $N_{\text{single,1}}$ and $N_{\text{single,2}}$ are directly proportional to the tracer activity, the random rates will increase with the square of the activity and decrease as $1/r^2$ for a full detector ring with radius r, whereas the true coincidences rate only will increase linearly with the activity and decrease as 1/r. The amount of randoms can be reduced by reducing the coincidence time window Δ_{window} at the cost of efficiency. If the width of the Δ_{window} is chosen too small, valid data will be lost, and if the width is too large, the number of random coincidences will be increased with no increase in the number of true events. The required length for Δ_{window} depends on the timing properties of the detector.

Random coincidences occur more or less uniformly across the field of view (FOV) of the scanner, causing a loss of image contrast as well as inaccuracies in quantification of the activity within the object of study. The amount of randoms can be estimated directly from (2.3) and subtracted by monitoring the single count rate in the individual detectors, given the width of the coincidence time window Δ_{window}. Randoms can also be estimated by adding a second delayed coincidence circuit. To ensure that no true coincidence is recorded, the time window from one detector is delayed with a time much bigger than the time resolution of the system. Each coincidence measured in this delayed time window must be a random coincidence This does not require any knowledge of the coincidence time window Δ_{window}, but is not as statistically accurate as the direct method. If N_{true} is the number of true coincidence events recorded, N_{scatter} the number of scatter coincidences and N_{random} is the number of random coincidences subtracted from the total, the uncertainty in the remaining (true plus scatter) coincidences is [12]

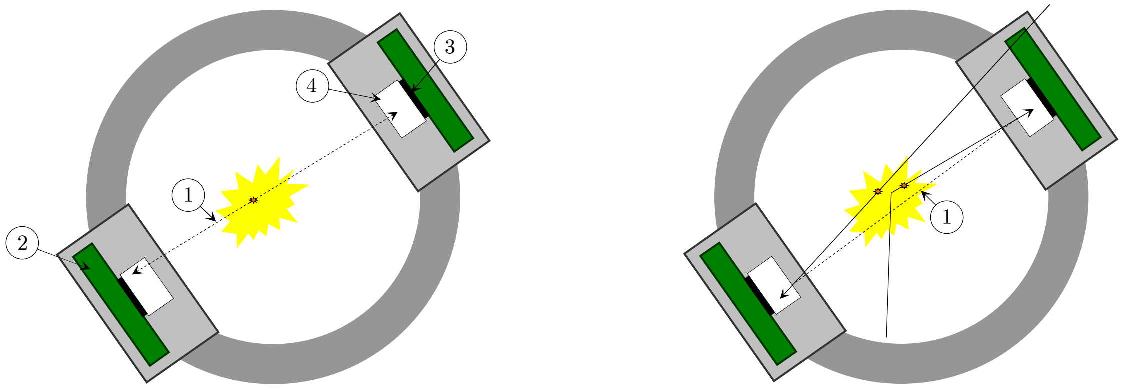
$$\sigma_{(\mathrm{N}_{\mathrm{true}}+\mathrm{N}_{\mathrm{scatter}})} = \sqrt{(\mathrm{N}_{\mathrm{true}}+\mathrm{N}_{\mathrm{scatter}}) + \left(2 \,\mathrm{x}\, \mathrm{N}_{\mathrm{random}}\right)} \tag{2.4}$$

Thus, even if accurate corrections can be made, the random coincidence rate should be minimized to avoid increasing the statistical noise level of the image. By using absorptive septa to restrict the active-containing region sampled by the coincidence detectors, the randoms-to-true count rate ratio can be reduced substantially. By using faster detectors with better timing properties, the randoms-to-true count rate ratio can be reduced further.

A second category of non-valid prompt coincidences are the scatter coincidences. These occur when one or both of the photons from an annihilation event are deviating from their original path by Compton scattering (Figure 2.3.c). The scattering event occurs within the patient, but it also can occur within components of the scanner and result in mispositioned events. The scatter count rate as well as the true count rate is proportional to the activity present and therefore the scatter-to-true count rate ratio is independent of the activity. Because true and scatter each result from single annihilation events, the scatter-to-true count rate ratio is also independent of the coincidence timing window.

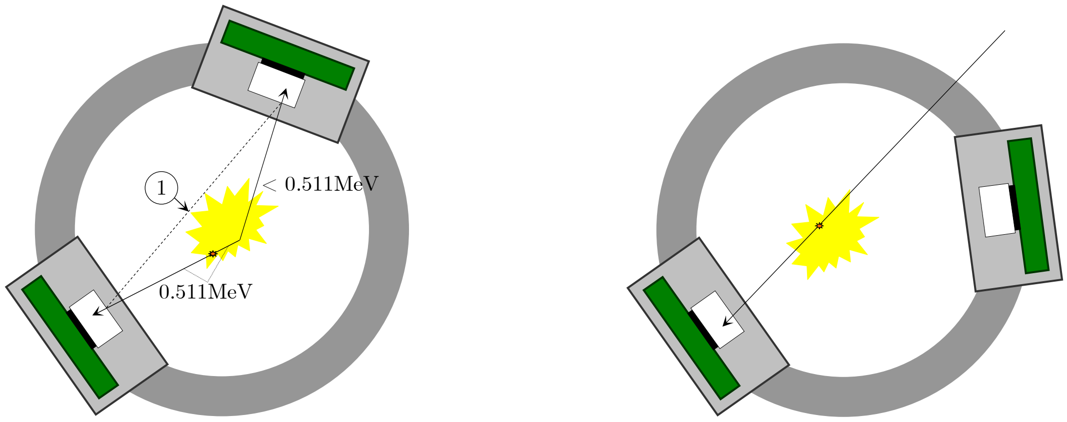
Like randoms, scatter results in generally diffuse background counts in reconstructed PET images, reducing contrast and distorting the relationship between image intensity and activity concentration. Scattered photons can, in principle, be identified from the energy lost in the scattering process and rejected by applying a simple energy threshold. However, as seen later, the energy resolution of current PET detectors is unable to accurately distinguish scattered from non-scattered photons above a certain energy threshold that may be as low as 350keV. Thus, in addition to a lower energy threshold, sophisticated scatter correction models have been developed to remove the residual scatter bias. The scatter background cannot be measured directly and must, instead, be estimated from the data. In a typical clinical imaging situation, even after applying an energy threshold, the fraction of the total events in the image that are scattered (SF) is 8%–10% in two dimensions (2D) and up to 45% or greater in three dimensions (3D) [14]. The 2D multiring PET scanners incorporate septa, lead or tungsten annular shields mounted between the detector rings. The purpose of the septa is to shield the detectors from photons that scattered out of the transverse plane. Only LORs in-plane or with small angles of incidence are measurable. The remaining LORs intersect the septa and the photons never reach the detectors. When the septa are retracted (3D PET), the number of active LORs is greatly increased, thereby improving the sensitivity.

In 2D PET, scatter correction is rather straightforward. Once the randoms correction has been applied, the peripheral “tails” in the projection-image count profiles, presumably due exclusively to scatter, are fit to a mathematical function and then subtracted (deconvolved) from the measured profile to yield scatter-corrected profiles for tomographic image reconstruction [15]. Scatter corrections for 3D PET include: dual energy window-based approaches; convolution/deconvolution- based approaches (analogous to the correction in 2D PET); direct estimation of scatter distribution (by Monte Carlo simulation of the imaging system); and iterative reconstruction based scatter compensation approaches (also employing Monte Carlo simulation) [15]. The Monte Carlo simulation and subtraction of scatter are now practical and have been implemented in commercial PET scanners A last category of detected event is the single event detection as seen in Figure 2.3.d. These events are rejected electronically because second annihilation photon was not detected in the time window. Unfortunately, they will add extra dead time during which the scanner is not sensitive to possible upcoming coincidence events. The way to minimize their effect is to reject them as early as possible in the detection process, since this type of events compromise a substantial part of all the events.



a: True coincidence

b: Random coincidence



c: Scatter coincidence

d: Single event

① = Measured LOR ③ = Photo detector

② = Preamplifier board ④ = Scintillation crystal

Figure 2.3: The four event types associated with PET. (a) True coincidence event, (b) random coincidence event and (c) scatter coincidences. Randoms and scatter coincidences yield incorrect LOR information and contribute to a relatively uniform background image that result in a loss of contrast. (d) The single event type detects only one of both annihilation photons and dominates the dead time losses of the scanner.

2.2.2 The interaction modes of photons in matter

After the tracer injection, the next step in the PET process is the detection of both annihilation photons. The detector should be able to detect and process several millions of photon events per second and sort out a small percentage of desired true coincidences for further processing. This means the detection material should be dense enough so that a significant fraction of the annihilation photons will interact with it. The detector materials that best meet these criteria are the inorganic crystalline materials called scintillators. In order to detect the annihilation photons, they must interact with matter in which a fraction or all of its energy is passed to an electron. When an photon passes through a scintillator, three interactions modes are probable: photoelectric effect, Compton scattering and pair production [16].

The process of pair production involves the transformation of a photon into an electron-positron pair. This effect only occurs for photons with energy higher than 1022keV and is thus not relevant in the detection of 511kev annihilation photons.

In photoelectric absorption the photon is fully absorbed by an atomic electron and all of its energy (minus the binding energy of the electron) is transferred to the electron that is ejected from the atom. When an annihilation photon undergoes a Compton interaction, the photon transfers only a part of its energy to the electron. Both the photon and electron then travel away from the scattering point with directions determined by the amount of energy transferred in the collision.

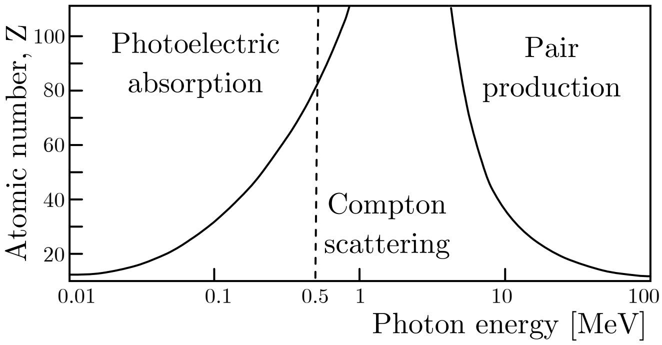


Figure 2.4: Predominating interaction versus photon energy for absorbers of different atomic numbers.

The relative probability of these three interaction mechanisms depends on the energy of the photon, but also on the atomic number Z_{eff} and the density ρ of the interaction medium (see Figure 2.4). The atomic cross-section for photoelectric effect is proportional to $\rho.Z_{eff}^{n}/E_{\gamma}^{m}$ where n and m are both function of the energy: n is about 4 at 100keV and gradually rises to 4.6 at 3MeV, whereas m decreases slowly from 3 at 100keV to 1 at 5MeV [11]. The ratio of the photoelectric cross-section over the total cross-section is defined as the photo fraction. This parameter is often used to characterize a scintillator. The cross-section for Compton scattering is proportional to $\rho.Z_{eff}/A$ where A is the mass number. Z_{eff}/A is almost constant, at 0.45 ± 0.05, for all elements except hydrogen. A high density ρ favors the interaction of photon in the crystal, whilst a higher Z_{eff} value increases the number of photoelectric occurrences with respect to Compton scattering. Therefore, high Z_{eff} crystals are to be preferred. The effective atomic number Z_{eff} is defined as [12]:

$$Z_{eff} = \sqrt[x]{\omega_1 Z_1^x + \omega_2 Z_2^x + \ldots + \omega_n Z_n^x} \quad \text{with } \omega_\iota = \frac{m_i.Z_i}{\sum_{i=1}^{n} m_i.Z_i} \tag{2.5}$$

where ω_i is a weighting factor proportional to the fractional number of electrons per gram for element i, m_i represents the number of atoms of element i present in the absorber. The power x is dependent on the energy of the photon. For photons in the 100-600kev range, x is typically between 3 and 3.5.

2.2.3 The scintillation process

The electron energy states of an isolated atom or molecule consists of a series of discrete energy levels defined by Schrödinger's equation. In an inorganic crystal lattice the outer electron energy levels are perturbed by mutual interactions between the atoms. The result is a broadening of the allowed energy levels into energy bands. In the lower band, called the valence band, the electrons are bound at lattice sites. In the upper band, called the conduction band, the electrons are free to migrate throughout the crystal. The energy band that separates the valence band and conduction band is called the forbidden band or band gap (Figure 2.5).

When a scintillator is irradiated a large number of electron-hole (e-h) pairs are formed. The electrons will be excited to the conduction band and can therefore migrate freely throughout the crystal. The electrons will eventually recombine with a hole in the valence band and de-excite back to the valence band, causing either a luminescent decay or a non-radioactive decay (quenching).

Inorganic scintillators are of two types. Activated scintillators such as NaI(Tl) or LSO(Ce) become radioluminescent through the introduction of a small amount of impurity dopant into the pure single host crystal. These impurities locally create activator levels. The second type is self-activated scintillators such as BGO, where the activator atoms are a major constituent of the crystal. The role of these activators is to produce quantum energy levels (ground and excited states) within the forbidden gap (see Figure 2.5).

When the electron-hole pairs are formed, the holes will migrate to the activator ground sites and ionize them, the electrons will be attracted to these charged activator sites and neutralize them. The activator site will have its own excited energy states within the forbidden gap. If the formed activator site is an excited configuration with an allowed transition to the ground state, it will rapidly de-excite with a high probability of emission of a light quantum or scintillation photon.

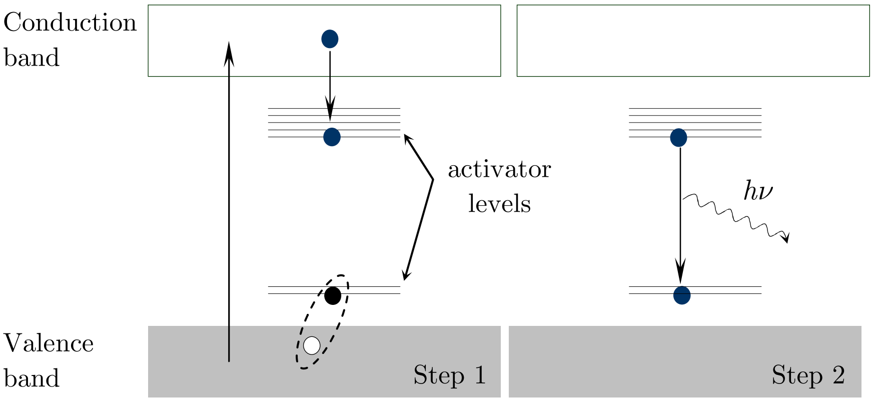


Figure 2.5: Radioluminescent decay with activator levels through introduction of small amount of impurities e.g. Lu_2SiO_5:Ce . denotes the ionization of the activator ground state. ↑ and ↓ indicate respectively the excitation and de-excitation of an electron.

2.2.4 Requirements for scintillators in PET scanners

An ideal scintillator would have a combination of several physical and scintillation properties as summarized in Table 2.3. As mentioned before, scintillator materials should have both a high effective atomic number Z_{eff} and high density ρ .

Parameters like decay time τ and light yield (LY) are also important physical properties of the absorber. When scintillators are struck by ionizing particle, they emit a small flash of light. The scintillation pulse can be described as:

$$N(t)=\frac{N_0}{\tau_d-\tau_r}\left(e^{-\frac{t}{\tau_d}} - e^{-\frac{t}{\tau_r}}\right) \tag{2.6}$$

with N_0 the total number of emitted scintillation photons, τ_r and τ_d respectively the rise time and the decay time of the scintillation process. Often the rise time τ_r is very short and can be neglected, such that N(t) can be approximated by a single exponential decay function. In many scintillators however, the decay cannot be described by a single exponential, but exhibit a more complex decay process which requires a more accurate description using a two component exponential:

$$N(t)=Ae^{-\frac{t}{\tau_{df}}}+Be^{-\frac{t}{\tau_{d_s}}} \tag{2.7}$$

where τ_{d_f} and τ_{d_s} denote respectively the fast and the slow decay constants since for many scintillators, one component is usually much faster than the other (i.e. BaF_2, $\tau_{d_f}=0.6$ns and $\tau_{d_s}=620$ns). A short decay time is highly recommended for PET applications. The rapid collection of the scintillation photons reduces the dead time of the scanner and increases the maximum data rate.

The light yield (LY) of a crystal refers to its efficiency for converting ionization energy into light photons. Usually, it is expressed as the number of photons N_{ph} produced per unit energy. A greater LY implies a better energy resolution as well as a more accurate spatial resolution. In both cases, a major source of noise in the measurement (leading to errors in positioning or energy) is the statistical fluctuation in the number of scintillation photons detected. These fluctuations are governed by Poisson counting statistics and reduce as $1/N_{ph}$. The LY and the decay time also affect the time resolution. A fast, bright scintillator will produce a signal with less timing variation than a slow, dim scintillator. This improves the time resolution of the scanner and allows a minimization of the coincidence time window Δ_{window} which results in a reduction of the random coincidences [12]. To understand this fact, consider a typical light pulse emitted by a scintillator, see Figure 2.6. The arrival time of a signal is taken to be the time at which it passes a preset threshold. Because of noise and statistical fluctuations, two identical signals will not always be triggered at the same point. This time variation is referred as time jitter. Beside the statistical fluctuations, the time jitter is also inversely proportional with the slope of the signal leading edge.

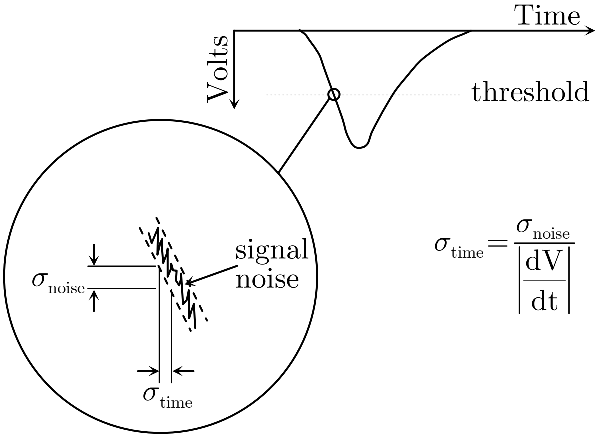


Figure 2.6: Illustration of the time jitter effect

The light yield, depends on the overall efficiency of the scintillation process η. This process can be characterized as the product of three factors [17]:

$$\eta=\beta.\mathrm{S}.\mathrm{Q} \tag{2.8}$$

where β is the number of e-h pairs produced in the conduction and valence bands respectively due to the absorption of radiation in the crystal, S is the transport efficiency of the e-h pairs to the luminescent centre (activator ion), and Q is the quantum efficiency for the luminescent centre (probability of a radiative transition to the ground state). Q is not unity because other competitive non-radiative processes can convert the energy to elastic vibrations, e.g. phonons, quantized modes of vibration occurring in a crystal lattice. The de-excitation of captured electron and holes, which result in radiationless transitions, is called quenching. This mechanism is determined by many factors such as other impurity atom dopants, structural defects during crystal growth, etc.

To collect as many light quanta as possible, each crystal is covered by reflective material at all surfaces except at the entrance window of the photo detector. Unfortunately, not all of the emitted fluorescence photons reach the sensitive area of the photo detector.

Table 2.3 Properties of ideal scintillation crystal for PET [17]

Crystal property	*Purpose*
High density ρ	High photon detection efficiency
High effective atomic number Z_{eff}	High photo fraction
Short decay time τ	Good coincidence timing, less random events, low dead time
High light yield (LY)	Good signal-to-noise ratio
Emission wavelength near 400 nm	Match spectral response of photo detector
Compatible refraction index	Good transmission at crystal-detector surface
Transparent at emission wavelength	Light can travel unimpeded to photo detector
Radiation hard	Stable crystal performance
Non hygroscopic	Simplifies packaging
Rugged	Allows fabrication of smaller crystal elements
Economic growth process	Reasonable cost

There are absorptive losses in the crystal due to self-absorption in the scintillator. Secondly, there are losses at the non-perfect reflective surfaces of the detector, with further losses of light quanta at the coupling between the crystal and the photo detector. Refraction causes light quanta to be lost or reflected back into the crystal for angles of incidence greater than the critical angle θ_c given by Snell's law $\theta_c = \sin^{-1}(n_1/n_0)$ with n_0 and n_1 the refractive index of respectively $medium_0$ and $medium_1$. Since the index of refraction of typical glass window is 1.5, scintillators which approximate this refraction index give a better coupling. Such matching is more closely achieved with some detector materials, e.g., LSO (n=1.82), than others, e.g., BGO (n=2.15).

2.2.5 Properties of commercial available scintillators

In the early stage of PET, NaI(Tl) was used for its very high light yield. But negative properties such as low density, low photo fraction and long absorption length make the crystal not optimal for photon detection. Beside, NaI(Tl) is hygroscopic, which requires that the crystals be hermetically sealed, usually in a thin aluminum container, to prevent the entrance of moisture. Moisture causes the crystal to develop yellow spots, which causes uneven light transmission.

Accordingly to these suboptimal properties, NaI(Tl) was gradually replaced by BGO that showed excellent stopping power thanks to its high density ρ and high effective atomic value Z_{eff} of 75. For example, the photoelectric cross-section at 511keV for BGO is 5.8 times higher than for NaI(Tl) [11]. The disadvantages of BGO, relative to other detector materials, include its low light yield, $\pm 20\%$ relative to NaI(Tl), as a result of which it exhibits an inferior energy resolution at 511keV of only 18% compared to 8% for NaI(Tl). Also the long decay time (τ=300ns) contribute to a further reduction in the energy resolution of BGO detectors. For scintillators with slow fluorescent emission characteristics, there is always a trade-off between energy resolution and dead time. Reducing the pulse integration time improves the dead time (and consequently the count rate performance), but results in a decrease in the amplitude of the signal, thereby reducing the energy resolution. Nevertheless, these properties combined with the widespread availability of BGO have rendered it the most widely used scintillator for commercial PET scanners.

With a density ρ of 7.4 g/cm^3, a photo fraction of 33%, a fast decay time of 42.3 ns and a fluorescence peak around 420nm, LSO is one of the most suitable scintillator materials for PET. Also the large brightness of 27000ph/MeV and the short absorption length of 1.13cm are strongly appreciated. A disadvantage of LSO is the presence of a naturally long-lived isotope of lutetium ^{176}Lu. It has been estimated that 2.59% of the lutetium in LSO is ^{176}Lu, which has a half-life of approximately 3.10^{10} years. This isotope emits β-particles with an endpoint energy of 565keV and γ-rays of 202keV (84% of disintegrations) and 307keV (93% of disintegrations). The expected background count rate due to ^{176}Lu is about 300 $\text{counts}/_{\text{s}}$ per cm^3 [18]. Since in PET coincidence counting is used, this background does not have a large effect on the spectrum. However, when the counting rate is low, background subtraction could become important. Although not hydroscopic, LSO fluoresces when exposed to ambient light, and therefore must be encased in a light-tight package. Nevertheless due to the best combination of properties for PET of any scintillator known today, LSO becomes the leading contender in this domain.

Table 2.4: Main characteristics of commonly used scintillators in PET [3, 10, 11]

	NaI(Tl)	BGO	LSO	LuAP
Chemical formula	NaI:Tl	$Bi_4Ge_30_{12}$	Lu_2SiO_5:Ce	$LuAlO_3$:Ce
Density ρ [g/cm^3]	3.67	7.13	7.4	8.34
Effective atomic number Z_{eff}	50	75	66	65
511keV Photo fraction [%]	18	41	33	32
Decay time τ [ns]	230	300	42	17
Light yield (LY)[ph/MeV]	38000	8200	27000	11400
Refractive index	1.85	2.15	1.82	1.97
Peak emission [nm]	415	480	420	365
Mean free path [cm]	2.91	1.04	1.14	1.05

2.3 *Photo detectors for the conversion of scintillation photons in an electrical signal*

The last step in the detection process prior to image reconstruction is the conversion of the light signal received from the scintillation crystal to a measurable electrical pulse by a photo detector. The working horse for the detection of photons is the photomultiplier tube (PMT) which has been a commercial product since 1936. It is an elaborated device but still, after 70 years, impressive improvements have recently been achieved [19]. PMTs however have a severe draw back. They are sensitive to magnetic fields which make them unworkable in combination with MRI. Also their large size along with the large dead space between individual PMT initiated the search for alternative devices. Photon sensitive semiconductor devices like avalanche photodiodes (APD) and recently Geiger-mode avalanche photodiodes (GM-APD) have been developed and have already replaced PMTs in many fields of research and will gain more ground in the near future. Geiger-mode avalanche photodiodes have an especially high potential because they have high gain and need no or only a simple amplifier and they can be produced with a standard and cost effective CMOS technique.

2.3.1 Photo multiplier tube

Already 73 years ago (1936) the photomultiplier tube (PMTs) was invented. This was only 8 years after Einstein proposed the concept of the work function.

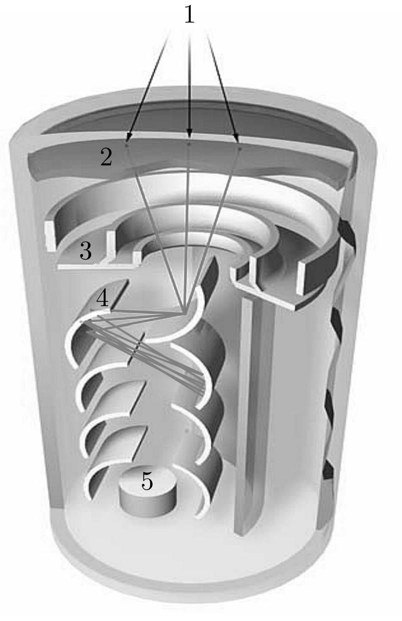


Figure 2.7: scheme of photomultiplier tube
1 incident light
2 semi-transparent photocathode
3 focusing electrodes
4 dynodes
5 anode
(figure from [20])

On commercial PET scanners, the light collected from the scintillator is converted into a measurable electrical voltage pulse by photomultiplier tubes (PMTs). PMTs convert the scintillation light into a very weak picoampere electrical signal at the photocathode. The charge produced on the photocathode as the result of interaction with incident light is not sufficient to operate electronic circuitry. For this reason the PMT is equipped with a charge multiplication system through an efficient low noise avalanche cascade process. They consist of a vacuum tube with a series of dynodes (electrodes) maintained under the control of a voltage divider (tube operating voltage between 1000V and 3000V). Photo electrons emitted from the photocathode are accelerated using an electric field produced by focusing electrodes to collide with the first dynode and produce emission of secondary electrons. From this dynode, the secondary electrons are accelerated in the direction of a second dynode from where an even greater number of electrons are ejected (Figure 2.7). This process repeats until high gain, up to 10^7, is achieved. The electric charge, collected from the anode is proportional to the light received by the PMT. This is useful in discriminating out lower-energy scattered coincidences in which one or both annihilation photons has undergone Compton scattering. The high gain, stability, and low noise of the PMT have rendered it the standard scintillation light amplifier.

2.3.2 Avalanche photodiode

Silicon PIN photodiode

An alternative to PMTs are the photon detectors based on the silicon photodiode. Semiconductor photodiodes are based on the p-n junction, which have a rather simple structure and is produced by standard semiconductor processes: boron diffusion on one side and phosphor diffusion on the other side of a high purity silicon wafer and at the ends contacts are made by aluminum deposition (Figure 2.8 right). By changing the thickness of the outer p-layer, substrate n-layer and bottom n^+-layer as well as the doping concentration, the characteristics of the photodiode can be controlled.

The formation of a pn-junction creates a special zone at the interface between the two materials, called the depletion zone. Due to the difference in concentration of electrons and holes between the two sides of the junction, holes from the p-region diffuse towards the n-region and similarly a diffusion of the electrons towards the p-region occurs until equilibrium is attained. This creates an electric field gradient across the junction which eventually halts the diffusion process leaving a region of immobile space charge.

Photodiodes are usually reverse biased with relatively large biasing voltages, because this lowers the diode capacitance C. The capacitance is proportional to the active area A and inversely proportional to the depletion zone width d. Since the depletion zone width is proportional to the product of the resistivity ρ of the substrate material and the reverse voltage V_R, the following relation can be made [21]:

$$C \propto A\left(\left(V_R+0.5\right)\rho\right)^{-\frac{1}{2}\text{to-}\frac{1}{3}} \tag{2.9}$$

However, there is a trade off between low capacitance and increased dark current (see below).

When light strikes the photodiode, electrons within the crystal structure become stimulated [21]. If the light energy is greater than the band gap energy E_g, electrons are pulled up into the conduction band, leaving holes in their place in the valence band (Figure 2.8 left). The holes have the properties of a net positive charge. Under the applied electric field in the depletion zone, the electrons drift towards the cathode and the holes drift towards the anode, constituting an electric photocurrent that can be measured.

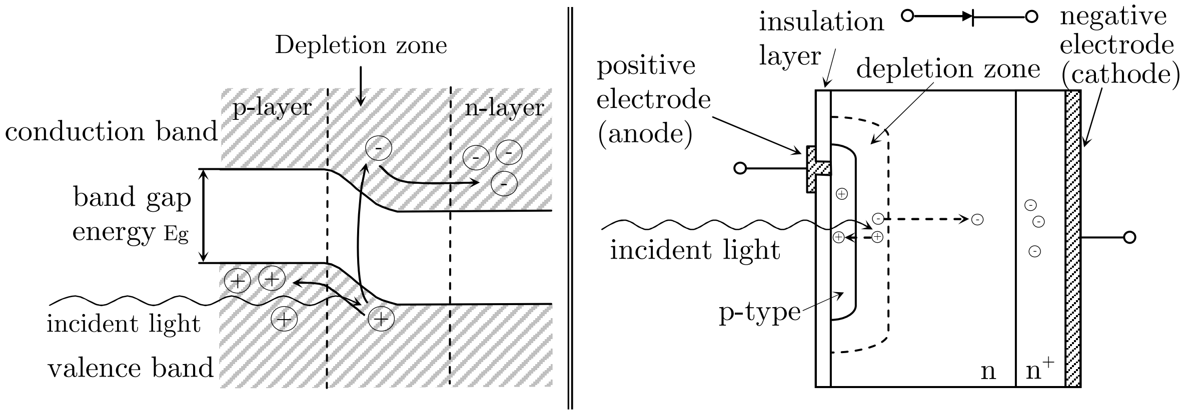


Figure 2.8: (left) Energy levels of a photodiode to illustrate the working principle of a photodiode. (right) Cross-sectional representation of a reverse biased photodiode with drifting electron and hole which are formed by an incident photon. (Based on Hamamatsu figures)

Avalanche photodiode

In common photodiodes, the electric field is not large enough to cause avalanche breakdown. Further development and improvement of semiconductor photo detectors has resulted in the avalanche photodiodes (APD). These devices produce an internal amplification of the induced charge. An APD is also a p-n junction diode, but compared with p-i-n diodes, APDs are operated under much higher reverse bias.

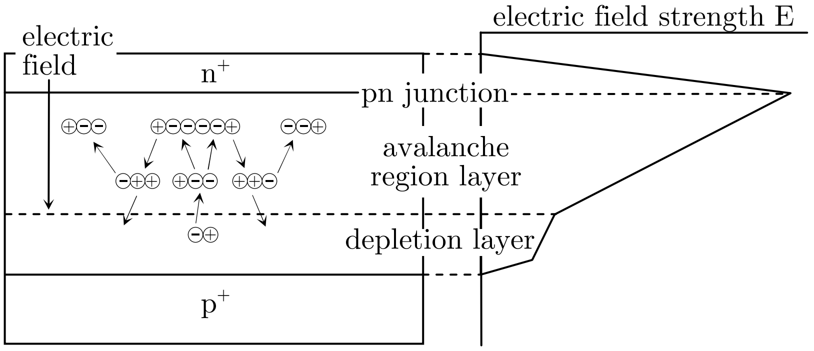


Figure 2.9: Schematic diagram of an avalanche process showing how the generated carriers are multiplied inside the APD. (Based on Hamamatsu figure)

The physics of operation is based on the impact ionization of Si atoms by the photoelectrons generated in the depletion region. These devices must be operated at a bias sufficient to accelerate the photoelectron to energies capable of ionizing a Si atom [11]. When electron-hole pairs are generated inside the depletion layer of an APD with a reverse voltage applied to the pn-junction, the electrons drift towards the n^+ side while the holes drift towards the p^+ side due to the electric field developed across the junction, see Figure 2.9. The drift speed of these electron hole pairs depends on the electric field strength. When the electric field is increased to a certain level, the carriers are more likely to collide with the crystal lattice so that their drift speed becomes saturated at an average speed. This phenomenon begins to occur when the electric field is typically in the vicinity of 10^4 V/cm, and the saturated drift speed at this point about 10^7 cm/s [22]. If the reverse voltage is increased even further, some of the carriers which escaped collision with the crystal lattice will have a great deal of energy. When these carriers undergo subsequent collisions with the crystal lattice, ionization in which electron-hole pairs are newly generated takes place. The electron-hole pairs then create additional pairs in a process just like a chain reaction. This phenomenon is referred to as avalanche multiplication of the photocurrent. It begins to take place when the electric field strength reaches typically $2x10^5$ V/cm. In general, the electric field is high enough for an avalanche only in a small part of the depletion region (avalanche region on Figure 2.9). Since the internal gain generally produces less noise than an equivalent external amplifier, APDs are useful for very low light conditions in which minimization of the noise is critical

APDs for PET applications have many advantages compared to conventional photomultiplier tubes because of their compactness, low-power consumption, ruggedness, insensitivity to magnetic field and high quantum efficiency (see below) for blue-UV light. APDs can also be produced as an array, greatly facilitating compact read-out systems. The limitations of APDs are that they have to operate at moderate multiplication between 50 and 200. A gain of 10^4 is possible but at values higher than a few hundreds, the environment (e.g. temperature and voltage supply) needs to be very stable [23]. Consequently, low noise amplifiers are needed.

Geiger-mode avalanche photodiode

When an APD is operated at a bias voltage higher than the breakdown voltage any photon or thermally liberated electron will start an avalanche which persists until the voltage is lowered actively or when the voltage drops on a properly chosen serial resistivity. The output signal is proportional to the overvoltage and the capacitance of the APD [23]. Clearly there is a long dead time of microseconds after each breakdown. This problem was overcome by subdividing the large area into many APD-cells and connect them all in parallel via an individual quenching resistor. The resulting device is called a Geiger-mode avalanche photodiode (GM-APD). The first devices of this type were developed in the late 1990s in Russia. GM-APDs have properties similar to PMTs and therefore some people call them Silicon Photomultiplier, SiPM. A photon impinging on one of the cells can create free carriers that give rise to a Geiger-type discharge. Since every cell is connected to the bias voltage via an individual integrated resistor, this discharge is quenched when the cell's voltage drops below the breakdown voltage. After a short recovery time, the time needed for recharging, the cell is ready to detect the next photon. The cell is a binary device since the signal from a cell always has the same shape and amplitude. The discharge currents from all cells are added on a common load resistor, therefore the output signal of a G-APD is the sum of the signals from all the cells firing at the same time. The high density of cells ($100–10000 mm^{-2}$) makes the response of a G-APD linear over a wide range of light intensities. Saturation effects do not set in until the flux of photons is comparable to the number of cells per unit area [24]. The gain is in the range of 10^5-10^7. There is no or at most a simple amplifier required, pickup noise is no more a concern and shielding is not needed [19]. Unfortunately, since thermally liberated electrons can trigger an avalanche, the APDs operated in Geiger mode have a high dark count rate of 1MHz per mm^2. The probability for accidental coincidences in PET would be high. This can be significantly reduced by operating them at low temperatures. A temperature of –50°C, easily achievable with Peltier elements, would be enough to reduce the dark count rate by 3 orders of magnitude [23]. However GM-APDs are still in a development stage, they already are very promising candidates for next generations PET modalities.

2.3.3 Properties of avalanche photodiodes

Quantum efficiency

The quantum efficiency (QE) η (or photon detection efficiency) of a radiation detector is defined as the probability that a photon incident on the detector will generate a charge carrier pair that may contribute to a current flow through the detector. If ζ represents the percentage of generated charge carriers that contributes to the current, we may write the quantum efficiency as follows [25]:

$$\eta = (1\text{-R}).\zeta.(1\text{-e}^{-\mu(\lambda).d_M}) \tag{2.10}$$

with R the reflectivity of the entrance face, $\mu(\lambda)$ the absorption coefficient of the detector and d_M the length of the conversion layer. The absorption coefficient $\mu(\lambda)$ is the inverse of the mean free path of the photon and is a strong function of the wavelength. In consequence, also the QE η is a strong function of the wavelength as shown in Figure 2.10 for the Hamamatsu S8550 APD.

At a wavelength of 420nm (LSO wavelength) and for an equal number of incident photons the number of created primary electron-hole pairs in an APD is 2 to 3 times larger than the created primary photo-electrons in the PMT. This is due to the high QE of APDs for the LSO scintillation light which is 60-70% compared to a QE of 20-30% for a PMT.

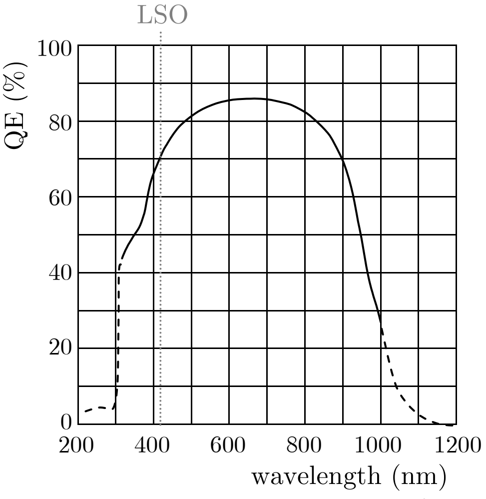


Figure 2.10: Wavelength dependence of the quantum efficiency η (QE) for the S8550 Hamamatsu APD. The QE for LSO scintillation light is typical 60-70%

Gain

APDs exhibit internal gain such that a single photon may produce hundreds of charge carrier pairs. As mentioned before, in an APD the field in the junction region is so high that the charge carriers can gain enough energy between collisions to produce further ionization and so produce an avalanche process which provides charge gain. In silicon the avalanche is mostly caused by electrons. In an APD the ionization coefficient for electrons will increase roughly in proportion to the electric field strength. Unfortunately the ionization rate has also inverse temperature dependence. As temperature increases, ionization rate decreases due to thermal agitation. The number of scattering collisions is increased and reduces the high-energy tail of the carrier energy distribution and accordingly the probability of ionization.

In silicon, the ionization coefficient α_h for hole-induced ionization is less than the coefficient α_e for electrons, but it is nonetheless significant and the ionization ratio $k = \alpha_h / \alpha_e$ increases toward unity with increasing field strength. The ionization ratio k represents the relative effectiveness of holes and electrons to ionize new carriers and is an important factor in determining the gain as well as the gain stability. The ideal APD would be one in which only one type of carrier is capable of having ionizing collisions, for instance electrons. Under the influence of the electric field within the depletion zone, the electrons drift toward the n^+ layer, gaining sufficient energy for ionization (see Figure 2.9). Thus as the multiplication process continues, the number of electrons would grow exponentially but remain finite. Using the ionization rates α_e, α_h and the depletion width d_M, calculation of the position-dependent multiplication factor M is given by [26]:

$$M(x) = \frac{\exp\left(-\int_0^x (\alpha_e(x') - \alpha_h(x'))\,dx'\right)}{1 - \int_{d_M} \alpha_e(x) . \exp\left(-\int_{d_M} (\alpha_e(y) - \alpha_h(y))dy\right) dx} \tag{2.11}$$

When α_e and α_h are about the same order of magnitude, both holes and electrons may ionize new carrier pairs, the multiplication exhibits positive feedback and the possibility of infinite gain. The voltage at which the APD enters in this mode is called the breakdown voltage. As electron and holes have opposite paths in the depletion zone, new electrons are created along the paths of ionizing holes at location such that these electrons will have time along their travel towards their respective electrode to ionize in their turn and create new electron–hole pairs from which holes will be accelerated in the opposite direction acquiring enough energy to create new electron-hole pairs, and so one, and so one. This leads to an incontrollable avalanche of charges causing the breakdown of the device. To eliminate all of the positive feedback, APD materials and bias voltages are chosen such that $k \ll 1$.

Noise

EXCESS NOISE FACTOR

One of the most important parameters of the APDs is their excess noise factor. The excess noise is due to the statistical nature of the multiplication process, which causes additional fluctuation of the measured signal. When the reverse voltage is constant, the gain becomes constant. However, the ionization of individual carriers is not uniform so that multiplication noise known as excess noise is added during the multiplication process [22]. The energy resolution obtained with APDs is mainly limited by three factors: the statistical contribution associated with the number of primary electron–hole pairs, the non-uniform avalanche gain process across the diode detection area and the noise of the preamplifier system [27]. The variance σ_{st}^2 associated to the statistical factor can be expressed as

$$\sigma_{\mathrm{st}}^2=\sigma_{\mathrm{N}}^2+\mathrm{N}\frac{\sigma_{\mathrm{M}}^2}{\mathrm{M}^2} \tag{2.12}$$

where N is the number of primary electron-hole pairs generated in the APD and σ_{N}^2 the correspondent variance, M is the APD gain and σ_{M}^2 the variance of the gain. In this equation, the first term describes the statistical fluctuation of the primary electrons and the second one the fluctuations of the avalanche gain in the APD. Defining the excess noise factor as

$$\mathrm{F}=1+\frac{\sigma_{\mathrm{M}}^2}{\mathrm{M}^2} \tag{2.13}$$

Equation (2.12) can be rewritten as

$$\sigma_{\mathrm{st}}^2=\sigma_{\mathrm{N}}^2+\mathrm{N}(\mathrm{F}\text{-}1) \tag{2.14}$$

In particular, for light pulse detection, the variance of primary electrons is described by Poisson statistics consequently $\sigma_{\mathrm{N}}^2=\mathrm{N}$. This way, the statistical limitation of the energy resolution is given by

$$\sigma_{\mathrm{st}}^2 = NF \tag{2.15}$$

The rms statistical noise σ_{st} for an APD is thus given by $\sqrt{\mathrm{NF}}$, whereas for PMT readout it is given by $\sqrt{\mathrm{N}'}$, where N' is the number of primary photoelectrons generated in the PMT. For the same number of photons incident, N will be 2 to 3 times larger than N' since the differences in quantum efficiency (see above). Unfortunately, this positive effect is cancelled by the excess noise caused by the multiplication process in the APD. Therefore the statistical SNR will be comparable reading signals with APDs or with PMTs [28].

DARK CURRENT

The dark current $\mathrm{I_D}$ is a current which arises when an inverse voltage is applied on the APD even in the absence of incident photons. It can be divided in two branches: a surface current $\mathrm{I_{Ds}}$ and a current $\mathrm{I_{Db}}$ generated in the bulk region of APD. The surface leakage current $\mathrm{I_{Ds}}$ flowing between the pn-junction and the Si-layer does not go into the avalanche region and can be described by a resistor connected in parallel to the APD. The surface current $\mathrm{I_{Ds}}$ increases linearly with the applied bias. On the other hand, the bulk current $\mathrm{I_{Db}}$ arises from electron-hole pairs created by thermal excitation inside the depletion zone. $\mathrm{I_{Db}}$ is fully amplified with the avalanche gain M:

$$\mathrm{I_D}=\mathrm{I_{Ds}}+\mathrm{MI_{Db}} \tag{2.16}$$

In an APD, the thermal generation probability P for excitation of an electron-hole pair across the band gap ε varies as $\mathrm{P}=\mathrm{e}^{-\varepsilon/(\mathrm{k_b T})}$ [29]. Accordingly, $\mathrm{I_{Db}}$ and also the dark current I_D will increase with increasing temperature T. The random arrival of these charges which cause fluctuations in an otherwise steady current is called shot noise.

EQUIVALENT NOISE CHARGE

As mentioned before, APDs have a moderated gain of 10^2-10^3, compared to the PMT gain of the order of 10^6. To overcome this lack, a charge sensitive preamplifier is used. For the best results, the preamplifier should be located as close as physical possible to the APD. This maximizes the electronic SNR by amplifying the signal before additional noise or signal distortion can occur. It is more common to express the electronic noise, generated by the APD and the preamplifier, as the equivalent noise charge (ENC). The ENC corresponds to the number of primary charge carriers required at the entrance of the APD to achieve a SNR of 1 at the output of the preamplifier. The ENC at the input of the APD, when it is followed by a charge sensitive preamplifier and a shaping stage, is given by [30]:

$$\mathrm{ENC}^2 = \left[2q_e I_{Db} F \tau a_1 + \frac{8k_B T}{3M^2} \left(\frac{(C_D + C_t)^2}{g_m} \right) \frac{1}{\tau} a_2 \right] \tag{2.17}$$

with:

C_D, C_t: detector and preamplifier capacitance

F: excess noise factor

g_m: transconductance of the preamplifier ($g_m = \Delta I_{out} / \Delta V_{in}$)

I_{Db}: bulk dark current

M: APD multiplication gain

T: absolute temperature (K)

τ : shaping time

q_e, k_B: electron charge, Boltzmann constant

$a_1 = \frac{1}{2\tau h_{max}^2} \int_{-\infty}^{+\infty} h^2(t)\,dt$ and $a_2 = \frac{\tau}{2h_{max}^2} \int_{-\infty}^{+\infty} \left(\frac{dh(t)}{dt} \right)^2 dt$ with h(t) the impulse response function depending on the shaping time. Both a_1 and a_2 are dimensionless coefficients of the order of 1 for the commonly used shaping functions.

The first term in Equation (2.17) is proportional with the shot noise due to the dark current I_D, an APD characteristic. Generally, the dark current is modeled as being composed of two part, see Equation (2.16). But since the contribution of the surface current I_{Ds} is not multiplied, it is much smaller than the bulk dark current and negligible in the expression for the noise. The second term of Equation (2.17) represents the amplifier noise. It is assumed that this noise is entirely due to the thermal noise in the conduction channel of the first FET transistor in the preamplifier, and that all other contributions to the amplifier noise are negligible [30].

Since both terms in ENC^2 are respectively proportional to τ and $1/\tau$, an adequate preamplifier with optimal shaping time and capacitance can be chosen to minimize the sum.

2.4 PET performances

An extensive series of parameters have been developed over the years to characterize PET scanner performance, and detailed data acquisition and analysis protocols have been published for this purpose. The discussion below, however, addresses only several key parameters: sensitivity, energy- and spatial resolution.

2.4.1 Sensitivity

System sensitivity is an important parameter since it determines the image quality per unit scan time. Sensitivity refers to the relationship between recorded true coincidences and the activity of a positron-emitting source. The two major elements of sensitivity are the absorption efficiency of the detector material and its solid angle of coverage of the imaged object. The efficiency of the material is based mainly on the density and thickness of the scintillation crystal (see paragraph 2.2.4). Notice that the detector sensitivity is proportional to the square of the individual detector efficiency for coincidence detection in PET. Thus, a difference in efficiency of detection of 2 leads to a global sensitivity difference of 4. The second component of sensitivity is the geometric factor, i.e. the solid angle available to the radioactive sources, and this is dependent on the size, distance and number of detector arrays surrounding the imaging volume.
In order to provide a physical measure to benchmark advances in PET camera performance, the concept of noise equivalent count rate (NECR) was introduced. The NECR is defined as:

$$\mathrm{NECR}=\frac{\mathrm{T}^2}{\mathrm{T+R+S}} \tag{2.18}$$

where T is the true coincidence rate, R is the random coincidence rate, and S is the scatter coincidence rate (see paragraph 2.2.1). An NECR curve provides a statistic which shows the dependence of the increase in trues on increasing activity in the field of view, yet is penalized by scatter and randoms, which provide inaccurate positional information about the source distribution and therefore degrade image contrast and quality. Because the random event rates increase as the square of the activity (Equation (2.3)) but the true events increase linearly with the activity, at some point there will be more randoms than trues. Thus, eventually the NECR will decrease as activity is increased. This is reflected in Figure 2.11 which shows a NECR curve of an uniform water phantom for a rectangular scanner geometry [31].

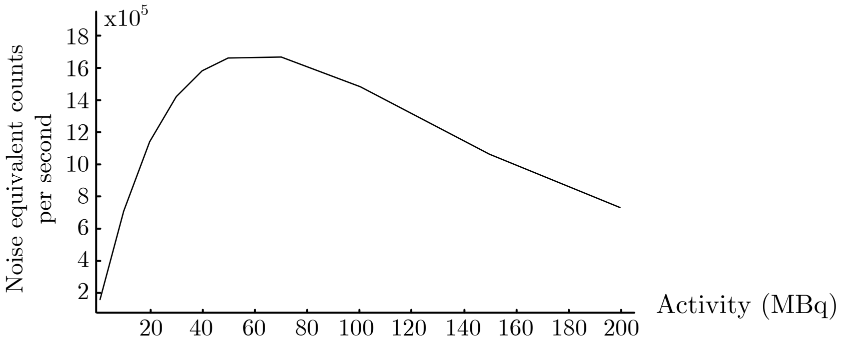


Figure 2.11: NECR curve for a simulated scanner geometry based on rectangular detector blocks with an uniform water phantom

2.4.2 Energy resolution

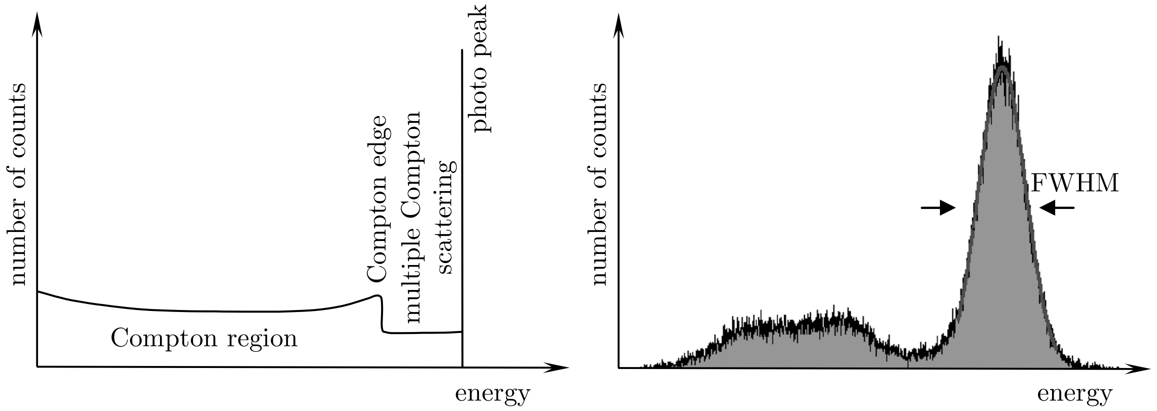


Figure 2.12: (left) Theoretical energy spectrum. (Right) Typical energy spectrum obtained by a LSO block coupled to an APD

Figure 2.12 (left) shows an ideal spectrum of a 511keV gamma ray source placed in front of a detector. With energy less than 1022keV, pair production interactions do not occur. As already explained in 2.2.2, the principle interactions with the detector will be photoelectric absorption and Compton scattering. Most of the photoelectric interactions result in full absorption of the gamma ray energy $\mathrm{E}_{511\mathrm{keV}}$. With an ideal detector, this would produce a single narrow line in the energy spectrum, known as the photo peak. In Compton scattering, only a part of the gamma ray energy is transferred to the detector, via the Compton recoil electron. If the scattered gamma ray also is absorbed in the detector, the event produces a pulse in the photo peak, whereas if the scattered gamma ray escapes, the energy deposition in the detector is less than $\mathrm{E}_{511\mathrm{keV}}$. The amount of energy transferred to the recoil electron in Compton scattering is defined by [12]

$$\mathrm{E}_{\mathrm{re}} = \mathrm{E}_{511\mathrm{keV}} \text{-} \mathrm{E}_{\mathrm{sc}} \text{ with } \mathrm{E}_{\mathrm{sc}} = \frac{\mathrm{E}_{511\mathrm{keV}}}{1 + \frac{\mathrm{E}_{511\mathrm{keV}}}{0.511}(1-\cos\theta)} \tag{2.19}$$

where E_{re} and E_{sc} are respectively the energy transferred to the recoil electron and the scattered photon energy. The amount of energy transferred to the recoil electron ranges from nearly zero for $\theta \approx 0$ degrees up to some maximum value $\mathrm{E}_{\mathrm{re}}^{\mathrm{max}}$ that occurs in 180-degree backscattering events. This distribution of pulse amplitudes ranging from nearly zero up to the maximum energy $\mathrm{E}_{\mathrm{re}}^{\mathrm{max}}$ forms the Compton region in the ideal energy spectrum. The sharp edge in the spectrum at $\mathrm{E}_{\mathrm{re}}^{\mathrm{max}}$ is called the Compton edge. Another possibility is that a Compton scattered gamma ray may experience additional Compton scatter interactions in the detector. Multiple Compton scattering events produce the distribution of pulses with amplitudes in the valley between the Compton edge and the photo peak [12].

In practice, the actual spectrum Figure 2.12 (right) measured with a LSO block coupled to an APD is quite different from the ideal one shown in Figure 2.12 (left). For instance, the spectrum is spread out. The photo peak is not a sharp line but a somewhat broadened peak and the Compton edge is rounded. This is caused by the imperfect energy resolution of the detector. The Compton region is cut due to the use of a hardware threshold. The relative variance σ_E^2 in the pulse height distribution of the photo peak measured with the scintillator coupled to the APD can be expressed as [32]:

$$\sigma_E^2 = \sigma_{intr}^2 + \sigma_{st}^2 + \sigma_{noise}^2 \tag{2.20}$$

with σ_{intr}^2 the intrinsic energy resolution of the scintillator, σ_{st}^2 the statistical contribution, and σ_{noise}^2 the electronic noise. The intrinsic resolution of a crystal is mainly associated with the non-proportional response of the scintillator and various effects, such as inhomogeneities in the scintillator causing local variations of the light output and nonuniform reflectivity of the crystal covering [25]. The statistical fluctuation on the emitted scintillation photons contributes to σ_{st}^2. It also depends on the statistical fluctuation of the APD gain, affected by the excess noise factor F (see Equation (2.15)). The electronic noise is not negligible in case an APD is used for the readout. It depends mainly on the dark current of the APD and the noise of the preamplifier (see Paragraph 2.3.3).

The energy resolution is determined as the FWHM of the photo peak divided by the peak centroid, and is given by

$$\frac{\Delta E}{E} = 2.35\frac{\sigma_E}{E} \tag{2.21}$$

2.4.3 Spatial resolution

As mentioned before (paragraph 2.1.1), positron range and non-colinearity set an ultimate resolution limit that can be achieved in PET due to the physics of the positron decay. In addition to this limit, the design and properties of the detector used in the PET scanner, and the system geometry, will also contribute to the final image resolution.

Intrinsic resolution

The intrinsic spatial resolution depicts the resolution of a detector pair in the system. It is usually given in terms of the coincidence response function (CRF) for a source located at the midpoint between the two detectors. The CRF is the count rate profile obtained by passing a point source or a narrow line source transversally between the two detector modules, while recording the LORs during a fixed time Δt for each position (Figure 2.13). The intrinsic spatial resolution is then referred to in terms of FWHM and FWTM of the CRF (source midway both detectors).

For discrete, pixilated detectors, the CRF is determined by the detector width w, degrading from w/2 midway between opposed coincidence detectors to w at the face of either detector. At midplane, the CRF is a triangle and becomes trapezoidal in shape, eventually becoming a rectangle of width w as the source is moved closer to one of the detectors (Figure 2.13).

For continuous detectors with an empirically determined detector resolution w_{det}, the CRF degrades from $w_{det}/\sqrt{2}$ midway between the opposed detectors to w_{det} at the face of either detector.

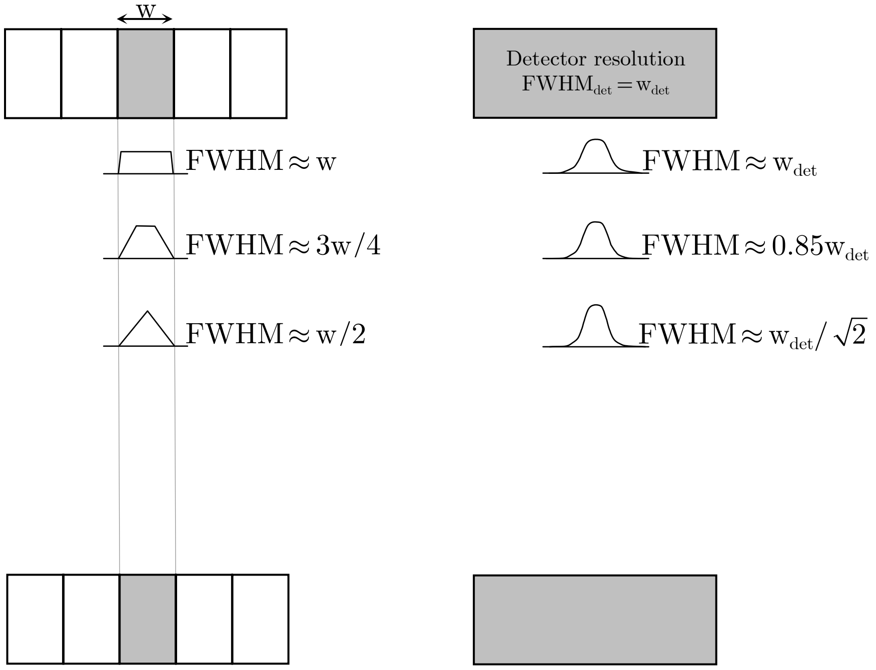


Figure 2.13: Coincidence response function (CRF) as a point source moved between two detectors. For pixilated detectors (left) the response function is determined by the width of the detector element, w. The intrinsic spatial resolution for a pixilated detector is w/2 FWHM (right) For continuous detectors CRF is determined by the intrinsic detector resolution $\mathrm{w_{det}}$. The intrinsic spatial resolution is $\mathrm{w_{det}}/\sqrt{2}$ FWHM.

System image resolution

Since the annihilation photons have a relative long mean free path in scintillators, a non-negligible percentage of them do not interact at all with the scintillation crystal. In order to partially overcome this problem, the crystal thickness should be increased. However, the thick scintillation detectors (typically 2–3 cm) used for PET imaging lead to another geometric effect that degrades the spatial resolution. This effect, which is referred to as detector parallax or the depth of interaction effect, is caused by the fact that the annihilation photons can interact at any depth in the scintillator material (Figure 2.14). Although the effect can also occur in the axial direction in the scanner, the primary effect is an increased uncertainty on the LORs along the radial direction plane [12]. The loss in resolution is dependent on the crystal density, the crystal length and the diameter of the scanner. For a fixed diameter and a given scintillation material, resolution can be improved by shortening the crystal. However, this results in a loss of sensitive material. There is thus a trade-off between sensitivity and spatial resolution.

Different methods have been presented to minimize the parallax effect. Depth-of-interaction information can be obtained by the “phoswich” technique. Instead of a single 20mm long crystal, two 10mm crystals are used with two different time constants. The analysis of the signal shape should allow distinguishing in which of the two longitudinal segments the interaction occurred. An alternative technique is to collect the crystal light in the front and back faces and to use the relative amplitude of the two signals to estimate the longitudinal coordinate of the interaction point.

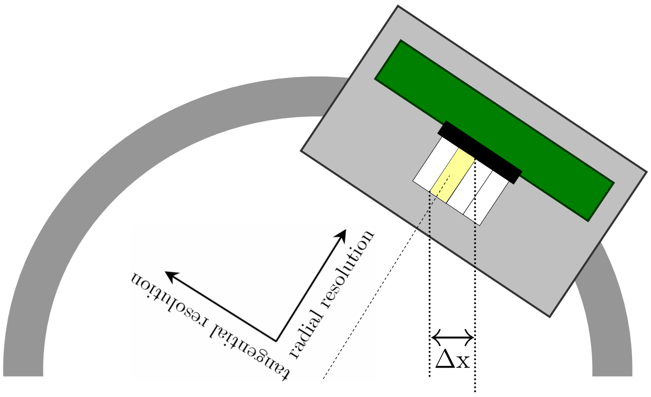


Figure 2.14: Illustration of the parallax effect. Because the depth at which the gamma rays interact within the scintillation crystal is unknown, the annihilation event for a pair of photons recorded in coincidence could have occurred anywhere within Δx .

The final system resolution for a particular system design is a convolution of all the resolution response functions, including the positron range, photon non-colinearity, geometric factors, intrinsic spatial resolution, and physical factors. A reasonable accurate estimate of the reconstructed image resolution (FWHM) Γ near the center of a PET camera can be made by using the formula [26]:

$$\Gamma = a.\sqrt{\left(\mathrm{FWHM}_{\mathrm{intr}}\right)^2 + (0.0022\mathrm{D})^2 + \mathrm{P}_{\mathrm{eff}}^2 + b^2} \qquad (2.22)$$

with $\mathrm{FWHM}_{\mathrm{intr}}$ the intrinsic spatial resolution at the centre, D the detector ring diameter and $\mathrm{P}_{\mathrm{eff}}$ the effective positron range. Factor b is due to the crystal decoding process. In case of a one-to-one coupling, b will be zero. For block detectors, b has to be determined experimentally. The factor $1.1 < a < 1.3$ relies on the reconstruction algorithm used. For example, using FBP with a ramp filter (see below), factor a is typically 1.25 [33].

Actually, the full description of the spatial resolution involves two components: a transaxial component in the planes perpendicular to the scanner axis and an axial component parallel to the axis (slice thickness). In addition, for off-center positions, transaxial resolution is usually given in terms of a radial (along the radius) and a tangential (perpendicularly to the radius) component (Figure 2.14). Transaxial spatial resolution can be determined by imaging a point source. To give a measure of resolution uniformity, this can be done for different locations within the FOV. The axial resolution of a PET system is defined by the intrinsic resolution of the detectors and is determined by passing a point or line source axially through the image plane.

2.5 Image reconstruction

Each pair of detectors in a scanner defines a possible emission path, represented by a line-of-response (LOR). Over the course of a PET scan, the system is counting how many times each pair of detectors is hit in coincidence. One way to represent the raw data is to group the LORs in a histogram depending of their position x_r and angle ϕ (see Figure 2.15). This type of histogram is called a sinogram. In consequence, each bin in the sinogram represents a LOR and the amount of each bin is the number of coincidences measured throughout this LOR.

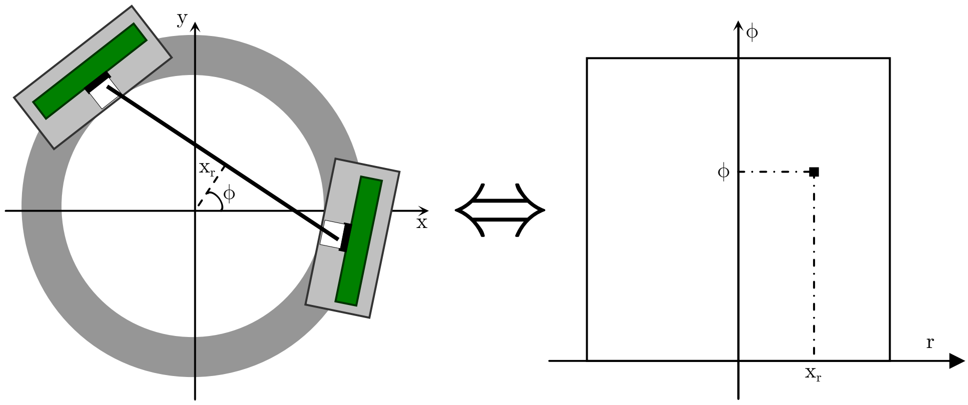


Figure 2.15: A line-of-response defines a point is the sinogram

Since the number of coincidences detected for a LOR is proportional to the concentration of the tracer along the entire LOR, this number is also proportional to the sum or integral of the tracer concentration along the LOR. It is conventional to say that the sinogram contains projections $p_\phi(x_r)$ of the tracer concentration.

The inverse problem of determining the unknown tracer distribution f(x,y) inside the scanner's FOV from the measured projections $p_\phi(x_r)$ is called image reconstruction. To do so, there are two widely used methods. Analytic reconstruction methods and iterative reconstruction methods minimizing the error between the measured and estimated data.

Analytic reconstruction methods are based on the central-section theorem and can be stated as follows, if an object described by f(x,y) has a two-dimensional Fourier transform $F(\nu_x,\nu_y) = \Im_{2D}\{f(x,y)\}$, and a projection, $p_\phi(x_r)$, then the one-dimensional Fourier transform of the projection $P_\phi(\nu_{x_r}) = \Im_{1D}\{p_\phi(x_r)\}$ is identical to a section of the two-dimensional Fourier transform through the origin at an angle ϕ, such that $P_\phi(\nu_{x_r}) = F(\nu_{x_r},\nu_{y_r})\big|_{\nu_{y_r}}$. An illustration of the central-section theorem is given in Figure 2.16. In consequence, if the projections in all direction $(0<\phi<\pi)$ are measured, the unknown tracer distribution can be calculated.

A frequently used inverse method is the filtered-back projection (FBP) and is defined as:

$$f(x,y) = \int_0^\pi d\phi \int_{-\infty}^{\infty} d\nu_{x_r} \left|\nu_{x_r}\right| e^{2\pi i \nu_{x_r}(x_r)} P_\phi(\nu_{x_r}) \tag{2.23}$$

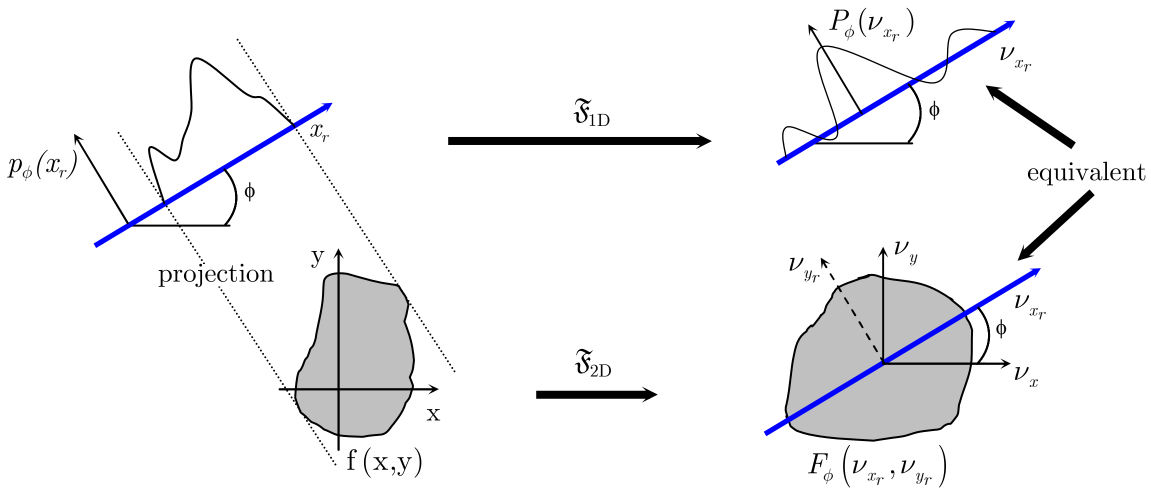


Figure 2.16: Illustration of the two-dimensional central-section theorem. The Fourier transformation of a 1D projection is equivalent with the 1D section of the two-dimensional Fourier transform through the origin at an angle ϕ

Equation (2.23) allows to determine the tracer distribution f(x,y) from the measured projections $p_\phi(x_r)$ in the following manner:

1. Filtering

The measured projection data $p_\phi(x_r)$ is Fourier transformed to $P_\phi(\nu_{x_r})$ and in Fourier space multiplied by a filter defined by $H\left(\nu_{x_r}\right)=\left|\nu_{x_r}\right|$. The result is inverse transformed to give the filtered projection

$$\begin{aligned} p^*_\phi(x_r) &= \mathfrak{F}^{-1}_{1D}\left\{H(\nu_x)P_\phi\left(\nu_{x_r}\right)\right\} \\ &= \int_{-\infty}^{\infty} d\nu_{x_r}\, e^{2\pi i\nu_{x_r}x_r}\, H\left(\nu_{x_r}\right)P_\phi\left(\nu_{x_r}\right) \end{aligned} \tag{2.24}$$

2. Back projection

Back projection of the filtered projection $p^*_\phi(x_r)$ provides an estimate $\hat{f}(x,y)$ for the tracer distribution f(x,y)

$$f(x,y) \approx \hat{f}(x,y)=\int_0^\pi d\phi\ p^*_\phi(x_r) \tag{2.25}$$

The term $H\left(\nu_{x_r}\right)=\left|\nu_{x_r}\right|$ in the filtering step is known as the ramp filter and can be seen as compensating for the increased density of sampling points of the Fourier transform as $\left|\nu_{x_r}\right| \to 0$. Due to the presence of a high-frequency statistical noise, it is usually desirable to roll-off the high frequency values of $H\left(\nu_{x_r}\right)$ with a windowing function $W\left(\nu_{x_r}\right)$ such that $H\left(\nu_{x_r}\right)=W(\nu_{x_r})\left|\nu_{x_r}\right|$. $W\left(\nu_{x_r}\right)$ is a function which avoids that the noise takes the upper hand by slowly suppressing higher frequencies and cut-off at certain frequency ν_c. A typical example of $W\left(\nu_{x_r}\right)$ is the Hamming window defined as:

$$W\left(\nu_{x_r}\right)=\begin{cases}\alpha+(1-\alpha)\cos\left(\dfrac{\pi\nu}{\nu_c}\right) & for\ |\nu|<\nu_c\\ 0 & for\ |\nu|\geq\nu_c\end{cases} \tag{2.26}$$

where ν_c is a cut-off frequency usually chosen such that for $\left|\nu_{x_r}\right|\geq\nu_c$, the signal power $\left|P_\phi(\nu_{x_r})\right|^2$ is less than the statistical noise power, and α controls the smoothness of the roll-off, a typical value of $\alpha=0.5$. Making $W\left(\nu_{x_r}\right)$ roll off to zero more quickly causes the reconstructed images to be less noisy and also smoother. The reduction in noise generally improves image quality, but the smoothing effect results in a loss of resolution. The shape of $W\left(\nu_{x_r}\right)$ can be modified to adjust the trade-off between noise and resolution, and for the Hamming window, this is done by adjusting α and ν_c.

The transform method discussed in the section above is based on inversion formulas that are converted to a discrete form. The inversion formulas are derived assuming noiseless and continuously sampled projections, which is not the case in practice. An alternative approach is to start with a model that takes into account the noise and the discrete nature of the projection data and reconstructed image. This is the approach taken by iterative algorithms, so named since they converge to an estimate of f_j, which is a sampled version of f(x,y), by a series of successively more accurate estimates $\hat{f}_j^{(k)}$, where k is the number of the iteration. The advantages of this technique versus analytic reconstruction are a more accurate acquisition model, including the ability to incorporate a priori information or constraints on the estimated images. In theory all effects can be corrected for. But iterative methods are much slower than the analytical methods due to the slow convergence of the algorithm and high computational demands.

Most of the algorithms start with the data collection process modeled as a set of equations of the form [34]

$$p_i=\sum_{j=1}^{N}L_{ij}.f_j \quad (i=1,...,M) \tag{2.27}$$

where N is the number of image pixels, M is the number of LOR measured, p_i are the measured coincidences in the i-th projection bin, f_j is the activity in the j-th pixel of the image matrix describing the activity distribution and L_{ij} is an MxN projection matrix describing the probability of an emission according to projection i to be detected in pixel j. The goal then, is to recover an estimate of f_j from the measured projection data p_i. This means solving the set of linear M equations for the N unknown f_j's. Typical values are on the order of M ~ 50000 and N ~ 15000, so that direct methods of inverting Equation (2.27) are not practical, or even possible in some cases. Iterative algorithms use successive approximations to ever more accurately estimate the f_j terms of Equation (2.27). The general procedure for iterative algorithms is as follows:

1. Start with an initial estimate, $\hat{f}_j^{(k=0)}$, usually uniform for all j
2. Calculate $\hat{p}_i^{(k)}=\sum_{j=1}^{N}L_{ij}.\hat{f}_j^{(k)}$
3. The calculated $\hat{p}_i^{(k)}$ is compared to p_i

4. If $\hat{p}_i^{(k)}$ is close enough to p_i then $\hat{f}_j^{*}=\hat{f}_j^{(k)}$ and the process is stopped
5. The result of this comparison is used to compute correction factors that are then used to generate $\hat{f}_j^{(k+1)}$ from $\hat{f}_j^{(k)}$
6. Step 2 is repeated with $k \rightarrow k+1$

The intent is that the series of estimates ($\hat{f}_j^{(0)}$, $\hat{f}_j^{(1)}$, $\hat{f}_j^{(2)}$, …, $\hat{f}_j^{(k)}$) converges to a final estimate, $\hat{f}_j^{*}$, such that $\hat{p}_i^{*}=\sum_{j=1}^{N} L_{ij}.\hat{f}_j^{*}$ is the closest possible to p_i.

Different iterative algorithms differ in the way they define the object parameterization (pixels, voxels, …), system model (scanner geometry, detector response, attenuation and scatter correction, …), statistical model (algebraic reconstruction technique (ART), maximum likelihood expectation maximization (ML-EM), least square (LS)), cost function (distance between $\hat{p}_i^{(k)}$ and p_i and how they calculate the correction factors used to update the image estimates), etc.

The most widely used iterative reconstruction approaches are based on maximum likelihood (ML) methods [34]. Likelihood, is a general statistical measure that is maximized when the difference between the measured and estimated projections is minimized. The expectation-maximization (EM) algorithm is an iterative algorithm that maximizes likelihood under a Poisson data model. It implicitly treats the projection data as having a Poisson distribution determined by the counting statistics in each projection bin and thus takes into account the statistical noise in the data.

The ML-EM algorithm for PET can be written as [34]:

$$\hat{f}_j^{(k+1)}=\frac{\hat{f}_j^{(k)}}{\sum_i L_{ij}}\sum_i \frac{L_{ij}p_i}{\sum_j L_{ij}\hat{f}_j^{(k)}} \tag{2.28}$$

This equation shows how the image pixel intensity $\hat{f}_j^{k+1}$ at iteration k+1 is calculated based on the estimated image pixel intensity $\hat{f}_j^{k}$ at iteration k and the measured projection counts p_i. With increasing iterations, ML-EM algorithms have been found to improve image quality and generally exhibit less noise than FBP, as well as improve lesion detection [35]. Notice that when the estimated projection data exactly equal the measured projection data p_i that (substituting from Equation (2.27)) $\hat{f}_j^{k+1} = \hat{f}_j^{k}$ and the image does not change any more. However, this never occurs in practice because of noise in the data and inevitable errors and approximations in L_{ij}. In addition, a further difficulty of iterative methods is to decide when an acceptable solution has been reached. As expected, with increasing iterations, the likelihood increases. However the image began to degenerate, becoming increasingly noisy [36]. It is therefore important to terminate the reconstruction before this degeneration begins, even though the likelihood function may not be a maximum. The origin of this effect lies in the fact that the measurements are Poisson random variables with noise, and that insistence on an exact fit to the data will result in an image dominated by this noise.

Chapter 3

Study of spatial resolution in block detectors

The image quality in high-resolution PET is mainly determined by two parameters: signal to noise ratio (SNR) and spatial resolution. In most of the existing scanners the 511keV annihilation photons emitted by the administered tracers are detected by matrices of small individual scintillation crystals. The accuracy to localize the impinging photons depends obviously on the size of the individual crystals. Hence there is a tendency to make these crystals smaller and smaller to maximize the spatial resolution. However this also has some detrimental effects:

- Increase fraction of dead space in the detector module (lower packing fraction) due to the presence of the material to optically separate the crystals. This reduces the sensitivity.
- Degraded time resolution and energy resolution due to the worse collection of the scintillation light.
- To maintain the better resolution outside the central part of the field of view, extra measures have to be taken to determine the depth of interaction (DOI) in order to minimize the parallax error of the 511 keV photons in the crystal matrix.

The result is a worse signal to noise ratio (SNR) in the image given the same dose of radio tracer and scan time. To maintain the same image quality while improving the spatial resolution it is hence essential to improve the sensitivity. Indeed, the squared signal-to-noise ratio SNR^2 is inversely proportional to the fourth power of the image pixel size [37]. Hence, when the image resolution is improved by a factor of 2, one needs 16 times more coincidence events to achieve the same variance in the reconstructed image.

The VUB research group has developed a new detector concept which allows to simultaneously enhance the spatial resolution and sensitivity of PET detectors compared to the current systems. To achieve this goal, we make use of undivided -or monolithic- LSO scintillator blocks. Most important is the larger amount of sensitive detector material for a similar volume due to the absence of inter-pixel material for optical separation. This results in an increased sensitivity. In addition the blocks can be made trapezoidal to create a gapless ring and further enhance the detection efficiency significantly. Another advantage is the enhanced energy resolution. The light piping effect seen in matrices of small individual scintillating crystals, results in a lower light output due to multiple reflections on the side surfaces. In a monolithic LSO block energy resolutions of 11.5% have been achieved compared to 20–25% in the classical pixilated scanner designs [37, 38].

The position information of the impinging 511 keV photon within the scintillator block is embedded in the shape of the scintillation light distribution. This principle of light spreading allows that the scintillator block can be larger than the sensitive area of the photo detector, avoiding the dead space due to the packaging of the photo detector. This again enhances the sensitivity.

Parallax correction in pixilated detector systems is usually based on a measurement of the interaction depth of the photon, either in a discrete way (e.g. pulse shape analysis in phoswich configurations or pixel-encoding schemes using multiple crystal layers which are displaced relative to each other) or in a continuous way (e.g. using the ratio of the signal measured on the top and bottom side of the scintillator).

Hence, the accuracy on the estimated true interaction position in the pixilated scintillator depends on the accuracy of the estimated position of the photon in the plane of the photo detector and on the accuracy of the DOI measurement.

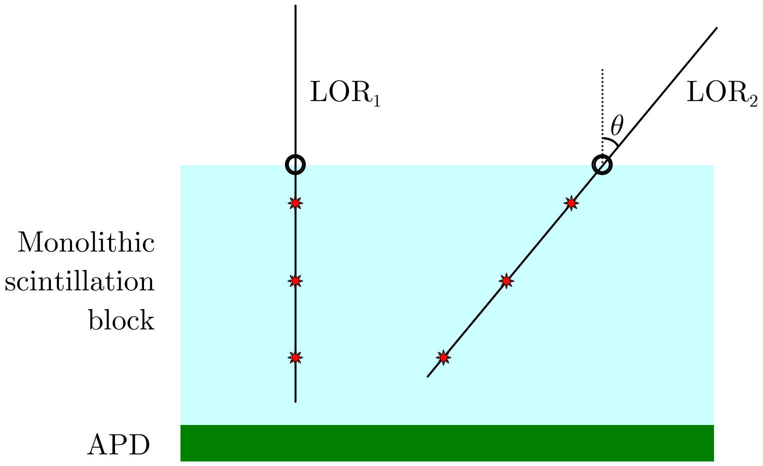


Figure 3.1: Illustration of a parallax free positioning determination. The lines-of-response needed for image reconstruction are determined from the incidence 2D photon position ○ and the incidence angle θ instead of photon interaction position ✱.

To overcome the parallax effect in our block approach, we will evaluate the photon incidence position instead of the interaction position within the block (Figure 3.1). Together with the incidence angle, we can define a correct LOR independent of the interaction depth. The advantage is that there is no need to measure the interaction depth separately, and hence, there is also no need for a separate DOI calibration of the detector module. However the relation between the measured scintillation light distribution and the incidence position depends on the incidence angle θ. In a scanner geometry this incidence angle θ can easily be estimated from the two detectors firing in coincidence, i.e. estimate the incidence angles from the line joining the centers of the two detector modules. If necessary, this method can be refined by using the calculated entry points to derive an improved estimate of the angle of incidence, and repeating the position estimation using this improved estimate.

3.1 Front-end detector designs

3.1.1 The S8550 avalanche photo diode

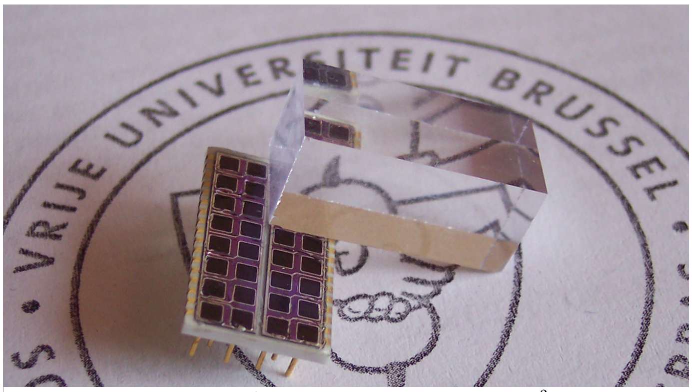


Figure 3.2: S8550 Si APD and a 20x10x10mm^3 LSO block

All detector configurations used in this study are based on the S8550 Si APD array [39] from Hamamatsu. The S8550 APD is designed for short wavelength detection, featuring low noise and low terminal capacitance. It also offers uniform gain and small cross-talk between each element [39]. The array consists of 32 APD pixels arranged in a 4x8 scheme (Figure 3.2). Two silicon wafer parts, each of them comprising 16 APDs, are housed in a ceramic package with a 0.5mm thick epoxy window completely covering the sensitive area. The common cathodes and the individual anodes of the 32 diodes are connected by a plastic grid array at the backside of the carrier plate [40]. The specifications of the device are listed in Table 3.1.

Table 3.1: Technical data of the S8550 Hamamatsu APD array [39, 40]

Parameter	Value
Device size	19.5x11.2mm^2 (218.4mm^2)
Total active area	81.92mm^2
Number of pixels	32
Element size	1.6x1.6mm^2 (2.56mm^2)
Element pitch	2.3 mm
Bias voltage range	100-400V
Operating gain range M	1-100
Spectral response range	320-1000nm
Quantum efficiency at 420nm	60-70%
Dark current per pixel (M=50)	10nA
Terminal capacitance per pixel (M=50)	10-15pF

The S8550 APD has a “reverse type” structure [41]. Consider the schematic cross-section shown in Figure 3.3. These APDs are designed to have a high-field multiplying layer restricted to a narrow portion of the depletion layer, and a relatively low electric field in the rest of the depletion layer [42]. This design permits the use of a wide depletion layer, which reduces the capacitance per unit area, while still permitting operation at quite low bias voltages (<500V). The multiplying p-n junction is located about 4µm below the p^+-layer so that only primary photo-electrons generated by short-wavelength light (i.e. strongly absorbed) are fully multiplied. Whereas for a pair generated within the wide drift region behind the multiplying region only the hole enters the multiplying region where it undergoes a much reduced amplification. This also means that, most of the dark current undergoes only hole multiplication, and so the contribution to the noise is reduced significantly [43]. The gain of the electrons and holes are related by [43]:

$$M_h = 1 + k_1 \left(M_e - 1 \right) \tag{3.1}$$

where k_1 is a weighted ratio of the ionization coefficients (close to, but slightly greater than $k = \alpha_h / \alpha_e$). k_1 is generally below 0.05, so that, for an electron gain of 100, the hole multiplication is at least a factor 20 less.

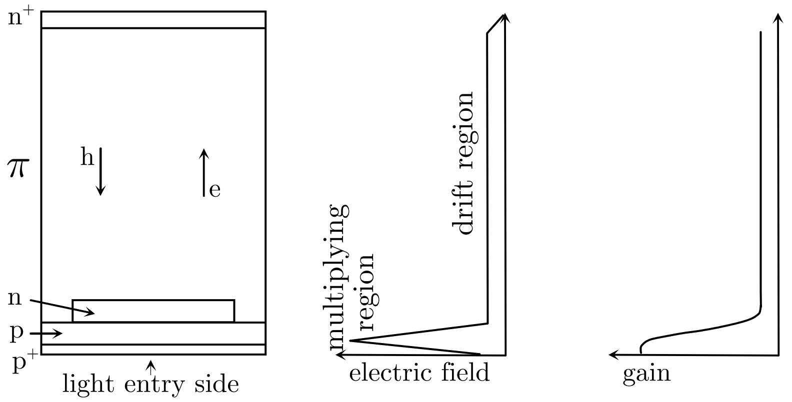


Figure 3.3: Schematic cross-section of a reverse type APD. The electric profile and the gain profile of the APD are also given.

3.1.2 Monolithic block geometries

Monolithic scintillator blocks provide an efficient way to significantly increase the sensitivity of high-resolution PET systems compared to the classically used matrix of small scintillation crystals coupled individually to the pixels of the APD array (Figure 3.4.c).

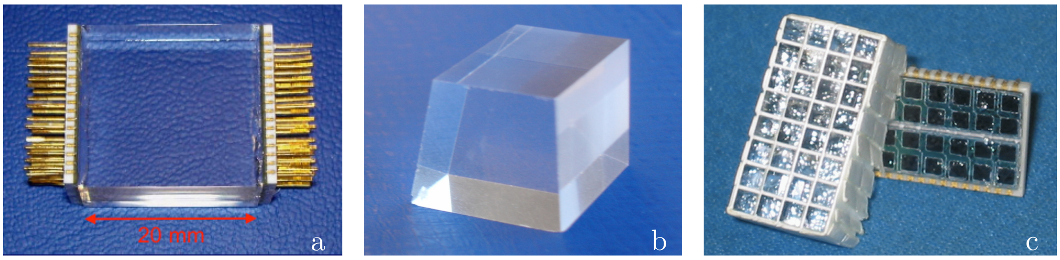


Figure 3.4: (a) a $20x10x20mm^3$ LSO block with a double APD readout, (b) a trapezoidal $20x15.4(11.5)x20mm^3$ LSO block which allows the construction of a gapless detector ring (c) a 4x8 crystal matrix of $2x2x10mm^3$ LSO pixels

Three different continuous LSO block geometries were used in the evaluation study: a rectangular $20x10x10mm^3$ block, a thicker rectangular $20x10x20mm^3$ block and a

trapezoidal 20x15.4(11.5)x20mm^3 block. The increase in sensitivity is mainly due to the absence of optical separation material between the individual scintillation pixels. In addition, the fact that the scintillation light distribution contains information about the impact coordinates removes the constraint to restrict the scintillator to the sensitive areas of the photo detector, i.e. the exit surface of the LSO block can also cover the packaging of the photo detector used and hence increase the sensitive detection volume. A further increase in sensitivity is possible when trapezoidal scintillator blocks are used to avoid the wedge-shaped gaps when the scintillator blocks are placed in a ring
Simulations, using the Monte-Carlo code GATE, also confirm the increase in sensitivity of a PET scanner using monolithic LSO blocks instead of small LSO pixels [31, 37]. The simulated scanner consisted of four rings with a diameter of 12.4 cm. Each ring had a thickness of 2 cm and contained 32 APDs. Three crystal configurations were simulated: a 4x8 matrix of 2x2x20mm^3 pixels with a pitch of 2.3mm, a 20x11.5x20mm^3 rectangular LSO block and a 20x15.4(11.5)x20mm^3 trapezoidal LSO block. The energy window was set at 250–750 keV.

Table 3.2 shows the sensitivity (%) for a point source at the center of the scanner. Replacing the individual LSO pixels with a rectangular monolithic block doubles the sensitivity of the scanner. Another 25% gain in sensitivity is achieved when the rectangular monolithic blocks are replaced by the trapezoidal blocks.

Table 3.2 Active crystal volume and simulated sensitivity for a 4 ring PET-scanner

	crystal matrix 8x4 matrix 2x2x20mm^3	rectangular block 20x11.5x20mm^3	trapezoidal block 20x15.4(11.5)x20mm^3
Volume (cc)	2560	4600	5380
Sensitivity (%)	8	17	21.5

3.2 *A bench set-up for the acquisition of training data*

When an annihilation photon interacts with the scintillator block, its entry point on the front surface of the crystal is estimated from the resulting distribution of the scintillation light measured by the APD array(s). In this study the entry point is extracted from the distribution using one of the three machine learning algorithm that will be introduced in Paragraph 3.3. The training process of these machine learning algorithms requires a reference set of samples that link the measured scintillation light distribution to the known photon incidence position. In order to acquire such training data a bench set-up was build up (Figure 3.5). The first tests were made at the VUB with a 20x10x10mm^3 LSO block. In parallel, the construction of a second set-up and the acquisition of training data were done by collaborators of the Delft University of Technology in The Netherlands.

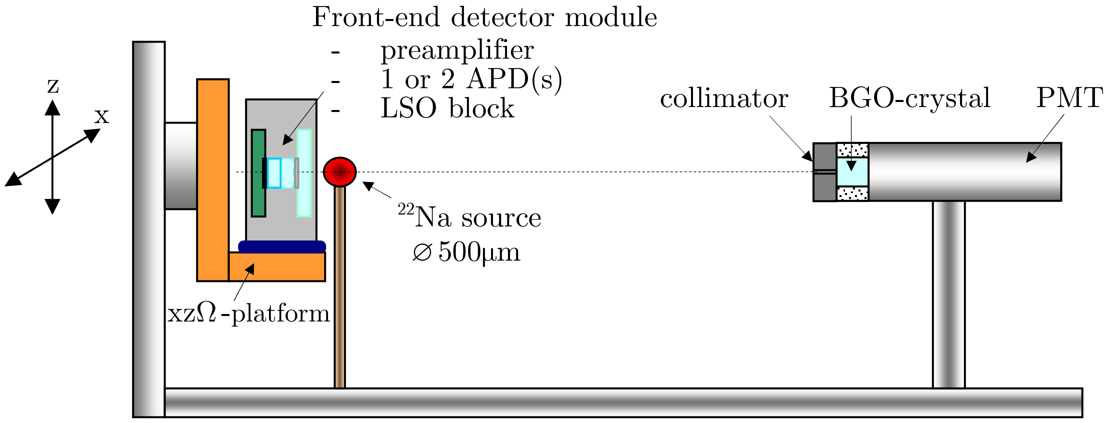


Figure 3.5: Schematic representation of the bench set-up

Training data for three different geometries were acquired on the Delft bench set-up, using a rectangular 20x10x10mm^3 block, a thicker rectangular 20x10x20mm^3 block and a trapezoidal 20x15.4(11.5)x20mm^3 block (see Figure 3.2 and Figure 3.4). The LSO blocks were wrapped with Teflon on five sides (10mm thick block) or four sides (20mm thick blocks) to maximize the light output. The scintillation light distribution emerging at the bottom of the 10mm thick block is sampled by one S8550 Hamamatsu APD array. In case of the 20mm thick blocks, the top and bottom surface were read out by an APD. Each APD array is mounted on a printed circuit board (PCB) holding 32 Cremat CR-110 preamplifiers [44]. The crystal, APDs and PCBs are placed in a light tight, cubic aluminum box, and form together the front-end detector module. A beam of 511 keV annihilation photons emitted by a ∅0.5mm ^{22}Na source irradiates the detector module. The beam is defined by a second detector in coincidence with the APD detector, consisting of a 50mm thick BGO crystal coupled to a PMT, with a ∅5.0mm Pb-collimator (Pb thickness 60mm). The detector can be scanned through the beam using a motorized xzΩ-stage, controlled by PC. The pulses from the preamplifiers are fed into two/four 16-channel spectroscopy amplifiers (CAEN N568B) through twisted pair flat cables. For each channel, these amplifiers have a fast output branch, consisting of a fixed gain single differentiation stage with a time constant of 100 ns, and a slow branch, providing a semi-Gaussian pulse shape with an adjustable gain and a shaping time of 0.1 µs , 0.2 µs , 1.0 µs or 3.0 µs . The signals from the slow outputs are transferred to one/two 32-channel, 12-bit, peak sensing ADCs (CAEN V785). Time pick-off on the APD signals is performed on the analog sum of the fast outputs using a constant fraction discriminator (CFD, Ortec 934). The PMT anode signal is passed through an Ortec 474 timing filter amplifier. Time pick-off on the resulting PMT signal is again performed by a CFD (Ortec 934). A coincidence occurs when both CFD signals trigger within a time window of 2x10ns. In all experiments, the APD arrays were kept at room temperature, and the shaping time of the slow branch of the amplifiers was set at 0.2 µs .

The resulting ~1mm FWHM electronically collimated 511 keV photon beam was stepped over the surface in 250 µm intervals along the long (20mm) central axis of the monolithic LSO blocks. At each beam position a number of events were measured and the 32/64 APD pixel values were stored together with the incidence position (i.e. the center of the known beam position).

The width of the photon beam and the position of the LSO block relative to the photon beam are determined by moving the photon beam over the edge of the block in 250 µm steps. Per position the number of events detected in a fixed period is recorded. The obtained sigmoid count rate profile is the result of a convolution between the photon beam profile and a step function representing the edge of the LSO block. Assuming the beam profile is Gaussian, the measured count rate profile is fitted with

$$\mathrm{g(x)} \otimes \mathrm{h(x)} = \int_{-\infty}^{\infty} \mathrm{h(y)g(x\text{-}y)dy} = \int_{\mathrm{m}}^{\infty} \mathrm{A}e^{-\frac{1}{2}\left(\frac{\mathrm{x\text{-}y}}{\sigma}\right)^{2}} \mathrm{dy} \tag{3.2}$$

with m the position of the crystal edge, σ the width of the photon beam, and A a scaling factor. Due to background events in the measurement, a constant term was added to the fit function. Figure 3.6 (left) shows the result of such a fit on a beam profile measurement. When the obtained sigmoid is halfway between its lower level and upper level, the center of the photon beam is exactly on the edge of the block.

The edge of the block can also be estimated from symmetry considerations. Given the count rate profile of the whole block (see Figure 3.6 (right)) and knowing the length of the crystal block (20mm), both edges of the block should have the same count rate. Accordingly, the two positions, 20mm apart, having the same rate correspond to the edges of the block.

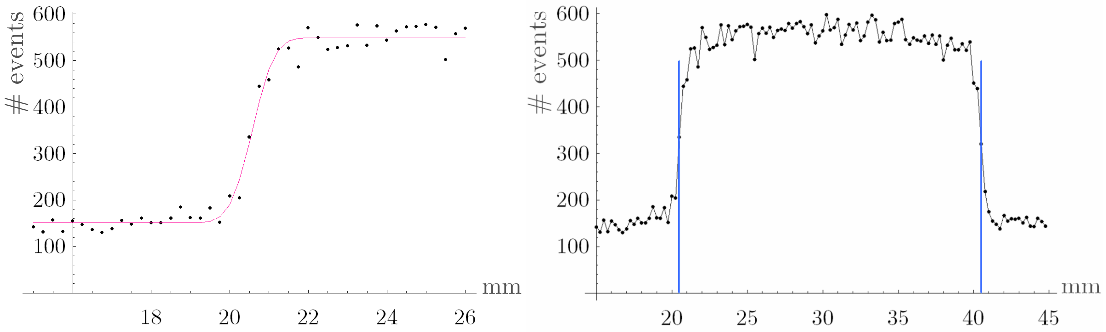


Figure 3.6: (left) Measured count rate profile over the edge of the scintillator block fitted with $\mathrm{g(x)} \otimes \mathrm{h(x)}$, the convolution between a Gaussian and a step function. (right) Count rate over the entire block. The edges are found by symmetry, i.e. positions 20mm apart with equal rates.

Both methods, the convolved Gaussian and the symmetry approach, give equivalent results ($\Delta < 0.2$mm). Using the convoluted Gaussian approach, the FWHM beam width can be derived. The beam had an estimated diameter of ~1.3mm FWHM.

Thus, training data for the positioning algorithms consists of light distribution profiles acquired by irradiating the crystal with a beam of annihilation photons at many different positions with known coordinates. As mentioned before the algorithms are trained to compute the photon incidence position instead of the interaction position in the block. Because the incidence position is independent of the interaction depth, it doesn't suffer from parallax errors. However the relation between the measured scintillation light distribution and the incidence position depends on the incidence angle (Figure 3.1). Therefore different positioning algorithms have to be trained as a function of the incidence angle.

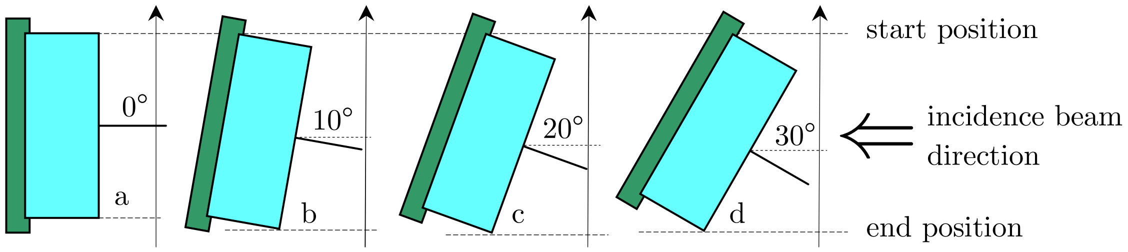


Figure 3.7: Top views of a frond-end detector module in the bench set-up. Four block angles are evaluated: 0, 10, 20 and 30 degree. The beam is stepped over the whole surface of the block (from start position to end position in steps of 250 μm)

In order to determine the best detector geometry/positioning algorithm combination, the intrinsic detector resolution of photons impinging at different angles e.g. 0°, $\pm 10°$, $\pm 20°$ and $\pm 30°$ are evaluated. With the use of the bench set-up, the detector resolution is determined experimentally along the long axis of the scintillator. First, the beam was scanned along the central x-axis of the APD array, perpendicular to its surface, from one edge of the crystal to the other in steps of 250 μm (see Figure 3.7 (a)). Since we intend to use nearly symmetric 21.4x18.4x10mm^3 LSO blocks read out by two APDs in a future implementation of this detector principle, it is sufficient to study the resolution along this axis. To study the influence of non perpendicular incidence photons, the detector box was also rotated over 10, 20 and 30 degree relative to the photon beam (see Figure 3.7 (b-d)).

Since we only determine the interaction coordinate along the long side of the block, the light distribution along the pixels in an APD column yields little or no extra information. Hence, we can sum the signals of all pixels for every column, $\sum_{j=a}^{d} x_{ij}$ i:A..H (Figure 3.8). This significantly reduces the number of parameters to be determined and speeds up the learning process. Since the position only depends on the shape of the light distribution, the vector containing the 8 combined APD signals is also normalized by $x_i / \sum_{j=A}^{H} x_j$ i:A..H . This makes the input signal distribution energy independent, resulting in a further simplification of the problem.

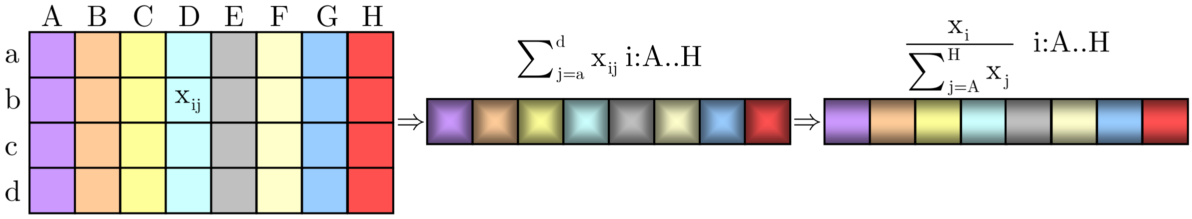


Figure 3.8: Schematic representation of the data preprocessing used to simplify the training parameters from 32 to 8 input variables.

Per incidence angle, the pre-processed data is then split into a training set and a validation set. The training set is used to train the positioning algorithms. For each event in the validation set, the positioning error, defined as the predicted photon position (using the trained positioning algorithm) minus the true beam position, is calculated. To obtain an estimate for the spatial resolution, the positioning errors are collected in histograms. The Full Width at Half Maximum (FWHM) and the Full Width at Tenth Maximum (FWTM) are used to characterize the spatial resolution.

3.3 Positioning algorithms

The photon interaction position and DOI information is embedded in the scintillation light distribution sampled by the APD array. The block detector can be considered as a non-linear system mapping a beam position coordinate onto a set of APD responses. The role of the positioning algorithms is just to inverse this mapping i.e. matching the set of APD responses with an impact gamma position coordinate. To extract this position three machine learning algorithms were evaluated:

- Neural network (NN) with Levenberg-Marquardt training
- Neural network with algebraic training (algNN)
- Support vector machines (SVM)

The approach of using NNs for the positioning problem in PET detectors has already been proposed by some groups and shown to be promising [20, 45-48]. Two fundamental properties make neural networks a powerful tool widely used for function approximation [20]. The property of universal approximation implies that neural networks can approximate any smooth function within any required accuracy. Another important property of neural networks is the learning ability. The universal approximation property of neural networks states that a particular set of weight values will result in a network that can approximate the function with the required accuracy (see Paragraph 3.3.1). However, these desired weights are unknown for most applications and must be learned by the neural network using training data.

The novel algebraic neural network training technique was developed by Ferrari et al. [49-51]. This approach suggests an innovative framework for analyzing neural approximation properties and for training neural networks in a much simplified way. The training process and the network approximation properties are solved via linear algebra. Algebraic training is characterized by faster execution speeds than contemporary NNs.

Finally, Support Vector Machines (SVM) algorithms combine the simplicity and computational efficiency of linear algorithms with the flexibility of non-linear systems. Their foundation in the principles of statistical learning theory makes them remarkably resistant to overfitting. A significant advantage of SVMs compared to NNs is that SVM regression always finds a global minimum, while NN can suffer from multiple local minima [52].

The aim of the following three subsections is to give a brief introduction of the parameters that characterize the different machine learning algorithms. It does not involve a study of the algorithms, but only the practical use of them. For more detail, references to the literature are made.

3.3.1 Neural networks with Levenberg-Marquardt training

Artificial neural networks are massively parallel computational units, based upon models of neurological structures and processing functions in the brain. They are used in a variety of applications because they can learn by example and provide excellent universal function approximation for multivariate input/output spaces [51]. They are good at solving problems that involve complexity, non-linearity and uncertainty. Similar to biological systems, neural networks receive their abilities, such as adaptation and error tolerance from the interconnected structure of individual simple processing units called neurons. Generally, a neural network is a structure involving weighted interconnections among neurons, which are most often nonlinear scalar transformations. The processing ability of the network is stored in the inter-unit connection strengths, or synaptic weights, obtained by learning from a set of training patterns, termed as training set [20].

A wide variety of network types exists, but one of the most popular and most widely used models is the feed forward neural network (FF).

Figure 3.9 shows an example of a one-hidden-layer FF neural network with n inputs, $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ that feed each of the s neurons comprising the hidden layer. The s outputs from this layer are then fed into the single output layer neuron, yielding the scalar output, $\hat{y}$. The layer of s neurons is called “hidden layer” because its outputs are not directly seen in the data. Each arrow in Figure 3.9 corresponds to a real-valued parameter, or a weight, of the network. The values of these parameters are tuned in the network training. Generally, a neuron is structured to process multiple inputs in a nonlinear way, producing a single output. Specifically, all inputs to a neuron are first augmented by multiplicative weights **W**. These weighted inputs are summed and then transformed via a nonlinear activation function, σ. As indicated in Figure 3.9, the neurons in the hidden layer of the network are nonlinear. In our study, the single output neuron is linear, since no activation function is used.

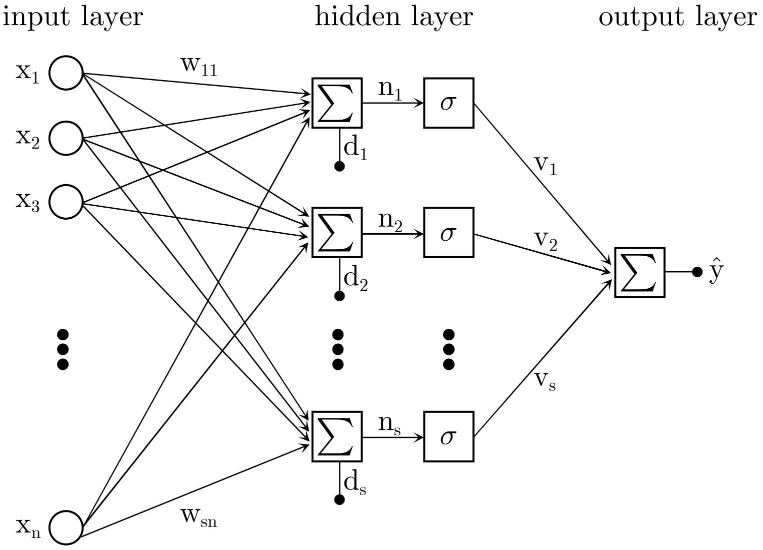


Figure 3.9: General structure of a feed forward neural network with one hidden layer

Mathematically the functionality of a hidden neuron is described by:

$$\sigma\left(\sum_{j=1}^{n} w_{ij}x_j + d_i\right) \text{ for each hidden neuron } (i = 1..s) \qquad (3.3)$$

where **x** is the input vector, n the dimension of the input vector **x**, s the number of neurons in the hidden layer, **W** the matrix of input-to-hidden layer weights and **d** the vector of hidden node biases. The network output is formed by another weighted summation of the outputs of the neurons in the hidden layer, vector **v**. This summation on the output is called the output layer. The output of this network is mathematically given by

$$\hat{y}(\boldsymbol{\theta}) = g(\boldsymbol{\theta},\mathbf{x}) = \mathbf{v}^{T}\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}] = \sum_{i=1}^{s} v_i\,\sigma\left[\sum_{j=1}^{n} w_{ij}x_j + d_i\right] \tag{3.4}$$

where **v** is a vector of the hidden-to-output layer weights, $\boldsymbol{\theta}$ the parameter containing all the variables $\{\mathbf{W}, \mathbf{d}, \mathbf{v}\}$ of the network model, g(.,.) the network function and **x** is the input to the network (see below).

The nonlinear activation function σ in the neuron is usually chosen to be a smooth differential function. In our study the standard sigmoid function was used [53]:

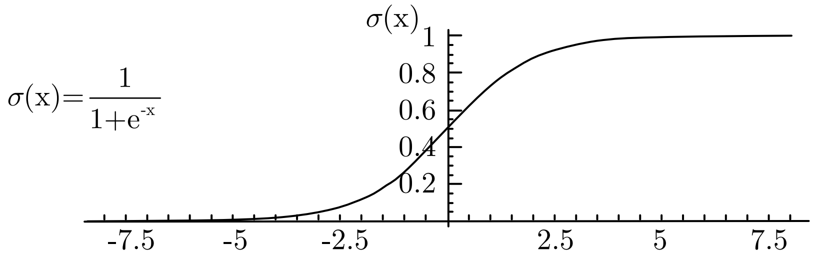


Figure 3.10: Graphical representation of the standard sigmoid function as the nonlinear activation function $\sigma(x)$

Once the structure g(.,.) of the NN is chosen, e.g. number of layers and the number of neurons in the different layers, it can be trained using a set of data containing N input-output pairs, $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$. Denote this network with $\hat{y}=g(\boldsymbol{\theta},\mathbf{x})$. Training the network means that all its parameters are adjusted incrementally until the training data satisfy the desired mapping as good as possible. That is, until $\hat{y}(\boldsymbol{\theta})$ matches the desired output y as closely as possible. The typical stop criterion, used in almost all supervised learning algorithms, consists of minimizing some measure of the error between the desired input/output and the actual network's performance. A good estimate for parameter $\boldsymbol{\theta}$ is one that minimizes this MSE or RMSE (root mean square error). Given the training data $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$ the mean square error (MSE) is defined by

$$V_N(\boldsymbol{\theta})=\frac{1}{N}\sum_{i=1}^{N}(y_i\text{-}g(\boldsymbol{\theta},\mathbf{x}_i))^2 \tag{3.5}$$

The various training algorithms that apply to FF have one thing in common: they are iterative. Starting at the initial parameter vector $\boldsymbol{\theta}^0$, the training algorithm iteratively decreases the MSE in Equation (3.5). Neural network minimization problems are often very ill-conditioned. This makes the minimization problem harder to solve and for such problems, the Levenberg-Marquardt algorithm is often a good choice because of its robustness and faster convergence [53]. In this study we used a neural network with two hidden layers, due to its faster convergence than 1 hidden layer [54]. Each hidden layer consists of five neurons. This choice was made via trial-and-error to achieve a good performance using a simple network topology.

3.3.2 Neural network with algebraic training

The nonlinear behavior of neural networks results from the sigmoid activation function in the neurons of the hidden layer. It was suggested in [51] that the neural network training can be converted into a linear problem if the input-to-hidden weights **W** and biases **d** have been fixed previously (see Equation (3.4)). When w_{ij} and d_i are known, the output of the nonlinear hidden layer $\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}]$ can be pre-computed for each event $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$ in the training set. Starting from Equation (3.4), the problem of finding the hidden-to-output weights **v** is now reduced to a simple linear algebra problem,

$$\mathbf{v}=\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}]^{-1}\,\mathbf{y} \tag{3.6}$$

When the matrix $\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}]$ is not a square matrix (i.e. the number of training samples **x** is larger than the number of nodes s in the hidden layer), a solution can still be obtained using the pseudo-inverse

$$\mathbf{v}=\left(\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}]^{T}\,\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}]\right)^{-1}\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}]^{T}\,y \tag{3.7}$$

However the question remains how to determine w_{ij} and d_i. One of the strategies to produce a well-conditioned sigmoid matrix $S\equiv\sigma[\mathbf{W}\cdot\mathbf{x}+\mathbf{d}]$ consists of generating the input-to-hidden weights **W** between the j^{th}-input and the i^{th}-node, according to the following rule [50]

$$w_{ij}=f.r_{ij} \tag{3.8}$$

where r_{ij} is randomly chosen from a normal distribution with zero mean and unit variance $N[0,1]$ and f is a tunable scaling parameter. It turns out that the choice of the free parameter f is not very critical and is of the order of 10 [51]. Finally, the input bias, **d**, is computed to center each sigmoid at one of the training pairs, $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$

$$\mathbf{d}=-\mathrm{diag}\left(\mathbf{X}\cdot\mathbf{W}^{T}\right) \tag{3.9}$$

where **X** is the matrix composed of all the input elements from the training set.

Using the above procedure, the matrix S can be computed and consequently the linear problem of Equation (3.6) solved. In the rare case that the computation of the pseudo-inverse leads to numerical instabilities, one just chooses new random w_{ij} and d_i's. Due to the simplicity of the whole procedure, training of a neural network is fast, even for larger sized networks.

3.3.3 Support vector machines

Support vector machines (SVM) comprise another class of learning algorithms, motivated by results of the statistical learning theory introduced by Vapnik in the sixties. The basic idea in support vector regression (SVR) is to map the input data **x** into a higher dimensional feature space *F* via a nonlinear mapping function ϕ. This approach reduces the learning step to a linear regression problem (in the higher dimensional feature space). This regression function is approximated by [55]:

$$f(\mathbf{x})=\sum_{i=1}^{N} w_i\phi(x_i)+b \tag{3.10}$$

where $\{\phi(x_i)\}_{i=1}^{N}$ are the features of the training inputs, $\{w_i\}_{i=1}^{N}$ determine the orientation of the hyperplane going through the features while the scalar b determines the offset of the hyperplane from the origin.

Standard SVR uses an ε-insensitive loss function. An example is shown in Figure 3.11. The aim in ε-SVR is to find a function $f(x_i)$ that has at most ε deviation from the outputs y_i for all the training data $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$, and at the same time is as flat as possible. In other words, if the deviation between the actual and predicted value is less than ε, the regression function is not considered to be in error. Figure 3.11 depicts the situation graphically, it can be visualize as a band or tube of size 2ε around the function f(x), any point outside this tube can be viewed as a training error [56].

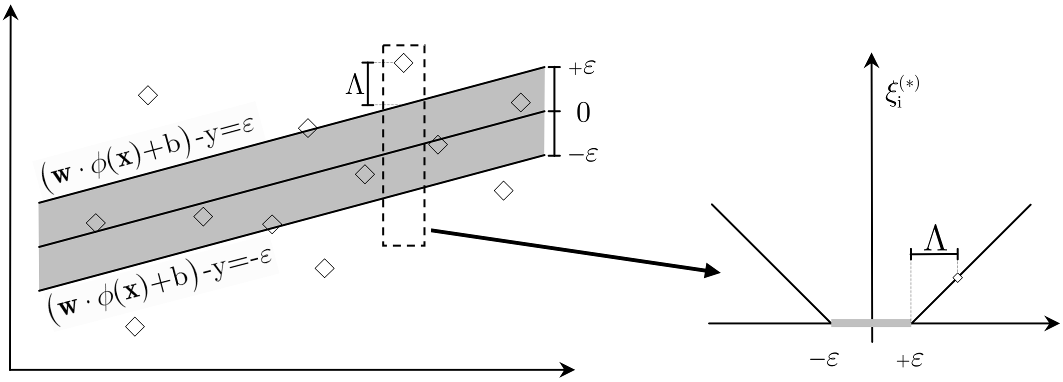


Figure 3.11: (left) Plot of $\mathbf{w}\cdot\phi(\mathbf{x})+b$ versus y with the ε-insensitive tube. Points outside the tube are penalized. (right) A typical linear ε-insensitive loss function.

The coefficients $\{w_i\}_{i=1}^{N}$ and b of Equation (3.10) can be estimated by minimizing the risk function $R(\mathbf{w},\xi)$ [56]:

$$R(\mathbf{w},\xi)=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{N}(\xi_i+\xi_i^*)$$
$$\text{subject to}\begin{cases} y_i-\mathbf{w}\phi(x_i)-b\le\varepsilon+\xi_i \\ \mathbf{w}\phi(x_i)+b-y_i\le\varepsilon+\xi_i^* \\ \xi_i,\ \xi_i^*\ge 0 \end{cases} \tag{3.11}$$

The first term in Equation (3.11), $(1/2).\|w\|^2$ is used as a complexity measurement of function (3.10). Complexity in this case means a "flatter" function and one way to ensure this, is to minimize the norm, i.e. $\|\mathbf{w}\|^2$ [56]. The regularization constant C, chosen a priori, is a constant determining the tradeoff between minimizing the training error and the model complexity. To account for training errors, slack variables ξ_i and ξ_i^* are introduced. The nonnegative error variables $\xi_i^{(*)}$ were added in each constraint of (3.11) and then added as a weighted penalty term to cope with otherwise infeasible constraints of the optimization problem. The first variable ξ_i computes the error for underestimating the function. The second ξ_i^* computes the error for an overestimation. The slack variables are zero for points inside the tube and progressively increase for points outside the tube according to the loss function used [57].

In practice, the optimization problem (3.11) is solved more easily in its dual formulation [56]. By introducing a dual set of variables, function f(x) of equation (3.10) becomes:

$$\begin{aligned} \mathrm{f}\left(\mathbf{x},\alpha_{\mathrm{i}},\alpha_{\mathrm{i}}^{*}\right) &= \sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w}_{\mathrm{i}}\phi(\mathrm{x}_{\mathrm{i}})+\mathrm{b} = \sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}\left(\alpha_{\mathrm{i}}\text{-}\alpha_{\mathrm{i}}^{*}\right)\phi(\mathrm{x}_{\mathrm{i}})^{\mathrm{T}}\phi(\mathrm{x})+\mathrm{b} \\ &= \sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}\left(\alpha_{\mathrm{i}}\text{-}\alpha_{\mathrm{i}}^{*}\right)\mathrm{k}\left(\mathrm{x}_{\mathrm{i}},\mathbf{x}\right)+\mathrm{b} \end{aligned} \tag{3.12}$$

Based on the nature of the dual formulation, only a number of coefficients among α_i and α_i^* are nonzero [58]. The set of training events x_i associated with them and thus used in the approximation function $f\left(\mathbf{x},\alpha_i,\alpha_i^*\right)$ are called support vectors.

For computational convenience, the form $\phi(x_i)^T\phi(\mathbf{x})$ in Equation (3.12) is replaced by a kernel function $k(x_i,\mathbf{x})$ [56]. Using a kernel function in the training algorithm, avoids computing $\phi(x_i)^T\phi(\mathbf{x})$ explicitly which is computationally infeasible for features of higher order and higher dimensionality. The question that arises now is, which functions $k(x_i,\mathbf{x})$ correspond to a dot product in some feature space *F*. In [56] it is proven that following kernels are suitable; the Gaussian kernel $k(x,y)=\exp\left(-\|x\text{-}y\|^2/2\sigma^2\right)$ and the polynomial kernel $k(x,y)=(x^Ty+1)^p$.

Besides standard model selection issues, i.e. how to specify the trade-off between empirical error and model complexity (parameter C in Equation (3.11)), there also exists the problem of an optimal choice of ε. There exists, however, a method to construct SVR that automatically adjust ε and moreover also have a predetermined fraction ν of the training data as support vectors. The accuracy parameter ε becomes a variable of the optimization problem, including an extra term in the risk function $R\left(\mathbf{w},\xi\right)$ which attempts to minimize ε. Equation (3.11) becomes [59]:

$$\mathrm{R}_{\nu}\left(\mathbf{w},\xi\right) = \frac{1}{2}\|\mathbf{w}\|^2 + \mathrm{C}\left(\sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}\left(\xi_{\mathrm{i}}+\xi_{\mathrm{i}}^{*}\right)+\mathrm{N}\nu\varepsilon\right) \tag{3.13}$$

The coefficients α_i and α_i^* in its dual formulation are computed by the following optimization problem [56]:

$$\begin{aligned} &\text{maximize } -\frac{1}{2}\sum\nolimits_{\mathrm{i,j}=1}^{\mathrm{N}}\left(\alpha_{\mathrm{i}}\text{-}\alpha_{\mathrm{i}}^{*}\right)\left(\alpha_{\mathrm{j}}\text{-}\alpha_{\mathrm{j}}^{*}\right)\mathrm{k}(\mathrm{x}_{\mathrm{i}},\mathrm{x}_{\mathrm{j}})+\sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}\mathrm{y}_{\mathrm{i}}\left(\alpha_{\mathrm{i}}\text{-}\alpha_{\mathrm{i}}^{*}\right) \\ &\text{subject to } \begin{cases} \sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}\left(\alpha_{\mathrm{i}}\text{-}\alpha_{\mathrm{i}}^{*}\right) = 0 \\ \sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}\left(\alpha_{\mathrm{i}}+\alpha_{\mathrm{i}}^{*}\right) \leq \mathrm{C}\nu\mathrm{N} \\ \alpha_{\mathrm{i}},\alpha_{\mathrm{i}}^{*} \in [0,\mathrm{C}] \end{cases} \end{aligned} \tag{3.14}$$

As mentioned before, the parameter C determines a trade-off between the flatness of $f\left(\mathbf{x},\alpha_i,\alpha_i^*\right)$ and the tolerance of deviations while ν is the fraction of the set of training events which can be used in the expansion of $f\left(\mathbf{x},\alpha_i,\alpha_i^*\right)$.

The SVM training was done using the ν-SVM algorithm implemented in LIBSVM [60]. In equation (3.12) a radial basis function (RBF) was used as a kernel function, i.e. $k(x_i,y)=\exp\left(-\gamma\left\|x_i\text{-}y\right\|^2\right)$ where γ is the width of the RBF kernel. This parameter has to be optimized in parallel with the parameter C.

3.4 Evaluation of the positioning algorithms on a 20x10x10mm^3 LSO block

To evaluate the machine learning algorithms, the spatial resolution of the detectors is estimated using data from one-dimensional scans. Since we intend to use nearly symmetric 21.4x18.4x10mm^3 LSO blocks read out by two adjacent APDs in a future implementation of this detector principle (e.g. BrainPet system), it is sufficient to study the resolution along this axis. The beam is scanned along the central axis of the APD array from one edge of the crystal to the other as described in Paragraph 3.2. At each position, a fixed number of light distributions are recorded. The pre-processed data is then split into a training set and a validation set. The positions of the events in the validation set are estimated using the algorithms introduced above. For each event in the validation set, the positioning error, defined as the estimated minus the true coordinate, is calculated. To obtain an estimate for the spatial resolution, the positioning errors are collected in histograms. The Full Width at Half Maximum (FWHM) and the Full Width at Tenth Maximum (FWTM) of these histograms are used as an estimate for the spatial resolution. Histograms of the position errors of all the beam positions yield an average resolution over the complete detector. The position dependent spatial resolution is estimated using histograms corresponding to only small intervals of neighboring beam positions. The contribution of the finite photon beam size (~1mm FWHM) is not subtracted from the results.

3.4.1 Levenberg-Marquardt NN (LM-NN)

The NN training procedure is an iterative process that adjusts the parameters of the NN by comparing the predicted output of each event in the training set with its desired output. The way the parameters are adjusted depends on the training algorithm used. We used the Levenberg–Marquardt (LM-NN) method available in the Neural Network package of Mathematica [61]. However, when the NN is trained too long, it can lead to overfitting of the data in the training set. The NN starts to fit the noise structure of the particular training set and hence loses generality. To avoid this, the training set is split, once more, in a training set and a separate test set. After each training cycle, the total RMSE on all events in the test set is computed.

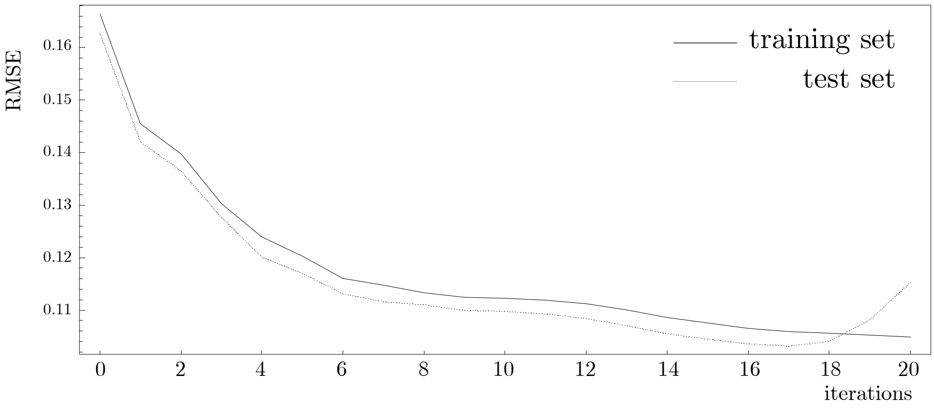


Figure 3.12: Typical RMSE minimization profiles during a NN training

While the total RMSE in the training set keeps diminishing as training progresses, the RMSE on the test events will come to a plateau or start to increase again. At this point, training is stopped. Figure 3.12 shows a typical decrease of the RMSE during the training process of the training set. In this example, the LM-NN parameters are found at the 17th training iteration. At which the RMSE of the test set starts to increase and the neural network loses its generality.

Figure 3.13 shows the global positioning error profile for the 20x10x10mm^3 LSO block obtained from the trained LM-NN using the separate validation data set.

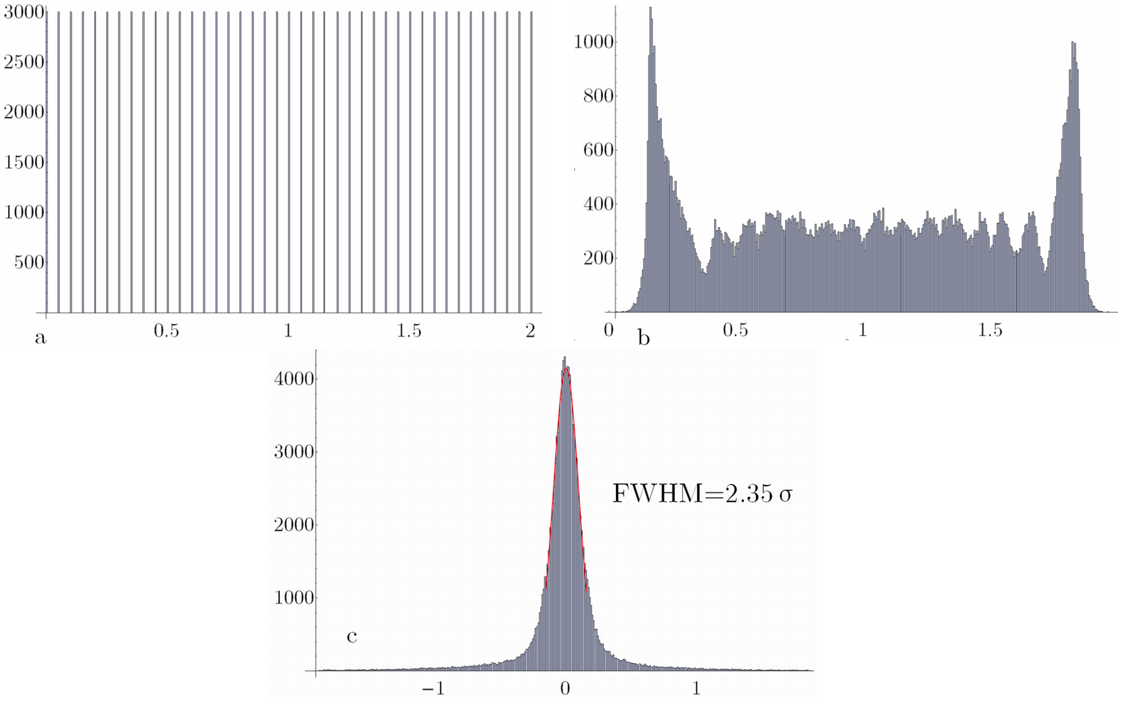


Figure 3.13: Histogram (a) illustrates the number of acquired events per beam position in the validation set. Histogram (b) summarizes the estimated impinging photon position by the LM-NN applied on the same validation data set. The global error profile (true minus estimated position) is given in histogram (c). A Gaussian fit on this histogram provides the FWHM.

Figure 3.14 shows the FWHM and FWTM of the global error profiles (or spatial resolution) for the 20x10x10mm^3 LSO block as a function of the incidence angle and energy threshold. The FWHM is found by a Gaussian fit $N[\mu, \sigma^2]$ on the positioning error histogram. The FWHM is defined as 2.35σ. The FWTM is acquired by an interpolation function to calculate the width at the tenth of the maximum. For perpendicular incident photons, the measured FWHM detector resolution is 1.60mm and 1.75mm for an energy threshold of respectively 380 keV and 100 keV. When the photon incidence angle increases to 30°, the spatial resolution degrades slightly to 1.85mm and 2.0mm FWHM. The FWTM of the photon position error distribution is between 4.05mm and 4.3mm depending on the energy threshold on the perpendicular incident photons and degrades to 5.0mm and 5.5mm at a 30° incidence angle.

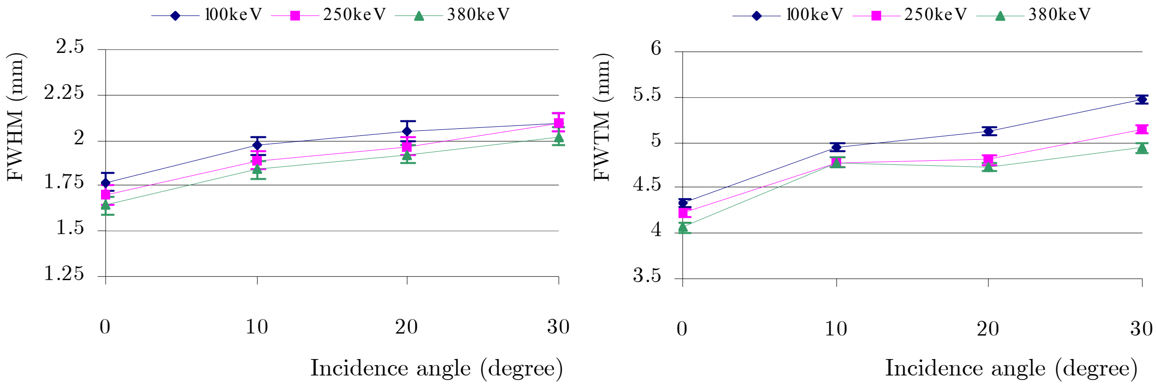


Figure 3.14: (left) the FWHM resolution and (right) the FWTM resolution achieved with Levenberg–Marquardt trained neural networks on a 20x10x10mm^3 LSO block as a function of the photon incidence angle and energy threshold.

Summing the 32 APD signal values yields the energy deposited in the crystal. A typical energy spectrum is shown in Figure 3.15 (left). Figure 3.15 (right) shows the measured energy resolution as a function of the beam position. Energy spectra were recorded for photons impinging on the LSO block in 2-mm intervals. The energy resolution is rather uniform. On average, the energy resolution was 11.5%. Only a slight degradation was observed near the edge of the block.

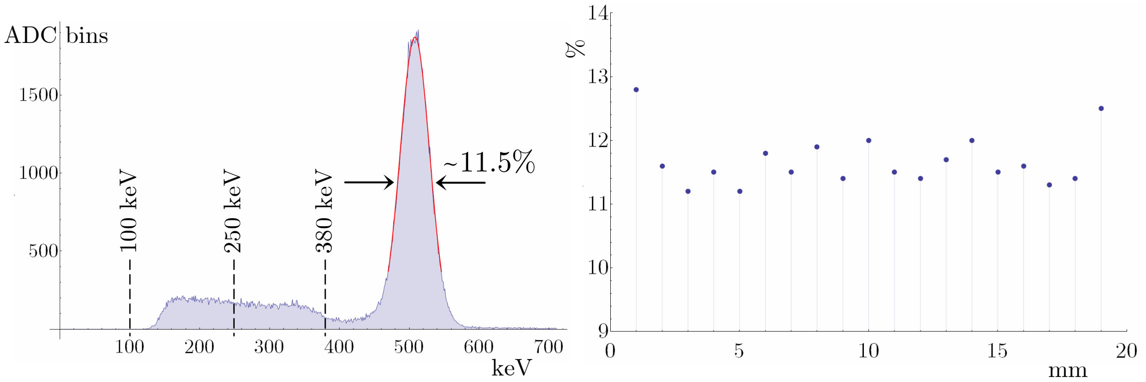


Figure 3.15: (left) The energy spectrum obtained with a 20x10x10mm^3 LSO block has an average FWHM energy resolution of 11.5%. (right) FWHM energy resolution measured in 2 mm intervals along the 20 mm side of the LSO block. The energy resolution is rather uniform except for a slight deterioration near the edges.

The dashed lines in Figure 3.15 (left) denote the three different energy threshold applied in this study. Increasing the threshold from 100 keV to 250 keV or 380 keV reduces the detector sensitivity by respectively 11.8% and 26.4% [47]. However, the signal-to-noise ratio (SNR) per pixel improves by increasing threshold, resulting in a better estimate of the impinging position.

Beside the global resolution, the local resolution and nonlinearity of LM trained NN for perpendicular incident photons are examined. This was accomplished by computing the positions of events impinging within 1mm intervals (i.e. combining events from four adjacent beam positions) and comparing them with their known incidence position. The FWHM and FWTM resolutions are again found by respectively a Gaussian and interpolation fit applied on the positioning error histogram. The spatial nonlinearity is defined as the deviation of the peak position of the local error profile from the true center of the photon beam. Figure 3.16 shows the FWHM resolution, FWTM resolution and nonlinearity for each of the twenty 1mm regions along the 20mm-long beam trajectory. The resolution remains fairly uniform across the set of measured positions. The FWHM fluctuation across the different regions is less than 0.3mm rms. However, a local deterioration of the resolution is observed when the photons impinge about 4-5mm from the edge. This phenomenon is also present in other positioning estimators for monolithic scintillators. The origin of this is still not understood [48, 62].

The nonlinearity of the trained neural networks is negligible except for photons impinging within 1mm from the edge. In this region the estimated positions are shifted towards the center of the block over a distance of ~1mm. The shifts can also be seen on histogram (b) in Figure 3.13. This behavior was also observed for the different block geometries and algorithms described above [37, 45, 47, 48, 63].

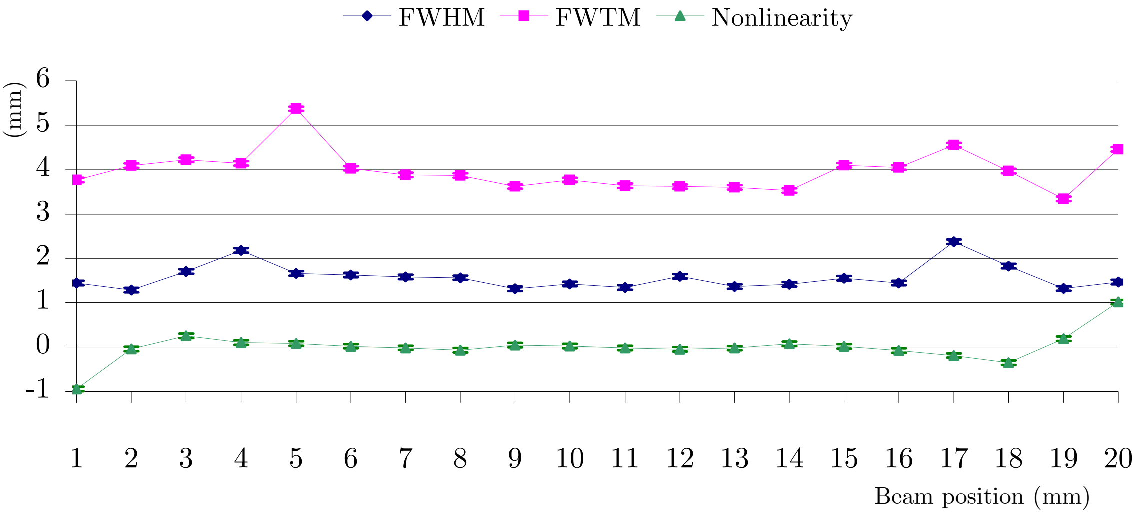


Figure 3.16: Local FWHM, FWTM resolutions and nonlinearity using LM trained neural network for 1mm intervals along the 20mm side of the 20x10x10mm^3 LSO block

3.4.2 Algebraic trained NN (Alg-NN)

In the algebraic training approach, only one parameter has to be fixed before training, i.e. the number of neurons in the hidden layer, see equation (3.8).

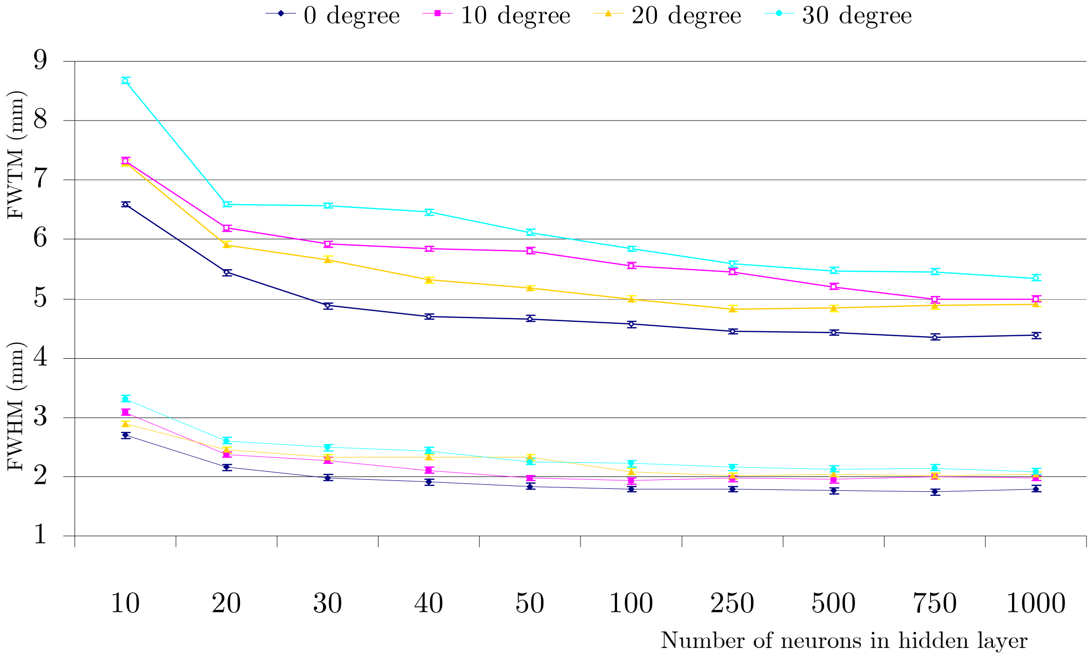


Figure 3.17: (lower 4 plots) the FWHM resolution and (upper 4 plots) the FWTM resolution achieved with algebraic trained neural networks on a 20x10x10mm^3 LSO block as a function of the photon incidence angle and the number of neurons in the hidden layer. The energy threshold was 100keV.

Figure 3.17 shows the FWHM and FWTM resolutions obtained with an algebraic trained neural network as a function of the incidence angle and the number of neurons used in the hidden layer. An energy threshold of 100keV was applied. An algebraic trained NN can achieve a similar performance to a LM trained NN but requires a much larger number (>500) of neurons in the hidden layer. If the energy threshold is increased to 250keV or 380keV, the FWHM and FWTM resolutions improve to values very comparable to those obtained with LM trained neural networks.

3.4.3 Support vector machines (SVM)

Before training the SVM using a training data set, we need to fix three parameters: C, ν and γ (see equation (3.14) for more information about parameters C and ν. γ is the width of the RBF kernel). Our initial training data set consisted of 10000 events acquired at 80 beam positions along the 20 mm side of the LSO block. Figure 3.18 shows the obtained FWHM resolution as a function of γ and C when we allow 5000 training vectors to be used as support vectors, i.e. ν=0.5. The resolution is not very sensitive to the choice of C while the value for γ should be larger than 20.

The measured FWHM and FWTM detector resolution as a function of energy threshold and photon incidence angle are given in Figure 3.19. These results were obtained with C=11 and γ=31. For perpendicular incident photons the SVM algorithm resulted in a 1.69mm FWHM and 4.31mm FWTM resolution in case of a 100keV threshold. Increasing the energy threshold to 380keV has a minimal beneficial effect, i.e. 1.62mm FWHM and 4.08mm FWTM. The parallax correction at larger incidence angles is slightly inferior compared to what can be achieved using LM trained neural networks.

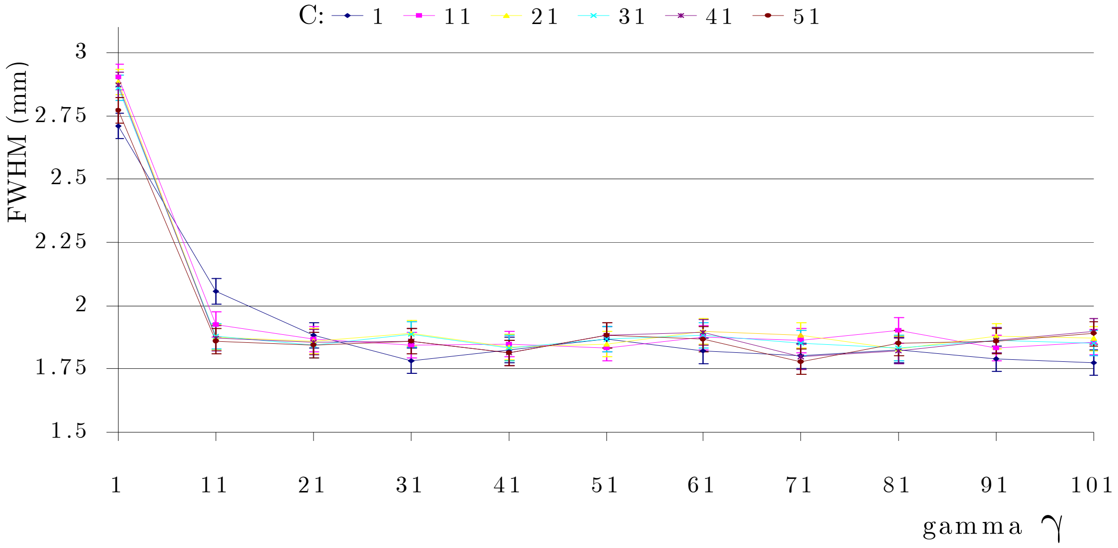


Figure 3.18: FWHM resolution as a function of the parameters γ and C. The parameter γ is the width of the RBF kernel while variable C determines the trade-off between the model complexity and the degree to which deviations larger than ε are tolerated

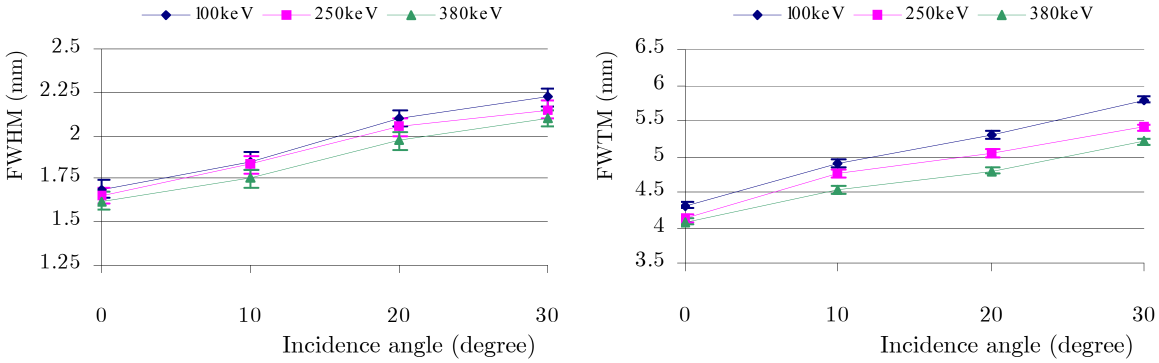
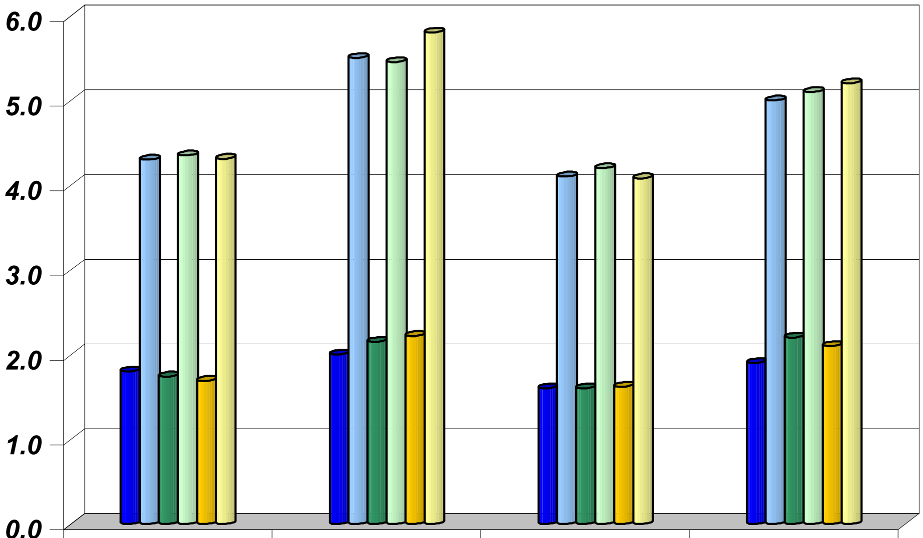


Figure 3.19: (left) the FWHM resolution and (right) the FWTM resolution achieved with SVM on a 20x10x10mm^3 LSO block as a function of the photon incidence angle and energy threshold. The ν-SVM parameters were: ν =0.5, C=11 and γ =31

3.4.4 Discussion and conclusion

The results obtained with the different positioning algorithms on the 20x10x10mm^3 LSO block at photon incidence angles of 0° and 30° and energy thresholds of 100keV and 380keV are summarized in Figure 3.20. The Levenberg-Marquardt neural network uses two hidden layers where each hidden layer consists of five neurons. For the algebraic trained NN, 500 neurons are used in the hidden layer. And finally, the results of the SVM approach were attained with C=11 and γ=31.



	0° at 100keV	30° at 100keV	0° at 380keV	30° at 380keV
🟦 FWHM LM-NN	1.8	2.0	1.6	1.9
🟦 FWTM LM-NN	4.3	5.5	4.1	5.0
🟩 FWHM Alg-NN	1.7	2.2	1.6	2.2
🟩 FWTM Alg-NN	4.4	5.5	4.2	5.1
🟧 FWHM SVM	1.7	2.2	1.6	2.1
🟨 FWTM SVM	4.3	5.8	4.1	5.2

Figure 3.20: Overview of the FWHM and FWTM detector resolutions achieved with the different positioning machine learning algorithms for the 20x10x10mm^3 LSO block at 0° and 30° photon incidence angles and energy thresholds of 100keV and 380keV.

All three algorithms can achieve similar resolutions but the LM trained neural networks do so most efficiently (a simpler network structure and hence a faster execution speed). For perpendicular incident photons they achieve a measured detector resolution of 1.8mm FWHM and 4.3mm FWTM on a 20x10x10mm^3 LSO block when a minimum threshold of 100 keV is applied, i.e. maximum sensitivity setting. Because the 511 keV photon beam size is about 1 mm FWHM, we expect that the true intrinsic detector resolution is closer to 1.4 mm FWHM.

Increasing the threshold improves the resolution. At a 380 keV threshold, the measured detector resolution is 1.6 mm FWHM and 4.1 mm FWTM. The sensitivity is decreased to 73.6% in comparison to the sensitivity at 100keV threshold [47].

When neural networks are trained for different incidence angles, the resolution degradation due to photon penetration for non-perpendicular incident photons can be limited. At an incidence angle of 30° the measured detector resolution degrades to 2.0mm FWHM and 5.5mm FWTM using a threshold of 100 keV. Again better performance can be achieved at the cost of a lower sensitivity by increasing the energy threshold.

The analytic trained neural networks achieve similar resolutions but need a much larger neural network to do so, i.e at least 500 neurons in the hidden layer. Given the simplicity of the algebraic training algorithm, training a neural network with 500 neurons is still much faster than LM training of a neural network with 2 hidden layers with 5 neurons each. However, the usage of such a large neural network will be slower in comparison to the small LM trained neural networks (Table 3.3).

Table 3.3: Training and execution speed of the different training algorithms

	Training speed	Execution speed
LM-Neural network	±	++
Alg-Neural network	++	±
SVM	+	-

The SVMs perform slightly better at the high sensitivity setting of the energy threshold (100 keV). At an energy threshold of 380keV they perform similarly. To obtain these results, a few thousand support vectors are required. The training speed is in between the algebraic training and the LM training of the neural networks. Again, given the large number of support vectors required, the usage of these SVMs is slower than the LM trained neural networks.

Although the three algorithms have a similar performance, only the LM trained neural networks manage to achieve this with a simple architecture. This makes them for the moment the only candidates for on-line processing of events since these simple LM-NN can be implemented in hardware.

3.5 Evaluation of alternative detector geometries with a LM-NN

3.5.1 20mm thick LSO block

To increase the detection efficiency the thickness of the rectangular LSO block was doubled to 20mm and read out by two APDs on the top and bottom surface of the 20x10x20mm^3 LSO block (Figure 3.4 (a)). The probability, for a 511 keV photon perpendicularly incident on the crystal, to undergo any interaction (photoelectric or Compton) in a 20x10x10mm^3 LSO block is ~58%. Increasing the crystal thickness to 20 mm increases this probability to ~83% [63].

The data reduction introduced at the end of Paragraph 3.2 is still applied; the 2x32 APD signals from the two APDs on the top and bottom surface of the 20x10x20mm^3 LSO block are reduced to 2x8 NN inputs. The FWHM and FWTM resolutions obtained with an LM trained NN are shown in Figure 3.21. The FWHM detector resolution only decreases slightly in function of the incident photon angle, starting from 2.07mm for perpendicular incident photons to 2.24mm FWHM for photons impinging at 30° with an energy threshold of 380keV. Nevertheless, the measured FWHM detector resolutions are worse compared to the thinner 20x10x10mm^3 LSO block. Also the FWTM (4.97mm at 0° to 5.45mm for 30°) are inferior compared to the 20x10x10mm^3 results. This could be due to the fact that the scintillation light is now spread over 64 APD pixels in stead of only 32 pixels. Hence the statistical fluctuation on the number of photons per pixel will be larger. In addition each of the 32 extra readout channels will also contribute electronic noise to the data.

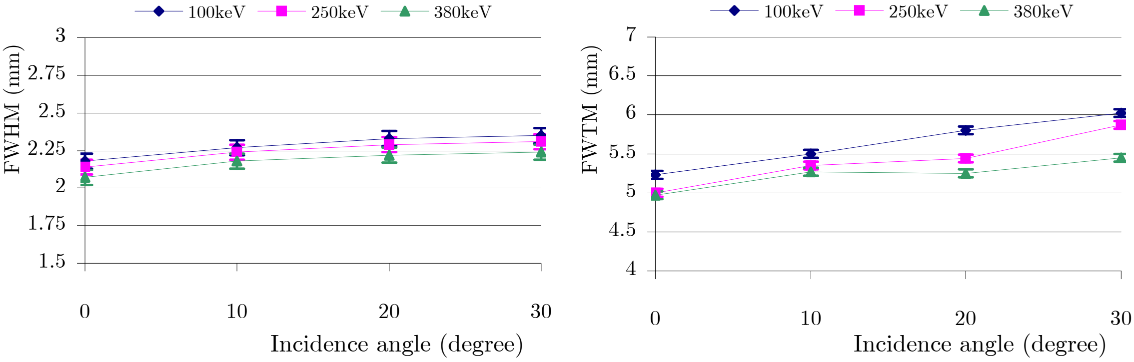


Figure 3.21: (left) the FWHM resolution and (right) the FWTM resolution achieved with Levenberg–Marquardt NN trained neural networks on a 20x10x20mm^3 LSO block as a function of the photon incidence angle and energy threshold.

To conclude, the detector resolutions achieved with a 20x10x20mm^3 LSO block are worse in comparison to a thinner 20x10x10mm^3 block. Consequently, a dual layer of 10mm thick LSO blocks is preferred to increase the overall sensitivity. In addition, since the two layers operate independently, a multiple photon interaction involving both blocks can be detected. This opens the possibility to determine the first interaction point on the basis of Compton kinematics. This is not possible in a single 20mm thick block where multiple scattered events cannot be discerned.

3.5.2 Trapezoidal LSO block

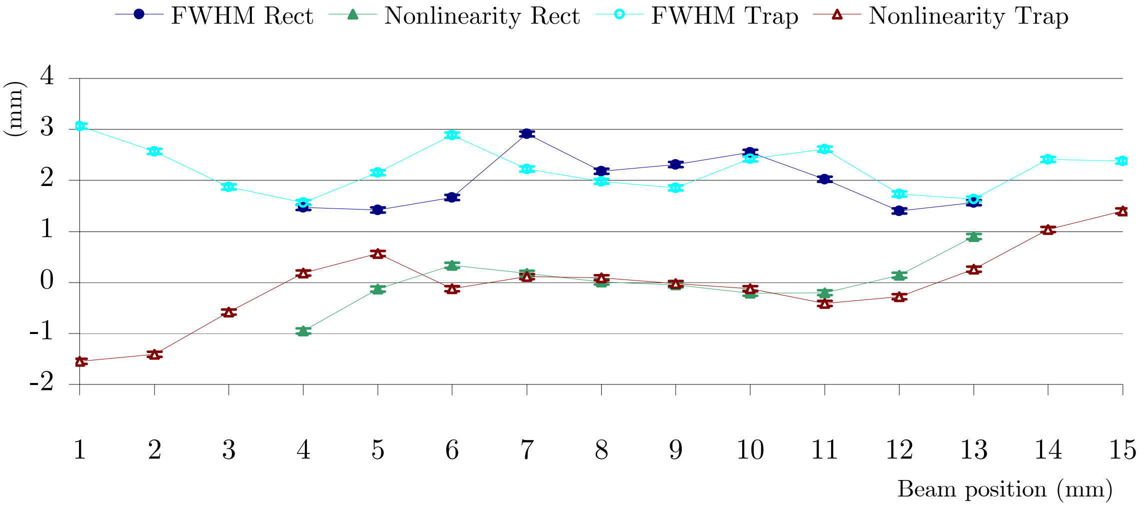


Figure 3.22: FWHM resolution and nonlinearity along the short side of a rectangular 20x10x20mm^3 LSO block and a 20x15.4(11.5)x20mm^3 trapezoidal LSO block

The 20x15.4(11.5)x20mm^3 trapezoidal LSO block has sloped sides on the short axis (Figure 3.4 (b)). When collecting training data for this block, the photon beam is also positioned on these slanted sides in order to obtain incidence positions for photons entering the crystal at those surfaces. Figure 3.22 shows the average FWHM resolution obtained using LM trained NNs along the short axis of a 20 mm thick rectangular LSO block and along the short axis of the trapezoidal LSO block. On the flat top surface the FWHM resolution is similar for both geometries. On the slanted sides of the trapezoidal block, the resolution degrades slightly towards the very end near the APD plane. The spatial nonlinearity also shows comparable properties. This could be due to the fact that photons impinging at the edges, interact very close to the bottom where the LSO block slightly overhangs the APD.

Chapter 4

Evaluation of block detectors on the PET prototype demonstrator

Previous chapter demonstrates that a detector module based on a 20x10x10mm^3 LSO block and APD readout, combined with a Levenberg-Marquardt neural network positioning algorithm can achieve a measured detector resolution better than 2mm FWHM for perpendicular incidence photons. When different LM-NNs are trained for several incidence angles, the resolution degradation due to photon penetration for non-perpendicular incident photons can be limited. For an incidence photon angle of 30°, the measured detector resolution is still better than 2.1mm FWHM.

However, the data acquired for this study was done on an "academic" bench set-up. For instance, the electronic collimation was done by a 35 mm thick BGO crystal coupled to a PMT, with a ∅5mm Pb-collimator (see Figure 3.5) instead of two LSO-APD detector modules. Also the use of the Cremat CR-110 preamplifiers (2.24x2.16cm) is not feasible in a real PET system due to their large dimensions and their big power dissipation (~100mW). To maximize the SNR, these preamplifiers are located at the output of the APD. Accordingly, Peltier elements are needed to keep the detector modules at room temperature.

In order to evaluate the block detectors in a real compact PET environment, a prototype PET demonstrator was built. The demonstrator consists of only two 20x10x10mm^3 LSO detector modules. To simulate a full-ring scanner, the detector modules are mounted on separate rotating platforms which allow the movement of both detector modules, also relative to each other.

In addition, since the detector characteristics may change in time, it is also appropriate to acquire new training data from time to time. The use of an auxiliary bench set-up for this calibration procedure implies the removal, calibration and re-mounting of all detector modules of the scanner. This would be a time consuming and tedious task. That's why an automated acquisition method of training data for the positioning algorithm is investigated. The implementation and validation of this procedure is also done on the demonstrator set-up.

4.1 New front-end electronics for APD readout

As mentioned before, the CR-110 Cremat preamplifiers in the detector modules used on the bench set-up were bulky and produced too much heat. Therefore new front-end electronics were designed.

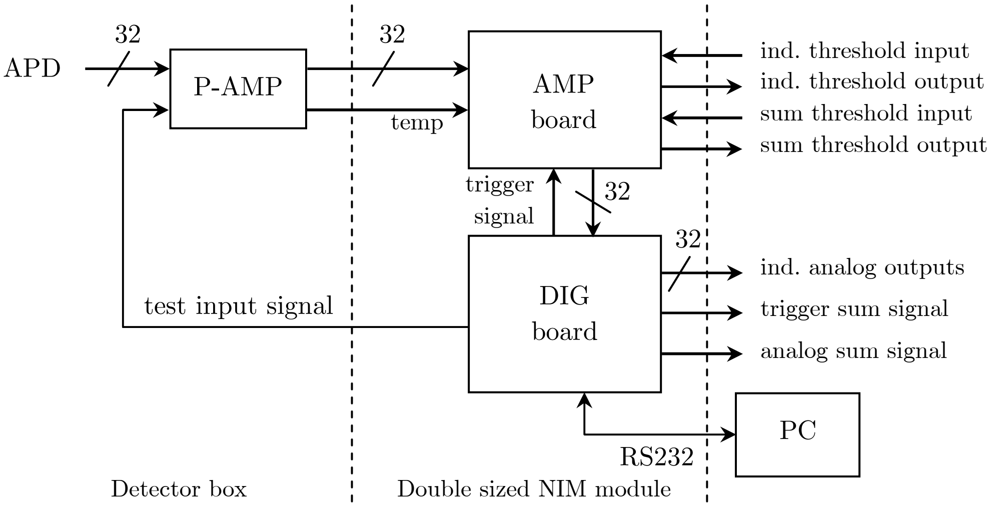


Figure 4.1: Scheme of the new front-end electronic built up of three parts: a pre-amplifier -, amplifier - and digital board.

The design consists of three boards: a preamplifier board (called P-AMP board), an amplifier board (called AMP board) and a digital board (called DIG board). The latter two boards were designed as NIM boards [64].

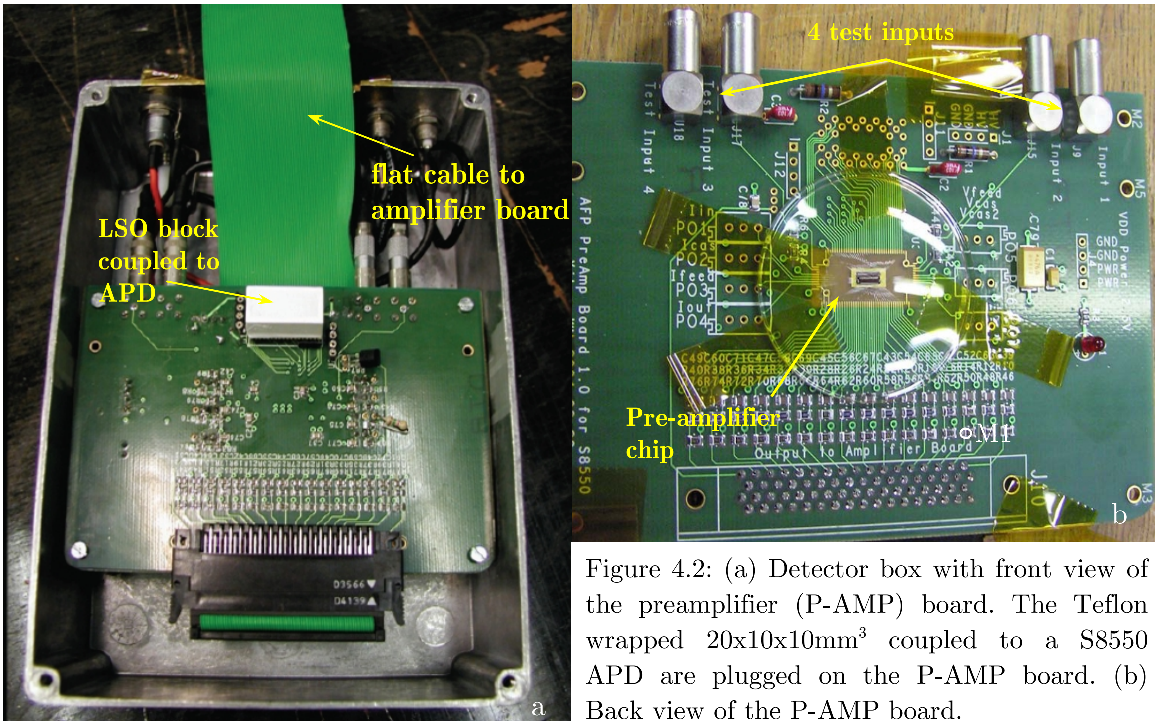


Figure 4.2: (a) Detector box with front view of the preamplifier (P-AMP) board. The Teflon wrapped 20x10x10mm^3 coupled to a S8550 APD are plugged on the P-AMP board. (b) Back view of the P-AMP board.

4.1.1 Pre-amplifier board (P-AMP)

The pre-amplifier board and the amplifier board are linked by a flat cable (Figure 4.2 (a)). The compact P-AMP board (11.4cmx8cm) houses the 20x10x10mm^3 LSO block, the S8550 APD and a pre-amplifier chip. The chip was manufactured in CERN for silicon strip readout in the NA60 experiment [65].

The design on the P-AMP board is such that the APD is directly coupled to the pre-amplifier chip to minimize the noise (Figure 4.3). The preamplifier chip is a transimpedance pre-amplifier (current-to-voltage converter). In general a transimpedance pre-amplifier is faster than an integrating pre-amplifier because the integrating amplifier needs time to discharge.

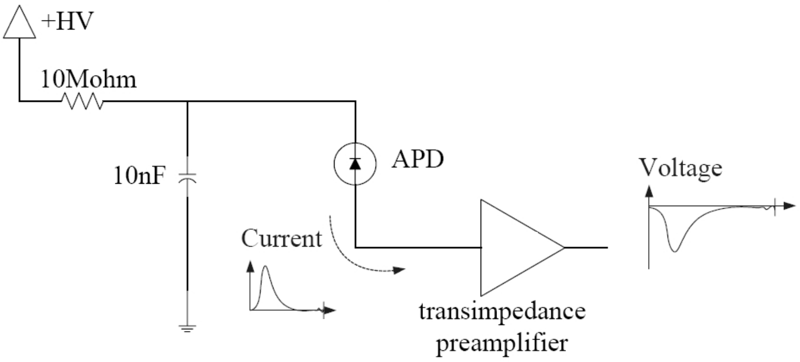


Figure 4.3: Pre-amplifier board diagram. The current-to-voltage pre-amplifier doubles the decay time due to its square root response function.

Four test input connections are available on the P-AMP board (Figure 4.2 (b)). They can be used to calibrate the square root response of each of the 32 APD channels. Besides, they are useful to verify the gain uniformity over the 32 APD channels.

4.1.2 Amplifier board (AMP)

This board includes two stages of amplification, a discriminator and a shaper for each channel.

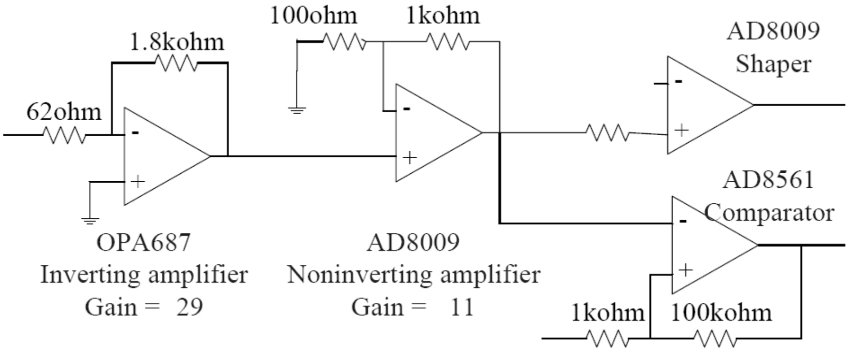


Figure 4.4: Amplifier board diagram. Each channel passes through two amplification stages, a discriminator and a shaping circuit.

Each of the 32 channels passes through a two stage amplification which provides a total gain of 319. An individual trigger per channel is generated by a discriminator when the amplified APD pixel signal crosses the preset threshold level. The shaping time can be switched from 30ns to 100ns. And finally, offset adjustment is possible through a potentiometer which is mounted at the start of each amplification channel (Figure 4.5).

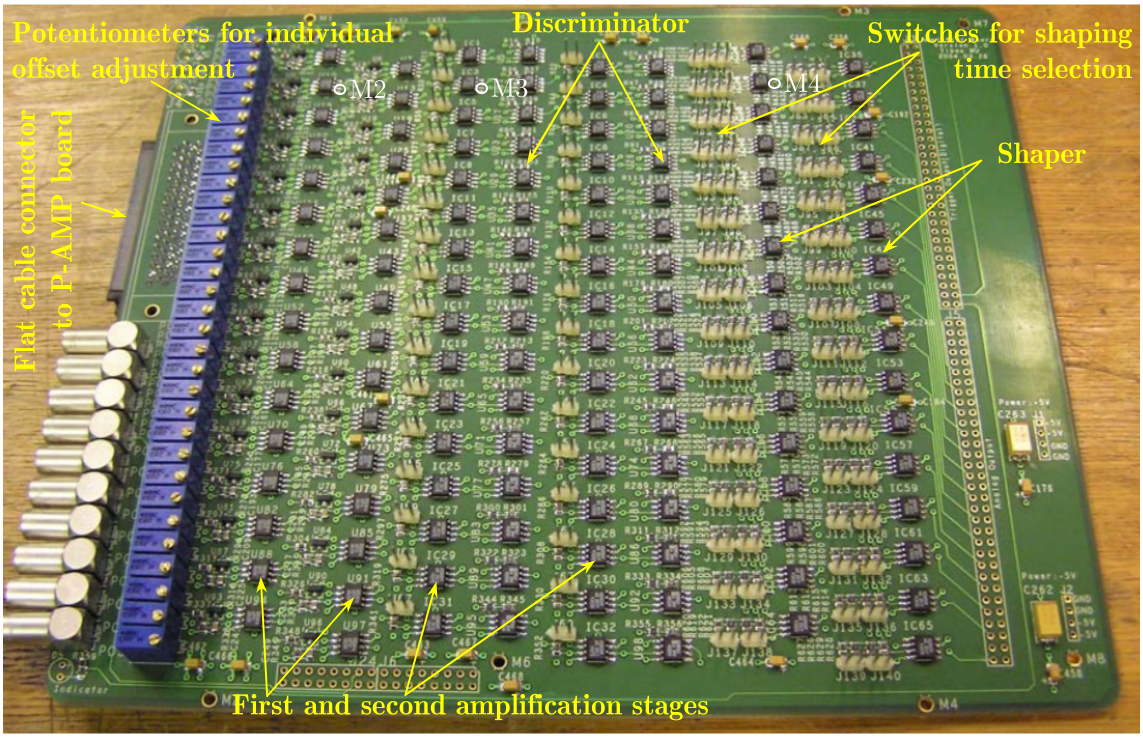


Figure 4.5: The amplifier board. Each channel passes two amplification stages, a discriminator and a shaping circuit. The shaping time can be selected by use of switches. Offsets can be modified by use of potentiometers.

4.1.3 Digital board (DIG)

This board includes an analog part as well as a digital part. The digital part involves a FPGA and a microcontroller. They perform threshold and test pulse generation in addition to system monitoring (Figure 4.6). A RS-232 interface provides the communication between the DIG board and computer. The analog part consists of an analog sum circuit that adds all 32 shaped APD signals into one sum signal. This sum and the corresponding trigger are sent out at the front of the DIG board. Also available at the front are the 32 analog shaped outputs which can be connected directly to a DAQ system using integrating or peak-sensing ADCs.

4.1.4 Test pulse evaluation

The test pulse is generated on the DIG board by the microcontroller and injected directly into the P-AMP chip. Figure 4.7 shows the test pulse outputs at different stages within the amplification process. M1-M4 indicates the reference points and can be found on Figure 4.2.b and Figure 4.5.

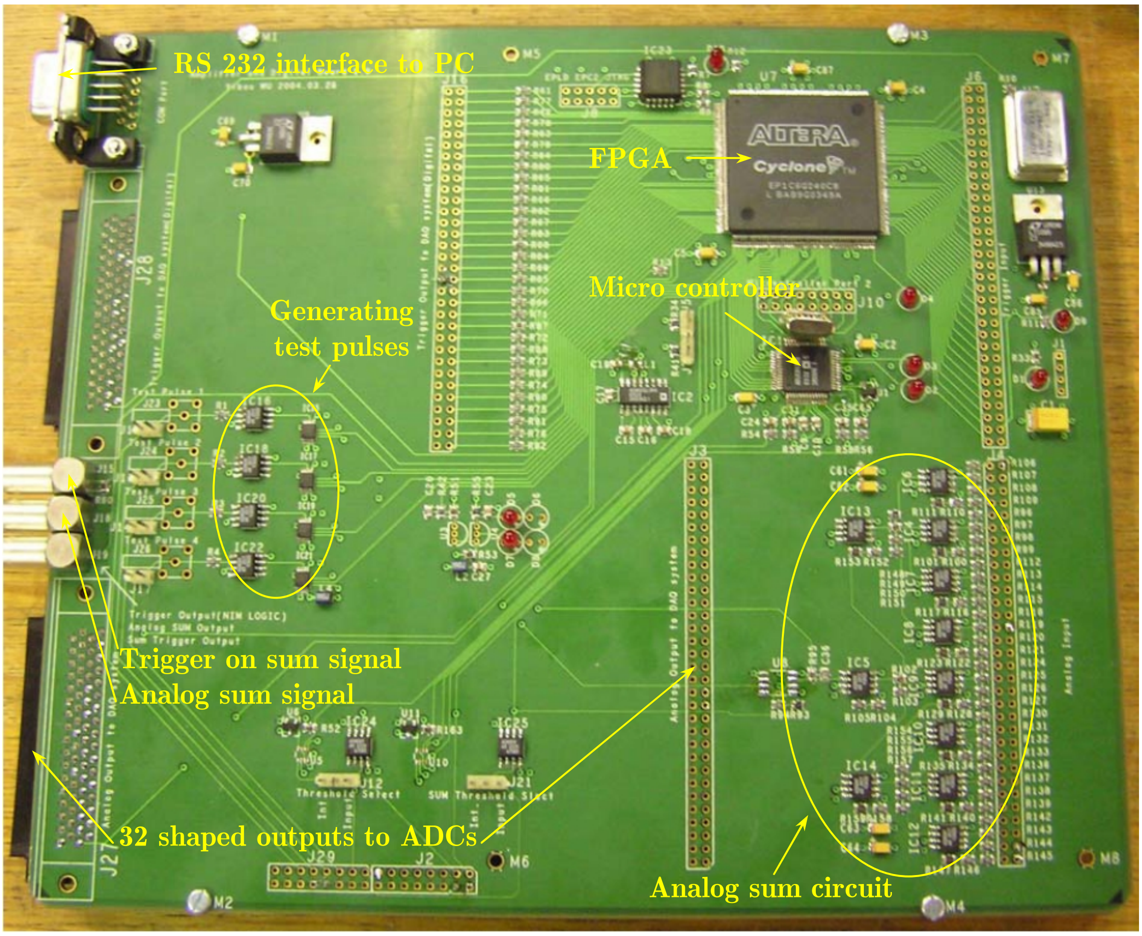


Figure 4.6: The digital board, incorporate analog circuits as well as a digital part.

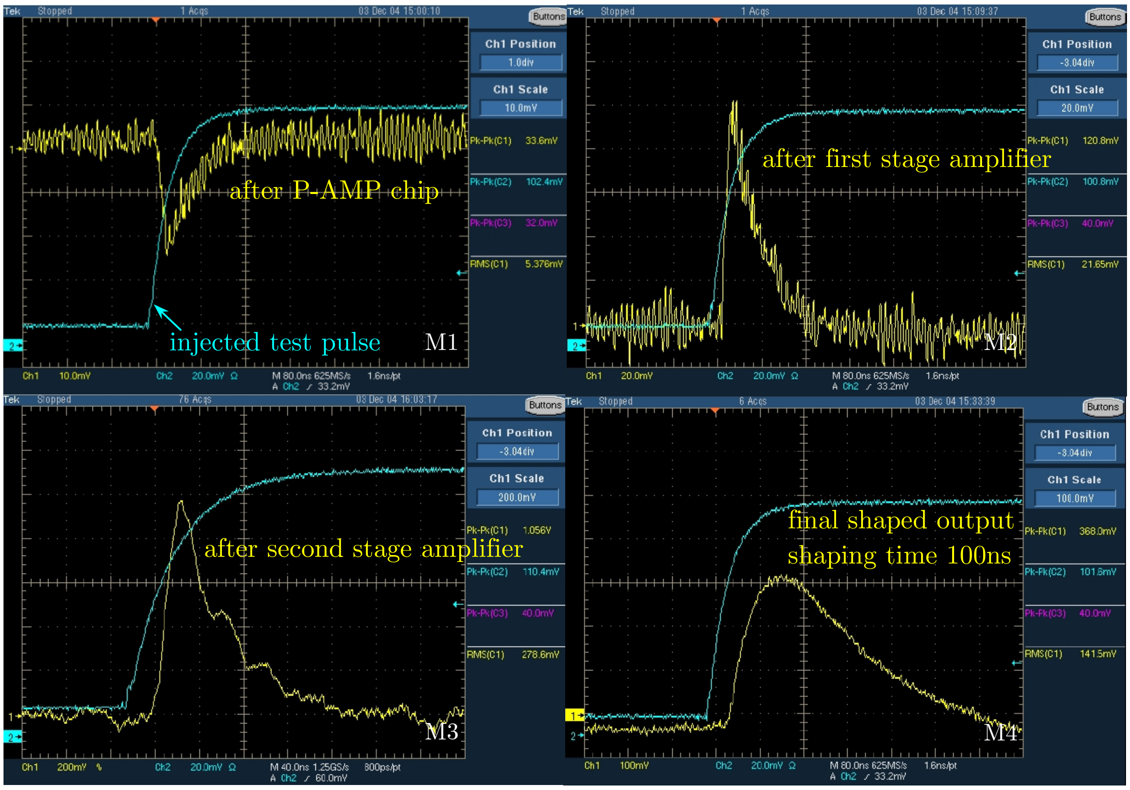


Figure 4.7: Test pulse output at different positions in the amplification process. The positions of M1-M4 can be found on Figure 4.2.b and Figure 4.5

The RMS noise without and with APD is measured and is 5mV and 7mV respectively. Since the preamplifier is not linear, the total system gain is also not linear. Due to the square root response of the P-AMP chip, larger inputs have lower gain compared to smaller inputs. Due to this dependency, it is not straight forward to calculate the equivalent noise charge (ENC). However, the RMS noise is defined at low inputs, i.e. high gain. The ENC can be estimated if the gain is assumed to be constant over a small input range. A small test pulse of 27.2 mV is injected and a shaped output peak value of 120mV is measured. The preamplifier chip is a transimpedance amplifier thus the voltage test pulse is converted to charge by a 100fF capacitor. So the gain is found by 120mV/(27.2mV x 100fF) $\approx$ 44mV/fC where F=C/V. The ENC can then be calculated as:

$$\begin{gathered}\mathrm{ENC}_{\mathrm{without}}=5\mathrm{mV}/(44\mathrm{mV/fC}) \approx 0.144\mathrm{fC} \approx 700\mathrm{e}^{-} \\ \text{and} \\ \mathrm{ENC}_{\mathrm{withAPD}}=7\mathrm{mV}/(44\mathrm{mV/fC}) \approx 0.159\mathrm{fC} \approx 1000\mathrm{e}^{-}\end{gathered} \tag{4.1}$$

4.2 Data acquisition system of the demonstrator

The data acquisition (DAQ) system is summarized in Figure 4.9. The demonstrator set-up consists of two rotating platforms onto which two detector boxes are mounted (see Paragraph 4.3, Detector movement for a full-ring simulation). The detector modules contain a 20x10x10 mm^3 Teflon wrapped LSO block mounted on a S8550 Hamamatsu APD array. As mentioned in previous section, the APD is coupled directly to the P-AMP chip to minimize the noise. Through a flat cable the 32 pre-amplified analog APD channels are fed into the AMP board (Figure 4.9:1). After the 2 stage amplification and shaping, the 32 semi-Gaussian pulses are split to a sum module and to the peak sensing ADCs via a twisted pair flat cable (Figure 4.9:2 and Figure 4.8). In previous section an onboard summing circuit with trigger capability was described on the DIG board. Unfortunately this circuit did not function properly. Therefore we used an external summing module which was already designed for a previous DAQ set-up. A trigger signal in the form of a 10ns pulse is send out (Figure 4.9:4) when the analog sum of the 32 APD signals (Figure 4.9:3) crosses a preset threshold level in the discriminator (Lecroy mod. 4608). If both triggers of the detector modules overlap, a 500ns gate activates the peak sensing ADCs (CERN AD811) digitizing the 2x32 APD channels from both detector modules (Figure 4.9:5). The interface between the ADCs and the computer is managed by a Wiener CC32 controller (Figure 4.9:6&7) and a Labview program running on the PC. Moreover, the program controls the whole set-up: motors and ADCs initialization, motor movement, spectrum analyses and software thresholding, event counting as well as event storage.

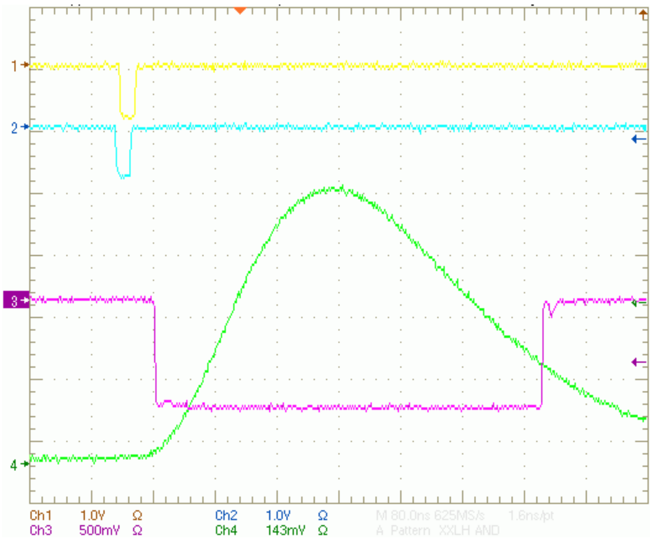
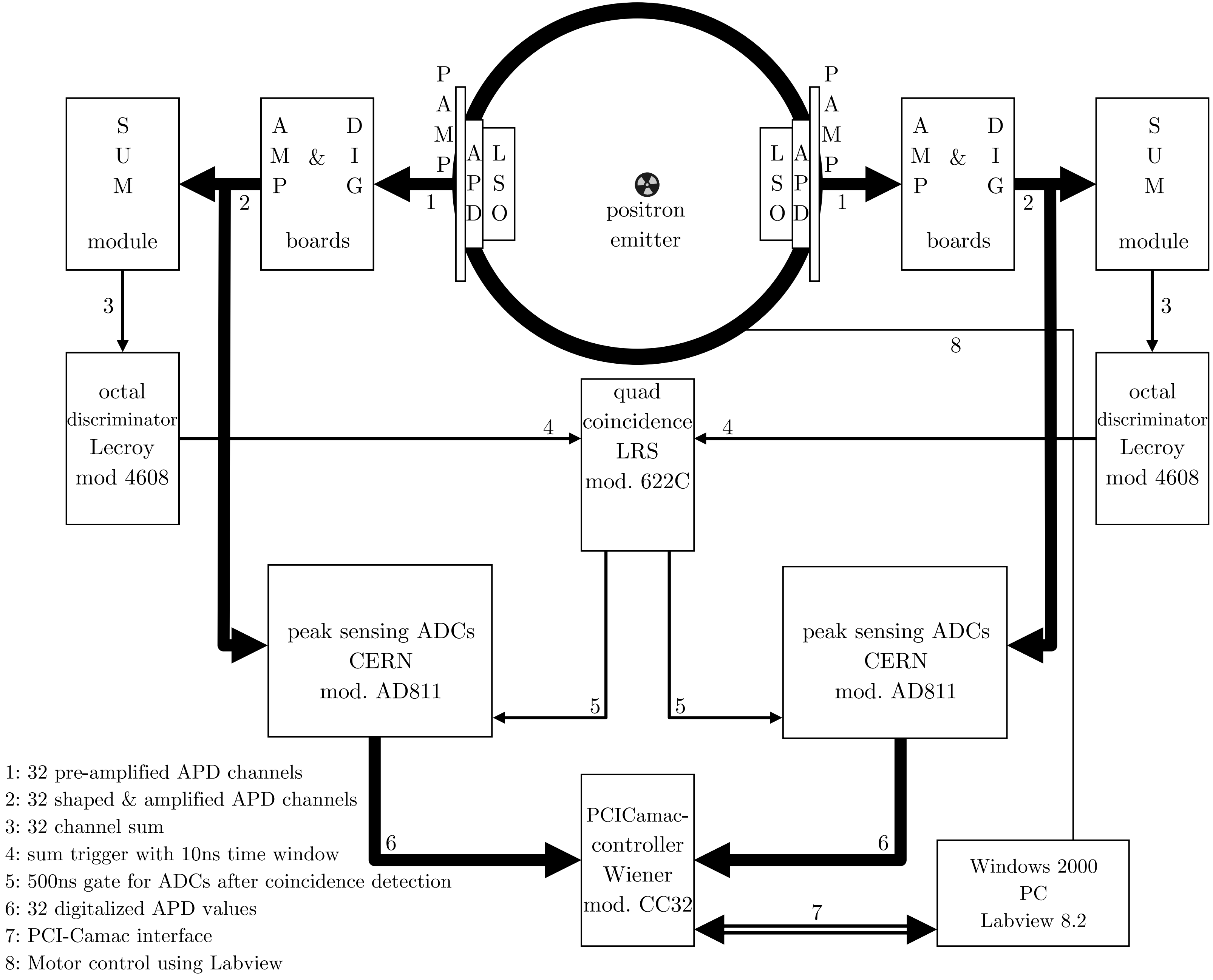


Figure 4.8: CH1 and CH2 are the sum triggers of both detector modules with a time window of 10ns each. If both triggers coincide, 500ns gates (CH3) activate the peak sensing ADCs digitizing the peak value per channel (CH4) within this 500ns.



1: 32 pre-amplified APD channels
2: 32 shaped & amplified APD channels
3: 32 channel sum
4: sum trigger with 10ns time window
5: 500ns gate for ADCs after coincidence detection
6: 32 digitalized APD values
7: PCI-Camac interface
8: Motor control using Labview

Figure 4.9: Schematic representation of the prototype demonstrator set-up.

4.3 Detector movement for a full-ring simulation

Consider a full ring PET system consisting of 48 $20x10x10mm^3$ LSO block detectors numbered from 1 to 48 as illustrated in Figure 4.10. A pair of detector modules operating in coincidence will acquire data from LORs lying within a diamond-shaped region in the sinogram, as clarified below. To recapitulate, a line-of-response or LOR is the line joining both incident photon positions measured in coincidence. And a sinogram is an ordered way of storing these LORs in function of their position x_r (e.g. distance of LOR to the center of the scanner) and orientation ϕ (e.g. angle between LOR and x-axis) (Figure 4.10 (left)).

The origin of the diamond shaped segment is also demonstrated in Figure 4.10 (right). The LORs a_1, b, a_5 and c form the boundaries of the sinogram segment, defined by coincidences between detectors 1 and 25. All possible LORs measured between both detector modules, such as LOR d, can be stored within this segment.

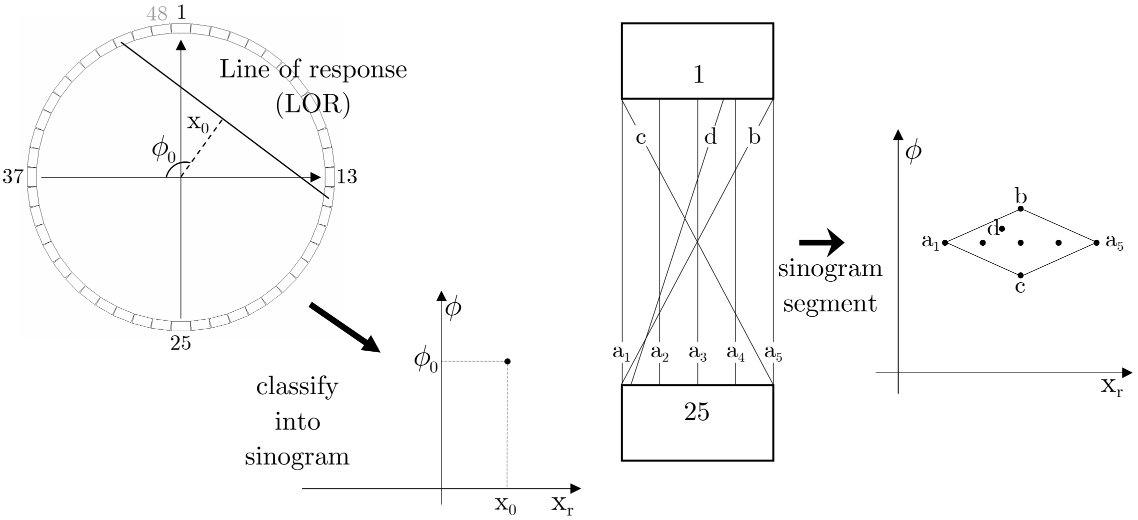


Figure 4.10: (left) A full-ring PET system with 48 detector modules. A LOR can be characterized by the parameters x_r and ϕ and organized into a sinogram. Each LOR (x_0, ϕ_0) defines a point in the sinogram. (right) Illustration of a diamond-shaped sinogram segment for the two opposite detector modules 1 and 25. The points a_1, b, a_5 and c form the boundaries of the segment. All LORs measured in coincidence by both detector modules will be within this diamond shape.

Figure 4.11 shows the central part of a sinogram where the sinogram is built up from the different segments associated with each possible detector combination. As demonstrated in Figure 4.10, the boundaries of the sinogram segment for detector combination 1-25 are a_1, b, a_5 and c. For detector combination 1-24, LORs c and c_1 outline the radial deviation. The angular limits are given by LORs a_5 and e as seen on Figure 4.11 (left). Finally, combining all possible detector pairs results in a sinogram covering the whole FOV of the scanner. Typical, the radial distance x_r will vary between $-x_{max}$ and $+x_{max}$ where x_{max} is given by the radius of the FOV while the angular ordinate ϕ is restricted from 0 to Pi radians. LORs with $\phi > \pi$ are convert to an angle smaller than π through the transformation $(x_r, \phi) \Leftrightarrow (-x_r, \phi - \pi)$

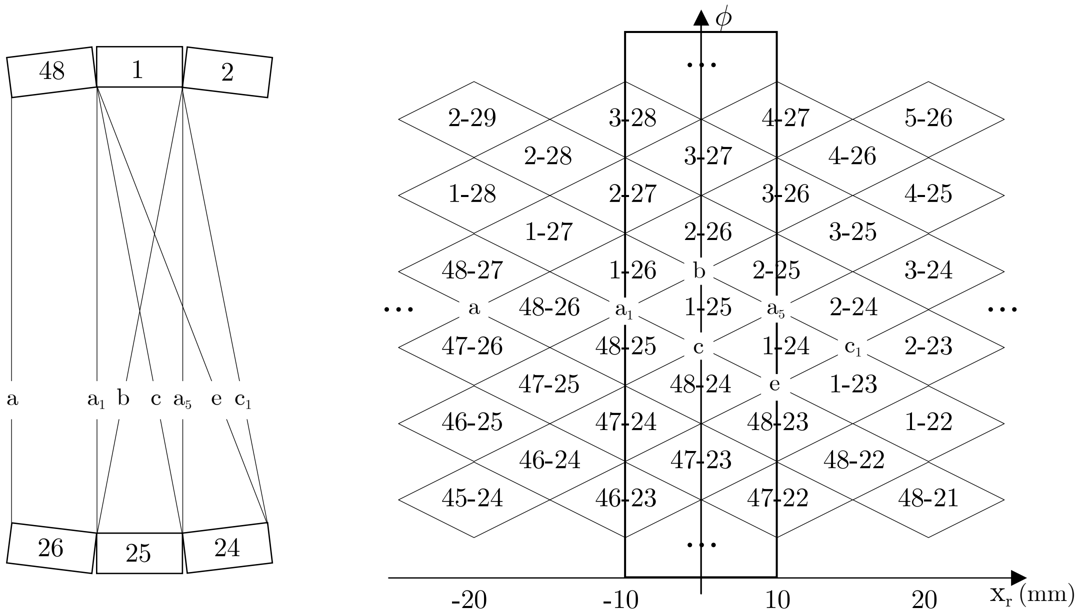


Figure 4.11: Central part of a sinogram where each detector combination is associated with a corresponding diamond segment within the sinogram. LORs a_1, b, c, a_5 and c, a_5, c_1, e respectively give the boundaries of the diamond shaped segments for detector combination 1-25 and 1-24. The rectangle encloses all LORs needed for a complete acquisition (all radial as well as angular samples) of a 2cm FOV.

On the demonstrator setup, the sinogram is progressively assembled by collecting data from all possible detector combination in a step-and-shoot mode. However, only LORs within the field-of-view (FOV) are mandatory. Nevertheless, for a correct image reconstruction, all radial samples are required at each angular projection [66]. For a single diamond region, this condition is only satisfied for the radial sample at the center of the sinogram (i.e. the line a_1-a_5 in diamond 1-25 on Figure 4.11). For a FOV of 2cm, where the radial samples x_r vary from -10mm to 10mm, data has to be acquired for all LOR within the rectangle on Figure 4.11. This means that all detector combinations with a detector difference of $\Delta_{D_i\text{-}D_j} = 23$, 24 and 25 have to be measured. In particular, full diamonds of the detector combinations 1-25 to 48-24 and half-diamonds of the detector combinations 1-24 to 48-23 and 1-26 to 48-25 are used for image reconstruction.

To simulate a full-ring scanner, two detector modules, D_0 and D_1, were mounted on two rotating platforms (Figure 4.12). The bottom platform rotates D_0 and D_1 simultaneously over 360°. The upper platform rotates detector D_1 relative to D_0. These two rotational movements allow the simulation of a full-ring system, i.e. acquire all necessary LORs. The diameter of the ring is 300mm. Each detector box contains a 20x10x10 mm^3 Teflon wrapped LSO block mounted on the S8850 APD array. The detector boxes also house the necessary front-end pre-amplifier electronics as described in section 4.1.1.

The alignment of both detector modules was accomplished by a 20x10x300mm^3 aluminum bar placed on the Ωz-stage as illustrated on Figure 4.12 (right). Centering both detector modules at each extremity of the bar, the alignment is assured. The

Ωz-stage comprises a compact linear stepping motor (DRL60, OrientalMotor) for axial displacement and a rotary stepping motor (DG60, OrientalMotor) for the rotation of the object under study. In fact, this angular motion is identical to the rotational movement of the lower demonstrator's platform that rotates both detector modules simultaneous. But due to its higher accuracy, the angular motion of the Ωz-stage is used to rotate the study object relative to both detectors instead of vice versa.

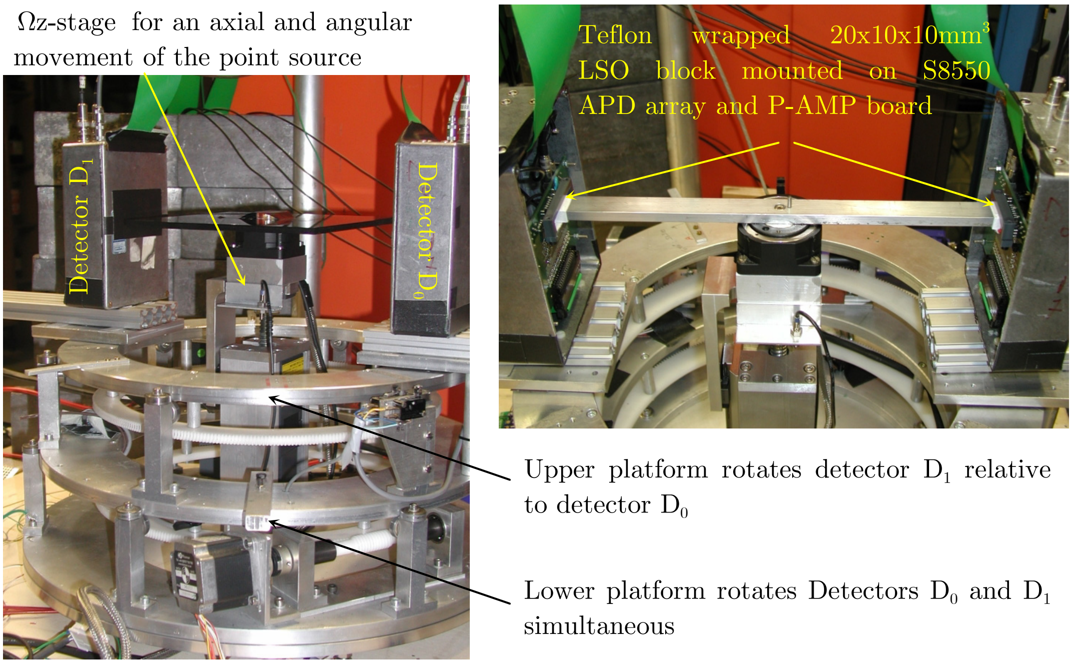


Figure 4.12: (left) PET prototype demonstrator consisting of two detector modules mounted on rotation platforms allowing the simulation of a full-ring system.
(right) A 20x10x300mm^3 aluminum bar used for the alignment of both detector modules.

As described above, one detector combination determines a diamond-shape segment in the sinogram. As the source rotates in a step-and-shoot mode, this diamond-shaped segment is displaced down the ordinate ϕ of the sinogram. The shift of the segment is defined by the angular step size α of the Ωz- stage. This is demonstrated in Figure 4.13. The shaded diamonds illustrate the different positions for the detector combination 1-25. The angular step α is equal to one detector width such that there is no overlap between the different positions. However, it is apparent that not all radial samples are acquired at each projection as is required for a correct image reconstruction. To overcome this issue and to attain a larger FOV, different detector combinations have to be incorporate. Indeed, the number of positions taken by D_1 relative D_0 determines the FOV. For example, Figure 4.13 shows the use of five detector combinations resulting in a FOV of approximately 40mm. To achieve a uniform acquisition over all radial samples only LORs within the dotted lines are accepted, leading to a parallelogram in the sinogram. Rotation of the source will again move this shape over all angular samples ϕ of the sinogram, needed for the image reconstruction.

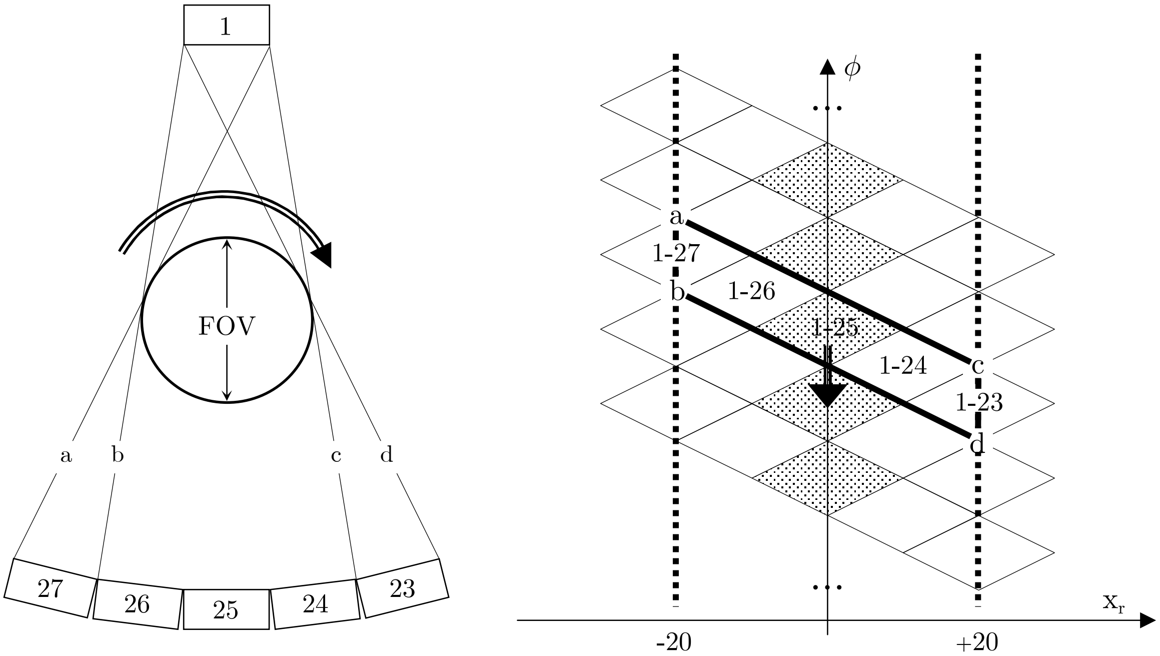


Figure 4.13: The number of detector combinations at which data are acquired determines the diameter of the FOV. In this example, five detector combinations result in a FOV of 40mm. In order to obtain a uniform acquisition over all radial samples, only data within the dotted lines are accepted. For the five detector combinations 1-27 to 1-23, a parallelogram shaped segment is formed. As the source rotates, this segment is displaced down the ordinate ϕ of the sinogram and fills all angular samples needed for image reconstruction.

4.4 An automated method of training data acquisition

The scan procedure described in previous paragraph results in the complete measurement of all possible projections within the FOV. Each recorded coincidence event consists of both actual detector positions, the angular position of the Ωz- stage and the 2x32 APD sampled light distribution of both LSO blocks. As introduced in Paragraph 3.4, neural networks trained at different angles can extract a nearly parallax-free incidence photon position from this scintillation light distribution if the impinging photon angle is known. However, in a scanner geometry, like a full-ring design, the selection of the appropriate neural network can also be based on the detector combination involved in the coincidence detection. In fact, if the centers of the scintillator blocks firing in coincidence are used to estimate the incidence angle, the maximum uncertainty on the estimated incidence angle depends on the scintillator block size and ring diameter. The biggest discrepancy with the true incidence photon angle is ± 3.8 degrees for a PET-system with a ring diameter of 300mm using 20x10x10mm^3 LSO blocks. Larger ring diameters result in a lower uncertainty as the uncertainty is given by the width b of the LSO block and the radius R of the scanner

$$\Delta = \tan^{-1}[b/R] \tag{4.2}$$

To study the influence of a mismatch between the true incidence angle and the estimated incidence angle, neural networks trained at a specific angle are evaluated with data obtained at incidence angles 5 degrees below and above the incidence angle the neural networks were trained for (more information about NN positioning estimation, see Paragraph 3.4). The effects of these inappropriate angle selections are shown in Table 4.1. The FWHM degradation due to the incidence angle mismatch is on average 0.1mm and is smaller when the incidence angle is underestimated. A similar trend is observed for the FWTM with an average degradation of 0.3mm. Consequently, no big deterioration is observed between the positioning capabilities of neural networks trained with the true incidence angle and neural networks trained with an incidence angle mismatch of 5 degrees.

Table 4.1: FWHM and FWTM resolutions for NNs evaluated with an incidence angle of -5°, 0° and 5° from the training incidence angle. The energy threshold was set at 100keV

	Angular mismatch					
Training angle	FWHM			FWTM		
	-5°	0°	+5°	-5°	0°	+5°
0°		1.8mm	2.0mm		4.3mm	5.0mm
5°	2.0mm	1.9mm	2.1mm	4.7mm	4.8mm	5.2mm
10°	2.0mm	2.0mm	2.2mm	4.9mm	5.0mm	5.3mm
15°	2.0mm	2.0mm	2.2mm	5.2mm	5.2mm	5.5mm
20°	2.1mm	2.1mm	2.2mm	5.3mm	5.1mm	5.6mm
25°	2.1mm	2.1mm	2.3mm	5.4mm	5.3mm	5.8mm
30°	2.2mm	2.0mm		5.7mm	5.5mm	

Assuming a ring geometry with 20x10x10mm^3 LSO blocks, an uncertainty on the estimated incidence angle less than ± 5 degree per block is already met if the ring radius is larger than 115mm and the center of the detector blocks are used to estimate the incidence angle. In consequence, on the demonstrator set-up only one neural network per detector combination has to be trained.

The training process of the neural networks requires a set of samples that link the measured scintillation light distribution to the known photon incidence position. Like on the bench set-up, (see Paragraph 3.2), the acquisition of such a data set is done by placing a 250 µm ^{22}Na point source (50 µCi) very close to the detector module that must be trained, i.e. detector D_{train}. The impact position of the impinging photon is confined to a very small region through electronic collimation with the second detector D_{opp} (see Figure 4.14 left, dotted lines). The beam width scales with the ratio of the distance D_{train}-source and the distance source-D_{opp}. On the demonstrator set-up, this ratio was about 1/29 resulting in a beam width of less than 1mm FWHM. This photon beam is stepped over the surface of the LSO block in steps of 500 µm in both x (coincidence interval, see Figure 4.14) and y direction (block height). At each position on this grid, 150 events were measured to train and evaluate the NN. For each coincidence, the source position and the 32 digitized APD pixel values of D_{train} are saved. The grid acquisition in 500 µm steps is repeated for each detector combination within the FOV. The different detector combinations are obtained by rotating detector D_{opp} relative to D_{train} as illustrated on Figure 4.14 (left).

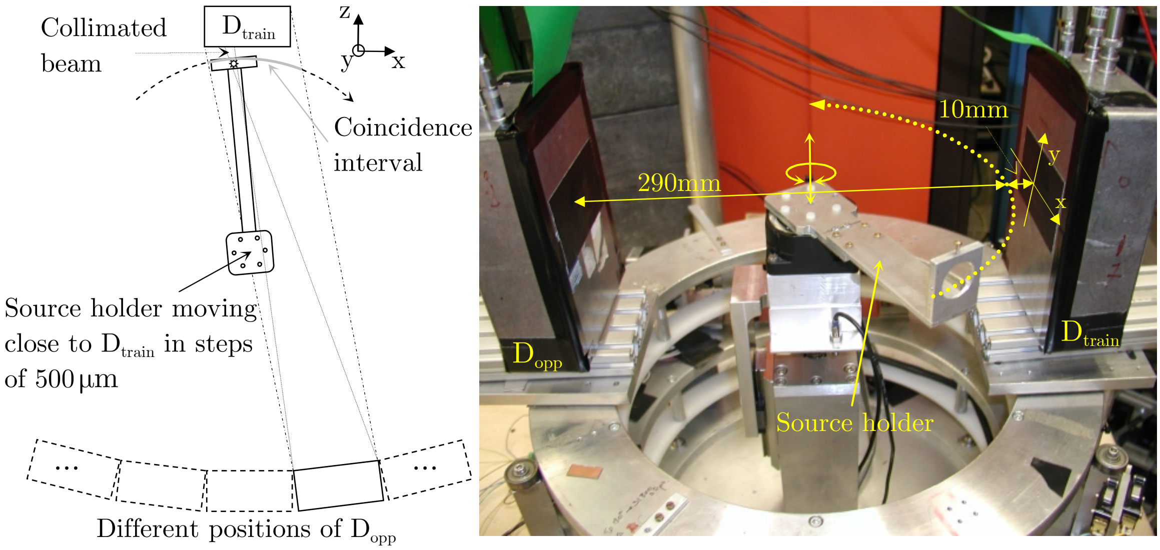


Figure 4.14: (left) Schematic representation of the demonstrator set-up for the automatic training data acquisition. The source is moving over a grid in steps of 500 µm through the coincidence interval (x direction) and the 10mm block height (y direction). The different detector combinations are obtained by moving D_{opp} relative to D_{train}. (right) Illustration of the demonstrator set-up. In order to minimize the collimated beam width, the source platform shown on Figure 4.12 is replaced by a source holder. The beam width scales with the ratio of the distances D_{train}-source and source-D_{opp} if the dimensions of the point source are negligible. In our setup, the ratio on the demonstrator is 1/29 which reduces the beam width to about 1mm FWHM.

The start and stop positions of the coincidence interval per detector combination are calculated relative to the measured start and stop positions of detector combination 1-25. Only for this detector combination the edges of the block can be estimated from symmetry considerations. Given the mean count rate profile of the whole block (dashed line in Figure 4.15, right) and knowing the length of the crystal block (20mm), both edges of the block should have the same count rate. Accordingly, the two positions, 20mm apart, having the same rate correspond to the edges of the block. Figure 4.15 illustrates the count rate profiles acquired during the raster scan for the acquisition of training data for detector D_0 and detector combination 1-25.

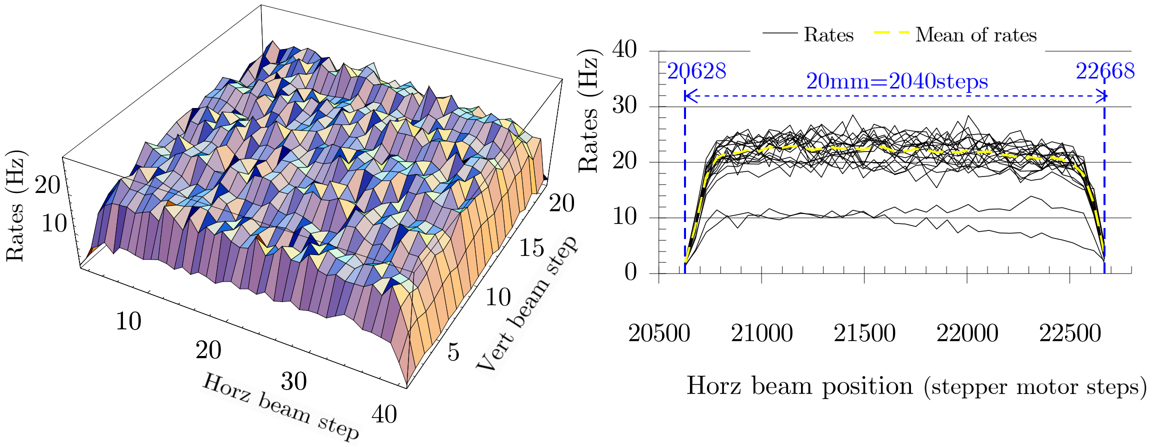


Figure 4.15: (left) 3D representation of the measured count rate profiles during the raster scan of detector D_0 for detector combination 1-25. (right) 2D plot of the same count rate profiles. The dashed line represents the mean value over all rates. The start and stop positions are the points, 20mm apart with an identical count rate.

The start and stop positions of detector combinations 1-25 to 1-29 are calculated as follow:

$$\begin{aligned} \text{Start} &= \text{start}_{1\text{-}25}\text{-}20.\text{Tan}[(\text{DetD}_{1_{\text{Nb}}}\text{-}25).\frac{\pi}{48}] \\ \text{Stop} &= \text{stop}_{1\text{-}25}\text{-}10.\text{Tan}[(\text{DetD}_{1_{\text{Nb}}}\text{-}25).\frac{\pi}{48}] \end{aligned} \qquad \text{for DetD}_{1_{\text{Nb}}}\text{: } 25..29 \tag{4.3}$$

and for detector combinations 1-21 to 1-25, the start and stop positions are computed with:

$$\begin{aligned} \text{Start} &= \text{start}_{1\text{-}25}+10.\text{Tan}[(25\text{-DetD}_{1_{\text{Nb}}}).\frac{\pi}{48}] \\ \text{Stop} &= \text{stop}_{1\text{-}25}+20.\text{Tan}[(25\text{-DetD}_{1_{\text{Nb}}}).\frac{\pi}{48}] \end{aligned} \qquad \text{for DetD}_{1_{\text{Nb}}}\text{: } 21..25 \tag{4.4}$$

with $\text{start}_{1\text{-}25}$ and $\text{stop}_{1\text{-}25}$ respectively the start and stop positions for detector module D_0 at detector combination 1-25. The factors 10 and 20 are clarified in Figure 4.16 and represent the distance between the front of the block and source or the distance between the backside of the block and source.

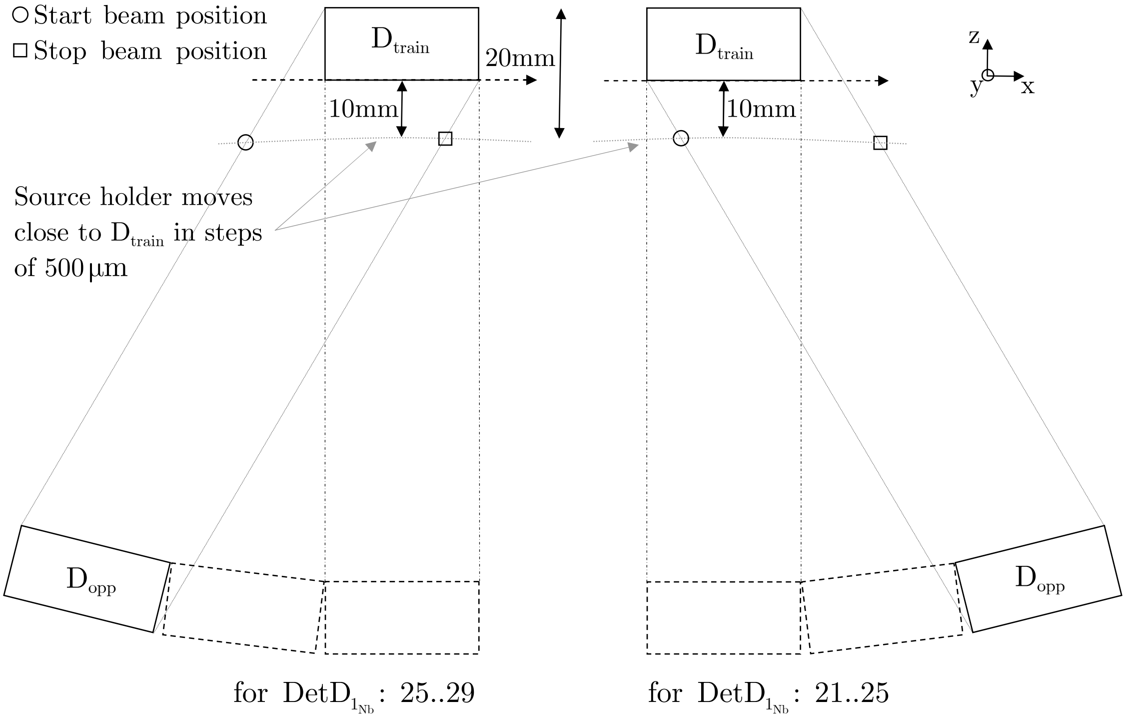


Figure 4.16: Schematic representation for the calculation of the start and stop positions for the different detector combinations relative to the detector combination 1-25.

Figure 4.17 summarizes the average count rate profiles for the different detector combinations 1-21 to 1-29 needed for the training of the NNs for detector module D_0 for all incidence angles needed. The start and stop positions are shifted to the right for detector combinations 1-21 to 1-24 and shifted to the left for the combinations 1-26 to 1-29. The calculation of the start and stop positions and the count rate profiles of the training data for detector module D_1 are totally identical.

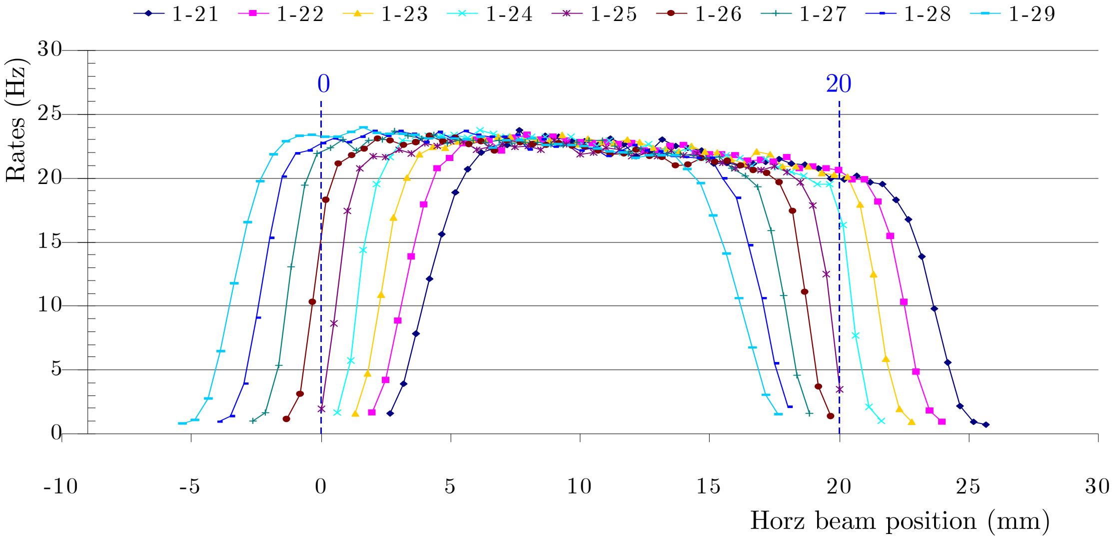


Figure 4.17: Summary of the average count rate profiles for all acquired detector combinations 1-21 to 1-29 needed for the training of detector module D_0.

Similar to the bench approach, the acquired data is then split into a training set and a validation set. The training set is used to train a neural network in the Mathematica Neural Network package [61], where the known source position is used as an approximation of the impinging photon position.

For each event in the validation set, the positioning error, defined as the difference between the estimated and the true position, is calculated. To obtain an estimate for the spatial resolution, these positioning errors are binned in a histogram. The FWHM and FWTM are again used to evaluate the spatial resolution. However, due to the statistical nature of the data sets and the random initialization of the NN prior to training, retraining the NN with a training set obtained in identical circumstances will result in a different NN with slightly different performances. Therefore, five different pairs of training sets and test sets were created. Each pair was used to train a NN. The resulting NNs were then evaluated using the same validation set. The averages of the FWHM and the FWTM of these five histograms are used as estimates for the spatial resolution. As an estimates of the uncertainty, the sample standard deviations of the mean FWHM and FWTM are used and are all smaller than 0.05mm.

Figure 4.18 shows the FWHM and FWTM of the spatial resolution for both detector modules D_0 and D_1 on the demonstrator set-up trained with data acquired via a fully automatic acquisition process written in Labview. For a FOV diameter of 80mm, nine consecutive detector combinations have to be trained per detector module. For detector module D_0, the combinations are 1-21, 1-22, 1-23, …, 1-29. The combinations 45-25, 46-25, 47-25, 48-25, 1-25, …, 5-25 are acquired to train detector module D_1. The corresponding incident photon angles per detector combination, defined as the line joining the centers of both detectors modules vary from -15degree to 15degree in steps of 3.75degree.

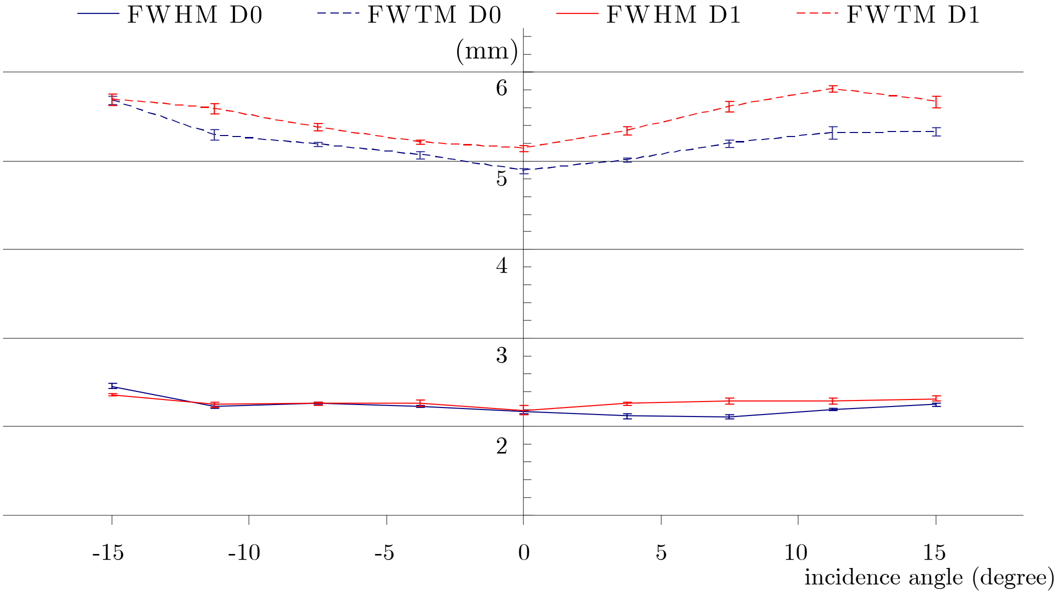


Figure 4.18: FWHM and FWTM resolutions achieved with Levenberg-Marquardt trained neural networks, in function of the different detector combinations of both detector modules D_0 and D_1 involved on the demonstrator set-up.

For detector combination 1-25 acquiring almost perpendicular incident photons, the measured FWHM detector resolution for detector modules D_0 and D_1 are respectively 2.17mm and 2.19mm. At the boundaries of the 80mm FOV (detector combinations 1-21 & 1-29 and 45-25 & 5-25) the spatial resolution degrades slightly to an average of 2.36mm FWHM for both detector modules. The FWTM resolution for D_0 degrades from 4.89mm to 5.5mm for respectively perpendicular incident photons to photons impinging at ± 15degree. For detector D_1 the FWTM resolution degrades from 5.14mm to 5.68mm depending on the incidence photon angle.

Hence these results show the possibility of an automatic acquisition of training data. In a real PET scanner this can be implemented in a fully automated procedure by slowly spiralling a point source close to the detectors and only saving coincidences with LORs within the FOV (Figure 4.19). Sorting the recorded coincidences per detector combination yields the required training data sets, similar to these obtained with the demonstrator setup.

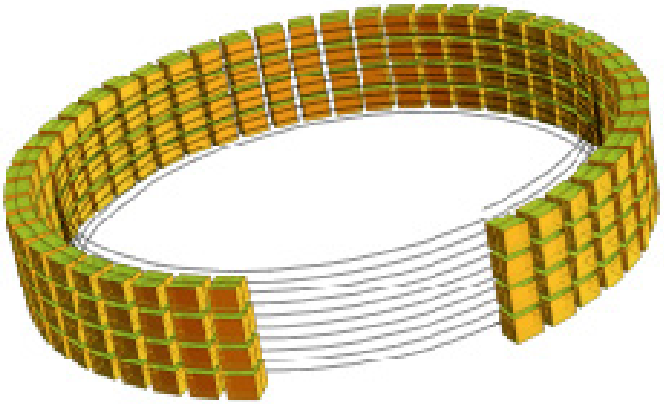


Figure 4.19: Illustration of the automated acquisition of training data on a full ring set-up using a spiralling point source moving close to the detector surface.

4.5 Influence of the acquisition parameters on the neural network performance

Table 4.2 compares the FWHM and FWTM spatial resolutions achieved on the bench set-up and demonstrator set-up for perpendicular incident photons at an energy threshold of 380keV.

Table 4.2 FWHM and FWTM spatial resolutions achieved on bench and demonstrator set-up for perpendicular incident photons at an energy threshold of 380keV

	Bench set-up	Demonstrator set-up
FWHM	1.6mm	2.2mm
FWTM	4.1mm	5.0mm

In order to understand the differences between both set-ups, a Monte Carlo simulation of a 20x10x10mm^3 LSO block detector module was implemented using GATE, a Geant4 Monte-Carlo toolkit [67]. (The implementation was done by Li Zhi, a PhD student also working at IIHE-VUB). The code simulated an experimental set-up where a narrow photon beam is positioned on the front surface of a 20x10x10mm^3 Teflon-coated LSO block. Upon interaction, the impinging 511 keV photons produce a number of optical photons with an average of 32000 ph/MeV [68]. Also the reflections at the edges of the Teflon wrapped block are simulated with a reflectivity factor of 95%. The output of such a simulation consists of the number of optical photons reaching each of the 32 APD pixels.

In order to study the influences of the different detector parameters on the spatial resolution, the simulated output per APD pixel was post-processed in Mathematica. The contributions of the quantum efficiency (QE), excess noise factor (F), gain (G_p) and pre-amplifier equivalent noise charge (ENC) are added to the simulated APD signals using the following formula for the equivalent optical photon noise (EOPN) per pixel:

$$\mathrm{EOPN_i} = \mathrm{N}_{\gamma_i}.\sqrt{\frac{\mathrm{F}}{\mathrm{N}_{\gamma_i}.\mathrm{QE}} + \left(\frac{\mathrm{ENC}}{\mathrm{N}_{\gamma_i}.\mathrm{QE}.\mathrm{G_p}}\right)^2 - \frac{1}{\mathrm{N}_{\gamma_i}}} \tag{4.5}$$

with N_{γ_i} the number of optical photons impinging on pixel i. The first term under the square root in Equation (4.5) reflects the statistical fluctuation of the APD gain and the optical photon statistic. For the Hamamatsu S8550 the APD quantum efficiency QE is 70% and the excess noise factor F is about 1.75 [39]. As this term only depends on the APD characteristics, a similar influence in both set-ups is expected. The second term describes the electronic noise, generated by the APD and the preamplifier, through the equivalent noise charge measured at the input of the amplifier. It depends on the dark current of the APD pixel capacitance and the noise of the preamplifier. On the bench set-up, the APD signals are read out by Cremat CR-110 preamplifiers which resulted in a measured 600e$^-$ ENC. The demonstrator set-up is based on a noisier pre-amplifier, i.e. the P-AMP chip introduced in Paragraph 4.1.1. As calculated in Equation

(4.1), the ENC of the APD plugged on the P-AMP chip is 1000e$^-$. Because the statistical fluctuation on the number of optical photons hitting the APD pixel is already included in the Monte-Carlo data, this contribution to the EOPN has to be removed by adding the last term in Equation (4.5).

The post-processing in Mathematica was achieved through the generation of a random sample of the underlying Gaussian distribution $N(N_{\gamma_i}, EOPN_i^2)$ with mean N_{γ_i} and variance $EOPN_i^2$ for each of the 32 APD pixels i.
To study the influence of the different detector parameters on the intrinsic resolution, training data were simulated using the parameters that represent the experimental set-up. After training, the NN was evaluated using simulated data generated with the same parameters except that the photon beam is assumed to be perfect now, i.e. a zero beam width. The resulting resolution will hence only reflect the influence of the detector components and the data acquisition method, i.e. it represents the intrinsic detector resolution.

Validation of Monte Carlo simulation

To validate the simulation, the photon incidence positions were estimated from the data (experimental bench set-up or simulated) using a neural network. The following parameters are used in the bench set-up simulation:

Table 4.3: Parameters used in the simulated bench set-up

APD QE	F	APD Gain	ENC	Beam width	Beam step
70%@420nm	1.75	50	600e$^-$	1mm FWHM	250 µm

The local resolutions and nonlinearity for the perpendicular incident photons are determined. This was accomplished by computing the positions of events impinging within 1mm regions, (i.e. combining events from four adjacent beam positions) and comparing them with their known incidence position. The FWHM resolutions and nonlinearities are in good agreement (see Figure 4.20). The experimental FWTM resolution is slightly larger than the simulated one. This is probably due to the fact that the photon beam profile is not exactly a Gaussian profile, i.e. the tails of the experimental photon beam are larger compared to the simulated Gaussian distribution. The average simulated resolution over the complete block was 1.8mm FWHM and 4.4mm FWTM at an energy threshold of 100keV. Besides, these resolutions still include the contribution of the 1mm beam width.

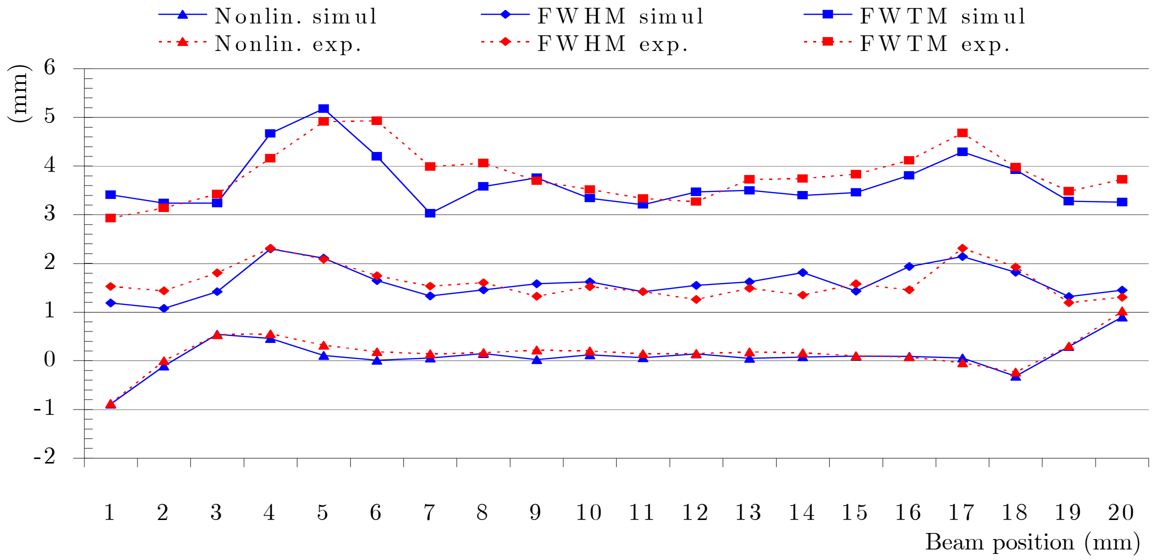


Figure 4.20: Comparison of the experimental bench data set and simulated bench data set for perpendicular incident photons. The local FWHM and FWTM resolutions and nonlinearity are plotted as function of the photon incidence position on the LSO block.

Origin of nonlinearity

The nonlinearity of the trained NNs is negligible except for photons impinging close to the edges (<2mm). In this region the estimated positions are shifted towards the center of the block. The origin is twofold. First, there is an intrinsic detector effect. When studying the resolution along the long side of the APD, the normalized signals from the four APD pixels in the same column are summed. This gives 8 input values to train the NN. Figure 4.21 illustrates per beam position the average responses of the normalized and summed APD channels of the training data set for perpendicular incident photons obtained on the experimental bench set-up. The analysis of the average APD responses per beam position indicates very similar trends for photons impinging within the first/last two millimeter from the crystal edge. In consequence the neural network which uses these APD responses as input for the position estimation will also project these different beam positions to similar positions introducing the nonlinear behavior.

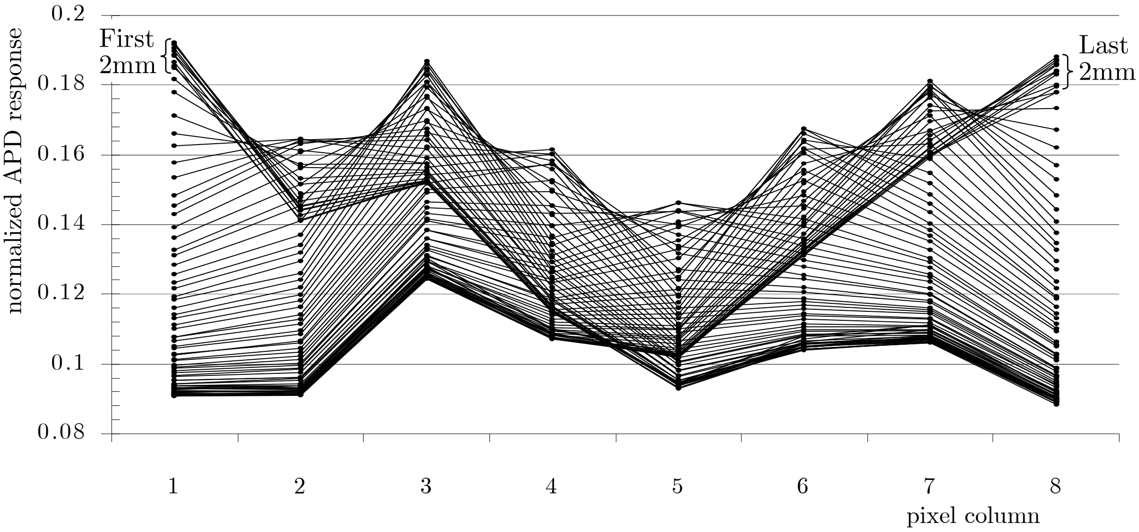


Figure 4.21: Average APD responses per beam position for perpendicular incident photons acquired on the bench set-up in steps of 250 µm. The 32 normalized APD pixels are reduced to 8 inputs by summing the 4 pixels within a same column.

Monte Carlo simulations which takes photon statistics and the optical photon transport into account, also point out that there is no big variation in APD responses within the first/last 2mm of the crystal edge. Figure 4.22 shows the average simulated response per beam position, after normalization and summation per column, for three 2mm intervals. The APD responses for the different beam positions in the central 2mm can clearly be differentiated. For beam positions in the outer interval, the light profiles are indiscernible.

The second cause of the systematic shifts comes from the NN training itself. The NN training introduces a nonlinearity at the boundaries of the NN output domain. To illustrate this, a NN was trained for data restricted to the interior 10mm of the LSO block. When all data are taken, i.e. all beam positions between 0mm and 20mm, no nonlinearity is observed within the interior 10mm. If a NN is trained with the same data set but only using events from beam positions restricted to the interval 6-15mm, nonlinearities are introduced at the boundaries of this interval as shown on Figure 4.23.

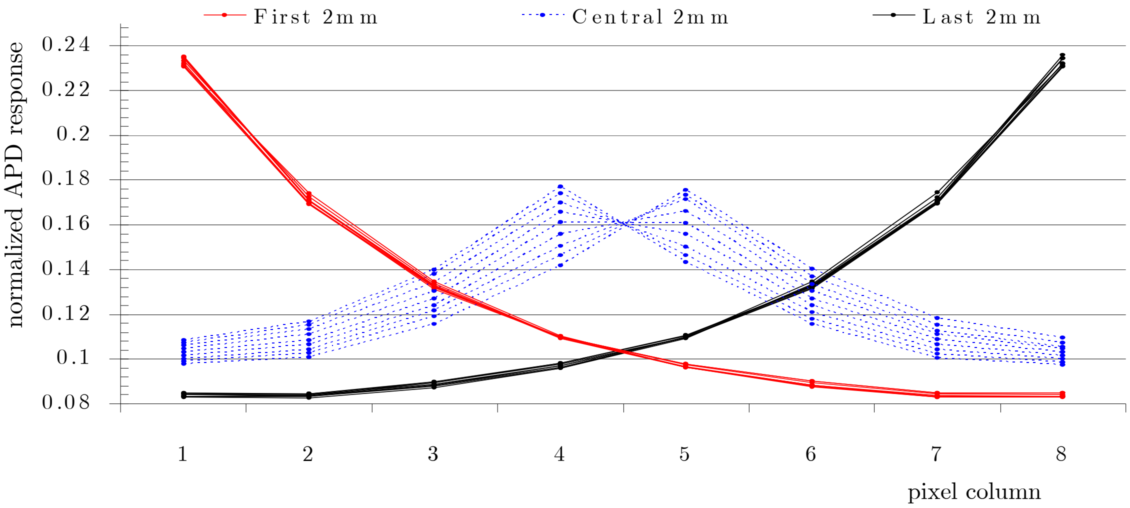


Figure 4.22: Average APD responses per beam position (250 µm step size) for perpendicular incident photons obtained with a Monte Carlo simulation. The 32 normalized APD pixels are reduced to 8 inputs by summing the 4 pixels within a same column. Only three 2mm intervals are shown, both 2mm extremities and the central 2mm.

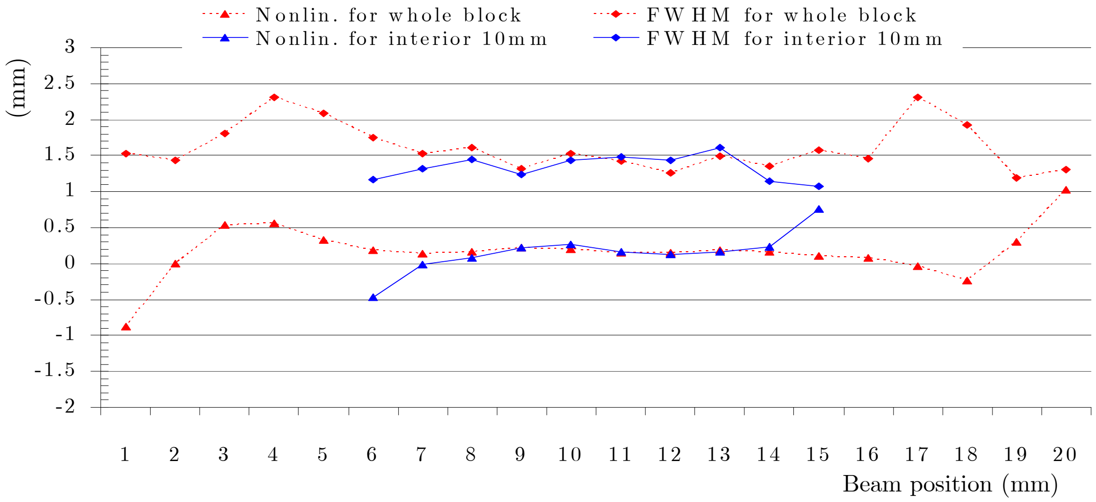


Figure 4.23: Illustration of the extra nonlinearity introduced by the NN at the boundaries of the NN output domain. The local FWHM resolution and nonlinearity are plotted as function of the photon incidence position.

Influence of the beam step size and the beam width

A photon beam steps over the block to acquire training data for the neural network. But at how many points on the LSO block surface the scintillation light distribution has to be acquired? And how narrow the beam has to be focused to minimize the uncertainty between the true photon impact point and the known beam position such that it does not deteriorate the intrinsic detector resolution? The effects of the step size and beam width on the intrinsic resolution are evaluated in two separate simulations. The default parameters of the simulation are given in Table 4.3. The influence of the beam step size is examined for a step size of 0.25mm, 0.5mm, 1mm, 1.5mm and 2mm. Figure 4.24 shows that the spatial resolution is rather independent of the beam step size. Measuring the scintillation light distribution every 1mm is sufficient to prevent any resolution loss.

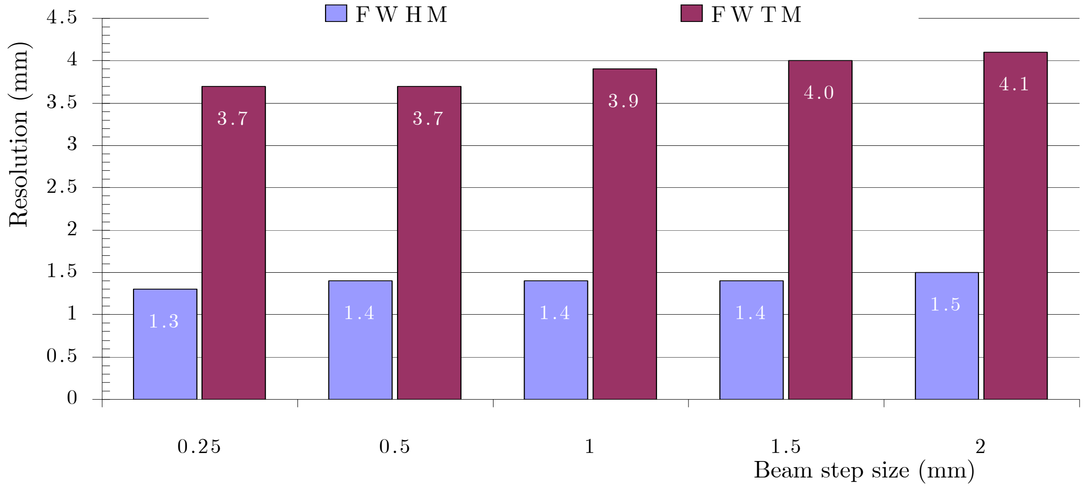


Figure 4.24: Simulated intrinsic resolution as function of the photon beam step size used to collect training data. A beam width of 1mm FWHM and an ENC of 600e$^-$ was used in the simulation (ENC of the bench set-up front-end electronics).

To study the beam width dependency, simulations with five different beam widths, from 0mm to 2mm FWHM in steps of 0.5mm, are evaluated. Figure 4.25 shows the spatial resolutions found for the different beam widths. The simulations show that the width of the photon beam during the training phase has a negligible influence on the spatial resolution if the beam width is less than 1mm FWHM.

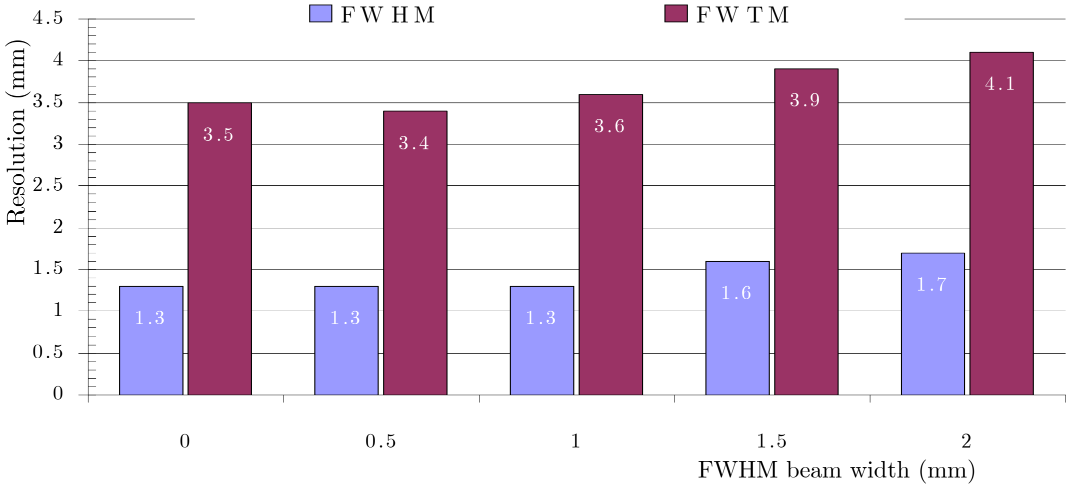


Figure 4.25: Simulated intrinsic resolution as function of the photon beam width used to collect training data. The beam was stepped in 250 µm over the length of the LSO block. An ENC of $600e^-$ was assumed in the simulation

Influence of electronic noise level
The front-end electronic contributes to the noise on the APD signals. The influence of the noise level on the spatial resolution was studied by varying the simulated ENC at the input of the preamplifier. The ENC was evaluated between 0e$^-$ and 1000e$^-$ in steps of 200e$^-$.

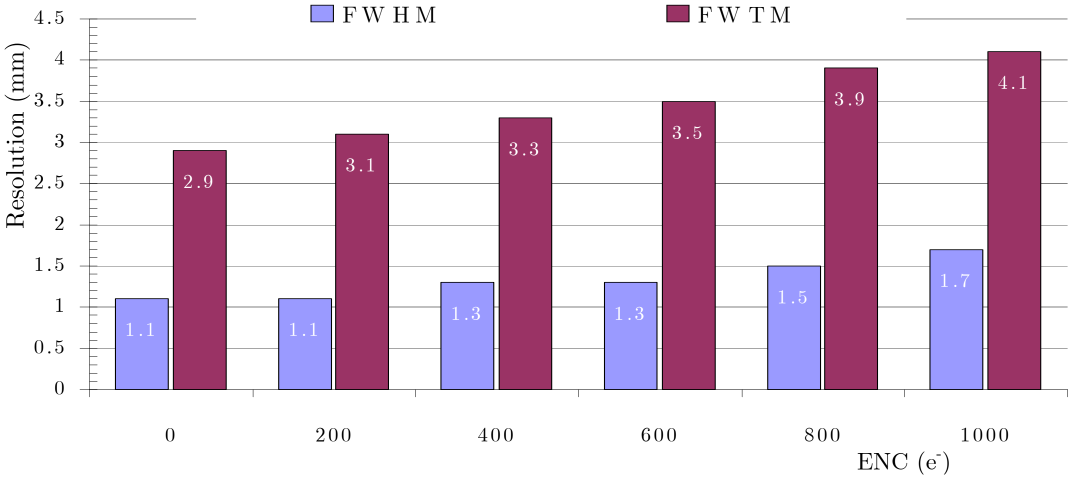


Figure 4.26: Simulated intrinsic resolution as function of the amplifier ENC. A 1mm FWHM photon beam was used to generate the training data. The beam was stepped in 250 µm steps over the length of the LSO block

The simulation predicts that the evaluation of a data set with a noise level of 600e$^-$ and a zero-beam width only adds 0.2mm to the FWHM and 0.6mm to the FWTM resolution compared with the evaluation of a perfect data set without noise and zero-beam width. The extra deterioration of the intrinsic FWHM resolution for a 1000e$^-$ ENC is 1.7/1.3 = 1.3 times worse than the 600e$^-$ ENC FWHM resolution. The FWTM degrades about 4.1/3.5 = 1.2 times.

Thus, three parameters were evaluated to understand the resolution differences found on the bench set-up and the demonstrator set-up. The influences of the beam step size and beam width differences are minimal as both set-ups has a beam step size below 1mm and a beam width around 1mm FWHM. However, the differences in resolution as summarized in Table 4.2 can be attributed to the different ENC values. The simulation calculates a degradation of 30% and 17% for respectively the intrinsic FWHM and FWTM resolutions when the ENC varies from 600e$^-$ to 1000e$^-$.

The degradations measured between both experimental set-ups are in agreement. Indeed, the spatial resolution deterioration discerned between both set-ups can be explained as follow:

$$\sqrt{\left(\sqrt{1.6^2 - 1}\cdot 1.3 \cdot 1.1\right)^2 + 1} = 2.1 \tag{4.6}$$

with 1.6 the Bench set-up resolution, 1 the estimated beam width profile, 1.3 the ENC contribution and 1.1 the effect of the beam step. Consequently the simulations confirm the implemented automated acquisition method of training data for the positioning algorithm on the demonstrator set-up. The discrepancy in the intrinsic spatial resolutions of both set-ups is mainly due to the different ENC values.

Chapter 5

Two-dimensional reconstructed images on the PET prototype demonstrator

The automated acquisition method for training data on the prototype demonstrator for neural networks depicted in Chapter 4 demonstrates that an intrinsic spatial detector resolution of 2.2mm FWHM can be achieved for nearly perpendicular incident photons (Figure 4.18). At the boundaries of the examined 80mm FOV the intrinsic spatial resolution degrades slightly to 2.4mm FWHM. In addition, simulations point out that an optimization of the pre-amplifier ENC from $1000e^{-}$ to $600e^{-}$ can improve this FWHM intrinsic resolution with 0.4mm.

In order to further evaluate these results on $20 \times 10 \times 10 mm^3$ LSO blocks, the spatial resolution in reconstructed 2D images is examined. Per detector combination the correspondingly trained NN is used to define the LOR of each coincidence event acquired on the prototype demonstrator.

5.1 Scanning process

On the demonstrator the tomographic data is acquired in step-and-shoot mode since there is no position encoder on the stepper motor moving the source platform (see Figure 5.1.a). As described in the previous chapter, one detector combination determines a diamond-shaped segment in the sinogram. The rotation of the source platform displaces this diamond-shaped segment down the ordinate ϕ of the sinogram. The shift of the segment is defined by the angular step size α of the Ωz- stage.

The width of the FOV depends upon the number of detector combinations used. Since all radial samples of a projection are required at each projection angle to reconstruct an image, the minimal number of detector combinations is three. The parallelogram formed by detector combinations 1-24, 1-25 and 1-26 result in a FOV of 20mm (see Figure 5.1.b and c). Rotation of the source platform over 360° will sweep this parallelogram through all angular samples ϕ of the sinogram, needed for the image reconstruction.

In the previous chapter nine different neural networks for detector combinations 1-21 to 1-29 were trained. Accordingly, with these nine different detector combinations a FOV of maximum 80mm diameter can be evaluated on the demonstrator (Figure 5.1.c). Again, for each detector combination the source platform has to rotate over 360° for the acquisition of all angular samples.

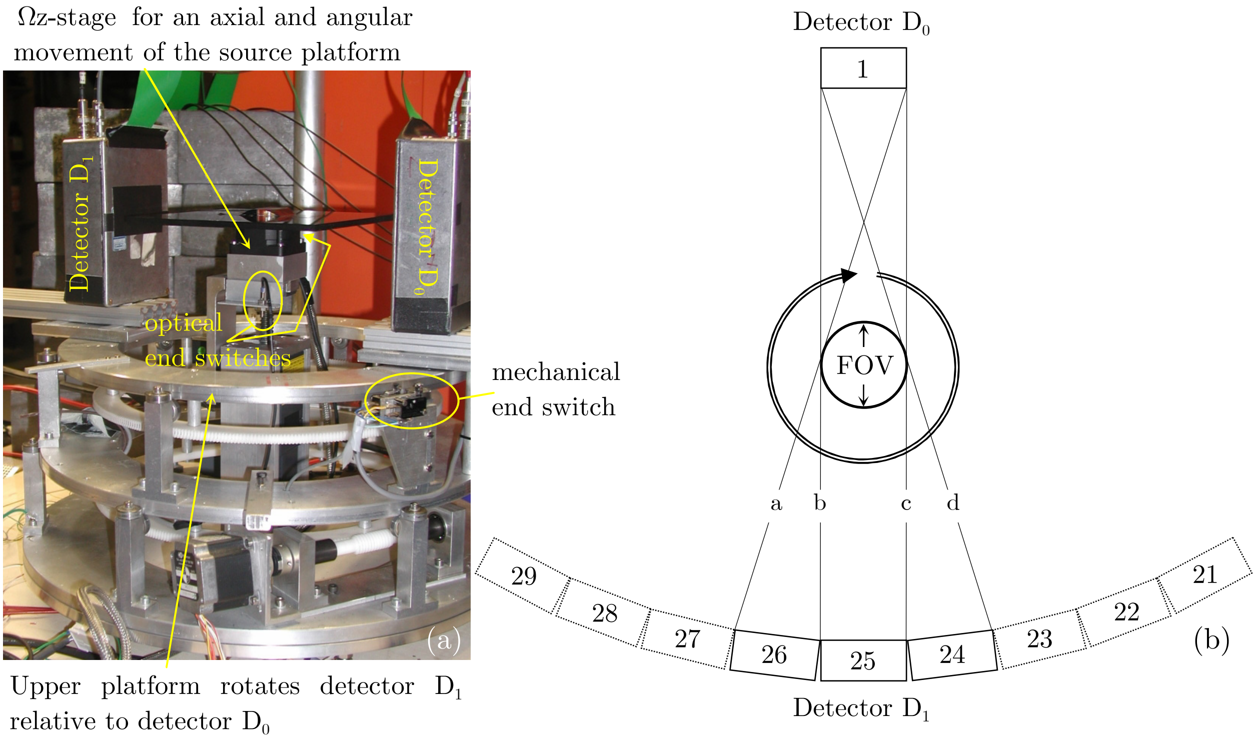
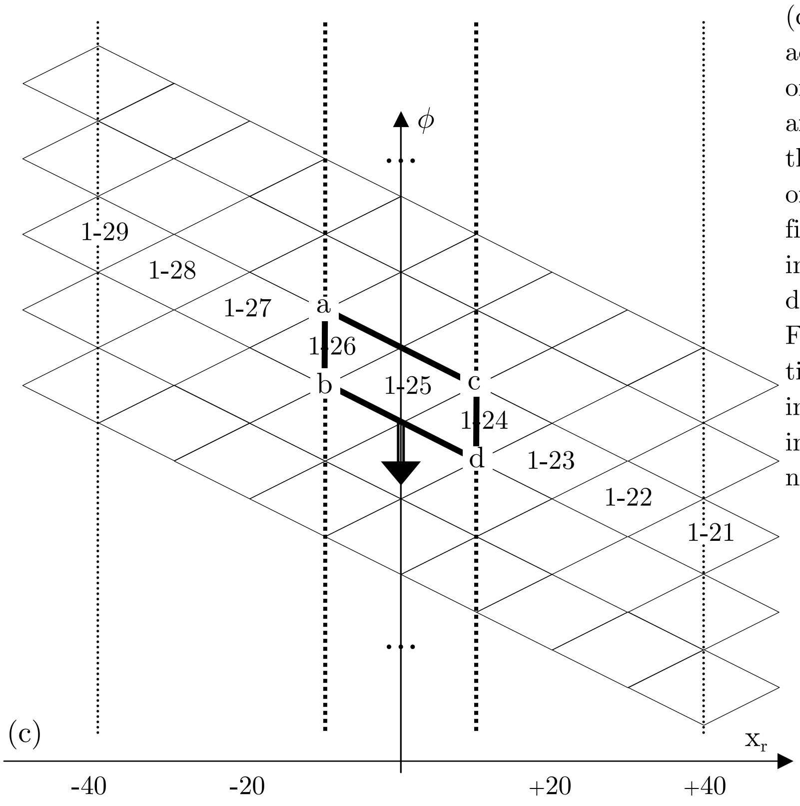


Figure 5.1: (a) Illustration of the PET prototype demonstrator. The axial and angular movement of the source platform is made by the Ωz-stage. The different detector combinations are obtained by moving detector D_1 relative to detector D_0 by use of the upper scanner platform.
(b) The diameter of the FOV depends on the number of detector combinations at which data is acquired. The FOV width achievable with three detector combinations is shown.



(c) In order to obtain a uniform acquisition over all radial samples, only data within the shaded lines are accepted. As the source rotates, this segment is displaced down the ordinate ϕ of the sinogram and fills all angular samples needed for image reconstruction. The bold dotted lines depict the limits of the FOV for three detector combinations. The other pair of dotted lines indicates the edges of the FOV including the nine detector combinations.

The acquisition of tomographic data is implemented in Labview. Due to the absence of position encoders the scan process starts with the movement of detector D_1 and the source platform to their initial positions defined by respectively a mechanical and two optical end switches (see Figure 5.1.a). From this known initial position detector D_1 leaves the mechanical switch and moves to its start position, determined by the required FOV diameter, e.g. for a FOV diameter of 20mm detector D_1 moves to position 26 as illustrated on Figure 5.1.b. The axial start position of the source platform is such that the object under study is centered along the 10mm side of the LSO block.

For the given detector combination, e.g. 1-26, and the initial angular source position, coincidences are recorded for a fixed amount of time. Then the Ωz-stage rotates over an angle α to its next position and coincidence events are acquired again for the fixed amount of time. This last step is repeated until the Ωz-stage has rotated over 2Π rad. This rotation results in the movement of the diamond-shaped segment over all angular samples in $2\Pi/\alpha$ steps. After the angular Ωz-cycle is accomplished, detector D_1 rotates to the next adjacent detector position and again the angular Ωz-cycle is completed. This procedure is repeated for all detector positions within the FOV, e.g. positions 25 and 24 in this example. Each saved coincidence event contains both detector positions D_0 and D_1, the angular source position and the 2x32 scintillation light distributions sampled by peak sensing ADCs.

5.2 Influence of angular source step α on angular sampling

Figure 5.3 and Figure 5.4 illustrate the effect of the angular step α of the Ωz- stage on the diamond-shaped segment displacement within a sinogram. Figure 5.3 shows the sinograms for respectively detector combination 1-24, 1-25 and 1-26 as well as the global sinogram for a point source 10mm off center. In this example the angular step α corresponds to one detector width such that there is no overlap of successive detector positions. As the demonstrator simulates a full ring of 48 detectors, one detector step equals 360°/48 = 7.5°. The dotted lines in sinograms 1-26, 1-25 and 1-24 delimit the maximal radial deviation of a LOR. The gaps in the sinograms between the different angular positions demonstrate that an angular step α of 7.5° is insufficient to acquire all angular samples. The gaps between the segments are due to the wedge shaped voids between two consecutive detector positions (Figure 5.2). In addition, the gaps are enlarged by the non-linear effect of the NN positioning estimator. Indeed the non-linear effect reduces the radial and angular positions of the LORs at the extremities of the diamond-shaped sinogram segments.

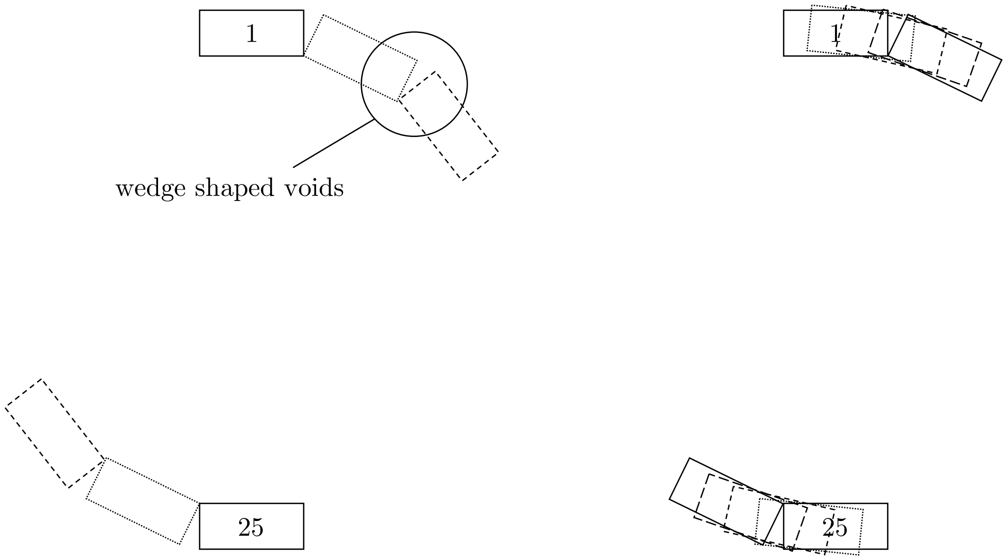


Figure 5.2: Scheme of both tested geometries, i.e. a non-overlapping and an overlapping scanning geometry. The angular steps α are respectively 7.5° and 1.875°. The overlapping scanning geometry allows the acquisition of LORs that were missing mainly due to voids in the non-overlapping scanning geometry.

In order to avoid the gaps in the sinogram an overlapping scanning geometry was investigated. For this lay-out the angular step α was reduced to one fourth of the non-overlapping approach as shown in Figure 5.2 (right). The gaps between the different angular positions as seen in Figure 5.3 disappear. The overlapping geometry allows the measurement of all LORs needed for a correct image reconstruction (Figure 5.4).

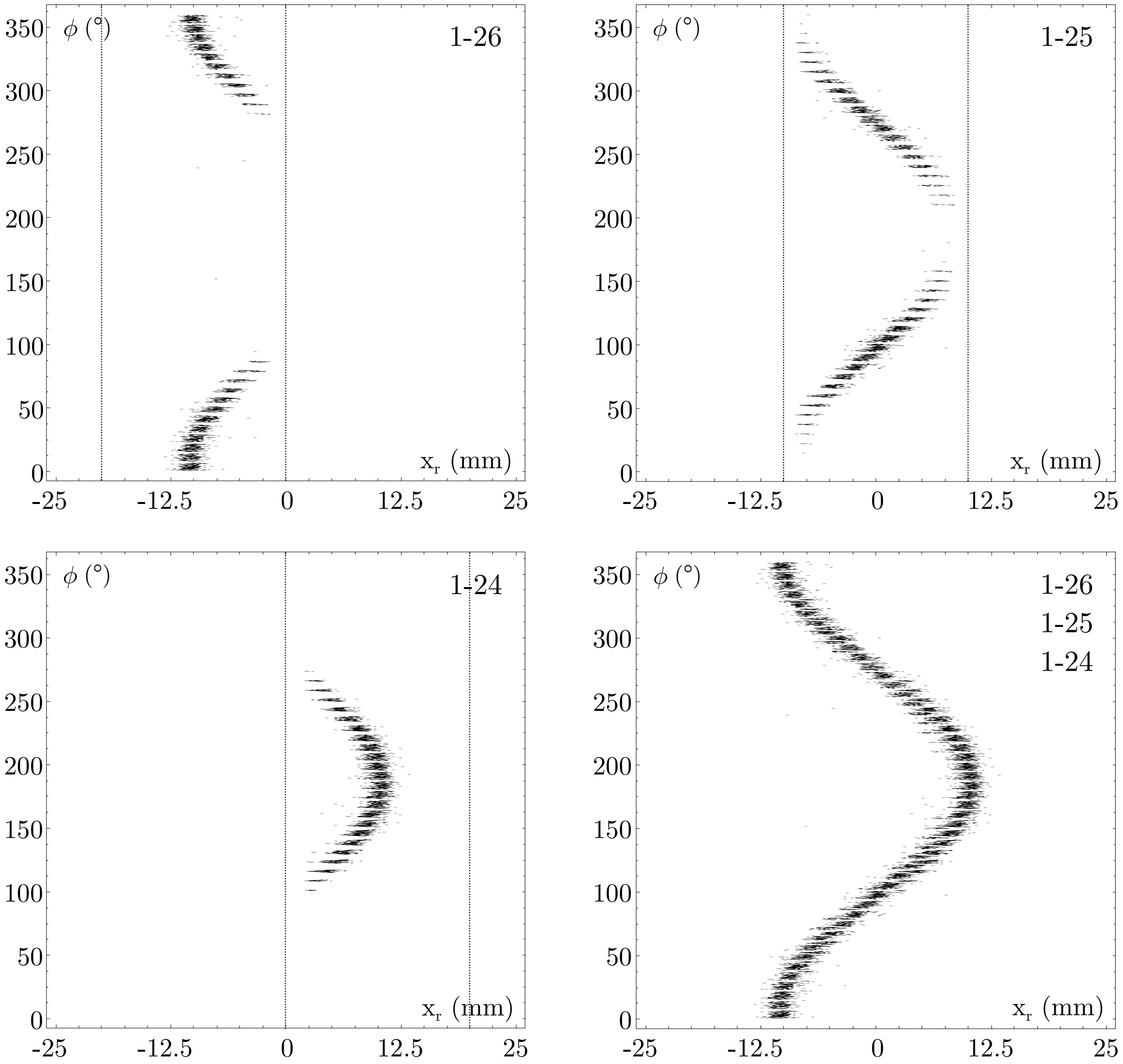


Figure 5.3: Sinograms per detector combination and the global sinogram for a point source 10mm off center. The acquisition is done with an angular source step α of 7.5°. As the demonstrator simulates a full ring of 48 detectors, 360°/48 equals 7.5° which corresponds to one detector width. The dotted lines in sinograms 1-26, 1-25 and 1-24 delimit the radial range measured by each detector combination. The gaps in the sinograms between the different angular positions demonstrate that an angular step α of 7.5° is insufficient to acquire all angular samples.

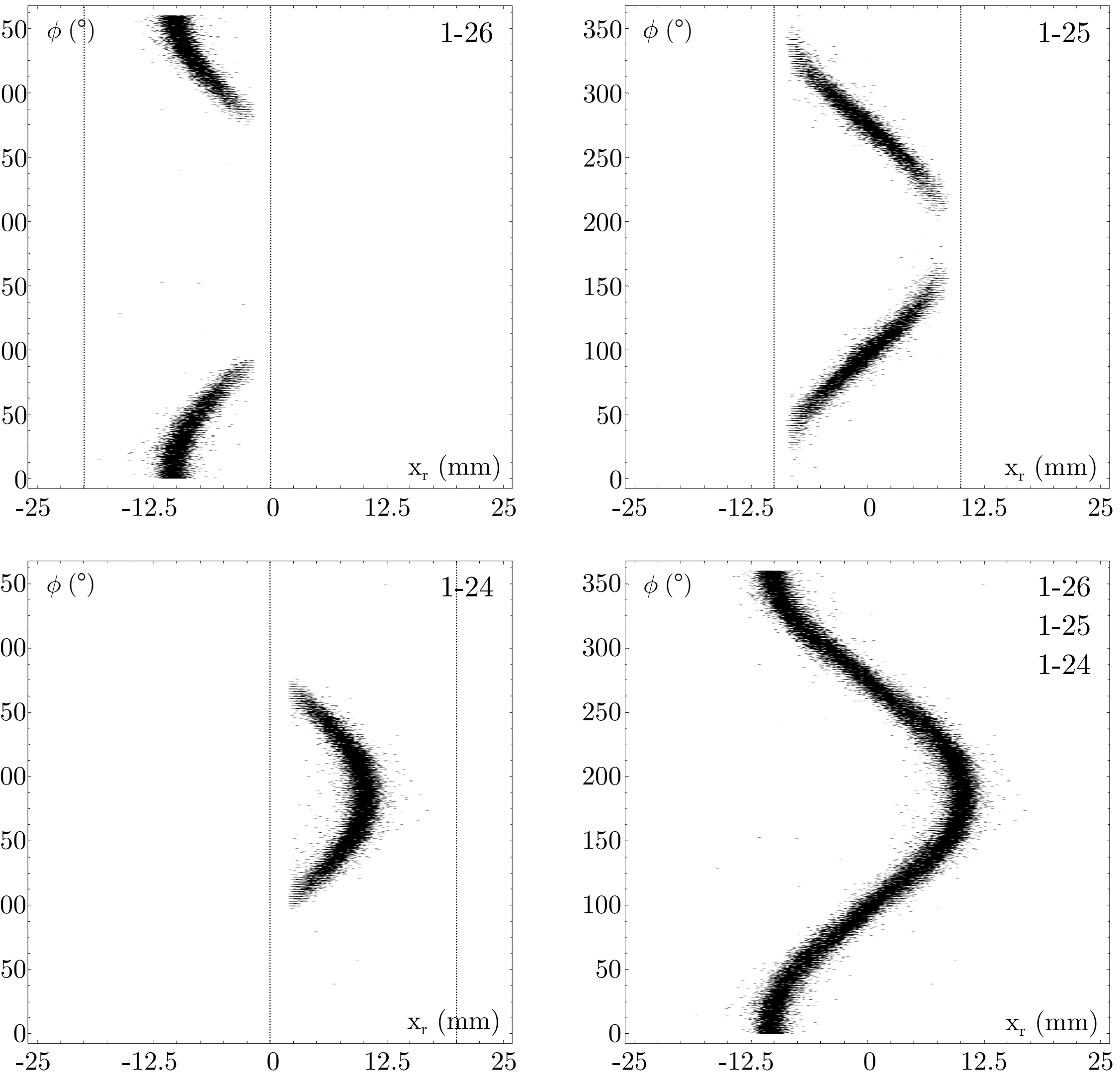


Figure 5.4 Sinograms per detector combination and the global sinogram for a point source 10mm off center. The angular source step α was set to $360°/192 = 1.875°$ which corresponds to a quarter of one detector width. The dotted lines in sinograms 1-26, 1-25 and 1-24 delimit the radial range measured by each detector combination. The gaps between the different angular positions as seen in Figure 5.3 disappear. The overlapping geometry allows the measurement of all LORs needed for a correct image reconstruction.

5.3 Position estimation of non-perpendicular incident photons with perpendicular trained NN

In Paragraph 4.4 the influence of a mismatch between the true and the estimated incidence angle was already examined. At incidence angles five degrees below and above the true incidence angle, the FWHM degradation due to this angle mismatch was on average 0.1mm. A similar trend was observed for the FWTM with an average deterioration of 0.3mm. In the present section the improvement in image resolution due to specifically trained NNs per detector combination is examined. To evaluate this improvement, a comparison with the usage of NN_{perp} for each detector combination is made. NN_{perp} is the NN trained for perpendicular incident photons.

For a point source 35mm off center, all trained detector combinations, from 1-21 to 1-29, are used to reconstruct the tomographic image. Prior to the analysis of the reconstructed image, the differences in local intrinsic detector resolution and nonlinearity achieved with a specifically trained NN and NN_{perp} are evaluated (Figure 5.5). The data set used for this evaluation is the data set acquired by detector combination 1-29 on the PET demonstrator. As mentioned before the specifically trained $NN_{1\text{-}29}$ achieved a global intrinsic detector resolution of 2.3mm FWHM and 5.3mm FWTM. The nonlinearity is negligible except for photons impinging close to the edges of the LSO block.

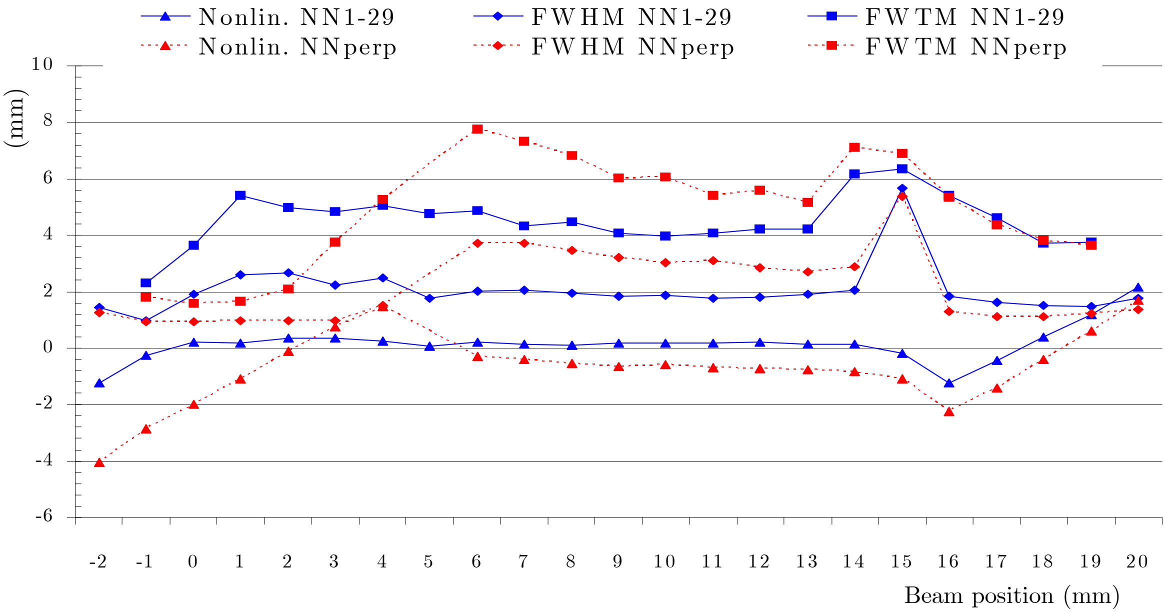


Figure 5.5: Comparison of intrinsic local resolutions and nonlinearities achieved with a specifically trained neural network $NN_{1\text{-}29}$ and N_{perp}, a NN only trained for perpendicular incident photons. The local FWHM and FWTM resolutions and nonlinearities are plotted as function of the photon incidence position on the LSO block. The results of NN_{perp} at 5mm are removed from the plot due to the incorrect fit of the local resolution. Also the FWTM results of $NN_{1\text{-}29}$ and NN_{perp} at -2mm and 20mm are dropped for the same reason.

If NN_{perp} is used to estimate the incident photon positions at detector combination 1-29 the global intrinsic detector resolution degrades to 3.9mm FWHM and 8.1mm FWTM. The nonlinearities of the local resolutions worsen for all beam positions. At the edges, the large shifts towards the center of the block result in an artificial compression of the local intrinsic detector resolutions.

The differences in position estimation per detector combination for specifically trained NNs versus NN_{perp} are summarized in Figure 5.6. The data used in this evaluation is the tomographic data acquired for a point source 35mm off center. As $NN_{1\text{-}25}$ and NN_{perp} are one and the same trained NN, it is obvious that all data on the scatter plot of $NN_{1\text{-}25}$ versus NN_{perp} are lying on the bisector as seen on Figure 5.6. All remaining plots show a strong linear dependency, however all with a small bias. The line of identity was shifted vertically by -0.4mm for 1-21 to 0.6mm at position 1-29. The positions of the different detector combinations 1-21 to 1-29 are illustrated on Figure 5.1.b. It also appears that the spread between the estimated positions of $NN_{1\text{-}X}$ and NN_{perp} broaden in function of the step differences between the detector position X and 25. The spread vary from 0.5mm FWHM for combination 1-24 (1 detector step) to 1.3mm FWHM for combination 1-29 (4 detector steps).

A last tendency coming out these plots is the inability of NN_{perp} to correctly estimate positions outside the trained 0-20mm interval where it was trained for. For detector combination 1-21 to 1-24, situated on the right-hand side of combination 1-25 the acquired incident photon interval varies from (as explained in Figure 4.16)

$$0\text{mm to } \left(20+10.\text{Tan}[\Delta_{25\text{-}X}.\frac{\Pi}{48}]\right)\text{mm} \tag{5.1}$$

with $\Delta_{25\text{-}X}$ the step difference between the detector combination 1-X and 1-25. Factors 10 and 20 are respectively the thickness and the length of the LSO block and 48 the number of detectors within the simulated full ring.

As demonstrated in Paragraph 4.5, the unavoidable non-linear effect on the specifically trained $NN_{1\text{-}25}$ already starts for photons impinging within the first/last two millimeter from the crystal edge. This trend persists outside the trained interval. The estimated position for photons impinging beyond 18mm is shifted towards the center of the block as illustrated on Figure 5.6. in scatter plot $NN_{1\text{-}22}$ versus NN_{perp}.

A similar results are found for detector combinations 1-26 to 1-29, located on the left side of detector combination 1-25. For these detector positions the possible incident photon interval is

$$\left(-10.\text{Tan}[\Delta_{X\text{-}25}.\frac{\Pi}{48}]\right)\text{mm to 20mm} \tag{5.2}$$

In this case, NN_{perp} will estimate positions beneath 2mm towards 2mm as shown in scatter plot $NN_{1\text{-}28}$ versus NN_{perp}..

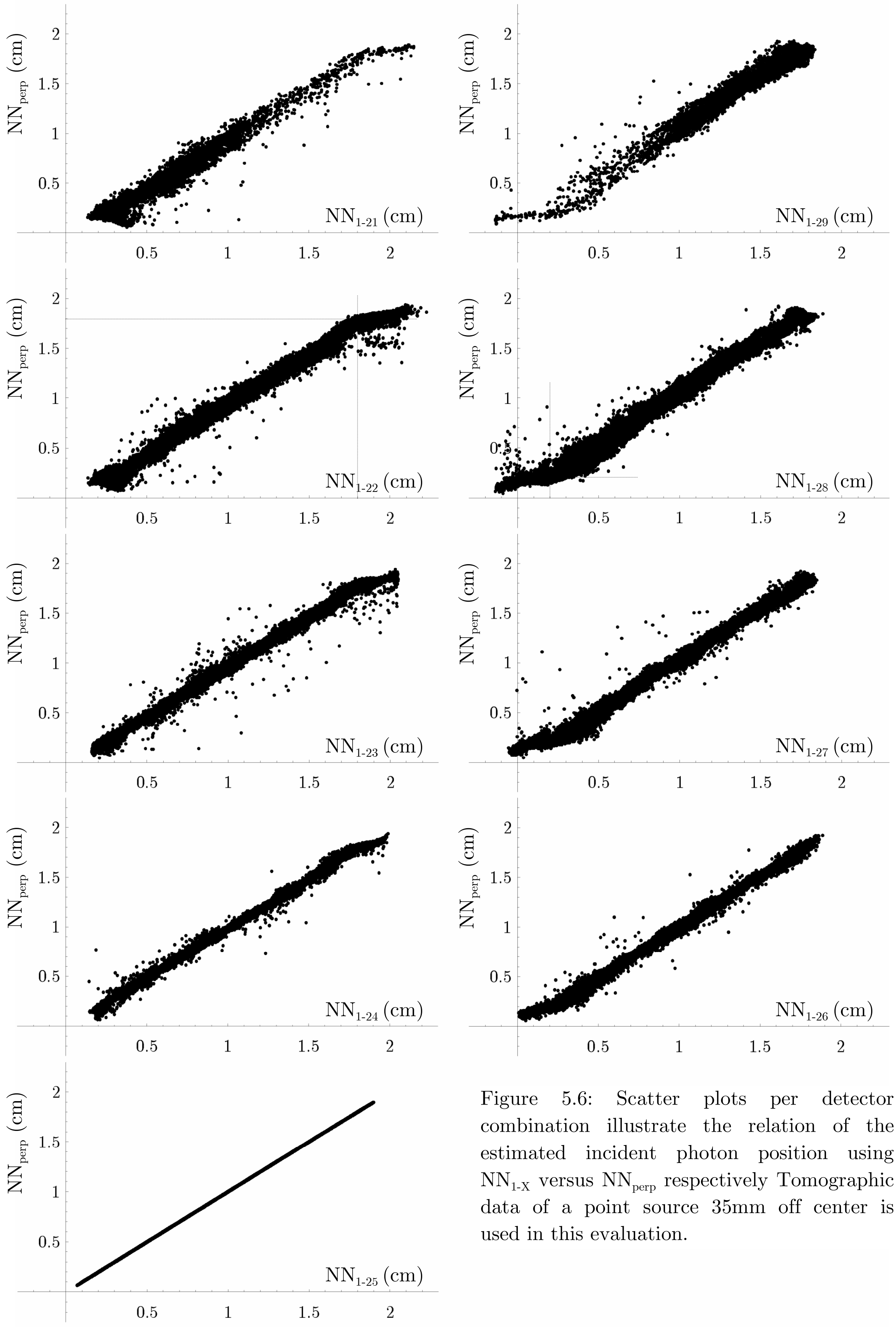


Figure 5.6: Scatter plots per detector combination illustrate the relation of the estimated incident photon position using NN$_{1\text{-}X}$ versus NN$_{perp}$ respectively Tomographic data of a point source 35mm off center is used in this evaluation.

Consequently, all previous results confirm the benefits of specific trained NNs in order to acquire optimal parallax-free position estimations. The effect of non-specifically trained NNs on reconstructed images is illustrated in Figure 5.7 and Figure 5.8. Figure 5.7 shows the substraction of $\text{sinogram}_{\text{spec}}$ with $\text{sinogram}_{\text{perp}}$ for a point source 35mm off center. The gray $\text{sinogram}_{\text{perp}}$ is the sinogram obtained from position estimation using only NN_{perp} whereas $\text{sinogram}_{\text{spec}}$ (black points) is obtained by using specific trained $\text{NN}_{\text{1-X}}$ per detector combination. The gap in the middle of the plot corresponds with the band of diamond segments 1-25. As NN_{perp} and $\text{NN}_{\text{1-25}}$ are identical the differences at this detector combination of both sinograms equals zero. The bias and non-linearity seen in Figure 5.6 result in a reduction of the amplitude of the gray $\text{sinogram}_{\text{perp}}$ curve compared to the black $\text{sinogram}_{\text{spec}}$. However the average radial resolutions over all angular projections of both sinograms are similar, with an average of 2.2mm FWHM and 5.0mm FWTM. Accordingly and as shown in Figure 5.8, the MLEM reconstructed images using one of both position estimation methods does not differ in their reconstructed image resolution. A FWHM of 2.1mm and a 3.9mm FWTM image resolution is achieved after 10 iterations. Nevertheless, as expected from the sinogram analysis, a displacement in the point source position occurs. The magnitude of the shift is 0.8mm.

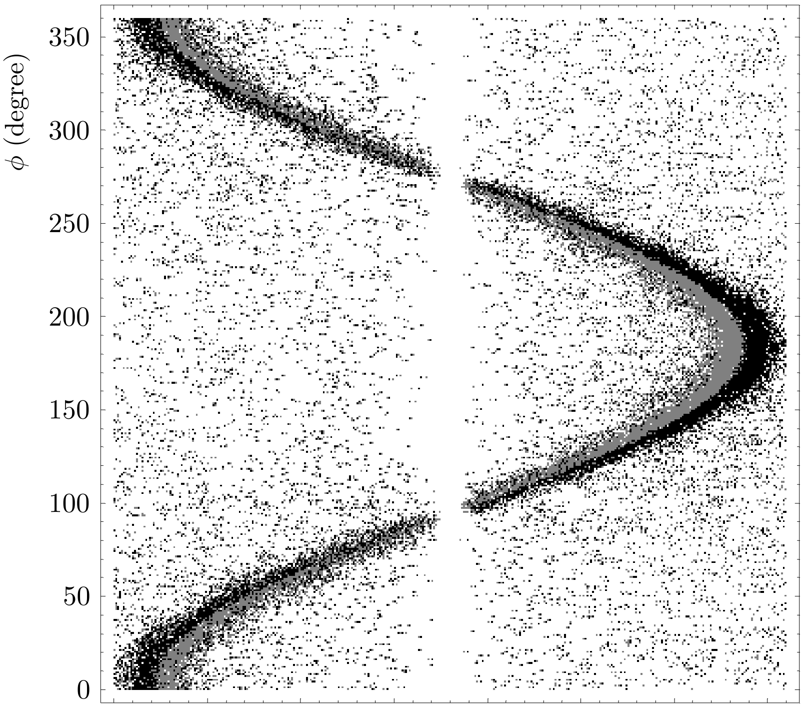


Figure 5.7: Scatter plot obtained by subtraction of $\text{sinogram}_{\text{perp}}$ from $\text{sinogram}_{\text{spec}}$ for a 250 µm ^{22}Na point source 35mm off center. The gray $\text{sinogram}_{\text{perp}}$ is the sinogram obtained via position estimation using only NN_{perp} whereas $\text{sinogram}_{\text{spec}}$ (black points) is achieved by using specific trained $\text{NN}_{\text{1-X}}$ per detector combination. The gap in the middle of the plot occurs because NN_{perp} and $\text{NN}_{\text{1-25}}$ are the same NN.

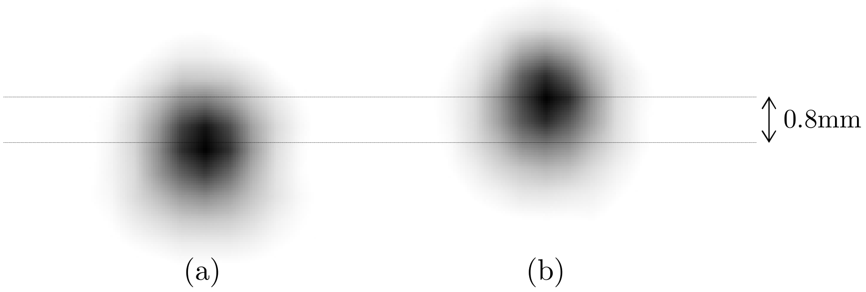


Figure 5.8: MLEM reconstructed image of a 35mm off center point source. (a) Reconstructed point source using specifically trained $NN_{1\text{-}X}$ positioned at 35 mm from the center while for (b) the reconstruction is done only with NN_{perp} estimated positions. A FWHM of 2.1mm and a 3.9mm FWTM resolution are achieved in both cases. However a displacement of 0.8mm appears between the reconstructed points, using the same data set but different trained NNs.

5.4 Point sources at different radial distances

The transaxial resolution of a ^{22}Na point source (250 µm diameter) was measured at different radial distances, from midpoint to 40mm off center in steps of 5mm (see Figure 5.9). Axially the point source was placed near the center of the FOV. The acquisition of the tomographic data is completed according to the scan process described in Paragraph 5.1. A two-dimensional (2-D) Maximum Likelihood Expectation Maximization (MLEM) algorithm is applied for the image reconstruction. The expectation maximization algorithm, introduced in Chapter 2, is an iterative algorithm that maximizes the likelihood under a Poisson data model. At each iteration, the current image estimate $\hat{f}_j^{(k)}$ is used to generate, by forward projection procedure, the projection data $\hat{p}_i^{(k)}=\sum_{j=1}^{N} L_{ij}.\hat{f}_j^{(k)}$ which is compared with the measurements p_i. The measured coincidences in the i-the projection bin [36]. This comparison is used to compute correction factors to make $\hat{p}_i^{(k)}$ consistent with p_i by modifying the image estimate $\hat{f}_j^{(k)}$ and generate a new estimate $\hat{f}_j^{(k+1)}$. In other words, the intent is that the series of estimates ($\hat{f}_j^{(0)}$, $\hat{f}_j^{(1)}$, $\hat{f}_j^{(2)}$, …, $\hat{f}_j^{(k)}$) converges to a final estimate, $\hat{f}_j^{*}$, such that $\hat{p}_i^{*}=\sum_{j=1}^{N} L_{ij}.\hat{f}_j^{*}$ is the closest possible to p_i. However, research showed that if the iterations continue, the likelihood increases, whereas the image begins to degenerate, becoming increasingly noisy [36]. The origin of this effect lies in the fact that the measurements p_i are Poisson random variables with noise, and that insistence on an exact fit to the data will result in an image dominated by noise. It is therefore important to terminate the reconstruction before this degeneration begins, even though the likelihood function may not be at maximum. However the choice of the stopping point is difficult, and research continues in order to define appropriate rules for use with real data [36].

Stefaan Vandenberghe from the Medical Imaging and Signal Processing (MEDISIP) research group at the Ghent University provided the iterative listmode MLEM reconstruction program used [69]. The MLEM algorithm stops after ten iterations and has an image pixel size of 0.5mm.

Figure 5.9 illustrates the nine reconstructed ^{22}Na point sources from midpoint to 40mm off center in steps of 5mm which characterize an 80mm FOV. No corrections such as normalization, scatter - or random correction were applied to the reconstructed images.

Figure 5.10 shows the radial and tangential resolution as a function of the radial source position. The radial and tangential resolutions remain constant around 2.0mm FWHM and 3.7mm FWTM when the source is moved from the CFOV to 40 mm off center. However, only the reconstructed point source at 0mm has a perfect round shape. This is probably due to the fact that for this CFOV position, all LORs can be measured by just one detector combination, i.e. the opposite detector combination 1-25. The slight asymmetry in the remaining positions is probably due to a non-perfect rotational movement of both detector modules and can also be enhanced by the non-linear effect of the neural network positioning algorithm. Nevertheless Figure 5.9 and Figure 5.10 demonstrate the effectiveness of the PET prototype demonstrator set-up within an

80mm FOV. This confirms that the parallax distortion due to the large incident photons angles is countered owing to the use of LSO blocks and the neural network positioning algorithm approach.

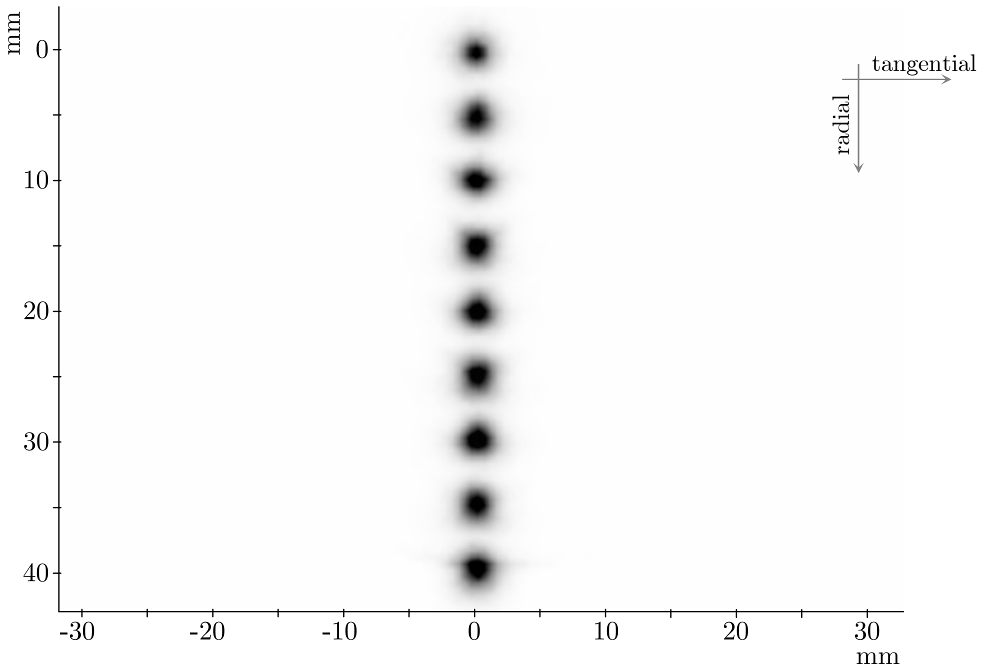


Figure 5.9: Illustration of the iterative listmode MLEM reconstructed 250µm ^{22}Na point source at the radial distances 0mm to 40mm off center in steps of 5mm.

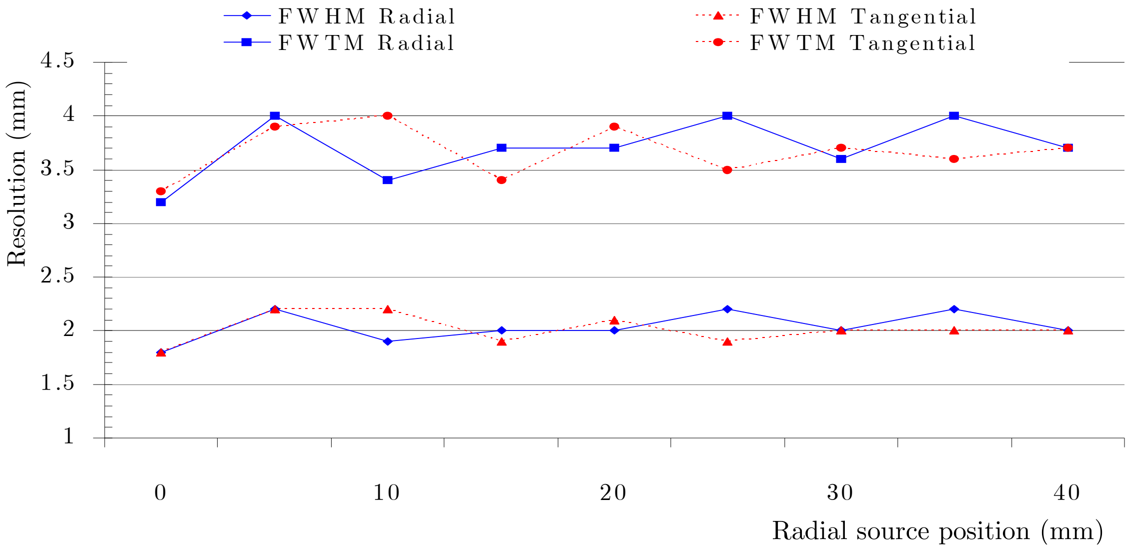


Figure 5.10: Radial and tangential 2D-MLEM reconstructed FWHM/FWTM resolution as a function of the radial position of the 250µm ^{22}Na point source.

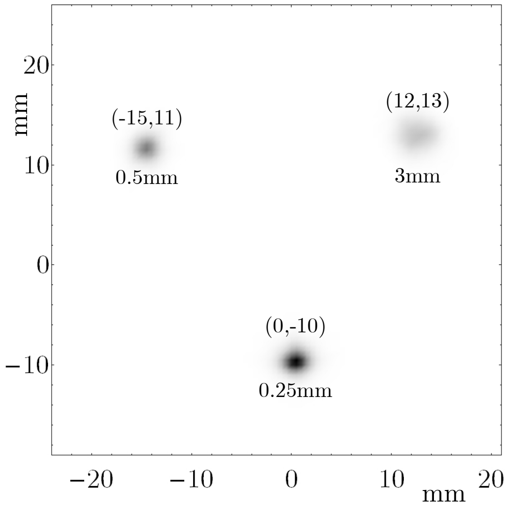
5.5 *Tomographic images of combined point sources and a Derenzo phantom*

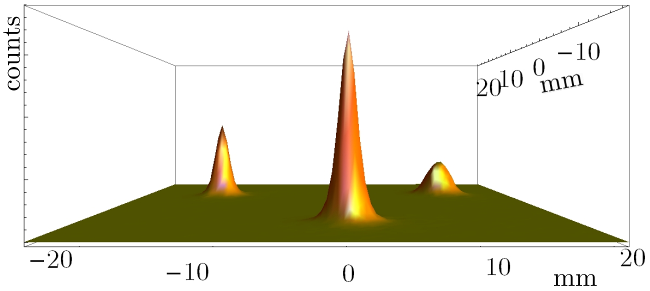
The performance of the prototype demonstrator has been studied in various analyses. For example in Paragraph 5.4, the spatial resolution was measured by imaging one point source at different positions. At the centre of the field of view the resolution is 1.8mm FWHM, and it remains constant around 2.0mm FWHM for radial positions within 40mm from the scanner axis.

Table 5.1 ^{22}Na point sources used on prototype demonstrator

Diameter	Activity on purchase	Calculated Activity during test	Ratio with max activity
3.0mm	93.59 μCi on 01/02/2002	19 μCi	0.5
0.5mm	100 μCi on 01/04/2002	22 μCi	0.58
0.25mm	100 μCi on 29/04/2004	38 μCi	1

In the next approach, three ^{22}Na point sources with different diameters and activities are scanned together. One scan was performed with the three sources arranged around the center. While in a second scan, two point sources were placed at 3.5mm and one point source at 1.5mm out center. The parameters of the three ^{22}Na point sources are summarized in Table 5.1. The acquisition of the tomographic data is again completed according to the scan process described previously. A FOV of 40mm (-20mm to 20mm) is obtained with five detector combinations, i.e. from detector combination 1-23 until combination 1-27. The 2D MLEM algorithm is applied for the image reconstruction. Figure 5.11 illustrates the three reconstructed point sources through a density plot and a 3D plot, showing the acquired spatial resolutions and variation in source activity.





Source diameter	FWHM radial	FWHM tang.	FWTM radial	FWTM tang.
0.25mm	1.8mm	2.1mm	3.3mm	3.8mm
0.5mm	2.0mm	2.2mm	3.7mm	3.9mm
3mm	3.1mm	3.2mm	5.7mm	5.9mm

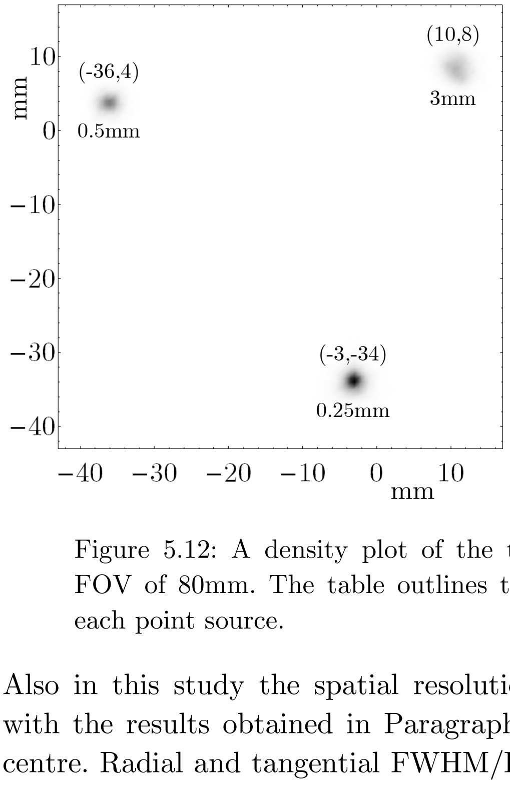
Figure 5.11: (left) A density plot of the three reconstructed point sources centered around the scanner axis. (right) A 3D representation of the reconstructed point sources. It clearly illustrates the differences in source activity. The table outlines the acquired FWHM and FWTM resolutions for each point source.

The spatial resolution obtained for the 250 µm (=0.25mm) point source matches the prior results for a point source placed 10mm off centre. As illustrated on Figure 5.9 and Figure 5.10, a spatial resolution of 1.9mm/3.4mm and 2.2mm/4.0mm was attained for respectively the radial and tangential FWHM/FWTM resolutions.
In order to compare the relative activities of the three sources, the volumes of the three Gaussian graphs are calculated. The ratio with the most active source is taken. Table 5.2 summarizes the results with a good agreement between expected and measured ratios. This result demonstrates a uniform photon detection of the demonstrator within the FOV.

Table 5.2: Ratio of expected and measured activities

Diameter	Expected activity ratio	Measured activity ratio
0.25mm	1	1
0.5mm	0.58	0.56
3mm	0.5	0.56

Similarly, Figure 5.12 shows the result obtained with the three point sources spread over the maximal FOV of 80mm such that all trained detector combinations are used (detector combinations 1-21 to 1-29).



Source diameter	FWHM		FWTM	
	radial	tang.	radial	tang.
0.25mm	2.0mm	1.9mm	3.6mm	3.5mm
0.5mm	2.2mm	1.9mm	4.1mm	3.4mm
3mm	2.8mm	3.3mm	5.1mm	6.0mm

Figure 5.12: A density plot of the three reconstructed point sources spread over a FOV of 80mm. The table outlines the acquired FWHM and FWTM resolutions for each point source.

Also in this study the spatial resolution of the 0.25mm point source is in agreement with the results obtained in Paragraph 5.4 for a single point source located 35mm off centre. Radial and tangential FWHM/FWTM resolutions of respectively 2.2mm/4.0mm and 2.0mm/3.6mm were achieved.

Another common way to characterize the spatial resolution of a PET scanner is to scan a Derenzo phantom. A home made Derenzo phantom was used, consisting of capillary tubes with diameters tubes 1.5, 2.0, 2.5 and 3.0mm, arranged like slices in a pie (Figure 5.13). A centre to centre separation of twice the tubes diameter was taken. Rods of the phantom were filled with 5mCi FDG.

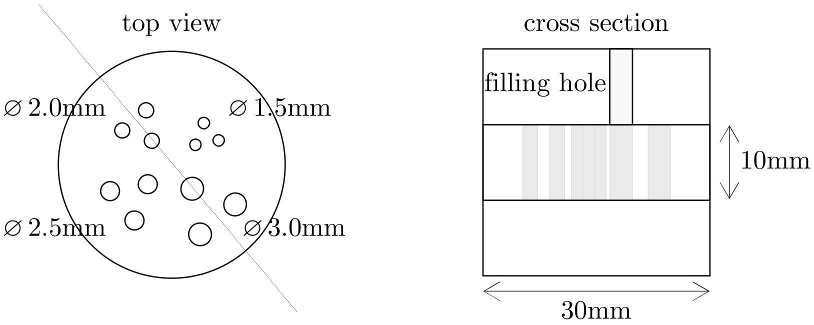


Figure 5.13: Top and cross sectional view of the home made Derenzo phantom. Diameters from 1.5mm to 3.0mm in steps of 0.5mm were evaluated.

FDG (Fluorodeoxyglucose) is one of the most commonly used positron emitter in PET. The fluorine in the FDG molecule is chosen to be the positron-emitting radioactive isotope fluorine-18. FDG, a glucose analogue, can be used for the assessment of glucose metabolism in the heart and the brain. It is also used for imaging tumors in oncology. FDG is taken up by cells and retained by tissues with high metabolic activity, such as most types of malignant tumors. As a result FDG-PET can be used for diagnosis, staging, and monitoring treatment of cancers [70].

Compared to ^{22}Na, FDG has a much shorter half-life, i.e. 110min. To overcome the variation in FDG activity during a scan, a new scan procedure was implemented. The FDG activity over time is illustrated in Figure 5.14. After 1080min (18hours) only a 0.1% fraction of the original activity remains.

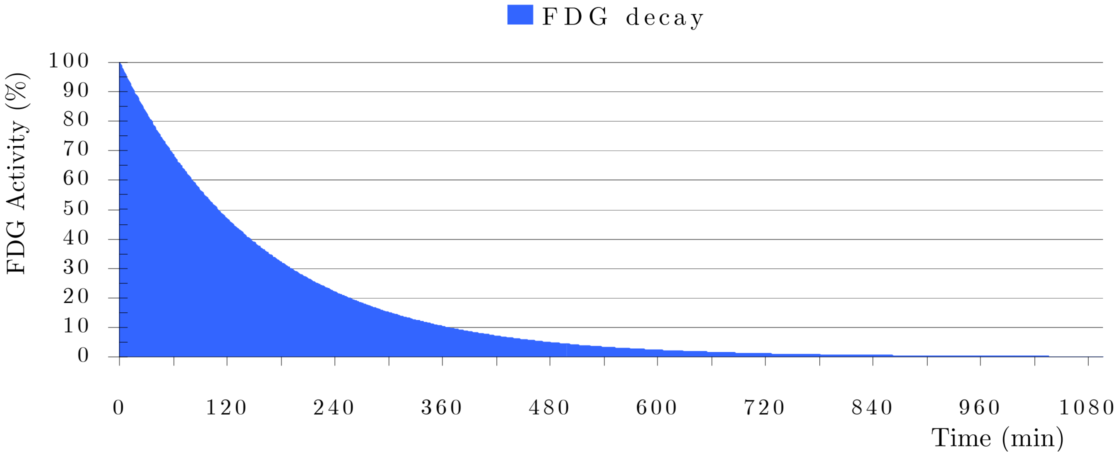


Figure 5.14: Illustration of the FDG decay over time

To measure the full 30mm diameter of the Derenzo phantom, 5 consecutive detector combinations are needed, i.e. detector combination 1-23 to 1-27. The angular source step α was kept to 360°/192 = 1.875° which corresponds to a quarter of one detector width. Accordingly, in the step-and-shoot mode, 960 different steps have to be measured.

Figure 5.15 shows the acquisition time per step such that each step is exposed to a same amount of FDG activity. In order to finish the scan within 6 hours, a start acquisition time of 7.5sec was taken. Taking into account the time needed for the 960 angular source movements and the time necessary to move to the 4 adjacent detector positions, the acquisition time of the last 960th step was extended to 61.5sec. The total scan completed after 5h48min.

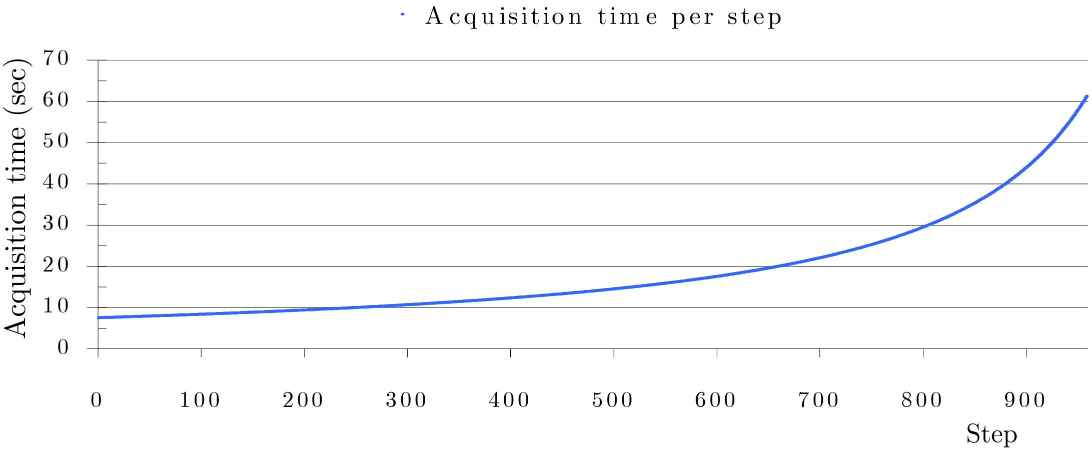


Figure 5.15: The acquisition time per step. Due to the short half-life of FDG the scan time per step is adjusted to ensure an equal amount of FDG exposure.

Assume a full ring PET scanner was available. In order to acquire all LORs measured in the prototype setup, the full ring scanner has only to rotate over 4 consecutive angular steps, i.e. positions 0°, 1.875°, 3.75° and 5.625°, reducing the total scan time to only 30sec (4x7.5sec).

1.045.568 coincidences were acquired during the scan. In order to obtain a uniform acquisition over all radial samples, only LORs within interval -20mm – 20mm are accepted (as shown in Figure 5.1. (c)) This reduces the number of acquired coincidence with 8% to 966.559.

Figure 5.16 and Figure 5.17 show the scanned Derenzo phantom reconstructed with the iterative listmode MLEM algorithm. Both plots give the FWHM/FWTM for each tube in the radial and tangential direction. An X denotes that it was impossible to define the resolution for the corresponding tube and direction. The first plot uses all 966.559 coincidences to reconstruct a 2D image of the Derenzo phantom. The tubes with diameters as small as 2.0 mm are still distinguishable. However there are symmetry issues as well as some tubes missing in the image. During an examination of the Derenzo phantom after the scan, it appears that some air bubbles were trapped into the small tubes. These air bubbles fully clarify the missing tubes due to the lack of FDG in these regions.

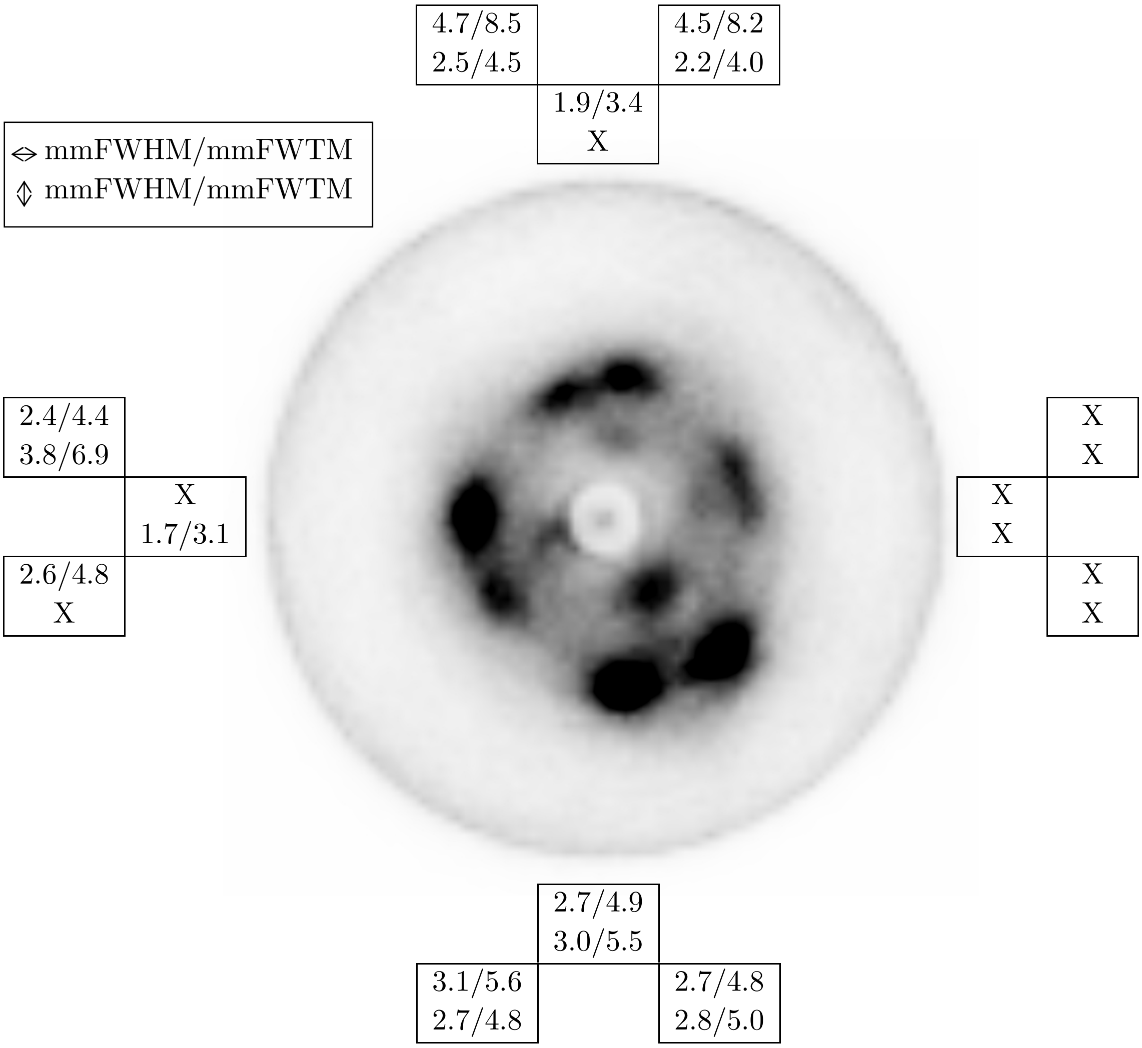


Figure 5.16: Reconstructed Derenzo phantom using all acquired 966.559 coincidences. The FWHM and FWTM resolutions are given for each cylinder in both directions. An X denotes that it was impossible to define the resolution for the corresponding tube and direction.

In all previous reconstructed images, no software energy cut was applied. The energy window corresponds to the hardware energy level of the discriminator fixed at approximately 300keV.

However the asymmetry reduces when an energy cut is applied as illustrated in Figure 5.17. The energy cut was such that only the photo peak remains. The FWTM resolutions slightly enhance but this at the expense of the sensitivity. The number of coincidences drops to 46% which results in 446.195 remaining coincidences.

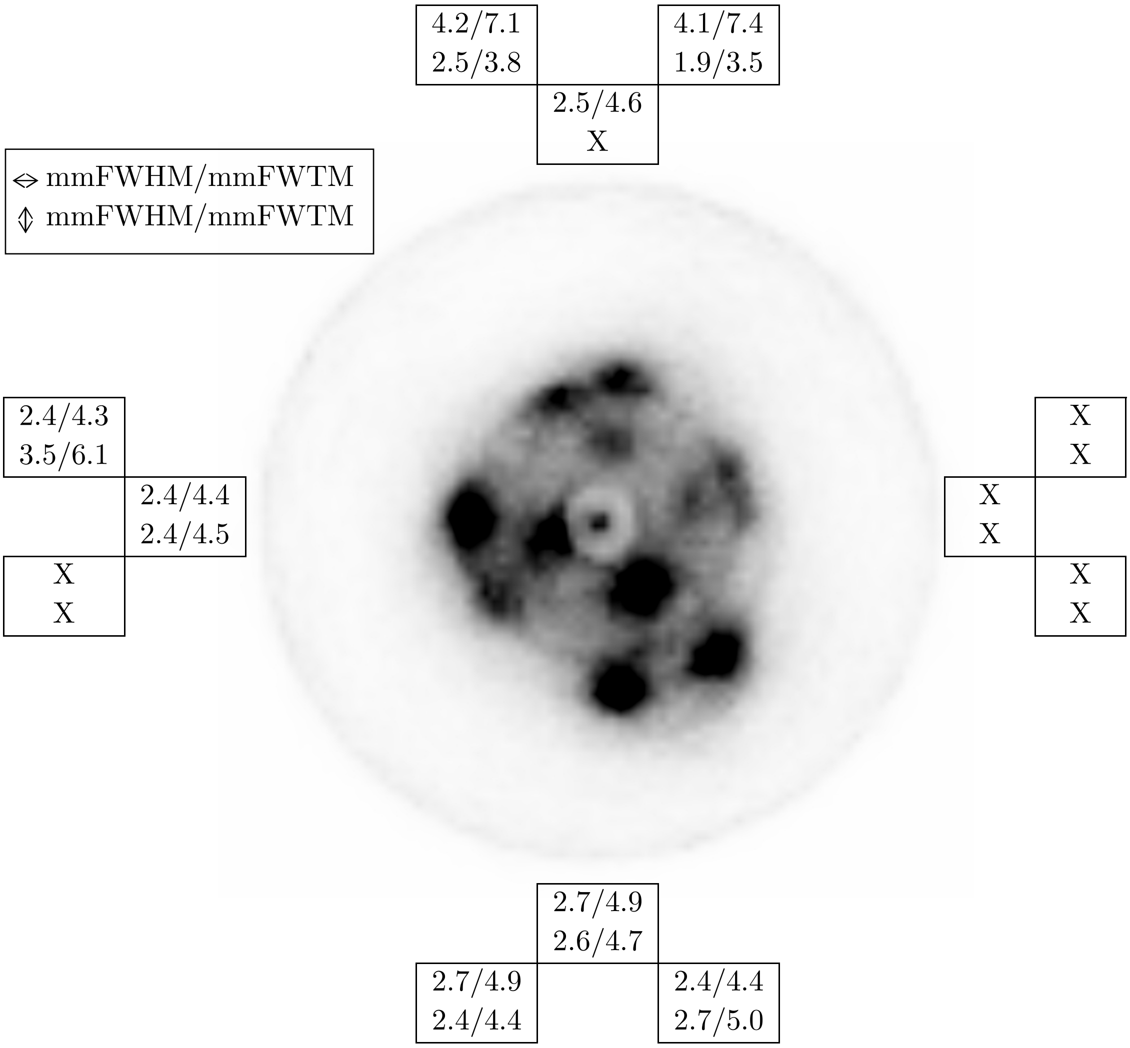


Figure 5.17: The FWHM and FWTM resolutions are given in both directions for each cylinder. An X denotes that it was impossible to define the resolution for the corresponding tube and direction. The reconstructed Derenzo phantom is based on only 446.195 coincidences. In order to enhance the FWTM resolutions, an energy cut was applied. Only coincidences within the photo peak are accepted.

Conclusion

Dedicated PET scanners such as small animal PET, Positron Emission Mammography (PEM) or Brain PET scanners, all require high spatial resolution and high sensitivity. Most current designs use small scintillation crystals. The general approach to improve the spatial resolution in such designs is to decrease the crystal size. However, the decreased pixel size results in loss of sensitivity because of the increased dead space between the pixels. If the sensitivity of the PET scanner is insufficient, the obtained images have to be smoothed to reduce the image variance. Obviously, this results in a loss of image resolution and hence the potential of the system is not fully exploited. To obtain a maximum coincidence rate, the sensitivity of the detectors in the PET system has to be optimized. This can be achieved by increasing the thickness of the scintillators used to stop the 511 keV annihilation photons and by minimizing the dead spaces in the detector design. However, these design changes should not degrade the spatial resolution of the scanner.

To achieve this goal we developed detectors based on monolithic scintillator blocks that are read out by avalanche photodiodes (APDs). This increases the sensitivity due to the absence of optical separation material between the individual scintillation pixels used in current PET designs. The position information within the scintillator block is embedded in the shape of the scintillation light distribution. This principle of light spreading allows the scintillator block to be larger than the sensitive area of the photo detector, avoiding dead space due to the packaging of the photo detector. This again enhances the sensitivity. In this perspective, the detector module had to be based on new technologies. For the scintillator part, Lutetium Orthosilicate (LSO) was chosen because of its high light yield, good stopping power and short decay time. The S8550 APDs were chosen as photo detector. These presented a number of advantages relatives to position sensitive photomultiplier tubes (PSPMTs) in the applications of interest.

In this thesis, the characteristics and implementation of the monolithic LSO scintillator blocks in combination with a machine learning positioning algorithm were evaluated, via simulations as well as experimentally on a bench set-up and on a prototype scanner. First three different positioning algorithms were tested experimentally on the bench set-up. The intrinsic detector resolutions of photons impinging at different angles e.g. 0°, ± 10°, ± 20°, ± 30° were studied. To this end, following positioning algorithms were evaluated: Levenberg-Marquardt Neural Networks (LM-NN), Neural Networks trained with an algebraic method (Alg-NN) and Support Vector Machines (SVM). The position information is extracted from the measured scintillation light distribution generated in monolithic LSO blocks of various shapes and read out by the Hamamatsu S8550 APD array.

The LM trained neural network has the best performance/algorithm complexity ratio of the three algorithms evaluated. For perpendicular incident photons the measured detector resolution on a 20x10x10mm^3 LSO block is 1.75mm FWHM and 4.3mm FWTM when a minimum threshold of 100 keV is applied, i.e. maximum sensitivity setting. Increasing the threshold improves the resolution. At a 380 keV threshold, the measured detector resolution is 1.6mm FWHM and 4.05mm FWTM. When neural networks are trained for different incidence angles, the localization procedure is capable of limiting the resolution degradation due to photon penetration for non perpendicular incident photons. At an incidence angle of 30° the measured detector resolution degrades to 2.0mm FWHM and 5.5mm FWTM using a threshold of 100 keV. Again better performance can be achieved at the cost of a lower sensitivity by increasing the energy threshold. The algebraic trained NNs can perform similarly but need a much larger neural network to do so, i.e at least 500 neurons in the hidden layer. Given the simplicity of the algebraic training algorithm, training a neural network with 500 neurons is still much faster than LM training of a neural network with 2 hidden layers with 5 neurons each. However, the computation of the incidence position using an algebraic trained NN with 500 neurons takes much longer. The SVM perform slightly better at the high sensitivity setting of the energy threshold (100 keV). To obtain this result, a few thousand support vectors are required which make this algorithm also rather slow to use on-line after it has been trained. The training speed is in between the LM training and the Alg-NN.

The LM training of a NN to find the incidence position in a thicker 20x10x20mm^3 LSO block performs worse in comparison to a thinner 20x10x10mm^3 LSO block. Consequently, a dual layer of 10mm thick LSO blocks is preferred to increase the overall sensitivity. Finally a comparison of LM trained NNs to find the position along the short axis of a rectangular 20x10x20mm^3 LSO block and a trapezoidal 20x15.4(11.5)x20mm3 LSO block shows that the resolution for photons impinging on the flat top surface is similar for both algorithm. The detector resolution for photons impinging on the slanted sides of the trapezoidal block is also similar except at the very edge.

The data acquired for previous positioning algorithm evaluations were done on an “academic” bench set-up. In order to evaluate the block detectors in a real compact PET environment, a prototype PET demonstrator was built. The demonstrator consists of only two 20x10x10mm^3 LSO detector modules. To simulate a full-ring scanner, the detector modules are mounted on separate rotating platforms which allow the movement of both detector modules, also relative to each other. In addition, since the detector characteristics may change in time, it is also appropriate to acquire new training data from time to time. The use of an auxiliary bench set-up for this calibration procedure implies the removal, calibration and re-mounting of all detector modules of the scanner. This would be a time consuming and tedious task. That’s why an automated acquisition method of training data for the positioning algorithm is investigated. The implementation and validation of this procedure was done on the demonstrator set-up. The results confirm the possibility of an automatic acquisition of training data. In a real PET scanner this can be implemented in a fully automated procedure by slowly spiralling a point source close to the detectors and only saving coincidences with LORs. Sorting the recorded coincidences per detector combination yields the required training data sets, similar to these obtained with the demonstrator setup.

However, a slight deterioration of the spatial resolution between the bench set-up and the demonstrator set-up was noticed. In order to study the origin of this difference and which instrumentation parameters limit the performance of the whole system, a GATE based Monte Carlo simulation was developed. Training data were simulated using the parameters that represent the experimental set-ups. After training, the NN was evaluated using simulated data generated with the same parameters except that the photon beam is assumed to be perfect now, i.e. a zero beam width. The resulting resolution will hence only reflect the influence of the detector components and the data acquisition method, i.e. it represents the intrinsic detector resolution.

The simulation was validated against the experimental bench set-up based on the 20x10x10mm^3 LSO block. The simulation shows that the photon beam used to generate a set of training events should have a width smaller than 1mm FWHM. It should also be stepped over the surface of the scintillator block with intervals not larger than 1mm. If these conditions are met, the resulting neural network will suffer no performance loss. However, the front-end amplifiers play the biggest role in the achievable resolution of the block detectors. Additional noise added by the front-end amplifiers degrades considerably the spatial resolution and causes the discrepancy between both experimental set-ups. In conclusion, improvements of the parameters in the existing set-up for the acquisition of training data will not improve the spatial resolution any more except for the preamplifier noise. Introducing a front-end amplifier with a lower noise will still enhance the resolution.

Finally; the spatial resolution in 2D reconstructed images of the monolithic front-end detectors in combination with the trained LM-NNs is also examined. The radial and tangential resolutions as function of the radial source position were tested. The radial and tangential resolutions remain constant around 2.0mm FWHM and 3.7mm FWTM in the 80mm FOV. Also uniform photon detection within the FOV was demonstrated. To conclude, a mini-Derenzo phantom filled with FDG showed very encouraging results and corresponds with the expectations according to the outcomes of the studied point sources.

Future outlook

The results from both experimental set-ups and Monte Carlo simulations allow a thorough evaluation of the performance of the detector blocks, illustrating their potential for high-sensitivity PET imaging of, e.g. the human brain. As a result, the monolithic blocks will be used on a research prototype for human brain PET/MRI imaging – the BrainPET scanner.

The BrainPET project is a joined effort of CIEMAT in collaboration with the VUB and Forschungzentrum Jülich. It will be a modular PET detector ring design capable of being inserted in existing clinical MRI scanners (Figure P1).

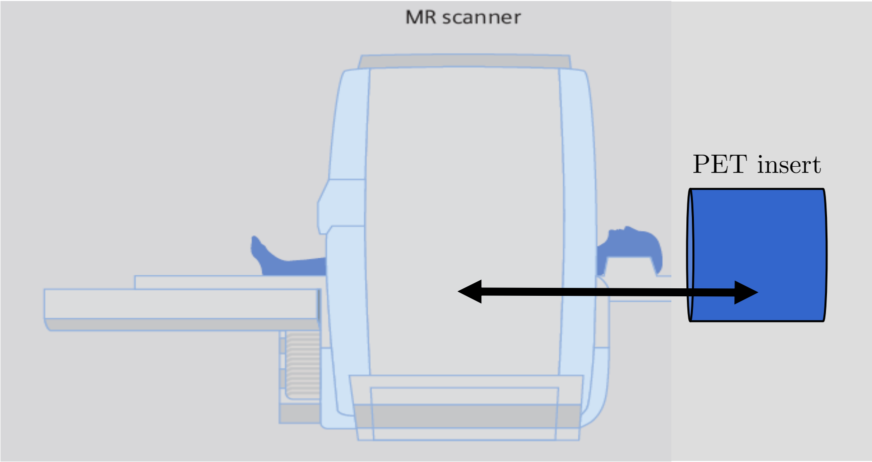


Figure P1: Illustration of the PET insert. PET only uses patient handling system of MR scanner. PET insert will be removed during MR imaging.

In order to maximize the sensitive volume, detector blocks are based on trapezoidal LSO scintillator crystals readout by the Hamamatsu S8550 APDs. A detector block is formed by two independent sub-detectors, each one coupled to a pair of APD matrices, with overall external dimensions determined by considering a 40cm scanner diameter (Figure P2). The crystals are encapsulated in $BaSO_4$ which acts both as an optical reflector and a mechanical stabilizer of the whole block. The pet insert consists of 4 rings (ø=40 cm) with 52 detector modules each

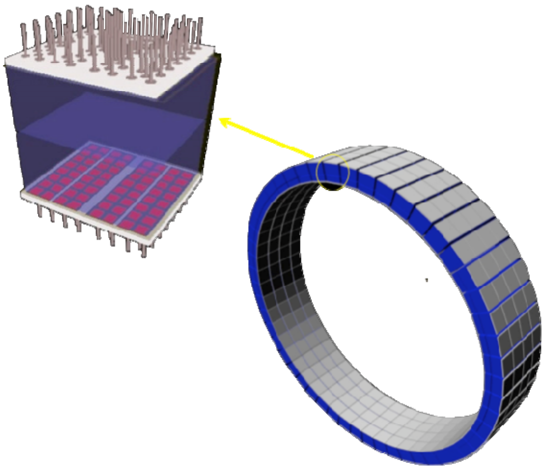


Figure P2: The proposed BrainPET trapezoidal detector block (left) and a full scanner composed of 4 rings of blocks (right).

A dedicated ASIC summing preamplifier has been developed. It will sum the 64 APD input signals to 16 differential output currents as the signals of the 8 pixels of each column and row provided by 2 APD arrays are summed internally by the preamplifier. Also the full sum of 64 channels will be provided, useful for the energy discrimination and normalization. The preamplifier has a RMS noise per sum channel of less than 1000e- when no APD array is connected to the inputs.

The incidence position on the detector surface will be derived from the sampled scintillation light distribution using Levenberg-Marquardt neural networks. In order to acquire the indispensable training data, the proposed automated "spiralling" technique will be used (Figure P3). A coincidence with an opposite block detector defines a narrow beam with known incidence position. This information together with the sampled light distribution is sufficient to train the NNs.

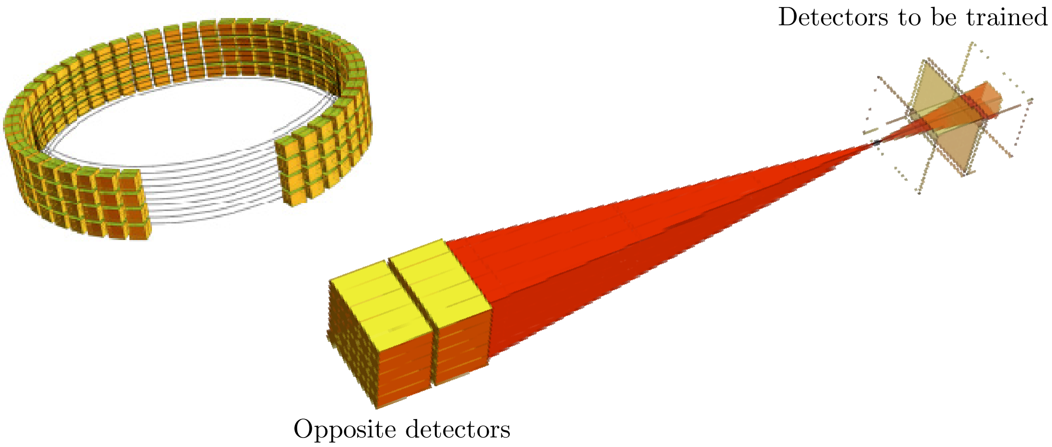


Figure P3: The "spiralling" method for the automatic acquisition of training data.

Bibliography

[1] F. Deconinck, "Nuclear imaging in the realm of medical imaging," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 509, pp. 213-228, 2003.

[2] M. V. Green, J. Seidel, C. A. Johnson, J. J. Vaquero, J. Pascau, and M. Desco, "Towards high performance small animal positron emission tomography," 2002, pp. 369-372.

[3] K. Ziemons, E. Auffray, R. Barbier, G. Brandenburg, P. Bruyndonckx, Y. Choi, D. Christ, N. Costes, Y. Declais, O. Devroede, C. Dujardin, A. Fedorovd, U. Heinrichs, M. Korjik, M. Krieguer, C. Kuntner, G. Largeron, C. Lartizien, H. Larue, P. Lecoq, S. Leonard, J. Marteau, C. Morel, J. B. Mosset, C. Parl, C. Pedrini, A. G. Petrosyan, U. Pietrzyk, M. Rey, S. Saladino, D. Sappey-Marinier, L. Simon, M. Streun, S. Tavernier, and J. M. Vieira, "The ClearPET(TM) project: development of a 2nd generation high-performance small animal PET scanner," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 537, pp. 307-311, 2005.

[4] P. Sempere Roldan, E. Chereul, O. Dietzel, L. Magnier, C. Pautrot, L. Rbah, D. Sappey-Marinier, A. Wagner, L. Zimmer, M. Janier, V. Tarazona, and G. Dietzel, "Raytest ClearPET(TM), a new generation small animal PET scanner," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 571, pp. 498-501, 2007.

[5] Raytest, "http://www.raytest.de/pet/clearPET/clearPET.html."

[6] P. A. Newcomb and P. M. Lantz, "Recent trends in breast cancer incidence, mortality, and mammography," *Breast Cancer Res Treat,* vol. 28, pp. 97-106, Nov 1993.

[7] J. Varela, "A PET imaging system dedicated to mammography," *Radiation Physics and Chemistry,* vol. 76, pp. 347-350, 2007.

[8] M. Zaidi, "The New Challenges of Brain PET Imaging Technology," *Current Medical Imaging Reviews,* vol. 2, pp. 3-13, 2006.

[9] Wikipedia, "Dirac equation," *Wikipedia The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=Dirac_equation&oldid=222665286, accessed 1 July 2008.

[10] C. D. Anderson, "The Positive Electron," *Physical Review,* vol. 43, p. 491, 1933.

[11] J. L. Humm, A. Rosenfeld, and A. Del Guerra, "From PET detectors to PET scanners," *Eur J Nucl Med Mol Imaging,* vol. 30, pp. 1574-97, Nov 2003.

[12] J. S. S.Cherry, M.Phelps *Physics in Nuclear Medicine: 3rd ed.; Saunders, 2003, 512 pp., ISBN 0-7216-8341-X*, 2003.

[13] A. Sánchez-Crespo, P. Andreo, and S. Larsson, "Positron flight in human tissues and its influence on PET image spatial resolution," *European Journal of Nuclear Medicine and Molecular Imaging,* vol. 31, pp. 44-51, 2004.

[14] D. Townsend, "Basic Science of PET and PET/CT," in *Positron Emission Tomography*, 2006, pp. 1-16.

[15] P. Zanzonico, "Positron emission tomography: a review of basic principles, scanner design and performance, and current systems," *Seminars in Nuclear Medicine,* vol. 34, pp. 87-111, 2004.

[16] W. R. Leo, *Techniques for nuclear and particle physics experiments*. United States: Springer-Verlag New York Inc.,New York, NY, 1987.

[17] C. L. Melcher, "Scintillation Crystals for PET," *J Nucl Med,* vol. 41, pp. 1051-1055, June 1, 2000 2000.

[18] K. S. Shah, P. Bennett, and M. R. Squillante, "Gamma ray detection properties of lutetium aluminate scintillators," *Nuclear Science, IEEE Transactions on,* vol. 43, pp. 1267-1270, 1996.

[19] D. Renker, "New trends on photodetectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 571, pp. 1-6, 2007.

[20] M. B. A. Bronstein, M. Zibulevsky, Y. Zeevi, "High-energy photon detection in positron emission tomography using adaptive non-linear parametric estimation algorithms," 2002.

[21] Hamamatsu, "Photodiode Technical Information," *www.hamamatsu.com*, 28 Feb 2003.

[22] Hamamatsu, "Characteristics and use of Si APD," *www.hamamatsu.com*.

[23] D. Renker, "Photosensors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 527, pp. 15-20, 2004.

[24] I. Britvitch, I. Johnson, D. Renker, A. Stoykov, and E. Lorenz, "Characterisation of Geiger-mode avalanche photodiodes for medical imaging applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 571, pp. 308-311, 2007.

[25] I. Tapan and R. S. Gilmore, "Simulation of signal generation for silicon avalanche photodiodes (APDs)," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 454, pp. 247-251, 2000.

[26] Y. Musienko, S. Reucroft, and J. Swain, "A simple model of EG&G reverse reach-through APDs," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 442, pp. 179-186, 2000.

[27] L. M. P. Fernandes, J. A. M. Lopes, and J. M. F. dos Santos, "Excess noise factor in large area avalanche photodiodes for different temperatures," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 531, pp. 566-568, 2004.

[28] A. Fremout, "Design study for a PET scanner based on the use of avalanche photodiodes and new scintillators," *Phd thesis,* 2002.

[29] I. Tapan, A. R. Duell, R. S. Gilmore, T. J. Llewellyn, S. Nash, and R. J. Tapper, "Avalanche photodiodes as proportional particle detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 388, pp. 79-90, 1997.

[30] R. Chen, A. Fremout, S. Tavernier, P. Bruyndonckx, D. Clement, J. F. Loude, and C. Morel, "Readout of scintillator light with avalanche photodiodes for positron emission tomography," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 433, pp. 637-645, 1999.

[31] D. J. van der Laan, M. C. Maas, H. W. A. M. de Jong, D. R. Schaart, P. Bruyndonckx, C. Lemaitre, and C. W. E. van Eijk, "Simulated performance of a small-animal PET scanner based on monolithic scintillation detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 571, pp. 227-230, 2007.

[32] M. Moszynski, M. Balcerzyk, W. Czarnacki, M. Kapusta, W. Klamra, A. Syntfeld, and M. Szawlowski, "Intrinsic energy resolution and light yield nonproportionality of BGO," *Nuclear Science, IEEE Transactions on,* vol. 51, pp. 1074-1079, 2004.

[33] W. W. Moses, P. R. G. Virador, S. E. Derenzo, R. H. Huesman, and T. F. Budinger, "Design of a high-resolution, high-sensitivity PET camera for human brains and small animals," *Nuclear Science, IEEE Transactions on,* vol. 44, pp. 1487-1491, 1997.

[34] S. Cherry and M. Dahlbom, "PET: Physics, Instrumentation, and Scanners," in *PET*, 2006, pp. 1-117.

[35] M.-J. Belanger, J. J. Mann, and R. V. Parsey, "OS-EM and FBP reconstructions at low count rates: effect on 3D PET studies of [11C] WAY-100635," *NeuroImage,* vol. 21, pp. 244-250, 2004.

[36] D. W. Townsend and M. Defrise, "Image-reconstruction methods in positron tomography," *CERN Report 93-02,* 1993.

[37] P. Bruyndonckx, C. Lemaitre, D. Schaart, M. Maas, D. J. van der Laan, M. Krieguer, O. Devroede, and S. Tavernier, "Towards a continuous crystal APD-based PET detector design," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 571, pp. 182-186, 2007.

[38] C. W. E. van Eijk, "Inorganic scintillators in medical imaging detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 509, pp. 17-25, 2003.

[39] Hamamatsu, "Si APD array S8550," *www.hamamatsu.com*, Datasheet.

[40] M. Kapusta, P. Crespo, D. Wolski, M. Moszynski, and W. Enghardt, "Hamamatsu S8550 APD arrays for high-resolution scintillator matrices readout," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 504, pp. 139-142, 2003.

[41] J. Kataoka, T. Saito, Y. Kuramoto, T. Ikagawa, Y. Yatsu, J. Kotoku, M. Arimoto, N. Kawai, Y. Ishikawa, and N. Kawabata, "Recent progress of avalanche photodiodes in high-resolution X-rays and [gamma]-rays detection," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 541, pp. 398-404, 2005.

[42] R. J. McIntyre, P. P. Webb, and H. Dautet, "A short-wavelength selective reach-through avalanche photodiode," *IEEE Transactions on Nuclear Science,* vol. 43, pp. 1341-1346, June 1996.

[43] R. Lecomte, C. Pepin, D. Rouleau, H. Dautet, R. J. McIntyre, D. McSween, and P. Webb, "Radiation detection measurements with a new "Buried Junction" silicon avalanche photodiode," *Nuclear Instruments and Methods in Physics*

Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 423, pp. 92-102, 1999.

[44] Cremat, "http://www.cremat.com/CR-110.pdf," Rev. 2 (Dec. 2006).

[45] P. Bruyndonckx, S. Leonard, C. Lemaitre, S. Tavernier, and Y. Wu, "Performance Study of a PET Detector Module Based on a Continuous Scintillator," *Nuclear Science, IEEE Transactions on,* vol. 53, pp. 2536-2542, 2006.

[46] P. Bruyndonckx, S. Leonard, S. Tavernier, C. Lemaitre, O. Devroede, W. Yibao, and M. Krieguer, "Neural network-based position estimators for PET detectors using monolithic LSO blocks," *Nuclear Science, IEEE Transactions on,* vol. 51, pp. 2520-2525, 2004.

[47] P. Bruyndonckx, C. Lemaitre, D. J. van der Laan, M. Maas, D. Schaart, O. Devroede, M. Krieguer, and S. Tavernier, "Comparison of Nonlinear Position Estimators For Continuous Scintillator Detectors In PET," in *Nuclear Science Symposium Conference Record, 2006. IEEE*, 2006, pp. 2518-2522.

[48] S. Léonard, "Spatial resolution study of PET detector modules based on LSO crystals an avalanche photodiode arrays," *Phd thesis,* 2005.

[49] S. Ferrari and R. F. Stengel, "Smooth function approximation using neural networks," *IEEE Trans Neural Netw,* vol. 16, pp. 24-38, Jan 2005.

[50] S. Ferrari and R. F. Stengel, "Classical/neural synthesis of nonlinear control systems," *Journal of Guidance Control and Dynamics,* vol. 25, pp. 442-448, May-Jun 2002.

[51] S. Ferrari and R. F. Stengel, "Algebraic training of a neural network," in *American Control Conference, 2001. Proceedings of the 2001*, 2001, pp. 1605-1610 vol.2.

[52] C. J. C. BURGES, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Min. Knowl. Discov.,* vol. 2, pp. 121-167, 1998.

[53] J. Sjöberg, "Mathematica, NEURAL NETWORKS, Train and analyze neural networks to fit your data," *Wolfram Research,* p. 406, 2005.

[54] D. L. Chester, "Why Two Hidden Layers are Better than One," *IJCNN-90-WASH-DC, Lawrence Erlbaum,* vol. 1, pp. 265-268, 1990.

[55] W. Wang and Z. Xu, "A heuristic training for support vector regression," *Neurocomputing,* vol. 61, pp. 259-275, 2004.

[56] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing,* vol. 14, pp. 199-222, 2004.

[57] K. Bennett and C. Campbell, "Support Vector Machines: Hype or Hallelujah?," *SIGKDD Explorations,* vol. 2, pp. 1-13, 2000.

[58] V. Cherkassky and Y. Ma, "Selecting the loss function for robust linear regression," *Neural Computation,* vol. Under review, 2002.

[59] A. Chalimourda, B. Scholkopf, and A. J. Smola, "Experimentally optimal [nu] in support vector regression for different noise models and parameter settings," *Neural Networks,* vol. 17, pp. 127-141, 2004.

[60] Chih-Chunag Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm,* 2001.

[61] WolframResearch, "Mathematica 5.2," *www.wolfram.com,* 2005.

[62] M. C. Maas, D. J. van der Laan, D. R. Schaart, J. Huizenga, J. C. Brouwer, P. Bruyndonckx, S. Leonard, C. Lemaitre, and C. W. E. van Eijk, "Experimental characterization of monolithic-crystal small animal PET detectors read out by APD arrays," *Nuclear Science, IEEE Transactions on,* vol. 53, pp. 1071-1077, 2006.

[63] D. J. van der Laan, M. C. Maas, D. R. Schaart, P. Bruyndonckx, S. Leonard, and C. W. E. van Eijk, "Using Cramer-Rao theory combined with Monte Carlo simulations for the optimization of monolithic scintillator PET detectors," *Nuclear Science, IEEE Transactions on,* vol. 53, pp. 1063-1070, 2006.

[64] Wikipedia, "NIM-standard," *Wikipedia The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=NIM&oldid=222731439, accessed 19-July-2008.

[65] G. Anelli, K. Borer, L. Casagrande, M. Despeisse, P. Jarron, N. Pelloux, and S. Saramad, "A high-speed low-noise transimpedance amplifier in a 0.25 [mu]m CMOS technology," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 512, pp. 117-128, 2003.

[66] D. W. Townsend, M. Wensveen, L. G. Byars, A. Geissbuhler, H. J. Tochon-Danguy, A. Christin, M. Defrise, D. L. Bailey, S. Grootoonk, A. Donath, and R. Nutt, "A Rotating PET Scanner Using BGO Block Detectors: Design, Performance and Applications," *J Nucl Med,* vol. 34, pp. 1367-1376, 1993.

[67] "Gate," *OpenGATE Collaboration,* http://opengatecollaboration.healthgrid.org/opengate/.

[68] G. Ren and et.al, "Scintillation characteristics of lutetium oxyorthosilicate (Lu2SiO5:Ce) crystals doped with cerium ions," *Nucl. Instr. and Meth.,* A 531 (2004) 560–565.

[69] S. Vandenberghe, Y. D'Asseler, M. Koole, L. Bouwens, R. Van de Walle, I. Lemahieu, and R. Dierckx, "Iterative list mode reconstruction for coincidence data of gamma camera.," *Proceedings of SPIE/Medical imaging 2000,* vol. 3979, pp. 1538-1546, 2000.

[70] Wikipedia, "Fluorodeoxyglucose," *Wikipedia The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=Fluorodeoxyglucose&oldid=219364751, accessed 21-June-2008.

List of abbreviations

ADC	analog-to-digital convertor
Alg-NN	algebraic trained NN
AMP	amplifier board
APD	avalanche photo diode
ART	algebraic reconstruction technique
BaSO4	barium sulfate
BGO	bismuth germanate
CCC	Crystal Clear Collaboration
CERN	Conseil Européen pour la Recherche Nucléaire
CFD	constant fraction discriminator
CFOV	Centre of FOV
CIEMAT	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
CRF	coincidence response function
CT	computed tomography
DAQ	data acquisition
DIG	digital board
DOI	depth-of-interaction
ECG	electrocardiogram
EEG	electroencephalogram
e-h pair	electron-hole pair
ENC	equivalent noise charge
EOPN	equivalent optical photon noise
F	excess noise factor
FBP	filtered-back projection
FDG	18F–fluoro-deoxy-glucose
FF	feed forward neural network
FOV	field-of-view
fRMI	functional magnetic resonance imaging
FW20M	full width at 20% of the maximum amplitude
FWHM	full width at half maximum
FWTM	full width at tenth maximum
FZJ	Forschungszentrum Julich
GATE	Geant4 application for tomographic emission
GM-APD	geiger-mode avalanche photo diode
Gp	photon gain
IIHE	Interuniversity Institute for High Energies

LIP	Laboratório de Instrumentação e Física Experimental de Partículas
LM-NN	Levenberg-Marquardt trained NN
LOR	line-of-response
LS	least square
LSO	lutetium orthosilicate
LuAP	lutetium aluminate perovskite
LY	light yield
MEDISIP	Imaging and Signal Processing
ML-EM	maximum likelihood expectation maximization
MRI	magnetic resonance imaging
MSE	mean square error
NECR	noise equivalent count rate
NN	neural network
P-AMP	preamplifier board
Pb	lead (plumbum)
PCB	printed circuit board
PEM	positron emission mamography
PET	positron emission tomography
PMT	photomultiplier tube
PSPMT	position sensitive photo multiplier tube
QE	quantum efficiency
RMS	root mean square
RMSE	root mean square error
SiPM	silicon photomultiplier
SNR	signal-to-noise-ratio
SPECT	single photon emission computed tomography
SVM	support vector machines
SVR	support vector regression
VUB	Vrije Universiteit Brussel

Promotoren :

Prof. Dr. Peter Bruyndonckx
Prof. Dr. Stefaan Tavernier

Departement Natuurkunde
Faculteit Wetenschappen
Vrije Universiteit Brussel

2009

Proefschrift ingediend met het oog op het behalen van de academische graad Doctor in de Wetenschappen

Vrije Universiteit Brussel – Faculteit Wetenschappen
Pleinlaan 2 – 1050 Brussel
Contact: - http://we.vub.ac.be/-