Feasibility study of the non-invasive estimation of the b+ arterial input function for human PET imaging

Xavier Hubert

École Centrale Paris
Laboratoire de Mathématiques Appliquées aux Systèmes

Thèse

présentée en vu d'obtenir le grade de Docteur de l'École Centrale Paris,
spécialité "mathématiques appliquées"

par

Xavier HUBERT

Etude de faisabilité de l'estimation non-invasive de la fonction d'entrée artérielle β^+ pour l'imagerie TEP chez l'homme

Thèse soutenue le 8 décembre 2009 devant le jury composé de :

Mme	Irène Buvat	IN2P3	Rapporteur
M.	Andrew Reader	McGill University	Rapporteur
M.	Nicholas Ayache	INRIA	Examinateur
Mme	Régine Trébossen	CEA-SHFJ	Examinateur
M.	Dominique Chambellan	CEA-DRT	Superviseur
M.	Nikos Paragios	ECP/INRIA	Directeur

2009ECAP0044

Un humanisme bien ordonné ne commence pas par soi-même,
mais place le monde avant la vie, la vie avant l'homme,
le respect des autres êtres avant l'amour-propre.
Claude Levi-Strauss

Rire souvent et beaucoup ;
Gagner le respect des gens intelligents et l'affection des enfants ;
Savoir qu'un être a respiré plus aisément parce que vous avez vécu.
C'est cela réussir sa vie.
Ralph Waldo Emerson

Remerciements

En premier lieu, je tiens à remercier chaleureusement les membres du jury qui ont accepté d'évaluer mon travail, tout particulièrement M^{mme} Irène Buvat et M^{r} Andrew Reader.

Ce travail n'aurait pas pu voir le jour sans les contributions petites et grandes d'un grand nombre de collègues. Éric Barat, Thierry Montagu et Thomas Dautremer, vous avez cette facilité à rentrer dans des solutions certes complexes mais toujours intéressantes et élégantes, c'est un plaisir de bénéficier de vos lumières. Merci aussi à Antoine Souloumiac pour les après-midis qu'il m'a consacré afin de me donner des pistes et des exemples que j'ai pu ensuite appliquer dans mes travaux. Merci aussi à Vincent et Jean-Pierre pour leurs trucs et astuces en tout genre.

Merci à Raphaël Boisgard, Bertrand Tavitian et toutes les personnes du SHFJ qui se sont intéressé à mes travaux et qui m'ont filé petit et grand coups de mains. En particulier, merci à Régine Trébossen, qui en plus d'avoir suivi mes travaux et fourni de nombreux conseils et avis éclairés, m'a aidé et soutenu dans les dernières étapes de la rédaction de ce manuscrit de thèse. Merci Régine.

Sans oublier François Tola, soi-disant jamais disponible mais au final toujours le premier à s'impliquer pour comprendre ce que j'ai pu observer au cours de mes travaux sans que je sache toujours l'analyser, mais aussi pour m'expliquer de nombreux points qui me paraissaient jusqu'alors pas bien clairs voire complètement obscurs.

Ces travaux n'auraient été que l'ombre de ce qu'ils sont sans les innombrables contributions de Juan. Sa disponibilité, sa patience et sa gentillesse ne rendent ses compétences que plus appréciables. Merci pour tout Juan.

Je ne serai jamais suffisamment reconnaissant à Fanny pour ses innombrables coups de pouce et la complicité qui s'est créée entre les voisins de bureau que nous étions.

Merci également à Chambellan pour ses idées lumineuses, il a été l'illustration vivante de l'expression "tout ce qui n'est pas interdit est permis". Je me souviendrai toute ma vie avec le sourire des quelques heures où nous nous sommes inventés 'métallo' avec Patrick Jousset dans l'atelier à fabriquer un alliage improbable. Sans parler des descentes au SHFJ d'Orsay avec nos boîtes à outil remplies de plomb, quelle rigolade ! Chambellan, merci pour ta joie de vivre et ton rire qui me mettait de bonne humeur même quand rien ne fonctionnait.

Samuel, tu as aiguillé les idées que j'ai pu développer aux cours de mes élucubrations et ça n'a pas du être facile tous les jours, je t'en remercie.

Je voudrais tout particulièrement remercier Delphine qui s'est plongé dans les brouillons de mon manuscrit et qui m'a permis par ses remarques de faire profondément évoluer le contenu de ce document, aussi bien sur le fond que sur la forme. Si ce document est aujourd'hui lisible, je te le dois pour beaucoup.

Mathieu et Eric, le séjour hawaiien restera un grand moment dans ma vie de globe-

trotteur, Eric, je garde au chaud ta photo de plongeur de la Hanauma Bay Nature Preserve Park, en espérant que tu en feras de même de la mienne.

Merci aux Fred pour les quelques footings que nous avons courus ensemble autour du centre, votre compagnie me faisait oublier que ça fait mal aux jambes de courir.

Merci à Ratto, Titi et Romu pour les franches rigolades que l'on a partagées, aussi bien dans les couloirs et les bureaux qu'à la cantine.

Chonchon, continue de faire peur aux stagiaires, je n'oublierai jamais les joutes verbales que l'on a pu se livrer, tu m'as fait énormément rire toutes ces années.

Merci aux voileux du CEA, tout particulièrement la fine équipe de la Neptunium 2007. Vous m'avez donné le goût du grand large et j'espère de tout coeur vous croiser sur les pontons dans les années à venir. Guitou, ta compagnie sur l'eau aussi bien que dans les bureaux est un plaisir, continue de distiller ta bonne humeur.

En un mot comme en mille, merci les collègues pour ces 3 années passées en votre compagnie.

Ce manuscrit n'aurait certainement pas vu le jour sans le soutien et les encouragements d'Évelyne, Romain et Marie, merci à vous les amis-chercheurs.

Enfin merci aux amis sans qui ces 3 années n'auraient été que l'ombre de ce qu'elles ont été, votre amitié m'est précieuse. En particulier, merci à Raphaël, Benoit et Clément, qui m'ont tellement entendu tenter d'expliquer mon sujet de recherche qu'ils étaient au final les meilleurs pour expliquer au grand public de quoi il en retournait. Tous les moments que l'on a pu passer ensemble m'ont changé les idées plus souvent que vous ne le soupçonnez.

Merci aux coloc' des premières années, Héloïse et Aurélie, notamment pour les petits muffins le soir en rentrant du boulot, mais aussi et surtout à Titom pour avoir été mon coloc' pendant trois ans, à la fois copain et complice, toujours curieux et parfois même force de proposition sur ma thèse, dans tous les cas toujours présent pour rigoler, sortir et passer des moments aujourd'hui inoubliables.

Merci aussi à Pierrot pour son inénarrable façon de percevoir ce que sont des années de thèse, et de ses improbables pretextes pour organiser de somptueuses fêtes. J'en profite pour saluer tout le 115 et les belles soirées qui s'y sont déroulées.

Merci enfin à mon frère et à mes parents, qui m'ont toujours soutenu et fait confiance dans mes choix, et ce dans cadre bien plus large que cette thèse. Etre soutenu permet d'avancer plus sereinement, et pour cela, je vous en serai éternellement reconnaissant.

Contents

List of Figures

Introduction

Positron emission tomography (PET) has become essential during the last decades to investigate human body mechanisms, to diagnose diseases and to develop new drugs which until then could not be assessed through imaging techniques. In this context, the quantification of medical images is fundamental to investigate *in-vivo* physiological mechanisms. Furthermore, the demand increases for providing dynamic information from dynamic sequences of nuclear imaging.

In oncology and neurology, it is primordial to assess the performance of drugs on patients and to monitor quantitatively the response of diseases to medicinal and non-medicinal therapies. The more accurate the quantification of the tissues response is, the earlier the treatments can be adapted to better cure either cancers or neurological diseases. Several actions have been initiated in France to address such an issue, through "Plan Cancer" in 2002 which should bring the number of PET-scans in France to 80. "Plan Alzheimer", initiated in 2008 in France, also aims at accelerating research programs on new radiotracers and the quantification of their impact on tissues affected by diseases. For instance, fluoromisonidazol (F-MISO) and fluorothymidine (FLT) are promising radiotracers under development. F-MISO allows measuring hypoxia level in tissues, which is related to angiogenesis in tumours. Angiogenesis is a much better indicator of tumour response to therapies than the measurement of the volume of the tumour with conventional tracers in PET. In this way, therapies are adapted earlier to better cure the tumour. This also applies to FLT with is a marker of cellular proliferation, which is particularly relevant in oncology.

Such new radiotracers will bring better monitoring of patients' diseases only if nuclear imaging fully takes advantage of quantification.

1 The role of radiotracer arterial input function

Improvements on scanner technologies allow the dynamic acquisition of temporal frames which gives access to dynamic information of physiological mechanisms. In order to estimate physiological parameters, concentrations of tracers in tissues and blood must be quantified and compared (Patlak 83; Muzic 01). The concentration of the injected molecules labelled with positron emitters in arteries is called "arterial input function" or "input function" and is currently obtained through arterial blood sampling (Phelps 79; Gallezot 05).

This is an invasive procedure which is medically not trivial, requires qualified medical staff and is painful for patients. These constraints make it inappropriate for clinical studies and consequently, arterial blood sampling is mostly applied in practice to pre-clinical investigations. However, the input function could be estimated using non-invasive methods. Some general features of input functions, obtained from blood sampling, were extracted from population-based studies (Wakita 00) as well as through analytically modelled input functions with a few parameters (Phillips 95), but the variability from patients to patients, particularly for investigating pathologies, makes it not suitable. Furthermore, investigating new molecules implies that no standardised or parametric arterial input function is available.

Research is led towards non-invasive estimation of the arterial input function from reconstructed images of spatial distribution of radiotracers in tissues (van der Weerdt 01;

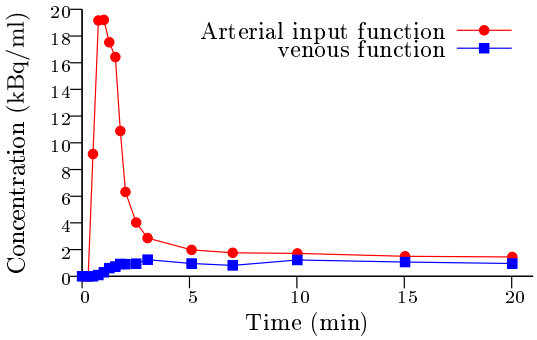


Figure 1 – Input function obtained from arterial blood sampling, first minutes. Radiotracers were administrated through bolus injection. The radiolabelled molecule is [18F]F-A85380 (Gallezot 05)

Naganawa 05). It is important to notice that nuclear imaging allows the estimation of the concentration of radiolabelled molecules in arteries, not specifically the concentration of the injected tracer. This means that for radiotracers which have metabolites in blood, the difference between the input function and the β^+-concentration, that will be called β^+ input function, can be significant. Consequently, the estimation of the input function with the β^+ input function is only relevant for tracers with a little amount of metabolites. This thesis addresses the estimation of the input function from the non-invasive estimation of the β^+ input function and it is assumed that they are similar in the rest of this thesis, which is realistic for tracers with no or a few metabolites in blood. For cardiac investigations, it is possible to extract the arterial input function from blood pool in vascular structures which stand in images. However for investigations on other organs, it is usually difficult to estimate the arterial input function from images: either the spatial resolution is poor, preventing the vascular structures from being identified, or the sensitivity is poor and the images are degraded, which leads to estimations of arterial input functions which are dominated by noise. It was also proposed to develop external detectors which are dedicated to the estimation of the input function from arteries whose surrounding tissues take up less radiotracers than the heart or the brain (Kriplani 06).

2 The problem

The injected dose does usually not exceed 370 MBq (10 mCi). This results in a very low radiotracer concentration in arterial blood as illustrated in Fig. 1. This plot is obtained from successive arterial blood samplings from the radial artery of a patient after the bolus injection of a ^{18}F-labelled molecule. This data was obtained at Frederic Joliot Hospital Service (SHFJ), Orsay, France. The amplitude of the peak depends on the radiolabelled molecules, but it is common that its average value after first minutes is below 4 kBq/ml. As the volume of arteries is small (vessel diameters are about a few millimetres), this represents a small amount of photons emitted by arteries during clinical acquisitions. When radiotracers are injected more slowly than as a bolus, the width of the peak increases and the amplitude decreases.

It has to be noticed that the concentrations in arteries and veins may only differ within the first minutes as shown in Fig. 1, but this depends strongly on the tracer which is injected. The time it takes before concentrations in arteries and veins are equal depends on the radiolabelled molecules. Only the concentration in arteries is of interest because this is the arterial blood which provides organs with the radiolabelled molecules.

Methods proposed until now in literature suffered from the fact that surrounding tissues contain a lot of activity (Chen 07), because their volume is large compared to arteries and they usually take up the radiolabelled molecules whereas arteries are just a transport means. As a result, we decided to investigate the possibility to perform a parallel acquisition: while PET acquisitions are achieved on the brain of patients, we would like to acquire another set of data at another location where surrounding tissues are far less radioactive. This location should contain large arteries in order to maximise the number of photons which are emitted from arteries. Arteries should also stand close to the surface of the body so that the geometric efficiency is good and the attenuation of 511-keV photons in the body is negligible. As much as possible these arteries should stand far from the most active organs, for instance the brain, the heart and the bladder. This led us to choose knees as the acquisition location, justifications being given in chapter 2. Indeed, the popliteal artery which stands in the knees are large (about 5 mm) as they are just in the continuity of the femoral artery. Furthermore they stand in the middle of the knee, which is close to the surface since the corresponding depth of the popliteal artery is about 50 mm from the skin. It was also considered to estimate the arterial input function from the brachial artery, the carotid or the aorta, but in view of the considerations which are exposed before, these solutions were rejected.

3 Outline

In order to increase the amount of collected photons, we took advantage of coded-aperture imaging which offer high-sensitivity and high-resolution (Accorsi 01). The methods developed during this thesis thus belongs to Single Photon Emission Computed Tomography (SPECT). Classical reconstruction methods which are used with coded-apertures (correlation-based algorithms) are not adapted to medical imaging because some hypotheses of the coded-aperture theory are not respected in medical imaging. We suggested using coded-aperture collimators combined with statistical reconstruction algorithms, instead of decorrelation techniques which are usually used in conjunction with coded apertures. Indeed we think that statistical reconstruction algorithms are complementary with coded-apertures for addressing low-activity issues in emission tomography. On one hand, coded apertures have a transparency reaching about 50%, which deeply increases the sensitivity of SPECT-systems with classical pinhole, parallel or cone-beam collimators. On the other hand, statistical reconstruction algorithms take into account the statistical modelling of Poisson variables, which is very appropriate for low-signal imaging.

The use of statistical algorithms with coded apertures at 511-keV raises, however, new problems. In particular, statistical algorithms require the computation of a system matrix which models forward projections. Such a matrix allows the computation of the expected projections of the spatial distribution of any objects in the field of view. The computation of system matrices is based on ray-tracing techniques, but it must be made very efficiently because the coded-aperture collimators have complex geometries. Furthermore, the more accurate the system matrix, the better the accuracy of reconstructions. However, improving the accuracy of system matrices usually involves more computer memory resources to compute them. As the system matrix is used for every update in statistical reconstructions, such a large matrix slows down the reconstruction time. Furthermore, as projections are intermixed because of the coded aperture, the convergence is slower than with classical collimators. Using coded aperture collimators with statistical reconstruction algorithms thus requires the development of new methods which efficiently deal with such system matrices and accelerate the convergence of reconstructions. The benefits of such imaging systems and new algorithms are demonstrated on analytical data, then on Monte-Carlo generated data and finally on experimental data. Eventually β^+-input function is estimated: a method is

introduced which aims at estimating kinetics in the two main vessels of the knee (popliteal artery and popliteal vein) from the whole data acquisitions, from blind source separation techniques. This takes advantage of the decomposition of the spatiotemporal distribution of the object on spatial distributions and temporal distributions.

Chapter 1 introduces concepts for understanding the choices and methods encountered in the other chapters of this thesis. Chapter 2 motivates the choice of the popliteal artery for the non-invasive measurement of the β^+ input function. It also supports the choice of coded-aperture imaging with a γ-imager, and provides new algorithms for efficiently computing system matrices required by statistical reconstruction algorithms. Chapter 3 describes new innovative methods to optimise computations of statistical reconstruction algorithms and to accelerate their convergence to solutions. Performances are illustrated by reconstructions of analytical datasets, then Monte-Carlo simulations and finally actual projections. Eventually chapter 4 addresses the reconstruction of several frames by blind source separation with positivity constraints, with the inclusion of the reconstruction step during the separation, which is an original method for dealing with source separation.

Bibliographical studies 1

This chapter presents concepts and methods useful to understand techniques that were developed in this PhD thesis as well as the choices about imaging devices and algorithms that were made. It also provides state-of-the-art reviews in fields that are related to the present work. First, a brief description of pharmacokinetics and compartmental models is done to illustrate the important role of arterial input function for investigations on new drugs. Then a review of the methods that were investigated to measure input functions is presented. It is then deduced that the non-invasive estimation of the β^+-input function requires that it is determined through nuclear imaging techniques, from reconstructed images. Consequently, some basics on nuclear physics and nuclear imaging are provided. Then an exhaustive description of coded-aperture collimators is presented for applications in medical imaging. This kind of collimators improves very significantly the trade-off between spatial resolution and sensitivity of the imaging system. 3D and 4D statistical reconstruction algorithms based on maximum-likelihood expectation-maximisation are then presented. This will be used to reconstruct acquisitions made through coded-aperture collimators. A recent method called non-negative matrix factorisation (NMF) is then introduced and will be used in the third chapter to perform 4D-reconstructions by decoupling spatial and temporal distributions in order to reconstruct spatial distributions from very low amount of detected photons. Finally, some gradient-based optimisation algorithms are presented and they will be used to optimise models that are encountered in NMF.

1.1 Pharmacokinetics and its models

Welling defines pharmacokinetics as "the study of the rate change of drug concentrations in the body" (Welling 86). Pharmacokinetics investigates the interactions of drugs after administration to patients: absorption, distribution, interaction and excretion. It plays a predominant role in drug development as it allows the determination of characteristics of new molecules, as their specificities to tissues and their dynamics in them. Such analyses are mostly based on compartmental or non-compartmental models, which are detailed in following sections.

In vivo nuclear imaging techniques are most often used for the characterisation of pharmacokinetics since they are almost the only quantitative methods (Zaidi 03) and their sensitivity is incomparable (up to 10^{-11}-10^{-12}mol/L) (Jain 03).

Compartmental and non-compartmental analyses require that the arterial input function is known. The arterial input function of a drug is the concentration of this drug in arterial blood, more precisely in plasma. It is a time-dependent function, since the injected molecules spread into the body, are taken up by tissues and eventually are cleared from the body.

1.1.1 Compartmental analysis

Compartmental analyses model the body as compartments which constitute physiologically separated pools. A compartmental model is characterised by its number of compartments and the dependencies between them. All these compartments depend on each other through rate constants k_i. In compartmental analysis, it is usually assumed that distribution and chemical interactions are first-order processes and that they do not saturate. From these rate parameters, some macro parameters are defined that provide essential information on the behaviour of the molecules under investigation and their physiological functions (Watabe 06). Common macro-parameters are the distribution volume and the binding potentials. The distribution volume is an indicator of the ability of the molecules to be taken up by tissues and is commonly used for the study of new drugs. The binding potential is a combined measure of the density of receptors and the affinity of the molecules to these receptors. This is of particular interest in drug development as the affinity of the drugs to some receptors is an indicator of the quality of the drug.

Compartmental analysis is a parametric analysis. Several types of models are available to perform the compartmental analysis that models *in vivo* mechanisms. The number of compartments in the model depends on the chemical and biological properties of the molecules under investigation. When compartmental models get complex, the number of unknown parameters increases, and it is usually not possible to estimate them because of considerations on statistical variability.

1.1.1.1 One-compartment model

The one-compartment model is the most simple pharmacokinetic model. It assumes that the drug is distributed in a single compartment. The most popular application of one-compartment models is the blood flow measurement with ^{15}O labelled molecules using PET. However, it can also be applied as a simplification of more complex models (Welling 86).

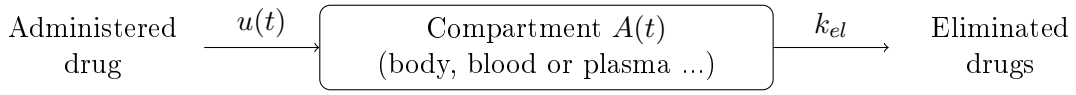


Figure 1.1 – One-compartment model

In Fig. 1.1, $u(t)$ is the quantity of molecules that enters the compartment as a function of time and k_{el} is the clearance. u is a mass (g) and k_{el} is a rate constant that has the dimension of the inverse of a time. On an infinitesimal duration, the variation of drugs amount $d\mathrm{m}$ in compartment is given by Eq. (1.1) as it is assumed that transfers inside and outside the compartment are first-order mechanisms.

$$d\mathrm{m} = du - k_{el} \cdot \mathrm{m} \cdot dt \tag{1.1}$$

The molecules are usually administrated to patients through bolus intravenous injections. In this case, the function $u(t)$ is close to a Dirac function, all the dose being injected within a few seconds. In the following, it is assumed that $u(t)$ is a Dirac function and then $du = 0$. The evolution of the quantity of molecules in the compartment can consequently be modelled as:

$$\frac{d\mathrm{m}}{\mathrm{m}} = -k_{el} \cdot dt \tag{1.2}$$

$$\mathrm{m}(t) = \mathrm{m}_0 \cdot e^{-k_{el} \cdot t} \tag{1.3}$$

In practice, it is not possible to achieve a bolus injection that can be mathematically modelled by a Dirac function, which makes the theoretical analytic model of the quantity of mass m to be more complicated than in Eq. (1.3).

When drug administration is performed continuously as for infusion, it can be assumed that $du = k_0 \cdot dt$ and the evolution of the quantity of molecules in the compartment is then given by Eq. (1.5)

$$\frac{d\mathrm{m}}{dt} = k_0 - k_{el} \cdot \mathrm{m} \tag{1.4}$$

$$\mathrm{m}(t) = (k_0/k_{el}) \cdot \left(1 - e^{-k_{el} \cdot t}\right) \tag{1.5}$$

The point of compartmental analyses is that it is possible to estimate from Eq. 1.3 or (1.5) the unknown physiological parameter k_{el} from experimental measurements of the quantity of mass m. This is performed for a series of time points. The measurement of m is made possible through labelling molecules under investigation with radiotracers, so that the amount of m can be quantified through nuclear measurements or imaging, since the quantity of mass m is proportional to the measured activity in the compartment of interest.

1.1.1.2 Model with more than one compartment

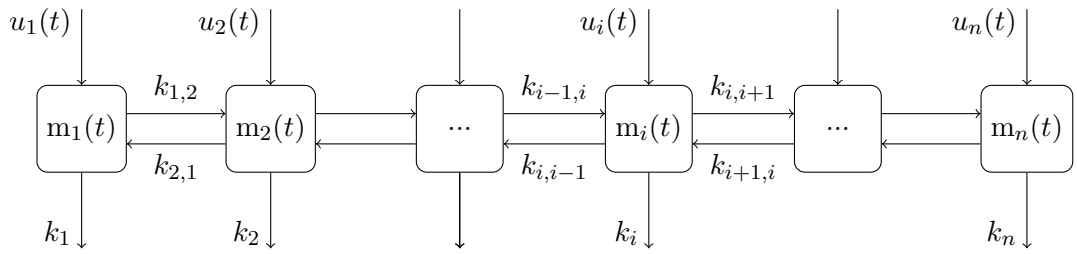


Figure 1.2 – Multi-compartment catenary model

More complex models are usually required to describe molecule behaviours in the human body. Models are classified according to the relations that exist between compartments. The main models are the catenary model and the mammillary model (Welling 86; Vicini 00). For the catenary model, it is assumed that the model is made of compartments that form a chain as described in Fig. 1.2. Mass of molecules m_i in every compartment i is ruled according to Eq. (1.6) as long as no saturation appears in compartments. This can be written in a matrix form as in Eq. (1.7) and can be solved through basis changes so that these ordinary differential equations are decoupled. Another method is to use Laplace transform (Mayersohn 70).

$$\frac{\partial \mathrm{m}_i}{\partial t} = k_{i-1,i} \cdot \mathrm{m}_{i-1} + k_{i+1,i} \cdot \mathrm{m}_{i+1} - (k_{i,i+1} + k_{i,i-1} + k_i) \cdot \mathrm{m}_i + u_i(t) \tag{1.6}$$

By denoting $\dot{m}_i = \dfrac{\partial m_i}{\partial t}$

$$\begin{bmatrix} \dot{m}_1 \\ \vdots \\ \dot{m}_i \\ \vdots \\ \dot{m}_n \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & & & \\ c_2 & a_2 & b_2 & & \\ & c_i & a_i & b_i & \\ & & & c_n & a_n \end{bmatrix} \cdot \begin{bmatrix} \mathrm{m}_1 \\ \vdots \\ \mathrm{m}_i \\ \vdots \\ \mathrm{m}_n \end{bmatrix} + \begin{bmatrix} u_1(t) \\ \vdots \\ u_i(t) \\ \vdots \\ u_n(t) \end{bmatrix} \tag{1.7}$$

This can be generalised into Eq. (1.8)

$$\frac{\partial \mathbf{m}}{\partial t} = K\mathbf{m} + \mathbf{u}(t) \tag{1.8}$$

with K being any real matrix, meaning that any compartments can interact with other compartments.

Fig. 1.3 addresses a three-compartment model, which is standard in pharmacokinetics since it models FDG interactions in tissues, which is an glucose analog and is consequently a good indicator of energy consumption by tissues. In this case, the distribution volume DV and the binding potential BP can be expressed as functions of rate constants between compartments:

- Binding potential $BP = \dfrac{k_{2,3}}{k_{3,2}}$
- Distribution volume $DV = \dfrac{k_{1,2}}{k_{2,1}}(1 + BP)$

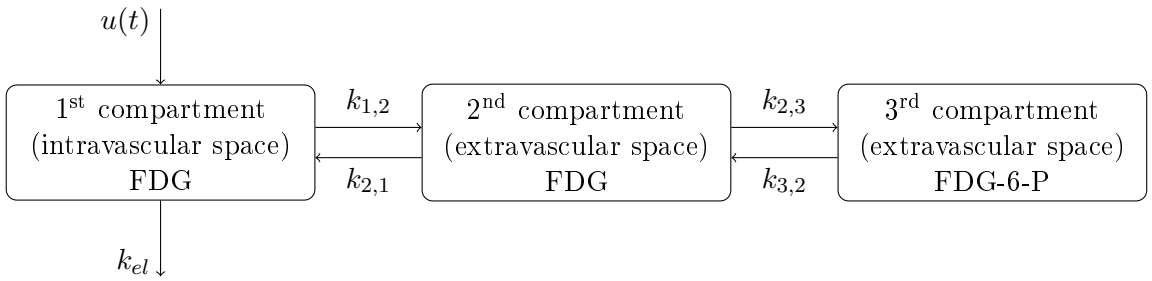


Figure 1.3 – Three-compartment model

The mammillary model is another model, which assumes that there is a central compartment, as blood, connected to peripheral compartments. Peripheral compartments are assumed to be unconnected one from another as in the example shown Fig. 1.4 (Schmidt 02). The first compartment is usually the arterial blood, the second is called the free compartment and corresponds to the unbound molecules in the extracellular space and the third compartment usually refers to the specific binding compartment, which is the compartment of interest. Indeed, it is usually made of the target tissues for which the new molecules are designed. Finally the fourth compartment commonly refers to the non-specific binding compartment, which is a compartment that also interacts with the drugs. It is in competition with the specific binding compartment.

Physiological parameters for such models are estimated by fitting model outputs to experimental data (Muzic 01). Depending on the molecules and on the kinetics models, the determination of the arterial input function can be indispensable. For instance it is necessary for the determination of cerebral metabolic rate of glucose by FDG-PET, which is usually denoted CMRGlu.

1.1.1.3 Quantitative imaging

PET provides both the spatial distribution and the quantitative measurements of the concentration of labelled compounds as time-dependent functions, which are called time activity curves (TAC). Several time frames can indeed be reconstructed from dynamic acquisitions.

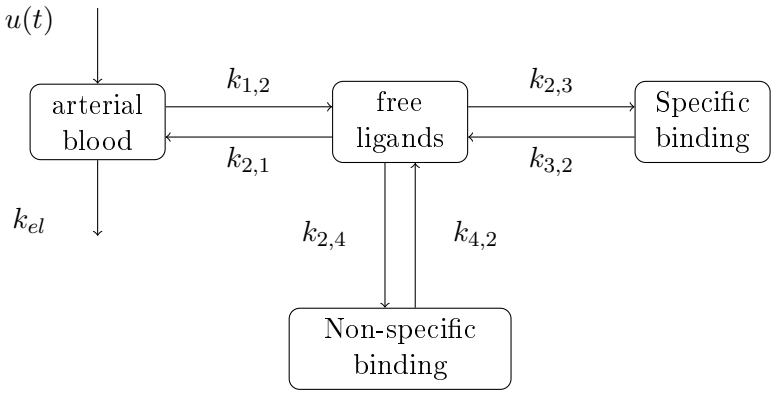


Figure 1.4 – 4-compartment model, with competitive binding potential

This allows the estimation of TACs in tissues for the drugs under investigation. The comprehension of physiological processes depends on the ability to quantify physiological variables as the rate constants of transfers from one compartment to another. By deriving the metabolic rate of the injected molecules as well as some other physiological parameters, some mechanisms and diseases can be investigated and their evolution can be monitored. For instance, the myocardial metabolic rate of glucose (MMRGlu) in cardiac studies allows the assessment of the myocardial viability after infarction (Brogsitter 05). For brain investigations, the local cerebral metabolic rate of glucose (LCMRGlu) is directly connected to cerebral functions and its estimation brings a lot of information on the activity of brain structures (Sokoloff 77). LCMRGlu is a functional indicator of the brain tissues and can be derived for a three-compartment model from Eq. (1.9), the first compartment being the plasmatic compartment, the second and the third being made respectively of unmetabolised and metabolised molecules in tissues (Phelps 79). This requires that the plasma arterial input function $\mathrm{C}_P(t)$ is known.

$$\mathrm{LCMRGlu} = \frac{\bar{\mathrm{C}}_P \cdot \left(\mathrm{C}_i(T) - \frac{k_{1,2}}{\alpha_2-\alpha_1}\left((k_{3,2}-\alpha_1)e^{-\alpha_1 t} + (\alpha_2 - k_{3,2})e^{-\alpha_2 t}\right) \star \mathrm{C}_P(t)\right)}{LC \cdot \left(\frac{k_{2,1}+k_{2,3}}{\alpha_2-\alpha_1}\right)\left(e^{-\alpha_1 t} - e^{-\alpha_2 t}\right) \star \mathrm{C}_P(t)} \tag{1.9}$$

where $\star$ denotes the convolution operator, C_i is the tracer concentration in the tissue of interest, C_P is the tracer concentration in plasma, $\bar{\mathrm{C}}_P$ is the mean tracer concentration in plasma over the acquisition, LC is a correction factor called the lumped constant, which compensates for the transport and chemical differences between FDG and glucose (Graham 02). Finally k_i are the rate constants of a 3-compartment model as the one shown in Fig. 1.3. α_1 and α_2 $(mol \cdot s^-1)$are given by (1.10) and (1.11).

$$\alpha_1 = \frac{k_{2,1} + k_{2,3} + k_{3,2} - \sqrt{(k_{2,1} + k_{2,3} + k_{3,2})^2 - 4k_{2,1}k_{3,2}}}{2} \tag{1.10}$$

$$\alpha_2 = \frac{k_{2,1} + k_{2,3} + k_{3,2} + \sqrt{(k_{2,1} + k_{2,3} + k_{3,2})^2 - 4k_{2,1}k_{3,2}}}{2} \tag{1.11}$$

The estimation of the CMRGlu requires the estimation of the plasma input function $\mathrm{C}_P(t)$ as well as the rate constants that govern the compartmental model, whose determinations also depend on the input function $\mathrm{C}_P(t)$.

1.1.2 Non-compartmental analysis: multiple-time graphical analysis

Multiple-time graphical analysis (MTGA) provides information about the radioligands and does not need that a particular compartmental model is specified. It is a non-compartmental

analysis. The only information which is required is whether there are some irreversible compartments in the model or not. It may provide a measure of the tracer uptake in tissues and on tracer binding. It was first introduced by Patlak for radioligands which were taken up irreversibly by some of the tissues (Patlak 83). It was then generalised for radioligands which bind reversibly to tissues (Logan 90).

1.1.2.1 Non-reversible ligands

After denoting $\mathrm{ROI}(t)$ and $\mathrm{C}_P(t)$ the concentrations of radiotracers in tissues of the region of interest (ROI) and in plasma at time t, respectively, the plot $\left(\mathrm{ROI}(t) \Big/ \mathrm{C}_P(t)\right)$ versus $\left(\int_0^t \mathrm{C}_P(t')\mathrm{d}t' \Big/ \mathrm{C}_P(t)\right)$ becomes linear after some time when the radiolabelled molecules are irreversibly trapped in at least one compartment (Patlak 83). This is usually referred as the steady-state condition. The linear portion of the plot can be written as:

$$\frac{\mathrm{ROI}(t)}{\mathrm{C}_P(t)} = K_i \frac{\int_0^t \mathrm{C}_P(t')\mathrm{d}t'}{\mathrm{C}_P(t)} + V \tag{1.12}$$

K_i and V can be determined graphically. K_i is called the influx constant and is a measure of the amount of radiotracers which were taken up in tissues in relation to the amount of available tracers in plasma. K_i depends on the tracer binding potential, the higher the binding potential is, the higher the influx constant is. V is a constant which depends on the regional blood volume and on the distribution volume of the reversible compartments. $\mathrm{ROI}(t)$ is obtained from nuclear imaging and currently $\mathrm{C}_P(t)$ is mostly obtained from arterial blood sampling.

This is an easy method to obtain properties of new radioligands, as its binding potential.

1.1.2.2 Reversible ligands

In section 1.1.1.2, an equation which describes exchanges between compartments was written. By considering that the plasma is a particular compartment and using concentrations instead of quantities of mass, it can be rewritten (Logan 90; Logan 03):

$$\frac{\partial \mathbf{C}}{\partial t} = \boldsymbol{K} \cdot \mathbf{C} + \boldsymbol{K}_1 \cdot \mathrm{C}_P(t) \tag{1.13}$$

with $\mathbf{C}$ being the concentrations in the tissue compartments, C_P being the concentration of unmetabolized radioligands in plasma, $\boldsymbol{K}$ the matrix of transfer constants between tissue compartments and $\boldsymbol{K}_1$ the vector describing the transfer from plasma to tissues. $\boldsymbol{K}_1$ is a vector usually made of a single nonzero coefficient. By defining a ROI in the reconstructed images, the concentration in this region can be modelled with:

$$\mathrm{ROI}(t) = \sum_i^n \mathrm{C}_i(t) + V_P \mathrm{C}_P \tag{1.14}$$

with C_i being the concentration of the compartment i in ROI, n the number of compartments in ROI and V_P being the blood volume contained in the ROI. As a result, $\left(\int_0^t \mathrm{ROI}(t')\mathrm{d}t' \Big/ \mathrm{ROI}(t)\right)$ can be written as in Eq. (1.15), which can be seen as a linear equation when the second term becomes constant, which is usually true for time t higher

than a cut-off time t^* which usually ranges from 20 to 90 min. The i^{th} coefficient of $\mathbb{U}_n$ is 1 if the i^{th} compartment stands inside ROI, 0 otherwise.

$$\left(\frac{\int_0^t \mathrm{ROI}(t')\mathrm{d}t'}{\mathrm{ROI}(t)} \right) = (-\mathbb{U}_n^{\mathrm{T}} \boldsymbol{K}^{-1} \boldsymbol{K}_1 + V_P) \left(\frac{\int_0^t \mathrm{C}_P(t')\mathrm{d}t'}{\mathrm{ROI}(t)} \right) + \frac{\mathbb{U}_n^{\mathrm{T}} \boldsymbol{K}^{-1} \mathbf{C}(t)}{\mathbb{U}_n^{\mathrm{T}} \mathbf{C}(t) + V_P \mathrm{C}_P(t)} \tag{1.15}$$

$$= DV \cdot \left(\frac{\int_0^t \mathrm{C}_P(t')\mathrm{d}t'}{\mathrm{ROI}(t)} \right) + int \tag{1.16}$$

By plotting $\left(\int_0^t \mathrm{ROI}(t')\mathrm{d}t' \Big/ \mathrm{ROI}(t) \right)$ versus $\left(\int_0^t \mathrm{C}_P(t')\mathrm{d}t' \Big/ \mathrm{ROI}(t) \right)$, it is possible to estimate the slope DV, also called the total distribution volume, and the intercept int of the linear portion of the curve. $\mathrm{ROI}(t)$ and $\int_0^t \mathrm{ROI}(t')\mathrm{d}t'$ are obtained from reconstructed images of the ROI and $\int_0^t \mathrm{C}_P(t')\mathrm{d}t'$ from arterial blood sampling. For a two tissue compartment model, the total distribution volume can be expressed with rate constants:

$$DV = \frac{K_1}{k_2} \left(1 + \frac{k_3}{k_4} \right) \tag{1.17}$$

This means that a large DV indicates either a large number of tracer binding sites, through $(k_3/\,k_4)$, or a large K_1, which depends on the blood flow in arteries and on the capillarity of the tissues. DV is consequently a good indicator whether a radioligand binds to a target area or not.

This method was evaluated on simulated data (Logan 03). A tissue TAC was generated from a measured input function and a two tissue compartment model whose constant rates are assumed to be known. This models the kinetics of $^{11}C-$racloprid, a radioligand which binds to basal ganglia in the brain. When only points from t=80 to t =120 min are used to perform the linear regression, DV is equal to 2.01 ml/ml, but when the regression is performed on points between t = 80 and t = 250 min, DV is equal to 2.26 ml/ml, which corresponds to bias of 20% and 9% respectively, the expected DV being 2.42 ml/ml. This illustrates that the cut-off time t^* plays an important role in the estimation of the distribution volume DV.

This method also requires that the arterial input function $\mathrm{C}_P(t)$ is determined.

1.1.2.3 Graphical analysis without blood sampling

Graphical analysis were generalised in order not to use the arterial input function $\mathrm{C}_P(t)$ but a reference region instead (Patlak 85; Logan 00; Logan 03). This works for both reversible and non-reversible ligands. A common reference region for brain investigations is the cerebellum.

When it is assumed that the radioligands are irreversibly trapped in at least one compartment, Eq. (1.12) can be replaced by Eq. (1.18) and the coefficients K_i and V can be estimated from acquisitions with only few time frames, and no input function (Patlak 85).

$$\frac{\mathrm{ROI}(t)}{\mathrm{REF}(t)} = \frac{K_i}{\mathrm{DV}^{\mathrm{REF}}} \frac{\int_0^t \mathrm{REF}(t')\mathrm{d}t'}{\mathrm{REF}(t)} + V' \tag{1.18}$$

When it is assumed that the radioligands are reversibly exchanged between compartments, Eq. (1.15) can be applied to a reference region:

$$\int_0^t \mathrm{C}_P(t')\mathrm{d}t' = \frac{1}{DV^{REF}} \left(\int_0^t \mathrm{REF}(t')\mathrm{d}t' - int^{REF}\mathrm{REF}(t) \right) \tag{1.19}$$

which can be substituted in Eq. (1.15), with $DVR = DV^{ROI}/DV^{REF}$:

$$\left(\frac{\int_0^t \mathrm{ROI}(t')\mathrm{d}t'}{\mathrm{ROI}(t)} \right) = DVR \left(\frac{\int_0^t \mathrm{REF}_P(t')\mathrm{d}t'}{\mathrm{ROI}(t)} \right) + DVR \cdot int^{REF} \frac{\mathrm{REF}(t)}{\mathrm{ROI}(t)} + int^{ROI} \tag{1.20}$$

DVR is called the distribution volume ratio. The second term of the right-hand side is usually negligible compared to the first term, which results in a linear expression when the assumption that int^{ROI} and int^{REF} are constant is valid. It can be observed that Eq. (1.20) does not depend on the plasma input function

For a two tissue compartment model and choosing a reference region without binding sites (which can be modelled with a 1 tissue-compartment model), the DVR is given by Eq. (1.21) and can be simplified to Eq. (1.22) if the ratios of transport constants are the same in the reference region and in the region of interest.

$$DVR = \frac{K_1}{k_2}\left(1 + \frac{k_3}{k_4}\right) \Big/ \frac{K_1^{REF}}{k_2^{REF}} \tag{1.21}$$

$$DVR \approx 1 + \frac{k_3}{k_4} \tag{1.22}$$

Simulations similar to Sec. 1.1.2.2 were performed in order to estimate the accuracy of the method. It was found that $DVR = 4.41$ ml/ml, the expected value being 4.43 ml/ml (error is lower than 1%) (Logan 03).

MTGA provides robust information on radioligands. A drawback of these methods is that they only provide information on macro-parameters, as the binding potential or the distribution volume for instance. Furthermore, they tend to underestimate DV and DVR, and consequently the measure of the binding potential of the radioligands. Nevertheless, because of their simplicity, they are particularly appropriate for early investigations of new tracers properties, all the more so as no detailed compartmental model is required to lead such an analysis.

1.2 Input function determination

It was explained in section 1.1 that the determination of the tracer-input function may be fundamental to perform quantitative estimation of kinetics rate constants and consequently estimate the sensitivity and specificity of new molecules in target-tissues. It is commonly measured through blood sampling, but this procedure may be painful and dangerous for patients (Hall 71). It is also a time-consuming protocol that is not adapted to daily clinical investigations. Furthermore it unnecessarily exposes the medical personal to radiation. Number of methods were developed to estimate it in a non-invasive way or with minimal invasion. Intending to measure the β^+- input function in a non-invasive way, it is clear that only the whole blood concentration can be estimated, not the plasma concentration. As a result, such methods only present an interest for molecules with a low metabolic rate in blood.

1.2.1 Arterial blood sampling and other invasive techniques

This method is the gold standard technique. It is invasive but it is the most accurate way of determining the β^+-input function. Samples of blood are drawn from the patient and measured with a β- or γ-counter. This procedure was automated (Eriksson 88), which presents the advantage to reduce the exposition of the medical staff to the radiation. Furthermore, as small volumes of blood are continuously drawn, the temporal resolution can go below the second. Blood is drawn through a catheter and a peristaltic pump brings it to detectors. The activity in blood samples were determined either by counting positrons with plastic scintillation detectors or by counting γ with BGO detectors. Measurements are performed in a lead box whose walls are 5 cm-thick, which results in a background lower than 1 ‰. When the blood activity is determined from counting positrons with a plastic scintillation detector, the method is very sensitive but requires calibrations as the sensitivity of such a system depends on the energy of the positrons, which varies depending on the positron energy and consequently on positron emitters. Positrons must go through the wall of tubings before being detected, the wall being 0.26 mm thick. The second detection method is based on the detection of the two γ that are emitted in opposite directions when a positron annihilates with an electron. This suppresses the problem of energy calibration as the energy of the γ-particles that are emitted from a positron annihilation is 511 keV. The two γ are detected in coincidence with two BGO scintillators. Both methods work fine but errors were not quantified. As the temporal resolution is very good (1 s), studies on the dispersion function of the blood sampling system were performed. Depending on the length of the catheter between its tip and the detectors, there is a delay in the activity measurement. This delay can be corrected by appropriately deconvolving the input function by the dispersion function. When detectors are placed 60 cm away from the tip of the catheter, the characteristic time τ varies from 17 min^{-1} to 40 min^{-1}, the speed of the peristaltic pump being respectively 2.5 ml/min and 5 ml/min, the inner diameter of the tubings being 0.94 mm. The input function which is derived with this method was used to estimate the rate constants k_1 and k_2 of a two-compartment model, as well as the cerebral blood volume CBV of a patient, which is the volume of blood in a given mass of brain tissues. The input function is corrected for dispersion. Radiolabelled molecules were ^{11}C-raclopride and the aim was to quantify the transport of these molecules through the blood-brain barrier, the two compartments being respectively the arteries and the brain. It was found that $k_1 = 0.15\ \text{ml} \cdot \text{g}^{-1} \cdot \text{min}^{-1}$, $k_2 = 0.48\ \text{ml} \cdot \text{g}^{-1} \cdot \text{min}^{-1}$ and CBV = 0.033 ml/g.

Similar methods were also investigated to estimate the arterial input function in rodents, the difficulty being to draw very small volumes of blood. First an experimental set-up with a blood counter and sampler was developed which continuously draws blood from rats or mice (Lapointe 98). This blood is passed into a counter which measures the activity of radiotracers in the blood by detecting the positrons emitted with a plastic detector. The counting rate can be as high as 1 measure/s. Blood is then separated into independent samples for further analyses. This experimental set-up uses an original technique to limit the dispersion of activity in the tube. It is based on the injection of air bubbles between samples, so that they do not mix one with another. This experimental system allows the formation of samples up to 1 sample/s and samples can be as small as 10 μL. Samples can be further processed, for instance their activity can be verified with a well counter. Results are in good agreement with the values measured in the counter. The volume of blood in samples can be accurately determined by measuring the length of blood in samples. From the activity and the volume of every sample, an accurate estimation of the radioactive concentration is obtained. The standard deviation of the blood activity measured from the counter or from the samples was lower than 4% for all experiments, whatever the drawing speed and the sampling

frequency were. The delay between the peak of the input function at the catheter tip and the sampler peak is well estimated, as the length of tubings between the tip of the counter and the sampler is known. The delay between the input function peak and the sampler peak was estimated and was 44.6 s which is in good agreement with the real delay which was 44 s, the length of the tubing being 24 cm. The input function can then be corrected for dispersion as the characteristic transport time from the artery to the sampler is known. The plasma-to-blood ratio can be determined by centrifugating the samples. The plasma time activity curve (TAC_P) is then deduced from this ratio and the whole blood TAC which is estimated either from the counter or from the samples. The amount of metabolites can also be determined from the samples, which can be useful to investigate the pharmacokinetics properties of the radiolabelled molecules in the blood. Although efforts were made to control the activity dispersion in tubings by placing the counter and the sampler close to the cannula site, it is expected that the TACs which are measured in the counter or from the samples are a flatten representation of the real arterial input function because of dispersion. Furthermore a main drawback of such a system is that a large amount of blood is drawn from the rodents, which corresponds to non-physiological conditions.

It was then proposed to use an integrated microfluidic platform to collect smaller blood samples (Wu 07) and consequently estimate kinetics parameters of mice under standard physiological conditions. Samples as small as 220 nL were collected. This corresponds to a total withdrawal of blood lower than 0.5% of the total weight of mice used in investigations. The transport time of blood from the tip of the catheter to the sample wells was taken into account by estimating the blood flow in the chip, as well as the dead space between the tip and the sample wells (7 μL). The blood curves which were obtained from the blood-sampling device were compared to samples manually drawn directly from the left ventricle. An average error of 6.0% was measured for the 4 mice which were observed. This article indicates that the plasma-to-blood concentration ratio $R_{P/B}$ may not be constant during time, and should be modelled with an exponential function as $R_{P/B} = 0.386 \cdot e^{-0.191t} + 1.165$ when the radiolabelled molecules are ^{18}FDG in mice. Kinetics rate constants for the myocardium and the brain of mice were estimated from the TACs in the heart and the brain and the arterial plasma input function, the tissue TAC being obtained from microPET images and the input function being estimated with the microfluidic blood-sampling device. Fitting the TACs in the myocardium and in the brain with these rate constants gives respectively coefficients of determination R^2 of 0.99 and 0.90, which means that the model of TACs that is derived from the estimated input function fits well to the observed TACs from the microPET. The goodness of Patlak fittings (Patlak 83) is also excellent ($R^2 > 0.99$). The cerebral glucose metabolic rates for every 4 mice are then calculated from rate constants as in (Phelps 79) and the result is 21.5 ± 4.3 μmol/min/100g. This method is consequently adapted to the estimation of the arterial input function in mice. However, the input function which is estimated is not corrected for tracer dispersion in the catheter, which results in an input function whose peak is larger than the one of the real arterial input function. It is important to notice that the dispersion of the arterial input function has a major impact since the metabolic blood flow can change by a factor 4, depending on whether the dispersion is taken into account or not (Convert 07a; Convert 07b) (no correction on dispersion: 1.12 ± 0.17 ml/g/min, with correction: 4.84 ± 0.8 ml/g/min). This correction is performed by modelling the dispersion characteristic time as a function of the catheter length and the blood flow in the catheter. The real arterial input function is then obtained by deconvolving the measured input function through Laplace transforms. This study points out the importance of correcting the measured input function for dispersion.

It was also proposed to insert a β-probe in a large artery of small animals to measure the activity of blood by direct detection of positrons (Pain 04). With this method, no blood is

drawn from animals, which means that rodents are under standard physiological conditions during investigations. Furthermore, no dispersion has to be corrected as the arterial input function is directly measured from an artery. This potentially provides a more accurate estimation of the input function than methods presented previously. However, a significant amount of the signal measured by the β-probe is due to the contribution of the tissues which surround the artery. As a result, it was proposed to use a second β-probe which estimates the background outside the artery. The arterial input function is then retrieved by subtracting both signals. The results were validated by comparing the area under the curve (AUC) of the input function estimated with the β-probe to the AUC of the input function which was obtained from blood samples. The mean difference was calculated from a 5-animal study and is 5%. From the input function measured with the two β-probes, the kinetics rate constants of a 3-compartment model can be estimated, as well as the ^{18}FDG metabolic rate in tissues. For striatum, the estimation of the ^{18}FDG metabolic rate in 3 rats shows a variation of $15\% \pm 18\%$ depending on whether the input function was measured from the β-probe or from manually-drawn blood samples. Main advantages of such a technique are that it can estimate the arterial input function without blood loss and it is efficient at detecting positrons ($> 10\%$) which allows an excellent temporal resolution. This is particularly interesting for short-lived radiotracers and for input functions which have a narrow peak. The main drawback is that the sensitivity of the probe depends on the position of the probe inside the artery. This limits the accuracy of the estimation of the input function because of reproducibility considerations.

Arterial blood sampling methods result in an accurate estimation of the plasmatic input funciton $C_P(t)$, but are invasive. Some investigations were done in order to estimate it in a non-invasive way.

1.2.2 Image-derived input function extraction

Instead of drawing blood samples from subjects or from animals, it was investigated whether the arterial input function could be estimated from reconstructed images. For that, a region of interest is drawn on vascular structures. This segmentation is most often done manually, but it may not be reproducible, because the results depend strongly on the operator who segments the data. Some methods were consequently developed to automate segmentation steps. They rely on factor analysis (Bazin 80) or independent component analysis (Hyvärinen 01). Factor Analysis of Dynamic Sequence will be developed in more details in section 1.6.1.

The arterial input function is usually denoted image-derived input function (IDIF) when it is obtained from these methods. The extraction of the IDIF is usually performed from sequences of cardiac images or brain images, as illustrated in the following.

1.2.2.1 Extraction of arterial input function from cardiac PET images

In cardiac imaging, it is possible to measure the arterial input function directly from the PET-images. One method to estimate the arterial input function is to reconstruct the whole field-of-view and to draw regions of interest on blood pools (van der Weerdt 01; de Geus-Oei 06). 4 vascular structures were compared in this study: the left atrium, the left ventricle, the ascending aorta and the descending aorta. Input functions were estimated from every structure and the 4 MRGlu were derived from these 4 input functions. All MRGlu were compared to the reference MRGlu obtained from the input function measured by arterial blood sampling. The ratios were respectively 0.81 ± 0.06, 0.79 ± 0.08, 0.97 ± 0.07 and 1.00 ± 0.11. The ascending aorta provides the best estimate of the input function, also the corresponding

standard deviation is larger than for the descending aorta. The ascending aorta is a large structure that consequently suffers slightly from partial-volume effect and statistical noise.

Factor analysis was also applied to cardiac imaging in order to extract the pure blood time-activity curves with ^{13}N-ammonia (Wu 95) and FDG in small animal (Wu 96) in order to estimate respectively the rate constants of a three-compartment model and the metabolic blood flow in dogs. This method has the advantage of dealing with the partial volume effect as every pixel of the reconstructed images is viewed as a linear combination of the contribution of every tissue. It also solves issues concerning segmentation of ROIs in images. The mean difference between the arterial input function measured from arterial blood sampling and from factor analysis was lower than 10% for monkey studies. The mean errors on the estimation of kinetic parameters for monkey studies was lower than 8%, the kinetic model being assumed to be made of three compartments.

Some non-invasive techniques were also tested on small animals (Laforest 05) but it is difficult to estimate the arterial input function because the spatial resolution of microPET scanner is not sufficient to produce images without severe partial volume effect and spillover effect when imaging mice's heart. However, for studies on rats, it was found that the error on the area under the curve (AUC) of the estimated input function when compared the AUC determined with blood sampling is $9.8\% \pm 2.7\%$. This means that the input function can be correctly estimated from a succession of cardiac PET images in rats.

1.2.2.2 Extraction of arterial input function from brain PET images

A few methods try to extract the arterial input function from the dynamic brain PET-images, but this is a difficult task because the vascular structures are small and the radiotracer concentration in surrounding tissues is usually high. As it is usually not possible to observe carotids in reconstructed PET-images, in particular in early frames, it was suggested to draw regions-of-interest from images obtained with MRI-scan whose spatial resolution is much better and allows an accurate drawing of the ROI (Litton 97). At early time points, the arterial input function seems to be well estimated although no error is given on the difference between this method and arterial blood sampling. For later time points, estimations of concentrations are not as good, due to important spillover effect from surrounding tissues. This is the main drawback of the method. Furthermore, this method raises the question of image registration from two very different imaging modalities: MRI and PET. It would be more appropriate to segment the vascular structures directly from the PET images. Automatic segmentation through factor analysis (see Sec. 1.6.1) was proposed to extract the shape of the input function and then to estimate physiological parameters, for instance CMRGlu (Bentourkia 05). This results in estimations of the CMRGlu obtained from automatic segmentation of brain arteries close to the CMRGlu obtained from arterial blood sampling, numerical results being respectively 41.77 ± 6.29 $\mu\text{mol} \cdot 100\text{g}^{-1} \cdot \text{ml}^{-1}$ and 41.89 ± 3.13 $\mu\text{mol} \cdot 100\text{g}^{-1} \cdot \text{ml}^{-1}$ (mean bias $< 1\%$, standard deviation $\approx 15\%$).

As an alternative to factor analysis, it was proposed to extract plasma time-activity curve from dynamic brain PET images with an ICA-based method (Naganawa 05). ICA allows the separation of sources that are not Gaussian-distributed, which is the case in the problem of plasma time-activity curve extraction. In Naganawa's paper, the input function is not assessed by identifying vascular structures in the brain images, but rather by estimating the input function from the blood volume image which is the volume of the extravascular space where blood is in. PET-images are pre-processed so that the extraction of the input function is possible. The pre-processing consists in whitening the images by appending their opposite at the end of the dataset: the dataset is twice as large as the PET-images and its average value is zero. The difference between the blood volume image and the tissue image is then

enhanced according to a standardisation that sharpens the distribution of the blood volume image into a peak that is centred around 0 and in the same time concentrates the values of the tissue image around 1 or -1. In this study the ICA was applied on the following objective function that includes a penalty terms that constrains the solution away from 0:

$$G(u) = u^6 - \frac{\lambda}{m} exp\left(-\frac{|u|}{m}\right) \tag{1.23}$$

where λ and m are two parameters in $[10; 100]$ and $[0.1; 0.5]$. Results are robust to the variations of these parameters. Source separation for such an objective function is based on the ability of u^6 to detect long tails of the distribution of the blood volume image and the tissue image. Beyond the problem that this method returns negative values for some voxels of the blood volume image, this method requires an arterial blood sampling because the solution requires scaling, whereas such a puncture should be avoided as much as possible in order to dispose of a non-invasive method. This method was applied on human FDG-studies and the influx constant K_i derived from a three-compartment model shows a good correlation with K_i obtained from arterial blood sampling, the correlation coefficient being $\rho = 0.999$ and the fitted line being $y = 0.984x - 3.34 \times 10^{-4}$. Instead of deriving the input function from the blood volume image with ICA, a method was also reported that uses ICA to identify the carotid and the surrounding tissues (Chen 07). It requires late venous samples. It allows the estimation of CMRGlu for human FDG studies which are very close to CMRGlu obtained from arterial blood sampling, the regression line being $y = 1.02x + 2 \times 10^{-4}$ and the coefficient of determination being $R^2 > 0.999$.

An original method was proposed to incorporate the estimation of elementary temporal functions as well as their factors in the reconstruction process (Reader 06). Among other advantages, this allows the reconstruction of first frames with a better accuracy. This method is further developed in section 1.5.2.2 that presents state-of-the-art method for dynamic reconstructions. This method was tested on simulations and on clinical data. The mean error was lower than 1%, with a mean noise error being about 20% as computed by authors. The interpretation of clinical results is difficult because temporal basis functions have no physiological meaning. The arterial input function could however be viewed as a mixture of temporal basis functions.

Image-derived input functions allow the estimation of macro-parameters as the CMRGlu or MRGlu with good accuracy, but the temporal resolution in first minutes still suffers from the low efficiency of imaging system. Most methods require to scale the input function with late venous or arterial samples. Recent studies (Reader 06) seem to have the potential to address the estimation of the input function accurately even in first minutes, although some efforts need still to be made.

1.2.3 External non-invasive measurement

Several non-invasive methods were proposed to estimate the arterial input function from external detectors that are independent from the PET scanner. First, a transcutaneous detector was developed that detects positrons emitted by ^{11}C-tracers (Litton 90). The detector is made of a plastic scintillator placed in contact with the carotid artery. It is not possible to separate the contributions of the carotid artery and the jugular vein that stands close to it. However, as the temporal resolution is excellent, the arterial peak can be observed before the contribution of the jugular vein becomes significant. The venous contribution then spoils the measurements, although arterial recirculation is observed. If compared to the method proposed by Eriksson in section 1.2.1, this method presents the advantage of no time shift

between the actual radioatracer concentration that floods in artery and the measurements. Nevertheless, it only gives information in the very first instant, which is usually not sufficient. Another plastic detector was designed (Watabe 95) that better rejects the background due to γ-particles. It intends at measuring water labelled with ^{15}O from wrist arteries. It is made of two detectors: one measures the contribution of both the artery and the surrounding soft tissues, the other detector measures the concentration in soft tissues only. The arterial input function is retrieved from subtraction of these two measurements. These methods that are based on the detection of positrons work better for positron emitters as ^{15}O or ^{11}C than for ^{18}F because their positron range is larger than that of ^{18}F. Thus, the probability of reaching the plastic detector is higher. These methods are less appropriate for ^{18}F-emitters, although ^{18}F is also frequently used for radiopharmaceutical investigations.

Instead of detecting positron, it was also proposed to detect γ-particles. This suppresses the issue of the mean free path of positrons in tissues which is usually lower than the distance between the artery and the detector, resulting in a low sensitivity of the counting systems. A first investigation proposed to detect with a NaI scintillator the photons emitted from ^{15}O and that come from the superior lobe of the right lung (Nelson 93). This technique can only be used with positron-emitters that are in gaseous state because it measures the amount of radiotracers in lungs, which restricts the choice of the emitters to ^{15}O in medical imaging. The injected dose ranged between 3.14 and 3.88 GBq (85 and 105 mCi). Such a dose can only be injected for very short-lived radiotracers as ^{15}O. The input functions obtained respectively with this method and arterial blood sampling are visually very similar, although no error on the non-invasive estimation of the arterial input function is given by Nelson. It results in a mean error on the global cerebral blow flow (CBF) of 6.8%, the standard deviation being 4.9%. This means that CBF is well estimated from this non-invasive technique and results are reproducible. It was also proposed to perform 3D-imaging on the human radial artery by designing a small PET scanner whose ring diameter is 90 mm-large so that the arm can fit through (Rajeswaran 92). γ-particles are detected in coincidence. The spatial resolution is above 3 mm when the artery lumen is expected to be 1 mm-large. As a result, this method suffers from the contribution of the veins and surrounding tissues that stand in the arm since it is not possible to accurately separate the artery contribution from other tissue contributions. Another arm scanner was recently developed by a group at Brookhaven. It is based on LSO-scintillators that are coupled to avalanche photodiodes (APD) (Villanueva 03; Shokouhi 03; Kriplani 06; Kriplani 07). The spatial resolution of this method is sufficient to discriminate the arterial and venous radiotracer distribution by reconstructing the planar coincidence images. The efficiency of this imaging system is 0.11 cps/Bq/cm^{-3} (4 cps/nCi/cm^{-3}), which limits the temporal resolution. Furthermore, this method strongly suffers from partial volume effect, the voxel size being of the same characteristic dimension as the diameter of the ulnar artery ($\approx$ 1 mm). Some input functions were estimated from this method (Kriplani 06), but no comparison to the real input function is available.

1.2.4 Statistical modelling on a population

It is expected to perform no arterial blood sampling for the estimation of the arterial input function, but non-invasive measurement gives inaccurate results. To overcome the main limitation of non invasive external and image-derived measurements (insufficient sensitivity and spatial resolution), it was proposed to compute the arterial input function from an analytical or statistical consideration. It was proposed to take the arterial input function through parametric models (Phillips 95). In this study, parameters are estimated from only 6 arterial samples. For FDG-compartmental model, the plasma time-activity curve was

modelled according to Eq. (1.24).

$$\mathrm{C}_P(t) = b\left(m \cdot t \cdot e^{-n \cdot a \cdot t} + e^{-a \cdot t}\right) \tag{1.24}$$

From this model, it was found that with only six arterial blood samples, it is possible to achieve a solution with an error around ± 5%. Another example of population-based analysis requires a unique arterial or venous blood sample (Wakita 00). It has the advantage to decrease the exposure of the staff to radiation and to simplify the drawing procedure since only one sample has to be dealt with, to be compared to several dozens in other studies. In this investigation, 120 patients underwent multiple blood sampling after FDG-injection. A mean arterial input function was deduced from these 120 arterial input functions, which will be called the standardised input function. It was observed that the variation between patient arterial input function was minimal at 12 min after the injection. With only one arterial blood sample 12 min after injection, it is then possible to estimate accurately the arterial input function of any patients. The error was calculated and was 1.7% if compared to multiple arterial blood sampling. The coefficient of correlation is 0.997, denoting a low standard deviation of results. The difference between the arterial concentration and the venous concentration disappears after 40 min. As venous sampling is preferred to arterial sampling, input functions are rather estimated from a late venous sample drawn 40 min after bolus injection. It gives an error on input functions of about 3.6%, the correlation coefficient being 0.996. This shows that the arterial input function can be estimated from a single sample, either arterial or venous, provided a standardised input function was previously determined from a large number of patients which underwent arterial blood sampling.

Finally, it was investigated whether the arterial input function for every patient could be retrieved with no sample at all. As in the previous method, a standardised input function is derived from a large population and it is is then scaled according to the body surface area (BSA) of the patient instead of a blood sample (Shiozaki 00). Once the standardised input function is derived, this method requires no blood sampling at all but the error in the estimation of CMRGlu in gray and white matter from input functions obtained with this method are non negligible, respectively $8.9 \pm 9.0\%$ and $10.9\% \pm 9.0\%$.

1.2.5 Hybrid methods

Hunter developed a simplified kinetics model (SKM) to estimate the glucose metabolic rate (MRGlu) of a tumour from a single venous sample and a single static image with ^{18}FDG (Hunter 96). It is particularly relevant for estimating the MRGlu from patients who can not tolerate prolonged imaging times or who can not undergo arterial blood sampling. It is also an interesting alternative to full kinetics analysis which requires much longer acquisition time than SKM, since the acquisition lasts only 15 min for SKM, which should be compared to 1 to 3 hours for full kinetics analysis. This method is illustrated on lung tumours which can be modelled with a 3-compartment 3-rate constant model (the fourth rate constant k_4 can be assumed to be negligible for such investigation). The concentration C_T in the tumour can be modelled as:

$$\mathrm{C}_T(t) = \frac{K_1 k_2}{k_2 + k_3} e^{-(k_2+k_3)t} \sum_{i=1}^{3} A_i \left[\frac{1 - e^{-(b_i-(k_2+k_3))t}}{b_i - (k_2 + k_3)}\right] + \frac{K_1 k_3}{k_2 + k_3} \sum_{i=1}^{3} \frac{A_i}{b_i}\left(1 - e^{-b_i t}\right) \tag{1.25}$$

if assuming that the arterial input function $\mathrm{C}_P(t)$ is well described by a tri-exponential function:

$$C_P(t) = \sum_{i=1}^{3} A_i e^{-b_i \cdot t} \tag{1.26}$$

First, MRGlu is calculated from the arterial input function measured from arterial blood sampling, which is the gold-standard method, in order to have reference MRGlu values. This will allow the assessment of the performance of SKM. For every patient, parameters $(A_1, A_2, A_3, b_1, b_2, b_3)$ are calculated by fitting Eq. (1.26) to blood concentrations obtained from arterial blood sampling. Then, parameters (K_1, k_2, k_3) are estimated by fitting Eq. (1.25) to the measured activity in tumours from dynamic sequences of tumour images obtained from PET. The reference MRGlu_{ref} is then derived from:

$$\mathrm{MRGlu}_{ref} = \mathrm{C}_{glu} \frac{K_1 k_3}{k_2 + k_3} \tag{1.27}$$

where C_{glu} is the concentration of glucose in blood. MRGlu_{ref} was calculated before and after treatment of a tumour for 13 patients, and results vary from 0.038 μmol/min/g to 0.37 μmol/min/g.

SKM intends at estimating MRGlu without performing arterial blood sampling on every patient. SKM assumes that the first term in the right-hand side of Eq. (1.25) is small (about 10% of C_T). If neglecting it, the MRGlu depends on the activity in tumour C_T and the arterial input function through:

$$\mathrm{MRGlu} = \mathrm{C}_{glu} \frac{\mathrm{C}_T(t)}{\sum_{i=1}^{3} \frac{A_i}{b_i} (1 - e^{-b_i t})} \tag{1.28}$$

$\mathrm{C}_T(t)$ is calculated from a single 15 min-long acquisition of the tumour (from 45 min to 60 min after FDG-injection). The denominator is the area under the curve of the arterial input function (which was modelled as a tri-exponential as mentioned before), as a result, it can be calculated from the parameters $(A_1, A_2, A_3, b_1, b_2, b_3)$ which determine the arterial input function. The arterial input functions of a control group were modelled with such a function and it was observed that characteristic times b_1, b_2 and b_3 were almost constant ($b_1 = 9.33$ min$^{-1} \pm 0.92$ min^{-1}, $b_2 = 0.289$ min$^{-1} \pm 0.022$ min^{-1} and $b_3 = 0.0125$ min$^{-1} \pm 0.0007$ min^{-1}). Amplitude coefficients A_1 and A_2 are then assumed to be equal and their value is supposed to be the ratio of the injected dose to the blood volume, which is estimated from the lean body weight (by humans, there is about 70 ml of blood per kilogram of lean body weight). The last amplitude coefficient A_3 is estimated from a late venous sample collected 55 min after injection, as the influence of the two first exponentials is negligible because of their much smaller characteristic times. MRGlu is also calculated before and after treatment of the tumour and results with SKM vary from 0.050 $\mu\mathrm{mol} \cdot \mathrm{min}^{-1} \cdot \mathrm{g}^{-1}$ to 0.35 $\mu\mathrm{mol} \cdot \mathrm{min}^{-1} \cdot \mathrm{g}^{-1}$. The coefficients of determination R^2 are respectively 0.97 and 0.94 for pre- and post-treatment, which indicates that the error between the reference MRGlu and the SKM-based MRGlu is small. Consequently SKM is an accurate method to estimate the MRGlu of tissues. It only requires a 15 min-long static PET-acquisition of the tissue and a single venous sample, provided that the input function can be modelled with a tri-exponential function.

The method of Hunter was then extended to better estimate MRGlu (Sundaram 04). It is called SKA-M and is also based on a single blood sample, the input function being obtained by scaling a population-based input function with the concentration of radiotracers measured in the blood sample. The main difference between SKM and SKA-M is that the SKA-M takes advantage of several PET-images of the tumours. MRGlu is estimated by fitting a regression line on Eq. (1.29)

$$\left(\frac{Tumour(t)}{A_{patient}(t)} \right) = K \left(\frac{\int_0^t A_{patient}(\tau) d\tau}{A_{patient}(t)} \right) + V_d \tag{1.29}$$

$Tumour(t)$ is estimated from dynamic PET-acquisitions of the tumour, $A_{patient}(t)$ is the concentration of radiotracers in blood at the times t where the dynamic images of the tumour were acquired, $A_{patient}(t)$ is derived from the scaled population-based input function. K and V_d are the parameters to be estimated through regression. The regression fits well with the data ($r > 0.99$, $P < 0.001$), the variance being lower than 4.6%. K correlates very well to MRGlu obtained with Patlak graphical analysis since the bias between the K and MRGlu is $1.0\% \pm 1.4\%$, whereas the bias is $15.1\% \pm 3.9\%$ when the method of Hunter is used instead. The improvement which is brought by SKA-M can also be seen in terms of the number of the estimations of MRGlu whose error is less than 20% compared to the MRGlu computed from Patlak analysis. Whereas the error on MRGlu estimations was above 20% for 10 out of 27 patients with Hunter's method (SKM), it decreases to 2 out of 27 with SKA-M. All these improvements can be explained by the fact that SKA-M takes into account dynamic information and consequently is able to model the unmetabolized ^{18}FDG, SKA-M being a measure of the rate of tracer uptakes and not a measure of tracer uptake. SKA-M is consequently a good method for estimating the glucose metabolic rate of tumours, since it is simpler than a full kinetics analysis (25 min-long dynamic PET acquisition without arterial blood sampling) but has a better accuracy than SKM.

All these methods have advantages and drawbacks as summarised in Table 1.2.5, but none is able to estimate non-invasively the arterial input function in the general case, that is to say for any radiotracers and any patients. Either estimations are invasive and must be corrected for dispersion, or they provide inaccurate input functions since macro-parameters, as MRGlu, calculated with these input functions have large variability, or they require that a hundred patients underwent arterial blood sampling before a standardised input function is available for a new radioligand. Furthermore, analytical or statistical-based input functions are not convenient for dealing with pathological patients, since the pathology can greatly impact the input function. Nevertheless, all studies underline that the estimation of arterial input functions can be non-invasive only if it is based on imaging techniques if it is expected to work on new radioligands (no standardised input function is available in such a case).

	non invasiveness	accuracy	temporal resolution	model complexity
arterial blood sampling	- -	++	+	++
Image-derived estimation from PET-images	++	+	-	+
Estimation from external detectors	+	+	+	+
Statistical model	-	- -	+	-
Hybrid methods	- -	+	++	-

Table 1.1 – Summary of methods for estimating arterial input functions: advantages and drawbacks

1.3 Some physics related to nuclear imaging

One of the alternatives that was presented in the previous section to estimate the arterial input function is to measure the concentration of tracers in an artery in the arm with an external detector (Kriplani 06). Moreover the determination of the TACs in tissues is based on nuclear imaging, either Single Photon Emission Computed Tomography (SPECT) or Positron Emission Tomography (PET). SPECT and PET are based on the use of pharmaceuticals labelled with radioelements. The way these imaging modalities work is developed in section 1.3.2. When it is intended to understand the way a tissue or a set of tissues work in the brain for instance, molecules are developed in order to bring to light specific mechanisms. They are labelled with radiotracers so that they emit high energy photons that can travel through tissues and consequently can be measured noninvasively with external detectors. It allows performing *in-vivo* investigations. It the following sections, some concepts on radioactivity will be presented as well as the way the photons are collected and transformed into a signal that can then be numerically processed.

1.3.1 Radioactive decay and interaction of radiation with matter

Some atoms are unstable and emit particles in order to lose energy and become more stable. This phenomenon is called radioactive decay and these atoms are said to be radioactive. Such atoms exist in nature but can also be produced artificially through nuclear reactions. The two main nuclear reactions in nuclear imaging are: β-decay where an electron e^- or positron e^+ is emitted from the nucleus and γ-decay where a photon is emitted. α-particle and β-particles travel a short range before interacting with matter, but γ-particles can travel several meters before interacting. Their range depends on their energy. Detectors are devices that have a high probability to detect these emitted particles when they are passing through them. It is possible to use atoms that are radioactive and emit γ-particles to label molecules and then estimate their quantity only from the amount of collected particles in detectors.

For the detection of high-energy photons, crystals are used because a high sensitivity is required. Photons are detected by interacting with electrons of the crystal of the detector, either through photoelectric effect or through Compton scattering. Crystal's properties are

presented in next section. If the energy of photons would be higher than 1022 keV, it could give rise to an electron and a positron through pair production but such high-energy photons are rarely used in medical imaging. The photoelectric effect is a physic phenomenon where a photon transfers all its energy to an electron. On the contrary, Compton scattering is the scatter of a photon on an electron, where only a fraction of the photon energy is transferred to the electron that recoils so that the energy conservation is respected. According to Klein-Nishina formula, the energy of the photon after a Compton scatter is given by:

$$E'_\gamma = \frac{E_\gamma}{1 + \frac{E_\gamma}{m_e c^2}(1 - \cos\theta)} \tag{1.30}$$

where θ is the new direction of the photon, m_e is the mass of an electron at rest and c is the speed of light. This means that a 511 keV photon can transfer at maximum 340 keV through Compton scattering, as shown in Eq. (1.32).

$$E_\Delta = E_\gamma - E'_\gamma = \frac{E_\gamma^2(1 - \cos\theta)}{m_e c^2 + E_\gamma \cdot (1 - \cos\theta)} \tag{1.31}$$

$$\max_\theta E_\Delta(511 \text{ keV}) = 340 \text{ keV} \tag{1.32}$$

511-keV photons are usually emitted from the annihilation of a positron with an electron, the positron being originated from the decay of a β^+-emitter. Principal β^+-emitters that are used for PET-imaging are listed in Table 1.2. The most important differences between these β^+-emitters are their half-life and their mean range in water, which is close to their range in tissues. For instance radiotracers labelled with $H_2^{15}O$ can only be used to observe fast physiological phenomena as cerebral blood flow, as the half-life of ^{15}O is about 2 min. On the contrary, ^{18}F can be used to understand slower phenomena as for example the uptake of ^{18}FDG in brain tissues.

Radio-isotope	Half-life	Mean-free path in water (Bailey 05)	Examples of radio-pharmaceuticals
^{11}C	20.38 min	1.1 mm	[^{11}C]raclopride
^{13}N	598 s	1.5 mm	[^{13}N] − ammonia
^{15}O	122 s	2.5 mm	[^{15}O]H_2O
^{18}F	109.7 min	0.6 mm	[^{18}F]FDG

Table 1.2 – Main β^+-emitters for medical imaging

1.3.2 Photon collection and image formation

When it is intended to detect high-energy photons as γ, a common technology is to use a crystal (Webb 03). Some of the photons that enter the crystal interact with electrons, either through photoelectric effect or Compton scattering. Crystals are scintillators that are adapted to detect high-energy photons. The atoms who receive the energy of impinging photons are in an excited state. When they return to their fundamental energy level, they emit visible photons. The number of visible photons emitted from a scintillation event is proportional to the energy of the photon that has excited atoms. Every scintillator is characterised by the wavelength spectrum of the visible photons produced through scintillations. At the back of the scintillator, a photocathode usually converts visible photons to electrons, called photoelectron. This electron is amplified through a photomultiplier and the resulting signal is collected to the anode of the photomultiplier which generates an electric pulse. A scintillator should have convenient properties:

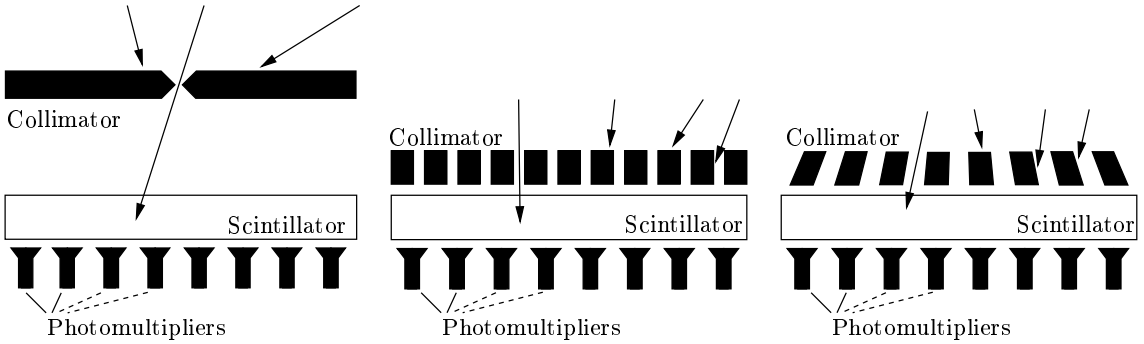
- a large attenuation coefficient so that the impinging photons do not pass through it without interacting. This means that scintillators should be dense and have a large number of protons Z since attenuation coefficients depend on the density ρ and Z^5 for high-energy photons (Leo 94).

- a large light output so that a large number of visible photons reach the photocathode, which improves the spatial and the energy resolution. Consequently, it should be transparent to visible photons so that scintillations can reach the photomultipliers.

In emission tomography, as long as the location of the photon is determined in the scintillator, it is known that the photon comes from a line that passes through this point as shown in Fig. 1.5 and 1.6. Several methods then allow knowing the direction where the photons come from.

1.3.2.1 SPECT

In SPECT, information is collected with γ-cameras, which include: a collimator, a detector and electronics. Detectors are either direct converters of high energy photons into electrons or consist in a scintillation crystal coupled to photomultipliers. The collimator is the most important part since it determines the spatial resolution and the statistics in images. Some collimators can have only one hole, in this case there are called pinhole collimators. An example is given in Fig. 1.5a. They can also have several holes that are parallel as in parallel-beam imaging or converge to a remote point or line as for cone-beam and fan-beam imaging. This is illustrated in Fig. 1.5b and 1.5c. Pinhole imaging allows the reconstruction of spatial distribution with an excellent spatial resolution, at the expense of the sensitivity that is poor since only the photons that pass through the pinhole can reach the detector and be used to estimate the object distribution. Pinhole collimators are not adapted to high energy photons such as those in PET-imaging because the edges of the pinhole are transparent at 511 keV. This degrades the spatial resolution. Parallel-collimators can be used instead of pinhole collimators, which improves sensitivity, but degrades spatial resolutions. Moreover, if it would be expected to collimate 511 keV photons, which is rare, these collimators would have to be thick and septas would have to be large so that only photons that have a direction normal to the detector surface would pass through the collimator. As a result, parallel-collimators would not be very sensitive at 511 keV. Instead of these collimators, cone-beam and fan-beam collimators can be used to detect 511 keV photons, as they focus on a point (respectively a line). This leads to very sensitive imaging systems which have a good spatial resolution around the region of interest, but the spatial resolution degrades rapidly in regions



(a) Pinhole collimator (b) Parallel-hole collimator (c) Cone-beam or fan-beam collimator

Figure 1.5 – Collimators for SPECT imaging

that are away from the focused region. These collimators are not adapted to reconstruct large volumes.

1.3.2.2 PET

In PET, the direction where the photons are coming is assessed without the need of physical collimators, although collimators can help discriminating random coincidences. After emission and a short range in matter, positrons annihilate with electrons into 2 511-keV γ-rays at 180 degrees. As a result, by detecting both, the line where the positron and the electron annihilate is known as it is illustrated in Fig. 1.6. This line is called line of response (LOR). By detecting a large number of LOR, the spatial distribution of the object is estimated. As the location of the emission of positrons and the location of their annihilation are different (characteristic distance is about 1 mm, see table 1.2 for more details), PET images provide a map of the annihilation locations, not of the radiotracers distribution. This can be viewed as an intrinsic spatial blurring, which limits the spatial resolution of PET.

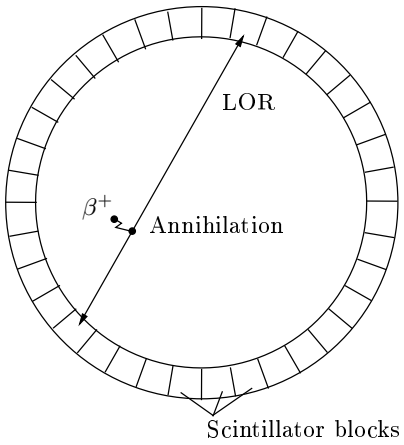


Figure 1.6 – Directions of photons in PET

1.4 Coded-Aperture Imaging and Medical Applications

In previous section (see 1.3.2.1), three collimator families were indicated for SPECT-imaging. The shape of the holes, their orientation as well as their distribution on the surface of the collimators impact the estimation of the radioactive distribution of the object. In this context, an original family of collimators appeared in late sixties, called coded-aperture collimators. Coded-aperture collimators were first developed for astrophysics applications (Dicke 68). The aim is to design imaging systems that have similar spatial resolution than pinhole-based imaging systems but with a much better sensitivity. It was then used for nuclear medical imaging (Ohyama 84; Groiselle 00; Schellingerhout 02). It is expected that coded-aperture collimators would sufficiently increase the sensitivity of an imaging SPECT system so that it will be possible to measure arterial input functions with such collimators.

1.4.1 Pinhole and coded-aperture imaging: physical and mathematical principles

This section presents the basis of pinhole imaging and how it has been extended to coded-aperture imaging.

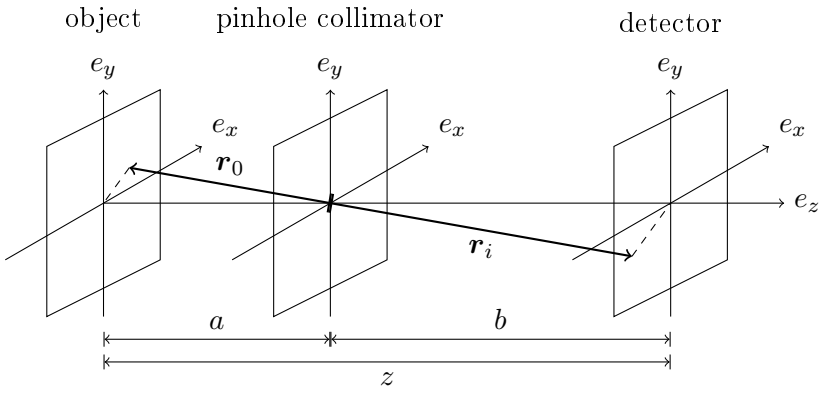


Figure 1.7 – Geometry of pinhole imaging

1.4.1.1 Pinhole imaging

Pinhole imaging denotes imaging with collimators consisting in a thin opaque plate with a single hole in it. When an object is placed in front of the collimator, only photons that travel through the hole can be detected by the detector on the other side of the collimator as illustrated in Fig. 1.7. With pinhole imaging, acquisitions on the detector are reversed and scaled copies of the spatial distribution of the object in the field-of-view (FOV). The position of a point that is located in the emitting object is denoted $\boldsymbol{r}_0$ whereas it is denoted $\boldsymbol{r}_i$ in the detector. There is a one-to-one correspondence between a point of the object and its projection on the detector through the pinhole collimator:

$$P(\boldsymbol{r}_i) \propto O(\boldsymbol{r}_0) = O\left(-\frac{a}{b}\boldsymbol{r}_i\right) \tag{1.33}$$

where O is the spatial representation of the object that emits photons and P is its projection onto the detector. The minus sign is due to the inversion of the image through the pinhole and the ratio $m = b/a$ is the magnification of the imaging system. If b is chosen larger than a, then the projection is an enlarged representation of the object. a and b can be chosen to achieve a magnification that compensates for the limited spatial resolution of the detector. The main drawback is the reduction of the FOV. For a given detector dimension d_{dtc}, the size of the FOV is:

$$FOV = d_{dtc} \cdot \frac{a}{b} \tag{1.34}$$

There is no reconstruction step, the projection being directly a representation of the object in the field-of-view. The previous considerations are applicable for perfect imaging systems, but in practice, pinholes have a finite dimension w_c. It directly impacts the spatial resolution R of the imaging system as illustrated in Fig. 1.8. Pinhole imaging can have very good spatial resolution, but this is achieved at the expense of the sensitivity. Indeed, the smaller the pinhole, the better the spatial resolution, but the lower the number of photons reaching the detector.

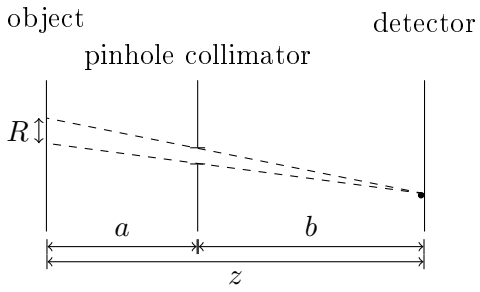


Figure 1.8 – Spatial resolution R is limited by the size of the pinhole

1.4.1.2 Coded-aperture imaging

Coded-aperture collimators are used in chapters 3 and 4 of this thesis to increase the statistics of image. The main idea of coded-aperture imaging is to use several pinholes in the collimator in order to improve the sensitivity. This leads to the acquisitions of several projections on the detector surface that usually overlap. The initial object is retrieved through image processing of these mixed projections. When only one point source, located in $\boldsymbol{r}_0$, is taken into account, the photon flux P on the detector is obtained through simple geometrical considerations:

$$P(\boldsymbol{r}_i) \propto A\left(\boldsymbol{r}_0 + \frac{a}{z}(\boldsymbol{r}_i - \boldsymbol{r}_0)\right) \tag{1.35}$$

$$\propto A\left(\frac{a}{z}\boldsymbol{r}_i + \frac{b}{z}\boldsymbol{r}_0\right) \tag{1.36}$$

where $\boldsymbol{r}_0$ represents the source position and $\boldsymbol{r}_i$ is a position on the detector. a is the distance between the collimator and the detector and z is the distance between the source and the detector. $A(r)$ is a binary function whose value is 1 if there is a hole in the mask at position $\boldsymbol{r}_i$, 0 otherwise. These notations are illustrated in Fig. 1.9. In coded-aperture imaging, some assumptions must be done so that the projections overlap does not deteriorate the estimation of the spatial distribution of the object. These assumptions are:

- Sources stand at infinity
- Collimator is infinitely thin and its septa are completely opaque.

According to the first assumption, all detected photons coming from a unique point source have the same incidence, as illustrated in Fig. 1.10. Then, the image of a point-source on the detector is not a point but the shifted pattern of the mask. In a sense, the coded

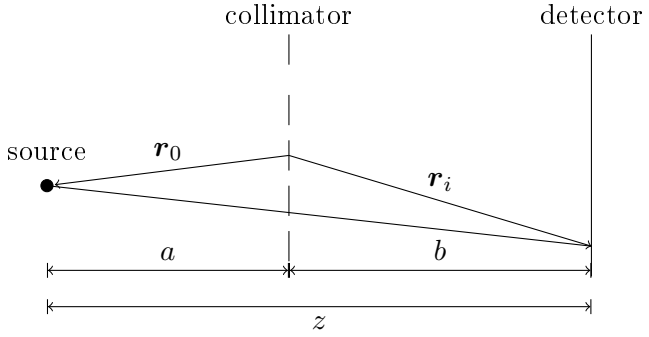


Figure 1.9 – Projection of a point-source on a detector through a multi-hole collimator

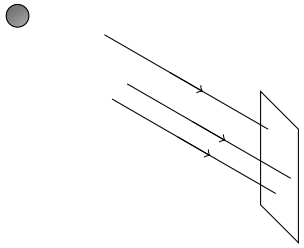


Figure 1.10 – Parallel incidence of photons impinging the collimator

aperture has encoded the position through the shift. As long as the hole size is larger than photon wavelengths, interference is negligible and contributions of several point sources can be viewed as the superposition of the independent contributions of every point source:

$$P(\boldsymbol{r}_i) \propto \int_{r_0} \mathrm{d}^2\boldsymbol{r}_0 \, O(\boldsymbol{r}_0) A\left(\boldsymbol{r}_0 + \frac{a}{z}(\boldsymbol{r}_i - \boldsymbol{r}_0)\right) \tag{1.37}$$

$$\propto \int_{\boldsymbol{r}'_0} \mathrm{d}^2\boldsymbol{r}'_0 \, O'(\boldsymbol{r}'_0) A'\left(\boldsymbol{r}_i - \boldsymbol{r}'_0\right) \tag{1.38}$$

$$P \propto O' \star A' \tag{1.39}$$

where $\star$ indicates the convolution operator and $\boldsymbol{r}'_0$, O' and A' are defined according to:

$$\boldsymbol{r}'_0 = -\frac{b}{a}\boldsymbol{r}_0 \tag{1.40}$$

$$O'(\boldsymbol{r}) = O\left(-\frac{a}{b}\boldsymbol{r}\right) \tag{1.41}$$

$$A'(\boldsymbol{r}) = A\left(\frac{a}{z}\boldsymbol{r}\right) \tag{1.42}$$

In practice, an additional term is added to model contributions of noise, as for instance Poisson fluctuations or electronic noise of the detector.

$$P \propto O \star A + N \tag{1.43}$$

When several sources are in the FOV, it is difficult to interpret directly the projections, because the superposition of the contributions through every pinhole prevents from recognising the mask pattern. However, as the original pattern of the coded aperture is known, it can be used to reconstruct the source spatial distribution.

Projection P must be 'decoded'. First papers (Mertz 61; Brown 72) that have investigated the possibility to take advantage of several pinholes showed artefacts in the estimated spatial resolution of objects because of wrong choices of the holes pattern. Then, Fenimore proposed patterns for coded-aperture imaging with perfect autocorrelation properties (Fenimore 78). The pattern of a collimator is said to have perfect imaging properties if the autocorrelation function of the mask function is a Dirac's delta function δ, as shown in Fig. 1.11, where the pattern of the coded-aperture collimator is denoted A and $\otimes$ is the periodic correlation operator. This means that the point spread function of the imaging system is a Dirac.

From a mathematical point of view, a mask has perfect imaging properties if:

$$A \otimes A = \delta + C \text{ , } C \text{ being a constant over all the image} \tag{1.44}$$

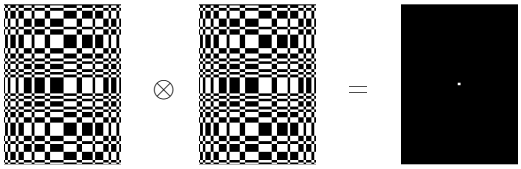
$A \otimes A = \sum_{k,l} A(i,j)A(i+k,j+l) \Leftrightarrow$  $\otimes$ $=$

Figure 1.11 – The autocorrelation function of coded-aperture pattern A is a Dirac-function

More generally, when a decoding matrix G that satisfies Eq. (1.45) exists, the imaging system whose collimator holes are defined by A is said to have perfect imaging properties.

$$A \otimes G = \delta + C' \text{ , C}' \text{ being a constant} \tag{1.45}$$

With such a decoding matrix, point source position O is estimated from Eq. (1.43) through:

$$\begin{aligned} \hat{O} &= P \otimes G && (1.46)\\ &= O \star A \otimes G + N \otimes G && (1.47)\\ &= O + N \otimes G + C' \text{ , C}' \text{ being a constant} && (1.48) \end{aligned}$$

Such a correlation procedure is known as 'match filtering' (Barrett 81). The interpretation is that the correlation tests all possible shifts and returns a spot when decoding mask matches the projection. Choice on G should minimise as much as possible the term $N \otimes G$. An example is given for a one-dimensional coded sequence in Fig. 1.12. In this figure, the projection P stands in the first row. It was periodically duplicated with 3 elements on left and 4 elements on the right for the clarity of the explanation of the decoding procedure. When the decoding matrix (7-element sequence) is in phase with the projection as in G_0-line, there are three pairs of holes (white boxes) that match. Otherwise, whatever the shift of the decoding matrix is, there is always one and only one pair of holes that matches. The position of such a pair is indicated with a coloured box in Fig. 1.12. From an arithmetic point of view, black boxes correspond to 0 and white boxes to 1. In the decoding procedure, the projection P and the decoding sequence G_i are multiplied term-by-term. The sum of the 7 multiplications is the value in the right column which is the decoded spatial distribution. In this example, the projection P corresponds to the situation where a unique point source is placed in the field of view of the collimator. This is what is observed in the decoded sequence, with an ideal point spread function. As the sidelobes are flat, it is simple to subtract the background. In return, it improves the signal-to-noise ratio (SNR). Such kinds of sequences also exist in 2D and coded sequences will be presented in section 1.4.2.

In practice, decoding is performed very efficiently through Fourier transforms:

$$\hat{O} \propto \mathcal{F}^{-1}\Big(\mathcal{F}(P)^H * \mathcal{F}(G)\Big) \tag{1.49}$$

where ^{H} denotes the conjugate operator and $*$ is the element-by-element matrix multiplication. $\mathcal{F}$ and $\mathcal{F}^{-1}$ are respectively the Fourier transform and its inverse.

1.4.1.3 Autocorrelation versus balanced decoding

When using autocorrelation for decoding projections, the spatial distribution of the sources is reconstructed over a non-zero constant background as shown in Fig. 1.12. For coded apertures whose autocorrelation function is a Dirac, it is possible to define a balanced decoding matrix G from the matrix A as it is explained in Eq. (1.50), where $(\lambda, N) = (1, 3)$. λ is the number of hole pairs which matches when the decoding sequence is not in phase with the

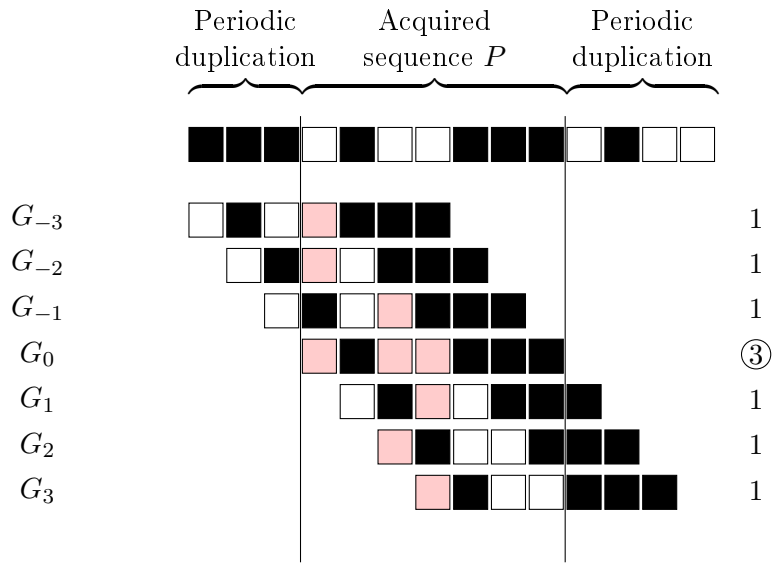


Figure 1.12 – Coded sequences and flat sidelobes

projection and N is the number of holes in the sequence. Using G instead of A as a decoding matrix leads to the direct reconstruction of the spatial distribution of the source over a null background.

$$G_i = \begin{cases} 1 & \text{if } A_i = 1 \\ -\dfrac{\lambda}{N-\lambda} (< 0) & \text{if } A_i = 0 \end{cases} \tag{1.50}$$

If applied to the example of Fig. 1.12, it leads to:

$$\mathcal{F}^{-1}\Big(\mathcal{F}([1\,0\,1\,1\,0\,0\,0])^H * \mathcal{F}([1,\text{-}1/2,1,1,\text{-}1/2,\text{-}1/2,\text{-}1/2])\Big) = [0,0,0,3,0,0,0] \tag{1.51}$$

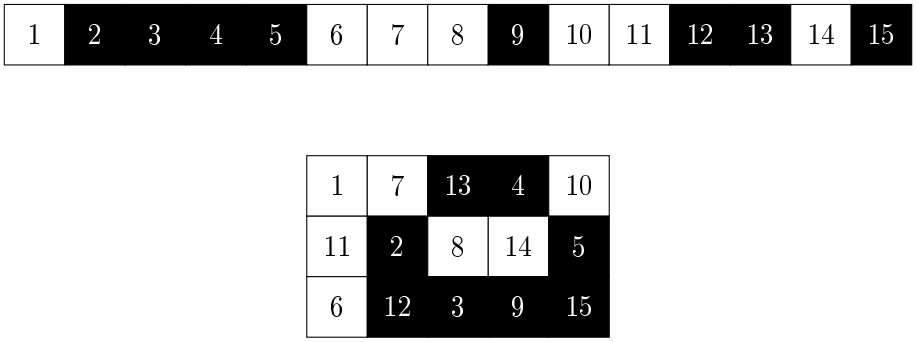
instead of

$$\mathcal{F}^{-1}\Big(\mathcal{F}([1,0,1,1,0,0,0])^H * \mathcal{F}([1,0,1,1,0,0,0])\Big) = [1,1,1,3,1,1,1] \tag{1.52}$$

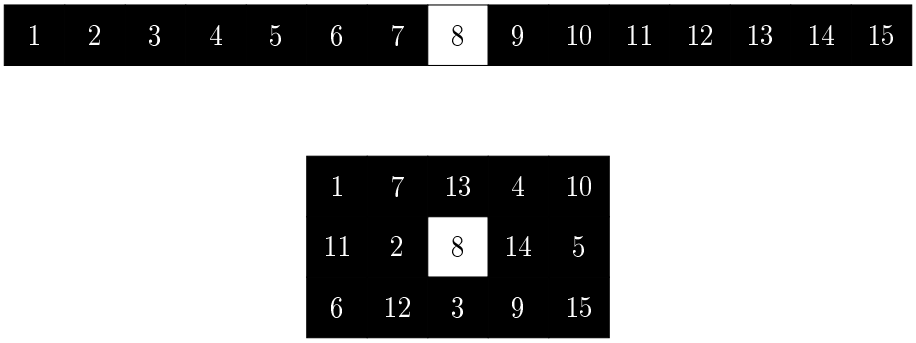
With this decoding method, background subtraction is avoided.

1.4.1.4 From 1D sequences to 2D arrays

2D arrays are usually created directly from 1D coded sequence. It introduces supplementary constraints, for instance the size of the sequence must not be prime if a 2D array of size $p \times q$ is expected. Every element of the sequence must indeed correspond to a unique element of the array. Furthermore, the 2D periodicity should also be taken into account as a periodic correlation is used in the decoding process. A way to fill a 2D array with 1D sequence without deteriorating the perfect imaging properties is to diagonally fill the 2D array as shown in Fig. 1.13a where all diagonals of the 2D pattern are incremented from one case to the other. The example in Fig. 1.13 shows that perfect imaging properties are maintained from 1D-sequence to 2D-array. Decoding is still achieved with Eq. (1.49), but $\mathcal{F}$ is now the 2D Fourier transform and $\mathcal{F}^{-1}$ its inverse. Balanced decoding works as for 1D-sequence, which results in a peak over no background constant.



(a) 1D sequence is diagonally folded in a 2D array



(b) The 1D-sequence and the 2D arrays have perfect imaging properties

Figure 1.13 – Perfect imaging properties are maintained from 1D sequence to 2D-array

1.4.1.5 Mask Mosaicking

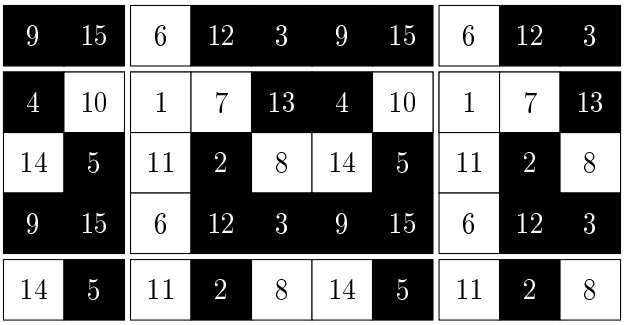
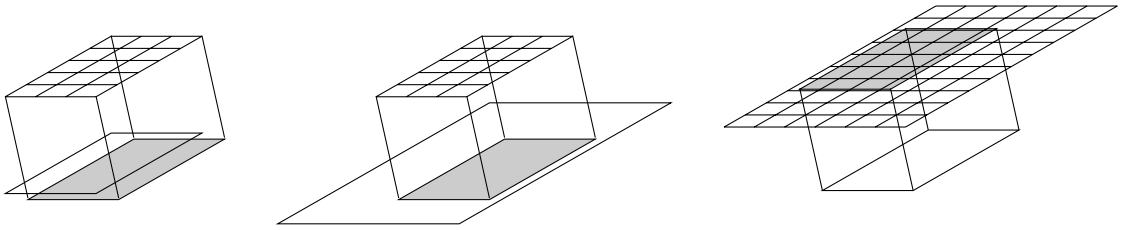


Figure 1.14 – The mask of Fig. 1.13a is mosaicked

Mosaicking the mask consists in replicating it through periodicity. For instance, Fig. 1.14 shows the mosaicked version of the mask of Fig.1.13a. Mosaicking the mask is intended to decrease the dimension of the detector. As it is shown in Fig. 1.15b, when the mask is not mosaicked, the detector must be larger than the mask so that the whole pattern of the mask (gray) is projected on the detector surface (white) when a source is out of axis. This is an expensive solution because it requires a large detector. It is preferred to mosaick the mask as in Fig. 1.15c so that the equivalent of the elementary pattern (gray) is acquired on the detector (white), although this is a periodically shifted projection. As decoding procedure is based on periodic correlation, the decoding result is similar to the case where the detector is larger than the mask.



(a) No mosaicking

(b) The detector is larger than the mask

(c) The mask is larger than the detector

Figure 1.15 – Mosaicking principle

1.4.2 Coded-aperture family

The previous section shows that some sequences are adapted to estimate source distributions in the FOV of an imaging system without artefacts and with a good sensitivity. Although such sequences are not very common, several methods are available to build them. Caroli wrote a complete review on coded apertures (Caroli 87). The point is to dispose of a large set of patterns that can be used with any geometrical and physical constraints of imaging systems. Main pattern parameters that influence the efficiency of imaging systems are:

- the open fraction ρ, which is the ratio of the open surface to the total surface. The larger the ratio, the larger the number of photons pass through the collimator.

- the dimension of the pattern, so that the projection of a spatial distribution through the collimator covers the whole surface of the detector, in order to maximise the efficiency.

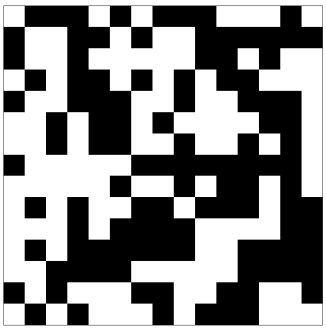
All the collimator patterns that are presented in this section are not mosaicked, for simplicity.

1.4.2.1 Random arrays

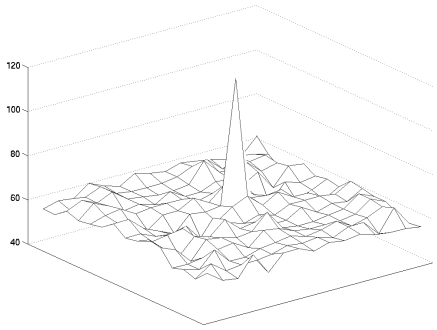
This kind of arrays does not present ideal imaging properties, they are just arrays where holes are randomly distributed on the elements. It is presented to illustrate the peculiar properties of coded apertures. The open fraction is chosen to be about $\rho = 50\%$, so that they can be compared to coded aperture array family that are presented in the next paragraphs. Their autocorrelation function is close to a Dirac's delta function, but is not a Dirac delta function, because of the random repartition of the holes of the mask elements. Any dimension and any open fractions are achievable. No constraint limits the realisation of such collimators, although one would usually prefer that the mask is self-supported, which means that all mask elements are connected altogether. Signal-over-noise ratio (SNR) dependence on open fraction ρ and on the number of mask-elements N_T was investigated (Accorsi 01) and it appears that SNR only depends on the number of holes according to:

$$SNR \propto \sqrt{N_T} \tag{1.53}$$

Two remarks are of interest on Eq. (1.53): first, the SNR does not depend on the open fraction ρ, second, a random mask with an infinite numbers of mask-elements would have ideal imaging properties. In practice, it is not feasible to have a very large number of very small mask elements. This explains why perfect imaging properties can not be achieved with random arrays.



(a) random mask pattern



(b) Autocorrelation function

Figure 1.16 – Example of a random array and its autocorrelation function. Sidelobes are not flat

1.4.2.2 Uniformly Redundant Arrays

Mathematical background is given by Johnsen (Johnsen 66): a framework for "cyclic difference sets" is presented, which was then used to generate first patterns that have perfect imaging properties. Coded apertures were obtained through the use of cyclic difference sets (Gunson 76; Fenimore 78). A cyclic difference set $\Delta(N_T, N, \lambda)$ is a set of N integers (α_i) that verifies:

$$\forall \rho \in [\![1, N_T - 1]\!], \quad card\left(\left\{(\alpha_i, \alpha_j) \in \Delta^2(N_T, N, \lambda) / \alpha_i - \alpha_j \equiv \rho \pmod{N_T}\right\}\right) = \lambda \tag{1.54}$$

N_T is the number of elements that constitute the mask, N is the number of holes in the mask and λ is the number of overlapping holes when decoding. Two main families are described in the literature:

- Singer sets, where main parameters N_T, N and λ are given by:

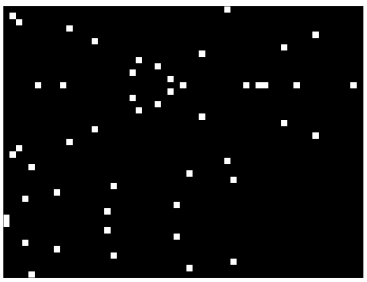
$$\{N_T, N, \lambda\} = \left\{\frac{t^{m+1}-1}{t-1}, \frac{t^m-1}{t-1}, \frac{t^{m-1}-1}{t-1}\right\}, \quad t \in \mathbb{N} \tag{1.55}$$

- Hadamard sets which satisfy:

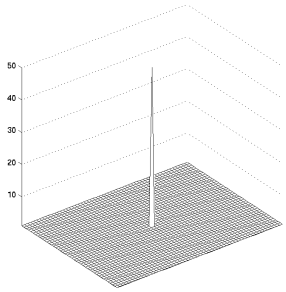
$$\{N_T, N, \lambda\} = \{4t-1, 2t-1, t-1\}, \quad t \in \mathbb{N} \tag{1.56}$$

An example of Singer arrays is presented in Fig. 1.17a. It was used for increasing the SNR in thyroidal imaging (Groiselle 00). The characteristic parameters are $\{t, m\} = \{49, 2\}$. It leads to an open fraction which is:

$$\rho \quad = \frac{N}{N_T} = \frac{49^2 - 1}{49^{2+1} - 1} \approx 2.0\% \tag{1.57}$$



(a) 2D Singer pattern, 50 holes out of 2541 elements ($\approx 2\%$)



(b) PSF after decoding procedure

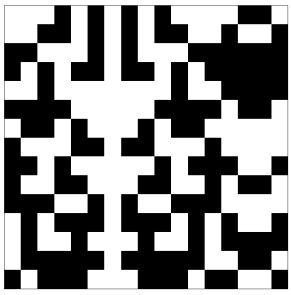
Figure 1.17 – 43×57 Singer 2D arrays

Among Hadamard sets, several methods allow the generation of coded apertures. Depending on the method, properties change as the dimension of the pattern or the lattice of elements. However all the patterns that belong to the Hadamard set have perfect imaging properties and their transparency N/N_T is about 50% as it can be derived from Eq. (1.56). Here are presented three methods for building coded-aperture collimators.

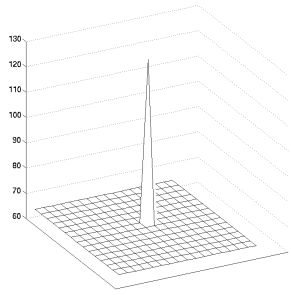
- Twin-prime arrays: the array that Fenimore used in his original paper on coded apertures for achieving perfect imaging belongs to this family. It is built on a $r \times s$ array with r and s being both prime and verifying $r - s = 2$. An example is given in Fig. 1.18, where with $r = 17$ and $s = 15$. The mask elements are defined according to:

$$A_{i,j} = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } j = 0, i \neq 0 \\ 1 & \text{if } c_r(i) \cdot c_s(j) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.58}$$

$$\text{where } \forall p, \quad c_p(i) = \begin{cases} 1 & \text{if } \exists x \in [\![1, p-1]\!] \,/\, i \equiv x^2 \pmod{p} \\ -1 & \text{otherwise} \end{cases} \tag{1.59}$$



(a) 15×17 URA pattern

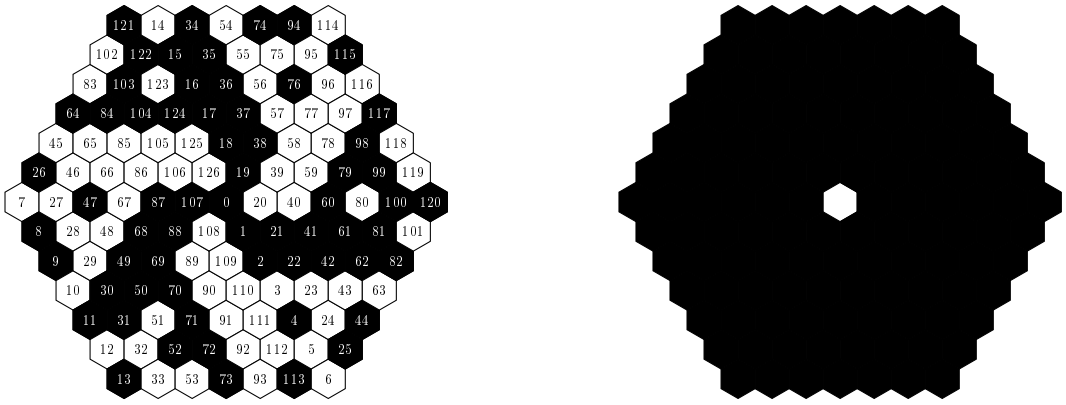


(b) Autocorrelation function

Figure 1.18 – Example of a uniformly redundant array and its autocorrelation function

- Quadratic residues: this corresponds to the case where N_T is prime. As a result, it can not be directly folded in a 2D rectangular array. However, such sequence can be folded on a hexagonal lattice, as in Fig. 1.19a. They are commonly called Hexagonal Uniformly Redundant Arrays, or HURA (Cook 84; Finger 85). They belong to a special class of URAs called the skew-Hadamard URAs. They are antisymmetry upon 60°. It presents the advantage that through a simple 60°-rotation, the mask turns into the anti-mask which allows background subtraction from two successive acquisitions (Gmar 04). It is defined from a few parameters called the order ν and the multiplier r. For HURA construction, the order ν and the multiplier r verify:

$$\begin{cases} \nu \text{ is prime and } \nu = 3 \text{ or } \nu = 12n + 7 \\ r^2 \equiv r - 1 \mod \nu \end{cases} \tag{1.60}$$



(a) Pattern on a hexagonal lattice (b) Decoded PSF

Figure 1.19 – Example of a Hexagonal Uniformly Redundant Array (HURA) (a) Coded sequence, obtained from quadratic residues, and folded on a 2D hexagonal lattice. (b) The autocorrelation function. Other HURA patterns are presented in Appendix 1

- m-sequences constitute the third sub-category of Hadamard arrays (Fenimore 83). It corresponds to the case where N_T can be written $N_T = 2^t - 1$, with $t > 1$. The number of open positions N is $2^{t-1} - 1$ which means that the open fraction is about 50% as the other URAs. The sequences are derived from polynomials of degrees m. A list is available in (Fenimore 83). Such m-sequences are shown in Fig. 1.12 and 1.13 for $m = 3$ and $m = 4$.

1.4.2.3 Pseudo-Noise Product arrays (PNP-arrays)

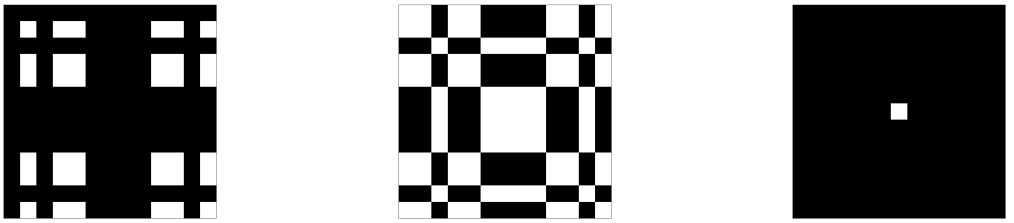
Pseudo-Noise Product arrays (PNP) (Gottesman 86) is a family of arrays that is more flexible than Uniformly Redundant Arrays since coded apertures can be produced with dimensions that were previously impossible. Furthermore, they are self-supported, which means that holes never isolate a portion of the collimator, which simplifies the mechanical construction of the mask. They are obtained through the direct product of two 1-dimensional coded-sequences. Indeed, it was proven (Luke 88) that the 2-dimensional arrays obtained from such a direct product has a decoding matrix G that verifies $A \otimes G = \delta$. If (a_i) and (b_j) are two 1D pseudo-noise sequences, then the coded pattern A and the decoding matrix G are

respectively given by Eq. (1.61) and (1.62).

$$A_{i,j} = a_i \cdot a_j \tag{1.61}$$

$$G_{i,j} = \begin{cases} -(-1)^{A_{i,j}} & \text{if } \exists k \text{ so that } A_{i,k} = 1 \\ (-1)^{A_{i,j}} & \text{otherwise} \end{cases} \tag{1.62}$$

Fig. 1.20 shows an example of such a mask, its decoding array and the result of the decoding procedure. Here the mask is square, which is very convenient as most detectors are also square.



(a) 13×13 PNP array (b) 13×13 PNP decoding array (c) 13×13 decoded projection

Figure 1.20 – PNP array

1.4.2.4 Modified Uniformly Redundant Arrays

This family was introduced in a study made by Gottesman (Gottesman 89). It is not based on cyclic difference sets nor pseudo-noise products, but still derives from quadratic residues. The main idea is that perfect imaging properties can be achieved although the autocorrelation function of the coded-aperture pattern A is not a delta function. The decoding matrix G has to be appropriately designed so that the periodic correlation $A \otimes G$ is a delta function. This new family of arrays increases by a factor of 3 the number of available coded apertures. Useful sequences can be generated for any length L that is prime and that verifies:

$$\exists m \in \mathbb{N}, \text{ so that } L = 4m + 1 \tag{1.63}$$

Sequences are given by:

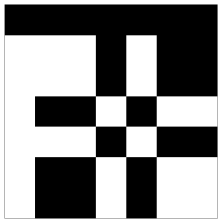
$$A_i = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } \exists m \text{ so that } i \equiv m^2 \pmod{L} \\ 0 & \text{otherwise} \end{cases} \tag{1.64}$$

These can be used to generate either square arrays or hexagonal arrays. Square arrays are generated according to Eq. (1.65). Two examples illustrate such kind of arrays in Fig. 1.21.

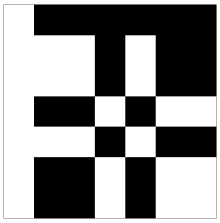
$$A_{i,j} = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } j = 0, i \neq 0 \\ 1 & \text{if } c(i) \cdot c(j) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.65}$$

where

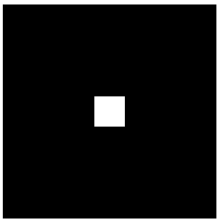
$$c(i) = \begin{cases} 1 & \text{if } \exists x \in [\![1, p-1]\!] \,/\, i \equiv x^2 \pmod{p} \\ -1 & \text{otherwise} \end{cases} \tag{1.66}$$



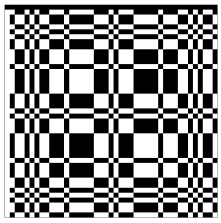
(a) 7×7 MURA array



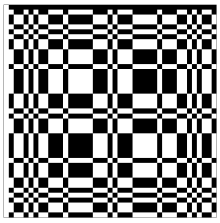
(b) 7×7 MURA decoding array



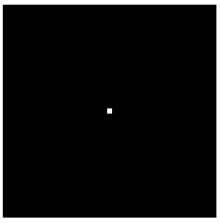
(c) 7×7 decoded projection



(d) 43×43 MURA array



(e) 43×43 MURA decoding array



(f) 43×43 MURA decoded projection

Figure 1.21 – MURA patterns for 7×7 array and 43×43 array, their associated decoding matrix and the resulting reconstructed PSF

The decoding matrix G is obtained from A according to:

$$G_{i,j} = \begin{cases} 1 & \text{if } i+j=0 \\ 1 & \text{if } A_{i,j}=1 \\ -1 & \text{if } A_{i,j}=0 \end{cases} \tag{1.67}$$

This family is very close to quadratic residue arrays because of condition in Eq. (1.66). Its coded sequences have a throughput of $\frac{L-1}{2L}$ which converges rapidly to 50% as the sequence length increases.

All the properties of these families are summed up in Table 1.3. There are some extra coded-aperture families that are rarely used, so no further description in this thesis is provided to the reader. It consists in geometric masks (Gourlay 83), dilute URAs (Wild 83).

	number of holes N_T	ρ =holes/elements	lattice
Singer sets	$\frac{t^{m+1}-1}{t-1}$	${}^{N}/_{N_T} \approx 1/t$	□
URA			
- Twin prime	$N_T = r \cdot s$ r and s are twin prime $r - s = 2$	$\frac{2^m-1}{2^{m+1}-1} \approx 50\%$	□
- Quadratic residues	N_T is prime	$\frac{2^m-1}{2^{m+1}-1} \approx 50\%$	⬡
(HURA)	$N_T = 3$ or $12m+7$		
- m-sequences	$N_T = 2^m - 1, t > 1$	$\frac{2^m-1}{2^{m+1}-1} \approx 50\%$	□
PNP			□
MURA	N_T is prime $N_T = 4m+1$	$\frac{N_T-1}{2N_T} \approx 50\%$	□⬡

Table 1.3 – Summary of the properties of every coded-aperture family

1.4.3 Medical imaging applications and near-field domain

Medical imaging held attracted by the performance of the coded-aperture collimators (Barrett 72; Rogers 80; Ohyama 84) since its early development in astrophysics. Indeed, the increase of SNR would allow low-signal imaging, and consequently low-dose imaging. As a larger choice of patterns that have perfect imaging properties were available, their performances for medical imaging were investigating (Meikle 01; Schellingerhout 02). Some assumptions are not valid anymore, such as the fact that sources stand at infinity away from the collimator. Some works have been done in order to analyse the artefacts that appear because the far-field approximation does not hold anymore. Eq. (1.37) has to be modified into Eq. (1.68) in order to take into account the geometric efficiency (Accorsi 01).

$$P(\boldsymbol{r}_i) \propto \int_{r_0} \mathrm{d}^2\boldsymbol{r}_0 \, \cos^3(\theta) O(\boldsymbol{r}_0) A\left(\boldsymbol{r}_0 + \frac{a}{z}(\boldsymbol{r}_i - \boldsymbol{r}_0)\right) \tag{1.68}$$

with θ being a function of $\boldsymbol{r}_0$ and $\boldsymbol{r}_i$:

$$\theta = \arctan\left(\frac{\|\boldsymbol{r}_i - \boldsymbol{r}_0\|}{z}\right) \tag{1.69}$$

where θ is the incidence angle of the ray and the detector. It is not possible to estimate the distribution $O(\boldsymbol{r}_0)$ as with the far-field approximation because of the term $\cos^3\theta$. In the far-field approximation, $\cos^3\theta = 1$. Accorsi proposed to expand the near-field term $\cos^3\theta$ through Taylor expansion. This approximation holds better when $\boldsymbol{r}_0 \ll z$, which corresponds to imaging systems that have a small magnification ratio.

Zero-order and first-order artefacts are not presented but details on are available in Appendice 2. It demonstrates that the mask pattern should balanced and the object spatial distribution should centred in the middle of the field-of-view so that artefacts are at their minimum.

In this section, it was shown that coded-aperture collimators have a larger open fraction than pinhole collimators. A large choice of patterns allows choosing a coded mask that fits

to detector dimensions. Although they were not primarily developed for medical imaging applications, it is possible to perform high-sensitivity high-resolution imaging by respecting a few rules in order to limit apparitions of artefacts. Next section will introduce reconstruction techniques used in emission tomography, since one can imagine that coded-aperture collimators could be decoded by other methods than correlations.

1.5 Statistical reconstruction techniques

In this section, statistical reconstruction algorithms are presented. It explains how they work, their advantages but also their drawbacks. Analytical reconstruction algorithms, as for instance filtered-backprojection, are not presented because it will not be possible to apply them in the rest of the thesis, because they are not adapted to address low-signal problems. In this thesis, it is intended at using such statistical reconstruction techniques on SPECT-scans acquired with coded-aperture collimators in order to dynamically quantify the concentration of tracers that flows inside arteries. Results of chapter 3 take advantages of such algorithms.

1.5.1 Statistical reconstruction through MLEM

First statistical reconstructions of a single acquisition are explained. They are based on an algorithm called "maximum-likelihood expectation-maximisation" (MLEM). Its application in medical imaging is due to Shepp & Vardi (Shepp 82). Statistical reconstruction techniques take into account the statistical distribution of the acquired datasets that is Poisson-distributed. Acquired observations are denoted $\boldsymbol{y}$ and the object is denoted $\boldsymbol{\lambda}$. $\boldsymbol{\lambda}$ represents the unobserved parameters to be estimated. The main idea is to maximise the likelihood $L(\boldsymbol{y}|\boldsymbol{\lambda})$ to have $\boldsymbol{y}$ given $\boldsymbol{\lambda}$ with respect to $\boldsymbol{\lambda}$. The maximisation is not trivial. It is usually performed through EM-algorithm: it employs a complete data set $\boldsymbol{z}$ (Dempster 77) which can not be observed but is related to observed data $\boldsymbol{y}$, called the "incomplete" data set. On one hand, the "complete" data set $\boldsymbol{z}$ is the set of observations that corresponds to the independent acquisitions of the contributions of every voxel, which is a N-element set if the object is made of N voxels. On the other hand, the "complete" data set $\boldsymbol{y}$ corresponds to the real acquisition with contributions from all voxels. EM-algorithm is an iterative algorithm and every iteration is made of two steps: expectation and maximisation.

1.5.1.1 Expectation step

The expectation step, also called E-step, derives the following conditional expectation:

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n) = E\left[\ln\Big(L(\boldsymbol{z}|\boldsymbol{\lambda}) \Big) \middle| \boldsymbol{y}, \boldsymbol{\lambda}^n \right] \tag{1.70}$$

where $L(\boldsymbol{z}|\boldsymbol{\lambda})$ denotes the likelihood on the complete data set $\boldsymbol{z}$ for a given vector $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^n$ is the current estimate at iteration n. In emission tomography, the likelihood of the "complete" data set and its logarithm are:

$$L\left(\boldsymbol{z}|\boldsymbol{\lambda}\right) = \prod_i^M \prod_j^N \frac{\bar{z}_{i,j}^{z_{i,j}} e^{-\bar{z}_{i,j}}}{z_{i,j}!} \tag{1.71}$$

$$\ln\left(\boldsymbol{z}|\boldsymbol{\lambda}\right) = \sum_i^M \sum_j^N z_{i,j} \ln \bar{z}_{i,j} - \bar{z}_{i,j} + C \tag{1.72}$$

with:

- $z_{i,j}$ being the number of photons emitted from voxel j and detected in pixel i. In particular

$$\forall i, \quad y_i = \sum_j z_{i,j} \tag{1.73}$$

- $\bar{z}_{i,j}$ is the conditional expectation $E[z_{i,j}|\boldsymbol{\lambda}]$: $\bar{z}_{i,j} = E[n_{i,j}|\boldsymbol{\lambda}] = a_{i,j} \cdot \lambda_j$
- $a_{i,j}$ is the probability that a photon emitted from voxel j would be detected in pixel i. $(a_{i,j})$ constitutes a matrix $\boldsymbol{A}$ which is commonly called system matrix or transition matrix.
- $C = \sum_{ij} \ln z_{ij}!$ does not depend on $\boldsymbol{\lambda}$

The result of the E-step for the $(n+1)^{th}$ iteration is then:

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n) = E\left[\ln\Big(\boldsymbol{z}|\boldsymbol{\lambda}\Big) \,\middle|\, \boldsymbol{y}, \boldsymbol{\lambda}^n \right] \tag{1.74}$$

$$= \sum_i^M \sum_j^N E[z_{i,j}|(y_i), \boldsymbol{\lambda}^n] \ln \bar{z}_{i,j} - \bar{z}_{i,j} + C \tag{1.75}$$

With the assumption that $(n_{i,j})$ are independent, it becomes:

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n) = \sum_i^M \sum_j^N \left(\frac{a_{ij}\lambda_j^n}{\sum_k a_{i,k}\lambda_k^n} \cdot y_i \right) \ln a_{i,j}\lambda_j - a_{i,j}\lambda_j + C \tag{1.76}$$

1.5.1.2 Maximisation step

The maximisation step, also called M-step, maximises the previous conditional expectation $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n)$ with respect to $\boldsymbol{\lambda}$. $\boldsymbol{\lambda}^n$ is the best estimate at iteration n and consequently is constant. Setting first derivative of $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n)$ to zero leads to:

$$\forall j, \quad \frac{\partial Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n)}{\partial \lambda_j} = 0 \quad \Leftrightarrow \quad \sum_i^N \left(\frac{a_{i,j}\lambda_j^n}{\sum_k a_{i,k}\lambda_k^n} \cdot y_i \right) \frac{a_{i,j}}{a_{i,j}\lambda_j} - a_{i,j} = 0 \tag{1.77}$$

$$\Leftrightarrow \quad \lambda_j^{n+1} = \frac{\lambda_j^n}{\sum_i a_{i,j}} \sum_i y_i \frac{a_{i,j}}{\sum_k a_{i,k}\lambda_k^n} \tag{1.78}$$

This optimisation is concave since $\boldsymbol{\lambda} \mapsto Q(\boldsymbol{\lambda})$ is semi-definite negative. Indeed, the second derivative of $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n)$ only has negative coefficients. At every iteration of this algorithm, λ^{n+1} verifies $\sum_{i,j} a_{i,j}\lambda_j^{n+1} = \sum_i y_i$ as demonstrated below:

$$\sum_{i,j} a_{i,j}\lambda_j^{n+1} = \sum_{i,j} a_{i,j} \frac{\lambda_j^n}{\sum_l a_{l,j}} \sum_l y_l \frac{a_{l,j}}{\sum_k a_{l,k}\lambda_k^n} \tag{1.79}$$

$$= \sum_{j,l} \sum_i \left(\frac{\sum_i a_{i,j}}{\sum_l a_{l,j}} \right) y_l \frac{\sum_j a_{l,j}\lambda_j^n}{\sum_k a_{l,k}\lambda_k^n} \tag{1.80}$$

$$\sum_{i,j} a_{i,j}\lambda_j^{n+1} = \sum_i y_i \tag{1.81}$$

In particular, when the initialisation of the algorithm is made from a uniform image, the activity λ_0 in every voxel is:

$$\lambda_0 = \frac{\sum_i y_i}{\sum_{i,j} a_{i,j}} \tag{1.82}$$

This algorithm allows the estimation of the spatial activity distribution. However, beyond a certain point, reconstructed images start deteriorating, essentially because the most likely solution is not the real activity distribution. This is why it is recommended to early stop this algorithm.

This suggests that *a priori* information should be added through regularisation so that no deterioration appears when the number of iterations increases.

1.5.1.3 Regularisation

Regularisation can be used to improve the robustness of reconstructions. Reconstructions are inverse problems that are ill-posed (Idier 01). As a result, a way to stabilise numerical solutions is to regularise the cost function. The main idea with regularisation is to add *a priori* information (Green 90). The *a priori* information $P(\boldsymbol{\lambda})$ is taken into account through Bayes formula:

$$P(\boldsymbol{\lambda}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\lambda}) \cdot P(\boldsymbol{\lambda})}{P(\boldsymbol{y})} \tag{1.83}$$

Logarithm-function is applied to this equation:

$$\ln\left(P(\boldsymbol{\lambda}|\boldsymbol{y})\right) = \ln\left(P(\boldsymbol{y}|\boldsymbol{\lambda})\right) + \ln\left(P(\boldsymbol{\lambda})\right) - \ln\left(P(\boldsymbol{y})\right) \tag{1.84}$$

The *a priori* information $P(\boldsymbol{\lambda})$ is usually defined through a potential function $\boldsymbol{\lambda} \mapsto V(\boldsymbol{\lambda})$ according to:

$$P(\boldsymbol{\lambda}) = \exp\left(-\beta \cdot V(\boldsymbol{\lambda})\right) \tag{1.85}$$

$$= \exp\left(-\beta \sum_s \sum_{r \in \mathcal{N}_s} w_{s,r} \cdot \varphi\left(\frac{\lambda_s - \lambda_r}{\delta}\right)\right) \tag{1.86}$$

where the weighting term β determines the strength of the prior and δ is a range parameter. $\mathcal{N}_s$ is the neighbourhood of pixel s and $w_{s,r}$ is a weight that indicates the neighbourhood relation between pixels r and s. φ is a prior function, some examples will be given below. In his work, Green proposed an algorithm called "one-step late" (OSL) which is derived from MLEM that maximises the maximum *a-posteriori* activity distribution $P(\boldsymbol{\lambda}|\boldsymbol{y})$ of Eq. (1.83). The update scheme for iteration $(n+1)$ is:

$$\lambda_j^{n+1} = \frac{\sum_i z_{ij}}{\sum_i a_{ij} + \beta \frac{\partial}{\partial \lambda_j} V(\boldsymbol{\lambda})_{|\boldsymbol{\lambda}=\boldsymbol{\lambda}^n}} \tag{1.87}$$

with z_{ij} being estimated according to:

$$z_{i,j} = \lambda_j^n a_{i,j} \frac{y_i}{\sum_{j'} a_{i,j'} \lambda_{j'}^n} \tag{1.88}$$

(z_{ij}) represents the "complete" data set. Numerous potential functions were proposed and every of them impacts solutions differently. The global idea is to smooth region which should have the same activity without smoothing edges where intensity variations have a real physical justification. Charbonnier reviews some of these priors (Charbonnier 97). *A priori* functions can be classified into two main categories: the potential functions that are semi-definite positive, which insures that the solution that maximises Eq. (1.83) exists and is unique. Some examples are the following potential functions and their first derivatives:

$$\varphi(t) = \ln(1+t^2) \qquad \varphi'(t) = \frac{2t}{1+t^2} \tag{1.89}$$

$$\varphi(t) = 2\sqrt{1+t^2} - 2 \qquad \varphi'(t) = \frac{2t}{\sqrt{1+t^2}} \tag{1.90}$$

$$\varphi(t) = 2\ln(\cosh(t)) \qquad \varphi'(t) = 2\tanh(t) \tag{1.91}$$

The second category is made of functions that are asymptotically convergent. Examples are:

$$\varphi(t) = \frac{t^2}{1+t^2} \qquad \varphi'(t) = \frac{2t}{(1+t^2)^2} \tag{1.92}$$

$$\varphi(t) = \frac{t^2\delta^n}{2(\sqrt{t^2/2+\delta^2/2})^n} \qquad \varphi'(t) = \frac{t\delta^n(t^2(1-n/2)+\delta^2)}{(t^2+\delta^2)^{n/2+1}} \tag{1.93}$$

MAP converges faster than MLEM and the result is equivalent to post-smoothed MLEM when the resolution is uniform in the whole image. MAP requires less iterations than MLEM but every iteration requires a larger computational load.

Application of MLEM algorithm to emission tomography results in accurate estimation of the spatial distribution of the object to be imaged. It allows the addition of *a priori* information. However it is slow because it requires a lot of iterations to converge to solutions. Furthermore, it requires that an accurate system matrix $\boldsymbol{A}$ was computed, which can be difficult because such a matrix should take into account several physical phenomena, as scattering and spatial blurring for instance.

1.5.2 Dynamic reconstruction techniques

Demand on dynamic information is increasing because it is more informative on physiological *in-vivo* mechanisms. 2D or 3D images are not sufficient to deeply understand the way organs work and interact with radiolabelled molecules; the time sequence must be taken into account. When independent 3D reconstruction methods are used, the better the time resolution is, the worse the reconstruction quality per frame is. The system matrix is denoted $\boldsymbol{A}_{M,N}$, the expected image is denoted $\boldsymbol{y^{exp}}$. The number of time frames is denoted T. Every time frame t can be reconstructed according to the model:

$$(\boldsymbol{y^{exp}})^t = \begin{bmatrix} (y^{exp})_1^t \\ \vdots \\ (y^{exp})_M^t \end{bmatrix} = \boldsymbol{A}_{M,N} \cdot \begin{bmatrix} \lambda_1^t \\ \vdots \\ \lambda_N^t \end{bmatrix} = \boldsymbol{A}_{M,N} \cdot \boldsymbol{\lambda}^t \tag{1.94}$$

Independent reconstructions of several sets of projections as in Eq. 1.94 can be modelled through a unique equation:

$$(\boldsymbol{y^{exp}})^{tot} = \begin{bmatrix} (\boldsymbol{y^{exp}})^1 \\ \vdots \\ \vdots \\ (\boldsymbol{y^{exp}})^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{M,N} & \boldsymbol{0}_{M,N} & \dots & \boldsymbol{0}_{M,N} \\ \boldsymbol{0}_{M,N} & \boldsymbol{A}_{M,N} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}_{M,N} \\ \boldsymbol{0}_{M,N} & \dots & \boldsymbol{0}_{M,N} & \boldsymbol{A}_{M,N} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\lambda}^1 \\ \vdots \\ \boldsymbol{\lambda}^T \end{bmatrix} = \boldsymbol{H} \cdot \boldsymbol{\lambda}^{tot} \tag{1.95}$$

$\boldsymbol{H}$ is a system matrix which takes into account several time frames. It depends on $\boldsymbol{A}_{M,N}$ which is the system matrix for the static model representation. The unknown parameters (λ_j^t) are the intensities in every voxel j for every time frame t. The number of unknown parameters increases linearly with the number of time frames, which is not convenient from a computational point of view. Furthermore, the better the time resolution is, the less counts every projection contains which leads to poor quality of reconstructions because of statistical considerations. As a result, it is usually not possible to properly reconstruct independent time frames. Some methods were proposed to deal with such a problem. They consist in modelling the spatiotemporal distribution with less parameters than the total number of parameters $\boldsymbol{\lambda}^{tot}$ is made of.

1.5.2.1 Model of the observations through inhomogeneous Poisson process

It was suggested to model dynamic data through inhomogeneous Poisson process in PET-imaging (Nichols 02), which means that the radioactive concentration (λ_i) for every voxel i in the object is modelled as a time-varying continuous function. Such a function is built from a linear combination of cubic B-splines. Thus this method fully exploits the list-mode by taking into account the exact detection time of every detected coincidence. This method is based on the maximisation of the likelihood of arrival times of detected photons pairs and does not rely on any physiological models. Inhomogeneous Poisson process is used to model radioactive concentrations for all voxel j as:

$$\lambda_j(t) = \sum_l w_{jl} B_l(t) \tag{1.96}$$

where $B_l(.)$ is the l^{th} spline basis function, $(w_{j,l})$ is the weight (also called "control vertex") of the l^{th} B-spline. Instead of estimating the $(N \times T)$ unknown parameters of $\boldsymbol{\lambda}^{tot}$, the $(N \times C)$ parameters $(w_{j,l})$ are estimated. This presents an interest only if $C < T$. This model assumes that the kinetics in all voxels is a linear combination of limited number C of functions B_l. From Eq. (1.96), the model for the measured data is deduced:

$$\eta_i(t) = \sum_j a_{i,j} \sum_l w_{jl} B_l(t) \tag{1.97}$$

where $a_{i,j}$ is the probability that an annihilation which happens in voxel j is detected in pair i and $\eta_i(t)$ is the expectation that an event would be detected in the detector pair i at time t. All parameters (w_{jk}) can be represented in a matrix $\boldsymbol{W}$. When N events are detected at time $(t_1, \ldots, t_N)$ during an acquisition which starts at T_i and ends at T_f, the likelihood function of a Poisson process, whose rate function is given by $\eta(t)$, is:

$$P(t_1, \ldots, t_N | \eta(t)) = \left(\prod_{k=1}^{N} \eta(t_k) \right) \cdot \exp\left(- \int_{T_i}^{T_f} \mathrm{d}u \; \eta(u) \right) \tag{1.98}$$

When using list-mode data, detection time is accessible for every detected coincidence. Measured data are denoted $\boldsymbol{\mathcal{D}} = \{\mathbf{y}, \mathbf{t}_1, \ldots, \mathbf{t}_M\}$ with $\mathbf{y}$ being the sinogram and $\mathbf{t}_i = \{t_{i,1}, \ldots, t_{i,n_i}\}$ being the time stamps of the n_i events detected in the detector pair i. As a result, the log-likelihood of obtaining $\boldsymbol{\mathcal{D}}$ is:

$$L(\boldsymbol{\mathcal{D}} | (w_{j,l})) = \sum_i \sum_k \ln \eta_i(t_{i,k}) - \sum_i \int_{T_i}^{T_f} \mathrm{d}u \; \eta_i(u) \tag{1.99}$$

Some regularisations are added to the log-likelihood $L(\mathcal{D}|\boldsymbol{W})$ in order to constrain smoothness on spatial and temporal variations ($\rho(\boldsymbol{W})$ and $\phi(\boldsymbol{W})$), as well as a negativity penalty $\nu(\boldsymbol{W})$. The objective function E is defined according to Eq. 1.100 to take them into account:

$$E = L(\mathcal{D}|\boldsymbol{W}) - \alpha\rho(\boldsymbol{W}) - \beta\phi(\boldsymbol{W}) - \gamma\nu(\boldsymbol{W}) \tag{1.100}$$

A difficult task is to maximise such a function because there is no reason it is concave. The maximisation is performed through a conjugate gradient method. After maximising the cost function E on $\boldsymbol{W}$, the spatiotemporal distribution is obtained through Eq. (1.96). This method assumes that kinetics of voxels can be decomposed on B-splines which is not necessarily the optimal decomposition.

1.5.2.2 Model of the observations without *a priori* on temporal functions

It was seen in the previous section that the spatiotemporal distribution $\boldsymbol{\lambda}^{tot}$ could be modelled from only a few parameters $\boldsymbol{W}$. In this section, the method that is presented investigates further this model, but without assuming that the kinetics of voxels should be decomposed on a particular type of functions. In this investigation, TACs of every voxel are not modelled from a predefined temporal basis functions, but rather through a linear combination of C temporal basis functions that will be estimated in the reconstructions, as well as their factors (Reader 06). The activity in voxel j at time t can be written:

$\forall(j \times t) \in ([\![1; N]\!] \times [\![1; C]\!])$,

$$\lambda_j(t) = \sum_{k=1}^{C} c_j^k \cdot f^k(t) = \sum_{k=1}^{C} c_j^k \cdot f_t^k \tag{1.101}$$

where (c_j^k) are weighting coefficients and $(f^k)_{k\in[\![1;C]\!]}$ is a set of C temporal functions. It is then expected to estimate the weighted coefficients (c_j^k) and the temporal functions (f^k) that lead to the most likely solution $\boldsymbol{\lambda}^{tot}$ in Eq. (1.95). Every temporal function f^k can be represented as a T-long vector. Every c^k can be represented as a N-long vector and has the dimension of a 3D-image. The number C of temporal basis functions is not estimated, it is an input parameter of the algorithm. All parameters c_j^k and all parameters f_t^k are stored respectively in vectors $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ that are defined as:

$$\boldsymbol{\theta} = [c_1^1 \dots c_N^1, \dots, c_1^C \dots c_N^C]^{\mathrm{T}} \tag{1.102}$$

$$\boldsymbol{\gamma} = [f_1^1 \dots f_T^1, \dots, f_1^C \dots f_T^C]^{\mathrm{T}} \tag{1.103}$$

where $^{\mathrm{T}}$ is the transpose operator. $\boldsymbol{\theta}$ is a $(N \times C)$-long vector and $\boldsymbol{\gamma}$ is a $(C \times T)$-long vector. It is proposed in (Reader 06) to jointly estimate temporal basis functions $\boldsymbol{\gamma}$ and their 3D weighting coefficients $\boldsymbol{\theta}$ through alternate MLEM optimisations. One set of unknown parameters (for instance the weighting coefficients $\boldsymbol{\theta}$) is kept constant when the other (the temporal functions $\boldsymbol{\gamma}$) is optimised and *vice versa*. This means that Eq. (1.95) must be put in the form of:

$$(\boldsymbol{y}^{\boldsymbol{exp}})^{tot} = \boldsymbol{H}^1 \cdot \boldsymbol{\theta} \tag{1.104}$$

$$(\boldsymbol{y}^{\boldsymbol{exp}})^{tot} = \boldsymbol{H}^2 \cdot \boldsymbol{\gamma} \tag{1.105}$$

so that MLEM-algorithm can be applied on $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$. It is clear that $\boldsymbol{H}^1$ depends on $\boldsymbol{H}$ and $\boldsymbol{\gamma}$ and that $\boldsymbol{H}^2$ depends on $\boldsymbol{H}$ and $\boldsymbol{\theta}$. In the following calculations, the set (f^k) is assumed to be made of 3 components ($C = 3$) for clarity. By denoting $\boldsymbol{\lambda}^{tot} = [\lambda_1(t_1) \dots \lambda_N(t_1), \dots, \lambda_1(t_T) \dots \lambda_N(t_T)]^T$, $\boldsymbol{\lambda}^{tot}$ can be written with matricial notations:

$$\boldsymbol{\lambda}^{tot} = \boldsymbol{B}_f \cdot \boldsymbol{\theta} \tag{1.106}$$

with

$$\boldsymbol{B}_f = \begin{bmatrix} \boldsymbol{F}_1^1 & \boldsymbol{F}_1^2 & \boldsymbol{F}_1^3 \\ \vdots & \vdots & \vdots \\ \boldsymbol{F}_T^1 & \boldsymbol{F}_T^2 & \boldsymbol{F}_T^3 \end{bmatrix}_{(T\cdot N)\times(3\cdot N)} \quad \text{and } \boldsymbol{F}_j^i = f_j^i \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}_{N\times N} \tag{1.107}$$

Indeed, from Eq. (1.101) and (1.107),

$$\begin{array}{l} \forall (i,j) \in [\![1;T]\!] \times [\![1;N]\!], \\ \quad \lambda_{i\cdot N+j}^{tot} \quad = \quad \lambda_j(t_i) \quad = \quad c_j^1 \cdot f_{t_i}^1 + c_j^2 \cdot f_{t_i}^2 + c_j^3 \cdot f_{t_i}^3 \\ \quad (\boldsymbol{B}_f \cdot \boldsymbol{\theta})_{i\cdot N+j} \quad = f_{t_i}^1 \cdot \theta_j + f_{t_i}^1 \cdot \theta_{N+j} + f_{t_i}^3 \cdot \theta_{2N+j} \quad = \quad c_j^1 \cdot f_{t_i}^1 + c_j^2 \cdot f_{t_i}^2 + c_j^3 \cdot f_{t_i}^3 \end{array}$$

It would be demonstrated in the same way that:

$$\boldsymbol{\lambda}^{tot} = \boldsymbol{B}_c \cdot \boldsymbol{\gamma} \tag{1.108}$$

with

$$\boldsymbol{B}_c = \begin{bmatrix} \boldsymbol{C}_1^1 & \boldsymbol{C}_1^2 & \boldsymbol{C}_1^3 \\ \vdots & \vdots & \vdots \\ \boldsymbol{C}_N^1 & \boldsymbol{C}_N^2 & \boldsymbol{C}_N^3 \end{bmatrix}_{(T\cdot N)\times(3\cdot T)} \quad \text{and } \boldsymbol{C}_j^i = c_j^i \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}_{T\times T} \tag{1.109}$$

As a result, Eq. (1.95) can be modelled as:

$$\boldsymbol{y}^{\boldsymbol{exp}} = (\boldsymbol{H} \cdot \boldsymbol{B}_f) \cdot \boldsymbol{\theta} + \boldsymbol{N}_1 \tag{1.110}$$

$$\boldsymbol{y}^{\boldsymbol{exp}} = (\boldsymbol{H} \cdot \boldsymbol{B}_c) \cdot \boldsymbol{\gamma} + \boldsymbol{N}_2 \tag{1.111}$$

where $\boldsymbol{B}_f$ is a matrix that depends on the temporal basis functions γ and $\boldsymbol{B}_c$ is a matrix that depends on the weighting coefficients $\boldsymbol{\theta}$. $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ are Poisson noise. Optimisation is then performed through alternate MLEM on Eq. 1.110 and 1.111 according to Shepp & Vardi update scheme shown in Eq. (1.78).

This method was validated on PET-data. The number of detected coincidences is 60 millions. Reconstruction volume is made of $64 \times 64 \times 50$ voxels whose side length is 1,2 mm. Spatiotemporal reconstruction is performed on 100 time frames. Post-reconstruction smoothing on both 3D spatial set of coefficients and temporal basis functions was investigated. Only spatial smoothing should be considered for this method, because temporal smoothing only introduces bias without impacting the mean noise error. This method does not exploit any *a priori* on the temporal basis functions and consequently, this method is very general and can be convenient for estimating TACs. Limitations are that solutions depend on the initialisation, on the choice of the number of functions that are taken into account for reconstructions and on the frequency of the alternations of optimisations on $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$.

1.5.3 Direct and semi-direct estimations of kinetic parameters within reconstructions

It was then proposed to add a prior on time-activity curves of every voxel of a region of interest in order to constrain the 4D solution (3D+time) so that it is also a correct solution for the compartment model for a given radioligand (Kadrmas 01). The prior is introduced through maximum *a posteriori*. At every iteration, the optimal kinetic parameters are computed, which means that the input function is available. The new kinetic parameters modify the prior in the next update of the solution, which results in a final solution

which complies to projections but at the same time allows a robust estimation of kinetic parameters. This method was applied for a two-compartment model on two patients, as well as on simulations in order to quantify improvements compared to independent reconstructions. The radioligands were ^{99m}Tc-teboroxime and activity distributions were reconstructed from SPECT-scans. It was shown that in most cases, the kinetic parameters k_{12} and k_{21} of the compartmental model are better estimated than with reconstruction algorithms without constraints on TACs of voxels. In simulations, the true value for k_{12} and k_{21} were respectively 1.4 and 0.5 $\mathrm{ml} \cdot \mathrm{min}^{-1} \cdot \mathrm{ml}^{-1}$. With this new algorithm, the estimations are respectively 1.402 ± 0.151 $\mathrm{ml} \cdot \mathrm{min}^{-1} \cdot \mathrm{ml}^{-1}$ and 0.535 ± 0.027 $\mathrm{ml} \cdot \mathrm{min}^{-1} \cdot \mathrm{ml}^{-1}$, whereas with OSEM and a compartmental analysis from the reconstructed frames, the estimations are 1.491 ± 0.129 $\mathrm{ml} \cdot \mathrm{min}^{-1} \cdot \mathrm{ml}^{-1}$ and 0.524 ± 0.025 $\mathrm{ml} \cdot \mathrm{min}^{-1} \cdot \mathrm{ml}^{-1}$. This shows that k_{12} is better estimated with the new method, but not k_{21}. The most interesting parameter in such analysis is usually the wash-in parameters k_{12} however, thus this method brings improvements. This method requires that the input function is available but does not estimate it by itself, which is a limitation of the method as it is difficult to estimate the input function. Furthermore, this method was only applied to a one-tissue compartment model, where only two parameters must be estimated.

The idea of directly estimating kinetic parameters from sinogram data was then applied to more complex kinetic models (Kamasak 05). This method, called parametric iterative coordinate descent (PICD), directly estimates the kinetic parameters for every voxel. The method was derived for a two-tissue compartment model but no restriction prevents this method from working on more complex models. As previously, the input function $\mathrm{C}_P(t)$ must be measured or estimated beforehand and be supplied to the algorithm. The activity λ_i in every voxel i is expressed as a function of kinetic parameters $\boldsymbol{\varphi}_i = [k_{1,i}\ k_{2,i}\ k_{3,i}\ k_{4,i}]^{\mathrm{T}}$as:

$$\lambda_i(\boldsymbol{\varphi}_i, t) = [a_i, b_i] \begin{bmatrix} \alpha(c_i, t) \\ \beta(d_i, t) \end{bmatrix} + \gamma(t) \tag{1.112}$$

with

$$a_i = \frac{k_{1,i}}{2\Delta_i}\left(k_{2,i} - k_{3,i} - k_{4,i} + \Delta_i\right) \qquad b_i = \frac{k_{1,i}}{2\Delta_i}\left(-k_{2,i} + k_{3,i} + k_{4,i} + \Delta_i\right) \tag{1.113}$$

$$c_i = \frac{1}{2}\left(k_{2,i} + k_{3,i} + k_{4,i} + \Delta_i\right) \qquad d_i = \frac{1}{2}\left(k_{2,i} + k_{3,i} + k_{4,i} - \Delta_i\right) \tag{1.114}$$

$$\Delta_i = \sqrt{(k_{2,i} + k_{3,i} + k_{4,i})^2 - 4k_{2,i}k_{4,i}} \tag{1.115}$$

$$\alpha(c_i, t) = \left(\mathrm{C}_P(t) \star \left[e^{-c_i(t)}\chi(t > 0)\right]\right) \cdot (1 - V_B)S_A e^{-t/\tau} \tag{1.116}$$

$$\beta(d_i, t) = \left(\mathrm{C}_P(t) \star \left[e^{-d_i(t)}\chi(t > 0)\right]\right) \cdot (1 - V_B)S_A e^{-t/\tau} \tag{1.117}$$

$$\gamma(t) = V_B C_{WB}(t) \tag{1.118}$$

where χ is the Heaviside unit step function, $\mathrm{C}_P(t)$ is the arterial input function, V_B is the fraction volume of blood per voxel, S_A is the initial specific activity of the tracer, C_{WB} is the tracer concentration in whole blood and τ is the characteristic decay time of the radiotracer. $\star$ is the convolution operator. The function defined as $(k_{1,i}, k_{2,i}, k_{3,i}, k_{4,i}) \mapsto (a_i, b_i, c_i, d_i)$ is a bijection and consequently all $k_{.,i}$ can be retrieved from (a_i, b_i, b_i, d_i). Writing Eq. (1.112)

with a matrix notation, the log-likelihood $L(\boldsymbol{Y}|\boldsymbol{\varphi})$ to have the sinogram for all time frames $\boldsymbol{Y}$ given all kinetic parameters $\boldsymbol{\varphi}$ is given by Eq. (1.120):

$$\boldsymbol{\Lambda}(\boldsymbol{\varphi}) = \begin{bmatrix} \lambda_i(\boldsymbol{\varphi}_0, t_0) & \dots & \lambda_i(\boldsymbol{\varphi}_0, t_{K-1}) \\ \vdots & & \vdots \\ \lambda_i(\boldsymbol{\varphi}_{N-1}, t_0) & \dots & \lambda_i(\boldsymbol{\varphi}_{N-1}, t_{K-1}) \end{bmatrix} \tag{1.119}$$

$$L(\boldsymbol{Y}|\boldsymbol{\varphi}) = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} Y_{m,k} \ln((\boldsymbol{A} \cdot \boldsymbol{\Lambda})_{m,k} + \mu) - ((\boldsymbol{A} \cdot \boldsymbol{\Lambda})_{m,k} + \mu) - \ln(Y_{m,k}!) \tag{1.120}$$

where μ may be an estimation of the number of accidental coincidences and $\boldsymbol{A}$ is a model matrix of the imaging system. The maximisation of this likelihood is not performed through a 2-step expectation then maximisation scheme, but rather through a 2-stage nested optimisation. This algorithm updates sequentially every voxel i by first looking for optimal a_i and b_i, with the constraint $(a_i, b_i \geq 0)$, and then by looking for c_i and d_i with the constraint $(c_i \geq d_i \geq 0)$. The objective function is a second-order Taylor expansion of Eq. (1.120). Some regularisations can be added but it is not discussed here because it is not the key contribution of this study. This method was evaluated from simulations of a rat brain phantom. The simulated radioligand was ^{11}C-raclopride. The simulation models a 60-min acquisition divided into 18 time frames. This method shows significant improvements on the estimation of all kinetic parameters compared to the estimation of kinetic parameters from reconstructions of independent frames. A normalised root-mean square error (nRMSE) was defined for every kinetic parameter k_i as:

$$nRMSE(k_i) = \frac{\sqrt{\sum_{s \in \mathcal{S}} \left(k_{i,s}^{true} - k_{i,s}^{estimated} \right)^2}}{\sqrt{\sum_{s \in \mathcal{S}} \left(k_{i,s}^{true} - k_{i,s}^{ind.frames} \right)^2}} \tag{1.121}$$

where $\mathcal{S}$ is the spatial domain where the nRMSE is computed. The reduction on the error on the estimation of k_1 and k_4 with PICD was about 40% and the error reduction was about 80% for k_2 and k_3. Estimating kinetic parameters for every voxel decreases the number of parameters to be estimated, as the kinetics of every voxel is completely defined from 4 parameters, instead of 18 estimations in this example. Nevertheless this method requires that the arterial input function was previously measured or estimated.

Methods which were developed for 4D-imaging intend at modelling the variation of concentrations in voxels in order to estimate them from a small number of parameters. Such kinds of methods should be investigated to attempt to estimate arterial input functions from imaging modalities. Next section presents a method called "non-negative matrix factorisation" (NMF), that could also be used to build a 4D model of the activity spatiotemporal distribution. The point of using such a method is that positivity constraints are handled directly. Furthermore, new algorithms that are used to perform NMF take advantage of projected gradient methods that are faster than other optimisation constrained algorithms.

1.6 Non-Negative Matrix Factorisation

Two techniques of the previous section intend at decomposing unknown parameters $\lambda_i(t)$ as a linear combination of weighted temporal functions. Furthermore, the new unknown parameters are expected to be positive. This leads to investigate the capability of a method

called "non-negative matrix factorisation" (NMF) that is presented in this section and will be used and modified in chapter 4 for estimating the β^+ input function. NMF intends at decomposing a matrix $\boldsymbol{V}$ on two non-negative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$. In this sense, it is a source separation technique. It will be used to decompose the spatiotemporal distribution of any object that is imaged into several spatial components and their associated kinetics. This is expected to improve the estimation of the arterial concentration for temporal frames where the amount of signal is low which prevents the spatial distribution of the artery from being estimated. NMF is very interesting because it leads to solutions whose physical interpretations are more intuitive than solutions that principal component analysis (PCA) or independent component analysis (ICA) provide. For instance, performing PCA or ICA on activity distributions return negative coefficients, which can not be interpreted. NMF avoids such situations. It was developed in the field of spectral data analysis and found applications in part recognition in computer vision (Lee 99). Attempts have also been done to apply NMF to medical imaging (Lee 01).

The main advantage if compared to PCA is that it is not assumed that the data is Gaussian. Furthermore the decomposition returns two non-negative matrices that may be interpreted in term of physical phenomena.

1.6.1 Factor Analysis of Medical Image Sequences

NMF is close to a method developed from 1980 and called the factor analysis of medical image sequences (FAMIS) (Di Paola 82) which aims at decomposing the data as a linear combination of elementary functions that are weighted by factors. For instance, a sequence of T images which is made of M pixels can be represented by a $(M \times T)$ matrix. Rows $(\boldsymbol{x}_i)$ are called trixels and represents the signal variation in pixel i for successive time frames. If it is assumed that there are T acquired images and M trixels $\boldsymbol{x}_i$ which are T-long vectors. Although the number of trixels is usually large in medical imaging, the kinetics of only a few tissues are sufficient to understand the variations in all trixels: every trixel can be viewed as a linear combination of a limited number of elementary kinetics. This can be written with matrix notations:

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{F} \tag{1.122}$$

where $\boldsymbol{X}$ is a $(M \times T)$-matrix that contains all the signal, $\boldsymbol{A}$ is a $(M \times C)$-matrix that represents weighting factors and $\boldsymbol{F}$ is a $(C \times T)$-matrix that represents the elementary functions. C is the number of tissues with independent kinetics that lay in the image, T the number of time frames and M the number of pixels or voxels in any images of the sequences. In practice, such a decomposition is very unlikely to exist because there are less coefficients in $\boldsymbol{A}$ and $\boldsymbol{F}$ than in $\boldsymbol{X}$ if the number of expected elementary functions is smaller than the number of time frames. Consequently an additional noise term $\boldsymbol{N}$ should be taken into account. In nuclear imaging, Poisson fluctuations are included in $\boldsymbol{N}$.

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{F} + \boldsymbol{N} \tag{1.123}$$

The first step to decompose $\boldsymbol{X}$ into $\boldsymbol{A}$ and $\boldsymbol{F}$ with FAMIS consists in performing a singular value decomposition according to a metric that is adapted to the statistical properties of $\boldsymbol{X}$, namely all the observations x_{ij} of $\boldsymbol{X}$ are random draws from Poisson distributions with unknown parameters ν_{ij} (Benali 93). After this step, an orthogonal decomposition of noise-free observations $\tilde{\boldsymbol{X}}$ is obtained:

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{N} = \boldsymbol{U} \cdot \boldsymbol{V}^T \tag{1.124}$$

with $\boldsymbol{U}$ and $\boldsymbol{V}$ being two orthogonal matrices. This is a valid decomposition of the noise-free matrix $\tilde{\boldsymbol{X}}$, but $\boldsymbol{U}$ and $\boldsymbol{V}$ have no physical meaning. An oblique analysis is performed on $\tilde{\boldsymbol{X}}$ to decompose it on two matrices $\boldsymbol{A}$ and $\boldsymbol{F}$ under positivity constraints (Buvat 93). The matrices $\boldsymbol{A}$ and $\boldsymbol{F}$ are estimated alternatively through regression according to χ^2-metric. At every iteration, negative coefficients are either set to 0 or are penalized so that they converge to 0 as the number of iterations increases. The algorithm stops when the number of negative coefficients is below a user-defined threshold. The result is a decomposition of the image sequences $\boldsymbol{X}$ on two positive matrices $\boldsymbol{A}$ and $\boldsymbol{F}$:

$$(\boldsymbol{A}_\infty, \boldsymbol{F}_\infty) = \arg \min_{(\boldsymbol{A},\boldsymbol{F})} \|\boldsymbol{X} - \boldsymbol{A} \cdot \boldsymbol{F}\|^2_{\boldsymbol{M}} \tag{1.125}$$

where $\boldsymbol{M}$ is a $(T \times T)$ matrix that defines a metric on $\mathbb{R}^T$.

FAMIS allows the decomposition of a set of observations into the factorisation of positive matrices through a succession of algebraic manipulations. The convergence of the regression in the oblique analysis does not guarantee that the algorithm converges to the optimal solution. Analysing convergence properties is easier with NMF as it will be shown in next section.

1.6.2 Model of NMF

NMF is also an approximate decomposition since usually, $\boldsymbol{V}$ is not exactly decomposable on two non-negative matrices, because $\boldsymbol{V}$ is usually made of more coefficients than $\boldsymbol{W}$ and $\boldsymbol{H}$. The decomposition can be written as in Eq. (1.126), where $\boldsymbol{R}$ is a residue, that is not necessarily non-negative.

$$\boldsymbol{V} = \boldsymbol{W} \cdot \boldsymbol{H} + \boldsymbol{R} \tag{1.126}$$

A first method to perform such a decomposition relies on multiplicative updates (Lee 00). In this paper, Lee proposes two types of algorithms depending on the cost function that is taken into account. He developed methods that minimise Euclidean distance between $\boldsymbol{V}$ and $\boldsymbol{W} \cdot \boldsymbol{H}$, as well as the Kullback-Leibler divergence $D(\boldsymbol{V}||\boldsymbol{W}\boldsymbol{H})$.

1.6.2.1 Observations are Gaussian variables

Least square methods minimise the distance between $\boldsymbol{V}$ and $\boldsymbol{W} \cdot \boldsymbol{H}$ that is defined according to:

$$D(\boldsymbol{W}, \boldsymbol{H}) = \|\boldsymbol{V} - \boldsymbol{W} \cdot \boldsymbol{H}\|^2_F \tag{1.127}$$

where $\|.\|_F$ is the Frobenius norm ((Golub 96)). This corresponds to assume that coefficients of matrix $\boldsymbol{V}$ are realisations of Gaussian variables. Indeed, the likelihood $L_G(\boldsymbol{W}, \boldsymbol{H}) \propto P(\boldsymbol{V}|\boldsymbol{W}, \boldsymbol{H})$ can be written under Gaussianity assumptions as:

$$L_G(\boldsymbol{W}, \boldsymbol{H}) \propto \prod_{i,j} \exp\left(-\frac{(v_{i,j} - (\boldsymbol{W}\boldsymbol{H})_{ij})^2}{2\sigma^2_{i,j}}\right) \tag{1.128}$$

When all (σ_{ij}) are equal to 1, maximising $L_G(\boldsymbol{W}, \boldsymbol{H})$ is equivalent to minimising the following objective function:

$$E = -\ln\left(L_G(\boldsymbol{W}, \boldsymbol{H})\right) = \sum_{i,j}(v_{i,j} - (\boldsymbol{W}\boldsymbol{H})_{ij})^2 \tag{1.129}$$

$$= -\|\boldsymbol{V} - \boldsymbol{W} \cdot \boldsymbol{H}\|^2_F \tag{1.130}$$

Minimising E is complex because it is not a quadratic function. However, it becomes quadratic if an alternate minimisation is performed on E, once with $\boldsymbol{W}$ being kept fixed, the other with $\boldsymbol{H}$ being kept fixed. Lee demonstrated that the following update procedure converges to a local minimum (Lee 00):

$$w_{i,j}^{n+1} = w_{i,j}^{n} \frac{(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{V})_{i,j}}{(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H})_{i,j}} \tag{1.131}$$

$$h_{j,k}^{n+1} = h_{j,k}^{n} \frac{(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{W})_{j,k}}{(\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})_{j,k}} \tag{1.132}$$

This is a simple update procedure that can be implemented easily. It was applied in spectral analysis and part recognition in computer vision. However, in nuclear imaging, it is usually not possible to assume that the observations $\boldsymbol{V}$ are Gaussian-distributed.

1.6.2.2 Observations are Poisson variables

Lee proposed a second algorithm that minimises the divergence $D(\boldsymbol{V}||\boldsymbol{W},\boldsymbol{H})$:

$$D(\boldsymbol{V}||\boldsymbol{W},\boldsymbol{H}) = \sum \left(v_{i,j} \ln \frac{v_{i,j}}{(\boldsymbol{W}\cdot\boldsymbol{H})_{i,j}} - v_{i,j} + (\boldsymbol{W}\cdot\boldsymbol{H})_{i,j} \right) \tag{1.133}$$

This is denoted a divergence instead of a distance because this measure is not symmetric. $D(\boldsymbol{V}||\boldsymbol{W},\boldsymbol{H})$ can be related to Kullback-Leibler divergence when $\|\boldsymbol{V}\|_1 = 1$ and $\|\ \boldsymbol{W}\cdot\boldsymbol{H}\|_1 = 1$. It is sometimes referred as generalised Kullback-Leibler divergence. Minimising this divergence is equivalent to maximise the likelihood $L_P(\boldsymbol{W},\boldsymbol{H}) \propto P(\boldsymbol{V}|\boldsymbol{W},\boldsymbol{H})$. Indeed, when $\boldsymbol{V}$ is Poisson-distributed, $L_P(\boldsymbol{W},\boldsymbol{H})$ can be written as:

$$L_P(\boldsymbol{W},\boldsymbol{H}) = \prod_{i,j} \frac{(\boldsymbol{W}\cdot\boldsymbol{H})_{i,j}^{v_{i,j}} \cdot e^{-(\boldsymbol{W}\cdot\boldsymbol{H})_{i,j}}}{v_{i,j}!} \tag{1.134}$$

Maximising $L_P(\boldsymbol{W},\boldsymbol{H})$ is equivalent to minimise the cost function E defined as:

$$E = \ln\left(L_P(\boldsymbol{W},\boldsymbol{H})\right) - \sum \ln v_{i,j}! \tag{1.135}$$

$$= \sum_{i,j} \left[-v_{i,j} \ln(\boldsymbol{W}\cdot\boldsymbol{H})_{i,j} + (\boldsymbol{W}\cdot\boldsymbol{H})_{i,j} \right] \tag{1.136}$$

$$= D(\boldsymbol{V}||\boldsymbol{W},\boldsymbol{H}) - \sum_{i,j} \left[v_{i,j} \ln v_{i,j} - v_{i,j} \right] \text{ where } \sum_{i,j} \left[v_{i,j} \ln v_{i,j} - v_{i,j} \right] \text{ is constant} \tag{1.137}$$

Lee proposed a multiplicative update algorithm for minimising the divergence $D(\boldsymbol{V}||\boldsymbol{W},\boldsymbol{H})$:

$$w_{i,j}^{n+1} = \frac{w_{i,j}}{\sum_l h_{j,l}} \sum_l h_{j,l} \frac{v_{i,l}}{(\boldsymbol{W}\boldsymbol{H})_{i,l}} \tag{1.138}$$

$$h_{j,k}^{n+1} = \frac{h_{j,k}}{\sum_l w_{l,j}} \sum_l w_{l,j} \frac{v_{l,k}}{(\boldsymbol{W}\boldsymbol{H})_{j,k}} \tag{1.139}$$

One can recognise the algorithm proposed by Shepp&Vardi (Shepp 82) which was applied to solve the systems of Eq. (1.140) on the columns of $\boldsymbol{H}$ and on the columns of $\boldsymbol{W}^{\mathrm{T}}$:

$$\boldsymbol{V} = \boldsymbol{W}\cdot\boldsymbol{H} + \boldsymbol{N} \text{ and } \boldsymbol{V}^{\mathrm{T}} = \boldsymbol{H}^{\mathrm{T}}\cdot\ \boldsymbol{W}^{\mathrm{T}} + \boldsymbol{N}^{\mathrm{T}} \tag{1.140}$$

1.6.3 Scale indetermination and order indetermination

The decomposition of $\boldsymbol{V}$ on $\boldsymbol{W}$ and $\boldsymbol{H}$ is not unique (Moussaoui 05). Indeed, different $\boldsymbol{W}$ and $\boldsymbol{H}$ can lead to similar ($\boldsymbol{W} \cdot \boldsymbol{H}$). Let choose $\boldsymbol{W}_0$ and $\boldsymbol{H}_0$. Assuming that a decomposition on C components is performed, any matrices $\boldsymbol{W}_R$ and $\boldsymbol{H}_R$ which are defined by $\boldsymbol{W}_R = \boldsymbol{W}_0 \times \boldsymbol{R}^{-1}$ and $\boldsymbol{H}_R = \boldsymbol{R} \times \boldsymbol{H}_0$ is also a solution if $\boldsymbol{R} \in GL(C, \mathbb{R})$ which is the group of invertible $(C \times C)$-matrices. Positivity constraint on physical components imposes that $\boldsymbol{W}\boldsymbol{R}$ and $\boldsymbol{R}^{-1}\boldsymbol{H}$ are real positive matrices. For instance, by multiplying any diagonal matrix $\boldsymbol{D}$ with strictly positive coefficients with a permutation $\boldsymbol{P}$ matrix, a matrix $\boldsymbol{R}$ that satisfies the previous positivity constraint can be built. The multiplication by the permutation matrix $\boldsymbol{P}$ means that the columns of matrix $\boldsymbol{W}$ can be exchanged as long as rows of matrix $\boldsymbol{H}$ are also exchanged without any influences on log-likelihood $\ln(\boldsymbol{W}, \boldsymbol{H})$. This is called order-invariance. Columns of matrix $\boldsymbol{W}$ can also be scaled by multiplying $\boldsymbol{W}$ by as strictly positive diagonal matrix as long the rows of matrix H are also scaled by matrix $\boldsymbol{D}^{-1}$. This is known as the scale invariance. These invariances have no influence on the calculation of $\boldsymbol{W}\boldsymbol{H}$.

NMF allows the decomposition of a spatiotemporal non-negative distribution on a non-negative spatial matrix and a non-negative temporal matrix. This decomposition is usually performed with iterative algorithms whose convergence is proven. This decomposition can take into account the nature of the statistical noise in the spatiotemporal matrix when it is either gaussian or Poisson. The main drawback of this method is that the solution depends on the initialisation point, the solution is consequently local.

1.7 Gradient methods for the estimation of non-negative matrix factorisation

The algorithms that Lee proposed can be viewed as a gradient method according to the following considerations. When the observations are Gaussian-distributed, the first-order Taylor expansion of any $h_{j,k}^{n+1}$ can be expressed as:

$$h_{j,k}^{n+1} = h_{j,k}^{n} + \eta_{i,j} \frac{\partial E}{\partial h_{i,j}} \tag{1.141}$$

$$= h_{j,k}^{n} + \eta_{j,k} \left[(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{W})_{j,k} - (\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})_{j,k} \right] \tag{1.142}$$

if setting $\eta_{j,k} = \dfrac{h_{j,k}^{n}}{(\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})_{j,k}}$, Lee's first multiplicative update algorithm is obtained:

$$h_{j,k}^{n+1} = h_{j,k}^{n} \frac{(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{W})_{j,k}}{(\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})_{j,k}} \tag{1.143}$$

With a similar method, the update algorithm for $\boldsymbol{W}$ could be viewed as a gradient-based algorithm. This could also be done when observations $\boldsymbol{V}$ are Poisson-distributed.

The algorithms that Lee proposed for performing NMF are straightforward to implement but their convergence rates are not optimal as it is equivalent to a ascent gradient where the size of the steps are fixed. New classes of constrained gradient optimisation can be used to optimise cost functions encountered in NMF. However, such methods have only been proposed in the literature to solve the class of problems where the data are assumed to be Gaussian variables (Lin 07). Chu proposes an overview of the optimisation methods for non-negative matrix factorisation (Chu 04).

The direct minimisation of all parameters $(w_{i,j}, h_{j,k}) = (x_1, \ldots, x_n)$ that is required to perform non-negative matrix factorisation is not quadratic, whatever the measure that is chosen. However, it becomes quadratic if the observations are assumed to be Gaussian and that the optimisation is performed alternatively on $\boldsymbol{W}$ and $\boldsymbol{H}$. In the context of this PhD thesis, the Gaussian approximation usually does not stand because the counts in pixels are very low. Furthermore, it is wanted to update all parameters at every iteration in order to reduce the risk to get trapped in a local minimum. Indeed, if $\boldsymbol{W}$ and $\boldsymbol{H}$ are respectively made of $M_{\boldsymbol{W}}$ and $M_{\boldsymbol{H}}$ variables, it can happen that the gradients $\boldsymbol{\nabla}_{S_{\boldsymbol{W}}} f(\boldsymbol{x}_k)$ and $\boldsymbol{\nabla}_{S_{\boldsymbol{H}}} f(\boldsymbol{x}_k)$ are null on every subspace $S_{\boldsymbol{W}}$ and $S_{\boldsymbol{H}}$, without gradient $\boldsymbol{\nabla}_{S_{\boldsymbol{H}} \oplus S_{\boldsymbol{W}}} f(\boldsymbol{x}_k)$ is null, with $S_{\boldsymbol{W}}$ and $S_{\boldsymbol{H}}$ being the convex spaces where $\boldsymbol{W}$ and $\boldsymbol{H}$ respectively stand in. As non-negative factorisations suffer from local minima, alternate optimisations seem not to be the best choice. As a result, optimisations that can performed non-negative matrix factorisations can not be viewed as a set quadratic problems because they are non-linear problems.

1.7.1 Optimisation through gradient descent

The general method to solve an optimisation problem on a multi-dimensional space is to search iteratively points $(\boldsymbol{x}_k)$ that makes the objective functional f decreasing. From an initial point $\boldsymbol{x}_0$, the next point is searched on the descending gradient direction $-\boldsymbol{\nabla} f(\boldsymbol{x}_0)$. This procedure is repeated iteratively according to Eq. (1.144), where $\boldsymbol{\nabla} f(\boldsymbol{x}_k)$ is the gradient of objective functional f at point $\boldsymbol{x}_k$ and α_k is the step size that minimises f on the line defined by $\alpha \mapsto \boldsymbol{x}_k - \alpha \boldsymbol{\nabla} f(\boldsymbol{x}_k)$, which can be written as in Eq. (1.145).

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{\nabla} f(\boldsymbol{x}_k) \tag{1.144}$$

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}_k - \alpha \boldsymbol{\nabla} f(\boldsymbol{x}_k)) \tag{1.145}$$

The gradient $\boldsymbol{\nabla} f(\boldsymbol{x}_k)$ being usually straightforward to derive, the tricky part is to choose an efficient step size α_k at very iteration k. The gradient descent technique looks for an optimal α through line search on line $\alpha \mapsto \boldsymbol{x}_k - \alpha \boldsymbol{\nabla} f(\boldsymbol{x}_k)$, but this is time-consuming since it requires numerous evaluations of f. Some more elaborated methods exist that provide step sizes α_k that make optimisations converge rapidly.

1.7.2 Newton and quasi-Newton methods

Non-linear optimisations mostly take advantages of Newton and quasi-Newton algorithm methods, that used the Hessian $\mathrm{Hess}(f)$ of f or an approximation of it to derive α_k efficiently. These algorithms assume that objective functions can be locally approximated by a quadratic function in the neighbourhood of an optimal point.

Newton's optimisation uses Newton's method on the gradient $\boldsymbol{\nabla} f(\boldsymbol{x}_k)$ in order to find its roots. Taylor expansion of $\boldsymbol{\nabla} f(\boldsymbol{x}_k)$ leads to:

$$\boldsymbol{\nabla} f(\boldsymbol{x}_k + \boldsymbol{h}) \approx \boldsymbol{\nabla} f(\boldsymbol{x}_k) + \mathrm{Hess} f(\boldsymbol{x}_k) \cdot \boldsymbol{h}^{\mathrm{T}} \tag{1.146}$$

When $\boldsymbol{x}_k + \boldsymbol{h}$ is an optimal point, $\boldsymbol{h}$ is obtained through:

$$\boldsymbol{\nabla} f(\boldsymbol{x}_k + \boldsymbol{h}) = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{\nabla} f(\boldsymbol{x}_k) + \mathrm{Hess} f(\boldsymbol{x}_k) \cdot \boldsymbol{h}^{\mathrm{T}} = \boldsymbol{0} \tag{1.147}$$

$$\Leftrightarrow \quad \boldsymbol{h} = -\left[\mathrm{Hess} f(\boldsymbol{x})\right]^{-1} \boldsymbol{\nabla} f(\boldsymbol{x}) \tag{1.148}$$

As a result, the sequence $\{\boldsymbol{x}_k\}$ given by Eq. (1.149) defines an iterative scheme that optimises objective functional f.

$$\boldsymbol{x}_{k+1} = \boldsymbol{x} - \left[\mathrm{Hess} f(\boldsymbol{x})\right]^{-1} \boldsymbol{\nabla} f(\boldsymbol{x}) \tag{1.149}$$

This method is very interesting because it has a quadratic convergence rate around the solution. In order to ease the local convergence when quite far from the solution, a line search can be added in $\left(\left(\text{Hess}f(\boldsymbol{x})\right)^{-1}\boldsymbol{\nabla}f(\boldsymbol{x})\right)$-direction:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x} - \rho_k \left[\text{Hess}f(\boldsymbol{x})\right]^{-1}\boldsymbol{\nabla}f(\boldsymbol{x}) \tag{1.150}$$

The main drawback of this method is that it requires to derive the Hessian of f and to invert it at every iteration k. When the objective function f describes a large-scale problem, the computation of the Hessian matrix $\text{H}f(\boldsymbol{x}_k)$ and its inversion are resource-demanding. Furthermore, the inversion is not necessarily stable depending on $\text{Hess}f(\boldsymbol{x}_k)$ condition number.

An alternative is to use quasi-Newton methods that replaces $\text{Hess}f(\boldsymbol{x}_k)$ by a matrix $\boldsymbol{B}_k$ that verifies :

- $\boldsymbol{B}_k$ is positive-definite
- $\boldsymbol{B}_k$ is symmetric
- $\boldsymbol{B}_k$ verifies: $\boldsymbol{B}_k\left(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\right) = \boldsymbol{\nabla}f(\boldsymbol{x}_{k+1}) - \boldsymbol{\nabla}f(\boldsymbol{x}_k)$

One of the most used quasi-Newton optimisation algorithms is the one called BFGS-algorithm, from the name of its authors: Broyden, Fletcher, Goldfarb and Shannon. It gives an update formula of the approximate Hessian $\boldsymbol{B}$ and the corresponding invert matrix $\boldsymbol{B}_k^{-1}$. Details on this method are reported in Appendix 3. The new update formula is:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \rho_k \boldsymbol{B}^{-1}\boldsymbol{\nabla}f(\boldsymbol{x}_k) \tag{1.151}$$

1.7.3 Optimisation on positive sets

BFGS method can be adapted to perform non-negative matrix factorisation by modifying the objective function of the non-negative matrix factorisation problem of Eq. (1.130) and (1.137) as:

$$E = \|\boldsymbol{V} - \boldsymbol{W}^2\boldsymbol{H}^2\|^2 \text{ for Gaussian observations} \tag{1.152}$$

$$\text{or } E = \sum\left(v_{i,j}\ln\frac{v_{i,j}}{(\boldsymbol{W}^2\cdot\boldsymbol{H}^2)_{i,j}} - v_{i,j} + (\boldsymbol{W}^2\cdot\boldsymbol{H}^2)_{i,j}\right) \text{ for Poisson observations} \tag{1.153}$$

where $\forall(i,j,k),\quad \left(W^2\right)_{i,j} = w_{i,j}^2$ and $\left(H^2\right)_{j,k} = h_{j,k}^2$. The drawback of such a method is that the objective function does not look like a quadratic function at all which leads to poor convergence rate.

1.7.4 Optimisation based on a projected gradient method

Projected gradient methods were developed to address constrained optimisations on convex sets. For a problem whose solution is constrained on convex set Ω, the successive iterations that approximate the solution have the form:

$$\boldsymbol{x}_{k+1} = P(\boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k) \tag{1.154}$$

where P is the projection operator on convex set Ω, $\boldsymbol{d}_k$ is the search direction and α_k is the step size that makes f decreasing on the line defined by $\alpha \mapsto \boldsymbol{x}_k + \alpha\boldsymbol{d}_k$. Different algorithms were proposed that suggest different step sizes α_k. A way to determine efficiently α_k was proposed in (Bertsekas 99). It sets $\alpha_k = \beta^t$ where $\beta \in [0,1]$ and t is the first non-negative

integer that verifies: $f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \leq \sigma \boldsymbol{\nabla} f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)$, σ being in $[0, 1]$. A common choice for σ is $\sigma = 0.01$.

A more refined method is the Spectral Projected gradient method (Birgin 00). It presents the advantage that the convergence is non-monotonic which allows escaping from some of the local minima that are encountered in non-negative matrix factorisation. The search direction is defined as:

$$\boldsymbol{d}_k = P\left(\boldsymbol{x}_k - \frac{1}{\sigma_k}\boldsymbol{\nabla} f(\boldsymbol{x}_k)\right) - \boldsymbol{x}_k \tag{1.155}$$

where σ_k is defined by $\sigma_k = \dfrac{(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{\nabla} f(\boldsymbol{x}_{k+1}) - \boldsymbol{\nabla} f(\boldsymbol{x}_k))}{(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)}$, which is deduced from quasi-Newton method where the Hessian is approximated by $\sigma_k \boldsymbol{I}$. It is the solution that minimises $\|\sigma(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) - (\boldsymbol{\nabla} f(\boldsymbol{x}_{k+1}) - \boldsymbol{\nabla} f(\boldsymbol{x}_k))\|_2^2$. At every iteration, the estimation of the solution is updated according to:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \tag{1.156}$$

Contrary to previous methods, the line search that estimates α_k is non-monotonic, which means that it is not imposed that the objective functional f decreases at every iteration (Grippo 86). Some local minima are then avoided.

1.8 Conclusions

Pharmacokinetics aims at understanding physiological mechanisms in tissues, at diagnosing pathologies and at developing drugs to cure patients. This requires the development of new molecules, whose *in-vivo* properties can be assessed through nuclear imaging by labelling these molecules with radioisotopes which emit positrons or photons. The dynamic behaviour of these molecules provides information on the uptake potential of tissues and on chemical mechanisms these molecules go through. Such phenomena can be quantified through pharmacokinetic studies, which can mainly be categorised into two kinds of analyses: compartmental and non-compartmental analyses. Both requires that the radiotracer concentration in plasma is dynamically measured or estimated beforehand for every analysis, as radiotracer kinetic properties can not be derived only from tissue TACs. A large tissue uptake can mean that radiotracers are very efficient at binding to the tissue although a small amount of radiotracers is brought to tissues by arteries, or on the contrary, it can mean that radiotracers poorly bind to tissue but they are massively brought to tissues by arteries. The binding potential of a new molecule to a target tissue being a key property, it is fundamental to estimate the arterial input function in order to know how arteries supply the target tissues with the new molecules.

The gold-standard technique for measuring the arterial input function is arterial blood sampling, Most often the plasmatic input function is required because the radiotracer binds to blood cells. Plasma is obtained by centrifugation. Arterial blood sampling is invasive and painful for patients. Non-invasive or minimally-invasive methods would rather be used to estimate input functions. Some methods for measuring or estimating input functions were presented in this chapter and it is deduced that a general technique for estimating input functions should take advantage of nuclear imaging, but improvements on sensitivity are expected so that arterial input functions can be estimated accurately even in first minutes. Such a technique should work for every β^+-emitter, at least for cardiac and brain imaging, and for newly-developed molecules whose input function shape is unknown, but it will not be able to extract useful information for radiolabelled molecules with a strong metabolism in blood.

An accurate estimation of the arterial input function allows complex pharmaceutical models which better diagnose patients' diseases and their response to therapy. Some notions on nuclear imaging were then introduced. Detailed information is provided on coded-aperture imaging, which has the potential to achieve low-signal high-resolution imaging, which is particularly relevant for estimating arterial input functions, since arteries emit a small amount of photons and their contribution must be separated from contributions of surrounding tissues, particularly from veins.

Statistical reconstruction algorithms were then introduced because it is expected that they better deal with decoding of medical images than correlation-based algorithms which are normally used in coded-aperture imaging. Methods were explained for reconstructing static and dynamic radioactive distributions. Dynamic statistical reconstructions mostly rely on models of the spatiotemporal distributions. As a new model for dynamic nuclear imaging, non-negative matrix factorisation was considered in order to decompose a spatiotemporal distribution on a spatial matrix and a temporal matrix. Such a decomposition is advantageously performed through gradient-based algorithms and some concepts on such optimisation algorithms are given, as it will be used in next chapters.

Imaging system design and system matrix calculation 2

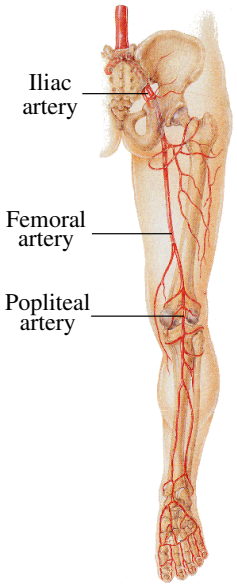
We aim at estimating the β^+-input function from non-invasive measurements of the concentration of radiotracers in patients' arteries using an external detector. First we select the most appropriate artery for such measurements. Then we propose an imaging system based on the experience of our laboratory in the field of coded-aperture imaging (Gmar 04). that takes into account geometric constraints (depth of the artery in tissues, distance of the detector to the artery) and physical considerations (photon attenuations and detection efficiencies at 511 keV (Berger 05)). It is then explained why correlation-based reconstruction algorithms which are usually used to decode projections acquired with coded apertures are not adapted for estimating 3D spatial distributions when thick coded collimators are used for collimating 511 keV-photons. As an alternative, it is suggested to perform reconstructions with a statistical reconstruction algorithm, as the maximum-likelihood expectation maximisation algorithm (MLEM) previously presented in section 1.5.1. Such an algorithm requires the computation of a system matrix which models the imaging system. It can be done through Monte-Carlo simulations, but more efficient methods would be expected, for example through ray-tracing. Three new techniques are presented to speed up the computation of system matrices. The accuracy and the computation time of these methods are quantified.

2.1 Definition of the device characteristics

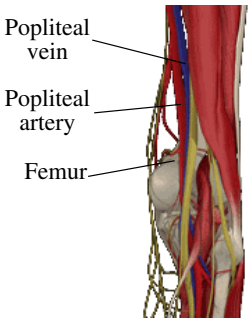
The imaging system which is developed has to comply with some constraints imposed by physiological considerations and by the devices which are available to perform the acquisitions.

2.1.1 Physiological considerations about the location of measurements

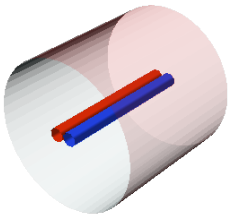
Usually non-invasive measurements take place either over the radial artery in the wrist (Shokouhi 03) or over the carotid in the neck (Litton 90). Other measurement locations could be the aorta, but also the brachial artery in the arm, the iliac, femoral or popliteal arteries in the leg. Pros and cons are investigated in order to determine a measurement location which is appropriate for the estimation of the β^+-input functions. The radial artery stands close from the surface ($<$1 cm) which means that more than 85% ($\mu_{tot}(H_20) = 0.17$ cm^{-1} at 511 keV) of the emitted photons leaves the wrist. The wrist can be isolated from signals coming from the rest of the body by making measurements behind a thin wall made of lead. The main drawback of making the measurements at the radial artery is its small cross section which is about 1 mm^2. There exist larger arteries such as for instance the brachial artery, the aorta or the femoral artery. The section of the brachial artery is larger (from



(a) Map of arteries in a leg, viewed from the back



(b) Description of the anatomy of a knee (from http://www.human-anatomy.net)



(c) Model of popliteal artery and popliteal vein

Figure 2.1 – Model of a knee

1 to 5 mm^2), which should result in a larger amount of photons emitted from this artery than from the radial artery. However, it is difficult to develop an imaging system adapted to this artery which does not measure the contribution of the heart and the brain. The carotid is a large artery whose section is about 30 mm^2 (Ferrara 95). It stands close to the surface. However it is close to the brain whose contribution would largely dominate the contribution of the carotid. The aorta is an artery which is much larger than the carotid since its section is between 300 and 700 mm^2 (Naidich 07) and consequently it would emit much more photons. However, it stands close to the heart which is much bigger (about 300 cm^3) than the aorta (the volume of a 10 cm-long portion of an aorta is between 3 cm^3 and 7 cm^3) and consequently it is difficult to extract the signal which comes from the aorta from the signal coming from the heart. Furthermore the aorta moves during acquisitions, which introduces large spatial blurring in reconstructions. It remains three arteries: the iliac artery, the femoral artery and the popliteal artery. The sections of the common iliac and the common femoral artery are between 30 mm^2 and 80 mm^2 (Radegran 00). The iliac artery stands close to the bladder which is usually very radioactive because of the biological clearance of the radiolabelled molecules. As a result, it is difficult to extract the signal which comes from the iliac artery from the signal coming from the bladder. The superficial femoral artery stands about 50 mm deep in the leg and it is surrounded by muscles, which may take up the radiotracers. Finally, the popliteal artery has a section which is between 20 mm^2 and 40 mm^2 (Debasso 04). It passes through the knee and stands about 50 mm from the surface of the knee. Moreover, surrounding tissues are mostly bones, cartilages and ligaments. The perfusion of radiotracers in these kinds of tissues is slow (0.08 $ml/cm^3/min$ (Bronzino 06)). By elimination, it is decided to select the popliteal artery for estimating the input function. Indeed the carotid, the aorta, the iliac and the femoral arteries can not be good candidates because of the organs in the neighbourhoods which contain a large amount of radiotracers which would spoil the estimation of the concentration in these arteries. The radial and the brachial arteries are not chosen because they are smaller than the popliteal artery.

Measuring the input function at the popliteal artery in the knee is an original choice and

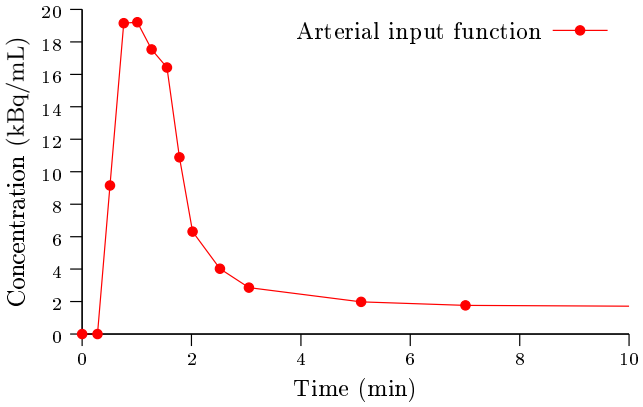


Figure 2.2 – Example of an input function obtained from arterial blood sampling, first 10 minutes. Radioligand is ^{18F}F-A85380 with a bolus injection

is not yet described in literature. The most often selected artery for external measurements is the radial artery in the wrist(see 1.2.3). The diameter of the popliteal artery which is shown in Fig. 2.1 is more than twice larger than the diameter of the radial artery. This means that for a given length and a given concentration of the artery, the activity is at least 4 times larger in the artery of the knee than in the artery of the wrist.

Measuring the arterial input function at the knee should induce almost no delay in the estimation of the arterial input function at the brain. While the heart ejects blood in the ascending aorta, a fraction goes to the brain, another to legs. As the blood flow in the arteries is high (about 1 m/s (Gabrielsen 92)), the time shift between the input function that enters the brain and the input function that is measured at the knee is negligible ($<$1 s). Furthermore, if the temporal resolution of the imaging system is higher than a few seconds, this delay will not be measured. This input function is obtained from arterial blood sampling, and details on the acquisition are given in section 2. Moreover it is important to notice that a vein always stands close to an artery, especially when they are large. This can be observed for the popliteal artery in Fig. 2.1b. The popliteal artery and vein stand in the middle of the knee. A knee is about 100 mm-large. As a result, the knee is modelled as in Fig. 2.1c in the examples presented in this thesis, with the vessels being at 50 mm from the surface of the knee. This is a simple model where tissues are modelled by water, but it is accurate enough to model the main vascular structures in the knee. In particular, no bone is modelled, which remains realistic as long as measurements are not performed from the front side of the knee. Tissues attenuate the flux of photons which leave the artery, which corresponds to a 60%-loss of the signal for a 50 mm-thick tissue. In the next section, the material which is used for designing the imaging system is described.

2.1.2 Material

The laboratory where this PhD-thesis took place has experience with coded-aperture imaging for industrial applications (Gmar 04). Such imaging modality has the potential to address the non-invasive estimation of arterial input functions and this PhD-thesis studies the feasibility of such a technological solution. The design of the imaging system is done in regard to the materials available in our laboratory and at the hospital we are working with. The aim is to develop an imaging system which is affordable. As a result it is decided to use the γ-imager which is manufactured by Biospace Lab [1]. This detector was coupled to a coded-aperture

[1] Biospace Lab, 10 rue Mercoeur, 75011 Paris, France

		Tungsten	CsI(Na)
ρ	$(g \cdot cm^{-3})$	19.3	4.51
μ_T/ρ	$(cm^2 \cdot g^{-1})$	1.34×10^{-1}	9.61×10^{-2}
μ_T	(cm^{-1})	2.580	0.433
μ_{PE}/ρ	$(cm^2 \cdot g^{-1})$	5.66×10^{-2}	2.00×10^{-2}
μ_{PE}	(cm^{-1})	1.090	0.090

Table 2.1 – Densities, photon cross-sections and linear attenuation coefficients for tungsten and CsI(Na) for photoelectric (PE) and total (T) interactions at 511 keV

collimator, which has a much better efficiency than classical collimators.

2.1.2.1 γ-imager

The detector used in this thesis is the bidimensional γ-imager manufactured at Biospace Lab. This is a detector made of a circular CsI-scintillator whose diameter is 120 mm-large. The scintillations are collected by a position-sensitive photomultiplier and this results in a 100 mm-large circular detection area (Russo-Marie 07). At 140 keV, the intrinsic spatial resolution is 2 mm and the energy resolution is 11%. As it was not possible to perform measurements at 511 keV as it will be explained in section 3.5, it is assumed that the intrinsic spatial resolution of the imager does not change at 511 keV and is still 2 mm. This assumption is plausible according to following considerations:

- On one hand, spatial resolution takes advantage of photons with higher energy since they induce brighter scintillations.
- On the other hand, scintillations occur at shorter distance in crystals for 511-keV photons than 140-keV photons, as detailed in Appendix 4. This means that scintillations occur further from photomultipliers, which impairs spatial resolution.

The energy resolution at 511 keV can be deduced analytically from the energy resolution at 140 keV through Eq. (2.1) and is 5.8%.

$$\mathrm{E_R}(511\ \mathrm{keV}) = \mathrm{E_R}(140\ \mathrm{keV})\sqrt{\frac{140}{511}} \tag{2.1}$$

$$= 5.8\% \tag{2.2}$$

This means that it is possible to discriminate photons which deposits less than 498 keV (1σ) in the scintillator. The photons which are not discriminated represents 84% of the photoelectric peak, since 1σ corresponds to 68% of a normal distribution and only the lower tail is rejected (84%=68%/2+100%/2). It is important that the imaging system has a good energy resolution because it allows the discrimination of photons that have undergone Compton scattering either in tissues or in the collimator.

The efficiency of the CsI-scintillator is derived from the total absorption μ_T/ρ of CsI-crystals obtained from NIST-database (Berger 05) and given in Table 2.1. Both photoelectric attenuation coefficients μ_{PE} and the total attenuation coefficients are provided in order to show that at 511 keV, a large amount of interactions are not photoelectric interactions. The total detection efficiency of a 4 mm-thick CsI-crystal at 511 keV is 16%, but it drops to 5.8% when only photons which deposit more than 472 keV (3σ of the photoelectric peak) are taken into account. This result is obtained from a Monte-Carlo simulation with an energy threshold set to 472 keV and an energy resolution set at 5.8% at 511 keV.

	rank	step (mm)	thickness (mm)	collimation efficiency at 511 keV	number of holes	opening fraction
mask 1	9	1.26	4	64%	256	50%
mask 2	9	1.26	6	79%	256	50%
mask 3	6	1.85	12	95%	544	50%

Table 2.2 – Characteristics of the available HURA coded-aperture collimators

Figure 2.3 – Rank-6 HURA tungsten collimator: thickness: 12 mm, hole-to-hole distance: 1.85 mm

2.1.2.2 Coded-aperture collimator

A large choice of coded apertures is available and they are presented in section 1.4.2. As the γ-imager used for measuring the input function is made of a circular scintillator, it is decided to use a hexagonal pattern in order to maximise the overlap of the collimator shadow on the scintillator. A few collimators were already designed in our laboratory (Gmar 04), and the manufacture of coded-aperture collimators being complex and expensive, it is chosen to use one of the available collimators. They were designed for photons with energy ranging from 662 keV to 1.25 MeV. Table 2.2 presents the characteristics of the 3 available masks. They are all made of tungsten, whose linear attenuation coefficients are given in table 2.1. The first mask is the most appropriate to perform imaging at 511 keV since it is the thickest. Indeed the collimator should be thick so that the collimation is efficient. 12 mm of tungsten stops more than 95% of the impinging photons at 511 keV.

In practice, due to considerations explained in section 3.5, the mask modelled in this thesis is a 9 mm-thick tungsten collimator instead of 12 mm-thick, so that results can be validated experimentally. The chosen mask is a rank-6 HURA whose characteristics are described in section 1.4.2.2. The inter-motif distance is 1.85 mm. Fig. 2.3 shows such a HURA collimator.

2.1.3 Imaging system design

As a single γ-imager is available for acquiring images at SHFJ department, it is not possible to develop a coincidence system to collect photons emitted from vessels and tissues. This section defends that a SPECT acquisition system with a coded-aperture collimator is able to achieve performances close to the ones of a coincidence system. Furthermore, the development of a SPECT system is cheaper than the development of a coincidence system, as a PET dedicated to small animals or a set of 2 bidimensional detectors as the γ-imager presented in Fig. 2.4.

The efficiency of the SPECT imaging system with a coded-aperture collimator is compared to the efficiency of a coincidence system made of 2 γ-imagers. Let us consider a coincidence system based on 2 γ-imagers as in Fig. 2.4. The efficiency ε_{coinc} of such a system can be approximatively given as a function of the geometric efficiency ε_G, the efficiency of the tissues ε_{att} and the detection efficiency ε_{dtc} by Eq. (2.6). The detection efficiency ε_{dtc}

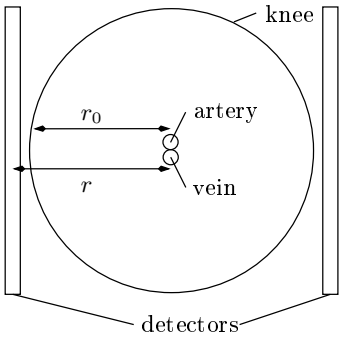


Figure 2.4 – A coincidence system based on 2 γ-imagers

is set to 5.8% which corresponds to the discrimination of photons which deposit more than 472 keV in the scintillator, as explained in the previous section.

$$\varepsilon_{coinc} = \varepsilon_G \cdot (\varepsilon_{att} \cdot \varepsilon_{dtc})^2 \tag{2.3}$$

$$= \frac{2 \times S_{dtc}}{4\pi r_0^2} \cdot \left(e^{-\mu_T \cdot r_0} \cdot \varepsilon_{dtc}\right)^2 \tag{2.4}$$

$$= \frac{2 \times \pi 50_{dtc}^2}{4\pi 50^2} \cdot \left(e^{-0.017 \cdot 50} \cdot 5.8\%\right)^2 \tag{2.5}$$

$$= 3.1 \times 10^{-4} = 3.1‰ \tag{2.6}$$

where S_{dtc} is the surface of each detector, r is the distance from the centre of the knee to the detectors, r_0 is the thickness of tissues crossed by the photon flux and μ_T is the total linear attenuation coefficient of the tissues. The efficiencies are respectively $\varepsilon_G = 50\%$, $\varepsilon_G = 43\%$, and $\varepsilon_{dtc} = 5.8\%$. The factor 2 in the geometric efficiency ε_G comes from the fact that 2 photons are emitted in opposite directions. The last term is squared because the detection probability of the two detectors are independent and a coincidence is registered only if both detectors detect a photon. In this example, $r = r_0$.

The efficiency ε_{SPECT} of a SPECT system based on a single γ-imager is given by Eq. 2.7:

$$\varepsilon_{SPECT} = \varepsilon_G \cdot \varepsilon_c \cdot \varepsilon_{att} \cdot \varepsilon_{dtc} \tag{2.7}$$

$$\varepsilon_{SPECT} = \frac{2 \cdot S_{dtc}}{4\pi r^2} \cdot \varepsilon_c \cdot e^{-\mu_T r_0} \cdot \varepsilon_{dtc} \tag{2.8}$$

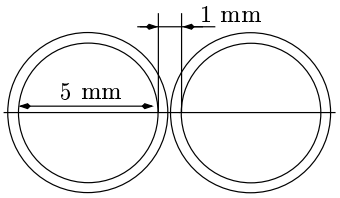
By setting the collimator efficiency ε_c at 50%, it is possible to derive the limit distance r so that $\varepsilon_{SPECT} \geq \varepsilon_{coinc}$:

$$r = r_0 \sqrt{\frac{\varepsilon_c}{\varepsilon_{att} \cdot \varepsilon_{dtc}}} \tag{2.9}$$

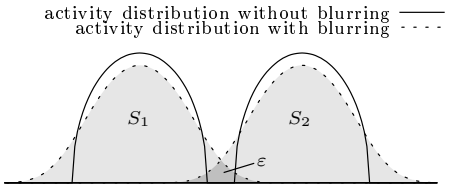
$$r = 50 \sqrt{\frac{50\%}{43\% \cdot 5.8\%}} \tag{2.10}$$

$$\leq 224 \text{ mm} \tag{2.11}$$

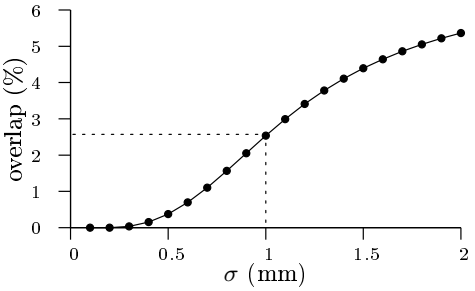
Contrary to classical high-energy collimators, most of the coded apertures are 50% transparent which increases the efficiency of the imaging system. Coded collimators are chosen instead of a random one because it was observed (Accorsi 01) that coded apertures reconstruct spatial distributions which are less noisy than other collimators. As far as efficiency is concerned, it is better to use the γ-imager in a SPECT-system than in a coincidence system as long as the γ-imager is placed at a distance smaller than 225 mm from the centre of the



(a) 2 cylinders



(b) Projected activity distribution on the line shown in Fig. 2.5a, with and without blurring



(c) Overlap of the first cylinder into the second as a function of the spatial resolution

Figure 2.5 – Impact of the spatial resolution of the imaging system

knee, as deduced from Eq. (2.11). However, quality of information collected from such a coincidence system is better than from such a SPECT system, despite the coded-aperture collimator, since coincidence systems provide directly lines of responses where the photon has been emitted from.

The spatial resolution of the imaging system must be good enough so that the artery and the vein can be spatially separated in the reconstructions. Let's assume that two 5 mm-large cylinders are placed one next to the other as in Fig. 2.5a. The distance between these two cylinders is 1 mm as the artery thickness, also called the intima-media thickness (IMT), is about 0.5 mm at the femoral artery (Vriend 06). Because of the spatial resolution of the imaging system, these two distributions overlap, as shown in Fig. 2.5b. The overlap coefficient between the two distributions is calculated according to:

$$\text{overlap}(\%) = 100 \times \frac{\varepsilon}{\varepsilon + S_1 + S_2} \tag{2.12}$$

where ε is the dark-grey area in Fig. 2.5b and S_1 and S_2 are the left and right light-grey areas. The spatial blurring is performed with a Gaussian kernel whose standard deviation is denoted σ. The overlap of the two distributions depends on the standard deviation σ of the kernel, as shown in Fig. 2.5c. According to this plot, it is chosen that the spatial resolution of the imaging system should be at maximum σ=1 mm. In this case, the overlap is 2.6%. Such a spatial resolution can be achieved if the γ-imager and the coded mask are placed so that the magnification of the imaging system compensates for the limited spatial resolution of the γ-imager. In section 2.1.2.1, it was assumed that the spatial resolution of the imager is 2 mm at 511 keV. As it is expected that the spatial resolution of the imaging system is 1 mm so that it is possible to separate the artery from the vein as mentioned earlier in this section, this means that a magnification factor of 2 is needed. By placing the collimator in contact to the skin of the knee, the distance between the vessels and the collimator is about 50 mm. This means that the γ-imager should be placed 150 mm away from the centre of the knee as it is illustrated in Fig. 2.6. This limits the field of view that can be imaged as

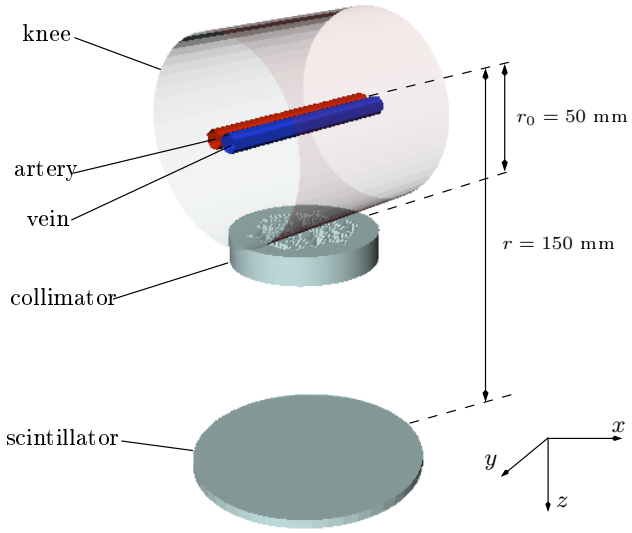


Figure 2.6 – Configuration of the imaging system

it is explained in section 1.4.1.1. From the central hole of the coded-aperture collimator, it is derived from Eq. (1.34) that the dimension of the field-of-view of this imaging system is 50 mm for a magnification factor 2. The efficiency of the imaging system is then derived from Eq. (2.13):

$$= \frac{\pi \cdot 50^2}{4\pi \cdot 150^2} \cdot 0.5 \cdot 43\% \cdot 5.8\% = 6.9 \times 10^{-4} = 6.9‰ \tag{2.13}$$

This efficiency is more than twice better than the efficiency of the coincidence system in Eq. (2.6), which justifies the choice of a SPECT-imaging system rather than a coincidence imaging system.

2.2 System matrix calculation

2.2.1 Limitations of correlation-based reconstruction method

With the proposed geometry, the far-field approximation is not valid anymore. In addition the collimation of 511-keV γ-rays requires thick collimators. In that condition, correlation-based algorithms fail.

Furthermore, when coded-aperture collimators are thick, there is an ambiguity on the decoding depth as illustrated in Fig. 2.7. Indeed the magnification of the projection of the object slice o_1 through the collimator slice c_1 is equal to the magnification of the projection of the object slice o_2 through the collimator slice c_2. Contrary to the case where the collimator is thin, the depth of the spatial distribution can not be estimated from the scaling factor that is applied to the decoding matrix before decoding the projection (Cannon 79). Moreover, every collimator slice is partially transparent, which degrades the coding of the signal on the detector. As decoding algorithms which are usually used in coded-aperture imaging suffer from limitations for high-energy photons, it is suggested to take advantage of a statistical algorithm such as MLEM which is presented in section 1.5.1 instead of correlation-based decoding algorithms. Such an algorithm extracts available information in signal in an optimal

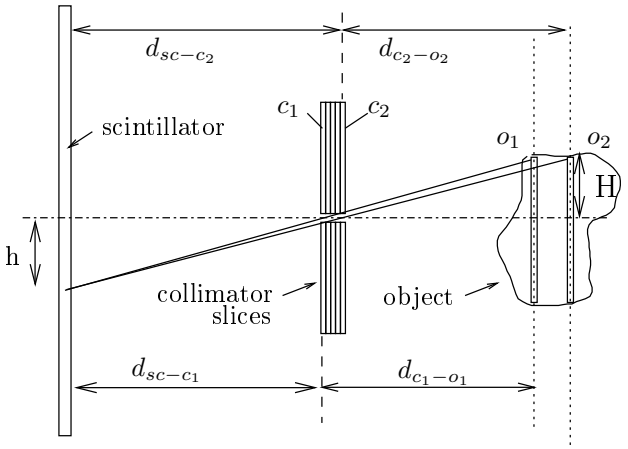


Figure 2.7 – Ambiguity on the decoding depth

way. In particular, Poisson fluctuations can be modelled in statistical algorithms, which is not possible with correlation algorithms. This is especially relevant when the activity of objects to be imaged is low. Some investigations have already been performed in this direction (Berrim 96; Groiselle 00), but the open fraction of the coded aperture that was used is lower than 1%. Although such a collimator increases the efficiency of the imaging system, it does not allow the estimation of the input function because the amount of detected photons is not sufficient.

Such an algorithm requires that the imaging system is modelled with a system matrix that allows the computation of the expected observations in detectors when any object is placed in the field of view of the imaging system. The calculation of the transition matrix is fundamental in the reconstruction process. Indeed, the better it models the imaging system, the better reconstructions are. Every coefficient $a_{i,j}$ of this matrix represents the probability that a photon which is emitted from a voxel v^j is detected in a pixel p^i. It depends on geometrical parameters such as the solid angle between the voxel and the pixel and the distance between them, as well as on physical parameters such as the attenuation and the scattering of photons in tissues and in the collimator, and the detection efficiency of the detector.

In this section, Monte-Carlo simulations are first used to investigate the computation of a system matrix, then an analytical derivation is proposed to speed up the computations, but requires that some physical phenomena are neglected.

2.2.2 Monte-Carlo simulations

As Monte-Carlo simulations are time-consuming, it is proposed to investigate the properties of the estimation of system matrices by investigating the properties of the estimation of only one of its column. When the system matrix models an object which is made of N voxels, this decreases the computation time by a factor N. The n^{th} column of the system matrix $\boldsymbol{A}$ models the expected projections which would be obtained for photons emitted from the n^{th} voxel of the object.

Monte-Carlo simulations are performed with the GATE software, version 2.2. Simulations are performed on a single 2.4 GHz 64-bit processor.

2.2.2.1 Configuration of the simulation

The configuration of the simulation is the following:

- The centre of the collimator is located at $(0, 0, 0)$. The axis of the collimator is oriented towards the z-direction. The HURA collimator that is used is made of tungsten and is 9 mm-thick as proposed in section 2.1.2.2.

- The centre of the detector is located at $(x, y, z) = (0\text{ mm}, 0\text{ mm}, 100\text{ mm})$ and the normal of its surface is oriented towards the z-direction. It is 4 mm-thick. The properties of the scintillator (dimensions, material, energy and spatial resolutions) are set to model the imager described in the section 2.1.2.1. The imaging system is assumed to have a perfect intrinsic spatial resolution in order to quantify limit performances of coded-aperture collimators. The energy resolution is set to 6% at 511 keV as justified in section 2.1.2.1.

- All particles were emitted from the voxel located at the centre of the object ($(X, Y, Z) = (0\text{ mm}, 0\text{ mm}, -50\text{ mm})$). This voxel is a box whose dimensions are $1 \times 1 \times 1$ mm^3. 2.1×10^9 photons are emitted.

- Compton scattering is taken into account and it is chosen that only singles with an energy above 485 keV contribute to the projection, which corresponds to 95% of the photoelectric peak, according to Eq. (2.14):

$$2\sigma \;=\; 2 \cdot \frac{FWHM}{2\sqrt{2\ln 2}} \;=\; 2 \cdot \frac{6\% \cdot 511\ keV}{2.35} \;=\; 26.1\text{ keV } (=\; 511 - 485) \tag{2.14}$$

- Emitted particles are photons whose energy is 511 keV. This means that the distance between the location of the emission of positrons and the location of their annihilation is not taken into account. It is chosen to do so in order to emit particles in a cone in order to accelerate computations. With the configuration described above, the detector completely stands inside a cone with a half-angle being lower than 20°, according to:

$$\arctan\left(50/(50+100)\right) = 18.4^\circ \tag{2.15}$$

The acceleration factor is then given by:

$$\left(\frac{1}{4\pi}\int_{\varphi=0^\circ}^{2\pi}\int_{\theta=0^\circ}^{20^\circ}\sin\theta \mathrm{d}\theta\mathrm{d}\varphi\right)^{-1} = \left(\frac{1}{2}(1-\cos(20^\circ))\right)^{-1} \tag{2.16}$$

$$= 33.2 \tag{2.17}$$

2.2.2.2 Results

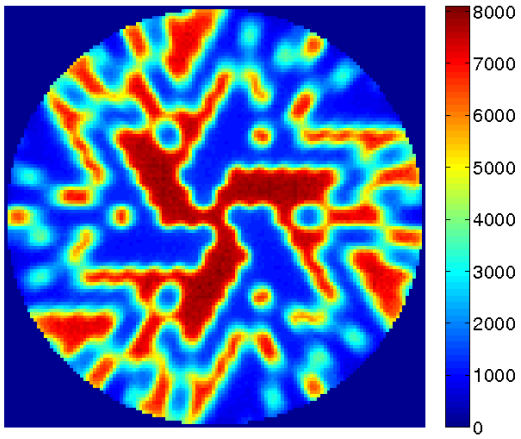
It takes 100 hours to obtain the projection of Fig. 2.8a. This projection is a (128×128)-pixel image. It presents the number of counts detected in every pixel. From this figure, it is possible to deduce the probability a_{ij} that a photon emitted from the voxel is detected by every pixel of the detector, according to Eq. (2.18). The map of detection probabilities is given in Fig. 2.8b.

$$\text{prob} = \frac{\text{counts}}{2.1 \times 10^9} \cdot \frac{1}{2}(1 - \cos(20^\circ)) \tag{2.18}$$

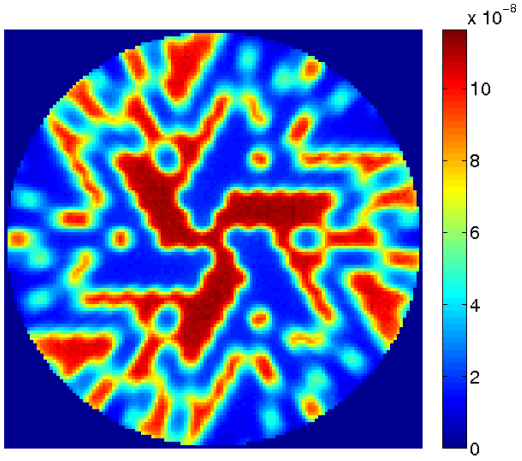
On the other hand, as counts are Poisson-distributed, it is possible to estimate the error which is associated with every pixel from Fig. 2.8a and Eq. (2.19). Errors for the whole image is presented in Fig. 2.8c according to:

$$\text{Error} = \frac{1}{\sqrt{\text{counts}}} \tag{2.19}$$

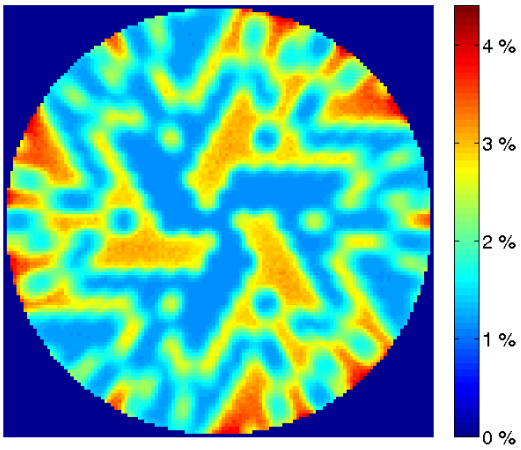
Fig. 2.8d shows the distribution of the error in the pixels. It is lower than 4%. The error decreases as a quadratic function of the number of detected photons, which means that an improvement on the error by a factor 2 (the error is then below 2%) requires simulations with 4 times more photons, which increases the computation time by a factor 4.



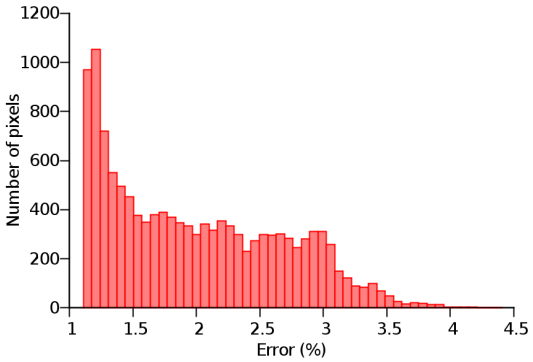
(a) Detected photons



(b) Probability map



(c) Error on the number of detected photons



(d) Distribution of statistical errors

Figure 2.8 – Derivation of the probability map and the associated error from a Monte-Carlo projection. Fig. 2.8b represents a single column of the system matrix $\boldsymbol{A}$

Such kind of simulations should be performed for every voxel of the reconstruction volume so that a system matrix can be built. This means that for a $(21\times21\times21)$-voxel object (it does not cover the whole field-of-view), 9.5×10^5 hours (105 years) would be required. Computing a whole system matrix with Monte-Carlo simulations would thus require huge computation resources. However, future investigations could take advantage of recent accelerations on Monte-Carlo simulations ((Martineau 09)).

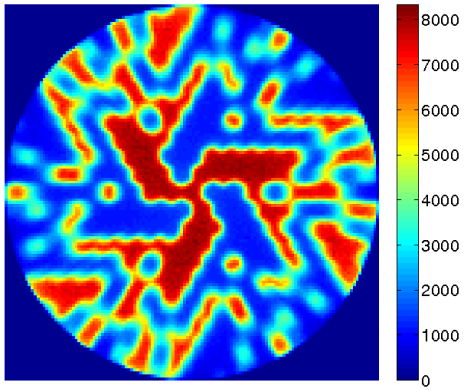
2.2.3 Analytical determination of system matrices

As it takes long to compute accurately system matrices through Monte-Carlo simulations, it is proposed to investigate whether this computation could be performed through analytical

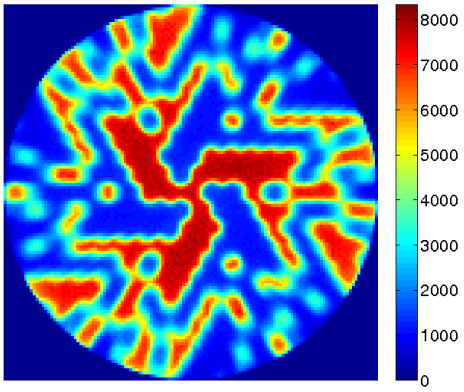
methods. This section explores this issue.

2.2.3.1 Hypotheses

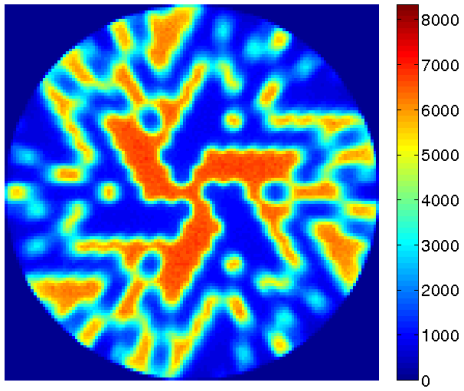
The computation of system matrices through analytical methods implies that some hypothesis are satisfied. First, the analytical determination of system matrices can only modelled the photoelectric effect and possibly first scattering of photons through Compton effect, not all Compton scattering. As a result, it is decided to compare results which are obtained in next sections with Monte-Carlo simulations where Compton scattering is inactive. Consequently, the linear attenuation coefficients should be set to the photoelectric contributions σ_{PE} in the analytical determinations, not to the total contributions σ_T.



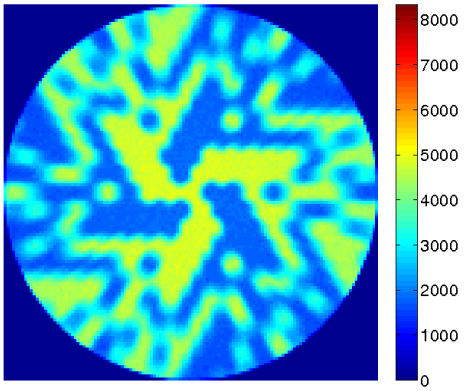
(a) The energy threshold is set to 472 keV (3σ)



(b) The energy threshold is set to 485 keV (2σ)



(c) The energy threshold is set to 498 keV (1σ)



(d) Compton scattering is not taken into account. No energy threshold is used

Figure 2.9 – Dected photons from simulation for 3 different energy cut-off (2.9a, 2.9b, 2.9c) and without Compton scattering (2.9d)

However, such an hypothesis results in a systematic error, which corresponds to the difference between the Monte-Carlo simulations with Compton scattering being taken into account or not. Fig. 2.9 shows such Monte-Carlo simulations for 3 energy cuts. The energy cuts for Fig. 2.9a, 2.9b and 2.9c are respectively 472 keV (3σ), 485 keV (2σ) and 498 keV (1σ), which means that only photons which deposit in the scintillator more energy than these thresholds are registered in the projections. Fig 2.9d shows the result of the same simulation with no Compton scattering, the threshold being set to 472 keV. It can be seen that the

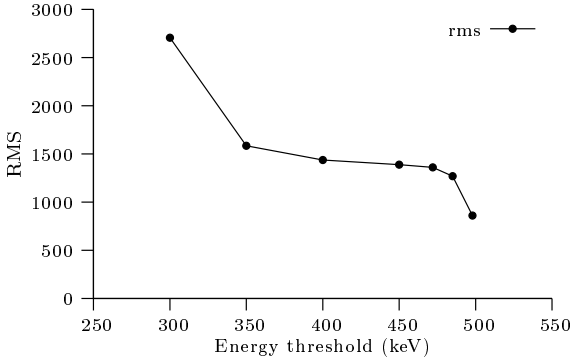


Figure 2.10 – Root mean square between the simulation without Compton scattering and simulations where Compton scattering is taken into account, for various energy cuts

projection which is the most similar to this last projection is the one with the energy threshold set at 498 keV, the mean factor between these two projections being 0.90. It is then confirmed by the root-mean-square (RMS) representation in Fig. 2.10. The RMS is calculated according to:

$$RMS(E) = \sqrt{\frac{1}{N}\sum_{i,j}\Big(I_{PE} - I_{PE+C}(E)\Big)^2} \tag{2.20}$$

where I_{PE} is the projection with no Compton scattering, in Fig. 2.9d. $I_{PE+C}(E)$ is the projection that takes into account Compton scattering with the energy threshold being set at E. N is the number of pixels in the circular portion of these projections. From the plot of Fig. 2.10, it is deduced that the energy threshold should be set close to the value of the photoelectric peak in order to minimise the impact of Compton scattering, which can not be fully modelled analytically. It can be noticed that the number of counts in the opaque portion of the collimator (blue) is higher for Fig. 2.9d than for the three other figures, although no Compton scattering is taken into account. This is due to the energy threshold in the three first figures which rejects photons which interact with the scintillator through photoelectric effect, whereas no threshold is used in Fig. 2.9d.

2.2.3.2 Theory: analytical formulation of system matrices

In this section, an analytical expression to compute every coefficient a_{ij} of the system matrix $\boldsymbol{A}$ is developped. It offers an alternative to Monte-Carlo simulations.

From Fig. 2.11, the probability that a photon emitted from A, which is located at $\boldsymbol{r}^j = (0,0,0)$, is detected in a volume element around B located at $\boldsymbol{r}^i = (r,\theta,\varphi)$ is given in spherical coordinates by:

$$\mathrm{d}P = \frac{r^2 sin\theta \mathrm{d}\theta \mathrm{d}\varphi}{4\pi r^2} \cdot \mu_d dr \cdot e^{-\mu_c l_c(\theta,\varphi)} \cdot e^{-\mu_d (r - z_{in}/\cos\theta)} \tag{2.21}$$

where notations are presented in Fig. 2.11 and can be interpreted as follows:

- $\dfrac{r^2 sin\theta \mathrm{d}\theta \mathrm{d}\varphi}{4\pi r^2}$ accounts for the geometric efficiency
- $\mu_d \,\mathrm{d}r$ is the probability that a photon which enters the volume element $\mathrm{d}V$ is detected
- $e^{-\mu_c l_c(\theta,\varphi)}$ is the probability that the photon passes through the collimator

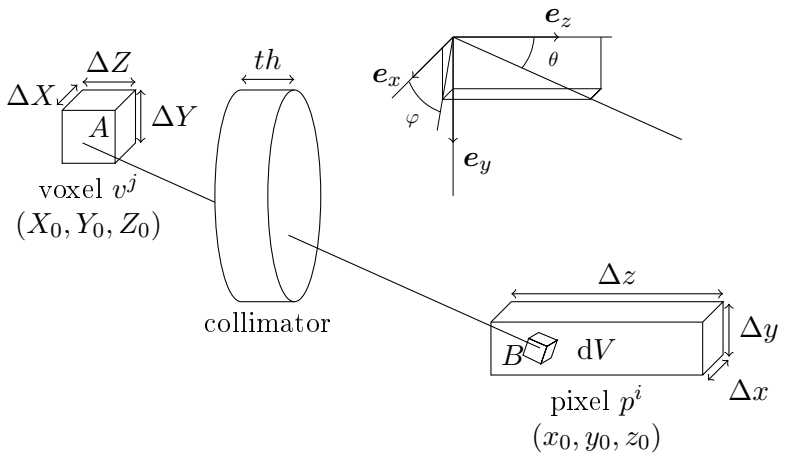


Figure 2.11 – Ray starting from the voxel v^j and being detected in the pixel p^i after crossing the collimator

- $e^{-\mu_d(r-z_{in}/\cos\theta)}$ is the probability that a photon which enters the detector reaches the volume element $\mathrm{d}V$
- z_{in} is the z-coordinate of the entrance surface of the detector
- $l_c(\theta, \varphi)$ is the length which is crossed by photons in the collimator as illustrated in Fig. 2.12.

A change of variables leads to:

$$\mathrm{d}P = \frac{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z}{4\pi(x^2+y^2+z^2)} \cdot \mu_d \cdot e^{-\mu_c th(x,y,z)\sqrt{x^2+y^2+z^2}/z} \cdot e^{-\mu_d(z-z_{in})\sqrt{x^2+y^2+z^2}/z} \tag{2.22}$$

This is the exact probability that a photon coming from A is detected in a volume element $\mathrm{d}V$ in the scintillator. Then, the probability that a photon emitted from a voxel v^j is detected in a pixel p^i is deduced in Eq. (2.24). The coordinate of a point A which stands in a voxel v^j is denoted $\boldsymbol{r}^j = (X, Y, Z)$ and the coordinate of a point B in a pixel p^i is denoted $\boldsymbol{r}^i = (x, y, z)$:

$$a_{i,j} = \iiint\limits_{(X,Y,Z)\in v^j} \frac{\mathrm{d}X\mathrm{d}Y\mathrm{d}Z}{\Delta X\Delta Y\Delta Z} \iiint\limits_{(x,y,z)\in p^i} \frac{\mu_d\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \cdot e^{-\mu_c l_c(x,y,z,X,Y,Z)}}{4\pi((X-x)^2+(Y-y)^2+(Z-z)^2)} \times e^{-\mu_d(z-z_{in})\sqrt{(X-x)^2+(Y-y)^2+(Z-z)^2}/(Z-z)} \tag{2.23}$$

$$= \iiint\limits_{\boldsymbol{r}^j\in v^j} \frac{\mathrm{d}X\mathrm{d}Y\mathrm{d}Z}{\Delta X\Delta Y\Delta Z} \iiint\limits_{\boldsymbol{r}^i\in p^i} \frac{\mu_d \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\cos^2\left(\theta(\boldsymbol{r}^i,\boldsymbol{r}^j)\right)}{4\pi(Z-z)^2} \cdot e^{-\mu_c l_c(\boldsymbol{r}^i,\boldsymbol{r}^j)} \cdot e^{-\mu_d(z-z_{in})/\cos\left(\theta(\boldsymbol{r}^i,\boldsymbol{r}^j)\right)} \tag{2.24}$$

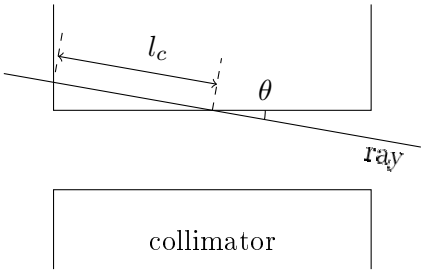


Figure 2.12 – Details on intersection between a ray and the collimator

2.2.4 Approximations for the analytical computation of system matrices

It is proposed to perform some approximations in order to compute Eq. (2.24) efficiently. As the integral in Eq.2.24 can not be derived analytically, it is proposed to perform some approximations to efficiently compute it. In this section, a point that is in a voxel v^j is located in $\{X_j \pm \Delta X/2; Y_j \pm \Delta Y/2; Z_j \pm \Delta Z/2\}$ and a point that stands in a pixel p^i is located in $\{x_i \pm \Delta x/2; y_i \pm \Delta y/2; z_i \pm \Delta z/2\}$. z_{in} is equal to $(z_i - \Delta z/2)$.

2.2.4.1 Derivation of probability $a_{i,j}$

From Fig. 2.13 some approximations can be done to compute Eq. 2.24. The first approximation is to assume that $\cos\theta$ is almost constant. $\cos\theta$ is calculated for every vertex of voxels $\boldsymbol{r}_v^j$ and every vertex of pixels $\boldsymbol{r}_p^i$. Neighbour values are compared and the maximum relative difference is less than 1 % (0.95%) for the imaging configuration described in section 2.1.3. This approximation is a zero-order approximation on $\theta = \theta_{ij} + \theta_h$, with θ_{ij} being the angle between the centre of a voxel v^j and the centre of a pixel p^i as shown in Fig. 2.13, Eq. (2.26) being the zero-order Taylor series expansion of the terms depending on θ. This justifies the approximation.

$$\cos^2(\theta) \cdot e^{-(a/\cos\theta)} \approx \left[\cos^2(\theta_{ij}) - \theta_h(\sin(2\theta_{ij}) + a\sin(\theta_{ij}))\right] \cdot e^{-(a/\cos\theta_{ij})} \tag{2.25}$$

$$\approx \cos^2\theta_{ij} \cdot e^{-(a/\cos\theta_{ij})} \tag{2.26}$$

Eq. 2.24 becomes:

$$a_{i,j} = \iiint_{v^j} \iiint_{p^i} \frac{\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z}{\Delta X \Delta Y \Delta Z} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \frac{\mu_d\,\cos^2\theta_{ij}}{4\pi(Z-z)^2} \cdot e^{-\mu_c l_c(x,y,z,X,Y,Z)} \cdot e^{-\mu_d(z-z_i+\Delta z/2)/\cos\theta_{ij}} \tag{2.27}$$

The second approximation is to assume that the intersection length of the collimator and every ray which starts from a voxel v^j and ends in a pixel p^i is constant and

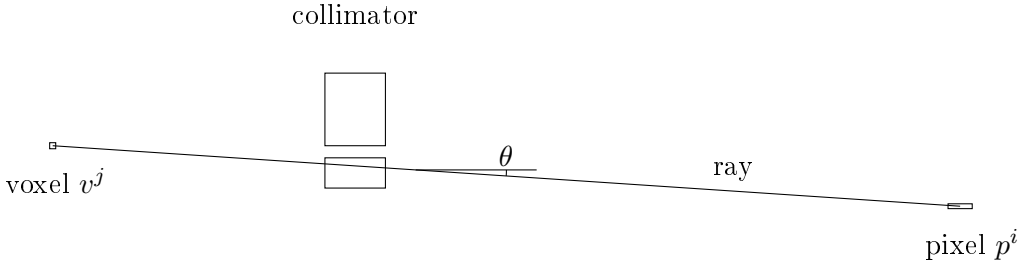


Figure 2.13 – 2D-view of the imaging system. Proportions are respected

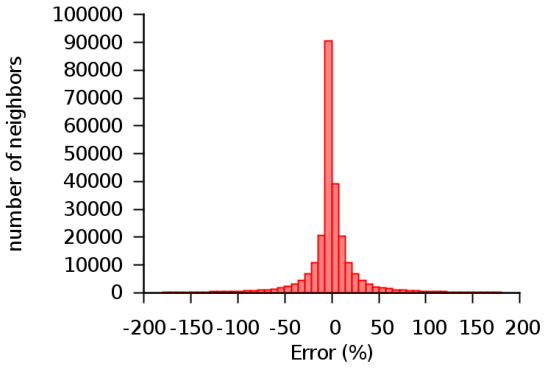


Figure 2.14 – Distribution of the relative difference between intersecting lengths of neighbour rays

its value is $l_c(x_i, y_i, z_i, X_j, Y_j, Z_j) = l_c(\boldsymbol{r}_j, \boldsymbol{r}_i)$. This approximation is justified by Fig. 2.14, which shows the distribution of the relative difference between neighbour rays. This distribution is derived from the computation of the intersecting lengths of an object which is $1 \times 1 \times 1$ mm^3 and made of $2 \times 2 \times 2$ voxels on the detector made of 100×100 mm^2 and 256×256 pixels. The intersecting length of any ray starting from the centre of a voxel of the object and ending in the centre of a pixel is stored in a tensor $L(nx, ny, i, j, k)$ where (n_x, n_y) are the indices of the concerned pixel and (i, j, k) are the indices of the concerned voxel. Two rays are said neighbours if their intersecting lengths with the collimator are stored in neighbour cells of the tensor L. For instance, the rays corresponding to $L(n_x, n_y, i, j, k)$ and $L(n_x + 1, n_y, i, j, k)$ are neighbours. With this definition, the rays corresponding to $L(n_x, n_y, i, j, k)$ and $L(n_x + 1, n_y, i, j, k + 1)$ are not neighbours, because oblique neighbourhoods are not considered. Any relative difference between the lengths of two neighbour rays l_1 and l_2 is computed according to Eq. (2.28):

$$\Delta_R(l_1, l_2) = 100 \cdot 2 \cdot \frac{l_1 - l_2}{l_1 + l_2}\% \tag{2.28}$$

Fig. 2.14 is built by computing all relative differences $\Delta_R(l_i, l_j)$ on neighbour lengths l_i and l_j and taking its histogram. This distribution is centred and its standard deviation is 32%, which justifies the second approximation.

By denoting $(Z_m, Z_M) = (Z_j - \Delta Z/2, Z_j + \Delta Z/2)$ and $(z_m, z_M) = (z_i - \Delta z/2, z_i + \Delta z/2)$, Eq. (2.27) becomes:

$$a_{i,j} = \frac{(\Delta x \Delta y)\mu_d \cos^2\theta_{ij}}{4\pi\Delta Z} e^{-\mu_c l_c(\boldsymbol{r}_j, \boldsymbol{r}_i)} \int_{Z=Z_m}^{Z_M} \int_{z=z_m}^{z_M} \mathrm{d}Z\, \mathrm{d}z \times \frac{e^{-\mu_d (z - z_i + \Delta z/2)/\cos\theta_{ij}}}{(Z - z)^2} \tag{2.29}$$

The third approximation is to assume that $(Z - z)$ is constant while Z and z vary respectively on a voxel v^j and a pixel p^i. Its mean value is $(Z_j - z_i)$. This approximation is justified by the fact that the standard deviation of all $(Z - z)$ is $\sigma = 1.2$ as given by Eq. 2.32. This corresponds to a relative difference of all $(Z - z)$ which is $\sigma/\mu = 7.9‰$. $(Z - z)$ can

consequently be considered as constant and equals to their mean $(Z_j - z_i)$.

$$\sigma = \sqrt{var(Z-z)} = \sqrt{var(Z) + var(z)} \text{ as } Z \text{ and } z \text{ are uncorrelated} \tag{2.30}$$

$$= \sqrt{\frac{1}{Z_M - Z_m} \int_{Z=Z_m}^{Z_M} (Z - Z_j)^2 \mathrm{d}Z + \frac{1}{z_M - z_m} \int_{z=z_m}^{z_M} (z - z_i)^2 \mathrm{d}z} \tag{2.31}$$

$$= 1.2 \tag{2.32}$$

with$(Z_m, Z_j, Z_M) = (98, 100, 102)$ and $(z_m, z_i, z_M) = (-50.5, -50, -49.5)$. var denotes the variance.

Eq. (2.29) becomes:

$$a_{i,j} = \frac{(\Delta x \Delta y)\mu_d \cos^2\theta_{ij}}{4\pi\Delta Z(Z_j - z_i)^2} \times e^{-\mu_c l_c(\boldsymbol{r}_j, \boldsymbol{r}_i)} \int_{Z=Z_m}^{Z_M} \int_{z=z_m}^{Z_M} \mathrm{d}Z\, \mathrm{d}z e^{-\mu_d(z - z_i + \Delta z/2)/\cos\theta_{ij}} \tag{2.33}$$

$$= \frac{(\Delta x \Delta y)\mu_d \cos^2\theta_{ij}}{4\pi(Z_j - z_i)^2} \times e^{-\mu_c l_c(\boldsymbol{r}_j, \boldsymbol{r}_i)} \left[-\frac{\cos\theta_{ij}}{\mu_d} e^{-\mu_d z/\cos\theta_{ij}} \right]_{z=z_m}^{z_M} \tag{2.34}$$

$$\boxed{a_{i,j} = \frac{(\Delta x \Delta y)\cos^3\theta_{ij}}{4\pi(Z_j - z_i)^2} \times e^{-\mu_c l_c(\boldsymbol{r}_j, \boldsymbol{r}_i)} \times \left(1 - e^{-\mu_d \Delta z/\cos\theta_{ij}}\right)} \tag{2.35}$$

This is a well-known result (Accorsi 01), where probability $a_{i,j}$ depends on:

- $\dfrac{\Delta x \Delta y}{4\pi((Z_j - z_i)/\cos\theta_{ij})^2}\cos\theta_{ij}$, the geometric efficiency
- $e^{-\mu_c l_c(\boldsymbol{r}_j, \boldsymbol{r}_i)}$, the attenuation due to the collimator
- $\left(1 - e^{-\mu_d \Delta z/\cos\theta_{ij}}\right)$, the detection probability of impinging photons by the detector

$(Z_j - z_i)/\cos\theta_{ij}$ is the centre-to-centre distance from the voxel to the pixel and $\Delta z/\cos\theta_{ij}$ is the intersection length between the scintillator and the ray passing through the centres of the voxel and the pixel. An example of detection probability map is shown in Fig. 2.15, where no collimator is taken into account, *i.e.* all $\left(l_c(\boldsymbol{r}_j, \boldsymbol{r}_i)\right)$ are set to zero.

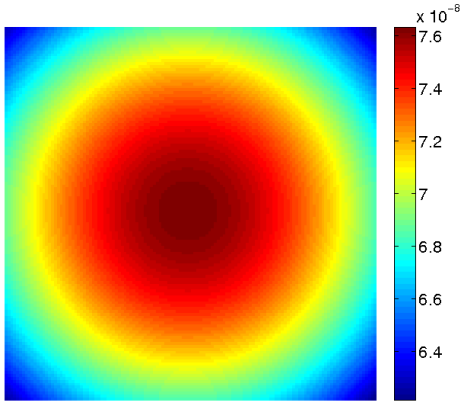


Figure 2.15 – Detection probabilities with no collimation according to Eq. 2.35

2.2.4.2 A more robust derivation of probability $a_{i,j}$

The previous section proposes a way to compute (a_{ij}) and the derived formula has a direct physical interpretation. It is possible not to use the third approximation, which leads to a result which is more accurate without increasing the computation time. The physical interpretation of the new formula is not as simple as the previous, and some part of this formula can not be interpreted physically. When the third approximation is not used, Eq. (2.29) can be written as:

$$a_{i,j} = \frac{(\Delta x \Delta y)\cdot \mu_d \cos^2\theta_{ij}}{4\pi\Delta Z} \cdot e^{-\mu_c l_c(\boldsymbol{r}_i, \boldsymbol{r}_j)} \cdot e^{\mu_d (z_i - \Delta z/2)/\cos\theta_{ij}} \times \int_{z=z_i-\Delta z/2}^{Z_j+\Delta Z/2} \mathrm{d}z\, e^{-\mu_d z/\cos\theta_{ij}} \int_{Z=Z_j-\Delta Z/2}^{Z_j+\Delta Z/2} \frac{\mathrm{d}Z}{(Z-z)^2} \quad (2.36)$$

For clarity, $\left(\frac{(\Delta x \Delta y)\cdot \mu_d \cos^2\theta_{ij}}{4\pi} \cdot e^{-\mu_c l_c(\boldsymbol{r}_i,\boldsymbol{r}_j)} \cdot e^{\mu_d (z_i-\Delta z/2)/\cos\theta_{ij}}\right)$ is denoted α. Integration on Z leads to:

$$a_{i,j} = \frac{\alpha}{\Delta Z} \cdot \int_{z=z_i-\Delta z/2}^{z_i+\Delta z/2} \mathrm{d}z\, e^{-\mu_d z/\cos\theta_{ij}} \left(-\frac{1}{Z_j+\Delta Z/2-z} + \frac{1}{Z_j-\Delta Z/2-z}\right) \tag{2.37}$$

$$= \frac{\alpha}{\Delta Z} \cdot \int_{z=z_i-\Delta z/2}^{z_i+\Delta z/2} \mathrm{d}z \frac{2\Delta Z/2}{(Z_j-z)^2-(\Delta Z/2)^2} \cdot e^{-\mu_d z/\cos\theta_{ij}} \tag{2.38}$$

$$= \alpha \int_{z=z_i-\Delta z/2}^{z_i+\Delta z/2} \mathrm{d}z \frac{e^{-\mu_d z/\cos\theta_{ij}}}{(Z_j-z)^2} \quad \text{because } \underbrace{(\Delta Z/2)^2}_{=(1/2)^2=0.25} \ll \underbrace{(Z_j-z)^2}_{\approx 150^2=22500} \tag{2.39}$$

Some changes of variables are performed in order to make appear incomplete gamma functions:

$$a_{ij} = \alpha \cdot e^{-\mu_d Z_j/\cos\theta_{ij}} \int_{z=z_i-\Delta z/2}^{z_i+\Delta z/2} \mathrm{d}z \frac{e^{-\mu_d (z-Z_j)/\cos\theta_{ij}}}{(Z_j-z)^2} \tag{2.40}$$

$$= \alpha \cdot e^{-\mu_d Z_j/\cos\theta_{ij}} \int_{w=w_1}^{w_2} dw \frac{e^{-\mu_d w/\cos\theta_{ij}}}{w^2} \tag{2.41}$$

$$\text{with } (w_1, w_2) = (z_i - Z_j - \Delta z/2; z_i - Z_j + \Delta z/2)$$

$$= \alpha \cdot \frac{\mu_d}{\cos\theta_{ij}} e^{-\mu_d Z_j/\cos\theta_{ij}} \int_{x=x_1}^{x_2} \mathrm{d}x\, e^{-x}\, x^{-2} \tag{2.42}$$

$$\text{with } (x_1, x_2) = \frac{\mu_d}{\cos\theta_{ij}}(z_i - Z_j - \Delta z/2; z_i - Z_j + \Delta z/2)$$

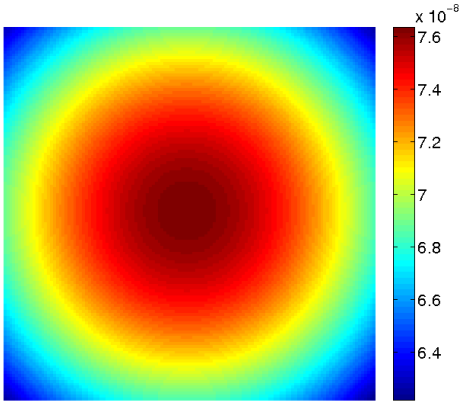
$$= \alpha \cdot \frac{\mu_d}{\cos\theta_{ij}} e^{-\mu_d Z_j/\cos\theta_{ij}} \left(\int_{x=x_1}^{+\infty} \mathrm{d}x\, e^{-x}\, x^{-2} - \int_{x=x_2}^{+\infty} \mathrm{d}x\, e^{-x}\, x^{-2} \right) \tag{2.43}$$

$$= \alpha \cdot \frac{\mu_d}{\cos\theta_{ij}} e^{-\mu_d Z_j/\cos\theta_{ij}} \Big(\Gamma(-1, x_1) - \Gamma(-1, x_2) \Big) \tag{2.44}$$

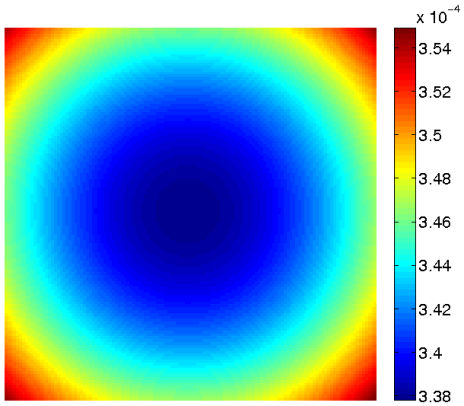
where Γ is the incomplete gamma function (Abramovitz 70). As a result, the probability that a photon emitted from a voxel v^j is detected in a pixel p^i is:

$$\begin{aligned} a_{i,j} = \; & \frac{(\Delta x \Delta y)\cdot \mu_d^2 \cos\theta_{ij}}{4\pi} \cdot e^{-\mu_c l_c(\boldsymbol{r}_i,\boldsymbol{r}_j)} \cdot e^{\mu_d (z_i - Z_j - \Delta z/2)/\cos\theta_{ij}} \\ & \times \left(\Gamma\left(-1, \tfrac{\mu_d}{\cos\theta_{ij}}(z_i - Z_j - \Delta z/2)\right) - \Gamma\left(-1, \tfrac{\mu_d}{\cos\theta_{ij}}(z_i - Z_j + \Delta z/2)\right) \right) \end{aligned} \tag{2.45}$$

A detection probability map, which is calculated according to Eq. (2.45), is shown in Fig. 2.16a. The difference between the terms a_{ij}, which is derived according to Eq. (2.35) and (2.45), is presented in Fig. 2.16b. Fig. 2.17 shows the impact of the parameters that

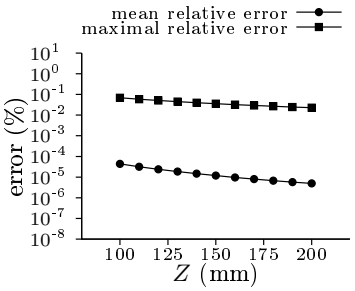


(a) Detection probabilities with no collimation and the new formula on a_{ij} given Eq. 2.45

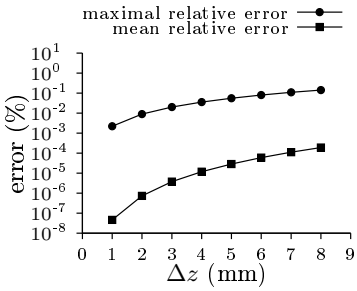


(b) Relative difference of the two methods. The new method impacts the estimation differently depending on the localisation

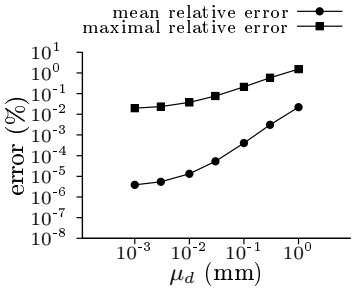
Figure 2.16 – Detection probabilities without the third approximation



(a) Influence of the distance between the voxel and the pixel



(b) Influence of the thickness of the scintillator

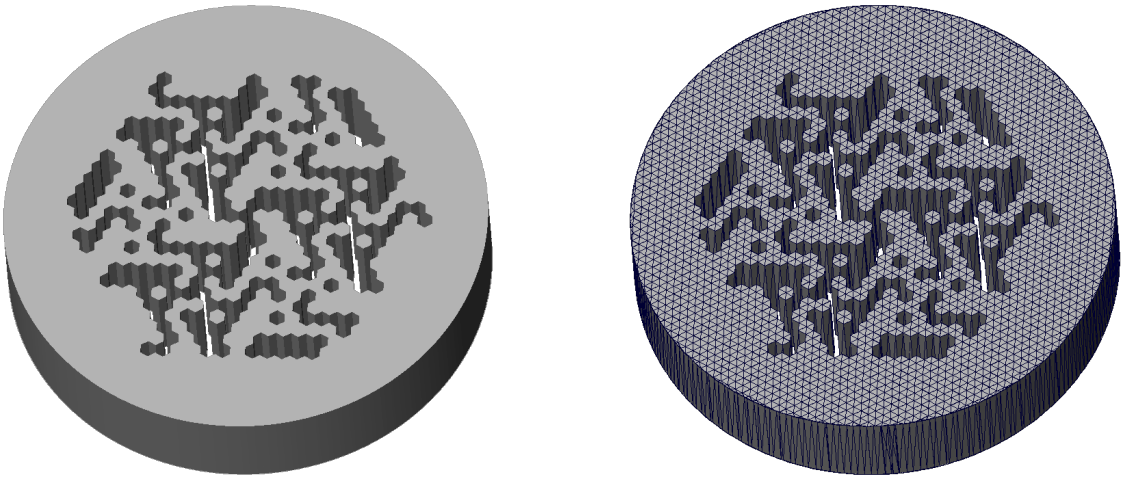


(c) Influence of the linear attenuation of the collimator

Figure 2.17 – Influence of parameters Z, Δz and μ_d on the derivation of the projection matrix

influence the derivation of the probabilities a_{ij}. The difference between the classical method of Eq. (2.35) and the new method given by Eq. (2.45) is negligible ($< 1\%$) for the configuration of the imaging system that is used in this thesis, but should be handled carefully while looking for increasing the sensitivity of the imaging system. Indeed, a better sensitivity can be achieved by either decreasing the distance from the object to the scintillator, or by increasing the thickness of the scintillator or by choosing a scintillator whose linear attenuation coefficient is high. They all amplify the relative difference between the classical and the new methods. Furthermore, it is observed that the new derivation of a_{ij} is twice longer to compute than the old derivation. For instance, it takes 360 ms to compute a (128×128) projection with Eq. (2.45) whereas it takes only 180 ms to compute it with Eq. 2.35.

Most of the computations of probabilities a_{ij} are fast, excepted the calculation of $l_c(\boldsymbol{r}_{i'}, \boldsymbol{r}_{j'})$. The collimator contributes through an exponential term. The computation of (a_{ij}) is implemented for every voxel v^j in two steps. First, $a_{i,j}$ is calculated for every pixel p^i without the contribution of the collimator. A first image is obtained. Then, all collimator contributions are calculated, which leads to a second image. Finally, both images are multiplied element-by-element as suggested by Eq. 2.45. The next section addresses methods to efficiently compute the contribution of the collimator to the detection probability map.



(a) HURA whose rank is 6 (b) Representation of the HURA with 12606 faces

Figure 2.18 – Representation of the collimator

2.3 Three different approximations and Results

In this section, three new ray-tracing methods are proposed to accelerate the computation time of system matrices. We call them face-driven ray tracing techniques, by analogy to ray-driven (Zeng 93) and pixel-driven methods (Zhuang 94). First, a few manipulations on the faces representing the coded-aperture collimators are performed in order to increase the performances of ray-tracing algorithms. Then a reference method which computes the contribution of the collimator to probabilities (a_{ij}) is presented. The new ray-tracing methods are then explained. The first method interpolates all intersection lengths from a short number of intersection lengths. The second method takes advantage of a simplified representation of the collimator. Finally the last method derives all intersection lengths by taking into account the successive contribution of every face.

The particularity of coded-aperture collimators is that a large number of triangles is required to model them. This is an issue when a large number of intersections has to be computed as for the computation of system matrices. In Eq. (2.35), all calculations of all terms are direct, excepted the one describing the collimator contribution $e^{-\mu_c l_c(\boldsymbol{r}_j,\boldsymbol{r}_i)}$, since the intersection length can not be known from the positions of the pixel and the voxel. It is usually computed through ray-tracing techniques. When the collimator is described by faces as shown in Fig. 2.18, intersections of every ray with all faces are calculated and intersection distances are deduced from intersecting points. Classical techniques (Beck 85; Tabary 07) are not adapted to coded-aperture collimators because the number of faces is very large (several thousands). Furthermore, 511-keV photons are very penetrating and consequently septa penetration can not be neglected. For instance, a 9 mm-thick tungsten shield only stops 90% of the impinging photons, the total linear attenuation coefficient at 511 keV being $\mu = 2.59$ cm^{-1}. As a result no portion of the collimator can be considered as opaque: all faces have to be taken into account. Computing intersections between the collimator and every ray crossing it is time-consuming. The computation time increases linearly with the number of pixels in the detector. Computing all the intersection lengths for the projection of a single voxel through the collimator made of 12606 faces shown in Fig. 2.18 onto a (128×128)-pixel detector takes about 13 s on a single 3.2 GHz 64-bit processor. As objects that are expected to be reconstructed are usually made of several thousands of voxels, the

total computation time is large. For instance, one of the matrices which is used in this thesis allows the reconstruction of objects made of $(30 \times 40 \times 30)$ voxels. With such a ray-tracing method, it takes more than 5 days to compute it. This is faster than Monte-Carlo simulations, but can still be improved.

2.3.1 Simplification of the representation of the coded-aperture collimator

The computation of the contribution of the collimator can be accelerated by simple considerations on the representation of the collimator.

2.3.1.1 Complementary aperture

We observed that the frame of the collimator can be obtained from the frame of the holes of the collimator. Indeed, intersection lengths of rays with the collimator that is shown in Fig. 2.18b can be deduced from the intersection lengths of rays with the collimator that is shown in Fig. 2.20a. The point is that the number of faces is smaller in this latter case, which speeds up the calculation of the system matrix $\boldsymbol{A}$. In Fig. 2.20a, the collimator is represented by 2150 vertices, 6408 edges and 4272 faces, which is about 3 times less faces than for the collimator in Fig. 2.18b. Intersection lengths which are calculated with the complementary collimator are denoted L_c. L_c is related to the intersecting lengths l_c of the collimator through:

$$L_c(\boldsymbol{r}_i, \boldsymbol{r}_j) = th_c / \cos\theta_{ij} - l_c(\boldsymbol{r}_i, \boldsymbol{r}_j) \tag{2.46}$$

where th_c is the thickness of the collimator and θ_{ij} is the incidence angle of the ray with the collimator as shown in Fig. 2.19. Eq. (2.49) provides a new equation for the computation of a_{ij}, which is only based on the intersecting lengths of the complementary collimator. ψ denotes all terms in Eq. (2.35) or Eq. (2.45) but the collimator contribution.

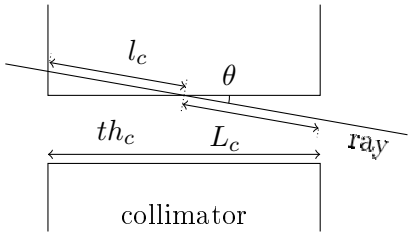
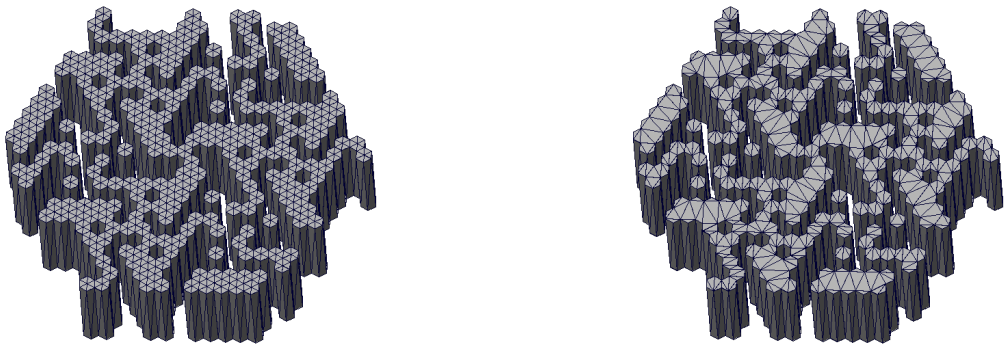


Figure 2.19 – Relation between l_c and L_c

$$\begin{aligned} a_{i,j} &= e^{-\mu_c l_c(\boldsymbol{r}_i, \boldsymbol{r}_j)}\psi & (2.47)\\ &= e^{-\mu_c (th_c/\cos\theta_{ij} - L_c(\boldsymbol{r}_i, \boldsymbol{r}_j))}\psi & (2.48)\\ &= e^{\mu_c L_c(\boldsymbol{r}_i, \boldsymbol{r}_j)} \left(e^{-\mu_c th_c/\cos\theta_{ij}}\psi\right) & (2.49) \end{aligned}$$



(a) The complementary collimator (4272 faces)

(b) The simplified complementary collimator (2876 faces)

Figure 2.20 – Representation of the complementary collimator. Fig. 2.20a is a basic triangulation of the complementary collimator. This triangulation is used to generate this collimator. Fig. 2.20b is an optimal Delaunay triangulation.

2.3.1.2 Delaunay simplification

The number of faces can be further reduced by representing the complementary collimator by a minimal number of faces. The result is presented in Fig. 2.20b where the complementary collimator is composed of 1452 vertices, 4314 edges and 2876 faces, which further decreases by 32 % the number of faces. Using this representation of the collimator speeds up the calculation of the system matrix $\boldsymbol{A}$ by 30%.

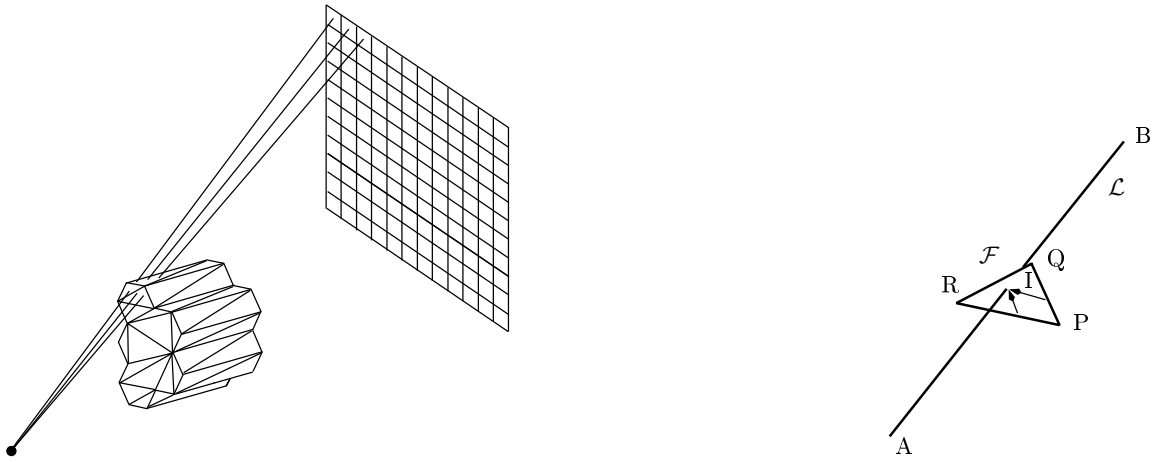
Although the computation time of system matrices can be accelerated by using complementary collimators, computing intersections is still very time-consuming. This slows down the computation of system matrices when the pattern of the collimators is made of a large number of faces. As much as possible, a short number of intersections should be computed.

2.3.2 Reference technique for computing intersection lengths

A reference technique is described in (Beck 85) to compute every l_c. This technique is tested and serves as a reference to evaluate the new techniques which are developed in this thesis.

2.3.2.1 Method

The reference technique for determining the intersecting lengths l_c is based on the computation of the intersecting points of every ray which starts from every voxel and ends in every pixel with every face of the collimator and to derive the intersecting lengths l_c from the intersecting points. Fig. 2.21a shows three rays starting from a unique voxel and ending in three different pixels. For an imaging system which is made of an object containing of $21 \times 21 \times 21$ voxels, a detector with 128×128 pixels and a collimator made of 2876 faces as in Fig. 2.20b, this results in the computations of $21^3 \times 128^2 \times 2876 = 4.4 \times 10^{11}$ intersection lengths l_c. It is obvious that the computation time linearly increases with the number of faces of the collimator, but also linearly increases with the number of pixels and with the number of voxels.



(a) Projection of the collimator frame on pixels

(b) Intersection of a line and a face

Figure 2.21 – Representation of rays crossing the collimator and the notations used to compute intersections

The intersection I between the face $\mathcal{F}$ and the line $\mathcal{L}$ is computed efficiently by solving the system according to the notations of Fig. 2.21b:

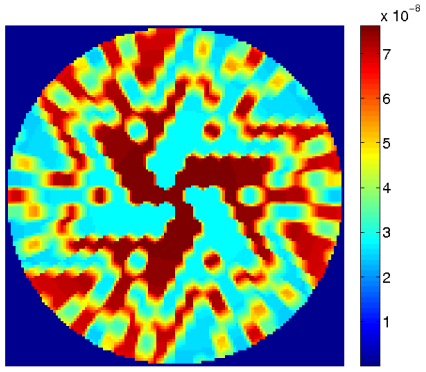
$$\begin{cases} I \in \mathcal{L} \\ I \in \mathcal{F} \end{cases} \Leftrightarrow \begin{cases} \overrightarrow{AI} = \lambda \cdot \overrightarrow{AB} \\ \overrightarrow{PI} = \alpha\overrightarrow{PQ} + \beta\overrightarrow{PR} \end{cases} \tag{2.50}$$

$$\Leftrightarrow \begin{cases} x_A + \lambda(x_B - x_A) = x_P + \alpha(x_Q - x_P) + \beta(y_Q - y_P) \\ y_A + \lambda(y_B - y_A) = y_P + \alpha(y_Q - y_P) + \beta(y_Q - y_P) \\ z_A + \lambda(z_B - z_A) = z_P + \alpha(z_Q - z_P) + \beta(y_Q - y_P) \end{cases} \tag{2.51}$$

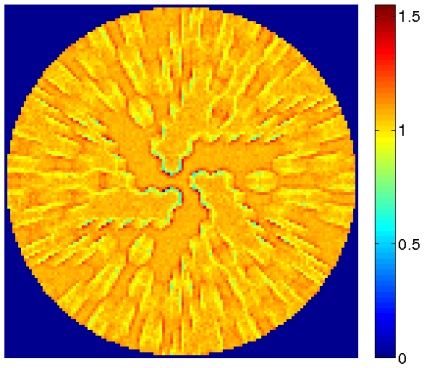
This is a linear system with 3 equations and 3 unknown variables (λ, α, β). As a result, it can be easily inverted and (λ, α, β) are retrieved. Because of the way α and β are defined in Eq. (2.50), the intersecting point I stands in the face $\mathcal{F}$ if $(\alpha, \beta) \in [0; 1]^2$ and $\alpha + \beta \leq 1$.

2.3.2.2 Results

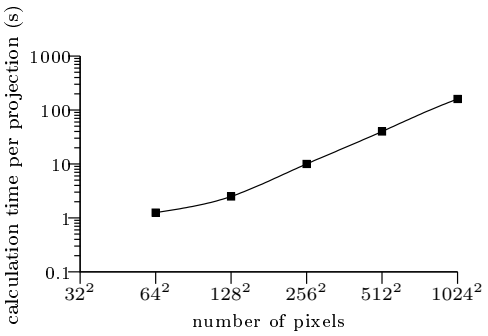
Fig. 2.22a presents the detection probabilities which are associated with the voxel located in $(0, 0, 0)$. These probabilities are derived from Eq. (2.45) with the intersecting lengths l_c computed according to Eq. (2.51). This probability map is compared to the Monte-Carlo simulation of Fig. 2.9d which is obtained without Compton scattering. The element-by-element ratio of these two probability maps is shown in Fig. 2.22b. Fig. 2.22d shows that the values spread around 1, as expected. However, it can be noticed that these two maps differ mostly on the edges of the collimators. The dispersion of the values of this image is represented in the histogram of Fig. 2.22d. The mean of this distribution is 1.057 and its standard deviation is $\sigma = 7.08 \times 10^{-2}$, which means that the probability map which is obtained with this ray-tracing method is biased by a factor 5.7%. The standard deviation tends to indicate that the result which is obtained through ray-tracing is close to the result which was obtained previously with Monte-Carlo simulations. The main difference is the computation time. Fig. 2.22c represents the time which is required to compute such probability maps, for a single voxel. It takes 2.5 s to compute a (128×128)-pixel projection. The computation times increase rapidly when the number of pixels increases, since Fig. 2.22c is represented in log-scales.



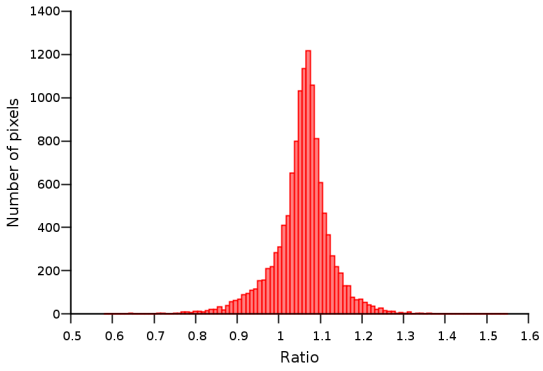
(a) Projection according to the reference ray-tracing method



(b) Ratio of projections of Fig. 2.22a and 2.9d



(c) Computation time



(d) Distribution of Fig. 2.22b

Figure 2.22 – Estimation of the probability map of a single voxel at 511 keV obtained with the reference technique and comparison with the GATE simulation of Fig. 2.8b(without scattering)

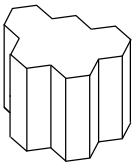
In conclusion, the reference analytical technique provides a detection probability map close to the map obtained using Monte-Carlo simulations. Next section presents a method using a smaller number of intersections for calculating system matrices. The number of projections does not depend on the number of pixels in the detector. This method takes advantage of interpolations to compute the whole system matrix.

2.3.3 First technique: interpolation-based ray-tracing

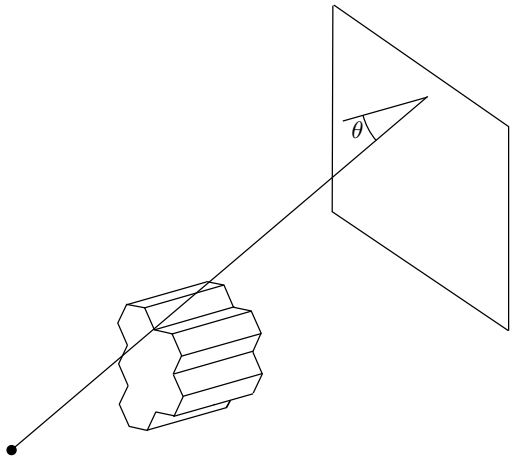
In this section, the ray-tracing method interpolates most of the coefficients from the computation of a limited number of intersection lengths. This aims at accelerating the computation time of system matrices, with minor deteriorations on the accuracy of the probabilities (a_{ij}).

2.3.3.1 Method

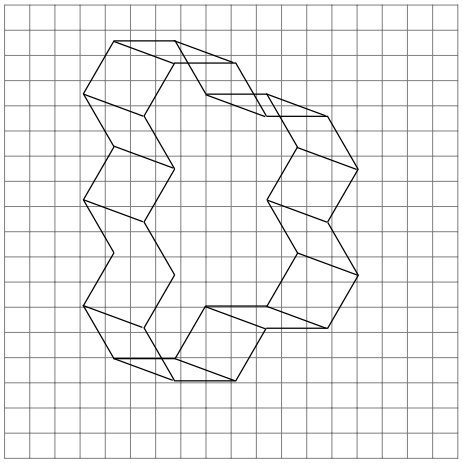
A collimator can be viewed as a surface made of edges. For instance, the object shown in Fig. 2.23a can be viewed as a piece of the collimator represented in Fig. 2.20b. Let's choose a voxel located at (X_j, Y_j, Z_j). The frame of the collimator is projected by the central projection whose centre is this voxel on the pixel plane ($z = z_i$) as shown in Fig. 2.23b. An example of a projection is shown in Fig. 2.23c. All the projected points and intersection points are used to build a Delaunay triangulation of this projection as shown in Fig. 2.23d. This tri-



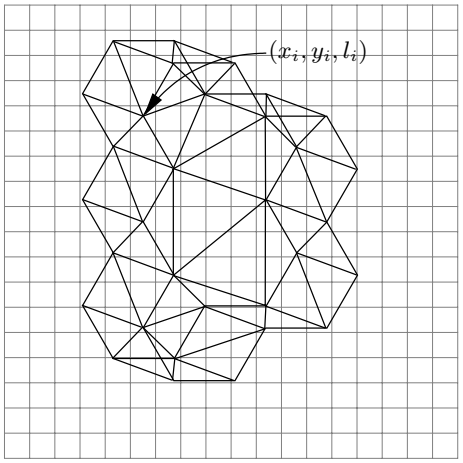
(a) Frame of a piece of a coded-aperture HURA collimator



(b) Projection of a point-source through a piece of collimator on a detector



(c) Projection of the collimator frame on pixels



(d) Constrained Delaunay triangulation of the projection of Fig. 2.23c

Figure 2.23 – Projection of the collimator frame on pixels

angulation is constrained, which means that all the edges which are present in Fig. 2.23c are also present in Fig. 2.23d. Vertices which are physical neighbours in Fig. 2.23a are still connected in Delaunay triangulation. Ray-tracing through the collimator is then performed on rays starting from (X_j, Y_j, Z_j) and ending on vertices of the constrained Delaunay triangulation. As shown in Fig. 2.23d, an intersecting length l_i is associated to every Delaunay node whose coordinates are (x_i, y_i). Let's choose one of the triangle. It usually covers several pixels as shown in Fig. 2.24. Intersecting lengths $l(x, y)$ for every position in the triangle are interpolated from a linear approximation. Indeed, (x_1, y_1, l_1), (x_2, y_2, l_2) and (x_3, y_3, l_3) constitute a plane $(\mathcal{P})$ which is a good approximation of the intersection distance for rays passing inside the triangle. The equation of the plane $\mathcal{P}$ is given by:

$$(\mathcal{P}): \quad a \cdot x + b \cdot y - l + c = 0 \tag{2.52}$$

As vertices of the triangle belong to $(\mathcal{P})$, the triplet (a, b, c) verifies:

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} - \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.53}$$

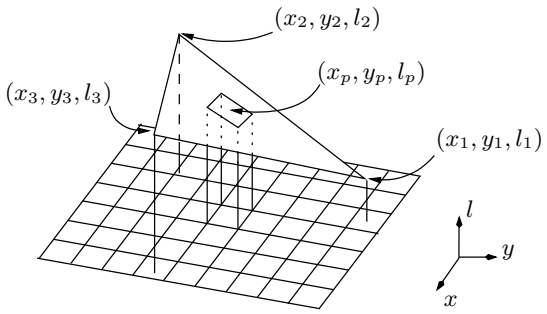


Figure 2.24 – The projection of a face $\mathcal{F}$ covers several pixels

(a, b, c) are then deduced according to:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \tag{2.54}$$

As a result, the intersecting length for every pixel p being in the shadow of the face $\mathcal{F}$ at (x_p, y_p) can be interpolated from an affine plane and the result is given by:

$$l_p(x_p, y_p) = a \cdot x_p + b \cdot y_p + c \tag{2.55}$$

$$= [x_p, y_p, 1] \cdot \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \tag{2.56}$$

The interpolation on a plane is motivated by the first-order approximation presented below. It is assumed that the values of the intersecting lengths l at extremities are known. The intersection length inside the triangle is a function of $1/\cos(\theta(x, y))$ as it was shown in Fig. 2.12. It is given by:

$$l(x, y) = \frac{\alpha}{\cos(\theta(x_i, y_i))} + \beta \tag{2.57}$$

Taylor expansion on a two-variable function $f(x, y) = (g \circ h)(x, y)$ is used:

$$f(x_i + x_\varepsilon, y_i + y_\varepsilon) \approx f(x_i, y_i) + x_\varepsilon \cdot f_x(x_i, y_i) + y_\varepsilon \cdot f_y(x_i, y_i) \tag{2.58}$$

$$\approx g \circ \cos(x_i, y_i) + x_\varepsilon \cdot (-sin)(x_i, y_i) \cdot g' \circ \cos(x_i, y_i) + y_\varepsilon \cdot (-sin)(x_i, y_i) \cdot g'(x_i, y_i) \tag{2.59}$$

When $g(x) = 1/x$ and $h(x, y) = \dfrac{z}{\sqrt{x^2 + y^2 + z^2}} = \cos(\theta(x, y))$, it comes:

$$l(x, y) = l(x_i + x_\varepsilon, y_i + y_\varepsilon) \qquad \text{where} \quad x_\varepsilon \ll x_i \text{ and } y_\varepsilon \ll y_i \tag{2.60}$$

$$= \beta + \frac{\alpha}{\cos(\theta(x_i + x_\varepsilon, y_i + y_\varepsilon))} \tag{2.61}$$

$$\approx \beta + \alpha \left(\frac{1}{\cos(\theta(x_i, y_i))} + x_\varepsilon \cdot \left(-\frac{x_i \cdot z_i}{\left(x_i^2 + y_i^2 + z_i^2\right)^{3/2}} \right) \cdot \left(-\frac{1}{\cos^2(\theta(x_i, y_i))} \right) + y_\varepsilon \cdot \left(-\frac{y_i \cdot z_i}{\left(x_i^2 + y_i^2 + z_i^2\right)^{3/2}} \right) \cdot \left(-\frac{1}{\cos^2(\theta(x_i, y_i))} \right) \right) \tag{2.62}$$

$$\approx a \cdot x_\varepsilon + b \cdot y_\varepsilon + c \tag{2.63}$$

with

$$a = \alpha \left(\frac{x_i}{\left(x_i^2 + y_i^2 + z_i^2\right) \cos^3(\theta(x_i, y_i))} \right); \qquad b = \alpha \left(\frac{y_i}{\left(x_i^2 + y_i^2 + z_i^2\right) \cos^3(\theta(x_i, y_i))} \right) \tag{2.64}$$

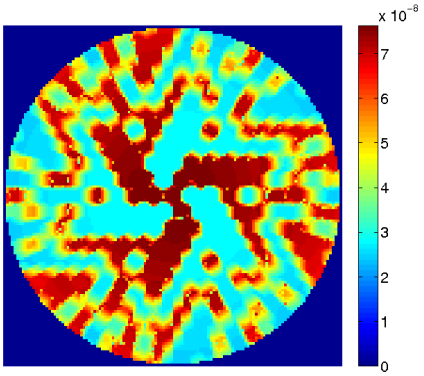
$$c = \beta + \alpha \left(\frac{1}{\cos(\theta(x_i, y_i))} - \frac{x_0^2}{\left(x_i^2 + y_i^2 + z_i^2\right) \cos^3(\theta(x_i, y_i))} - \frac{y_0^2}{\left(x_i^2 + y_i^2 + z_i^2\right) \cos^3(\theta(x_i, y_i))} \right) \tag{2.65}$$

A nearest neighbour approximation is applied to pixels covered by several triangles, only centre of pixels matters. This is its centre which determines from which face the intersecting length should be interpolated from. Once all the intersection lengths are interpolated, the detection probabilities are deduced from Eq. (2.49).

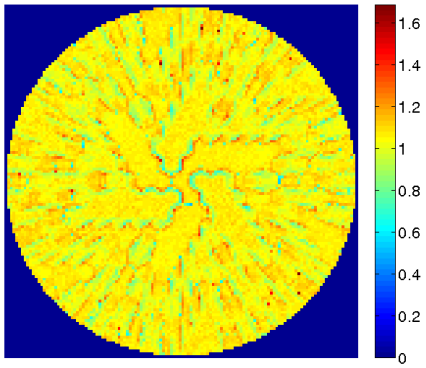
2.3.3.2 Results

An example of a projection generated with such an algorithm is shown in Fig. 2.25a. Some pixels are wrongly estimated. It takes 30 s to compute it. Fig. 2.25c shows the calculation time of a single projection as a function of the number of pixels in the detector. It was performed on a single 1.7GHz processor. The computation time for the reference technique is shown for comparison. The point of this method is that the number of intersection calculations depends principally on the collimator, not on the number of pixels. It becomes advantageous when the number of pixels in the detector is higher than 512×512. When the number of pixels is lower, the constrained Delaunay triangulation impacts strongly the computation time, since it takes about 30 s to be performed. The number of calculations of intersection lengths is constant when using this method. This is the number of interpolations which increases linearly with the number of pixels. As interpolations are much faster to compute than intersecting distances, the new algorithm is faster than ray-driven algorithms for detectors with a large number of pixels. Furthermore, special care should be taken with rays which are aligned with edges, because the detection probability map is then discontinuous, which is difficult to take into account with this method. Furthermore, this triangulation generates a very large number of triangles that are used for interpolating values of pixels. The larger the number of triangles is, the slower the computation is.

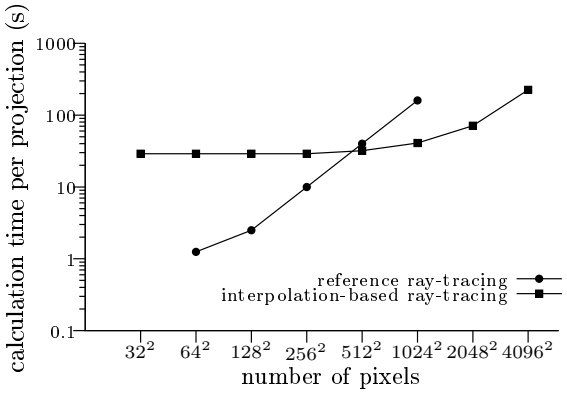
It is chosen to quantify the error between the projection that is simulated with the Monte-Carlo code GATE in Fig. 2.8b and the projection which is obtained with the interpolation technique by calculating the ratio between these two projections. The result is shown in Fig. 2.25b and it can be observed that the largest errors are located on the edges of the collimator. Fig. 2.25d is then derived from Fig. 2.25b in order to quantify the error. The mean of this distribution is 1.042, which means that there is a bias between both projections that is lower than 5%. The standard deviation is $\sigma = 8.47 \times 10^{-2}$, which should be compared to the standard deviation of the reference technique which was 7.08×10^{-2}. The interpolation-based method is consequently 16% worse than the reference technique. This means that the error is larger with this method than with the reference method. This was expected because most of the intersecting lengths are interpolated and consequently are not the exact lengths of the intersection of rays with the collimator.



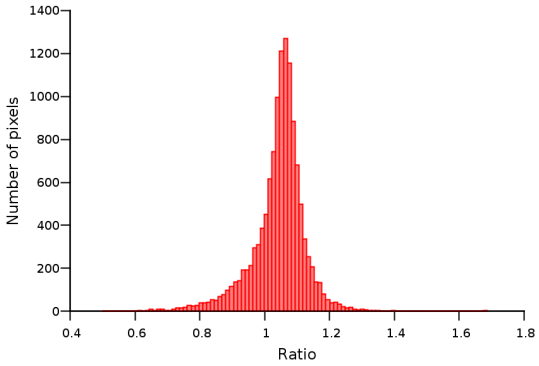
(a) Detection probabilities of the photons that are emitted from the centre voxel of the 3D spatial distribution



(b) Ratio of projections of Fig. 2.25a and 2.8b



(c) Computation time



(d) Distribution of Fig. 2.25b

Figure 2.25 – Estimation of the probability map of a single voxel at 511 keV obtained with the interpolation-based technique and comparison with the GATE simulation of Fig. 2.8b

2.3.4 Second technique: computing intersection lengths from collimator slices

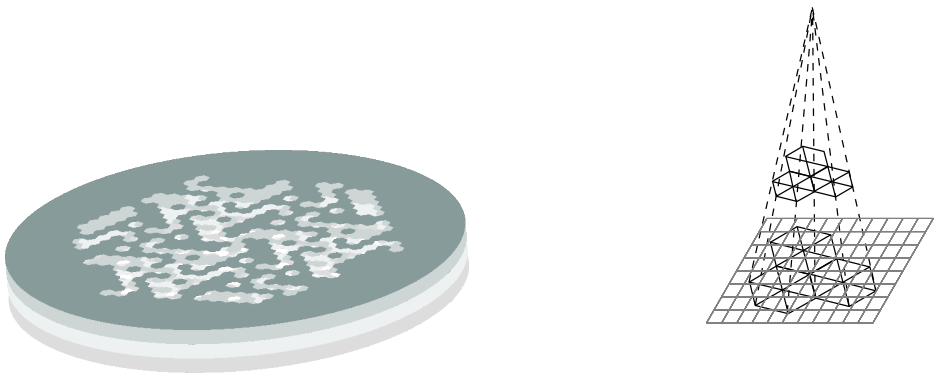
A second method is proposed to compute system matrices, which does not rely on constrained Delaunay triangulation.

2.3.4.1 Method

In this section it is assumed that the collimator is made of s thin slices which can be represented as plane surfaces as in Fig. 2.26a. This hypothesis implies that rays either pass completely through a slice or not at all. No intersection calculation is required anymore. The contribution of the collimator in the calculations of $(a_{i,j})$ becomes:

$$e^{-\mu_c l_{i,j}} = e^{-\mu_c \sum_s l_{i,j,s}} \tag{2.66}$$

where $l_{i,j,s}$ is the length of the contribution of the slice s to the attenuation of the collimator. For each slice, all the triangles are projected from the centre of the voxel j to the detector, as it shown in Fig. 2.26b. As it can be observed in Fig. 2.27a, several pixels are usually covered by projected slice triangles. If the pixel i is not covered by any triangles from the slice s, then $l_{i,j,s} = 0$. From Fig. 2.27b, a relation is derived between the intersecting length



(a) Thin slice of a coded-aperture collimator

(b) Projection of point source through the collimator on the detector

Figure 2.26 – Collimator is made of S slices that are projected successively on the detector

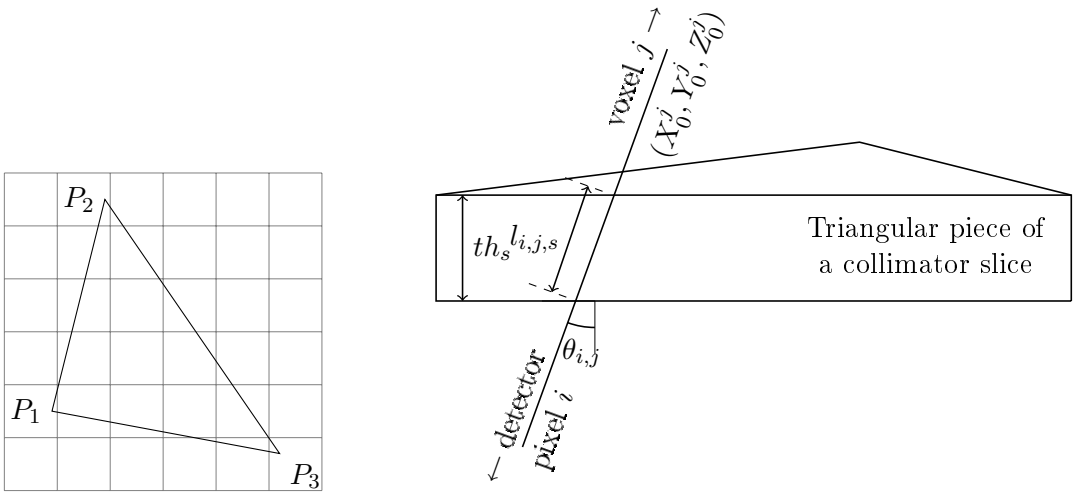
$l_{i,j,s}$, the incidence angle $\theta_{i,j}$ and the thickness th_s. $\forall i$ so that $(x_i, y_i) \in (P_1\, P_2\, P_3)$,

$$l_{i,j,s} = \frac{th_s}{\cos(\theta_{i,j})} \tag{2.67}$$

Finally, the contribution of the collimator is obtained through:

$$e^{-\mu_c l_{i,j}} = e^{-\mu_c / \cos\theta_{i,j} \sum_s \left(th_s \cdot \delta^s_{i,j}\right)} \tag{2.68}$$

where $\delta^s_{i,j}$ is 1 if the ray starting from the centre of a voxel j and ending at the centre of a pixel i crosses the slice s and 0 otherwise.

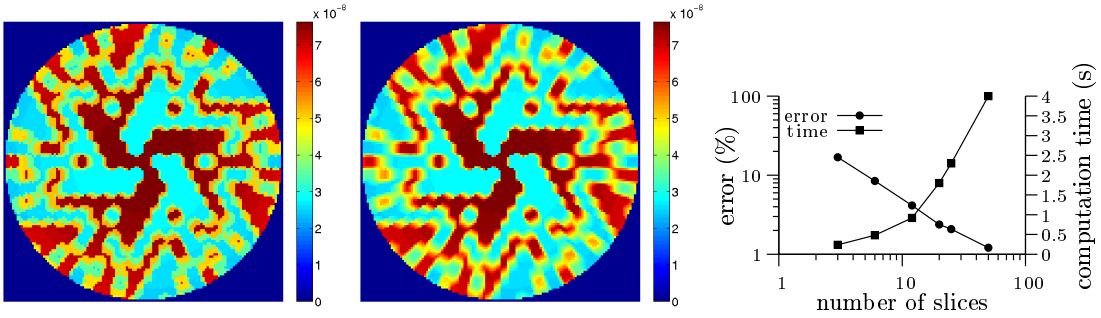


(a) Projection of a collimator face on pixels

(b) Intersection of ray and a triangular piece of a thin collimator slice

Figure 2.27 – Projection of a triangular piece of collimator slice on a pixel grid

2.3.4.2 Results



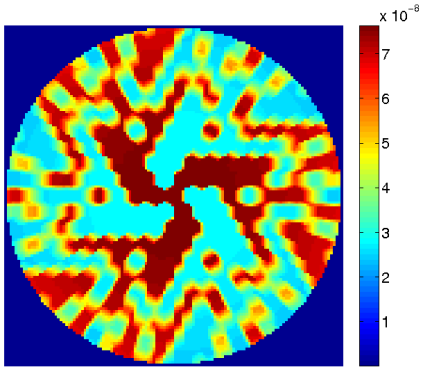
(a) Projection with 3 slices (b) Projection with 100 slices (c) The number of slices impacts the accuracy of the projections and its computation time

Figure 2.28 – Determination of an appropriate number of slices

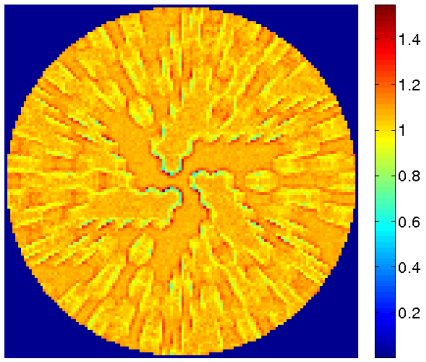
Fig. 2.28 presents the impact of the number of slices which is used to model the collimator on the accuracy of the projection of the collimator on the detector as well as the computation time. When the collimator is modelled with a short number of slices as in Fig. 2.28a where 3 slices are used, discontinuities are observed on the edges of the collimator in the projection. Conversely, when a large number of slices is used to model the collimator as for instance in Fig. 2.28b where 100 slices are used, no discontinuities are visible in the projection but the computation time becomes large. Fig. 2.28c shows the accuracy of the projection and the associated computation time as a function of the number of slices. From this plot, it is decided to model the collimator with 20 slices as it corresponds to an error in the projection which is lower than 1% while the computation time per voxel is 0.65 s for a 128×128 projection. Computation times are obtained on a single 2.4 GHz processor.

The projection that is obtained for 20 slices is shown in Fig. 2.29a and this projection is compared to the result which is obtained with the Monte-Carlo simulation of Fig. 2.8b The ratio of these two projections is shown in Fig. 2.29. The mean of the distribution of the pixel values of this images, which is shown in Fig. 2.29d, is 1.057 with a standard deviation $\sigma = 7.23 \times 10^{-2}$, which is 2.1% worse than with the reference technique (σ was 7.08×10^{-2}). However, the computation time is much better as it is 4 times faster than the reference technique for a 128×128 projection. Fig. 2.29c shows the computation time of a single projection for various number of pixels in the detector. This method is always faster than the reference technique.

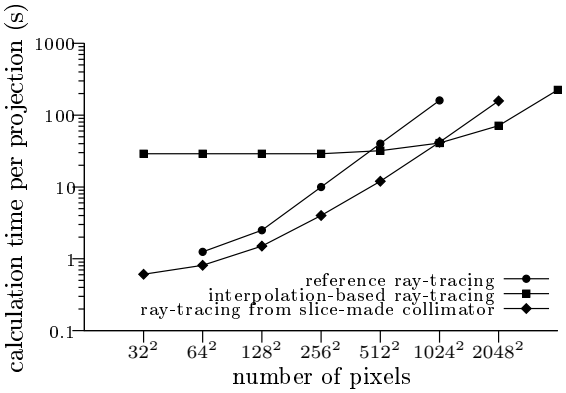
The limitation of this method is that the accuracy of the computation of the system matrices depends on the number of slices used to model the collimator. An exact method would be preferred, as long as the computation time is comparable to this method.



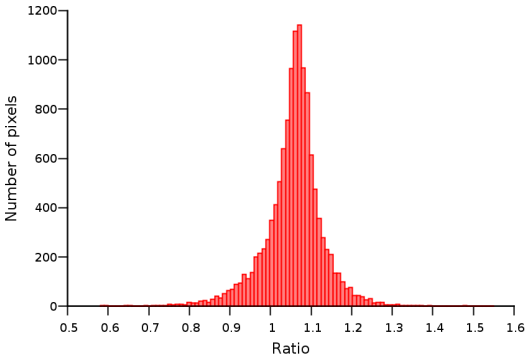
(a) Projection of the centre voxel on the detector, which is made of 128×128 pixels



(b) Ratio of projections of Fig. 2.29a and 2.8b



(c) Performance of the computation of a the face-based ray-tracing algorithm



(d) Distribution of Fig. 2.29b

Figure 2.29 – Estimation of the probability map of a single voxel at 511 keV obtained with the collimator viewed as 20 slices and comparison with the GATE simulation of Fig. 2.8b

2.3.5 Third technique: face-based ray-tracing

A third and last method is now presented. It aims at computing efficiently and accurately system matrices when the collimator is made of a large number of faces. Contrary to the two previous methods, it does not require any interpolations or approximations. Indeed, it is possible to exactly derive the intersecting length between any ray and the collimator on the base of a face-based method. First, this method is exposed in 2D for clarity, where the ray intersects a unique piece of collimator. Then, it is extended to the 3D case. The general method is then developed where a ray can intersect several pieces of collimator.

2.3.5.1 Method

First, this technique is explained from a 2D example. It is demonstrated that the exact distance between A and B in Fig. 2.30 can be expressed as an function of (x_d, y_d). The distance AB depends on the parameters (a_1, b_1, c_1), (a_2, b_2, c_2), (x_i, y_i) and (x_d, y_d). This

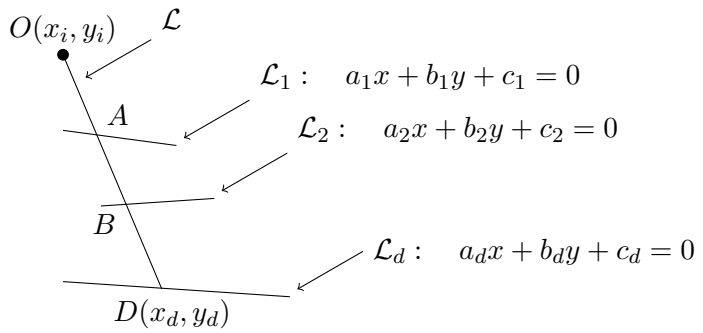


Figure 2.30 – Projection of a point through a plane-surface slice

distance is derived from the following system of equations:

$$\begin{cases} A \in \mathcal{L} \\ A \in \mathcal{L}_1 \\ B \in \mathcal{L} \\ B \in \mathcal{L}_2 \end{cases} \Leftrightarrow \begin{cases} (x_A, y_A) = (x_i, y_i) + \lambda_A(x_d - x_i, y_d - y_i) \\ a_1 \cdot x_A + b_1 \cdot y_A + c_1 = 0 \\ (x_B, y_B) = (x_i, y_i) + \lambda_B(x_d - x_i, y_d - y_i) \\ a_2 \cdot x_B + b_2 \cdot y_B + c_2 = 0 \end{cases}$$

By denoting $(X, Y) = (x_d - x_i, y_d - y_i)$

$$\Leftrightarrow \begin{cases} (x_A, y_A) = (x_i, y_i) + \lambda_A(X, Y) \\ (x_B, y_B) = (x_i, y_i) + \lambda_B(X, Y) \\ \\ \lambda_A = -\dfrac{a_1 \cdot x_i + b_1 \cdot y_i + c_1}{a_1 \cdot X + b_1 \cdot Y} \\ \\ \lambda_B = -\dfrac{a_2 \cdot x_i + b_2 \cdot y_i + c_2}{a_2 \cdot X + b_2 \cdot Y} \end{cases}$$

$$\Rightarrow AB = \left(-\frac{a_2x_i + b_2y_i + c_2}{a_2X + b_2Y} + \frac{a_1x_i + b_1y_i + c_1}{a_1X + b_1Y} \right) \sqrt{X^2 + Y^2} \tag{2.69}$$

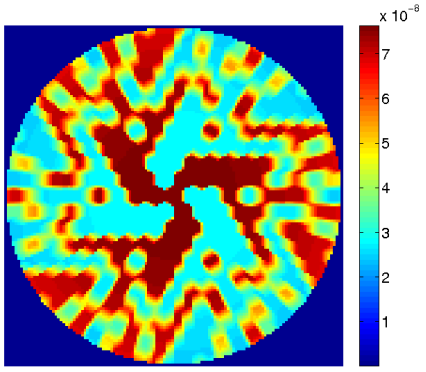
This is an exact derivation of the distance AB. This extends directly to 3D. When a ray intersects two planes, the distance between the two intersection points is given by:

$$AB = \left(-\frac{a_2x_i + b_2y_i + c_2z_i + d_2}{a_2X + b_2Y + c_2Z} + \frac{a_1x_i + b_1y_i + c_1z_i + d_1}{a_1X + b_1Y + c_1Z} \right) \sqrt{X^2 + Y^2 + Z^2} \tag{2.70}$$

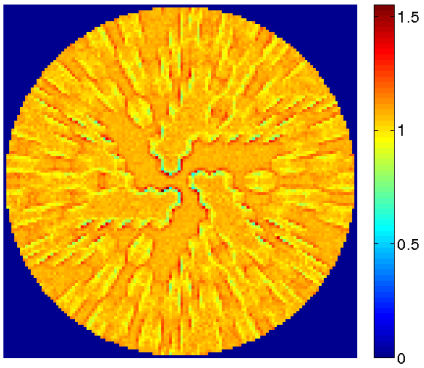
where $Z = z_d - z_i$. In the parenthesis of Eq. 2.70, the first term corresponds to the point A which is the point where the ray enters into the collimator, whereas the second term corresponds to the intersection point where the ray goes out from the collimator. Instead of computing intersections of every ray originating from $0(x_i, y_i, z_i)$ and ending in every pixel centre $D(x_d, y_d, z_d)$, the computation of the intersecting distances will be made more directly according to Eq. (2.70). Every face describing the collimator is projected as in Fig. 2.23b, and the pixels standing in its shadow are updated. Actually, if the ray enters the collimator volume through face $\mathcal{F}$, its contribution C is added to the pixel in the shadow, and if its goes out, it is subtracted, with C being:

$$C = \frac{a_Fx_i + b_Fy_i + c_Fz_i + d_F}{a_F(x_d - x_i) + b_F(y_d - y_i) + c_F(z_d - z_i)} \sqrt{(x_d - x_i)^2 + (y_d - y_i)^2 + (z_d - z_i)^2} \tag{2.71}$$

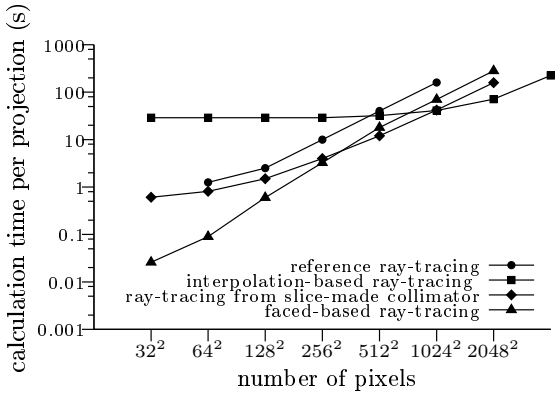
where parameters (a_F, b_F, c_F, d_F) describe the plane where face F is in. This is applied to all faces of a triangulated representation of the collimator as in Fig. 2.20b. Normal vectors of faces are assumed to be pointing outside the collimator.



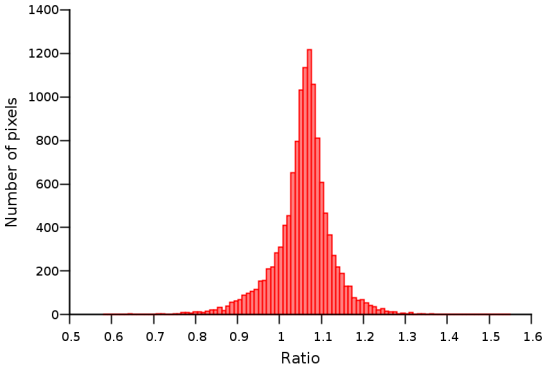
(a) Projection of the centre voxel on the detector, which is made of 128×128 pixels



(b) Ratio of projections of Fig. 2.31a and 2.8b



(c) Performance of the computation of a the face-based ray-tracing algorithm



(d) Distribution of Fig. 2.31b

Figure 2.31 – Estimation of the probability map of a single voxel at 511 keV obtained with the face-based technique and comparison with the GATE simulation of Fig. 2.8b

2.3.5.2 Results

The example of Fig. 2.31a models the detection probabilities obtained with the face-driven technique. Fig. 2.31b is obtained by taking the ratio of Fig. 2.31a and Fig. 2.8b. It shows that errors are still mostly distributed on the edges of the collimator. The distribution of the values of this figure are represented in Fig. 2.31d. The mean value of this distribution is 1.057 and its standard deviation is $\sigma = 7.08 \times 10^{-2}$, which is equal to the standard deviation of the reference technique. This is due to the fact that this technique does not take advantage of interpolations or approximations. Fig. 2.31c shows that the computation time increases logarithmically as the detector is refined in terms of number of pixels. However, the performances are excellent since it takes about 800 ms to perform the projection of a 2900-face collimator on a 128×128-pixel detector.

2.3.6 Oversampling

Although it is possible to calculate efficiently and exactly the intersection lengths $l_c(\boldsymbol{r}_i, \boldsymbol{r}_j)$, the computation of (a_{ij}) still suffers from the approximations of Eq. (2.27) and (2.29). As a result, it is proposed to investigate the error on the (a_{ij}) by oversampling the pixels and the voxels. The refinement frequency is denoted for voxels $M' = M'_X \cdot M'_Y \cdot M'_Z$ and for

pixels $N' = N'_x \cdot N'_y \cdot N'_z$. The centres of subvoxels are denoted $\{X_{j'}, Y_{j'}, Z_{j'}\}$ and subvoxels stand inside $\{X_j \pm \Delta X/2, Y_j \pm \Delta Y/2, Z_j \pm \Delta Z/2\}$. The centres of subpixels are denoted $\{x_{i'}, y_{i'}, z_{i'}\}$ and subpixels stand inside $\{x_i \pm \Delta x/2, y_i \pm \Delta y/2, z_i \pm \Delta z/2\}$.

2.3.6.1 Method

The probability that a photon emitted from a voxel v^j is detected in a pixel p^i is equal to the sum of probabilities that it is emitted from subvoxels $v^{j'}$ in v^j weighted with the probability that a photon emitted from v^j is emitted from $v^{j'}$. This can be expressed according to:

$$a_{i,j} = \frac{1}{M'} \sum_{v^{j'} \subset v^j} a_{i,j'} \tag{2.72}$$

Furthermore, the probability that a photon emitted from $v^{j'}$ is detected in p^i is equal to the sum of the probabilities that it would be detected in subpixels $p^{i'}$:

$$a_{i,j'} = \sum_{p^{i'} \subset p^i} a_{i',j'} \cdot e^{-\mu_d((z_{i'} - \Delta z/2N'_z) - (z_i - \Delta z/2))/\cos\theta_{i'j'}} \tag{2.73}$$

The exponential term accounts for the attenuation of the scintillator: by over-sampling the thickness of the scintillator, it is necessary to take into account the attenuation of the first scintillator slices. $z_{i'} - \Delta z/2N'_z$ is the z-coordinate of the slice where $p^{i'}$ is in and $z_i - \Delta z/2$ is the z-coordinate of the surface of scintillator. The probability $a_{i,j}$ is then derived from Eq. (2.45), (2.72) and (2.73):

$$a_{ij} = \frac{1}{M'} \sum_{v^{j'} \subset v^j} \sum_{p^{i'} \subset p^i} a_{i'j'} \cdot e^{-\mu_d(z_{i'} - z_i + (1 - 1/N'_z)\cdot \Delta z/2)/\cos\theta_{i'j'}} \tag{2.74}$$

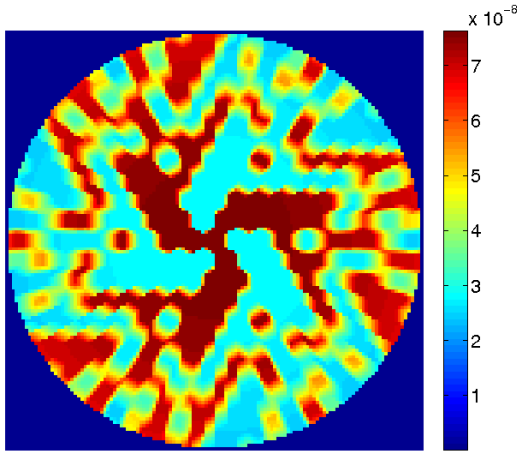
$$\begin{aligned} &= \sum_{v^{j'} \subset v^j} \sum_{p^{i'} \subset p^i} \frac{(\Delta x/N'_x)(\Delta y/N'_y) \cdot \mu_d^2 \cos\theta_{i'j'}}{4\pi M'} \cdot e^{-\mu_c l_c(\boldsymbol{r}_{i'}, \boldsymbol{r}_{j'})} \\ &\qquad \times e^{\mu_d(z_{i'} - Z_{j'} - \Delta z/2N'_z)/\cos\theta_{i'j'}} \cdot e^{-\mu_d(z_{i'} - z_i + (1-1/N'_z)\cdot\Delta z/2)/\cos\theta_{i'j'}} \\ &\quad \times \left(\Gamma\left(-1, \frac{\mu_d\left(z_{i'} - Z_{j'} - \Delta z/2N'_z\right)}{\cos\theta_{i'j'}}\right) - \Gamma\left(-1, \frac{\mu_d\left(z_{i'} - Z_{j'} + \Delta z/2N'_z\right)}{\cos\theta_{i'j'}}\right)\right) \end{aligned} \tag{2.75}$$

$$\begin{aligned} &= \sum_{v^{j'} \subset v^j} \sum_{p^{i'} \subset p^i} \frac{(\Delta x/N'_x)(\Delta y/N'_y) \cdot \mu_d^2 \cos\theta_{i'j'}}{4\pi M'} \cdot e^{-\mu_d(Z_{j'} - z_i + \Delta z/2)/\cos\theta_{i'j'}} \cdot e^{-\mu_c l_c(\boldsymbol{r}_{i'}, \boldsymbol{r}_{j'})} \\ &\quad \times \left(\Gamma\left(-1, \frac{\mu_d\left(z_{i'} - Z_{j'} - \Delta z/2N'_z\right)}{\cos\theta_{i'j'}}\right) - \Gamma\left(-1, \frac{\mu_d\left(z_{i'} - Z_{j'} + \Delta z/2N'_z\right)}{\cos\theta_{i'j'}}\right)\right) \end{aligned} \tag{2.76}$$

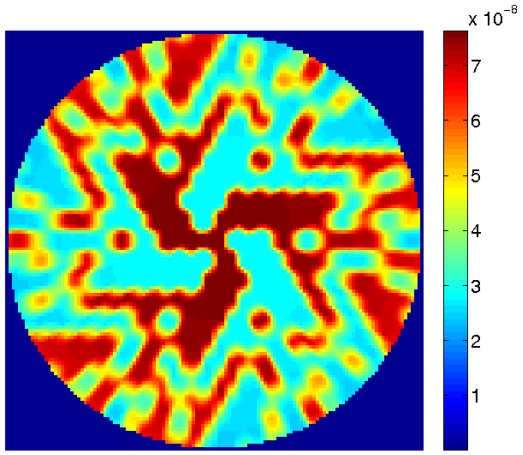
where i' describes subpixels of a pixel p^i and j' describes subvoxels of a voxel v^j. As volumes of subvoxels and subpixels are smaller than the volume of voxels and pixels, zero-order approximations on $\cos\theta(\boldsymbol{r}^j, \boldsymbol{r}^i)$ and $l_c(\boldsymbol{r}^j, \boldsymbol{r}^i)$ as explained at page 71 are more consistent. As a result, the approximations induce smaller errors for calculations of $a_{i',j'}$ than for the calculations of $a_{i,j}$. However, calculation time increases.

2.3.6.2 Results

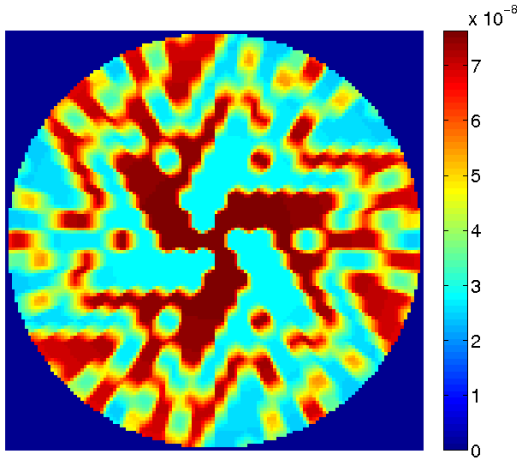
Fig. 2.32 shows the impact of oversampling either the object or the detector or both on the estimation of the probability map corresponding to a single box-shaped voxel. It can be observed that the edges are not visible anymore in Fig. 2.32d than in the other projections



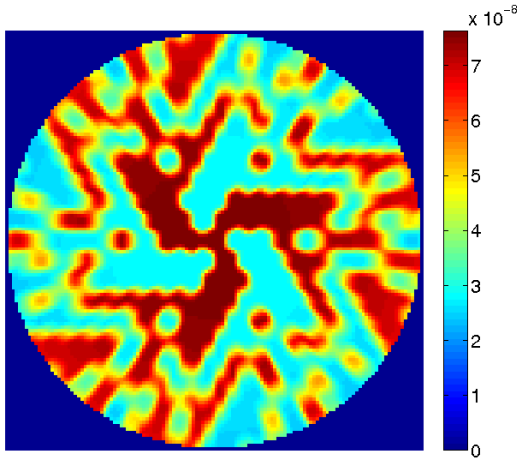
(a) No oversampling



(b) Oversampling of the detector



(c) Oversampling of the object

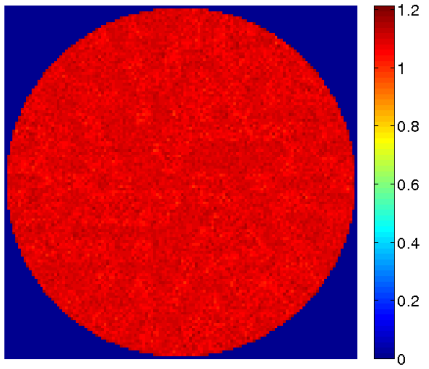


(d) Oversampling of both the detector and the object

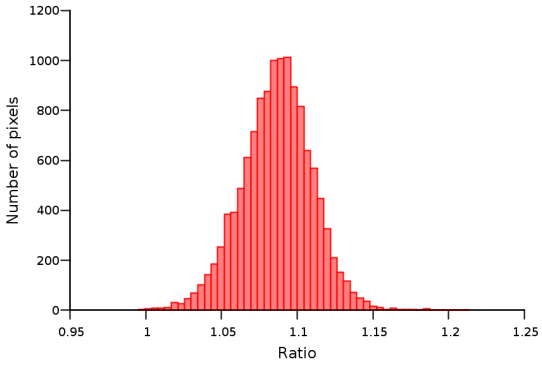
Figure 2.32 – Projection of the centre voxel on the detector with oversampling. Fig 2.32a: no oversampling. Fig 2.32b: the detector is oversampled ($2 \times 2 \times 2$). Fig 2.32c: the object is oversampled ($2 \times 2 \times 2$). Fig 2.32d: Both the detector and the object are oversampled ($2 \times 2 \times 2$) $\times$ ($2 \times 2 \times 2$)

of this figure. Fig. 2.33a shows the ratio between the oversampled analytic method and the Monte-Carlo simulation. It can be observed that the error on edges are not visible anymore. The distribution of this figure is plotted in Fig. 2.33b. Although this distribution is expected to be centred around 1, it is centred around 1.087. The standard deviation of this distribution is $\sigma = 0.0193$, which is better than any standard deviation presented before. The important point to notice is that the standard deviation is decreased by oversampling, which means that the projection suffers less from errors. The bias that is observed can be easily corrected by applying a correction factor.

Dimensions of voxels are $1 \times 1 \times 1$ mm^3. They could be subdivided in 8 smaller voxels whose dimensions are $0.5 \times 0.5 \times 0.5$ mm^3. In the same way, pixels could be subdivided in smaller volumes. The detector, which is modelled in this thesis, is made of pixels whose volume is $0.8 \times 0.8 \times 4$ mm^3. Each of them could be subdivided into smaller volumes, for instance into 16 volumes whose dimensions would be $0.4 \times 0.4 \times 1$ mm^3. In this case, the calculation



(a) Ratio of projections of Fig. 2.32d and 2.8b



(b) Distribution of Fig. 2.33a

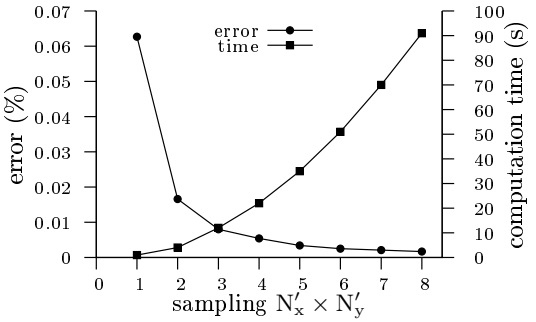
Figure 2.33 – Error estimation

time would increase by a factor $8 \times 16 = 128$, which would result in a prohibitive computation time. As a result, the impact of oversampling parameters $(M'_x, M'_y, M'_z, N'_x, N'_y, N'_z)$ is investigated independently to determine which dimensions should be oversampled in priority to optimally improve the accuracy of system matrices. First, the detector is oversampled in the transverse direction, which is equivalent to the calculation of the probability matrix on a detector with $(N'_x \times N'_y)$ times more pixels, which are $(N'_x \times N'_y)$ times smaller. Then, the detector is oversampled along the depth of the detector. Secondly, the object is oversampled, first in the transversal plane, then along its depth. By oversampling the voxels and the pixels, the accuracy of the matrix increases, but also the computation time. The errors in Fig. 2.34 are calculated according to:

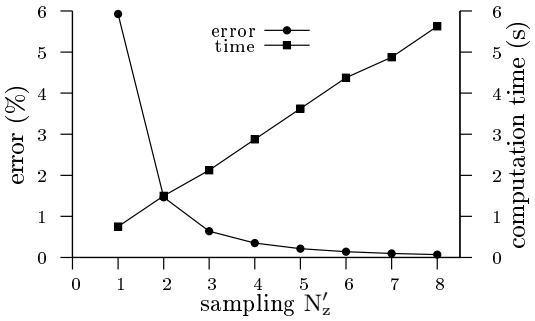
$$\text{error} = \sqrt{\frac{1}{N} \sum_{i,j} \left(2 \cdot \frac{I_{ij} - I_{ij}^{ref}}{I_{ij} + I_{ij}^{ref}} \right)^2} \tag{2.77}$$

where I is the image obtained through oversampling, for instance Fig. 2.32b, and I^{ref} is the image obtained by taking a large oversampling factor. For the plots shown in Fig. 2.34, the sampling parameters which are used for computing the reference image I_{ref} are all set to 1, excepted the sampling parameters which are under investigation which are set to 16. N is the number of pixels in the circular area. It is observed that oversampling the pixel depth or the voxel depth results in less improvements on the accuracy than oversampling the transverse directions of the voxels and the pixels. Furthermore, oversampling the transverse directions of the voxels $(M'_x \times M'_y)$ leads to a better improvement of the accuracy than oversampling the transverse directions of pixels $(N'_x \times N'_y)$.

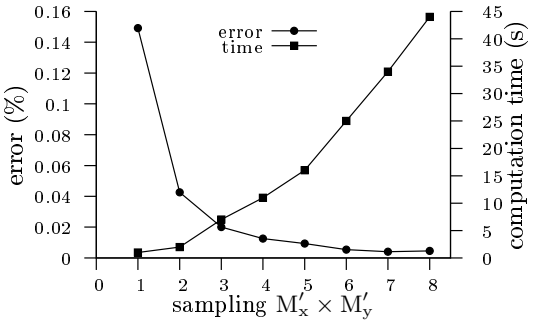
It would seem better to calculate a refined system matrix instead of oversampling pixels and voxels, but the computer's RAM requirement becomes too large. The typical sizes of system matrices that are used in this thesis vary from 1 to 4 GBytes. By refining voxels and pixels by a factor 2, the memory requirement increases by a factor $(2^3) \times (2^2) = 32$, which is complicated in practice because it would lead to matrices too large to be handled. Moreover, it slows down reconstructions, which are already quite slow because information is very intricate due to the use of coded aperture collimators.



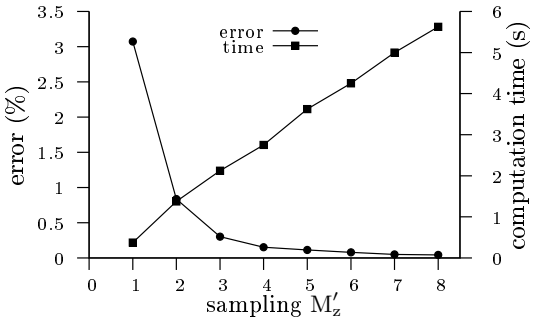
(a) Influence of the transverse sampling parameters N'_x and N'_y. The reference projection corresponds to $N'_x = 16$ and $N'y = 16$



(b) Influence of the axial sampling parameter N'_z. The reference projection corresponds to $N'_z = 16$



(c) Influence of the transverse sampling parameters M'_x and M'_y. The reference projection corresponds to $M'_x = 16$ and $M'y = 16$



(d) Influence of the axial sampling parameters M'_z. The reference projection corresponds to $M'_z = 16$

Figure 2.34 – Influence of sampling parameters on the determination of the probability matrix

2.4 Discussion and conclusion

Estimating the β^+ input function is a low-signal imaging problem. As a result, appropriate choices must be made concerning the measurement location and the imaging system. It is suggested to perform the measurement of the β^+ input function at the popliteal artery, which is an original choice. The diameter of this artery is large (about 5 mm) which results in a large amount of emitted photons. The popliteal artery is located about 50 mm from the surface of the knee, which results in an attenuation of the signal by tissues of 60%. Furthermore, no organ with a large amount of radiotracers (heart, brain, bladder) is located close to the knee. Other candidates are arteries which are either smaller arteries or arteries with a very active organ in its surrounding whose signal dominates the signal coming from the artery. Depending on the radiotracers, it could be interesting to perform the measurement at the common femoral artery, which is larger than the popliteal artery, but the proportion of radiotracers which is taken up by muscles should be low otherwise it would be complicated to extract the signal coming from the artery.

The non-invasive estimation of the β^+-input function requires the development of an imaging system with a good efficiency because of the low activity concentration during typical PET acquisitions. A γ-imager is used to collect photons. Efficiency would be better if some modifications could be made on the scintillator. Replacing the CsI scintillator by a BGO scintillator would increase the efficiency by a factor 2 (from 16% to 32% for a 4 mm-thick scintillator at 511 keV). Furthermore, it was not possible to calibrate the imager for acquisitions at 511 keV. When placing a 511 keV source in the field of view, only photons which undergo Compton scattering and deposit less than 350 keV can be registered, because

of an electronic collimation which rejects events whose energy was higher than 350 keV. This would be useful to experimentally determine the energy resolution and the spatial resolution of this imager at 511 keV, so that it can then be modelled into Monte-Carlo simulations. It was chosen to develop a SPECT imaging system, because one unique γ-imager is available at Service Hospitalier Frederic Joliot (Orsay), where acquisitions take place. A coincidence system could be developed with a second γ-imager since it disposes of a LIST-mode which allows the registration of the time stamp to every detected photon. The imaging system which is proposed has an efficiency which might be better than the efficiency of a coincidence system based on 2 γ-imagers, thanks to the coded-aperture collimator, which is about 50% transparent to 511-keV photons. It is a rank-6 HURA collimator. It should be reminded that information quality is better with a coincidence system than with a SPECT system, although a coded aperture is used, since coincidence systems provide lines of responses which are close from where photons were emitted, whereas SPECT systems do not.

The imaging system must have a good spatial resolution so that contributions of vessels can be separated. This results in a choice of the magnification of the imaging system which is set to 2, but a worse overlap could be accepted (up to 10%), which results in a smaller magnification, a imager placed closer to the knee and consequently a better efficiency of the imaging system. As correlation-based decoding algorithms reconstruct spatial distributions with strong artefacts, it is suggested to take advantage of a statistical reconstruction algorithm as maximum-likelihood expectation-maximisation algorithm (MLEM). Such an algorithm requires the computation of a matrix which models the imaging system. For 511 keV photons in SPECT, this matrix is large (several gigabytes), because no portion of the collimator is completely opaque to radiation. As a result, the computation of such a system matrix is time-consuming. Ideally, it should be computed through Monte-Carlo simulations, that models all the physics (photoelectric effect, Compton scattering, spatial and energy resolutions, energy cuts). However, performing Monte-Carlo simulations on a single processor takes a prohibitive time. The computation time could be reduced by several orders of magnitude thanks to recent improvements made for accelerating GATE (Beenhouwer 08). Such a version of GATE was not available yet. As an alternative, we proposed to use ray-tracing methods to estimate system matrices within a reasonable computation time. The new methods deeply accelerate the computation of every voxel to the system matrices, but no every physical phenomena can be modelled with ray-tracing methods. In particular, Compton scattering is not taken into account. This results in a systematic error of about 6% between the Monte-Carlo simulations and the projections obtained through ray-tracing. It is shown that this error is minimal when the energy threshold in Monte-Carlo simulations is close to the energy of the photoelectric peak. This means that simulations of datasets with Monte-Carlo simulations in the next chapter should be made with an energy cut set to 500 keV (1σ of the photoelectric peak).

Instead of the Monte-Carlo simulations, an analytical approach is developed. First an exact multi-dimensional integral expression of every coefficient a_{ij} of the system matrix is derived. As this integral can not be analytically solved, some approximations are applied. This results in an expression which can be decomposed into two terms: a term which models the geometry of the imaging system as well as the efficiency of the detector and a term which contains the contribution of the collimator to the system matrix. This two terms are computed independently. An original expression is derived for the computation of the first term. The improvement is not significant when compared to the classical expression for the imaging configuration of this thesis, but should result in a better accuracy of the first term when a better efficiency of the imaging system is looked for. The second term is derived through ray-tracing methods. Several ray-tracing methods are developed. They are all compared to the reference ray-tracing method that consists in computing all the

intersecting lengths of the rays starting from a voxel and ending in a pixel. Although this method is much faster than Monte-Carlo simulations, it can be accelerated. The first accelerating method is based on interpolations. Interpolations deteriorate the accuracy of system matrices by 16%, however it speeds up the computation of system matrices when the detector is made of a large number of pixels ($> 1000 \times 1000$). This method is not faster than the reference technique for detectors with a smaller number of pixels because it requires a constrained Delaunay triangulation which is time-consuming. Once this triangulation is performed, the computation of the contribution of the collimator is fast because it consists only in interpolations. The second method does not use constraint Delaunay triangulation. It is faster than the reference technique and the accuracy can be controlled by the number of slices which models the collimator. It is shown that describing a 9-mm-thick collimator with 20 slices results in a small error (2%) when compared to the reference technique. It does not require the computation of any intersection nor any constrained Delaunay triangulation. Finally a last technique is proposed. It is as accurate as the reference technique but faster. Instead of computing system matrices ray-by-ray, this method computes system matrices face-by-face. The acceleration factor varies from 2 to 4 depending on the number of pixels in the detector. Finally it is presented how to include oversampling in the computation of system matrices. Oversampling the voxels and the pixels decreases the error between the Monte-Carlo simulations and ray-tracing methods. However it increases the computation time. The oversampling directions do not impact similarly the reduction of the error. It is demonstrated that voxels and pixels should only be oversampled in transaxial directions.

In conclusion we have selected the region of the knee for measuring the external measurement of the β^+ arterial input function using a gamma camera equipped with a coded aperture. This chapter then presents methods to efficiently and accurately compute system matrices associated with this imaging configuration in order to use maximum-likelihood expectation-maximisation algorithm. The face-based method is chosen for generating system based matrices for reconstructing spatial distribution in the next chapter.

Statistical reconstruction algorithm with coded-aperture collimators

3

In this chapter, methods are presented to reconstruct object distributions with a statistical reconstruction algorithm as MLEM when projections are acquired with coded-aperture collimators. These methods include a better handling of large system matrix as well as accelerations of convergence rates.

Objects which are reconstructed in this chapter are either point-sources or a set of 2 cylinders. This allows the investigation of the impact of the concentration of radiotracers in objects and the spatial resolution of the detector on the reconstructed distributions. Reconstructions are performed on analytical projections, then on projections generated with Monte-Carlo simulations and finally on real projections obtained by placing phantoms in the field of view of an experimental set-up.

3.1 Handling of MLEM-algorithm with large system matrices

3.1.1 Matrix representation of MLEM-algorithm

In section 1.5.1, the iterative algorithm of Shepp and Vardi is presented. It maximises the likelihood $L(\boldsymbol{\lambda}) \propto P(\boldsymbol{y}|\boldsymbol{\lambda})$, where $\boldsymbol{y}$ represents the observations in detectors and $\boldsymbol{\lambda}$ is the object spatial distribution. This algorithm can be written through a matrix formulation in order to efficiently calculate updates in all voxels for every iteration:

$$\boldsymbol{\lambda}^{h+1} = \frac{\boldsymbol{\lambda}^{h}}{(\begin{bmatrix}1 & \dots & 1\end{bmatrix} \cdot \boldsymbol{A})^{\mathrm{T}}} * \left(\boldsymbol{A}^{\mathrm{T}} \cdot \frac{\boldsymbol{y}}{\boldsymbol{A}\boldsymbol{\lambda}^{h}} \right) \tag{3.1}$$

Divisions, as well as the multiplication indicated with the symbol $*$, are element-by-element operations. The matrix $\boldsymbol{A}$ is a system matrix which is computed according to the face-based method described in the previous chapter. It models the geometry and the physics of the imaging system.

3.1.2 Benefits of symmetries

With high energy photons, collimators are not completely opaque. This results in system matrices which are very dense, unlike system matrices which are used for low-energy SPECT-imaging. Computational resources limit the number of voxels which can be reconstructed since the memory requirement for using system matrices increases linearly with the number of voxels in objects. As a result, if symmetries exist, it is expected to take advantage of them. This section describes how to handle symmetries to reduce memory load during reconstructions.

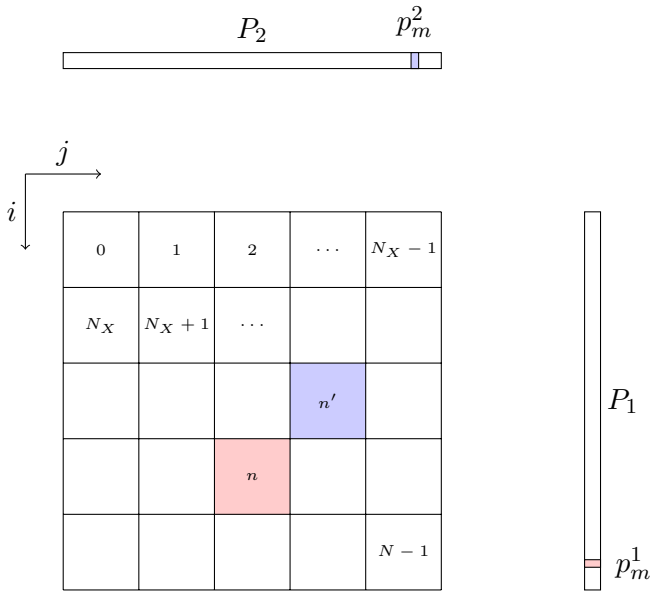


Figure 3.1 – Invariants with Cartesian coordinates. Collimators have not been represented, but this does not change the invariants

Usually, an acquisition $\boldsymbol{y}$ is made of several projections $\{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^P\}$. Let's assume that every projection is made of M pixels and that the object is made of N voxels. Some projections may be orthogonal one to the other. Let's also assume for simplicity that the acquisition is made of only two orthogonal projections $\{\boldsymbol{y}^1, \boldsymbol{y}^2\}$. For clarity, the m^{th} pixel of the whole acquisition is denoted p_m whereas the m^{th} pixel of any projection k is denoted p_m^k. This means that $p_m^k = p_{(k-1)\cdot M+m}$. In Fig. 3.1, where coordinates are Cartesian, the probability $a_{n,m}$ that a photon emitted from a voxel n is detected in a pixel p_m^1 in the first projection P_1 is equal to the probability $a_{n',M+m}$ that a photon emitted from a voxel n' is detected in a pixel p_i^2 in the second orthogonal projection P_2. This means that matrix $\boldsymbol{A}$ is redundant. Let's define $\boldsymbol{A}^1$ and $\boldsymbol{A}^2$ the system matrices which model respectively projections P_1 and P_2, $\boldsymbol{y}$ can then be modelled as:

$$\begin{bmatrix} y_0 \\ \\ \vdots \\ \\ \\ \\ y_{2M-1} \end{bmatrix} = \begin{bmatrix} y_0^1 \\ \vdots \\ y_{M-1}^1 \\ \\ y_0^2 \\ \vdots \\ y_{M-1}^2 \end{bmatrix} = \begin{bmatrix} a_{0,0} & \cdots & a_{0,N-1} \\ \vdots & \ddots & \vdots \\ a_{M-1,0} & \cdots & a_{M-1,N-1} \\ \\ a_{M,0} & \cdots & a_{M,N-1} \\ \vdots & \ddots & \vdots \\ a_{2M-1,0} & \cdots & a_{2M-1,N-1} \end{bmatrix} \times \begin{bmatrix} \lambda_0 \\ \vdots \\ \lambda_{N-1} \end{bmatrix} + \boldsymbol{N}_P \tag{3.2}$$

$$\boldsymbol{y^1} = \boldsymbol{A}^1\boldsymbol{\lambda} + \boldsymbol{N}_P^1 \tag{3.3}$$

$$\boldsymbol{y^2} = \boldsymbol{A}^2\boldsymbol{\lambda} + \boldsymbol{N}_P^2 \tag{3.4}$$

with $\boldsymbol{N}_P^1$, $\boldsymbol{N}_P^2$ and $\boldsymbol{N}_P$ being respectively Poisson noise in projections P_1, P_2 and the Poisson noise in the whole observations, and $a_{i,j}$ being the probability that a photon emitted from voxel v_j is detected in pixel p_i. $\boldsymbol{A}^1$ is made of the first M rows of $\boldsymbol{A}$ and $\boldsymbol{A}^2$ is made of the last M rows, as it can be observed in Eq. (3.2).

We are now looking for the coefficient $a_{m,n'}$ which is redundant with coefficient $a_{m,n}$ in order to express $\boldsymbol{A}^2$ as a function of $\boldsymbol{A}^1$. $\boldsymbol{A}^2$ is redundant with $\boldsymbol{A}^1$ only if it is chosen to work

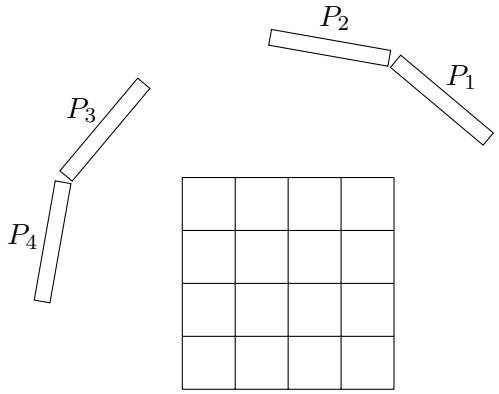


Figure 3.2 – Two orthogonal sets of two projections each. The first set is made of P_1 and P_2, the second set is made of P_3 and P_4

with a square voxel grid as in Fig. 3.1 and 3.2. The number of voxels per row is denoted N_X and consequently, the total number of voxels is $N = N_X^2$. It is then possible to represent voxel indices (i, j) as a 1D-vector $\boldsymbol{v}$ where $\boldsymbol{v}(n)$ is given by Eq. (3.5). n' is then deduced from Eq. (3.5) and Fig. 3.1:

$$n = i \cdot N_X + j \tag{3.5}$$
$$n' = (N_X - 1 - j) \cdot N_X + i \tag{3.6}$$

Half columns $\begin{bmatrix} a_{0,n} & \dots & a_{M-1,n} \end{bmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} a_{M,n'} & \dots & a_{2M-1,n'} \end{bmatrix}^{\mathrm{T}}$ in Eq. 3.2 are then equal. Consequently $\boldsymbol{A}^2$ is obtained from the permutation of columns of $\boldsymbol{A}^1$. This corresponds to:

$$\boldsymbol{A}^2 = \boldsymbol{A}^1 \cdot \boldsymbol{P} \text{ where } \boldsymbol{P} \text{ is a permutation matrix} \tag{3.7}$$

From a computational point of view, it is more efficient to exchange the columns of $\boldsymbol{A}^1$ through a permutation vector $\boldsymbol{v}$ rather than multiplying $\boldsymbol{A}^1$ by the corresponding permutation matrix $\boldsymbol{P}$. When the number of projections is larger than 2, every part of the transition matrix $\boldsymbol{A}$ which corresponds to orthogonal sets of projections S_k can be obtained from $\boldsymbol{A}^1$ and permutation matrices $(\boldsymbol{P}_{1,k})_{k\in[\![1;K]\!]}$ which can be viewed as permutation vectors $(\boldsymbol{v}_k)_{k\in[\![1;K]\!]}$. Fig. 3.2 explains the difference between orthogonal projections and orthogonal sets. The first set is made of projections P_1 and P_2 and the second is made of projections P_3 and P_4. Both sets are orthogonal whereas projections are not two-by-two orthogonal. $\boldsymbol{y}^1$ is here made of projections P_1 and P_2 and $\boldsymbol{y}^2$ is made of P_3 and P_4 and M is the number of pixels per set. The method which is presented in this section also deals with such kind of configurations.

When the acquisition is made of K sets of projections which permute by rotation,

Eq. (3.2) becomes:

$$\begin{bmatrix} \begin{pmatrix} y_0^1 \\ \vdots \\ y_{M-1}^1 \end{pmatrix} \\ \begin{pmatrix} y_0^k \\ \vdots \\ y_{M-1}^k \end{pmatrix} \\ \begin{pmatrix} y_0^K \\ \vdots \\ y_{M-1}^K \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^1 \times \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \\ \boldsymbol{A}^1 \times \quad \boldsymbol{P}^{1,k} \\ \\ \boldsymbol{A}^1 \times \quad \boldsymbol{P}^{1,K} \\ \end{bmatrix} \times \begin{bmatrix} \lambda_0 \\ \vdots \\ \vdots \\ \lambda_{N-1} \end{bmatrix} + \boldsymbol{N}_P \tag{3.8}$$

This statistical model is only based on matrix $\boldsymbol{A}^1$ which is K times smaller than matrix $\boldsymbol{A}$. In this thesis, two projections at maximum are orthogonally acquired. This reduces by a factor of 2 the size of system matrices which are used. When 360°-tomography is performed, this method reduces up to a factor of 4 the size of the system matrix if reconstructions are performed on a Cartesian grid. It can be further extended in cylindrical coordinates as described in Appendix 6 and save a lot of RAM requirement. The question is then about performing multiplications with $\boldsymbol{A}$ and with $\boldsymbol{A}^{\mathrm{T}}$ by only using matrix $\boldsymbol{A}^1$ and the permutation vectors $\boldsymbol{v}_k$, so that the memory requirement is kept low when updating the spatial distribution $\boldsymbol{\lambda}$ according to Eq. (3.1):

$$\boldsymbol{\lambda}^{h+1} = \frac{\boldsymbol{\lambda}^h}{([1\ldots1]\cdot\boldsymbol{A})^{\mathrm{T}}} * \left(\boldsymbol{A}^{\mathrm{T}} \cdot \frac{\boldsymbol{y}}{\boldsymbol{A}\cdot\boldsymbol{\lambda}^h}\right) \tag{3.9}$$

3.1.2.1 Multiplication with $\boldsymbol{A}$

We assume that matrix $\boldsymbol{A}$ is the model of K orthogonal sets of projections. $\boldsymbol{A}$ is a $(KM\times N)$-matrix based on $\boldsymbol{A}^1$ as in Eq. (3.8). Every set is represented by M rows in $\boldsymbol{A}$. We want to calculate $\boldsymbol{c} = \boldsymbol{A}\cdot\boldsymbol{b}$.

$\forall i, \quad \exists!\ (q,r)/ \quad i = q\cdot M + r$ and $0 \le r < M$. It then comes,

$$c_i = (\boldsymbol{A}\cdot\boldsymbol{b})_i = (\boldsymbol{A}^1\cdot\boldsymbol{P}^{1,q}\cdot\boldsymbol{b})_r \tag{3.10}$$

In practice, k copies, called $\boldsymbol{b}^k$ of $\boldsymbol{b}$ are created and every permutation $\boldsymbol{v}_k$ is applied on every copy $\boldsymbol{b^k}$. Every $\boldsymbol{b}^k$ is a permuted copy of $\boldsymbol{b}$. We then multiply $\boldsymbol{A}^1$ with every $\boldsymbol{b}^k$. This gives the k^{th} portion of $\boldsymbol{c}$ as shown in Eq. (3.11).

$$\boldsymbol{A}^1 \begin{bmatrix} b_0^k \\ \\ b_{N-1}^k \end{bmatrix} = \boldsymbol{A}^1\boldsymbol{P}^{1,k} \begin{bmatrix} b_0 \\ \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} c_0^k \\ \\ c_{M-1}^k \end{bmatrix} \tag{3.11}$$

With matrix notations, $\boldsymbol{A}$ can be written as:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}^1 \\ \boldsymbol{A}^1\cdot\boldsymbol{P}^{1,2} \\ \vdots \\ \boldsymbol{A}^1\cdot\boldsymbol{P}^{1,K} \end{bmatrix} \tag{3.12}$$

3.1.2.2 Multiplication with $\boldsymbol{A}^{\mathrm{T}}$

Similar assumptions as in the previous section are done. $\boldsymbol{A}$ is a $(KM \times N)$-matrix. We want to multiply matrix $\boldsymbol{A}^{\mathrm{T}}$ with a vector $\boldsymbol{c}$ by only using $\boldsymbol{A}^1$, which is a $(M \times N)$-matrix, and all permutations $(\boldsymbol{v^k})_{k\in[\![1;K]\!]}$. The result of this multiplication is denoted $\boldsymbol{b}$: $\boldsymbol{b} = \boldsymbol{A}^{\mathrm{T}} \cdot \boldsymbol{c}$. The block transposition of Eq. (3.12) leads to:

$$A^{\mathrm{T}} = \left[\left(\boldsymbol{A}^1\right)^{\mathrm{T}}, \quad \left(\boldsymbol{A}^1 \times \boldsymbol{P}^{1,2}\right)^{\mathrm{T}}, \quad \ldots \quad, \left(\boldsymbol{A}^1 \times \boldsymbol{P}^{1,K}\right)^{\mathrm{T}}\right] \tag{3.13}$$

$$= \left[\left(\boldsymbol{A}^1\right)^{\mathrm{T}}, \quad (\boldsymbol{P}^{1,2})^{\mathrm{T}} \times \left(\boldsymbol{A}^1\right)^{\mathrm{T}}, \quad \ldots \quad, (\boldsymbol{P}^{1,K})^{\mathrm{T}} \times \left(\boldsymbol{A}^1\right)^{\mathrm{T}}\right] \tag{3.14}$$

$\boldsymbol{c}$ is divided in K parts $\boldsymbol{c}^k$ whose size is M. We apply $(\boldsymbol{A}^1)^{\mathrm{T}}$ on every $\boldsymbol{c}^k$ and the result is denoted $\boldsymbol{b}^k$:

$$\boldsymbol{b}^k = (\boldsymbol{A}^1)^{\mathrm{T}} \cdot \boldsymbol{c}^k$$

We then permute every $\boldsymbol{b}^k$ according to matrix $P_{1,k}^{\mathrm{T}} = P_{k,1}$ which is also a permutation matrix and we sum all $\boldsymbol{b}^k$ together:

$$\boldsymbol{b} = \sum_k \boldsymbol{P}^{1,k} \cdot \boldsymbol{b}^k$$

As a result, it is possible to compute MLEM-updates with just $\boldsymbol{A}^1$ and the permutation vectors $\boldsymbol{v}_k$ loaded in RAM, which decreases RAM by a factor K since the RAM requirements of $(\boldsymbol{v}_k)_{k\in[\![1;K]\!]}$ is negligible when compared to $\boldsymbol{A}^1$. A lot of memory is saved by performing calculations in this way. However, this does not improve calculation performance, as calculation time increases linearly with the number of projections. Slight improvements could be observed because CPU performances depend non linearly on RAM load, especially with large RAM.

3.2 Acceleration of MLEM convergence

MLEM is known for its low convergence (Chuang 04). Coded-aperture collimators slow down even more the algorithm since projections are very intricate. Among accelerating algorithms, ordered-subsets expectation-maximisation (OSEM) (Hudson 94) is the most popular algorithm. We do not encourage to use it because it is not a convex algorithm and some convergence troubles are expected because of the low number of projections and because the projections are very intricate due to the use of coded-aperture collimators.

3.2.1 Principle

A new algorithm is proposed to speed up reconstructions. The aim is to efficiently maximise the log-likelihood E^1 which is defined in Eq. 3.15.

$$E^1 = \sum_i \left(-(\boldsymbol{A\lambda})_i + y_i \ln((\boldsymbol{A\lambda})_i) \right) \tag{3.15}$$

The main idea to accelerate reconstructions is to adaptively perform reconstructions, starting from a coarse representation of projections and adaptively refining the observations to finally use the finest representation available. Let's say that projections are $(M_x \times M_y)$ images and there are 2 projections. The system matrix is then made of $(2 \times M_x \times M_y)$ rows. By gathering neighbour pixels, the number of observations and consequently the number of rows of the system matrix are decreased. Using a smaller matrix decreases the time required to perform

an iteration of MLEM. When neighbour pixels $\left(p^{i_k}\right)_{k\in\mathcal{N}}$, which follow Poisson distributions with parameter m_{i_k}, are summed together by $(2^{q-1}\times 2^{q-1})$-packet into a new pixel $p^{i'}$, $p^{i'}$ follows a Poisson distribution whose mean is $\sum_{k=1}^{2^{2(q-1)}} m_{i_k}$. The new observation vector $\boldsymbol{y}^q$ has $2^{2(q-1)}$ less pixels than the initial observation $\boldsymbol{y}$. A new log-likelihood function E^q is defined from $\boldsymbol{y}^q$:

$$E^q = \sum_i \left(-(\boldsymbol{A}^q\boldsymbol{\lambda})_i + y_i^q \ln((\boldsymbol{A}^q\boldsymbol{\lambda})_i) \right) \tag{3.16}$$

where $\boldsymbol{A}^q$ is derived from $\boldsymbol{A}$ by appropriately summing rows. The log-likelihood E^1 is maximised by successively maximising log-likelihoods E^q, q being decreased as the algorithm converges to a solution. However, updating $\boldsymbol{\lambda}$ so that $E^q(\boldsymbol{\lambda}^{h+1})$ is greater than $E^q(\boldsymbol{\lambda}^h)$ does not guaranty that $E^1(\boldsymbol{\lambda}^{h+1})$ is greater than $E^1(\boldsymbol{\lambda}^h)$. As a result, controls are performed on E^1. When $E^1(\boldsymbol{\lambda}^h)$ starts decreasing, next updates are performed to maximise the log-likelihood E^{q-1} instead of E^q. A finer representation is also used if the log-likelihood E^1 does not improve better than a factor 1.1 between two controls. This factor was chosen empirically. This process is repeated until it is not possible to obtain a finer representation of the system matrix.

3.2.2 Binning procedure

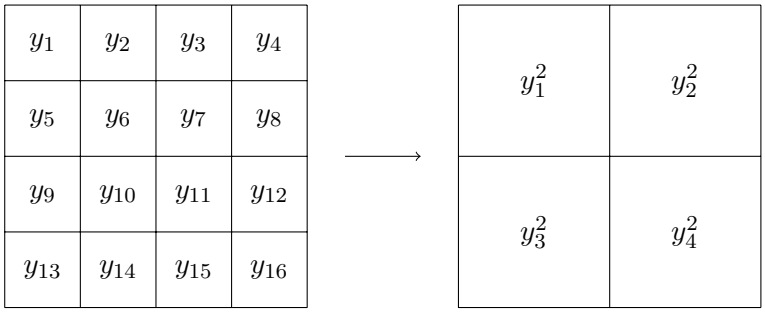


Figure 3.3 – A (4×4)-image binned into a (2×2)-image

Binning procedure is not direct because observations are handled in a vector. As a result, pixels which are neighbours in detectors are not necessarily neighbours in the vector-representation. Let's have a look to the example shown in Fig. 3.3. $\boldsymbol{y}^2$ is deduced from $\boldsymbol{y}$ according to:

$$\begin{bmatrix} y_1^2 \\ \vdots \\ y_4^2 \end{bmatrix} = \boldsymbol{Q}\cdot \begin{bmatrix} y_1 \\ \vdots \\ y_{16} \end{bmatrix} \tag{3.17}$$

where $\boldsymbol{Q}$ is a binning matrix made of 0 and 1 which indicates whether y_i contributes to $y_{i'}^2$. For $q=2$, the vector $\boldsymbol{y}^2$ is 2^2 times smaller than the vector $\boldsymbol{y}$. System matrix $\boldsymbol{A}$ can also be multiplied by binning matrix $\boldsymbol{Q}$, which also results in a matrix which is denoted $\boldsymbol{A}^2$ which is also 2^2 times smaller than the matrix $\boldsymbol{A}$. This can be generalised to higher levels of binning q, in which case, the new system matrix $\boldsymbol{A}^q$ is $2^{2(q-1)} = 4^{q-1}$ times smaller than matrix $\boldsymbol{A}$. Consequently it also accelerates the computation time per iteration by a factor 4^{q-1}. As a large number of iterations must be done to unmixed information in projections acquired with a coded-aperture mask, this accelerates the total reconstruction time.

3.2.3 Implementation

Let's assume that observations are made of M pixels.

- The algorithm is initialised with a homogeneous vector $\boldsymbol{\lambda}^0$ which satisfies:

$$\sum_{i=0}^{M-1} y_i = \sum_{i=0}^{M-1} (\boldsymbol{A}\boldsymbol{\lambda}^0)_i \quad , i.e \quad \forall j, \quad \lambda_j^0 = \frac{\sum_i y_i}{\sum_{i',j'} a_{i',j'}} \tag{3.18}$$

  Whatever the rebinning pattern q is, if $\boldsymbol{\lambda}^0$ satisfies Eq. (3.18), it also satisfies:

$$\sum_{i=0}^{M/4^{q-1}-1} y_i^q = \sum_{i=0}^{M/4^{q-1}-1} (\boldsymbol{A}^q\boldsymbol{\lambda}^0)_i \tag{3.19}$$

- Log-likelihood is calculated with the finest data representation. From Eq. (3.16) and (3.18), it leads to:

$$E^1(\boldsymbol{\lambda}_0) = -\sum_i y_i + \sum_i y_i \ln\left(\frac{\sum_i y_i}{\sum_{i',j'} a_{i,j}} \cdot \sum_j a_{i,j}\right) \tag{3.20}$$

- MLEM iterations are performed on $\boldsymbol{\lambda}$ from the binned observations $\boldsymbol{y}^q$ so that iterations are fast. Log-likelihood $(E^q(\boldsymbol{\lambda}^h))_{k\in\mathbb{N}}$ increases strictly when the algorithm iterates according to:

$$\boldsymbol{\lambda}^{h+1} = \frac{\boldsymbol{\lambda}^h}{([1\ldots1]\cdot(\boldsymbol{A}^q))^{\mathrm{T}}} * (\boldsymbol{A}^q)^{\mathrm{T}} \frac{\boldsymbol{y}^q}{\boldsymbol{A}^q\boldsymbol{\lambda}^h} \tag{3.21}$$

- Every 4^{q-1} iteration, a control is performed on the complete data representation $\boldsymbol{y}$ to check if the likelihood E^1 keeps on increasing. If so, MLEM iterations continue with the same observation $\boldsymbol{y}^q$. Otherwise, finer observation $\boldsymbol{y}^{q-1}$ is used with a new system matrix $\boldsymbol{A}^{q-1}$ derived from $\boldsymbol{A}$. Controls are performed every 4^{q-1} iterations because matrix A^q is 4^{q-1} times larger than matrix A. This means that the frequency of controls is constant over time all along reconstructions.

Fig. 3.4 is an example of the computation times required to reconstruct a spatial distribution with and without acceleration, the object being made of two cylinders. It shows that the adaptive binning method significantly accelerates the reconstruction time. This algorithm is convex and consequently converge to a unique solution. The convexity comes from the controls which are achieved on the log-likelihood E^1. Updates from binned projections only accelerate reconstructions, the objective function which is optimised is always the log-likelihood E^1, from the beginning of the algorithm.

In the following sections, the objects which are reconstructed are either made of a point-source or made of 2 cylinders whose axis-to-axis distance is 6 mm and their diameters are 5 mm-large, which is the expected configuration for the popliteal artery and vein in a knee. It is assumed in the rest of this thesis that the mean radiotracer concentration in vessels is 1.1 kBq/ml (30 nCi/ml), which is coherent with the concentrations measured from blood samples in several investigations at Service Hospitalier Frederic Joliot, Orsay (Gallezot 05). For a two-hour clinical acquisition, this means that every voxel (1 mm^3) standing in a vessel

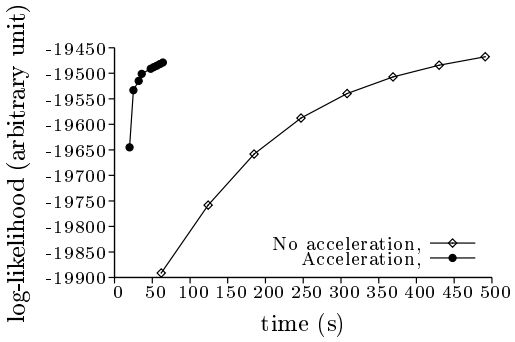


Figure 3.4 – Acceleration of the reconstruction

emits 8000 photons. Radioactive decay is not taken into account because it is radiotracer-dependent. However, considering the radioactive decay would decrease this amount of photons.

$$\frac{1.11 \text{ kBq}}{1000 \text{ mm}^3} \times 2 \cdot 3600 \text{ s} = 8000 \text{ photons/mm}^3$$

Next sections investigate whether it is possible to accurately reconstruct spatial distributions with such a small amount of emitted photons. It is assumed in the following reconstructions that the tissues are not modelled which means that the photons are emitted in air and do not have to cross and scatter in 50 mm of tissues. Reconstructions are first performed on projections which are generated analytically, then on Monte-Carlo simulations and finally on real projections which are acquired with the γ-imager described in section 2.1.2.1. Redundancies in the system matrix and the acceleration technique are applied to efficiently reconstruct objects from projections which are acquired through a coded-aperture.

3.3 Reconstruction of analytical projections

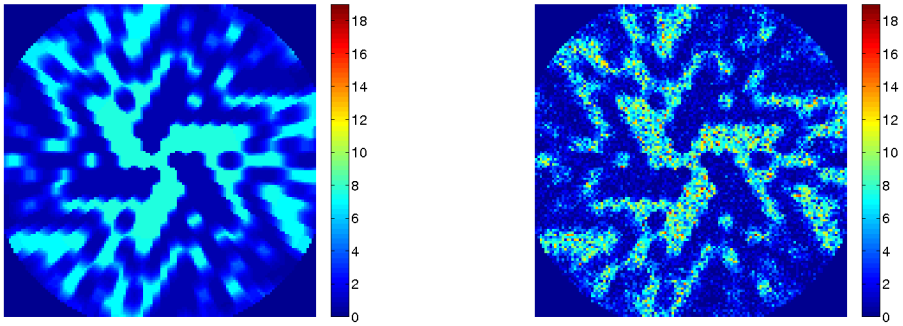
In this section, projections are generated from the system matrix which is used to reconstruct objects. This presents the advantage that the imaging system is perfectly modelled by the system matrix. As a result it will provide the limits of the performances which can be expected by the imaging system.

3.3.1 Generation of noise-free analytical projections and analytical projections with Poisson fluctuations

Let's compute a system matrix $\boldsymbol{A}$ which models the imaging system defined in the previous chapter. A reference object $\boldsymbol{\lambda}_{ref}$ is then generated on a voxel grid by setting activities in its voxels according to the expected spatial distribution. For instance, only one voxel is non-null when it is wanted to investigate the characteristics of a reconstructed single point. Mean expected projections $\boldsymbol{y}^{exp}$ are then generated analytically by multiplying this reference distribution with matrix $\boldsymbol{A}$. The mean expected projection is then given by Eq.(3.22):

$$\boldsymbol{y}^{exp} = \boldsymbol{A} \cdot \boldsymbol{\lambda}_{ref} \tag{3.22}$$

When $\boldsymbol{\lambda}_{ref}$ is a point-source and the system matrix $\boldsymbol{A}$ models the imaging system of Fig. 2.6, the expected projection is the one shown in Fig. 3.5a. It is then possible to generate a projection $\boldsymbol{y}$ which takes into account Poisson fluctuations by drawing for every pixel p^i of $\boldsymbol{y}$ a sample from a Poisson distribution whose parameter is y_i^{exp}. An example is shown



(a) Expected noise-free projection (b) Expected projection with Poisson noise

Figure 3.5 – Analytical generation of a projection for a point-source (a) without noise, (b) with Poisson noise. First, the expected projection is computed from Eq. (3.22), then Poisson noise is taken into account

in Fig. 3.5b. As the relative error of a sample which is drawn from a Poisson distribution whose mean is m is given by $1/\sqrt{m}$, the larger y_i^{exp} is, the smaller the influence of Poisson noise is.

3.3.2 Reconstruction of a point-source

The first step is to observe the reconstructions of a point-source. A single projection is first taken into account as for the configuration which is presented in Fig. 3.6. Reconstructions from projections without Poisson noise will be investigated, then, Poisson noise will be taken into account in the projection. Finally reconstructions are performed from 2 projections.

3.3.2.1 Reconstruction of a point-source from a noise-free projection

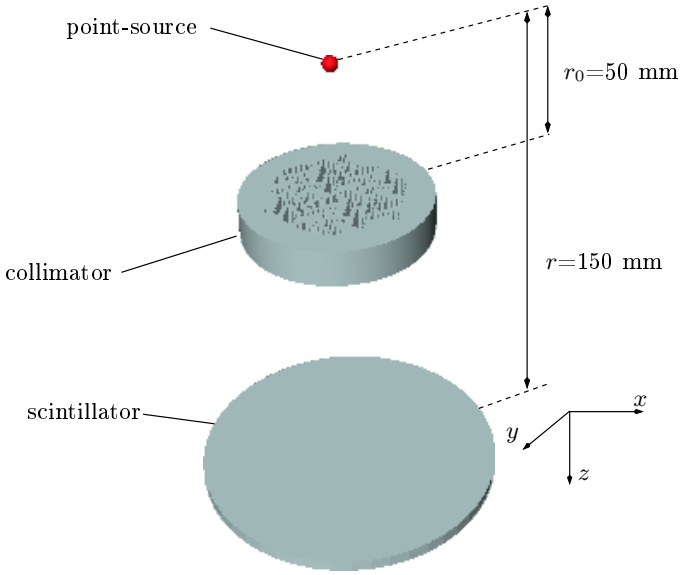
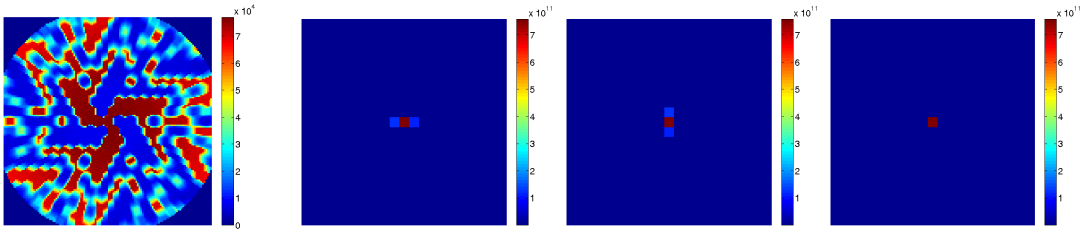


Figure 3.6 – Configuration for acquiring a single projection

First, a point-source is reconstructed from a single noise-free projection. This allows the investigation of the systematic error, as no statistical noise impacts the reconstruction. The



(a) 128×128 initial noise-free projection

(b) Central slices of the estimated distribution in the zx-, yz- and xy-planes

Figure 3.7 – Reconstruction of a point-source from a single noise-free projection

binning level	projection size	iterations	time (s)	log-likelihood
5	8×8	102400	1098	3337911180
4	16×16	128832	2187	3337923092
3	32×32	139024	3874	3337928403
2	64×64	140356	4837	3337929310
1	128×128	140360	4842	3337929312

Table 3.1 – Evolution of the reconstruction of the spatial distribution of a point-source from a noise-free projection

fact that the object is made of a point-source results in a projection shown in Fig. 3.7a. From this single projection, the distribution of the object is reconstructed and the result is presented in Fig. 3.7b. The reconstruction is performed with the accelerated MLEM-algorithm presented in section 3.2. The reconstruction is initialised with a binning level equals to 5, which means that first iterations of the estimation were performed on a 8×8-image which is a rough representation of Fig. 3.7a. Table 3.1 presents the number of iterations performed for every binning level, as well as the time it takes. The evolution of the log-likelihood based on the finest representation of the projection is also presented. This shows that a very large number of iterations are performed within a short time, when the projection is coarse (for instance it takes 10 ms to perform one iteration when the binning level is 5), whereas it takes about 1 s to perform an iteration with the whole projection. The reconstruction is stopped when the likelihood does not improve better than a factor 1.1 between two successive iterations. Calculations are performed on a 2.4GHz 64-bit processor whose RAM is 16 GBytes. As the projection is made of 128×128 pixels and the objects is made of $21 \times 21 \times 21$ voxels, this results in a system matrix of 1.15 GBytes, probabilities being stored with double-precision (8 bytes). It was tried to use system matrices with single-precision (4 bytes), but numerical instabilities appeared and the likelihood computed with Shepp and Vardi algorithm decreased after a certain number of iterations, which is theoretically impossible.

The characteristics of the reconstructed point-source are estimated by fitting Gaussian functions on principal directions as shown in Fig. 3.8, where reconstructed curves correspond to the main directions of Fig. 3.7. Fittings are performed through least-square minimisation. From the estimated standard deviations σ_x, σ_y and σ_z of the Gaussian functions, the full width at half maxima (FWHM) of the reconstructed point are calculated and are respectively 0.37 mm, 0.33 mm and 1.22 mm in the x-, y- and z-directions. The standard deviation and the FWHM are related for a Gaussian function through:

$$\text{FWHM} = 2\sigma\sqrt{2\ln(2)} \tag{3.23}$$

The FWHM are excellent when compared to the size of the voxels ($1 \times 1 \times 1$ mm^3), as

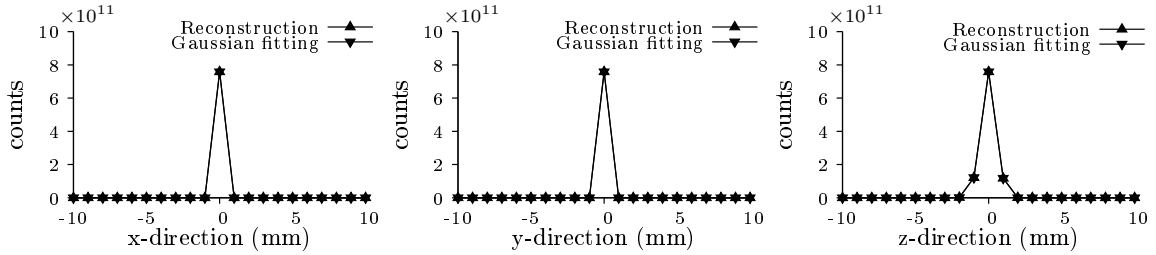


Figure 3.8 – Example of the fitting of Gaussian functions on a reconstructed 3D point

it is first observed qualitatively in Fig. 3.9b. It is important to notice that the FWHM is worse in the z-direction which is the main axis of the imaging system. Furthermore, the estimated activity in the most active voxel is 0.76×10^{12} photons, whereas the input activity is 1.0×10^{12} photons, which means that the error on the activity estimation is biased (24%). When the activity in the region of interest (ROI) around the peak is computed (3σ of the Gaussian fits), all the activity is retrieved. As a result, it appears that a point-source distribution can be accurately estimated from a noise-free projection. The error between the initial spatial distribution and the reconstructed distribution is 3.0×10^{-3}. It is calculated according to:

$$\text{error} = \frac{\sqrt{\frac{1}{N}\sum_{i,j,k}\Big(\lambda(i,j,k) - \lambda_{ref}(i,j,k)\Big)^2}}{\max_{i,j,k}\Big(\lambda_{ref}(i,j,k)\Big)} \tag{3.24}$$

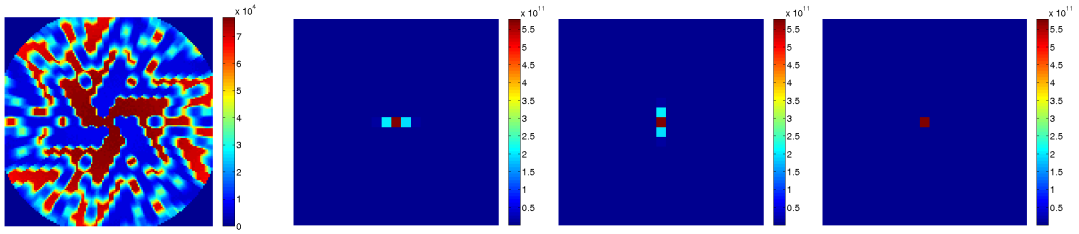
with λ_{ref} being the spatial distribution of the reference object (the point-source), λ being the reconstructed distribution and N being the number of voxels in the object.

The proposed imaging system is well-adapted to reconstruct a point-source from a single noise-free projection when the reconstruction algorithm is a MLEM.

3.3.2.2 Impact of Poisson fluctuations on reconstructions of point-sources from a single projection

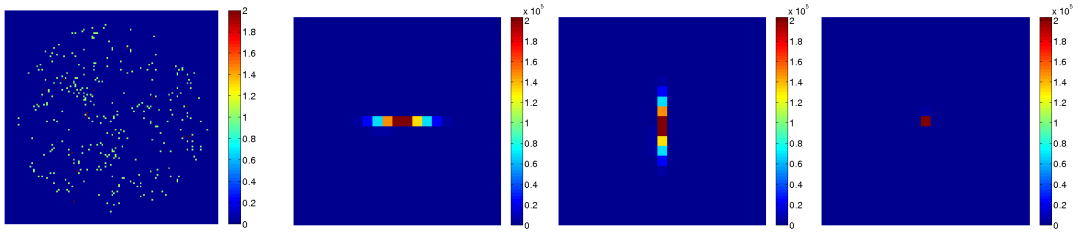
In practice, the projections are Poisson-distributed and it cannot be considered that the projection is noise-free. This section investigates the impact of Poisson fluctuations on reconstructions. Fig. 3.9a presents an example of a projection which takes into account Poisson fluctuations. The object is a point-source which emits 10^{12} photons. The value in every pixel of the projection is obtained by randomly drawing a sample from a Poisson distribution whose mean is the value of the same pixel in Fig. 3.7a. The spatial distribution of the point-source is estimated from this projection and the reconstructed distribution is shown in Fig. 3.9b. This figure represents the three slices (transverse, sagittal and coronal planes) containing the voxel with the highest activity. It can be noticed that Poisson fluctuations impact the amplitude of the reconstructed peak as well as its FWHM. When the activity of the point-source is smaller than 10^{12} photons, for instance when it emits 10^6 photons, less photons are detected in the projection, as shown in Fig. 3.9c. The amplitude of the reconstructed point under-estimates the real activity of the point-source and the reconstructed spatial distribution spreads on several voxels, as illustrated by Fig. 3.9d.

The ratio between the reconstructed amplitudes and the initial amplitudes is quantitatively investigated by reconstructing multiple noisy datasets: for 7 different activities of the



(a) Initial projection. 343×10^6 photons are detected

(b) Central slices of the estimated distribution obtained from Fig. 3.9a, in the zx-, yz- and xy-planes. The point-source emits 10^{12} photons.

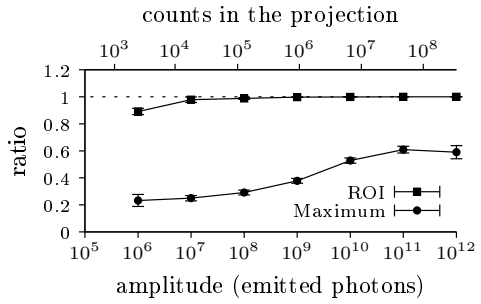


(c) Initial projection. 321 photons are detected

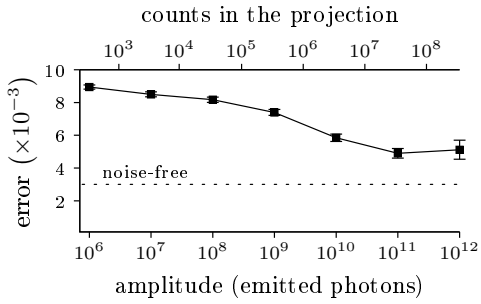
(d) Central slices of the estimated distribution obtained from Fig. 3.9c, in the zx-, yz- and xy-planes. The point-source emits 10^6 photons.

Figure 3.9 – Reconstruction of a point-source from a single projection, with Poisson fluctuations

point-source (from 10^6 to 10^{12} photons), 5 projections are constructed each time in order to obtain the mean and the standard deviation of the ratio for every activity. Fig. 3.10a shows that the maximum reconstructed amplitudes are biased. The maximal reconstructed activities are between 2 and 5 times smaller than the real simulated activities. This is due to the fact that the activity is distributed to several voxels because of the non-perfect spatial resolution of the imaging system. This is commonly called the partial volume effect. This bias is more severe for low activities of the point-source. However, by summing the activities of voxels in the region-of-interest (ROI) around the voxel with the highest intensity, the total estimated activity is close to the simulated activities (the ratios of the initial activities and the activities in the ROI are close to 1). The ROI is defined by taking 2σ (95%) of the spatial distribution of the 3D reconstructed point. There are two abscissa on this figure and on the following ones: the top abscissa indicates the number of detected photons in the projection, whereas the bottom abscissa indicates the number of photons emitted from the source.



(a) Reconstructed maximum intensity and reconstructed intensity in the region of interest



(b) Error between the reconstructed spatial distribution and the reference object, calculated according to Eq. (3.24)

Figure 3.10 – Accuracy of the reconstructed spatial distributions

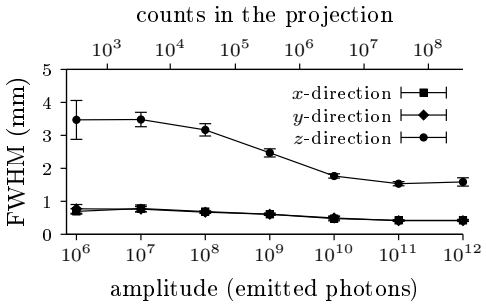


Figure 3.11 – Impact of the activity of the point-source on the FWHM of reconstructed points

In Fig. 3.10b, the error between the reference spatial distribution and the reconstructed distributions is shown. The error decreases when the activity in the point-source increases, which is expected since high activities correspond to small Poisson fluctuations. It converges to 5.0×10^{-3}, whereas the error for the noise-free reconstruction was 3.0×10^{-3}. This result is consequently worse but comparable to the noise-free reconstruction. Errors are calculated according to Eq. (3.24).

Contrary to Fig. 3.7b, the reconstructed point has tails around its peak. Every reconstructed spatial distribution is fitted with Gaussian functions on the 3 main directions as in Fig. 3.8. From the standard deviations σ of the Gaussian functions which are estimated for all these 5×7 reconstructions, the FWHM are derived according to Eq. (3.23) and represented in Fig. 3.11. It illustrates that the FWHM in the z-direction, which is the normal direction to the detector surface, is worse than in the two other directions. The FWHM remain almost constant (0.4 mm) in the x- and y-directions, whereas it decreases down to 1.6 mm in the z-direction when the activity increases. The difference with reconstructions from a noise-free projection are small, the difference being respectively 8%, 21% and 31%.

The reconstruction of point-sources estimates accurately the activity of the initial point-source as long as it is measured from a ROI around reconstructed points. The FWHM in the z-direction is deteriorated when a small amount of photons are detected ($< 10^6$ detected photons) and is always larger than in x and y-directions.

3.3.2.3 Reconstruction of a point-source from 2 projections

We suggest to add a second projection in order to improve the spatial resolution in the z-direction. The configuration of the imaging system becomes the one presented in Fig. 3.12. The second acquisition system is obtained by rotating the first one by 90° around the y-direction. Acquisitions with a point-source in the field of view are reconstructed for 7 activities, ranging from 10^6 to 10^{12} emitted photons. Independent noisy datasets are reconstructed for each of the 7 activities in order to evaluate statistics of estimated parameters. Fig. 3.13b shows the spatial distribution reconstructed from Fig. 3.13a. It illustrates qualitatively the fact that reconstructions of 2 coded projections with the MLEM-algorithm result in an accurate estimation of a point-source although the amount of detected photons is very low ($<$1000 photons).

Reconstructed distributions estimate correctly initial distributions as it is shown in Fig. 3.14a and 3.14b. Contrary to the case where a single projection is taken into account, the maximum amplitude of reconstructed distributions is well estimated, excepted when the point-source does not emit enough photons ($<10^9$ emitted photons) but the total activity on the ROI around reconstructed points is always the expected activity since the

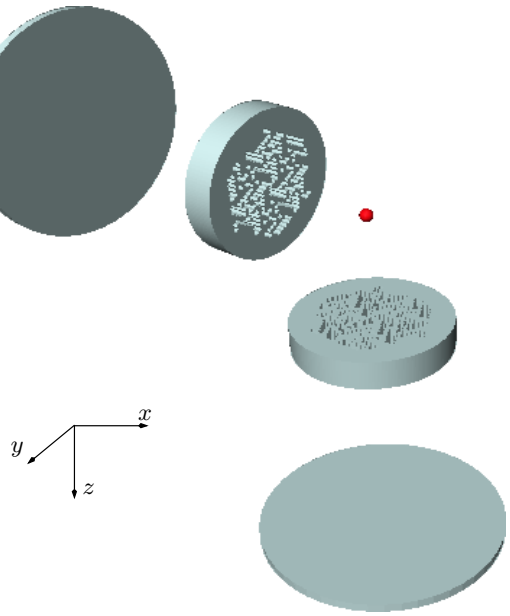
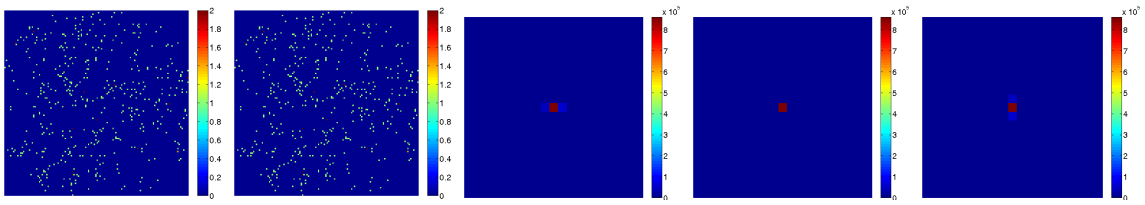


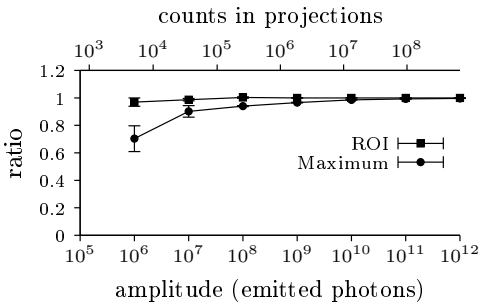
Figure 3.12 – Configuration for acquiring two orthogonal projections



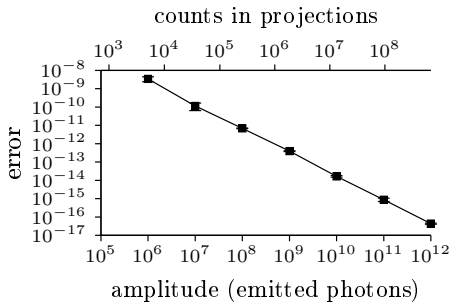
(a) Two orthogonal projections (b) Central slices in the transverse, coronal and sagittal planes

Figure 3.13 – Reconstruction of a point-source from 2 projections with strong Poisson fluctuations due to low amount of signal in the point-source (10^6 emitted photons). 470 and 484 photons are respectively detected in projections of Fig. 3.13a.

ratio is always 1. The error goes rapidly to zero as the simulated activity in the point-source increases as shown in Fig. 3.14b.



(a) Reconstructed maximum intensity and reconstructed intensity in the region of interest



(b) Error between the reconstructed spatial distribution and the reference object, calculated from Eq. (3.24)

Figure 3.14 – Accuracy of the reconstructed spatial distributions

FWHM of reconstructed distributions are estimated in x-, y- and z-directions. This

results in the Fig. 3.15, where it is shown that the FWHM of the 3D reconstructed points decreases as the activity increases. When compared to Fig. 3.11 where reconstructions are performed from a single projection, it is important to underline that the estimation of the FWHM in the z-direction is much better (from 3.5 mm to 0.9 mm for a point-source emitting 10^6 photons). Moreover, the FWHM in the y-direction is better than in the two other directions. This can be explained by the fact that the y-direction is a transverse direction for both projections, while other directions are the axial direction of at least one of the projections.

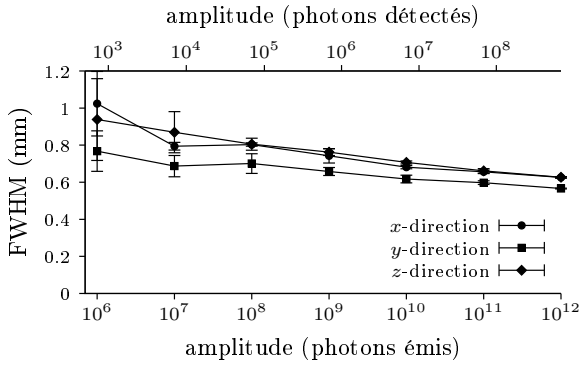
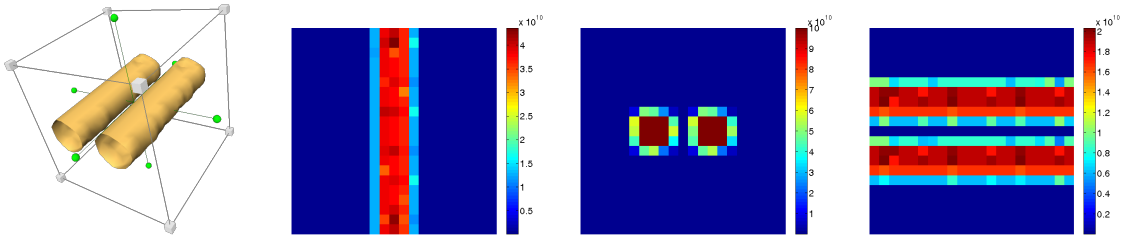


Figure 3.15 – The FWHM of reconstructed points depend on the activity of the point-source

Adding a second projection, orthogonal to the first one, results in a better reconstructions of point-sources, especially for low-activity point-sources. The quantification of the activity of the point-source is excellent, the spread of the reconstructed point is smaller than a voxel in all three directions.

3.3.3 Reconstruction of 2 cylinders



(a) 3D representation with an isosurface

(b) Mean view of the distribution in the (yz)-plane

(c) Mean view of the distribution in the (zx)-plane

(d) Mean view of the distribution in the (xy)-plane

Figure 3.16 – Reference activity distribution

It is then investigated whether 2 cylinders could be reconstructed. The goal is to quantify the concentration of radiotracers in cylinders, since the two cylinders model an artery and a vein. First the spatial distribution $\boldsymbol{\lambda}_{ref}$ of two cylinders is generated on a voxel grid as shown in Fig. 3.16 which represents the mean view on (xy), (xz) and (yz) planes, as well as a 3D representation of the two vessels. If the whole voxel stands inside the cylinder, all the activity is taken into account but if the voxel stands on the border of the cylinder, its activity is weighted by the intersecting volume of the voxel with the cylinder. The object which is

generated is made of two cylinders whose diameter is 5 mm and their length is 21 mm. The axis-to-axis distance is 6 mm, which means that only 1 mm separates the surface of these two cylinders. According to results obtained from reconstructions of point-sources, it should be possible to separate them. In the example of Fig.3.16, every voxel which is inside a cylinder emits 10^{11} photons.

3.3.3.1 Reconstruction of 2 cylinders from a single noise-free projection

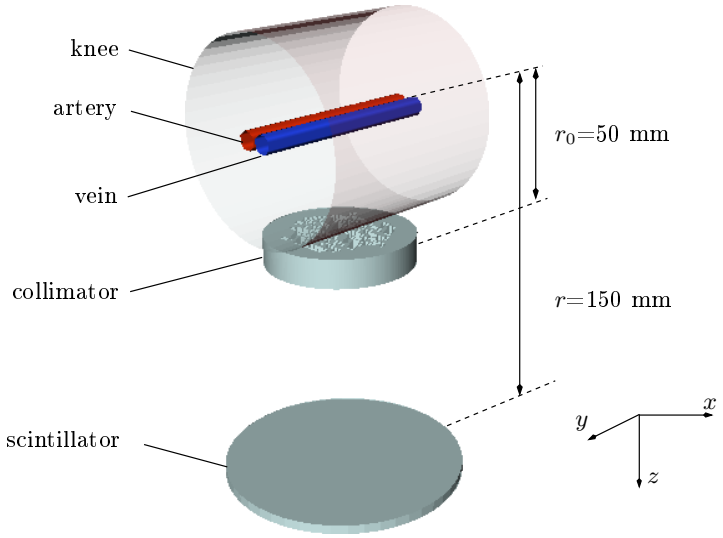
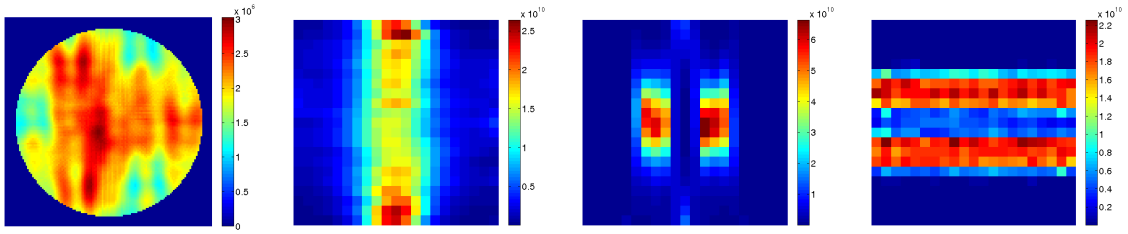


Figure 3.17 – Configuration for acquiring a single projection



(a) Noise-free projection (b) Mean view in the (yz)-plane (c) Mean view in the (zx)-plane (d) Mean view in the (xy)-plane

Figure 3.18 – Reconstruction of 2 cylinders from a single noise-free projection

A projection is generated from this spatial distribution by multiplying the reference distribution $\boldsymbol{\lambda}_{ref}$ with the system matrix $\boldsymbol{A}$ of the imaging system, as for generating the projection corresponding to a point-source in the previous section. The noise-free projection which is generated from this object is presented in Fig. 3.18a and it allows the reconstruction which is presented in Fig. 3.18b, 3.18c and 3.18d. Table 3.2 presents the number of iterations, the computation time and the log-likelihood at every binning level of the reconstruction. These three images are obtained by computing the mean of all the slices of the reconstructed distribution in directions x, y and z. As a result, estimated activities in voxels should be invariant along the axis of the cylinders, which is not the case: the reconstruction suffers from a poor homogeneity of the reconstruction inside cylinders, especially in the (yz)-plane. This raises a question about the quantification ability of the imaging system. Fig. 3.19 is a histogram of the concentration in the voxels which stand inside the cylinders.

binning level	projection size	iterations	time (s)	log-likelihood
8	8×8	339200	3538	3327563117259
4	16×16	403072	6321	3327696329579
3	32×32	431744	10177	3327710233799
2	64×64	437556	14025	3327711270260
1	128×128	437557	14028	3327711270413

Table 3.2 – Evolution of the reconstruction of 2 cylinders from a single noise-free projection

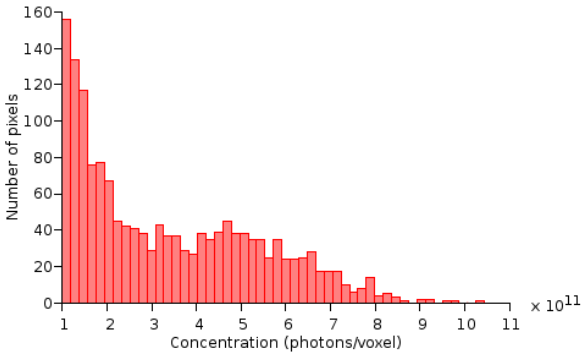


Figure 3.19 – Distribution of the activity in the cylinders

It would be expected that a large amount of voxels has an activity concentration which is 10^{11} photons/voxel rather than such a distribution where only a short number of voxels reaches this concentration. Nevertheless, there is as much activities in the reconstructed cylinders than in the initial distribution since 30.0×10^{12} photons are emitted per cylinder in the initial distribution in Fig. 3.16 and it is estimated that the two cylinders in Fig. 3.18 emit respectively 30.0×10^{12} and 31.1×10^{12} photons. The error between the reconstructed and the expected spatial distribution is 0.135, as derived from Eq. (3.24).

The cylindrical structures are observable and well separated in Fig. 3.18. The cylinders in Fig. 3.18c have an ellipsoidal shape, which underlines a poor spatial resolution in the z-direction. FWHM are estimated from this figure. Its mean activity in the two main directions (horizontal and vertical) are represented in Fig. 3.20 and they are fitted with Gaussian functions in order to retrieve the standard deviations of these two curves and consequently their FWHM. For instance, the FWHM in the transverse and in the axial direction are respectively 3.0 and 6.8 mm in Fig. 3.20. The transverse FWHM is the mean of the FWHM of the two peaks on the left figure. This should be compared to 3.1 and 3.1 mm

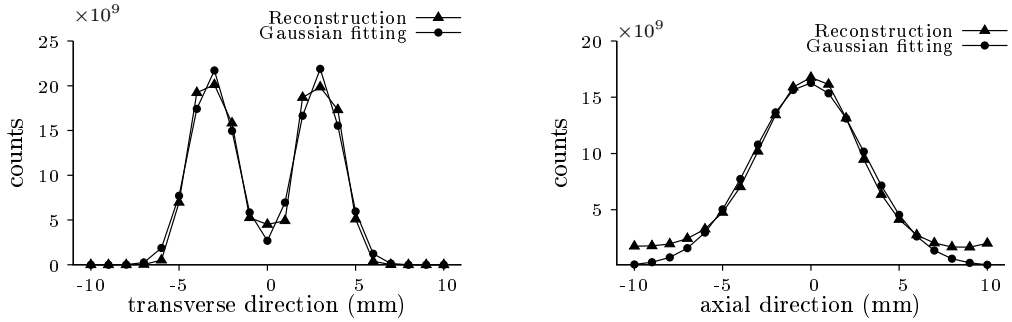
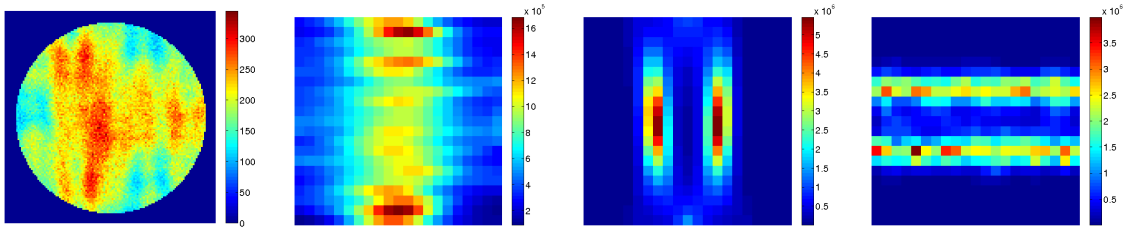


Figure 3.20 – Estimation of the FWHM in the transverse and axial directions from Fig. 3.18c

for FWHM in x and z-directions in Fig. 3.16c, the differences being respectively 3% and 120%. This means that the spatial distribution of the two cylinders are well estimated in the transverse direction, but suffers from a poor spatial resolution in the axial direction. These FWHM should not be confused with the resolution of the imaging system, these FWHM are the profiles of cylinders.

As the projection is noise-free, these results on the quantification ability of the imaging system, on the reconstruction error and on the FWHM present systematic errors which are the limits of this imaging system when associated with a statistical reconstruction algorithm as MLEM.

3.3.3.2 Impact of Poisson fluctuations on reconstructions of 2 cylinders



(a) Initial projection. 2.1×10^6 photons are detected

(b) Mean view of the distribution in the yz-plane

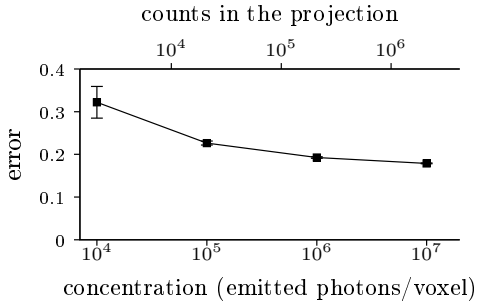
(c) Mean view of the distribution in the zx-plane

(d) Mean view of the distribution in the xy-plane

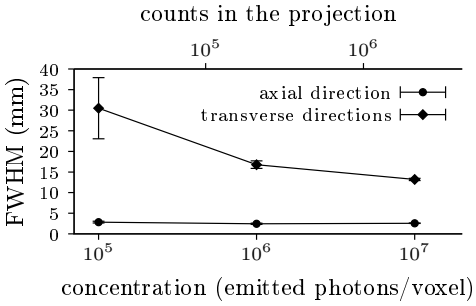
Figure 3.21 – 3D spatial distribution obtained from a unique projection. Distributions in Fig. 3.21b, 3.21c and 3.21d are the mean of all slices in the (yz), (zx) and (xy)-planes. The concentration in cylinders is 10^7 photons/voxel.

An object made of two cylinders is simulated in the field of view of the imaging system. Its spatial distribution is similar to Fig. 3.16. The projection in Fig. 3.21a is reconstructed with the accelerated MLEM of section 3.2. It leads to the 3D spatial distribution presented in Fig. 3.21b, 3.21c and 3.21d. Qualitatively it is observed that the transversal resolution is correct, but the resolution in the direction normal to the surface of the collimator is far above the millimetre. This is quantified by reconstructing spatial distributions for concentrations ranging from 10^4 to 10^7 photons/voxel. For every concentration, 5 independent projections are generated in order to obtain the mean and the standard deviation of every investigated parameters. The influence of the concentration in cylinders on the reconstruction errors, which are calculated according to Eq. (3.24), is presented in Fig. 3.22a. As expected, the error decreases for large concentrations, down to 0.18, the error from noise-free projection being 0.135. When the concentration is set to 10^4 photons/voxel, the amount of detected photons is not sufficient to accurately reconstruct the spatial distribution. This can be observed in Fig. 3.22a, 3.23a and 3.23b, where the error bars are large for the left-hand point. Reconstructions with this concentration are not presented in Fig. 3.22b since it is not possible to fit Gaussian functions in Fig. 3.21c to estimate the FWHM of reconstructed distributions. When the concentration increases, the FWHM is almost constant in the transverse direction (2.5 mm) whereas it decreases in the axial direction (from 30 to 13 mm). The error bars of the left-hand point of the axial FWHM is large (6 mm), which indicates this result is not reproducible.

Quantification of the concentration in cylinders is then performed. Fig. 3.23a shows the ratio of the maximum reconstructed concentration in the two cylinders over the concentration in the reference object. It is observed that this ratio is far above 1 for low concentration. This

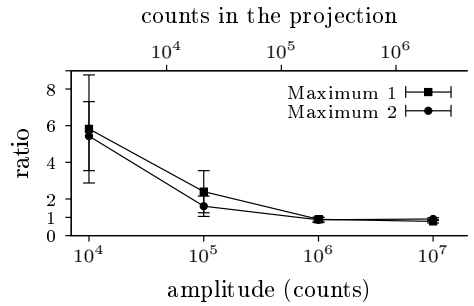


(a) Error between the reconstructed spatial distribution and the reference object

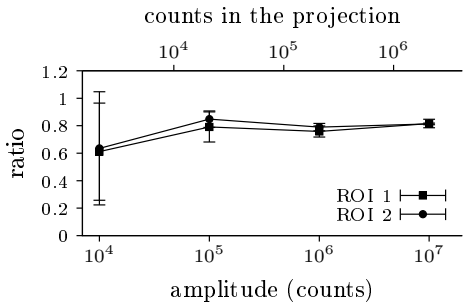


(b) Impact of the Poisson statistics on the FWHM of reconstructed points for a single projection

Figure 3.22 – Characteristics of the reconstructed cylinders



(a) Estimated maximum concentration in each cylinder



(b) Estimated mean concentration in each cylinder

Figure 3.23 – Characteristics of the reconstructed cylinders

can be explained by important inhomogeneities for low concentration which results in over-estimations of some voxels in cylinders, as shown for instance in Fig. 3.21d. It converges to 0.85 for high concentrations, which is an under-estimation of the real concentration. It can be explained by the spatial resolution of the imaging system which is not perfect, which results in a spread of the cylinders with a diminution of the amplitude of the concentration. Instead of quantifying the concentration from the maximum reconstructed concentration, it can be derived from the mean concentration in a ROI defined on every cylinder. Fig. 3.23b shows that the concentration of cylinders is better estimated from the ROI-based ratio than from the maximum concentration. Indeed it is less sensitive to variation of concentration in cylinders and the estimations are more reproducible since error bars are in general smaller. The ratio when the concentration in cylinders is 10^7 photons/voxel is 0.82, which is 3% worse than for Fig. 3.23a, but it can be easily calibrated since it does not depend on the concentration for concentration higher than 10^5 photons/voxel.

Reconstructions suffer from a poor resolution in the axial direction, especially for low concentrations. The quantification performances are also affected by low concentrations. As a result, it is suggested to consider a second projection. It should be orthogonal to the first one so that the two projections are complementary.

3.3.3.3 Reconstructions of 2 cylinders from 2 analytical projections

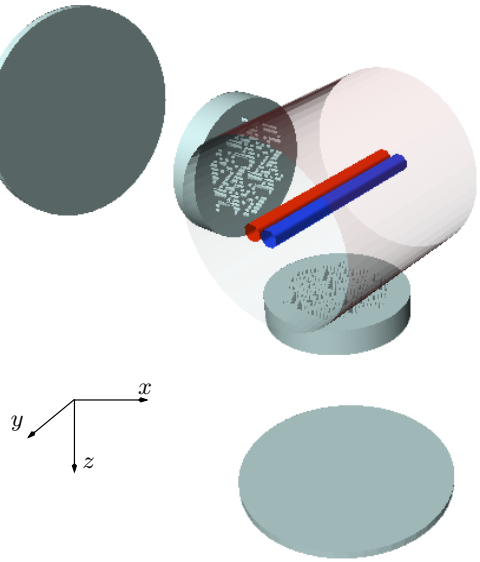
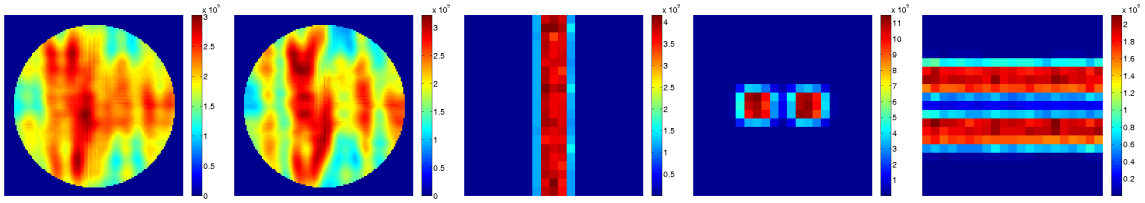


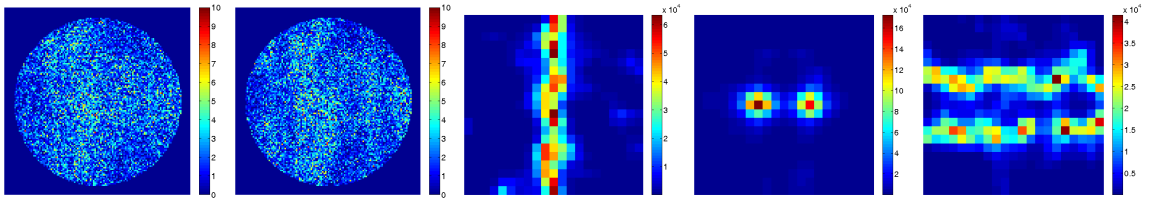
Figure 3.24 – Configuration for acquiring two orthogonal projections.



(a) Noise-free projections (b) Mean views of the reconstructed distribution

Figure 3.25 – Reconstruction of 2 cylinders from 2 noise-free projections

The second set of detector and coded-aperture collimator could not have stood in the axis of the vessels because this can not be achieved in practice, the leg continuing in the same direction as the vessels. The axis of the leg is denoted $\boldsymbol{e}_y$ and consequently the second acquisition system is obtained by rotating the first one by 90° around the y-direction. The distribution of the two cylinders are reconstructed from two noise-free projections, shown in Fig. 3.25a. The diameter of the cylinders is 5 mm and the axis-to-axis distance is still 6 mm. The mean views are presented in Fig. 3.25b. This distribution is very similar to the expected distribution of Fig. 3.16. The error between these two distributions is 4.4×10^{-2}. It is the limit error which can be expected from the reconstruction of 2 cylinders from 2 projections. Results which are presented in next sections can not have errors below this limit, because reconstructions from analytical projections do not suffer from modelling issues of the imaging system by the system matrix, projections being generated with the same matrix as the one used to reconstruct spatial distributions. The maximum reconstructed concentration in the two cylinders are respectively 12.6×10^9 and 12.7×10^9 photons/voxel, which are overestimations of the initial distribution of the two cylinders since their concentration is in reality 10^{10} photons/voxel for this example. The total activity on a ROI defined on 95% (2σ) of the distribution for the two cylinders are respectively 3.03×10^{12} and 3.09×10^{12} photons. As it is expected from the initial object of Fig. 3.16 that each cylinders emits



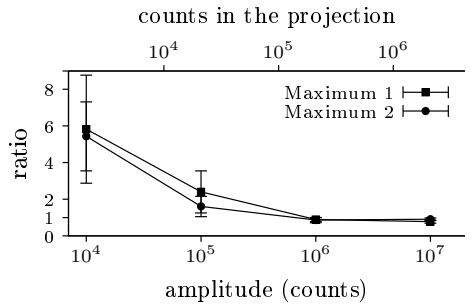
(a) Two orthogonal analytic projections (b) Mean views of the reconstructed distribution

Figure 3.26 – Reconstructed distribution of 2 cylinders from 2 analytical projections. The number of detected photons are respectively 21308 and 21223 photons.

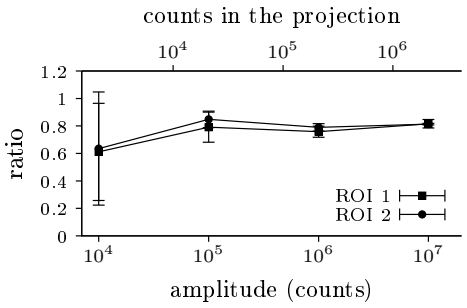
3.06×10^{12} photons, these results show that the concentration in cylinders can be quantified accurately. The FWHM in the x and z-directions are respectively 3.09 and 3.07 mm and it was expected that it was 3.12 mm (<2%) and 3.07 mm (<4‰), which confirms the good performances of reconstructions from two noise-free projections. Poisson noise is then added to projections and an example is shown in Fig. 3.26a for the concentration in cylinders being 10^5 photons/voxel. The result of the reconstruction from these two projections is shown in Fig. 3.26b.

As it is expected to estimate the concentration in vessels, the quantification ability of the imaging system is investigated for 4 activities (from 10^4 to 10^7 photons/voxel). First, it is observed in Fig. 3.27a that the maximum reconstructed concentration in cylinders is not an appropriate estimator since it largely overestimates the reference concentration, the best ratio being 1.9. This is due to inhomogeneities in the reconstructed volumes, as it can be observed in Fig. 3.26b. A better estimator can be derived from the total activity which is reconstructed in every cylinder. Fig. 3.27b shows the ratio of the reconstructed total activity in each cylinder over the total activity in each cylinder of the reference object in Fig. 3.16. The reconstructed total activities for cylinders 1 and 2 are derived from a 3σ-ROI around each cylinders. Ratios are almost constant (ratio=0.8 ± 0.2) excepted for the first point which corresponds to a concentration in cylinders which is 10^4 photons/voxel. The reconstruction error is plotted in Fig. 3.28a. It decreases when the concentration increases, without converging to a limit. For a concentration of 10^7 photons/voxel in cylinders, the error is 7.7×10^{-2}, which should be compared to 4.4×10^{-2} when reconstructions are performed from 2 noise-free projections. The quantification of the concentration in cylinders is reliable from concentration being larger than 10^5 photons/voxel. It was expected that the quantification should be extracted from reconstructions where the concentration is 8000 photons/mm^3, which means that concentrations in vessels can not be quantified with this imaging system, unless the activity which is injected to patients is 12 times higher, which is not feasible because of radioprotection considerations.

Fig. 3.28b shows the FWHM in the x and z-directions. When compared to Fig. 3.22b where only one projection was used to reconstruct the two cylinders, the spatial resolution in the z-direction is improved. For instance, when the concentration in cylinders is 10^7 photons/voxel, the FWHM in the z-direction decreases from 13 to 2.8 mm. The FWHM in the x-direction is 2.8 mm, which is larger than the reconstruction from a single projection, but closer to reality as the expected FWHM in the x and z-directions is 3.1 mm (see section 3.3.3.1). This shows that it is possible to estimate the spatial distribution of the two vessels from a very limited amount of detected photons if two orthogonal projections are acquired.

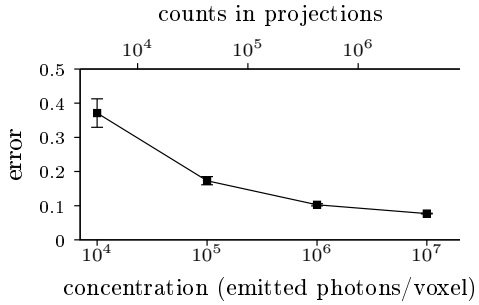


(a) Maximum of the activity concentration in each cylinder

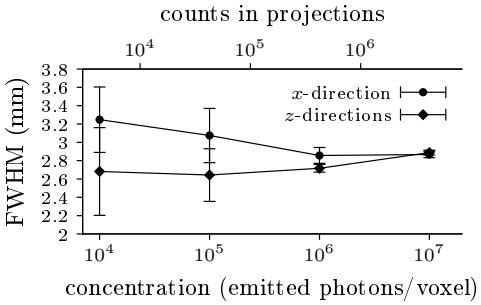


(b) Mean activity concentration in each cylinder

Figure 3.27 – Quantification ability of the imaging system for 2 cylinders in the field of view



(a) Error between the reconstructed spatial distribution and the reconstructed object



(b) Estimated FWHM

Figure 3.28 – Accuracy of reconstructions

Using analytically-generated projections allows the investigation of the impact of the concentration of the vessels on the quality of the reconstructed spatial distribution, as no modelling problem exists since the observations are generated from the system matrix used to reconstruct the spatial distributions. It was demonstrated that it is possible to reconstruct accurately spatial distributions from noise-free projections. The impact of Poisson fluctuations affects strongly the spatial resolution in the z-direction, but this can be compensated by using a second projection, orthogonal to the first one. Such a configuration with two projections allows the accurate reconstruction of point-sources and objects made of 2 cylinders. These results are the limit performances, since the projections are generated from the system matrix used to reconstruct objects, there is consequently no discrepancy between the projections and the system matrix.

3.4 Reconstructions of Monte-Carlo simulated projections

Results which are presented in this section are obtained from the simulations of two orthogonal projections with the particle transport code GATE (Jan 04). The objective is to demonstrate that spatial ditributions can be accurately reconstructed although the system matrix does not model exactly the imaging system. Every projection is simulated independently according to Fig. 3.29, which models the configuration with two projections already shown in Fig. 3.24. For all Monte-Carlo simulations presented below, the energy resolution is set to 5.8% at 511 keV and the energy threshold is set to 470 keV which corresponds to 99% (3σ) of the photoelectric peak.

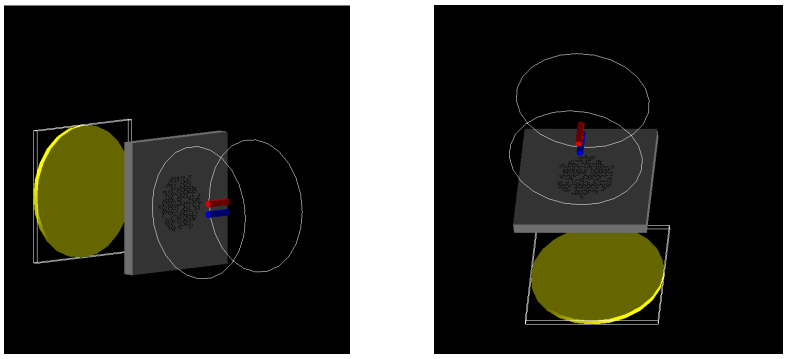
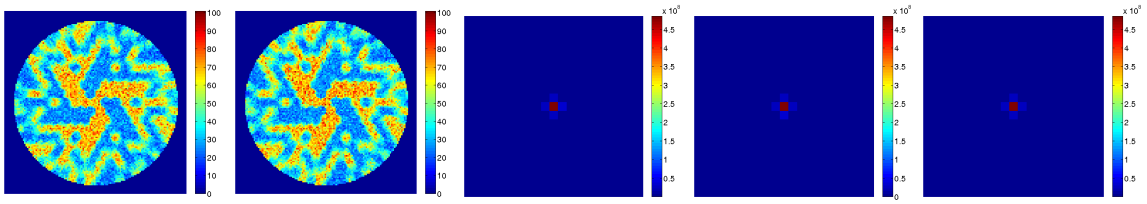


Figure 3.29 – Acquisition of two orthogonal projections

3.4.1 Reconstruction of a point-source from 2 Monte-Carlo projections

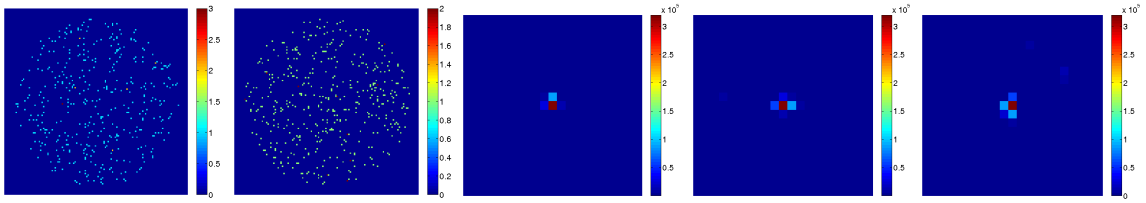
The characteristics of reconstructed points are evaluated from reconstructions of point-sources. First, Compton scattering is switched off in Monte-Carlo simulations, in order to validate results which were previously obtained with analytical projections. Then Compton scattering is added to simulations and finally characteristics of reconstructions are performed for 4 different intrinsic spatial resolutions of the detector (from 0 mm to 3 mm).

3.4.1.1 Coherence with analytical projections



(a) Monte-Carlo projections (b) Middle slices

Figure 3.30 – Example of projections and the corresponding reconstructed point. The point-source emits 10^9 photons. The number of detected photons are respectively 456213 and 460665 in Fig. 3.30a.



(a) Monte-Carlo projections (b) Middle slices

Figure 3.31 – Example of projections and the corresponding reconstructed point. The point-source emits 10^6 photons. The number of detected photons are respectively 425 and 452 in Fig. 3.31a.

First, no Compton scattering is taken into account into Monte-Carlo simulations. Projections are simulated for activities of the point-source varying from 10^6 to 10^9 photons. The intrinsic spatial resolution of the detector is for the moment set to 0 mm. Two examples

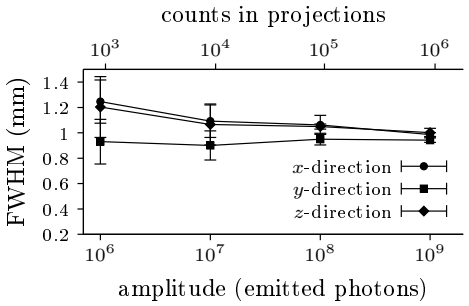


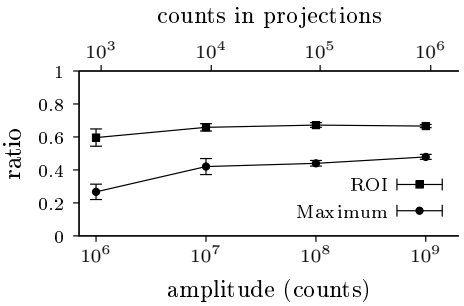
Figure 3.32 – The FWHM of reconstructed points depend on the activity of the point-source

are shown in Fig. 3.30a and 3.31a corresponding to these limit activities. The result of the reconstructions associated with these projections are respectively presented in Fig. 3.30b and 3.31b. Qualitatively, it is observed that the reconstructed point is very sharp for activity being 10^9 photons in the point-source, but tails appear for smaller activities, indicating a degradation of the spatial resolution of the imaging system. It also induces a stronger bias on the maximum reconstructed activity. This is investigated in more details in the next paragraph.

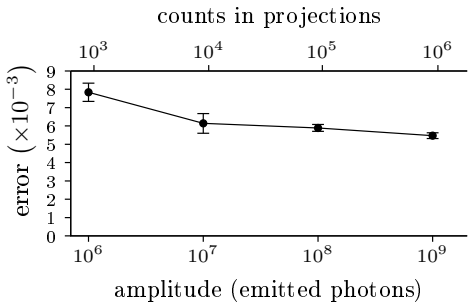
The FWHM of the 3D reconstructed points are larger than for analytical projections as shown in Fig. 3.32, but does not exceed 1.3 mm even for low activity point-sources. The FWHM in the y-direction is slightly better than in the two other directions which can be explained by the fact that the y-direction is always a transverse direction to the simulated orthogonal projections. All FWHM converge to 1 mm for point-sources with high activity whereas it was 0.74, 0.66 and 0.76 mm for analytical projections (see section 3.3.2.3). Although it corresponds to a degradation of the spatial resolution of the imaging system, these FWHM are comparable and close to 1 mm which is the size of voxels and the spatial resolution which is expected to obtain.

As it can be observed in Fig. 3.33a, the reconstructed amplitude of peaks is biased, since the ratio between the maximum reconstructed activity and the real activity in the point-source should be 1. Biases decrease as the number of photons in the simulations increases, but is at best 0.48. As for reconstructions from analytical projections, this can be explained by the spatial spread of the reconstructed point-source. By defining a ROI around the peak (2σ), and summing all activities of voxels which stand in this ROI, a better estimation of the activity of the point-source is obtained, the ratio reaching 0.67 at best, instead of 1. This may be explained by the fact that the energy resolution was set to 5.8% at 511 keV and the energy cut-off was set to 470 keV. The main difference between the ratio based on the maximum reconstructed activity and the ratio based on a ROI is that the latter is almost constant (ratio=0.65 ± 0.04) for a large range of activities of the point-source. It was investigated why this ratio is not equal to 1, but no explanation has been found that would explain this discrepancy. If knowing this ratio, reconstructions can be easily calibrated: a correction factor of 1.5 might be applied in order to retrieve the amplitude of the point-source from the total reconstructed activities in the ROI around the reconstructed point. In this case, reconstructions might become quantitative, althougth the estimation of the correction factor would not be trivial.

Fig. 3.33b shows the reconstruction errors as a function of the activity of the point-source. Although the error decreases as the activity increases, the error is far larger for reconstructions which are based on Monte-Carlo simulations than for reconstructions based on analytical projections shown in Fig. 3.28a. For instance, when the activity of the point-source is 10^9 photons, the error is 3.9×10^{-4} for reconstructions from two analytical projections



(a) Reconstructed maximum intensity and reconstructed intensity in the region of interest



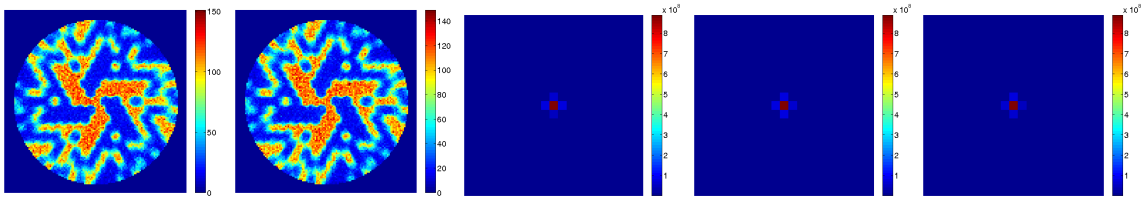
(b) Error between the reconstructed spatial distribution and the reference object

Figure 3.33 – Accuracy of the reconstructed spatial distributions

whereas it is 5.3×10^{-3} from Monte-Carlo simulations.

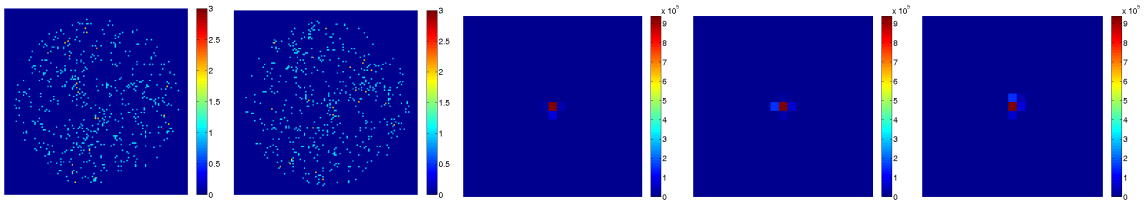
Although performances are not as good as for reconstructions based on analytical projections, the results presented here are still correct. FWHM are 20% worse and a correction factor $F = 1.5$ might be used to quantify the activity of point-sources. Nevertheless results are reproducible even for point-sources with low activity.

3.4.1.2 Impact of Compton scattering



(a) Monte-Carlo projections (b) Middle slices

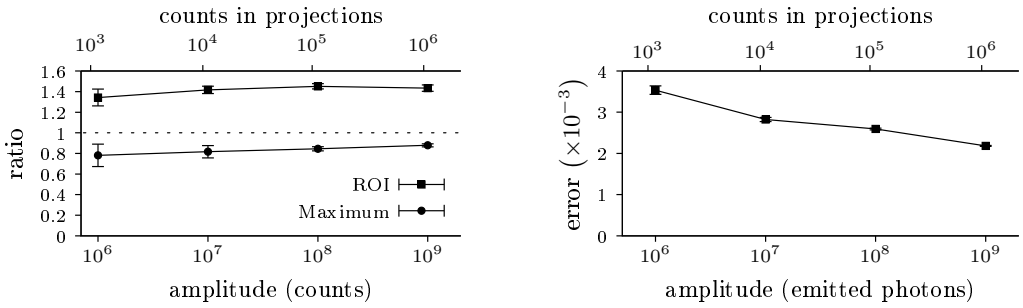
Figure 3.34 – Example of projections and the corresponding reconstructed point for the spatial resolution being SR=3 mm. The point-source emits 10^9 photons. The number of detected photons are respectively 563600 and 556785 in Fig. 3.34a



(a) Monte-Carlo projections (b) Middle slices

Figure 3.35 – Example of projections and the corresponding reconstructed point for the spatial resolution being SR=3 mm. The point-source emits 10^6 photons. The number of detected photons are respectively 592 and 577 in Fig. 3.35a

In reality, it is not possible to discriminate all photons which undergo Compton scattering. Compton scattering is consequently added to Monte-Carlo simulations in order to investigate



(a) Reconstructed maximum intensity and reconstructed intensity in the region of interest

(b) Error between the reconstructed spatial distribution and the reference object

Figure 3.36 – Accuracy of the reconstructed spatial distributions

its impact on reconstructions. Two examples are shown in Fig. 3.34 and 3.35 for high and low-activity point-sources (10^9 and 10^6 emitted photons). For both activities, reconstructions appear to be correct. The error, the FWHM and the amplitude of the reconstructed point are systematically investigated for every reconstruction. Reconstructions are performed for 4 activities (from 10^6 to 10^9 emitted photons). For every activity, 5 independent pairs of projections are generated through Monte-Carlo simulations, in order to estimate the mean and the standard deviation of every parameter of interest.

Neither the maximum reconstructed activity nor the total activity in a ROI defined around the reconstructed points succeed in quantifying the amplitude of the point-source, as shown in Fig. 3.36a which represents theses quantities over the expected activities. Contrary to previous reconstructions, it seems that the maximum reconstructed activities better estimate the expected concentration than the total activity on a ROI defined around the point-source. In both cases, reconstructions should be calibrated, which seems easily feasible as the ratios are almost constant (0.8 for the ratio corresponding to maximum activities and 1.4 for the ratio calculated from the ROI). Up to now, ratios were always smaller than 1 for reconstructions of point-sources. The correction factors would be respectively 1.25 and 0.7 to correct the two ratios of Fig. 3.36a. This difference can be explained by approximations made for computing the system matrix. It was indeed assumed that system matrices would only be modelled by photons which would deposit all their energy at once in the detector, which means in practice that the linear attenuation coefficient for the collimator and the detector are the photoelectric attenuation coefficients μ_W^{PE} and μ_{CsI}^{PE} (see section 2.2.3.1). In reality, some photons deposit all their energy in the scintillator through several interactions and it is not possible to discriminate them. As a consequence, too many photons are detected, which results in an overestimation of the total activity in the peak. Fig. 3.36b shows the reconstruction errors. Reconstructions are very reproducible since the error bars are small when compared to the mean errors. Furthermore, mean errors decrease as the activity of the point-source increases. Adding Compton scattering results in smaller reconstruction errors than for without it. This can be explained by the fact that correction factors are not included in the calculation of errors. As the maximum reconstructed activities are closer to the real activity when Compton scattering is modelled, this results in smaller errors.

Finally, the FWHM are represented in Fig. 3.37. It ranges from 0.88 to 1.17 mm. For the activity of the point-source being 10^9 photons, the FWHM in the x, y and z-directions are respectively 1.05, 1.00 and 1.06 mm, whereas it was about 1 mm for projections generated without Compton scattering. This means that Compton scattering does not affect a lot the spatial resolution of the imaging system.

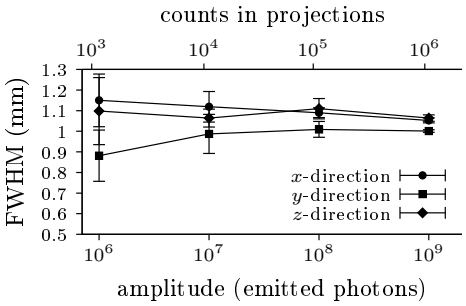


Figure 3.37 – The FWHM of reconstructed points depend on the activity of the point-source

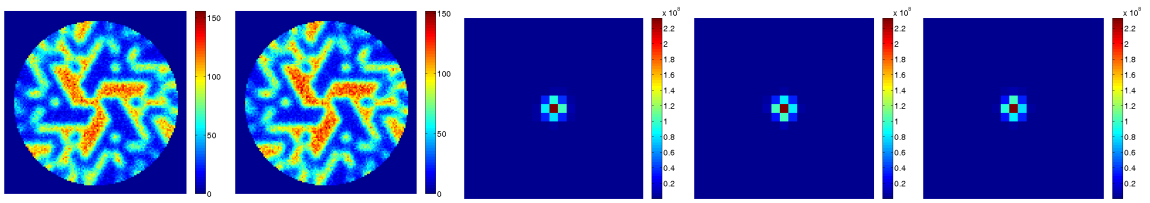
Adding Compton scattering does not affect significantly reconstructions. The correction factor might be modified from $F = 1.5$ to $F = 0.7$ but it is still possible to quantify the activity of point-sources. The FWHM are not impacted by Compton scattering.

3.4.1.3 Impact of the intrinsic spatial resolution of detectors

Up to this point, it is assumed that the intrinsic spatial resolution (SR) of the detector is perfect, although this resolution is 2 mm at 140 keV (Russo-Marie 07). The intrinsic spatial resolution is the standard deviation of the spot which is measured when a beam of photons is impinging the detector with a normal incidence. It is expected that the intrinsic spatial resolution of the detector impacts the spatial resolution of the imaging system. With Monte-Carlo simulations, it is possible to generate projections with a defined intrinsic spatial resolution. As a result, projections are generated with the detector spatial resolution varying from 0 mm to 3 mm.

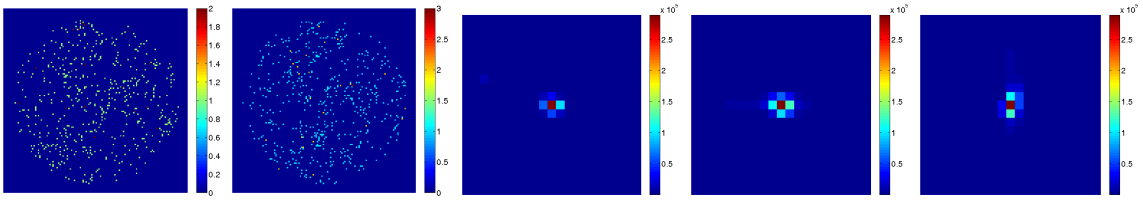
Fig. 3.38a and 3.39a show 4 projections where the spatial resolution of the detector is set to 3 mm in the Monte-Carlo simulations. Fig. 3.38b and 3.39b are the results of the reconstructions based on these two sets of projections. It is observed that reconstructed point-sources have tails, which were almost invisible in previous examples. Fig. 3.40a shows the impact of SR on the error of the reconstructed distributions, according to Eq. (3.24). It is observed that the error increases as SR decreases, but the order of magnitude is kept similar (from 2.2×10^{-3} for SR=0 mm to 8.4×10^{-3} for SR=3 mm, the activity of the point-source being 10^9 photons). The mean error does not decrease when the activity increases, which means that these errors are systematic errors and does not depend on the activity of the point-source. Reproducibility of reconstructions is better for high-activity point-sources as indicated by the small error bars.

Fig. 3.40b shows the influence of SR on FWHM of the reconstructed point-sources. For



(a) Monte-Carlo projections (b) Middle slices

Figure 3.38 – Example of projections and the corresponding reconstructed point-source for the spatial resolution being SR=3 mm. The point-source emits 10^9 photons. The number of detected photons are respectively 571388 and 577765

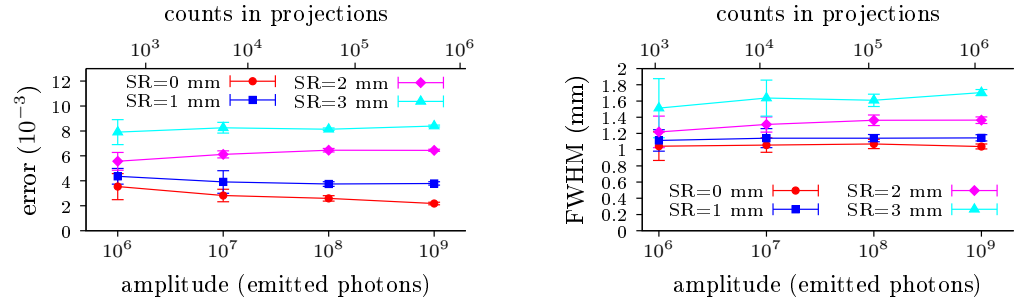


(a) Monte-Carlo projections (b) Middle slices

Figure 3.39 – Example of projections and the corresponding reconstructed point-source for the spatial resolution being SR=3 mm. The point-source emits 10^6 photons. The number of detected photons are respectively 539 and 572

every reconstruction, the FWHM is computed by taking the mean of the FWHM in the three direction x, y and z. The FWHM is 1 mm when the detector spatial resolution is 0 mm, but it increases to 1.15, 1.30 and 1.60 mm for detector spatial resolution being respectively 1 mm, 2 mm and 3 mm.

Finally, the impact of the detector spatial resolution on the quantification of the point-source is investigated. Fig. 3.41a shows that the maximum reconstructed activities is very dependent on the detector spatial resolution whereas the total activity on a ROI defined on the reconstructed point-source is far less sensitive, as shown in Fig. 3.41b. Although these two curves are biased since they should be equal to 1, it might be possible to apply a correction factor to the ROI-ratio, which is almost constant (1.4±0.03). This factor does not depend on the detector spatial resolution. The correction factor would be 1/1.4=0.7, as in section 3.4.1.1.

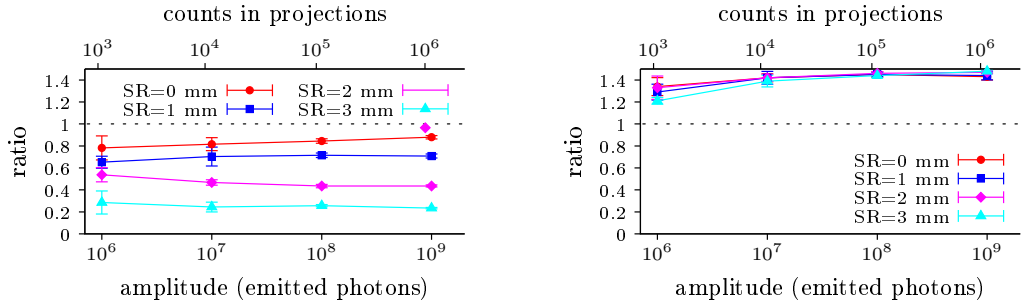


(a) Error between the reconstructed spatial distribution and the reference object

(b) FWHM of the reconstructed point-sources do not depend on the intrinsic spatial resolution of the detectors

Figure 3.40 – Impact of the intrinsic spatial resolution of the detector

The reconstructions of point-sources from 2 Monte-Carlo projections result in small errors. The main difference with reconstructions based on analytically-generated projections is the bias between the maximum reconstructed activities and the simulated activities. This is explained by the fact that the system matrix does not model Compton scattering, which results in a larger number of detected photons in Monte-Carlo projections than in analytically-generated projections. However, this difference can be corrected since the bias is very similar for all activities of the point-source. Moreover, the impact of the detector spatial resolution



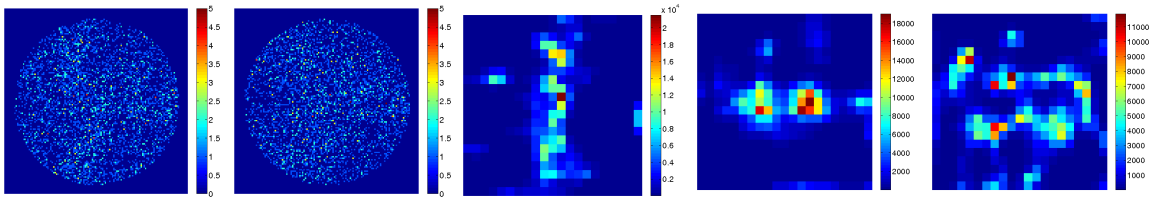
(a) Ratio of the reconstructed maximum activity over the activity of the point-source in simulations

(b) Ratio of the reconstructed maximum activity over the activity of the point-source in simulations

Figure 3.41 – Impact of the intrinsic spatial resolution on quantification

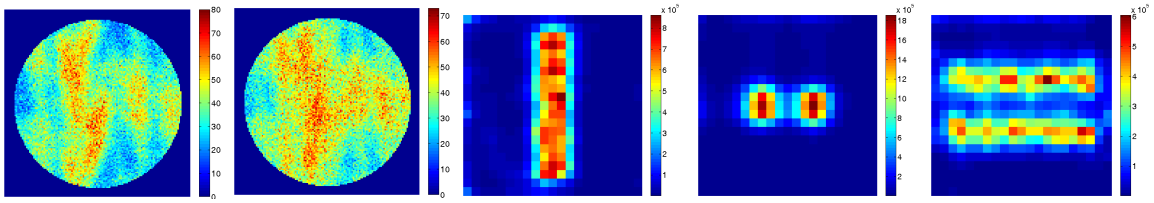
on the reconstructions was investigated. Large spatial resolutions of the detector induce a broadening of the reconstructed point-source and smaller maximum reconstructed activities. However the total activity in the reconstructed point-source does not depend on the detector spatial resolution, consequently it might possible to quantify the activity of the point-source from this estimator by applying a correction factor F which depend neither on the resolution of the detector nor the activity of the simulated point-sources.

3.4.2 Reconstruction of 2 cylinders from 2 Monte-Carlo projections



(a) 2 Monte-Carlo projections (b) Mean yz-plane (c) Mean zx-plane (d) Mean xy-plane

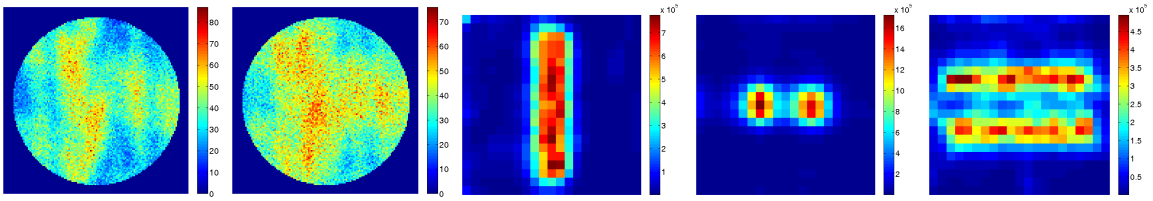
Figure 3.42 – Reconstructed distribution for concentration being 10^4 photons/voxel in Monte-Carlo simulations, the number of detected photons in projections being respectively 4070 and 3901 (Fig. 3.42a). The detector spatial resolution is 0 mm.



(a) 2 Monte-Carlo projections (b) Mean yz-plane (c) Mean zx-plane (d) Mean xy-plane

Figure 3.43 – Reconstructed distribution for concentration being 10^6 photons/voxel in Monte-Carlo simulations, the number of detected photons in projections being respectively 393979 and 409403. The detector spatial resolution is 0 mm.

The performance of the imaging system is then investigated with Monte-Carlo simulations of 2 cylinders emitting 511-keV photons. The axis-to-axis distance is 6 mm. Results



(a) 2 Monte-Carlo projections (b) Mean yz-plane (c) Mean zx-plane (d) Mean xy-plane

Figure 3.44 – Reconstructed distribution for concentration being 10^6 photons/voxel in Monte-Carlo simulations, the number of detected photons in projections being respectively 393314 and 408781. The detector spatial resolution is 3 mm.

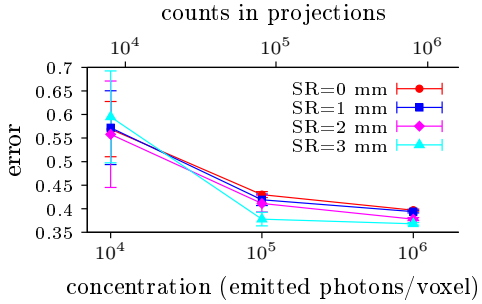
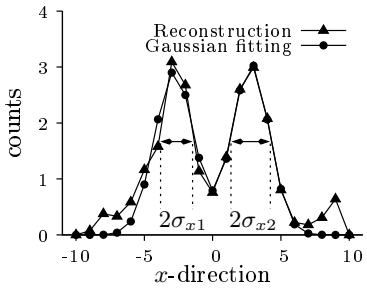


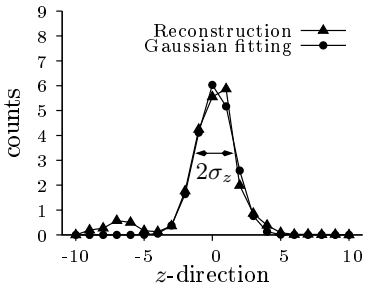
Figure 3.45 – Impact of the concentration in cylinders on reconstruction errors, with the intrinsic spatial resolution (SR) of the detector varying from 0 mm to 3 mm

are presented for 3 concentrations in cylinders (from 10^4 to 10^6 photons/voxel) and for 4 intrinsic spatial resolutions of the detectors (from 0 to 3 mm). Every projection is simulated 5 times in order to obtain the mean and the standard deviation of every parameter of interest of the correpsonding reconstructed point-sources (FWHM, error, estimated concentrations). Fig. 3.42a shows an example of projections obtained from Monte-Carlo simulations, where the concentration in cylinders is 10^4 photons/voxel. The number of detected photons in projections are respectively 4070 and 3901. For this example, the intrinsic spatial resolution of the detector is supposed to be perfect (SR=0 mm). The reconstruction from these projections leads to Fig. 3.42b, 3.42c and 3.42d. The reconstructed distribution is not homogeneous inside the cylinders, furthermore cylindrical structures are hardly visible. When the concentration in cylinders is 100 times larger, the reconstructed spatial distribution estimates correctly the original distribution as it is qualitatively shown in Fig. 3.43. A last example shows the impact of the detector spatial resolution in Fig. 3.44, where it is set to 3 mm. When compared to the previous example where the spatial resolution was set to 0 mm, it can be seen that the two cylinders are less separated. The reconstruction error is shown in Fig. 3.45. It is observed that the mean error decreases when the concentration increases. Furthermore, results become more and more reproducible as the error bars also decreases when the concentration increases. It is surprising that the errors associated to large spatial resolutions of the detector are better than errors associated to good spatial resolutions, especially for high concentrations. It would be expected that a large spatial resolution of the detector would deteriorate the reconstructions.

As reconstructions may suffer from important inhomogeneities, the FWHM of the cylinders and their concentrations are estimated from the mean view of the (xz)-plane. The mean views corresponding to Fig. 3.42c are shown in Fig. 3.46. Such a figure also shows the fitting of Gaussian functions on the mean projected concentrations of Fig. 3.42c. The

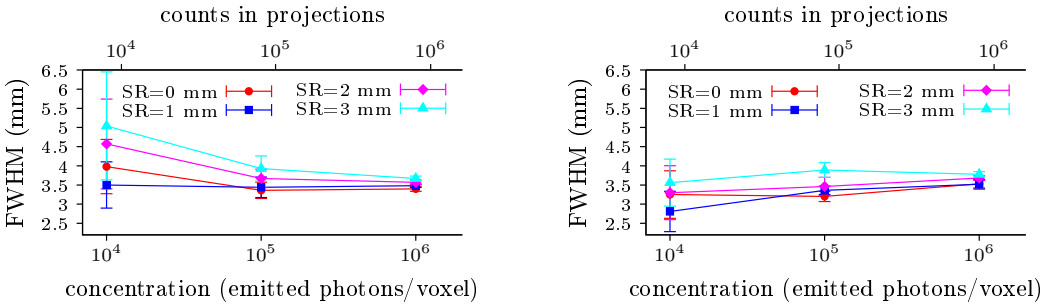


(a) Fit of 2 Gaussian functions on the mean distribution in the x-direction



(b) Fit of a Gaussian function on the mean distribution in z-direction

Figure 3.46 – Mean views of Fig. 3.42c in the x and z-directions



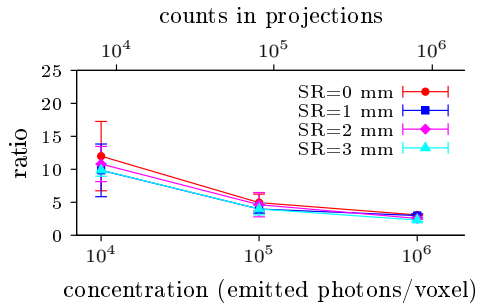
(a) x-direction (b) z-direction

Figure 3.47 – Estimated FWHM of cylinders in the x- and z-directions, with the intrinsic spatial resolution (SR) of the detector varying from 0 mm to 3 mm

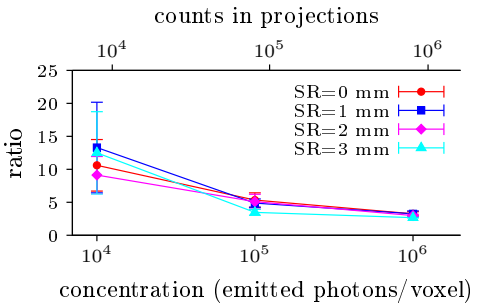
reconstructed points of Fig. 3.46a can be viewed as the mean row of Fig. 3.42c or as the mean column of Fig. 3.42d. The reconstructed points of Fig. 3.46b can be viewed as the mean row of Fig. 3.42b or as the mean columns of Fig. 3.42c. Although the reconstructed distribution is not homogeneous, the fitting of Gaussian functions on mean views in x and z-directions is accurate, although some discrepancies are observable at borders. The FWHM is calculated from the standard deviation of the fitted Gaussian functions σ_{x1}, σ_{x2} and σ_z and the result is presented in Fig. 3.47. It shows that the detector spatial resolution tends to increase FWHM. The impact of SR on FWHM is stronger for low concentrations. Increasing concentrations has opposite effects on the FWHM in the x and z-directions. It tends to decrease the FWHM in the x-direction whereas it increases the FWHM in the z-direction. For high concentration (10^9 photons/voxel), FWHM in x and z-directions respectively converge to 3.6 and 3.5 mm, whatever the detector spatial resolution is. These results are larger than the expected FWHM, which are 3.1 and 3.1 mm (see section 3.3.3.1).

It is then investigated whether the concentration in cylinders could be accurately estimated from reconstructed distributions. It requires that the total activity in every cylinder as well as their diameter are estimated. The maximum reconstructed concentration in each cylinder is not a good estimator of the concentration in cylinders because of important inhomogeneities in the reconstructions, which results in an over-estimation of concentrations in some voxels whereas in some other voxels, the concentration is under-estimated. Fig. 3.48a and 3.48b present the maximal estimated concentration in cylinder 1 and in cylinder 2 for 4 spatial resolution of the detector. It shows that the maximum concentrations largely overestimate the real simulated concentrations but does not strongly depend on the detector spatial resolution. The best ratios are respectively 2.3 and 2.7 for the first and

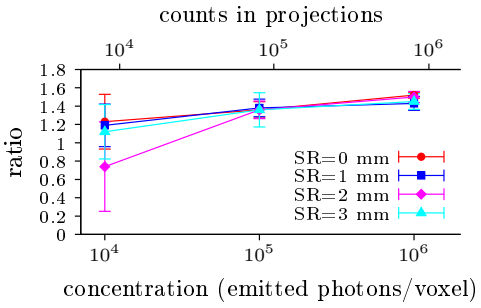
the second cylinders and it corresponds to the largest concentration (10^6 photons/voxel). No interpretation can be done from the lowest concentration (10^4 photons/voxel) since the reconstructions are not reproducible as suggested by the large error bars. These ratios depend on the concentration in the two cylinders and consequently it seems inappropriate to attempt to use a correction factor on such reconstructions, even with a calibration procedure. A more robust estimator of the initial concentration is the mean activity in the region of interest which corresponds to 95% (2σ) of the Gaussian distributions. All the activities in this ROI are summed together and the total activity for every cylinder is then normalised by the volume of cylinders to obtain the mean concentrations. Here it is assumed than the volume of cylinders is known, but in reality it should be estimated from reconstructions. The ratios of these mean concentrations and the expected concentrations are then plotted in Fig. 3.48c and 3.48d. These ratios do not depend strongly on the detector spatial resolution either. As for the study of the reconstructions of point-sources from Monte-Carlo simulations in section 3.4.1.3, it is observed that the concentrations are over-estimated since it would be expected that the ratio would be around 1 whereas its value is 1.4 for high concentrations.



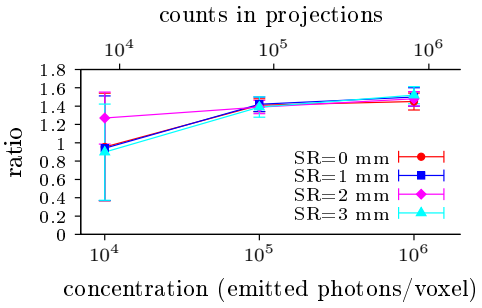
(a) Ratio of the maximum concentration in the first cylinder to the expected concentration



(b) Ratio of the maximum concentration in the second cylinder to the expected concentration



(c) Ratio of the mean concentration in the first cylinder to the expected concentration



(d) Ratio of the mean concentration in the second cylinder to the expected concentration

Figure 3.48 – Quantification ability of the imaging system with the accelerated MLEM-algorithm, the intrinsic spatial resolution (SR) of the detector being varied from 0 mm to 3 mm

This section confirms the results obtained from the reconstructions of two cylinders from analytical projections: the concentration in cylinders can be estimated from reconstructions

as long as a correction factor is applied, which should be set 0.7. This correction factor does not depend on the concentration in cylinders. Furthermore, the impact of the detector spatial resolution was investigated. It deteriorates the estimation of FWHM for low concentration but does not affect the quantification of the concentration in cylinders. Reconstructions corresponding to concentration below 10^5 emitted photons/voxel are not reproducible. This confirms what was noticed for reconstructions based on analytical projections. Under clinical conditions, it is not possible to quantify the concentration in cylinders which is expected to be 8000 photons/voxel, unless patients are injected with 12 times larger activity (2.2 GBq, 60 mCi), which is not possible because of radioactive protection considerations.

3.5 Reconstructions of real projections

An experimental set-up was designed in order to validate the previous numerical results. We used the γ-imager described in section 2.1.2.1. This imager can measure the energy of the detected photons so that it is possible to discriminate some of the photons which undergo Compton scattering either in tissues or in the collimator. This camera is calibrated so that it is not possible to detect photons which have an energy higher than 350 keV because of the electronic discrimination. This is a problem for validating our method since it intends at working with photons whose energy is 511 keV.

First, it is explained how we built a collimator with the same transparency for 140 keV photons as tungsten for 511 keV photons. Then point-sources are reconstructed from 2 projections acquired with this experimental set-up and finally, reconstructions of 2 cylinders are performed.

3.5.1 Model of an experiment based on 511-keV photons with an experiment based on 140-keV photons

The available imager was not set to detect 511-keV photons. As a consequence, we built a coded-aperture collimator which has the same attenuation coefficient at 140 keV as tungsten at 511 keV, namely $\mu_{\mathrm{W}}(511\ \mathrm{keV}) = 2.57\ \mathrm{cm}^{-1}$, and we made experiments with either 122-keV or 140-keV photons. The aim is to validate that it is possible to reconstruct the 3D spatial distribution of two vessels with a good spatial resolution although the activity is very low. With such a collimator, we can partially validate our method at 511 keV with a γ-camera which is not set to work at 511 keV, but works well at 140 keV.

The new coded aperture collimator whose total linear attenuation coefficient is expected to be $2.57\ \mathrm{cm}^{-1}$ at 140 keV is made of an alloy of zinc (Zn) and tin (Sn). These metals have a fusion temperature which are respectively 420°C and 232°C, consequently melting them into an alloy is quite simple. From Bragg's rule, the attenuation coefficient of the alloy μ_a can be calculated according to Eq. (3.25) (Leo 94).

$$\mu_a = \rho_a \left(w_{Zn} \frac{\mu_{Zn}}{\rho_{Zn}} + w_{Sn} \frac{\mu_{Sn}}{\rho_{Sn}} \right) \tag{3.25}$$

where μ_a, μ_{Zn} and μ_{Sn} are the attenuation coefficients of the new alloy, zinc and tin at 140 keV, ρ_a, ρ_{Zn} and ρ_{Sn} their densities, and w_{Zn} and w_{Sn} are the weighting fractions of zinc and tin. A good approximation of the density ρ_a is $\rho_a = w_{Zn}\rho_{Zn} + w_{Zn}\rho_{Zn}$. As a result, as $w_{Sn} = 1 - w_{Zn}$, the weighting fraction w_{Zn} which leads to a compound that has a linear attenuation coefficient $\mu_a = 2.57\ \mathrm{cm}^{-1}$ at 140 keV is the solution of a second-order polynomial in w_{Zn}:

$$\mu_a = (w_Z\rho_Z + (1 - w_Z)\rho_S)(w_Z \frac{\mu_Z}{\rho_Z} + (1 - w_Z)\frac{\mu_S}{\rho_S}) \tag{3.26}$$

Figure 3.49 – ZnSn coded collimator

and its solution is:

$$w_{Zn} = \frac{1}{2}\sqrt{\left(\frac{\mu_{Sn}\rho_{Zn}}{\mu_{Zn}\rho_{Sn}-\mu_{Sn}\rho_{Zn}}+\frac{\rho_{Sn}}{\rho_{Zn}-\rho_{Sn}}\right)^2-4\frac{(\mu_{Sn}-\mu_a)\rho_{Zn}\rho_{Sn}}{(\rho_{Zn}-\rho_{Sn})(\mu_{Zn}\rho_{Sn}-\mu_{Sn}\rho_{Zn})}} \quad (3.27)$$

$$-\frac{1}{2}\left(\frac{\mu_{Sn}\rho_{Zn}}{\mu_{Zn}\rho_{Sn}-\mu_{Sn}\rho_{Zn}}+\frac{\rho_{Sn}}{\rho_{Zn}-\rho_{Sn}}\right) \quad (3.28)$$

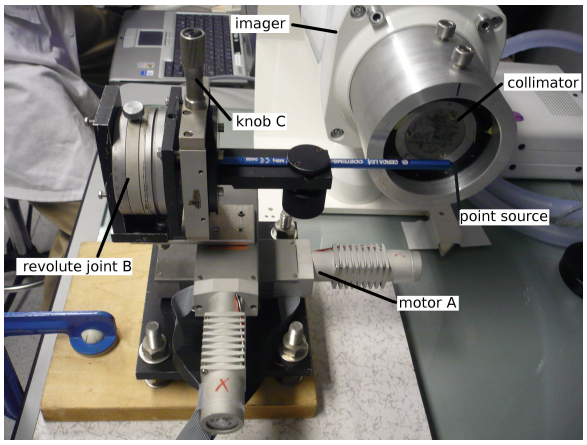
Numerically, it gives: w_{Zn}=80% and w_{Sn}=20%, with $\frac{\mu_{Zn}}{\rho_{Zn}} = 0.261$ cm$^2 \cdot$g^{-1} and $\frac{\mu_{Sn}}{\rho_{Sn}} = 0.717$ cm$^2 \cdot$g^{-1} at 140 keV, $\rho_{Zn} = 7.36$ g$\cdot$cm^{-3} and $\rho_{Sn} = 7.14$ g$\cdot$cm^{-3}. Attenuation coefficients are obtained from NIST database of photon attenuations (Berger 05).

A ZnSn-alloy is melted into a cylinder from 177 g of tin and 532 g of zinc (25%-75%), which results in a linear attenuation coefficient $\mu_a = 2.74$ cm^{-1} according to Eq. 3.27, and confirmed by the software XCOM (Berger 05): $\mu_a = 2.74$ cm^{-1}. The diameter of the cylinder is 60 mm-large and its height is 35 mm. It is cut into 1 mm-thick slices in order to machine the pattern of the coded aperture by laser etching. If the slices would be thicker than 1 mm, the pattern would not have straight sides. All the slices were then stacked together to form a 9 mm-thick HURA collimator whose attenuation coefficient at 140 keV is the one of tungsten at 511 keV. This means that this collimator has the same transparency for 140 keV photons as tungsten for 511 keV photons. It was expected to build a 12 mm-thick collimator as the one available in section 2.1.2.2, but only 9 slices were available after the etching process. This collimator is shown in Fig. 3.49. This collimator is adapted to the γ-imager as shown in Fig. 3.50.

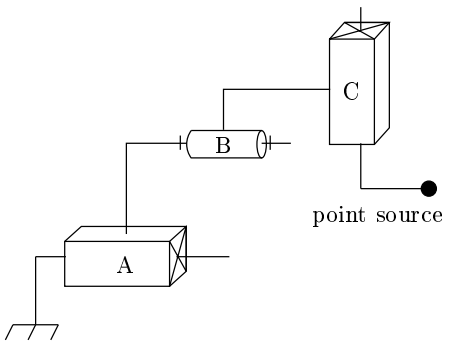
The system matrix has to take into account the fact that the detection efficiency of the scintillators is higher at 140 keV than at 511 keV. At this energy, the photoelectric attenuation coefficient of CsI-crystals is 3.14 cm^{-1} and the intrinsic spatial resolution of the detector is 2 mm. A new system matrix is computed in order to reconstruct projections which are acquired at 140 keV with this collimator.

3.5.2 Reconstruction of a point-source

First, a point-source is placed in the field-of-view of the imaging system in order to validate results which were previously obtained with Monte-Carlo simulations. The point-source which is a sphere made of cobalt 57 (^{57}Co). Its diameter is 1 mm. 85% of the decays which occur correspond to the emission of 122 keV photons. The activity of the point-source at the moment of the acquisition was 793 kBq (December 4$^{\text{th}}$, 2007) since its activity was 822 kBq 14 days before (November 21$^{\text{st}}$,2007). Its half-life is 271.79 days, which means that its activity can be assumed to be constant for all acquisitions made on December 4$^{\text{th}}$, 2007.



(a) Description of the experimental set-up



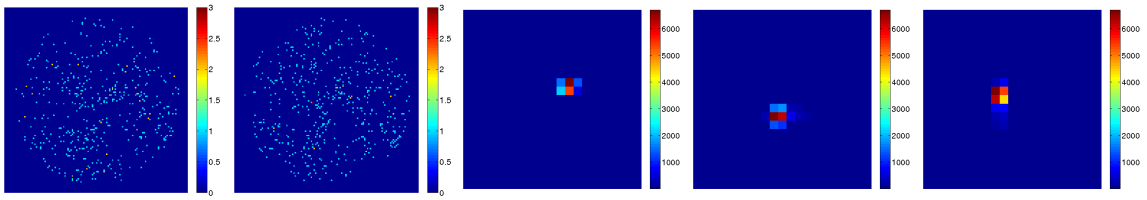
(b) Kinematics scheme

Figure 3.50 – Experimental set-up

3.5.2.1 Experimental set-up

The experimental set-up used to acquire projections is shown in Fig. 3.50. The point-source can be displaced thanks to 2 sliders and a revolute joint as shown in Fig. 3.50b. A first projection is acquired, then the point-source is rotated by 90° thanks to the revolute joint B to acquire a second projection, which is equivalent to keep the point-source still and rotate the imaging system. It is important to notice that in the kinematics scheme, the slider joint C stands after the revolving joint B, otherwise it would not be possible to simulate the acquisition of the second projection. Each acquisition lasts 37 s. Projections were acquired for several positions of the point-source, the position of the point source being changed with the knob C. Some acquisitions were also performed while the point-source was moving in the field-of-view thanks to the slider joint A which was connected to a step-by-step motor, in order to model a line-source. These results are not presented here, since simpler experiments were then performed with real line-sources, as presented later.

3.5.2.2 Results



(a) Projections acquired at 0° and 90°

(b) Central slices of of the estimated distribution in the zx-, yz- and xy-planes

Figure 3.51 – Reconstruction of a point-source from experimental projections. The number of detected photons in projections are respectively 485 and 498, which corresponds to a point-source emitting 10^5 photons

The energy windows of the detector is [103 keV – 150 keV]. Two projections are acquired which last 37 s each. The number of detected photons for these projections are respectively 136698 and 136351, which corresponds to efficiencies 6.38‰ and 6.36‰. Every projection is acquired in LIST-mode, which means that the detection location of every

detected photons (counts)	Activity of the point-source (emitted photons)	number of independent projections
10^3	1.075×10^5	5
10^4	1.075×10^6	5
10^5	1.075×10^7	2
2.73×10^5	2.93×10^7	1

Table 3.3 – Initial activities in the object and corresponding number of detected photons.

photons in the pixel is registered individually. This allows the generation of several projections from a unique acquisition and the number of counts per projection can be chosen. It is decided to generate projections which have the same number of counts as the projections generated with Monte-Carlo simulations. These sets of projections are reconstructed with the appropriate system matrix. The initial activities in the object are given in Table 3.3. Fig. 3.51 shows an example of a reconstructed point-source from two projections which contain respectively 485 and 498 detected photons. Although the number of detected photons is small, the reconstructed point-source only spreads on a few voxels. The influence of the activity of the point-source on the reconstruction error, on the FWHM and on the quantification performances of the imaging system is then investigated.

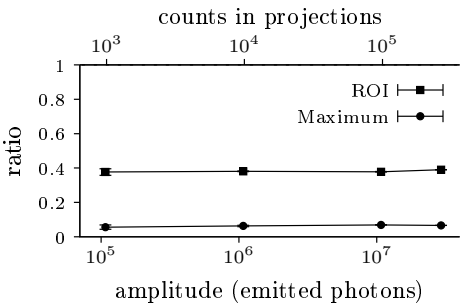
Quantification performances can be derived either from the maximum reconstructed activities or from the total activities in a ROI defined around the reconstructed points. Fig. 3.52a shows the ratio of these quantities over the expected activities. It illustrates that reconstructions are biased, however the bias does not depend on the activity (bias=0.38±0.006 for the ratio based on the maximum reconstructed activities and bias=0.06±0.006 for the ratio based on the ROI). As a result, a correction factor might be applied to retrieve the expected activity of the point-source. It should be set to 1/0.38=2.6 so that the ratio based on the ROI becomes almost equal to 1. In this case, reconstructions might become quantitative. The correction factor would not be derived from the ratio based on the maximum reconstructed activities because it does not account for the tails of the reconstructed points, which also contain activity. The ROI-based ratio is a more robust estimator of the activity of the point-source than the maximum-based ratio. The reconstruction errors are then observed in Fig. 3.52b. As no reference object is available, the reference $\boldsymbol{\lambda}_{ref}$ is chosen to be the result of the reconstruction with the maximum number of detected photons. This allows the investigation of the statistical error. $\boldsymbol{\lambda}_{ref}$ is weighted to account for the activity in point-sources with lower activities for the computation of the error:

$$\text{error} = \frac{\sqrt{\frac{1}{N}\sum_{i,j,k}\Big(\lambda(i,j,k) - c\cdot\lambda_{ref}(i,j,k)\Big)^2}}{\max_{i,j,k}\Big(c\cdot\lambda_{ref}(i,j,k)\Big)} \tag{3.29}$$

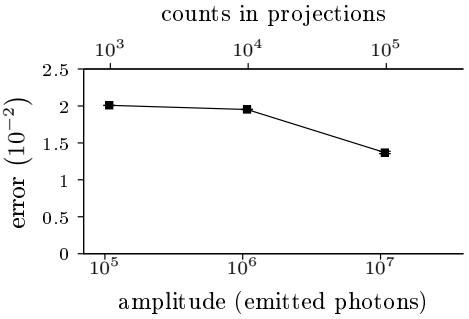
where c is the weighting coefficient that makes possible to compare two reconstructions with different activities. For instance, the weighting coefficient for the left-hand point in Fig. 3.52b is derived from Table 3.3 and is:

$$c = \frac{10^3}{2.73 \times 10^5} = 3.6 \times 10^{-3}$$

The error for the case where all the detected photons in LIST-mode are taken into account is not represented since it is the reference object and consequently the error is null. It is observed that the reconstruction error decreases as the activity in the point-source increases,



(a) Reconstructed maximum intensity and reconstructed intensity in the region of interest



(b) Error between the reconstructed spatial distribution and the reference object

Figure 3.52 – Accuracy of the reconstructed spatial distributions

as expected. The error bars are small when compared to the mean errors, which means that the reconstructions are very reproducible even for low activities.

For analytical and Monte-Carlo projections, the three FWHM depend on the activity of the point-source. Fig. 3.53 shows how the activity impacts the FWHM when the projections are obtained through the experimental set-up presented above. The FWHM are almost constant, from 1.8 to 1.4 mm. This is worse than for Monte-Carlo simulations where the spatial resolution of the detector was set to 2 mm. However, in Monte-Carlo simulations, the point-source was simulated as a real point with no volume whereas the point-source here is a 1 mm-sphere, which can explain that the three FWHM are in average 20% larger. As before, the FWHM is smaller in the y-direction. The reason is that the y-direction is a transverse direction for both projections and it was observed that the spatial resolution of the imaging system is better in the transverse direction than in the axial direction.

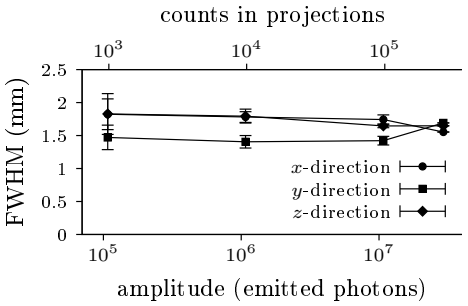


Figure 3.53 – The FWHM of reconstructed points depend on the activity of the point-source

Reconstructions of point-sources from real projections confirm that the activity of a point-source can be accurately estimated with our imaging system and a statistical reconstruction algorithm, as long as a correction factor F is applied to the reconstructions. The correction factor F is larger for reconstructions of real projections ($F = 2.6$) than for Monte-Carlo simulated projections ($F = 0.7$). The three FWHM are not degraded when compared to Monte-Carlo simulations.

3.5.3 Reconstruction of 2 cylinders

3.5.3.1 Experimental set-up

In this set of acquisitions, a syringe is filled with 740 kBq of ^{99m}Tc (0.02 mCi). ^{99m}Tc emits 140 keV photons which can be detected by the γ-imager. The diameters of the syringe is 4.6 mm. The active length of the syringe is 30 mm. As a result, the volume of reconstruction must be increased in order to reconstruct such an object. It is now made of $30\times\ 40\times\ 30$ voxels which are $1 \times 1 \times 1$ mm^3. The syringe is placed in the field of view with the experimental set-up shown in Fig. 3.54a. The experimental set-up allows the vertical translation of the syringe so that the location of the syringe can be modified. Furthermore, this experimental set-up allows the rotation of the translation system and the syringe, and consequently a second projection can be acquired, orthogonally to the first projection. For any vertical translation of the syringe, two orthogonal projections are acquired. The kinematics scheme of this set-up is shown in Fig. 3.54b. It is made of a slider joint controlled by the knob C, in order to control the vertical positions of the syringe, and a revolute joint B to turn the source and then acquire a second projection.

The syringe is vertically translated 2 mm by 2 mm and at every location, 2 orthogonal projections are acquired. Acquisitions last 1 min each. In the results presented below, two cylinders are placed in the field-of-view. It is obtained by acquiring orthogonal projections, the syringe being shifted by 6 mm between the first set of acquisitions and the second. The two projections are summed together and the two orthogonal projections are also summed together in order to obtain projections that would correspond to an acquisition where two syringes would be placed in the field of view, the distance between their axis being 6 mm. The two pairs of projections which are used to reconstruct the results below are acquired 97 min and 148 min after that the syringe was filled with ^{99m}Tc. This means that at the time of acquisitions, the number of emitted photons per voxel over 1 min is respectively $n_1 = 6.8 \times 10^4$ and $n_2 = 5.9 \times 10^4$ photons. It is therefore expected to reconstruct a concentration about 7.0×10^4 photons/voxel in cylinders, as summarised in Table 3.4.

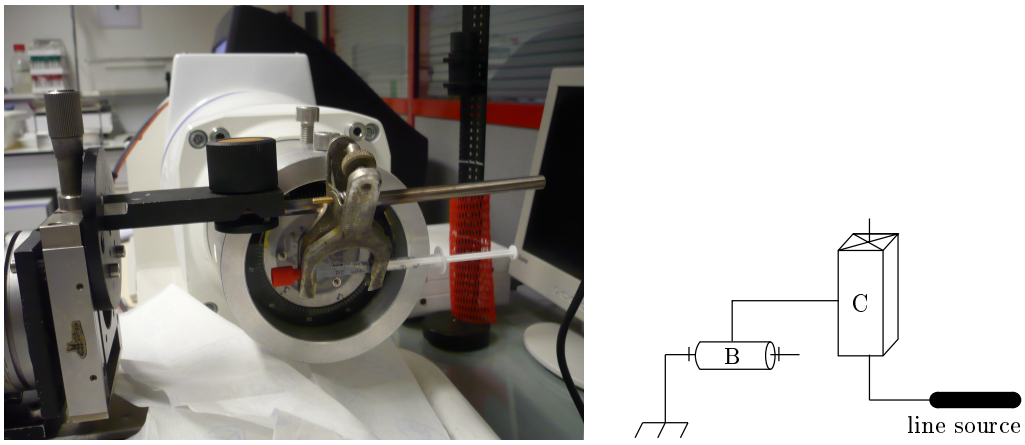
$$n = \frac{\text{activity}}{\text{volume}} \times \text{acquisition time} \times \exp\left(-\ln 2 \frac{\text{decay time}}{\text{half-life}}\right) \tag{3.30}$$

$$n_1 = \frac{0.02 \cdot 10^{-3} \cdot 37 \text{ GBq}}{\pi \cdot 2.3^2 \cdot 30 \text{ mm}^3} \times 60 \text{ s} \times \exp\left(-\ln 2 \frac{97 \cdot 60 \text{ s}}{6.01 \cdot 3600 \text{ s}}\right) = 6.8 \times 10^4 \tag{3.31}$$

$$n_2 = \frac{0.02 \cdot 10^{-3} \cdot 37 \text{ GBq}}{\pi \cdot 2.3^2 \cdot 30 \text{ mm}^3} \times 60 \text{ s} \times \exp\left(-\ln 2 \frac{148 \cdot 60 \text{ s}}{6.01 \cdot 3600 \text{ s}}\right) = 5.9 \times 10^4 \tag{3.32}$$

detected photons (counts)	Activity of a 1-mm^3 box inside syringe (emitted photons)	number of independent projections
1.75×10^4	1.95×10^3	5
1.75×10^5	1.95×10^4	3
6.3×10^5	7.0×10^4	1

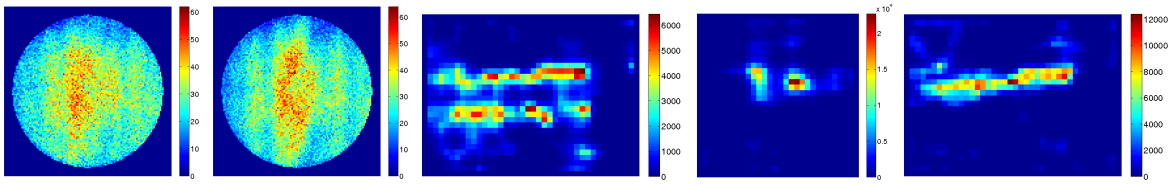
Table 3.4 – Initial activities in the object and corresponding number of detected photons



(a) Description of the experimental set-up (b) Kinematics scheme

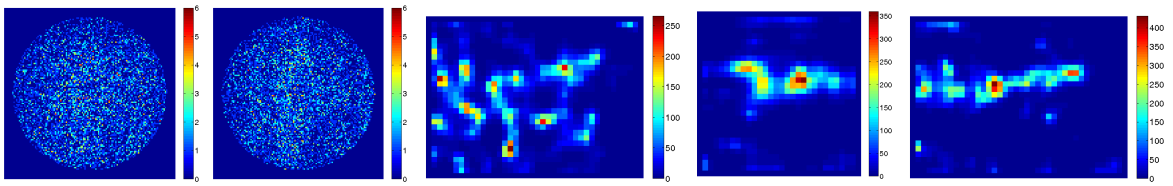
Figure 3.54 – Experimental set-up

3.5.3.2 Results



(a) Orthogonal projections (b) Mean xy-plane (c) Mean zx-plane (d) Mean zy-plane

Figure 3.55 – Reconstructed spatial distributions from two real projections, acquired at 0° and 90°. The number of detected photons in projections are respectively 290432 and 284158.

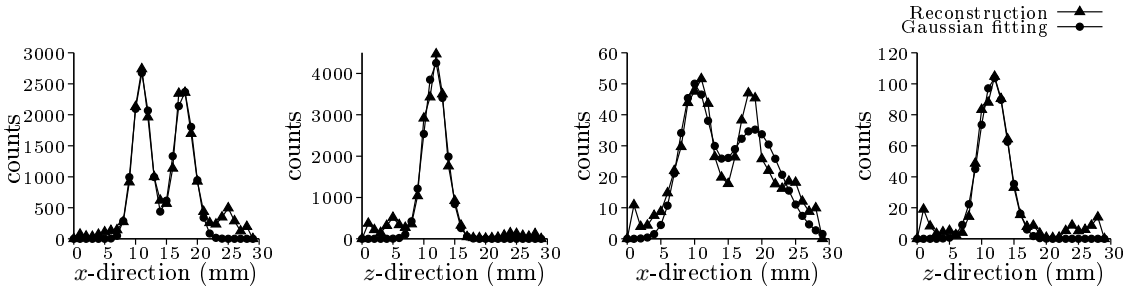


(a) Orthogonal projections (b) Mean xy-plane (c) Mean zx-plane (d) Mean zy-plane

Figure 3.56 – Reconstructed spatial distributions from two real projections, acquired at 0° and 90°. The number of detected photons in projections are respectively 7891 and 7703.

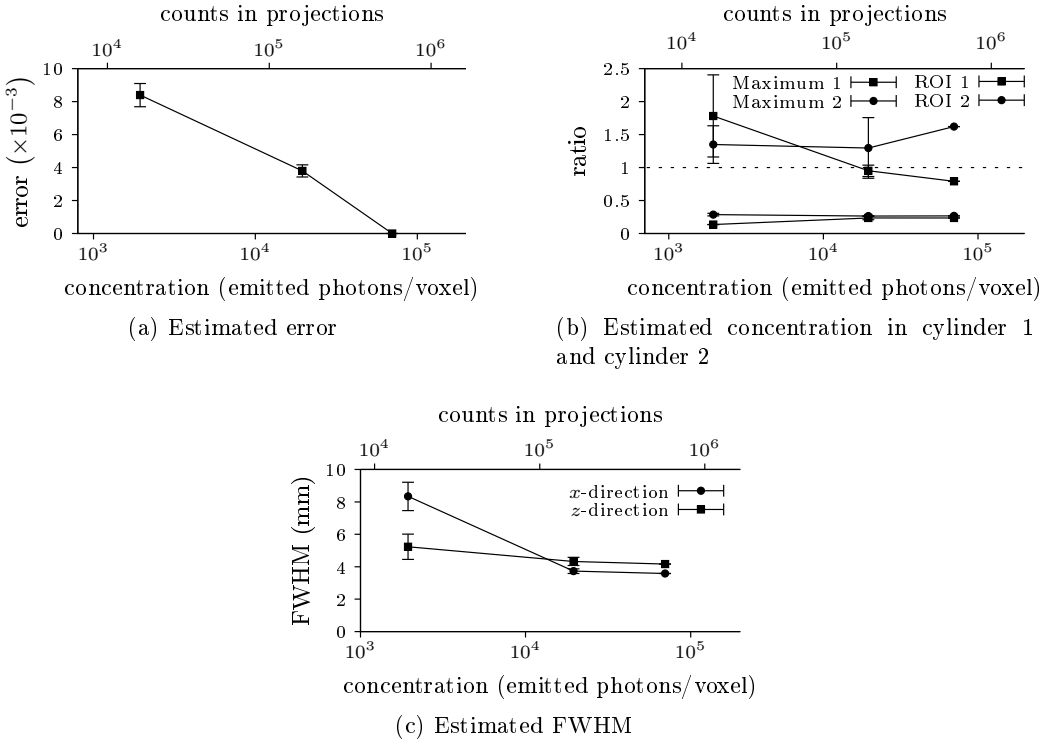
Fig. 3.55a and 3.56a show two examples of pairs of projections obtained from the experimental set-up described above. The mean reconstructed views in Fig. 3.55 and 3.56 illustrate that the concentration in cylinders plays an important role in reconstructions. In Fig. 3.55, two cylinders are clearly visible, although their homogeneity is poor, but in Fig. 3.56, cylinders can not be seen. However, in both cases, it is possible to extract information from the mean vertical and horizontal views of Fig. 3.55c and 3.56c which are represented in Fig. 3.57a and 3.57b. Gaussian functions fit well with mean spatial distributions when the concentration in the cylinder is high, but the fittings are not as good for cylinders whose concentration is low.

Contrary to previous investigations with analytical and Monte-Carlo projections, the reference distribution is not directly available. It is known that the object is made of 2 cylinders



(a) Horizontal and vertical views of Fig. 3.55c (b) Horizontal and vertical views of Fig. 3.56c

Figure 3.57 – Fit of Gaussian functions for the quantification of FWHM in x and z-directions



(a) Estimated error

(b) Estimated concentration in cylinder 1 and cylinder 2

(c) Estimated FWHM

Figure 3.58 – Impact of the concentration in cylinders on reconstructions, from 9 reconstructed distributions

whose diameter and length are respectively 4.6 mm and 30 mm, but their locations is subject to uncertainty because the experimental set-up does not allow the exact positioning of the syringe. In particular, the syringe is not perfectly parallel to the collimator and the imager, as it can be observed from Fig. 3.55. As a result, the error can not be computed from a reference spatial distribution. Consequently, the reconstruction based on the projections with the largest number of detected photons is chosen as a reference, and reconstructions with lower counts are compared to this reference distribution. The result is shown in Fig. 3.58a and it can be observed that the error for the right-hand point is null, since this reconstruction is the distribution which is taken as a reference. The concentration in cylinders is then quantified from the ratio between the total reconstructed activity in each cylinder and the expected total activity in the cylinders. It is shown in Fig. 3.58b. The estimation of the total activity is biased but can be corrected by applying a correction factor of $1/0.26=3.8$. The ratio based on the maximum reconstructed concentrations is shown to illustrate that no quantification can be expected from this estimator. It strongly depends on the concentration in cylinders and varies from one cylinder to another.

Finally, the FWHM are estimated and it is almost constant, excepted for the reconstructions which correspond to the lowest concentrations of the cylinders. The FWHM in the x-direction converges to 3.5 mm and the FWHM in the z-direction converges to 4.2 mm. The expected FWHM are 3.1 mm, which means that the estimated FWHM are over-estimated but still close to the expected values. The differences are respectively 13% and 35%.

The concentration in cylinders can be quantified from reconstructions based on real projections which contain more than 16000 photons. This corresponds to a concentration in the cylinders of this experiment of 2000 emitted photons/voxel, but for 511 keV photons, it would correspond to a concentration of 20000 emitted photons/voxel. Although this imaging system is able to reconstruct the 3D spatial distribution of 2 cylinders which emit a small amount of photons, it is not sufficient to consider estimating the concentration in the popliteal artery of patients with activities which are commonly injected (185MBq, 5 mCi).

3.6 Discussion and conclusion

This chapter provides methods to deal with system matrices associated with imaging systems containing coded-aperture collimators. These matrices are large (several giga-bytes) and consequently their redundancies should be exploited. It is illustrated that smaller system matrices can be loaded in memory as long as the imaging system is made of orthogonal sets of projections. However, this does not work if the attenuation of the object is not invariant by 90°-rotation, since it would affect the probabilities a_{ij}. Inhomogeneous attenuations can be taken into account by introducing an attenuation matrix which pre-corrects the spatial distribution $\boldsymbol{\lambda}$ at every iteration. Spatial distributions can then be reconstructed from a system matrix which is redundant. A new method to accelerate reconstructions is also proposed. It adaptively uses several levels of refinements of the projections and of system matrices in order to speed up first iterations of reconstructions. The transition from one level of refinements to another should be discussed. Either the control log-likelihood E^1 starts decreasing and the refinement must be decreased, or E^1 keeps on increasing but tends to converge to a limit. In this case, it was chosen to decrease the refinement level when improvements were not better than a factor 1.1 between two controls, but this factor was chosen from empirical tests. By the end, the finest representation of projections is used and the algorithm stops when E^1 does not improve better than a factor 1.1. More elaborated stopping criteria could be used (Veklerov 87). The reconstruction of the spatial distribution is convex, however, the algorithm is stopped before convergence because it would result in

solutions extremely noisy because the solution is not regularised. As a result, the solution depends on the solutions computed at each iteration and consequently, it depends on the criteria used to change binning levels and to stop the algorithm. However the difference on the log-likelihood and the obtained solutions is very small. A question raised by this adaptive method concerns the regularisation of this algorithm. It is based on the successive maximisation of different log-likelihood functions, as a result it would be difficult to introduce a prior which has the same influence on all log-likelihoods. *A priori* information could be added to reconstructions in order to compensate for the low amount of signal. Smoothing should correct for most of inhomogeneities, but more elaborate regularizations should be very adapted to the reconstructions of 2 vessels in a knee, as Gauss-Markov-Potts prior models (Mohammad-Djafari 06).

A way to decrease reconstruction errors would be to model the detector spatial resolution into system matrices. It can be achieved by convolving every column, which corresponds to expected projection from a single voxel, by a point-spread function. The point-spread function should not be a 2D Gaussian function but a more elaborate point-spread function because the incidence of photons is not necessarily normal to the detector surface: the point-spread function should be a ellipse-like function. This corresponds to the addition of some blurring in the system matrix, which induces a worse condition number of the system matrix. From this point of view, improving the model of the matrix does not automatically necessarily result in better reconstructions. The system matrices which are used for reconstructing spatial distributions correspond to a collimator with perfect imaging properties for correlation-based algorithms, not for the MLEM-algorithm. This means that coded-aperture collimators may not be optimal for low-activity imaging, although their performances offer an excellent trade-off between a good resolution and a high sensitivity. Collimators with 'perfect imaging properties' for statistical reconstruction algorithms should be looked for, as for correlation-based reconstruction algorithm.

Furthermore, no tissue are modelled around the vessels for the examples presented in this chapter, although it is expected that the vessels stand 50 mm-deep in the knee. Adding these tissues would result in a diminution of the signal as well as an increase of the fraction of the photons which undergo Compton scattering. It would deteriorate reconstructions. Another phenomenon which would deteriorate reconstructions is the detection noise. The detector noise was measured and it was 1.7 counts/s (cps) at 140 keV when no source was placed in the field-of-view of the imager. This means that for a clinical acquisition which lasts 2 hours, 12240 photons are detected per detector because of the detector noise. It is expected that this noise decreases at 511 keV as detectors are usually less sensitive when the energy threshold of the imager is increased. For the real projections which were presented in this chapter, this detector noise is negligible because acquisitions last at maximum 1 min with sources which are much more radioactive than the vessels in patients. As a result, detector noise accounts for about 100 photons in projections whereas about 15000 photons are detected per acquisition. Reconstructions with real projections from the experimental set-up should be further validated with 511 keV and the tungsten coded-aperture collimator. In particular, it would be of interest to estimate the detector spatial resolution at this energy.

Moreover there is no background activity in the results presented in this chapter, although it is expected that radiotracers spread in the whole body, at least in capillaries. The section of popliteal vessels being small ($\pi \times 0.25^2 = 0.20\ cm^2$) compared to the section of the knee ($\pi \times 50^2 = 78.5\ cm^2$), even a low concentration outside the vessels brings an important contribution to projections. For instance, if the concentration in tissues is 200 times smaller than in vessels ((78.5/0.2)/2=200), there are as many detected photons which come from one of the two vessels than from the rest of the knee. In examples presented in this chapter, the reconstruction volume was limited because of RAM requirements, this is why the length

of modelled cylinders is at maximum 30 mm. In reality, vessels pass through the whole field-of-view of the detector and longer field of view should be modelled. Photons starting from 85 mm away from the axis of the imaging system can pass through a hole of the coded-aperture collimator and reach the detector, although the longest field-of-view was 40 mm-long.

Better performances would be expected if the CsI-scintillator would be replaced by BGO-scintillators. Indeed, BGO-crystals are more sensitive than CsI which means that more photons would be detected. BGO is 8 times more efficient than CsI at detecting 511 keV photons when the scintillator is 4 mm-thick, and 3 times more efficient when only counting photons which are detected at once by photoelectric effect. The fact that BGO is slower than CsI is not a problem for measuring the β^+ input function, since just a few thousand photons are detected in 2 hours. The lower light output is not a problem because with 511 keV photons, the light output is still correct. LSO would be a less interesting candidate because of the natural radioactivity of lutetium, which adds a supplementary source of noise. It exists coded-aperture collimators with larger opening fraction than 50%. It would allow the collection of a larger amount of signal but projections would be more intricate. A trade-off between the quantity of information and its quality should be looked for in order to find an optimal opening fraction.

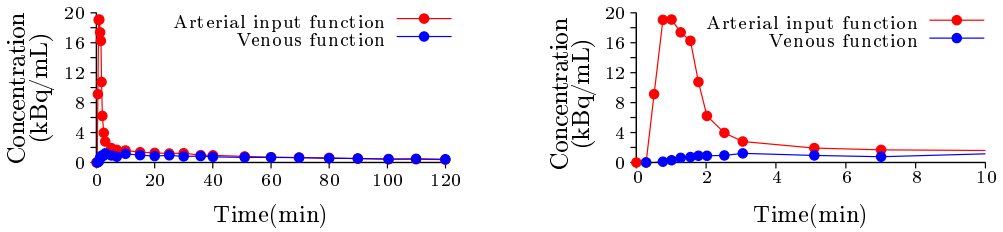
In this chapter, it is shown that coded-aperture collimators and a statistical algorithm such as MLEM are complementary to address low-activity imaging. Examples are taken from analytical projections, then from Monte-Carlo simulations. Finally, reconstructions from real projections are performed. As the system matrices which are used to reconstruct the spatial distributions shown in this chapter are large (several giga-bytes), symmetries are used to decrease the memory load during the reconstruction iterations. Furthermore, information which is contained in coded projections are very intricate, which leads to poor convergence rates of the reconstruction algorithm. A new method is developed to accelerate the reconstructions. It is based on an adaptive refinement of the observations. Results indicate that the improvement on the sensitivity of the imaging system are not sufficient. Concentrations at least 10 times higher than concentrations in patients' vessels would be required to expect an estimation of the mean concentration of the artery and the vein over a clinical acquisition. Furthermore it is expected to separate signals coming from the artery and from the vein every 20 s, not for a whole clinical acquisition. Next chapter investigates how to estimate concentrations for such small time frames from the reconstructions of projections acquired from the whole acquisition.

Spatiotemporal reconstruction

4

This chapter presents new methods which deal with the reconstruction of spatiotemporal distributions. It takes advantages of temporal correlations of the spatial distributions.

In the following, it is assumed that objects in the field of view are still. Only their kinetics changes along time. As it is expected to estimate the kinetics of an artery and a vein, the assumption is mostly made that there are at maximum two components to separate. Fig. 4.1 presents concentrations in an artery and a vein, obtained from blood samples. This is the ground-truth kinetics that it is intended to estimate. These radioactive concentration are obtained from an investigation where 174 MBq (4.7 mCi) of ^{18}F-radiolabelled molecules are injected to a patient. Blood samples are taken from an artery and a vein every 15 seconds just after injection and 10 minutes by the end. These samples are then placed in a γ-counter and their radioactive concentration is measured. These data are corrected from radioactive decay that corresponds to the time it takes to bring it to the γ-counter, however, it is not corrected for the radioactive decay of the tracer since the injection time. This means that these concentrations are the real concentrations of the blood that floods in the artery and the vein at the time the samples are drawn. The area under the curve is 7.57×10^6 counts/ml for this investigation that lasts 120 min. This means that the average concentration is about 1 kBq/ml (about 30 nCi/ml). In the following development, experimental validations are based on cylinders that are 30 mm-long and whose diameters are 4.6 mm-large, as a result most of the cylinders in Monte-Carlo simulations are cylinders with such dimensions so that it is possible to compare reconstructions from Monte-Carlo simulations with experimental acquisitions. It can then be deduced from Eq. (4.1) that every cylinder emits about 3.8×10^6 photons during a clinical acquisitions that lasts 2 hours. Radioactive decay is not taken into account because it is radiotracer-dependent, but this means that the following amount of



(a) Kinetics over 2 hours

(b) Zoom on the first minutes

Figure 4.1 – Ground-truth kinetics. (4.1a) Kinetics evolutions over the whole clinical acquisition. (4.1b) Kinetics evolution in first minutes

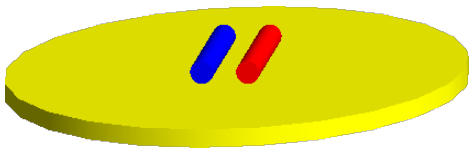
emitted photons is the highest expected limit.

$$7.57 \cdot 10^6 \text{ counts/ml} \times \left(\pi \times \frac{4.6^2}{4} \times 30 \right) \text{mm}^3 = 3.8 \times 10^6 \text{ photons} \tag{4.1}$$

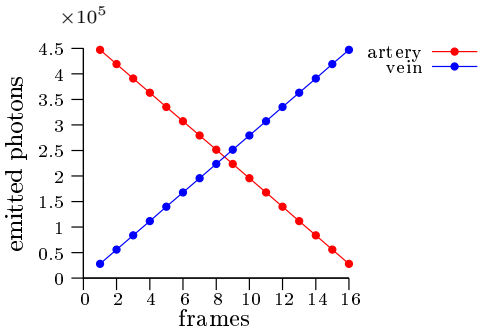
This is the amount of photons emitted during an acquisition. In Sec. 4.1, a method is described to estimate the kinetics of every object in the field of view from observations coming from the detector. Section 4.2 improves this method through the incorporation of a reconstruction step. These methods are blind-source separation methods and take advantage of nonnegative-matrix factorisation, which is explained in section 1.6. Both principles of the methods and validations are given. Validations are obtained from either Monte-Carlo simulations or real projections that are obtained from the experimental set-up of Fig. 3.54.

4.1 Source separation in detector-space

First we investigate the simple case where there is no reconstruction of the spatial distribution of the source and where the kinetics are very distinguishable as the one presented in Fig. 4.2b. The basic idea is to separate components on the basis of their spatiotemporal distribution. Let's have C objects which emit 511-keV photons in front of a pixellised detector as presented in Fig. 4.2a. No collimator is used. Every pixel value in the detector is a realisation of a Poisson distribution.



(a) Simulation configurations. Cylinders are 4.6mm-large and 30mm-long. Scintillator dimensions are $(100 \times 100 \times 4)$ mm^3

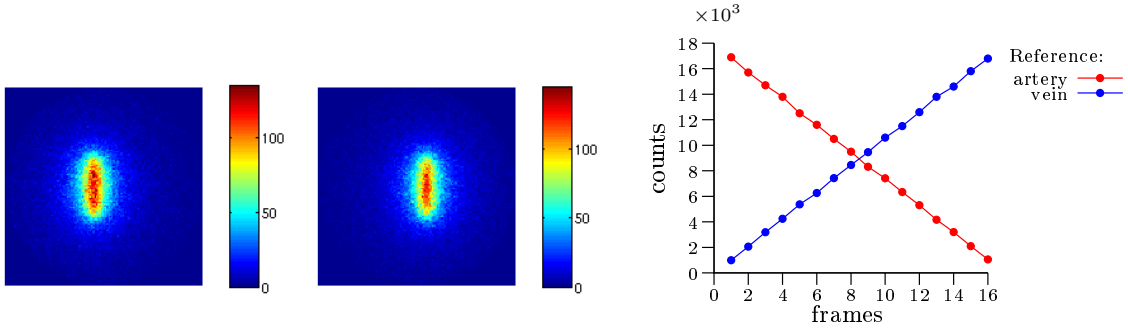


(b) Simulated activities of vessels

Figure 4.2 – Simulation of the spatiotemporal distribution of two cylinders, which represent an artery (red) and a vein (blue)

Let's have a series of (128×128)-images, corresponding to the dynamic acquisition of these sources whose activities vary along 16 time frames. We first generate 16 frames using Monte-Carlo simulations, where two cylinders are placed in the field-of-view. The cylinders are 30 mm-long and their diameter is 4.6 mm-large. The axes of these cylinders are located 3 mm above the surface of the scintillator, which means that the distance between the border of the cylinder and the surface of the scintillator is 0.5 mm: there are almost in contact. This is not possible in reality because the depth of vessels in the knee is about 50 mm, but this configuration is taken to illustrate source separation steps. Every cylinder contribution is simulated independently so that it is possible to compare the solutions given by source-separation algorithms to the real spatio-temporal distribution. They are generated with GATE (Jan 04) according to Fig. 4.2. No collimation is taken into account. Positions of both sources remain unchanged but their intensities (brightness) vary according to Fig. 4.2b. The total projection over all frames of every vessel is presented in Fig. 4.3a where the left

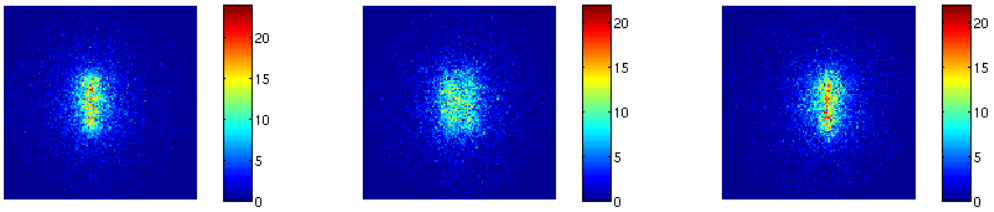
figure corresponds to the artery and the right figure corresponds to the vein. Three frames are shown in Fig. 4.4. Contributions of the artery and the vein vary according to Fig. 4.3b and are subjects to Poisson fluctuations. These curves are known only because the two components are generated independently. These are the curves estimated from the projections containing both the arterial and venous contributions. The red and blue cylinders in this example have similar average intensity.



(a) Simulated projection of the artery and the vein over all time frames

(b) Kinetics of artery and vein

Figure 4.3 – Total contributions of the artery and the vein (Fig 4.3a) while they are in contact with the scintillator surface. These are Monte-Carlo simulations. Fig. 4.3b shows their measured kinetics in the detector as if they are acquired independently.



(a) frame 1

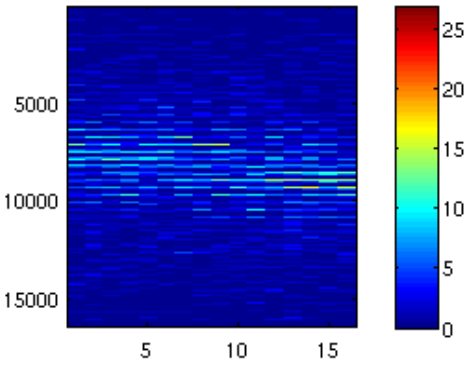
(b) frame 8

(c) frame 16

Figure 4.4 – 3 temporal frames. Fig. 4.4a represents the first frame where contribution of the artery largely dominates the contribution of the vein, while in Fig. 4.4c, the contrary happens. In the middle frame, both vessels contribute similarly

The whole acquired data can be represented by a $(M \times T)$-matrix $\boldsymbol{S}$ as illustrated in Eq. (4.2), where every row corresponds to a pixel m of the image/detector and every column of $\boldsymbol{S}$ corresponds to a time frame t. The number of observations and time frames are respectively denoted M and T. The point is to retrieve the kinetics of every source from spatiotemporal matrix $\boldsymbol{S}$. The separation of the contributions of every source is based on matrix $\boldsymbol{S}$.

The acquired frames are the contributions of both sources weighted by their kinetics. As a result we want to factorize this matrix so that we can estimate the kinetics of every source and their intensity evolution as a function of time. The factorisation would look like:



$$\boldsymbol{S} = \begin{bmatrix} \vdots & & \vdots \\ \cdots & s_{i,j} & \cdots \\ \vdots & & \vdots \end{bmatrix} \begin{matrix} \\ \\ \\ \end{matrix} \overset{T}{\longleftrightarrow} \quad \updownarrow M \tag{4.2}$$

Figure 4.5 – Visualisation of matrix $\boldsymbol{S}$. Every column t is a lexicographic representation of simulated projections at frame t as in Fig. 4.4a, 4.4b and 4.4c

$\boldsymbol{S} = \boldsymbol{X} \cdot \boldsymbol{K}$ with $\boldsymbol{X}$ being the spatial components and $\boldsymbol{K}$ the kinetics components:

$$\begin{bmatrix} \\ s_{i,j} \\ \\ \end{bmatrix} = \overset{C}{\longleftrightarrow} \quad \begin{matrix} \\ M \\ \\ \end{matrix} \updownarrow \begin{bmatrix} \vdots & \vdots \\ x_{i,1} & x_{i,2} \\ \vdots & \vdots \end{bmatrix} \times \overset{T}{\longleftrightarrow} \begin{bmatrix} \cdots & k_{1,j} & \cdots \\ \cdots & k_{2,j} & \cdots \end{bmatrix} \updownarrow C \tag{4.3}$$

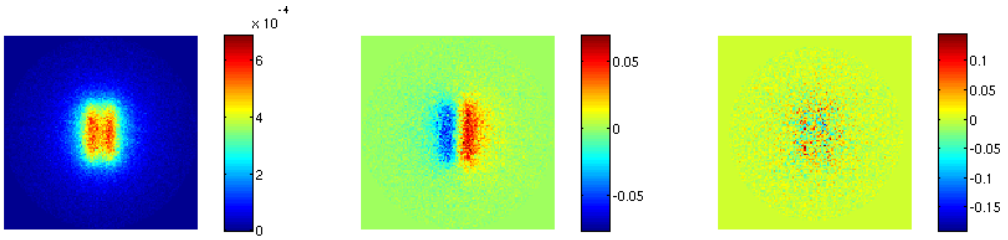
While the number of components C is small when compared to the dimensions of the matrix $\boldsymbol{S}$, it is very unlikely that such a matrix factorization exists, since $\boldsymbol{S}$ is made of $16 \times (128 \times 128) = 245760$ coefficients, while $\boldsymbol{X}$ and $\boldsymbol{K}$ are respectively made of $(128 \times 128) \times 2 = 32768$ and $16 \times 2 = 32$ coefficients, for a total of 32800 coefficients, which reduces by a factor of ≈ 8 the number of variables to estimate. This matrix factorization is largely over constrained. As a result, we are looking for an approximation instead of an exact decomposition:

$$\boldsymbol{S} \approx \boldsymbol{X} \cdot \boldsymbol{K} \tag{4.4}$$

A first immediate solution is obtained from a Singular Value Decomposition (SVD) where only the C first components are taken into account. Any matrix can be decomposed into two orthonormal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ and a diagonal matrix $\boldsymbol{D}$ as:

$$\boldsymbol{S}_{(M,N)} = \boldsymbol{U}_{(M,N)} \cdot \boldsymbol{D}_{(N,N)} \cdot \boldsymbol{V}^{\mathrm{T}}_{(N,N)} \tag{4.5}$$

where $\boldsymbol{S}_{M,N}$ indicates that $\boldsymbol{S}$ is $(M \times N)$-matrix. By setting all the diagonal coefficients of



(a) First component (b) Second component (c) Third component

Figure 4.6 – First components of SVD decomposition

$\boldsymbol{D}$ to zero but the C first ones, an approximation of $\boldsymbol{S}$ is obtained:

$$\boldsymbol{S}_{(M,N)} = \boldsymbol{U}_{(M,C)} \cdot \boldsymbol{D}_{(C,C)} \cdot \boldsymbol{V}^{\mathrm{T}}_{(N,C)} \tag{4.6}$$

By denoting $\boldsymbol{X} = \boldsymbol{U}$ and $\boldsymbol{K} = \boldsymbol{D} \cdot \boldsymbol{V}^{\mathrm{T}}$, an approximation of $\boldsymbol{S}$ on two matrices is obtained, but this approximation is not non-negative. Fig. 4.6 presents the three first spatial components corresponding to the SVD decomposition. The third component is shown only to demonstrate that this component contains no significant information, only noise. The first component in Fig. 4.6a is the average spatial distribution and is non-negative. However, the second component contains both positive and negative coefficients. By using an appropriate mixing matrix $\boldsymbol{R}$, a linear combination of these two components are obtained, which is mostly non-negative.

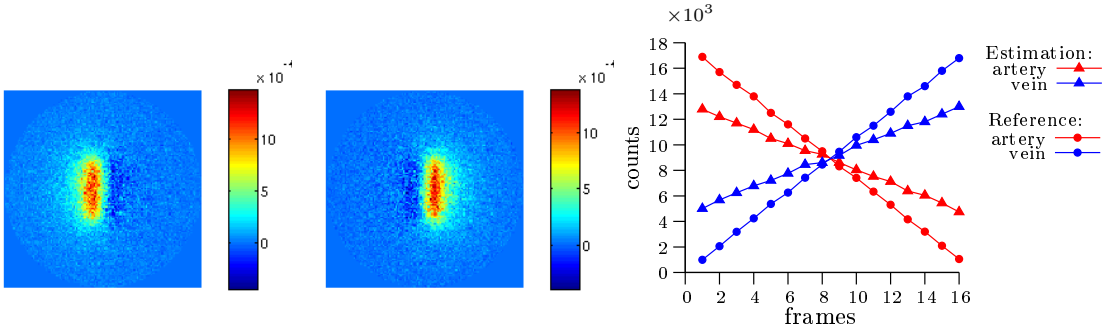


Figure 4.7 – Linear combination of $\{\boldsymbol{X}, \boldsymbol{K}\}$ solution obtained from SVD

For instance, by applying $\boldsymbol{R} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ as in Eq. (4.7), components are almost non-negative as shown in Fig. 4.8. Matrices $\boldsymbol{X}$ and $\boldsymbol{K}$ are normalised so that the sum of every spatial component is 1. In this case, $\boldsymbol{K}$ represents the real estimated counts coming from every vessel.

$$\boldsymbol{X}' = \boldsymbol{X} \cdot \boldsymbol{R} \qquad\qquad \boldsymbol{K}' = \boldsymbol{R}^{-1}\boldsymbol{K} \tag{4.7}$$

It can be noted that spatial components are not non-negative, but only a few pixels are negative. Furthermore, the spatial components are not locally symmetric around the main activity distribution as it can be observed in Fig. 4.7. On the other hand, activity is lower than the real activity since in Fig. 4.3a, maximum activity is higher than 6×10^6 counts while in Fig. 4.7, it does not exceed 5×10^6 counts.

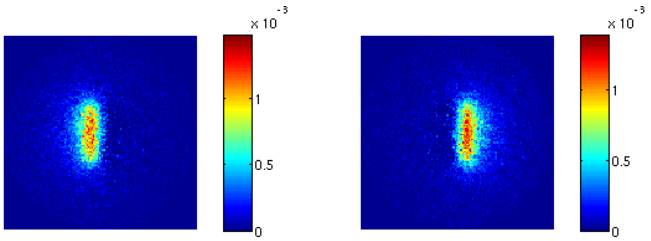


Figure 4.8 – Candidate for non-negative matrix factorisation, from SVD

By setting negative values to zero, a non-negative solution is obtained in Fig 4.8. The corresponding kinetics is still the one shown in Fig. 4.7.

As it is not possible to exactly decompose the matrix $\boldsymbol{S}$ on two non-negative matrices $\boldsymbol{X}$ and $\boldsymbol{K}$, a question arises about the nature of the residual that can be defined as $\boldsymbol{N} = \boldsymbol{S} - \boldsymbol{X} \cdot \boldsymbol{K}$. In radiation physics, observations are Poisson-distributed which encourages the residual $\boldsymbol{N}$ to be modelled with Poisson noise. Instead of looking for a matrix factorization as in Eq. (4.3), we are now looking for $\{\boldsymbol{X}, \boldsymbol{K}\}$ that maximises the likelihood $L(\boldsymbol{X}, \boldsymbol{K}) \propto P(\boldsymbol{S}|\boldsymbol{X}, \boldsymbol{K})$ with the constrains that $\boldsymbol{X}$ and $\boldsymbol{K}$ are non-negative. As all $(s_{i,j})$ are realisations of Poisson-distributed variables, the likelihood $L(\boldsymbol{X}, \boldsymbol{K})$ (and its logarithm) can be written:

$$L(\boldsymbol{X}, \boldsymbol{K}) = \prod_{i,j} \frac{e^{-(\boldsymbol{XK})_{i,j}} (\boldsymbol{XK})_{i,j}^{s_{i,j}}}{s_{i,j}!} \tag{4.8}$$

$$\ln(L(\boldsymbol{X}, \boldsymbol{K})) = -\sum_{i,j} (\boldsymbol{XK})_{i,j} + \sum_{i,j} s_{i,j} \ln((\boldsymbol{XK})_{i,j}) - \ln(s_{i,j}!) \tag{4.9}$$

In order to maximise the likelihood to have $\boldsymbol{X}$ and $\boldsymbol{K}$, the spectral projected gradient method described in Sec 1.7.4 is applied on the cost function E which is closely related to the likelihood $L(\boldsymbol{X}, \boldsymbol{K})$ using Eq. (4.10).

$$E = -\ln(L(\boldsymbol{X}, \boldsymbol{K})) + \ln(s_{ij}!) = \sum_{i,j} \Big((\boldsymbol{XK})_{i,j} - s_{i,j} \ln((\boldsymbol{XK})_{i,j}) \Big) \tag{4.10}$$

$$= \sum_{i,j,r} x_{ir} k_{rj} - \sum_{ij} s_{i,j} \ln \left(\sum_r x_{ir} k_{rj} \right) \tag{4.11}$$

The minimisation of E with the spectral projected gradient method requires the derivation of its partial differentiations on x_{nc} and k_{ct}:

$$\frac{\partial E}{\partial x_{n,c}} = \sum_j k_{c,j} - \sum_j s_{n,j} \frac{k_{cj}}{\sum_{r'} x_{ir'} k_{r't}} = \left(\left(\mathbb{U} - \frac{\boldsymbol{S}}{\boldsymbol{XK}} \right) \boldsymbol{K}^{\mathrm{T}} \right)_{c,t} \tag{4.12}$$

$$\frac{\partial E}{\partial k_{ct}} = \sum_i x_{ic} - \sum_i s_{it} \frac{x_{ic}}{\sum_{r'} x_{ir'} k_{r't}} = \left(\boldsymbol{X}^{\mathrm{T}} \left(\mathbb{U} - \frac{\boldsymbol{S}}{\boldsymbol{XK}} \right) \right)_{c,t} \tag{4.13}$$

where $\mathbb{U}$ is the unitary matrix: $\forall (i,j), \quad \mathbb{U}_{ij} = 1$. When applied the constrained minimisation with the gradient derived in Eq. (4.12) and (4.13), the algorithm is not stable because some terms in $\dfrac{1}{AXK}$ tends to $+\infty$ when AXK is close to 0. As a result, we propose to make a modification on the likelihood in Eq. 4.8 by adding a term $\varepsilon = 10^{-200}$ to every $(XK)_{i,j}$:

$$L(\boldsymbol{X}, \boldsymbol{K}) = \prod_{i,j} \frac{e^{-(\boldsymbol{XK})_{i,j} + \varepsilon} ((\boldsymbol{XK})_{i,j} + \varepsilon)^{s_{i,j}}}{s_{i,j}!} \tag{4.14}$$

The objective function E and its gradient becomes:

$$E = \sum_{i,j} \Big((\boldsymbol{XK})_{i,j} + \varepsilon - s_{i,j} \ln((\boldsymbol{XK})_{i,j} + \varepsilon) \Big) \tag{4.15}$$

$$\frac{\partial E}{\partial k_{ct}} = \left(\boldsymbol{X}^{\mathrm{T}} \left(\mathbb{U} - \frac{\boldsymbol{S}}{\boldsymbol{XK} + \varepsilon} \right) \right)_{c,t} \tag{4.16}$$

$$\frac{\partial E}{\partial x_{n,c}} = \left(\left(\mathbb{U} - \frac{\boldsymbol{S}}{\boldsymbol{XK} + \varepsilon} \right) \boldsymbol{K}^{\mathrm{T}} \right)_{c,t} \tag{4.17}$$

The solution $(\boldsymbol{X}, \boldsymbol{K})$ which is looked for and the intermediate steps $(\boldsymbol{X}^n, \boldsymbol{K}^n)$ should not result in $(\boldsymbol{XK})_{ij} = 0$ since this minimises the likelihood instead of maximising it as it can be seen from Eq. 4.8. Adding ε is a security which prevents such a situation from happening.

The spatiotemporal distribution of Fig. 4.8 is used to initialise the projected gradient minimisation, with a slight modification: the kinetics is rescaled so that the subproblem defined by the objective function $E(\alpha)$ in Eq. (4.18) is optimal:

$$E(\alpha) = \sum_{i,j,r} x_{ir}(\alpha k_{rj}) - \sum_{ij} s_{i,j} \ln\left(\sum_{r} x_{ir}(\alpha k_{rj})\right) \tag{4.18}$$

$$\frac{\partial E}{\partial \alpha} = 0 \qquad \Leftrightarrow \qquad \sum_{ij}(XK)_{ij} = \frac{\sum_{ij} s_{ij}}{\alpha} \tag{4.19}$$

The solution is obtained when the gradient is null which happens when $\alpha = \dfrac{\sum_{ij} s_{ij}}{\sum_{ij}(XK)_{ij}}$.

This leads to solution presented in Fig. 4.9. The coefficient α is found to be close to 1 as it is expected: $\alpha = 1.09$.

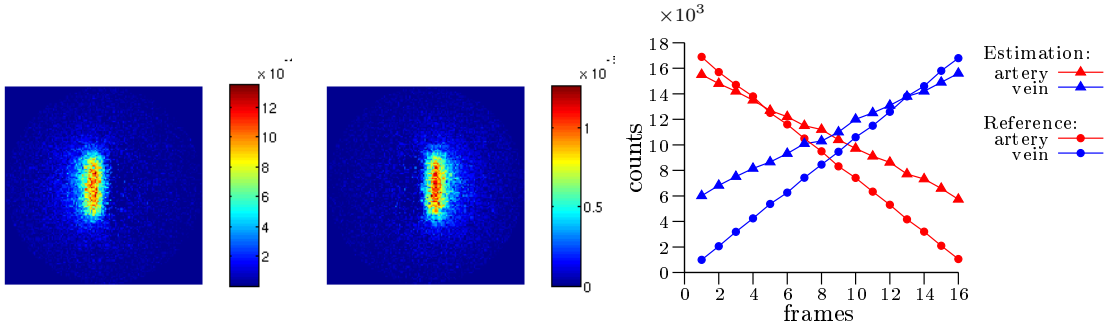


Figure 4.9 – Spatial and temporal components that are used to initialised the NMF-algorithm

The spectral projected gradient algorithm is then applied on the objective function E in Eq. 4.15, with the initial point being the spatio-temporal distribution of Fig. 4.9. It converges to the solution shown in Fig. 4.10. The difference between the initial log-likelihood and the log-likelihood at convergence is about 8800 as shown in Fig. 4.11. This means that the solution obtained after the optimisation is about exp(8800) more likely. The kinetics is very similar to the one obtained in Fig. 4.7, but the spatial components are now non-negative. The convergence rate of the algorithm, which is shown in Fig. 4.11, illustrates the fact that the

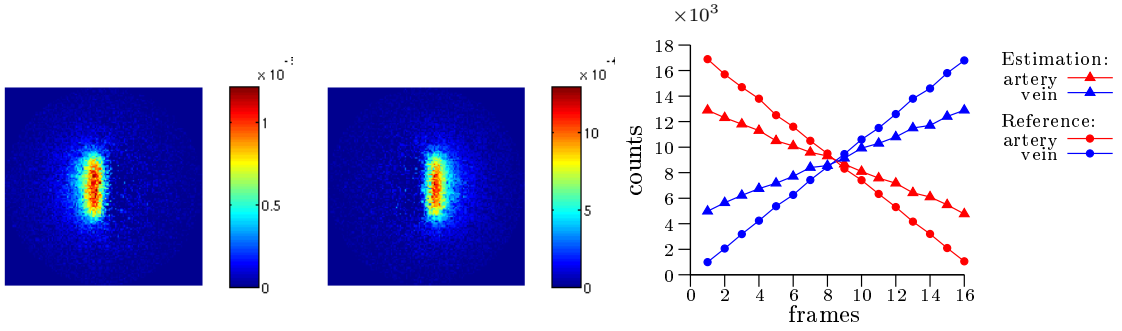


Figure 4.10 – Result of the source separation based on the maximum likelihood

algorithm converges within a few hundred iterations. For instance, it took 30 s to estimate this factorisation.

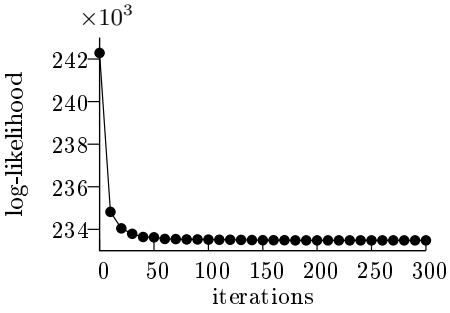


Figure 4.11 – Convergence of the non-negative matrix factorisation

When the scintillator is not in contact with the vessels as in the previous example, but rather 50 mm away as it is expected when acquisitions measure vessel activities at the knee location, it is difficult to retrieve the initial kinetics. Fig. 4.12 shows the initial kinetics (circle points) simulated with GATE, as well as the estimation of the spatio-temporal distribution obtained when the estimation is based on Eq. (4.7). The estimated kinetics is very different from the expected one. Fig 4.13 which shows the result obtained with non-negative matrix factorisation does not give significantly better result, although this solution is exp(4220) times more likely. This means that from a statistical point of view, there is not enough counts to separate the two sources.

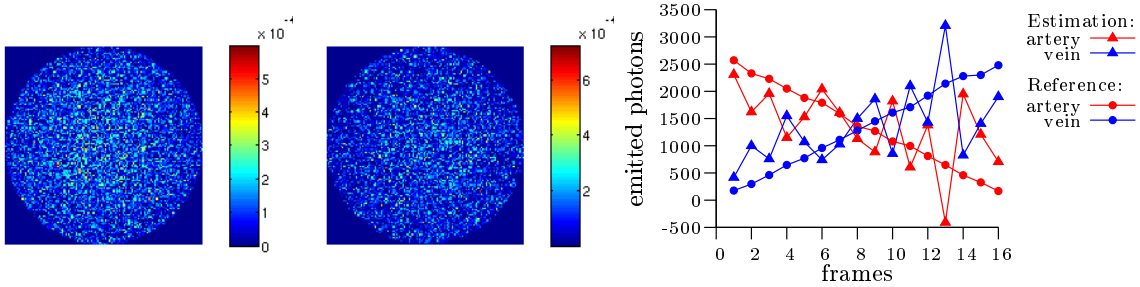


Figure 4.12 – Estimation of the spatiotemporal distribution on two components through SVD-based method when the scintillator is 50 mm-away from the vessels

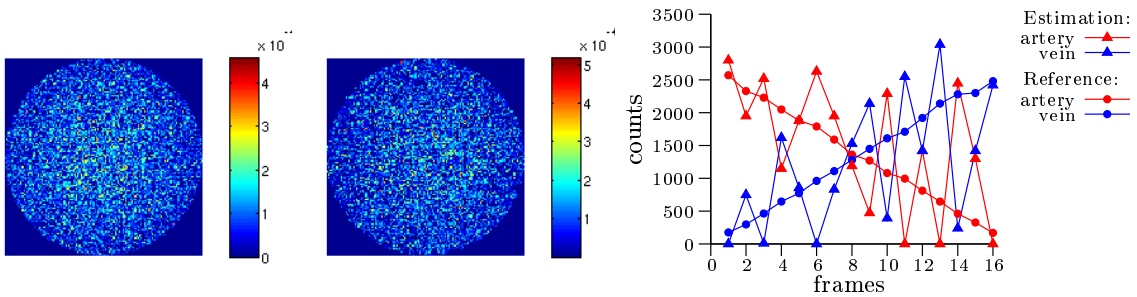


Figure 4.13 – Refinement of the spatio-temporal estimation through minimisation of the objective function of Eq (4.15)

In the next section, it will be shown how spatiotemporal distributions can be better improved, by estimating the spatial components in the object-space rather than in the projection space.

4.2 Components lay in the object-space

As it mentioned in the discussion of article (Aykac 01), the "TAC (should) be derived from sinograms and not images". This section presents a method based on this idea, although there is no sinogram in coded-aperture imaging but coded observations. First a new model for non-negative matrix factorisation is introduced to estimate the spatial components in the object-space instead of the image-space. This requires that reconstructions are incorporated in the non-negative matrix factorisation. The method is illustrated with examples with one component, then with two components. As for classical NMF, this method has several local minima. As a result, special care should be taken to initialise the algorithm. It is initialised from a SVD-based method as in the previous section, but also from the segmentation of the spatial component obtained from the 1-component model which admits a unique maximum.

4.2.1 Model

It is intended to perform non-negative matrix factorization. $\boldsymbol{A}$ is supposed to be known. We use Poisson statistics for modelling observations $\boldsymbol{S}$:

$$\boldsymbol{S} = \boldsymbol{AXK} + \boldsymbol{N} \tag{4.20}$$

The likelihood $L(\boldsymbol{X}, \boldsymbol{K}) = \text{Prob}(\boldsymbol{S}|\boldsymbol{X}, \boldsymbol{K})$ is:

$$L(\boldsymbol{X}, \boldsymbol{K}) = \prod_{i,j} \frac{e^{-(\boldsymbol{AXK})_{i,j}} (\boldsymbol{AXK})_{i,j}^{s_{i,j}}}{s_{i,j}!} \tag{4.21}$$

An objective function E is derived from likelihood $L(\boldsymbol{X}, \boldsymbol{K})$. An additional term ε is added for the robustness of the algorithm as in Eq. (4.14):

$$E = \sum_{i,j} \left[(\boldsymbol{AXK})_{i,j} + \varepsilon - s_{i,j} \ln\Big((\boldsymbol{AXK})_{i,j} + \varepsilon \Big) \right] \tag{4.22}$$

It is needed to compute $\vec{\nabla} E$ on every $x_{a,b}$ and $k_{b,c}$ so that the spectral projected gradient minimisation can be applied to estimate the most likely non-negative matrix factorisation:

$$\forall (a,b,c), \quad \frac{\partial E}{\partial x_{a,b}} = \left(\boldsymbol{A}^{\mathrm{T}} \left(\mathbb{U} - \frac{\boldsymbol{S}}{\boldsymbol{AXK} + \varepsilon} \right) \boldsymbol{K}^{\mathrm{T}} \right)_{a,b} \tag{4.23}$$

$$\frac{\partial E}{\partial k_{b,c}} = \left((\boldsymbol{AX})^{\mathrm{T}} \left(\mathbb{U} - \frac{\boldsymbol{S}}{\boldsymbol{AXK} + \varepsilon} \right) \right)_{b,c} \tag{4.24}$$

Calculations are detailed in appendix 5. This is very different from performing NMF in the image-space. Indeed, $\boldsymbol{A}$ acts as a filter and it constrains the spatial components $\boldsymbol{Y}$ of the factorisation of $\boldsymbol{S}$ to be in the image $\mathcal{I}m(\boldsymbol{A})$ of $\boldsymbol{A}$:

$$\boldsymbol{S} = \boldsymbol{YK} = (\boldsymbol{A} \cdot \boldsymbol{X}) \cdot \boldsymbol{K} \tag{4.25}$$

In the following developments, the term "spatial components" refers to the spatial components $\boldsymbol{X}$ in object-space, not to the spatial components $\boldsymbol{Y}$ in the image-space.

4.2.2 One-component model

This section presents how NMF can be used to reconstruct several time frames simultaneously. Examples which are presented here are based on real data acquired according to the

experimental set-up described in section 3.5.3. Observations are registered in list-mode, as a result it is possible to generate any kinetics by creating temporal frames with as many detected photons as wanted. Modelling the spatiotemporal distribution with a single component means that it is intended to factorise $\boldsymbol{S}$ as:

$$\boldsymbol{S} = \boldsymbol{A} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} k_1 & \dots & k_T \end{bmatrix} + \boldsymbol{N} \tag{4.26}$$

where $\boldsymbol{N}$ is a noise matrix. When a single component is taken into account, partial differentiations on temporal components (k_j) can be solved analytically, which leads to a new objective function which only depends on the spatial components (x_i):

$$E = \sum_{i,j} (\boldsymbol{AX})_i k_j - \sum_{i,j} s_{i,j} \ln (\boldsymbol{AX})_i k_j \tag{4.27}$$

$$\forall \beta, \quad \frac{\partial E}{\partial k_\beta} = \sum_i (\boldsymbol{AX})_i - \sum_i \frac{s_{i,\beta}}{k_\beta} \tag{4.28}$$

From Eq. 4.27, we can solve $\boldsymbol{K}$ that maximises E:

$$\forall \beta, \quad \frac{\partial E}{\partial k_\beta} = 0 \quad \Leftrightarrow \quad \forall \beta, \quad k_\beta = \frac{\sum_i s_{i,\beta}}{\sum_i (\boldsymbol{AX})_i} \quad \Leftrightarrow \quad \boldsymbol{K} = \frac{\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \boldsymbol{S}}{\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \cdot \boldsymbol{A} \cdot \boldsymbol{X}} \tag{4.29}$$

$\sum_i (\boldsymbol{AX})_i$ is a scalar which only depends on $\boldsymbol{A}$ and $\boldsymbol{X}$. As for any spatiotemporal distribution $\boldsymbol{S}$, the optimal temporal parameters $\boldsymbol{K}$ are known, it is only needed to optimise on spatial component $\boldsymbol{X}$. Eq 4.27 and 4.29 lead to:

$$E = \sum_{i,j} \left((\boldsymbol{AX})_i \frac{\sum_{i'} s_{i',j}}{\sum_{i'} (\boldsymbol{AX})_{i'}} \right) - \sum_{i,j} s_{i,j} \ln \left((\boldsymbol{AX})_i \frac{\sum_{i'} s_{i',j}}{\sum_{i'} (\boldsymbol{AX})_{i'}} \right) \tag{4.30}$$

$$= \frac{1}{\sum_{i'} (\boldsymbol{AX})_{i'}} \sum_i (\boldsymbol{AX})_i \sum_{i',j} s_{i',j}$$

$$- \sum_{i,j} \left(s_{i,j} \ln (\boldsymbol{AX})_i + s_{i,j} \ln \sum_{i'} s_{i',j} - s_{i,j} \ln \sum_{i'} (\boldsymbol{AX})_{i'} \right) \tag{4.31}$$

$$= \sum_{i',j} s_{i',j} - \sum_{i,j} s_{i,j} \ln \sum_{i'} s_{i',j} - \sum_{i,j} s_{i,j} \left(\ln (\boldsymbol{AX})_i - \ln \sum_{i'} (\boldsymbol{AX})_{i'} \right) \tag{4.32}$$

Partial differentiation on x_α gives:

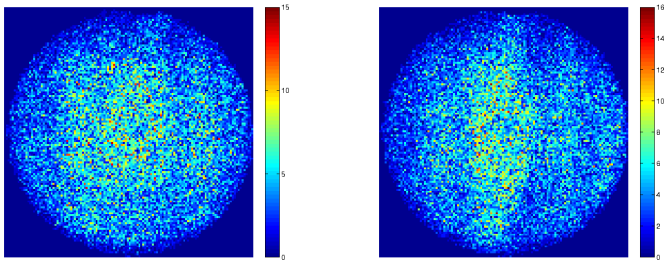
$$\frac{\partial E}{\partial x_\alpha} = - \sum_{i,j} s_{i,j} \left(\frac{a_{i,\alpha}}{(\boldsymbol{AX})_i} - \frac{\sum_{i'} a_{i',\alpha}}{\sum_{i'} (\boldsymbol{AX})_{i'}} \right) \tag{4.33}$$

$$= - \sum_{i,j} s_{i,j} \left(\frac{a_{i,\alpha}}{(\boldsymbol{AX})_i} \right) + \left(\sum_{i,j} s_{i,j} \right) \frac{\sum_{i'} a_{i',\alpha}}{\sum_{i'} (\boldsymbol{AX})_{i'}} \tag{4.34}$$

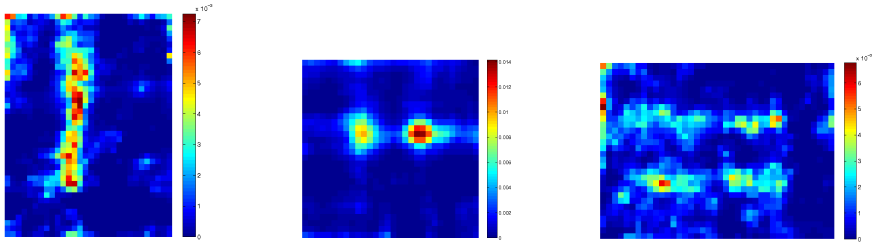
$$= - \left(\boldsymbol{A}^{\mathrm{T}} \frac{\boldsymbol{S} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}}}{\boldsymbol{AX}} \right)_\alpha + \left(\frac{\sum_{i,j} s_{i,j}}{\sum_{i'} (\boldsymbol{AX})_{i'}} \right) (\boldsymbol{A}^{\mathrm{T}} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}})_\alpha \tag{4.35}$$

$$= \left(-\boldsymbol{A}^{\mathrm{T}} \left(\frac{\boldsymbol{S} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}}}{\boldsymbol{AX}} - \frac{\sum_{i,j} s_{i,j}}{\sum_{i'} (\boldsymbol{AX})_{i'}} \right) \right)_\alpha \tag{4.36}$$

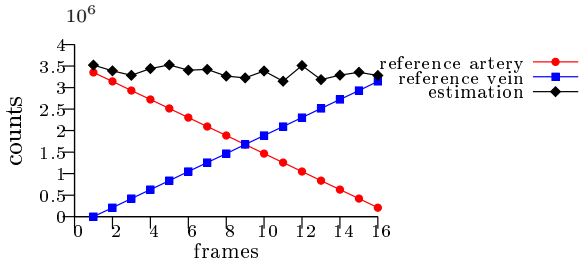
The algorithm is initialised with a positive homogeneous image $\boldsymbol{X}^0$: $\forall i, \quad (\boldsymbol{X}^0)_i = x_0 > 0$. The value of x_0 does not impact the algorithm. Indeed applying Eq. (4.36) to a homogeneous spatial component results in a null gradient. The objective function E is minimised with the spectral projected gradient method that uses Eq. (4.36) as the gradient of the objective function. The estimation of the spatio-temporal distribution for a one-component model is shown in Fig. 4.14. The algorithm estimates the kinetics in every frame and the spatial distribution of the object.



(a) Integration of all time frames for every projection



(b) Integration of all the projections



(c) Representation of the spatial distribution in (xz), (xz) and (yz) planes

Figure 4.14 – Reconstruction of all frames simultaneously

Identical solution would have been obtained by performing the following two-step procedure: first, the total spatial distribution $\boldsymbol{\lambda}^{tot}$ is estimated from the observations which are acquired for the whole time frames. Secondly, the spatial distribution $\boldsymbol{\lambda}^t$ for every time frame t is obtained by weighted the total distribution $\boldsymbol{\lambda}^{tot}$ with the number of photons detected in the time frame t over the total number of detected photons.

4.2.3 Model with several components

The estimation of the β^+-input function requires that at least two components are taken into account in the NMF-model, because it is expected to separate the contribution of the artery from the contribution of the vein. This section shows several procedures to estimate the spatio-temporal distribution of every of the components. The objective function and its gradient are determined previously in Eq. (4.22), (4.23) and (4.24).

4.2.3.1 Initialisation

The initialisation is important because:

- This algorithm is not convex, consequently it gets trapped in local minima
- The convergence rate is faster if the initialisation point is not too far from the convergence point

From Eq. 4.22, it is possible to set an initial point which is quite good. It is assumed that for the initialisation, the temporal components in $\boldsymbol{K}$ are identical, not necessarily homogeneous. This assumption means that:

$$\forall j, j', k, \quad k_{jk} = k_{j'k} \tag{4.37}$$

The objective function can then be written:

$$E = \sum_{i,j} \left((\boldsymbol{AXK})_{i,j} - s_{i,j} \ln \left((\boldsymbol{AXK})_{i,j} \right) \right) \tag{4.38}$$

$$= \sum_{i,j} \left(\boldsymbol{AX} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} k_1 & \dots & k_T \end{bmatrix} \right)_{i,j} - \sum_{i,j} s_{i,j} \ln \left(\left(\boldsymbol{AX} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} k_1 & \dots & k_T \end{bmatrix} \right)_{i,j} \right) \tag{4.39}$$

$$= \sum_{i,j} \left(\boldsymbol{AX} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \right)_i k_j - \sum_{i,j} s_{i,j} \ln \left(\left(\boldsymbol{AX} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \right)_i k_j \right) \tag{4.40}$$

$$= \sum_{i,j} \left(\boldsymbol{AX} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \right)_i k_j - \sum_{i,j} s_{i,j} \ln \left((\boldsymbol{AX} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}})_i \right) - \sum_{i,j} s_{i,j} \ln k_j \tag{4.41}$$

Differentiation on k_t results in:

$$\frac{\partial E}{\partial k_t} = \sum_i \left(\boldsymbol{AX} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \right)_i - \sum_i \frac{s_{i,t}}{k_t} \tag{4.42}$$

which means that:

$$\forall t, \quad \frac{\partial E}{\partial k_t} = 0 \quad \Leftrightarrow \quad \forall t, \quad k_t = \frac{\sum_i s_{i,t}}{\sum_i \left(\boldsymbol{AX} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \right)_i} = \frac{\sum_i s_{i,t}}{\sum_{i,j} (\boldsymbol{AX})_{i,j}} \tag{4.43}$$

$$\Leftrightarrow \quad \boldsymbol{k} = \frac{\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \boldsymbol{S}}{\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \boldsymbol{AX} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}}} \tag{4.44}$$

$\boldsymbol{K}$ is then retrieved using:

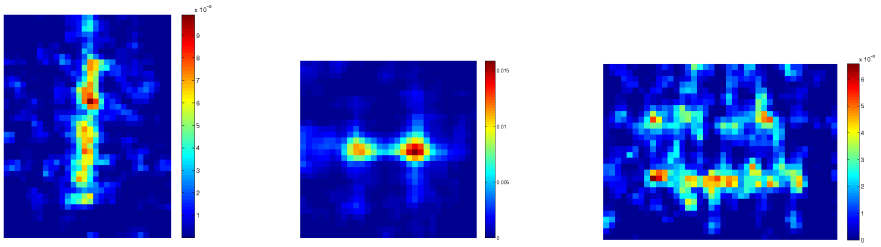
$$\boldsymbol{K} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \cdot \boldsymbol{k} \tag{4.45}$$

This means that whatever the spatial distribution of components is, by assuming that the temporal components are initialised from identical components, an optimal initialisation point for $\boldsymbol{K}$ is found. If it is also assumed that all coefficients of $\boldsymbol{X}$ are equal, it comes:

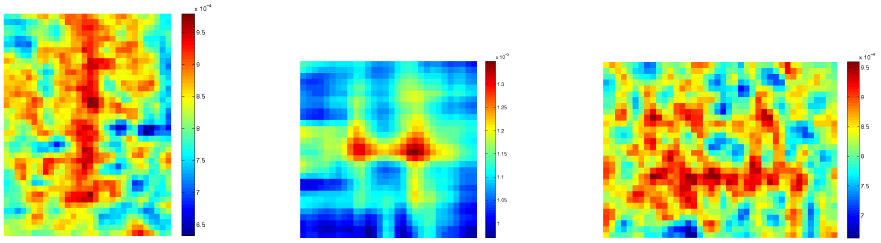
$$k_t = \frac{\sum_i s_{i,t}}{x_0 \cdot \sum_{i,j} a_{i,j}} \tag{4.46}$$

However, this leads to poor estimations of the spatiotemporal distribution. Indeed, the algorithm is trapped in local minima far from the optimal solution. In fact, the one-component

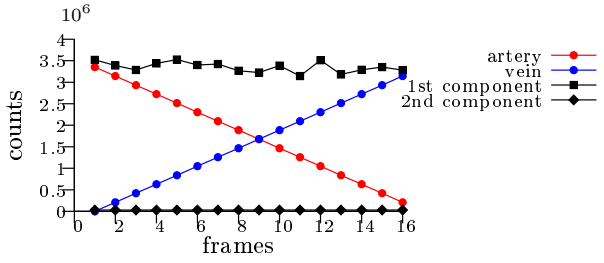
model converges to a more likely solution than the solution given by the two-component model which is initialised with homogeneous $\boldsymbol{X}$ and identical temporal components in $\boldsymbol{K}$. This is due to the fact that the two-component model has several local minima whereas the one-component has none. As a result, attempts are done to initialise the two-component model from the one-component solution. The second spatial component and the second temporal component are chosen to be homogeneous and small compared to their respective first components in order not to impact too much the objective function. The two components are shown in Fig. 4.15. It can be seen that the algorithm does not succeed in separating the two components. The initialisation should be more accurate so that it is possible to estimate the input function.



(a) Integration of all the projections for the first component



(b) Integration of all the projections for the second component



(c) Temporal components

Figure 4.15 – Reconstruction of all frames altogether

4.2.3.2 SVD-based initialisation

It is expected that a better initialisation point can be estimated by performing a SVD of the total reconstructed distribution. Before performing SVD, reconstructions are performed on every time frame. All time frames can be efficiently reconstructed by estimating the most likely parameters of the model:

$$\boldsymbol{S} = \boldsymbol{A} \cdot \boldsymbol{\Lambda} + \boldsymbol{N}_P \tag{4.47}$$

where $\boldsymbol{\Lambda}$ is $(N \times T)$ matrix. The likelihood that a matrix $\boldsymbol{\Lambda}$ gives rise to the observation matrix $\boldsymbol{S}$ is:

$$E = \ln(P(\boldsymbol{S}|\Lambda)) \tag{4.48}$$

$$= \sum_{i,j} -(\boldsymbol{A} \cdot \Lambda)_{i,j} + s_{i,j} \ln((\boldsymbol{A} \cdot \Lambda)_{i,j}) + Cst \tag{4.49}$$

The most likely solution is obtained through a maximisation procedure which requires the calculation of the gradient of E in Eq.(4.49):

$$\forall m, n, \quad \frac{\partial E}{\partial \lambda_{m,n}} = \sum_{i} -a_{i,m} + s_{i,n} \frac{a_{i,m}}{(\boldsymbol{A}\Lambda)_{i,n}} \tag{4.50}$$

$$= \left(\boldsymbol{A}^{\mathrm{T}} \left(\frac{\boldsymbol{S}}{\boldsymbol{A}\Lambda} - \mathbb{U} \right) \right)_{m,n} \tag{4.51}$$

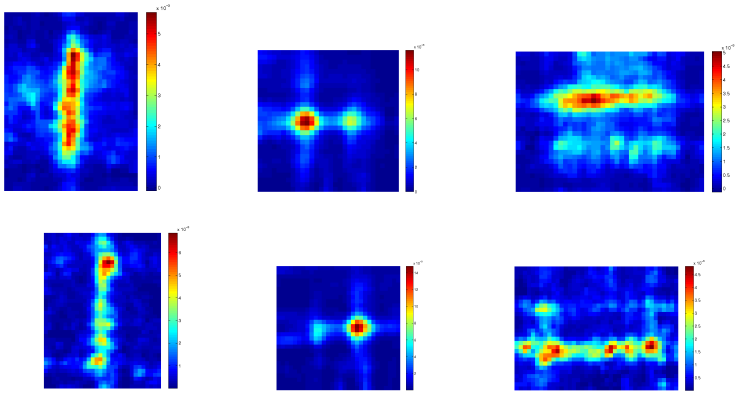
Constrained optimisation is then performed in order to maximise Eqs. (4.49). This optimisation is concave because the second derivatives of E are negative, as it can be derived from Eq. (4.50). SVD is then applied on the matrix $\boldsymbol{\Lambda}$. Every column t of $\boldsymbol{\Lambda}$ is the most likely solution to have the t^{th} column of $\boldsymbol{S}$. By applying a linear combination to the components that are obtained with SVD, as in Eq. (4.7), a quite positive estimation of the spatiotemporal distribution of the artery and the vein is obtained, as shown in Fig. 4.16. Although it separates the components, this separation is not complete as the shape of the second component is visible in the first component in Fig. 4.16a and *vice-versa*. Furthermore, the error between the estimated kinetics and the reference kinetics is large, as shown in Fig. 4.16b. As a result it is proposed to perform a NMF by initialising the algorithm with this solution, in order to improve the estimation of the spatiotemporal distribution. First, the negative coefficients of the spatial and temporal components are set to zero. Then the NMF-algorithm is applied in order to find the most likely solution which models $\boldsymbol{S}$. The obtained solution is shown in Fig. 4.17. Although the spatial components are almost identical to the one obtained in Fig. 4.16a, the estimation of the temporal components is slightly better as shown in Fig. 4.17b. Fig. 4.17c shows the convergence of the algorithm. It is observed that the solution obtained with NMF is about exp(5000) times more likely than the solution obtained using SVD. It can be noticed that there is a bump around the tenth iteration. This is due tot the fact that the spectral projected gradient method does not impose that the objective function decreases from one iteration to the other (Birgin 00). This allows escaping some local minima.

The accuracy on the estimation of the kinetics components is not sufficient to determine the β^+-input function, especially when the kinetics are not as simple as in the previous examples. Furthermore, the kinetics that are simulated in the two previous examples are 30 times larger than what is expected to measure in real acquisitions. As a result, it is suggested to help the algorithm by segmenting the spatial components from the spatial solution of the one-component model.

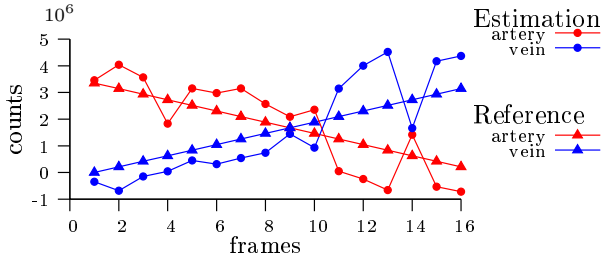
4.2.3.3 Reconstructions with either known $\boldsymbol{X}$ or known $\boldsymbol{K}$

Two more alternatives exist to improve the estimation of the components of the spatiotemporal distribution:

- if the spatial components in $\boldsymbol{X}$ are known, for example by segmenting the spatial component of the solution of the one-component reconstruction

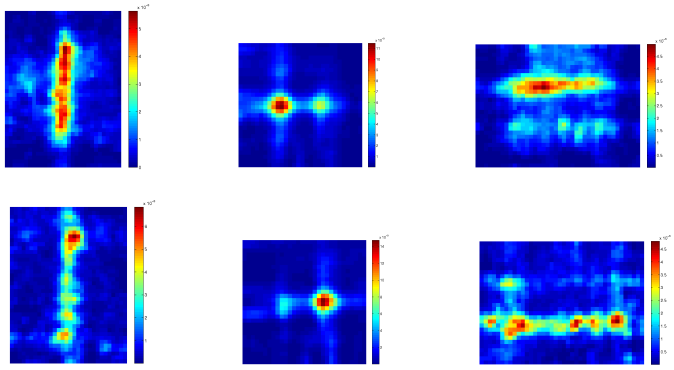


(a) Integration of the three projections. The first component is shown in the first row, the second component in the second row

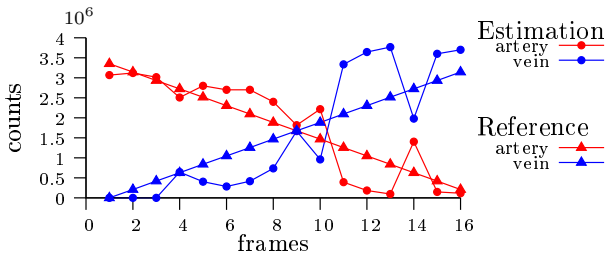


(b) Temporal components

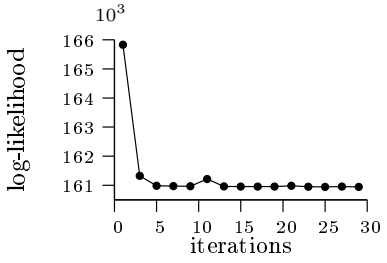
Figure 4.16 – Representation of the two first components that are obtained by applying a linear combination to the SVD of the estimated matrix $\mathbf{\Lambda}$



(a) Integration of the three projections. The first component is shown in the first row, the second component in the second row



(b) Temporal components



(c) Convergence of the NMF algorithm in object-space

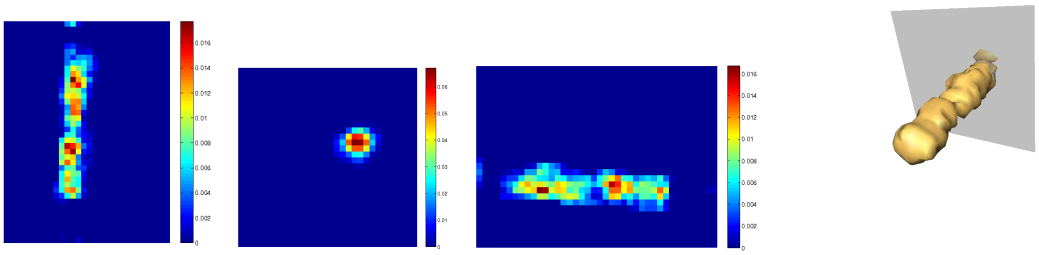
Figure 4.17 – Estimation of the spatiotemporal distribution of the object, initialisation from the solution shown in Fig. 4.16

- if kinetics in $\boldsymbol{K}$ are known, for example from the counts measured in detectors and *a priori* information on the kinetics of tissues which are in the field of view.

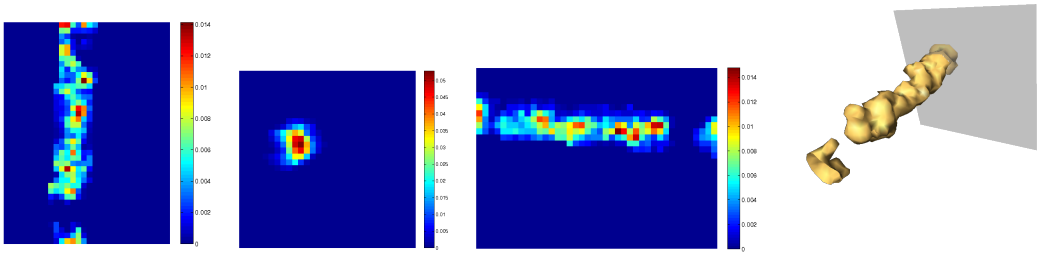
In these cases, reconstructions are convex, because E becomes positive semi-definite. Indeed, if it is assumed that the spatial matrix $\boldsymbol{X}$ is known, the second derivative of E in Eq. (4.22) is obtained by differentiating Eq. (4.24):

$$\frac{\partial^2 E}{\partial k_{b',c'} \partial k_{b,c}} = \sum_i (\boldsymbol{AX})_{ib} \frac{s_{ic'} (\boldsymbol{AX})_{ib'}}{(AXK + \varepsilon)^2_{ic'}} \geq 0 \tag{4.52}$$

By differentiating $\boldsymbol{\nabla}_{\boldsymbol{X}} E$ in Eq. (4.23), it would be shown that if $\boldsymbol{K}$ is assumed to be known, the optimisation on $\boldsymbol{X}$ is also convex. This is applied to estimate the β^+-input function. In the following examples, the estimations of kinetics are performed from real data. The centre-to-centre distance between the two cylinders is 10 mm.



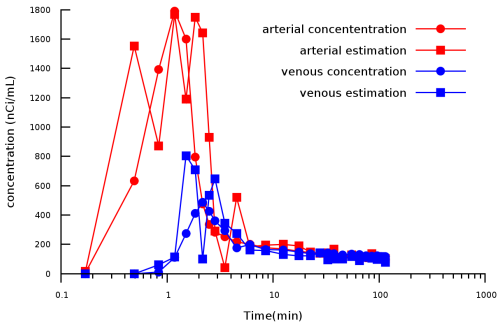
(a) Three projected views of the arterial spatial distribution



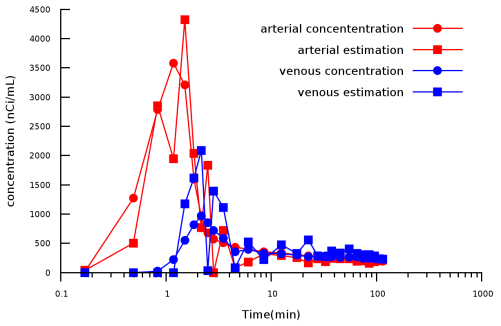
(b) Three projected views of the venous spatial distribution

Figure 4.18 – The two spatial components which are obtained using the segmentation of the 1-component distribution shown in Fig. 4.14a

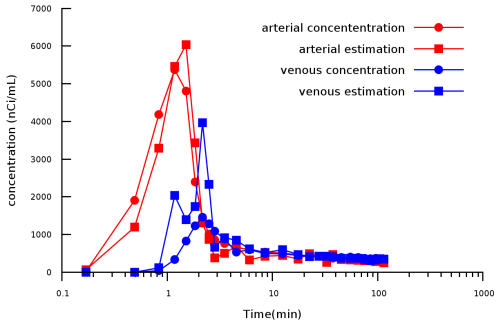
First, the spatiotemporal distribution is estimated with the one-component model. The spatial component is then segmented and Fig. 4.18 shows the two obtained components. Then, the optimal initial temporal matrix $\boldsymbol{K}$ is estimated with Eq. (4.45), with the assumptions that both temporal components are identical in the initial point $\boldsymbol{K}^0$. The objective function E in Eq. (4.22) is then optimised on parameters $\boldsymbol{K}$ and the estimated kinetics are shown in Fig. 4.19. Depending on the amount of information, the estimation of β^+-input function can be correct. In Fig. 4.19a, the amount of detected photons which is taken into account is 3 times larger than what is expected in real clinical investigations. The error on the estimation of the kinetics is large, which means that the amount of detected photons is not sufficient. Nevertheless, when the amount of detected photons is about 10 times larger than what is expected from clinical acquisitions, the kinetics are well estimated, as shown in Fig. 4.19c and 4.19d



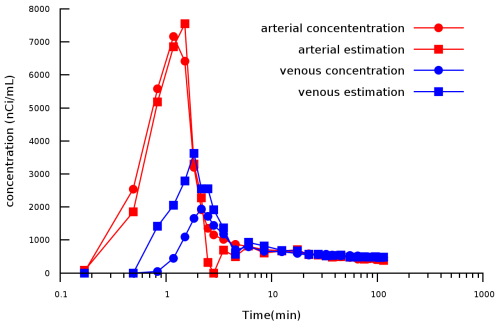
(a) Estimation of concentration from kinetics which are 3 times larger than the expected concentration in vessels



(b) Estimation of concentration from kinetics which are 6 times larger than the expected concentration in vessels



(c) Estimation of concentration from kinetics which are 9 times larger than the expected concentration in vessels



(d) Estimation of concentration from kinetics which are 12 times larger than the expected concentration in vessels

Figure 4.19 – Estimation of the arterial and venous concentrations from 4 various levels of concentrations

4.3 Discussion

As the amount of detected photons is low, especially in the first temporal frames, it is not possible to reconstruct accurately the spatial distribution of the vessels from a single frame, and consequently, it is not possible to discriminate the contribution of the artery from the contribution of the vein. In this chapter, it is proposed to model the spatiotemporal distribution of either the projection or the object as the product of two decoupled distributions: a distribution describing the spatial components which lay in the object and a distribution of the temporal components describing the kinetics of the spatial components. This decomposition is based on non-negative matrix factorisation. It is first applied on projections in the detector. A more likely solution $(\boldsymbol{X}, \boldsymbol{K})$ than the solution obtained with the SVD-based method is found. However, the solutions are accurate only if the detector is in contact with the vessels, which is in practice not possible because vessels stand several centimetres inside the knee.

It is then proposed to apply NMF in the object-space, which requires that reconstructions must be incorporated to NMF. Reconstructions are taken into account by modifying models of NMF so that the system matrix is taken into account. When the estimation is performed on only one component, the results are good, that is to say the mean spatial distribution and the mean kinetics are properly estimated. The one-component algorithm is not convex, however it seems that it admits a unique optimum, although no proof is found. When several components are taken into account in the model of NMF, problems with local minima prevent

from converging to acceptable solutions and it is not possible to separate the contribution of the artery from the one of the vein. As a result, it is suggested to estimate the spatial distribution of the object from the one-component model which gives an acceptable solution, then to segment it in order to build two spatial components and finally, the kinetics are estimated by optimising the NMF-model while keeping the spatial components unchanged. This results in a good estimation of the kinetics components, as long as the concentration in the object is about ten times larger than what is expected in clinical investigations.

As it is not possible to inject a dose which is about ten times higher because the dose deposited in the patients should not be too large, the temporal resolution should be decreased. Further investigations should be led towards an adaptive algorithm that would bin the detected photons in frames so that the estimated kinetics would be below a threshold which would be supplied by the user. This seems possible to be achieved by taking advantage of the list-mode of the γ-imager which registers detected photons one-by-one.

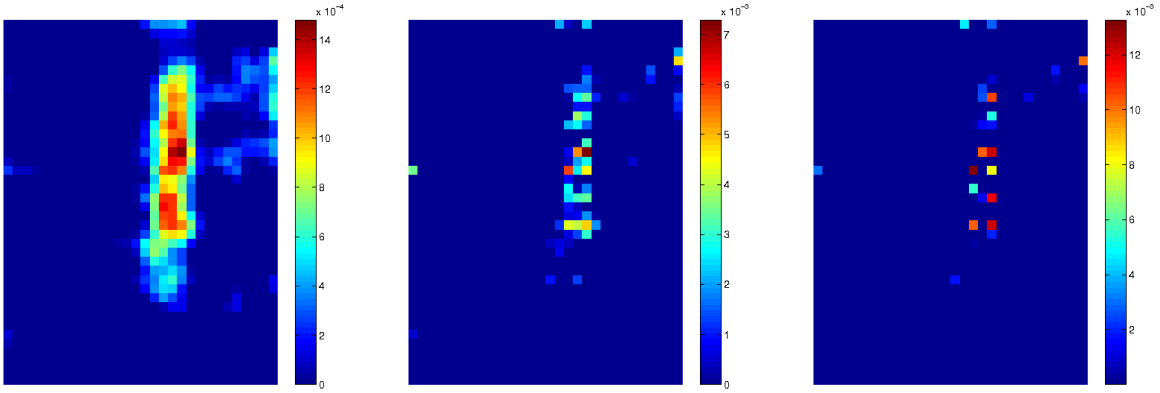
On the other hand, the methods which are developed in this chapter should be regularised in order to decrease the risk to get trapped in local minima, as it is done for other spatio-temporal decompositions (Ou 08). Regularisations can be introduced into non-negative matrix factorisation. Two cases occur: either the observed matrix $\boldsymbol{S}$ is decomposed on one component, or it is decomposed on more than one component, which requires the introduction of a new *a priori* function.

4.3.1 Regularisation of the one-component model

Regularisation appears to be necessary because the AXK-algorithm reconstructs high frequencies, which leads to reconstructions having no physical meaning. Fig. 4.20 presents a comparison of the reconstruction of a single cylinder at three successive iterations. Reconstruction of high frequencies may be explained by the system matrix imperfections. This can be avoided by regularising the spatial distribution $\boldsymbol{X}$. This can be done by adding a term to Eq. (4.41):

$$E = -\sum_{i,j} s_{i,j} \left(\ln(\boldsymbol{AX})_i - \ln \sum_{i'} (\boldsymbol{AX})_{i'} \right) + \beta \cdot \Phi(\boldsymbol{X}) \tag{4.53}$$

where β is a parameter controlling the impact of the prior on reconstructions. Φ is called an *a priori* function. *A priori* functions rely on Gibbs prior as explained in section 1.5.1.3. As the one-component model only depends on $\boldsymbol{X}$, regularisations can be applied as if it was a static reconstruction. Regularisation is more complicated when several components are taken into account in the model.



(a) iteration 20 (b) iteration 100 (c) iteration 700

Figure 4.20 – Comparison of reconstructions as iterations increase

4.3.2 Regularisation of the N-component model

The way the regularisation is applied is slightly more complicated than for the one-component model. Some adaptations have to be made on classical regularisation methods so that they can be applied on NMF-algorithms. Indeed, because of considerations of section 1.6.3, it exists a class of pairs $\{\boldsymbol{X}, \boldsymbol{K}\}$ which gives constant result for the calculation of the likelihood. Let's assume that the *a posteriori* objective function which takes into account the spatial regularization can be written as:

$$E = \sum_{i,j} \left[(\boldsymbol{AXK})_{i,j} - s_{i,j} \ln\Big((\boldsymbol{AXK})_{i,j} \Big) \right] + \beta \cdot \Phi(\boldsymbol{X}) \tag{4.54}$$

Let's also assume that $(\boldsymbol{X}^0, \boldsymbol{K}^0)$ is a solution of Eq. (4.54) and $\alpha > 0$. Then $(\frac{\boldsymbol{X}^0}{\alpha}, \frac{\boldsymbol{K}^0}{\alpha})$ is a better solution since the first likelihood term is constant and the regularisation term decreases. In order to avoid this effect, regularization term has to be modified, but in a way that makes its gradient still easily analytically derivable. It is proposed to write the regularisation term as in Eq. (4.55):

$$\Phi(\boldsymbol{X}, \boldsymbol{K}) = \sum_{c=1}^{C} \sum_{i=0}^{n_x-1} \sum_{j=0}^{n_y-1} \sum_{k=0}^{n_z-1} \left[\begin{array}{r} \varphi\left(s^c \cdot \left(x^c_{i+1,j,k} - x^c_{i,j,k} \right) \right) + \varphi\left(s^c \cdot \left(x^c_{i,j+1,k} - x^c_{i,j,k} \right) \right) \\ + \varphi\left(s^c \cdot \left(x^c_{i,j,k+1} - x^c_{i,j,k} \right) \right) \end{array} \right] \tag{4.55}$$

where $s^c = \sum_{t=0}^{T-1} k_{c,t}$ is the sum of the c^{th} component of matrix $\boldsymbol{K}$ which is represented by its c^{th} row. The gradient of Eq.(4.55) is derived in Eq. (4.57) and (4.58).

$$\frac{\partial \Phi(\boldsymbol{X},\boldsymbol{K})}{\partial x^{\xi}_{\alpha,\beta,\gamma}} = \sum_{i,j,k,c} \delta_{\xi,c} \cdot s^c \cdot \begin{bmatrix} \delta_{\beta,j}\delta_{\gamma,k}\left(\delta_{\alpha,i+1}-\delta_{\alpha,i}\right)\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{i+1,j,k}-x^{\xi}_{i,j,k}\right)\right) \\ +\delta_{\alpha,i}\delta_{\gamma,k}\left(\delta_{\beta,j+1}-\delta_{\beta,j}\right)\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{i,j+1,k}-x^{\xi}_{i,j,k}\right)\right) \\ +\delta_{\alpha,i}\delta_{\beta,j}\left(\delta_{\gamma,k+1}-\delta_{\gamma,k}\right)\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{i,j,k+1}-x^{\xi}_{i,j,k}\right)\right) \end{bmatrix} \tag{4.56}$$

$$= s^{\xi} \cdot \begin{bmatrix} \varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{\alpha,\beta,\gamma}-x^{\xi}_{\alpha-1,\beta,\gamma}\right)\right)-\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{\alpha+1,\beta,\gamma}-x^{\xi}_{\alpha,\beta,\gamma}\right)\right) \\ +\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{\alpha,\beta,\gamma}-x^{\xi}_{\alpha,\beta-1,\gamma}\right)\right)-\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{\alpha,\beta+1,\gamma}-x^{\xi}_{\alpha,\beta,\gamma}\right)\right) \\ +\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{\alpha,\beta,\gamma}-x^{\xi}_{\alpha,\beta,\gamma-1}\right)\right)-\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{\alpha,\beta,\gamma+1}-x^{\xi}_{\alpha,\beta,\gamma}\right)\right) \end{bmatrix} \tag{4.57}$$

$$\frac{\partial \Phi(\boldsymbol{X},\boldsymbol{K})}{\partial k_{\xi,\tau}} = \sum_{i,j,k} \begin{bmatrix} \left(x^{\xi}_{i+1,j,k}-x^{\xi}_{i,j,k}\right)\cdot\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{i+1,j,k}-x^{\xi}_{i,j,k}\right)\right) \\ \left(x^{\xi}_{i+1,j,k}-x^{\xi}_{i,j,k}\right)\cdot\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{i,j+1,k}-x^{\xi}_{i,j,k}\right)\right) \\ \left(x^{\xi}_{i+1,j,k}-x^{\xi}_{i,j,k}\right)\cdot\varphi'\left(s^{\xi}\cdot\left(x^{\xi}_{i,j,k+1}-x^{\xi}_{i,j,k}\right)\right) \end{bmatrix} \tag{4.58}$$

where δ is the Kronecker symbol. This new *a priori* function does not suffer from the effect described at the beginning of this section because (s^c) prevents the spatial components from collapsing. This function is convex as long as the potential function φ is convex.

Conclusion & Perspectives

The development of new drugs require that their interactions with tissues are well understood. The physiological mechanisms these new molecules are involved in are investigated in pharmacokinetics with compartmental analysis by labelling these molecules with radioelements, so that their distribution can be imaged *in-vivo*. Such analyses require that the radiotracer concentration in the blood is known as a function of time, so that the uptake of the molecules by tissues can be compared to the amount of molecules they receive. When the amount of metabolite is small, the concentration of the radiotracers in arterial blood is well estimated by the concentration of positron emitters in arterial blood, also called "β^+ arterial input function". Currently, it is estimated by drawing blood samples from patients, but a non-invasive method would be preferred. This can be achieved by imaging vessels and measuring the activity in the reconstructed images. However, the low concentration and the small volume of arteries lead to low activities whose spatial distribution is difficult to estimate. Furthermore a good spatial resolution is required because there is always a vein that stands close to any artery. The signal that comes from the vein can be viewed as a noise and should be separated from the signal that comes from the artery.

In the first chapter, it was chosen to measure the $\beta+$ arterial input function from the popliteal artery that stands in the knee, because it is a large artery that is close to the surface of the body. Furthermore, it stands far from the main active organs which decreases the background noise. It was decided to develop an imaging system with coded-aperture collimators rather than a coincidence imaging system. As the conditions for perfect coded-aperture imaging can not be fulfilled in medical imaging with high-energy photons, it was suggested to reconstruct spatial distributions that are acquired with coded-aperture masks with statistical algorithms that maximise the likelihood of the observations. Such methods require the computation of a system matrix that models the imaging system. For high-energy photons, this matrix is dense and the computation time is long because it takes time to take into account the contribution of complex collimators as coded-aperture masks. Consequently, some methods are proposed to efficiently compute the system matrix. It results in improvements of a factor up to 4 on the computation time, without deteriorating on the accuracy of the system matrix. These methods are faced-based ray-tracing algorithms.

Reconstructions were performed in the second chapter. Reconstructions of spatial distributions were first investigated with just one projection, then with two orthogonal projections. Performances of such an imaging system were investigated and the impact of the activity on the error of the reconstructed distribution was estimated, as well as the impact of the concentration on the spatial resolution between two cylinders. Reconstructions were performed from projections that were first generated analytically, then from Monte-Carlo simulations and finally from real projections. These investigations show that it is possible to separate the artery from the vein with the amount of detected photons that is expected to collect for clinical acquisitions. Some methods were proposed to take advantage of symmetries in the configuration of the imaging system, in order to decrease the memory load for reconstructions, which can be a limitation when the objects are made of a large number of voxels. Furthermore, an accelerated MLEM-algorithm was proposed because the convergence rate of MLEM is slow. Indeed, in coded-aperture imaging, the observations are very intricate and

a large number of iterations is required to estimate spatial distributions. This acceleration is based on an adaptive use of the observations, where a rough representation of the projections is used for first iterations whereas the finest available representation is used when the algorithm is close to converge.

Although the second chapter shows that it is possible to separate the spatial distribution of the artery from the vein, this can not be done for every temporal frame, because there are not enough detected photons, in particular in first frames. Consequently, it was proposed to look for non-negative matrix factorisations that accurately model spatiotemporal observations in detectors. Such factorisations were estimated from projected gradient methods that constrain the factorisation to be positive. It was first applied in the detector-space, which returns more likely solutions than SVD-based methods. A new method was then introduced that incorporate the reconstruction in non-negative matrix factorisations. This new method achieves accurate decompositions when the factorisation is performed on one component, but when it is expected to model spatio-temporal distributions with more components, the algorithm converges to local minima and the factorisation does not model properly the underlying structures of the distributions. As an alternative, it was proposed to segment the two vessels from the spatial component that was estimated from the one-component model and then to estimate the most likely associated kinetics. This results in accurate estimations of temporal components if the radiotracer in vessels is at least ten times larger than from clinical investigations. Otherwise, kinetics estimations are noisy.

Some limitations have to be mentioned concerning this thesis. First, no background activity is taken into account: although it was decided to measure the β^+-input function from the popliteal artery in the knee, which is mostly made of bones and cartilages, some muscles are in the field of view and should uptake some of the radiotracers. Secondly, some validations should be performed with 511 keV photons, by appropriately calibrating the γ-imager. Finally it seems important to notice that coded-aperture collimators that were used for this thesis are not necessarily optimal for reconstructing spatial distributions with statistical algorithms. Some investigations should be led to determine an optimal collimator that results in an optimal system matrix for addressing low-dose imaging. A larger number of projections should also improve reconstructions but this requires that more symmetries are taken into account in order to keep the system matrix with dimensions that can be handled. This suggests that the spatial distribution should be reconstructed in cylindrical voxels as it is mentioned in Appendix 6. In such a case, reconstructions should also be parallelised so that the reconstruction time become short enough for routine use.

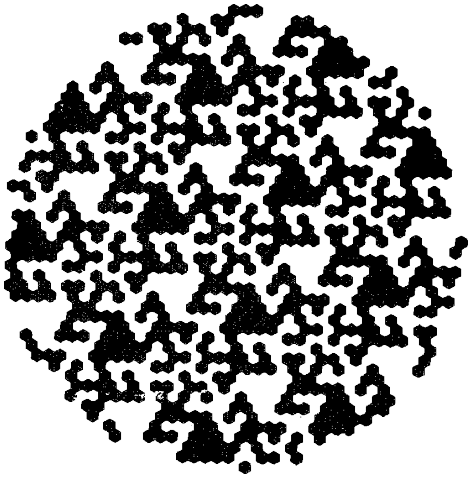
The results that this thesis provides can have applications for other low-dose imaging. For instance, some cameras are developed for detecting tumours in cancer surgery (Menard 98) and it would be of interest to increase the sensitivity of such an imaging system. It could also be applied for building high-energy high-resolution and high-sensitivity imaging systems, which are for the moment only addressed by PET-systems which are expensive. Finally, it could be applied to scintigraphic imaging in order to lower the dose that is administrated to patients.

Appendices

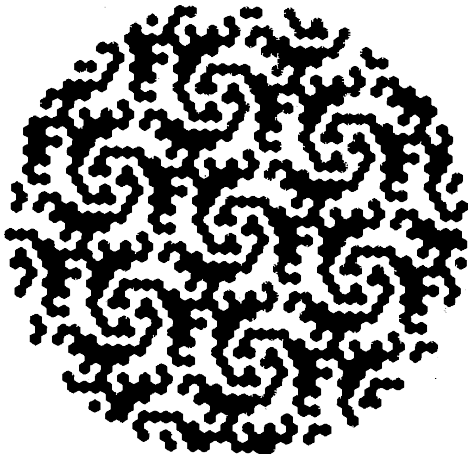
A

1 HURA PATTERNS

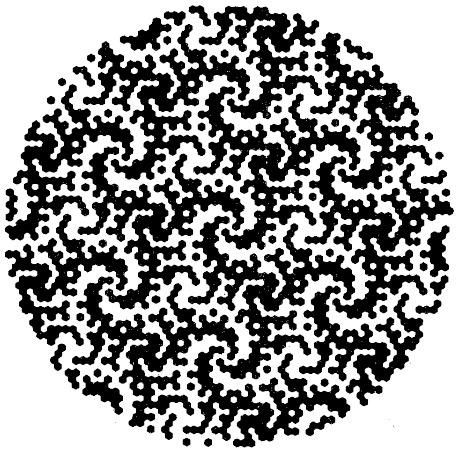
Here are shown some HURA patterns that could be used for coded-aperture collimators. For a given surface, it offers choices on the number and the dimensions of mask elements.



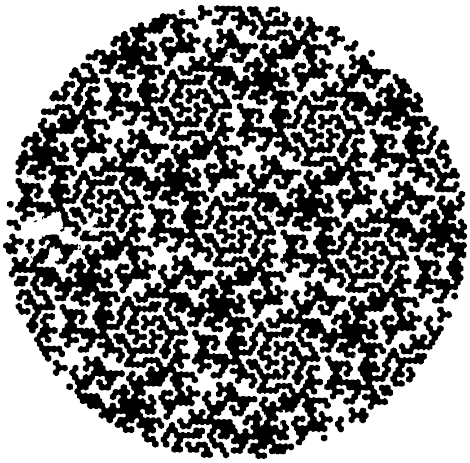
(a) Order $\nu = 139$



(b) Order $\nu = 151$



(c) Order $\nu = 331$



(d) Order $\nu = 619$

2 Minimisation of near-field artefacts for coded-aperture imaging

Zero-order and first-order artefacts will be investigated in next sections, and some methods are presented to correct those artefacts (Accorsi 01).

2.1 Zero-order artefact

Taylor expansion is performed on $\cos^3\left(\arctan\left(\frac{\|\boldsymbol{r}_i-\boldsymbol{r}_0\|}{z}\right)\right)$. When only the zero-order term is taken into account, expression (1.68) becomes:

$$P(\boldsymbol{r}_i) = \cos^3\left(\arctan\left(\frac{\|\boldsymbol{r}_i\|}{z}\right)\right) \int_{r'_0} \mathrm{d}^2\boldsymbol{r}'_0\, O'(\boldsymbol{r}'_0) A'\left(\boldsymbol{r}_i - \boldsymbol{r}'_0\right) \tag{A.1}$$

where $\boldsymbol{r}'_0$, O' and A' are respectively defined in Eq. (1.40), (1.41) and (1.42) . It is not possible to estimate directly the spatial distribution of the source $O(\boldsymbol{r}_0)$ because $P(\boldsymbol{r}_i)$ is not a correlation of O and A. However, by dividing $P(\boldsymbol{r}_i)$ by $\cos^3\left(\arctan\left(\frac{\|\boldsymbol{r}_i\|}{z}\right)\right)$, the object is retrieved from Eq. (A.2) which is very similar to the solution obtained with the far-field approximation in Eq. (1.49). The projection $P(\boldsymbol{r}_i)$ was just corrected before decoding by array G. The spatial distribution of the object is derived by adapting Eq. 1.49 into:

$$\hat{O} = \mathcal{F}^{-1}\left(\mathcal{F}\left(\frac{P(\boldsymbol{r}_i)}{\cos^3\left(\arctan\left(\frac{\|\boldsymbol{r}_i\|}{z}\right)\right)}\right)^H \star \mathcal{F}(G)\right) \tag{A.2}$$

This approximation physically means that it is assumed that all the source distribution stands in the centre of the field of view, as the cos-term does not depend on $\boldsymbol{r}_0$ in this approximation.

2.2 First-order artefact

The previous approximation can be refined by taking into account the first-order term of the Taylor expansion of $\cos^3\left(\arctan\left(\frac{\|\boldsymbol{r}_i-\boldsymbol{r}_0\|}{z}\right)\right)$, which is shown in Eq. (A.3)

$$\cos^3\arctan\left(\frac{\|\boldsymbol{r}_i-\boldsymbol{r}_0\|}{z}\right)_{|1} = \cos^3\arctan\left(\frac{\|\boldsymbol{r}_i\|}{z}\right)\frac{3}{z^2+\|\boldsymbol{r}_i\|^2}\left(\boldsymbol{r}_i\cdot\boldsymbol{r}_0\right) \tag{A.3}$$

After applying the zero-order correction, Eq. (A.4) is obtained.

$$\frac{P(\boldsymbol{r}_i)}{\cos^3\left(\arctan\left(\frac{\|\boldsymbol{r}_i\|}{z}\right)\right)} = O\star A - \frac{3a}{b}\frac{\boldsymbol{r}_i}{z^2+\|\boldsymbol{r}_i\|^2}\cdot\int_{\boldsymbol{r}'_0}\mathrm{d}^2\boldsymbol{r}'_0\;\boldsymbol{r}'_0 O'(\boldsymbol{r}'_0)A'\left(\boldsymbol{r}_i-\boldsymbol{r}'_0\right) \tag{A.4}$$

The second term on the right-hand side of Eq. (A.4) should be as small as possible so that O can be retrieved from deconvolution without too much artefacts. Accorsi analysed what these first-order artefacts $\hat{O}_1$ look like (Accorsi 01). The integral can be viewed as the first moment of inertia of the object distribution O' that is weighted by pattern A. The integral can be viewed as a constant, denoted O'', because for coded aperture masks, binary

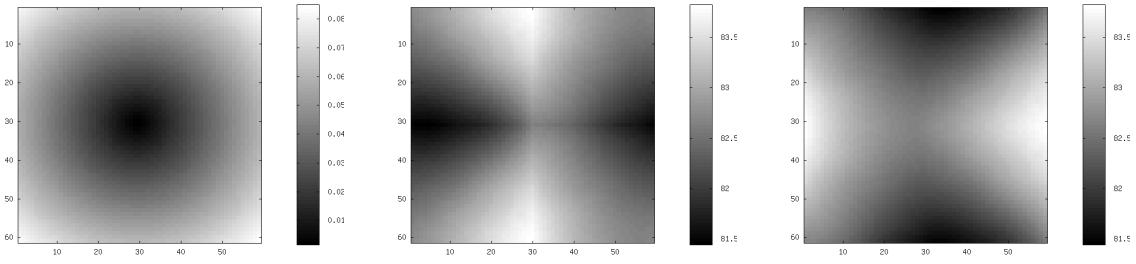
functions A distribute rather uniformly 0 and 1 : the convolution consequently depends slightly on shifts and the integrals are rather similar for every decoded point $\boldsymbol{r}_i$.

$$\frac{P(\boldsymbol{r}_i)}{\cos^3\left(\arctan\left(\frac{\|\boldsymbol{r}_i\|}{z}\right)\right)} \approx O \star A - \frac{3a}{b}\frac{\boldsymbol{r}_i}{z^2+\|\boldsymbol{r}_i\|^2}\cdot O'' \tag{A.5}$$

Decoding the second term which characterises the first-order artefact $\hat{O}_1$ leads to:

$$\hat{O}_1(\boldsymbol{r}_0) = -\frac{3a}{z}O''\cdot\int_{\boldsymbol{r}_i} \mathrm{d}^2\boldsymbol{r}_i \frac{\boldsymbol{r}_i}{z^2+\|\boldsymbol{r}_i\|^2}G(\boldsymbol{r}_i+\boldsymbol{r}_0) \tag{A.6}$$

From Eq. (A.6), it can be noted that the shape of the first-order artefact only depends on decoding matrix G. Accorsi proposed to shift the decoding pattern G so that it is centred. This method is known as "pattern centring". An example is presented in Fig. A.1. Artefacts are less sharp in Fig. A.1c than in Fig. A.1b. In order to further decrease this term, Accorsi recommends to acquire a first projection of the object and moves the imaging system laterally in order to centre the acquired distribution on the detector so that O'' goes to zero. In this case, the first-order artefact disappears because O'' then goes to zero.



(a) $\frac{\boldsymbol{r}_i}{\boldsymbol{r}_i^2+z^2}$ term

(b) Decorrelation of first-order term as the integrand of Eq. (A.6)

(c) Decorrelation based on a centred-pattern decoding matrix

Figure A.1 – 1st-order artefact and impact of pattern centring for a 59×61 twin prime URA

2.3 Second-order artefact

When zero-order and first-order artefacts are corrected, second-order artefacts become visible. It was also studied by Accorsi, but it only gives a prediction of the way artefacts look like. No method is proposed to reduce or suppress them. As a result, this family of artefacts is not presented here, all the more so as the physical interpretations are unclear.

3 BFGS-method for non-linear optimisation

Newton's method gives an update scheme to the minimisation problem $\min_{\mathbb{R}^n} f$:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x} - \rho_k \left[\text{Hess} f(\boldsymbol{x})\right]^{-1} \boldsymbol{\nabla} f(\boldsymbol{x}) \tag{A.7}$$

While it is intended at minimising an objective function f that is non-linear, derivation of $[\text{Hess} f(\boldsymbol{x})]^{-1}$ is time-consuming and possibly unstable. BFGS-method takes advantages of approximations of the hessian and its inverse to update efficiently the best current points $\boldsymbol{x}_k$. It is quasi-Newton method which means that the solution is looked for as the root of $\boldsymbol{\nabla f(x_k)}$. In quasi-Newton method, the hessian $\text{Hess} f(\boldsymbol{x}_k)$ is approximated by a matrix $\boldsymbol{B}_k$. f attains its minimum when expressions (A.8) and (A.9) are verified.

$$\boldsymbol{\nabla} f(\boldsymbol{x}_k) + \text{Hess} f(\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = \boldsymbol{0} \tag{A.8}$$

$$f \text{ is definite positive} \tag{A.9}$$

At every iteration, first step is to determine the search direction $\boldsymbol{p}_k$. Setting Eq. (A.8) to $\boldsymbol{0}$ and approximating the hessian $\text{Hess} f(\boldsymbol{x}_k)$ by $\boldsymbol{B}_k$, the new search direction $\boldsymbol{p}_k$ is given by:

$$\boldsymbol{p}_k = -\boldsymbol{B}_k^{-1} \boldsymbol{\nabla} f(\boldsymbol{x}_k) \tag{A.10}$$

Line search on $\alpha \mapsto f(\boldsymbol{x}_k - \alpha \boldsymbol{p}_k)$ is then performed and leads to a new point $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{p}_k$. After denoting $\boldsymbol{s}_k = \alpha_k \boldsymbol{p}_k$ and $\boldsymbol{y}_k = \boldsymbol{\nabla} f(\boldsymbol{x}_{k+1}) - \boldsymbol{\nabla} f(\boldsymbol{x}_k)$, the hessian at iteration $k+1$ is updated with the addition of two rank-1 matrices :

$$\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \frac{1}{\boldsymbol{y}_k^T \boldsymbol{s}_k} \boldsymbol{y}_k \boldsymbol{y}_k^T + \frac{1}{\boldsymbol{s}_k^T \boldsymbol{B}_k \boldsymbol{s}} \boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^T \boldsymbol{B}_k \tag{A.11}$$

The inverse of $\boldsymbol{B}_k$ in Eq. (A.10) is calculated very efficiently through Sherman-Morrison formula:

$$\boldsymbol{B}_{k+1}^{-1} = \boldsymbol{B}_k^{-1} + \frac{\boldsymbol{s}_k^T \boldsymbol{y}_k + \boldsymbol{y}_k^T \boldsymbol{B}_k^{-1} \boldsymbol{y}_k}{\left(\boldsymbol{s}_k^T \boldsymbol{y}_k\right)^2} (\boldsymbol{s}_k \boldsymbol{s}_k^T) - \frac{1}{\boldsymbol{s}_k^T \boldsymbol{y}_k} \left(\boldsymbol{B}_k^{-1} \boldsymbol{y}_k \boldsymbol{s}_k + \boldsymbol{s}_k \boldsymbol{y}_k^T \boldsymbol{B}_k^{-1}\right) \tag{A.12}$$

4 Impact of the energy of photons on the spatial resolution

The mean probability detection height is given by Eq. (A.13) where $p_{through}(z)$ is the probability that a photon reaches depth z and $d\,p_{detection}(z)$ is the probability that a photon is detected between z and $z+dz$. l is the intersecting length of the trajectory of photons with the scintillator.

$$h = \int_{z=0}^{l} z \cdot p_{through}(z) \cdot d\,p_{detection}(z) = \int_{z=0}^{l} z \cdot e^{-\mu_d z} \cdot \mu_d dz \tag{A.13}$$

$$= \frac{1}{\mu_d} \int_{w=0}^{\mu_d l} dw\, w\, e^{-w} = \frac{1}{\mu_d}\gamma(2, \mu_d l) \tag{A.14}$$

where γ is the lower incomplete gamma function (Abramovitz 70).

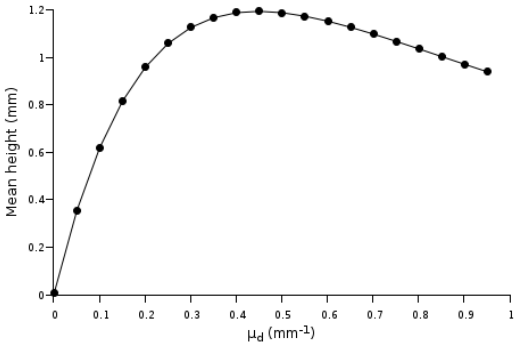


Figure A.2 – Mean height of detection in the 4 mm-thick scintillator

At 511 keV, the linear attenuation coefficient μ_d is 9×10^{-3} mm^{-1} while it is 31×10^{-2} mm^{-1} at 140 keV. This means that the average detection height in the scintillator is deeper at 140 keV than at 511 keV, which tends to deteriorate the spatial resolution. Indeed, it decreases the full surface at half maximum (FSHM) of the point spread function of the scintillator process $PSF(z)$.

5 Derivation of the gradient of the likelihood for 'AXK'-method

$$\frac{\partial E}{\partial x_{n,c}} = \frac{\partial}{\partial x_{n,c}} \left[\sum_{i,j,r,s} a_{i,r} x_{r,s} k_{s,j} - \sum_{i,j} s_{i,j} \ln \left(\sum_{r,s} a_{i,r} x_{r,s} k_{s,j} \right) \right] \tag{A.15}$$

$$= \sum_{i,j} a_{i,n} k_{c,j} - \sum_{i,j} s_{i,j} \frac{a_{i,n} k_{c,j}}{\sum_{r,s} a_{i,r} x_{r,s} k_{s,j}} \tag{A.16}$$

$$= \sum_{i,j} a_{i,n} \left(\frac{s_{i,j}}{(AXK)_{i,j}} - 1 \right) k_{c,j} \tag{A.17}$$

$$= \left(A^T \left(\frac{S}{AXK} - \mathbb{U} \right) K^T \right)_{n,c} \tag{A.18}$$

$$\boxed{\frac{\partial E}{\partial x_{n,c}} = \left(A^T \left(\frac{S}{AXK} - \mathbb{U} \right) K^T \right)_{n,c}} \tag{A.19}$$

$$\frac{\partial E}{\partial k_{c,t}} = \frac{\partial}{\partial k_{c,t}} \left[\sum_{i,j,s} (AX)_{i,s} k_{s,j} - \sum_{i,j} s_{i,j} \ln \left(\sum_{s} (AX)_{i,s} k_{s,j} \right) \right] \tag{A.20}$$

$$= \sum_{i} (AX)_{i,c} - \sum_{i} s_{i,t} \frac{(AX)_{i,c}}{\sum_{s} (AX)_{i,s} k_{s,t}} \tag{A.21}$$

$$= \sum_{i} (AX)_{i,c} \left(\frac{s_{i,t}}{(AXK)_{i,t}} - 1 \right) \tag{A.22}$$

$$= \left(X^T A^T \left(\frac{S}{AXK} - \mathbb{U} \right) \right)_{c,t} \tag{A.23}$$

$$\boxed{\frac{\partial E}{\partial k_{c,t}} = \left(X^T A^T \left(\frac{S}{AXK} - \mathbb{U} \right) \right)_{c,t}} \tag{A.24}$$

Eq. (A.18)

6 Cylindrical image representation

As coded aperture collimators generates large system matrix for statistical reconstruction algorithm, it is important to avoid redundancy in its calculation. Furthermore, it seems possible to increase number of projections that are used in the reconstruction without increasing the required RAM. This method takes advantage of symmetries that can be observed in transition matrix. While using Cartesian coordinates, just two symmetries can be exploited, while using cylindrical coordinates, number of symmetries increase as illustrated in Fig. A.3. While calculating the transition matrix, we can take advantage of invariants. It is then not required to calculate transition matrix for all n projections p_k since it can be directly deduced from projection p_1. First, let calculate the transition matrix just for p_1, that is to say we calculate the probability that a photon which is emitted from any voxel j is detected in pixel i of p_1. Let then say that there are v voxels per sector, then coefficient $a_{i',j}$ of transition matrix A is equal to coefficient $a_{i,j}$, with $i' = v \times (k-1) + i$. k is the index number of the projection. Consequently, it is as fast to calculate the transition matrix for one projection as for n-projections.

Figure A.3 – Invariants in cylindrical image representation

6.1 Construction of the cylindrical image representation

Among specifications, we would like that:

- voxel size is almost constant (about 1 mm^3)
- characteristics dimensions are almost equal (about $1 \times 1 \times 1$ mm^3)

There are several solutions to mesh the image in cylindrical coordinates. If voxel volume is constant, we can still choose:

- radial sampling distance to be constant ($R_n = (n+1) * R_0$), which leads to an increase of voxels in crowns as $p_n = 2 \cdot n - 1$. Problem is that $r_n / l_n \to +\infty$ as n increases.
- increase of voxels in crowns is regular, *i.e* $p_n = n$, in this case r_n is proportional to l_n and proportional ratio depends on the number of sectors m.

If incrementing by 1 the number of voxel in every sector of a m-sector circle (which is m-invariant, i.e m-rotation of $2\pi/m$ is the identity transformation) and choosing to have :

$$r_1 : \quad \pi \cdot r_0^2 = \pi/m \cdot \left((r_1 + r_0)^2 - r_0^2\right) \tag{A.25}$$

$$r_2 : \quad \pi \cdot r_0^2 = \pi/(2 \cdot m) \cdot \left((r_2 + r_1 + r_0)^2 - (r_1 + r_0)^2\right) \tag{A.26}$$

$$r_n : \quad \pi \cdot r_0^2 = \pi/(n \cdot m) \cdot \left(\left(\sum_{k=0}^{n} r_k\right)^2 - \left(\sum_{k=0}^{n-1} r_k\right)^2\right) \tag{A.27}$$

$$n \cdot m \cdot r_0^2 = R_n^2 - R_{n-1}^2 \tag{A.28}$$

$$R_n^2 = n \cdot m \cdot r_0^2 + R_{n-1}^2 \tag{A.29}$$

As a result, $R_1 = r_0 \cdot \sqrt{1+m}$, $R_2 = r_0 \cdot \sqrt{1+(1+2)\cdot m}$, $R_3 = r_0 \cdot \sqrt{1+(1+2+3)\cdot m}$. We demonstrate through induction that:

$$R_n = r_0 \cdot \sqrt{1 + m \cdot \sum_{k=0}^{n} k}$$

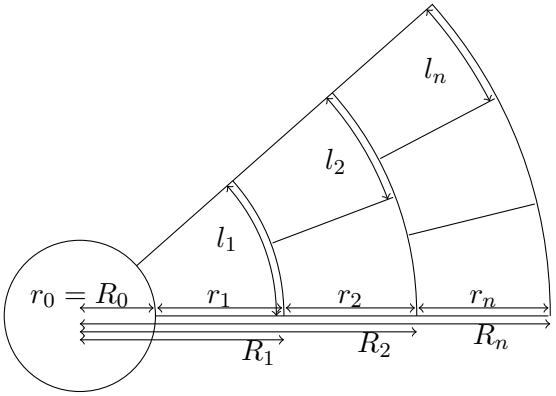


Figure A.4 – Voxel distribution and notations

We then deduce the step r_n and the curvilinear abscissa l_n:

$$r_n = r_0 \left(\sqrt{1 + m \cdot \sum_{k=0}^{n} k} - \sqrt{1 + m \cdot \sum_{k=0}^{n-1} k} \right) \tag{A.30}$$

$$l_n = \frac{2\pi}{m \times n} R_n \tag{A.31}$$

$$= \frac{2\pi r_0}{m \times n} \sqrt{1 + m \cdot \sum_{k=0}^{n} k} \tag{A.32}$$

This respects the first specification, that is to say that every voxel has the same volume. However, we will see that second specification is not respected, indeed:

$$r_n / l_n = \frac{m \cdot n}{2\pi} \left(1 - \sqrt{\frac{1 + m \cdot \sum_{k=0}^{n-1} k}{1 + m \cdot \sum_{k=0}^{n} k}} \right) \tag{A.33}$$

$$= \frac{m \cdot n}{2\pi} \left(1 - \sqrt{1 - \frac{m \cdot n}{1 + m \cdot \sum_{k=0}^{n} k}} \right) \tag{A.34}$$

$$= \frac{m \cdot n}{2\pi} \left(1 - \sqrt{1 - \frac{2 \cdot m \cdot n}{2 + m \cdot n \cdot (n+1)}} \right) \tag{A.35}$$

$$\underset{n \to \infty}{\sim} \frac{m \cdot n}{2\pi} \left(1 - \sqrt{1 - \frac{2}{n}} \right) \tag{A.36}$$

$$\underset{n \to \infty}{\sim} \frac{m}{2\pi} \tag{A.37}$$

As a result, depending on the number of sectors which are used in the cylindrical image representation, the ratio between the curvilinear abscissa l_n and the side of the voxel r_n is constant, but in general not equal to 1, which leads to rectangular voxels. If choosing $m = 6$, then $r_n \simeq l_n$. Problem is that having just 6-sector is very limiting. Indeed, 90° is not one of the invariant: we can not perform orthogonal acquisitions.

As far as I'm concerned, having 8 sectors could be nice, since it takes into account the most common projections: 0°, 45° and 90°-invariants. 12-sectors representation is good to since it offers to have 0°, 30°, 60°, 90°, 120°, 150°, 180° and so on.

Best trade-off would be to use 24-sector representations since it has invariants along most common angles: 0°, 15°, 30°, 45°, 60° 75°, 90° and so on.

The idea to deal with large number of sectors is to decrease the frequency of evolution of the number of voxels per crown. Problem is that if the frequency is not the inverse of an integer (this is never the case), then the number of voxels per crown per sector is not integer. As a result, we round it to the closest integer. We can define the radii in terms of the frequency of the evolution of the number of voxels per crown per sector:

$$R_n = r_0 \cdot \sqrt{1 + f \cdot m \cdot \sum_{k=0}^{n} k} \tag{A.38}$$

$$= r_0 \cdot \sqrt{1 + 2\pi \cdot \sum_{k=0}^{n} k}, \text{ with } f = 2\pi/m \tag{A.39}$$

r_n becomes:

$$r_n = r_0 \cdot \left(\sqrt{1 + 2\pi \cdot \sum_{k=0}^{n} k} - \sqrt{1 + 2\pi \cdot \sum_{k=0}^{n-1} k} \right) \tag{A.40}$$

It converges to $r_0 \cdot \sqrt{\pi} = 1$ for $n \to \infty$:

$$r_n = r_0 \cdot \sqrt{1 + 2\pi \cdot \sum_{k=0}^{n} k} \left(1 - \sqrt{\frac{1 + 2\pi \cdot \sum_{k=0}^{n-1} k}{1 + 2\pi \cdot \sum_{k=0}^{n} k}} \right) \tag{A.41}$$

$$= r_0 \cdot \sqrt{1 + \pi \cdot n(n+1)} \left(1 - \sqrt{1 - \frac{2\pi n}{1 + 2\pi \cdot \sum_{k=0}^{n} k}} \right) \tag{A.42}$$

$$= r_0 \cdot \sqrt{1 + \pi \cdot n(n+1)} \left(1 - \sqrt{1 - \frac{2\pi n}{1 + \pi \cdot n(n+1)}} \right) \tag{A.43}$$

$$\underset{n \to \infty}{\sim} r_0 \cdot n \cdot \sqrt{\pi} \left(1 - \sqrt{1 - \frac{2}{n}} \right) \tag{A.44}$$

$$\underset{n \to \infty}{\sim} r_0 \cdot n \cdot \sqrt{\pi} \left(1 - (1 - \frac{1}{2} \cdot \frac{2}{n}) \right) \tag{A.45}$$

$$\underset{n \to \infty}{\sim} r_0 \cdot \sqrt{\pi} \tag{A.46}$$

$$\underset{n \to \infty}{\sim} 1 \tag{A.47}$$

If we want to be sure that voxels are smaller than a given surface, we should choose the following formulation: $ceil(2 * \pi/m * [1 : 7])$. Reference surface is given by $\pi \cdot r_0^2$. The equivalence of a $1 \times 1 \times 1$ mm^3 voxel in Cartesian coordinates corresponds to a radius r_0 that is equal to $1/\sqrt{\pi} \simeq 0,564$ mm.

Radii do not depend on the number of sectors m. However, the number of crowns will be different, depending on the number of symmetries/sectors. In table 6.1, the number of voxels per crowns is given by: $round(2 * \pi/m * [1 : 7])$.

It is noticeable that all voxels have more or less equal surfaces. The point is that at infinity (but it works well far before), sides of any voxels are equal.

6.2 General remarks

It has to be noticed that we can have more projections than the number of invariants in the cylindrical image representation. Let say we want 60 projections and the object representation is made of 12 invariants. Transition matrix will be calculated for the first $60/12 = 5$

		crowns						
nb sectors	frq	1	2	3	4	5	6	7
m=3	$2\pi/3 \simeq 2$	2 (4.5%)	4 (4.5%)	6 (4.5%)	8 (4.5%)	10 (4.5%)	12 (3.4%)	14 (2.3%)
m=4	$2\pi/4$	2 (27%)	3 (4.5%)	5 (6.1%)	6 (4.5%)	8 (1.9%)	9 (4.5%)	11 (0.04%)
m=6	$2\pi/6 \simeq 1$	1 (4.5%)	2 (4.5%)	3 (4.5%)	4 (4.5%)	5 (4.5%)	6 (4.5%)	7 (4.5%)
m=8	$2\pi/8$	1 (27%)	2 (27%)	2 (15%)	3 (4.5%)	4 (1.9%)	5 (6.1%)	5 (9.1%)
m=12	$2\pi/12 \simeq 0.5$	1 (91%)	1 (4.5%)	2 (27%)	2 (4.5%)	3 (14.6%)	3 (4.5%)	4 (9.1%)
m=24	$2\pi/24 \simeq 0.25$	1 (280%)	1 (91%)	1 (27%)	1 (4.5%)	1 (23%)	2 (27%)	2 (9.1%)

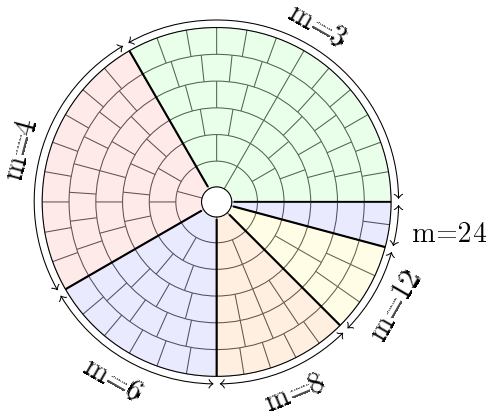


Figure A.5 – Voxel distribution, depending on the number of sectors m

projections, and the rest of the transition matrix will be deducted from this elementary matrix. This reduces by a factor of 12 the calculation time of the transition matrix, and by a factor of 12 the RAM requirement.

6.3 Discussion

Similar analysis could be done in spherical coordinates. The point is that the observations would be more informative because less correlated since they are taken from wider points of view than in classic tomographies where detectors acquire information around a unique axis.

Notations

$\boldsymbol{v}$	vector notations are bold, with lower case
$\boldsymbol{M}$	matrix notations are bold with upper case
$\boldsymbol{A}$	static transition matrix, alias system matrix
$\boldsymbol{y}$	measured data for one acquisition
$\boldsymbol{\lambda}$	spatial variables to be estimated through reconstruction algorithms
$\boldsymbol{X}$	Spatial components in the spatiotemporal reconstructions
	Every column of $\boldsymbol{X}$ corresponds to one component
$\boldsymbol{K}$	Kinetics components in the spatiotemporal reconstructions
	Every row of $\boldsymbol{K}$ corresponds to one component
$\boldsymbol{\eta}$	model of acquired data
$^{\mathrm{T}}$	the transposition operator
M	number of pixels in all detectors
N	number of voxels in reconstructed
T	number of time frames in acquisitions
m	subscript of one of the M pixels, $m \in [1, M]$
n	subscript of one of the N voxels, $t \in [1, N]$
t	subscript of one of the T time frames, $t \in [1, T]$
FOV	Field-of-View
m	Quantity of mass (g)
$\mathcal{F}$	Fourier transform, either 1D or 2D transformations depending on the sequences it is applied
$\mathcal{F}^{-1}$	Inverse Fourier transform
A	Mathematical representation of a coded aperture
G	Decoding matrix
$\otimes$	periodic correlation
$\star$	convolution operator
$*$	element-by-element multiplication
$\mathrm{C}_P(t)$	radiotracer concentration in plasma at time t
$k_{i,j}$	rate constant from compartment i to compartment j
CBV	cerebral blood volume
TAC	time activity curve
AUC	area under the curve
DV	distribution volume
BSA	blood surface ratio
MRGlu	metabolic rate of glucose
CMRGlu	cerebral metabolic rate of glucose
ICA	independent component analysis
NMF	non-negative matrix factorisation
MLEM	maximum-likelihood expectation-maximisation

Bibliography

[Abramovitz 70] M. Abramovitz & I.A. Stegun. Handbook of mathematical functions - with formulas, graphs, and mathematical tables. Dover publications, 1970. 75, 169

[Accorsi 01] R. Accorsi. *Design of Near-Field Coded Aperture Cameras for High-Resolution Medical and Industrial Gamma-Ray Imaging.* PhD thesis, Massachusetts Institute of Technology, 2001. 3, 32, 38, 62, 74, 166

[Aykac 01] M. Aykac, R.D. Hichwa & G.L. Watkins. *Investigation of a noninvasive detector system for quantitative[O-15]water blood flow studies in PET.* Nuclear Science, IEEE Transactions on, vol. 48, no. 1, 2001. 149

[Bailey 05] P.D.L. Bailey, D.W. Townsend, P.E. Valk*, et al.*, editeurs. Positron emission tomography: basic sciences, page 22. Springer, 2005. 23

[Barrett 72] H.H. Barrett. *Fresnel Zone Plate Imaging in Nuclear Medicine.* J Nucl Med, vol. 13, no. 6, 1972. 38

[Barrett 81] H.H. Barrett & W. Swindell. Radiological imaging: The theory of image formation, detection, and processing, 1981. 29

[Bazin 80] J.P. Bazin, R. Di Paola, B. Gibaud*, et al. Factor analysis of dynamic scintigraphic data as a modelling method. An application to the detection of the metastases.* Information Processing in Medical Imaging, 1980. 15

[Beck 85] R.N. Beck & L.D. Redtung. *Collimator Design Using Ray-Tracing Techniques.* Nuclear Science, IEEE Transactions on, vol. 32, no. 1, 1985. 77, 79

[Beenhouwer 08] J. De Beenhouwer, S. Staelens, S. Vandenberghe*, et al. Acceleration of GATE SPECT simulations.* Medical Physics, vol. 35, no. 4, 2008. 95

[Benali 93] H. Benali, I. Buvat, F. Frouin*, et al. A statistical model for the determination of the optimal metric in factor analysis of medical image sequences (FAMIS).* Physics in Medicine and Biology, vol. 38, no. 8, 1993. 48

[Bentourkia 05] M. Bentourkia. *Kinetic modeling of PET data without blood sampling.* Nuclear Science, IEEE Transactions on, vol. 52, no. 3, 2005. 16

[Berger 05] M.J. Berger, J.H. Hubbell, S.M. Seltzer, *et al. XCOM: Photon Cross Section Database (version 1.3)*, 2005. 57, 60, 130

[Berrim 96] S. Berrim, A. Lansiart & J.-L. Moretti. *Implementing of maximum likelihood in tomographical coded aperture.* Image Processing, 1996. Proceedings., International Conference on, vol. 1, 1996. 65

[Bertsekas 99] D.P. Bertsekas. Nonlinear programming. Athena Scientific, 1999. 53

[Birgin 00] E.G. Birgin, J.M. Martínez & M. Raydan. *Nonmonotone Spectral Projected Gradient Methods on Convex Sets.* SIAM J. on Optimization, vol. 10, no. 4, 2000. 54, 154

[Brogsitter 05] C. Brogsitter, T. Gruning, R. Weise, *et al. 18F-FDG PET for Detecting Myocardial Viability: Validation of 3D Data Acquisition.* J Nucl Med, vol. 46, no. 1, 2005. 9

[Bronzino 06] J.D. Bronzino. Tissue engineering and artificial organs, chapitre 10. CRC Press, 2006. 58

[Brown 72] C.M. Brown. Multiplex imaging and random arrays. Chicago, 1972., 1972. 28

[Buvat 93] I. Buvat, H. Benali, F. Frouin, *et al. Target apex-seeking in factor analysis of medical image sequences.* Physics in Medicine and Biology, vol. 38, no. 1, 1993. 49

[Cannon 79] T.M. Cannon & E.E. Fenimore. *Tomographical imaging using uniformly redundant arrays.* Applied Optics, vol. 18, no. 7, 1979. 64

[Caroli 87] E. Caroli, J.B. Stephen, G. di Cocco, *et al. Coded aperture imaging in X- and gamma-ray astronomy.* Space Science Reviews, vol. 45, 1987. 32

[Charbonnier 97] P. Charbonnier, L. Blanc-Feraud, G. Aubert, *et al. Deterministic edge-preserving regularization in computed imaging.* Image Processing, IEEE Transactions on, vol. 6, no. 2, 1997. 42

[Chen 07] K. Chen, X. Chen, R. Renaut, *et al. Characterization of the image-derived carotid artery input function using independent component analysis for the quantitation of [18F] fluorodeoxyglucose positron emission tomography images.* Physics in Medicine and Biology, vol. 52, no. 23, 2007. 3, 17

[Chu 04] M. Chu, F. Diele, R. Plemmons, *et al. Optimality, computation and interpretation of nonnegative matrix factorizations.* SIAM Journal on Matrix Analysis, 2004. 51

[Chuang 04] K.-S. Chuang, M.-L. Jan, J. Wu, *et al. The Thresholding MLEM Algorithm.* Journal of Medical and Biological Engineering, vol. 24, no. 2, 2004. 101

[Convert 07a] L. Convert, G. Morin-Brassard, J. Cadorette, *et al. A New Tool for Molecular Imaging: The Microvolumetric beta Blood Counter.* J Nucl Med, vol. 48, no. 7, 2007. 14

[Convert 07b] L. Convert, G. Morin-Brassard, J. Cadorette, *et al. A Microvolumetric β Blood Counter for Pharmacokinetic PET Studies in Small Animals.* Nuclear Science, IEEE Transactions on, vol. 54, no. 1, 2007. 14

[Cook 84] W.R. Cook, M. Finger, T.A. Prince, *et al. Gamma-ray imaging with a rotating hexagonal uniformly redundant array.* Nuclear Science, IEEE Transactions on, vol. 31, 1984. 35

[de Geus-Oei 06] L.-F. de Geus-Oei, E.P. Visser, P.F.M. Krabbe, *et al. Comparison of Image-Derived and Arterial Input Functions for Estimating the Rate of Glucose Metabolism in Therapy-Monitoring 18F-FDG PET Studies.* J Nucl Med, vol. 47, no. 6, 2006. 15

[Debasso 04] R. Debasso, H. Åstrand, N. Bjarnegård, *et al. The popliteal artery, an unusual muscular artery with wall properties similar to the aorta: implications for susceptibility to aneurysm formation?* Journal of Vascular Surgery, vol. 39, no. 4, 2004. 58

[Dempster 77] A.P. Dempster, N.M. Laird & D.B. Rubin. *Maximum Likelihood from Incomplete Data via the EM Algorithm.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, 1977. 39

[Di Paola 82] R. Di Paola, J.P. Bazin, F. Aubry, *et al. Handling of Dynamic Sequences in Nuclear Medicine.* Nuclear Science, IEEE Transactions on, 1982. 48

[Dicke 68] R.H. Dicke. *Scatter-Hole Cameras for X-Rays and Gamma Rays.* Astrophysical Journal, 1968. 25

[Eriksson 88] L. Eriksson, S. Holte, C. Bohm, *et al. Automated blood sampling systems for positron emission tomography.* Nuclear Science, IEEE Transactions on, 1988. 13

[Fenimore 78] E.E. Fenimore & T.M. Cannon. *Coded aperture imaging with uniformly redundant arrays.* Appl.Opt., vol. 17, no. 3, 1978. 28, 33

[Fenimore 83] E. E. Fenimore. *Large symmetric π transformations for Hadamard transforms.* Appl. Opt., vol. 22, no. 6, 1983. 35

[Ferrara 95] L.A. Ferrara, M. Mancini, R. Iannuzzi, *et al. Carotid Diameter and Blood Flow Velocities in Cerebral Circulation in Hypertensive Patients.* Stroke, vol. 26, no. 3, 1995. 58

[Finger 85] M.H. Finger & T.A. Prince. *Hexagonal uniformly redundant arrays for coded-aperture imaging.* In F. C. Jones, editeur, International Cosmic Ray Conference, volume 3 of *International Cosmic Ray Conference*, 1985. 35

[Gabrielsen 92] F.G. Gabrielsen, A. Bonnekoh, C. Felder, *et al. Determination of normal values of the aortic blood flow profile using continuous Doppler echocardiography from apical and suprasternal echo position.* Z Kardiol, vol. 81, no. 1, 1992. 59

[Gallezot 05] J.-D. Gallezot, M. Bottlaender, M.-C. Gregoire, *et al.* *In Vivo Imaging of Human Cerebral Nicotinic Acetylcholine Receptors with 2-18F-Fluoro-A-85380 and PET.* J Nucl Med, vol. 46, no. 2, 2005. 1, 2, 103

[Gmar 04] M. Gmar, O. Gal, C. Le Goaller, *et al.* *Development of coded-aperture imaging with a compact gamma camera.* Nuclear Science, IEEE Transactions on, vol. 51, no. 4, 2004. 35, 57, 59, 61

[Golub 96] G. H. Golub & C. F. Van Loan. Matrix computations. Johns Hopkins University Press, 3rd edition, 1996. 49

[Gottesman 86] S.R. Gottesman & E.J. Schneid. *PNP - A new class of coded aperture arrays.* Nuclear Science, IEEE Transactions on, vol. 33, 1986. 35

[Gottesman 89] S.R. Gottesman & E.E. Fenimore. *New family of binary arrays for coded aperture imaging.* Appl. Opt., vol. 28, no. 20, 1989. 36

[Gourlay 83] A.R. Gourlay & J.B. Stephen. *Geometric coded aperture masks.* Appl. Opt., vol. 22, no. 24, 1983. 37

[Graham 02] M.M. Graham, M. Muzi, A.M. Spence, *et al.* *The FDG Lumped Constant in Normal Human Brain.* J Nucl Med, vol. 43, no. 9, 2002. 9

[Green 90] P.J. Green. *Bayesian reconstructions from emission tomography data using amodified EM algorithm.* Medical Imaging, IEEE Transactions on, 1990. 41

[Grippo 86] L. Grippo, F. Lampariello & S. Lucidi. *A Nonmonotone Line Search Technique for Newton's Method.* SIAM Journal on Numerical Analysis, vol. 23, no. 4, 1986. 54

[Groiselle 00] C. Groiselle. *Tomographie par projections orthogonales et ouverture codée.* PhD thesis, Université Paris-Nord -Villetaneuse, 2000. 25, 33, 65

[Gunson 76] J. Gunson & B. Polychronopulos. *Optimum design of a coded mask X-ray telescope for rocket applications.* Monthly Notices of the Royal Astronomical Society, vol. 177, 1976. 33

[Hall 71] R. Hall. *Vascular injuries resulting from arterial puncture or catheterization.* British Journal of Surgery, 1971. 12

[Hudson 94] H.M. Hudson & R.S. Larkin. *Accelerated image reconstruction using ordered subsets of projection data.* Medical Imaging, IEEE Transactions on, 1994. 101

[Hunter 96] G.J. Hunter, L.M. Hamberg, N.M. Alpert, *et al.* *Simplified Measurement of Deoxyglucose Utilization Rate.* J Nucl Med, vol. 37, no. 6, 1996. 19

[Hyvärinen 01] A. Hyvärinen, J. Karhunen & E. Oja. Independent component analysis. John Wiley & Sons, 2001. 15

[Idier 01] J. Idier & L. Blanc-Féraud. *Déconvolution en imagerie*. In J. Idier, editeur, Approche bayésienne pour les problèmes inverses. Hermes, 2001. 41

[Jain 03] M. Jain & S.K. Batra. *Genetically Engineered Antibody Fragments and PET Imaging: A New Era of Radioimmunodiagnosis.* J Nucl Med, vol. 44, no. 12, 2003. 5

[Jan 04] S. Jan, G. Santin, D. Strul*, et al. GATE: a simulation toolkit for PET and SPECT.* Physics in Medicine and Biology, vol. 49, no. 19, 2004. 118, 142

[Johnsen 66] E.C. Johnsen. *Skew-Hadamard Abelian Group Difference Sets.* Journal of Algebra, 1966. 33

[Kadrmas 01] D.J. Kadrmas & G.T. Gullberg. *4D maximum a posteriori reconstruction in dynamic SPECT using a compartmental model-based prior.* Physics in Medicine and Biology, vol. 46, no. 5, 2001. 45

[Kamasak 05] M.E. Kamasak, C.A. Bouman, E.D. Morris*, et al. Direct Reconstruction of Kinetic Parameter Images From Dynamic PET Data.* Medical Imaging, IEEE Transactions on, vol. 24, no. 5, 2005. 46

[Kriplani 06] A. Kriplani, D.J. Schlyer, P. Vaska*, et al. Non-invasive and selective measurement of the arterial input function using a PET Wrist Scanner.* Nuclear Science, Symposium Conference Record, 2006. IEEE, vol. 6, no. 3, 2006. 2, 18, 22

[Kriplani 07] A. Kriplani, D. Schlyer, P. Vaska*, et al. Measurement of the arterial input function using a non-invasive wrist scanner.* J NUCL MED MEETING ABSTRACTS, vol. 48, 2007. 18

[Laforest 05] R. Laforest, T.L. Sharp, J.A. Engelbach*, et al. Measurement of input functions in rodents: challenges and solutions.* Nuclear Medicine and Biology, vol. 32, no. 7, 2005. 16

[Lapointe 98] D. Lapointe, J. Cadorette, S. Rodrigue*, et al. A microvolumetric blood counter/sampler for metabolic PET studies in small animals.* Nuclear Science, IEEE Transactions on, vol. 45, no. 4, 1998. 13

[Lee 99] D.D. Lee & H.S. Seung. *Learning the parts of objects by non-negative matrix factorization.* Nature, vol. 401, no. 6755, 1999. 48

[Lee 00] D.D. Lee & H.S. Seung. *Algorithms for Non-negative Matrix Factorization.* In NIPS, 2000. 49, 50

[Lee 01] J.-S. Lee, D.D. Lee, S. Choi*, et al. Application of nonnegative matrix factorization to dynamic positron emission tomography.* In In 3rd International Conference on Independent Component Analysis and Blind Signal Separation, 2001. 48

[Leo 94] W.R. Leo. Techniques for nuclear and particle physics experiments. Springer, 1994. 24, 129

[Lin 07] C.-J. Lin. *Projected Gradient Methods for Nonnegative Matrix Factorization.* Neural Comput., vol. 19, no. 10, 2007. 51

[Litton 90] J.E. Litton & L. Eriksson. *Transcutaneous measurement of the arterial input function inpositron emission tomography.* Nuclear Science, IEEE Transactions on, vol. 37, no. 2, 1990. 17, 57

[Litton 97] J.E. Litton. *Input function in PET brain studies using MR-defined arteries.* Journal of computer assisted tomography, vol. 21, no. 6, 1997. 16

[Logan 90] J. Logan, J.S. Fowler, N.D. Volkow*, et al. Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects.* J Cereb Blood Flow Metab, vol. 1, 1990. 10

[Logan 00] J. Logan. *Graphical analysis of PET data applied to reversible and irreversible tracers.* Nucl Med Biol, vol. 27, 2000. 11

[Logan 03] J. Logan. *A review of graphical methods for tracer studies and strategies to reduce bias. Nucl Med Biol.* Nucl Med Biol, 2003. 10, 11, 12

[Luke 88] H.D. Luke. *Sequences and arrays with perfect periodic correlation.* Aerospace and Electronic Systems, IEEE Transactions on, vol. 24, no. 3, 1988. 35

[Martineau 09] A. Martineau. *Optimisation d'un collimateur à ouverture codée avec la plate-forme de simulation Monte-Carlo GATE.* PhD thesis, Université Paris 7, 2009. 67

[Mayersohn 70] M. Mayersohn & M Gibaldi. *Mathematical Methods in Pharmacokinetics. I. Use of the Laplace Transform in Solving Differential Rate Equations.* Amer. J. Pharm. Ed., vol. 34, 1970. 7

[Meikle 01] S.R. Meikle, P. Kench, A.G. Weisenberger*, et al. A prototype coded aperture detector for small animal SPECT.* Nuclear Science Symposium Conference Record, 2001 IEEE, vol. 3, 2001. 38

[Menard 98] L. Menard, Y. Charon, M. Solal*, et al. POCI: a compact high resolution γ camera for intra-operative surgical use.* Nuclear Science, IEEE Transactions on, vol. 45, no. 3, 1998. 162

[Mertz 61] L. Mertz & N. O. Young. *Fresnel transformations of images.* In Proceedings of the International Conference on Optical Instruments and Techniques, 1961. 28

[Mohammad-Djafari 06] A. Mohammad-Djafari & L. Robillard. *Hierarchical Markovian Models for 3D Computed Tomography in Non Destructive Testing Applications.* In EUSIPCO 2006. EUSIPCO 2006, September 4-8, Florence, Italy, 2006. 138

[Moussaoui 05] S. Moussaoui. *Séparation de sources non-négatives – Application au traitement des signaux de spectroscopie.* PhD thesis, université H. Poincaré, 2005. 51

[Muzic 01] Jr. Muzic, R.F. & S. Cornelius. *COMKAT: Compartment Model Kinetic Analysis Tool.* J Nucl Med, vol. 42, no. 4, 2001. 1, 8

[Naganawa 05] M. Naganawa, Y. Kimura, K. Ishii, *et al. Extraction of a plasma time-activity curve from dynamic brain PET images based on independent component analysis.* Biomedical Engineering, IEEE Transactions on, vol. 52, no. 2, 2005. 2, 16

[Naidich 07] D.P. Naidich, W.R. Webb, Nestor L.M., *et al.*, editeurs. Computed tomography and magnetic resonance of the thorax, chapitre 2. Lippincott Williams & Wilkins, 4 edition, 2007. 58

[Nelson 93] A.D. Nelson, F. Miraldi, Jr. Muzic R.F., *et al. Noninvasive Arterial Monitor for Quantitative Oxygen-15-Water Blood Flow Studies.* J Nucl Med, vol. 34, no. 6, 1993. 18

[Nichols 02] T. E. Nichols, J. Qi, E. Asma, *et al. Spatiotemporal recinstruction of list-mode PET data.* Medical Imaging, IEEE Transactions on, vol. 21, no. 4, 2002. 43

[Ohyama 84] N. Ohyama, T. Endo, T. Honda, *et al. Coded-aperture imaging system for reconstructing tomograms of human myocardium.* Applied Optics, vol. 23, no. 18, 1984. 25, 38

[Ou 08] W. Ou, P. Golland & M. Hämäläinen. *A Distributed Spatio-temporal EEG/MEG Inverse Solver.* In MICCAI (1), 2008. 158

[Pain 04] F. Pain, P. Laniece, R. Mastrippolito, *et al. Arterial Input Function Measurement Without Blood Sampling Using a beta-Microprobe in Rats.* J Nucl Med, vol. 45, no. 9, 2004. 14

[Patlak 83] C.S. Patlak, R.G. Blasberg & J.D. Fenstermacher. *Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data.* J Cereb Blood Flow Metab, vol. 3, no. 1, 1983. 1, 10, 14

[Patlak 85] C.S. Patlak & R.G. Blasberg. *Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. Generalizations.* J Cereb Blood Flow Metab, vol. 5, 1985. 11

[Phelps 79] M.E. Phelps, S.C. Huang, E.J. Hoffman, *et al. Tomographic measurement of local cerebral glucose metabolic rate in humans with [F-18]2-fluoro-2-deoxy-D-glucose: validation of method.* Annals of Neurology, 1979. 1, 9, 14

[Phillips 95] R.L. Phillips, C.Y. Chen, D.F. Wong, *et al. An improved method to calculate cerebral metabolic rates of glucose using PET.* J Nucl Med, 1995. 1, 18

[Radegran 00] G. Radegran & B. Saltin. *Human femoral artery diameter in relation to knee extensor muscle mass, peak blood flow, and oxygen uptake.* Am J Physiol Heart Circ Physiol, vol. 278, no. 1, 2000. 58

[Rajeswaran 92] S. Rajeswaran, D.L. Bailey, S.P. Hume, *et al. 2D and 3D imaging of small animals and the human radial artery witha high resolution detector for PET.* Medical Imaging, IEEE Transactions on, vol. 11, no. 3, 1992. 18

[Reader 06] A.J. Reader, F.C. Sureau, C. Comtat, *et al. Joint estimation of dynamic PET images and temporal basis functions using fully 4D ML-EM.* Physics in Medicine and Biology, vol. 51, 2006. 17, 44

[Rogers 80] W.L. Rogers, K.F. Koral, R. Mayans, *et al. Coded-Aperture Imaging of the Heart.* J Nucl Med, vol. 21, no. 4, 1980. 38

[Russo-Marie 07] F. Russo-Marie, A. Samson & S. Teysseyre. *Improved apoptosis imaging in tumor bearing mice by* ^{99m}Tc-*AFIM planar scintigraphy*, 2007. 60, 123

[Schellingerhout 02] D. Schellingerhout, R. Accorsi, U. Mahmood, *et al. Coded aperture nuclear scintigraphy: a novel small animal imaging technique.* Mol Imaging., 2002. 25, 38

[Schmidt 02] K.C. Schmidt & F.E. Turkheimer. *Kinetic modeling in positron emission tomography.* Q J Nucl Med, vol. 46, 2002. 8

[Shepp 82] L. A. Shepp & Y. Vardi. *Maximum Likelihood Reconstruction in Positron Emission Tomography.* Medical Imaging, IEEE Transactions on, vol. 1, no. 2, 1982. 39, 50

[Shiozaki 00] T. Shiozaki, N. Sadato, M. Senda, *et al. Noninvasive Estimation of FDG Input Function for Quantification of Cerebral Metabolic Rate of Glucose: Optimization and Multicenter Evaluation.* J Nucl Med, vol. 41, no. 10, 2000. 19

[Shokouhi 03] S. Shokouhi, S.P. Stoll, A. Villanueva, *et al. A noninvasive LSO-APD blood radioactivity monitor for PET imaging studies.* Nuclear Science, IEEE Transactions on, vol. 50, no. 5, 2003. 18, 57

[Sokoloff 77] M. Sokoloff, M. Reivich, C. Kennedy, *et al. The [14C] deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat.* Journal of Neurochemistry, vol. 28, no. 5, 1977. 9

[Sundaram 04] S.K. Sundaram, N.M.T. Freedman, J.A. Carrasquillo, *et al. Simplified Kinetic Analysis of Tumor 18F-FDG Uptake: A Dynamic Approach.* J Nucl Med, vol. 45, no. 8, 2004. 20

[Tabary 07] J. Tabary, P. Hugonnard & F. Mathy. *SINDBAD : a realistic multi-purpose and scalable X-ray simulation tool for NDT applications.* In International Symposium on Digital industrial Radiology and Computed Tomography, 2007. 77

[van der Weerdt 01] A.P. van der Weerdt, L.J. Klein, R. Boellaard, *et al. Image-Derived Input Functions for Determination of MRGlu in Cardiac 18F-FDG PET Scans.* J Nucl Med, vol. 42, no. 11, 2001. 1, 15

[Veklerov 87] E. Veklerov & J. Llacer. *Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing.* Medical Imaging, IEEE Transactions on, vol. 6, no. 4, 1987. 137

[Vicini 00] P. Vicini, H.-T. Su & J.J. DiStefano III. *Identifiability and interval identifiability of mammillary and catenary compartmental models with some known rate constants.* Mathematical Biosciences, vol. 167, no. 2, 2000. 7

[Villanueva 03] Jr. Villanueva, A., S.P. Stoll, D.J. Schlyer, *et al. Spatial resolution of a noninvasive measurement of the arterial and venous input function using a wrist monitor.* Nuclear Science Symposium Conference Record, 2003 IEEE, vol. 4, 2003. 18

[Vriend 06] J.W.J. Vriend, E. de Groot, T.T. de Waal, *et al. Increased carotid and femoral intima-media thickness in patients after repair of aortic coarctation: Influence of early repair.* American Heart Journal, vol. 151, no. 1, 2006. 63

[Wakita 00] K. Wakita, Y. Imahori, T. Ido, *et al. Simplification for Measuring Input Function of FDG PET: Investigation of 1-Point Blood Sampling Method.* J Nucl Med, vol. 41, no. 9, 2000. 1, 19

[Watabe 95] H. Watabe, M. Miyake, Y. Narita, *et al. Development of skin surface radiation detector system to monitorradioactivity in arterial blood along with positron emission tomography.* Nuclear Science, IEEE Transactions on, vol. 42, no. 4, 1995. 18

[Watabe 06] H. Watabe, Y. Ikoma, Y. Kimura, *et al. PET kinetic analysis-compartmental model.* Ann Nucl Med, 2006. 6

[Webb 03] S. Webb. The physics of medical imaging, chapitre 6. Medical Science Series. Institute of Physics Publishing, 2003. 23

[Welling 86] P.G. Welling. Pharmacokinetics processes and mathematics. American Chemical Society, 1986. 5, 6, 7

[Wild 83] W.J. Wild. *Dilute uniformly redundant sequences for use in coded-aperture imaging.* Opt. Lett., vol. 8, no. 5, 1983. 37

[Wu 95] H.-M. Wu, C.K. Hoh, Y. Choi, *et al. Factor Analysis for Extraction of Blood Time-Activity Curves in Dynamic FDG-PET Studies.* J Nucl Med, vol. 36, no. 9, 1995. 16

[Wu 96] H.-M. Wu, S.-C. Huang, V. Allada, *et al. Derivation of Input Function from FDG-PET Studies in Small Hearts.* J Nucl Med, vol. 37, no. 10, 1996. 16

[Wu 07] H.-M. Wu, G. Sui, C.-C. Lee, *et al. In Vivo Quantitation of Glucose Metabolism in Mice Using Small-Animal PET and a Microfluidic Device.* J Nucl Med, vol. 48, no. 5, 2007. 14

[Zaidi 03] H. Zaidi, editeur. Quantitative analysis in nuclear medicine imaging. Springer, 2003. 5

[Zeng 93] G.L. Zeng & G.T. Gullberg. *A Ray-driven Backprojector For Backprojection Filtering And Filtered Backprojection Algorithms.* Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record., 1993. 77

[Zhuang 94] W. Zhuang, S.S. Gopal & T.J. Hebert. *Numerical evaluation of methods for computing tomographic projections.* Nuclear Science, IEEE Transactions on, vol. 41, no. 4, 1994. 77

Résumé Cette thèse traite de l'estimation de la concentration dans le sang artériel de molécules marquées par un radioélément émettant des positons. Cette concentration est appelée "fonction d'entrée artérielle β^+". Elle doit être déterminée dans de nombreuses analyses en pharmacocinétique. Actuellement, elle est mesurée à l'aide d'une série de prélèvements artériels, méthode précise mais nécessitant un protocole contraignant. Des complications liées au caractère invasif de la méthode peuvent survenir (hématomes, infections nosocomiales).

L'objectif de cette thèse est de s'affranchir de ses prélèvements artériels par l'estimation non-invasive de la fonction d'entrée β^+ à l'aide d'un détecteur externe et d'un collimateur. Cela permet la reconstruction des vaisseaux sanguins afin de discriminer le signal artériel du signal contenu dans les autres tissus avoisinants. Les collimateurs utilisés en imagerie médicale ne sont pas adaptés à l'estimation de la fonction d'entrée artérielle β^+ car leur sensibilité est très faible. Pour cette thèse, ils sont remplacés par des collimateurs codés, issus de la recherche en astronomie. De nouvelles méthodes pour utiliser des collimateurs à ouverture codée avec des algorithmes statistiques de reconstruction sont présentées.

Des techniques de lancer de rayons et une méthode d'accélération de la convergence des reconstructions sont proposées. Une méthode de décomposition spatio-temporelle est également mise au point pour estimer efficacement la fonction d'entrée artérielle à partir d'une série d'acquisitions temporelles.

Cette thèse montre qu'il est possible d'améliorer le compromis entre sensibilité et résolution spatiale en tomographie d'émission à l'aide de masques codés et d'algorithmes statistiques de reconstruction ; elle fournit également les outils nécessaires à la réalisation de telles reconstructions.

Abstract This work deals with the estimation of the concentration of molecules in arterial blood which are labelled with positron-emitting radioelements. This concentration is called "β^+ arterial input function". This concentration has to be estimated for a large number of pharmacokinetic analyses. Nowadays it is measured through series of arterial sampling, which is an accurate method but requiring a stringent protocol. Complications might occur during arterial blood sampling because this method is invasive (hematomes, nosocomial infections).

The objective of this work is to overcome this risk through a non-invasive estimation of β^+ input function with an external detector and a collimator. This allows the reconstruction of blood vessels and thus the discrimination of arterial signal from signals in other tissues. Collimators in medical imaging are not adapted to estimate β^+ input function because their sensitivity is very low. During this work, they are replaced by coded-aperture collimators, originally developed for astronomy.

New methods where coded apertures are used with statistical reconstruction algorithms are presented. Techniques for analytical ray-tracing and for the acceleration of reconstructions are proposed. A new method which decomposes reconstructions on temporal sets and on spatial sets is also developped to efficiently estimate arterial input function from series of temporal acquisitions.

This work demonstrates that the trade-off between sensitivity and spatial resolution in PET can be improved thanks to coded aperture collimators and statistical reconstruction algorithm; it also provides new tools to implement such improvements.