UNIVERSIDAD AUTONOMA
DE MADRID

Escuela Politécnica Superior
Departamento de Ingeniería Informática

HIGH RESOLUTION THREE-DIMENSIONAL RECONSTRUCTION FOR SMALL ANIMAL PET CAMERAS

PHD DISSERTATION

Ana Iriarte ruiz

Madrid, Junio 2012

A Carlos Óscar, Roberto, Robert y Samuel

Contents

Motivación y Objetivos

Antecedentes del grupo

Ésta tesis ha sido concebida en un grupo de investigación con una amplia experiencia en la aplicación de metodologías computacionales y experimentales a la solución de problemas claves en biología (La Unidad de Biocomputación (BCU) del Centro Nacional de Biotecnología CNB): http://biocomp.cnb.csic.es/). Para un grupo de estas características, el potencial de la tomografía por emisión de positrones (PET) en la biología moderna (Cherry and Gambhir (2001)) y la necesidad de herramientas para la gestión de datos PET (Cherry and Chatziioannou (2004)) no pasaron desapercibidos. Por estos motivos, PET fue una de las fuentes de datos que se consideraron en la propuesta del proyecto "De la información al conocimiento: La gestión y análisis de grandes conjuntos de datos de objetos complejos en biología estructural y ñgenómica funcional". Tres grupos de investigación distintos, con un un núcleo común en el campo del desarrollo de nuevos métodos para la gestión y el análisis de grandes conjuntos de datos de objetos biológicos complejos, participaban en el proyecto. El grupo de la Universidad Autónoma de Madrid (http://www.ii.uam.es/esp/investigacion/index.php?siglas=GTSB, vinculado a la BCU) estaba a cargo del subproyecto titulado "Hacia la Proteómica Visual: herramientas para el cálculo y gestión de estructuras tridimensionales". El objetivo de este subproyecto era

el desarrollo de nuevos modelos de datos y entornos gráficos que integraran imagen funcional PET con datos de genómica y proteómica, con el fin de situar el conocimiento de los mecanismos moleculares en un contexto fisiológico.

El grupo nunca había trabajado con PET y la familiarización con una nueva y compleja técnica se imponía como tarea antes de empezar el proyecto. A cambio de la inexperiencia en el campo de PET, se poseía una experiencia importante en reconstrucción de datos procedentes de otras técnicas (Marabini et al. (1997), Marabini et al. (1998), Sorzano et al. (2001), Marabini et al. (2004), Sorzano et al. (2008)) que podía ser trasladado al entorno de PET. Por ello, el grupo de la UAM dirigió un proyecto de reconstrucción para PET de animales pequeõs que coordinaba el trabajo de seis laboratorios (FIS PI040683). Este proyecto proponía la adaptación de varios algoritmos de reconstrucciń tridimensional a las peculiaridades de las cámaras PET para animales pequeños, y la realización de una comparacin objetiva de su rendimiento en condiciones realistas. El proyecto tenía, para el grupo de la UAM, un doble propósito. Por un lado, desde el punto de vista de contribución al campo, la aplicación de los conocimientos que se poseía en reconstrución de imagen al desarrollo de algoritmos de reconstrucción para PET. Es sabido que los algoritmos de reconstrución tienen un papel crucial en el futuro de PET para animales pequeños (Lewitt and Matej (2003)), y por ello cualquier contribución en este aspecto está bien justificada. Por otro lado, desde el punto de vista del aprendizaje, el proyecto proporcionaba un excelente primer contacto con los datos y el entorno PET, antes de centrarse en las aplicaciones biológicas de la técnica. Este trabajo se ha desarrollado dentro del marco de este segundo proyecto. Esta es por lo tanto una tesis sobre reconstrucción de imágenes PET, que tiene la función subyacente de familiarizar a un grupo con esta tecnología, que constituye una herramienta que encaja perfectamente con su linea de investigación de técnicas aplicadas a la extracción de conocimiento biológico.

Estado del arte y objetivos

Una vez centrados los objetivos en la reconstrucción de imágenes PET, se encontró en la literatura una clara tendencia hacia los algoritmos conocidos como "iterativos" (Qi and Leahy (2006)), y por lo tanto este trabajo se ha orientado hacia este tipo de reconstrucción. Es bien sabido que lo que hace que los algoritmos iterativos obtengan imágenes con mejores propiedades de resolución y ruido que otros métodos, es el modelo estadístico de los datos y de su proceso de adquisición en el que se basan. Con respecto al modelo de datos, la distribución de Poisson es un modelo ampliamente aceptado para el proceso de emisión de positrones, que es la fuente de los datos PET. La familia más popular de algoritmos de reconstrucción iterativos (basados en el algoritmo Maximum Likelihood Expectation Maximization MLEM (Vardi et al. (1985)) se basa de hecho en un modelo de Poisson de la emisión de positrones. Por esta razón, el algoritmo MLEM es el esquema iterativo en el que se basan las reconstrucciones de esta tesis. Con respecto al modelo del proceso de adquisición de datos, en la familia de algoritmos MLEM se compone de un conjunto de términos estadísticos organizados en una estructura matricial conocida como la "matriz del sistema" o la "matriz de respuesta del sistema". Los términos de la matriz del sistema resultan de la combinación de una gran variedad de factores muy complejos, muchos de los cuales pueden depender de las características particulares de cada escáner o de las condiciones en las que se realice el experimento. Además de suponer un problema matemático complicado, el cálculo de la Matriz del Sistema requiere grandes recursos computacionales, tanto en términos de tiempo como de almacenamiento. (Herraiz et al. (2006)).

El cálculo de los términos de la matriz del sistema es, por lo tanto, una tarea compleja que al mismo tiempo constituye un punto crucial en la obtención de reconstrucciones de alta calidad. Esta afirmación está en concordancia con el enorme número de trabajos sobre este campo que se publican continuamente (Veklerov et al. (1988), Mumcuoglu et al.

(1996), Qi et al. (1998), Kudrolli et al. (2002), Rafecas et al. (2004), Alessio et al. (2006), Herraiz et al. (2006), Panin et al. (2006a),Panin et al. (2006b), Scheins et al. (2006), Moehrs et al. (2008), Tohme and Qi (2009), Ortuño et al. (2010), Aguiar et al. (2010), Pratx and Levin (2011), Herraiz et al. (2011)), todos ellos centrados en el cálculo de la matriz del sistema de una manera precisa y eficiente. Sin embargo, de entre todos estos trabajos, no se ha encontrado ninguno que reúna y tenga en cuenta las peculiaridades de la tecnología conocida como de "detectores continuos". Siendo el modelo del sistema un aspecto clave en la reconstrucción estadística, la ausencia de modelos especficos implica que las reconstrucciones obtenidas hasta ahora con datos de estos detectores no aprovechan al máximo los beneficios de trabajar con esta tecnología. Esta observación es importante, en particular para el desarrollo del proyecto FIS, ya que uno de los grupos participantes poseía datos provenientes de una de estas cámaras (Balcerzyk et al. (2009), Sánchez et al. (2012)) y para la comunidad PET en general, donde un interés renovado por este tipo de cámaras ha emergido en los últimos años Hun et al. (2002), (Tavernier et al. (2005)).

El primer problema que surge cuando las implementaciones de la matriz del sistema existentes quieren ser usadas con la tecnología de detectores continuos es que no son capaces de retener toda la precisión de los datos que este tipo de detectores proporciona. Las implementaciones analíticas que serían capaces de retener esta precisión (Vardi et al. (1985)), han sido tradicionalmente descartadas en beneficio de otras metodologías experimentales (mediciones en las cámaras, simulaciones Monte Carlo) (Rafecas et al. (2004), Panin et al. (2006a), Alessio et al. (2006), Ortuño et al. (2010)). Además se ser especialmente adecuadas para la reconstrucción de datos de detectores continuos, las aproximaciones analíticas al cálculo de la matriz del sistema tienen otros beneficios (ausencia de ruido, precisión, flexibilidad...) para la reconstruccin PET en general. La razón por la cual estas ventajas no han sido totalmente explotadas aún es que las metodologías analíticas dan lugar a implementaciones muy costosas en cuanto a tiempo de computación, por lo que

sólo las aproximaciones analíticas basadas en modelos simplificados del proceso PET son viables. Sin embargo, ahora que los recientes avances tecnológicos (paralelización de procesos, uso de unidades de procesamiento gráfico (GPUs)) proporcionan recursos computacionales que permiten implementaciones cada vez más complejas de la matrix del sistema (Pratx and Levin (2011), Herraiz et al. (2011)), se abre una puerta al estudio de las ventajas que una implementación analítica adecuada de la matriz del sistema puede tener.

El estudio del estado del arte ha conducido, por lo tanto, a dos problemas abiertos en el campo de la reconstrucción para PET: la falta de métodos de recostrucción adaptados a los requisitos de la tecnología de detectores continuos, y, en un plano más general, la ausencia de metodologías analíticas elaboradas para el cálculo de la matriz del sistema. Estos han sido los dos aspectos principales en los que se ha centrado esta tesis.

Estructura del documento

De acuerdo a la discusión previa, y después de una introduccin general a la técnica PET proporcionada en el Capítulo 1, esta tesis se ha construido en torno a los capítulos 2 and 3, cuyos contenidos se describen brevemente a continuación:

1. El Capítulo 2 aborda la aplicación de metodologías analíticas al cálculo de la matriz del sistema. Como ya se ha mencionado, este tipo de metologías no ha sido suficientemente explotado hasta ahora, debido a su elevado coste computacional. Ahora que la tecnología está permitendo acelerar cada vez más las reconstrucciones, es el momento de comprobar las mejoras que estas metologías pueden conseguir. Sin embargo, aunque es de esperar que las aproximaciones analíticas tengan un gran número de ventajas, ha de tenerse en cuenta que la matriz del sistema es un modelo estadístico de un proceso muy complejo, en el cual intervienen factores de naturalezas muy distintas. Algunos de estos factores estarán indudablemente bien modelados

con una aproximación analítica, pero pueden existir otros efectos para los cuales un tratamiento analítico puede no ser tan apropiado. El objetivo del capítulo no es, por lo tanto desarrollar una implementación púramente analítica de la matriz del sistema, sino explorar como las ventajas de estas aproximaciones, que ha sido tradicionamente descartadas debido a su ineficiencia, pueden ser explotadas. En el Capítulo 2, se ha realizado un estudio exhaustivo del proceso de adquisición de datos PET y de como las diferentes metodologías son apropiadas para modelar este proceso en la matriz del sistema. El resultado de este estudio es una nueva aproximación para el cálculo de la matriz del sistema basada en la sinergia entre la metodologías analíticas y otras metodologías experimentales. Las principales ventajas y limitaciones de cada uno de los trabajos previos en este campo se han analizado de manera que justifican las elecciones hechas en el diseño de la nueva matriz del sistema. Además, las implicaciones que cada una de estas decisiones tiene en el dominio de la imagen se ha estudiado para evaluar las ventajas que el nuevo esquema aporta a las reconstrucciones.

2. El Capítulo 3 aborda la reconstrucción de datos de detectores continuos. La razón subyacente que ha conducido al desarrollo de la nueva metodología presentada en el Capítulo 2 es la ausencia de aproximaciones específicas que son capaces de retener la precisión de los datos de la tecnología de detectores continuos. Las reconstrucciones de datos de este tipo de cámaras se realiza actualmente con matrices del sistema que no tienen en cuenta ni esta ni otras particularidades de esta técnica. Por lo tanto, las reconstrucciones obtenidas hasta ahora no aprovechan al máximo los beneficios de trabajar con la tecnología de detectores continuos. Una vez que la nueva aproximación para el cálculo de la matriz del sistema se ha desarrollado en el Capítulo 2, se mostrará en el Captulo 3 que no sólo puede utilizarse para retener la precisión de los datos de detectores continuos, sino para tener en cuenta otras peculiaridades de

este tipo de dispositivos en el modelo estadístico. Además de la nueva metodología para calcular la matriz del sistema, otros aspectos de la reconstrucción (estructura de los datos, cálculo de la sensibilidad) se han adaptado al uso con detectores continuos, dando lugar a un esquema de reconstrucción totalmente adaptado a este tipo de detectores, cuyas ventajas en el dominio de la imagen han sido probadas.

Motivation and Objectives

Group background

This thesis has been conceived within a research group with large experience in the application of computational and experimental methodologies to the solution of key problems in biology (The Biocomputing Unit (BCU) of the National Center for Biotechnology (CNB): http://biocomp.cnb.csic.es/). For a group with these characteristics, the potential of positron emission tomography (PET) in modern biology (Cherry and Gambhir (2001)) and the need of tools for the management of PET data (Cherry and Chatziioannou (2004)) did not go unnoticed. For those reasons, PET was one of the sources of data considered in the proposal of the project "From information to knowledge: The management and analysis of large data set of complex objects in structural biology and functional genomics". Three research different groups with a common core in the topic of the development of new methods for the management and analysis of large data sets of complex biological objects, participated in the project. The Universidad Autonoma group (http://www.ii.uam.es/esp/investigacion/index.php?siglas=GTSB, attached to the BCU) was in charge of the subproject entitled "Toward Visual Proteomics: tools for 3D-Map Calculation and Management(BIO2007-67150-C03-03)". The aim of this subproject was the development of new data models and graphics environments that take into account PET

functional image data on small animals together with genomics and proteomics data, in order to place the knowledge on molecular mechanism in a physiological context. The group had never worked with PET and the important challenge of familiarization with a fairly complex technique had to be addressed before starting with the project. In exchange for the inexperience in the PET field, an important background in reconstruction of data from other image techniques existed (Marabini et al. (1997), Marabini et al. (1998), Sorzano et al. (2001), Marabini et al. (2004), Sorzano et al. (2008)) that was able to be transferred to the PET environment. For these reasons, the UAM group coordinated a Network-related project of six laboratories on PET reconstruction for small animals (FIS PI040683). This project proposed to adapt several reconstruction algorithms to the singularities of small animal scanners and to perform an objective quantitative comparison of their performance under realistic conditions. The project had, for the UAM group, a double purpose. First, from the point of view of contribution to the field, the application of the knowledge owned on image reconstruction to the development of PET reconstruction algorithms. The reconstruction algorithms are known to have an important role in the success of the small animal PET (Lewitt and Matej (2003)), and any chance of improvement in the reconstruction field is well justified. Second, from the point of view of learning, the project provided an excellent first contact with the PET data and environment, before focusing on the biological applications of the technique. The present work has been developed in the framework of the second mentioned project. This is thus a thesis on PET image reconstruction that has the underlying function of familiarize a group with PET; a tool that fits in perfectly with its line of research in techniques applied to the extraction of biological knowledge.

State of the art and objectives

Once focused on PET reconstruction, a clear trend towards the algorithms known as "iterative" was found in the literature (Qi and Leahy (2006)) and this work has been consequently devoted to this type of reconstruction. It is well know that what makes iterative reconstruction to obtain images with better resolution and noise properties than other approaches is the accurate statistical model of the data formation and acquisition process they are based in. Concerning the data model, the Poisson distribution is a widespread accepted model for the positron emission process, which is the source of data in PET. The most popular family of iterative reconstruction algorithms (based on the Maximum Likelihood Expectation Maximization MLEM (Vardi et al. (1985)) algorithm) is indeed based on a Poisson model of the positron emissions. For that reason the MLEM has been chosen as the iterative scheme in which the reconstructions of this thesis are based in. Regarding the acquisition process model, in the MLEM algorithms family, it is composed of a collection of statistical terms, arranged in a matrix structure known as the "system matrix" or the "system response matrix". The system matrix terms result from the combination of a great variety of complex factors, many of which can be dependent on the particular scanner features and/or on the experiment conditions. Apart from being a complicated mathematical problem, the system matrix calculation requires high computational resources both in terms of time and storage (Herraiz et al. (2006)).

The computation of the system matrix terms is thus a challenging task that at the same time constitutes a crucial step in the achievement of high quality reconstructions. This assertion is in agreement with the large number of works that are continuously published on the topic (Veklerov et al. (1988), Mumcuoglu et al. (1996), Qi et al. (1998), Kudrolli et al. (2002), Rafecas et al. (2004), Alessio et al. (2006), Herraiz et al. (2006), Panin et al. (2006a),Panin et al. (2006b), Scheins et al. (2006), Moehrs et al. (2008), Tohme and Qi

(2009), Ortuño et al. (2010), Aguiar et al. (2010), Pratx and Levin (2011), Herraiz et al. (2011)), all of them focused on the computation of the system matrix in an accurate and efficient way. However, among all these works, no one has been found that gathers and takes into account all the peculiarities of the so-called ”continuous detectors“ technology (Karp et al. (1990)). Being the system modeling a key issue in statistical reconstructions, the lack of specific models implies that the reconstructions obtained so far from continuous detector data do not take fully advantage of the benefits of working with this technology. This is an important issue, in particular for the development of the FIS project, since one of the groups collaborating owned data from one of these scanners (Balcerzyk et al. (2009), Sánchez et al. (2012)), and in general for the PET community, where a renewed interest for this type of scanners has emerged in the last years (Hun et al. (2002), Tavernier et al. (2005))

The main problem that arises when the existing system matrix implementations want to be used with continuous detectors technology is that they are not able to retain the whole precision of the data this type of detectors provides. The analytical implementations of the system matrix that would be able to retain such precision (Vardi et al. (1985)), have been traditionally discarded in pursuit of other experimental (scanner measurements, Monte Carlo simulations) methodologies (Rafecas et al. (2004), Panin et al. (2006a), Alessio et al. (2006), Ortuño et al. (2010)). Apart from being specially suited for reconstruction of continuous detectors data, the analytical approaches to the system matrix computation have other benefits (noise free, precision, flexibility) for the PET reconstruction in general. The reason these benefits have not been fully explored so far is that elaborated analytical methodologies give place to high time consuming implementations. Consequently, only those analytical approaches based in simplified models of the PET process are feasible. However, now that the recent technological advances (computer clustering, use of

Graphical Processing Units (GPUs)) provide computational resources that allow increasingly complex implementations of the system matrix (Pratx and Levin (2011), Herraiz et al. (2011)), a door is open to check to what extent a proper analytical implementation of the system matrix can lead to improved reconstructions.

The study of the state of the art has lead therefore to two open problems in the field of PET reconstruction: the lack of reconstruction methods adapted to the requirements of the continuous detector technology and, more generally, the lack of elaborated analytical methodologies for the system matrix calculation. These have been the two main topics this work has been focused in.

Structure of the document

According to the previous discussion, and after a general introduction to the PET technique provided in Chapter 1, this dissertation is built around Chapters 2 and 3, whose contents are briefly described bellow:

1. Chapter 2 deals with the application of analytical methodologies to the computation of the system matrix. As it has been mentioned, such methodologies have not been fully exploded so far, due to its slow performance. Now that the technology is allowing increasingly faster implementations, is the moment to check the improvements these methodologies can achieve. But, even if the analytical approaches are expected to have a great number of advantages, it has to be taken into account that the system matrix is an statistical model of a very complex process, in which effects of very different nature are involved. Some of these effects will undoubtedly be well modeled by an analytical approach, but there may be other effects for which an analytical treatment may not be appropriate. The objective of the chapter is not, therefore, to develop a purely anatitical system matrix implementation, but to explore how the

advantages of these approaches, that had been traditionally discarded due to its inefficiency, can be exploited. In Chapter 2, a thorough study of the PET data acquisition and of how the different methodologies are appropriate to model this process in the system matrix is performed. The result of this study is a new approach for the calculation of the system matrix based on the synergy between the analytical and other experimental approaches. The main advantages and limitations of each of the previous works on this topic are analyzed so that they justify the choices made in the design of the new system matrix. Moreover, the implication each of these choices has in the image domain has been studied in order to evaluate to what extent the new scheme improves the reconstructions.

2. Chapter 3 deals with the reconstruction for continuous detectors data. The underlying reason that has led to the development of the new methodology introduced in Chapter 2 is the absence of specific approaches that are able to retain the precision of the data of the continuous detectors technology. Reconstructions of data from this kind of scanners are currently performed with system matrices that do not take into account this or other particularities of the technique. Consequently, the reconstructions obtained so far do not take advantage of the benefits of working with continuous detectors. Once the new approach to the system matrix computations has been developed in Chapter 2, it will be proved in Chapter 3 how it can be adapted not just to retain the precision of the continuous detector data, but to account for other peculiarities of this type of devices. Apart from the new methodology to compute the system matrix, other issues of the reconstruction (data format, sensitivity calculation) have been adapted to the continuous detectors requirements, giving place to a reconstruction scheme totally suited to these type of detectors, whose advantages in the image domain have been proved.

Chapter 1

Introduction to Positron Emission Tomography (PET)

1.1 Functional imaging

Medical imaging has traditionally been thought of as a way of viewing the body's anatomy. Indeed, x-ray computed tomography and magnetic resonance imaging yield exquisitely detailed images of the anatomical structures. However, in some cases it is useful to acquire images of physiologic function instead of (or in addition to) images of anatomy. Such images can be acquired by the class of imaging techniques known as nuclear medicine imaging, whose two major subbranches are the single-photon emission computed tomography (SPECT), and the positron emission tomography (PET).

Both techniques involve the injection of a radiopharmaceutical into the subject under study to image properties of the body's physiology. A radiopharmaceutical consits of two parts: a tracer compound that interacts with the body and a radioactive label. The tracer principle (George de Hevesy, early 1900s) states that the radioactive labeled compounds are

incorporated to biochemical pathways in the same way as nonradiactive materials. Therefore, by way of the emission of gamma rays, the radioactive labels can be used to track the flow and distribution of important substances in the body. The tracer principle has two powerful benefits as a basis for imaging biological processes. First, it can be used to measure molecular concentrations with tremendous sensitivity, as one can readily detect even minute quantities of radioactive material. Second, tracer measurements are noninvasive, since the concentration of tracer is deduced from counts of gamma rays emitted from within the body as the result of radioactive decay of the administered tracer.

A wide range of biologically interesting compounds can and have been synthesized with radionuclide tags thus permitting the measurement of quantities of interest going from glucose metabolism to gene expression. Among other applications, the nuclear imaging techniques can be used to detect tumors, locate areas of the heart affected by coronary artery disease or identify brain regions influenced by drugs (Cherry and Chatziioannou (2004)).

1.2 Positron emission tomography

1.2.1 Electronic collimation

SPECT uses radiouclides that emit single gamma rays in the $80 - 350$ keV range. These gamma rays can be detected externally by position-sensitive detectors, also known as gamma cameras. A collimator is inserted between the patient and the detector that lets pass only photons that approach the detector with certain angles while discarding the other photons (see Figure 1.1). In order to image the patient from many points of view, the process is repeated by positioning the gamma camera at many orientations about the patient. Although the majority of SPECT scans at the present time are done for the diagnosis of

cardiac disease, there are many other clinical and research applications where SPECT is used.

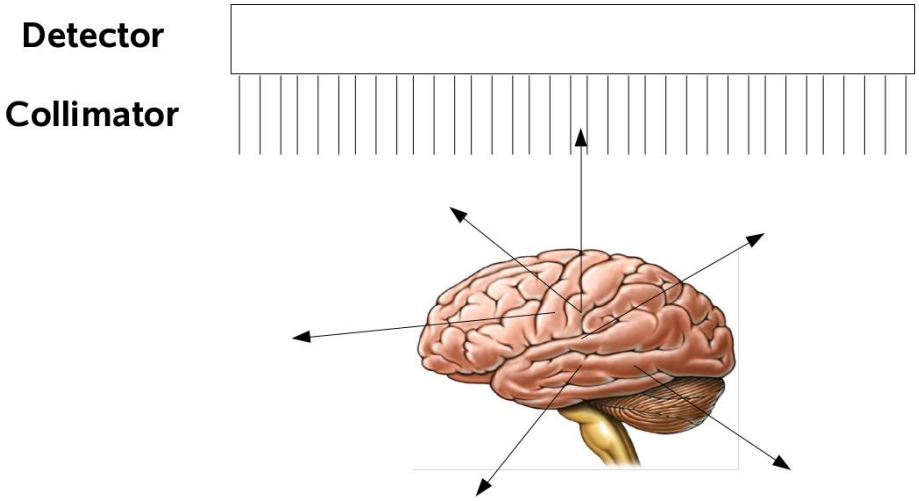

Figure 1.1: Parallel-hole collimator SPECT. Although photons are emitted isotropically at each point, only those emitted along lines parallel to the channels of the collimator reach the detector (drawings are not to scale).

SPECT collimation does an inefficient use of the emitted radiation, as the gamma rays not perpendicular to the camera face are absorbed by the lead walls between the holes. However, the collimator is necessary to define the direction from which the gamma rays are coming, thereby providing spatial information. PET studies do not require a physical collimator, but use an scheme of detection called “electronic collimation”. Electronic collimation is based on the use of positron emitting radioisotopes as radioactive labels. Positrons are the antiparticles to the electrons, with the same mass but with opposite electric charge. When the positrons reach thermal energies, they interact with nearby electrons

by the formation of a hydrogen-like orbiting pair called positronium. Positronium is unstable and eventually decays, via annihilation, into a pair of 511 keV gamma photons emitted at 180 degrees relative to one another with completely random orientantion. A PET scanner consists basically of a collection of gamma-ray detectors connected to circuitry that senses the timing of the gamma-ray detections. When two gamma rays are detected roughly simultaneously (typically an acceptance time window of a few nanoseconds is imposed), the coordinates of each photon interception are recorded by the detector system. The segment line defined by the two detected points is usually referred to as line of response (LOR).

1.2.2 PET data

The data provided by a PET scanner is not directly in the form of a tomographic image, but consist of a collection of lines of response (LORs). Consequently, some computer reconstruction method must be applied in order to estimate the radiotracer distribution that gave place to the measurements.

The collection of LORs can be stored before the reconstruction in different formats. Data can be binned as a function of the LOR spatial orientations, through structures called sinograms. The sinogram is a representation of the detections measured at a given plane. For each plane, a sinogram is a two-dimensional matrix with the vertical columns representing projection angles and the horizontal rows representing spatial positions within the projections, as shown in Figure 1.3. An alternative option is the storage mode known as list-mode. List-mode acquisition is achieved by storing information regarding the acquired events as they are detected one-by-one in the form of a list.

There are some sources of error in the recorded data. The errors come from the assumption that an annihilation event occurred along each of the LORs that make up the PET data. Not all the coincidence events accepted by the tomograph meet this assumption.

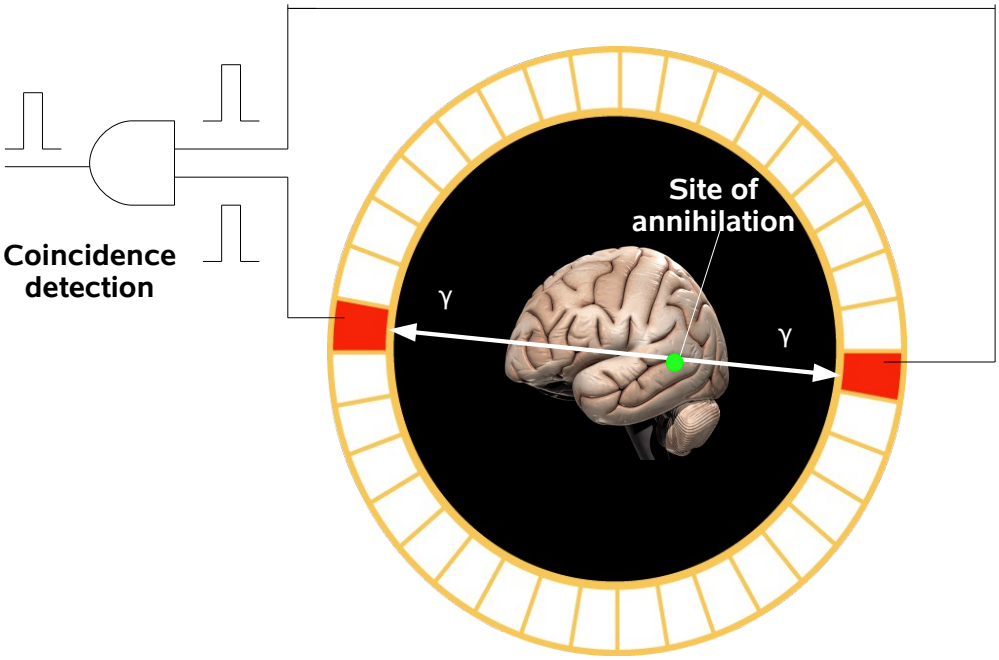

Figure 1.2: Schematic diagram of electronic collimation in PET. Two gamma rays emitted as a result of positron annihilation are sensed at two detectors at roughly the same instant. Thus, it can be inferred that a positron was generated along the line segment connecting the detectors involved.

It can happen that one or both annihilation photons interact in the object being scanned. Consequently, photons lose all (this effect is referred to as “attenuation”) or a part of their energy and can be scattered in a new direction. The interaction of the gamma rays with the object can give place to the existence of three different situations that cause coincidence events (see Figure 1.4):

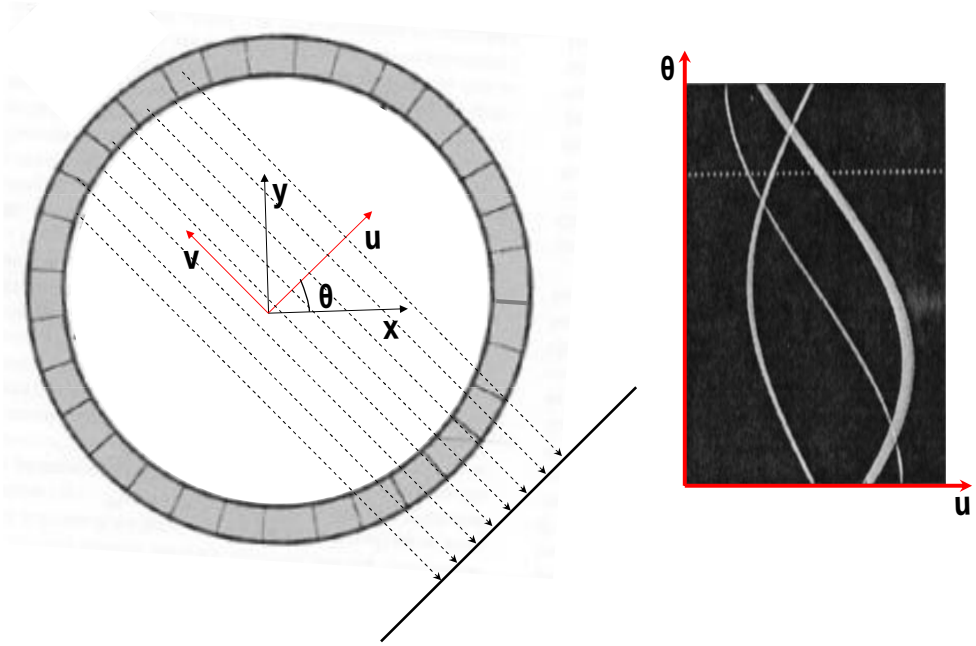

Figure 1.3: Sinogram formation. Coincidence events in PET scanner are categorized by plotting each LOR as function of its angular orientation versus its displacement from center of gantry

- A “True coincidence” occurs when a positron annihilates and both of the gamma rays are detected without either of them scattering in the object to be scanned.

- If one or both photons are scattered, but both photons are detected in coincidence, the result is a “Scattered coincidence” This is a source of error in the data, since the original annihilation is not on the line joining the coordinates of the detected photons and the resulting LOR will be misplaced.

- If one of the photons misses the detector (this can be due to an interaction or to the relatively small solid angle subtended by the detector ring), the time window will usually contain only one detected photon and the coordinates will be discarded, unless the same situation arises close enough in time with another annihilation. This case is called a "Random Coincidence" (or "Accidental coincidence") and is a source of error since the two detected photons come from different annihilations, neither of which is on the line in space joining the coordinates of the detected photons.

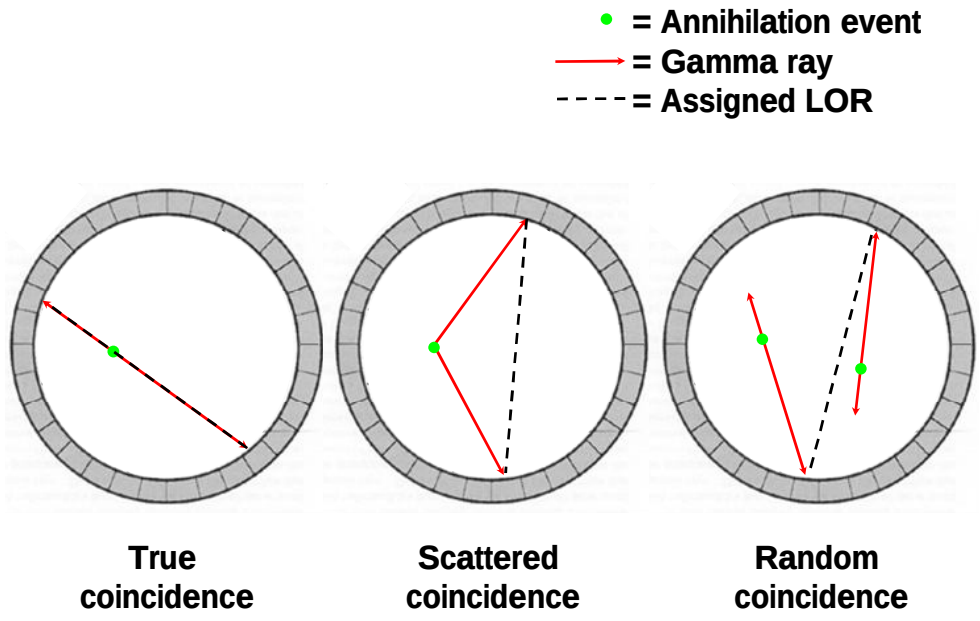

Figure 1.4: The three different situations that can give rise to a coincidence event measured in a PET scanner

During image reconstruction, corrections can be made for both randoms and scatter (Lewellen (2008)). There are two standard approaches for measuring the randoms. The delayed-window approach is based in applying a timing acceptance window that is delayed from the prompt window and only samples random events. The second approach calculates the randoms by measuring the single event rates in each detector. Scatter detection can be reduced by applying an energy acceptance criteria or estimated and corrected after data acquisition. Several analytical methods used with this purpose can be found in (Lewellen and Karp (2004)).

1.2.3 PET detectors

Although various alternatives continue to be considered (Lewellen (2008)), the most common PET detector block scheme is based on some form of scintillation detector coupled to photomultiplier (PMT) tubes. The overall function of the detection system is to convert the interaction of a 511 keV gamma ray in the scintillator into a robust current pulse that can be detected and processed by relatively standard electronics. Specifically, when a photon interacts in the scintillator, electrons are moved from the valence band to the conduction band. The electrons return to the valence band at impurities in the crystal, emitting many optical-wavelength photons in the process. The light obtained at the scintillator is collected by the PMTs, which convert the light photons into electrons and amplifies the signal. This current is sensed by accompanying electronics, which compute the spatial coordinates of the gamma-ray event relative to the face of the camera and register the occurrence of an event.

In order to optimize the detection process, a scintillator should thus be fast (for accurate timing), dense (so that the probability level of the 511 keV gamma rays interacting in them is somewhat high), have high light output (for position accuracy) and be cheap to produce.

Currently the high-end PET scanners being offered are mainly based on bismuth germinate (BGO), lutetium oxyorthosilicate (LSO) and gadolinium oxyorthosilicate (GSO).

Scintillator-based Detector designs

The first PET scanners used one PMT per scintillator crystal (Phelps et al. (1978)). As designs began to reduce the crystal cross section to obtain higher spatial resolution, several limitations to this approach were encountered: the existing PMTs were too large to pack them together with a full block of very small crystals. Moreover, the large number of PMTs and electronic channels increased the cost of the system. As a result, two alternative scintillator detector designs (see Figure 1.5), emerged in the 1980s:

- The first scheme is based on large area continuous crystals viewed by an array of PMTs (Karp et al. (1990)). Traditionally, the position of an event in a continuous detector is determined by calculating the centroid of the emitted light by means of a modified resistor network amongst all PMT array pads. In a continuous crystal detector, the main characteristics of the detector (especially energy and spatial resolution) are expected to be strongly related to the crystal surface treatment. For this reason, the surface treatment of the crystal must be optimized before mounting the crystal on the final module .

- The second approach uses smaller discrete crystals that shape the light response function (LRF) to allow the decoding of crystal positions with a small number of PMTs (Casey and Nutt (1986)). The LRF can be controlled with different coupling compounds at the interface and surface finishing or by using different lengths of reflector between the crystals. Typically, four PMTs were placed over the crystals in a rectangular pattern and ratios were formed from the PMT signals to provide a transverse and axial position signal.

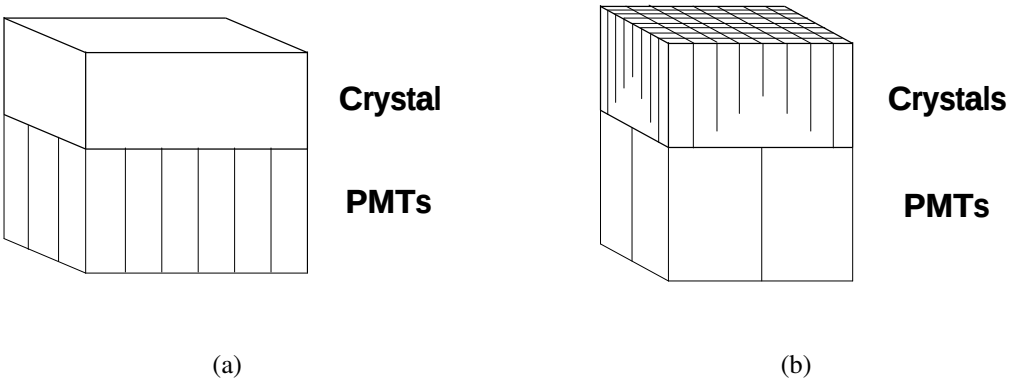

Figure 1.5: The major scintillator-PMT schemes options currently in use. (a) Continuous large crystal read by an array of PMTs. (b) Array of crystals with different treatments and/or reflectors viewed by a small number of PMTs.

The advent of new technical developments such as position sensitive photomultipliers PS-PMTs (Kume et al. (1986)) has led the PET scintillator-based detectors to evolve in a variety of design approaches, whose detailed description is out of the scope of this document. However, all these designs are based in the two above mentioned basic approaches, and their performance is still strongly associated to the continuous/pixelated nature of the scintillators:

- The use of continuous scintillators instead of pixellated blocks allows, in principle to improve the spatial resolution (since the detected positions are not associated to the centers of pixelated components), while avoiding the problems of light collection efficiency that are related to fine pixellation of the crystals. Additionally, the cost and complexity of the detector are reduced. On the other hand, the large continuous crystal approach requires scintillators with higher stopping power, in order to prevent the light from spreading too far in the crystal and achieve accurate spatial localization of

the events. Unfortunately, the scintillators with higher stopping power can't currently be grown into large crystals. An exception is LSO, with high scintillation efficiency, high cross-section for 511keV gamma rays and fast decay time, which results in excellent count-rate performance (Siegel et al. (1995)). Other disadvantages currently associated to the use of these kind of scintillators are the non-uniformity and the nonlinearity in the camera response, but their effects can be minimized during the process of camera calibration (Sánchez et al. (2004)).

- The discrete crystal machines offer higher sensitivity due to the higher stopping power of the scintillators used (BGO, LSO or GSO) and much higher count-rate performance. On the other hand, as every scintillator pixel must be treated individually before being inserted in the detector block, the cost and complexity of the PET system increases. Moreover, in a pixelated detector, energy resolution is a function not only of the intrinsic scintillation efficiency of the crystals, but also of the crystal size (Giménez et al. (2004)). This is due to the fact that smaller cross section crystals exhibit greater light loss and, consequently, lower energy resolution (a high light yield from a scintillator normally improves the energy resolution because it reduces the Poisson noise of the PMT signal).

Detector geometry

A PET scanner is composed by several of the detector modules described in the previous section. The blocks can be configured as full static rings that completely surround the patient (Surti et al. (2005)) or as partial rings in a rotation device that allows the obtaining of the needed angular sampling (Del Guerra et al. (2006)).

In two-dimensional scanners, multiple planes of detectors surround the patient with dense material, or "septa", separating each plane, as shown in Figure 1.6 (left). The septa

stop photons traveling between planes so that coincidence events are collected only between pairs of detectors in a single plane. Using this configuration, the data are separable and the image can be reconstructed as a series of two-dimensional sections. In contrast, the three-dimensional scanners (Figure 1.6 (right) have no septa so that coincidence photons can be detected between planes. In this case, the reconstruction problem is not separable and must be treated directly in three-dimensional. The septa removal in the three-dimensional mode increases the total number of detected photons and hence increases the signal to noise ratio. However, the number of scatter coincidences increases considerably when working in this mode (Townsend (1991)).

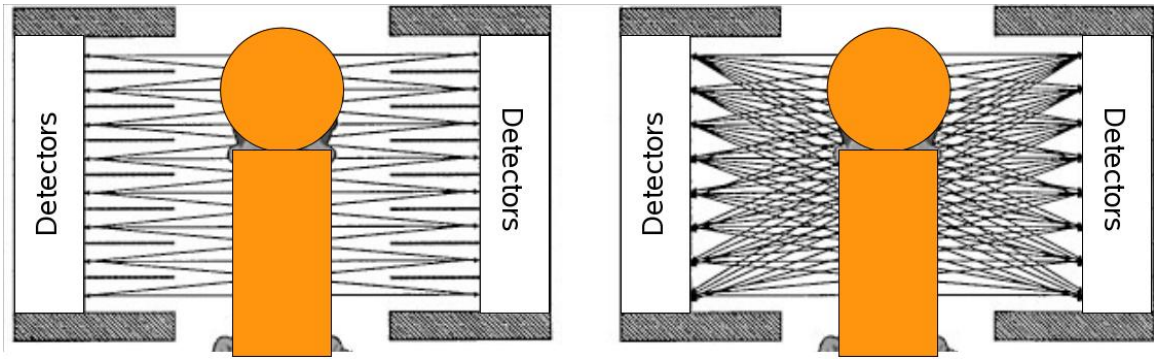

Figure 1.6: Schematic diagram of an axial cross section through (left) a 2D and (right) a 3D PET scanner. The septa in the 2D scanner stop out-of-plane photons while the 3D scanner detects these events as additional data

1.3 Small animal PET

PET scanners for human imaging have improved dramatically since their introduction in the mid-1970s. Clinical PET scanners typically produce reconstructed tomographic images with a spatial resolution in the 5 to 8 mm range. These clinical systems have also been used for animal research, predominantly in the larger laboratory animals such as non-human primates, dogs and pigs, where the spatial resolution is often adequate. There has been attempts to image smaller animals such as rats on these clinical scanners (Agon et al. (1988), Ingvar et al. (1991)) but the spatial resolution is sufficient for only a narrow range of applications. The first dedicated animal PET scanners were also designed for larger research animals, particularly for brain imaging in non-human primates (Cutler et al. (1992), Watanabe et al. (1997)). These systems were designed to obtain somewhat higher spatial resolution that clinical PET systems.

With the tremendous advances in mouse genomics and the wide range of animal models of human disease based on mice and rats, there has been significant motivation to extend the PET technique to imaging of small animals. The first dedicated small animal PET scanner (Bloomfield et al. (1995), Bloomfield et al. (1997)) took the standard detector technology being developed for clinical PET systems but placed these detector units in smaller diameter rings than common clinical PET scanners to form a compact PET system for rodent imaging. Although the spatial resolution was not superior to that found in clinical systems, it did elegantly demonstrate the concept of a dedicated small animal PET system and a tremendous amount of useful research was carried out with it. In the mid-1990s numerous groups began to develop small animal PET scanners with detector technology developed specifically for that application and with improved spatial resolution (Marriott et al. (1994), Lecomte et al. (1996), Bruyndonckx et al. (1996), Pichler et al. (1998), Chatziioannou et al. (1999)). It must be remarked that two Spanish research groups have been involved in the

design, development and manufacture of three of the commercially available small animal PET systems:

- The rPET (Vaquero et al. (2004)) and the ARGUS PET (Vaquero et al. (2005)) (commercialized under the name of eXplore VISTA by General Electric Health Care) are two pixelated small animal PET scanners developed by people of the Medical Imaging Laboratory at Gregorio Maraõn Hospital (Madrid, Spain).

- The Albira PET (Balcerzyk et al. (2009)) is a small animall PET scanner with monolithic crystals developed in the Instituto de Física Corpuscular (IFIC) and manufactured by Oncovisión (Valencia, Spain)

1.3.1 Comparison of small animal imaging techniques

One of the main advantages of PET is its noninvasive character. It is clear that noninvasive techniques providing the same or similar information than other techniques in which it is needed to euthanize the animal (as autoradiography or tissue dissection do) have great value. It permits longitudinal, within-subjet study designs to follow disease models and interventions over periods of days, weeks and even months. Because the same animal is used at every time point, each animal serves as its own control and variability due to interanimal differences is effectively removed. Therefore, a single animal studied multiple times by PET may in some instances provide the same data that would have needed tens of animals using traditional invasive techniques requiring sacrifice of the animal. This means a great reduction in cost and speed results. It may also improve the quality of the data (because of the within-subject design), although this has yet to be unequivocally demonstrated.

There exist other non-invasive techniques such as magnetic resonance (MRI) and x-ray

computed tomography providing exquisite high resolution images that are largely reflective of anatomy. Although they have an important place in small animall imaging, these techniques do not provide functional data and they are in general more suited to address a different set of questions. Some limited functional information can be obtained with modern magnetic resonance imaging methods, although the sensitivity levels are much lower than those provided by PET.

There also exist optically based techniques for *in vivo* functional imaging. These are also non-invasive techniques, that provide functional image at high levels of sensitivity. The problem with these techniques is that the emitted light can penetrate just several millimeters of tissue. The depth limitation along with the lack of tomographic methods for accurately computing the 3D emission distribution limit optical imaging methods to the smaller laboratory animals (mice) or to superficial imaging in larger animal models.

Another important advantage of PET over the optical techniques that it enables the distribution of the radionuclide (and therefore the molecule to which it is attached) to be measured in quantitative units, assuming that appropriate corrections are made for physical factors such as gamma ray attenuation and scatter in the tissue. In optical imaging, light scatter makes accurate quantification extremely challenging.

Finally, As has already been stated, compared with nuclear imaging techniques that use single gamma ray-emitting radionuclides (SPECT), PET has at least one order of magnitude advantage in sensitivity because the direction of an incident gamma ray is defined electronically by the detection of the opposing gamma ray. The physical collimation used by SPECT dramatically reduces the detected gamma ray event rate per unit of injected dose of radioactivity.

1.3.2 Applications

Dedicated small animal PET scanners are already being used for many different applications. These include measurements of the glucose metabolism in the rat brain and heart (Moore et al. (2000), Kornblum et al. (2000)) studies of the dopaminergic system in the rat (Hume et al. (1996)) and mouse (Chatziioannou et al. (1999)) brain , or investigation of the effect of photodynamic therapy in a mouse tumor model (Lapointe et al. (1999)).

Gene expression

In the environment of structural biology and functional genomics in which this dissertation has been conceived, an area of application that attracts great interest is the merger of PET with molecular biology to create methods to measure gene expression *in vivo* (Gambhir et al. (2000)). Genes direct the physical development and behavior of an organism. The long strand of nucleotides of which a gene consist of contains a coding sequence, which determines what a gene produces (such as a protein). This sequence is controlled by a promoter region, which regulates when, where, and to what extent the gene generates the product which it encodes. Gene expression is the process by which, under the promoter control, the coding sequence is made manifest as a physical and biologically functional gene product. Research into gene expression will enable scientists to decipher the functions of genes and their protein products, and to get a clearer picture of the complex regulatory networks that control fundamental biological processes. To understand these complex processes, there is a growing interest in studying the conditions under which each gene in the DNA sequence is expressed. The use of non-destructive imaging procedures like PET that allow to follow individual subjects of a same population over an extended period during which various procedures are performed, are especially valuable for the timely advancement of research in gene expression.

A PET reporter gene has been successfully used (Tjuvajev et al. (1998), Gambhir et al. (1999), Gambhir et al. (2000) that is able to produce a protein that is capable of trapping a positron-labeled compound. The reporter gene is driven by the same promoter (the promoter can be thought of as a switch that controls the level of expression of the gene) as the gene of interest, such that when the gene of interest is expressed, the reporter gene is also expressed. The retention of the positron-labeled probe by the protein product of the PET reporter gene has been shown to be proportional to the level of reporter gene expression, which in turns reflects the level of expression of the gene of interest (Gambhir et al. (2000)). In this way, the location, magnitude of expression and time course of expression levels of any gene that is introduced in a mouse can be monitored *in vivo*. The same PET reporter gene approach can be used in transgenic mice where every cell in the mouse carries the PET reporter gene, but the signal is only detected when the promoter driving the PET reporter gene is switched on. This enables endogenous gene expression to be studied in mouse models. There are widespread applications for these PET reporter gene methods (Gambhir et al. (1999)). For example, genetic tagging of tumor cells that can then be followed over time after injection in an animal, studies of the efficiency of gene therapy vectors for delivering genes into experimental animals, interactions between cancer cells and the immune system, or studies of gene expression patterns during development to see when certain genes are switched on or off. This opens up many powerful research opportunities that take advantage of the ability of PET to longitudinally measure gene expression in an entire mouse.

1.3.3 Resolution and sensitivity in small animall PET

The previous sections have introduced the advantages and the benefits of the use of PET in the field of biomedical research. However, it must be taken into account that, due to

the differences in size between rodents and humans, small animal PET imaging imposes challenging performance requirements, particularly on system sensitivity and on image resolution (Vaquero and Desco (2005)).

Resolution

There are many examples in the small animal PET field where the ability to visualize and accurately measure radiopharmaceutical accumulation in structures that have dimensions of a millimeter or less in size is important (Stickel et al. (2007)). The typical image resolution of many of PET systems is in the 1-to 2.5-mm range (Cherry and Gambhir (2001)). Whether the submillimiter range is close to be reached for small animal PET is not all clear since there are several complex factors interacting to limit spatial resolution during data formation and collection:

- ***Positron physics***. The assumption that the annihilation that gave place to the true coincidences occurred somewhere along the line connecting the two involved detectors does not take into account the effect of two positron physics factors:
  - *Positron range*: Before reaching thermal energies, positrons travel through tissue giving up their kinetic energy principally by Coulomb interactions with electrons. The distance the positron travels before annihilating is termed positron range. The magnitude of this range depends on the positron energy, which varies widely among isotopes and with the fraction of air in the tissue. The only known way to reduce this effect is using a strong magnetic field, an attractive approach since the PET scanner can be potentially merged with a Magnetic Resonance Imaging (MRI) scanner (Blanco (2006)).
  - *Non-collinearity*: One would normally expect the annihilation gamma rays to be antiparallel. However, the residual kinetic energy and momentum of the

positron and electron at the time of annihilation results in an angular uncertainty in the direction of the 511 keV photons that can be modelled as a Gaussian distribution with a standard deviation of 0.212 degrees (DeBenedetti et al. (1950)). Scanner design can minimize non-collinearity by minimizing the separation between detectors.

- ***Crystal penetration***. Once the gamma rays reach the scintillator, they may travel some distance in the scintillator before being absorbed. As a result, if the gamma ray enters the crystal at an oblique angle, the location of the interaction will not be the same as the point of entry into the scintillator. Thus, an incorrect line of response will be assigned to the interaction because the LOR is normally assigned to a position at the front of the crystal interaction. This source of error worsens as the source position moves radially away from the center of the scanner because a larger fraction of the gamma rays enter the crystals at oblique angles. The consequence of this effect is a non uniform resolution response that degrades radially across the field of view.

  One solution that is a very active area of development is to add the ability to determine how deep in the crystal an event actually occurs (Lewellen et al. (2004), Inadama et al. (2006)). Moreover, crystals with high density help to reduce the crystal penetration effect, since its high stopping power reduces the variability of the depth of interaction (DOI) of incident gamma rays in the detectors.

- ***Detector scatter***. A 511 keV photon can undergo two main interactions in the scintillator: Compton scattering and photoelectric absorption. A Compton scattering interaction results in a scattered photon and a recoil electron. Photoelectric absorption results in a photoelectron and in a characteristic x-ray. The x-ray produced in photoelectric absorption and the Compton scattered photon may interact again at some distance from the original interaction site or alternatively they may escape from the

detector. As a result, each 511 keV gamma photon emitted can cause interactions at different points in the scintillator crystal. The position of the final event depends on the particular detector readout and signal processing of the system. The center of mass of the deposited energy is the positioning approach underlying most current PET detector technology, although other positioning approaches exist (Tavernier et al. (2005)). Independently of the positioning method, the increase in the number of interactions in the scintillator leads to an increase in the error of the location of the initial photon interaction (Stickel and Cherry (2005)).

The degradation of the spatial resolution due to the detector scatter can be fought using narrow detectors that reduce the room for secondary interactions and a good energy discriminator that rejects the low energy events coming from secondary interactions (Stickel and Cherry (2005)). However, both the detector thickness reduction and energy discrimination have the cost of a possible reduction in the overall scanner sensitivity (as will be seen next. Rafecas et al. (2003) show how to overcome this problem by using individual crystal readout along with an appropriate identification scheme to select the primary crystal. Other techniques applied during calibrations (Surti et al. (2009)) have shown to reduce the effect of the detector scatter in the reconstructed spatial resolution.

- ***Light emission***. Unlike one to one scintillator-photomultiplier tube coupled detectors, detector blocks like those shown in Figure 1.5, have additional degradation of their spatial resolution, since the shared scintillation photons within the block create an uncertainty in positioning the events. In linear positioning algorithms, as the energy centroid computation, the result is a loss of resolution due to misidentification of the interaction point. An alternative, is to use positioning techniques based on statistical models of the light behavior within the scintillators (Joung et al. (2000))

rather than linear algorithms.

- ***Detector design***.Each of the detector designs described in Subsection 1.2.3 deals with particular issues that contribute to the resolution degradation:
  - For a detector composed of small discrete crystals, all interactions are assumed to occur at the center of individual crystals and consequently the spatial resolution of a pixelated scanner is limited to approximately half the width of an individual crystal (Lewellen (2008))
  - In the case of continuous designs, where the light response function (LRF) is not shaped, the statistics of the light emission and collection as well as the distortions in the shape of the LRF due to reflections from the sides and back surfaces work to degrade the positioning of the event (Lewellen (2008)). The LRF can depend strongly on the DOI if the crystal is thick enough and the light reflections are not controlled. This DOI dependence can be either an hindrance or an added parameter to extract and utilize to address the parallax problem (Lerche et al. (2005)).
- ***Statistics of the light***. The resolution is affected as well by the statistical fluctuations of the phototube signals, which depend upon the light output of the crystal and the conversion efficiency of the photocathodes. High light output scintillators are chosen to minimise the statistical fluctuations.

Sensitivity

Sensitivity in the PET environment refers to the fraction of radioactive decays that result in a detected event. Being the amount of injected dose rather limited for small animals (in order to not perturb their biological systems), high levels of sensitivity are needed to

achieve acceptable numbers of detected events and consequently acceptable signal to noise ratios in the reconstructions (Cherry and Chatziioannou (2004)). However, the sensitivity of most current animal PET scanners is in the range of $0.5 - 2.5\%$ at the center of the scanner (Chatziioannou (2002)), indicating that a large number of decays do not lead to recorded events. There are three major ways in which events are lost (Stickel and Cherry (2005)). First, one or both of the 511 keV photons may not intersect the detector system. This is remedied by designing PET systems with good solid angle coverage. Second, if a photon intersects a detector it may not interact with it. This requires that detectors have reasonable efficiency. Typical efficiencies are in the range of $20 - 70\%$, and depend on the detector material and thickness. Finally, events are not detected if they fall outside the energy windows set in order to reject secondary interactions of the particles with the scintillator. Tight energy windows (e.g $350 - 650$ keV) can reject a very significant fraction of events. It is therefore important to set energy thresholds that can capture all possible events.

The path towards much higher sensitivity animal PET systems, without increasing cost, is to design high efficiency ($> 60\%$) detectors with adequate depth of interaction determination to compensate for the resolution degradation, as will be discussed in next section. They can be brought closer to the animal to reduce the detector area required per unit solid angle coverage. Using this approach, along with optimized energy selection, should yield system sensitivities in the range of $10 - 20\%$ (Stickel and Cherry (2005)).

1.4 Reconstruction

It has been mentioned that the small animal PET studies require higher sensitivity and resolution capabilities than the clinical applications. It is clear that, in spite of its advantages, these performance issues (along with the scanner cost, access to PET tracers, and user

friendliness) will ultimately dictate the level and extent of PET's participation in the future of the biology (Cherry and Chatziioannou (2004)). Initially, the effort to address these challenges had been focused largely on instrumentation (some outlines on this topic have been provided in Section 1.3.3). But as animal PET scanner technology started to mature, it encountered certain limitations to overcome these problems. For example, the requirements of high resolution and sensitivity can be satisfied using detector modules composed of narrow and long crystals, placed in a gantry with the smallest possible diameter. However, long pixelated crystals lead to significant detection uncertainties due to the crystal penetration. The use of short pixelated crystals reduce this effect but at the same time leads to a reduced detection efficiency. On the other hand, the increase of the gantry diameter can reduce the DOI effect but at the same time leads to a reduction of sensitivity and a degradation of the spatial resolution due to non-collinearity of the pairs (Ortuño et al. (2010)). Under, these paradoxical circumstances, other areas apart from the instrumentation have started to be given their due attention. A particularly important area is the reconstruction step, in which the acquired data is converted to tomographic images. There is actually strong evidence that the use of appropriate reconstruction algorithms will be required to achieve submillimeter reconstructed images at acceptable signal-to-noise values in small animal PET studies (Cherry and Chatziioannou (2004)).

1.4.1 Analytical reconstruction

The early techniques of image reconstruction from PET data were analytical approaches based on the method of Filtered Back Projection (FBP) (Shepp and Logan (1974)) . FBP is a mathematical technique based on an idealized model of PET that ignores many significant features of real data. Specifically, FBP assumes that the number of detected gamma-ray events traveling a particular direction approximates the line-integral of the radio-isotope

distribution along that line, from which the image can be reconstructed using analytical inversion formulas. In spite of its approximate nature, FBP has enjoyed widespread use and great longevity largely because of its computational simplicity. Unfortunately, FBP amplifies the signal noise when applied to the low-count data of nuclear molecular imaging (it is still a good method for applications where the number of measured counts is high, as computed tomography), and there is a growing interest in the development of reconstruction alternatives that overcome this problem (Kinahan and Rogers (1989), Ollinger and Fessler (1997), Lewitt and Matej (2003)). These techniques represent an important contribution to the results obtained traditionally with FBP, providing better resolution and noise characteristics.

1.4.2 Iterative reconstruction

Analytical methods typically neglect noise and complicating physical factors in an effort to obtain frameworks that yield explicit inversion formulas for the reconstruction problem. The introduction of iterative methods allowed for the explicit inclusion of realistic factors in the reconstruction process. These type of algorithms achieve better spatial resolution and improved signal-to-noise ratio than the analytic reconstruction methods, while maintaining the quantitative nature of the data (Johnson et al. (1997), Chatziioannou et al. (2000)). The price of this added refinement is that the resulting set of equations describing the problem becomes very large and non-linear, and solving for the tracer distribution by direct inversion of the forward problem becomes intractable. In this case, the equations must be solved using iterative methods (Lee et al. (2004)).

All the iterative reconstruction methods share a number of common traits. The general model behind statistical reconstruction involves repeating the process of projecting an image estimate, comparing the estimated projections to the measured data to compute some

form of error, backprojecting the error and using the error to update the image estimate.

Two wide classes of iterative algorithm exists, namely algebraic and statistical. Classical algebraic iterative algorithms, such as ART and MART (Gordon et al. (1970)) are based on the Kaczmarz method of solving systems of linear equations (Kaczmarz (1937)). Although not widely used in nuclear medicine, they form an important basis for understanding the statistical algorithms that were developed later and for which iterative methods are mainly used.

Statistical algorithms have two basic parts: a statistical criterion (the basis for determining which image among the many possible is to be the solution) and a numerical algorithm (the method for finding the solution prescribed by the criterion). The most successfull early statistical algorithm, MLEM (Vardi et al. (1985)) uses a Poisson model of the produced data, related to the maximum likelihood statistical criterion, along with the EM algorithm as numerical method.

A serious disadvantage of MLEM is its slow convergence, which can lead to prohibitive reconstruction times on standard computer platforms (Lewitt and Matej (2003)). Several acceleration techniques have been proposed for MLEM algorithm. Lewitt and Muehllehner (1986) improved convergence speed by incorporating an over-relaxation parameter. Tanaka (1987) used a frequency amplification method to accelerate estimation of the higher-frequency components in the image. The ordered subsets expectation maximization (OSEM) method (Hudson and Larkin (1994)) divides the data into subsets, giving a factor speed increase proportional to the number of subsets chosen. Browne and DePierro (1996) proposed the row-action maximum-likelihood algorithm (RAMLA), which uses as many subsets as there are projections. However, these block methods don't necessarily converge, as reported empirically by Byrne (1997). The space-alternating generalized EM (SAGE) (Fessler and Hero (1994)) improves the convergence rate by updating each pixel individually and using a matrix-based projection model. Rebinning methods such as single

slice rebinning (SSRB) (Daube-Witherspoon and Muehllehner (1987)) or Fourier rebinning (FORE) (Defrise et al. (1997)) can reduce the dimensionality of 3D acquisitions by performing a set of 2D reconstructions to obtain volumetric data, although there is a loss of image quality with respect to the more time consuming fully 3D implementation.

In addition to the slow convergence issue, there is another important drawback related to the MLEM derived methods: at high iteration numbers, images exhibit high-variance behavior (Leahy and Qi (2000)). This is usually attributed to either the fact that there is no stopping rule in this kind of iterative reconstruction or to the statistical (noisy) nature of the detection process and reconstruction method (Herraiz et al. (2006)). The usual approach to overcome this drawback is either to use stopping rules (Veklerov and Llacer (1987), Johnson (1994), Coakley (1991)) or to smooth the images with kernels (Snyder et al. (1987), Liow and Strother (1991)) filters (Slijpen and Beekman (1999)) or wavelet based methods (Mair et al. (1996)).

The Maximum a Posteriori (MAP) methods (Green (1990)) offer a more flexible and principled method of encouraging desirable properties in the reconstructed imagen by incorporating *a priori* information (priors) that models the distribution of activity and noise in the acquired data. The regularizing influence of the prior, controls the variance of the reconstruction, and hence the MAP methods do not exhibit the instabilities at higher iterations encountered using MLEM and OSEM (Qi et al. (1998))

There exist also statistical algorithms based on a Gaussian model of the produced data, related to the weighted least squares (WLS) criteria, which result in quadratic objective functions. These can efficiently use many established numerical algorithms, such as coordinate gradient (Tsui et al. (1991), Kaufman (1993)) or coordinate descent (Fessler (1994)), and result in faster reconstruction algorithms.

The algorithms mentioned above use binned data (e.g., projection sets) to reconstruct the image. Concerning the List-mode format, it is not amenable to analytic reconstruction

methods such as FBP without first histogramming the data into a standard sinogram or projection data format, which can be an inefficient procedure. However, there exists several versions of the MLEM and OSEM algorithms (Reader et al. (1998), Reader et al. (2002b), Rahmim et al. (2004)) for the list-mode data.

The MLEM algorithm

As mentioned in the previous section, the MLEM algorithm is a golden standard of the statistical reconstruction. Since its publication in 1985, most of the work in this field has been based on finding alternatives to the MLEM that offer faster convergence, improved resolution, contrast and/or signal-to-noise ratio. The MLEM reconstruction algorithm (Vardi et al. (1985)) is indeed a golden standard in the statistical reconstruction field. It seems worth to analyze in detail the work by Vardi et al. (1985) in order to have a clear idea about where the advantages of statistical reconstruction stem from.

The mathematical development of Vardi et al. (1985) starts from the well known statement that positron emissions from a large number of radioactive nuclei occur according to a Poisson distribution. The reconstruction problem consists of finding out the unknown intensity function $\lambda(\mathbf{r}) : \mathbf{r} \in \mathbb{R}^3$ of the emission spatial Poisson point process in a certain region H (the patient's body) of $\mathbb{R}^3$.

The measured data set $n^*(d)$, from which the Poisson intensity function must be estimated, are the total number of coincidences in each tube d formed by a pair of detectors. Since classifying the annihilations according to the discrete detectors pair that detected them amounts to a thinning of the Poisson point process, it can be shown that the data $n^*(d)$ (with $d = 1..D$, D being the the maximum number of detector bins), constitutes a collection of independent Poisson random variables with mean:

$$\lambda^*(d) = \int_{\mathbb{R}^3} \lambda(\mathbf{r}) c(\mathbf{r}, d) \mathbf{dr} \quad d = 1, ..., D \tag{1.1}$$

where $c(\mathbf{r}, d)$ is the probability that a line originated at the spatial point $\mathbf{r}$ is detected at the discrete tube d.

The manipulation of images in digital form is an essential part of virtually all scientific disciplines. In applications in which the physical magnitude is not inherently discrete (as is the density function that PET reconstruction algorithms try to estimate), a digital image is used to represent the continuous image that, in turn, represents the physical magnitude. When implementing the reconstruction algorithm described in Vardi et al. (1985), it was assumed that the radiotracer volume can be approximated by a fine grid of B basis functions. If $v(\mathbf{r} - \mathbf{r}_b)$ stands for the general mathematical expression of the basis function centered at point $\mathbf{r}_b$, then:

$$\lambda(\mathbf{r}) \approx \sum_{b=1}^{B} \lambda(b) v(\mathbf{r} - \mathbf{r}_b) \tag{1.2}$$

where the weight $\lambda(b)$ will be the image value at the b-th basis function, which is proportional to the total number of positron-emitting nuclei contained in the volume spanned by the basis function. Substituting (1.2) into (1.1), the following expression is obtained for the mean of the measured data in each tube d:

$$\lambda^*(d) \approx \sum_{b=1}^{B} \lambda(b) p(b, d) \tag{1.3}$$

where

$$p(b, d) = \int_{\mathbb{R}^3} v(\mathbf{r} - \mathbf{r}_b) c(\mathbf{r}, d) d\mathbf{r} \tag{1.4}$$

is the probability that an event generated in the region defined by the b-th function is detected in the tube d. Coefficients $p(b,d)$ can be arranged in a matrix called the system response matrix or simply, the system matrix.

At this point, the activity distribution $\lambda(\mathbf{r})$ that maximizes the probability of obtaining the measured data $n^*(d)$ must be found. Since the collected data follow a Poisson model, this corresponds to the maximization of the probability function:

$$P(n^*) = \prod_{d=1}^{D} e^{-\lambda^*(d)} \frac{\lambda^*(d)^{n^*(d)}}{n^*(d)!} \tag{1.5}$$

The expression in (1.5) can be rewritten in terms of the set of basis functions by means of substituting $\lambda^*(d)$ by the expression given by (1.3). As the resulting expression is concave, hence it follows that sufficient conditions for λ to be a maximizer of the likelihood are the Karush-Kuhn-Tucker (KKT) conditions (Karush (1939)), which can be satisfied with many iterative schemes. Particularly appealing is the one given by the expectation maximization algorithm. The final expression given in Vardi et al. (1985) is an instance of this algorithm and constitutes the basis for much of the work in statistically based algorithms in medical image reconstruction over the last years:

$$\lambda^{new}(b) = \frac{\lambda^{old}(b)}{p(b,.)} \sum_{d=1}^{D} \frac{n^*(d)p(b,d)}{\sum_{b=1}^{B} \lambda^{old}(b)\, p(b,d)} \tag{1.6}$$

where

$$0 < p(b,.) = \sum_{d=1}^{D} p(b,d) \tag{1.7}$$

The term $p(b,.)$ is a normalization factor representing the probability of detecting (considering all possible detectors) a pair of photons arbitrarily emitted from within function b.