POLITECNICO DI MILANO

Facolta' di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Ingegneria Elettronica

EVALUATION AND IMPLEMENTATION OF A MOTION CORRECTION METHOD FOR HIGH RESOLUTION TOMOGRAPHS

Relatore: **Prof. Carlo E. FIORINI**
Correlatore: **Prof. Christer Halldin**

Tesi di Laurea Magistrale di
Matteo Ferrante
matr. 767122

Anno Accademico 2012/13

Contents

List of Figures

List of Tables

List of abbreviations

2D Two Dimensional
3D Three Dimensional
ADC Analog-to-Digital Converter
APD Avalance Photo Diode
BGO Bismuth Germanate
Bq Bequerel
C Carbon
C Coulomb
CCD Charge-Coupled Device
Ce Cerium
Ci Curie
cm Centimeter
Cs Cesium
CT Computed Tomography
DOI Depth Of Interaction
dps Decays Per Second
E Energy
EM Expectation Maximization
eV electronVolt
F Fluorine
f frequency
FBP Filtered Back Projection
FOV Field Of View
FPGA Field Programmable Gate Array
FWHM Full Width at Half Maximum
FWTM Full Width at Tenth Maximum
GSO Cerium
Ci Curie
cm Centimeter
Cs Cesium

CT Computed Tomography
DOI Depth Of Interaction
dps Decays Per Second
GSO(Ce) Cerium activated Gadolinium Silicate
Gy Gray
HRRT High Resolution Research Tomograph
Hz Hertz
I Iodium
IR infrared
IRLED Infrared Light Emitting Diodes
J Joule
Kg Kilogram
l Attenuation length
LOR Line of response
LSO(Ce) Cerium activated Luthetium Oxyorthosilicate
MAF Multiple-Acquisition-Frames
MLEM Maximum-Likelihood-Expectation-Maximization
mm Millimeter
Mo Molibdenum
MRI Magnetic resonance imaging
N Nitrogen
NaI(Tl) Thallium activated Sodium Iodide
O Oxygen
OSEM Ordered-Subset-Expectation-Maximization
P Probability
PET Positron Emission Tomography
PHA Pulse Height Analyzer
PMT photomultiplier tube
PSF Point Spread Function
PVE Partial Volume Effect
QE Quantum Efficiency
R Roentgen
RAD Radiation Absorbed Dose
RMS Root Mean Square
SNR Signal to Noise Ratio
SO Spillover
Spect Single Photon Emission Computed Tomography
Sv Sievert
Tc Tecnetium

Th Thorium
Tl Tallium
TOF Time Of Flight
U Uranium
UV Ultra Violet
V Volt

List of symbols

A atomic mass
α alpha particle
β beta particle
ν neutrino
Φ output light-yield for scintillators
c light speed
D gantry diameter
$\Delta\phi$ differential angle
Δx differential length
κ attenuation coefficient due to Pair production
λ wavelength
m_e electrom mass
μ attenuation coefficient
N radionuclide population
n refractive index
p momentum
ψ inverse of radioactive decay time constant
R_{energy} energy resolution for scintillators
ρ mass density
σ attenuation coefficientent due to Compton effect
$\mathcal{R}$ reflection coefficient for the energies
t time variable
τ_{decay} scintillator decay constant
τ attenuation coefficient due to PE effect
W radioactive activity
X mother atoms
Z atomic number

Riassunto

Uno dei problemi piú rilevanti nella diagnostica per immagine é rappresentato dalla necessitá di minimizzare l'interferenza esercitata sui dati finali, e quindi sulla diagnosi, dai movimenti dei pazienti sottoposti ad indagine.
Ci siamo quindi proposti con questo studio di sviluppare un metodo di motion correction per la High Resolution Research Tomograph (HRRT), impiegata per l'imaging cerebrale presso il Dipartimento di Neuroscienze del Karolinska Institutet, a Stoccolma.
In prima istanza si é indagato sulla risoluzione spaziale dello scanner, posizionando nel centro del Field of View della camera un capillare di dimensioni ridotte, riempito con ^{18}F, al fine di studiare come il sistema fosse in grado di riprodurre sorgenti puntiformi di attivitá. A fronte di valori oscillanti tra i 1.5 mm e i 3.2 mm, in termini di Full Width at Half Maximum (FWHM) dell'impulso riprodotto, si é confermata la necessitá assoluta di ricorrere alla motion correction. Si é quindi sviluppato un algoritmo di motion correction su immagini ricostruite, optando per l'adozione di una strategia di correzione Multiple-Acquisition-Frame e di un dispositivo esterno capace di registrare posizione e orientamento dell'oggetto: il Polaris Vicra. Il metodo di motion correction implementato ha mostrato risultati promettenti, nella correzione sia dei movimenti di traslazione (77% dei frame corretti), sia dei movimenti di rotazione (solo 2 dei frame corretti hanno valori oltre la soglia). Inoltre, la ulteriore possibilitá futura di manipolare la singola coincidenza consentirá al nostro algoritmo di effettuare una correzione su singolo evento e sfruttare cosí scenari che saranno preclusi ai metodi di correzione su immagini giá ricostruite, come il metodo a cross-correlazione fino ad ora utilizzato.
Con questo studio si é quindi evidenziata la necessitá di sviluppare una metodologia solida di motion correction e si é dimostrato come anche metodi di correzione offline, che si avvalgono di dispositivi esterni, siano in grado di fornire buone prestazioni, nonostante operino su dati giá ricostruiti. Lo sviluppo in questo settore é dunque fondamentale per usufruire di indagini diagnostiche dall'alta risoluzione e accuratezza.

Abstract

A very important issue in modern Nuclear Medicine is to reduce the effect of motion artifacts on reconstructed data: the very high resolutions achieved can be severely lowered by patient's movements, affecting the precision of the diagnosis. Therefore, in the present work, a method for motion correction has been presented for the Siemens High Resolution Research Tomograph (HRRT), installed at the Neuroscience department of Karolinska Institutet, in Stockholm.
As a first step, we have wanted to prove the effective need for motion correction on such a high resolution tomograph: an experiment to assess camera spatial resolution has been conducted and the results have been compared with patients' typical movements. A capillar filled with ^{18}F has been positioned in the field of view to observe how activity point sources are reconstructed in the final images; the Full Width at Half Maximum (FWHM) of the retrieved Point Spread Functions ranged from 1.5 mm to 3.2 mm, showing the necessity for correction of motion artifacts.
An image-driven approach has been developed and implemented following the Multiple-Acquisition-Frame strategy and using NDI Polaris Vicra to obtain motion data, needed for the algorithm. Corrected frames show promising results both in translation (77% of the frames is below the set threshold) and in rotation correction (only 2 frames are above the set threshold). However, the future possibility provided by external trackers to perform online corrections, once raw data can be accessed, gives our algorithm the advantage to manipulate the single events and, subsequently, to obtain accuracies unreachable for post-reconstruction corrections, like the cross-correlation method employed so far, in the department.
The role of the manuscript is then to uphold the need to motion correct HRRT studies and to show how offline corrections, using external trackers, can provide good performances, despite the impossibilty to access the single events; further developments in the field will therefore help in obtaining the best resolutions and accuracies achievable for high precision investigations.

Chapter 1

Introduction

Nowadays, the acquisition time required for a medical imaging study typically varies from 60 up to 120 minutes [1]: for this reason, it is impossible to assume that a normal person does not move throughout the scans.
The issue becomes even more relevant when the acquisition is performed on people affected by mental disorders like Dementia[2] or neurological disease such as Parkinson; in these cases, indeed, movements are quite frequent and the effect on the collected data is dramatic.
Many researchers have tried to estimate the normal unconscious movements (translations and rotations) performed by patients during a Nuclear Medicine exam and, in literature[3], common values for the translation along the three axis go from 1 to 6 millimeters, while common rotations range from decimals of a degree to 5 degrees; very few cases show translations and rotations exceeding those values.
The problem with motion artifacts in medical images has become a very important matter only in the last years, because of the continuous developments in the field, and the subsequent better resolution achieved: referring to the imaging system that we used for our study, modern positron emission high resolution research tomographs are able to provide a resolution of about 1.2 mm[4], which is now comparable with the normal motion of patients during a scan. The subsequent deterioration of the resolution, due to movement, can be better understood considering the whole square resolution "Full Width at Half Maximum" (FWHM) of the Positron Emission Tomograph (PET) as the sum of the square resolution of the tomograph and the square resolution due to motion, according to[1]:

$$FWHM^2_{effective} = FWHM^2_{tomograph} + FWHM^2_{motion} \tag{1.1}$$

The greater influence introduced by motion artifacts on the overall resolution can be viewed in the graphical plot of the squareroot of equation 1.1 against $FWHM_{tomograph}$ (Figure 1.1), where the limit due to movements becomes dominant, as the scanner resolution approaches very low values.

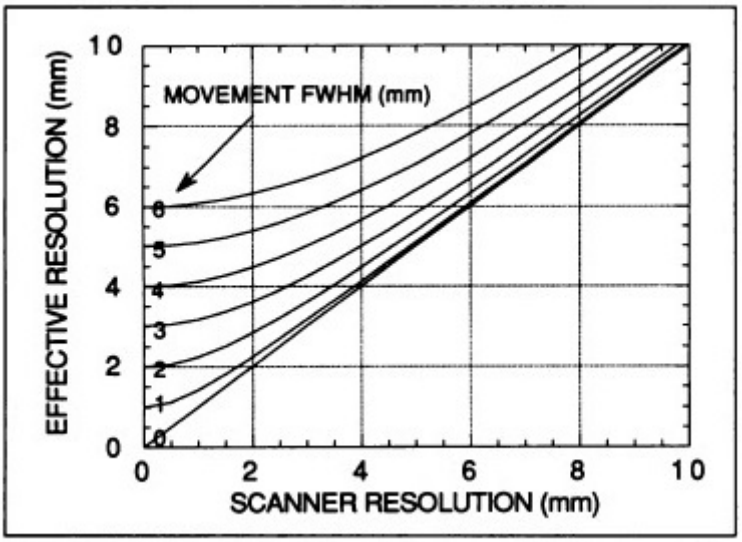

Figure 1.1: Overall resolution against camera resolution[5]

For a better understanding of motion artifact causes in PET, where we focused our research, it is necessary to know basic concepts of data reconstruction: when a PET study is performed, in addition to the emission scans, a transmission scan is performed, which after reconstruction is used to determine the attenuation map of the subject; the transmission scan is acquired with a process very similar to the Computed Tomography (CT), using either a rotating X-ray tube (working at the energy of 120 KeV like in the CT) or a rotating radioactive point source. The common used radioactive elements for the transmission scan are ^{68}Ge and ^{137}Cs: they both have long half-time constants, which makes them long-lasting, but in the last period ^{137}Cs has been more widely used than ^{68}Ge, because, even though it is not a positron emitter like ^{68}Ge, its spectrum has a narrow peak at the energy of 661.7 KeV, differing from the very wide spectrum of Germanium, and the use of a well defined energy spectra for the transmission scan definitely helps in determining patient's attenuation map[6].
The key role of the transmission scan is to build the FOV attenuation map, which takes into account the variations in radiation absorption among differ-

ent tissue types, and allows for a correct reconstruction of the activity inside the field of view.
A correct attenuation map is of outmost importance when reconstructing the emission data. Movements, causing a positional mismatch between transmission and emission data, will introduce artifacts. Some of these artifacts may be obvious when looking at the images, but it is just as likely that the artifacts only will result in incorrect values, causing a faulty data analysis.
Motion artifacts arise mainly from two phenomena[2]:

- **Intra-frame movements** : the patient moves within the acquisition time of a frame (from seconds to few minutes) and the effect is a blurring of the image, very similar to the one that we experience taking a photo of a moving object. One solution could be shortening the frame acquisition time but the resulting image would become increasingly more noisy due to worse statistics;

- **Inter-frames movements** : time-activity compartmental analysis of PET data requires that the patient's position stays the same during the full time of the scan. Any movement will cause positional shift between the frames, resulting in incorrect data. To avoid major movements, several restraints have been proposed and used[5] but, as the camera resolution is improving, new methods must be introduced to fight even very small movements.

Those movements create well-known artifacts in Nuclear Medicine images; among the most common ones in clinical routine, the cardiac cycles induce displacements noticeable with the present scanner resolutions, while respiration and brain movements cause the artifacts shown in Figure 1.2: the curvilinear cold artifacts, shown on the left, are related to patient's respiratory cycles, during whole body PET scans, while the discontinuities and white spots, which are evident at the border between white and gray matter or at the outer border of gray matter of the figure on the right, are due to head motion in brain imaging studies[2][7].

Motion correction has become a very hot topic in Nuclear Medicine modalities like Single Photon Emission Computed Tomography (SPECT), Magnetic Resonance Imaging (MRI) and PET; the initial idea to reduce patient's movements during the scan was to use restraints able to limit motion, but as soon as tomograph resolutions have started approaching millimeter values, this strategy has not been enough anymore.
In SPECT, different motion correction studies have been conducted, mostly

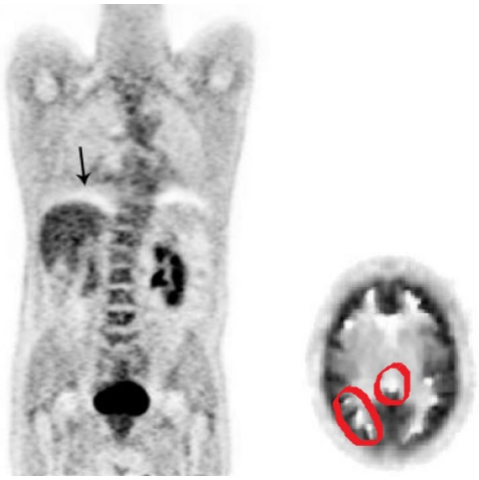

Figure 1.2: Curvilinear cold artifacts(on the left[6]) and head motion artifcacts(on the right[3])

using near-infrared tracking systems[8][9], to be able to scan small animals without the need of anesthesia or other chemical restraints that might influence the parameters of interest.
In MRI, the research on motion correction has led to different solutions: the use of an external tracker, which records motion information, allows online correction[10], while other strategies implemented exploit scanner gradient fields to localize the position inside the camera (field detection methods)[11].
In PET too, many solutions have been investigated: cardiac and/or respiratory gating are methods that ensure a certain liability, but they require long processing times and are unfit for brain imaging purposes[11].
The motion correction techniques developed for PET can be basically divided in two big categories[12]: strategies based on similarity criteria, and the ones using optical external trackers.
Similarity criteria are methods where the parameters for the correction are obtained comparing the reference frame and the moved frame, using functions that search for the geometrical transformation providing the highest similarity between the datasets; examples of similarity criteria are the cross-correlation function, the best modality when it can be assumed that a linear function links the intensity values of the two images, and the mutual information, where the parameters for the correction are retrieved checking for the statistical dependence between the two volumes. Usually, the cross-

correlation method is employed when the target and the source volume are from the same imaging modality (for example both from PET), while the mutual information is best employed when the two volumes come from different modalities (for example PET and MRI)[13][12].
The problem of motion corrction techniques relying on the nature of the emission values (like the previous ones) is that the correction is dependent on the quality of the data, and, then , noisy data can affect substantially the precision of the correction[7].
The limitation of the above strategies is overcome using external optical trackers, which follow patient's movements and permit to use motion data to compute the needed geometrical transformation. Two are the different approaches that can be employed, exploiting motion data:

- **Image-driven approach** : when the motion tracker records a displacement above a fixed threshold, the upcoming events will be framed in a new image, that will be corrected with the recorded motion data;

- **Event-driven approach** : every single annihilation (and therefore the connected line of response) is positioned in the Field Of View (FOV), according to the information output by the tracker, and without the need to define a priori the framing interval.

The biggest example of image-driven approach is the Multiple-Acquisition-Frame (MAF) strategy performed on reconstructed volumes, where the moved frames are brought back in the reference frame position [14], correcting for the big inter-frames movements but not for the intra-frame movements, since every time the threshold is overcome, a new frame is created.
The impossibility of correcting intra-frame movements with image-driven approaches is the reason why, in the last years, motion correction research has moved toward the event-driven methodologies; at the cost of a higher computational demand and a higher complexity, every single line of response (LOR) can be manipulated in order to solve problematics arising both from inter-frames and intra-frame movements and to ensure loyal activity values[16][17].
The object of this manuscript is to describe the motion correction technique implemented for brain imaging, at the Karolinska Institutet, in Stockholm: the treated method is an image-driven approach implemented to automatic motion correct datasets already reconstructed and reduce the effect of inter-frames movements on the High Resolution Research Tomograph (HRRT).
In the present work, important issues like the physical backgrounds needed

for PET are discussed; subsequently, the layout and the electronics are introduced first for PET and, then, for the HRRT, the scanner upon which we implemented our correction. In addition to this, an experiment to assess camera resolution is performed to highline the need for motion correction and, afterward, our algorithm is presented and tested.
In more detail, in chapter 2, the physical basis of PET are treated along with the electronics involved; in chapter 3, the HRRT camera is analyzed in every particular design choice of its and, in chapter 4, the experiment to retrieve scanner resolution is explained. In chapter 5, our motion correction algorithm is theoretically discussed, while, in chapter 6, the experimental data obtained with our method are presented together with the ones deriving from the employment of the pre-existing motion correction algorithm used in the department; finally, in chapter 7, the two methods are compared to underline pros and cons and the present analysis is exploited in chapter 8 and 9, to present, corrispondingly, the conclusions coming from our work and the future solutions able to reach higher precisions in the correction.
The results obtained with our method were satisfactory, with good performances in the correction of the translation and excellent performances in the correction of the rotation.
Unfortunately, our algorithm is an image-driven approach, but at the state of art, on this camera, an event-driven approach has not been validated yet and many are the equipes, in the world, working on such an algorithm.
The expertises developed in the field will be then of great value when the emission data will be accessed before of reconstruction, at the list-mode level, with the great possibility of implementing an event-driven approach for motion correction, even on the HRRT.

Chapter 2

PET basics

The underlying hypothesis used in PET is that a positron, ejected by the nucleus of an unstable nuclide, when captured by an electron, emits two 511 KeV photons, that are emitted nearly in opposite directions, with a 180^o angle between them, if the annihilation takes place with electron and positron "at rest". Working under this assumption, when the two photons are detected in coincidence by two detectors, within a very short time window, they are considered to be originated from the same annihilation and, in the reconstruction, it is assumed that the event took place on the straight line connecting the two detectors, which, for this reason, is named line of response; with enough of those coincidence detections, we will have enough data to determine precisely the position of the annihilations and, then, we will be able to reconstruct an image showing the activity inside the FOV[18][19]. As we will see in the subsequent paragraphs, this hypothesis, under which we work, is not always true, but there is a slight deviation from the 180^o angle between the two photons (they are not perfectly collinear), and in terms of the annihilation position (it is not fully true that the annihilation occurs at the same position where the positron was emitted): errors introduced by the above issues are small and can be taken into account by the reconstruction software.

2.1 Gamma radiation

In PET imaging, it is the annihilation radiation, gamma rays, that is detected. Unlike X-rays, which arise from interactions with orbital electrons, the annihilation gamma radiation (γ-rays) is emitted after a reaction taking place within the nucleus; this is also the reason why this branch of medicine,

that deals with γ-rays, is called Nuclear Medicine.
The reaction, which occurs within the nucleus, is due to the unstable nature of a radioactive substance. All matter tries to reach an as stable level as possible. An unstable isotope of an element is referred to as a radionuclide. There are different types of radionuclide, whose classification is made according to the feature of the corresponding stable isotope: radionuclides with same mass number of the stable nuclide, but different energies are called Isomers (i.e. ^{99}Tc and ^{99m}Tc); radionuclides having the same mass number of the stable nuclide are called Isobars (i.e. ^{131}I and ^{131}Xe); radionuclides with the same number of neutrons of the stable nuclide are called Isotones (i.e. $^{15}_{7}\mathrm{N}_{8}$ and $^{16}_{8}\mathrm{O}_{8}$) and finally, radionuclides having the same numbers of protons of the stable nuclide are the Isotopes, which are the ones we use extensively in PET (i.e. $^{14}_{6}$C and $^{12}_{6}$C)[18].
Radioactive decay is the process where an unstable substance spontaneously will decompose to form nuclei with a higher stability.
The decay process towards more stable nuclei will be in the form of emission of photons or subatomic particles, or a combination of photons and particles. The most important, for applications in the Nuclear Medicine field, are:

- **Alpha decay** : alpha particles (α) are Helium atoms, consisting of two neutrons and two protons, but void of the two K-orbital electrons. The decay consists of the emission of an α particle and then, in the nuclide reducing the atomic number (Z) of two, and the atomic mass (A) of four; the general chemical description of the process is:

$$^{A}_{Z}X \longrightarrow ^{A-4}_{Z-2}X + \alpha \tag{2.1}$$

  This decay is typical of heavy nuclei radionuclides and one of the most well known example of α decay is the Uranium decaying into Thorium ($^{238}_{92}\mathrm{U} \longrightarrow ^{234}_{90}\mathrm{Th}+\alpha$).

- **Negative Beta decay** : characteristic of neutron rich radionuclides, it occurs with the conversion of a neutron into a proton within the nucleus, and the emission of a neutrino(ν) along with a negative beta (β^-) particle; the general description of the process is:

$$^{A}_{Z}X \longrightarrow ^{A}_{Z+1}X + \beta^- + \nu \tag{2.2}$$

  One of the most important isotopes in Nuclear Medicine is the β^- decay of Molybdenum into metastable Technetium and, then, into stable

Technetium through an isomeric transition [18]; its importance is due to the fact that ^{99}Tc has a well defined energy spectra, has a half-life ideal for medical applications, is easy to produce and, in addition, ^{99}Tc can be labeled onto many bio-molecules targeting specific organs, making it one of the most used radioisotopes in SPECT, and its discovery allowed a big step forward in the field of radiotracers.

- **Positive Beta decay** : this process takes place in proton rich radionuclides, where one proton is converted into a neutron, and subsequently a neutrino and a positron (β^+) are emitted, to take away the positive charge from the nucleus. The overall decay can be then summarized this way:

$$ {}^{A}_{Z}X \longrightarrow {}^{A}_{Z-1}X + \beta^+ + \nu \tag{2.3} $$

The β^+ decay is the process, which we exploit in PET, since the recombination of the positron with an electron (annihilation) will cause the emission of the two γ-rays: the signal we use for the reconstruction of the information.

There are also other types of radioactive decay, like electron capture, where a K-shell electron is captured by a proton in the nucleus to produce a neutron, or isomeric transition, where the nucleus decays from an excited (metastable) state, or finally the internal conversion, where the energy provided by the unstability of the nucleus is given to inner shell electrons, ejecting them with great kinetic energies to cause further ionizations[18]; anyway they are not of great interest in PET and, for this reason, they will not be accurately explained in this manuscript.
Gamma radiation is generated by the different processes, which take place throughout the radioactive decays previously described: the 140 KeV photon, used in some SPECT studies, is indeed generated by the β^- decay of ^{98}Mo into ^{99m}Tc, and the subsequent isomeric transition of ^{99m}Tc into ^{99}Tc, where this energy transition is followed by the emission of the wanted γ-ray; the two collinear 511 KeV photons, exploited in PET, arising from the annihilation of an electron with a positron, emitted during a β^+ decay.
Focusing more deeply on PET, the positron is the antimatter equivalent of the electron, and the knowledge of the various mechanisms, which stand behind its annihilation, is fundamental to understand the performances of the scanner, that we are working with.
As soon as the positron is emitted with not null kinetic energy, it starts travelling through the tissues; Coulombian interactions with matter make the β^+

particle slow down and lose its momentum. When the lost energy is almost equal to the kinetic energy initially held, the positron annihilates with one of the nearby electrons, resulting in the emission of two γ photons.
The annihilation process is very simple and clear, but it is important to note the resolution limiting factors, which arise from positron behaviour:

1. The positron range: being emitted with not null kinetic energy, the positron travels a certain distance before undergoing annihilation; it means that the emission did not occur where the annihilation happened, and therefore the LOR corresponding to the detected event does not represent where the radiotracer is, but where the positron met the electron. The result is a deterioration of the camera resolution. Instead of a perfect point, the source of decay will be presented as a Gauss-shaped area around the radiotracer. Usually the path of the β^+ particle is very turtuous, and highly depending on its initial kinetic energy: this is why, to study it, histograms based on the initial kinetic energy and the position of annihilation for different radiotracers (Figure 2.1) are needed; positron range values, analyzed in water, are around $0.1 \div 1$mm[20].

2. The angular uncertainty: it is not always true that the annihilation takes place with zero momentum. Indeed, the reason why the two photons are emitted in opposite directions derives from the momentum conservation principle: on one hand, if the positron annihilates with the electron at zero momentum, then the two γ-rays are emitted in opposite directions, in order to have the system overall momentum equal zero; on the other hand, if the annihilation occurs with the positron having a residual momentum, the system must hold the same momentum, and the angle between the two photons must differ from 180^o, otherwise the overall momentum would be null.
   The probability for a positron to annihilate with zero momentum is around 35%[21]: only one out of three positrons annihilates emitting two collinear γ photons; this produces a loss of precision in the estimated position of the radiotracer, which, for small angle deviations, roughly can be approximated by:

$$\Delta\phi = \frac{p}{m_e c} \tag{2.4}$$

$$\Delta x \approx D \cdot \Delta\phi \tag{2.5}$$

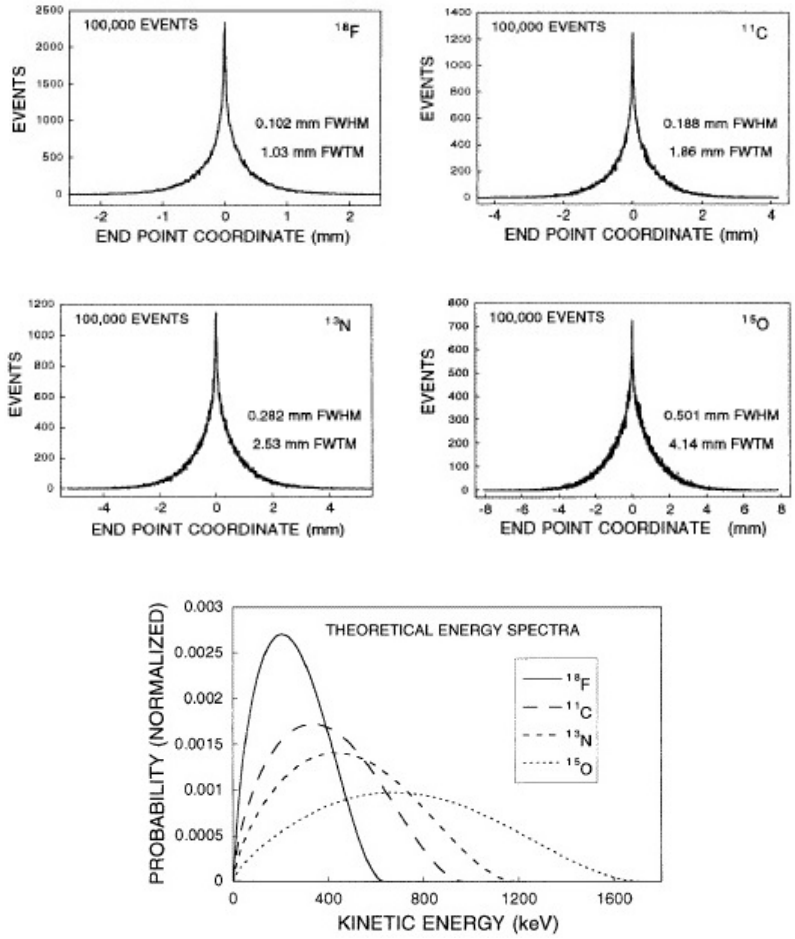

Figure 2.1: Annihilation position histogram (upper parts) and kinetic energy histogram (lower part) for the most used radiotracers[20]

In equations 2.4 and 2.5, $\Delta\phi$ stands for the angle deviation from the collinear value of 180^o, p stands for the positron momentum, m_e for the electron rest mass, Δx for the resolution error and D for the FOV diameter; then, the angular uncertainty produces an error depending on FOV size, and measured values for $\Delta\phi$ in water are of $0.4^o - 0.5^o$[22].

2.2 Radioactivity

To better undestand PET performances, it is necessary to quantitatively know the administered radioactivity.

Radioactive decay is a stocastic process; then, we cannot predict the exact number of radiotracers decaying at a certain time, but only predict the average behaviour. This average behaviour can be derived, observing that the variation (δN: in our case, it is obviously a decrement) in the number of the radionuclides, within a short time interval (δt), is proportional to the number of unstable nuclei in the sample[23], resulting in:

$$\delta N = -\psi N \delta t \tag{2.6}$$

The ψ parameter in 2.6 is a constant depending on the radiotracer, and its value depends on another important parameter: the half-life value. The half-life constant, carachteristic for every radionuclide, is the time interval required for half the atoms to decay. The relationship between half-life and ψ comes from:

$$t_{half-life} = \frac{\ln(2)}{\psi} \tag{2.7}$$

Considering differential equation 2.6, its solution, given population at $t = 0$ (N_o), is:

$$N(t) = N_o e^{-\psi t} \tag{2.8}$$

From equation 2.8, it is straightforward to understand that the activity, defined as number of decays per unit time ($W = \frac{\delta N}{\delta t}$), is maximum at $t = 0$ and it decays with an exponential shape toward zero.
Units describing radioactive decay and radioactive interaction with tissue, and therefore important in PET, are:

- **Unit for radioactivity** : it gives the number of decays per second(dps) and it can be expressed either in Becquerel(1 Bq=1 dps) or in Curie(1 Ci=$3.7 \cdot 10^{10} dps$); common values, in clinical applications, are KBq÷MBq or equivalently μCi÷mCi[23].

- **Unit for activity in the environment** : it gives the strength of γ or X rays in the air, and it is measured either in Roentgen or in $\frac{Coulomb}{Kg}$ ($1R = 2.58 \cdot 10^{-4} \frac{C}{Kg}$).

- **Unit for absorbed dose** : it gives the absorbed energy from radiation in a tissue unit mass and it is expressed either in Gray($\frac{J}{Kg}$) or in "Radiation Absorbed Dose" ($1Gy = 100 rads$).

- **Unit for biological damage from radiation** : it gives the equivalent dose absorbed by the tissue and is calculated as the product of the absorbed dose with a factor equal or greater than one. The factor is based on the amount of energy deposited in tissue per distance traveled; for gamma and X radiation, this factor is equal to one and the used unit of measure, whose name is the Sievert(Sv), is then coincident to the absorbed dose($1Sv = 1Gy$). Common values, in clinical environments, ranges from units up to tens of mSv[24].

2.3 γ-ray interactions with matter

As radiation travels through matter, it interacts with the electrons of the atoms and γ photons are scattered from the beam; to have the knowledge of the main interactions, whom the gamma radiation can undergo crossing the matter, gives us not only the tools to fully understand the process, which delivers us the signal we use in PET, but also the chance to implement methods to compensate for faulty information, arising from some of these phenomena, corrupting the real information, that we are interested of.
The three main types of γ-ray interaction with matter are the photoelectric absorption, Compton scattering and pair production

- **Photoelectric absorption** : in the photoelectric absorption, the energy of the γ photon is almost totally absorbed by one of the electrons of the inner shells (mostly the K-shell), which is ejected with a kinetic energy equal to the difference between the incoming photon energy and the binding energy of the electron ($E_{kin,e^-} = E_\gamma - E_{binding}$); the fast photoelectron is then able to ionize other atoms, while the vacancy left in the K-shell is filled by an electron from an outer shell, with the subsequent emission of an X-ray.
  The probability for a photoelectric absorption to occur is expressed, through a rough approximation, by $P \approx constant \cdot \frac{Z^n}{E_\gamma^{3.5}}$, where Z stands for the absorbing material atomic number, E_γ for the incoming gamma photon energy and n for a value, that varies between 4 and 5, depending on the photon energy[25].
  The photoelectric absorption has not a corruptive effect on the information we get with the scanner, but it weaks down the signal, because it subtracts gamma photons from the radiation beam we will detect; on the other hand, in the choice of detector material, it is highly desired

to have a predominance of photoelectric absorption over the other phenomena, and, mostly for this reason, high Z materials are preferred. This will be explained in detail later.

- **Compton scattering** : Compton scattering is probably the most unwanted interaction, the gamma photon can undergo in PET. It occurs when the photon interacts with the outer shells electrons and, as a result of this interaction, an electron with a kinetic energy equal to $E_{kin,e^-} = E_\gamma - E_{binding}$ and a scattered photon are emitted; this outcoming photon will have not only a different direction of propagation, but its energy will be lower than the one of the incoming photon.
  The Compton scattered photon will transfer some of its energy to the electron and, in accordance with the law on conservation of momentum, its direction will change. Even a small amount of energy loss will result in a quite large deflection of the photon. When a scattered photon is detected by the PET detector system, the resulting LOR will no longer represent the position of the annihilation, resulting in a blurring of the reconstructed image. To minimize the effect of Compton scatter in the final image, a lower limit of the detected photon energy is defined so that photons below this threshold are discarded. Setting the threshold too low will result in a blurring of the PET image due to too many events from scattered photons. Setting the threshold too high will result in low statistics.
  Another important tool to reduce the effect of Compton scattering is given by the Klein-Nishina formula (equation 2.9), which gives us the possibility to study the probability of the scattered photon to be emitted with a certain angle and then allows us to reconstruct the original path of the photon.

$$hf^{'} = \frac{hf}{1 + \frac{hf}{m_0 c^2} \cdot (1 - \cos\theta)} \tag{2.9}$$

- **Pair production** : the last interaction, that will be discussed, is pair production. Differently from the two previous cases, this interaction is very unlikely at PET working energies (511 KeV), since the energy required for the pair production is close to the rest-mass energy of an electron (1.022 MeV[25]); what happens in pair production is that the γ-ray dissapears and a positron-electron pair is created: soon after the

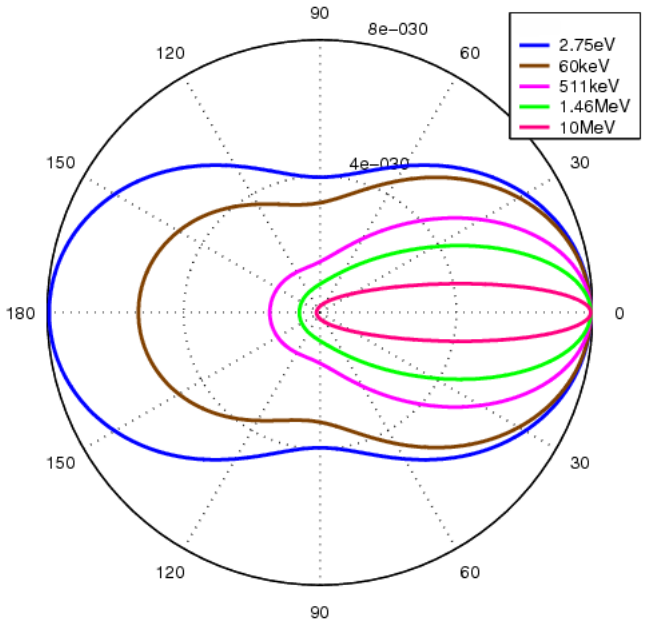

Figure 2.2: The Klein-Nishina distribution of photon scattering angles over a range of energies[26]

production, the positron will annihilate with an electron emitting two γ photons, exactly in the same way as with radiotracers.
Along with Rayleigh scattering, which is a scattering occurring at low γ-ray energies (below a few hundreds of KeV[25]), and consists only in the deflection of the photon, with no energy loss or fast electrons, the pair production does not occur very often and it is not of interest in PET.

At the microscopic level, the probability of interaction with matter is described by the atomic cross section (expressed in $\frac{cm^2}{atom}$), where the overall atomic cross section is given by the sum of all atomic cross sections, for all kinds of interaction.
In a macroscopic and more simplified approach, the interaction with matter is described through the use of the mass attenuation coefficient μ ($\frac{cm^2}{g}$), which plays an important role in the Lambert-Beer law:

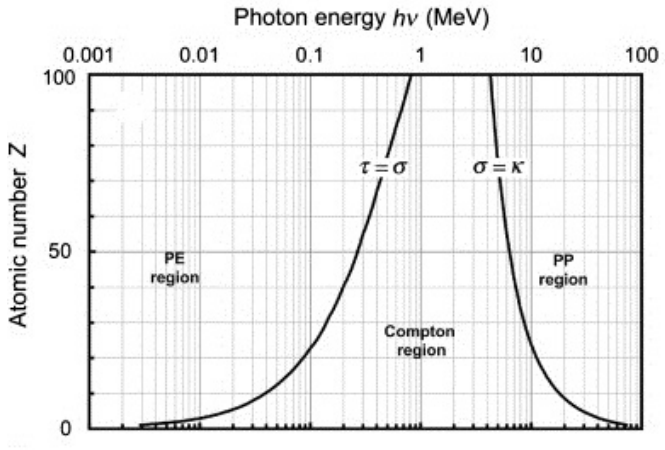

Figure 2.3: Regions of predominance of the three described interactions depending on the Z of the absorber and on the energy of the incoming photon[27]

$$I(x) = I_0 \cdot e^{-\int_0^x \mu(x)\varrho(x)dx} \tag{2.10}$$

In equation 2.10, I_0 is the incoming photon beam, $I(x)$ is the photon beam at coordinate x, ϱ (x) is the materal density at coordinate x, and the integral over the line $[0, x]$ is employed to take into consideration material dishomogeneities, which cause sudden changes in the mass attenuation coefficient value. The mass attenuation coefficient can be thought as the sum of the three possible interactions previously described, since its value is given by sum of the mass attenuation coefficient due to photoelectric absorption(τ), the mass attenuation coefficient due to Compton scattering(σ) and the mass attenuation coefficient due to pair production(κ), according to:

$$\mu = \tau + \sigma + \kappa \tag{2.11}$$

2.4 PET layout

Nowadays PET has a consolidated geometry: a ring of detectors is arranged around the bed where the patient lies, and collects all the possible photons emitted by positron annihilations.
To go more into the details of the scanner layout, each detector block, which

is part of the ring, is composed of a scintillator block, converting the γ photon into visible light photons, and of a PMT, amplifying the incoming light photons into an electric signal; back in the early years of PET, a scintillator crystal was coupled to a single PMT, and this design was the cause of a poor spatial resolution, since the minimum detectable dimension could not go lower than PMT size.

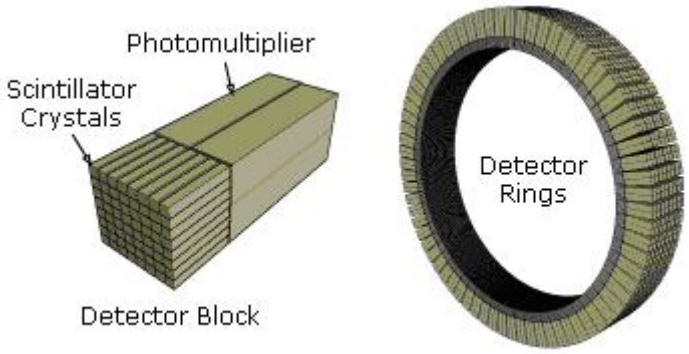

Figure 2.4: The crystals-to-PMT coupling and the overall detectors ring[28]

Thanks to big improvements in scintillator materials, modern cameras can have a much better spatial resolution than the early ones, using more scintillator crystals coupled to a single PMT; this design not only permits a higher precision in detecting the annihilation position, exploiting Anger logic to retrieve the position of interaction in the crystal array, but allows also a significant reduction of costs. The number of scintillator crystals, which can be coupled to a single PMT, varies depending on the chosen scintillator material, because of its output light-yield, and with Bismuth Germanate (BGO), currently the most employed scintillator in clinical PETs, the coupling ratio is 16:1[18].
The modern adopted coupling design is the "quadrant sharing design", where every quarter of PMT photocathode sensitive area is coupled to a different scintillator block, giving the best performances achievable in spatial resolution and reduction of costs, but also having deadtime issues, with lower maximum detection frequency, due to the sharing of the PMT among different blocks. The position of interaction is retrieved using Anger Logic, after that all detectors are calibrated to compensate for the different performances they have to the same event, as it will be explained in the HRRT chapter.

When the first photon, arising from the annihilation, hits the detector, a coincidence circuit, with a time window of $6 \div 20$ nsec[18], activates those

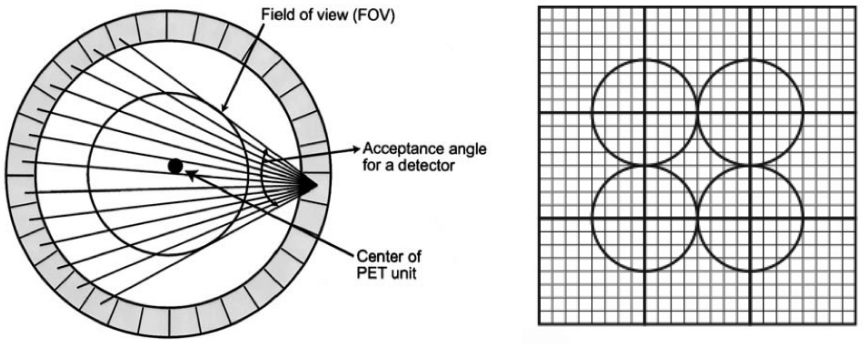

Figure 2.5: The acceptance angle concept (on the left) and the quadrant sharing coupling design (on the right)[18]

detectors allowed for the detection of the second photon; the activated detectors compose a fan-beam pattern, defined by the FOV, which makes us able not to use the whole hardware, but to activate part of it, reducing power consumption and limiting the maximum angle of acceptance, as seen by the first involved detector. Once the coincidence is detected, the event validity is analyzed by the Pulse Height Analyzer (PHA) circuit, which, relying on the energy information given by all the crystals involved, verifies if the coincidence is true or it is due to a Compton scattering (thanks to the energy window previously introduced and discussed also in scintillator paragraph); if the validity is confirmed, the signal is digitized by a $10 \div 12$ bits Analog-to-Digital Converter(ADC), and processed by a Field Programmable Gate Array(FPGA) with all the details of the LOR and time stamp; the event is then stored in memory, before being used in the reconstruction[29].
The present paragraph has been an introduction to the processing chain of a coincidence event in PET; the next paragraphs will treat more in detail issues related to scintillators, Photo Multiplier Tubes (PMTs) and reconstruction methods.

2.5 Scintillator

One of the most important part of components in the PET camera are the scintillators; a complete understanding of what happens in a scintillator crystal, during a coincidence detection, is then fundamental to achieve the best quality possible in the reconstructed data.

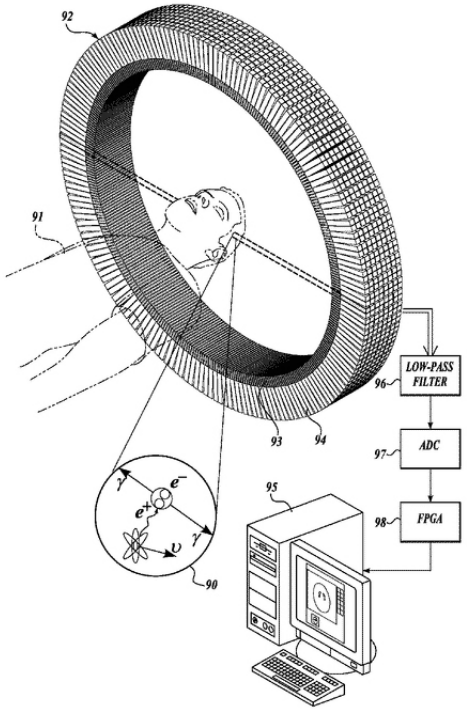

Figure 2.6: The data acquisition chain in PET[30]

The key role of the scintillator is to convert as much as possible γ photons into visible light photons (with energies close to 3 eV), in order to make the PMT output a strong electrical signal; the stronger and the faster the signal provided by the scintillator will be, the less noisy and more trustful information, we will have.
Scintillators used in PET cameras are exclusively inorganic, because of their better performance in terms of output light-yield, even though the organic crystals are able to provide better time resolution: the reason is that the pulse output from organic scintillators is too weak to allow acceptable signal-to-noise ratio (SNR) [25].
There are several factors to take into account when choosing detector material. Primarily, a material with high cross-section for photons in the energy window used is essential and the light-output should be well defined, both in energy and time. For each incoming annihilation photon, a large number of scintillation photons is desirable. Also, the self-absorption of the scintillation light should be low and the refraction index of the scintillator should be as close to glass as possible to allow for good optical coupling to the PMT.

There are several factors that must be taken into account when testing and deciding on scintillator material. Scintillator performance is usually evaluated looking at the following parameters[25][31][32]:

- **Material density** $[\varrho = \frac{g}{cm^3}]$ **and atomic number** $[Z]$: as already stated, the probability of photoelectric absorption is strongly increased by an higher atomic number and/or an higher density value; for this reason, researchers look for high Z and high density material, for the conversion of γ photons into visible light photons.

- **Attenuation length** $[l = mm]$: this parameter, which is the reciprocal of the attenuation coefficient $[\alpha = \mu\varrho = \frac{1}{cm}]$, gives us the needed length, for the material, to absorb the 63.21% of the incoming radiation intensity (from Lambert-Beer law: $I(\frac{1}{\alpha}) = I_0 e^{-1}$); a short value allows to reduce crystal sizes, keeping the same absorption efficiency.

- **Decay time constant** $[\tau = nsec]$: with a first order approximation, it defines the time interval required for the scintillation process to attenuate the incoming photons of 63.21% (according to: $I(\tau) = I_0 e^{-1}$). A short value gives us the chance to shorten the coincidence time window, and then reduce the probability of random coincidences, arising from detection of photon-pairs generated by different annihilation events; optimum values also permit to implement TOF methods, to achieve better accuracies.

- **Energy resolution** $[R = \%]$: it describes the capability of the material to discriminate different energy photons, and its expression is given by $R = \frac{FWHM_{working-energy}}{Working-energy}$, where $FWHM$ is the full width at half maximum of the material sensitivity peak, at the working energy; a low percentage permits us to adopt a narrower energy window, which will allow rejection of more Compton photons, detected with energies below 511 KeV.

- **Output light-yield** $[\Phi = \frac{photons}{MeV}]$: it gives the number of visible photons generated in the scintillator per MeV absorbed energy. A high value is required to couple more scintillator crystals to a single PMT, in order to reduce cost and obtain better spatial resolution.

- **Maximum sensitivity wavelength** $[\lambda = nm]$: it is the wavelength at

which the number of visible photons emitted by the scintillator is at maximum, and a value very close to 400 nm will give us the possibility to exploit the best PMT performances, since its peak efficiency is usually located around that wavelength.

- **Refractive index** $[n]$: the refractive index is very important, because we want to minimize photon loss due to reflection at scintillator-PMT interface; since the reflection coefficient for the energies can be, with a fast approximation, referred to $\mathcal{R} = (\frac{n_1 - n_2}{n_1 + n_2})^2$, where, in our case, n_1 and n_2 are corrispondigly the scintillator and the PMT photocatode refractive indeces, it is desirable to use scintillators with n_1 very close to 1.5, the value of Bialkali photocatode refractive index.

- **Hygroscopy** : the material tendency to absorb humidity can lead to decreased efficiency as well as erratic behavior of the crystals. Even though the crystal material is well isolated, it must be taken into account.

The material, which shows the best overall performance, according to the above described parameters, will be the right one for PET applications; anyway, before describing and comparing the different scintillator materials, it is useful to investigate the physic behind the scintillation process, using the models provided by the band theory.

2.5.1 Scintillator band theory

Band theory is a powerful tool to understand the phenomena in solid state electronics, and it may be a useful tool even to investigate more deeply for the scintillation process. As in semiconductors used in electronic devices, also in scintillator materials, carriers are allowed to populate a discrete number of energy states and, more precisely, at rest the electrons fill in the valence band, while the holes stay in the conduction band; the two bands are divided by the forbidden band, the band gap, which, in almost all scintillators, ranges from 4 to 13 eV[33].
Intrinsic scintillators are, anyway, never employed: the absolute absence of energy states in the forbidden band, causes the electron-hole recombination to produce a photon with energies in the Ultra Violet (UV), far away from the 400 nm value we would like to have; this is the main reason why all the scintillators are doped with elements like Tallium(Tl), Cesium(Cs) or

Cerium(Ce), which introduce dishomogeneities in the crystal lattice, evident, in the energy-momentum diagrams, as energy states in the band gap (the so called "activator sites"). The appearence of these energy states in the forbidden band makes most of the excited electrons, first relax from the bottom of the conduction band to these states, and then the electron-hole recombination emits a photon with a wavelength close to the desired value.
A detailed description of the scintillation process is however needed: the absorption of the γ photon by the material causes several electrons to be excited up to the conduction band, and, dually, several holes to be promoted in the valence band; immediately after the excitation, all these carriers have an energy that allows them to overcome the band gap, and also with kinetic energy, which causes those carriers to be called "hot carriers".
The extra kinetic energy held by carriers is lost, through collisions with crystal lattice, and the time required for this relaxation is negligible[25]; when the relaxation of the carriers has taken place, the electron is at the bottom of the conduction band, while the hole is at the top of the valence band; then, the electron continues to relax, but, this time, the relaxation drives it to the energy states, introduced through doping, in the band gap. Once here, the electron will recombinate with a hole on the top of the valence band, with a half-life of $10^{-7} \div 10^{-8}$ seconds and a visible-light photon will be emitted[25].
The process just described would be the ideal event, that should occur in scintillators; unfortunately, quite often, real materials are filled with impurities, which worsen the performance of scintillation.
Material impurities cause the existence of the so called "killer states", which are located in the middle of the band gap, and introduce further activator sites: those additional states can give rise to recombinations with lower energy transitions, which means photons emitted with higher wavelength.
Other impurities give birth to energy states called "electron traps", just below the conduction band (not more than 0.5 eV below the conduction band[33]), whose effect is to trap the electron temporarily and then release it, only after the carrier absorbed the necessary thermal energy for the promotion; soon after the excitation back in the conduction band, the charge will be able to undergo the recombination process and all this will lead to an afterglow not useful for our "fast" purposes, but exploited in dosimeters, where the measurement of such afterglow gives us the chance to quantify absorbed radiation[33].
With this simplified explanation of the scintillation process through the band theory, we can now relate, with a first order approximation, the decay time constant to the time required to the carrier to recombinate once in the acti-

vator site, or the emitted photon wavelength to the energy transition in the electron-hole recombination, or the output light-yield to the number of valid recombinations.
Finally, a comparison between different scintillator materials used in γ-ray detection, highlighting pros and cons, is now possible.

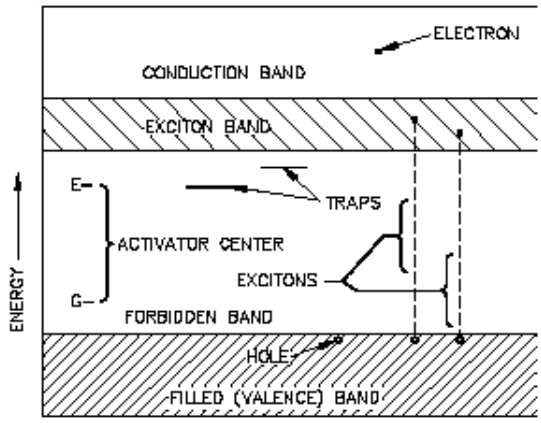

Figure 2.7: The simplified band diagram for scintillators; the excitation bound has not been treated, because it is not fundamental to the scintillation process[34]

2.5.2 Scintillator materials for PET

In the following paragraph, the performance of the most used scintillator materials in PET will be analyzed, with particular attention on the new solutions, which have been proposed in the last years.

- **NaI(Tl)** : Thallium activated Sodium Iodide is the scintillator material, which, more than others, have changed Medical Imaging history, and still is the most used crystal in SPECT; the major pro of this material is related to its big output light-yield ($\Phi = 38 \cdot 10^3 \frac{pho}{MeV}$), which allows high SNR in the signal processing, but its long decay time constant (τ = 230 nsec), its hygroscopy and a not negligible afterglow have made NaI(Tl) a scintillator unfit for high-performance PET cameras.

- **BGO** : for long time, Bismuth-Germanate (chemical composition: $Bi_4Ge_3O_{12}$) was the chosen scintillator for coincidence detection; its high density ($\varrho = 7.13 \frac{g}{cm^3}$), coupled to a high effective atomic number (Z =

75), ensured a high probability of absorption due to phoelectric effect ($P_{photoelectric} \approx 40\%$). Unfortunately, this material has a very long time decay constant ($\tau = 300$ nsec) and a cosiderably low output light-yield ($\Phi = 8 \cdot 10^3 \frac{pho}{MeV}$), which do not allow for high count rates and cost reductions; all those factors, along with the peak emission wavelength located at 480 nm, have made the material lose its appeal to PET applications, even though it can be easily fabricated with the Czochralski process.

- **GSO(Ce)** : Cerium activated Gadolinium Silicate (chemical composition: $Gd_2SiO_5 : Ce$) presents, more or less, the same value of BGO, in terms of density ($\varrho = 6.7 \frac{g}{cm^3}$) and effective atomic number ($Z = 59$), but, theoretically better time resolution performances ($\tau = 56$ nsec); despite this, the scintillator is not widely used for PET detectors, due to the time decay constant dependence on doping levels (τ for this reason can range from 56 to 400 nsec) and due to its fragility, which does not allow economic fabrication process.

- **LSO(Ce)** : Cerium doped Luthetium Oxyorthosilicate (chemical composition: $Lu_2SiO_5 : Ce$) is today the most interesting scintilllator material, in the field; its performance is comparable to that of BGO, in terms of density ($\varrho = 7.4 \frac{g}{cm^3}$) and effective atomic number ($Z = 66$), but the high output light-yield ($\Phi = 27.3 \cdot 10^3 \frac{pho}{MeV}$), the fast decay time constant ($\tau = 40 nsec$) and the peak emission wavelength ($\lambda_{peak} = 420$ nm), close to the required value, have made it the best material currently used in coincidence detections. The presence of the radioactive ^{176}Lu, in a low percentage (2.6%), introduces naturally occuring β^- decays in the scintillator, whose effect can be neglected, since the decay rate ($300 \frac{decays}{sec \cdot cm^3}$) is much lower than the ones commonly encountered in PET studies (10^6). Its fabrication follows the Czochralski process, and also with this material, we experience an unstable τ, whose value is usually in the interval between 29 and 43 nsec; the major drawback for this scintillator, is its extremely high cost ($50 \frac{\$}{cm^3}$).

- **LYSO(Ce)** : it has basically the same structure as that of the previous scintillator (its chemical composition is $Lu_{0.6}Y_{1.4}SiO_5 : Ce$) and, indeed, it ensures similar performance, with a decay time constant of 53 nsec; the fabrication process is much more stable and its cost is lower than LSO. The deficiency is in the energy resolution ($R = 12\%$), and

this is the reason, why it is commonly used as a second layer scintillator, in the Depth of Interaction (DOI) detection.

Below, the parameters for each of the major scintillation materials are presented. [25][31][32][33]

Figure of merit	$NaI(Tl)$	BGO	GSO	$LSO(Ce)$	$LYSO(Ce)$
$\varrho(\frac{g}{cm^3})$	3.67	7.1	6.7	7.4	5.37
Z	51	75	59	66	54
$l(mm)$	29.1	10.4	14.1	11.6	20
$\tau_{decay}(nsec)$	230	300	$56 \div 400$	40	53
R_{energy}@$662KeV$	6.5%	9%	7.8%	8.2%	12%
$\Phi(\frac{ph}{MeV})$	$38 \cdot 10^3$	$8.2 \cdot 10^3$	$8 \cdot 10^3$	$27.3 \cdot 10^3$	$26 \cdot 10^3$
$\lambda_{peak}(nm)$	415	480	440	420	420
n	1.85	2.15	1.85	1.82	1.81
$P_{photoelectric}$@$511KeV$	17%	40%	25%	32%	18%
Hygroscopic	Yes	No	No	No	No

Table 2.1: Detailed performances of each scintillator discussed

Lately, thanks to the big improvements made in electronics processing times, the scientific community has put great effort in incorporating Time Of Flgiht (TOF) data detection and reconstruction processes; for scintillators, it is necessary to find materials able to match the speed of light for the photon ($30\frac{cm}{nsec}$) and great interest has been shown to crystals like CsF and BaF_2, which provide τ, corrispondigly, in the order of 4 nsec and 0.8 nsec (BaF fast component)[32], but not all their performance parameters are sufficient and this is also the reason, why they are not widely used yet.

2.6 PMT

After the γ photon has been converted to visible light photons, the signal is ready to be amplified by the PMT.
The role of PMT is to convert the weak photon signal into a strong electrical signal, without adding too much noise, in order to be able to process the information correctly.
The working principle of this detector is simple: when the visible light photons hit the PMT sensitive layer (the photocatode), a number of electrons

are released by the interaction; this is due to the energy delivered by each one of the incident photons, which, if everything works as supposed, creates electrons free to migrate within the tube.
To be more specific and using, once again, the band theory, when the visible light photon is absorbed by the material, it supplies its energy (approximately 3 eV[25]) to one electron in the valence band, which is excited to a higher energetic level; the energy, given from the photon, will consist of kinetic energy that the carrier will partially lose through collisions with the photocatode crystal lattice, and in the energy requested to overcome the material band gap. The process just described is named the photoemission process and it is what happens during the first steps of the conversion into an electrical signal; subsequently, those emitted electrons are accelerated by a stong electric field, which will cause them to acquire the necessary kinetic energy to make the next step possible: the secondary electron emission. The secondary electron emission is the process in the PMT that multiplies the number of electrons released and thereby amplifies the signal.
The electrons that escape the photocathode, have an average kinetic energy of 1 eV[25] and, thanks to the present high voltage drop (around 100 Volts), they acquire the needed energy (approximately 100 eV) to ionize atoms, when they will hit the dynode material.
Applying voltage drops equal to 100 Volts, the secondary electron emission is a process, which, theoretically, would be able to produce 30 carriers per incident electron ($\frac{kineticenergyheld}{energytoionizeanatom} = \frac{100}{3}$), but unideal factors lower the number to $5\frac{ionization}{electron}$ and require this process to be repeated at least ten times, to obtain overall gain in the order of 10^7[25].
The previously described phenomena explain why a PMT is made up of a photocathode connected to a vacuum tube, which, on the inside is divided in multiple stages (a PMT has usually 10-12 dynodes): this way the photocatode, when hit by visible light photons, will allow the photoemission, while the voltage drop, applied between each (in the same way as between the photocatode and the first dynode), will make the secondary electron emission take place, within adjacent stages, and electron multiplication possible.

The overall gain of the PMT is then given by the product of the gain of every stage, and this results in[25]:

$$G = \alpha\delta^N \tag{2.12}$$

In equation 2.12, the overall gain (G) is equal to the product of a parameter depending on PMT geometry (α, it is usually very close to one) and

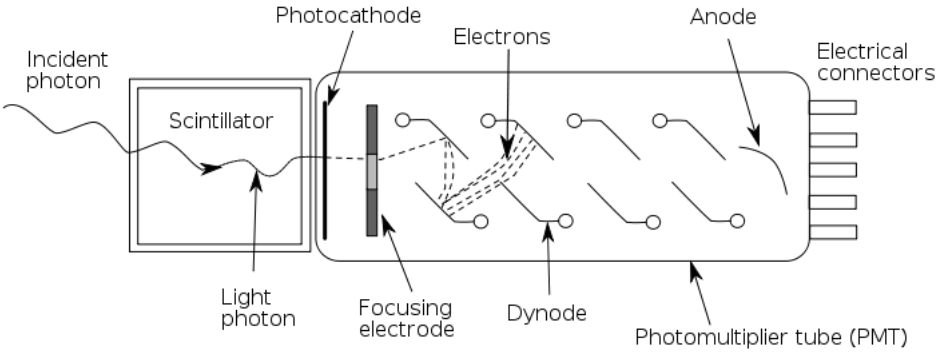

Figure 2.8: The general layout of a scintillator coupled to a PMT[35]

the gain of one stage (δ) raised to a power equal to the numbers of dynodes (N), assuming the gain at each stage being identical. With a feasible gain between two dynodes close to 5, it can be computed that, to achieve a gain of 10^7, 10 multiplication stages are needed (since $5^{10} \approx 9.76 \cdot 10^6$[25]).
It is also vey important to consider PMT noise characteristics: since the electron multiplication is a stocastic process, the number of carriers emitted, per single incident electron, fluctuates around the expected value, at a first order approximation following the Poisson statistics. The fluctuations consist of a continuously slightly varying overall gain that will add to the signal noise at each dynode. It will not be investigated more deeply in this manuscript, but it can be demonstrated that noise performance is strongly influenced by the first multiplication stage, and, usually PMTs have an higher voltage drop between the photocathode and the first dynode, in order to minimize noise variance.
As discussed in the scintillator paragraph, the PMT peak sensitivity is located at a wavelength around 400 nm, since the quantum efficiency (QE), for Bialkali photocathodes, reaches the maximum of 20-30% at that wavelength, as shown in Figure 2.9; it means that only 3 out of 10 visible light photons are converted into electrons, and this makes us understand the need of a strong signal amplification.
The PMT is polarized with a voltage divider, providing, at the same time, the voltage drop to accelerate the free electrons and the current to supply the dynodes the electrons taken away; to avoid sudden peak signals to reduce PMT gain, which is strictly depending on the applied voltage drop, capacitors are often put in parallel to resistors, to provide extra electrons to the

dynodes.
To conclude the section about PMTs, it is important to say that in the last period greater interest has been moved toward APDs than toward PMTs: the photodiode performance is not affected by magnetic fields, while, in the tube, the presence of a magnetic component makes the electrons deflect, due to Lorentz law, worsening the electron optics, which had been carefully designed, to have the best performance possible: then, the use of PMTs is extremely limiting for future systems, aiming to combine PET and MRI in a single scanner.

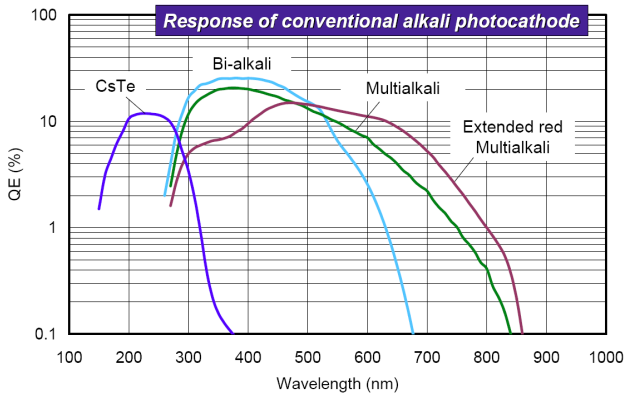

Figure 2.9: Photocathode QE for different materials[36]

2.7 Reconstruction algorithms

Once the information has been detected, it needs to be reconstructed to obtain the 3D images, which show us the radiotracer distribution within patient's body.
The reconstruction is done using algorithms, that will be introduced later on in this paragraph, but before being able to discuss their characteristics, it is very useful to point out how data are organized in the scanner.
When an annihilation is detected, it can be recorded and stored using two different strategies: either the frame duration is fixed before the acquisition, and the events are sorted into the appropriate LOR bin, depending on the frame and detector pair, or frame duration is decided after the scan, depending on count statistics, and coincidences are organized in a list mode file, where each event is recorded with timestamp and detector information.

In the first case, since the frame duration is known a priori, we have the advantage of knowing the number of frames we will have, but, at the same time, we will have to accept frames with very different SNRs; on the other hand, the list-mode data acquisition (the one with no fixed frame duration) will allow for a better and more thorough data handling, possibly better SNR, overcoming statistics-time resolution tradeoff, typical of dynamic studies[37]. In both cases, subsequently, data are organized in sinograms, which are matrices where all the projections from the same transaxial plane are stored, depending on the angle of projection and on the position in that projection. The name "sinogram" comes from the observation that a point in the transaxial plane follows, throughout the matrix, a sinusoidal path in the position coordinate, depending on the projection angle, as shown in Figure 2.10.

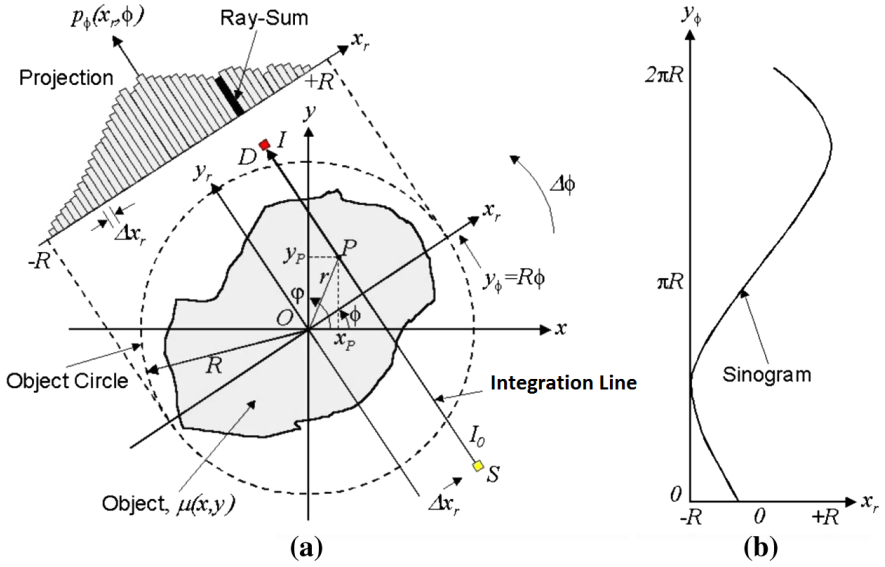

Figure 2.10: The projection corresponding to angle ϕ (on the left) and the sinogram, focused on point P position, built up with projections for all angles (on the right)[38]

Once data organization has been discussed, only one more thing needs to be introduced, prior of proceeding with reconstruction algorithms: data acquisition modalities; when we want to detect coincidences, we have to choose between the 2D and the 3D data acquisition modalities. In the 2D configuration, lead collimators, called "septa", are interposed between different transaxial planes, with the purpose of detecting only LORs belonging

to the same layer, perpendicular to the patient's main axis, and discarding all oblique coincidences; in the 3D modalities, septa are removed and coincidence events from oblique planes are accepted. In the past, the 2D data acquisition was the preferred one, because 3D modalities needed prohibitive processing times, required impossible memory allocation and scattered photons used to corrupt substantially the reconstructed information, but, as electronics heavily reduced processing times, improved storage capabilities, and scatter rejection has become efficient, 3D protocols have become the normal mode of operation, allowing much higher sensitivity and, then, much higher SNRs[39].
After having introduced all the preliminar needed knowledge, we are now ready to treat the main reconstruction algorithms for PET; nowadays, the methods to reconstruct the detected data are two: through analytical algorithms or through iterative algortihms.

- **Analytical algorithms** : they are generally faster but less accurate compared to the iterative ones, and they rely on a very simple mathematical model of the imaging process[40].
  Without any doubt, the most known and used one is the Filtered-Backprojection (FBP), whose concept for the reconstruction is to consider the number of photons, recorded in any given detector bin, as the sum of contributions of the activity located on the straight line perpendicular to the surface of both detectors (LOR); the recorded data is backprojected, redistributing all the measured photons along the correspondant LOR, but prior of doing this, the projections are convolved with a sinc filter, in the space domain, in order to avoid star artifacts in the final image[40].
  The problems with this class of algorithms is that they do not take into account statistical noise in the data, which has a dramatic impact on the information, and, to limit noise at high frequencies, they require low pass filtering, which degrades image spatial resolution (the detailed reconstruction algorithm for FBP is treated in Appendix A). All this results in lack of accuracy and artifacts, which degrade the information we have and can lead to erroneous conclusions.

- **Iterative algorithms** : this family of algorithms takes into account the statistical nature of the process and, mainly for this reason, converges to the solution much more slowly than analytical methods, but with a much higher quality of the image[40].

The name of the category comes from the fact that the reconstruction involves iterations, to update the resulting distribution within the FOV, in an attempt to obtain the optimization of a criterium, which reflects reconstruction trustfulness.
Almost in all the iterative algorithms, five basic concepts are present[39]: the image model, which is the structure of the final reconstruction (grid of voxels); the system model, which is nothing else than the carachterization of the imaging system (the probability of a photon from a certain voxel to be detected by a certain LOR); the model for the data, which gives the relationship between the measured data and the expected values (the statistics of fluctuation around the expected value); the cost function, which is the criterium to optimize (for example the likelihood function), and, finally, the algorithm to implement, which determines the steps to be followed at each iteration.
The system model of the algorithm is most of the times the so called "system matrix", a matrix stored in the scanner memory, which contains the details of the camera capability to detect photons emitted from a certain position, accounting for attenuation within the FOV, photon detection statistics and models for random events[40]. The content of this matrix, which can also be obtained through Monte Carlo simulations of scanner behaviour[41], is the main cause of the big step forward in image quality compared to analytical methods, since it gives the reconstruction a way to take into consideration the fluctuating nature of the data generation process.
Considering the data model, the statistical model often used in these kind of algorithms is the Poisson statistics, able to describe well photon detection variations, and/or the Bayesian method[39], to exploit the information a priori we have on radiotracer distribution (non negative value and low variations among neighboring voxels).
The cost function typically used is the likelihood function, while the algorithm usually employed is the expectation maximation (EM); indeed, currently the iterative algorithm most widely implemented is the Ordered Subset Expectation Maximization (OSEM), which is simply a development of the Maximum Likelihood Expectation Maximization (MLEM).
The MLEM is a reconstruction method, which searches for the best solution iteratively, starting from a very unlikely radiotracer distribution (for example uniform within the FOV); at each iteration, the estimation is updated, with a multiplicative correction factor, obtained from the

comparison between the real measured data and the data, which would be measured if we really would have the estimated distribution[40]. The iterations will stop whenever the maximization of the likelihood function is achieved, but anyway a detailed description of the algorithm is given in the Appendix B.
The problems of MLEM algorithm are in the number of iterations required to arrive at the convergence, and the nature of the final solution: on one hand, the number of cycles required to reach the optimum is usually within the interval $[20 \div 50]$[39], which would require long times to be calculated, while, on the other hand, the retrieved distribution might be a local minimum of the error variance, that we are trying to minimize, not a global minimum, and, then, the solution, we have obtained, would not be the best one achievable.
To solve the first of those two problems, the OSEM algorithm has been introduced; with OSEM, it is possible to divide the dataset in different subsets, applying the MLEM algorithm to each one of them, and make the final solution built up by the single subset solutions. In such a way, we are able to have the algorithm converging to the solution, in a number of iterations equal to the ones needed in MLEM, divided by the number of subsets, but, still, we are not able to solve the convergence problem, and this uncertainty is the reason why iterative algorithms are not the only reconstruction methods implemented.

Chapter 3

HRRT

The HRRT is currently the most complex positron annihilation based scanner, and it has been developed at CPS Innovations, Knoxville, TN[42].
The HRRT is a camera designed for studies on brain metabolism and radioligand binding to neuroreceptors[4], using human and primate brains, and its high performances in reconstructing radiotracer distribution are achieved with a particular design, able to set up $4.5 \cdot 10^9$ LORs and to provide great efficiency in annihilation detection[42].
In the next paragraphs, the camera layout, the materials and the devices used will be treated in an attempt to explain every design solution and the reason why that solution has been chosen.

3.1 HRRT layout

The HRRT camera is a scanner, which is based on the PET working principle, but with a different layout: being used only for brain studies, and not for whole body scans, this tomograph has smaller sizes, but also design solutions completely different from the ones previously implemented on PET cameras. First of all, the volume, where we want to carachterize radiotracer distribution, has been reduced compared to the one commonly experienced in clinical PET: indeed, HRRT FOV is $312 \times 312 \times 252\ mm^3$, much smaller than the dimensions of usual whole body scanners[43].
Apart from scanner dimensions, the major change is in the detector layout, which is no more circular around patient's body, but consists of 8 heads, arranged in an octagonal layout surrounding the FOV; this design solution has been chosen to reduce the parallax problem, arising from the angle of incidence of the photon with the scintillator surface. If, after the positron

capture by the electron, the photon is absorbed by the detector with a non-perpendicular angle, it can penetrate different neighboring crystals, not remaining confined only in one, and it can give the system a wrong interpretation of the LOR, along which the annihilation took place; the parallax problem results in an incorrect spatial information, and so the octagonal arrangement is conceived to reduce the effect of an angle of incidence smaller than 90^o, which is the source for this resolution degradation.

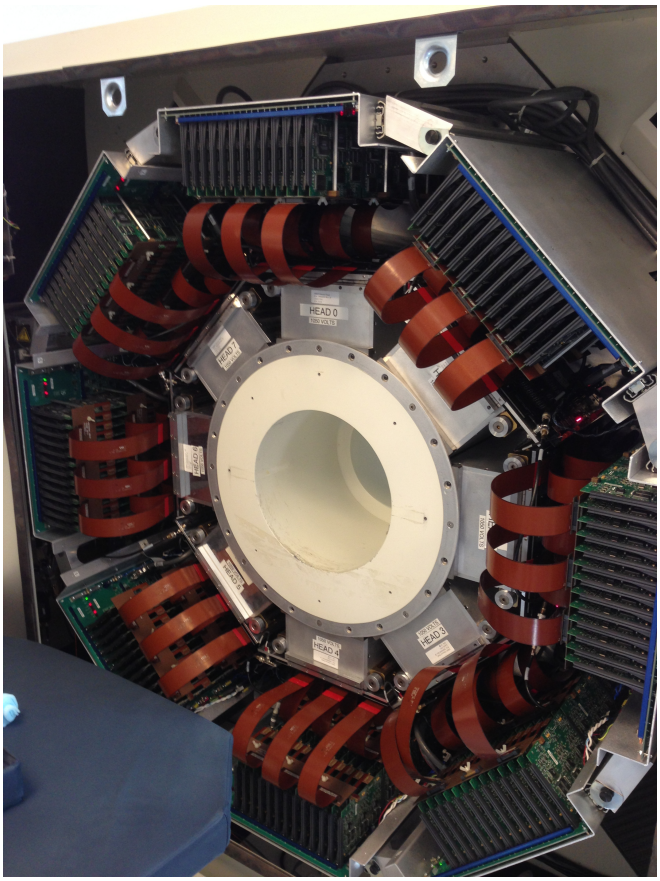

Figure 3.1: The camera layout, once the external protection has been removed

The octagonal layout is not the only way employed to solve the parallax problem: to reduce its corrupting effect, each head is designed with two different scintillator layers, with the scintillator materials differing in the decay time constant. The use of two layers allows for the implementation of DOI, a technique able to reconstruct the real LOR of the detection, analyzing the

trajectory held by the radiation beam within the 2 scintillators and differentiating the detections at different depths, just looking at the different decay times of the output electrical signals[44].
In order to achieve a very high spatial resolution, every head of detectors is composed of a matrix of 13×9 crystal blocks, repeated on both scintillator layers; each crystal block is given by 8×8 crystals, where the size of every single crystal is $2.2 \times 2.2 \times 10\ mm^3$. The separation between neighboring crystals is 0.2 mm, while the 0.5 mm space between neighboring crystal blocks is filled with a reflecting material, in order to minimize photon scatter into adjacent blocks. Each detector block is coupled to a 2×2 array of PMTs, following a quadrant sharing design and resulting in a 14×10 PMT array for the whole head; the precised position of detection is retrieved, merging the information coming from the 4 PMTs of one block and using the typical calculations from Anger logic.
However, the choice of adopting an octagonal design is the source of a particular problem during the reconstruction: unlike PET circular gantry design, the use of eight separated heads results in missing spots in the sinograms, due to the projection angles corresponding to the gaps present between adjacent heads. Obviously, missing projection angles will make the FBP reconstruction very hard to implement, and, indeed, the main used reconstruction algorithm is a development of the OSEM, as it will be explained later, though many studies have been led to have the chance to use the analytical algorithm even on the HRRT.

3.2 Scintillator

In the HRRT, two different scintillator materials are used. As already stated previously, the camera detector system is built up by two scintillator layers, and those two layers can be either $LSO_{fast} - LSO_{slow}$ or $LSO - LYSO$; in the first case, we exploit the unstability of LSO fabrication process to create the fast layer and the slow layer, while, in the second case, we use, explicitly, the difference in the decay time constant between the two materials. Usually the preferred solution is the $LSO - LYSO$ coupling, which allows to reduce costs, without losing considerably performances, and the LSO, employed on HRRT, must have a decay time constant not longer than $43 \div 44$ nsec, in order not to be too close to LYSO value.
The DOI information is retrieved in a very simple way. The output of the detector is connected to a pulse-shape discriminator, able to separate the

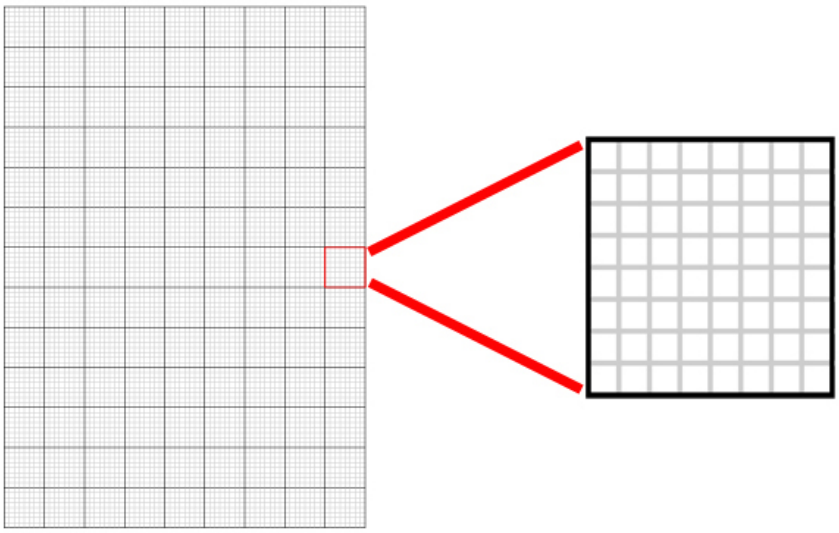

Figure 3.2: One of the 8 heads delimiting the gantry (on the left) and detail of the crystal bock (on the right)

fast component, belonging to the first scintillator layer, from the slow component, belonging to the second one. If we are able to differentiate the two components, we are also able to determine the fraction absorbed both in the first and in the second layer; with the present information, we can then determine the trajectory of the γ-ray and compute the correct LOR, observing, for example, that in case of a perpendicular angle of incidence, the fraction of gamma photons absorbed by the second layer will be much higher than in case of oblique beams.
The energy window used accepts the coincidence event as valid, only if the energy of the detected photons is within the interval $[400, 650KeV]$, with the value of 511 KeV almost exactly in the middle; the energy window width depends on the energy resolution of the scintillator, and, in effect, LSO and LYSO show, corrispondingly, a resolution of 8.2% and of 12% at 662 KeV.

3.3 PMT

Coupled with the second scintillator layer, through a waveguide, Hamatsu R1450-12 PMTs are the detectors chosen, to output the electrical signal, and this particular model of PMT is deliberately designed for Positron CT and high energy physics[45].

The device has a bialkali photocatode, whose spectral response (Figure 3.3) ranges from 300 to 650 nm, with the maximum response wavelength at 420 nm: the ideal situation, if we want to couple it with LSO/LYSO scintillators. The signal amplification is achieved through 12 multiplication stages, and the chosen voltage drop, applied across the detector is 1050 Volts, with the purpose of obtaining an overall gain higher than 10^5, as shown in Figure 3.3. The polarization of the detector is done with a voltage divider, where the resistor between the photocathode and the first dynode is twice as the other resistors used; this way, the first multiplication stage, at the same time, will have an higher multiplication factor and will provide the detector better noise performances, since we know that the first stages have a dominant role, in determining the dark current at the output.

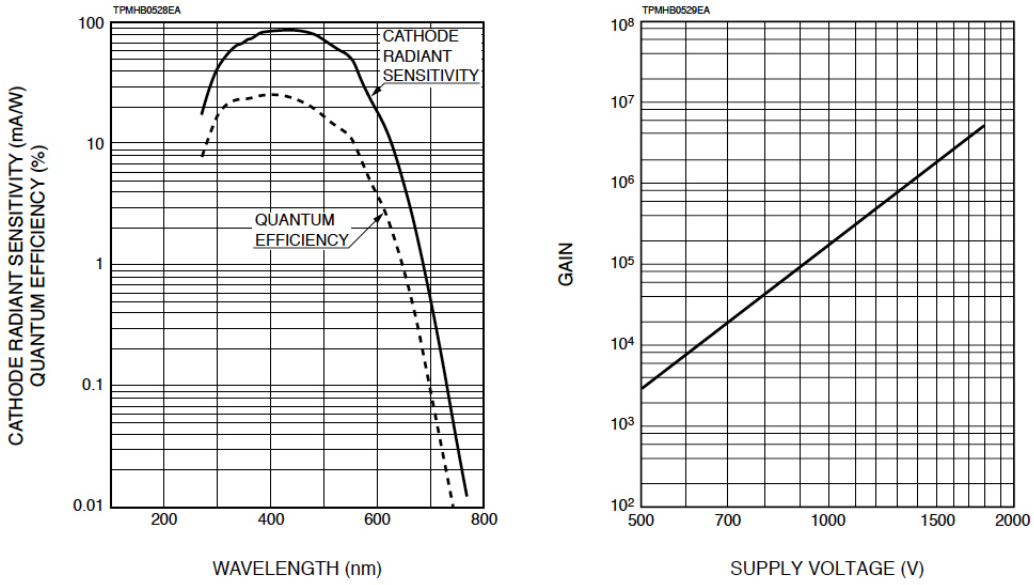

Figure 3.3: Spectral response of the photocathode (on the left) and the gain-voltage drop plot (on the right)[45]

3.4 Camera calibration

Some calibration procedures have to be performed periodically to make the information, output by the scanner, reliable; the behaviour, which we have described so far, is only ideal, and unexpected situations have to be compensated for, to obtain the desired data. In particular, scintillator and PMT

performances affect heavily the retrieved position of interaction and different outputs for the same event, among detectors, result in the wrong spatial information. To avoid that all these factors corrupt image information, the camera calibration is carried out once/twice a year, performing some corrections needed to optimize HRRT performances.
The first setup operation relates to the scintillators: the crystal absorption spectrum is a tool used indirectly by the PHA, when it analyzes for the energy of the detected photons and decides wheter the detection is valid or related to a Compton interaction; if the energy window is used but crystal spectral performance is not optimal, the system may accept events that should be discarded. It follows that, during calibration, the spectral performance of each crystal is tested and the thresholds of the energy window for that scintillator are manually shifted from the default values set by the system (Figure 3.4), in order to make the camera accept only photons with energies included in the desired interval.

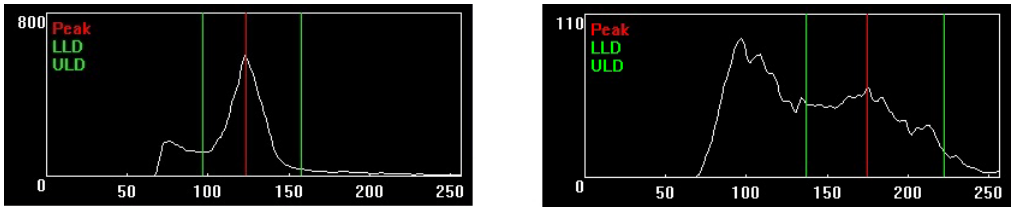

Figure 3.4: Cristal absorption spectra: the crystal on the left does not need correction while the one on the right has been corrected (energy window: red line=center value, green lines=lower and upper thresholds); energy units shown are directly proportional to the real ones

It is also very important to check out for PMT responsivities, to avoid spatial distortion: indeed, if one tube has a spectral response which differs from the spectral response of an adjacent PMT, since Anger Logic retrieves the position assuming the same performance for all detectors and signal intensity inversely proportional to the distance from the position of interaction, a photon, absorbed in a point equidistant from the two PMTs, will cause a stronger signal in one of them, and the reconstructed point will not be placed equidistant from the two detectors. To correct for this kind of mismatch, spectral performances of the PMT have to be investigated and adjusted for, and a correction is applied to mimic correct response, as shown in Figure 3.5.

Figure 3.5: The spectral response for four neighboring PMTs (A-B-C-D), before (on the left) and after (on the right) correction

Once this is done, the improvements are evident: the matrix of point sources detected by the heads was, prior of camera calibration, spatially distorted and what should have been similar to a square, was clearly deformed; after correction, the point sources are viewed as a less distorted matrix but it is still not enough. For this reason, a manual correction for every single crystal block is carried out, with an operator moving the detection peaks (the brightest spots in Figure 3.6) to the positions where they would be supposed to stay; in such a way, the system records the needed geometrical transformation to map the real output into the desired output, and, in future reconstructions, the camera will apply it to the raw data, in order to correct for the distortions introduced by the imaging system.

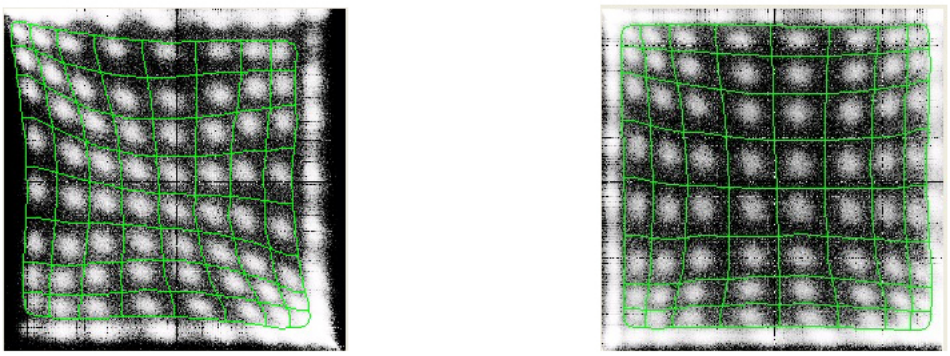

Figure 3.6: Example of a crystal block which requires correction (on the left) and of the desired output (on the right)

3.5 Reconstruction

On the HRRT, once data are organized in sinograms, the reconstruction follows the 3D Ordinary Poisson OSEM (3D-OP-OSEM).
The 3D-OP-OSEM has been introduced to reduce the problem arising from negative values in recostrunctions: since only positive values can be physically accepted (the output datum is a distribution), when, at the last iteration, a negative value is encountered, in the 3D-OSEM, the voxel is truncated to zero and this is the source of positive bias in the final image[46][47].
The reconstruction method, implemented on the HRRT, is just an evolution of the 3D-OSEM already treated, and the step forward consists in performing attenuation, random and scatter correction, not at the end of the reconstruction, but at each cycle, with an additive term in the iterative equation of the algorithm[46].
To be more specific, the attenuation correction is done by acquiring a transmission scan, with a ^{137}Cs point source, which rotates around the FOV and defines the attenuation coefficient for every voxel. On the other hand, for what concerns random detections, the coincidence time window to detect a valid annihilation is set to 6 nsec (approximately twice the time needed for a photon to cross the FOV), from the first photon detection, and the correction for random events consists of delaying, with a time lag much longer than the concidence time window, the channel, which detected the first photon; in such a way, a 6 nsec time window will record all the coincidences with that detector, starting from the delayed one, but they will be used to build a sinogram, to be subtracted from the one acquired without delays, and the so called "delayed coincidence window" method will permit to reduce the impact of random detections. Last but not least, the scatter correction is performed with the known energy window implemented in the PHA, in order to discard all coincidences, which may corrupt our information.
In this paragraph, some important information for data reconstruction on HRRT has been treated. In the next chapter, an experiment will be described to assess camera Point Spread Function (PSF); in the last period, great interest has been given in obtaining scanner PSF, because new algorithms, which exploit the information present in the PSF, have been able to achieve higher resolutions and, then, allow better investigations[4].

Chapter 4

Resolution assessment

As discussed in chapter 1, the spatial resolution of reconstructed data is depending on camera performance and patient's movements: motion correction is required, only when the scanner resolution is high enough to make the patient's movements a limiting factor.
For this reason, in the following chapter, an experiment to assess camera resolution will be described and its results will be used to underline the need for motion correction and to explain how the knowledge of this spatial performance can be used, if properly retrieved, to enhance image quality during reconstruction.

4.1 The Experiment

Normally, the procedures followed to determine camera resolution are incorporated in NEMA NU2-2001: the standard protocols employed to test the performance of positron emission tomographs.
According to the previoulsy introduced standards, the measurements to test spatial resolution are taken by imaging point sources in the air and reconstructing the data with a FBP algorithm, void of any smoothing or apodization; the results of this procedure are used to compare performance between different scanners, but it does not give us the real resolution that we will experience with human or animal samples[48].

The measurements are done by filling a capillar (carachteristics: 2 mm outer diameter, 1 mm inner diameter) using the ^{18}F radionuclide and taking acquisition of six different points along the radial direction, holding capillar main axis parallel to the tomograph main axis: the purpose is to obtain camera resolution through its Point Spread Function (PSF), which is dependent

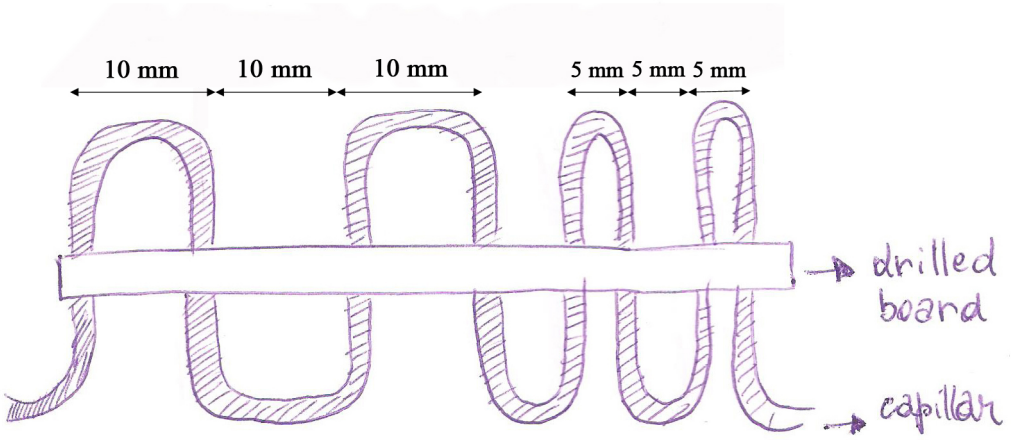

Figure 4.1: Drawing representing the setup used in the resolution assessmente experiment

on the distance from the center of the FOV, and to discuss performances of different scanners using the FWHM or the Full Width at Tenth Maximum (FWTM) of the retrieved PSFs. The procedure is then repeated on many transaxial planes, until the camera PSFs, along the three dimensions, can be reconstructed.
However, we decided to lead a different experiment to investigate the spatial performance of the HRRT, and, in particular, we have been interested in testing resolution trend in the radial direction and in proving that spatial resolution of modern brain scanners is comparable with the magnitude of patient's movements, during scans.
The experiment was very straightforward: we filled a capillar (carachteristic: 0.7 mm inner diameter) with 60 MBq of ^{18}F and positioned it in a plexiglass board, drilled on both sides: on one half, the board was drilled every 5 mm, while on the other half it was drilled every 10 mm (Figure 4.1). We collocated the radioactive sample in the HRRT FOV; then we took the emission scans and reconstructed the data with the usual 3D-OP-OSEM algorithm.

Once reconstructed, we studied the data in an attempt to find the best strategy to investigate spatial resolution.
The data were pretty satisfying and we chose a software like MATLAB to process the image and obtain the information that we were interested to; the choice of this development environment was strictly related to the possibility of handling the output volumes as 3D matrices.
In particular, from the data that we obtained after the reconstruction, we

Figure 4.2: Configuration adopted to test HRRT resolution

saw that the most reliable configuration was the one given by the board half, drilled every 5 mm: indeed, it permitted us to have a more frequent spatial sampling of the FOV in the radial direction and, as it appears clear in Figure 4.2, the assumption of having tube orientation parallel to tomograph main axis had more probability of being verified, within a certain tollerance. The 2D matrix, from which we derived our results, was obtained averaging the three adjacent transverse planes where the reconstructed activity was stronger, in terms of voxel value; then, to reconduct the 2D matrix to one dimensional array of values, we averaged the 5 lines of pixels (Figure 4.4), which had the detection peaks and plotted the pixel values from this array, obtaining the shape of Figure 4.3.

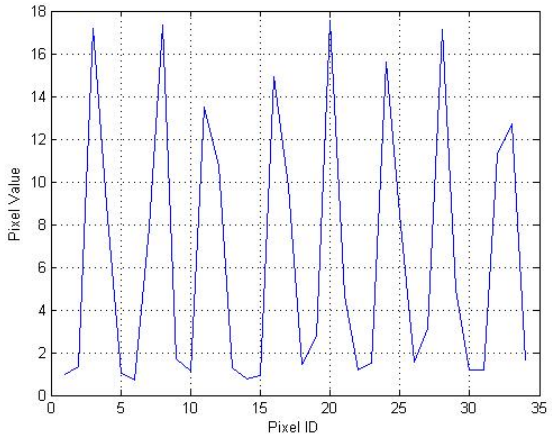

Figure 4.3: One of the plots displaying the detection peaks

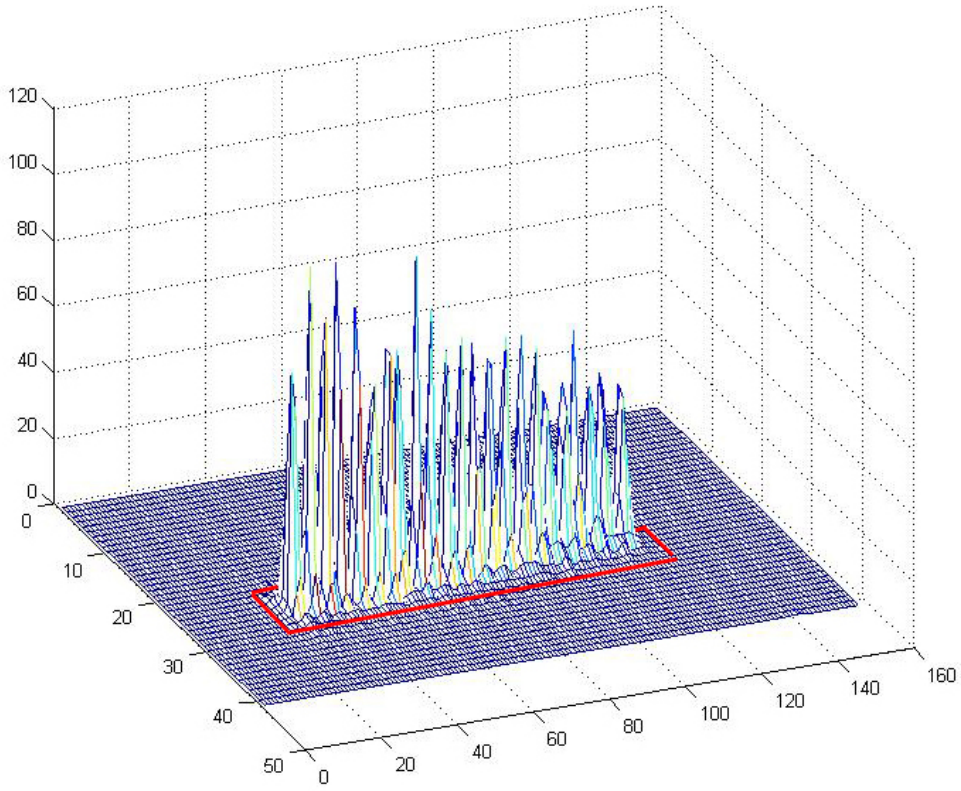

Figure 4.4: the 3D rappresentation of the intermediate step between the initial volume and the array of peaks: here is dispayed the 2D matrix containing the result of the average on the three transverse planes of the initial volume. Subsequently, the 5 lines of pixels in the red rectangle have been averaged to obtain the final array of peaks

4.2 Results and discussion

From the monodimensional array of values, we were then ready to retrieve the FWHM of the peaks, using the FWHM.m script, implemented to obtain the distance in millimeters between the two half values of the same peak: this way, we were able to study the data, analyzing the 22 FWHM values reported in correspondance of the distance from the center of the FOV, and we plotted this dependence to highlight the trend of the camera of having poorer spatial resolution as distance from the center increases, as shown in Figure 4.5.

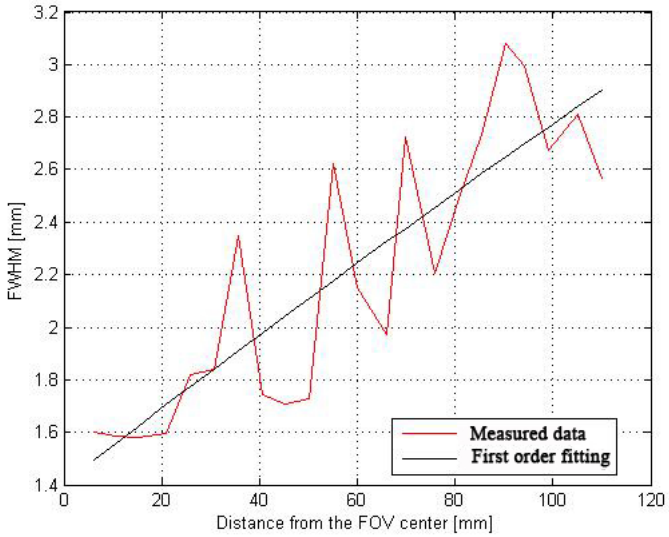

Figure 4.5: HRRT FWHM dependence on the distance from the center

In this graph, data appear to be very oscillating, while the first order fitting gives us the general trend of camera spatial resolution, along the radial direction. The oscillating nature of the real sampled data is due to the space discretization (voxel size is 1.22 mm in the three dimensions) and amplitude quantization performed by the volume elements; the problem consisted in those adjacent voxels, which shared the same detection peak and broadened inevitably the retrieved FWHM. This fact induced us to assess the resolution with a first order analysis, more prone in getting the order of magnitude of camera resolution than in retrieving very precised values.
If we observe carefully the graph in Figure 4.5, we can see that the FWHM of the different peaks is always comprised in the interval between 1.5 mm and 3.2 mm; the importance of these results is better perceived putting them in equation 1.1, along with the common patient's movements, discussed in chapter 1. Using the best achievable resolution, from experimental results, and a patient's translation of 1 mm, we can obtain that the FWHM of the output data would be equal to 1.8 mm, a resolution 20% lower than the one provided by the camera. The situation just computed consists in the best case, since, for the calculation, we considered the smallest movement possible among the ones introduced in chapter 1; in the worst case, it might happen that patient's movement is around 5 mm, and, then, with the previously considered scanner accuracy, the overall resolution would be of 5.22 mm: three

times and a half worse than the tomograph resolution.
Such a degradation of the resolution cannot be accepted, if we want to exploit the best performance offered by camera development, and this is the main reason why motion correction is one of the hottest topics in Nuclear Medicine.
Experiments to obtain camera PSFs are not useful only to underline the need for motion correction, but the information held by that tool can be employed during reconstruction, in order to enhance image quality. For instance, the knowledge of scanner PSFs has been included in the expectation step of the MLEM algorithm: indeed, the PSF of a certain region of the FOV can be interpreted as the distribution of probability that an event, from a precised image voxel, contributes to a precised sinogram bin. Another method to understand the reasons why PSF has a positive effect on data quality is by considering the role it plays in the reconstruction: it acts as a matched filter, enhancing the data, which match its detection model, and rejecting unwanted data, like, for example, high frequency noise[49].
Finally, the inclusion of PSF in the reconstruction has shown good results in reducing the impact of Partial Volume Effect (PVE) and Spillover (SO) on image quality: partial volume is an effect due to the discrete nature of voxels and caused by the fact that one voxel contains not only one tissue compartment, but many of them, and, then, the final voxel value will be the average of the activity from all the present compartments; spillover, on the other hand, is the result of the positron range and the angular uncertainty, treated in section 2.1, and its effect is to produce a blurring of the image, spreading the activity on a larger region than the real one. The influence of both these effects is heavier on the information at high frequencies than at low frequencies, and it can be shown that the use of PSF, in the reconstruction, helps in reducing their effects more efficiently[50].
Anyway, our test has not the needed precision to assess scanner performance in every point of the gantry and its aim was to provide the order of magnitude and trend of camera resolution, not to provide the tools to enhance data quality in the reconstruction. Obviously, the required precision in the assessment of scanner PSF is very high, if we want to use that information to compensate for the distortions introduced by the imaging system; therefore, the needed experiment must be able to determine the PSF for every point of the FOV and the protocol to be followed has to be universally accepted by the scientific community.

Chapter 5

Motion correction method

In the previous chapters, we have introduced all the knowledge and background required to understand how the HRRT works, so we have now all the basics required to treat and discuss the motion correction technique, implemented at the Neuroscience department of Karolinska Institutet, in Stockholm.
In the next paragraphs, great interest will be focused in giving the reasons why the adopted method was chosen and in pointing out advantages and disadvantages, deriving from its implementation; the required tools to perform the correction will be discussed and the algorithm itself will be analyzed step by step, in order to justify every single design choice.

5.1 MAF

Probably, one of the most important decision to take, in designing a motion correction system, is the one concerning the choice of the algorithm.; the right decision can make the system perform as desired, while the wrong decision can render all the subsequent efforts useless.
Studying all the situations that we had to face and the tools that we were equipped with, we understood from the beginning that the best approach was offered by the Multiple-Acquistion-Frames method; in this method, one frame from the PET study is chosen as reference, and all the other frames are corrected to realign them to the reference position. In such a way, we are able to re-estabilish the spatial correspondence between the other frames of a dynamic study and compartmental time activity analysis can be carried out, despite patient's movements.
From the previous description, it is straightforward to notice that MAF is an

algorithm capable to correct inter-frames movements and, ideally, to remove or at least reduce the effects of movement on reconstructed data, but this is not the only advantage: if properly designed, it can also correct for intra-frame movements. Indeed, if we are able to access the information before the reconstruction step, at the list-mode level, then we can associate the single events to the motion data available at the detection time and correct the LORs, according to the movement made by the patient at that instant. This option gives us the great possibility to perform an online correction, where the term online underlines the possibility to correct for motion, while the emission acquisition is still running.
At present we do not have the processing power needed for an online reconstruction. To be able to process the huge amount of data created in list mode, the only option is a GPU solution, preferably GPU CUDA. Contacts have been taken with research centers that will help with this, but at this time we do not have a working solution.
As already discussed in chapter 1, before of the emission scans, a transmission scan is acquired, set up to define the attenuation map; if, for simplicity sake, we choose, as reference frame, the first emission scan frame and assume that the patient does not move between the transmission scan and the first emission frame, then, that frame, and the position, will be the origin to where all the motion affected frames will be realigned.
The problem, however, consists in when the correction is performed: on one hand, if we can access the list-mode data, when the patient is moving, the movement will be recorded by the external tracker, introduced in the next paragraphs, and the correction will be performed on the single LOR; if, on the other hand, the correction can be done only on reconstructed volumes, there are other problems.
In the latter case, patient's movements will make the attenuation map data invalid: this means that every voxel in the emission data will be rototranslated compared to the corresponding voxels in the attenuation map, and when data will be reconstructed, the corrections will not be done in a proper way. Then, applying motion correction techniques, on already processed images, produces data, whose spatial orientation is corrected, but with numerical values that are wrong.
Unfortunately, this was our only option, and we decided, despite this, to proceed in implementing the motion correction method, because, once it will be possible to manipulate single events, the knowledge and experience developed from this offline correction, will be irredeemable.

5.2 Tools

We have already discussed the advantages deriving from the use of an external tracker, to perform the correction. Today, the market is full of motion tracking devices, and their performances are improving, thanks to the unstoppable development in the field of electronics; we decided to use the NDI Polaris Vicra, because of its availability at Karolinska Institutet, and because of the simple data interface.
The system consists of different pieces, but the two main parts are the tools and the position sensor.
The position tools are sensor monitored by Polaris Vicra, to obtain the spatial coordinates of the body that we want to follow; in our case, they consist of a rigid body incorporating at least 3 fixed markers, with no relative movements between them. These tools are called "passive tools". The position sensor is able to get the information of the position and orientation even from the so called "active tools", infrared light emitting diodes (IRLED), which emit electromagnetic waves to communicate their positions to the system.
The passive tools are composed of a rigid body, on which are mounted reflecting sphere markers: those spheres are arranged in an assymetrical geometry, with unique distances and angles between the markers, which allow the system to understand the orientation of the tool.
The position sensor is the main body where illuminators, detectors and the microprocessor are stored: as shown in Figure 5.1, the illuminator (an array of IRLEDs) and a matrix of Charge Coupled Devices (CCDs) are positioned at either side of the device. The illuminators flood the surrounding space through infrared waves, with wavelengths in the range between 800 and 1100 nm; then, the marker spheres, which are coated with a retro reflective material, reflect this electromagnetic radiation back to the sensor, where the arrays of CCDs detect the signal.

In this way, the system is able to retrieve the data needed, and outputs the information of the position, given in millimeters, and orientation, given in unit quaternions, which will be treated in the Appendix C; the position information is given as a distance from a fixed reference, which is, by default, the origin of the coordinate system built-in into the sensor, but we decided to set as reference point the origin of the coordinate system of the tool fixed to the camera (reference tool), in order to obtain the position of the patient's tool (the other tool) relative to a fixed point, and in order to be free to move the sensor, in different positions, with the purpose of achieving the best tracking conditions.

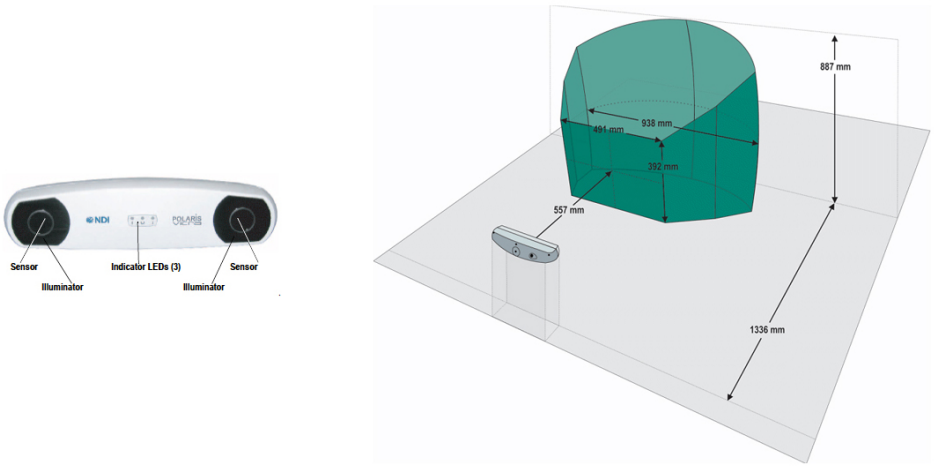

Figure 5.1: The Polaris Vicra position sensor (on the left) and the carachterized measurement volume (on the right)[51]

Both tools must be positioned within a certain specified measurement volume, which is a subset of the FOV of the position sensor, and it is where the performances of the tracker are well known; the precision of the position is 0.25 mm RMS, where this value is obtained through a least squares minimization of the difference between the detected position of all the markers and the position that the other three markers should occupy, according to the tool definition file (the "data sheet" of the tool), if the sensor detects only one of 4 markers in the tool[51].
The position and orientation data are updated by the system 20 times per second (20 Hz), which is a frequency much higher than the low frequencies involved in a patient's movement (usually much lower than 1 Hz), and the acquisitions took place in an environment lighted by fluorescent lamps, which allowed us to work with an information not corrupted by the interference that incandescent bulbs would have caused.
The scans were performed of a " Hoffman 3-D brain phantom", filled with ^{18}F. This phantom is accurately designed to simulate radioactivity distribution in a human brain for PET and brain SPECT studies[52]; the procedure followed was to attach the patient's tool onto the phantom, to collocate it within the FOV of the HRRT, as shown in Figure 5.2, and to acquire several scans with the phantom in different positions, recorded by Polaris Vicra: this way, we had all the information that we needed to perform our correction algorithm, which will be finally introduced, step by step, in the next paragraph.

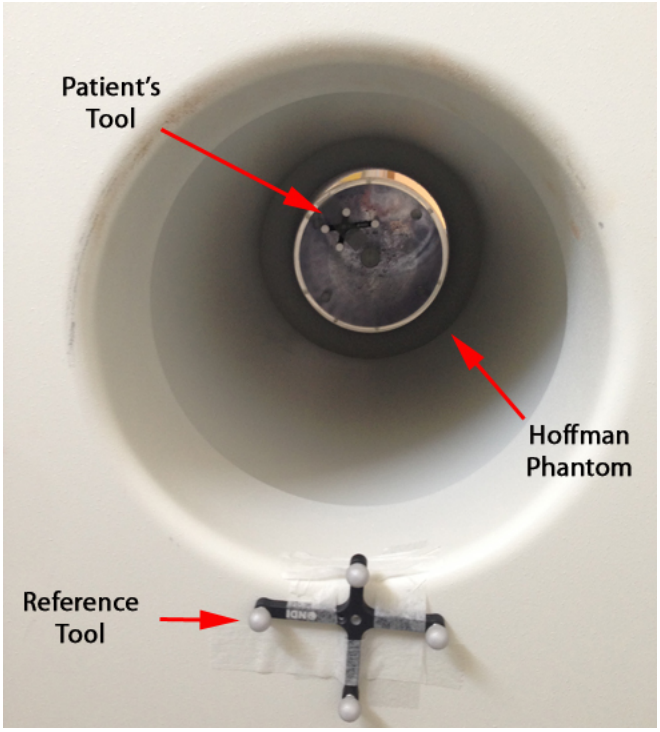

Figure 5.2: The phantom, the patient tool and the reference tool employed on the HRRT

5.3 The algorithm

The algorithm, implemented at the Neuroscience department of Karolinska Institutet, is composed of two fundamental steps: the motion correction, which uses Polaris Vicra output, and the coordinate migration, the transfer function between the coordinate systems, that will be discussed in the following section.
While the motion correction part is the real correction that we perform on the HRRT data, the coordinate migrations are the tools necessary to make this correction possible, using the information supplied by an external device. Starting with the exposition of the theory, which stands behind the correction, as explained in paragraph 5.1, the implemented algorithm takes one of the frames from the emission study as reference and corrects, according to

motion data provided by the Polaris Vicra, all the other frames to this reference frame. For practical reasons, assuming there are no movements during the transmission and first frame, this first frame is taken as reference frame. In theory any frame can be used, but that would force us to realign also the transmission data.
To correct for rotation and translation, following the MAF approach, we need to have the information on the origin position and on the orientation of patient's tool coordinate system, both at the reference frame and at the frame where the patient moved, data provided by the Polaris Vicra.
For each sampled measurement by the Polaris Vicra, three Cartesian coordinates for the origin position $(x_{origin}, y_{origin}, z_{origin})$ and four components for the unit quaternion (q_0, q_x, q_y, q_z), describing the orientation of patient's tool in the reference tool space, are recorded; with these data, we are able to reconstruct the position of all the four markers from patient's tool (whose relative position in the patient's tool space is given by $x_{def-file}, y_{def-file}, z_{def-file}$) in the reference tool space, according to:

$$\begin{pmatrix} x_f \\ y_f \\ z_f \end{pmatrix} = R \begin{pmatrix} x_{def-file} \\ y_{def-file} \\ z_{def-file} \end{pmatrix} + \begin{pmatrix} x_{origin} \\ y_{origin} \\ z_{origin} \end{pmatrix} \tag{5.1}$$

where:

$$R = \begin{pmatrix} q_0^2 + q_x^2 - q_y^2 - q_z^2 & 2 \cdot (q_x \cdot q_y - q_0 \cdot q_z) & 2 \cdot (q_x \cdot q_z + q_0 \cdot q_y) \\ 2 \cdot (q_0 \cdot q_z + q_x \cdot q_y) & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2 \cdot (q_y \cdot q_z - q_0 \cdot q_x) \\ 2 \cdot (q_x \cdot q_z - q_0 \cdot q_y) & 2 \cdot (q_y \cdot q_z + q_0 \cdot q_x) & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{pmatrix} \tag{5.2}$$

As shown in equation 5.1, the acquired data allow us to track the position of all the four markers from patient's tool, and more importantly, this concept can be extended not only to those markers, but also to all the points within the scanner FOV: this gives us the chance to track points from the patient's head and, using the tracking information, to correct their position, in order to correct for movements between the frames.
The formulae, necessary to perform motion correction, are derived from the following argument: since patient's tool is the same tool throughout the scans, its definition file too will remain the same and will be involved in determining the position in all the acquired frames.
The previous concept is really important because, although position will change among different frames, equation 5.1 is valid for all of them and the

factors defining the different positions in the scans are related to the variations in time of the rotation matrix (R in equation 5.1) and of patient's tool origin position (the vector with the subscript "origin" in equation 5.1), relative to reference tool coordinate system; then, the vector carachterized by the subscript "def-file" is the same one during the whole exam. Knowing that, we can obtain the coordinates present in the tool definition file, reversing the Polaris Vicra data of the moved frame, as shown in equation 5.3, and substitute them in the corresponding equation 5.1 for the reference frame, to arrive at expression 5.4.
In this way, we have just found the needed relation to rototranslate every voxel in the moved frame $(x_{mov-frame}, y_{mov-frame}, z_{mov-frame})$ in its correct position $(x_{ref-frame}, y_{ref-frame}, z_{ref-frame})$, according to the reference frame, and then we have just performed the required motion correction that we are working for.

$$\begin{pmatrix} x_{def-file} \\ y_{def-file} \\ z_{def-file} \end{pmatrix} = R_{mov}^{-1} \begin{pmatrix} x_{mov-frame} - x_{origin-mov} \\ y_{mov-frame} - y_{origin-mov} \\ z_{mov-frame} - z_{origin-mov} \end{pmatrix} \tag{5.3}$$

$$\begin{pmatrix} x_{ref-frame} \\ y_{ref-frame} \\ z_{ref-frame} \end{pmatrix} = R_{mov}^{-1} \cdot R_{ref} \begin{pmatrix} x_{mov-frame} - x_{origin-mov} \\ y_{mov-frame} - y_{origin-mov} \\ z_{mov-frame} - z_{origin-mov} \end{pmatrix} + \begin{pmatrix} x_{origin-ref} \\ y_{origin-ref} \\ z_{origin-ref} \end{pmatrix} \tag{5.4}$$

Once we have analytically retrieved the correction formula, we can now try to interpret the algorithm even from a conceptual point of view; previously, we have said that the orientation of the tool (and then of the patient) is given by the rotation matrix built up, using the quaternion components, and to better understand what will be explained next, we also need to know that the inverse of a rotation matrix expresses a rotation around the same axis as the one defined by the original matrix, but in the opposite direction. Those notions are fundamental in order to see how the term $R_{mov}^{-1} \cdot R_{ref}$ in equation 5.4 is the correction of the rotation, where matrix R_{mov}^{-1} nullifies the orientation of patient's tool at the moved frame, while matrix R_{ref} gives the tool the orientation of the patient at the reference frame: by multiplying the two matrices, we neutralize the effect of the wrong orientation at the moved frame, and we set the orientation of the volume to be the same orientation

held by the patient at the reference frame.
As we can see in equation 5.4, the rotation correction is performed, for every voxel "$x_{mov-frame}, y_{mov-frame}, z_{mov-frame}$", around the center of rotation "$x_{origin-mov}, y_{origin-mov}, z_{origin-mov}$", because that was the position of the origin of patient's tool coordinate system at the moved frame: this issue of accounting for the different centers of rotation at the two frames causes the addition too of the vector "$x_{origin-ref}, y_{origin-ref}, z_{origin-ref}$" in the motion correction expression and those two corrective terms compensate for patient's translations, along the three dimensions, occured between the two considered frames.

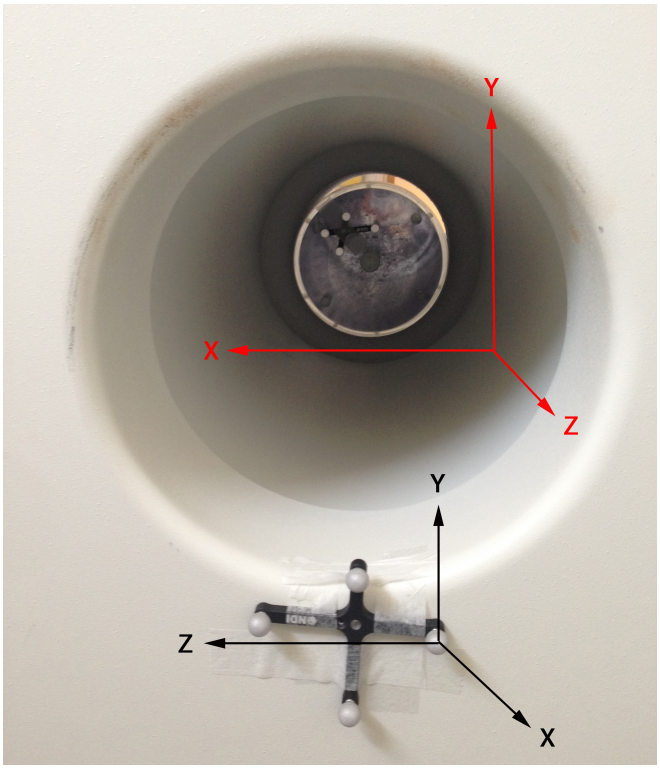

Figure 5.3: The HRRT coordinate system (in red) and the reference tool coordinate system (in black)

However, as briefly introduced at the beginning of this paragraph, to be able to perform the correction described so far, we need the two systems, Polaris Vicra and HRRT, to communicate with each other. When we obtain the list mode data from the HRRT camera, each frame is organized in a three dimensional matrix with the indeces of the matrix corresponding to

the three orthogonal axis of the HRRT; when we track the patient during the scans, the data are given according to the reference tool coordinate system, which is obviously different from the one of the scanner. In order to make the information obtained through Polaris Vicra, useful even in the HRRT coordinate space, it is necessary to compute a transformation matrix similar to that of the motion correction, but able to migrate the information of the orientation and position output by the tracker, into the camera coordinate space.
With this in place, we will be able to perform the transformation: so what we decided to do was to migrate each single voxel in the moved frame, into the Polaris Vicra coordinate system, to perform the motion correction, and, once the voxel would have been corrected, to migrate it back to the native coordinate system; following this procedure, we would have dodged manipulations of motion data and we would have used the output of Polaris Vicra in its native coordinate space, ensuring the highest level of liability possible for these data.
To make this transition between coordinate systems possible, we collected 37 transmission scans, to have the position of the four markers from patient's tool, both in the HRRT and in the Polaris Vicra spaces, in 37 different configurations; subsequently, we followed a least squares optimization[53] with the purpose of optimizing the coordinate migration needed to take the volume first to the Polaris Vicra space, and then to the native space. In order to avoid to deal with the problem of handling negative matrix indeces in Matlab, we migrated the three indeces of the single voxel into the Polaris Vicra space, performed the correction only on that voxel, and took the voxel back to the HRRT coordinate space, where it corresponded to a new motion corrected position in the final 3D matrix of the scan; we iterated this procedure for every voxel in the moved frame, and we finally obtained the new volume with the voxels from the original, rototranslated in the position that they would have occupied, if the patient had not moved.
A transformation able to make two systems exchangeable is possible only if both the coordinate spaces are either right-handed or left-handed, and since the HRRT had a left-handed coordinate system, while the Polaris Vicra had a right-handed one, we needed to reverse x-axis of the tracker coordinate system in order to make it compliant with the hypothesis for the transformation. Once we have treated the two main steps required in the algorithm, we are now able to summarize the procedure that we have to follow to perform the correction voxelwise, which consists in applying:

1. the coordinate migration to take the volume from the HRRT space to the Polaris Vicra space;
2. the motion correction, according to Polaris Vicra data,
3. the coordinate migration to take the volume back from the Polaris Vicra to HRRT space.

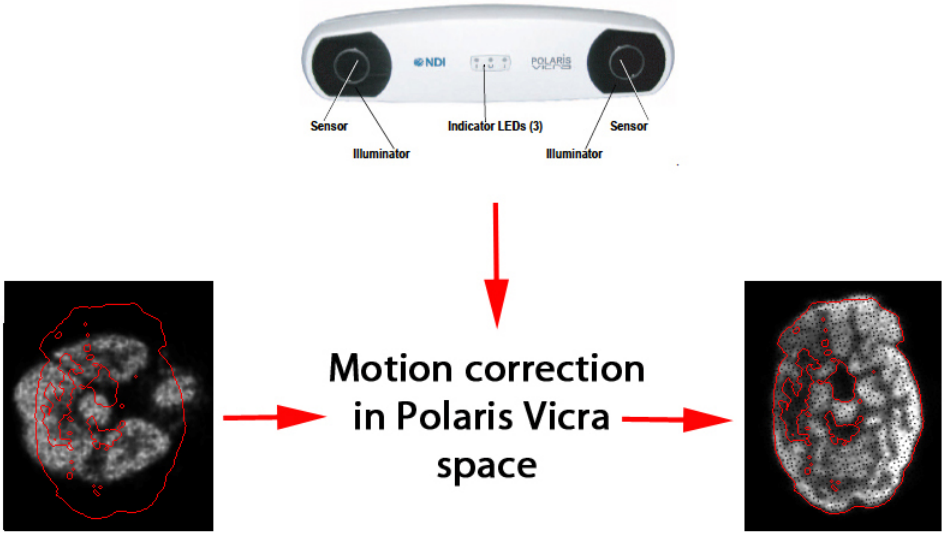

Figure 5.4: The three main steps of the motion correction method implemented

This procedure must be iterated for every voxel in the image.
Unfortunately, dealing with 3D matrices, we had to deal with a discrete volume and then with the existance of a rigid 3D grid: this issue creates a problem when corrected voxels are positioned halfway between grid slots in the final frame and a rounding function is needed to assign the value of the element to one of the shared voxels. The use of such a rounding function resulted in the appearence of artifacts in the corrected volume due to zero elements and also overlapped elements: When rounding the coordinate location of the voxels it occurs often that either more than one voxel are assigned to the same spot or that some voxels remain empty. With the aim of trying to alleviate the influence of this effect, some interpolation schemes may be used to assign a numerical value to every voxel, but we voluntarily decided not to use any of them in our method, in order to avoid manipulations on emission data.

All of the above explains the motion correction algorithm used during our tests and we are now able to show the experimental results that we achieved using our method on emission studies, performed at Karolinska Institutet.

Chapter 6

Experimental results

In this chapter, the results obtained performing our correction method are presented; first the general procedure followed to carry out the correction is introduced, then the numerical values are analyzed to better understand the outcome. In addition to this, the standard correction method currently used at Karolinska Institute is discussed and its performances are observed, using the same HRRT data as to test our algorithm.

6.1 The procedure

Each multi-frame emission study is preceded by a transmission scan of approximately 6 minutes to determine the attenuation map in the FOV; to ensure good statistics in our initial studies, the reference frame was acquired during approximately 10 minutes. Once the reference has been acquired, we move the phantom in order to rototranslate it randomly. As for the reference frame, every moved frame is acquired for approximately 10 minutes, and all the acquisition data are reconstructed after the last frame is finished.
After the reconstruction, information of the rotations (in radians) and translations (in millimeters) between the reference frame and the following frames are calculated using the "within-modality co-registration command" from Vinci's tools: Vinci is a freeware developed at The Max Planck Institute for Neurological Research in Cologne, Germany, for use with the analysis of medical imaging data from brain imaging modalities[54].
Once the rototranslation between the two frames have been calculated, the corrections can be performed following the steps described in paragraph 5.3. The correction is evaluated using the Vinci's co-registration tool and look-

ing at the differences left in terms of position and orientation between the two volumes; The same co-registration calculations are done with the volume corrected according to the current procedure at Karolinska Institutet, in order to be able to numerically compare the performance of the two motion correction algorithms.

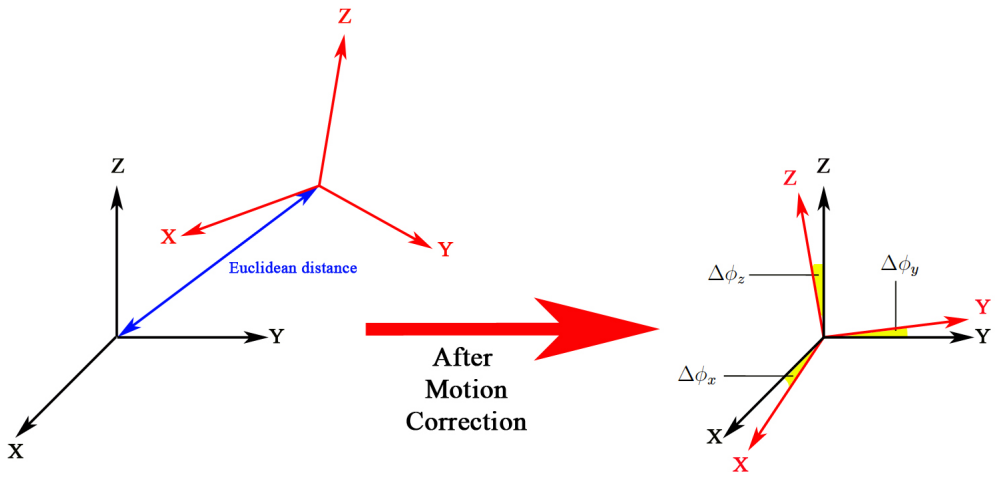

Figure 6.1: Graphical representation of the parameters used to evaluate the motion correction method. The black coordinate system refers to the reference frame, while the red one to the moved frame

6.2 Results

We tested our algorithm on three emission studies, performed on the HRRT PET camera; we performed those tests on March the 26th 2013, on April the 17th 2013 and on June the 12th 2013.
The first emission study was divided into 9 frames (1 reference frame + 8 moved frames) with an initial activity of 80 MBq; in the second study we had 8 frames (1 reference frame + 7 moved frames) with an initial activity of 30 MBq and the last study was divided into 14 frames (1 reference frame + 13 moved frames) with an initial activity of 50 MBq.
In the following tables, the Euclidean distance (in millimeters) and the difference (in degrees) in the orientation around the three HRRT Cartesian axis are reported for every frame in the study, prior and after the correction, in order to underline the positive effect that the correction had on the data.

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1	2 mm	0.1^o	-0.34^o	-10.1^o
2	4.28 mm	0.23^o	-0.33^o	-21.47^o
3	16.50 mm	-1.31^o	-1.27^o	-57.28^o
4	17.18 mm	0.41^o	-0.93^o	-0.75^o
5	1.41 mm	0.15^o	-1.33^o	0.73^o
6	4.36 mm	-0.07^o	-0.16^o	19.67^o

Table 6.1: Euclidean distances and differences on the orientations (angles) prior of the correction, for the test of 26/3/2013

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1 corrected	0.83 mm	-0.09^o	-0.09^o	0.04^o
2 corrected	0.93 mm	-0.05^o	0^o	0^o
3 corrected	5.30 mm	0.18^o	0.77^o	-0.11^o
4 corrected	1.22 mm	-0.10^o	-0.11^o	-0.06^o
5 corrected	1.19 mm	-0.11^o	-0.12^o	-0.02^o
6 corrected	0.95 mm	0.03^o	0.02^o	-0.01^o

Table 6.2: Euclidean distances and differences on the orientations (angles) after the correction, for the test of 26/3/2013

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1	1.82 mm	-0.02^o	0.01^o	-7.83^o
2	11.18 mm	0.88^o	1.50^o	-5.89^o
3	11.38 mm	0.63^o	0.57^o	-9.16^o
4	2.55 mm	0.37^o	1.47^o	-10.51^o
5	3.21 mm	0.26^o	1.56^o	2.07^o
6	17.70 mm	0.04^o	0.08^o	-2.93^o
7	8.11 mm	0^o	0.03^o	-8.37^o

Table 6.3: Euclidean distances and differences on the orientations (angles) prior of the correction, for the test of 17/4/2013

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1 corrected	0.89 mm	-0.08^o	-0.03^o	0.02^o
2 corrected	1.04 mm	-0.05^o	0.03^o	0.02^o
3 corrected	1.00 mm	-0.04^o	0.01^o	0.02^o
4 corrected	0.96 mm	-0.03^o	-0.01^o	0.01^o
5 corrected	0.97 mm	-0.10^o	0.85^o	0.04^o
6 corrected	1.07 mm	-0.03^o	0.04^o	0.02^o
7 corrected	0.61 mm	-0.67^o	-0.03^o	0.04^o

Table 6.4: Euclidean distances and differences on the orientations (angles) after the correction, for the test of 17/4/2013

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1	1.47 mm	-0.03^o	0.02^o	5.29^o
2	2.93 mm	0^o	0.08^o	9.73^o
3	1.87 mm	0.40^o	0.43^o	-6.45^o
4	6.82 mm	0.08^o	0.08^o	-5.16^o
5	0.69 mm	0.04^o	0.03^o	-0.58^o
6	5.60 mm	0.05^o	0.11^o	4.33^o
7	4.99 mm	0.12^o	-0.07^o	17.46^o
8	4.97 mm	0.07^o	0.03^o	17.21^o
9	4.83 mm	-0.01^o	0.11^o	9.11^o
10	5.13 mm	0.31^o	0.58^o	14.64^o
11	10.08 mm	0.06^o	0.87^o	4.03^o
12	4.79 mm	-0.02^o	0.11^o	-18.24^o
13	6.01 mm	0.69^o	2.04^o	-4.84^o

Table 6.5: Euclidean distances and differences on the orientations (angles) prior of the correction, for the test of 12/6/2013

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1 corrected	0.90 mm	0.01^o	0.06^o	-0.09^o
2 corrected	0.48 mm	-0.47^o	0.06^o	-0.37^o
3 corrected	0.86 mm	-0.12^o	0.09^o	0.97^o
4 corrected	1.98 mm	0.03^o	0.01^o	0.09^o
5 corrected	1.11 mm	-0.01^o	0.06^o	0.93^o
6 corrected	2.04 mm	0.05^o	0.09^o	-0.52^o
7 corrected	0.98 mm	0.95^o	0.01^o	-0.01^o
8 corrected	0.75 mm	0.11^o	-0^o	-0.94^o
9 corrected	1.36 mm	-0.02^o	0.11^o	-0.97^o
10 corrected	0.91 mm	0.58^o	0.05^o	-1.04^o
11 corrected	2.13 mm	0.03^o	0.08^o	-1.38^o
12 corrected	1.21 mm	-0.05^o	0.21^o	-0.37^o
13 corrected	2.63 mm	0^o	0.10^o	0.01^o

Table 6.6: Euclidean distances and differences on the orientations (angles) after the correction, for the test of 12/6/2013

As we can see from the previous tables, only in one case (frame 5 of 12/6/2013), the volume orientation and position have been worsened by our method, while in all the other cases the difference between the reference frame and the corrected frame has been considerably reduced. In particular, the initial purpose was to keep the difference in the position of the two volumes below the value of the voxel size (1.22 mm) and the difference in the orientation around the three dimensions below the value of 1^o.
The results show, in approximately the 77% of the 26 moved frames, that the distance after correction is below the threshold value, while the difference in the orientation is below the 1^o value in all the corrected datasets, except for frames 10 and 11 of the scan of 12/6/2013. In addition to this, it is interesting to point out how the corrected volumes, whose distance from the reference frame was above the voxel size, are frames with an initial uncorrected distance between 5 and 20 mm from the reference frame, a value of translation hardly experienced, as already said in chapter 1, in PET studies, but anyway useful to take into account for a complete presentation of the performance.
From the previous analysis we can also understand how the weak parts of our algorithm are the coordinate migrations. We use a least squares approach to compute the transformation matrix, and even though we collected 37 paired measurements of the 4 markers, an error will always exist on the data in the two coordinate systems, affecting the precision of our correction in a way

that must be considered, in the future, to improve the algorithm.
After having collected all these data, we can now analyze the performance of the already existing motion correction method used at the Karolinska Institutet, in order to be able, in the next chapter, to understand advantages and disadvantages arising from the choice of each one of the two correction strategies.

6.3 motion correction using cross-correlation function

The current method used for motion correction at Karolinska Institutet is straight forward rotation/translation using the parameters from a standard cross-correlation function as implemented in Matlab.
The cross-correlation function is used in a great variety of applications, ranging from velocimetry determination to pattern recognition and sonar, and it is an important tool to investigate the degree of similarity between two datasets[55]. In our case, we had to use the three dimensional version of the cross-correlation function (in the previous cited cases the cross-correlation function was a function either of the time variable or of the space coordinates of an image), since the datasets to compare were 3D matrices.
The normal procedure followed at Karolinska Institutet is to supply the two datasets that we want to realign to a modified version of the matlab script *realignEcat7_static.m*: this function takes the first input frame as the reference frame and the second input frame as the one, which needs to be realigned; then, it computes the cross-correlation function of the two volumes, working under the assumption that the two datasets are related by linear, rigid body displacement. A rigid body transformation is a geometrical transformation that preserves the initial distance among all the points and holds invariant the angles between every pair of lines and planes in the dataset. In brain imaging, rigid body transformation is the main method employed for the co-registration of volumes, since the shape of human brain changes very little with head motion; the estimation of the six parameters (three parameters for the translation and three parameters for the rotation) and their subsequent application to the dataset to move, therefore allows us to model different head positions[56].
According to the above assumption, the cross-correlation function will give the degree of conformity between the two volumes; the maximum of this function will occur at minimum difference between the two datasets. Once

the maximum has been found, the three parameters for the rotation and the three parameters for the translation are used in a rigid body transformation to output the corrected frame.
Vinci's co-registration tool was then used to obtain the distances and differences in the orientation still existing between the two volumes, after the correction; the results achieved are the following:

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1 corrected	0.30 mm	0^o	0.05^o	-0.31^o
2 corrected	0.36 mm	0^o	0.06^o	-0.31^o
3 corrected	14.36 mm	1.51^o	-1.92^o	-57.29^o
4 corrected	0.74 mm	0.06^o	-0.08^o	-0.08^o
5 corrected	0.76 mm	0^o	0^o	-0.08^o
6 corrected	0.75 mm	-0.02^o	0^o	-0.07^o

Table 6.7: Euclidean distances and differences on the orientations (angles) after the correction, for the test of 26/3/2013

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1 corrected	0.58 mm	0.01^o	-0.68^o	0^o
2 corrected	0.81 mm	0.02^o	0.02^o	-0.01^o
3 corrected	0.90 mm	0.01^o	0.03^o	-0.01^o
4 corrected	0.93 mm	0.02^o	0.03^o	-0.03^o
5 corrected	0.59 mm	0^o	-0.06^o	-0.60^o
6 corrected	0.98 mm	0.01^o	0.03^o	-0.02^o
7 corrected	0.75 mm	-0.07^o	-0.83^o	0.02^o

Table 6.8: Euclidean distances and differences on the orientations (angles) after the correction, for the test of 17/4/2013

Frame id	distance	$\Delta\phi_x$	$\Delta\phi_y$	$\Delta\phi_z$
1 corrected	0.45 mm	-0.38^o	0.04^o	0.58^o
2 corrected	0.54 mm	-0.14^o	0.03^o	0.57^o
3 corrected	0.29 mm	-0.50^o	0.27^o	-0.54^o
4 corrected	1.06 mm	-0.54^o	0.41^o	-0.50^o
5 corrected	0.52 mm	-0.50^o	0.20^o	0.45^o
6 corrected	0.64 mm	-0.08^o	0.09^o	0.56^o
7 corrected	0.65 mm	-0.54^o	0.25^o	0.51^o
8 corrected	0.97 mm	-0.56^o	0.16^o	1.39^o
9 corrected	1.43 mm	-0.98^o	-0.03^o	2.24^o
10 corrected	1.83 mm	-1.47^o	0^o	1.56^o
11 corrected	1.27 mm	-1.01^o	-0.07^o	1.40^o
12 corrected	0.84 mm	-0.09^o	0^o	-0.07^o
13 corrected	0.61 mm	-0.06^o	0.08^o	-0.57^o

Table 6.9: Euclidean distances and differences on the orientations (angles) after the correction, for the test of 12/6/2013

Using the method already implemented in the department we can see how, in approximately the 85% of the 26 moved frames, the translational correction of the volume to the reference frame is below the voxel size; With the rotational correction we can see how, in the 19% of the cases, the difference is above the threshold value of 1^o.
Finally, the case of frame 3, from the scan of the 26/3/2013, needs to be emphasized, because it is the case of a large rotation of near 60^o and, in this case, the cross-correlation method is not able to correct the dataset. In any practical setting, such a rotation of a patient's brain around the scanner axis cannot occur.
In this chapter, the results of two different motion correction algorithms have been presented, to familiarize with the orders of magnitude that we have dealt with; in the next chapter, a discussion of those data will be performed, focusing our attention mostly on advantages and disadvantages of using one method instead of the other, and on the practical needs that the clinical application of this method would require.

Chapter 7

Discussion

In this chapter, finally, we will compare the performance of the two methods, in an attempt to understand why we should choose the method using an external tracker, instead of the already implemented cross-correlation method. In particular, we will compare the performance of the two algorithms, frame by frame, trying to point out advantages and disadvantages of both the methods; in the second part of the chapter, we will propose an intermediate way between the two motion correction algorithms, drifting then toward the real reason why our method should be chosen in future applications.
We will conclude the chapter, considering what the efficiency limiting factors are and thinking about solutions that can overcome these limitations.

7.1 Performance comparisons

As we have seen in the previous chapter, the performance of the two correction methods are acceptable in both rotation and translation, since the corrected volumes are aligned at least in the 77% of the cases.
Although the results are acceptable, we want to determine which of the two methods performs better in rotation correction and in translation correction; from this analysis, we will be able to determine the best correction strategy for each case and the reason why that correction works better in that case.
Starting with the translation correction, we can see in Figure 7.1, 7.2 and 7.3 that, in the case of the motion correction using Polaris Vicra, the distance between the two volumes is always consistently reduced, while, with the cross-correlation method, we can see how for frame 3 from the scan of 26/3/2013, there is no a significant improvement, underlining how this method can be inefficient for large movements.

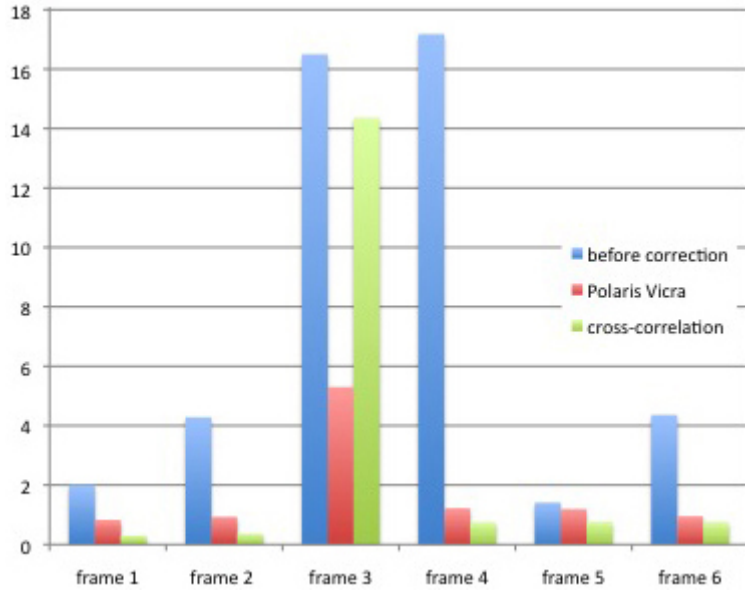

Figure 7.1: Comparison of the Euclidean distance between the moved and the reference frame, before correction and after correction with the two methods for the scan of 26/3/2013

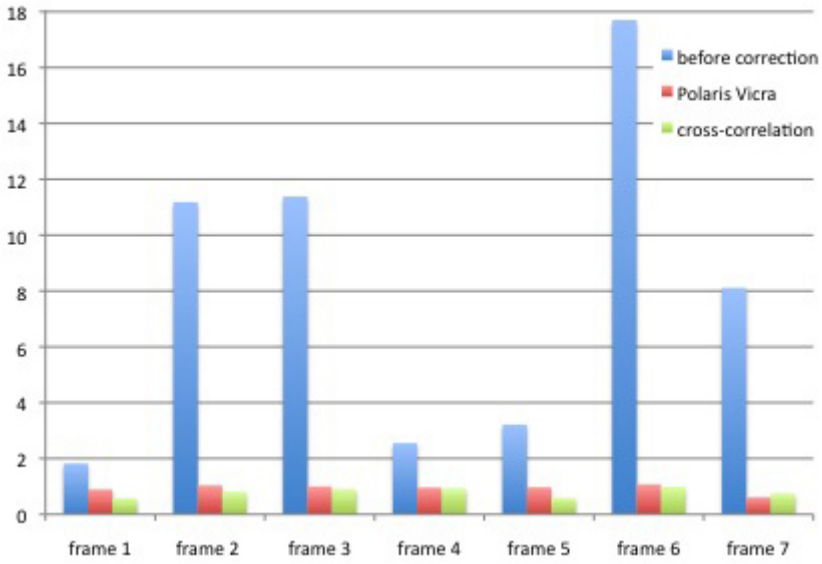

Figure 7.2: Comparison of the Euclidean distance between the moved and the reference frame, before correction and after correction with the two methods for the scan of 17/4/2013

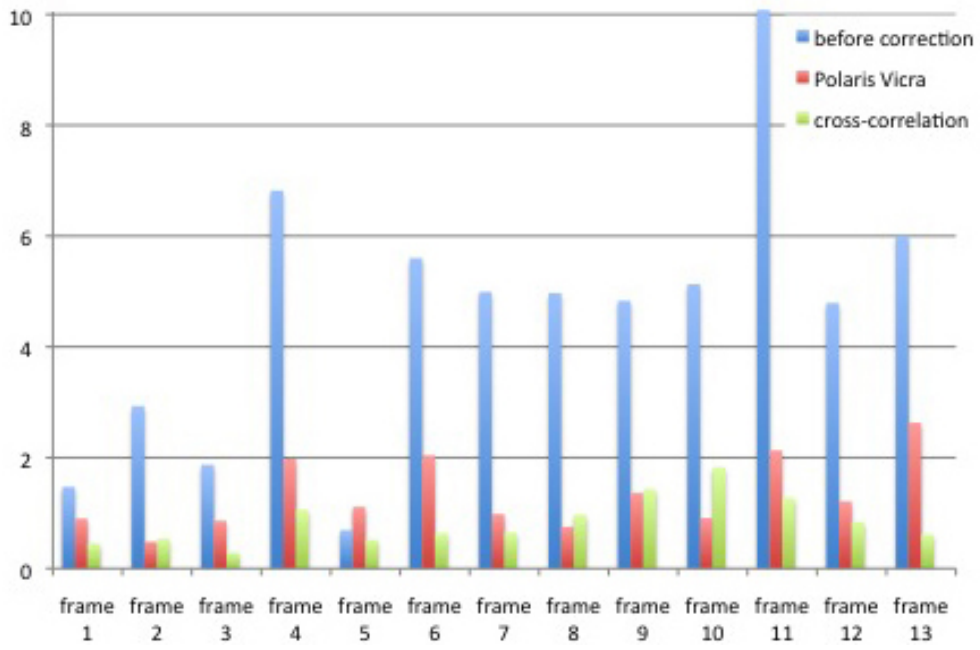

Figure 7.3: Comparison of the Euclidean distance between the moved and the reference frame, before correction and after correction with the two methods for the scan of 12/6/2013

However both strategies end with almost the same distances after correction, with performance slightly better for the algorithm already implemented and used in the department; this can be referred to the fact that the cross-correlation method does not need a coordinate migration to perform the correction, since the coordinate system where everything is carried out is the native one. Differently, in the method that we implemented, we first take the volume from the HRRT coordinate space to the Polaris Vicra coordinate space, where we motion correct the dataset, then we migrate it back to the native coordinate system, where the volume will be analyzed: those coordinate migrations are optimized but never perfect, so everytime we take the measure to another coordinate space, we add some error to it, that can be retrieved in the distance still existing after correction. In this way, we can understand why the percentage of corrected frames, whose distance from the reference one is below the 1.22 mm value, is higher using the cross-correlation method (85%) than using our method (77%). The situation with the orientation is the opposite; in Figure 7.4, 7.5 and 7.6, it is displayed the worst case among the three angles that are mirror for the different orientation between

the two volumes: it means that an angle is associated to each one of the three spatial dimensions, defining the degrees between the two datasets that must be applied to the moved frame in order to have it overlapping the reference frame, along that direction. The information that we wanted to show in the histograms is the angle with the biggest numerical value among the three provided in each comparison; in this way, we wanted to prove that, with our method, even along the dimension where the correction performances are the worst, the difference in the orientation is always below the 1^o value, except for frames 10 and 11 from the scan of 12/6/2013. Then our method ensures a very high precision in the correction of the rotation and shows much better results, in these terms, than the cross-correlation method, which, in approximately the 19% of the cases, left the biggest difference in the orientation above the threshold value.
Looking at the data that we obtained from our experiments, we can then briefly summarize that the cross-correlation method performs better in the correction for the translation, while the motion correction strategy, that we followed, achieves better results in the correction of the rotations.

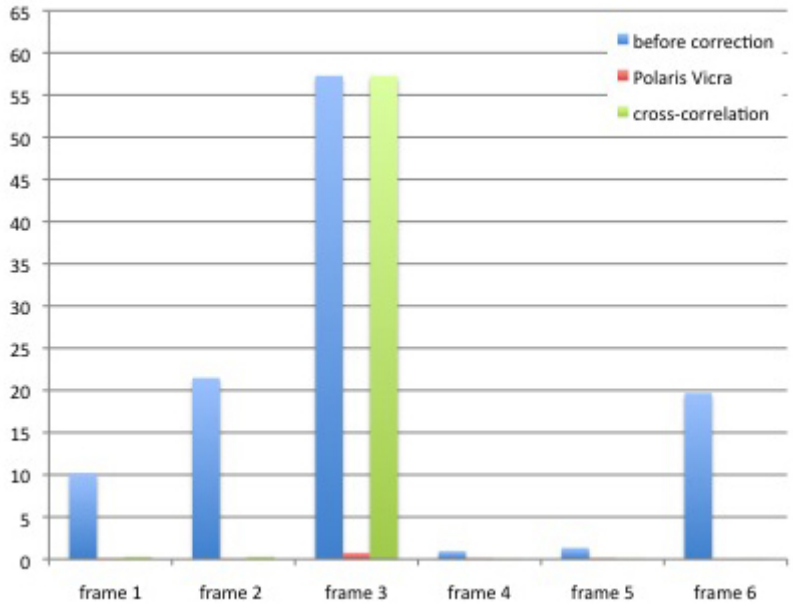

Figure 7.4: Comparison of the worst angle distance between the moved and the reference frame, before correction and after correction with the two methods for the scan of 26/3/2013

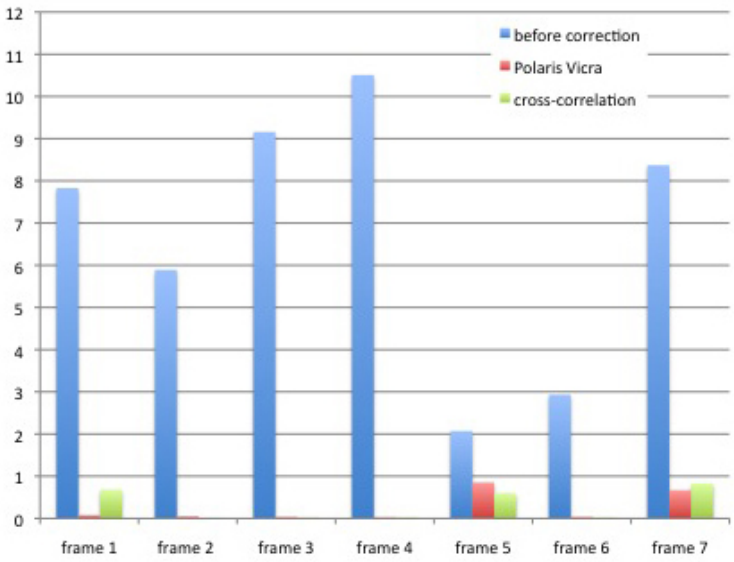

Figure 7.5: Comparison of the worst angle distance between the moved and the reference frame, before correction and after correction with the two methods for the scan of 17/4/2013

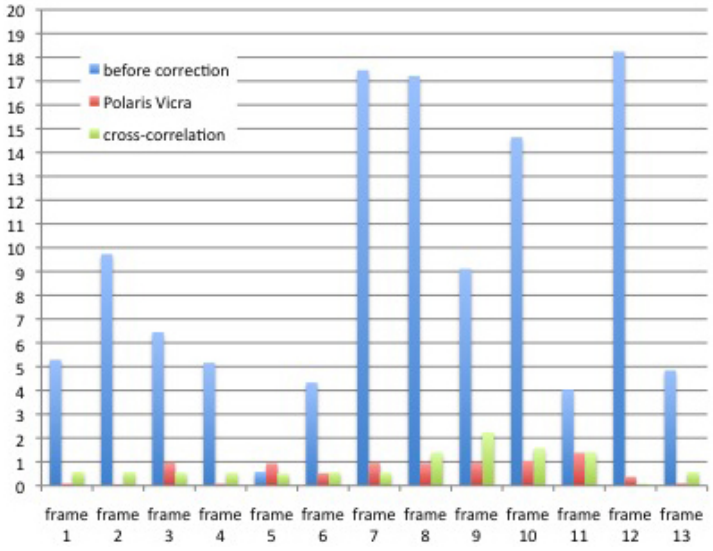

Figure 7.6: Comparison of the worst angle distance between the moved and the reference frame, before correction and after correction with the two methods for the scan of 12/6/2013

7.2 Alternative solutions

This quick analysis can give us the background to propose a new motion correction algorithm, which mixes both the studied methods: the idea could be to perform a first correction using the external tracker information and, then, to perform a second correction on the just corrected volume, using the cross-correlation method; in such a way, we would be able to exploit the great precision, provided by our method, in the correction of rotations, and the greater efficiency, offered by the algorithm used in the department, in reducing the differences in the position between the two volumes. Unfortunately, even though the motion correction strategy of using both the algorithms in succession on the moved frame would give us great results, it would not be able to correct the mismatches existing between the attenuation map and the emission data, and then the numerical values would still be not 100% loyal to reality.
The last concept discussed is a very central topic to understand why the motion correction developped with Polaris Vicra is much more powerful than the cross-correlation method; as we said in the previous chapters, we had not the chance to access the data before reconstruction and so, we had not the chance to motion correct the single LORs. Once we point out this issue, we can examine a little deeper the two algorithms to notice that, while the method using the cross-correlation function can work only on the already reconstructed volumes, the motion correction algorithms, that use external trackers, can work on list mode data. Providing a connection between the clock of the two systems (the Imaging tomograph and the external tracker) and being able to manipulate the single events, we can motion correct every annihilation event, according to the instant at which it occured and to the information on patient's movements output by the tracker; correcting the emission data before reconstruction, we avoid to create the positional mismatch between the attenuation map and the point of annihilation, and to compute wrong values for the activity within the FOV.
In addition, the possibility to correct the single events makes us also able to use all the collected frames, even those affected by intra-frame movements within the acquisition time. Indeed, provided of the motion information every 0.05 sec, we can rototranslate the LOR connected to the single annihilation and re-estabilish its right spatial correspondence with all the other events detected before of the intra-frame movement.
Keeping all this in mind, we can definitely say that the cross-correlation method showed its performance at its best, while, on the other hand, our al-

gorithm showed only the precision of the data that the Polaris Vicra outputs, but not the full potentials that this method can offer; the online correction is not only able to solve the problem between transmission data and emission data and the one linked to intra-frame movements, but it can also overcome the issue arising from the discrete nature of voxels, making us capable to deal with all the possible LORs, which cover the FOV in a more continuous way and do not require for an interpolation algorithm, as discussed in paragraph 5.3, to assign the numerical value to the voxels, after the correction.

7.3 The limiting factors

An analysis of all the data collected is necessary to understand which were the limiting factors in our method, and to try to suggest some corrections, with the purpose to enhance its performances.
During our study, we noticed two big problems, that need to be solved, if we want to apply our motion correction to the daily clinical routine.
As we have already seen in the previous paragraphs, one of the most precision limiting factors in our algorithm concerns the coordinate migrations to make the two systems communicate with each other: this needed step, repeated twice every time we want to motion correct a volume, is extremely bad for the precision of our correction, since the transformation to move freely between the two coordinate systems is not known a priori, but it is sought through a least squares optimization[53], which minimizes the squared sum of the errors; unfortunately, minimization is not synonym of cancellation and then every measure, that we used in the correction, was affected by a systematic error, due to these migrations. The only way to avoid such a big problem would be to have the two coordinate systems coincident, in order not to need any migration and to access the data directly as output by the tracker, on the camera; obviously this was impossible to us, since it would have meant to position the reference tool on the front-end of the scanner FOV and not on its back-end, losing the chance to follow the movements because of the new orientation of the reflective markers, which would have made the tracking, by the position sensor, impossible, regardless of position.
The second major limiting factor is something that must be taken into account for a future clinical application on living patients; during our study we fixed the rigid body tools with adhesive tape to the HRRT back-end (reference tool) and to the phantom back-end (patient's tool).
This worked fine for the purpose of our study, since we worked with an unani-

mated phantom, but it can hardly be proposed for human or animal patients. The issue concerning coordinate migrations can be overcome, considering to merge the tomograph and the tracker, in order to have the same coordinate system for both the applications and losing the need for transformations; referred to the problem of tool fixation, various solutions have been proposed in the last period, and one of the most interestings consists in tracking patient's movements, using a device, like Microsoft Kinect for Windows sensor, able to embed in the same body an infrared camera and a normal photo camera: the IR camera will give us the data of the depth, while the information along the two other dimensions will be retrieved, using complex algorithms able to obtain numerical values of the position from a picture, output 50 times per second (50 hertz)[57]; currently, markerless trackers are far away to be used, mainly because the quantitative information, obtained through the previously cited methods, suffers from low precision. Research in the field of markerless trackers is very active, and once their performance, in terms of precision, will be comparable to the devices using markers, they will provide great opportunities to improve the motion correction methods and to achieve better results.

.

Chapter 8

Conclusions

The need of a motion correction strategy is increasing as resolution and sensitivity of Medical Imaging tomographs improve: as resolution approaches and exceeds one millimeter, problems concerning patient's movements will increase and it is of outmost importance that a solution is found.
This study is focused on the HRRT PET camera, with a spatial resolution better than 2 millimeters in the center of the FOV and we have shown the importance of a motion correction strategy in order to avoid letting patient's movements be the limiting factor in scanner performance.
We performed the analysis motion correcting 26 moved frames, both with our method and with a method already used in the department, to demonstrate how the data retrieved with these 2 offline correction strategies show comparable acceptable results; the big advantage offered by our algorithm is in providing the possibility to perform online corrections, manipulating the single LORs, and in solving the mismatch between transmission and emission data, the problem of intra-frame movements and the issue concerning the discrete nature of voxels.
It is necessary to access and correct the emission data before reconstruction in order to fully utilize the advantages given by our algorithm and to be able to present images with a quality limited only by the scanner design.

Chapter 9

Further developments

As already discussed in chapter 7, one of the field in motion correction, where the reseach is more active, is the one concerning the external tracking using markerless devices, that do not require tools to be fixed on patient's head and the problems related to their fixation and precision.
In the last years, great efforts have been put in the attempt to design scanners able to merge together PET and MRI functionalities; indeed, new scanners are designed with APD detectors instead of PMTs, because of their performance unaffected by magnetic fields, that, on the other hand, seem to reduce the positron range and to further enhance the camera spatial resolution[58].
Finally, both in the field of scanner design and correction methods new solutions are continuously presented improving overall quality not least concerning spatial resolution, with performance approaching values unbelievable just a few years ago. This will allow for the introduction of new useful applications in the field of medical imaging of tomorrow.

Bibliography

[1] Justin Baba, William Wills, Abdel K Tahari, Dean F Wong, and Arman Rahmim. *Generalized Dynamic PET Inter-Frame and Intra-Frame Motion Correction-Phantom and Human Validation Studies.*

[2] Mirwais Wardak, Koon-Pong Wong, Weber Shao, Magnus Dahlbom, Vladimir Kepe, Nagichettiar Satyamurthy, Gary W Small, Jorge R Barrio, and Sung-Cheng Huang. *Movement correction method for human brain PET images: application to quantitative analysis of dynamic 18F-FDDNP scans*, volume 51. Soc Nuclear Med, 2010.

[3] Hans Herzog, Lutz Tellmann, Roger Fulton, Isabelle Stangier, Elena Rota Kops, Kay Bente, Christian Boy, Rene Hurlemann, and Uwe Pietrzyk. *Motion artifact reduction on parametric pet images of neuroreceptor binding*, volume 46. Soc Nuclear Med, 2005.

[4] Andrea Varrone, Nils Sjöholm, Lars Eriksson, Balazs Gulyás, Christer Halldin, and Lars Farde. *Advancement in PET quantification using 3D-OP-OSEM point spread function reconstruction with the HRRT*, volume 36. Springer, 2009.

[5] Michael V Green, Jurgen Seidel, Stacey D Stein, Thomas E Tedder, Kenneth M Kempner, Caroline Kertzman, and Tom A Zeffiro. *Head movement in normal subjects during simulated PET brain imaging with and without head restraint.*, volume 35. 1994.

[6] Waheeda Sureshbabu and Osama Mawlawi. *PET/CT imaging artifacts*, volume 33. Soc Nuclear Med, 2005.

[7] Arman Rahmim. *Advanced Motion Correction Methods in PET (Review Article)*, volume 13. 2005.

[8] JS Goddard, Justin S Baba, Seung Joon Lee, AG Weisenberger, A Stolin, J McKisson, and MF Smith. Real-time awake animal motion tracking system for spect imaging. In *Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE*, pages 4707–4710. IEEE, 2008.

[9] Justin S Baba, Christopher J Endres, Catherine A Foss, Sridhar Nimmagadda, Hyeyun Jung, James S Goddard, Seungjoon Lee, John McKisson, Mark F Smith, Alexander V Stolin, et al. Molecular imaging of conscious, unrestrained mice with awakespect. *Journal of Nuclear Medicine*, 54(6):969–976, 2013.

[10] Julian Maclaren, Brian SR Armstrong, Robert T Barrows, KA Danishad, Thomas Ernst, Colin L Foster, Kazim Gumus, Michael Herbst, Ilja Y Kadashevich, Todd P Kusik, et al. Measurement and correction of microscopic head motion during magnetic resonance imaging of the brain. *PloS one*, 7(11):e48088, 2012.

[11] Julian Maclaren, Michael Herbst, Oliver Speck, and Maxim Zaitsev. Prospective motion correction in brain imaging: a review. *Magnetic Resonance in Medicine*, 2012.

[12] Nicolas Costes, Alain Dagher, Kevin Larcher, Alan C Evans, D Louis Collins, and Anthonin Reilhac. Motion correction of multi-frame pet data in neuroreceptor mapping: simulation based validation. *Neuroimage*, 47(4):1496–1505, 2009.

[13] Michael K O'CONNOR, Kalpana M Kanal, Mark W Gebhard, and Philip J Rossman. *Comparison of four motion correction techniques in SPECT imaging of the heart: a cardiac phantom study*, volume 39. Society of Nuclear Medicine, 1998.

[14] Yani Picard and Christopher J Thompson. Motion correction of pet images using multiple acquisition frames. *Medical Imaging, IEEE Transactions on*, 16(2):137–144, 1997.

[15] Marcus Görge Ullisch, Jürgen Johann Scheins, Christoph Weirich, Elena Rota Kops, Abdullah Celik, Lutz Tellmann, Tony Stöcker, Hans Herzog, and Nadim Jon Shah. Mr-based pet motion correction procedure for simultaneous mr-pet neuroimaging of human brain. *PloS one*, 7(11):e48149, 2012.

[16] Peter M Bloomfield, Terry J Spinks, Johnny Reed, Leonard Schnorr, Anthony M Westrip, Lefteris Livieratos, Roger Fulton, and Terry Jones. The design and implementation of a motion correction scheme for neurological pet. *Physics in medicine and biology*, 48(8):959, 2003.

[17] WF Jones. Real-time event stream correction for patient motion in clinical 3-d pet. In *Nuclear Science Symposium Conference Record, 2001 IEEE*, volume 4, pages 2062–2064. IEEE, 2001.

[18] Gopal B Saha. *Basics of PET imaging: physics, chemistry, and regulations*. Springer, 2010.

[19] Miles N Wernick and John N Aarsvold. *Emission tomography: the fundamentals of PET and SPECT*. Academic Press, 2004.

[20] Craig S Levin and Edward J Hoffman. *Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution*, volume 44. IOP Publishing, 1999.

[21] Peter E Valk, Dale L Bailey, David W Townsend, and Michael N Maisey. *Positron emission tomography: basic science and clinical practice*. Springer London, 2003.

[22] S DeBenedetti, CE Cowan, WR Konneker, and H Primakoff. *On the angular distribution of two-photon annihilation radiation*, volume 77. APS, 1950.

[23] Steve Webb. *Webb's Physics of Medical Imaging*. Taylor & Francis, 2012.

[24] Fred A Mettler, Walter Huda, Terry T Yoshizumi, and Mahadevappa Mahesh. *Effective doses in radiology and diagnostic nuclear medicine: a catalog1*, volume 248. Radiological Society of North America, 2008.

[25] Glenn F Knoll. *Radiation detection and measurement*. Wiley, 2010.

[26] Wikipedia. *KleinNishina formula*. 2013.

[27] W Abdel-Rahman and EB Podgorsak. *Energy transfer and energy absorption in photon interactions with matter revisited: A step-by-step illustrated approach*, volume 79. Elsevier, 2010.

[28] Bryon Moyer. *An Overview of Medical Imaging*.

[29] MAXIM Integrated. *Positron emission tomography (PET) imaging.* www.maxim-ic.com, 2013.

[30] Michael Haselman, Robert Miyaoka, Thomas K Lewellen, and Scott Hauck. *Data acquisition for positron emission tomography*, volume Pub. No.: US 2011/0301918 A1. 2011.

[31] Carel WE Van Eijk et al. *Inorganic scintillators in medical imaging*, volume 47. IOP PUBLISHING LTD, 2002.

[32] Charles L Melcher. *Scintillation crystals for PET.*, volume 41. 2000.

[33] P. Dorenbos. *Scintillators for the detection X-rays, gamma rays, and thermal neutrons.* 2008.

[34] Expertsmind.com. *Scintillation counter.* 2013.

[35] Standford university. *What are scintillator materials?* 2012.

[36] University of Chicago. *Large-Area Picosecond Photo-Detectors project.* 2012.

[37] Thomas E. Nichols, Jinyi Qi, Evren Asma, and Richard M. Leahy. *Spatiotemporal reconstruction of list-mode PET data*, volume 21. IEEE, 2002.

[38] Jaafar Abdullah, Hearie Hassan, Mohamad Rabaie Shari, Salzali Mohd, Mahadi Mustapha, Airwan Affendi Mahmood, Shahrizan Jamaludin, Mohd Rosdi Ngah, and Noor Hisham Hamid. *GammaScorpion: mobile gamma-ray tomography system for early detection of basal stem rot in oil palm plantations*, volume 52. International Society for Optics and Photonics, 2013.

[39] Adam Alessio and Paul Kinahan. *PET image reconstruction*, volume 1. Philadelphia, PA: Mosby-Elsevier, 2006.

[40] Krzysztof Iniewski. *Medical imaging: principles, detectors, and electronics.* Wiley-Interscience, 2009.

[41] Juan E Ortuño, Pedro Guerra-Gutiérrez, José L Rubio, George Kontaxakis, and Andrés Santos. *3D-OSEM iterative image reconstruction for high-resolution PET using precalculated system matrix*, volume 569. Elsevier, 2006.

[42] Vesna Sossi, M-L Camborde, Stephan Blinder, A Rahmim, KJ-C Cheng, Ken R Buckley, Doris J Doudet, and Thomas J Ruth. *Dynamic imaging on the high resolution research tomograph (HRRT): non-human primate studies*, volume 4. 2005.

[43] Takehiko Fujiwara, Shoichi Watanuki, Seiichi Yamamoto, Masayasu Miyake, Shinya Seo, Masatoshi Itoh, Keizou Ishii, Hikonojyou Orihara, Hiroshi Fukuda, Tomohiko Satoh, et al. *Performance evaluation of a large axial field-of-view PET scanner: SET-2400W*, volume 11. Springer, 1997.

[44] A Braem, E Chesi, C Joram, F Garibaldi, C Joram, S Mathot, E Nappi, F Schoenahl, J Séguinot, P Weilhammer, et al. *Novel design of high-resolution parallax-free Compton enhanced PET scanner dedicated to brain research.* 2003.

[45] Hamatsu Photonics. *Photomultiplier tube R1450.* www.datasheecatalog.com, 1998.

[46] C Comtat, F Bataille, C Michel, JP Jones, M Sibomana, L Janeiro, and R Trebossen. *OSEM-3D reconstruction strategies for the ECAT HRRT*, volume 6. 2004.

[47] Floris HP van Velden, Reina W Kloet, Bart NM van Berckel, Adriaan A Lammertsma, and Ronald Boellaard. *Accuracy of 3-dimensional reconstruction algorithms for the high-resolution research tomograph*, volume 50. Soc Nuclear Med, 2009.

[48] Margaret E Daube-Witherspoon, Joel S Karp, Michael E Casey, Frank P DiFilippo, Horace Hines, Gerd Muehllehner, Vilim Simcic, Charles W Stearns, Lars-Eric Adam, Steve Kohlmyer, et al. *PET performance measurements using the NEMA NU 2-2001 standard*, volume 43. Soc Nuclear Med, 2002.

[49] Michael E Casey. *Point spread function reconstruction in PET.* 2007.

[50] Marine Soret, Stephen L Bacharach, and Irene Buvat. *Partial-volume effect in PET tumor imaging*, volume 48. Soc Nuclear Med, 2007.

[51] Northern Digital Inc. *Polaris Vicra User Guide.* http://www.ndigital.com/, 2005.

[52] EJ Hoffman, PD Cutler, WM Digby, and JC Mazziotta. *3-D phantom to simulate cerebral blood flow and metabolic images for PET*, volume 37. IEEE, 1990.

[53] Berthold KP Horn. *Closed-form solution of absolute orientation using unit quaternions*, volume 4. Optical Society of America, 1987.

[54] Stefan Vollmar, J Cizek, Michael Sué, Johannes Klein, AH Jacobs, and Karl Herholz. *VINCI-volume imaging in neurological research, co-registration and ROIs included*, volume 2004. Citeseer, 2003.

[55] Lisa Gottesfeld Brown. *A survey of image registration techniques*, volume 24. ACM, 1992.

[56] John Ashburner and Karl J Friston. *Rigid body registration*. 2007.

[57] M Alnowami, B Alnwaimi, F Tahavori, M Copland, and K Wells. *A quantitative assessment of using the Kinect for Xbox360 for respiratory surface motion tracking*. 2012.

[58] Bruce E Hammer, Nelson L Christensen, and Brian G Heil. *Use of a magnetic field to increase the spatial resolution of positron emission tomography*, volume 21. 1994.

[59] University of Washington. *Introduction to PET Physics*. 1999.

[60] Guido Valli and Giuseppe Coppini. *Bioimmagini*. Pàtron, 2005.

[61] MathWorks Inc. *Inverse Radon transform*. 2013.

[62] Klaus Mueller. *Introduction to Medical Imaging: Iterative Reconstruction with ML-EM*. Stony Brook University, 2012.

[63] C Michelet-Habchi, N Gordillo, S Bourret, Ph Barberet, C Jovet, Ph Moretto, and H Seznec. *Beyond filtered backprojection: A reconstruction software package for ion beam microtomography data*. Elsevier, 2012.

[64] Jack B Kuipers. *Quaternions and rotation sequences*. Princeton university press Princeton, 1999.

[65] Ken Shoemake. *Animating rotation with quaternion curves*, volume 19. ACM, 1985.

Appendix A

Filtered Backprojection

The need for tomographies arises from the impossibility of differentiating overlapping structures belonging to the same line of integration; the possibility of discerning all the different structures within the FOV has offered the chance to perform more accurate investigations and to overcome the limit set up by the projective nature of radiographies. Nowadays, FBP is the principal analytical method employed for the reconstruction and it is the result of the researches started by Radon in 1917, and continued by Housenfield in 1972, to reconstruct a 2 variables function, from its projections.
In 1970s, computational powers were not the ones that we have now and, then, volume reconstruction was achieved only through the two dimensional reconstruction of each transaxial plane, and not through a complete three dimensional reconstruction; in this appendix, we will treat the method used to reconstruct a two dimensional function, corresponding to a complete transaxial layer, starting from the measurements detected in PET.
In PET, when we collect the measurements belonging to a certain LOR, the value that we obtained is the result of the following equation:

$$c_n = \sum_{i,j} a_{ij} \cdot w_{ijn} \tag{A.1}$$

In equation A.1, c_n are the counts recorded along the nth-LOR, a_{ij} is the amount of radionuclide in ijth-element and w_{ijn} is the weighting factor of the ijth-element along the nth-LOR.
From all the LORs detected, we must be able to reconstruct a 2D-image whose values have a physical meaning: they give us the density of the radioligand in each spot of the 2D matrix (the image) and, therefore, the amount of activity in that region.

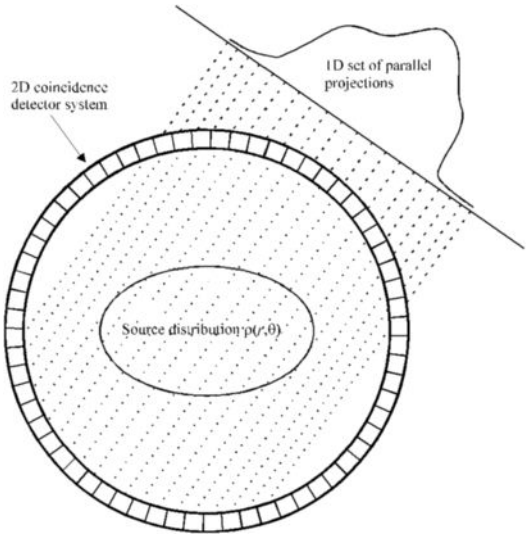

Figure A.1: The mechanism of projections in PET[59]

Once acquired, the projections are organized in sinograms and ready to be processed, as discussed in paragraph 2.7 and shown in Figure 2.10; however to explain the analytical method employed in the reconstruction, we refer to continuous variables and the projection that we detect can be expressed in the following way:

$$p_\theta(t) = \int_{LOR} f(x,y)dl \tag{A.2}$$

where

$$t = x \cdot \cos\theta + y \cdot \sin\theta \tag{A.3}$$

In equation A.2, $p_\theta(t)$ is the projection performed at an agle θ relative to the tomograph horizontal axis and corresponding to position t in the projection; this is equal to the integral along the line of integration of the image that we want to reconstruct ($f(x,y)$).
The projection just described is the Radon transform of the function $f(x,y)$, a transform able to correlate a function in the image plane to a function, periodic of 2π, in the $\theta - t$ plane; the knowledge of the Radon transform in the interval $[0,\pi]$ gives us also the knowledge of its behaviour in the interval $[-\pi,0]$ and this transform is provided of an important operator called "backprojection", which attributes to all the points belong to the line of integration the value of the projection. We will exploit its simmetry and the

backprojection tool to solve analytically the problem of the reconstruction, but before of doing so, we need to show what led to the current reconstruction method.
Indeed, it can be shown analytically that just backprojecting all the Radon transforms, we retrieve the original function convolved with another function that blurs the image, instead of obtaining the function that generated those projections:

$$\int_0^{\pi} p_\theta(x \cdot \cos\theta + y \cdot \sin\theta)d\theta = f(x,y) * \frac{1}{\sqrt{x^2+y^2}} \tag{A.4}$$

In equation A.4, the problematic of the backprojection method is evident: the image blurring cannot be accepted, since it lowers the image resolution in such a way that we cannot individuate anymore the small structures and see the evidences that can be fundamental for the diagnosis.
Therefore, it is really important to introduce a fundamental tool, like the Fourier transform, to investigate the frequency components present in the final image, in order to understand the best ways to avoid any distortion of the information.
We know that the Fourier transform enables us to study the harmonics present in the image and to see the effect that some operations has on those harmonics, in the frequency domain; each two dimensional image has a 2D-Fourier transform only if:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x,y)|^2 dxdy < \infty \tag{A.5}$$

If the previous relation is verified, the Fourier transform of the function $f(x,y)$ is:

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) \cdot e^{-j2\pi(ux+vy)} dxdy \tag{A.6}$$

while its Fourier antitrasform is the image itself, as shown in the following equation:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v) \cdot e^{j2\pi(ux+vy)} dudv \tag{A.7}$$

where u and v are the two spatial frequencies of the image. After having introduced the Fourier transform, we can notice how the only thing missing to reconstruct correctly the function $f(x,y)$ is the connection between the Radon transform and the Fourier transform of the image; the link is provided

by the central section theorem, which shows the existent connection between the data that we have measured (the Radon transform) and the tool, that allows us to retrieve the desired tomographic layer. It says that the Fourier transform of a projection obtained at an angle θ is nothing else than the line of the image Fourier transform, which forms an angle of θ with the u-axis.
Anyway, the fact that we have to deal with discrete data makes us unable to implement directly Fourier-based methods to reconstruct the layer, since they would not be able to handle spatial and angular sampling without interpolations. This is the reason why FBP does exploit the big help given by the central section theorem, but rearranging the order of the operations and without antitransforming, from the frequency to the spatial domain, the whole image.
In the FBP, the Fourier transform of each projection is first filtered in the frequency domain, subsequently antitransformed in the space domain and, then, backprojected, in order to reconstruct the 2D-function, combining together all the filtered backprojections. The mathematical steps behind this reconstruction method are the following:

$$\begin{aligned} f(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\cdot e^{j2\pi(ux+vy)}dudv \\ &= \int_{-\pi}^{\pi}\int_{-\infty}^{\infty} |\omega|\cdot F(\omega,\theta)\cdot e^{j2\pi\omega(x\cdot\cos\theta+y\cdot\sin\theta)}d\omega d\theta \\ &= \int_{-\pi}^{\pi}\int_{-\infty}^{\infty} |\omega|\cdot F(\omega,\theta)\cdot e^{j2\pi\omega t}d\omega d\theta \\ &= \int_{0}^{\pi} p_{\theta}^{filtered}(t)d\theta \\ &= \int_{0}^{\pi} h_{ramp}(t) * p_{\theta}(t)d\theta \end{aligned} \tag{A.8}$$

In equation A.8: in the first step, we have just written the definition of the Fourier antitrasform; in the second step, we have migrated from Cartesian spatial frequencies (u,v) to polar spatial frequencies (ω,θ) with all the changes in the differentials involved; in the third step, we have just applied the mathematical definition of variable "t"; in the fourth step, we have antitransformed the line of the Fourier transform, after that we applied the ramp filter $|\omega|$ to it and, finally, in the fifth step, we have shown that the antitransformed term is equal to the convolution of a ramp filter with the projection at that angle.

With the introduced ramp filtering, we can avoid the blurring effect that we have noticed just backprojecting the mesured data and obtain, ideally, the desired real image.
Unfortunately, practical problems like aliasing, due to the discrete number of detectors, and non-linearities, due to the different attenuation paths followed by the photons, lowers the quality of the reconstruction and, in addition to this, to reduce the variance of the omnipresent noise, we have to apply to the fourier transform of each projection a low-pass filter, which furtherly degrades the resolution of the final image.
Therefore to conclude this appendix, we can say that the FBP, on one hand, reconstructs the tomographic layer with great efficiencies, filtering the projection transform in the frequency domain and giving the chance to process separately all the different projections, but, on the other hand, does not take into account, in the reconstruction, the corrupting effects that occur in photon detection, causing, this way, the reconstructed data not to be 100% loyal to reality[60].

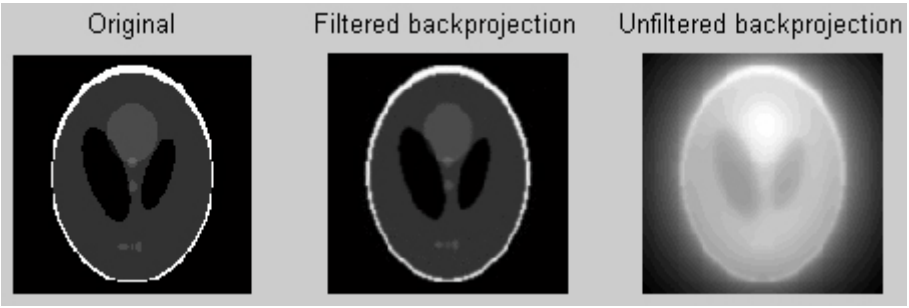

Figure A.2: Comparison between different methods for the reconstruction[61]

Appendix B

Iterative algorithms

Iterative reconstruction methods have been introduce in Medical Imaging, in the attempt to model and account for all the phenomena that take place throughout the photon detection.
As described in Appendix A, the most employed analytical reconstruction method, the FBP, does not take into consideration the statistical nature of the process and so, it retrieves the image not considering the noise and the corrupting effects that occur soon after positron annihilation.
Iterative algorithms, like MLEM and OSEM, reconstruct the data through iterations, set up to minimize the variance of the reconstruction error and those iterations lead to a real accurate estimation of the activity in the FOV[39], since statistical models, like the Poisson distribution, are employed to describe, with an acceptable approximation, the highly fluctuating nature of all the phenomena involved in PET.
Referring particularly to the MLEM algorithm, the estimation of the metabolic properties of the tissue (the model parameters) is achieved using the observed data, which are the counts recorded by all the LORs during the exam; those model parameters are initially assumed to follow a certain fixed behaviour, then, an iterative solver is employed to coverge to an optimal solution, which will be the final reconstructed distribution. In the case of the MLEM method, the expectation-maximization process is chosen because of its stability that, after a determined number of cycles, will take us to the maximum-likely solution for the distribution in the gantry.
The MLEM algorithm follows precised steps: as a first step, the initial setting of model parameters is decided, where usual adopted solutions consist in assuming uniform the activity across the whole FOV; in the second step, the expectation-maximization process is executed iteratively until, as a thrid

step, the iterations are stopped because the convergence has been reached, as displayed in the flowchart in figure B.1.

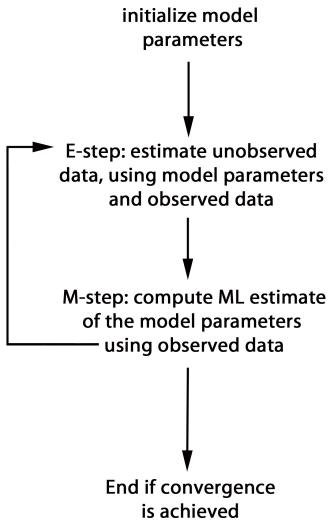

Figure B.1: The main steps of the MLEM algorithm

Focusing more our attention to the expectation-maximization process, in the expectation step, the unobserved data (the photon emission activities in the tissue, sources of detector readings) are evaluated considering the current estimation of the model parameters and the observed data, while, in the maximization step, the maximum likelihood estimate of the model is computed, using the retrieved unobserved data. Practically, the whole algorithm consists in assuming a certain default activity distribution in the gantry, for which we compute the corresponding projections, accounting for the various statistics involved, then we compare those computed projections with the real projections that we measured and we update the estimation of the activity distribution maximizing the log-likelihood function, which is the "index" that must be optimized when iterations end.
However, in the MLEM method, the inner assumption is that the mathematical model that relates the observed data, the model parameters and the unobserved data is the Poisson distribution; this means that, for each voxel "b" of the volume, there is a Poisson distributed variable which describes the emission activity fluctuations and which is carachterized by an expected value "$m(b)$". Therefore, the probability of having "n" emissions in that voxel is given by the following expression:

$$P(n(b)) = e^{-m(b)} \cdot \frac{(m(b))^n}{n!} \tag{B.1}$$

where the expected value is also the model parameter for the b-th voxel. When we want to reconstruct the activity in the FOV, in iterative methods, it is necessary to know the probability that a photon emitted in the b-th voxel can be detected by the d-th LOR; this very important information is present in the "system matrix" $P(d|b)$, a fundamental matrix, stored in scanner memory, that accounts for its geometry, for scattering and attenuating phenomena and for all the possible factors involved in photon detection. Such a matrix is one of the main reasons why iterative reconstructions can take into consideration non-idealities and, so, it is one of the tools that allow us to make a big step ahead in computing more loyal numerical values than with analytical methods.
If all the previous assumptions are verified, it can also be assumed that the number of detections in the d-th LOR (the observed datum) is a Poisson distributed process, since it is linear combination of Poisson distributed variables. This means that the probability of having "y" counts in the d-th LOR is given by the following formula:

$$P(y) = e^{-m(d)} \cdot \frac{(m(d))^y}{y!} \tag{B.2}$$

where "$m(d)$" is the expected value of counts in that LOR. The last argument allows us to finally find the fundamental connection between the observed data and the model parameters, since the expected value "$m(d)$" can be expressed in terms of the expected value of emissions in the b-th voxel and in terms of the probability that a photon, emitted in that voxel, is detected by d-th LOR, according to:

$$m(d) = \sum_{b=1}^{B} m(b) \cdot P(d|b) \tag{B.3}$$

To evidentiate the connection between the expected number of emissions occured in "b", counted by "d", and the expected number of emissions in "b", part of expression B.3 can also be rewritten as it follows:

$$m(d|b) = m(b) \cdot P(d|b) \tag{B.4}$$

Then, once we have proved the existing connection, it is important to introduce the definition of the log-likelihood function, which will be useful,

in the maximization step, to obtain the general equation of model parameter, at the k-th iteration.
The log-likelihood function is the logarithm of likelihood function and its expression is:

$$L_y(m) = \log l_y(m) = \sum_{b=1,..,B;d=1,..,D} -m(b)P(d|b) + x(d|b)\log m(b) + x(d|b)\log P(d|b) - \log P(d|b)! \tag{B.5}$$

Now, we have all the tools to describe mathematically the steps of the algorithm. In the expectation step, we obtain the number of emissions in the b-th voxel detected by the d-th LOR, at the $(k+1)$th iteration, computing its expected value, accounting for observed data and for model parameters at the previous iteration, as it follows:

$$x^{k+1}(d|b) = E[x(d|b)|y(d), m^k(b)] \tag{B.6}$$

After some mathematical manipulation of equation B.6, we arrive at:

$$x^{k+1}(d|b) = \frac{y(d) \cdot m^k(d|b)}{\sum_{b=1}^{B} m^k(d|b)} \tag{B.7}$$

Whenever we will have obtained the value computed in equation B.7, we will proceed with the maximization step, which will consist in maximizing the log-likelihood function and, then, in finding the model parameters which set to zero the derivative of the log-likelihood function. To be more precised, the mathematical steps involved are:

$$\frac{\partial L_y(m)}{\partial m(b)} = 0 \tag{B.8}$$

which leads to:

$$m^{k+1}(b) = m^k(b) \sum_{d=1}^{D} \frac{y(d) \cdot P(d|b)}{\sum_{b'=1}^{B} m^k(b') \cdot P(d|b')} \tag{B.9}$$

We have finally obtained the expression of the model parameter, for voxel "b", calculated at the $(k+1)$th iteration; we can see that in equation B.9, the value of the model parameter at the $(k+1)$th iteration is updated using

the value of this variable at the previous iteration, multiplied by a corrective term that employs the value of all the measured LORs, the information contained in the system matrix and the value of all the model parameters obtained at iteration k. The cycle of iterations will stop whenever the condition of minimum variance of the error in the reconstruction will be achieved, but the real risk that we run with iterative reconstruction methods is that the minimum reached is a local minimum, not a global minimum.
However, usually, the number of iterations required to reach the minimum is pretty high and, for this reason, new algorithms, like the OSEM, have been proposed to speed up the convergence.
To conclude the appendix, then, we can say that if, one hand, with iterative methods, we are able to consider not-ideal factors that can not be considered in the FBP, on the other hand, their computational demand is much higher than the one required by analytical algorithms and it can easily happen that the solution retrieved with MLEM (or OSEM) is not the global optimum, but just a local optimum; this is the main reason why both, iterative and analytical methods, are still employed and none of them has totally disappeared as modality for the reconstruction of PET studies[62].

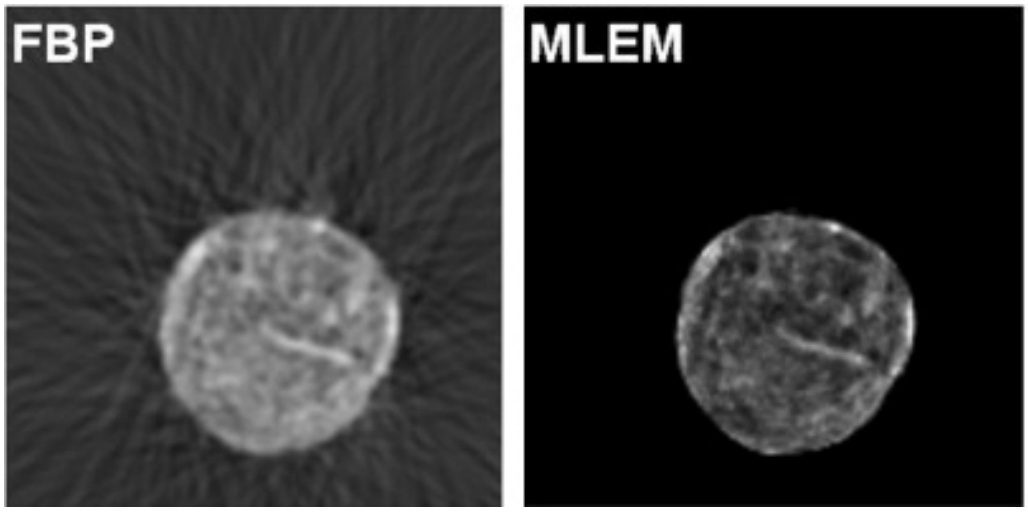

Figure B.2: Comparison between the FBP and the MLEM method, which shows how tomographic layer reconstructed with the iterative method are provided of greater accuracy and resolution[63]

Appendix C

Quaternions

In the following Appendix, we will discuss the role played by quaternions in describing rotations: indeed, in many of the applications (computer graphics and control of the position in the space) where rotations occur quite frequently, the representation of them is, most of the times, given by quaternions, for their simplicity and because they are really straightforward.
The description of the algebra of quaternions is reserved to specialized texts like [64]; in this section, we will focus our attention mainly on unit quaternions, the subset of quaternions which are employed in rotations.
However, a brief introduction of quaternions is needed.
The general expression of a quaternion is given by:

$$Q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = a + \mathbf{v} \tag{C.1}$$

In expression C.1, we can see that these numbers are provided of two parts: a scalar part (a) and a vector part ($b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$); the existence of these two parts allows us to say that quaternions are an extension of complex numbers, since we can obtain them setting $c = d = 0$ and, in addition to this, the possibility of obtaining spatial vectors, setting to zero the scalar part, will be something that we will exploit later on, in the appendix.
Before of moving toward unit quaternions, it is important, for the next demonstrations, to say that in the vector field of quaternions, the conjugate of a quaternion is the quaternion obtained reversing all the signs of the vector part of the initial quaternion, while the norm of a quaternion is given by:

$$N(Q) = \sqrt{a^2 + b^2 + c^2 + d^2} \tag{C.2}$$

It is then fundamental to underline that, with quaternions, the multiplication is not commutative (the order of unit quaternion multiplications does define the final rotation) and the general expression for multiplication between 2 quaternions, $Q_0 = a + \mathbf{v}$ and $Q_1 = a' + \mathbf{v'}$, is:

$$QQ' = (aa' - \mathbf{v} \cdot \mathbf{v'}) + (\mathbf{v} \times \mathbf{v'}' + a\mathbf{v'} + a'\mathbf{v}) \tag{C.3}$$

where, in expression C.3, the first term is the scalar part, while the second term is the vector part. Once we have clarified those points, we can focus more our attention on unit quaternions, which are quaternions having the norm equal to one, and this important property allows us to represent them as:

$$q = \cos\Omega + \mathbf{u}\sin\Omega \tag{C.4}$$

since:

$$N(q) = \sqrt{(\cos\Omega)^2 + (\sin\Omega)^2} = 1 \tag{C.5}$$

where $\mathbf{u}$ is a unit vector. Then, to show that quaternions are useful tools to describe rotations, we consider the two spatial unit vectors $\mathbf{v}_0$ and $\mathbf{v}_1$ (or equivalently quaternions with zero scalar part) and we compute the unit quaternion as the multiplication between $\mathbf{v}_1$ and the conjugate of $\mathbf{v}_0$, as the following equation shows:

$$q = \mathbf{v}_1\mathbf{v}_0^* = \mathbf{v}_0 \cdot \mathbf{v}_1 + \mathbf{v}_0 \times \mathbf{v}_1 = \cos\Omega + \frac{\mathbf{v}_0 \times \mathbf{v}_1}{N(\mathbf{v}_0 \times \mathbf{v}_1)}\sin\Omega \tag{C.6}$$

It is also needed to say that the conjugate of a unit quaternion is also the inverse of the unit quaternion, since:

$$q^{-1} = \frac{q^*}{N(q)} = q^* \tag{C.7}$$

Gathering all those properties together, we want to demonstrate that the expression $q\mathbf{v}_0q^{-1}$ rotates vector $\mathbf{v}_0$ of an angle equal to 2Ω[65]. Indeed, considering vector $\mathbf{v}_2 = q\mathbf{v}_0q^{-1}$, we just need to prove that $\mathbf{v}_2\mathbf{v}_1^* = \mathbf{v}_1\mathbf{v}_0^*$, because the equivalence of the scalar part will means that the angle between $\mathbf{v}_0$ and $\mathbf{v}_1$ is the same of the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$, while the equivalence between the vector parts will result in the three vectors belonging to the same plane. The demonstration is the following:

$$\begin{aligned} \mathbf{v}_2\mathbf{v}_1^* &= (q\mathbf{v}_0q^*)\mathbf{v}_1^* \\ &= (q\mathbf{v}_0(\mathbf{v}_1\mathbf{v}_0^*)^*)\mathbf{v}_1^* \\ &= q\mathbf{v}_0\mathbf{v}_0^*\mathbf{v}_1\mathbf{v}_1^* \\ &= q(-1)(-1) = \mathbf{v}_1\mathbf{v}_0^* \end{aligned} \tag{C.8}$$

The previous equivalence gives us the chance to summarize the situation in the following figure, where we can see that $q\mathbf{v}_0q^{-1}$ rotated vector $\mathbf{v}_0$ of an angle of 2Ω around the axis individuated by the direction of vector $\mathbf{v}_0 \times \mathbf{v}_1$.

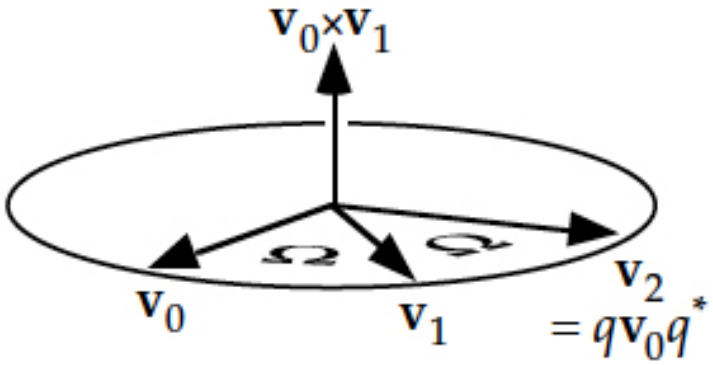

Figure C.1: Representation of the equivalence previously demonstrated[65]

The previous demonstration and the graphical representation of it make us understand why quaternions are so widely used to describe rotations: the possibility to extrapolate the angle and the axis of rotation, corrispondingly from the scalar and the vector part of the unit quaternion, are very useful carachteristics that make quaternions a very straightforward tool to implement rotations and to avoid complicate conventions to individuate the axis around which the rotation took place.