POLITECNICO DI MILANO

School of Industrial and Information Engineering

Department of Electronics, Information and Bioengineering (DEIB)

Master of Science in Biomedical Engineering

MACHINE LEARNING-BASED EVENT RECONSTRUCTION IN GAMMA CAMERAS FOR MEDICAL IMAGING

Supervisor: Prof. Carlo Ettore Fiorini

Co-Supervisors: Ing. Luca Buonanno

Ing. Ilenia D'Adda

Master Thesis by:

Carlo Carmelo Alaimo

Matr. 918791

Academic Year 2019 - 2020

Contents

List of Figures

List of Tables

Sommario

Al giorno d'oggi, l'uso di strumenti di machine learning e deep learning è stato esteso a innumerevoli applicazioni, grazie anche al progresso tecnologico e al conseguente aumento della potenza di calcolo disponibile, che rende possibile elaborare grandi quantità di dati in un tempo notevolmente ridotto.

La new wave del machine learning ha investito anche il mondo dell'imaging medico, dimostrando di rappresentare un potente strumento per diverse applicazioni; in particolare, nel caso delle tecniche di gamma imaging come PET e SPECT, i principali campi di applicazione sono stati la Computer-aided Detection (CAD), dove il machine learning è coinvolto come strumento di supporto decisionale per i medici al fine di eseguire diagnosi precoci e più accurate a partire da immagini mediche, l'elaborazione di immagini, principalmente come strumento per la riduzione del rumore, e, infine, la localizzazione della posizione di scintillazione di fotoni γ per rivelatori monolitici e pixelati.

Questo progetto di tesi può essere collocato esattamente nell'ultimo contesto; il lavoro è consistito nell'implementazione e nella valutazione di due diversi approcci basati sul machine learning per la stima delle coordinate di interazione dei fotoni γ in una gamma camera per SPECT imaging.

Il primo approccio è consistito in una tecnica di unsupervised learning, denominata Principal Component Analysis (PCA). La PCA è una tecnica di riduzione della dimensionalità, che consiste nell'eseguire una trasformazione lineare dello spazio originale dei dati in un nuovo spazio con dimensionalità ridotta. Questa trasformazione è eseguita in modo

tale che l'informazione contenuta nei dati originali venga mantenuta pressoché invariata. La riduzione con PCA è stata integrata all'interno di due metodi di ricostruzione statistica già in uso per la ricostruzione di eventi gamma, il metodo della Maximum Likelihood (ML) e il metodo Least Squares (LS); tuttavia, a differenza di questi ultimi, i metodi proposti, ML-PCA e LS-PCA, non operano nello spazio originario dei dati, ovvero quello costituito dai segnali dei fotorivelatori, ma in uno spazio a bassa dimensionalità ottenuto dopo la riduzione con PCA.

Il secondo metodo ha utilizzato un'altra tecnica di machine learning, costituita dai Decision Trees. I Decision Trees sono una tecnica di supervised learning utilizzata per risolvere problemi di classificazione e regressione. In particolare, nell'ambito del progetto di tesi, il problema della localizzazione delle coordinate di scintillazione (x,y) del fotone gamma all'interno dello scintillatore è stato risolto dopo un'opportuna conversione in un problema di classificazione discreto, che è stato poi affrontato implementando una cascata di Decision Trees.

Le misure di validazione simulate e sperimentali per le due tecniche sono state eseguite su una gamma camera continua, disponibile presso il Politecnico di Milano e sviluppata nell'ambito di un precedente progetto di ricerca, denominato INSERT.

INSERT è un progetto di ricerca finanziato dal "Seventh Framework Program " della Commissione Europea e avviato l'1 Marzo del 2013; l'obiettivo era lo sviluppo di un compatto inserto SPECT da integrare all'interno di scanner commerciali per la risonanza magnetica, al fine di una migliore stratificazione del tumore al cervello e una precoce valutazione dell'efficacia del trattamento. Nel contesto del progetto INSERT, il gruppo del Politecnico di Milano è stato incaricato dello sviluppo dei moduli di rilevamento gamma dello scanner SPECT e della relativa elettronica di lettura. Sono stati sviluppati due tipi di sistemi di rilevamento per diverse applicazioni: un sistema preclinico, destinato all'imaging di piccoli animali, e uno clinico, con dimensioni superiori al preclinico e utilizzato per l'imaging della testa e del collo nell'uomo.

In questo lavoro di tesi, è stato utilizzato il modulo clinico di INSERT,

sia per le misure simulate che per quelle sperimentali.

La discussione è organizzata in sei capitoli.

Il primo capitolo fornirà una panoramica sulle radiazioni gamma, le loro caratteristiche e le loro applicazioni nell'imaging medico. Inoltre, verranno descritte le caratteristiche e i componenti di una gamma camera, con un focus particolare sullo scanner INSERT SPECT.

Il secondo capitolo sarà incentrato sui SiPMs, i fotorilevatori implementati nel modulo INSERT, i loro principi di funzionamento e le loro principali cifre di merito. Nell'ultima parte del capitolo, verrà fornita una breve descrizione delle modalità di lettura dei SiPMs e, in particolare, della strategia utilizzata in INSERT.

Il terzo capitolo descriverà i principali metodi di ricostruzione che sono comunemente utilizzati per la ricostruzione planare dell'immagine SPECT, sottolineando i corrispondenti vantaggi e limiti. Inoltre, verranno descritti i metodi di ricostruzione implementati nel sistema INSERT.

Il quarto e il quinto capitolo, invece, introdurranno rispettivamente il metodo di ricostruzione basato sulla PCA e quello basato sui DT; in entrambi i capitoli, inizialmente, verranno introdotti i principi teorici e le applicazioni di questi due metodi nel campo dell'imaging medico. Successivamente, verrà descritto il processo utilizzato per implementare i due metodi di ricostruzione e, infine, verranno mostrati i corrispondenti risultati ottenuti sulle simulazioni e sulle misure sperimentali.

L'ultimo capitolo, invece, trarrà alcune conclusioni finali sul lavoro e possibili sviluppi futuri.

Abstract

Nowadays, the use of machine learning and deep learning tools has been extended to innumerable applications, accomplished by the advancement in technologies and the consequent increase in the available computational power, which makes possible to process big amount of data in a reduced time.

The wave of machine learning invested medical imaging world too, proving to represent a powerful instrument for different applications; in particular, in the case of γ-ray imaging techniques like PET and SPECT, the main fields of application have been Computer-aided Detection (CAD), where machine learning is involved as a decisional support for physicians in order to perform early or more accurate diagnosis from medical images, image processing, mostly as tool for the reduction of noise in the image, and, finally, the localization of the scintillation position of γ photons for both monolithic and pixelated detectors.

This thesis project can be placed exactly in the last framework; the work consisted in the implementation and evaluation of two different machine learning-based approaches for the estimation of the interaction coordinates of γ photons in a gamma camera for SPECT imaging.

The first approach consisted in a unsupervised learning technique, called Principal Component Analysis (PCA). PCA is a dimensionality reduction technique, which consists in performing a linear transformation of the original data space into a new space with lower dimensionality. This transformation is implemented in a way that the information contained in the original data is mostly preserved. PCA reduction has been integrated within two statistical reconstruction methods already in use

for event reconstruction, Maximum Likelihood (ML) and Least Square (LS); however, differently from these latter, the proposed methods, ML-PCA and LS-PCA, do not operate in the original space, that is the one constituted by the photodetectors signals, but in a low-dimensional space obtained after the PCA reduction.

The second method made use of another machine learning technique, which is represented by Decision Trees. Decision Trees are a supervised learning method, which is used in order to solve classification and regression problems. In particular, in the framework of the current thesis project, the problem of localization of the (x,y) scintillation coordinates of the gamma photon inside the scintillator has been solved by converting it into a discrete classification problem, which has been then addressed by implementing a cascade of decision trees.

The simulated and experimental validation measurements for the two techniques have been carried out on a continuous gamma camera available at Politecnico di Milano and developed in the framework of a previous research project, called INSERT.

INSERT is a research project funded by the Seventh Framework Program of the European Commission and started on March 1st, 2013; the goal was the development of a compact SPECT insert to be integrated inside commercial MR scanners for enhanced stratification of brain tumor and early assessment of treatment efficacy. In the context of INSERT project, Politecnico di Milano group was responsible for the development of the gamma ray detection modules of the SPECT scanner and the related readout electronics. Two types of detection systems have been developed for different applications: a preclinical system, aimed at small animals imaging, and a clinical one, with higher dimensions than the preclinical and used for human head/neck imaging. In this work, the clinical INSERT detection module has been used for running both simulated and experimental measurements.

The discussion is organized into six chapters.

The first one will provide an overview on gamma radiations, their characteristics and their applications in medical imaging. Furthermore,

the characteristics and components of a gamma camera will be described, with a particular focus on INSERT SPECT scanner.

The second chapter will focus on SiPMs, the photodetectors implemented in INSERT module, their working principles and figures of merit. In the last part of the chapter, a brief description of the SiPMs readout strategy implemented in INSERT will be also provided.

The third chapter will describe the main reconstruction methods that are commonly used for the planar reconstruction of the image in SPECT, showing their strengths and limitations. Furthermore, the reconstruction methods implemented in INSERT system will be highlighted.

The fourth and fifth chapters, instead, will introduce respectively the PCA-based reconstruction and the DT-based reconstruction; in both chapters, in the first place, the theoretical principles and the applications of these two methods in medical imaging will be introduced. Successively, the process followed in order to implement the two reconstruction methods will be described and, finally, the corresponding results on simulations and experimental measurements will be shown.

The last chapter, instead, will draw some final conclusions about the work and possible future developments.

Chapter 1

Gamma imaging and INSERT

Functional imaging techniques, like PET and SPECT, constitute nowadays a major instrument in the field of nuclear medicine and diagnostics. Both of them are based on the detection of gamma radiations, in order to image body tissues. This introductory chapter will provide an overview of gamma radiation characteristics, their detection modalities and their main applications in the imaging field. Moreover, the concept of multimodality will be addressed, with a particular attention on INSERT project, the multimodal imaging system on which the proposed reconstruction methods have been validated.

1.1 Gamma rays

Gamma (or γ) radiations are a type of electromagnetic radiation, whose energy range is tipically higher than tens of keV. Looking at the electromagnetic spectrum (figure 1.1), it is possible to observe that there is a region where γ-rays and x-rays, the other type of radiation commonly used in medical imaging, coexist: indeed, γ-radiations which are emitted by radioactive decay usually possess energies between some keV to tens of MeV. However, both the mechanisms of generation and interaction with the matter of these two radiations are substantially different.

γ-rays can originate by two phenomenons: the first one is the decay

of the nucleus of a radionuclide, while the other one is represented by positron annihilation.

x-rays, instead, are emitted by electrons outside the nucleus via fluorescence, Bremsstrahlung phenomenon and synchrotron light.

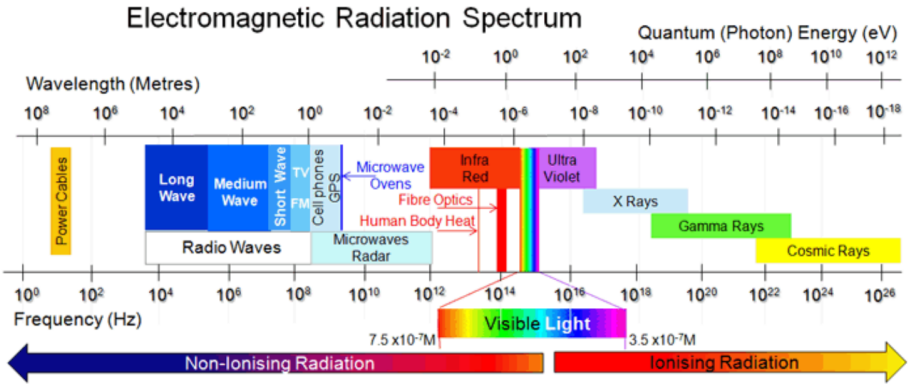


Figure 1.1: *Electromagnetic spectrum*

1.1.1 Mechanisms of generation

γ-radiations can be generated by different phenomena:

- emission from the nucleus of a radionuclide after radioactive decay
- generation of two γ-rays, travelling in opposite directions, caused by positron annihilation

The first phenomenon consists in the emission of γ radiations from radioactive isotopes, which are atoms whose inner core, their nucleus, is unstable. By decaying, the nuclei of these atoms change their composition and properties to reach a less energetic and more stable state, and simultaneously they emit radiations (not only γ, but also α and β).

The positron annihilation, instead, consists in the emission of two opposite γ-radiations, following the interaction between an electron and a positron. In β^+ (positron) decay, a nuclide transforms one of its core protons (p) into a neutron (n) and emits a positron (β^+), which is a particle

with the same identical mass of an electron but positively charged [1], and a neutrino (ν):

$$p \rightarrow n + \beta^{+} + \nu \tag{1.1}$$

The average positron range in matter depends on the positron's energy and material characteristics, such as the density and the atomic number. At the end of its path, the positron, being anti-matter to electrons, will recombine with an atomic electron, giving rise to the annihilation phenomenon. By annihilating, the electron and the positron convert their mass into energy and produce a pair of 511 keV annihilation photons traveling in (almost) opposite directions. The 511 keV photon energy (E) comes from Einstein's famous equation:

$$E = mc^2 \tag{1.2}$$

where m is the mass of the electron or positron and c is the speed of light.

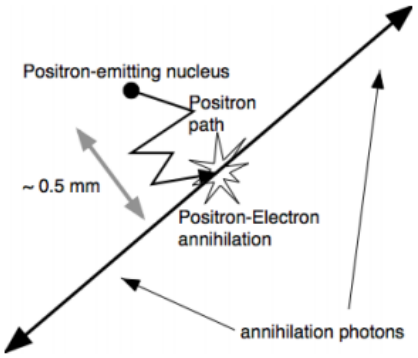


Figure 1.2: *Annihilation mechanism*

1.1.2 Interactions of gamma-rays with matter

In general, there are three main phenomena in nature which allow the absorption of a radiation by a material [2]:

a) **Photoelectric absorption** consists in the interaction between an incident photon and an inner shell electron in the atom. The electron

acquires energy from the photon, and so it reaches an energetic level which allows it to be removed from its shell.

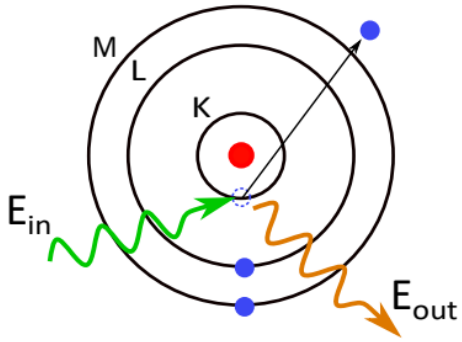


Figure 1.3: *Photoelectric absorption*

It is important to underline that the incident photon is completely absorbed in the process; this explains how the photoelectric effect contributes to the attenuation of the radiations beam as it passes through matter.

The kinetic energy possessed by the photoelectron, which is equal to the amount of energy lost by the beam, is given by the following equation:

$$E_{loss} = E_k = E_{ph} - E_{bind} \tag{1.3}$$

where E_{ph} is the energy of the incident photon, while E_{bind} is the binding energy of the shell. Indeed, when the electron is kicked out from its shell, it leaves a vacancy (a positive charge), which is quickly occupied by an electron from a close more energetic shell; this phenomenon can be accompanied by the emission of a radiation (fluorescence) or not. The x-rays emitted by fluorescence are usually reabsorbed by photoelectric interactions, but, if their energy is too high, they can escape the material, thus appearing subsequently in the reconstructed emission spectrum.

b) **Compton scattering** is an interaction process characterized by the fact that not all the energy of the incoming photon is released during the interaction. Indeed, in this case, the photon does not kick out an internal electron; it just interacts with a weakly bound electron, loosing part of its energy and deviating from its original trajectory (scattering).

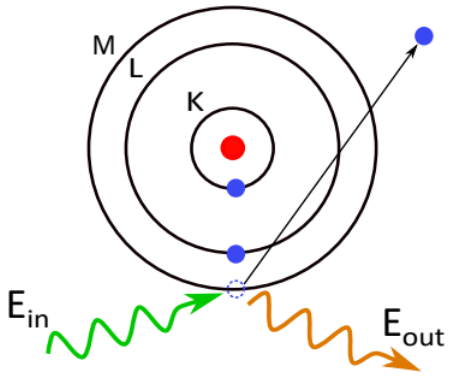


Figure 1.4: *Compton scattering*

More precisely, the residual energy after the scattering is given by the difference between the energy of the impinging photon and the energy of the emitted electron. Since not only energy but also momentum has to be conserved, the amount of energy loss during the interaction and the interaction probability itself, depend on the angle of collision, according to the Klein-Nishina formula [3]:

$$E_{loss} = E_{ph} \cdot \left(1 - \frac{1}{1 + \frac{E_{ph}}{m_0 c^2} \cdot (1 - \cos\theta)}\right) \tag{1.4}$$

Note that a particular case of Compton interaction is *Rayleigh scattering* or *elastic scattering* , which is characterized by a deflection of the original photon with no loss of energy; this happens because the energy possessed by the photon is not sufficiently high to kick-out an electron.

This phenomenon is particularly "dangerous" in medical imaging

techniques like PET and SPECT. Indeed, elastic scattering can also happen inside the body of the patient; when the elastically scattered photons leave the body of the patient, they are detected by the camera, which is not able to discriminate an unscattered photon from a photon which has been elastically scattered, being their corresponding energies equal.

Thus, elastically scattered γ photons are associated to a wrong scintillation position, and this phenomenon usually worsens the quality of the image.

c) **Pair production** is an interaction process which happens generally at higher energy than the other two: indeed, in pair production, the impinging radiation is absorbed through the generation of an electron-positron pair. In other words, if the gamma photon has a energy at least equal to 1,022 MeV (twice the energy of an electron, which is given by $2m_0c^2$) and if it's located in the coulomb field of a nucleus, it loses the previously mentioned amount of energy, generating a electron-positron pair. Of course this phenomenon can happen also at energy higher than 1,022 MeV.

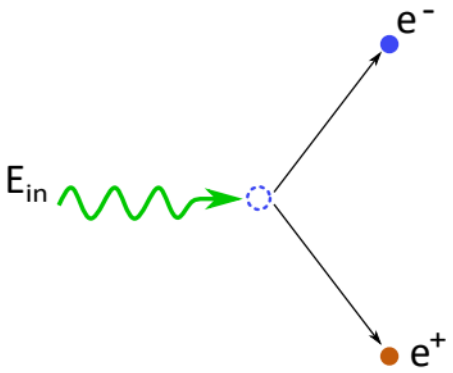


Figure 1.5: *Pair production*

Each of the three mechanisms described above gives its contribution to the overall attenuation coefficient, which is obtained by the following

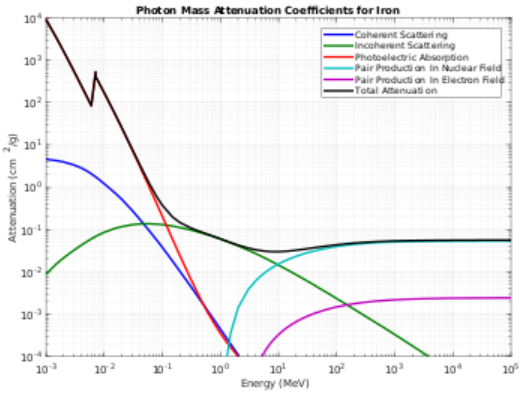


Figure 1.6: *Attenuation coefficient of iron*

relationship:

$$\mu = \mu_{photoelectric} + \mu_{compton} + \mu_{pairs} \tag{1.5}$$

Photon attenuation coefficient has a crucial importance in the absorption process, because it defines the rate of the exponential decay of electromagnetic radiation intensity as a function of the depth of penetration, according to the Beer-Lambert equation:

$$I(x) = I_0 \cdot e^{-\mu \cdot x} \tag{1.6}$$

where x is the depth of penetration and I_0 is the intensity of the incident beam.

In figure 1.6, by way of an example, the attenuation coefficient of iron is plotted in a log-log scale, in function of the energy of the radiation. It can be noticed how, at lower energies, the predominant contribution to the total attenuation coefficient comes from photoelectric interactions, while at higher energy values Compton and pair productions become more significant. Since the attenuation coefficient depends also on the atomic number of the absorbing material, this plot is characteristic of each material.

1.2 Gamma imaging

Throughout the years, the interest of the scientific community in γ radiation detectors has grown more and more, since they have proved to have applications in different fields, from nuclear physics to medical imaging. In the field of nuclear physics, the main application is spectroscopy, which consists in measuring the emission spectrum of a sample, in order to characterize it by individuating, for example, the type of present γ-emitter, or estimating its concentration.

Another major application of γ-rays is medical diagnostic. Indeed, x-rays imaging techniques like CT or radiography provide informations about the morphology and anatomy of a tissue, which, often, are not sufficient in order to obtain a complete clinical framework required for diagnosing and staging a disease. Many times, functional or metabolic changes can happen independently from anatomical ones.

The introduction of γ-rays in medical imaging made possible to individuate and monitor functional and metabolic changes of a tissue, with high resolution [4]. The principle exploited by both PET and SPECT, which are the two main medical imaging techniques employing γ-radiations, is to obtain images representing the distribution of γ-rays emitters inside the body of the patient: a substance containing specific receptors, tagged with a radionuclide is injected in the patient's body and then the tracer spreads and accumulates in different regions proportionally to the rate of delivery of nutrients to tissues.

1.2.1 PET/SPECT

Nowadays, SPECT (Single-Photon Emission Computed Tomography) and PET (Positron Emission Tomography) represent the most used functional imaging techniques. Even if they are both based on the detection of γ radiations and, consequently, they both exploit a gamma camera, their working principle is quite different.

In SPECT, spatial resolution is introduced in the system by the use of collimators, that absorb γ-photons not coming from a specific direction

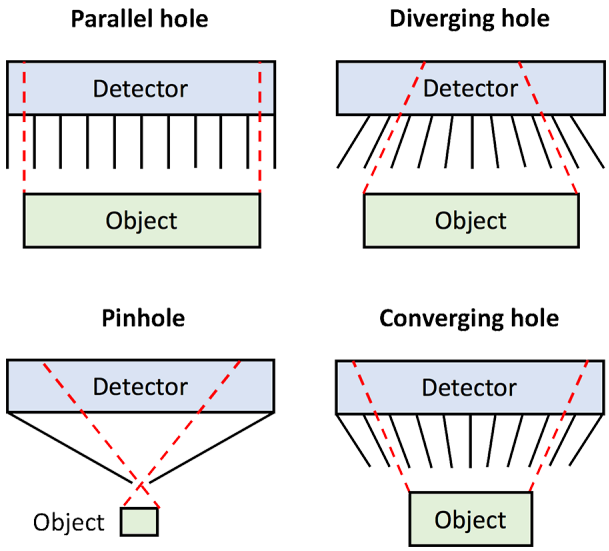


Figure 1.7: *Different types of collimators that accept photons incoming from different directions*

[5]. As a result, only photons emitted from a specific direction are detected by the system and the emission position can be ascertained. The main drawback of using a collimator is the reduced efficiency of the detector, since many emitted photons are absorbed by the collimator itself and not detected. The consequence of the efficiency reduction is that, in order to collect a significant amount of signal, the dose injected into the patient has to be be increased, thus limiting SPECT application to children or at-risk subjects.

Furthermore, when using a collimator, a trade-off between the geometrical efficiency of the collimator and the achievable spatial resolution is introduced; in order to improve the spatial resolution, geometrical efficiency unavoidably has to be decreased.

Differently from SPECT, PET detects the interaction position of the two annihilation photons that are produced back-to-back after positron emission from a radionuclide tagged tracer molecule, which is chosen to mark a specific function in the body on a biochemistry level. Due to the positron annihilation, we expect to observe two photons at roughly the same time (in coincidence) in the detector ring; the annihilation event,

i.e. the radioactive tracer, will be then located somewhere on the line connecting the two photon-detection points.

This "electronic collimation" represents an advantage respect to SPECT, since the knowledge of the photon direction prevents from the use of collimators in order to restrict all the possible directions of the photons reaching the detectors; this leads to an improvement of the detection efficiency, reducing also the dose to be injected.

The reasons why photons detections do not happen at the exact same time are different. The first is that the annihilation event may occur closer to one detector surface than the other, leading to a slight but measurable delay of one photon respect to the other. Another crucial factor which causes the temporal delay between the two detections is the finite timing resolution of the detector, i.e. its timing uncertainty, which arises from the decay time of the scintillation in the crystal and the processing time of the photodetectors signals.

These effects lead to the use of a coincidence time window on the order of 6-10 ns [5]. If the two detections of the photons happen within each other's coincidence window, they are assumed to be originating from the same annihilation event, which is attributed to the line-of-response (LOR) that connects the two detection points.

The problem of measuring the temporal delay between the two detections with a high accuracy becomes central in the Time-of-flight (TOF) PET: with TOF-PET imaging the relative time difference between the detection of the two annihilation photons is used to determine the most likely location of the annihilation event along the LOR [6].

A second important difference between SPECT and PET is the type of radioactive tracers used for the two techniques. Tracers consist in radiolabelled biomarkers that are introduced into the patient's organism, mainly to examine his organ or tissues functions. The labelling procedure consists in chemical linking of a radionuclide to the specific structure of a biomarker [7].

The main figures of merit to consider when dealing with radioactive sources are:

- **Activity:** is defined as the number of decaying atoms per unit time in a radioactive sample and measured in Becquerel [s^{-1}], equivalent to one disintegration per second.

  The activity of a radioactive source decreases exponentially, in accordance to the exponential-decay law:

  $$A(t) = A_0 e^{-\lambda(t-t_0)} \tag{1.7}$$

  where A_0 is the number of decaying atoms per unit time at the starting time t_0, $A(t)$ is the decays rate at time t and λ [s^{-1}] is the decay constant of the radioactive source.

  Therefore, the initial population of atoms decays exponentially at a rate that depends on the decay constant. A low activity leads to measurements that last for a long time, but it is less stressful for the acquisition system.

- **Half life:** is the time required by half of the original population of radioactive nuclides to decay or, referring to activity, the time required by the starting activity A_0 to decrease by half. The relationship between the half-life and the decay constant of a radioactive source can be described by the equation:

  $$\lambda = \frac{\ln 2}{\tau_{1/2}} \tag{1.8}$$

  where $\tau_{1/2}$ [s] is the half-life.

- **Energy:** is characteristic of each radioactive source. The emitted γ-rays interact with the patient's tissues before leaving the body; the type of interaction is mainly Compton scattering. This results in a broadening of the energy spectrum before they reach the detector. The energy spectrum constitutes a unique signature of the radioactive element: there is a one-to-one correlation between them.

For what regards SPECT applications, lots of radiopharmaceuticals

SPECT Radionuclide	Half-life	Principal γ emission [keV]
^{99m}Tc	6.01 h	140
^{123}I	13.27 h	159
^{67}Ga	3.26 d	93.3 - 184.6
^{111}In	2.80 d	171.3 - 245.4
^{201}Tl	3.04 d	167.4
^{155}Tb	5.32 d	86.5 - 105.3

Table 1.1: *Main SPECT radioisotopes: half-life and energy of emitted gammas [8].*

are available: indeed, radioactive isotopes emitting single photons can be bound to a high number of biomarkers. A list of the more common radioactive sources employed for SPECT is showed in table 1.1. Radiotracers commonly used in SPECT are characterized by a good trade-off between half-life and energy of the emitted photons. For Technetium, which is the most widely used radio-tracer for SPECT, half life is equal to about 6 hours, thus providing enough time to be able to transport the source and perform examinations but also keeping the patient exposure limited. The decay curve for Technetium is depicted in figure 1.8.

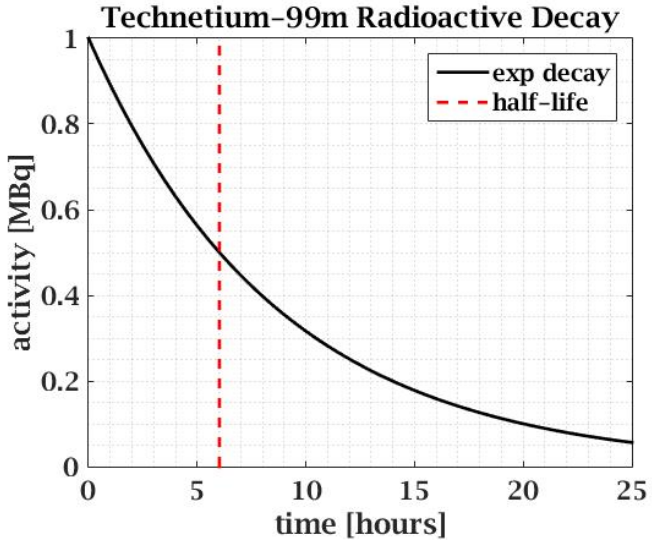


Figure 1.8: *Radioactive decay law for Technetium-99m. At the beginning the activity is equal to* 1 MBq*; after 6 hours (half-life) it decreases to* 0.5 MBq.

Radioisotopes for SPECT, however, may modify the biological properties of biomarkers they are bounded to, since they are atoms not naturally present in biological compounds. If radioisotopes for SPECT are characterized by a relatively long half-time and a low inertia towards biomarkers, the ones employed in PET show opposite characteristics. The main radiotracers for PET are derived by carbon (^{11}C), nitrogen (^{13}N) , and oxygen (^{15}O), and so, being elements commonly present in biological systems, they can be bound to specific molecules without altering them. On the other side, they are characterized by very short half-lives and, thus, require cyclotrons on site to be produced.

Finally, one of SPECT useful properties is the possibility of imaging more than one radionuclide simultaneously [9]. In order to be performed, this so-called dual-isotope imaging requires sufficient energy resolution of the scanner to separately identify the energies of each employed radionuclide. In this way, events produced by different sources can be employed to create separate images. The clinical relevance of this imaging modality consists in the possibility to simultaneously assess more than one functional biological property.

Dual-isotope imaging is not possible in PET, since all positrons emitted by radionuclides produce a couple of 511 keV during annihilation with electrons.

1.2.2 Gamma camera

The key component of a system for gamma imaging is the gamma camera, that is the module aimed to detect single γ-rays and then convert them into electronic signals to be processed, in order to finally estimate the position and the energy of the interaction between each γ photon and the module itself.

The conversion from radiation to electronic signals can be executed following two main strategies:

- **direct conversion:** in a direct conversion device, the γ radiations are absorbed and directly converted into a number of electrons pro-

portional to the energy of the incident beam.

- **indirect conversion**: the problem of the conversion from γ-photons to electric signals is decoupled. The first step, which consists in absorbing the radiation, is performed by inorganic scintillators, dense crystals made by elements at high atomic number, which, for this reason, have a high capability to absorb the radiation. When excited by γ-rays, they undergo a "scintillation", namely they emit optical photons, which isotropically spread inside the crystal. The following step consists, instead, in converting the released optical photons in charge carriers, which can be subsequently processed by a read-out circuitry.

The indirect conversion is characterized by a greater detection efficiency, thanks to the use of scintillation crystals, but the decoupling of the detection process amplifies the error on output signals. The indirect-type conversion is usually preferred over the direct one, because of its flexibility and the possibility to optimize the absorption efficiency using materials at high atomic number.

In general, depending on how the scintillator crystals and the photodetectors are coupled, two different families of gamma-cameras, illustrated in figure 1.9, can be distinguished:

- **pixelated camera:** a pixelated camera is constituted by an array of photodetectors, each one individually coupled with its own scintillation crystal, forming an independent detection unit, which is defined as pixel, and whose dimension defines the spatial resolution of the camera.

- **Anger camera:** this type of configuration is based on a single, continuous, monolithic scintillator crystal, which is coupled to an array of photodetectors, which provide an output current proportional to the number of incident photons [10]. In this case, the spatial resolution is no longer limited to the dimension of the pixel;

indeed, considering the same detection surface, it is possible to obtain the same spatial resolution using a lower number of photodetectors. Furthermore, using less photodetectors implies of course less readout electronics channels to be implemented, and consequently, less costs.

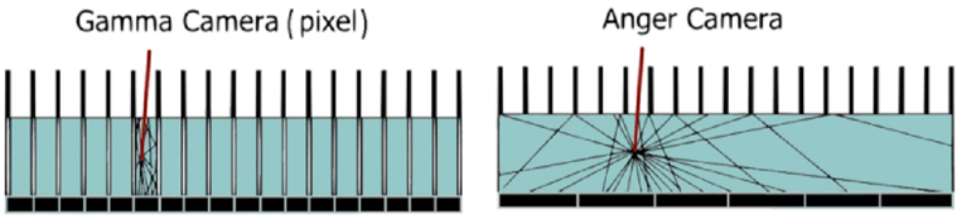


Figure 1.9: *Pixelated and Anger cameras*

While for pixelated detectors each pixel works as an independent detection module and is coupled with its proper scintillation crystal, in the case of Anger cameras, a photon interacting inside the continuous scintillator activates simultaneously different photodetectors; for this reason, Anger cameras require dedicated reconstruction algorithms in order to estimate the exact position of interaction with the crystal, on the basis of the signal detected by each photodetector.

Hereby, the following discussion will focus on continous detection modules rather than pixelated detectors.

The key elements composing a continuous gamma camera are:

- Scintillation crystal
- Reflective coverings
- Optical coupling
- Photodetectors

Scintillation crystal

The scintillator is the component designated to the absorption and conversion of the γ-radiations into a bundle of lower energy photons (typically in the UV or visible range).

Scintillators can be divided in two main categories: inorganic scintillators, that have a high conversion efficiency but are generally slow, and organic ones, that have opposite characteristics. Inorganic scintillators are the most widely used in nuclear medicine; they are solid crystals composed of two or more atoms with high atomic number, thus providing dense material, with a high absorption probability. Lambert-Beer law, already defined in equation 1.6 describes the probability for a certain material to absorb radiations at a fixed energy, with dependence to the depth of penetration.

The main physical parameters to account for, when dealing with scintillation crystals, are:

- **Density:** it gives an indication of the gamma radiation absorption capability of the scintillator. Since γ-rays interact primarily with atomic electrons, the attenuation coefficient is proportional to the electron density, which, in turn, is proportional to the bulk density of the absorbing material. As a consequence, the higher the density of the material, the higher will be its capability to absorb the radiation.

- **Photon Yield:** it defines the average number of photons generated in a scintillation process as function of the impinging gamma ray energy. Therefore, it represents the conversion efficiency of the crystal and its unit of measure is number of photons over keV ($\frac{\text{ph}}{\text{keV}}$). Since scintillation photons represent the useful signal to be read, higher values for this parameter allow better sensitivity and SNR (Signal to Noise Ratio).

- **Refractive index:** it is an optical characteristic of the medium, expressed as a dimensionless number, describing light propagation in a medium. Considering the interface of two objects with different

refractive indexes, it is possible to calculate the reflection probability of an impinging photon knowing its energy and the incidence angle with respect to the interface plane. For what concern scintillators, ideally, one would like the refractive index to be as close as possible to the value of the materials which are optically coupled to the scintillator itself. In such a way, collection of light from the crystal to the photodetector array is not affected by reflection phenomena.

- **Scintillation decay time:** generation of photons in the crystal exhibits an exponential time distribution, which is related to the cascade of processes involved in scintillation.

  Therefore, this characteristic is described by a decay time constant, which ranges from few tens of nanoseconds to tens of microseconds. In some cases, the time distribution is better described by a combination of two or more exponentials, with an equivalent number of time decay constants.

  Decay time affects signal collection by the readout electronics; indeed, in order to correctly acquire all the charge generated by a certain event, an integration of the photodetectors signal is needed.

- **Maximum wavelength:** this parameter resumes a more complex characteristic of scintillators, that is the light emission spectrum. Visible photons generated during the scintillation show wavelengths continuously distributed around a value corresponding to the maximum probability of emission. This value is considered as maximum wavelength or peak wavelength. When developing a gamma-camera, the emission spectrum of the scintillator must be matched with the absorption spectrum of the employed photodetectors, in order to optimize the detection efficiency.

- **Intrinsic Energy Resolution:** it is expressed in terms of percentage and represents the intrinsic scintillator crystal contribution to the energy resolution capability of a radiation detector. Indeed, local variations in the provided scintillation light output can be caused

by material inhomogeneities, non-proportional response to energy or non-uniform reflectivity of the reflecting covering.

Reflective coverings

The amount of scintillation light collected on photodetectors is a crucial parameter for both imaging and energy resolution performances, since it represents the signal of interest. Light produced in a scintillation event spreads in all directions inside the crystal from the gamma interaction site: the dispersion of these photons bundle is isotropic and only a percentage of them, roughly estimated by a solid angle calculation, directly hits the photodetection array. The optimization of the gamma detector requires to enhance the light collection also for those photons whose initial trajectory is not directed to the photodetection plane. For this reason, reflective materials are employed, with the objective to redirect photons towards the sensitive array. However, it has to be noticed that, although the presence of reflective coverings on the lateral walls is important for improving the energy resolution of the detector, it introduces some distorsions in the reconstruction of events interacting along the borders of the scintillator.

The most important parameters for reflectors are:

- **Refractive index:** it is fundamental to calculate the amount of reflected light at an interface between two media as function of the incident angle.

- **Reflectivity:** it is the average percentage of light reflected by a material given a light bundle with uniform angular impact direction. For light collection applications in scintillators, a high value of reflectivity is required from the covering materials to increase collection efficiency of photons.

Optical coupling

Differently from the lateral sides, where reflection is enhanced to improve energy resolution, reflection phenomenon must be minimized at the sur-

face from which the scintillator is coupled to photodetectors to prevent internal trapping of the light. For this reason, the crystal is coupled to the photodetectors array by means of materials with particular optical properties, generally a transparent grease or glue (resin).

Photodetectors

An indirect conversion detector has to convert optical photons into electric signals. The principles which can allow this conversion are different, depending on the type of detector, but, in general, each photodetector can be evaluated by some key parameters:

- **Photodetection efficiency:** represents the conversion efficiency, namely the percentage of incident photons which are actually converted into charge carriers.

  This characteristic depends on several factors:

  - the physical conversion process laying for the specific photodetector
  - geometrical features of the photodetector, such as the Fill Factor (FF), namely the ratio between the active sensitive area of a photodetector and its total area
  - the wavelength of the impinging photon
  - the Quantum Efficiency (QE) of the composing material, defined as the ratio between the number of photoelectrons generated and the number of incident photons (which ideally should be 100 %)

- **DCR (Dark Count Rate):** it expresses, in Hz/mm^2, the number of electron-hole pairs spontaneously generated in photodetectors because of temperature.

  This phenomenon introduces a random noise contribution that adds to useful output signals.

- **Gain:** a photodetector has to possess a certain internal gain, in order to create appreciable output electrical signals even in presence of few incoming optical photons.

- **Response time:** the conversion of photons into carriers and the following collection of charges at the electrodes present a time development that depends on the physics and on the geometry of the detector. The ideal detector would have a short pulse duration, close to a Dirac's delta impulse response.

- **Temperature stability:** it is necessary to ensure a proper functioning of the detector in different environmental conditions.

- **Bias voltages:** lower values of biasing volage are desirable, meaning less power dissipation and heat.

1.2.3 Figures of merit

In order to quantify the detection and imaging performances of the global system, several meaningful parameters are usually employed:

- **Spatial resolution:** it is defined as the minimum distance required between two distinct sources of gamma radiation to visualize them as distinct objects on the output image.

  Spatial resolution is associated to the concept of Point Spread Function (PSF), representing the spatial response of the gamma imaging system to a point source. PSF is usually described in the image plane with a Gaussian distribution; hence, spatial resolution is frequently measured as the full width at half maximum (FWHM) of the PSF characterizing an imaging device.

- **Sensitivity:** it is defined as the ratio between the number of events acquired by the instrument in the unit of time and the activity of the source (expressed in Bq). The sensitivity is closely related to the Count Rate (CR) , which is the maximum temporal frequency

of radioactive decays that the instrument is able to acquire. Pulses pileup, namely the superposition of light flashes produced by distinguished events, represents the main bottleneck for the count rate, preventing the gamma camera to have satisfactory performances when the activity of the radioactive source is increased.

- **Energy resolution:** it is defined as the minimum energy gap existing between two gamma-sources which can be distinguished by the detection system. Even if a population of detected monoenergetic γ photons is considered, the resulting energy distribution measured by a gamma camera is well represented by a Gaussian fuction. The reason for this is attributable to the intrinsic stochastic processes regarding scintillation and electric signal production.

- **Fields of view (FOV):** it is defined as the extent of the active detection surface of the system able to detect gamma events.

  Usually, two sub-sections of this area are defined:

  - Useful FOV (UFOV), which is the portion where events are reconstructed without showing particular non-linear effects.
  - Central FOV (CFOV), which corresponds to the more central area of the detector usually providing better spatial resolution.

1.3 Multimodal imaging and INSERT

1.3.1 Multimodality

Today, medical imaging technologies are expected to provide an adequate grade of correlated information between anatomical structure and functional processes. Indeed, anatomical imaging does not always provide the complete clinical framework required for diagnosing and staging diseases or monitoring response to therapies, since functional or metabolic changes can occur independently of anatomical ones. This new paradigm regarding

combination of functional and anatomical information in medical images found its actualization in *multimodal techniques*.

This expression is employed referring to the development of technological systems or computational approaches allowing to merge images from different medical modalities by performing their co-registration in space and time.

Many attempts of developing multimodal imaging systems have taken place during the years.

Initially, multimodality was intended just like retrospective software registration of volumetric datasets of patient's body districts [11]. Of course, this approach presents many unconvenients to face; changes in patients positioning passing from an imaging devices to the other, or simply organ motions during the acquisitions make the precise overlapping of the two recordings particularly challenging.

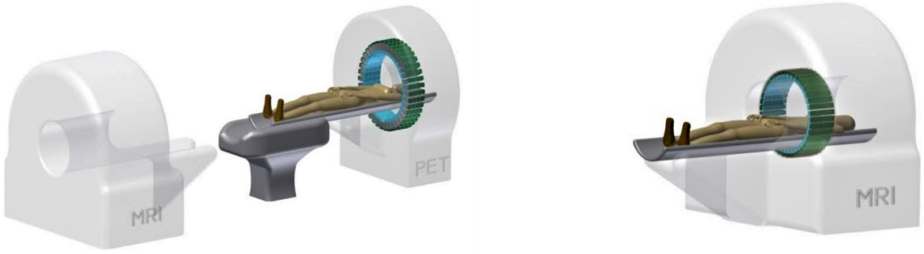
In order to cope with these problems, the attention of the research moved to the development of systems, which integrate on a hardware basis the two types of imaging modalities.

In the early 2000s, the first commercial implementations of devices allowing multimodality on a hardware basis were introduced. These were SPECT/CT and PET/CT scanners, in which nuclear medicine provided functional information by using radioactive tracers and gamma-photon detectors whereas CT was employed to image anatomical structures.

The introduction of combination of modalities through dedicated hardware instruments proved to possess the ability to address scientific or clinical questions that would be impossible on separate systems.

Different approaches can be used in order to design the architecture of a multimodal scanner [12]:

- **separate system approach**: the two imaging scanners are placed in two separate adjacent rooms, the patient is positioned on a imaging table which move from one scanner to the other, through two separate gantries. In this way, the time interval between the two acquisitions is reduced, and, being the two imaging systems inde-



(a) Sequential approach (b) Integrated approach

Figure 1.10: *Scanner architectures of two types of multimodal systems.*

pendent, their imaging performances are preserved and no redesign of the system is needed.

- **sequential approach**: differently from separate-systems approach, the two imaging systems are placed one next to the other, further minimizing the possibility of patient movements from one recording to the other. In these type of systems, a fixed coordinate transformation between the two imaging devices is applied, in order to allow the spatial-overlapping of the two images.

- **integrated approach**: even if it is undoubtedly the more challenging, integrated approach is also the more promising and precise multimodal technique, since it allows a direct co-registration, thus preventing errors due to the non simultaneous recording of the images. The main problem is the need to re-adapt the design of the two imaging systems, in order to prevent mutual interferences and to cope with geometrical and encumbrance constraints.

1.3.2 INSERT project

The present thesis project has been aimed at the exploration of new machine learning-based image reconstruction methods, whose validity may be extended, in principle, to any type of gamma camera.

The simulated and experimental validation of these techniques has

been performed on a multimodal imaging system, which has been developed by a previous research project, called INSERT [13].

INSERT project has been funded by the Seventh Framework Program of the European Commission and started on March 1st, 2013 (INSERT - development of an integrated SPECT/MRI system, s.d.). The acronym refers to the project aim of developing compact SPECT insert to be fit inside commercial MR scanners for enhanced stratification of brain tumor and early assessment of treatment efficacy.

In particular, the project wanted to address the need of a more powerful tool for the diagnosis and therapy of gliomas, a common type of brain and spine tumor, occurring in the 33 % of the overall central nervous system tumors and representing the 80 % of all malignant brain tumors. Indeed, one of the techniques producing outstanding results in terms of brain cancer treatment is radiotherapy, which requires a precise identification and localization of the tumor and its environment. Therefore, the integration of a new SPECT system in an existent MRI scanner, allowing the two imaging modalities to operate simultaneous acquisitions, gives the possibility to obtain multiple parameters to better define not only the tumor position inside the patient-specific anatomy, but also its biological characteristics.

1.3.3 INSERT clinical scanner architecture

In the framework of INSERT project, two different imaging system have been designed and implemented: a clinical scanner, for human neck imaging, and a pre-clinical, which has smaller dimension and it is finalized to small animals imaging.

The experimental measurements which will be introduced in the following chapters have been collected by using one of the detection modules of the Clinical system, which will be now briefly described in order to provide a better understanding of the functioning of the whole detection system.

Clinical INSERT SPECT/MRI system, depicted in figure 1.11, has

Figure 1.11: *Clinical INSERT SPECT Scanner*

been designed to obtain 3D images of the distribution of radiotracers in the whole human brain.

The system consists in three main components:

- **SPECT scanner:** composed by 20 detection modules (namely 20 Anger cameras each one featuring 72 channels), the collimator and ancillary systems (laptop, cooling unit, electronic boards for communication and power supply).
- **RF coil:** a RF coil designed in order to be inserted in the rear part of the SPECT ring. The transmitter coil generates the oscillating magnetic field necessary to bring the nuclear spins to the excited state. The receiver coil measures the radiofrequency (RF) electromagnetic radiation produced by the spontaneous relaxation of the nuclear spins orientation inside the subject.
- **Commercial MRI:** a standard magnetic resonance scanner.

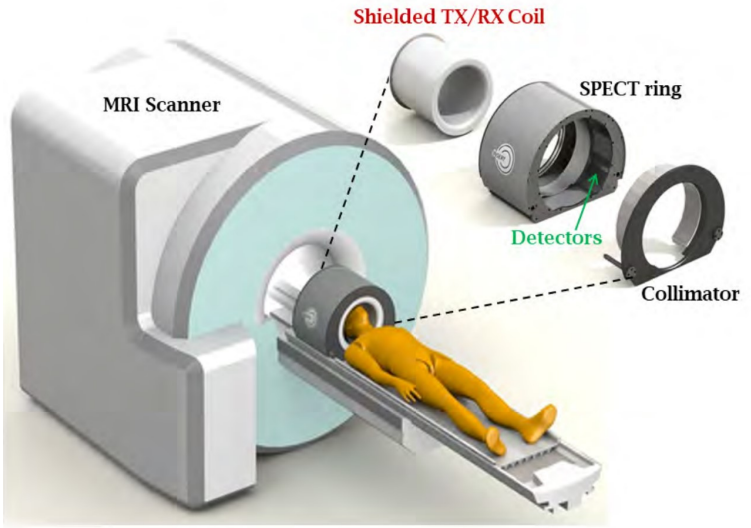


Figure 1.12: *INSERT system*

The three components are illustrated in figure 1.12

1.3.4 SPECT scanner

In order to form a 3D tomographic volume, providing functional information about the patient, it is necessary to build 2D projections and then to combine them together by dedicated algorithms.

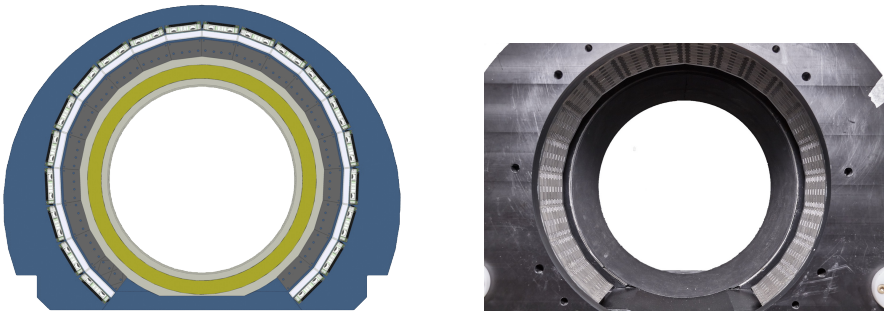
In general, there are the main possible approaches to obtain multiple planes acquisitions:

- **Rotational approach:** a set of gamma cameras (typically 4) are located around the region to be imaged and mechanically connected to a rotational engine that changes their angular position in order to acquire subsequent projection planes along a circular trajectory.

- **Stationary approach:** several gamma cameras are orientated around the longitudinal axis of the scanner in a ring configuration. The number of projection planes covered corresponds to the number of gamma cameras employed. The final geometry of the system can be a closed or open ring. The planar images are recorded within

the same time interval, opening the possibility for dynamic tracking of radionuclides and reducing the total examination time at the expense of a higher cost.

In order to be MR-compatible, INSERT clinical system adopts a stationary approach with an **open ring geometry**, as depicted in figure 1.13.



(a) Clinical INSERT geometric schematic

(b) Clinical INSERT ring, real picture

Figure 1.13: *INSERT SPECT scanner adopts a partial (or open) ring geometry: the bottom part of the ring is not covered by sensitive elements. (a) Conceptual draw showing the 20 detector modules disposition inside the instrument. (b) Real picture of the scanner, only the collimator is visible from the outside, while the 20 gamma cameras are hidden behind it.*

Collimator

Collimators are made of materials with high density and high attenuation coefficients; INSERT clinical scanner uses a Multi-mini Slit-Slat collimator (MSS) made of Tungsten. This type of collimator, illustrated in figure 1.14, is made of slits orientated along the axial direction and thus influencing the transaxial resolution of the detector, and slats, orientated in the opposite direction in a way to influence the axial resolution [14].

Scintillation crystal

Clinical INSERT single module has a **CsI(Tl) (Thallium-doped Cesium Iodide)** scintillation crystal. The crystal, depicted in figure 1.15

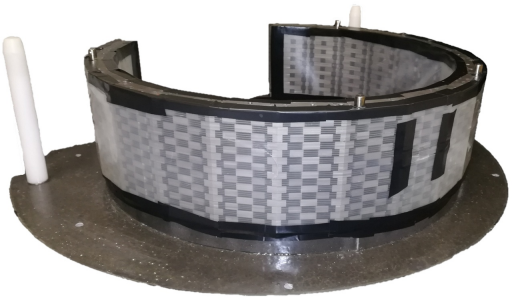
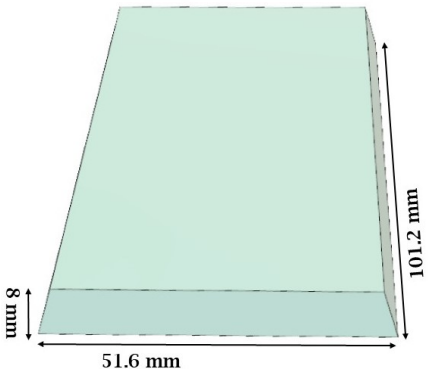
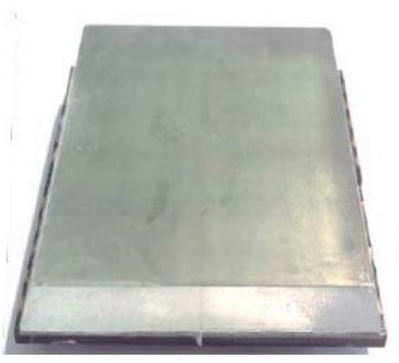


Figure 1.14: *MSS collimator*

has base dimensions approximately equal to 100 mm x 50 mm, while its thickness is 8 mm. Thickness choice derives from a compromise between absorption efficiency, directly proportional to thickness, and intrinsic spatial resolution, which instead improves for thinner scintillators. The reason for that is related to the distribution of scintillation light on the photodetection plane. As described by Lambert-Beer law (eq. 1.7), scintillation light is likely generated in the first millimeters of the material, then it is seen by the photodetectors through a solid angle.



(a) Clinical crystal, geometry



(b) Clinical CsI crystal, photo

Figure 1.15: *Clinical INSERT crystal. Two edges of the scintillator are slanted in order to fix the detector modules in the final SPECT architecture in a ring shape.*

How it is possible to observe from figure 1.16, the distribution of light

on the detection plane is sharper when the crystal is thin, while the spread increases for thick scintillators.

CsI(Tl) has high conversion efficiency (yield), which is important for low gamma energy detection (from 100 keV to 200 keV), but the time distribution of the emitted light photons is described by a slow bi-exponential function. The crystal is slightly hygroscopic, therefore it has to work in a dry environment but it does not require encapsulation (unlike NaI), making it easier to obtain a compact SPECT scanner.

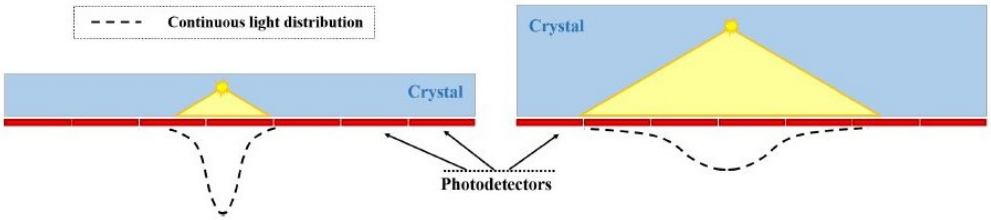


Figure 1.16: *Light distribution for a generic gamma event in a thin crystal (left) and in a thicker one (right).*

Parameter	Value for CsI
Density [g/cm^3]	4.52
Light Yield [photons/keV]	65
Refractive Index	1.79
Decay Time [µs] (room temperature)	0.68 (64 %) 3.34 (36 %)
λ **of max. emission** [nm]	540

Table 1.2: *CsI parameters regarding interaction with gamma rays and optical behaviour [3].*

The described properties of the scintillation crystal are listed in table 1.2.

Reflective Covering

In order to preserve as much as possible the amount of scintillation light reaching the photodetectors, INSERT CsI(Tl) crystal is covered with a

4-layers Teflon reflective covering, with the exception of the lower photosensitive plane. The choice of Teflon as reflective covering material derived from its high reflectivity (94 %) and its diffusive reflection properties.

Optical Coupling

In INSERT detection module the coupling between the scintillator and the array of photodetectors is provided by Meltmount, a transparent mounting media similar to a resin. When heated up, the viscosity of the material decreases and the photodetectors can be glued to the crystal surface. The target refractive index for the coupling material was chosen as an intermediate value between the refractive index of the scintillator crystal ($n_{crystal} = 1.79$) and the one of the optical protection resin covering the photodetectors ($n_{resin} = 1.51$). However, Meltmount can have different refractive indexes corresponding to different mechanical properties; it was choosen $n_{Meltmount} = 1.539$ because of the appropriateness of the corresponding mechanical properties [15].

Photodetectors

Photodetectors used in INSERT scanner are Silicon photomultipliers (SiPMs), which, how it will be described in the next chapter, present some important advantages in terms of low-voltage operation (with respect to PMTs), insensitivity to magnetic fields, mechanical robustness and compactness. A further description of the characteristics of SiPMs detector used in INSERT will be provided in the following chapter.

Chapter 2

INSERT photodetectors and signals readout

Within this chapter, the photodetector used in INSERT detection module, the Silicon Photomultiplier or SiPM, will be described in terms of structure, working principle and figures of merit. In the second part of the chapter, the SiPMs signals readout strategy implemented in INSERT clinical module will be addressed.

2.1 Photodetectors for gamma detection

Recent research and technological advancement has allowed to offer a wide variety of photodetectors for gamma radiations, each characterized by a different structure and working principle.

One of the most common is the **PIN photodiode**; it consists of a reverse p-n junction, where the p^+ and n^+ doped regions are separated by an intrinsic region, as shown in figure 2.1. This type of p-n junction is reversely biased with a biasing voltage in the order of 70 V.

The principle exploited by a PIN photodiode is the following: given the wide gap between the p^+ and n^+ regions, the optical photons are supposed to be absorbed mostly into the intrinsic depleted region, which is subjected to a strong electrical field. Consequently, photoelectrons gen-

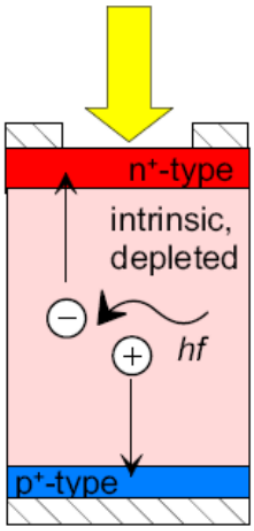


Figure 2.1: *PIN photodiode*

erated by absorption will migrate towards the n-region (the holes will do the opposite).

PIN photodiodes allow to obtain high resolution images, being possible to implement highly dense detector pixellation. On the other hand, substancial limits of the PIN photodiode are the absence of a multiplication stage, which introduces the need of designing a proper electronic amplifying stage, and the high dark current, due to thermal generation of carriers in the depletion layer.

Another family of detectors commonly employed is the one of **Silicon Drift Detectors** (SDDs). In SDDs, the depletion of the silicon bulk is achieved exploiting the sideward depletion principle; in addition, an electric field parallel to the surface of the wafer is superimposed in order to let the electrons to drift towards a small collecting anode (figure 2.2).

This additional electric field is obtained by means of segmentation of one or both of the p^+ electrodes at the wafer's surface and by suitably biasing these electrodes with a voltage gradient[16]. Like a PIN diode, an SDD does not provide internal multiplication, but it is charachterized by a very low anode capacitance, which is independent on the active area

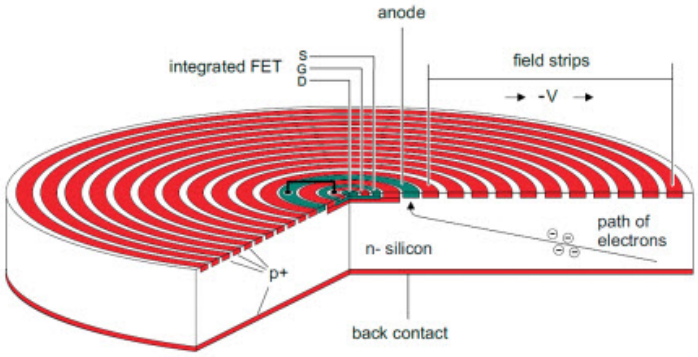


Figure 2.2: *SDD configuration*

size; thus, dark current is reduced respect to PIN [17].

Avalanche Photodiodes (APDs) are, instead, a particular type of photodiodes, where an high intensity electric field region (in the range of $5 \cdot 10^5 \ \frac{\mathrm{V}}{\mathrm{cm}}$ is added between the p and n regions (figure 2.3). This region is obtained reverse-biasing the junction at a voltage slightly below the breakdown voltage, that is the value beyond which the current in the diode increases exponentially.

This strong electrical field represents an internal gain for the signal. Indeed, when a photon is absorbed inside the depletion region, an electron-hole pair is created; the hole is collected at the anode, while the electron is accelerated by the electric field acquiring enough energy to create another electron-hole pair by ionization. Then, the primary and secondary electrons are accelerated again and generate other e-h pairs. Also holes can produce pairs by impact, generating a positive feedback effect.

Thanks to the fact that the device is biased just below breakdown voltage, the output current signal is proportional to the number of incident photons through a multiplication factor, which is function of the applied voltage, that ranges from 50 to 300 V. Quantum Efficiency of APDs is generally higher than 80% in the visible region. The main drawbacks are the high ENF (Excess Noise Factor), due to the cascade of ionizing collisions, their low stability to temperature, and their low gain, which make them unusable for single photons detection applications. For this

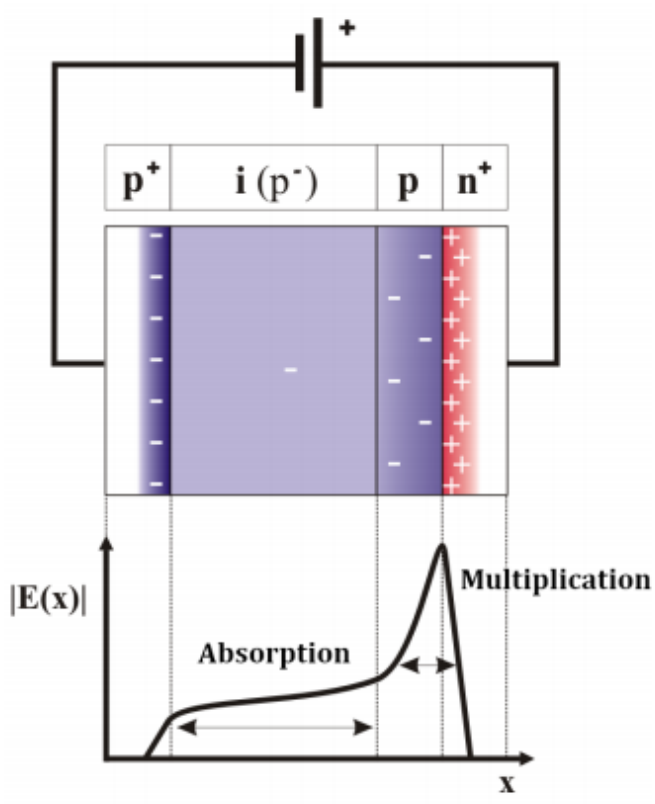


Figure 2.3: *APDs configuration*

last reason, they are chosen much more in spectroscopy than imaging.

Another type of detector commonly used for gamma imaging is the **Photomultiplier Tube** (PMT). PMTs are typically made of an evacuated glass housing which contains a photocathode, several electrodes and an anode as depicted in fig 2.4. The incoming optical photons are absorbed by the photocatode, which is covered with a thin photosensitive layer, and converted in low energy electrons by photoelectric effect. These so called photoelectrons are directed by the focusing electrode toward the multiplying stage which consists of a number of electrodes, called dynodes, each one held at more positive potential than the preceding one.

Each of the dynodes exploits the energy of the incoming electron (primary electron) to generate several secondary electrons. The gain M (typ-

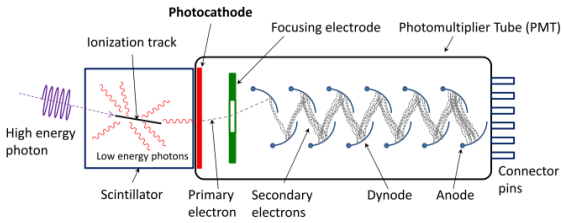


Figure 2.4: *PMT configuration*

ically in the order of $10^5 \div 10^8$) of a PMT with N dynodes is the total number of electrons produced by a single incident photon which can be expressed as:

$$M = \delta^N \tag{2.1}$$

where δ is the multiplication factor of each dynode.

The main strength of PMTs is the presence of a significative multiplication factor, which relaxes the constraints on the electronic noise. Furthermore, the time response of SDDs is very fast (shorther than 1ns). On the other hand, PMTs are very bulky, they require high biasing voltage (generally between 300 V and 2000V, depending on the number of dynodes), they usually have a low quantum efficiency ($< 30\%$) and they are sensitive to the presence of magnetic fields.

2.2 Silicon Photo-multiplier

The Silicon Photomultipliers (SiPMs), firstly developed in 1997 [18], attracted more and more the attention of researchers, in particular for nuclear and medical applications, because of their benefits compared to PMTs, mainly in terms of magnetic compatibility and reduced biasing voltages.

2.2.1 SiPM structure

A SiPM is constituted by a parallel array of photon counting microcells. Each microcell consists of a Geiger-Mode Avalanche Photodiode

(GMAPD) (also called Single Photon Avalanche Photodiode (SPAD)), with an integrated quenching element, as represented in figure 2.5.

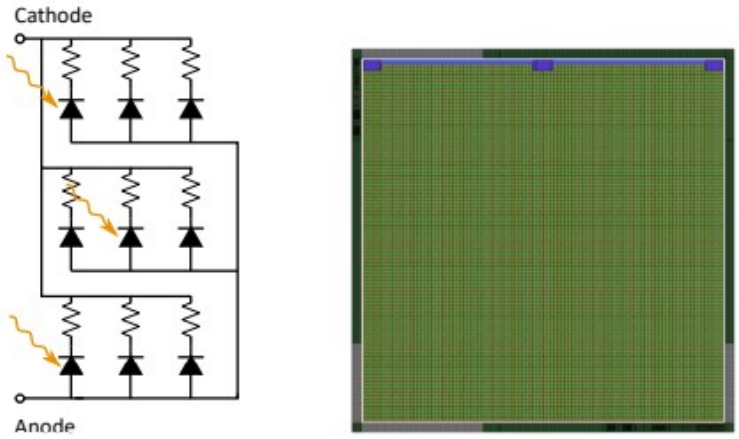
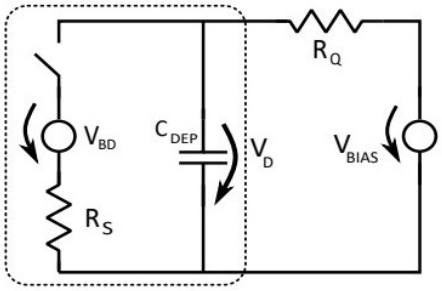


Figure 2.5: *SiPM configuration*

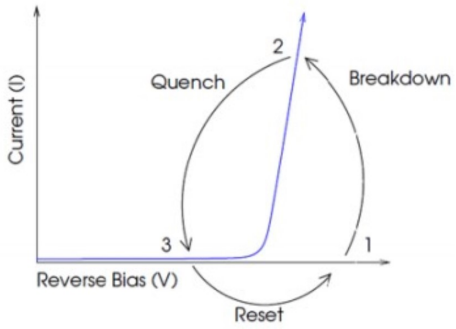
As already explained previously, APDs are biased below the breakdown voltage and thus they feature a much smaller gain than SPADs or PMTs, since the devices are operated to have a linear amplification generated by the multiplication of free carriers. The multiplication process is related to subsequent ionizing collisions, originated by both electrons and holes. This cause a high Excess Noise Factor (ENF) and, together with the small amplification provided, prevents the device from being able to detect single photons. However, the current signal provided from an APD is proportional to the number of triggered avalanches, therefore the output current is proportional to the number of absorbed photons.
On the contrary, SPADs are devices which operate above the breakdown voltage: an electron-hole pair can be generated in the depleted region through the absorption of a photon, and these free carriers, accelerated by a strong electric field, originate a self-sustaining current by means of impact ionization. In order to stop the avalanche, the approach used in a SiPM consists in implementing, in series to the avalanche diode, a passive quenching resistor, that produces a voltage drop proportional to the avalanche current and thus reduces the voltage across the junction below the breakdown voltage, quenching the avalanche. As no more current

flows, the diode depletion capacitance is recharged to the bias voltage, and a new avalanche can be triggered.

This behaviour can be modeled at first order with a simple equivalent circuit, depicted in figure 2.6, along with its operation cycle; the main parameters are the voltage across the junction V_D, the photodetector series resistance R_S, the photodetector depletion capacitance C_{DEP}, in the order of tens of fF, the breakdown voltage V_{BD}, the quenching resistance R_Q, typically in the order of hundreds of kΩ to some MΩ [19][20], and the bias voltage V_{BIAS}, higher than V_{BD}, that generates the high electric field that sustains the avalanche.



(a) SiPM microcell equivalent model



(b) SiPM microcell operating cycle view

Figure 2.6: *Equivalent model of SiPM single cell (top) and operating cycle (bottom).*

With no incoming photons, the switch is open, the voltage across the

depletion capacitance C_{DEP} is equal to V_{BIAS} and no current flows.

When the microcell absorbs a photon, an electron-hole pair forms; one of the charge carriers drifts to the avalanche region, where it can initiate an avalanche. As soon as the avalanche is triggered, the switch closes causing C_{DEP} to be discharging through R_S (being $R_Q \gg R_S$), sustained by a current I_{DET} .

In particular, from the moment of triggering of the avalanche, the current I_{DET} rises exponentially with a time constant τ_{rise} given by:

$$\tau_{rise} = C_{DEP} \cdot (R_S \| R_Q) \approx C_{DEP} \cdot R_S \tag{2.2}$$

The peak value of the current I_{DET} is given by:

$$I_{DET,peak} \approx \frac{V_{BIAS} - V_{BD}}{R_S + R_Q} \tag{2.3}$$

The quantity $V_{BIAS} - V_{BD}$ is commonly referred to as *excess voltage* V_{EX}, or *overvoltage*.

The junction voltage V_D, which before the avalanche was almost equal to V_{BD} (being $R_Q \gg R_S$), after the avalanche drops, following the same time constant τ_{rise} defined in equation 2.2. As the electric field across the junction decreases, due to the voltage drop on R_Q that reduces V_D, the avalanche extinguishes and no more current flows. This behaviour is modelled by opening the switch, and the voltage V_D is reset to V_{BIAS} with the slower time constant:

$$\tau_{reset} = C_{DEP} \cdot R_Q \tag{2.4}$$

Only when V_D reaches V_{BIAS} the microcell is ready to trigger another avalanche. The current pulse supplied by the photodetector, according to the model, is represented in figure 2.7.

Each microcell features a bi-stable behaviour, as the avalanche itself is not proportional to the number of absorbed photons, since the cell is no more sensitive till the bias voltage has been restored on the photodetector depletion capacitance, and no further events will generate signals. The

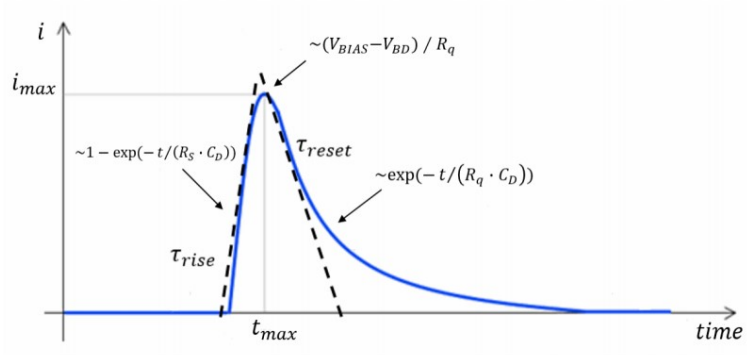


Figure 2.7: *Graph of the current flowing through the terminals of the SiPM from the start of the triggering to the recovery of the bias voltage. C_D in the figure is equal to C_{DEP}.*

principle exploited by SiPMs is triggering an avalanche breakdown when an incident photon generates an Electron-Hole Pair (EHP) inside the depleted region; pulses which are triggered by non-photo-generated carriers constitute noise sources.

Common noise sources are thermal generation of EHPs or generation of electrons by tunnelling effect; these effects are undesired and referred to as *dark counts*.

The typical response of a silicon photomultiplier is shown in figure 2.8, together with the distribution of the acquired signal amplitudes. The signal is quantized, and the contributions of single photons are clearly visible.

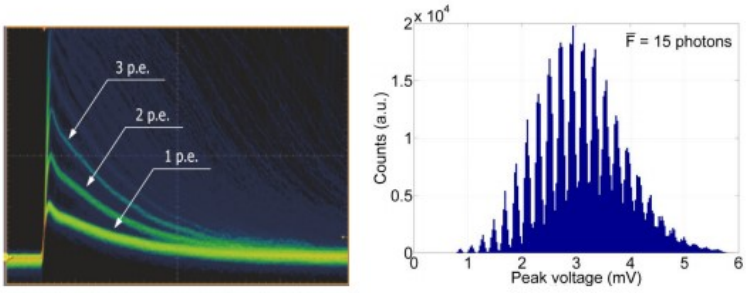


Figure 2.8: *Typical SiPM response (on the left) and spectrum (on the right)*

2.2.2 Figures of merit of a SiPM

Photodetection efficiency

The Photon Detection Efficiency (PDE) is defined as the probability for an impinging photon to develop an electrical signal. It is intuitive to express this quantity as the joint probability for a photon to be impinging on the active area of the photodetector, to be absorbed in silicon, and to effectively trigger the avalanche, leading to the following expression:

$$PDE = QE \cdot FF \cdot P_{trigger}(V) \tag{2.5}$$

where $QE(\lambda)$ represents the Quantum Efficiency, FF the Fill Factor and $P_{trigger}(V)$ the avalanche triggering probability.

The *Fill Factor* is a geometrical parameter, which depends on the layout of the microcell, defined as the ratio between the active area and the overall area of a microcell. Indeed, not the entire area of the detector is able to be activated: the dead area is mainly attributable to the quenching resistor and guard rings. However, increasing the size of the cell, for the same quenching resistor the fill factor improves, as the relative weight of the dead area reduces, but a larger active area lead to a larger depletion capacitance, thus to a slower response of the device.

Moreover, for SiPMs of the same overall size, larger cells lead to a lower total number of cells, thus decreasing the dynamic range of the photodetector.

The *Quantum Efficiency* is the probability for a photon impinging on the active area to be absorbed generating an electron-hole pair in the medium. This quantity is related to the promotion of valence electrons in conduction band and therefore is a function of the wavelength of the impinging photon and the absorption material (Silicon in this case).

In order to absorb the incident light, the absorbing material has to be sufficiently thick, according to the Beer-Lambert equation. Quantum efficiency in SiPMs can be maximized, until reaching values as high as 98%, by depositing proper anti-reflecting coating layers over the SiPM

active area.

The *Triggering probability* depends on how likely is for a carrier to trigger an avalanche and it can be expressed as [21] :

$$P_{trigger} = P_e + P_h - P_e \cdot P_h \tag{2.6}$$

where P_e and P_h are the probability to trigger an avalanche for an electron and a hole respectively.

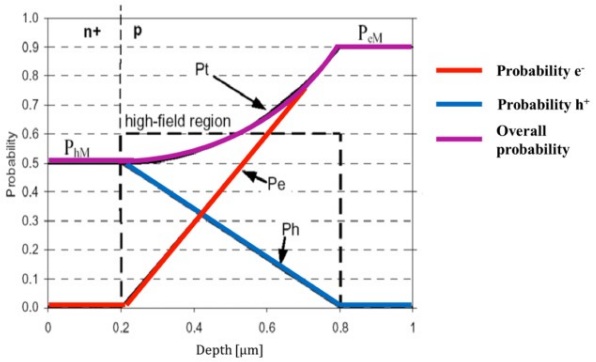


Figure 2.9: *Avalanche triggering probability as a function of the position of the free carrier generation: if the generation of e-h pairs occurs deeply in the high field region, only electrons can generate an avalanche, if instead the generation occurs near the n + region, the avalanche is triggered only by holes.*

Being related to the electric field in the depletion region, this quantity is a function of the SiPM biasing voltage. In silicon $P_e > P_h$, but the respective probabilities also depend on the position of generation of the electron-hole pair, as represented in figure 2.9.

For instance, if an electron-hole pair is generated close to the n^+ region, the electron will be accelerated for a very small time, and reaching a relatively small speed, it will have a small probability to trigger an avalanche, while the opposite holds for an hole. The electron avalanche triggering probability is maximized if the electron is generated at the limit of the depletion region, so that the electric field will accelerate the electron for the longest possible time.

Dynamic Range and Linearity

As previously explained, SiPMs are constituted by an array of elementary microcells.

The maximum signal that can be provided by a SiPM photodetector corresponds to the case in which every cell has been triggered, while the minimum signal consists in the output signal of a single cell. However the detection of photons is a statistical process, based on the probability of detecting a certain number of photons by a limited number of sensitive microcells. For this reason, the output signal of a SiPM is influenced by the statistical fluctuations that two or more photons hit the same cell.

The average number of firing microcells N_{fired}, as a function of the number of impinging photons N_{ph}, given a certain number of cells N_{cell} and PDE, can be computed as [22] :

$$N_{fired} = N_{cell} \cdot \left(1 - e^{-\frac{PDE \cdot N_{ph}}{N_{cell}}}\right) \tag{2.7}$$

According to this formula, it is evident that the output signal of a photodetector is proportional to the number of impinging photons only as far as $N_{ph} << N_{cell}$.

Saturation effects are explained from the fact that for a large number of photons, comparable to the number of the microcells, the probability of multiple photons hitting the same cell becomes significant.

In figure 2.10 are shown experimental results obtained with three different SiPMs with 576, 1024 and 4096 cells respectively, that shows the number of fired microcells as a function of the generated photo-electrons (obtained by knowing the emitted optical power, thus the average number of impinging photons and the PDE of the SiPMs) [22].

Gain

The gain of a SiPM sensor is defined as the number of carriers involved in the avalanche current for a single microcell. SiPMs generate a highly uniform number of carriers each time an avalanche event happens.

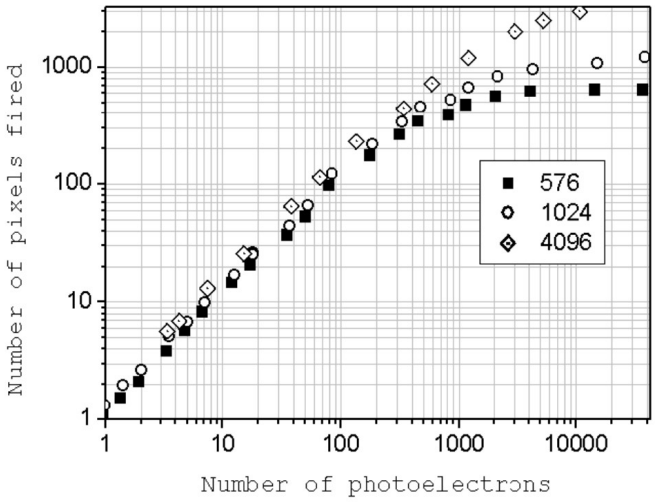


Figure 2.10: *The SiPM output signal saturation for three different numbers of microcells (576, 1024, 4096): the number of fired microcells is proportional to the number of incident photons if this is well below the number of cells, otherwise the output signal shows saturation. Figure taken from [22].*

The average gain can be computed as the charge delivered from the photodetector with respect to the elementary electron charge, according to [23]:

$$G = \frac{C_{DEP} \cdot (V_{BIAS} - V_{BD})}{q} = \frac{C_{DEP} \cdot V_{EX}}{q} \tag{2.8}$$

where $C_{DEP} \cdot V_{EX}$ represents the amount of charge needed to discharge the single cell to the voltage V_{BIAS}, namely the voltage corresponding to absence of triggering.

The gain of a SiPM can be enhanced by increasing the overvoltage or using large microcell photodetectors, which feature larger depletion capacitance. Typical available gain values for commercial SiPMs are larger than $1 \cdot 10^6$ [23][24], even larger than that of some high quality PMT.

Usually gain is in trade-off with response speed and noise performances, as increasing the photodetector depletion capacitance will cause a slower response, and increasing the overvoltage leads to a higher Dark Count Rate, as will be discussed in 2.2.2. The dependence of the gain on cell size (capacitance) at different overvoltages is shown in figure 2.11 [25].

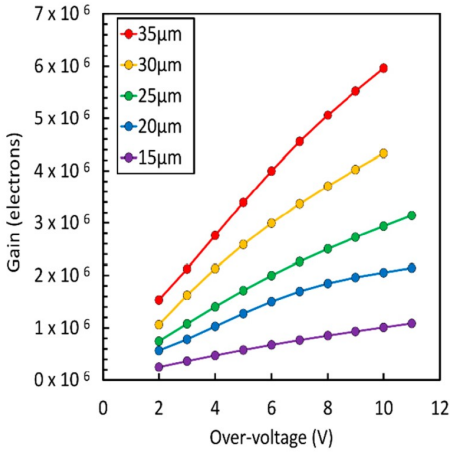


Figure 2.11: *Dependence of NUV-HD SiPM gain on cell size at different overvoltages*

SiPM gain strongly depends on temperature through the Breakdown Voltage (V_{BD}) [26]: if temperature increases, also V_{BD} increases, leading to a reduction of SiPM gain if V_{BIAS} is kept constant. It is therefore of utmost importance to operate SiPM at stable temperature or compensate for temperature variations by changing the biasing voltage.

Typical values of temperature coefficient ($\partial V/\partial T$) of commercially available SiPMs are in the order of 20 mV/°C to 30 mV/°C [23][24]. In figure 2.12 are reported experimental data from FBK NUV SiPMs [27] and for Hamamatsu S13360-3050CS SiPMs [19].

Dark Count

In parallel with photon absorption, also thermal agitation and tunnelling effect can generate free carriers by promoting them from valence to conduction band, which, accelerated by the electric field, can eventually trigger the avalanche [28]. All the processes cause the same output signals and hence they are not distinguishable. Thermal promotion and field-assisted charge generation by tunneling are distributed in time following the Poisson statistic, and are independent on the irradiation condition of the photodetector.

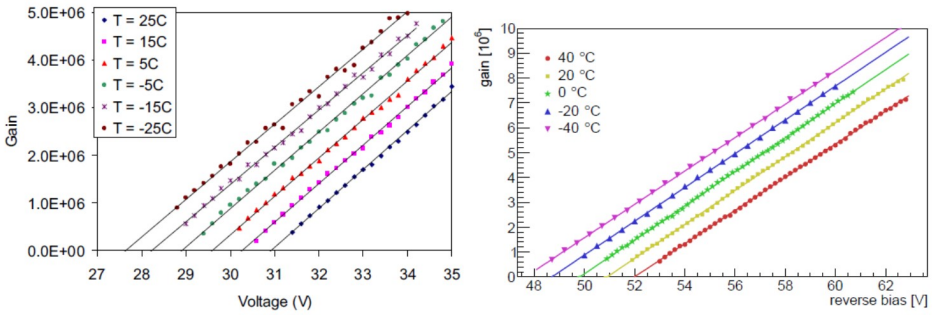


Figure 2.12: *Dependence of gain on temperature and overvoltage in FBK 50 µm×50 µm NUV SiPMs (on the left) and Hamamatsu S13360-3050CS SiPMs (on the right)*

Usually, datasheets refer to these effects as Dark Count Rate (DCR), which is the mean frequency at which avalanches are randomly triggered without light sources, and it is quoted in kHz/mm^2. The two effects depend respectively on temperature and on the overvoltage applied to the photodetector, as it can be seen from figure 2.13.

Correlated Noise

Along with dark count, optical crosstalk and afterpulses are phenomena contributing to the overall noise of SiPMs. They are referred to as correlated noise, since they happen with a certain probability when a photon is absorbed in silicon, and so they are correlated with the signal. In particular, optical crosstalk is caused by the finite probability of photon emission during the avalanche breakdown, and can take place through two different mechanisms: Direct Crosstalk (DiCT) and Delayed Crosstalk (DeCT).

The former happens when the generated photon moves directly towards the depletion region of an adjacent microcell and triggers new avalanches, almost simultaneously to the signal generated one, thus a single photon can lead to an erroneous output signal, equivalent to a number of photons greater than one.

Delayed crosstalk happens, instead, when the generated photon creates an electron-hole pair outside the depletion region (the substrate for

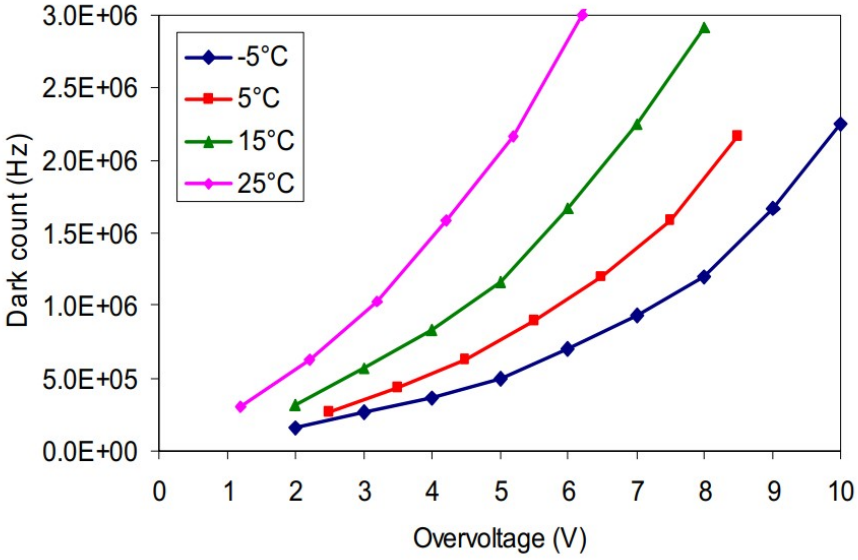


Figure 2.13: *Dark Count Rate as function of the overvoltage at different temperatures in a FBK SiPM.*

example) of an adjacent microcell. This electron-hole pair can trigger an avalanche only if it reaches the active region of the cell by diffusion; therefore, delayed crosstalk leads to the appearance of a succession delayed pulses, randomly distributed in time.

Optical crosstalk can be reduced increasing the distance between microcells (but at the cost of reducing the Fill Factor) or inserting optical insulators between them (trenches) [29].

A third source of correlated noise are afterpulses. Afterpulses are caused by impurities in the lattice that introduce trapping states corresponding to energy level close to the valence and conduction band, so that free carrier can be trapped in these states and released after a certain time proportional to the state energy. These carriers can then trigger other avalanches and thus, afterpulses (figure 2.14).

They can also occur as a consequence of the generation of an electron-hole pair in the substrate by a photon emitted during the avalanche that diffuses toward the substrate: if the generated carrier reaches the same microcell active region by diffusion, a new avalanche is triggered and an

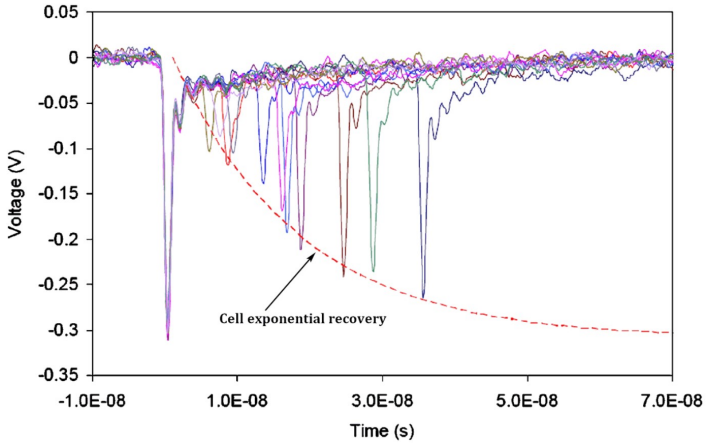


Figure 2.14: *Afterpulse waveform amplitude depend on the delay between the afterpulse and the original avalanche: if the cell is not completely restored, the second avalanche is characterized by a lower amplitude*

afterpulse is produced (in DeCT, instead, the avalanche is produced in a different cell). This phenomenon is referred to as Diffusive Afterpulsing. The delay of afterpulse generation is usually less than 100 ns and the amplitude is always lower than the one of the primary pulse, due to the fact that cell has not yet fully recharged.

All these effects are clearly visible through an oscilloscope, as reported in figure 2.15.

Excess Noise Factor

In photodetectors with internal multiplication, like APDs and PMTs, the internal gain is obtained by a multiplication mechanism, which is a statistical process: it is characterized by a mean value, i.e. the nominal gain, and a variance. Gain fluctuations enhances the noise on the overall measurement as they add uncertainty in the measure.

The ENF parameter is used to account for a fluctuation in the charge of the output signal. Despite having an internal multiplication process, SiPMs provide a highly uniform and quantized amount of charge in response to the absorption of a photon, leading to ENF very close to unity and mainly limited by optical crosstalk and afterpulses [30].

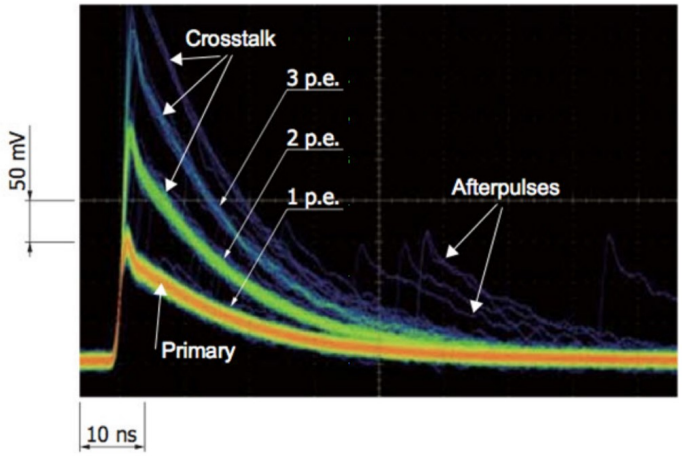


Figure 2.15: *Correlated noise waveforms acquired with an oscilloscope.*

$$ENF = 1 + \frac{\sigma^2_{cell}}{S^2_{cell}} \tag{2.9}$$

where σ^2_{cell} is the variance in gain between different cells (pixels) and S^2_{cell} is the average output signal from a single cell, and their ratio is smaller than 0.1 [31].

2.3 SiPMs readout

In literature, three are the most common approaches for SiPMs read-out [32]:

- **voltage mode:** the current signal coming from the SiPMs is converted into a voltage signal, by means of a small resistance, typically 50 Ω. The voltage signal is then amplified and filtered through a shaper to optimize the signal to noise ratio and to perform the charge measurement.

- **charge mode:** this approach consists in connecting one of the detector terminals directly on the virtual ground of an operational amplifier; the charge is obtained by direct integration of the current

signal coming from the detectors.

- **current mode:** it can be considered a development of the charge mode input stage, that adds a current scaling circuit in order to reduce the capacitance required to integrate the whole charge delivered from the detector.

The last method offers some important advantages respect to the charge mode. First of all, the use of a current feedback allows to provide a small input impedance and, consequently, a large bandwidth. Furthermore, the current mode approach does not suffer from possible voltage limitations due to deep submicron implementations; this enhances the dynamic range, allowing to exploit the whole SiPM signal [33].

2.3.1 SiPMs readout in INSERT

Clinical INSERT gamma module implements 8 tiles of SiPMs mounted on supporting PCBs (fig. 2.16). Each tile is constituted by 6x6 RGB-HD SiPMs (FBK, Trento). The bias is directly carried to all the SiPMs cathodes through bridge bondings between neighbouring photodetectors. The output anodes of four neighbouring SiPMs are connected together to form a merged channel, therefore, each acquisition channel reads the current signal provided by the sum of four SiPMs. The active area of the virtual SiPM (obtained by the merging) is 8 mm x 8 mm. Technical characteristics are reported in table 2.1.

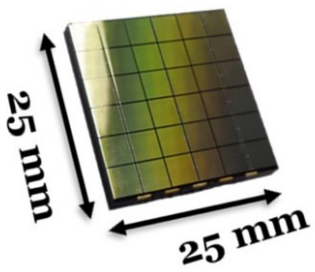


Figure 2.16: *Single SiPM tile. 6x6 SiPMs, each one with an active area of 4* mm^2.

RGB-HD characteristics	
microcell size	25 µm
SiPM active area	4 mm^2
effective PDE	35%
DCR (0 °C)	<50 $\frac{kHz}{mm^2}$
breakdown voltage (0°C)	29 V

Table 2.1: *Technical parameters of RGB-HD SiPMs.*

The high number of output channels, 72, requires a multi-channel readout ASIC and, furthermore, each channel should be able to manage a high input capacitance ($C_{DEP} > 10nF$) by keeping low the power dissipation. For this purpose, two ANGUS 36-channel readout ASICs are used to read and process the SiPMs output currents [34].

ANGUS implements a *current mode* approach at the input stage. As it possible to observe from the schematic structure of the ANGUS single-channel, depicted in figure 2.17, the first stage is constituted by a current conveyor. The low-impedance current input buffer takes the SiPM current and mirrors it to the filtering section. Switches B0,..,B3 can be set in order to regulate the gain of the input stage, in order to suit the dynamic range of the RC filter.

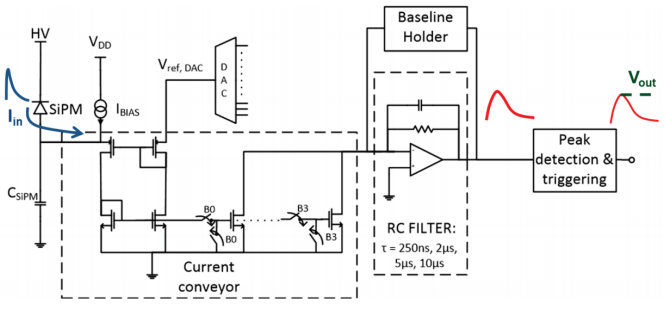


Figure 2.17: *Schematic structure of the channel of ANGUS. The first dashed box represents the current conveyor stage, made of a low-impedance buffer which mirrors the SiPM current to the filtering stage. This latter, contained into the second dashed box, is constituted by an RC filter, which converts the current pulse into a voltage pulse. Finally, the peak is detected by a peak detection stage and the triggering is provided. Image taken from [34]*

The DAC, instead, allows to tune the input voltage of the channel permitting to regulate the overvoltage of the connected SiPM. The shaping amplifier shapes the current pulse into a voltage signal. The amplifier has the function of filtering the signal and optimizing the SNR.

According to the optimal filter theory, in absence of a prior knowledge of the temporal distribution of the signal, the highest SNR is obtained by using as a shaper an infinite cusp [3]. However, the practical implementation of this shaping filter is not possible; a simple RC filter (with the RC network in the feedback loop) is used instead. The feedback resistance can be programmed to set different shaping times (from 200 ns to 10 µs, in order to match to the temporal shape of the input signal. The latter is mainly affected by the scintillation decay time of the crystal (under the assumptions that the SiPM contribution to the time distribution is negligible). CsI(Tl) is a slow scintillator, thus only shaping times greater than 1 µs are selected.

Fluctuations of SiPM dark current, bias current and input count-rate could lead to fluctuations of the DC current; in order to cope with this effect, a baseline holder circuit, in parallel to the RC filter, is used to stabilize the baseline at the input of the shaping filter to a fixed value.

Finally, the shaped pulse, which now consists of a voltage pulse, enters into the block indicated as *peak detection and triggering*. This block is constituted by a peak stretcher circuit and a triggering circuit: when the pulse overcomes a programmable threshold, the triggering circuit, which consists of a simple discriminator, enables the peak stretcher which tracks the peak of the pulse until an external ADC digitizes the voltage value.

Multi-channel acquisition and A/D conversion

The eletronics chain just described is identical for all the 72 SiPMs signals of a clinical INSERT module.

Given the high number of signals which have to be simultaneously acquired, processed and digitized, it is necessary to adopt an adequate *multi-channel approach*.

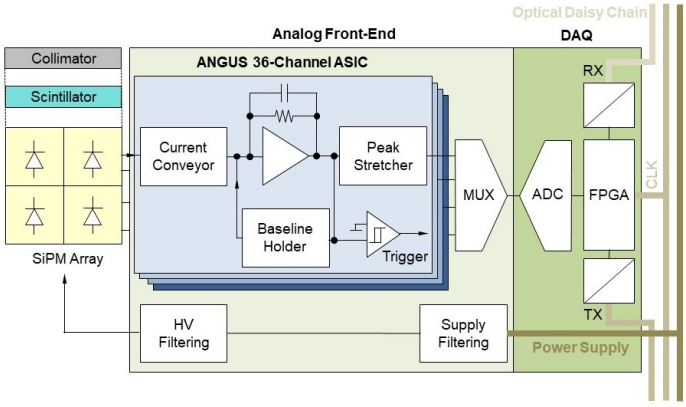


Figure 2.18: *Schematic of the single detection module. The output of each SiPM (actually four merged) is read by the single acquisition channel (36 acquisition channels for each ASIC). The parallel to serial conversion is performed by the MUX. The DAQ board on the right digitizes the analog values.*

The channel which detects the highest signal is the first to overcome the comparator threshold and, thus, the one which fires the output trigger. The generated trigger acts as a global trigger for all the other channels; in fact, it is sent to a global OR block, physically implemented outside the ASIC chip in order to act as a global trigger for both the ASICs.

In this way, a single trigger (the one coming from the highest channel) enables the peak stretcher blocks of all the 72 channels; then, each peak stretcher tracks its corresponding voltage signal up to the maximum value, holds it and provide it as output to the respective MUX for parallel-to-serial sequencing of the channels analog values.

After the acquisition of all the signals by the MUX, it sends a reset signal to all the peak stretchers. Two 18-channels MUX are used for each of the two ASICs.

The final block of the acquisition chain consists in the conversion from analog voltage signal to digital value; this step is performed on a separate board (DAQ board) by a 12 bit ADC, part of a complete FPGA-based data acquisition (DAQ) system (developed by Mediso Medical Imaging Systems, Hungary) [15]. This board provides the power supply for the ASIC board and for the SiPMs, manages the communication between the two ASICs (e.g. provides the global trigger to sample the 72 channels at

the same time) and programs the ASICs internal registers. Each signal between ASIC and DAQ is differential, to reduce the reciprocal interference between MRI and the SPECT module electronics. In particular, for digital signals between ASIC and DAQ, the LVDS standard is adopted.

ASIC and DAQ boards, with the attached scintillator crystal and SiPMs array, are shown in figure 2.19.

Figure 2.19: *(top) CsI(Tl) crystal wrapped in Teflon (white). The SiPMs coupled to the crystal are hidden below. The ASIC board under the heat sink is connected to the DAQ board on the right. (bottom) Top view of the ASIC board (left) and DAQ board (right).*

From figure 2.19, it is possible to observe also a heat sink, which is located between the photodetection array and the ASIC board, composing a compact "sandwich" geometry.

The purpose of this MR compatible heat sink is to provide a stable and moderate cooling for the reduction of the thermal noise, and thus the DCR, on the SiPMs.

Inside the heat sink, there is a serpentine in which flows a water-glycol mixture (respectively 40 % - 60 %); the cooling strategy is based on the heat extraction from the SiPMs thermal pads (placed on the bottom part of the tiles) by means of the heat sink directly in contact with them.

The refrigerant, cooled to -10 °C with a commercial chiller, flows inside the cooling block and then comes back to the chiller.

Furthermore, a small flux of nitrogen is pumped inside the module to prevent water droplets generated by condensation from falling on the electronic boards and cause short-circuits. Nitrogen is used because of its low dewpoint, compatible with the negative working temperature of the system.

Chapter 3

Planar reconstruction

A SPECT imaging system is based on detection modules able to detect γ-rays, which are emitted by pharmaceuticals containing radio-tracers and injected into the patient's body. The following step after the detection of the γ radiations consists in estimating the energy and the interaction coordinates of each event, in order to build 2D projections, which will be finally combined together in order to form a 3D tomographic volume, providing functional information about the patient. Different planar reconstruction methods for estimating interaction position and energy of gamma events have been proposed throughout the years. This chapter will describe the state-of-the-art methods of reconstruction and, in particular, the methods implemented in INSERT system.

3.1 Introduction to reconstruction methods

Any SPECT imaging system requires the acquisition of multiple projections, each one from a different angle; these planar acquisitions are then combined and processed by dedicated algorithms, such as Filtered Back-Projection (FBP), in order to reconstruct the 3D image of the scanned volume.

The present work, however, will be focused on the reconstruction of the planar acquisition rather than the 3D elaboration.

For each planar acquisition, the following informations need to be achieved:

- **(X,Y) coordinates of interaction**: the planar localization of the γ event is essential in order to obtain the final 3D image.
- **Energy of the gamma event**: the knowledge of the energy carried by each detected gamma photon is necessary, especially in the case of a multi-tracer examination, where different radioactive sources (each with its characteristic energy) are used simultaneously to study different biological processes. A common way to estimate the energy of a gamma event is to consider the sum of all the channels values for that event.
- **Z coordinate or DOI (Depth of Interaction):** the fact that the Z coordinate may be useful for a planar reconstruction seems counterintuitive. In reality, even if it is not strictly necessary for the planar image reconstruction, the DOI information can be crucial in order to improve spatial resolution. Indeed, a SPECT camera does not absorb only perpendicularly incident radiations, but also tilted ones, especially when collimators different from the standard parallel-hole are used. In the case of a tilted ray, since events can be absorbed at different DOI according to Lambert-Beer law, uncertainty about their depth of interaction results in an uncertainty about its (X,Y) scintillation coordinates. This effect, knowns as *parallax error*, causes a worsening of the spatial resolution. The possibility to distinguish the DOI of an event can significantly reduce the uncertainty in the determination of the scintillation point (figure 3.1).

3.2 Centroid-based methods

Centroid or Center-of-Gravity (CoG) algorithm is a linear reconstruction method, where the scintillation coordinates are calculated as the centroid

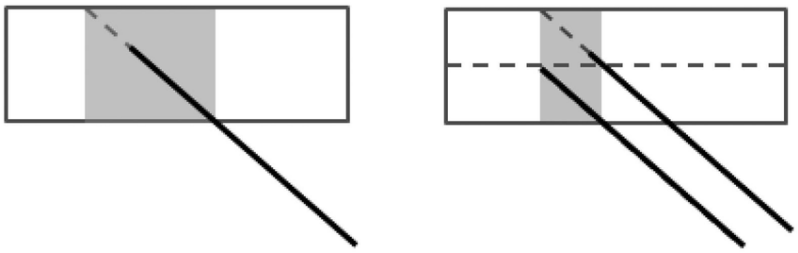


Figure 3.1: *Parallax error: on the left, the grey region represents the uncertainty region in absence of DOI information, while on the right, just knowing in which half (top or bottom) the photon is absorbed, the uncertainty region is halved*

of the light distribution over the photodetection matrix. In other words, the X_i and Y_i coordinates of the i-th gamma event are obtained by computing a weighted average of the positions of the D photodetectors, where each weight is represented by the corresponding signal measured by that photodetector:

$$X_i = \frac{\sum_{j=1}^{D} x_j \cdot s_{ij}}{\sum_{j=1}^{D} s_{ij}}; Y_i = \frac{\sum_{j=1}^{D} y_j \cdot s_{ij}}{\sum_{j=1}^{D} s_{ij}} \tag{3.1}$$

where s_{ij} is the signal acquired by the j-th photodetector in correspondence to the i-th event and D is the total number of photodetectors.

In other words each photodetector "attracts" the reconstructed point towards its center, with a weight dependent on the acquired signal.

The centroid method is the traditional reconstruction strategy; Anger logic, which was the very first implemented method for position and energy estimation in continuous scintillator-based gamma cameras is nothing but an hardware implementation of the centroid approach, where analog output signals, proportional to collected scintillation photons, are summed by a resistive network with proper weights. These weights are obtained by a resistor divider, which encodes photodetector positions, expressed as distances from the centre of the scintillator crystal [35] [36] .

Even if centroid method is still widely implemented in commercial

gamma cameras, thanks to its simplicity, it presents different limitations:

- **Compression of the image to the centre:** while gamma events absorbed near the center of the crystal are correctly reconstructed, centroid method is not able to reconstruct properly events interacting along the borders. This phenomenon, showed in figure 3.2, is due to the fact that light diffusion in the crystal and the electronic offsets of the acquisition channels add an almost constant baseline to the signals of all the channels, while useful signals are characterized by a bell-shaped spatial distribution. This leads to a shift in the reconstruction of lateral events towards the center of the image. This type of error is (partially) solved by the implementation of a "modified centroid" method, which will be discussed later.

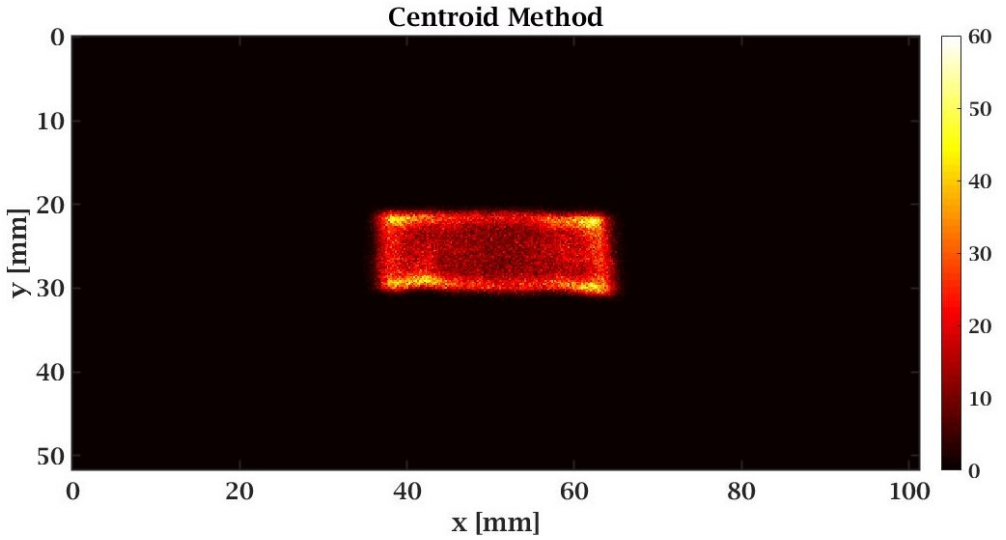


Figure 3.2: *Uncollimated irradiation (flood) reconstructed by Centroid method. In a flood irradiation, events should be reconstructed all over the detector surface (100mm × 50mm); instead, it is clearly visible the compression of the image towards the centre of the detector surface.*

- **Sensitivity non-uniformities:** another issue of centroid method is constituted by the introduction of fake density distribution of reconstructed events over the detector surface. These can be caused by the unavoidable slight differences between the photodetectors gain

or also by a narrow energy filtering around the photopeak; indeed, sensitivity non-uniformities produce estimated energy values, which present a spatial dependence and, therefore, the filter application may remove different amounts of events depending on the position [37]. There are different ways to cope with this type of distorsions. Some of them consist of hardware solutions, for example setting different voltage supplies to the photodetectors in order to compensate the gain differences, while others act at a software level and consist in using a weighted form of the centroid method, which weights each detector with a proper value, or in building correction maps of the detector which are estimated after a non-collimated irradiation of the detector surface.

- **Faulty channels:** this issue can be considered as an extreme case of sensitivity non-uniformities, where the gain of a detector is zero, because it is not working properly or it is broken. Centroid method is very sensitive to broken detectors, and this phenomenon generally creates huge distorsion of the image.

- **Nonlinearities-related distorsions:** Centroid method is based on a linearity assumption concerning the response of the camera. Even if this hypothesis stands at the centre of the detector, it is much less realistic along the borders, where a discontinuity is necessarily present due to the finite dimensions of the camera [38]. Furthermore, the reflection of photons by the lateral walls, enhanced by the use of reflective coverings, accentuates the non linearity of the camera response. Linearity corrections are usually executed by means of correction maps, which are obtained after grid irradiations covering the detector surface.

- **DOI-dependent artefacts:** These artefacts are related to the depth of interaction (DOI) of a γ photon. Indeed, when a γ photon is absorbed very close to the photodetector matrix, the generated optical photons are mainly collected by a single or few photodetec-

tors, which will consequently "attract" the reconstructed position towards its centre coordinates. An example of DOI-dependent artefacts is shown in figure 3.3.

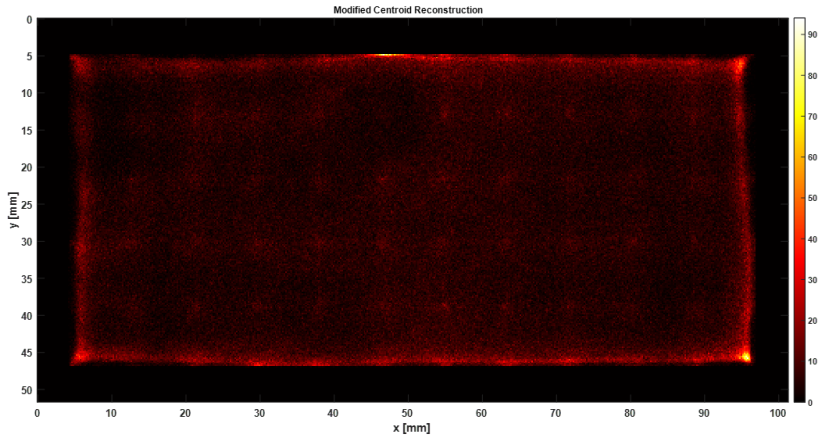


Figure 3.3: *Uncollimated irradiation of the crystal reconstructed by modified Centroid method, where DOI-dependent artefacts are clearly visible; the regions with higher density of events correspond to the SiPMs centers.*

3.2.1 Modified-Centroid methods

The compression effect typical of the Centroid reconstruction can be partially overcome by implementing **Modified Centroid** methods. One of them consists in subtracting a common baseline from the signal of each channel before reconstructing X_i and Y_i coordinates of the i-th event [39]:

$$X_i = \frac{\sum_{j=1}^{D} x_j \cdot (s_{ij} - B)}{\sum_{j=1}^{D} (s_{ij} - B)}; Y_i = \frac{\sum_{j=1}^{D} y_j \cdot (s_{ij} - B)}{\sum_{j=1}^{D} (s_{ij} - B)} \tag{3.2}$$

Since calculating the baseline B, which is mainly related to light diffusion, can be a complex task, it is more common to carry out an empirical research of a proper baseline value, based on the amount of FOV recovered, or just setting B equal to the mean signal calculated considering

all the events and all the channels. In figure 3.4, it is clearly visible the improvement apported by Modified Centroid method for what concerns the compression effect.

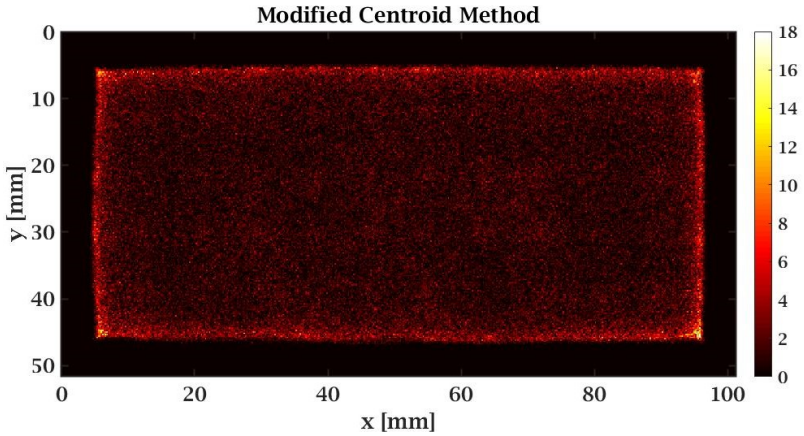


Figure 3.4: *Modified-centroid image: UFOV has been recovered almost totally from the compression*

Other approaches to obtain a linear "expansion" of the image consist in the application of a stretch to centroid reconstructed-coordinates, by multiplying each of them for a proper factor, or employing only photodetectors showing signals greater than a given threshold for a single event position computation (that means descarding low signals, usually not providing useful information as dominated by noise) [40].

3.3 Statistical methods

Differently from Centroid method or Anger Logic, statistical methods are based on a prior formulation of a probability model, which is estimated by measuring properties and response of the camera. For this reason, they provide better imaging reconstruction than Centroid-based methods, since they are able to take into account the random nature of gamma detection processes.

3.3.1 Maximum Likelihood reconstruction

Maximum Likelihood Estimation, MLE or simply ML, consists of a statistical estimation method that uses observed data to provide the values of some unknown parameters characterizing a given process described by a known statistical model. This method was first introduced in the field of position and energy reconstruction in gamma cameras by Gray and Macovski [41].

ML theory

Let $X_1, X_2, .., X_D$, with $X_j \sim M(\theta)$ be a set of D independent identically distributed discrete random variables. The probability model M describing them is known apart from the parameter θ to be estimated. The probability to observe a specific sample $\{x_1, x_2, .., x_D\}$ assuming that it was generated by a distribution $M(\theta)$ with θ unknown, is given by:

$$Pr(x_1, x_2, .., x_D|\theta) = Pr((X_1 = x_1) \cap (X_2 = x_2) \cap ... \cap (X_D = x_D)) \quad (3.3)$$

The variables X_i are independent between each other, thus the joint probability in eq.3.3 can be expressed as the product of marginal probabilities:

$$Pr(x_1, x_2, .., x_D|\theta) = \prod_{j=1}^{D} Pr(X_j = x_j) \quad (3.4)$$

The joint probability is a function of θ, the unknown parameter to be estimated; in particular, this function of θ represents the likelihood of the sample $\{x_1, x_2, .., x_D\}$:

$$L(\theta|x_1, x_2, .., x_D) = Pr(x_1, x_2, .., x_D|\theta) \quad (3.5)$$

where L indicates the likelihood.

The unknown parameter θ does not have to be necessarily a scalar; it can be a vector, composed by a series of parameters to be estimated, which univocally defines the probability model. So, generalizing to the

case of a vector $\boldsymbol{\theta}$ and an observation vector $\boldsymbol{x}$, we obtain:

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = Pr(\boldsymbol{x}|\boldsymbol{\theta}) \tag{3.6}$$

An estimator is, in other words, a function that maps a data vector $\boldsymbol{x}$ to an estimate of the unknown parameter, $\hat{\boldsymbol{\theta}}_{ML}$.

In MLE, the estimation rule consists in maximizing the likelihood function:

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{ L(\boldsymbol{\theta}|\boldsymbol{x}) \right\} \tag{3.7}$$

Thus, the ML estimate is the value of the unknown parameter for which the observation is the most likely result to have occurred. Note that an equivalent frequently applied rule consists in maximizing the logarithm of the likelihood:

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{ ln[L(\boldsymbol{\theta}|\boldsymbol{x}]) \right\} \tag{3.8}$$

ML reconstruction of gamma energy and interaction position

In the specific case of gamma event reconstruction, the $\boldsymbol{\theta}$ unknown parameter to be estimated consists in two main elements:

- coordinates of the gamma event absorption in the scintillation crystal : (X, Y) in the case of planar reconstruction, or (X, Y, Z) in the case of DOI-reconstruction.
- event energy ϵ_i (proportional to the total number of light photons N_{ph_i}).

Consequently, the unknown parameter vector results:

$$\boldsymbol{\theta} = (X, Y, Z, N_{ph})^T \tag{3.9}$$

while the observed data for the i-th event can be represented either by the number of detected photoelectrons for each of the D detectors:

$$\boldsymbol{x} = (n_{i1}, n_{i2}, ..., n_{iD}) \tag{3.10}$$

or by the D electrical signals produced by each channel:

$$\boldsymbol{x} = (s_{i1}, s_{i2}, ..., s_{iD}) \tag{3.11}$$

The parameters estimation needs an a-priori definition of a statistical model. A good model should be able to include all the possible random effects that can influence the data, once fixed the $\boldsymbol{\theta}$ parameter.

In a gamma camera, these random factors are constituted by:

- the random number of optical photons produced in each event
- random propagation of light to photodetectors
- the random number of photoelectrons produced
- random gain applied by photodetectors
- electronic noise

Given an incident photon with energy ϵ, a common assumption for linear scintillator-based detectors is that the number N_{ph} of optical photons produced by local energy deposition processes is a Poisson random variable. N_{ph} is usually considered as an estimate of ϵ since the number of produced optical photons is proportional to the energy according to the scintillation efficiency (or photon yield) of the crystal. If N_{ph} is a Poisson random variable, the number of photoelectrons produced in each photodetector is also a Poisson random variable because it results from a binomial selection applied to the starting N_{ph} photons [42]. Indeed, each of these ones can either hit a photodetector or not and, after that, produce a photoelectron or not.

Therefore, given a certain energy ϵ (represented by N_{ph}) and a certain scintillation position $\boldsymbol{r} = [X, Y, Z]^T$ of the gamma photon, the probability of the j-th photodetector to detect n_j photons is given by [43]:

$$Pr_j(n_j|\boldsymbol{r}, N_{ph}) = \frac{\mu_j(\boldsymbol{r}, N_{ph})^{n_j} \cdot e^{-\mu_j(\boldsymbol{r}, N_{ph})}}{n_j!} \tag{3.12}$$

where $\mu_j(\boldsymbol{r}, N_{ph})$ represents the expected number of optical photons which result in photoelectrons in the j-th photodetector, out of the N_{ph} initial ones.

As suggested by the notation, μ is function of both $\boldsymbol{r}$ and N_{ph}:

$$\mu_j(\boldsymbol{r}, N_{ph}) = N_{ph} \cdot \eta_j(\boldsymbol{r}) \tag{3.13}$$

with:

$$\eta_j(\boldsymbol{r}) = f_j(\boldsymbol{r}) \cdot \xi_j \tag{3.14}$$

where $f_j(\boldsymbol{r})$ represents the fraction of optical photons which, in average, are able to reach the j-th photodetector because of geometrical reason, while ξ_j represents the Quantum Efficiency of j-th photodetector.

So, if photoelectrons are considered as the observed data for position and energy estimation, the Poisson's model (eq. 3.12) applies. Through equation 3.4 the likelihood function (eq. 3.5) for the observation vector $\boldsymbol{x} = [n_1, n_2, ..., n_D]$ becomes:

$$L(\boldsymbol{r}, N_{ph}|n_1, n_2, ...n_D) = \prod_{j=1}^{D} \frac{\mu_j(\boldsymbol{r}, N_{ph})^{n_j} \cdot e^{-\mu_j(\boldsymbol{r}, N_{ph})}}{n_j!} \tag{3.15}$$

while the log-likelihood, that represents the actual function to be maximized (see equation 3.8), reduces to:

$$\ln(L(\boldsymbol{r}, N_{ph}|n_1, n_2, ...n_D)) = \sum_{j=1}^{D}(n_j \cdot \ln(\mu_j(\boldsymbol{r}, N_{ph})) - \mu_j(\boldsymbol{r}, N_{ph})) - \sum_{j=1}^{D} \ln(n_j)! \tag{3.16}$$

taking into account that the the last term of equation 3.16 is a constant and can be neglect in the optimization phase, the estimation of $\boldsymbol{\theta} = [\boldsymbol{r}, N_{ph}]^T$ is given by:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \Big\{ \sum_{j=1}^{D}(n_j \cdot \ln(\mu_j(\boldsymbol{r}, N_{ph})) - \mu_j(\boldsymbol{r}, N_{ph})) \Big\} \tag{3.17}$$

After the detection of a gamma photon, the electronic read-out chain provides as output the corresponding vector of signals $s_1, s_2, ..., s_D$.

In real systems, the detected signals are affected by electronic noise. The signal s_j measured by the j-th channel is thus a random variable, whose expected value is proportional to n_j by means of q_j, the average single photoelectron response of the j-th photodetector. If n_j is large (> 25 or more) and the single photoelectron distribution of the photodetector is reasonably symmetric, then, according to the central limit theorem, s_j is distributed following a Normal probability distribution with mean $\lambda_j(\boldsymbol{r}, N_{ph})$ and standard deviation $\sigma_j(\boldsymbol{r}, N_{ph})$, rather than a Poissonian one [44].

Even if in the case of $x = (s_{i1}, s_{i2}, ..., s_{iD})$ the Gaussian approximation would be more correct from a statistical point of view, the Poisson's model still applies and it is often used to describe also the signals output from the acquisition chain.

For this reason, the ML reconstruction algorithm implemented in INSERT applies the Poisson's model to both simulated data, measured in photoelectrons n_i, and INSERT measured data, measured in ADC bins s_i [15].

3.3.2 Least Squares reconstruction

Least Squares (LS) estimation is another member of the family of statistical methods which represents an alternative to ML and constitutes a more flexible solution when the photodetector signals s_j are considered. Differently from ML, where the estimated $\hat{\boldsymbol{\theta}}$ parameter is obtained by solving a maximization problem, in LS $\hat{\boldsymbol{\theta}}$ is found by minimizing χ^2, which is defined as the sum squared difference between the considered model and observed data.

LS theory

Given a set of D data points $(x_1, y_1), (x_2, y_2), ..., (x_D, y_D)$ and a known model function $y = f(x, \boldsymbol{\theta})$ where $\hat{\boldsymbol{\theta}}_{LS}$ represents the vector of parameters

to be estimated (with $D >=$ dimension of $\boldsymbol{\theta}$), it is possible to define, for the i-th observation, its residual as:

$$res_i = f(x_i, \boldsymbol{\theta}) - y_i \tag{3.18}$$

The optimal vector of parameters is the one for which the difference between the curve $y = f(x, \boldsymbol{\theta})$ and the observed data is minimum.

In LS, the quantity to be minimized is so represented by:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{M} res_i^2 \tag{3.19}$$

The estimated vector of parameters will be:

$$\hat{\boldsymbol{\theta}}_{LS} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}[\chi^2(\boldsymbol{\theta})] \tag{3.20}$$

LS reconstruction of gamma energy and interaction position

Let $s_1, s_2, ..., s_D$ be the photodetector signals produced after a gamma interaction with the detector, LS estimation of event position and energy is obtained by minimizing the following quantity:

$$\chi^2 = \sum_{j=1}^{D} (\lambda_j - s_j)^2 \tag{3.21}$$

where λ_j represents the expected value for the signal detected by the j-th channel , which corresponds to the curve model function to which the real measured j-th signal s_j is compared.

Under the hypothesis that photodetector signals are normally distributed, it is possible to state that:

$$\lambda_j(\boldsymbol{r}, N_{ph}) = \mu_j(\boldsymbol{r}, N_{ph}) \cdot q_j \tag{3.22}$$

where q_j indicates the average single photoelectron response of the j-th photodetector while μ_j equals to the average number of generated photoelectrons which depends on both N_{ph} and $\boldsymbol{r}$ according to equation (3.13).

Consequently, 3.21 can be rewritten as:

$$\chi^2(\boldsymbol{r}, N_{ph}) = \sum_{j=1}^{D}(N_{ph} \cdot \eta_j(\boldsymbol{r}) \cdot q_j - s_j)^2 \tag{3.23}$$

Thus, the LS estimate of $\boldsymbol{\theta}$ is obtained by solving the following minimization problem:

$$\hat{\boldsymbol{\theta}}_{LS} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left[\sum_{j=1}^{D}(N_{ph} \cdot \eta_j(\boldsymbol{r}) \cdot q_j - s_j)^2 \right] \tag{3.24}$$

3.3.3 Light Response Functions (LRFs)

Statistical methods are based on the prior definition of a mathematical model of the detector. After the model definition, the estimated position and energy of a new event is the one which provides the best match between the measured photodetector signals and the expected response of the camera provided by the mathematical model. As described previously, the best match can be obtained by maximizing (for MLE) or minimizing (for LS) a particular cost-function.

Therefore, these methods require the knowledge of the average response of each individual photodetector as a function of the event position, the so-called Light Response Function (LRF) .

The $LRF_j(\boldsymbol{r})$ is a function which represents the signal recorded by the j-th detector, normalized by the energy of the event, as a function of the (x,y) coordinates of the scintillation event (in principle, the LRFs can be defined as a function of z-coordinate too, but in this thesis work the reconstruction of the z-coordinate has been neglected). By way of an example, it is quite intuitive that a detector located in the up-left corner of the crystal will be characterized by a higher response when activated by a gamma photon interacting in the up-left corner and so forth.

$LRF_j(\boldsymbol{r})$ is nothing but $\eta_j(\boldsymbol{r})$, already defined in 3.14 as the average fraction of photons, emitted by a scintillation event with energy ϵ at position $\boldsymbol{r}$, that produces a signal in the j-th photodetector:

$$LRF_j(\boldsymbol{r}) = \eta_j(\boldsymbol{r}) \tag{3.25}$$

As a consequence, knowing the $LRFs$ of the D photodetectors allows to characterize the average response of the whole detector to an event with a given energy, as a function of its interaction coordinates.

LRFs estimation

Light Response Functions can be estimated by different methods. The most direct approach consists in a experimental characterization of the crystal, by means of a highly collimated radiation beam, which is used to scan the entire scintillator surface; the movement of the beam is usually generated by a robotic mechanism [42][45]. After the scanning phase, the LRFs can be computed for the beam coordinates and, then, missing positions can be obtained through data interpolation. Even if this technique is the one characterized by the higher reliability of the computed response, being estimated by experimental data and at controlled known spatial positions, the accuracy of the procedure is determined by the spatial sampling frequency of the scanning beam; this unavoidably introduces a trade-off between the number of positions to be covered and the acquisition time and complexity of the procedure. Furthermore, this approach becomes particularly challenging when detectors are mounted on the SPECT scanner.

An alternative to the previous method is the estimation of LRFs from simulated datasets obtained after a proper definition of a model of the camera [46]. In this case, the accuracy strongly depends on the exactness of the defined simulated model; furthermore, this strategy is not able to implicitly incorporate non-idealities of real detectors, differently from the previous method.

The last strategy, which is the one used in INSERT, is based on a iterative procedure, which exploits measured scintillation events produced by an uncollimated gamma source that uniformly irradiates the camera [44],[47],[48]. Even if this last approach takes advantage of experimental

measurements, it does not need long acquisitions and complex mechanical set-ups.

LRFs computation in INSERT

LRFs computation in INSERT, implemented by a MATLAB code called PERA [49], implements the last strategy described in the previous paragraph. In particular, the LRFs are estimated by means of an iterative procedure which consists in alternating multiple times two steps: reconstructing the interaction coordinates of the calibration events and using those coordinates to refine (through a proper curve fitting) the LRFs estimated at the previous iteration.

In order to apply the iterative process implemented in PERA, two main assumptions are required:

- the LRFs depend smoothly on $\boldsymbol{r}$
- all the events produce the same amount of light

The calibration measurement needed for the LRFs estimation is a *flood field irradation*, which corresponds to an uncollimated gamma source that uniformly irradiates the camera; in order to obtain an accurate characterization of the detector, a large population of calibration events should be acquired (70.000÷100.000 events).

The iterative process, illustrated in figure 3.5 , can be summarized into the following steps:

- **1$^{\mathbf{st}}$ - Filtering :** the first step consists in removing events whose estimated energy fall outside the chosen energy window around the photopeak (usually set at ±5% or ±10% around the photopeak), and, in addition, the remotion of events for which the number of activated SiPMs is under a given threshold. This second operation is done in order to filter out events interacting at high depth, which give rise to DOI artefacts.

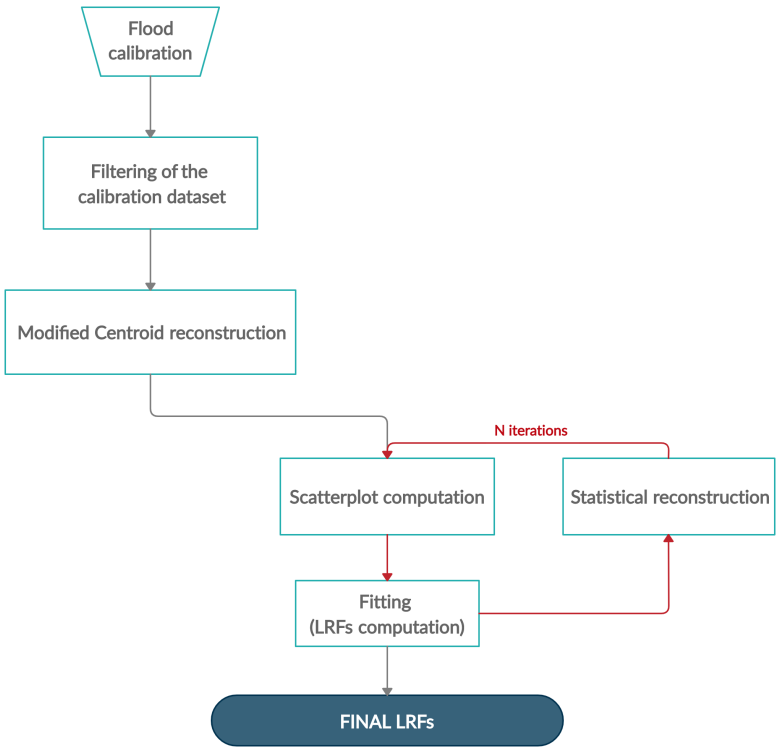


Figure 3.5: *Flow chart of the LRFs estimation process implemented by PERA*

- **2nd - Modified Centroid reconstruction :** the (X, Y) interaction coordinates of each calibration event are computed by using the modified centroid method.

- **3rd - Scatterplot fitting :** for each j-th SiPM, a 3D scatterplot, representing the acquired signal by that SiPM in function of the (X,Y) reconstructed coordinates, is computed. Then, each of the D scatterplots is fitted by using a parametric smooth function ($f(X, Y)$). The results of these fittings represent the first approximation of the LRFs ($LRFs^{(0)}$). It is important to underline that before computing the 3D scatterplot, each of the event is normalized by its energy, in a way that the LRFs obtained at the end of the process do not depend on energy.

- **4th - LRFs update :** This step constitutes the iterative part of

the process: events coordinates are reconstructed through a statistical method (both ML and LS reconstructions can be used), by using the LRFs estimated in the previous iteration. After the reconstruction, again the reconstructed coordinates are used to fit a new set of LRFs ($LRFs^{(iter)}$). These two operations are repeated till the stop condition (which can be maximum number of iterations, density threshold on the spatial distribution of reconstructed events or minimum change between the LRFs of two consequent iterations) is reached.

Iterating the LRFs estimation by means of statistical reconstruction methods allows to overcome the typical limitations of the centroid methods: the UFOV, originally compressed because of centroid reconstruction, enlarges and the reconstructed positions converge to a good approximation of the real ones, as it is possible to observe in figure 3.6.

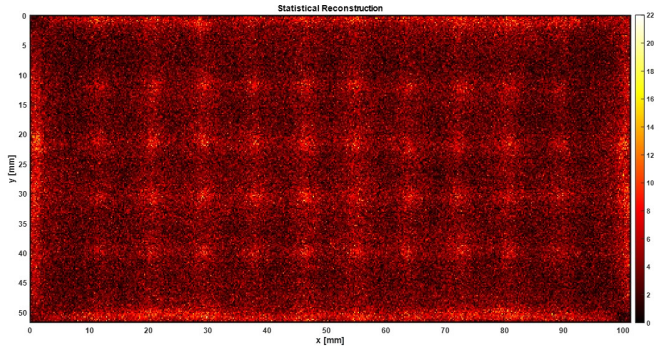


Figure 3.6: *The same flood irradiation shown in 3.2 and 3.4 reconstructed by using the ML method.*

Some technical "tricks" are useful in order to help the convergence of the algorithm, such as constraining the shape of the function used to fit the LRFs or adding a random shift (sampled from a Gaussian distribution with zero mean and a small standard deviation) to the reconstructed (X,Y) positions in order to avoid to be trapped in local minima.

A 2D Gaussian function [49] has been defined as fitting curve, as shown in figure 3.7.

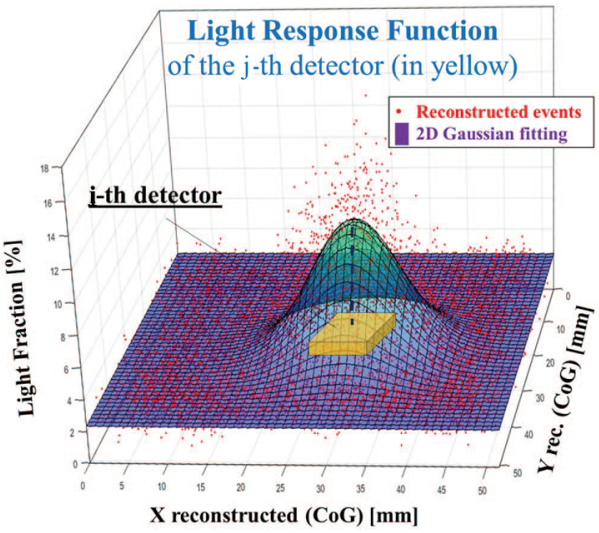


Figure 3.7: *2D Gaussian fitting of the calibration dataset after CoG reconstruction, for one detection channel positioned at the centre of the detection matrix. The bidimensional analytical interpolation result represents the Light Response Function (LRF) for the j-th detection channel. Figure taken from [49]*

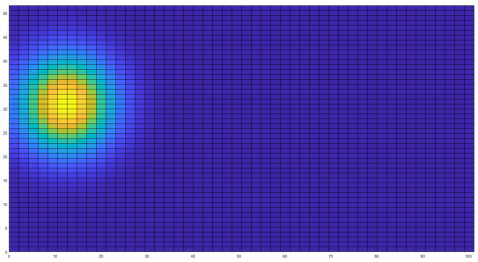
The parametric equation describing this family of functions is the following:

$$LRF(x,y) = Ae^{-b(x-x_0)^2+c(y-y_0)^2} + \phi \tag{3.26}$$

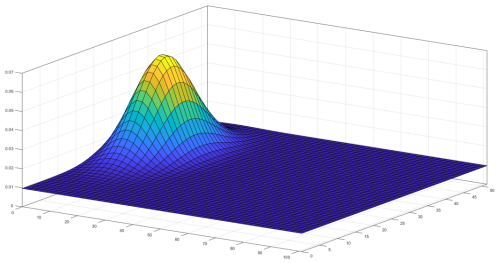
where A represents the amplitude of the Gaussian height, b and c are parameters inversely proportional to the Gaussian variance (respectively for the x and y directions), x_0 and y_0 are the coordinates of the centre of the Gaussian bell and ϕ is an offset proportional to the baseline threshold in the modified centroid method. The final LRFs obtained after the iterative estimation process are shown in figure 3.8

In conclusion, in order to estimate the planar interaction coordinates of an event with Maximum Likelihood, the following equation is applied:

$$(\hat{X},\hat{Y}) = \underset{(X,Y)}{\operatorname{argmax}} \Big\{ \sum_{j=1}^{D} [S_j \ln\left(LRF_j(X,Y)\cdot\hat{N}_{ph}\right) - LRF_j(X,Y)\cdot\hat{N}_{ph}] \Big\} \tag{3.27}$$



(a) LRF - Top view



(b) LRF - 3D view

Figure 3.8: *Light Response Function of a single channel at the last iteration of PERA estimation algorithm, seen from a top view(a) and a 3D view(b).*

where $\hat{N}_{ph}$ is estimated from:

$$\hat{N}_{ph}(X,Y) = \frac{\sum_{j=1}^{D} S_j}{\sum_{j=1}^{D} LRF_j(X,Y)} \tag{3.28}$$

With Least Square reconstruction, instead:

$$(\hat{X},\hat{Y}) = \underset{(X,Y)}{\operatorname{argmin}} \Big\{ \sum_{j=1}^{D} (\hat{N}_{ph} \cdot LRF_j(X,Y) - S_j)^2 \Big\} \tag{3.29}$$

where $\hat{N}_{ph}$ is estimated from the 3.28.

Chapter 4

PCA-based event reconstruction

Among the different tools usually employed in machine learning and, more in general, in data analysis, one important framework is constituted by dimensionality reduction techniques, of which probably the most common is Principal Component Analysis (PCA). The present thesis work has investigated and evaluated the possibility of integrating the PCA technique as a tool in order to accomplish the event reconstruction in a gamma camera. The two statistical methods defined in the previous chapter, ML and LS, have been modified so that they could operate in a new features space with lower dimensionality rather than the original one, where instead each feature corresponds to a given detection channel. This chapter will provide, at first, an overview of the dimensionality reduction techniques, their theoretical principles and their past applications in gamma imaging. Then, the whole process followed in order to integrate the PCA transformation in ML and LS statistical methods will be introduced. In the final part of the chapter, the obtained results on simulations and experimental measurements, conducted on INSERT clinical module, will be provided.

4.1 Introduction to dimensionality reduction techniques

Dimensionality reduction techniques are the set of methods allowing the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. Ideally, the reduced representation should have a dimensionality that corresponds to the *intrinsic dimensionality* of the data; the data are said to possess intrinsic dimensionality q if they are lying on or near a manifold with dimensionality q that is embedded inside the original space [50].

Given a $n \times D$ dataset $\boldsymbol{X}$, made of n observations $\boldsymbol{x_i}$ ($i \in \{1, 2, ..., n\}$) each defined in a D-dimensional space, reducing the dimensionality means transforming the dataset $\boldsymbol{X}$ into a new dataset $\boldsymbol{Y}$ with dimensionality d (with $d < D$), by preserving as much as possible the geometry of the data. In other words, each original observation $\boldsymbol{x_i}$ is replaced by $\boldsymbol{y_i}$, which is its corresponding in the d-dimensional space.

It is important to underline that, being the intrinsic dimensionality of the dataset unknown, the value of d, namely the dimension of the new space, is totally arbitrary (as long it is lower than D). This means that the truthfulness of the assumptions made about the intrinsic dimensionality and, in general, about the geometry of the data, will determine the goodness of the obtained transformation [50].

4.1.1 Principal Component Analysis

Principal Component Analysis (PCA) constitutes the most known dimensionality reduction method; it is a member of the family of linear techniques, since it performs reduction by embedding the data into a linear subspace of lower dimensionality.

Theory

The strategy adopted by PCA is to build up a low-dimensional representation of the data, by keeping as much as possible unchanged the informative

content, represented by the variance in the data. Concretely, PCA finds a new set of axes, named *principal components* (PCs), for which the variance contained in the data is maximum; in other words, applying this transformation consists in observing the data from a new reference system, obtained by rotating the original one, for which the distance between the observations (i.e. the variance) is higher. Figure 4.1 illustrates, by way of an example, the principal components of a dataset which is originally defined in a 2-dimensional space.

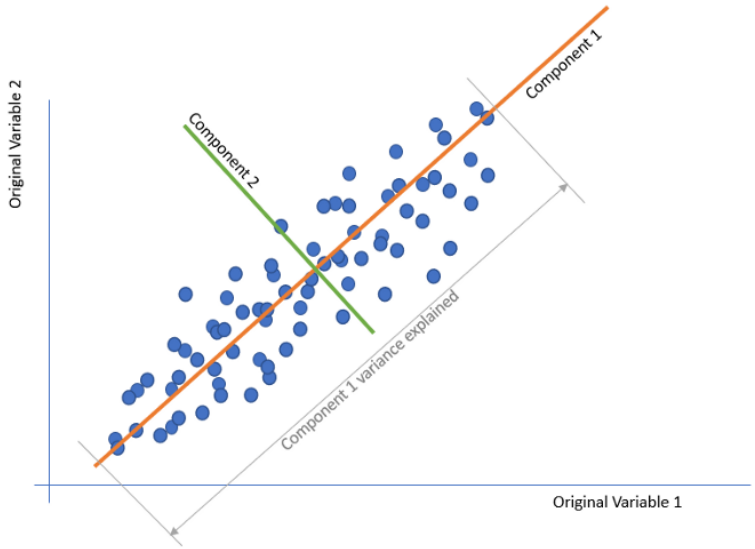


Figure 4.1: *Illustration representing the principal components of a dataset defined in a 2-dimensional space. Figure reproduced from [51].*

Each principal component, which in the rest of the discussion will be indicated as m_j (with $j = 1, ..., D$), is constituted by a linear combination of the original variables, by means of a specific set of weights, also known as *loadings*:

$$m_j = l_{j,1}X_1 + l_{j,2}X_2 + ...l_{j,D}X_D \tag{4.1}$$

where $X_1, X_2, ..., X_D$ are the original features, D is the dimension of the original space and $l_{j,k}$ represents the loading of the $k-$th feature on the j-th principal component.

Loadings describe how much each variable contributes to a particular principal component; large loadings (positive or negative) indicate a strong correlation to a particular principal component, while the sign of a loading indicates whether a variable and a principal component are positively or negatively correlated.

In mathematical terms, the optimization problem solved by PCA consists in finding the mapping $\boldsymbol{M}$ which leads to the maximum variance inside the data; $\boldsymbol{M}$ is a $D \times D$ matrix whose columns represent the principal components. Finding the mapping which solves this optimization problem is equivalent to solve the following maximization problem:

$$\boldsymbol{M} = \underset{\boldsymbol{M}}{\operatorname{argmax}}[\boldsymbol{M}^T \operatorname{cov}(\boldsymbol{X})\boldsymbol{M}] \tag{4.2}$$

$$\text{s.t.} \quad \|\boldsymbol{m}_j\|^2 = 1$$

where $\|\boldsymbol{m}_j\|^2$ represents the L2-norm of the j-th column of $\boldsymbol{M}$.

Notice that $cov(\boldsymbol{X})$ represents the sample covariance matrix of the "zero-mean" $\boldsymbol{X}$ matrix (the mean values of each feature are subtracted to the $\boldsymbol{X}$ original matrix before computing the covariance matrix).

Solving the constrained problem in 4.2 is equivalent to solve an unconstrained problem, if a Lagrange multiplier λ is defined. The generic j-th column of $\boldsymbol{M}$ can be found as:

$$\boldsymbol{m}_j = \underset{\boldsymbol{m}_j}{\operatorname{argmax}}[\boldsymbol{m}_j^T \operatorname{cov}(\boldsymbol{X})\boldsymbol{m}_j + \lambda(1 - \boldsymbol{m}_j^T \boldsymbol{m}_j)] \tag{4.3}$$

The stationary points are found when $cov(\boldsymbol{X})\boldsymbol{m}_j = \lambda \boldsymbol{m}_j$.

Hence, PCA basically solves the following eigenproblem:

$$cov(\boldsymbol{X})\boldsymbol{M} = \lambda \boldsymbol{M} \tag{4.4}$$

In other words, the linear mapping $\boldsymbol{M}$ is constituted by the set of eigenvectors of the covariance matrix of $\boldsymbol{X}$; the first principal component $\boldsymbol{m}_1$ will be the eigenvector associated to the greatest eigenvalue λ_1, and

the last one $\boldsymbol{m}_D$ will be the eigenvector associated to the lowest eigenvalue λ_D.

The positions of each observation in the new coordinate system of principal components, which are also called *scores*, can be calculated as the linear combination of the values of the original variables and the loadings l_{ij}.

For example, $y_{i,1}$, namely the score for the i-th observation on the first principal component can be computed as:

$$y_{i,1} = l_{1,1}x_{i,1} + l_{2,1}x_{i,2} + ... + l_{1,D}x_{i,D} \tag{4.5}$$

Passing in a matricial form, we obtain:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{M} \tag{4.6}$$

where $\boldsymbol{X}$ is the original dataset and $\boldsymbol{Y}$ the dataset projected in the new space.

Number of principal components

The computation of the principal components is based on an iterative algorithm, which starts finding the first component as the one which maximizes the variance contained in the data. At each iteration, the next principal component is computed as the one which maximizes the residual variance (namely the variance not yet explained by the principal components previously found) but adding the constraint of orthogonality with respect to all the principal components previously found (the principal components have to constitute a basis in $\mathbb{R}^D$).

Proceeding in this way iteratively, it is possible to find all the D principal components; in that case, the sum of the D eigenvalues would be equal to the overall variance contained in the original dataset, because:

$$Var(\boldsymbol{m}_j) = \lambda_j \tag{4.7}$$

i.e. the j-th eigenvalue is equal to the variance explained by its corre-

sponding eigenvector.

However, being the principal components more suitable to describe the data variability respect to the original features, a number d (with $d << D$) of principal components brings an informative content (almost) equivalent to the one of the originary features.

In particular, it is possible to calculate the percentage of the total variance which is explained by the first d principal components, as follows [52]:

$$Percentage_d = \frac{\lambda_1 + \lambda_2 + .. + \lambda_d}{\lambda_1 + \lambda_2 + .. + \lambda_D} \tag{4.8}$$

The estimation of the optimal number of principal components is usually performed by looking at a *scree plot*, which plots the percentage of explained variance as a function of the number of principal components; the number of components corresponding to the "elbow" of the scree plot is usually choosen as dimension of the new space. An example of scree plot is shown in figure 4.2

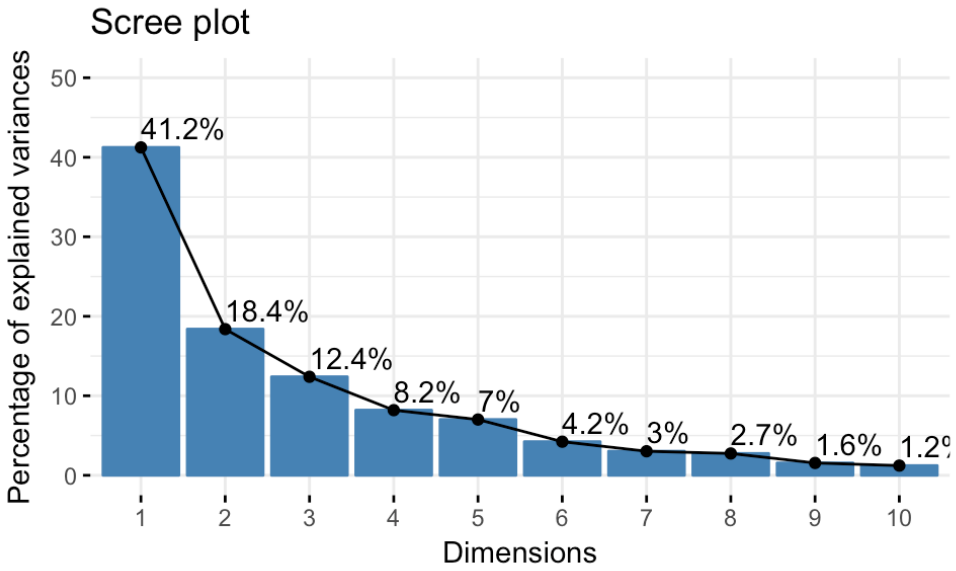


Figure 4.2: *Example of a scree plot.*

Advantages and limitations

PCA, and more in general dimensionality reduction techniques, are usually referred to as *unsupervised learning* techniques, meaning that, differently from problems like classification and regression, they are not characterized by a target variable. They are used mainly in a pre-processing stage, before applying supervised learning tasks; in this sense, their main application is *feature reduction*.

Feature reduction is a very common operation in machine learning, which consists in eliminating irrelevant features, by simply selecting a subset of them or creating a new set of features different from the original ones [52] (that is the strategy implemented by PCA). This practice is very common because, besides making possible to operate with smaller size problems, it has proved to improve, in some cases, the accuracy of the consequent prediction model [53] [54].

PCA also improves the interpretability and visualizability of the problem, which is a crucial factor, especially when dealing with high-dimensional datasets.

However, being a linear reduction technique, PCA cannot adequately handle data which lie on a non linear manifold; for this kind of data, non linear methods like LLE (Locally Linear Embedding) are more suitable.

4.1.2 Locally Linear Embedding:

There are some cases in which the data lie on highly non-linear manifolds, like the one illustrated in figure 4.3.

In these cases, one efficient method to map the data into a lower dimension is Locally Linear Embedding (LLE), which consists in writing the high-dimensional datapoints as a linear combination of their nearest neighbors [55].

Let $\boldsymbol{X}$ be a set of n data points $\boldsymbol{x_i}$ ($i \in \{1, 2, ..., n\}$), each with dimensionality D, sampled from some smooth underlying manifold. Provided the number of data points is sufficiently high (such that the manifold is well-sampled), we expect each data point and its neighbors to lie on or

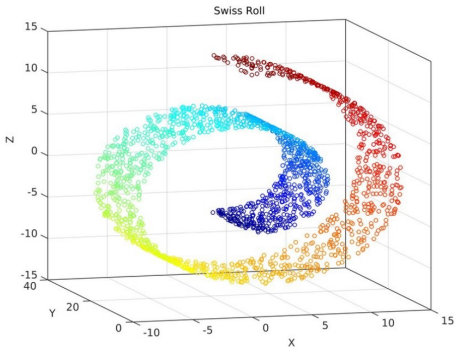


Figure 4.3: *Example of non linear 3-dimensional manifold. Even though the manifold is intrinsically non-linear, each point can be rewritten as linear combination of its k neighbors. Reproduced from [56].*

close to a locally linear patch of the manifold.

In LLE algorithm, the k nearest neighbors of each data point are identified, by using an Euclidean metric or, alternatively, by selecting all the points within a sphere of a given radius.

The total reconstruction error is computed as the sum of the squared distances between all the data points and their reconstructions:

$$E(W) = \sum_i \|x_i - \sum_j w_{ij} x_j\|^2 \tag{4.9}$$

where w_{ij} represents the contribution of the j-th data point to the i-th reconstruction.

The optimal weight matrix $\hat{W}$ can be computed as the one which minimizes the reconstruction error:

$$\hat{W} = \underset{W}{\operatorname{argmin}}\, E(W) \tag{4.10}$$

$$\begin{aligned} s.t. \quad & w_{ij} = 0 \quad \text{if } x_j \text{is not a neighbor of } x_i \\ & \textstyle\sum_j w_{ij} = 1 \end{aligned}$$

It is important to observe that, for any data point, the optimal weights which reconstruct it are invariant to rotations, rescalings, and translations

of that data point and its neighbor.

Assuming that the dataset has intrinsic dimensionality $q << D$, there exists a linear mapping (which consists of a translation, rotation, and rescaling) that maps the high dimensional coordinates of each neighborhood to global internal coordinates on the manifold; by design, the optimal reconstruction weights reflect intrinsic geometric properties of the data that are invariant to exactly such transformations.

Thus it is expected that their local geometry in the original data space is equally valid for local regions of the manifold; in other words, the same weights w_{ij} that reconstruct the i-th data point in D dimensions should also reconstruct its embedded manifold coordinates in q dimensions.

After the computation of the matrix of optimal weights $\hat{W}$, the mapping of each high dimensional observation x_i into the corresponding low dimensional vector y_i is computed by solving the following minimization problem:

$$\hat{Y} = \underset{Y}{\operatorname{argmin}} \sum_i \|y_i - \sum_j \hat{w}_{ij} y_j\|^2 \tag{4.11}$$

where $\hat{Y}$ represents the $n \times d$ matrix of the mapped dataset .

4.1.3 Applications to medical imaging

Spatial resolution constitutes a major requirement for a gamma camera both for PET and SPECT applications.

While in a pixellated camera, the resolution depends on the dimension of the pixel, in a continuous detector the achievable resolution is due to the specific algorithm used and it is not lower bounded by the photodetector dimension.

For both pixelated and continuous detectors, a common strategy to improve the achievable spatial resolution is using a larger number of smaller photodetectors coupled to the scintillator (or to the matrix of scintillators in the case of a pixellated camera). Increasing the number of detectors, however, can drammatically increase the complexity of the processing of

the output signals and the related electronics.

Furthermore, there are some applications where the high number of photodetectors is related to the size of the field-of-view which has to be covered; during the years, the interest of the research has moved more and more toward the development of high-FoV PET scanner, eventually leading to whole-body systems [57].

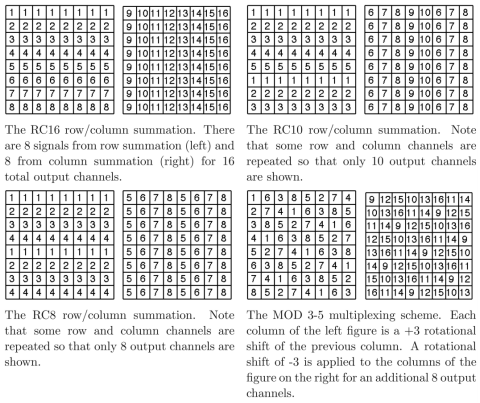
In this context, throughout the years different multiplexing techniques have been proposed, in order to reduce the number of output channels to process [58][59][60][61] [62]. However, it has been observed that reducing the number of output signals from the photodetectors can reduce the amount of available information to the positioning algorithm and could have a negative impact on the spatial resolution of the detector module [61].

It is precisely in this framework that dimensionality reduction techniques can constitute a powerful and useful instrument, since they offer the possibility to reduce the number of output channels in a "smart" manner, which allows to keep the original informative content almost unchanged.

In a work conducted by Pierce et al. in 2014 [62], PCA transformation of the original channels proved to guarantee the best positioning performances among different proposed multiplexing strategies, which are illustrated in fig 4.4.

LLE, instead, has been proposed as part of a new procedure to characterize the response of a DOI-sensitive detector from experimental pencilbeam measurements by means of an hybrid technique, which combined the Locally Linear Embedding with a finite mixture model [63]; in that case, LLE was applied individually to each spot-dataset (a laser beam was scanned in a grid-like pattern along the scintillator surface in order to obtain different spot-datasets) and, the spot-datasets have been projected, one by one, into a 1-dimensional embedding space, where, however, the depth information of each event was preserved, suggesting that the data were lying "intrinsically" on a 1-dimensional manifold inside the D-dimensional original space (figure 4.5).

In a following work [64], PCA and LLE techniques were combined to-



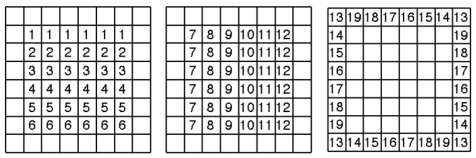
The RC16 row/column summation. There are 8 signals from row summation (left) and 8 from column summation (right) for 16 total output channels.

The RC10 row/column summation. Note that some row and column channels are repeated so that only 10 output channels are shown.

The RC8 row/column summation. Note that some row and column channels are repeated so that only 8 output channels are shown.

The MOD 3-5 multiplexing scheme. Each column of the left figure is a +3 rotational shift of the previous column. A rotational shift of -3 is applied to the columns of the figure on the right for an additional 8 output channels.

(a)



The MOD RC+edge multiplexing scheme. The left figure shows the six channels that sum the interior rows. The center figure shows the six channels that sum the interior columns. The figure at right shows seven additional channels in a ring around the edge of the detector module for a total of 19 output channels.

(b)

Figure 4.4: *Illustration of the tested multiplexing strategies. Each element of the 8*8 matrix represents a PMT output channel; its number, instead, indicates the corresponding multiplexing channel (PMT channels with the same label are summed together to achieve a single output channel). Reproduced from [62].*

gether. After the estimation of the principal components from a calibration simulated dataset, the corresponding PCA-multiplexing scheme was encoded in hardware in a resistive circuit and attached to the continuous detector module; in addition, a LLE-based calibration procedure able to operate directly on the multiplexed signals has been proposed, preventing from the need of performing scatter rejection and depth estimation before the attachment of the hardware multiplexing scheme.

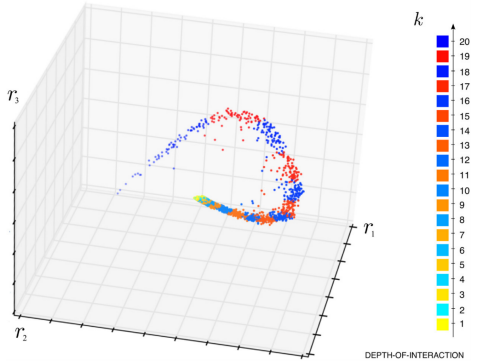


Figure 4.5: *Scatter-plot of the events projected onto a manifold of dimension d=3 (d=3 has been chosen just for visualization) for a given spot position. It is possible to observe that the data points are lying on a 1-dimensional manifold. Reproduced from [63]*

4.2 PCA-based reconstruction

The first part of the present thesis project was aimed at evaluating the possibility of integrating the PCA reduction inside the event reconstruction process for monolithic detectors.

The two statistical methods introduced in chapter 3, ML and LS estimation, require as input the D-dimensional vector of signals from all the photodetectors.

However, it is possible to abstract these algorithms to a more general formulation, which is able to operate in a d-dimensional space (with $d < D$) obtained by a PCA reduction; in the following discussion, the two proposed PCA-based algorithms will be referred to respectively as ML-PCA and LS-PCA.

Let x_i be the event whose (X, Y) planar interaction coordinates and energy have to be reconstructed:

$$x_i = [s_1, s_2, ..., s_D] \tag{4.12}$$

Let's suppose for the moment that the linear mapping $\boldsymbol{M}$, which defines the transformation from the original features space from which x_i has been drawn, to the new d-dimensional space, is known.

The knowledge of the linear mapping $\boldsymbol{M}$ allows to transform both the observation x_i and the $LRFs$ into the new space. In particular, y_i represents the d-dimensional representation of x_i :

$$y_i = [s_{PCA,1}, s_{PCA,2}, ..., s_{PCA,d}] \tag{4.13}$$

Instead, $LRFs_{PCA}$ represent the equivalent of the original LRFs into the new space .

At this point, it is possible to proceed in a very similar way to the classical ML and LS estimation to reconstruct the scintillation coordinates and energy of the y_i event.

Energy reconstruction

Differently from the original features space, in the principal components space the term $\sum_{j=1}^{d} s_{PCA,j}$, namely the sum of the features values observed for a given event, is not proportional anymore to the energy of the event.

This is the reason why the estimation of the energy of each event is still implemented according to equation 3.28 and, thus, requires all the signals by the photodetectors. This aspect is in accordance to a similar work by Pierce [62], where the energy estimation and the subsequent scatter rejection are applied by using all the original signals measured by the photodetectors.

(X,Y) scintillation coordinates reconstruction

The formulation of the optimization problem which leads to the estimation of the interaction (X, Y) coordinates is very similar to the one described for traditional statistical methods (subsection 3.3.3).

In particular, the LS-PCA estimation solves the following minimization problem:

$$(\hat{X}, \hat{Y}) = \underset{(X,Y)}{\operatorname{argmin}} \Big\{ \sum_{j=1}^{d} (LRFs_{,PCA}(X, Y) - s_{PCA,j})^2 \Big\} \tag{4.14}$$

Instead, the ML-PCA algorithm implements the following maximization problem:

$$(\hat{X},\hat{Y}) = \underset{(X,Y)}{\operatorname{argmax}} \Big\{ \sum_{j=1}^{d} [s_{PCA,j} \ln LRFs_{,PCA}(X,Y) - LRFs_{,PCA}(X,Y)] \Big\} \tag{4.15}$$

Both the expressions above, differently from equations 3.27 and 3.29, do not include the energetic term N_{ph} . This is due to the fact that each event is normalized respect to its own energy before undergoing PCA transformation.

4.3 Computation of the mapping and LRF estimation

In the previous paragraph, the whole formulation of the event reconstruction problem was based on the assumption that the linear mapping $\boldsymbol{M}$, namely the set of d principal components, was known.

However, in a real scenario, the computation of principal components has to be performed on a proper calibration dataset. The calibration dataset needs to contain events interacting all over the detector surface in order to guarantee that its principal components are able to represent as faithfully as possible the original features. This is the reason why a flood irradiation is used as calibration dataset.

In the present work, the same flood calibration dataset has been used both for the $LRFs$ estimation (by following the exact same procedure reported in 3.3.3) and for the computation of the principal components. Before performing these two operations, the calibration dataset has undergone an energy filtering; a window of $\pm 5\%$ around the photopeak, corresponding to the 122 keV of the Co-57, has been defined; all the events outside this window have been filtered out from the dataset. The adopted energy window is reported in figure 4.6.

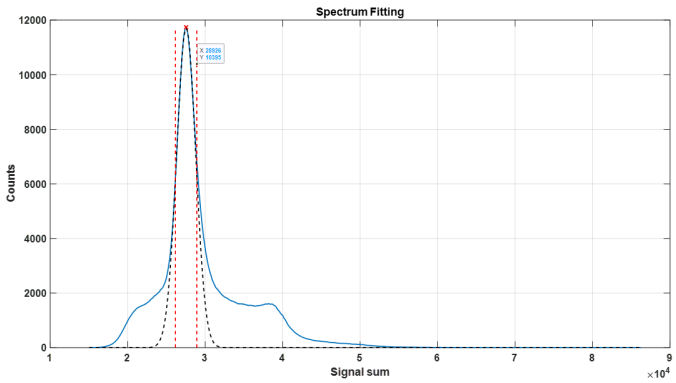


Figure 4.6: *Energy spectrum of the experimental calibration dataset; dark blue plot represents the rough computed spectrum, dashed black plot constitutes a gaussian fitting of the spectrum and the dashed red vertical lines represent the boundaries of the defined energy window.*

Computation of the principal components

The energy-filtered calibration dataset has been then used in order to compute the mapping $\boldsymbol{M}$, namely the set of d principal components.

Principal components computation consisted in the following operations:

- Normalization of all the calibration events with respect to their energy, computed as sum of all the detected signals for that event
- Subtraction of the mean of each channel
- Extraction of the d principal components, as the eigenvectors of the covariance matrix of the dataset, corresponding to the d greatest eigenvalues

It is important to underline that the value of d, namely the dimension of the new space, is totally arbitrary, and it is a tunable parameter. For this reason, as it will be described later, different values of d have been tested in order to evaluate how the positioning performances could change by changing the number of principal components used.

Projection of LRFs in the principal components space

In order to be able to apply the LS-PCA and ML-PCA estimations introduced respectively in equations 4.14 and 4.15, both the LRFs and the event to reconstruct have to be projected into the new space.

While the projection of an event in the principal components space is obtained by a simple scalar product between the observation and the mapping M, in order to obtain $LRFs_{PCA}$, which represent the equivalent of the $LRFs$ in the transformed space, a different approach has been required.

As introduced in chapter 3, the j-th LRF represents the expected signal measured by the j-th photodetector as a function of the (X,Y) interaction coordinates of the photon on the scintillator surface. In order to make feasible the computation and the allocation of these 3D continuous curves, a proper binning of the (X,Y) coordinates is performed; the length of the crystal along the X and Y direction is divided respectively into 506 and 258 discrete bins, with a pixel dimension of 0.2 mm. Consequently, the $LRFs$ estimation process implemented in PERA leads to the definition of D different matrices with dimension 258×506, where D is the number of channels of the module.

In order to obtain the $LRFs_{PCA}$, a new dataset has been created, where each row represented one of the (258×506) bins while the column represented the value assumed by each LRF for that specific bin; this operation allowed to build-up a D-dimensional dataset which could be then projected into the new principal components space.

After the projection in the principal components space, the $LRFs$ have been composed again into a d-dimensional object defined inside the same 258×506 bins matrix, which finally constitutes the $LRFs_{PCA}$.

It is important to observe that the ML-PCA optimization problem in equation 4.15 requires to compute the logarithm of $LRFs_{PCA}(X, Y)$; however, after the projection of the $LRFs$ in the new space, they contain both positive and negative values. For this reason, in order to be able to use the logarithmic formula, an offset value is summed to both the

$LRFs_{PCA}(X, Y)$ and the event to reconstruct, in order to obtain only non negative values.

This step, on the contrary, is not required for the LS implementation.

4.4 Validation on simulations

In a first phase of the work, the reconstruction performances of the two PCA-based methods have been validated on simulations, which have been obtained by means of a specific simulation package, called ANTS2.

4.4.1 ANTS2

ANTS2 is a simulation and experimental data processing package for Anger camera-type detectors. It is an open source and multiplatform package, developed using CERN ROOT and Qt framework (C++). It represents the second release of the software package ANTS (Anger camera type Neutron detector: Toolkit for Simulations) [65], originally developed by Coimbra LIP for simulation and event reconstruction of Anger-type gaseous detectors for thermal neutron imaging.

Besides providing tools for statistical reconstruction and for LRF estimation by iterative procedure, ANTS2 gives the possibility to model scintillator-based detectors and to simulate the processes of particle interaction, production and propagation of the scintillation light and generation of the detector output.

In order to generate a dataset on ANTS2, two steps are required:

- **Generation of the model of the camera:** a model of the Anger camera has to be defined, with its geometrical parameters and optical properties.

- **Definitions of the simulation parameters:** ANTS2 package allows to operate following two different modalities:

- *Photon source* modality consists in skipping the process of gamma-ray interaction with the materials; just emissions of optical photons at user-defined positions in the scintillator are simulated.
- *Particle source* modality, instead, allows the simulation of gamma photons from a radioactive source and their interaction with the detection medium, followed by emission of scintillation photons. Energy, shape (point-like, circular or rectangular), size and position of the source can be set. For this modality the distribution of events is in accordance with the real physics of gamma interaction: Lambert-Beer law is respected and the real Poisson's photon statistics stands.

For both the modalities, the simulation output consists of a table where each rows represent the simulated event and the column values represent the set of photodetector signals; furthermore, the true (x,y,z) scintillation coordinates and energy of each event are provided. Simulated datasets can be exported in the form of text files to be processed by means of other software (i.e. MATLAB). Differently from the real system, where the output of each channel is expressed in ADC bins, the simulator provides as signal output the number of detected photoelectrons.

Generation of the model of the Anger camera

Generating the model of a Anger camera means defining the geometrical and optical parameters of the camera itself.

The ANTS2 model of the clinical INSERT detection module, developed in a previous thesis work [66], requires the definition of the following components:

- Scintillator crystal
- SiPM detection matrix
- Crystal-SiPMs optical interface (Meltmount)
- Reflective wrapping of the crystal (Teflon)

- Scintillator-Teflon interface

An illustration of the defined structure is shown in figure 4.7.

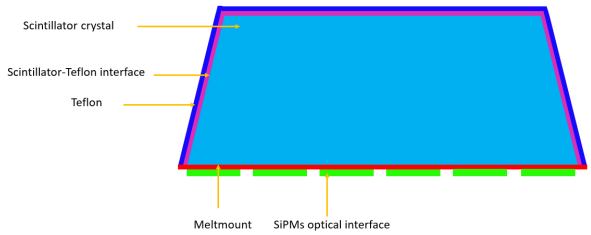


Figure 4.7: *Illustrative scheme representing the different components defined to model INSERT clinical module.*

It is important to clarify that the component indicated as Scintillator-Teflon interface does not correspond to a real physical component but it was defined to take into account the optical effects of the non-attachment of the Teflon layer to the crystal surface in the real module (instead of being glued to the crystal, it was just tightly wrapped around it and fixed by a black tape).

Dimensions of the components have been set in order to reproduce INSERT clinical module as faithfully as possible, and thus:

- Scintillator has been defined as a trapezoid with squared base faces (101.2 × 51.6 mm for the lower face, 101.2 × 48.9 mm for the upper face and 8 mm of thickness)

- Interfaces between different materials have been defined with reasonably realistic values, being impossible to directly measure them (0.1 mm thickness for both the interfaces)

- Teflon layer has been set to 0.2 mm thickness

- Photodetector matrix has been defined as a 12×6 matrix of square-shaped SiPMs with size 8.2 × 8.2 mm

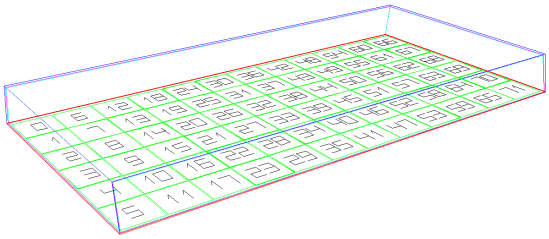


Figure 4.8: *Model of clinical INSERT single detection module, developed through ANTS2 toolkit. The SiPM 2D array (green) lies at the bottom, the Meltmount (red) is used to couple the SiPMs to the CsI(Tl) crystal (light blue). Teflon-wrapping (blue) is the most external layer.*

Besides defining the geometrical parameters of each component, it is necessary to characterize each component in terms of optical properties, in particular regarding their interaction with gamma particles:

- **SiPMs:** effective photon detection efficiency (PDE) was set to 0.35 (resulting from the coupling between the scintillator emission spectrum and the PDE-vs-wavelength response of the photodetectors) and dark count rate (DCR) was set to 25 $\frac{\text{kHz}}{\text{mm}^2}$ (DCR at T=-10 °C) .
- **SiPMs optical coupling :** Refractive index was set to 1.515.
- **Meltmount:** Meltmount, refractive index was set equal to 1.539 and the bulk absorption to 0.08 mm^{-1}
- **CsI scintillator:** The crystal has a density of 4.52 $\frac{\text{g}}{\text{cm}^3}$, a photon yield of 65 $\frac{\text{ph}}{\text{keV}}$ and an intrinsic energy resolution of 6 %. The refractive index is equal to 1.79. The scintillator was set as the only material able to interact with gamma rays (mass attenuation coefficient 1.425 $\frac{\text{cm}^2}{\text{g}}$ at 122 keV, mean free path 1.556 mm). The interaction properties of photoelectric and Compton effects for CsI were directly loaded in ANTS2 from NIST XCOM web database [67].
- **Teflon wrapping**: the optical properties at the interface with Teflon were characterized by 6 % of photons absorption and 94 % of Lam-

bertian scattering (diffusive reflection). For Teflon, the refractive index is set to 1.35 and the density to 2.2 $\frac{g}{cm^3}$.

Definitions of the simulation parameters

All the simulations have been run by using the *particle source* modality. This means that all the parameters regarding the gamma ray source had to be set: type of gamma ray source (a Co-57 source has been used, in accordance with the experimental measurements), distance from the detector, shape of the source, orientation and collimation. Details about the specific simulation parameters will be provided into the following sections.

4.4.2 Impact of the value of *d* on the reconstruction performances

The first step performed in the process of evaluation of the PCA-based statistical methods was the investigation of the reconstruction performances of the algorithms by varying the number of principal components and, thus, the dimension d of the new features space.

The knowledge of the real (X, Y) interaction coordinates of each detected gamma event allowed the definition of an *absolute error*. In particular, the reconstruction error for the i-th event is given by the Euclidean distance between the real interaction position of the event and the reconstructed one:

$$\epsilon_i = \sqrt{(x_{true} - x_{reconstructed})^2 + (y_{true} - y_{reconstructed})^2} \tag{4.16}$$

The evaluation of the error required the generation of two different simulated datasets, a *calibration* dataset and a *validation* dataset, in order to be able to test the reconstruction performances on "fresh" data.

A flood irradiation was simulated in order to generate each of the two datasets; an uncollimated point-like Co-57 source has been defined at a distance of 30 cm from the crystal surface.

After undergoing the energetic filtering stage and the $LRFs$ estimation, already described in section 4.3, the calibration dataset was used for the principal components extraction by varying the value of d.

A value of d from 2 to 50 has been tested and the corresponding Root Mean Square Error (RMSE) has been computed on the validation dataset and plotted against the RMSE of the traditional statistical methods.

The Root Mean Square Error is commonly employed as index of the spatial resolution of a camera, and it is defined as:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{\epsilon_i^2}{N}} \tag{4.17}$$

where ϵ_i is defined in equation 4.16 and N is the number of events.

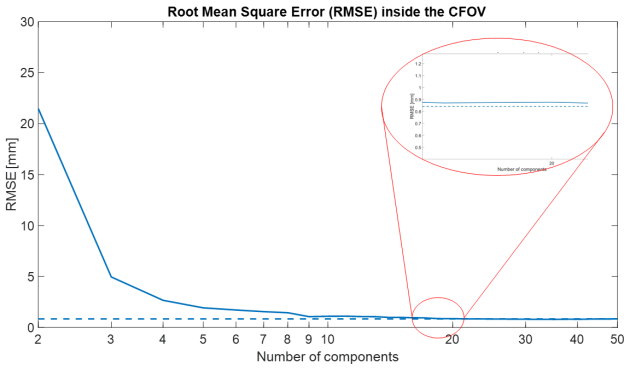
Notice that, before reconstructing the validation events, they have undergone the same energetic filtering stage adopted for the calibration dataset ($\pm 5\%$ around the photopeak) in order to filter out scattering events.

The spatial resolution generally worsens along the borders of the detector respect to the centre; for this reason, the computation of the RMSE has been carried out separately for the events whose real interaction coordinates laid inside the CFOV and the ones outside the CFOV. The CFOV practically excludes the first 10 mm from the borders of crystal surface.

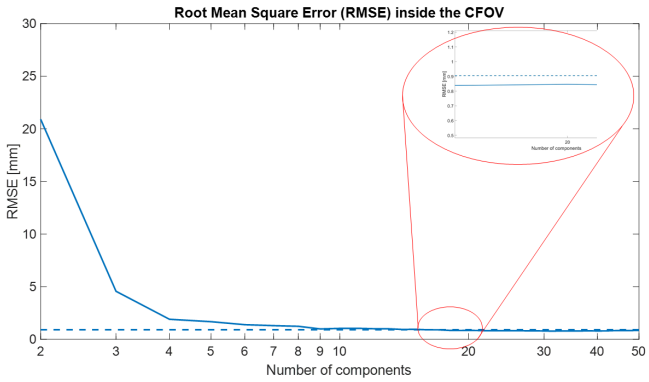
Figure 4.9 shows the RMSE inside the CFOV as a function of the value of d, both for ML-PCA and LS-PCA algorithms.

It is evident from figure 4.9 that, independently from the specific statistical algorithm used (ML or LS), increasing the number of principal components brings to a decrease of the RMSE, as a consequence of the higher informative content available. Both the curves show an elbow-like trend, which reflects the shape of the scree plot of figure 4.2.

With a value of $d = 10$, the error starts to be comparable to the one of ML and LS algorithms, respectively equal to 0.84 and 0.90 mm; by the way, increasing even more the value of d does not lead to a further improvement of the spatial resolution, meaning that the next components do not bring useful informations.



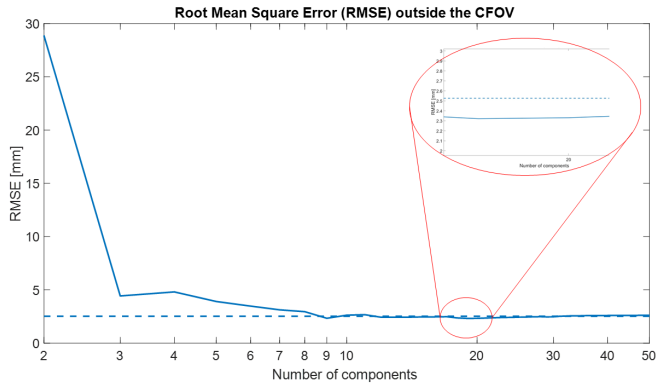
(a) ML-PCA Root Mean Square Error (CFOV)



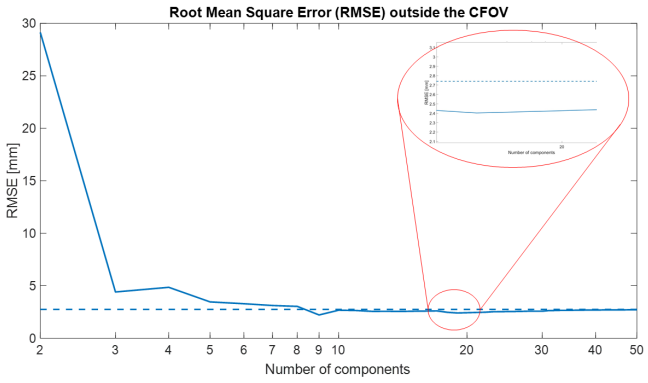
(b) LS-PCA Root Mean Square Error (CFOV)

Figure 4.9: *RMSE of ML-PCA (a) and LS-PCA (b) methods as a function of the number of principal components on a semi-log scale (continuous line), plotted against the RMSE of the traditional statistical method (dotted line). The error is calculated only for events interacting inside the CFOV*

The same plots are reported in figure 4.10 for events outside the CFOV; it can be observed that, even though the RMSE settles to a higher value than the one inside the CFOV ($\sim$ 2.5 mm), also in this case 10 principal components seem to contain enough information to ensure a spatial resolution practically equal to the one of standard ML and LS methods.



(a) ML-PCA Root Mean Square Error (borders)



(b) LS-PCA Root Mean Square Error (borders)

Figure 4.10: *RMSE of ML-PCA (a) and LS-PCA (b) methods as a function of the number of principal components on a semi-log scale (continuous line), plotted against the RMSE of the traditional statistical method (dotted line). The error is calculated only for events interacting outside the CFOV.*

4.4.3 Quantization of the principal components loadings

Each of the principal components is constituted by a set of D weights (or loadings), each one expressing the weight of a given channel for that specific principal component.

In the perspective of a hardware implementation of the PCA-based multiplexing of the original channels through resistive (or capacitive) components, like the one implemented in Pierce work [64], the values of the

weights associated to each channel have to be quantized with a finite number of bits N_{bits} (and thus limited to a finite set of possible $2^{N_{bits}}$ values).

For this reason, one concern of the evaluation on simulated data has been to estimate the impact of the loadings quantization on the accuracy in the reconstruction. For this purpose, once fixed a value for the number of quantization bits N_{bits}, each eigenvector j ($j = 1, ..., d$) has been quantized by admitting only equally spaced values in the interval $[\alpha_j, \beta_j]$, where α_j and β_j represent respectively the minimum and the maximum value assumed by the weights of the j-th eigenvector.

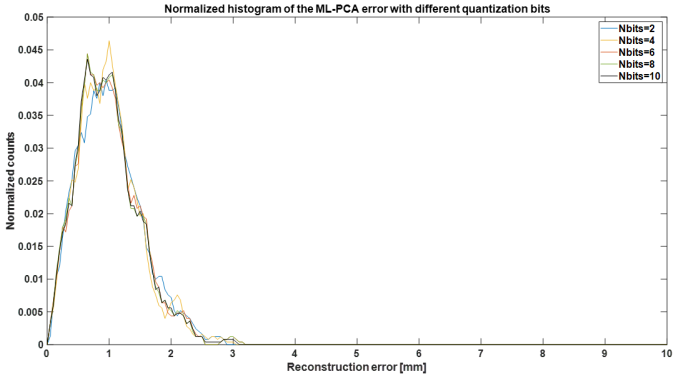
This quantization has been repeated with different values of N_{bits} (a value of bits equal to 2,4,6,8,10 have been set).

For this evaluation, the number of principal components d has been fixed to 10.

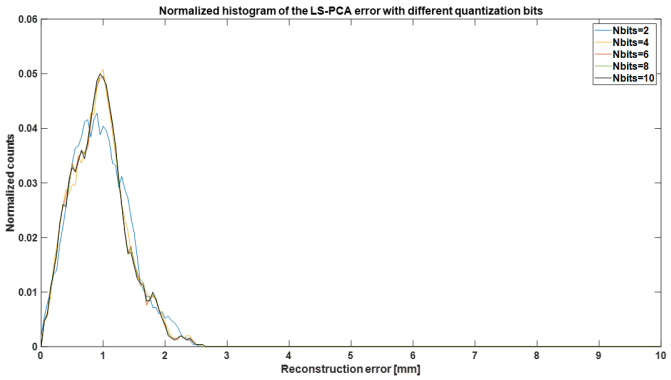
The smoothed histograms of the reconstruction error inside the CFOV, by varying the number of quantization bits, are shown in figure 4.11, respectively for ML-PCA and LS-PCA. It is important to notice that, independently on the number of bits used for the quantization of the loadings of the principal components, the histogram of the reconstruction error are Gaussian-shaped with a mean around 0.9 mm and a standard deviation of 0.4 mm.

The fact that the encoding of the computed weights into a set of discrete values does not require an high degree of precision (2 bits seem to provide a spatial resolution practically equal to the case of 10 bits) seems to be very promising for an hardware implementation of the weights through resistive networks or other components.

Figure 4.12 depicts the 3D barplots of the loadings of the first 10 principal components extracted from the calibration dataset ($d = 10$). In particular, on the top, it is shown the case where the weights have not undergone a quantization, while, on the bottom, they have been quantized on 2^2 intervals ($N_{bits} = 2$). It is possible to observe how the effect of the quantization on the values of the weights seems minimal.



(a) ML-PCA



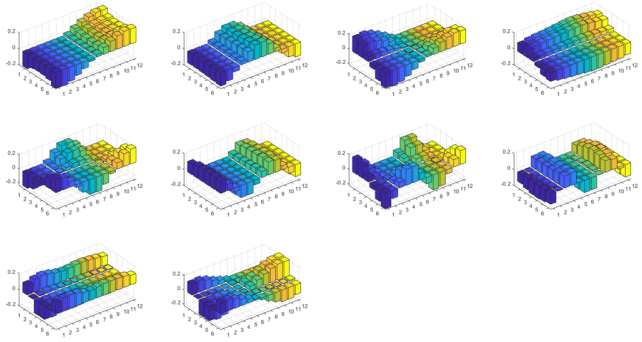
(b) LS-PCA

Figure 4.11: *Normalized histograms of the reconstruction error inside the CFOV with varying the number of quantization bits, for ML-PCA (a) and LS-PCA (b).*

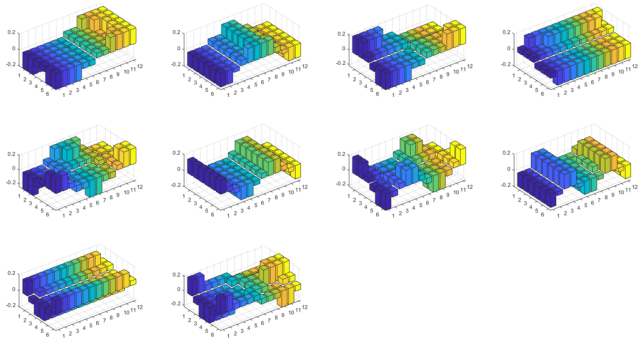
4.4.4 Validation on different SiPMs geometries

In both monolithic and pixellated detectors, the dimension of the SiPM is a crucial factor affecting the minimum achievable spatial resolution. A good strategy to improve the spatial resolution is the use of smaller photodetectors, but at the cost of increasing the complexity of the read-out electronics; however, the capability of PCA of scaling down the dimensionality of the data may be exploited to overcome this drawback, by keeping low the number of signals required in input to the positioning algorithm.

This is the reason why, besides the real INSERT clinical module, other



(a) No quantization



(b) Quantization on 2 bits

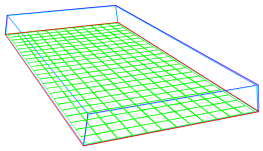
Figure 4.12: *Barplots representing the value of the weights of each channel for the first 10 principal components of the calibration dataset, respectively with no quantization (top) and 2 bits quantization (bottom).*

3 models of the camera have been defined in ANTS2 and, consequently, evaluated in terms of the corresponding spatial resolution.

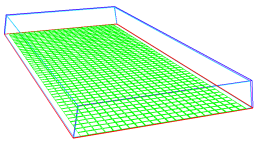
The geometrical and optical parameters previously described in 4.4.1 have been kept almost unchanged, with the only exception of the SiPMs array. The 3 proposed and evaluated geometries have been:

- 24×12 matrix of square-shaped SiPMs with size 4×4 mm
- 46×22 matrix of square-shaped SiPMs with size 2×2 mm
- 84×42 matrix of square-shaped SiPMs with size 1×1 mm

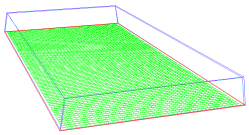
The defined geometries are showed in figure 4.13.



(a) Model with $4mm \times 4mm$ SiPMs



(b) Model with $2mm \times 2mm$ SiPMs



(c) Model with $1mm \times 1mm$ SiPMs

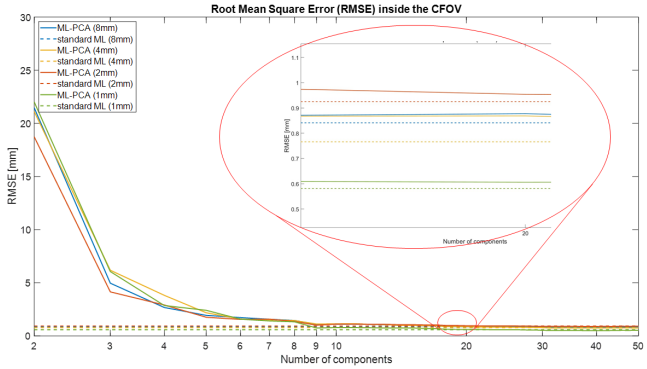
Figure 4.13: *The 3 different evaluated geometries for INSERT clinical module.*

Similarly to what has been done for INSERT clinical module simulations, for each defined geometry, the RMSE made by ML-PCA and LS-PCA algorithms have been computed by varying the number of principal components.

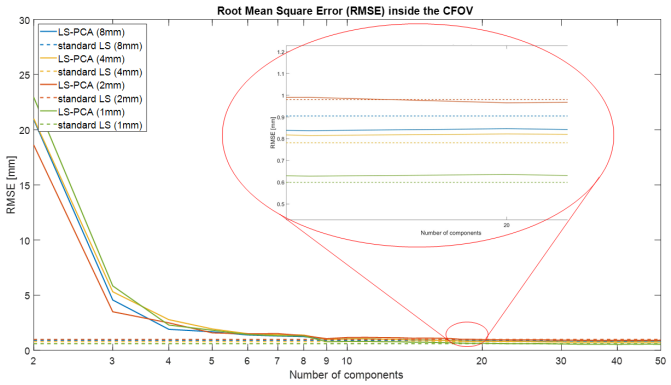
In relation to the results of these simulations, shown in figure 4.14 and 4.15 respectively for the CFOV and the border regions, two considerations can be made:

- the RMSE seems to decrease when decreasing the SiPMs dimension, especially when using SiPMs with 1 mm size.
- the number of principal components to extract in order to achieve a spatial resolution comparable to the one of the corresponding statistical method does not depend on the total number of SiPMs.

Table 4.1 and table 4.2 show the RMSE computed for each of the simulated geometries, in the case of ML and LS reconstructions.



(a) ML-PCA

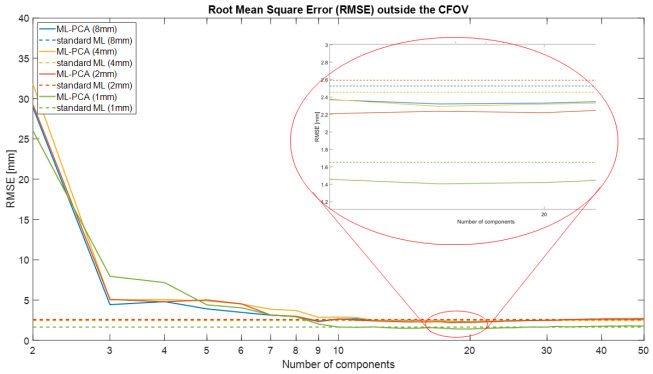


(b) LS-PCA

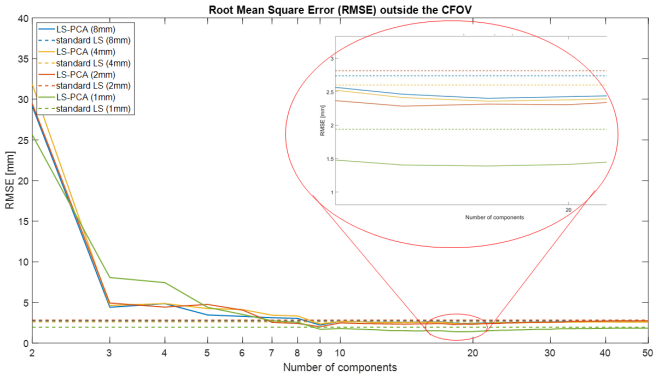
Figure 4.14: *RMSE of ML-PCA (a) and LS-PCA (b) methods as a function of the number of principal components and for different SiPMs sizes. The RMSE is calculated only for events interacting inside the CFOV.*

In the perspective of implementing a higher number of smaller SiPMs to improve spatial resolution, especially outside the CFOV, PCA may allow to damp the unavoidable increase in the computational effort made by the reconstruction algorithm, since it would allow to use a much smaller number of features instead of the original ones.

However, this advantage would not be achieved for free, because it would mean also increasing the number of physical channels to acquire and to weight for the principal components hardware implementation.



(a) ML-PCA



(b) LS-PCA

Figure 4.15: *RMSE of ML-PCA (a) and LS-PCA (b) methods as a function of the number of principal components and for different SiPMs sizes. The RMSE is calculated only for events interacting outside the CFOV.*

4.5 Validation on experimental measurements

The experimental measurements have been carried out directly on the clinical INSERT single detection module, avalaible in Politecnico di Milano.

The same workflow described for simulated datasets has been followed, by acquiring a calibration dataset, used for LRFs estimation and extraction of the principal components, and then validating the ML-PCA and LS-PCA algorithms on a different validation dataset.

SiPMs size	ML	LS
8 mm	0.84 mm	0.90 mm
4 mm	0.77 mm	0.78 mm
2 mm	0.92 mm	0.98 mm
1 mm	0.58 mm	0.60 mm

Table 4.1: *RMSE inside the CFOV varying the size of SiPMs*

SiPMs size	ML	LS
8 mm	2.53 mm	2.74 mm
4 mm	2.45 mm	2.60 mm
2 mm	2.59 mm	2.82 mm
1 mm	1.65 mm	1.94 mm

Table 4.2: *Values of RMSE outside the CFOV varying the size of SiPMs*

4.5.1 Calibration phase

The Co-57 source was positioned at a distance of 30 cm from the box containing INSERT module, in a way to obtain a flood field irradiation. After the energy filtering stage, the same procedure descibed in 3.3.3 has been then followed in order to generate the LRF of every SiPM.

The same calibration dataset has been then used for the principal components estimation.

4.5.2 Validation phase

Being the real interaction coordinates of each detected event unknown, differently for simulations, the reconstruction performances of the ML-PCA and LS-PCA methods have been evaluated in terms of FWHM and image quality, by comparison to the ML and LS methods.

With regard to the number of extracted principal components, simulations results had suggested that a value of d greater than 10 could provide a spatial resolution comparable to the one of classical statistical methods. In order to assess if these results may be consistent or not with the real case scenario, different values of d have been tested, starting from $d = 10$

($d = 10, 20, 30$). In this case, the number of quantization bits N_{bits} has been initially fixed to 10.

The validation dataset consisted of a grid irradiation, obtained by using a Lead collimator, which is showed in figure 4.16.

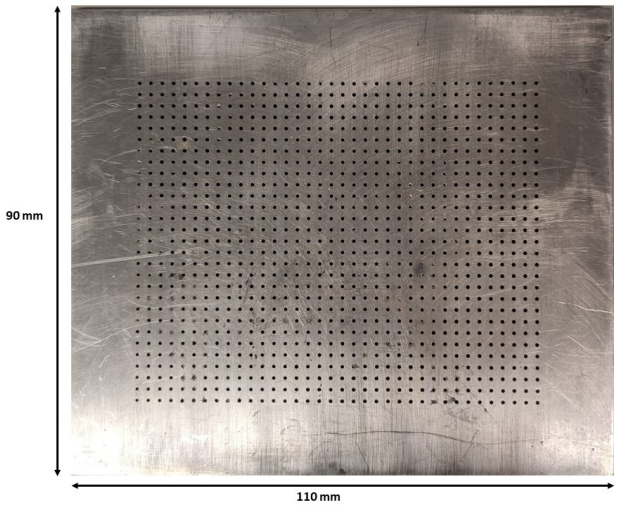


Figure 4.16: *Photo of the Lead collimator used for the experimental validation. The holes of the collimator had 1 mm diameter and the distance between the holes was 3 mm along the two directions x and y.*

In the first three rows of figure 4.17, it is possible to observe the validation grid reconstructed by ML-PCA and LS-PCA algorithms, selecting respectively 10, 20 and 30 principal components, while the row at the bottom shows the same image reconstructed by classical ML and LS algorithms, thus exploiting all the 72 original channels of INSERT clinical module.

It is possible to observe that the first 10 principal components adequately describe only the CFOV of the detector; indeed, the image undergoes a compression and events interacting along the borders are not correctly reconstructed.

Increasing the number of components up to 20 allows to overcome this distorsion and to obtain a grid image practically comparable to the one reconstructed by ML with 72 channels.

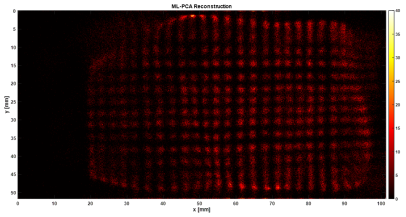
It can also be observed that passing from $d = 20$ to $d = 30$ does not

provide significative improvements on the quality of the image in the more central region of the detector, but has the effect of increasing the density of events for the spots placed immediately near the borders.

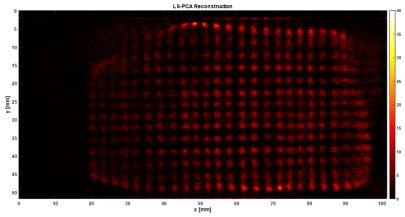
A more quantitative assessment of the spatial resolution provided by the two PCA-based algorithms is shown in figure 4.18. The figure represents the histograms of the reconstructed x and y coordinates for a subset of spots from the validation grid, taken inside the CFOV. Here, the ML-PCA and LS-PCA reconstructions have been run by extracting the first 30 principal components and quantizing them with 10 quantization bits. It is evident how the $FWHM_x$ and $FWHM_y$ are practically equal to the ones of ML statistical reconstruction which employees all the 72 original channels.

Finally, in order to assess if the reconstruction performances in the case of experimental data was robust to a coarser quantization of the loadings of the principal components, the same reconstruction has been run by quantizing the estimated principal components on just 2 bits, thus allowing for each principal components only 4 possible values for its weights.

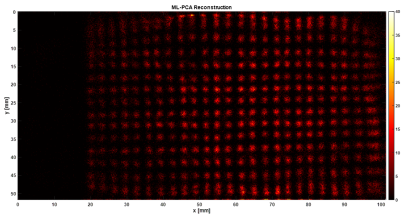
Like for simulations, the spatial resolution seems to be totally unaffected, as it is possible to observe from the histograms depicted in figure 4.19.



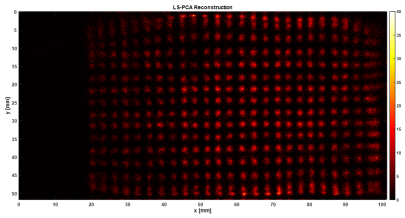
(a) ML-PCA reconstruction, 10 components



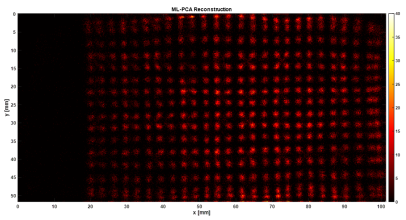
(b) LS-PCA reconstruction, 10 components



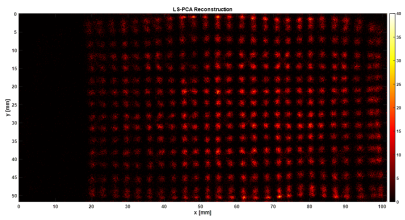
(c) ML-PCA reconstruction, 20 components



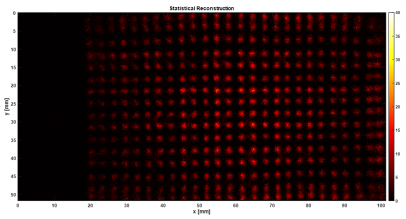
(d) LS-PCA reconstruction, 20 components



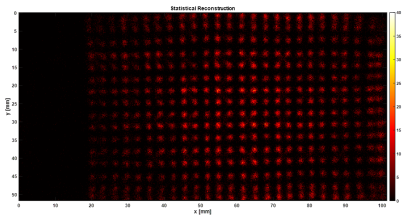
(e) ML-PCA reconstruction, 30 components



(f) LS-PCA reconstruction, 30 components

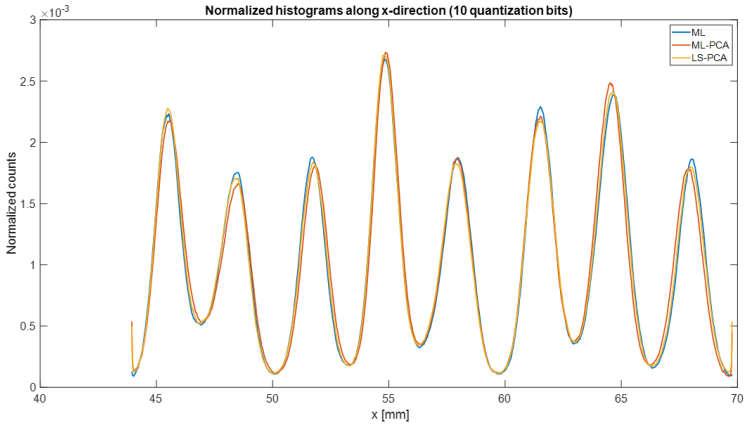


(g) ML reconstruction, 72 components

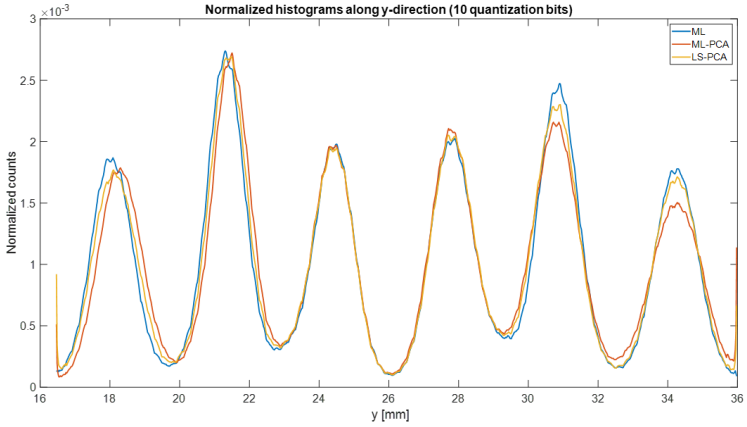


(h) LS reconstruction, 72 components

Figure 4.17: *Grid irradiation reconstructed by ML-PCA (left column) and LS-PCA (right column) with varying the number of principal components d used to run the reconstruction. The last row represents, instead, the same grid irradiation reconstructed by ML and LS without applying the PCA reduction.*

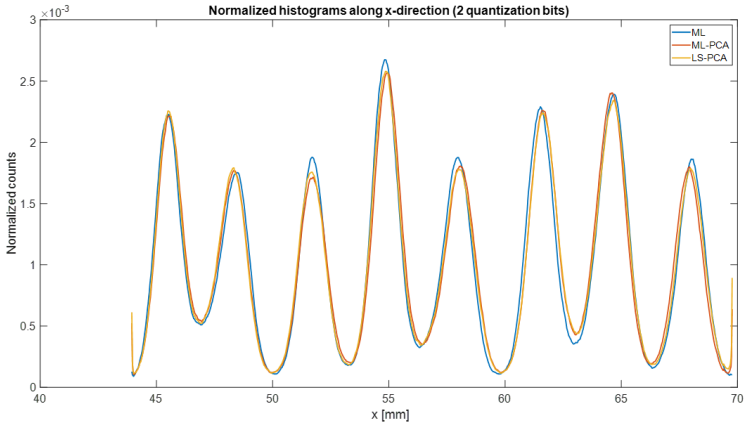


(a) Histograms along x-direction

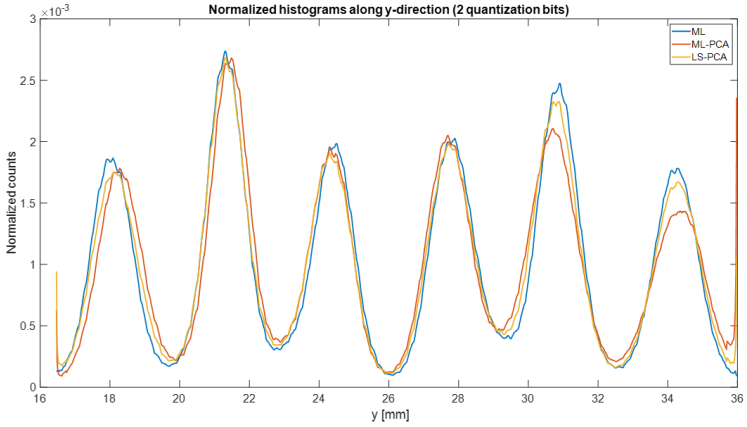


(b) Histograms along y-direction

Figure 4.18: *Smoothed histograms of the reconstructed coordinates respectively along the x and y direction. 30 principal components have been used to run the reconstruction, and the weights of those principal components have been quantized using 10 bits.*



(a) Histograms along x-direction



(b) Histograms along y-direction

Figure 4.19: *Smoothed histograms of the reconstructed coordinates respectively along the x and y direction. 30 principal components have been used to run the reconstruction, and the weights of those principal components have been quantized using 2 bits.*

Chapter 5

Decision trees

One of the most common methods for solving supervised learning problems is constituted by decision trees. In general, decision trees are used for both classification and regression problems; they owe their success to their high flexibility and, above all, their interpretability. In the second part of the present thesis work, a Decision Trees-based reconstruction method has been proposed, by converting the problem of localizing the (x,y) scintillation coordinates of a γ photon into a discrete classification problem, which is then addressed by implementing a cascade of decision trees. The method has been evaluated on both simulations and experimental data. Into this chapter, after an introduction on the theory on which decision trees are based and their applications in medical imaging, the whole process followed in order to create the classification model will be described, from the training process to the hyperparameters optimization and the final results on some validation datasets.

5.1 Introduction to Decision Trees

Decision Trees (DT) constitute a wide family of machine-learning methods, which are used in order to generate predictive models from data, in supervised learning problems. Typical supervised learning problems are classification and regression, where the input dataset includes a certain

number of observations, each consisting in a set of attributes, or features, plus a target variable. The target variable can assume a set of discrete values, in a classification problem, or it can be a continuous variable defined in a certain domain, in a regression problem.

In the following, the discussion will focus on classification trees; indeed, the problem of reconstructing the position of interaction in a gamma-camera can be seen as a classification problem, if the (x,y) position are discretized in a finite set of possible positions.

5.1.1 Theory

In a typical classification problem, the *training dataset*, that is the set of observations needed to make the model "learn" from the past events, consists in N observations.

Each of this observations is made by D predictor variables (or features), X_1,X_2,...,X_D, which can be categorical or numerical, plus a categorical target variable Y, namely a variable which can assume k different values, called classes.

Looking at the problem from a more intuitive point of view, each of the N observations is represented by a point, defined in a D- dimensional space, and characterized by a possible label value from 1 to k.

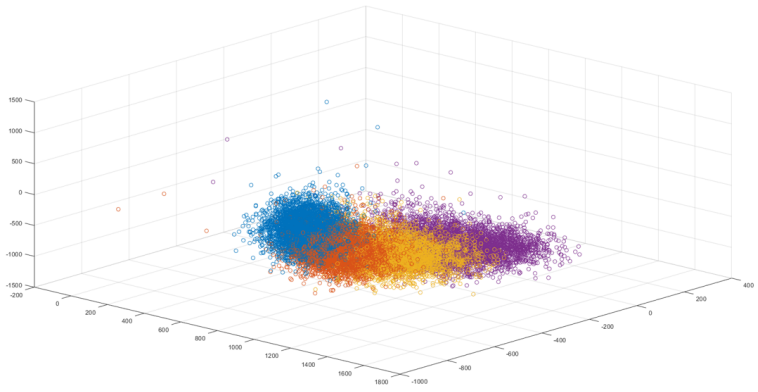


Figure 5.1: *Typical classification problem; each point represents an observation , while its color represents a given class.*

In figure 5.1, it is shown, just as an example, a typical classification problem; each observation is characterized by 3 features values ($D = 3$) and a given class, which is represented by the color of the point.

Each classification method aims to separate as accurately as possible points having different label values. After the generation of the classification model, it is possible to predict the label value of a new observation, just knowing the values of its features.

The simple idea on which a decision tree is based is to apply a recursive partition of the features space into a set of disjoint regions, simply generating some *splits*, which represent tests based on the values of one or more attributes.

A decision tree, whose general structure is depicted in fig.5.2, is made by one node called *root*, with no incoming edges, some *internal nodes*, which are reached by one incoming edge and spread two (or more) outcoming edges, and some *leaf nodes*, which have one incoming edge but no outcoming edges.

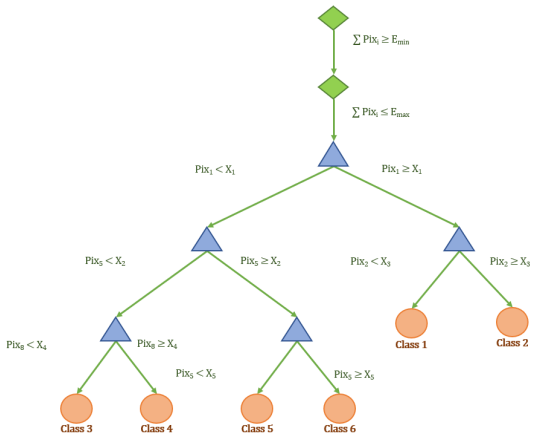


Figure 5.2: *Decision Tree structure*

Each of the nodes of the tree represents a partition of the instances space into two (or more) non overlapping regions of the original space, according to a certain function of the input features, called *splitting rule*. DTs owe their success to the fact that, being characterized by a set of rules structured as a tree, they are easily interpretable and close to the

human reasoning approach [68].

In its simplest form, called *univariate tree*, each split considers a single feature at time, thus the boundary of the obtained sub-regions are rectangular; in univariate trees, the conditions tested by the nodes are also more intuitive and clear to understand.

In the particular case of a dataset with just numerical attributes, the geometrical interpretation of a decision tree becomes a set of hyperplanes, each orthogonal to one of the axes.

When a certain stop condition is reached, the generation of the tree interrupts; all the current branches become leaf nodes and each leaf is assigned to one of the k classes, generally according to a majority criterion (the k-th leaf is assigned to class i if the majority class between the instances which have fallen in leaf k is class i).

After the generation of the tree, the classification of a new unseen observation is straightforward: the tree is navigated from its root node downward, according to the outcome of each tested condition along the path, till a leaf node is reached, and thus the corresponding label will be assigned to the observation. [69]

5.1.2 Training of a decision tree

Training, or induction, of a decision tree is the process which consists in growing the tree, given as input a *training dataset*, namely a dataset containing observations whose target value is known. Starting from a given training set, there is not only a possible decision tree which can be constructed and it has been shown that finding the optimal one, namely the minimum tree consistent with the training set, is a NP-hard problem [70].

This is the reason why the existing training algorithms limit themselves to find a reasonably accurate, despite sub-optimal, decision tree.

In order to train this sub-optimal classifier, empirical methods are used; in particular, these can be distinguished in top-down and bottom-up induction methods.

ID3 [71] , C4.5 [72] and CART [73] are some of the most known top-down induction algorithms, and they are all characterized by a greedy, "divide and conquer" strategy, where at each iteration the most appropriate partition function is choosen according to some *splitting rules* and the procedure stops when a certain *stopping rule* is satisfied.

Splitting rules

Many different criterions have been employed in literature in order to determine the goodness of an attribute test condition, but almost all of them have in common the concept of minimization of the *impurity* of the generated partitions, and so they are called *impurity-based methods*. A perfectly pure node is a node whose observations are all characterized by the same label value; intuitively, a good classifier will be the one which is able to generate,with its splitting rules, purer and purer child nodes.

Given a random variable x, which can assume k discrete values according to the discrete probability function $P = (p_1, p_2, .., p_k)$, an impurity measure is a function $\Lambda : [0, 1]^k \to R$, which satisfies these conditions:

- $\Lambda(P) \geq 0$
- $\Lambda(P)$ is minimum if $\exists\ i$ such that $p_i = 1$
- $\Lambda(P)$ is maximum if $\forall\ i$, $1 \leq i \leq k, p_i = 1/k$
- $\Lambda(P)$ is symmetric with respect to components of P
- $\Lambda(P)$ is differentiable everywhere in its range

The most used impurity measures are listed here in the following [52]:

$$Entropy = -\sum_{i=1}^{k} p_i(t) \log_2 p_i(t) \tag{5.1}$$

$$Gini\ index = 1 - \sum_{i=1}^{k} p_i(t)^2 \tag{5.2}$$

$$Misclassification\ error = 1 - \max_i [p_i(t)] \tag{5.3}$$

where $p_i(t)$ represents the relative frequency of instances of node t belonging to class i and k is the number of total classes.

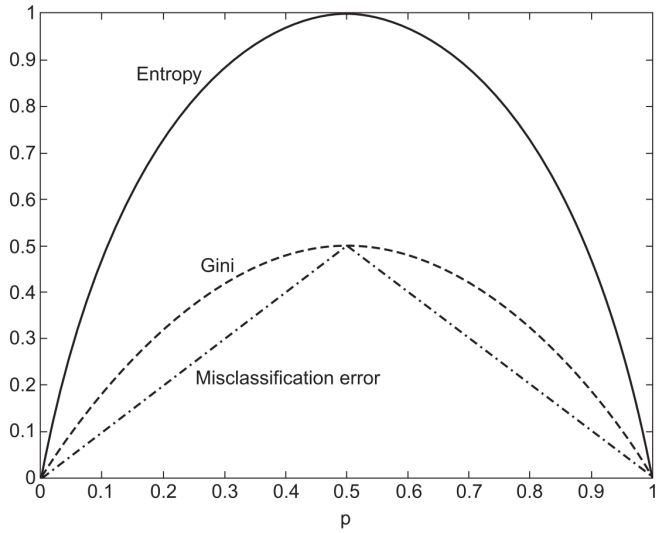


Figure 5.3: *Entropy, Gini and misclassification error for a binary classification problem: it is straightforward to verify that each of the three measures assumes a zero value if the node contains instances from a single class and maximum impurity if the node has equal proportion of instances from the two classes*

In order to evaluate which is the best split to perform for a given node t, the approach is to choose the split condition which leads to the greatest reduction of one of the three impurity indexes. This reduction, also termed *gain* being reflective of the gain in purity achieved after the split, is defined in terms of the difference between the impurity of the parent node t, and the impurity of the child nodes, which are the nodes generated by the partition :

$$\Delta = I(parent) - I(children) \tag{5.4}$$

where $I(children)$ is given by:

$$I(children) = \sum_{j=1}^{c} \frac{N(v_j)}{N} I(v_j) \tag{5.5}$$

where N is the total number of instances contained in the parent node, c is the number of children, $v_1, v_2, .., v_c$, generated by the split, $N(v_j)$ is the number of instances fallen into the child node v_j , and $I(v_j)$ is the impurity related to children node v_j . In other words, the total impurity of the children nodes is given by a weighted sum of the impurities of child nodes. The gain, in the case when entropy is used as impurity measure, is also termed *information gain*.

Another common splitting criterion which is also based on impurity measures is *twoing rule*. Unlike Gini rule which searches in training sample for the largest class and isolate it from the rest of the data, twoing basically looks for two classes that make up together more then 50 % of the data. Twoing splitting rule maximizes the following change of impurity measure [74] :

$$\Delta i(t) = \frac{P_L P_R}{4}\left[\sum_{i=1}^{k}\left|p_i(t_L) - p_i(t_R)\right|\right]^2 \tag{5.6}$$

where P_L and P_R are the probabilities of the left and right nodes.

Impurity measures suffer of one important limitation: they are biased towards features which can assume a large number of distinct values; in other words, they generally tend to choose as splitting features the ones which can assume more possible values, since they bring to a higher information gain, although they result in a low generalized accuracy [71].

For this reason, some ways to "normalize" the impurity-based measures have been proposed, like the *gain ratio* [72].

Gain ratio for the attribute i is defined as:

$$Gain\ Ratio_i = \frac{Information\ Gain_{,i}}{Entropy_{,i}} \tag{5.7}$$

Stopping rules

Stopping rules constitute a set of rules used to determine if it is advisable to stop the grow of a certain branch of the tree or not. Of course, building a too branched tree is counterproductive for mainly two reasons: it

causes overfitting, meaning that it generates a model with a poor generalization capability because too much reflective of training observations, and, secondly, it makes the model difficult to interpret.

Generally, an internal node becomes a leaf node if one of the following conditions is verified:

- the node contains a number of instances lower than a certain fixed treshold
- the percentage of node observations pertaining to the same class is higher to a certain fixed treshold
- the information gain which may derive by an eventual further partition is lower than a certain fixed treshold

Pruning

Finding the ideal stopping rules parameters in order to obtain a balanced tree, neither under- or over-fitted, is not always an easy task.

For this reason, another technique is often used after the generation of the tree, the *pruning*. The tree is initially let grown, even overfitted, but, in a next phase, the tree is cut back into a smaller tree by removing those branches which do not contribute to the generalization accuracy; many different pruning criterions have been proposed throughout the years, like cost-complexity pruning [72], reduced error pruning [75][76] and all these attempts have proved to be useful to reduce the complexity model without loosing in accuracy.

Techniques for increasing the predictive accuracy

Even if properly optimized by tuning the values of some hyperparameters like maximum number of allowed splits or minimum size of the leafs, a single tree may tend to overfit the training data, especially when they are characterized by a limited size.

In order to compensate overfitting and increase the predictive accuracy, different techniques have been investigated throughout the years.

One of them is **Bootstrap aggregating**: Bootstrap aggregating, also known as Bagging, is a machine learning ensemble technique conceived for improving the stability and accuracy of machine learning algorithms used in both classification and regression problems [77][78].

The idea on which bagging is based is quite simple: given a training set $\boldsymbol{X}$ of size N, bagging generates p new training sets $\boldsymbol{X}_i$ (with $i = 1, ..., p$), each having a size N_i (which does not have to be necessarily equal to N). Each of these new training datasets is obtained by sampling N_i observations from $\boldsymbol{X}$ uniformly and with replacement. The strategy implemented in order to generate the p new datasets, also called bootstrap samples, is shown in figure 5.4.

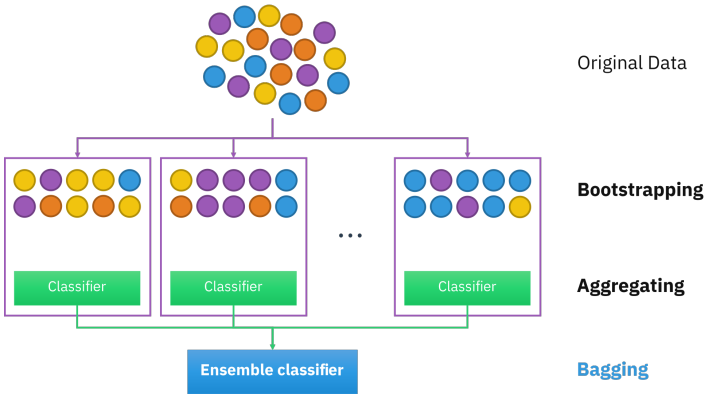


Figure 5.4: *Generation of bootstrap samples. The initial dataset, placed in the top is composed by 20 observations; each bootstrap dataset is obtained by sampling, uniformly and with replacement, 10 observations. The procedure is repeated p times in order to generate p new training datasets.*

By sampling with replacement, some observations will be repeated in each $\boldsymbol{X}_i$ (it can be demonstrated that, with $N_i = N$, almost the 60 % of the observations of a bootstrap sample $\boldsymbol{X}_i$ are unique, while the rest is duplicated [79]).

Besides reducing the variance, sampling with replacement ensures that every bootstrap sample is independent from the others.

After the generation of the p bootstrap samples, p different trees are grown, instead of 1. Finally, in order to classify a new observation, each of the p classification trees will predict its expected class and the final class

to which the observation will be assigned will be chosen according to a majority voting among the p fitted trees.

Another crucial factor influencing the predictive performances of a classification model is the dimension of the training set; it is intuitive that disposing of a small number of training events usually leads to overfit the training data, lacking of generalization capability. An even worse situation is when the number of training observations available for each class is not balanced.

It is also true that not always it is possible to collect a sufficiently high number of training events, especially for applications like nuclear imaging systems, where the generation of training events requires proper calibration procedures on the detector and, thus, increasing the number of training data would require longer calibrations.

One common technique adopted in order to compensate the unbalancement of the size of training events for different classes is **data augmentation**. Data augmentation includes a wide family of techniques, which are aimed at simulating new training data from the available ones, without altering the informations contained in the original data.

Even if the main application of data augmentation is related to images [80], in the sense that some operations like flipping, crossing and rotations are applied on the pixels of the image), in some cases it has been succesfully applied to numerical datasets too [81] [82]. In this context, data augmentation refers to the possibility of generating new training observations coherently to the sample already available, in order to not modify the information contained in the original data.

How it will be further described later, bagging and data augmentation techniques have been applied during the training of the DTs for the experimental measurements, with the purpose of increasing the generalization capability of the classification model.

5.1.3 Applications to medical imaging

Machine learning techniques have found unique applications in nuclear medicine field. Their use in medical imaging can be grouped along three main frameworks:

- **Computer-aided Detection (CAD):** classification models can be used as a decisional support for physicians in order to perform early-stage diagnosis of pathologies. Indeed, the use of classification and pattern recognition algorithms has proved to be a very efficient way to identify suspicious features of a medical image and bring them to the attention of the physician, leading to a decrease in the false negative rate, speeding-up the diagnosis process and, most of all, allowing an early diagnosis of a pathology, which may not be easily detectable by the human eye [83][84] .

- **Image enhancement:** another possible application involves the improvement of the "quality" of medical images, for example the correction for the object attenuation of annihilation photons, scatter correction and noise reduction in PET imaging [85][86]

- **Localization of the (x,y,z) scintillation coordinates:** ML-based classification models can be employed in order to estimate position and energy of the interaction of a γ photon inside monolithic PET/SPECT detectors (in case of PET detector, also the time of arrival of the photon can be estimated) [85].

The reason why ML algorithms have aroused such interest for the estimation of the scintillation coordinates of γ-photons in monolithic detectors is quite simple; solving the inverse problem that maps the light distribution to the position-of-interaction is often quite challenging in the presence of limited-statistics noise and the highly nonlinear behavior near the edges of the crystal. Compared to other estimators such as center-of-gravity or fitting methods, machine learning-based approaches, in general, resulted in superior detector spatial resolution, mainly due to reduced positioning

bias at the edges of the detector, where linear estimation methods such as Anger logic fail to accurately decode the nonlinear light distribution [85].

On the contrary, ML-based methods are based on a "blind" logic, which does not require the definition of particularly rigid statistical models and prescinds from the hypothesis of a linear distribution of the light over the detector.

In particular, considering Decision Trees, DT-based reconstruction techniques have been proposed for the estimation of the scintillation position in PET monolithic detectors [87], even introducing the DOI information [88]. Among the numerous types of classifiers which have been proposed for the estimation of scintillation position in continuous detectors, DTs proved to be an optimal candidate in the perspective of implementing the positioning algorithm into the system electronics, such as FPGA implementations. This is due to the fact that decision trees algorithms rely only on binary decision operations, making them a relatively straightforward and a computationally relaxed algorithm for fast event processing. However, the main limitation to the FPGA implementation of these algorithms proved to be the memory allocation required [87].

5.2 DT-based planar reconstruction

The present thesis work aimed at implementing a DT-based classifier able to reconstruct the (X,Y) scintillation coordinates of a γ photon in a monolithic detector; however, this approach may be equally applied to pixelated detectors too.

5.2.1 Basic principle

The basic idea consists in converting the problem of the reconstruction of the (x, y) interaction coordinates into a discrete classification problem, where each class corresponds to a specific (x, y) position on the crystal, as illustrated in figure 5.5.

The scintillator surface is virtually divided into a number C of classes.

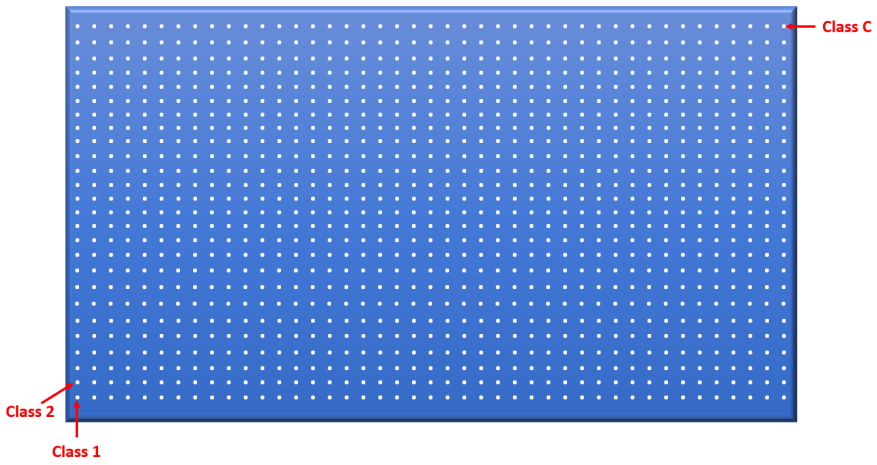


Figure 5.5: *Illustration representing the subdivision of the crystal surface (in blue) into a grid of C discrete (x, y) positions (white dots).*

In order to train the model, for each class, a set of events interacting in the corresponding position needs to be obtained, so that the classifier can "learn" from those data which are the most recurrent patterns of signals values for that specific (x, y) position. The generation of the datasets corresponding to each of the C classes (the white dots in figure 5.5) required different approaches for simulations and experimental measurements, which will be described later.

5.2.2 Cascade of Decision Trees

Differently from statistical methods, which handle the problem of estimating the interaction coordinates in a continuous way, being the $LRFs$ the result of a fitting in the (X, Y) domain, in the case of the DT reconstruction, the possible interaction coordinates are limited to just C possible positions.

As a consequence, the spatial resolution theoretically achievable by the method strictly depends on the inter-classes distance and, therefore, on the number C of classes defined.

Thus, one important point to deal with is the choice of a reasonable number of classes. Indeed, if from one side the minimum spatial resolu-

tion achievable by the algorithm decreases when decreasing the distance between two adjacent classes, from the other side, choosing a too dense pixellation unavoidably leads to an increase of the computational effort and time required for the training of the model and to a worsening in the accuracy of the classifier, which is not able anymore to discriminate correctly events corresponding to two adjacent classes.

Furthermore, implementing a classifier with too many classes proved to be a sub-optimal solution because of the high memory resources required to store a too complex classification tree.

A possible approach to overcome this issue may be, instead, splitting the problem of localizing the (X,Y) scintillation coordinates into two different classification steps:

1. the event is assigned by a first decision tree, which we will refer to as **Global Decision Tree**, to a specific macro-region of the crystal

2. the same event enters into a second decision tree, called **Local Decision Tree**, which is defined inside the macroregion to which the event has been assigned, and this second tree finally assigns the event to a more specific (x, y) position on the crystal among the ones defined inside that region

This strategy is illustrated in figure 5.6.

Thus, according to this strategy, instead of assigning a certain event directly to one of the C classes by means of a huge single tree, the Global Decision Tree (GDT) assigns the event to one of the M macro-regions, providing a first "coarse" classification, while the Local Decision Tree (LDT) of that specific macroregion operates a more fine classification assigning the event to one of the local classes defined within that area.

However, even this second approach presents some limitations. Making the classification a two-steps process may result in a worsening of the classification accuracy: an event which is assigned to a "wrong" macroregion will be unavoidably misclassified at the end of the process, even if the LDT made the most accurate possible classification.

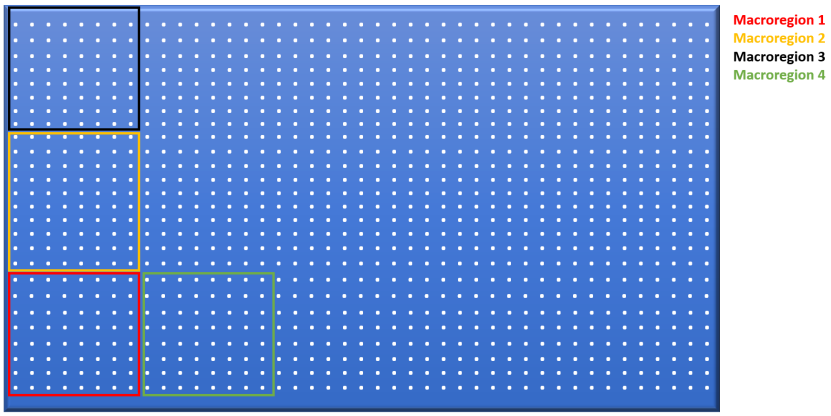


Figure 5.6: *Illustration representing the subdivision of the crystal surface in two different layers: the macroregions, defined by the colored boxes , and the local classes,represented by the white dots, which correspond to the final* (x, y) *allowed positions.*

In order to compensate for this issue, the solution adopted has been to admit overlaps between the macroregions; this practically means that some (x, y) coordinates which are located between two adjacent regions A and B are defined as local classes inside the Local Decision Tree of both A and B regions.

The final architecture of the model is shown in figure 5.7

5.3 DT reconstruction on simulations

The process required in order to generate the training datasets and build the classification model are different for simulations and experimental measurements.

ANTS2 simulation package allows to set all the parameters regarding the radioactive source: shape (point-like, linear or surface) dimension (i.e. the diameter of a round surface), position, orientation and collimation.

This made possible to generate with high flexibility and precision the datasets required to train both the GDT and the LDTs.

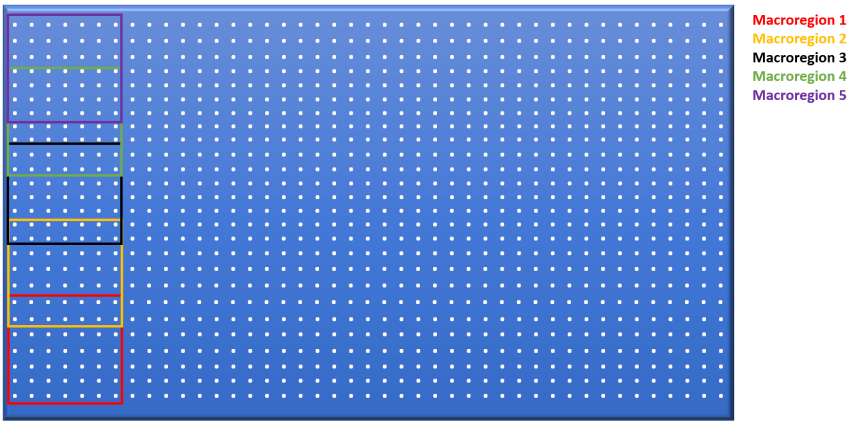


Figure 5.7: *Illustration representing the final architecture used for the classification model. The macroregions, defined by the colored boxes, are overlapped between each other in the vicinity of their respective boundaries, along both x and y directions.*

5.3.1 Generation of the training datasets

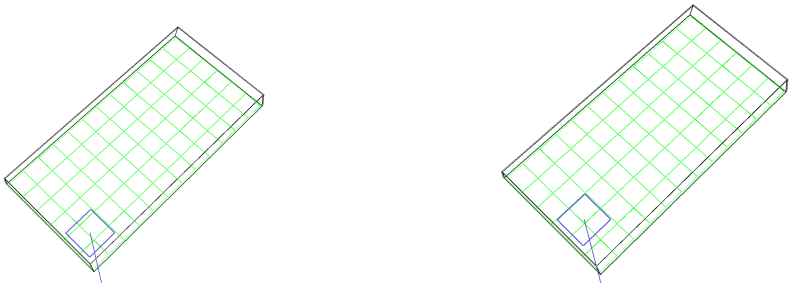
The architecture choosen in the case of simulations consisted in 105 square-shaped macroregions, with size 12.6 mm and with a reciprocal overlapping equal to half the size of each macroregion. The number of classes inside each macroregion has been set to 42 along the x direction and 42 along the y. The distance between the local classes has been set to 0.3 mm.

In order to train the whole classification model, two different types of training datasets need to be obtained:

- training dataset of the **Global Decision Tree**
- training datasets of the **Local Decision Trees**

In order to train the GDT, for each of the 105 macroregions (colored boxes in figure 5.7), a set of events interacting inside that specific region needed to be simulated.

At this purpose, a square-shaped γ source with size 12.6 mm, shown in figure 5.8, has been defined on ANTS2 and 105 simulations have been run, in order to obtain the training events for each region.



(a) Positions of the source for simulating events of macroregion 1

(b) Positions of the source for simulating events of macroregion 2

Figure 5.8: *Events simulated for the macroregions 1 and 2. The square-shaped source is shifted every time of steps equal to 6.3 mm, namely half the size of each region, in a way that all the 105 final macroregions are overlapped between each other for half their size.*

The training process for the 105 Local Decision Trees is, instead, quite different.

Each class defined inside a local tree represents a given (x, y) position (white dots in figure 5.7); in order to simulate the events corresponding to each class, a circular γ source with 0.3 mm diameter has been defined on ANTS2 and shifted along the x and y directions with a 0.3 mm step size. For each position of the beam, 500 events have been simulated.

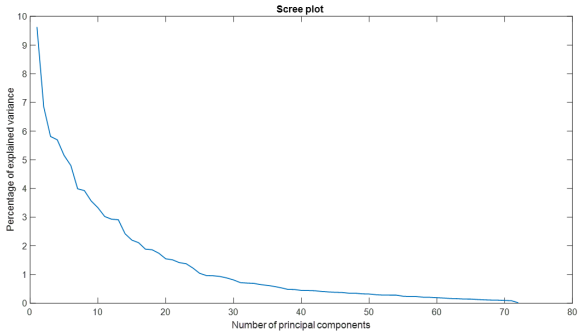
In all the simulations run on ANTS2, dark count rate (DCR) was set to 25 $\frac{\text{kHz}}{\text{mm}^2}$ (DCR at T=-10°C).

5.3.2 Feature reduction and PCA

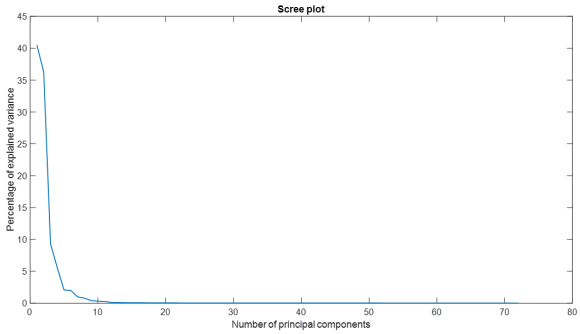
As already described in chapter 4, Principal Component Analysis represents an eligible tool for performing *feature reduction*.

In order to slim the training process, a PCA feature reduction has been performed on the training datasets of the GDT and the 105 LDTs.

In other words, before growing each tree, the corresponding input training dataset has undergone feature reduction, by obtaining a new



(a) Scree plot for the GDT



(b) Scree plot for the LDT

Figure 5.9: *Scree plots for the training dataset of the Global Decision Tree (a) and one of the Local Decision Trees. It is possible to observe that, in the case of the Local Tree, much less principal components are needed in order to describe the same amount of variance.*

set of features. The number of principal components has been choosen according to the *elbow method*, that consists in looking for an "elbow", namely an inflexion point, in the scree-plot [89].

In figure 5.9 are shown the scree plots respectively for the training sets of the GDT and one of the LDTs (the plot in figure 5.9(b) refers to macroregion 5, but the same curve has been plotted for each of the 105 macroregions, leading to equal results). It is interesting to observe how the number of principal components corresponding to the elbow of the curve is significantly different; in the case of LDTs, the highest percentage of the total variance contained in the training set is concentrated in the

very first principal components (up to 10), while for the training set of the global tree, the elbow corresponds to a higher number of components.

This is reasonable if we think that the training events of a Local Tree are events interacting in a very limited portion of the crystal surface (12×12 mm); this means that, among the original 72 detection channels, only few of them are carrying a really useful information about the data.

After the evaluation of the two scree plots, the number of principal components employed in order to train the trees has been set to 35 and 10, respectively for the GDT and the LDTs.

5.3.3 Training of the DTs

Once simulated the training datasets and performed the feature reduction, the process of generation of the GDT and the LDTs has been carried out.

DTs have been trained by using MATLAB "Statistics and Machine Learning Toolbox".

When training a DT, the critical parameters to tune in order to maximize the accuracy, also called *hyperparameters* are:

- maximum number of splits allowed during the growth of the tree; this parameter must be limited in order to prevent overfitting.
- splitting criterion, which can be choosen between Gini index, maximum deviance and twoing rule.
- minimum leaf size, which represents the minimum number of observations for a leaf node. Also this parameter needs to be tuned in order to prevent overfitting.

For both the GDT and the 105 LDTs, the type of splitting criterion proved not to be influential in terms of accuracy; **twoing rule** has been choosen as splitting rule.

Instead, the choice of the other two optimal hyperparameters, **maximum number of splits** and **minimum leaf size**, has been determined by a Bayesian optimization process, aimed at minimizing a 10-fold *cross-validation loss* function.

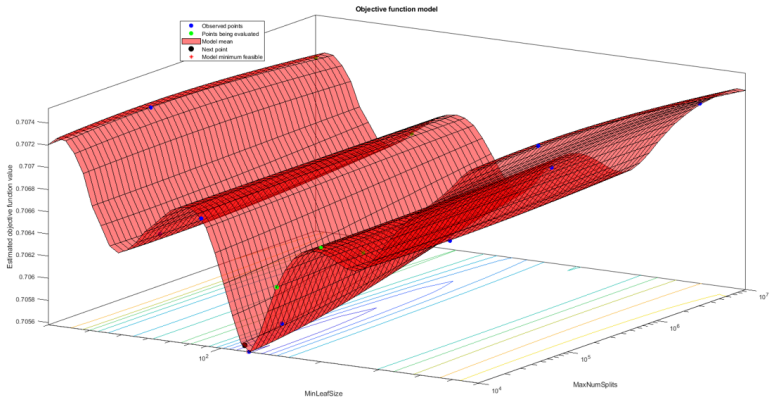


Figure 5.10: *10-fold cross validation loss function evaluated as function of the maximum number of splits and the minimum leaf size, in the case of the Global Decision Tree. The estimated optimal hyperparameters are MaxNumSplits=10206 and MinLeafSize=121*

When setting up a cross validation, the training observations are split into K partitions, the model is trained on $K-1$ partitions, and the test error is predicted on the left out partition k. The process is repeated for $k = 1, 2 \dots K$ and the result is then averaged.

This average classification error takes the name of cross-validation loss.

In figure 5.10 is depicted the 10-fold cross validation loss function computed during the hyperparameters optimization process for the Global Decision Tree.

The hyperparameters optimization process is performed in the same manner for each of the 105 Local Decision Trees, by obtaining analogue curves.

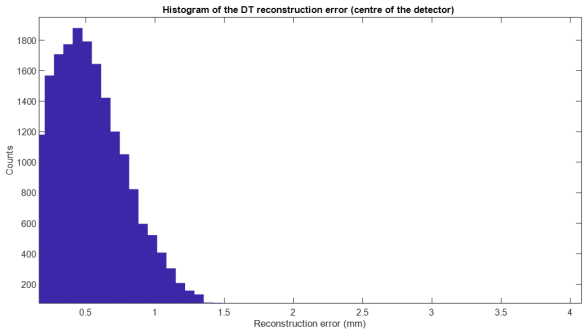
5.3.4 Validation

In order to evaluate the spatial resolution achieved by the algorithm on a new validation dataset, a flood irradiation has been simulated on ANTS2 and each event has been reconstructed with the cascade of decision trees. Being the real interaction coordinates of each validation event known, the absolute error, namely the Euclidean distance between the real coordinates

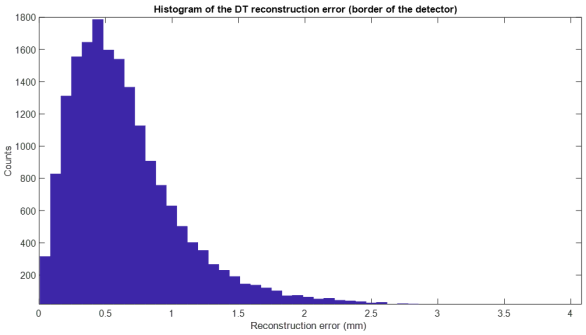
of each event and the reconstructed ones, have been defined and plotted in a histogram.

This has been done separately for the CFOV region and the one outside the CFOV.

The two histograms are depicted in figure 5.11



(a) Error inside the CFOV



(b) Error outside the CFOV

Figure 5.11: *Histograms of the DT reconstruction error inside (a) and outside the CFOV (b)*

It is possible to observe from 5.11(a) that the distribution of the absolute error inside the CFOV has a mean value equal to 0.6 mm (against the 0.8 mm provided by the ML reconstruction).

Instead, the distribution of the absolute error outside the CFOV has a mean value equal to 1.1 mm (while ML provided a mean error of 2 mm).

5.4 DT reconstruction on experimental measurements

In the final part of the work, DT reconstruction method has been implemented for the reconstruction of experimental measurements performed on INSERT clinical module. Even if the principle on which the method is based is exactly the same of simulations (section 5.2), the procedure required in order to generate the training datasets from the experimental measurements on INSERT module required a proper calibration procedure which will be introduced in the following subsection.

5.4.1 Generation of the training datasets

The generation of the training events for each of the C defined (x, y) coordinates may be, in principle, executed by scanning a collimated thin beam across the scintillator surface, which, by the way, is a common calibration procedure for the characterization of gamma detectors response [63].

However, scanning the beam to cover all the C positions defined in the classification model would be expensive in terms of required calibration time and, thus, hardly doable when prospecting a clinical application.

A possible alternative to this type of calibration is using, instead, a grid collimator; this solution would allow to obtain with a single irradiation the training events of multiple classes.

However, the use of a grid collimator instead of a single spot-beam, introduces two important drawbacks:

- the spatial resolution ideally achievable depends on the pitch of the collimator; if a collimator with step k is used, even with an accuracy of 100% , the classifier would not be able to achieve a spatial resolution lower than k mm.

  In order to achieve a spatial resolution feasible with medical diagnostic applications (< 1 mm), collimators with very low pitch should be used.

- in the second place, when performing a grid irradiation, the matrix of detected events does not contain any information about the correspondence between each event and the hole of the collimator from which it has hit the crystal surface.

  This information is crucial in order to assign that event to the training set for a specific (x, y) position on the detector surface.

In the present work, the two issues have been addressed with the following approaches:

- instead of using an extremely low-pitch collimator, a collimator with pitch 2 mm has been used and the classes sampling has been thicken up by performing, instead of a single grid irradiation, multiple irradiations. For each of the irradiations, the collimator has been shifted with step 0.5 mm in order to cover the intermediate positions between two adjacent holes of the collimator, as depicted in figure 5.12. A total of 16 (4 steps along x and 4 along y) different acquisitions have been collected.

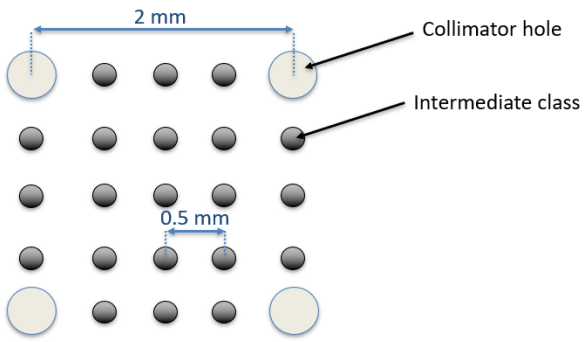


Figure 5.12: *Scheme representing the relative positions of the classes which have been obtained shifting the collimator along the* (x, y) *plane.*

- for each of the 16 grid datasets collected, each detected event has been assigned to a given hole of the collimator with a hybrid technique, combining a statistical reconstruction and a k-means clustering.

The Lead collimator used for the measurements had a size of 60×60 mm, with thickness 0.5mm; its holes had 0.5 mm diameter and a pitch of 2 mm along the x and y directions. It has to be specified that the collimator was able to cover above half of INSERT clinical module. Consequently, it has not been possible to define the classification model on the whole surface of the crystal, but just on one half; in every case, the results obtained on the validated area would be easily repeatable on the whole surface.

Moving the collimator with a high accuracy is crucial in order to ensure a 0.5 mm inter-classes distance. For this reason, the collimator has been shifted along the (x, y) plane by using a linear translator with micrometric resolution [90].

An illustration of the whole setup used for the calibration process is shown in figure 5.13.

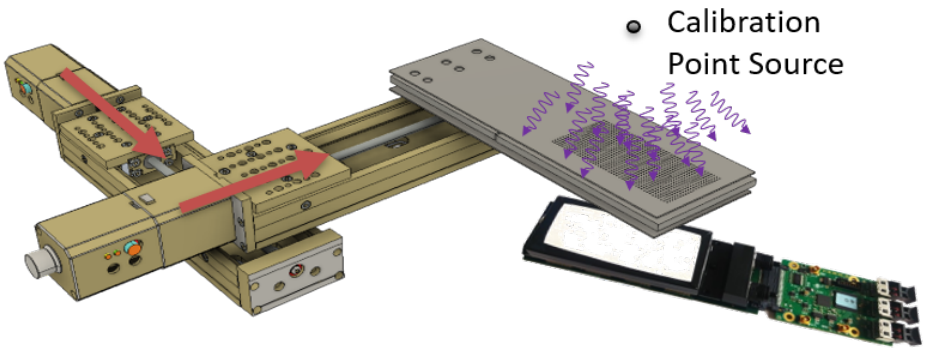


Figure 5.13: *Illustrations representing the experimental setup used for the calibration, with the $x - y$ linear translator and the collimator, which is placed between the Co57 point source and INSERT clinical module.*

Assignment of each calibration event to a (x, y) position

After the collection of the 16 grid-collimated irradiations, the following step has consisted in assigning each calibration event to the training set of a specific (x, y) position on the crystal surface.

In a first step, each acquisition has been energy-filtered, in order to discard Compton interactions due to the presence of the Lead collimator, by

selecting an energy window centered around the Co57 122 keV photopeak.

Then, each calibration dataset has been reconstructed by means of Maximum Likelihood statistical reconstruction (after having estimated the LRFs according to the traditional process already described in 3.3.3). Events interacting at high DOI have been identified, by defining a threshold on the number of activated SiPMs, and filtered out. This last operation was aimed at removing from the grid image the DOI-dependent artefacts.

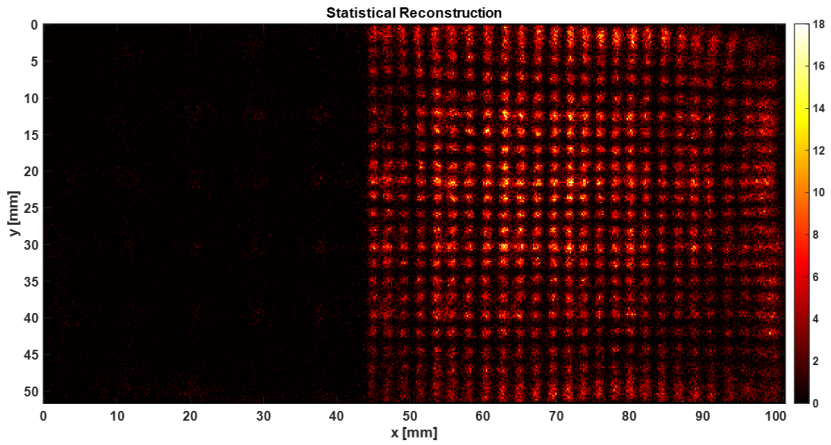


Figure 5.14: *Grid irradiation reconstructed by ML*

Figure 5.14 shows the ML-reconstructed image of one of the 16 calibration datasets. In order to separate calibration events interacting through different holes of the collimator, the (x, y) coordinates corresponding to the peaks of the 2D histogram have been identified. It has to be underlined that the holes coordinates identified from the ML reconstructed image do not reflect exactly the same 2mm inter-holes distance of the physical collimator, especially when moving outside the more central region of the detector. This is due to an intrinsic characteristic of ML reconstruction, which provokes a slight but noticeable stretching of the image towards the borders, as a consequence of the attempt to recover the FOV lost by centroid method reconstruction.

After the identification of the position of each hole from the ML image,

each calibration event has been assigned to one of the holes, and consequently one of the (x, y) coordinates defined in the classification model, by using a k-means clustering method (where k has been set equal to the number of holes).

Figure 5.15 shows the result of the clustering operation, which finally lead to the assignment of each grid event to a specific local class.

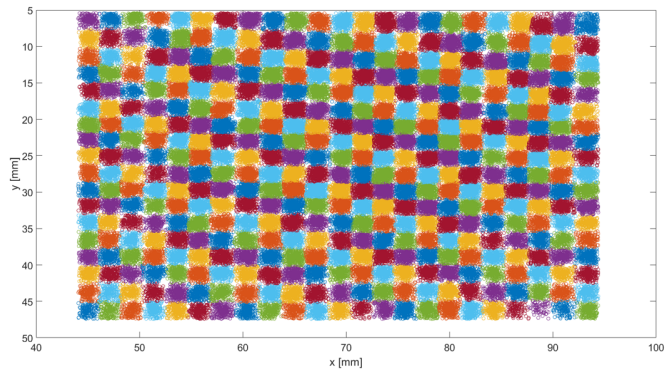


Figure 5.15: *Assignment of each calibration event to a specific hole of the collimator, according to a k-means clustering method. Each color represents a different cluster (colors are repeated only for visualization issues).*

Data augmentation

After the generation of the spot-datasets, one concern has been compensating for the unbalancements in the number of training events for each class. Indeed, how it is possible to observe from figure 5.14, the density of events for different spot positions is not equal; this is due to the different source collimation angle for a hole in the center of the detector and one along the border, or simply to partial occlusions of some holes of the collimator.

However, having the same number of training observations for each class proved to be crucial to avoid overfitting in classification problems.

For this reason, in order to dispose of the same number of training events for each class (1000 events for class) **data augmentation** has been exploited.

For each class, additional training events have been generated according to the following steps:

- computation of the means μ_j and standard deviations σ_j of each channel ($j = 1, ..., 72$) among all the events contained in that specific class
- generation of a new training event x_{new}, where the value of the signal detected by the j-th channel is randomly sampled from a Gaussian distribution with mean μ_j and standard deviation σ_j

Generation of the training set for the GDT

The organization of the two layers of trees in the case of experimental measurements was different from the one adopted for simulations.

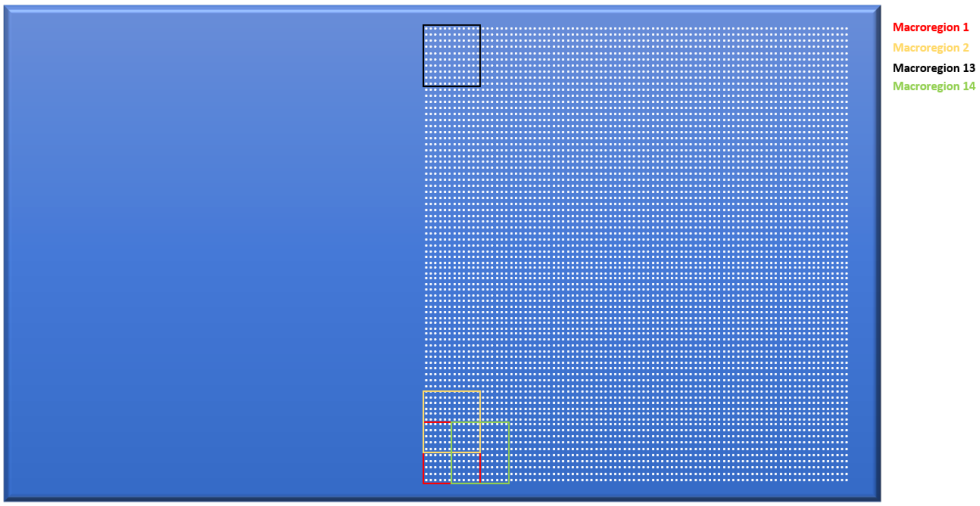


Figure 5.16: *Organization of the two layers of trees. The GDT's classes are represented by the 221 macroregions (colored boxes in the figure). Inside each macroregion, 100 local classes are defined. Adjacent macroregions are overlapped between each other, having in common 5 rows of local classes. Notice that the classification model is defined only on half of the crystal surface.*

In this case, 221 macroregions (13 along the y direction and 17 along the x direction) have been defined; each macroregion included 100 local classes (10 along the y direction and 10 along the x direction). Being the

distance between two local classes 0.5 mm, the dimension of each macroregion was 5mm × 5mm. Furthermore, each macroregion was overlapped with the adjacent one for half of its local classes, corresponding to 2.5 mm.

The architecture of the trees in the case of experimental measurements is represented in figure 5.16.

The procedure followed in order to obtain the training set for the Global Decision Tree consisted, instead, simply in summing up all the spot-datasets defined inside the 221 macroregions and assigning them the same label.

5.4.2 Feature reduction and PCA

Analogously to the case of simulations, after the collection of the training datasets, a PCA feature reduction has been implemented.

The evaluation of the scree plots, depicted in figure 5.17 suggested the adoption of the first 20 and 7 principal components, respectively for the GDT and the LDTs.

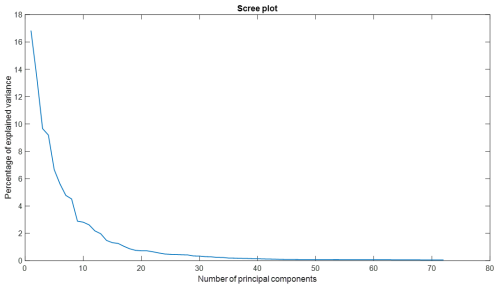
5.4.3 Training of the DTs

Like in the case of simulations, once fixed the number of principal components to use in input to the GDT and the LDTs, the phase of training and hyper-parameters optimization has been carried out.

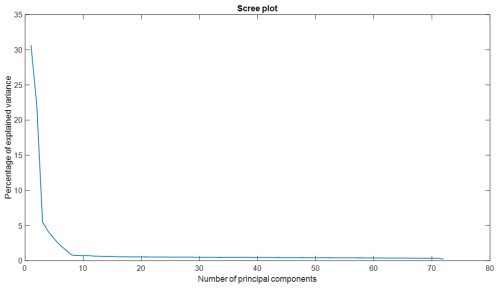
The choice of the two main hyperparameters, **maximum number of splits** and **minimum leaf size**, has been determined by a Bayesian optimization process, aimed at minimizing a 10-fold *cross-validation loss* function.

In order to improve the predictive accuracy of the model, the **bagging** technique has been implemented for growing the LDTs.

In particular, once found the set of optimal hyperparameters with the optimization process described above, instead of growing one single LDT for each macroregion, 20 bootstrapped samples have been obtained sampling with replacement from the original training set, and each of them has been used to grow a distinct LDT for that specific macroregion.



(a) Scree plot for the GDT



(b) Scree plot for the LDT

Figure 5.17: *Scree plots for the training dataset of the Global Decision Tree (a) and one of the Local Decision Trees, in the case of experimental data. Similarly to simulations, in the case of the Local Tree, much less principal components are needed in order to describe the same amount of variance.*

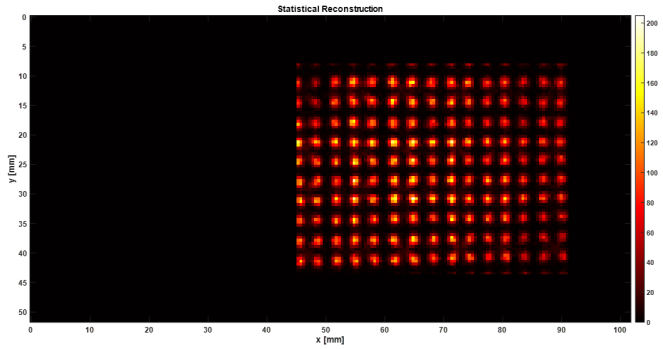
Consequently, in phase of prediction, the event i is classified separately by each of the 20 bagged trees and finally it is assigned to a given class, according to a majority voting among the 20 bagged trees. This technique proved to improve the generalization capability and, thus, the accuracy of the classification.

5.4.4 Validation

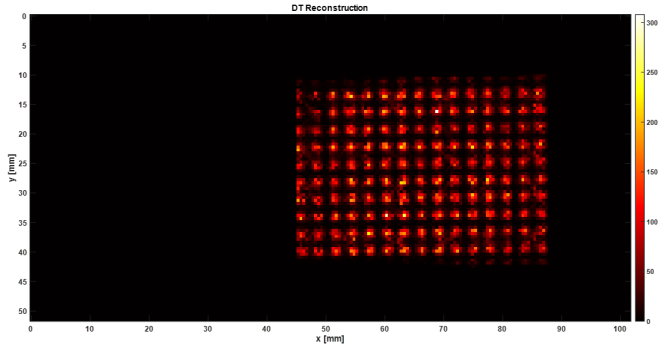
In order to validate the spatial resolution of the technique, the DT-reconstruction has been run on different validation datasets. The reconstructed image has been compared with the one reconstructed by ML statistical method.

The first validation dataset consisted in a grid irradiation performed by using a Lead collimator with pitch 3 mm along the x and y directions,

with hole diameter 1mm (illustrated in figure 4.16). Figure 5.18 shows the image reconstructed by ML and DT reconstruction.



(a) ML reconstruction

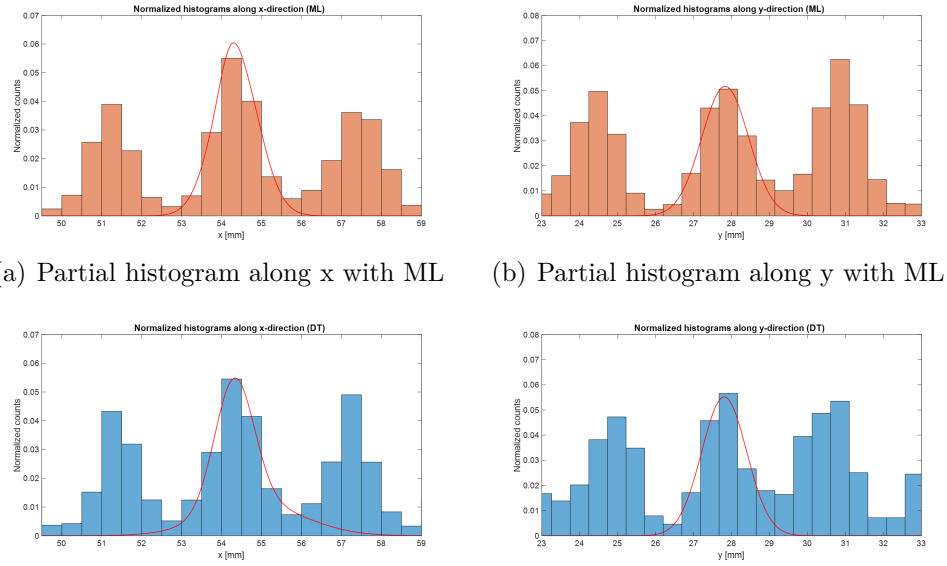


(b) DT reconstruction

Figure 5.18: *Grid irradiation reconstructed with ML and DT.*

The pixel dimension in each of the images is set to 0.5 mm, which corresponds to the distance between the classes defined in the classification model.

It is possible to observe that the DT-reconstructed image is able to identify correctly the pattern of the grid; furthermore, differently from ML, it does not stretch the image toward the borders. The physical distance of 3 mm between the holes of the collimator is preserved. From a qualitative assessment of the image, it is possible to observe that the width of the



(a) Partial histogram along x with ML

(b) Partial histogram along y with ML

(c) Partial histogram along x with DT

(d) Partial histogram along y with DT components

Figure 5.19: *Normalized partial histograms of the reconstructed x (left column) and y (right column) coordinates for a subset of spots of the grid. ML histograms are depicted in orange and the DT histograms in blue.*

reconstructed spot is practically equal in the two reconstructions.

This can be assessed also from figure 5.19, which shows the partial histograms of the reconstructed x and y coordinates for a subset of spots taken inside the CFOV.

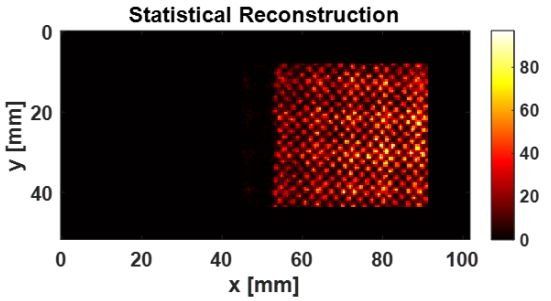
In order to be able to properly compare the $FWHM_x$ and $FWHM_y$ of the two techniques, the displacement between the spot positions detected by ML and DT, due to the stretching effect of ML, has been corrected. The $FWHM$ along the x and y directions proved to be practically equal in the two types of reconstruction and, respectively, equal to 1.4 mm and 1.3 mm.

Finally, a second validation dataset has been collected, by placing the same collimator employed for the calibration (2mm pitch) with a tilted orientation respect to the x and y direction.

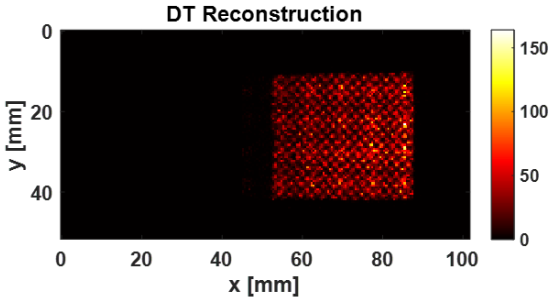
The reconstructed image is depicted in figure 5.20.

It is possible to observe that DT-reconstruction is able to reconstruct

correctly the grid pattern, even though the (x, y) positions defined in the classification model lay on axes parallel to the x and y directions.



(a) ML reconstruction



(b) DT reconstruction

Figure 5.20: *Grid irradiation reconstructed with ML and DT.*

Chapter 6

Conclusions

This thesis project places itself in a framework consisting in the exploration of two strategies for the event reconstruction of γ photons in Anger cameras:

- the former consisted in the integration of Principal Component Analysis, a well-known dimensionality reduction technique, with two statistical reconstruction methods already in use for event reconstruction in Anger cameras, the Maximum Likelihood (ML) and the Least Square estimation (LS) estimation.

- the latter, instead, solves the problem of estimating the planar interaction coordinates of the γ photon inside the scintillation crystal by defining a classification problem, where each class corresponds to a specific (x, y) coordinate.

As to the first framework, the obtained results suggest that PCA constitutes a valid and efficient strategy to reduce the number of features to process by the reconstruction algorithm, without affecting the spatial resolution performances. Indeed, the spatial resolution of ML and LS statistical methods operating in the space of the principal components is practically the same of the classical ML and LS estimation, if a number of principal components sufficiently high is extracted; this stands not only in the case of simulated data, but also for experimental measurements.

In particular, in the specific case of INSERT clinical detection module, already 20 principal components proved to contain enough information to equalize the performances corresponding to the 72 original channels.

This represents a promising result in the perspective of a hardware implementation of a PCA-based multiplexing scheme of the photodetectors signals; furthermore, the PCA-reconstruction proved to be robust to a coarser quantization of the weights of the principal components.

In addition, simulations suggested that the spatial resolution theoretically achievable could be improved by using smaller SiPMs, especially along the borders; in that case, the advantage of scaling down the number of channels in input to the DAQs could be even more evident.

Equally encouraging results came from the event reconstruction implemented with Decision Trees. DTs classifiers provided a spatial resolution comparable to the one obtained by statistical methods. Moreover, using a decision tree for the event reconstruction may significantly reduce the computational time needed to reconstruct the position of an event, eventually making possible even a real-time implementation.

Even though the DTs training procedure involves a first step which employes the reconstruction of the calibration datasets by means of a statistical reconstruction, the latter is used only in order to assign a label to each training event. After the generation of the training dataset for each class, the DT classifier operates autonomously according to a totally different logic in order to reconstruct the position of an event and, consequently, it is able to overcome the intrinsic limitations of statistical methods, such as the phenomenon of stretching of the image toward the borders, which instead is an intrinsic limit of ML reconstruction.

However, the dependance of the training procedure on the statistical reconstruction stage presents some limitations, for example the classification of events along the borders. Indeed, the DT classification model has not been defined in the region immediately close to the borders of the detector, since it is impossible to discriminate events interacting in different holes of the collimator, by using the ML reconstruction.

Thus, a possible goal for the near future may be to investigate possible

solutions for carrying out the training of the trees along the borders.

In conclusion, data augmentation proved to be an efficient strategy in order to obtain new training data on the basis of the data effectively measured, allowing to improve the generalization capability of the model. This technique deserves further investigations, because it may allow to decrease the duration of the calibration procedure, which currently constitutes maybe the main limitation of the reconstruction method for clinical applications.

Bibliography

[1] Koji Iwata, R. G. Greaves, and C. M. Surko. "γ-ray spectra from positron annihilation on atoms and molecules". In: *Physical Review A - Atomic, Molecular, and Optical Physics* (1997). ISSN: 10941622. DOI: 10.1103/PhysRevA.55.3586 (cit. on p. 3).

[2] G. Nelson and D. Reilly. "Gamma-ray interactions with matter". In: *Passive Nondestructive Analysis of Nuclear Materials* (1991) (cit. on p. 3).

[3] Glenn F Knoll. *Radiation detection and measurement*. John Wiley & Sons, 2010 (cit. on pp. 5, 29, 51).

[4] D. A. Torigian et al. "Functional Imaging of Cancer with Emphasis on Molecular Techniques". In: *CA: A Cancer Journal for Clinicians* (2007). ISSN: 0007-9235. DOI: 10.3322/canjclin.57.4.206 (cit. on p. 8).

[5] D. W. Townsend. *Multimodality imaging of structure and function*. 2008. DOI: 10.1088/0031-9155/53/4/R01 (cit. on pp. 9, 10).

[6] S. Vandenberghe et al. *Recent developments in time-of-flight PET*. 2016. DOI: 10.1186/s40658-016-0138-3 (cit. on p. 10).

[7] Magdy M. Khalil et al. "Molecular SPECT Imaging: An Overview". In: *International Journal of Molecular Imaging* (2011). ISSN: 2090-1712. DOI: 10.1155/2011/796025 (cit. on p. 10).

[8] IAEA (International Atomic Energy Agency). *X-ray and Gamma-ray decay data*. URL: https://www-nds.iaea.org/xgamma_standards/ (visited on Mar. 2021) (cit. on p. 12).

[9] Penelope Bouziotis and Carlo Fiorini. *SPECT/MRI: dreams or reality?* 2014. DOI: 10.1007/s40336-014-0095-6 (cit. on p. 13).

[10] Hal O. Anger. "Scintillation camera". In: *Review of Scientific Instruments* (1958). ISSN: 00346748. DOI: 10.1063/1.1715998 (cit. on p. 14).

[11] Simon R. Cherry. *Multimodality Imaging: Beyond PET/CT and SPECT/CT*. 2009. DOI: 10.1053/j.semnuclmed.2009.03.001 (cit. on p. 22).

[12] Osman Ratib and Thomas Beyer. *Whole-body hybrid PET/MRI: Ready for clinical use?* 2011. DOI: 10.1007/s00259-011-1790-4 (cit. on p. 22).

[13] European Union Seventh Framework Programme. *Development of an integrated SPECT/MRI system*. URL: http://www.insert-project.eu/ (visited on June 2019) (cit. on p. 24).

[14] Debora Salvado et al. "Development of a Practical Calibration Procedure for a Clinical SPECT/MRI System Using a Single INSERT Prototype Detector and Multimini Slit-Slat Collimator". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* (2018). ISSN: 2469-7311. DOI: 10.1109/trpms.2018.2828163 (cit. on p. 27).

[15] Michele Occhipinti. "Development of a gamma-ray detection module for multimodal SPECT/MR imaging". PhD thesis. Italy, 2015 (cit. on pp. 30, 52, 66).

[16] M. Marisaldi et al. "Silicon drift detectors coupled to CsI(Tl) scintillators for spaceborne gamma-ray detectors". In: *Nuclear Physics B - Proceedings Supplements* (2006). ISSN: 09205632. DOI: 10.1016/j.nuclphysbps.2004.06.008 (cit. on p. 32).

[17] M Moszynski et al. "A Comparative Study of Silicon Drift Detectors With Photomultipliers, Avalanche Photodiodes and PIN Photodiodes in Gamma Spectrometry With LaBr _{3} Crystals". In: *IEEE Transactions on Nuclear Science* 56.3 (2009), pp. 1006–1011 (cit. on p. 33).

[18] A. V. Akindinov et al. “New results on MRS APDs”. In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (1997). ISSN: 01689002. DOI: 10.1016/S0168-9002(96)01201-6 (cit. on p. 35).

[19] Adam Nepomuk Otte et al. “Characterization of three high efficiency and blue sensitive silicon photomultipliers”. In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (2017). ISSN: 01689002. DOI: 10.1016/j.nima.2016.09.053. arXiv: 1606.05186 (cit. on pp. 37, 44).

[20] Claudio Piemonte et al. “Performance of NUV-HD Silicon Photomultiplier Technology”. In: *IEEE Transactions on Electron Devices* (2016). ISSN: 00189383. DOI: 10.1109/TED.2016.2516641 (cit. on p. 37).

[21] William G. Oldham, Reid R. Samuelson, and Paolo Antognetti. “Triggering Phenomena in Avalanche Diodes”. In: *IEEE Transactions on Electron Devices* (1972). ISSN: 15579646. DOI: 10.1109/T-ED.1972.17544 (cit. on p. 41).

[22] N Otte. “The Silicon Photomultiplier-A new device for High Energy Physics, Astroparticle Physics, Industrial and Medical Applications”. In: *Proceedings to SNIC symposium (SLAC, Stanford, ...* (2006) (cit. on pp. 42, 43).

[23] ON Semiconductor (formerly SensL). *C-Series SiPM Sensors datasheet*. URL: https://www.onsemi.com/pub/Collateral/MICROC-SERIES-D.PDF (visited on Dec. 2020) (cit. on pp. 43, 44).

[24] AdvanSiD. *NUV SiPMs datasheet*. URL: http://advansid.com/attachment/get/up_28_1432731773.pdf (visited on Dec. 2020) (cit. on pp. 43, 44).

[25] Alberto Gola et al. *NUV-sensitive silicon photomultiplier technologies developed at fondazione Bruno Kessler*. 2019. DOI: 10.3390/s19020308 (cit. on p. 43).

[26] S. M. Sze and Kwok K. Ng. *Physics of Semiconductor Devices: Third Edition*. 2006. ISBN: 0471143235. DOI: 10.1002/9780470068328 (cit. on p. 44).

[27] C. Piemonte et al. "Recent developments on silicon photomultipliers produced at FBK-irst". In: *IEEE Nuclear Science Symposium Conference Record*. 2007. ISBN: 1424409233. DOI: 10.1109/NSSMIC.2007.4436565 (cit. on p. 44).

[28] G. A.M. Hurkx, D. B.M. Klaassen, and M. P.G. Knuvers. "A New Recombination Model for Device Simulation Including Tunneling". In: *IEEE Transactions on Electron Devices* (1992). ISSN: 15579646. DOI: 10.1109/16.121690 (cit. on p. 44).

[29] Claudio Piemonte et al. "Performance of a novel, small-cell, high-fill-factor SiPM for TOF-PET". In: *IEEE Nuclear Science Symposium Conference Record*. 2013. ISBN: 9781479905348. DOI: 10.1109/NSSMIC.2013.6829170 (cit. on p. 46).

[30] S. Vinogradov. "Analytical models of probability distribution and excess noise factor of solid state photomultiplier signals with crosstalk". In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2012. DOI: 10.1016/j.nima.2011.11.086 (cit. on p. 47).

[31] Claude Leroy and Pier Giorgio Rancoita. *Silicon solid state devices and radiation detection*. 2012. ISBN: 9789814390057. DOI: 10.1142/8383 (cit. on p. 48).

[32] Pietro P. Calò et al. "SiPM readout electronics". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 926 (2019). Silicon Photomultipliers: Technology, Characterisation and Applications, pp. 57–68. ISSN: 0168-9002. DOI: https://doi.org/10.

1016/j.nima.2018.09.030. URL: https://www.sciencedirect.com/science/article/pii/S0168900218311756 (cit. on p. 48).

[33] F. Corsi et al. "Current-mode front-end electronics for silicon photomultiplier detectors". In: *Proceedings of the 2nd IEEE International Workshop on Advances in Sensors and Interfaces, IWASI*. 2007. ISBN: 1424412455. DOI: 10.1109/IWASI.2007.4420025 (cit. on p. 49).

[34] Paolo Trigilio et al. "ANGUS: A multichannel CMOS circuit for large capacitance silicon photomultiplier detectors for SPECT applications". In: *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2014*. 2016. ISBN: 9781479960972. DOI: 10.1109/NSSMIC.2014.7431125 (cit. on p. 50).

[35] Hal O. Anger. "Sensitivity, resolution, and linearity of the scintillation camera". In: *IEEE Transactions on Nuclear Science* (1966). ISSN: 15581578. DOI: 10.1109/TNS.1966.4324123 (cit. on p. 57).

[36] Guen Bae Ko et al. "Development of a front-end analog circuit for multi-channel SiPM readout and performance verification for various PET detector designs". In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (2013). ISSN: 01689002. DOI: 10.1016/j.nima.2012.11.087 (cit. on p. 57).

[37] G. F. Knoll and M. E. Schrader. "Computer correction of camera nonidealities in gamma ray imaging". In: *IEEE Transactions on Nuclear Science* (1982). ISSN: 15581578. DOI: 10.1109/TNS.1982.4332184 (cit. on p. 59).

[38] J. R. Mallard and M. J. Myers. "The performance of a gamma camera for the visualization of radioactive isotopes in vivo". In: *Physics in Medicine and Biology* (1963). ISSN: 00319155. DOI: 10.1088/0031-9155/8/2/304 (cit. on p. 59).

[39] A. Fabbri et al. “Study of position reconstruction of a LaBr3:Ce continuous scintillation crystal for medical applications”. In: *Journal of Instrumentation* (2013). ISSN: 17480221. DOI: 10.1088/1748-0221/8/12/P12010 (cit. on p. 60).

[40] A. Morozov et al. “ANTS2 package: Simulation and experimental data processing for Anger camera type detectors”. In: *Journal of Instrumentation* (2016). ISSN: 17480221. DOI: 10.1088/1748-0221/11/04/P04022 (cit. on p. 61).

[41] Robert M. Gray and Albert Macovski. “Maximum a posteriori estimation of position in scintillation cameras”. In: *IEEE Transactions on Nuclear Science* (1976). ISSN: 15581578. DOI: 10.1109/TNS.1976.4328354 (cit. on p. 62).

[42] Harrison H. Barrett, Kyle J. Myers, and Satyapal Rathee. “Foundations of Image Science”. In: *Medical Physics* (2004). ISSN: 0094-2405. DOI: 10.1118/1.1677252 (cit. on pp. 64, 69).

[43] Harrison H. Barrett et al. “Maximum-likelihood methods for processing signals from gamma-ray detectors”. In: *IEEE Transactions on Nuclear Science* (2009). ISSN: 00189499. DOI: 10.1109/TNS.2009.2015308 (cit. on p. 64).

[44] V. N. Solovov et al. “Position reconstruction in a dual phase xenon scintillation detector”. In: *IEEE Transactions on Nuclear Science* (2012). ISSN: 00189499. DOI: 10.1109/TNS.2012.2221742. arXiv: 1112.1481 (cit. on pp. 66, 69).

[45] T. D. Milster et al. “Digital position estimation for the modular scintillation camera”. In: *IEEE Transactions on Nuclear Science* (1985). ISSN: 15581578. DOI: 10.1109/TNS.1985.4336935 (cit. on p. 69).

[46] F. Neves et al. “Position reconstruction in a liquid xenon scintillation chamber for low-energy nuclear recoils and γ-rays”. In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (2007). ISSN: 01689002. DOI: 10.1016/j.nima.2006.10.390 (cit. on p. 69).

[47] A. Morozov et al. “Iterative reconstruction of detector response of an Anger gamma camera”. In: *Physics in Medicine and Biology* (2015). ISSN: 13616560. DOI: 10.1088/0031-9155/60/10/4169 (cit. on p. 69).

[48] A. Morozov et al. “Iterative reconstruction of SiPM light response functions in a square-shaped compact gamma camera”. In: *Physics in Medicine and Biology* (2017). ISSN: 13616560. DOI: 10.1088/1361-6560/aa6029 (cit. on p. 69).

[49] M. Occhipinti, P. Busca, and C. Fiorini. “Light response estimation and gamma events reconstruction in gamma-detectors based on continuous scintillators and SiPMs”. In: *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop, NSS/MIC/RTSD 2016*. 2017. ISBN: 9781509016426. DOI: 10.1109/NSSMIC.2016.8069405 (cit. on pp. 70, 72, 73).

[50] L J P Van Der Maaten, E O Postma, and H J Van Den Herik. “Dimensionality Reduction: A Comparative Review”. In: *Journal of Machine Learning Research* (2009). ISSN: 0169328X. DOI: 10.1080/13506280444000102 (cit. on p. 76).

[51] Joos Korstanje. *What is the difference between PCA and Factor Analysis?* URL: https://towardsdatascience.com/what-is-the-difference-between-pca-and-factor-analysis-5362ef6fa6f9 (visited on Dec. 2020) (cit. on p. 77).

[52] Carlo Vercellis. *Business Intelligence: Data Mining and Optimization for Decision Making*. 2009. ISBN: 9780470511381. DOI: 10.1002/9780470753866 (cit. on pp. 80, 81, 115).

[53] Andreas Janecek et al. “On the Relationship Between Feature Selection and Classification Accuracy.” In: *Fsdm* (2008) (cit. on p. 81).

[54] Tom Howley et al. “The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data”. In:

Knowledge-Based Systems (2006). ISSN: 09507051. DOI: 10.1016/j.knosys.2005.11.014 (cit. on p. 81).

[55] Lk Saul and St Roweis. "An introduction to locally linear embedding". In: *unpublished. Available at: http://www. cs. toronto. ...* (2000). ISSN: <null> (cit. on p. 81).

[56] Suchismita Das and Nikhil R. Pal. *Nonlinear Dimensionality Reduction for Data Visualization: An Unsupervised Fuzzy Rule-based Approach*. 2020. arXiv: 2004.03922 [cs.LG] (cit. on p. 82).

[57] Ramsey D. Badawi et al. "First Human Imaging Studies with the EXPLORER Total-Body PET Scanner*". In: *Journal of Nuclear Medicine* 60.3 (2019), pp. 299–303. ISSN: 0161-5505. DOI: 10.2967/jnumed.119.226498. eprint: https://jnm.snmjournals.org/content/60/3/299.full.pdf. URL: https://jnm.snmjournals.org/content/60/3/299 (cit. on p. 84).

[58] Stefan Siegel et al. "Simple charge division readouts for imaging scintillator arrays using a multi-channel PMT". In: *IEEE Transactions on Nuclear Science* (1996). ISSN: 00189499. DOI: 10.1109/23.507162 (cit. on p. 84).

[59] Xiaoli Li et al. "Study of PET detector performance with varying SiPM parameters and readout schemes". In: *IEEE Transactions on Nuclear Science* (2011). ISSN: 00189499. DOI: 10.1109/TNS.2011.2119378 (cit. on p. 84).

[60] Y. C. Shih et al. "An 8x8 row-column summing readout electronics for preclinical positron emission tomography scanners". In: *IEEE Nuclear Science Symposium Conference Record*. 2009. ISBN: 9781424439621. DOI: 10.1109/NSSMIC.2009.5402200 (cit. on p. 84).

[61] Robert S. Miyaoka, Tao Ling, and Tom K. Lewellen. "Effect of number of readout channels on the performance of a continuous miniature crystal element (cMiCE) detector". In: *IEEE Nuclear Science Symposium Conference Record*. 2006. ISBN: 1424405610. DOI: 10.1109/NSSMIC.2006.354261 (cit. on p. 84).

[62] L. A. Pierce et al. “Multiplexing strategies for monolithic crystal PET detector modules”. In: *Physics in Medicine and Biology* (2014). ISSN: 13616560. DOI: 10.1088/0031-9155/59/18/5347 (cit. on pp. 84, 85, 87).

[63] Stefano Pedemonte, Larry Pierce, and Koen Van Leemput. “A machine learning method for fast and accurate characterization of depth-of-interaction gamma cameras”. In: *Physics in Medicine and Biology* (2017). ISSN: 13616560. DOI: 10.1088/1361-6560/aa6ee5 (cit. on pp. 84, 86, 132).

[64] L. A. Pierce et al. “Characterization of highly multiplexed monolithic PET / gamma camera detector modules”. In: *Physics in Medicine and Biology* (2018). ISSN: 13616560. DOI: 10.1088/1361-6560/aab380 (cit. on pp. 84, 98).

[65] A. Morozov et al. “ANTS - A simulation package for secondary scintillation Anger-camera type detector in thermal neutron imaging”. In: *Journal of Instrumentation* (2012). ISSN: 17480221. DOI: 10.1088/1748-0221/7/08/P08010 (cit. on p. 91).

[66] Michela Massara. “Processing methods for reconstruction of detected gamma events in a SPECT-MRI imaging system”. In: (2018) (cit. on p. 92).

[67] MJ Berger. “XCOM: Photon cross section database (version 1.3)”. In: *http://physics. nist. gov/xcom* (2005) (cit. on p. 94).

[68] Rodrigo Coelho Barros et al. *A survey of evolutionary algorithms for decision-tree induction*. 2012. DOI: 10.1109/TSMCC.2011.2157494 (cit. on p. 114).

[69] Lior Rokach and Oded Maimon. “Decision Trees”. In: *Data Mining and Knowledge Discovery Handbook*. Ed. by Oded Maimon and Lior Rokach. Boston, MA: Springer US, 2005, pp. 165–192. ISBN: 978-0-387-25465-4. DOI: 10.1007/0-387-25465-X_9. URL: https://doi.org/10.1007/0-387-25465-X_9 (cit. on p. 114).

[70] Thomas Hancock et al. “Lower Bounds on Learning Decision Lists and Trees”. In: *Information and Computation* (1996). ISSN: 08905401. DOI: 10.1006/inco.1996.0040 (cit. on p. 114).

[71] J. R. Quinlan. “Induction of Decision Trees”. In: *Machine Learning* (1986). ISSN: 15730565. DOI: 10.1023/A:1022643204877 (cit. on pp. 115, 117).

[72] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN: 1558602380 (cit. on pp. 115, 117, 118).

[73] Leo Breiman et al. *Classification and regression trees*. 2017. ISBN: 9781351460491. DOI: 10.1201/9781315139470 (cit. on p. 115).

[74] Murat Kayri and İsmail Kayri. “The Comparison of Gini and Twoing Algorithms in Terms of Predictive Ability and Misclassification Cost in Data Mining: An Empirical Study”. In: *International Journal of Computer Trends and Technology* 27 (Sept. 2015), pp. 21–30. DOI: 10.14445/22312803/IJCTT-V27P105 (cit. on p. 117).

[75] J. R. Quinlan. “Simplifying decision trees”. In: *International Journal of Man-Machine Studies* (1987). ISSN: 00207373. DOI: 10.1016/S0020-7373(87)80053-6 (cit. on p. 118).

[76] Cristina Olaru and Louis Wehenkel. “A complete fuzzy decision tree technique”. In: *Fuzzy Sets and Systems* (2003). ISSN: 01650114. DOI: 10.1016/S0165-0114(03)00089-7 (cit. on p. 118).

[77] J. R. Quinlan. “Bagging, boosting, and C4.5”. In: *Proceedings of the National Conference on Artificial Intelligence*. 1996 (cit. on p. 119).

[78] Leo Breiman. “Bagging predictors”. In: *Machine Learning* (1996). ISSN: 08856125. DOI: 10.1007/bf00058655 (cit. on p. 119).

[79] Javed A. Aslam, Raluca A. Popa, and Ronald L. Rivest. “On estimating the size and confidence of a statistical audit”. In: *EVT 2007 - 2007 USENIX/ACCURATE Electronic Voting Technology Workshop*. 2007 (cit. on p. 119).

[80] Luke Taylor and Geoff Nitschke. *Improving Deep Learning using Generic Data Augmentation*. 2017. arXiv: 1708.06020 [cs.LG] (cit. on p. 120).

[81] Chuan Wang and J. C. Principe. "Training neural networks with additive noise in the desired signal". In: *IEEE Transactions on Neural Networks* 10.6 (1999), pp. 1511–1517. DOI: 10.1109/72.809097 (cit. on p. 120).

[82] Chinmay R. Parikh, Michael J. Pont, and N. Barrie Jones. "Improving the performance of multi-layer Perceptrons where limited training data are available for some classes". In: *IEE Conference Publication*. 1999. ISBN: 0852967217. DOI: 10.1049/cp:19991113 (cit. on p. 120).

[83] Imène Garali et al. "Region-based brain selection and classification on pet images for Alzheimer's disease computer aided diagnosis". In: *Proceedings - International Conference on Image Processing, ICIP*. 2015. ISBN: 9781479983391. DOI: 10.1109/ICIP.2015.7351045 (cit. on p. 121).

[84] D. Mudali et al. "Classification of Parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features". In: *Computational and Mathematical Methods in Medicine* (2015). ISSN: 17486718. DOI: 10.1155/2015/136921 (cit. on p. 121).

[85] Kuang Gong et al. "Machine Learning in PET: From Photon Detection to Quantitative Image Reconstruction". In: *Proceedings of the IEEE* (2020). ISSN: 15582256. DOI: 10.1109/JPROC.2019.2936809 (cit. on pp. 121, 122).

[86] Ying Wang et al. "An improved PET image reconstruction method based on super-resolution". In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (2019). ISSN: 01689002. DOI: 10.1016/j.nima.2019.162677 (cit. on p. 121).

[87] Florian Muller et al. “Gradient Tree Boosting-Based Positioning Method for Monolithic Scintillator Crystals in Positron Emission Tomography”. In: *IEEE Transactions on Radiation and Plasma Medical Sciences* (2018). ISSN: 2469-7311. DOI: 10.1109/trpms.2018.2837738 (cit. on p. 122).

[88] Florian Müller et al. *A novel DOI positioning algorithm for monolithic scintillator crystals in PET based on gradient tree boosting*. 2018. DOI: 10.1109/trpms.2018.2884320. arXiv: 1808.06385 (cit. on p. 122).

[89] Louis Ferré. “Selection of components in principal component analysis: A comparison of methods”. In: *Computational Statistics & Data Analysis* 19.6 (1995), pp. 669–682. ISSN: 0167-9473. DOI: https://doi.org/10.1016/0167-9473(94)00020-J. URL: https://www.sciencedirect.com/science/article/pii/016794739400020J (cit. on p. 128).

[90] Zaber. *Zaber T-LSR datasheet*. URL: https://www.zaber.com/manuals/T-LSR (visited on 2021) (cit. on p. 134).