High throughput digital β autoradiography imaging

J. Cabello Velasco

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

Centre for Vision, Speech and Signal Processing
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

July 2009

Summary

This thesis presents three main strands of work concerned with developing digital imaging for high throughput β autoradiography. These three strands comprise work with the image sensor technology, Monte Carlo simulation and the use of post-acquisition image analysis based on image registration. In this way, the complete autoradiography imaging chain is addressed.

CCD and CMOS imaging technologies are presented as potential imaging alternatives to using conventional film in autoradiography. These digital technologies exhibit enhanced sensitivity, dynamic range and linearity compared to film using imaging methods developed at Surrey. These imaging methods address the different sources of noise typically present in CCD and CMOS technologies. Tissue imaging using ^{3}H, ^{35}S and ^{121}I, the typical radioisotopes used by the Drug Addiction Group in the School of Biomedical and Biological Sciences, is presented. The first successful images of ^{3}H-labelled tissue sections using CCD and CMOS technologies operating at room temperature are presented as one of the main achievements of this work.

To better understand the image creation process some prelimiary Monte Carlo simulations, using the GEANT4 toolkit, have been undertaken, demonstrating intrinsic and extrinsic key parameters of these digital sensors that can be used to optimise spatial resolution. These simulations demonstrate that each radioisotope requires a different optimum detector architecture. In this work these optimum architectures are analyzed.

To support the high sensitivity (i.e. fast) imaging produced by the sensor technology, automated post-acquisition analysis is also considered, using an atlas-based image registration approach, by previously aligning automatically segmented biological landmarks using a feature-based extraction approach, region growing. This has the potential to speed up the post-acquisition analysis aspects of the imaging chain. Thus a computer-based tool designed to semi-automatically elastically register a radiogram with an atlas has been developed.

Key words: Digital Autoradiography, CCD Technology, CMOS Technology, Monte Carlo, Region Growing, Image Registration.

Email: j.cabello@surrey.ac.uk
WWW: http://www.ee.surrey.ac.uk/

Acknowledgements

I would like to thank my supervisor Dr. Kevin Wells for his guidance and his understanding during this long time. Secondly I would like to dedicate this work to my family for all their support during these four years, thanks for making me as I am, otherwise I would not have been able to finish this. Also I would like to greatly thank Rebeca (aka the parrot) for being next to me during this time, sharing with me the good and bad moments, you are also partly responsible for this achievement. Thank also to the guys from Medical Imaging for their long and constructive discussions about work, believe me, I have learnt a lot from all of you. In addition, I would like to thank to my colleagues from CVSSP and the FHMS for their support and help in different aspects of this work. I do not want to forget to my friends and all the people that have passed through my life during my time in UK. Thanks to all of you!

Contents

Chapter 1

Introduction

Autoradiography (AR) is a method used to map the distribution of radiolabelled biomolecules, known as tracers, deposited in thin tissue specimens. There are many different applications for this technique that measure the pathways of many different biomolecules (see for example Figure 1.1). The particular application considered in this work is based on the identification, localisation and quantification of neurotransmitter receptors in brain tissue sections (see Figure 1.2). Studying the binding mechanism of dopamine ligands with dopamine receptors, these can be inhibited with new drugs avoiding the addiction. This is important in studies of the mechanism of addiction and the action of drugs on specific brain regions. The aforementioned neurotransmitters are targeted with specific drugs. For imaging purposes these drugs, or ligands, are labelled with a certain radioisotope. Thus when the ligand under study binds with a specific receptor, the location of this receptor binding is identified by detection of the radioactivity emitted by the radioisotope.

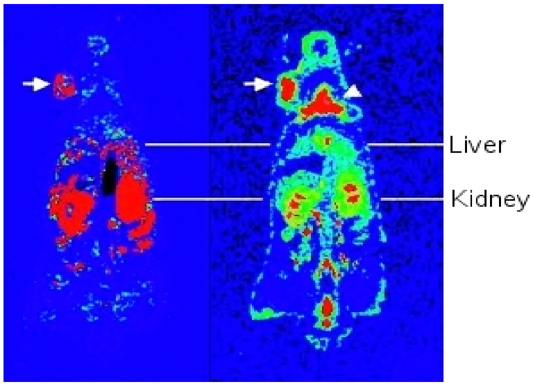


Figure 1.1: Exemplar whole body autoradiographic images of a whole body mouse with ^{3}H-TTP[1] (left) and ^{18}F-FDG[2] (right) of a tumour in the shoulder (arrow) exposed to conventional film. The TTP distribution demonstrates a more selective binding to glucose consumer centres such as the cancerous tissue, liver and kidneys, whereas the FDG image demonstrates a less selective binding, showing presence of the ligand in the background, making more difficult the correct delineation of the tissues under study.

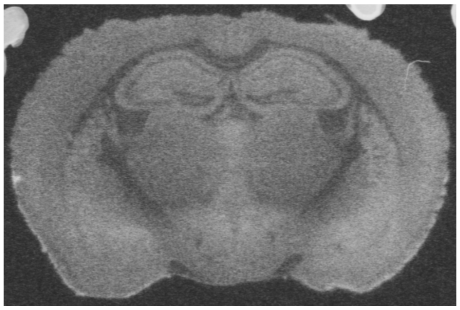


Figure 1.2: Autoradiographic image of a coronal mouse brain section bound with ^{35}S-GTP[3]γS demonstrating ligand binding to μ-opioid receptors exposed to conventional film for 4 days.

The most typical radioisotopes used to label the aforementioned ligands in autoradiography are ^{35}S, ^{14}C and ^{3}H. These radioisotopes emit β particles which are fast electrons. These β electrons exhibit tortuous paths in tissue and the detection medium because of their low mass. Thus, a collision with a nucleus or even with another electron will change their trajectories losing energy in the process. This loss of energy depends on the distance of the β particle to the orbital electron and its kinetic energy. This will have a critical effect on the spatial resolution of the final image.

Film emulsion has been the traditional detection medium used since Lacassagne and collaborators developed the first autoradiographic method to localise radioactive Polonium in biological specimens in 1924 [91]. Film has some special characteristics that make it suitable for autoradiography. autoradiography film is similar to that used in traditional photography: it is basically comprised of a layer with tiny halide silver bromide grains or crystals (0.4-1.0 μm diameter) contained in a gel emulsion. The proportion of crystals per unit area controls the sensitivity of the film and the background noise. The number of crystals should not exceed 10 crystals per 10,000 square micron, higher counts of crystals may obscure low signals and corrupt the evaluation of experimental data [164]. These grains are ionised by the β particles emitted from the tissue sample leaving a latent image. The extremely small size of the aforementioned grains means this technique exhibits excellent intrinsic spatial resolution (∼1-50 μm for ^{14}C, ^{3}H and ^{35}S imaging, due mainly to scattering) for low cost (∼\$3 per sheet for standard film and ∼\$50 per sheet for hypersensitive film). Typical autoradiographic film has an anti-scratch protection layer on both sides, although hypersensitive autoradiographic film lacks these layers to enable the detection of lower energy charged particles such as those emitted from ^{3}H. This obviously makes this film more delicate to handle.

On the other hand this technology presents some disadvantages such as poor linearity, especially at low activity levels, limited dynamic range (∼10^2) and in particular low sensitivity producing long exposure times (∼days or ∼weeks depending on the radioisotope).

Non-linear response may be observed in regions with extreme (high or low) activity,

---

[1] Thymidine-triphosphate

[2] Fluoro-2-deoxy-D-glucose

[3] Guanosine 5-Triphosphate

i.e. high or low uptake of drugs. In these areas the film can start to saturate and the number of ionised halide crystals is no longer proportional to the number of emitted β particles, i.e. activity. Thus the subsequent analysis, in which the exact level of uptake in particular regions of the drug under study is measured, needs to be corrected each time an experiment is undertaken.

The lack of sensitivity is due to the small diameter and low intrinsic density of the halide crystals suspended in the gel emulsion. The probability of interaction (cross section) of a β electron with a halide crystal is very low, leading to extended exposure times. This will depend also on the energy of the radioisotope. With higher energy β electrons there is less probability these will be stopped in the overlying structures of the film (anti-scratch coating) and thus proportionally more electrons will pass through to, and ionise the halide crystals.

There are also some other secondary issues such as the cost of consumables needed to develop autoradiography film, the necessity of having a dark room, the sensitivity to accidental damage (and the subsequent waste of time after an experiment). Moreover there is also a significant time overhead to analyse the resulting data, typically representing ∼70 hours of man effort. Previous efforts [108, 61] to address this area have corrected or enhanced the detection technology, and as this is where the main issue resides, this approach will also be considered here.

To better understand the physical mechanisms that affect autoradiography imaging performance Monte Carlo methods have been used. A set of simulations has been prepared, with different sensor geometries and different radioactive sources, to find the optimal geometry of a digital sensor for application in digital autoradiography.

However, as mentioned above, there is also a significant overhead associated with manual post-acquisition Region of Interest (ROI) analysis. Therefore, this bottleneck in the experimental process has also been addressed, to fully encompass the over-arching aim in this thesis: achieving high throughput autoradiography. As a result a computer-based tool has been developed in quantitative autoradiography (QAR) that registers an autoradiogram with a brain atlas, thus producing automated ROI analysis of arbitrary anatomical brain regions.

Chapter 2

Technology Review

A number of alternative imaging technologies to conventional film, indicated in Figure 2.1 in chronological order, have appeared in the last few years claiming better performance than film in some aspects (mainly sensitivity) to the point that may make these viable options to replace autoradiographic film. Among the technologies harnessed for this application are Microchannel Plates (MCP), gaseous wire chambers, optical-based approaches, such as Phosphor Imaging Plates (IP) or scintillators, and finally the most prolific alternative, silicon-based detectors. Among the most interesting of this latter family are the Depleted FET (DEPFET) BIOSCOPE system, silicon strip detectors, Charge-Coupled Devices (CCD) and a variety of designs utilising cost-effective Complementary Metal-Oxide-Semiconductor (CMOS) technology. These are described below, with exemplar results, summarising the current *state of the art* imaging technology applied to autoradiography. However, as this area of development has matured over the last few decades, inevitably different types of technologies have also been combined and different imaging methods adopted. For the purposes of this chapter, there are several approaches: indirect methods, where an intermediate step is used to convert the deposited energy to some other form (often optical), which is then registered using an optical imaging detector, and direct detection, where the β electrons produce charge directly.

2.1 Indirect Detection Methods

2.1.1 Optical Systems

The two main alternatives using optical photons are storage phosphors and scintillators. The Imaging Plate (IP) was invented by Fujifilm and was first used for X-ray imaging [161]. It was later used for radioisotope and electron detection. The IP is a storage film comprised of a protective layer on top, a support layer at the bottom and a photo-stimulable phosphor layer in between. The photo-stimulable layer contains phosphors of barium fluorobromide (BaFBr) containing an amount of bivalent europium as a luminescence centre, $\sim$5 μm diameter, that partially traps and stores the energy of

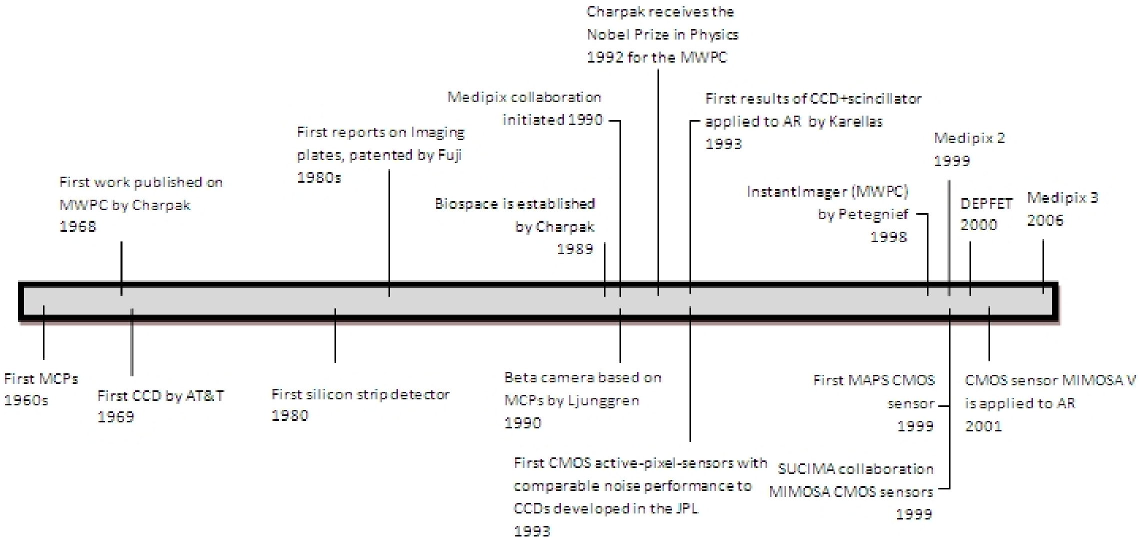


Figure 2.1: Time line showing the evolution of various imaging technologies that have been used in the development of digital autoradiography. Note the dominance of silicon-based imaging sensors in recent years.

particles while they traverse the detection medium. This energy is trapped in the phosphor until scanned with a laser. When the phosphor molecules are photostimulated, they release the stored energy as luminescence that is detected by a photomultiplier tube (PMT). This technology has the advantages of 10-1000 times greater sensitivity than conventional film, depending on the sample, it is reusable and it has a wider dynamic range and superior linearity. This technology is commercialised by FujiFilm Life Science [57] with the *Bio-imaging analyser system* (BAS), the BAS-5000 being the most recent release, and by GE Healthcare with the *Storm* gel and blot imaging system. Examples of autoradiographic data obtained with the BAS-5000 are shown in Figure 2.2.

(a) (b)

Figure 2.2: Examples of autoradiographic data acquired with BAS-5000. Rat brain section labelled with ^{35}S exposed for 36 hours (a) and whole body autoradiogram where expression of the apoptosis gene labelled with ^{33}P exposed for 60 hours is shown (b) [57].

Results using ^{14}C [80] and ^{3}H [182], with an exposure time of 2 and 15 hours respectively, have been reported, claiming a spatial resolution of 50 μm for ^{3}H [182]. This alternative has been the most attractive approach to replace conventional film for the life science community due to its excellent sensitivity so far, but this is still not of general use due to the open discussion about the best spatial resolution achievable by this method. Surrey's Drug-Addiction Group, that has collaborated in this thesis work, and other groups that use autoradiography as detection method for their research, certainly discard IP technology for high resolution rodent imaging due to the low spatial resolution exhibited by this technology.

The aforementioned scintillation materials have widespread use in broader medical imaging applications such as X-ray computed tomography (CT), single photon emission computed tomography (SPECT) and positron emission tomography (PET). Scintillators absorb ionising particles and produce fluorescing visible photons, subsequently detected by an optical detector, such as CCDs or PMTs.

The first work in scintillators applied to autoradiography was reported by Karellas [81] where a thin sheet of scintillator was applied to a low noise CCD. A much more successful realisation of this approach is the μ-Imager (μ-Imager 2000^{TM}), a commercial system developed by BiospaceLab, that is based on a scintillator sheet, an image intensifier tube and a cooled CCD camera, as shown in Figure 2.3. The ionising particles emitted by the sample generate an instantaneous cloud of photons in the thin (∼5 μm thick) scintillator, that are collected by an intensifier tube and guided to a CCD detector. Although good sensitivity can be achieved, there will be no discernable energy resolution because so few optical photons are detected. Dual tracer imaging has to rely on other phenomenon (such as the size of the observed scintillation light splash for each event) to separate different tracers. Results with ^{3}H and ^{35}S are shown in Figure 2.4 claiming 15 μm spatial resolution [12].

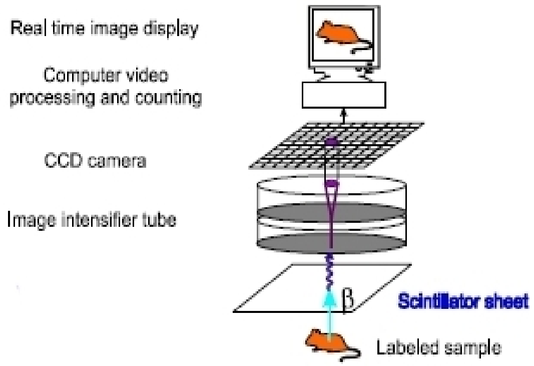


Figure 2.3: General operating principles of μ-Imager [12].

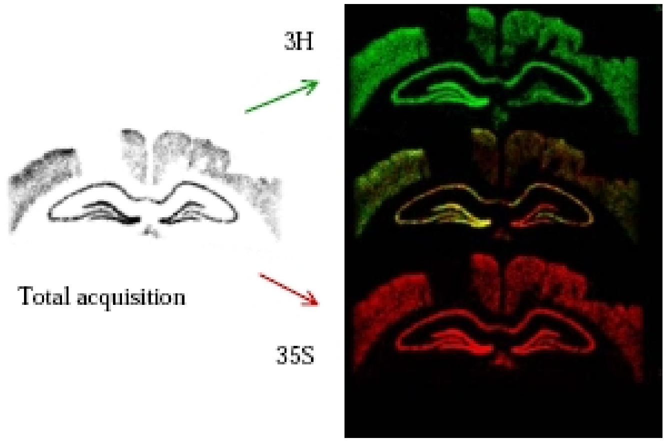


Figure 2.4: Rat brain section labelled with ^{3}H and ^{35}S obtained with the μ-Imager [12].

Another approach presented in [35] uses a lens-coupled cooled CCD camera (20 μm pixel size) with a 3 μm thin phosphor film placed next to the sample under study. In-vivo experiments with ^{99m}Tc labelled glandular tissue, exposing mandibular glands in direct contact with the thin phosphor film, were undertaken. The spatial resolution (FWHM) is estimated at 60 μm assuming a ^{90}Y/^{90}Sr point source, and a theoretical 50 μm FWHM (2.5 pixel size) is obtained assuming an ideal 70 μm FWHM ^{99m}Tc disc. The spatial resolution presented in this work can not be extrapolated to β- sources given their different manner of interaction in silicon. Moreover the dimensions of the point sources assumed in this work were not clearly described producing doubts over the aforementioned results. Nonetheless this spatial resolution is still insufficient for use in thin tissue β- autoradiography.

The limiting issue with using scintillators is the effect of the diffusion of the secondary electrons (see Section 3.1). Although this is a known issue, it has to be highlighted the remarkable claimed spatial resolution obtained with the μ-Imager (15 μm for ^{3}H). This is without doubt the best spatial resolution reported and thus, an aim of this study to be as close as possible to this figure.

2.2 Direct Detection Methods

2.2.1 Gaseous Detectors

Gaseous detectors are based on the principle of applying an electric field across a gas volume sensitive to incident ionising radiation. When a charged particle traverses a gas, free electrons and positive ions are generated. These electrons can be accelerated by an electric field producing secondary ionisation in an avalanche process, wherein a gain of x10^3-10^6 is obtained [22]. This is the basis of a proportional gas chamber. The focus of the research with this type of technology has been on developing a multiplicity of geometries for drifting, amplifying and collecting this charge, principally for applications in high energy physics.

In the sixties, charged particle detection in particle physics was mainly undertaken by examining millions of photographs from bubble chambers or spark chambers. In 1968 Georges Charpak, who joined CERN in 1959, revolutionised the particle detection community by inventing the Multiwire Proportional Chamber (MWPC) [33]. This device is comprised of a mesh of equally spaced thin wires (anode) between two orthogonal cathode planes enclosed in a gas mixture (of e.g. argon and ethane), as shown in Figure 2.5. The anode wires are grounded and the cathode planes have a negative voltage. The electrons that traverse the field will therefore be attracted towards the anode wires. As the field is higher near the wires the attracted electrons have an avalanche effect as they approach the wires. This development has spawned several attempts to harness this approach to autoradiography [14, 109], and ultimately led to Charpak's Nobel Prize in Physics in 1992, for services to Physics and the invention of the MWPC.

This idea has spawned several commercial systems such as the InstantImager[1], which is based on a microchannel array detector (MICAD) [137] and a MWPC. The electrons are detected by a microchannel array consisting of an array of holes, in which a drift field is established. Electrons (betas) interacting in the gas within the holes are multiplied

---

[1] Packard Instruments, Meriden, USA

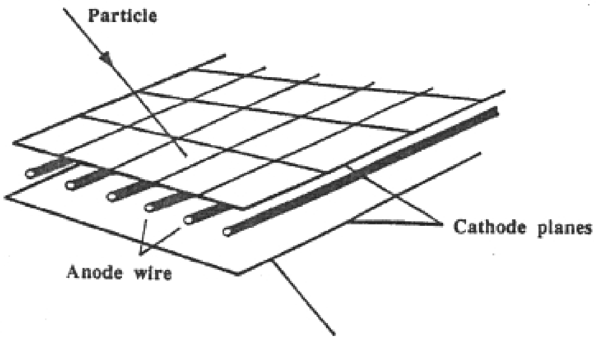


Figure 2.5: Schematic diagram of a MWPC showing the cathode plans and the anode wires. The chamber is filled with a gas mixture of argon and ethane [127].

by an argon based gas mixture. Then this shower of electrons is attracted towards an anode wire within an MWPC located beneath, where an avalanche is generated. Only results with ^{14}C and ^{32}P have been reported with a spatial resolution of 400 μm for ^{14}C. An example of a radiogram obtained with the InstantImager with ^{99m}Tc after 20 hours is shown in Figure 2.6 [137]. This system is no longer commercially available, in part due to its poor spatial resolution, and the difficulties in producing reliable gas mixtures in a life science setting.

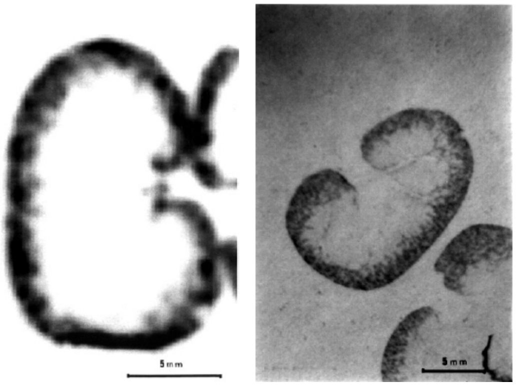


Figure 2.6: Radiogram of ^{99m}Tc in a rat kidney obtained with the InstantImager after 20 hours of exposure time (left) and adjacent section exposed to film after 1 day (right) [137].

More recently, Charpak, [168] has described the use of a gaseous detector to amplify emitted β particles. However, instead of using the charge to determine location, the scintillation light produced during the amplification process was detected and localised using a high performance CCD camera. Further developments with MWPC technology led to a commercial system known as the β-Imager (β-Imager 2000^{TM}), developed by BiospaceLab, based on a parallel plate avalanche chamber within which the sample is placed, as shown in Figure 2.7. Results using ^{3}H (Figure 2.8) claiming 50 μm of spatial resolution [12], and unrivalled sensitivity of ~80.00 10^{-6} cps mm^{-2} kBq^{-1} g (computed from [168]), have been published using this system.

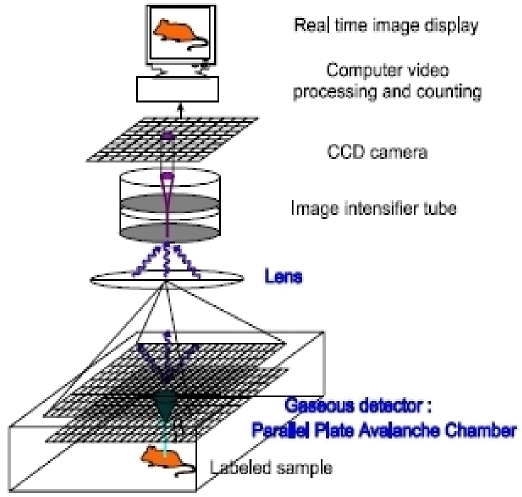


Figure 2.7: General operating principles of β-Imager [12].

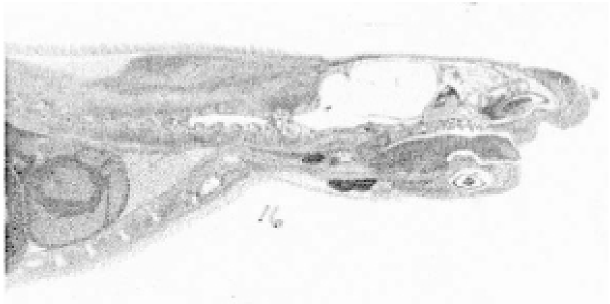


Figure 2.8: Whole body section imaging with β imager and ^{3}H labelling in a rat. Detection area is of 75mmx100mm with 8 h acquisition [12].

The main problem of all the systems based on MWPC is the unacceptable spatial resolution obtained due to the diffusion of the secondary electrons in the gas chamber, this being a critical problem due to the sub-millimetric (and ideally micron-level) spatial resolution demands of routine autoradiography studies.

2.2.2 Microchannel Plates

Microchannel Plate Detectors (MCP) were originally developed as amplification elements for image intensification. The first MCPs were implemented in USSR, USA and UK in the early sixties almost simultaneously [46]. The basic architecture of an MCP is constructed from a planar slab of glass, 2 mm thick usually, made of a high resistive material such as lead glass, tessellated with an array of tiny tubes (microchannels) that behave as electron multipliers (see Figure 2.9). A voltage in the range of 400-2000 volts is applied between the two extremes, which establishes a uniform electric field inside the channels. Most modern MCP detectors consist of two MCPs in a chevron (v-like) shape. The MCPs are usually placed in a vacuum chamber before applying the voltage between the two ends in order to reduce the noise inside the channels. This also helps to extend the lifetime of the MCPs. These are parallel to one another and typically biased at a small angle (~8°) away from the perpendicular face. Typical hole diameters are 10-100 μm with an approximate distance between them of 15 μm. The charge is amplified twice offering higher gain compared to the conventional straight configuration MCP. If a particulate radioactive source is placed close enough to the array of microchannels (of the order of $\sim \mu$m), the emitted particles (β electrons) hit the wall of the channels. The impacts generate showers of electrons that propagate through the channels, and thus the microchannels can be considered as continuous dynodes. There are several different read out possibilities used on the microchannels to detect the amplified β electrons such as photomultiplier tubes or a single segmented metal anode. The spatial resolution of this technique is clearly limited by the channel diameter and pitch, and the detection efficiency is strongly limited by the angle at which the tubes are tilted.

The first attempt of applying this technology to autoradiography was the β camera [106, 107]. This placed an MCP with a scintillator and a fibre-optic window between the source and the plate. The best resolution achieved by this system was ∼500 μm and was only tested with medium energy radioisotopes such as ^{14}C and ^{201}Tl in brain tissue sections. No results were reported with low energy radionuclides such as ^{3}H. The spatial resolution shown by this system is inferior to the minimum needs of autoradiography to replace conventional film. Nonetheless this extensive pioneering work undertaken by Ljunggren and Strand [106, 107] demonstrated the potential of this technology for replacing film.

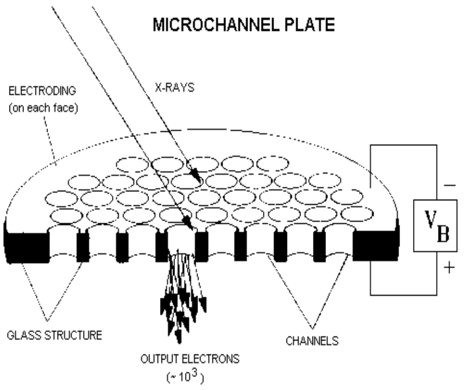


Figure 2.9: Schematic of a single channel of a MCP showing the mechanism of secondary electron production [46].

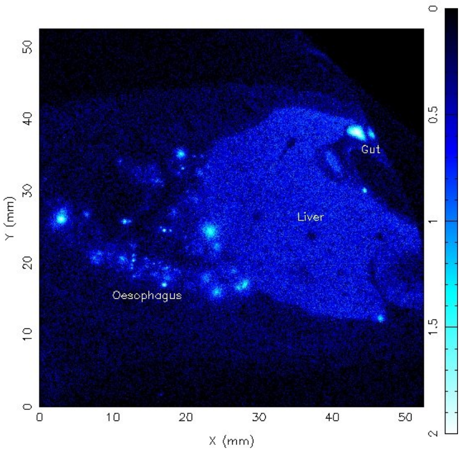


Figure 2.10: Whole body section of ^{14}C labelled rat exposed for 2 hours to an MCP. Note that the intensity values are shown in logarithmic scale. The spatial resolution measured in this section is ∼80 μm [98].

Other approaches using low noise MCPs applied to autoradiography have been extensively investigated for different radioisotopes such as ^{3}H [99, 101], ^{14}C [97] and other high energy radioisotopes such as ^{199m}Tc and ^{188}Re [100] . The best spatial resolution published with this technology so far is ∼60 μm using an X-ray source [98]. An example of a whole body rat section labelled with ^{14}C and exposed to an MCP for 2 hours is shown in Figure 2.10. It is indicated in this experiment how the spatial resolution (∼80 μm) is too low for this application. The sensitivity, measured from data published in [98], is estimated at ∼6 10^{-3} cps mm^{-2} kBq^{-1} g measured with a ^{14}C microscale, and ∼24 10^{-6} cps mm^{-2} kBq^{-1} g measured with a ^{3}H microscale.

A way to improve the spatial resolution with a MCP is the reduction of diameter of the microchannels and the reduction of pitch between them. One approach to increase the detection efficiency is including a *CsI* coating (scintillator) in the input surface of the MCP. A factor of x2 for ^{3}H and x3 for ^{14}C in detection efficiency improvement are reported in [96] using such approach.

It has been clearly shown the strong dependence of spatial resolution with the microchannel diameter and the pitch. If smaller microchannels are used this will have a negative impact on the detection efficiency, given that less particles will be detected. On the other hand a thin sheet of scintillator can be placed on the input surface of the MCP to increase detection efficiency, but this will have a negative impact on the spatial resolution. These sensitivity measurements shown here demonstrate comparable figures to other alternatives in the literature, but the spatial resolution is still far below that of traditional film.

2.3 Silicon Technology

The use of silicon-based technology for autoradiography and other radiation detection and imaging applications has become a prolific activity in the physical sciences.

The physical properties of silicon make this material highly attractive for sensing ionising radiation. Especially due to its low ionising energy to create one e-/hole pair (3.6 eV), compared to that in film (7 eV) [188].

The way in which the generated charge within the silicon is collected and read out from the sensor is where most of the research has been carried out. A variety of different silicon-based technologies can be found in the imaging arena, each fighting for better performance and a stronger place in the competition for scientific and commercial imaging applications. CCDs are the most popular choice as imaging sensors at the moment, as their performance currently represents the *gold standard* in solid-state digital imaging. Nevertheless there is a growing field of research being undertaken using CMOS technology which is starting to threaten the dominant position of CCD technology. During the course of this research, E2V Technologies plc, the UK main manufacturer of scientific CCD detectors has now started to release its first CMOS imaging sensors in late 2008. Besides CCD and CMOS technologies, there exist other alternatives such as DEPFET, silicon strip detectors and hybrid technology, initially designed for high energy physics projects, but also applied to medical applications. As there is such a high level of activity in this area, this aspect of direct imaging performance has been given its own subsection.

2.3.1 DEPFET pixel BIOSCOPE system

The Depleted Field Effect Transistor (DEPFET) pixel Bioscope is based on a 64x64 array of DEPFET pixels with two versions, square 50 μm x 50 μm pixel pitch and hexagonal 50 μm x 42 μm pixel pitch and with 3.2 x 3.2 mm^2 of sensitive area [173].

The structure from bottom to top, shown in Figure 2.11, is comprised of a layer of SiO_2 (30 nm thick), nitride (100 nm thick) and a shallow p+ implantation. Summing up the window entrance is only 200 nm thick. Beneath the entrance window is the fully depleted substrate of the sensor which is 300 μm thick, here is where the ionising particles generate the charge due to ionisation while they traverse the sensor, and are quickly collected by the internal gate. Finally, beneath the bulk is a field effect transistor (FET). The charge carriers (electrons) originated by the ionising particle in the substrate are collected by the gate of the transistor (internal gate in Figure 2.11) thus modulating the transistor current. By sensing this current the information about the impinged particle can be extracted. This measurement has to be undertaken under a very low noise read out scheme because the standing transistor current is 100 μA and charge of 100 electrons will originate a change of only 10 nA in this current. The total noise is 225 e- for the total read out chain [147]. In order to reduce the input capacitance to reduce the thermal noise of the device, the gate is split in two, one on the surface, as usual, and an internal gate to collect all the charge.

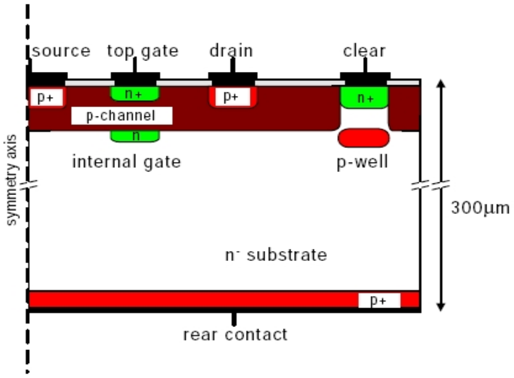


Figure 2.11: Cross section of a DEPFET pixel. The structure is symmetric along the symmetry axis shown [125].

The results published in [125] demonstrate the capability of this device at detecting ^{3}H due to the thin entrance of the rear side (200 nm), as shown in Figure 2.12 [173].

Figure 2.12: Optical image of a leaf labelled with ^{3}H (left) and autoradiogram of the same leaf (right) after 2 hours exposure [173].

An experiment using a ^{55}Fe source and using the η-function reconstruction algorithm [17] produced a measured spatial resolution for 6 keV γ photons of 6.7±0.7 μm [173].

By extension, the authors claim that the achievable resolution with ^{3}H is ~7 μm but no experimental results proving this resolution have, as yet, been presented.

The lack of biological results with ^{3}H demonstrating the excellent spatial resolution claimed by [173], and the very specialised process necessary to make this detector are the two main problems to use this approach as a general method for autoradiography.

2.3.2 Silicon strip detectors

A silicon strip detector is a detector comprised of an n-type bulk, as usual for silicon-based detectors, and an aluminium contact on the rear side. An n-type-based semiconductor has better conductivity and longer lifetime than a p-type-based semiconductor. The front face is comprised of aluminium strips and p-type silicon strips. These strips are separated between them with a thin layer of insulator. When a charged particle traverses the detector it generates charge in the bulk. The positive carriers (holes) drift towards the negatively charged p-type silicon strips at the top of the sensor whereas the negative carriers drift towards the positively charged aluminium contact at the back of the detector. When the holes, or positive carriers, reach the p-type strips this charge is measured through sensitive electronic read out channels.

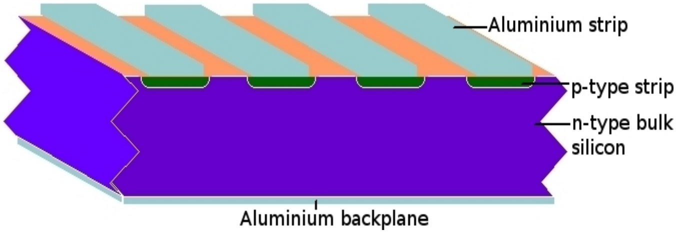


Figure 2.13: Typical structure of a single sided silicon strip detector [158].

Read-out strips can be added on the other side of the detector perpendicular to the existing strips on the top to produce 2D information. These are then called Double Sided Silicon Strip Detectors (DSSSD). Therefore when a ionising particle deposits charge in the bulk of the detector the orthogonal coordinates of the impact can be read out. The spatial resolution is strongly dependant on the separation between strips.

One of the first application of silicon strip detectors to autoradiography can be found in [155], where a 1D silicon strip detector was used. To gain 2D information the source was rotated automatically on the sensor and a subsequent reconstruction using Fore Back Projection was undertaken.

Use of double sided silicon strip detectors (DSSSD), due to the availability of 2D information, has been more extensively investigated. A DSSSD with a 300 μm thick bulk and 100 μm pitch between strips is presented in [18]. ^{32}P labelled human mammary epithelial cells were imaged claiming a relative efficiency of 20% in comparison to a scintillator counter (considered as ground truth, i.e. what this measured is considered to be the absolute activity). In [131] a 300 μm thick DSSSD with 50 μm pitch on both p and n sides was studied as part of the Bioscope system. This sensor has 640 strips on each side having an effective area of 32 x 32 mm^2. This study presented results obtained with ^{45}Ca, ^{35}S and ^{14}C (none of them biological samples), claiming a spatial resolution below 50 μm, as the best possible with this detector. In [90] the use of DSSSD is demonstrated to have good energy resolution being able to distinguish two drops of ^{33}P and ^{35}S (Figure 2.14).

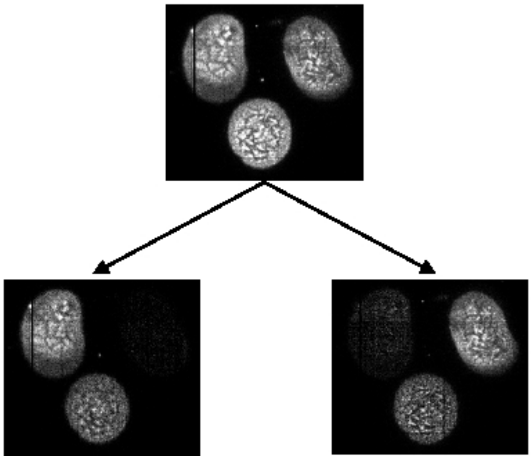


Figure 2.14: Autoradiogram of a drop of ^{33}P (top figure left corner), a drop of ^{35}S (top figure right corner) and a drop of a mixture of ^{33}P and ^{35}S (top figure bottom), same autoradiogram discarding energies out of the range corresponding to ^{33}P (bottom left figure) and same autoradiogram discarding energies out of the range corresponding to ^{35}S (bottom right figure) [90].

A more recent work uses silicon strip detectors to image serial sections of activity volumes, in order to image radiolabelled antibodies [129]. This work studied the non-uniform uptake of anti-CD20 and anti-CEA antibodies labelled with ^{131}I in tumours (Figure 2.15).

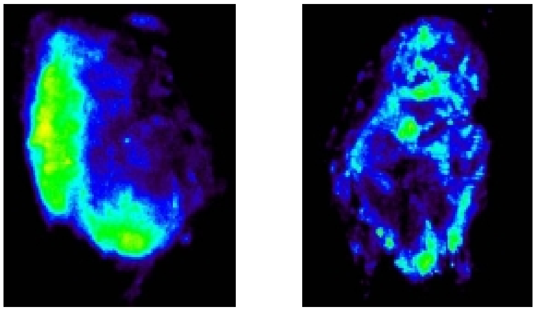


Figure 2.15: Reconstructed activity volumes of CD20 expressing tumours, acquired with a silicon strip detector. Left: 11h post injection of ^{131}I-labelled minibody, max tissue diameter: 10 mm. Right: 48h post injection of ^{131}I-labelled intact antibody, max tissue diameter: 25 mm [129].

Silicon strip detectors have been shifted out of the arena for autoradiography imaging due to the greater acceptance of CCD technology and entrance in the field of CMOS technology for scientific applications. The spatial resolution is again the main drawback of this approach being much easier to manufacture small pixels than narrow strips.

2.3.3 Hybrid CMOS technology

Hybrid CMOS technology is based on the concept of separation of the semiconductor sensor and the read out circuitry. There are some techniques for bonding such as direct bonding and bump bonding (see Figure 2.16). The most common way to connect these two different devices is bump bonding [20]. The pixel pitch strongly depends on the method used to grow the array of bumps and the material of these. This architecture presents some advantages compared to monolithic integration as higher fill factor and possible optimisation of both parts separately. For instance the semiconductor part can be changed depending on the application (silicon, GaAs, CdTe, CdZn for X-ray, SiC for UV or PbSe for infrared radiation) whereas the read out electronics is the same.

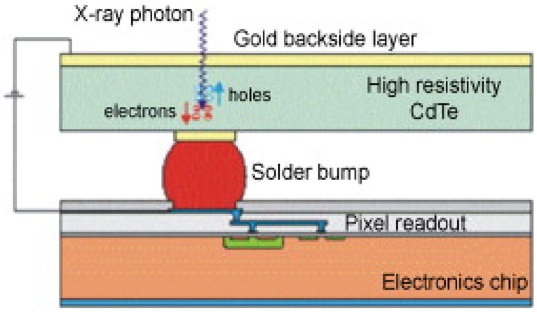


Figure 2.16: Example of bump bonding.

The most significant exponent of this technology is the Medipix consortium (CERN, Geneva, Switzerland) that has produced Medipix1 (Betaview) and Medipix2 (Timepix and Medipix3 are under development), characterised by an excellent low noise performance.

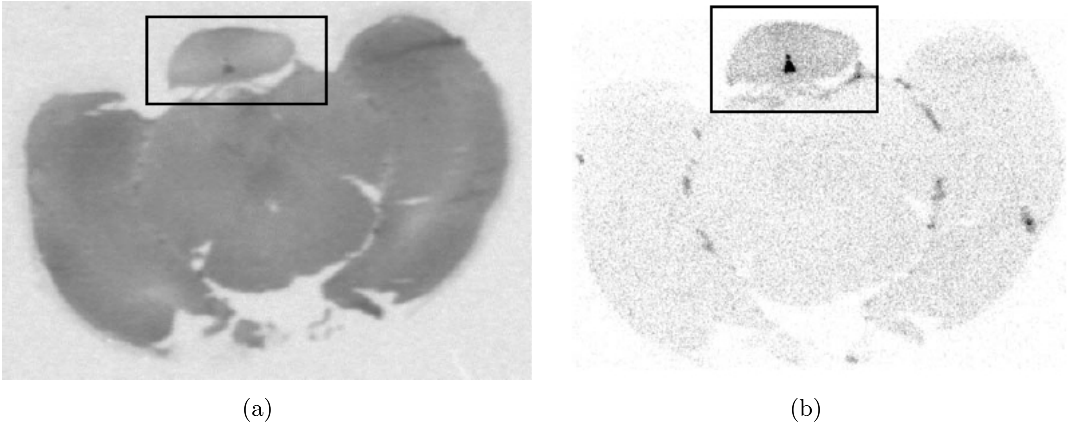
Medipix1 [1, 19] is a 64x64 array of pixels 150 μm size square pixels and a pixel separation of 20 μm (170 μm pixel size). The semiconductor used, GaAs or Si, is 200-300 μm thick and is bump bonded to a single-Particle Counting Chip (PCC). This design has in each cell a charge amplifier, a pixel-by-pixel adjustable discriminator to adjust the uniformity of the images and a 15-bit counter. GaAs is used as semiconductor in [1], where the frame rate is set at 10 mins/frame due to its poor sensitivity. The threshold to detect β events is set to 25-35 keV to keep the noise count rate low, so that ^{14}C, ^{35}S and ^{32}P imaging is possible. The background count rate measured with these settings is 2.50 10^{-3} cps mm^{-2}. On the other hand [19] used silicon-based Medipix1 to detect β electrons emitted from a ^{14}C source. In this case the threshold was decreased to 15 keV getting a lower background count rate of 0.35 10^{-3} cps mm^{-2} and a higher efficiency, 12%, compared 10% obtained with GaAs [1].

Medipix2 [118, 119, 120] is a 256x256 array of pixels with a 55 μm pixel pitch, hence the sensitive area is 14x14 mm^2. The semiconductor is a 300 μm thick slab of silicon and is Indium bump bonded to the counting chip. Each cell has a charge amplifier, a double threshold discriminator and a 13-bit counter. In [118] standard microscales were used to calibrate the sensor with a global threshold of 7 keV. The microscales consist of known amounts of radioactivity homogeneously distributed in a plastic tissue equivalent polymer of 120 μm thickness. Each microscale has 8 cells of varying activity from 1.13 kBq/g to 31.86 kBq/g for the low activity microscale (RPA504) and from 0.00381 kBq/g to 3.7740 kBq/g for the high activity microscale (RPA511) (GE Healthcare[1]).

The potential of this approach is tremendous, given the flexibility of this design, being possible to optimise the electronics independently on the detector semiconductor used. Future versions of Medipix will probably represent a good choice for autoradiography, but the very specialised process necessary to make these detectors and the high cost of the facilities necessary do not make this approach a good alternative to be extended in the autoradiography community.

2.3.4 Charge Coupled Devices technology

The application of CCD technology to autoradiography has already been presented using indirect approaches in [81, 12], as explained in the Section 2.1.1. In contrast, the first direct detection approach using CCDs was demonstrated by [130]. Direct irradiation was used under cooled conditions, using a cryostat system with liquid nitrogen. Results with tissue labelled with ^{14}C-deoxyglucose are shown in Figure 2.17 demonstrating higher dynamic range compared to film.


(a)

(b)

Figure 2.17: Images of mouse brain sections labelled with ^{14}[C]deoxyglucose obtained with film after three weeks exposure (a), and by the cooled CCD-based system at -150 C (121 K) after 24 hours exposure (b). The CCD image defines the distribution of the labelled glucose more accurately than the film autoradiogram which seems to show saturation in the region of the hypothalamus, enclosed in a square in the figure [130].

This work was further developed to extend its use to room temperature imaging [88] with ^{35}S, as shown in Figure 2.18, after fixed pattern noise is removed by using simple

---

[1] GE Healthcare UK Limited, Pollards Wood, Nightingales Lane, Chalfont St.Giles, Bucks, UK, HP8 4SP

image correction schemes. The autoradiograms were acquired at x36 times faster than film, but with almost an order of magnitude reduction in sensitivity compared to low temperature performance observed in [130]. Some corresponding structures in both autoradiograms are clearly visible in Figure 2.18 (e.g. see boxed region). However although this work represented the first attempt to produce direct autoradiograms at room temperature using CCD technology, the image quality is far inferior for that needed for routine use in life sciences applications. It may have been that this simply required longer exposure times, as total exposure time was limited by the hard disk storage available at the time.

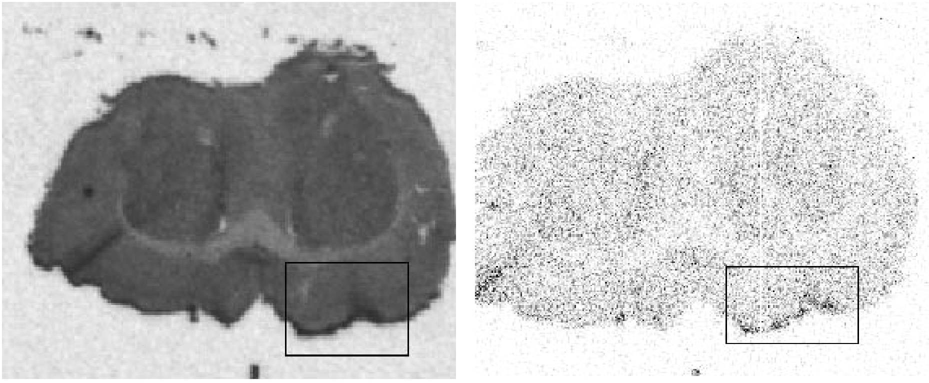

(a) (b)

Figure 2.18: Images of mouse brain sections labelled with ^{35}S obtained with film after four days exposure (a), and by the CCD-based system at 22 C after 2 hours exposure (b), after correction of dark current and fixed pattern noise. The squares indicate where selective local hot spots are more easily distinguished in the CCD system, compared to the film autoradiograms [88].

2.3.5 Complementary Metal-Oxide-Semiconductor technology

CMOS technology is still an emerging technology in scientific applications, where CCD technology is still considered to be a gold standard. CMOS technology is at the moment dominating the consumer-based low-cost imaging market, due to the ability of CMOS manufacturers to produce CMOS imaging sensors at the same rate as CMOS circuitry, as they are both based on the same standard fabrication processes. It is being used in medical imaging in X-ray detection with the use of scintillators such as Cs-I or Gadox [9, 23, 16]. A significant example this technology is starting to be used in mammography where CMOS technology is the only technology that can be easily packaged like a mammography film cassette without cooling, and also because of its low power consumption [16] (see Figure 2.19).

Typically detectors used in Computed Tomography (CT) have been flat panels, based on amorphous silicon or amorphous selenium, or CCD imaging sensors. CMOS technology is itself a low-cost alternative compared to other approaches. There are arguments relative to the sensitive area and the spatial resolution achievable by CMOS and CCD technologies, but undoubtedly both technologies offer competitive performance. The

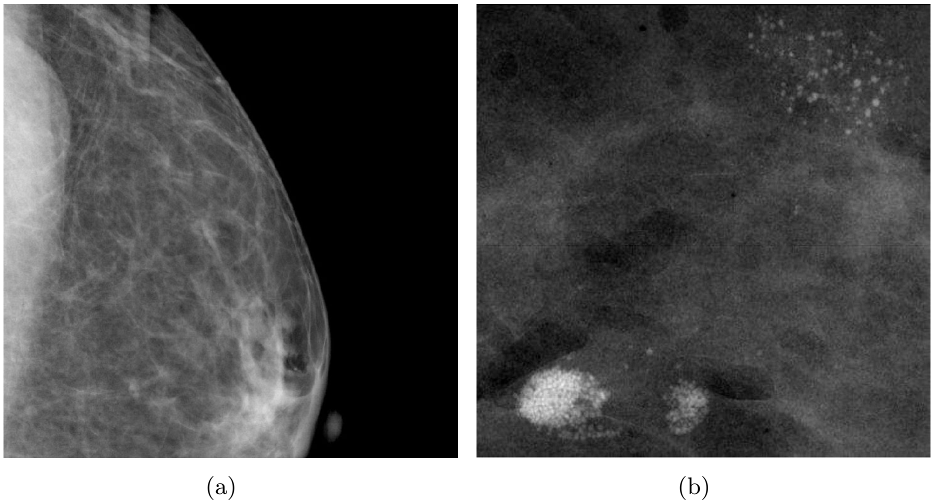

(a) (b)

Figure 2.19: Mammogram acquired with a digital sensor comprised of eight (2 x 4) CMOS tiled sensors (48 μm pixel size) with 100mm x 100mm field of view (a), and surgical 15mm x 20mm biopsy specimen acquired with the same CMOS sensor where dense (bottom left corner) and subtle (top right corner) calcifications are observed (b) [16].

low cost of CMOS technology makes this alternative very attractive to be used on regular basis in CT [94, 29, 185].

The use of this technology in autoradiography is still scant. MIMOSA [171] and Vanilla [31] are the only CMOS MAPS sensors applied to autoradiography published in the literature so far. In [43] MIMOSA V demonstrated its capability for detecting ^{3}H under cooled conditions but to date, no biological assessment has been published. The noise of this sensor is impressive compared to other non-CMOS sensors applied to autoradiography [1] but the sensitivity to detect ^{3}H is low.

A review of the most important alternatives applied to autoradiography has been exposed above. Silicon-based detectors are a very active area of research, thus a significant variety of designs is found in the literature. Some of these designs were initially design for high energy physics, therefore they have been recycled to be used in medical imaging.

In summary, hybrid technology has demonstrated high sensitivity and low noise, but the spatial resolution is still not good enough for thin tissue autoradiography. Moreover this is an expensive technology making it unattractive for large area development. Silicon strip technology, although being commercially available, again does not offer the spatial resolution that autoradiography demands. DEPFET has been demonstrated to have high spatial resolution and low intrinsic noise, but work on sensitivity has not been published yet to be able to compare with other approaches. In this regard, one may consider its applicability to be unproven, but not discounted. CCD and CMOS technologies are the most prolific approaches used by the scientific community at the moment. It has already been mentioned the main differences between these technologies, highlighting that CMOS technology is starting to reach a similar performance compared to CCD technology, but this is still a matter for debate. Among the most important advantages of CMOS technology compared with CCD technology are high-speed imaging and low cost, therefore most of the R&D work at the moment is undertaken using CMOS technology. Part of this work will examine the relative merits of both of these competing technologies for thin tissue autoradiography.

Chapter 3

Digital Autoradiography Imaging Principles with Solid-State Silicon Detectors

3.1 Signal Generation in Silicon-based Detectors

When an ionising particle is incident on a silicon-based imaging sensor, it may be either backscattered or absorbed in the overlying (insensitive) layers of the detector, traverse through these layers until it reaches the sensitive part of the detector, or traverse the whole detector without interacting, depending on the kinetic energy of the particle.

Assuming a standard front illumination geometry, the top layers of the detector are typically comprised of protective layers of SiO_2 and Si_3N_4 and interlaced electrodes of polysilicon and SiO_2 layers in CCD detectors (∼1 μm thick) [74], or protective layers of SiO_2 and Si_3N_4 in CMOS detectors (∼1-5 μm thick). The latter also includes a mixture of transistors (placed in a corner of the pixel) and interconnection lines. These are also commonly called dead layers because they are not sensitive to electromagnetic radiation. If the particle reaches the sensitive epitaxial layer it may interact with the valence electrons of the silicon lattice and/or with the silicon atoms nuclei (see Figure 3.2(a)). If the impinging particle has sufficient kinetic energy, it may produce ionisation in the silicon, generating charge carriers at a mean rate of ∼3.6 eV/electron, which corresponds to three times the energy band gap of silicon [52]. The sensitive region is usually called the epitaxial layer, referring to the method used to make it, by depositing monocrystalline film on a monocrystalline substrate (from Greek *above in ordered manner*). In CCD technology, this is divided between the depleted layer, affected by an electric field, and the undepleted field free layer, while in CMOS technology the entire epitaxial layer is field free. Figure 3.1 shows typical electron ranges induced in silicon for β- autoradiography. These energies cover the average and maximum energies of the most typical radioisotopes used in β- autoradiography, 5-1710 keV. These electron ranges were obtained by simulating a monoenergetic point source of a specific radioisotope in a block of silicon, being the electron ranges expressed as the straight line distances between the initial and final points of the trajectories of each simulated electron.

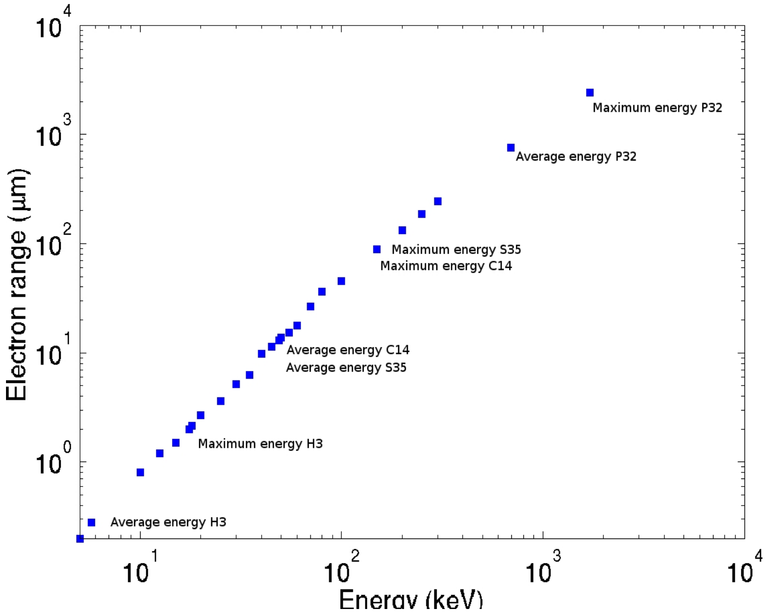


Figure 3.1: Electron range obtained from Monte Carlo simulations for a range of energies 5-1710 keV, covering the average and maximum energies of ^{3}H, ^{14}C, ^{35}S and ^{32}P, typical radioisotopes used in β- autoradiography.

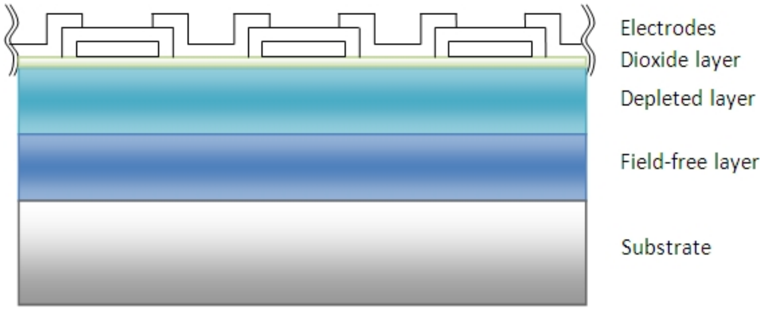
To maximise sensitivity, it is necessary to minimise absorption, back-scattering and transmission losses of the incident radiation field. These processes depend on the physical characteristics of silicon and the detector dead layer composition [75].

One method used to improve optical sensitivity in the near IR and UV spectrum in CCD and CMOS technologies, is to use a back-illuminated back-thinned geometry. Some groups are starting to apply this fabrication process to CMOS detectors [43, 15, 31], but it is still emerging. This fabrication process is based on bonding a new substrate (purely for mechanical support) on top of the front face and removing the underlying structures of the sensor below the sensitive volume. This process is graphically shown in Figure 3.2.

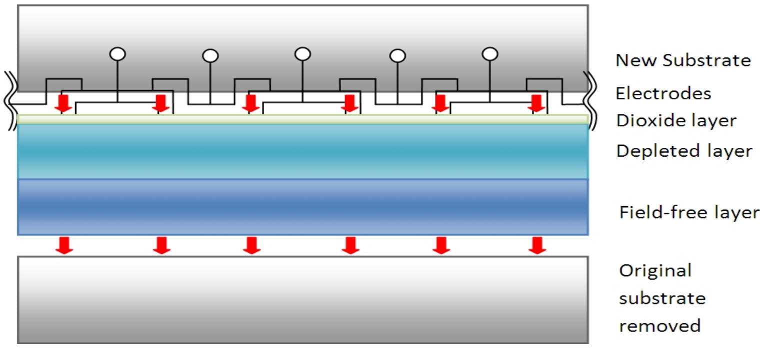
Electrons liberated from ionised silicon atoms mentioned above, initially generate a spherical electron cloud with initial radius σ_i, defined by equation 3.1 [167]:

$$\sigma_i = 0.0062E^{1.75}\mu m \tag{3.1}$$

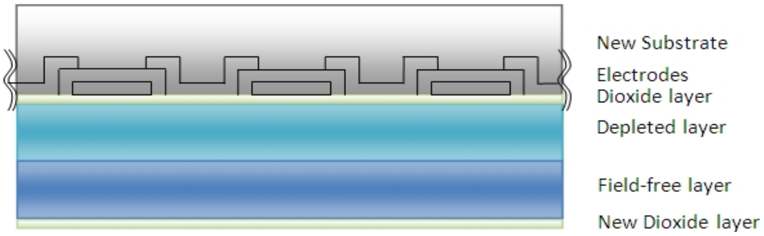
This parameter is derived from the well known energy/depth relationship $R = kE^n$, where E is energy in keV, R is range in μm and k and n are constants dependant on the material [51]. This charge may suffer two different effects: (1) In presence of an electric field the deposited charge drifts towards the charge collection points, i.e. photodiodes or photogates. (2) Deposited charge also suffers thermal diffusion. Depending on the



(a) Cross-section of a typical front-illuminated CCD detector



(b) Cross-section of a typical front-illuminated CCD detector, showing the new substrate bonded on top and the removal of the original substrate on the back side.



(c) Cross-section of a typical back-illuminated CCD detector, with a thin (~100 nm) dioxide layer on the back side. This is a residual layer produced by the chemical etching process applied to remove the substrate.

Figure 3.2: Process of back-thinning showing the initial typical cross-section of a front-illuminated device (a), the fabrication processes applied to the detector (b) and the final cross-section produced after back-thinning (c).

distance to be travelled, doping level and temperature, the degree of diffusion may also affect the final destination of the charge cloud, and may spread to neighbouring pixels.

In the case of CCD technology, the charge generated in the depletion region (affected by an electric field) drifts quickly towards the closest polarised gate of a pixel, confining most of the charge to the pixel where the charge was generated. The magnitude of the electric field is depth dependant, as shown in equation 3.2 [134]:

$$E(z) = \frac{qN}{\varepsilon}(z - z_d) \tag{3.2}$$

where ε is the electric permittivity of silicon ($1.044\ 10^{-12}$ F cm^{-1}), q is the elementary charge (C), N is the doping level (cm^{-3}), z_d is the thickness of the depleted layer and z is the silicon depth of the sensitive region. The depth of the depletion zone is governed by the doping level and the voltage applied, and it is defined by equation 3.3 [134]:

$$z_d = \sqrt{\frac{2V\varepsilon}{qN}} \tag{3.3}$$

where V is the mean of the surface voltage.

This charge deposited in the depletion region also suffers a low component of diffusion, defined by the standard deviation of a Gaussian projected on the collection point σ_d, mathematically expressed as shown in equation 3.4 [86]:

$$\sigma_d = 4\sqrt{\frac{\varepsilon kT}{q^2N} ln \frac{z_d}{z_d - z}} \tag{3.4}$$

where k is the Boltzmann constant ($8.62\ 10^{-5}$ eV/K), T is the temperature (K) and z is the depth of interaction. The most important parameter here is N the doping level, showing that higher doping levels reduces the thickness of the depleted layer (equation 3.3) and increases the electric field (equation 3.2), therefore producing a smaller charge cloud diameter. For example, the variation of σ_d with the depth of interaction in a 7 μm thick silicon depletion layer, for a wide range of doping levels, is shown in Figure 3.3. It has to be noted that the ordinate axis is presented in logarithmic scale. Typical doping levels for CCD sensors is $\sim 10^{12}$ cm^{-3} [74], and for CMOS sensors is $\sim 10^{15}$ cm^{-3} [44].

On the other hand, the charge generated in the field free layer diffuses until it eventually recombines or reaches the depleted region. A similar process is suffered by charge in the epitaxial layer of a CMOS detector due to the low (or non-existent) electric field, where deposited charge suffers diffusion until it recombines or reaches a collection point.

As mentioned above, if the deposited charge is not collected it eventually thermally recombines. This process is governed by the collection time t_c and the life time or recombination time t_r. If t_r is long enough ($t_c < t_r$) the charge will be collected before it recombines. This is controlled by the doping level in the silicon, with higher doping

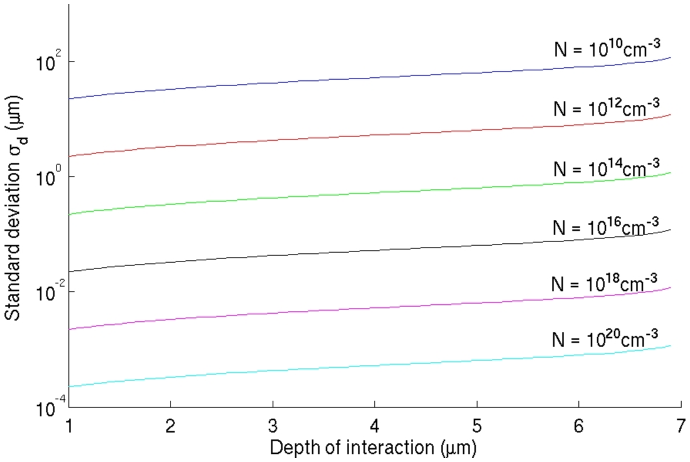


Figure 3.3: Variation of σ_d with the depth of interaction in a 7 μm thick depletion layer for different values of impurities concentration.

levels the recombination time is shorter. This is why if there is charge deposition in the substrate, that usually has very high doping levels, it recombines quickly. The average length that the charge travels before it recombines is called diffusion length [54]. The diffusion length is governed by the diffusion coefficient and the recombination time, as shown in equation 3.5:

$$L = \sqrt{D\, t_r} \tag{3.5}$$

where L is the diffusion length, D is the diffusion coefficient and t_r is the recombination life time, usually of the order of several hundred μs in today's epitaxial layers [154]. This parameter is usually measured experimentally [154, 3]. The diffusion coefficient D depends on the mobility as demonstrated by the Einstein relation shown in equation 3.6, which ultimately depends on the doping level as shown in equation 3.7 [54].

$$D = \frac{kT}{q}\mu \tag{3.6}$$

where μ is the mobility of the minority carriers (cm^2/V s), i.e. electrons in this case where the sensitive layer is usually p+ doped with acceptor dopants such as Boron. The term $\frac{kT}{q}$ is typically given a value of 25 mV at room temperature.

Some studies have tried to establish an empirical correspondence between majority carriers mobility and doping level. The mobility of the minority carriers can be approximated by that for majority carriers with the same doping density. An accepted empirical formula is shown in equation 3.7 [175]:

$$\mu = \frac{\mu_{max} - \mu_{min}}{1 + (\frac{N}{N_r})^{\alpha}} + \mu_{min} \tag{3.7}$$

where μ_{min}, μ_{max}, α and N_r are fitting parameters. A table with values for Boron (p type) and Phosphorus (n type) for the different fitting parameters is shown in Table 3.1. A graph with results obtained by different studies of majority carriers mobility vs. doping level for Boron in silicon at room temperature is shown in Figure 3.4.

Table 3.1: Parameters for calculation of the mobility as a function of the doping density

	Boron	**Phosphorus**
μ_{min}	44.9	68.5
μ_{max}	470.5	1414
N_r	$2.23\ 10^{17}$	$9.20\ 10^{16}$
α	0.719	0.711

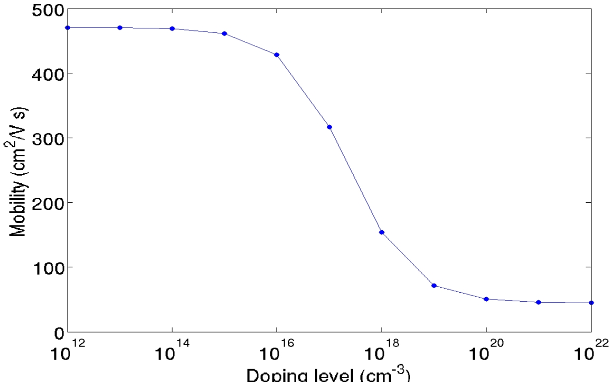


Figure 3.4: Variation of carriers mobility (cm^2/V s) in the ordinate axis with dopant level (atoms/cm^{-3}) in the abscissa axis for Boron in silicon at room temperature (300K), given by equation 3.7.

The depleted region is to first order, absent of free carriers, therefore the charge generated in the depleted region does not recombine. It drifts to the collection points due the electric field as mentioned above. In those regions where the electric field is too low or non-existent the charge will diffuse as explained above. This charge diffuses over 4π being possible to approximate the projection on the depleted region to a Gaussian shape [78]. By analogy the same process is described for charge deposited in the epitaxial region in a CMOS detector. A well accepted model for the width of this Gaussian shape σ_{ff} is defined empirically by equation 3.8, presented for first time in the seminal paper [78], and quoted by other authors [25, 115, 134]. This equation was obtained by using Monte Carlo simulations, where 2000 electrons randomly wander in silicon until they are collected in the depleted layer. In this work different absorption depths are simulated obtaining different cloud diameters.

$$\sigma_{ff} = \frac{z_{ff}}{2}\sqrt{1-(\frac{z_a}{z_{ff}})^2} \tag{3.8}$$

where z_{ff} is the thickness of the field free layer and z_a is the distance between the z

coordinate of the interaction and the bottom of the field free region. It is observed from equation 3.8 how the worst case situation where $z_a = 0$ (produces the largest cloud diameter) which is when the charge is generated at the bottom of the field free layer, while the best case is when the charge is generated near the surface with the depleted region ($z_a = z_{ff}$), where the charge does not suffer field free diffusion. A graph showing the dependence of σ_{ff} with z_a (equation 3.8) in a 20 μm thick field free layer is shown in Figure 3.5

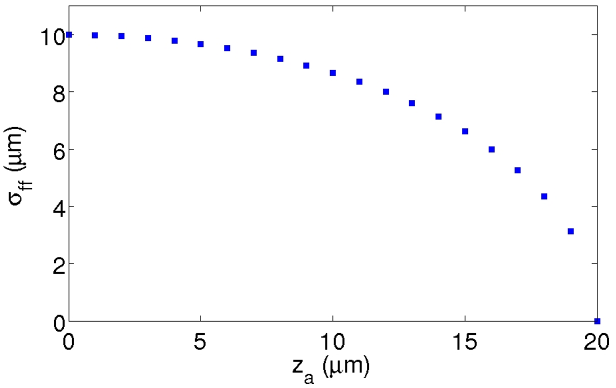


Figure 3.5: Evolution of the width of the Gaussian obtained in the field free region σ_{ff} with the distance between the point where the interaction took place and the rear side of the field free layer z_a, calculated in a 20 μm thick field free layer.

Subsequent work [70, 134, 25] undertook thorough analytical studies of charge diffusion, demonstrated that the assumption of a Gaussian distribution in the field free layer in fact underestimates charge in the tails of the distribution. Therefore, the real distribution affects more pixels than if a perfect Gaussian distribution were assumed. On the other hand the actual distribution has a more pronounced peak compared to a perfect Gaussian. This matter will be further discussed in Chapter 6.

The width of the final Gaussian distribution (σ_{tot}), taking into account the initial charge spread (σ_i), the diffusion in the depleted volume (σ_d) and the diffusion in the undepleted volume (σ_{ff}) is described by equation 3.9:

$$\sigma_{tot} = \sqrt{\sigma_i^2 + \sigma_d^2 + \sigma_{ff}^2} \tag{3.9}$$

An example of the Gaussian shape projections of charge diffused in a 20 μm thick active volume at several depths of interaction, from the upper surface to the bottom surface with the substrate, is shown in Figure 3.6. The standard deviation of the projected Gaussians on the top surface σ_{ff} are 2 μm, 4 μm, 6 μm, 8 μm and 10 μm from left to right, corresponding to depths of interaction 0.4 μm, 1.6 μm, 4 μm, 8 μm and 20 μm respectively.

Most scientific CCD imagers are fabricated on 30-50 Ω-cm epitaxial silicon (corresponding with 10^{15} cm^{-3} of doping level), but in order to mitigate the diffusion effect more and more CCD detectors are manufactured using high resistivity silicon wafers (~5-12

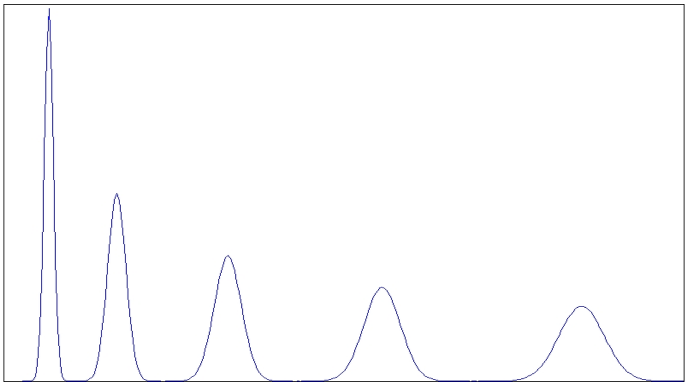


Figure 3.6: Example of several Gaussian shape projections of charge diffused in a 20 μm thick active volume at different depths of interaction, at 0.4 μm, 1.6 μm, 4 μm, 8 μm and 20 μm from left to right, with standard deviations 2 μm, 4 μm, 6 μm, 8 μm and 10 μm respectively.

KΩ-cm corresponding with $\sim 10^{12}$ cm^{-3} of doping level) [162]. This allows the electric field present in the depletion region, that usually reaches 7 to 10 μm deep in the silicon, to be extended. This will make the electrons drift towards the top of the layer, where the electrodes are, while the holes will be attracted towards the substrate, therefore mitigating diffusion or recombination. In CMOS technology low resistivity silicon is used so the electric field is very low compared to CCD technology reaching only to 1-2 μm (it can be considered null for certain studies) [75]. These physical processes define the charge collection efficiency of the detector, and have a direct impact in the spatial resolution achievable, depending on the doping level, thickness of the sensitive layer, pixel size and electric field in the case of CCD technology [148].

As has been observed there exists a wide variety of parameters that affect the performance of a silicon-based detector in different ways. Depending on the application a trade off has to be established to obtain the most suitable performance in each case. At the end of this chapter a discussion connecting the concepts explained in this section, with the different sources of noise and specific features of CCD and CMOS technologies is presented.

3.2 Sources of Noise in Silicon-based Detectors

Having discussed the principles of charge generation in silicon imaging detectors, it is necessary to also consider the different sources of noise found in CCD and CMOS technologies that stand in the way of obtaining excellent performance required in scientific applications.

3.2.1 Shot Noise

The arrival of ionising particles to a detector is a random process governed by a Poisson distribution. When the number of particles that reach the detector is small, this randomness adds a certain noise (i.e. uncertainty) component to the signal measured in a pixel given by Poisson statistics, as shown in equation 3.10:

$$\sigma_{shot} = \sqrt{S} \tag{3.10}$$

where S is the number of electrons that reach the detector and σ_{shot} is the shot noise component in electrons. Therefore this noise is relatively more significant when collecting small numbers of particles, such as the application considered in this work. There also exists dark current shot noise. This component represents the uncertainty of dark current generation in a pixel under low illumination conditions.

3.2.2 Thermal Noise

Also known as *dark current*, this is created by electrons that have jumped from the valence band to interface states and from these to the conduction band by means of thermal energy. These interface states are intermediate states between the valence and the conduction bands, generated by impurities or defects in the silicon's crystal lattice. The probability of this random process has a strong relation with temperature. These thermally generated electrons can be found, in a CCD sensor, in the bulk, in the depletion region and in the surface between the oxide and the silicon. These can be found mainly in the defective sidewalls and the edges of the photodiodes in a CMOS sensor. The relation between temperature and dark current is given by the equation 3.11 [74]:

$$D = RCA_pT^{1.5}e^{-E_g/2kT} \tag{3.11}$$

where D is the dark current (electrons/pixels/second), R is the dark current at 300K (usually specified by the manufacturer in electrons/secs cm^2), C is a unitless constant (2.5 10^{15}), A_p is the pixel area (cm^2/pixel), T is the absolute temperature (K) and E_g is the energy band gap, which is 1.12 eV at room temperature (300 K).

In CCD technology the typical approach to minimise this effect is by adding free carriers that occupy the interface states, often referred to as *inverted mode* operation. In an *inverted* CCD the potential in the substrate is higher than in the surface, therefore holes from the channel stop will migrate to the interface and hopping of electrons from the interface states to the conduction band will be inhibited [116]. This same approach has been adopted by some manufacturers of CMOS imagers. Another approach used in CMOS technology is reducing the number of crystal defects from the surface of the substrate by improving manufacturing processes. The most simple approach to reduce the effect of dark current is by cooling down the sensor, thus harnessing the exponential dependence on temperature (see equation 3.11).

3.2.3 Fixed Pattern Noise

Silicon wafers are fabricated using epitaxial growing. This process causes certain nonuniform doping profiles, i.e. crystal defects, in the silicon structure. These crystal defects and metal impurities act as generation and recombination sites for hole-electron pairs. This problem, and slight size variations in the pixel geometries, means that pixels across a sensor have generally nonuniform sensitivity, better known as pixel Fixed Pattern Noise (FPN).

In the absence of ionising particles, dark current at room temperature means unwanted electrons are generated and stored in a potential well. These are observed as isolated high intensity pixels or lines in an image. As these defects are unevenly distributed across the sensor, this produces nonuniformity in the image in terms of broken lines and *salt and pepper* noise.

This nonuniformity is approximately 1-2 % of the average signal in CCD and CMOS technologies, affecting in a similar way to both technologies. The advantage of CMOS technology is that, as the read out step is faster in CMOS detectors, the integration time can be much reduced in comparison to CCD sensors, therefore the FPN component should theoretically be less significant in CMOS sensors compared to CCDs.

It is also worth noting that there also exists column FPN, caused by read out circuitry (ADC, multiplexers,...) mismatches between different columns, however this is not considered in this work due to the minor effect observed in the sensors considered here.

3.2.4 Latch-up Effect

This name is used to describe the effect that appears when parasitic n-p-n and p-n-p junctions, present in the circuitry of the CMOS sensor as shown in Figure 3.7 [141], are triggered by undesirable currents between the transistors. These undesirable currents may appear when the sensitive layer of a CMOS detector is too thick. This will provoke malfunctioning in the detector and, if these currents become uncontrolled, thermal damage may occur. To avoid this problem CMOS detectors are not thicker than 20 μm. The latch-up effect is a well known problem present in integrated circuits (ICs) fabricated using standard CMOS processes, possibly causing catastrophic failures under extreme conditions.

The thin sensitive layer, necessary to avoid the latch-up, will concomitantly collect little charge compared to thicker detectors. This produces a concomitant problem, in that the noise that may mask the desired deposited charge (from an ionising particle in this case) in the thin sensitive area has to be low. This does not happen in CCD sensors as can be observed from the nature of this problem.

3.2.5 Read Noise

When the charge in a pixel is read out the analogue voltage may suffers several amplification steps, and finally it is converted to a digital value by an Analogue-to-Digital

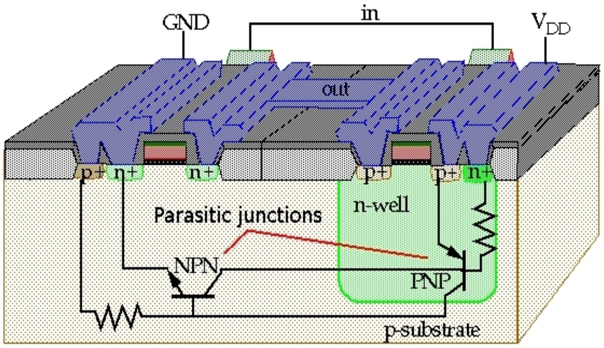


Figure 3.7: Basic p-substrate CMOS cross-section with latch-up circuit model.

Converter (ADC). All these processes introduce noise from the inherent electronic noise in the active read out circuitry. There exists a wide variety of low noise electronic designs for amplifiers and ADCs to reduce this effect.

The principle difference between CCD and CMOS technology on this aspect, is that CCD technology utilises amplification and digitisation steps off-chip, i.e. in an external electronic board in the acquisition system. Therefore the charge is driven by a cable from the image sensor to the board. By contrast CMOS imaging technology commonly incorporates the amplification and digitisation steps on-chip, i.e. in a pixel or immediately adjacent to the chip. The impact of this difference is that CCD technology is more susceptible to noise while it is being driven to the external board. In CMOS technology this risk is reduced.

On the other hand CCD technology has pushed the read noise floor to below 1 electron rms by careful amplifier design and design of digital filtering circuits that process the video signal. In CMOS technology, where the analog process is on-chip, it is more difficult to achieve this optimisation of low noise electronics design [75]. For instance it is easier in CCD technology to add capacitors to control the electrical bandwidth of the analog signal to reduce the white noise, while in CMOS technology this is not possible because adding filters would have a negative effect on the size of the chip.

CCDs manufacturers usually supply the on-chip read out noise figure in electrons root mean square (rms), referring to the standard deviation of the unwanted charge in a given pixel.

3.2.6 Reset Noise

Every time charge within a pixel is read out (i.e. transferred to some read out node for digitisation), the pixel is subsequently reset to a base line potential in order to allow accumulation of charge again, given that not all the charge is transferred. This problem happens because the clocking signals used to read out the charge are not able to attract all the deposited charge within the pixel. The reset noise arises when the pixel is reset but not all the remaining charge is emptied. Therefore the level of charge measured in the next cycle does not correspond exactly to what has been deposited by the ionising particles in that pixel: the actual value of charge has an offset.

The typical solution to this problem, originally developed for CCD technology, is denominated Correlated Double Sampling (CDS). CDS samples the photodiode/gate after reset and then after the signal acquisition, subsequently these two samples are subtracted, hence the reset offset is removed. This technique is easy to implement in CCD technology because of the serial mode read out scheme.

Reset noise is the main source of noise for CMOS technology. Also known as kTC noise, this is of the order of 30-50 electrons rms. This source of noise can be largely eliminated by CDS, as in CCD technology, or by active reset schemes explained below.

CDS is very common in CCD technology but it is still a matter under study in CMOS technology due to the extra difficulty of the parallel read out architecture. CDS can be applied on-pixel or off-pixel, externally. One design choice in state-of-the-art CMOS sensors involves adding an extra capacitor, acting as memory, to store the level of charge after the reset. When the level of charge is measured after integrating this is then subtracted from the level of charge stored in the extra capacitor, therefore mitigating the effect of the reset noise. This extra capacitor in each pixel reduces the fill factor of each pixel, resulting on reduced sensitivity and increased read noise by a factor of $\sqrt{2}$, making the read out of the charge from the pixel slower, as now the charge has to be measured twice.

Another possibility is to implement active reset schemes. Typically, using this approach, the pixel is reset applying $\mathrm{V}_{Reset} < \mathrm{V}_{DD}$ to the *Reset* transistor (see Figure 3.14), being V_{DD} the voltage at which the transistors are powered. This is commonly known as soft reset. With this choice the reset noise is reduced to $\sqrt{kTC}/2$ but two problems arise: FPN and image lag. Image lag is a problem related to the persistence of the previous frame in the current frame, given that the charge of the photodiodes has not been entirely emptied.

An alternative to soft reset is applying hard reset by replacing the n type *Reset* transistor by a p type transistor, which occupies more space, therefore reducing the fill factor. A more common choice is to increase the voltage applied to the n type *Reset* transistor gate above V_{DD} ($\mathrm{V}_{Reset} > \mathrm{V}_{DD}$). Under this scenario image lag is eliminated but the reset noise is still $\sqrt{kTC}$.

Most CMOS designers use a technique called Flushed reset, which is a combination of hard and soft reset. Flushed reset gets rid of the problems of both soft and hard reset and gains their advantages: the reset noise is reduced to $\sqrt{kTC}/2$ and the image lag effect is eliminated. This technique is implemented by applying two different voltage levels sequentially to the *Reset* transistor gate, first applying $\mathrm{V}_{Reset} > \mathrm{V}_{DD}$ (hard reset), and secondly applying $\mathrm{V}_{Reset} < \mathrm{V}_{DD}$ (soft reset).

3.3 Charge-Coupled Devices Technology

Having described the principles of charge generation and read out in pixellated silicon imaging sensors, the next 2 sections describe the detailed architecture and read out mechanisms involved in the two most prolific types of solid state imaging sensors: CCDs and CMOS sensors.

3.3.1 Typical Structure of a CCD Detector

The typical structure of a CCD detector is shown in Figure 3.8. At the very top is a layer of electrodes or gates, made of polysilicon (material consisting of multiple small silicon crystals), used to transfer the accumulated charge from each pixel to the global output buffer in serial mode. Below is a layer made of SiO_2 to isolate the electrode layer and the silicon structure. These layers comprise the dead layers responsible, in this application, for stopping the least energetic particles. Below is the epitaxial sensitive region, made of crystalline doped silicon, and below this is the substrate made of high resistivity and highly doped silicon. The substrate provides a mechanically rigid base upon which all the other layers are formed.

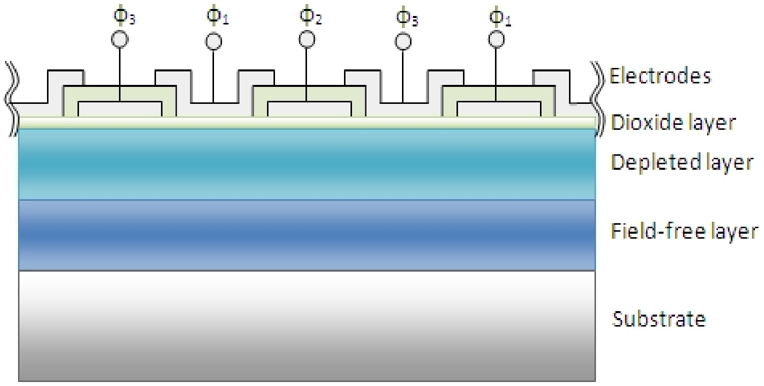


Figure 3.8: Anatomy of a Charge-Coupled Device (CCD).

The sensitive region is comprised of two layers, a depleted layer (i.e. charge free in the quiescent state), produced by application of an electric field, and a field free layer (see Figure 3.9).

A single pixel can be composed of two electrodes (two phase CCD), three electrodes (three phase CCD) or four electrodes (four phase CCD). Between the pixels there exist channel stops, implanted potential barriers made of doped silicon, to prevent the transfer of charge between pixels due to diffusion. In the first CCD detectors these channel stops were implanted on the surface, but the semiconductor/insulator interface was found to trap some charge while the well was being filled that was difficult to release when being emptied, so the transfer efficiency was low. This motivated the design change making the channel stops underneath the dioxide, i.e. buried channel stops [74].

From a more general point of view there exist two different CCD architectures to transfer the deposited charge out of the detector: full frame and frame-transfer architectures.

Frame-transfer devices have the active area divided in two halves, one of them is sensitive to incident radiation (image section) and the other half (storage section) is physically masked to optical radiation, as shown in Figure 3.10(a). Thus, after accumulating

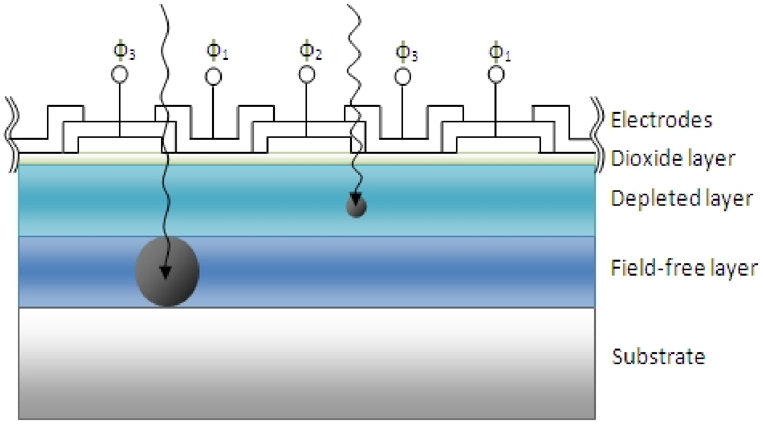
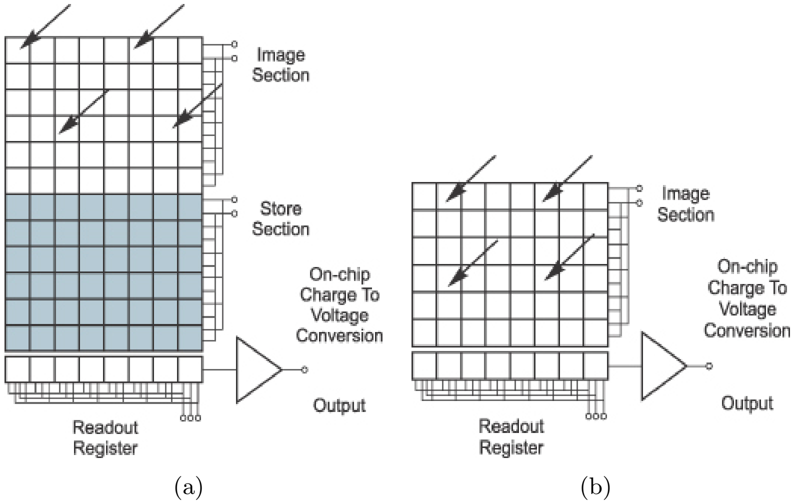


Figure 3.9: CCD cross-section showing charge diffusion generated in each layer of the sensitive region.

charge in the sensitive half for a certain period of time, charge is quickly transferred (∼msecs) from the sensitive half to the storage half. Subsequently the charge in the storage half is serially read out through the read out register to the output amplifier, while the sensitive half is acquiring the next subsequent image frame. The advantage of this approach is that read out and image acquisition processes are no longer dependent on one another, and can be independently optimised for a particular application. The major disadvantage is that the sensitive region is reduced by a half [140].

Full frame devices are more common, due to their simpler design, compared to frame-transfer devices. These detectors comprise of the image section only, thus either accumulate charge or read out the charge. The detailed description of this architecture is presented in the next section.



(a) (b)

Figure 3.10: Frame-transfer (a) and full frame (b) architectures.

3.3.2 Read out Process in a CCD Detector

The way in which a full frame CCD operates is divided in two steps. First it collects and integrates the charge produced by the incident radiation field, and secondly it reads out the deposited charge serially. The typical read out architecture of a CCD is shown in Figure 3.11.

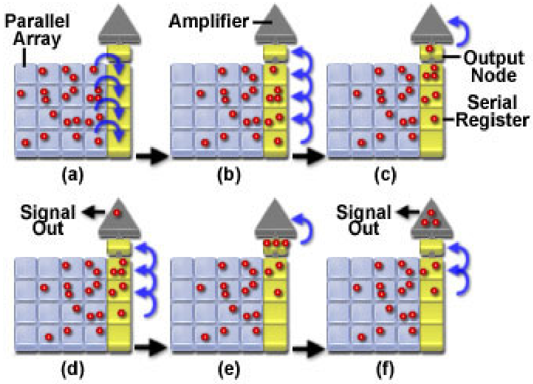


Figure 3.11: Typical read out architecture of a CCD sensor [53].

This architecture shows how one single column is read out in a clock cycle and saved in a vertical register. Subsequently each pixel in the vertical register is read out serially through the common output amplifier. This process takes as many clock cycles as number of pixels to read out, being therefore slow. This is a very important limiting factor for the integration time of the sensor because the integration time can not be shorter than the read out time. To understand this effect, the fact that while a sensor is reading out the charge, incident radiation will keep on impinging on the sensor, unless the read out cycle can be controlled by physically blocking the incident radiation with a shutter. If the integration time and the read out time are similar the last pixels being read out will comprise charge from their original position plus charge due to the integration of the incident radiation field across the entire column, and any associated dark current for that column.

To read out the accumulated charge in the sensitive region, the charge is transferred between pixels within rows, before reaching the vertical shift register, by applying voltages to the electrodes. There exists three different architectures to transfer the charge between pixels: two phase, three phase and four phase CCD clocking. The process of shifting the charge from one electrode to another in a three phase device is depicted in Figure 3.12.

First a positive voltage is applied to the P(1) electrode in each pixel, therefore all the deposited charge within each pixel is confined beneath this electrode (t_1 in Figure 3.12). Secondly the same voltage is applied to the P(2) electrode, so the charge confined beneath the P(1) electrode is now shared between these two electrodes, P(1) and P(2) (t_2 in Figure 3.12). Thirdly the voltage applied to P(1) drops so all the charge is beneath the second electrode P(2) (t_3 in Figure 3.12). This process is subsequently

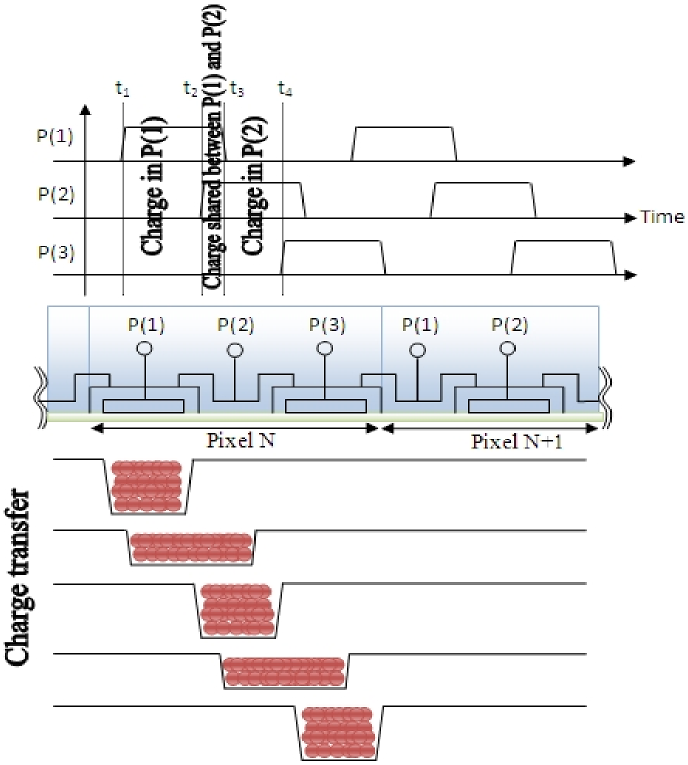


Figure 3.12: Three phase CCD clocking.

repeated to transfer the charge to the third electrode P(3), and then the the electrode P(1) of the adjacent pixel, and so on. This combination of signals is repeated across the whole sensor until the charge of every pixel has been read out. This is why this read out architecture is that slow.

In the three phase architecture observed in Figure 3.12 each pixel is comprised of three phases, each of them connected to one electrode. Therefore one pixel is comprised of three electrodes. In a two phase architecture each pixel is comprised of two phases, each of them connected to two electrodes, shortcircuited between them in pairs, therefore one pixel is comprised of four physical electrodes, which are effectively two. This architecture, used initially due to its simple fabrication process, requires more complex clocking sequences to shift the charge. In a four phase architecture each pixel is comprised of four phases, each of them connected to one electrode, therefore one pixel is comprised of four electrodes. This architecture, used by Philips, is able to use 50% of the pixel for storage obtaining higher charge capacity compared to the two and three phase designs used by other semiconductor manufacturers. The three phase architecture is the most common choice of manufacturers due to some relative advantages obtained with this architecture compared to the other two, such as better spatial resolution and higher frame rates [42].

3.4 Complementary Metal-Oxide-Semiconductor Technology

The continuous advances in CMOS technology motivated by the latest advances in microprocessors and DRAM memories in the semiconductors industry, have now started to make this technology a viable alternative to the popular CCD technology. In the early 1990s CMOS technology made possible the manufacture of imaging sensors based on this technology. The read out is achieved via row and column decoders, as shown in Figure 3.16, so these devices are fast, and by using low voltage control, consume low power. This read out architecture also allows random access to pixels, very useful for applications where certain regions of interest can be read out, instead of the entire active area, speeding up the read out process.

3.4.1 Typical Structure of a CMOS Detector

This technology, similar to CCD technology, is comprised of a two dimensional array of pixels, each of them sensitive to ionising particles accumulating the charge deposited within. There exist two main choices for pixel architectures: Passive Pixel Sensors (PPS) and Active Pixel Sensors (APS) shown in Figures 3.13 and 3.14 respectively.

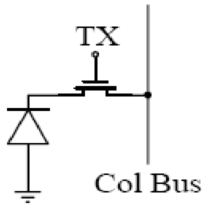


Figure 3.13: Typical pixel architecture of a Passive Pixel.

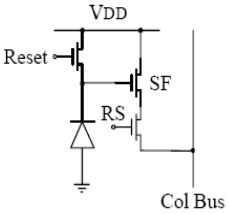


Figure 3.14: Typical pixel architecture of an Active Pixel.

As can be observed, PPS are comprised of just a charge collector (photodiode or photogate) and a transistor to enable the read out of the charge. Conversely, the APS is comprised usually of three transistors, better known as a 3T architecture. The Reset transistor is used to reset the pixel clearing the charge remaining in the pixel after being read out, the Source Follower (SF) transistor is used to convert the charge accumulated in the pixel from charge to current and finally the Read Selection (RS) transistor is used to pump the current out when enabled into a shift register.

In Figure 3.15 a surface profile of a typical CMOS detector is observed. This picture was taken with a surface profiler from the Department of Physics and Astronomy, in the University of Glasgow, UK.

It can be observed in this figure highlighted in red, the profile of the connections between the p-n junctions, and the circuitry and connections with the rest of the electronics contained within each pixel. It can also be observed the large tracks, corresponding with the main connections between pixels and the largest tracks corresponding with the ground, common to every pixel.

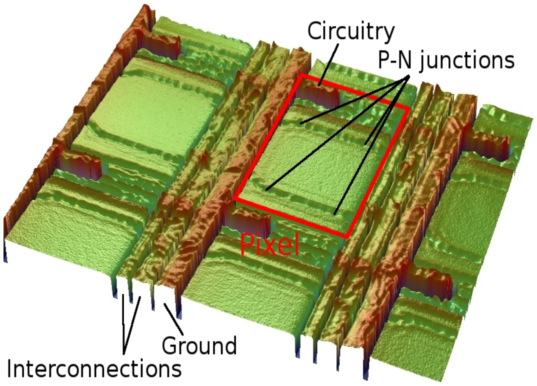


Figure 3.15: Surface profile scan of a CMOS detector array of pixels taken with a surface profiler.

There exists some variations to the 3T architecture to increase the functionality of the pixels, such as the inclusion of ADC converters or N-bits RAM memories as examples. All these additional structures, together with the metallic paths necessary to interconnect all the electronics, affect the fill factor of the detector, i.e. the ratio of the total area of a pixel and the sensitive area of a pixel.

This parameter is of extra importance for imaging applications where the activity and energy is specially low, as the transistors and paths, usually ∼1 μm thick, comprising the aforementioned structures will absorb the incident particles before these reach the sensitive layer. For the specific case of ^{3}H β- autoradiography, where the β- particles have an average energy of 5.7 keV, these will be absorbed in the first 1-2 μm of material. In this case, the higher the fill factor the better sensitivity, therefore too many transistors is not a desirable characteristic in the pixel. Compared to CCD sensors where the fill factor is 100%, due to the absence of transistors, the CMOS sensor suitable for autoradiography should have fill factor as high as possible to obtain acceptable sensitivity.

For higher energy radioisotopes such as ^{14}C or ^{35}S, with average energies ∼50 keV, the circuitry contained in each pixel does not represent a significant problem given that these β particles are absorbed in a depth of ∼10-15 μm.

The thickness of the epitaxial layer in CMOS technology does not exceed 20 μm due to limitations in the standard CMOS processes involved in the fabrication, and to avoid the latch-up effect (see Section 3.2.4). One of the basic concepts of CMOS technology is that it exhibits low energy consumption. Therefore the electric field used to collect the charge in the collection points usually extends to only ∼1 μm deep in the active volume. The rest of the epitaxial layer is not under the effect of any electric field, so that liberated charge behaves in a similar manner to that in the field free region of a CCD sensor.

Most CMOS sensors are APS sensors. As these sensors are fabricated on a single substrate, including detector and read out electronics, these are called Monolithic APS (MAPS), given that Monolithic means *consisting of one piece.*

3.4.2 Read out Process in a CMOS Detector

The read out architecture of CMOS sensors is usually column-parallel. There is usually an Analogue-to-Digital (A/D) converter at the end of each column so the charge in the pixels are digitised in parallel. There exist other architectures that use less A/D converters but still digitise several pixels in parallel (e.g. 1 A/D converter per 4 columns), and other more complex architectures with A/D converters in each pixel, thus the conversion of the whole array is undertaken in parallel, making the read out step very fast. This architecture allows faster read out compared to CCD detectors, making possible shorter integration times. As shown in equation 3.11, dark current is linearly dependant on integration time, therefore this architecture used in CMOS imagers can potentially produce lower thermal noise. A typical structure of CMOS read out architecture is shown in Figure 3.16. It can be observed how a single row is read out in parallel through the column amplifiers, and then outside through the output amplifiers.

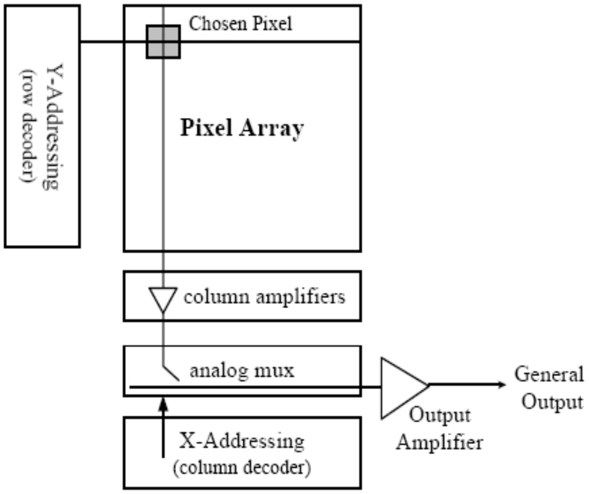


Figure 3.16: Typical read out architecture of a CMOS sensor.

3.5 Designing the Optimum Silicon-based Detector

The above text has briefly outlined the principles of CCD and CMOS technologies. From this discussion the relative advantages and disadvantages of each technology can be compared, but it is clear that no single technology outperforms clearly over the other. Using silicon imaging technology (CCD or CMOS) for direct $\beta\pm$ detection in β autoradiography, there are some requirements to fulfil, apart from those intrinsically obtained by using silicon-based detectors such as linearity and high dynamic range. These are high spatial resolution, high sensitivity and large area. In this section the effect of some key design parameters and intrinsic technological parameters are analysed from the perspective of the spatial resolution and the Signal-to-Noise ratio (SNR), for different energy ranges corresponding to traditional β- radioisotopes.

Conventional autoradiography film has been traditionally used as it offers an intrinsic spatial resolution of ~1 nm, imposed by the diameter of the silver grains that work as the active component of film. However, scattering of the β particles in the film deteriorates the resolution, and this is further affected by the resolution limit of the digitisation process. Thus, the final spatial resolution is typically in the range of 10-30 μm for tissue imaging.

These aforementioned key design parameters, common for CCD and CMOS technologies, are (1) the depth of the depleted region, (2) the pixel size, (3) the thickness of the active area and (4) the collection time.

3.5.1 Spatial Resolution Optimisation

Considering only the spatial resolution as the main feature to optimise, when a particle traverses through the detector, the longest the particle range the more pixels will be affected by this particle, producing a negative effect on the spatial resolution. These particles deposit charge after ionising atoms in the silicon, which subsequently diffuses spherically affecting more pixels. The conjunction of these two effects greatly imposes the final spatial resolution of the detector.

This effect can be mitigated by using optimised design parameters depending on the initial kinetic energy of the particle. In general the spatial resolution is improved by using a thin active area, so the particles escape the detector, after depositing some charge, avoiding further scattering. If deposited charge is quickly collected, the diffusion process is then mitigated. This is achieved by having deep depletion regions and reduced collection time.

CCD detectors usually have thicker sensitive volumes than CMOS detectors (20-100 μm compared to 4-20 μm respectively), so more diffusion is potentially generated, assuming that the incident radiation reaches the field free area. By making CCD detectors with high resistivity silicon (~5-12 kΩ-cm) the depth of the depleted region is extended covering the entire thickness of the active area. This is a very important property of CCD detectors that allows the deposited charge to be collected quickly, minimising diffusion effects. In CMOS detectors the charge diffuses largely due to the absence of electric field in most part of the active volume.

The location of the collection points within a CMOS pixel can also help to mitigate the diffusion process. CMOS pixels usually have a single collection point in the centre of the pixel, with typical dimensions of ~1x1x1 μm^3. Other approaches place more than one collection point distributed in the pixel to collect more charge and to reduce the collection time [144].

The spatial resolution is strongly governed by the pixel size. If short range particles are considered, a small pixel size will provide better spatial resolution. If the application requires to resolve small structures, a small pixel size will obtain more detailed structures than larger pixel sizes. If, on the other hand, longer range particles are considered, small pixel sizes do not represent a practical improvement given that the spatial resolution is ultimately imposed by the particle range.

3.5.2 Signal-to-Noise Ratio Optimisation

Considering only the SNR as the main feature to optimise, the idea would be to maximise the sensitivity, the amount of charge deposited in the active area and to reduce the noise of the detector.

The sensitivity is partly imposed by the dead layers located on top of the detector and by the fill factor. This is 100% for CCD technology, but in the case of CMOS technology this is reduced due to the presence of transistors and interconnection lines in each pixel. This consideration is applicable mainly to visible radiation. Some recent CMOS developments applied to charged particles claim to have 100% fill factor [40, 48, 144] by reasoning that any deposited charge in the epitaxial layer is collected, assuming that the charged particles traverse the pixel electronic layers ($\sim$1 μm thick). For low range β-, such as those emitted by ^{3}H, this is not the case, therefore fill factor is an important parameter for low energy radioisotopes.

A typical method used to increase the sensitivity is by applying the aforementioned back-thinning process. Back-thinning is a mature process in CCD technology where CCD experts understand the process well, mitigating the addition of the aforementioned imperfections. These same experts are starting to apply this process to CMOS technology, but as this is just the beginning there are some not well understood factors that make this process rather complicated. The first few attempts of back-thinning a CMOS detector [31, 43] so far have shown higher figures of noise than their front-illuminated counterparts.

To maximise the amount of charge that can be deposited in the active area the detector should have a thick active volume ($\geq$20μm). Charge recombination should be minimised to collect as much deposited charge as possible.

3.5.3 Detector Noise

To keep the detector noise low is also a very important parameter to maximise the previously discussed SNR. The main sources of noise are reset noise, FPN and dark current noise for CCD technology, and in addition read out noise and reset noise for CMOS technology.

Typical values of dark current at room temperature for CCD technology are 3-50 pA/cm^2 (117-1950 electrons/pixel/second given a 25 μm pixel), whereas for CMOS technology such figures increase up to 100-2000 pA/cm^2 (3900-78000 electrons/pixel/second given a 25 μm pixel) [76]. To counteract this disadvantage CMOS detectors usually acquire faster, using shorter integration times, reducing significantly this source of noise [79]. If faster frame rates are used in CCD technology, to match those rates used in CMOS technology, this will be by using higher electric fields. This increases the consumption and heats up the detector, therefore dark current increases. These problems can be easily removed by cooling down the sensor, which is not always a desirable option. The problem of FPN is also more significant for longer integration times. If the detector has high levels of noise, having concomitantly low SNR, this will probably mask partially, if not totally, the deposited charge by the incident particles. Therefore, to obtain similar quantitative results the radioactive samples will have to be exposed for longer.

CCD technology has had traditionally lower read out and reset noise figures compared to CMOS technology, having a typical value of 3-20 electrons rms of read out noise [50, 105] and reducing significantly the reset noise by using CDS. CMOS technology has traditionally suffered from high levels of read out noise until they were reinvented at the Jet Propulsion Laboratory at the beginning of the 90s, offering since then figures around 10-50 electrons rms. Nowadays figures such as <5 electrons rms can be found in commercial CMOS detectors [7].

CMOS detectors are able to read out at up to ∼1 Gb/s consuming power in the order of tens of mW, therefore dark current is significantly reduced. The efforts of most CMOS research groups to reduce the noise for this technology are concentrated on reducing the reset noise. Different approaches have been briefly described above. Typical reset noise figures in CMOS detectors are of the order of 30-50 electrons rms [172], and by using active reset schemes this can be reduced by a factor of 2 [56] or even completely removed using true CDS. This alternative increases other sources of noise though, as it has been mentioned.

Another problem related with the noise of each technology is related with the aforementioned back-thinning process. In order to back-thin a detector this has to be exposed to additional processes where the silicon is heated up. This has an effect on the silicon's crystal lattice, adding further imperfections to the silicon, generating a higher rate of dark current.

3.5.4 Detector Size

Another perspective that has attracted little attention traditionally is the size of the detector. There have appeared in recent years some applications in the medical and life sciences fields where the size of the detector is a critical factor. Digital detectors being manufactured nowdays are in the range of tens of mm^2, while the life sciences community is asking for tens of cm^2. Large area sensors is a problem for both technologies. The size of the detector is ultimately imposed by the wafer, but the yield, i.e. ratio of working sensors with total number of sensors in a wafer, is also important, as this is directly translated into process costs. A small detector will have higher yield than a large detector, therefore the cost of manufacturing a small detector is lower. However large detectors can be realised by merging some smaller detectors, by a technique called *stitching*. Some large area sensors are already being developed obtaining areas in the range of 30-60 cm^2, but still far from the usual 18cm x 24cm size (432 cm^2) that traditional emulsion film exhibits.

There are some large area CCD detectors in the market, but to read out the charge of these detectors high electric fields are applied (>1 amp consumption) and, as it has been mentioned, this heats up the detector. There are just a few of large area CMOS detectors known so far [10, 16, 123, 153], but as more interest on this matter arises, more groups are working on making larger detectors.

3.5.5 Conclusions

To summarise, assuming an ideal situation, where *anything* could be designed, the ideal detector should fulfil these characteristics:

1. The detector should be as thick as the maximum range of the β- particles emitted by the radioisotope used to maximise the signal deposited in the active area. Around 20 μm should be sufficient for low and medium energy β- autoradiography radioisotopes. To minimise charge diffusion deposited charge should be quickly collected. In CMOS technology several collection points can be distributed on the top layer. In CCD technology the active layer can be completely covered by the electric field, reducing charge diffusion largely.

2. It has been demonstrated in the literature that a back-thinned device is desirable for β- autoradiography, to enhance the sensitivity, therefore increasing SNR.

3. Regarding the noise, given that ^{3}H has an average energy of 5.7 keV, each β electron will release $\sim$1500 electrons. Assuming a desired SNR of 1 dB (signal is 10 times higher than the sum of all the noise sources) the overall noise should be as low as $\sim$150 electrons/pixel (thermal noise included). Some recipes to push down the noise level include working at low temperatures or using short integration times, among some of the most popular approaches. For medium energy radioisotopes with $\sim$50 keV, the signal deposited would be of $\sim$14000 electrons, leading to a less restrictive ground noise of 1400 electrons/event. In this case these 14000 electrons might be split between several pixels, therefore these 14000 electrons should be divided by the number of pixels affected.

4. Small pixels is desirable to obtain high spatial resolution. By considering the mean range of β- particles emitted by ^{3}H ($\sim$0.2 μm), the pixel size should be as close as possible to this measure. A pixel size of 1 μm is actually available in CMOS technology. On the other hand, small pixels have a direct negative impact on the SNR, given that a smaller pixel will collect less signal than a larger pixel in the case where a β electron interacts with more than one pixel, case of ^{14}C and ^{35}S. In this case charge splitting between the affected pixels would have a negative impact in the SNR.

As the overview of technology given in this chapter has exposed, there is no definitive answer to what technology is more suitable for autoradiography. CCD technology is well established in the field, but private and public capitals are pushing CMOS technology basically because of economic reasons. Senior scientists in the field claim that CCD technology is still better positioned than CMOS technology, but performance is improving so fast that CMOS technology can be expected to eventually subsume CCD technology from scientific applications.

Chapter 4

Experimental Materials and Methods

CCD technology is a mature technology of popular choice for scientific applications. Due to some inherent advantages of CMOS technology, such as low power consumption and cost, much effort has been expended to improve the overall performance of CMOS technology in order to compete with CCD technology. There are many groups in the scientific and commercial communities trying to improve such performance to make CMOS technology a serious competitor of CCD technology. There exists quite a lot of discussions on what imaging sensor technology is most suitable for different applications. Some work has been undertaken in autoradiography using CCD detectors, including prior work at Surrey. Rather fewer attempts at using CMOS detectors in autoradiography have been reported in the literature. As one of the aims of this work is to find out what technology is currently more suitable for autoradiography, several sensors have been used in this work: one commercial CCD detector and one off-the-shelf CMOS detector. The basic architecture of these sensors was described in the previous chapter. Their main features are detailed below. As the work shown here is undertaken at room temperature, in contrast with other works undertaken under cooled conditions, some sources of noise, such as dark current and fixed pattern noise (FPN), are exacerbated making observation of β- particle signals in the raw images impossible. Therefore post-processing of these raw images, to counteract the aforementioned sources of noise, is necessary. The post-processing applied to the images acquired with each sensor is described below.

When a β electron impinges on silicon and deposits energy, the resulting free charge diffuses in the silicon often resulting in charge sharing with the neighbouring pixels. Due to the different sources of noise described in Section 3.2, this charge is masked. Part of the work described here, as part of the post-processing step, also details a novel dual threshold that obtains a better sampling of a β- cluster hit, improving the accuracy of the number of pixels affected by a β- ionisation event. By accurately classifying more pixels as part of an actual β- event in a cluster hit, different alternatives aimed to improve the spatial resolution can potentially be applied.

4.1 CCD Sensor Description

The CCD detector used in this work is an E2V[1] CCD55-20 inverted mode sensor. The main features of this detector are detailed in Table 4.1. For ^{3}H imaging, a back-thinned 55-20 sensor was used, with the same pixel array geometry as the front-illuminated device, but with an anti-reflection layer added (∼100 nm thick), which in this case only serves to protect the surface. Being back illuminated, the ionised particles first encounter the residual field-free region rather than the depletion region.

E2V classifies its detector quality by grades from 0 to 2, 0 being the best quality and 2 the worst. This quality level represents the number of dead columns defects, white/black spots and traps that the customer should expect from the CCD detector. The CCD detector chosen for this work is grade 2 as it offers an acceptable number of defective columns/pixels for a reasonable cost of $\sim$ £300, compared to $\sim$ £2000 for a grade 0 CCD detector.

The CCD acquisition system used is a CDB01-X CCD Driver Assembly previously manufactured by E2V. This comprises an analogue read out board, with several amplification stages, Correlated Double Sampling (to remove reset noise) and a digital control board controlled by an ERA60100PBA FPGA that generates all the clock pulse waveforms to control the CCD detector. The resultant video signal is applied to a 12 bit ADC integrated in a National Instruments[2] Data Acquisition card within the PC under the control of a bespoke LabView application which saves each image frame to file. These images are subsequently processed off-line using in-house Matlab code.

4.2 CMOS Sensor Description

The RC-UK Basic Technology Multidimensional Integrated Intelligent Imaging (MI^3) programme [123], which has funded the design and manufacture of the CMOS sensor presented in this work, produced several CMOS detectors. Only the CMOS detector Vanilla is presented in this work as the most suitable detector for autoradiography. The main features of Vanilla, alongside with those of the CCD detector used in this work, are shown in Table 4.1. Vanilla is driven by an acquisition system called OptoDAQ, developed by the Rutherford Appleton Laboratory (Science and Technology Facilities Council), which is based around a board using a Virtex-II Pro^{TM} 20FF1152 FPGA shown in Figure 4.1. OptoDAQ drives and controls the CMOS sensor. The digitised image is then sent to the acquisition PC via fibre optic connection. The image acquisition protocol is accessible using Matlab, LabView or C++ through a dynamic link library (dll) which interacts directly with the hardware.

The Vanilla detector is shown in Figure 4.2, with its corresponding schematic in Figure 4.3. Each pixel has a 3T structure, as that shown in Figure 3.14, with one single charge collection point (1x1x1 μm^3) in the centre of the pixel structure.

---

[1] E2V Technologies Ltd, Chelmsford, Essex, UK

[2] National Instruments UK and Ireland, Newbury, Berkshire, UK

Table 4.1: Main features of the CCD and CMOS detectors

	CCD^{FI}	Vanilla^{FI}	Vanilla^{BI}
Array size	770x1152	520x520	520x520
Pixel size	22.5 μm	25 μ	25 μ
Active area	17.3x25.9 mm^2	13x13 mm^2	13x13 mm^2
Thickness of active area	7 μm Depleted 13 μm Field-free	$\sim$1 μm Depleted 20 μm Field-free	$\sim$1 μm Depleted 20 μm Field-free
Voltage conversion	3 μV/e-	8.9 μV/e-	8.9 μV/e-
Full well	450000 e-	100000 e-	100000 e-
Read out noise	3 e- rms	<25 e- rms	<25 e- rms
Integration time	10 secs/frame*	1 secs/frame	1 secs/frame
Digitisation resolution	11 bits ADC	12 bits ADC	12 bits ADC
Passivation layers	Poly-Si 0.5 μm SiO_2 0.1 μm Si_3N_4 0.1 μm	SiO_2 4 μm Si_3N_4 1 μm	SiO_2 $\sim$100 nm

* governed by serial read out electronics and maximum available clock speed of 1MHz.

^{FI} Front-illuminated.

^{BI} Back-illuminated.

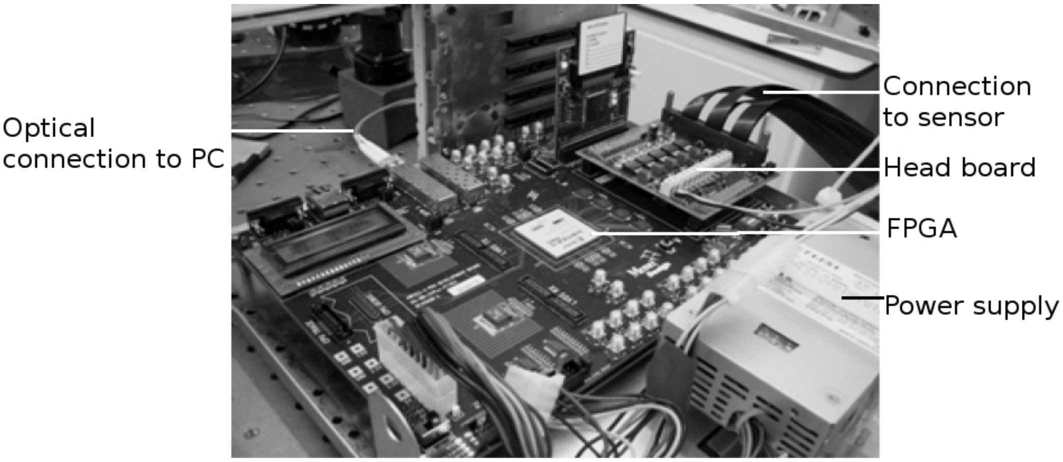


Figure 4.1: FPGA-based OptoDAQ system.

Regarding the cross-section of the sensor, several versions of the Vanilla CMOS sensor have been fabricated. Two different detectors with different epitaxial thicknesses have been fabricated, one with 14 μm thickness and another one with 20 μm thickness.

Due to the typical thickness of the passivation layers placed on top of the CMOS sensors (1-5 μm) a back-thinned version has also been fabricated. Back-thinning is a mature technique applied in CCD technology but relatively new to CMOS technology. This process, described in Section 3.1, is basically based on flipping the sensor and applying a chemical etch. A resulting thin overlying oxide layer will allow detection of low energy

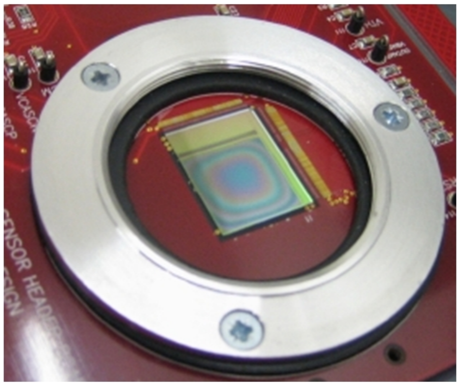


Figure 4.2: Vanilla sensor

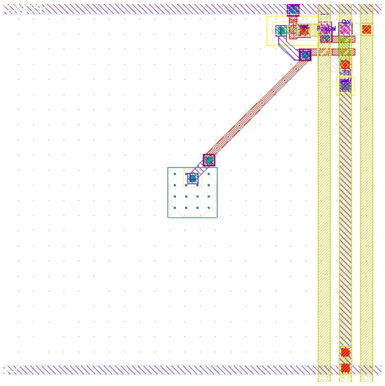


Figure 4.3: 3T pixel architecture used in Vanilla. The photodiode can be observed in the centre of the pixel, the transistors at the top right corner with interconnections.

electrons, such as those emitted by ^3H. See Figure 3.2 to see a block diagram of this process.

4.3 Experimental Methodology for CCD Imaging

In order to correct for dark current and inter-pixel nonuniformities (FPN), which are the main sources of noise masking the signal to be detected, every image acquired on each system has to be somehow *cleaned*. Due to the variety of sensors used in this work, their different features and different measures in which each type of noise affects each sensor a different methodology for each sensor has to be elaborated. In the next subsections the methodology applied to each technology sensor is described.

4.3.1 CCD Pattern Noise Correction

Due to the low activity that samples usually have in β- autoradiography, the count rate is poor, and due to the limited performance of the acquisition system used in this CCD work, relatively long integration times are used. This will concomitantly produce relatively high dark current in the pixels and will also aggravate the FPN (explained in Section 3.2). The typical way of addressing this problem is by cooling down the sensor using liquid nitrogen, CO_2 pellets or other methods. One of the goals of this project is to work at room temperature, to avoid using a cooling system or vacuum the detector chamber, both processes being very inconvenient for undertaking regular experiments in a life sciences laboratory. In the case of this work ten seconds of integration time per image frame was found appropriate to obtain a reasonable level of noise and deposited charge. This will be addressed later.

Long experiments are usually required in β- autoradiography, due to the low activity used in these experiments. As high spatial resolution is also a general requirement low/medium energy radioisotopes are used, which concomitantly makes these β- particles difficult to detect due to their short path lengths (~1-10 μm), therefore increasing the exposure time. This makes temporal stability of the detection medium an important key parameter. Ideally the behaviour of the sensor should be constant with time while it is acquiring. In practise this is not the case, where changes in the detector behaviour have been observed here in this and previous works [88]. These changes are sometimes slow, which is explained by slow changes of temperature, or there may be sudden jumps in individual and overall pixel dark current levels, which are more difficult to explain. These two problems are tackled as detailed in [178], where an adaptive FPN correction is applied to acquired images at room temperature, to mitigate dark current and FPN effects. This noise correction algorithm is briefly described below.

First a set of blank frames is acquired to compute for each pixel located at location (i, j), the mean dark current offset or pedestal ($\mu_{i,j}$), the standard deviation ($\sigma_{i,j}$), and the global mode of a set of blank frames (m_{ref}). This set of blank frames is used to correct the dark current and to apply the adaptive FPN correction. As dark current corruption is principally a bias effect, $\mu_{i,j}$ attempts to model this bias level for each pixel.

Examples of the mean and standard deviation computed from a 562 frames blank dataset are shown in Figures 4.4(a) and 4.4(c) respectively, with their corresponding intensity histograms. Figure 4.4(d) shows how the standard deviation exhibited by most of the pixels (~60 %) in the CCD detector have a value of 3 Digital Numbers (DN).

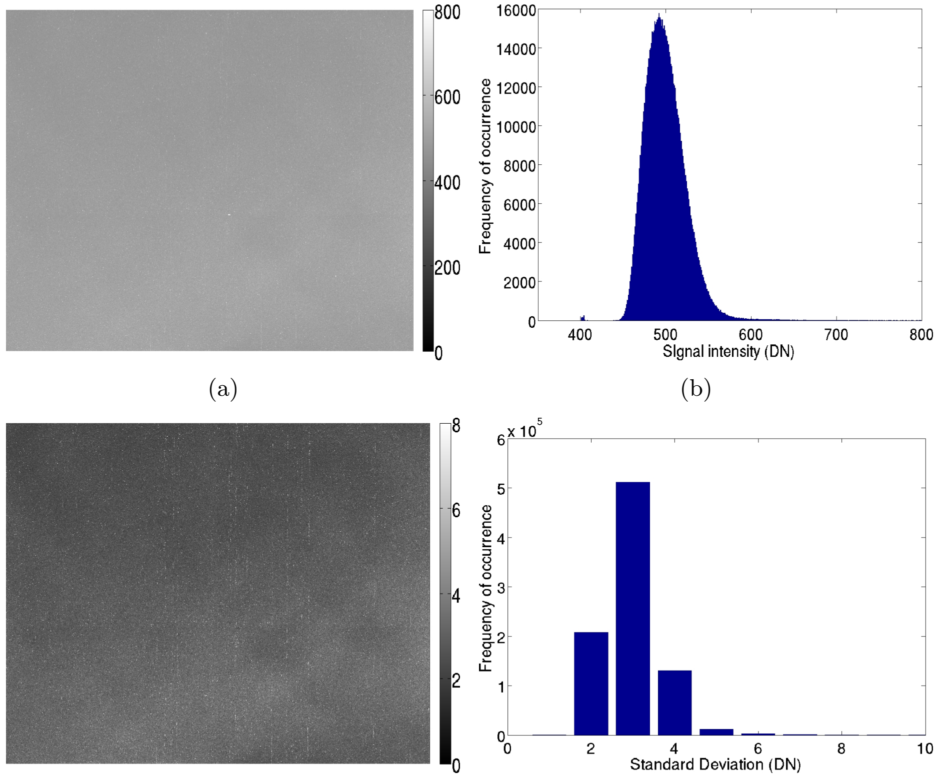
To demonstrate the temporal stability of the CCD detector, the evolution of the mode measured in every single frame from the set of blank frames is shown in Figure 4.5. The mode of this set is 495 DN with a standard deviation of 3.2 DN. This standard deviation gives an idea of how stable this detector is; it will be compared with the CMOS sensor in the following sections. It is observed an initial increase of the mode from 475 DN to 495 DN until the detector temperature stabilises.

A selection of random pixels have been independently studied to observe how accurate they can be modeled with the aforementioned mode and standard deviation. The intensity histograms of six random pixels obtained from 575 samples are shown in Figure 4.6.

The reason why the mode is chosen to model a pixel, instead of the mean, is because the mean can potentially be affected by outlier DN values measured in certain frames, while the mode is not affected. The coordinates, mode, mean and standard deviation of each pixel are shown in Table 4.2. It is observed how the mode and the mean are not significantly different, besides a slight skew of the mean. Figure 4.6(d) shows a bimodal distribution exhibited by pixel 4, representing a typical example of a hot pixel that shows a sudden change in its dark current level behaviour. This issue does not represent a problem for the detection of true particles as will be explained later.

After the set of blank frames has been acquired, new images are acquired and the correction, defined by equation 4.1, is applied.

$$y_{i,j} = x_{i,j} - [\mu_{i,j} - (m_{ref} - m_c)] \tag{4.1}$$



(a) (b)

(c) (d)

Figure 4.4: Mean image (top left), corresponding intensity histogram (top right), standard deviation of each pixel (bottom left) and corresponding histogram (bottom right), calculated with 562 frames acquired using ten seconds of integration time at room temperature with the CCD detector.

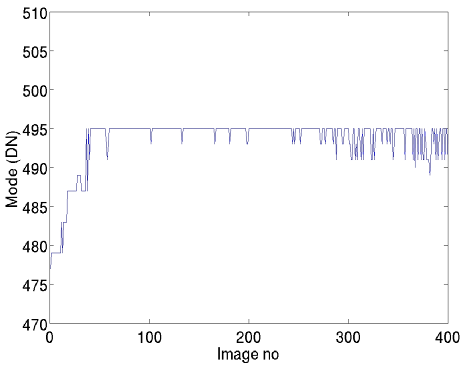
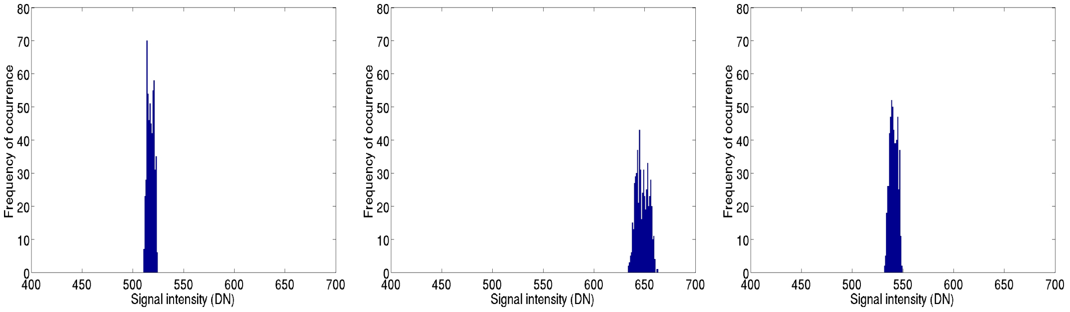
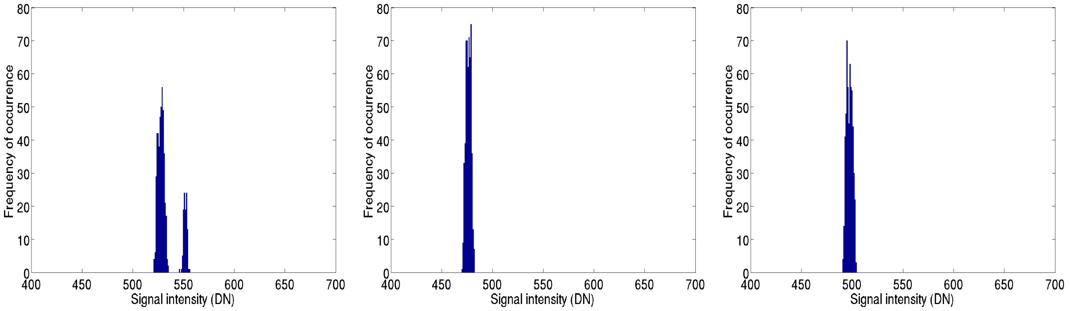


Figure 4.5: Evolution of the mode of each of the first 400 acquired images for the set of blank frames. The mode of this set of modes is 495 DN and the standard deviation is 3.2 DN, but there are obvious jumps present in the data. Each frame was acquired after ten seconds of integration time.



(a) Pixel 1 (b) Pixel 2 (c) Pixel 3



(d) Pixel 4 (e) Pixel 5 (f) Pixel 6

Figure 4.6: Intensity histograms of random pixels measured in the CCD detector for 575 frames.

Table 4.2: Main parameters of the pixels shown in Figure 4.6

	X	Y	Mode	Mean	Standard Deviation
Pixel 1	538	136	645	647	6.1
Pixel 2	897	258	539	540	3.8
Pixel 3	411	523	529	532	9.9
Pixel 4	166	207	479	476	2.5
Pixel 5	942	180	495	497	3.0
Pixel 6	291	588	514	517	3.2

where $x_{i,j}$ is the raw pixel value, $\mu_{i,j}$ is the mean value of the pixel (i,j) obtained from the reference set of blank images, m_{ref} is the mode of all the images in the reference data set, m_c is the mode of the current frame and $y_{i,j}$ is the new pixel value. The difference $m_{ref} - m_c$ counteracts drifts in $\mu_{i,j}$ when $m_c \neq m_{ref}$, i.e. when the behaviour of the sensor changes as explained above.

After this first step, a threshold proportional to the standard deviation of each pixel, defined by equation 4.2, is then applied to each pixel:

$$T_{i,j} = M_{i,j} + k\sigma_{i,j} \tag{4.2}$$

where $M_{i,j}$ is the mode dark signal of the pixel, averaged over the blank set of images, and $\sigma_{i,j}$ is the dark level standard deviation for each pixel obtained from the aforementioned blank frames, which defines the pixel threshold individually. The pixel threshold is then globally defined in terms of the coefficient k. This parameter is experimentally set for a certain radioisotope, hence k does not have to be changed between experiments given similar environmental conditions. A flow chart describing the whole image correction process and thresholding is shown in Figure 4.7.

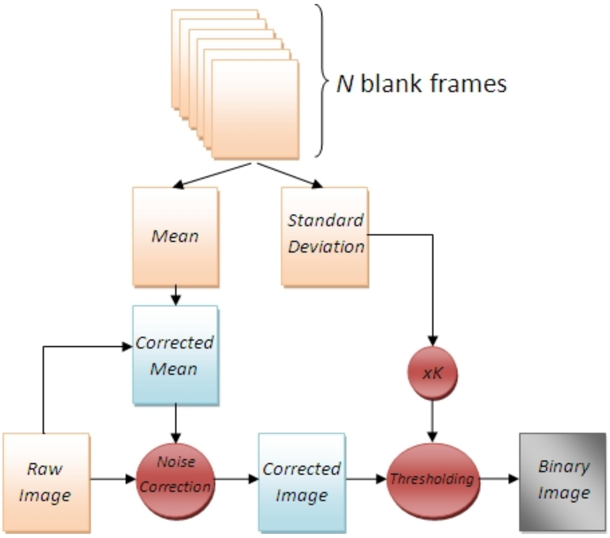
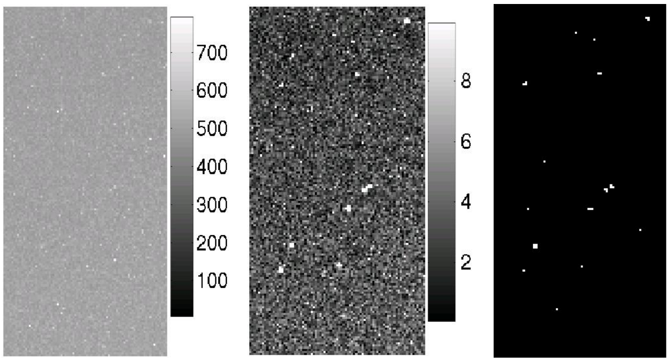


Figure 4.7: Flow chart of the image correction applied to the CCD images.

The resulting binary images are then labelled using 8-connectivity analysis and the location, size and intensity of each event cluster are saved. In order to obtain a useful composite image, the thresholded images, described above, are accumulated and summed. The result of this process applied to the image data is shown in Figure 4.8. Note the scale of each image.



(a) (b) (c)

Figure 4.8: Raw (left), corrected (middle) and thresholded (right) images of the front illuminated CCD sensor, showing the stages used to correct for pattern noise. (b) and (c) represent the image data prior and post application of the thresholding. This shows a single frame from a ^{14}C exposure for ten seconds of integration time.

4.4 Experimental Methodology for CMOS Imaging

In the case of the CMOS sensor Vanilla, the most important sources of noise are reset noise (kTC noise) and residual inter-pixel FPN. In the case of FPN, due to the short integration time, it is not as significant as with the CCD detector described previously.

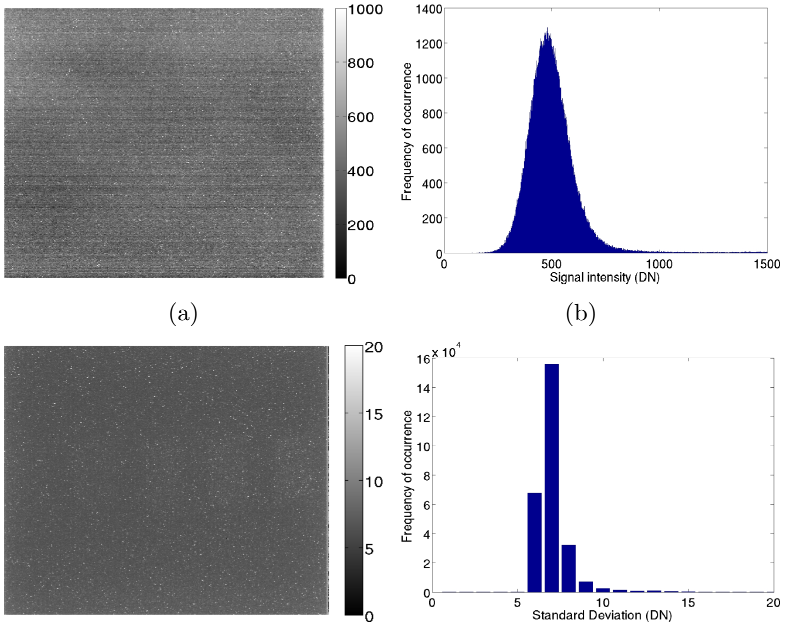
4.4.1 CMOS Pattern Noise Correction

The correction applied to the images acquired with Vanilla is the same as that explained in Section 4.3.1, applied to the CCD detector. However, the integration time used in all the experiments presented here is one second. This is the shortest integration time achievable by the OptoDAQ system with the current version of the software that shows long term reliability. But it is worth reiterating that the read out here is column-parallel, so that a complete frame only requires 520 pixels to be read out serially, compared to 887040 (770x1152) pixels in the CCD detector using single node serial read out. Dark current noise is directly proportional to integration time, therefore the longer the integration time the higher the dark current noise. It was shown in Section 3.5 that CMOS technology shows typically figures ~30-40 times higher of dark current compared to CCD technology. Therefore, to mitigate this effect, the integration time used was the shortest that the technology allows offering reliability with long term experiments.

As has been mentioned before, a set of blank frames is necessary to statistically model each pixel. Usually 3400 frames are used for this purpose. An example of the mean intensity image (μ), corresponding histogram, standard deviation for each pixel (σ) and corresponding histogram are shown in Figure 4.9. It can be seen how approximately 60 % of the pixels have a σ value of 7 DN. This shows that most pixels exhibit a similar performance but, in comparison to Figure 4.4, the device is noisier than the CCD sensor used in this work. This will affect significantly the threshold applied to each pixel (equation 4.2). In the case where a pixel changes significantly its signal, due maybe to a faulty crystalline silicon structure in that specific pixel, this will originate false positives in the detection process. Using the thresholding method defined by equation 4.2, a pixel under this situation will exhibit high $\sigma_{i,j}$, resulting in a high threshold $T_{i,j}$. Therefore its digital value will be below the threshold unless it has a significant high digital value (caused by deposited charge) to be considered as a true event. This explains the robustness of the thresholding method used here when noisy pixels are present in the image.

As has been mentioned, the temporal stability of a detector is a very important parameter to consider when undertaking long experiments such as those necessary in autoradiography. The evolution of the mode for the 3400 frames acquired for the set of blank frames is shown in Figure 4.10.

Once again, to demonstrate that a pixel can be modeled with the mode and the standard deviation, the histograms of six random pixels are presented in Figure 4.11 calculated with 42840 samples. The coordinates, mode, mean and standard deviation of each pixel are shown in Table 4.3. It can be observed how the mode and the mean are not



(a) (b)

(c) (d)

Figure 4.9: Mean image (top left), corresponding intensity histogram (top right), standard deviation of each pixel (bottom left) and corresponding histogram (bottom right) calculated with 3400 frames acquired with Vanilla using 1 second integration time.

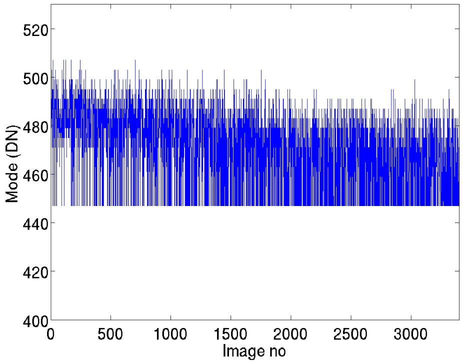
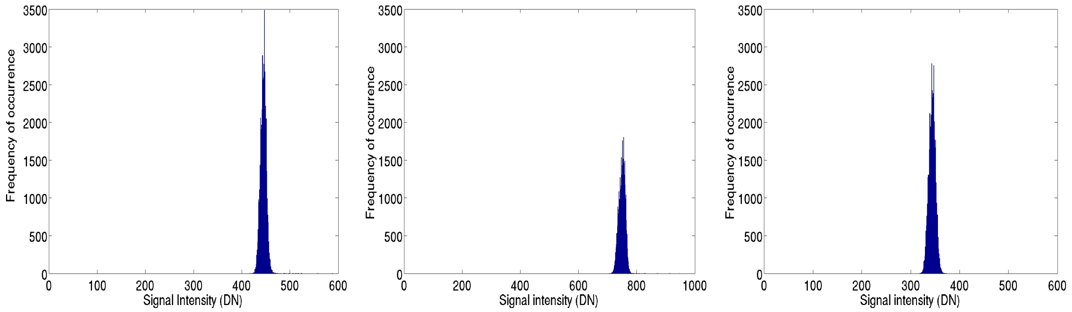


Figure 4.10: Evolution of the mode of each of the 3400 acquired images for the set of blank frames.

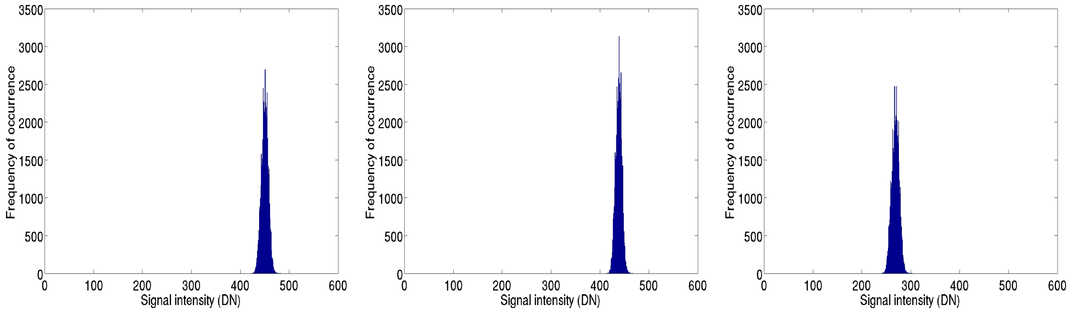
significantly different. Note how pixel 2 has the highest difference between mode and mean and also the highest standard deviation.

An example of a raw image obtained with Vanilla with one second of integration time and its corresponding corrected and thresholded images are shown in Figure 4.12.

Their corresponding histograms, presented in Figure 4.13, show a reduction of intensity, corresponding mainly to dark current noise, of ten times. It is observed how the pixel intensity values from the raw image (blue) are compressed in lower values in the corrected image (red) in the histogram.



(a) Pixel 1 (b) Pixel 2 (c) Pixel 3



(d) Pixel 4 (e) Pixel 5 (f) Pixel 6

Figure 4.11: Intensity histograms of random pixels measured in the Vanilla detector for 42840 images.

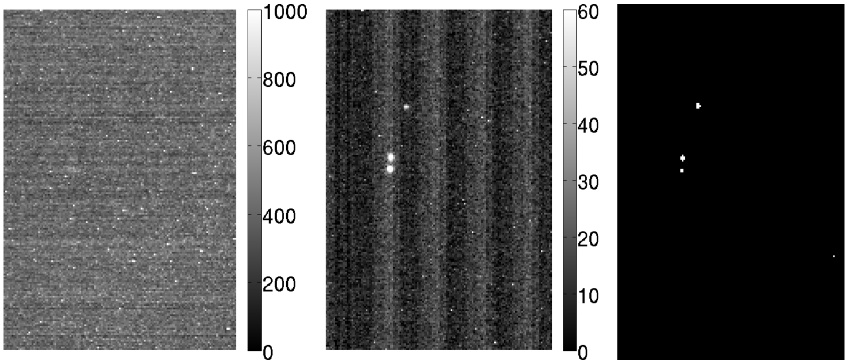
Table 4.3: Main parameters of the pixels shown in Figure 4.11

	X	Y	Mode	Mean	Standard Deviation
Pixel 1	274	207	447	444	6.1
Pixel 2	308	316	755	749	10.9
Pixel 3	105	290	343	343	6.8
Pixel 4	54	34	451	450	7.0
Pixel 5	106	106	439	438	6.3
Pixel 6	433	433	269	268	7.6

Other observations can be extracted from the histograms shown in Figure 4.13, such as few saturated pixels, with 4095 value (12 bits ADCs) in the intensity histogram of the raw image, although these are not shown in the histogram. This is attributed to a defect in this particular sensor, corresponding to excessive dark current noise in these pixels.

4.5 Hot Pixels

As has been mentioned, all pixels in a detector have certain level of leakage current, also called dark current. Hot pixels are considered to be those pixels that exhibit a higher rate of leakage current than the rest. Hot pixels are caused by a defective lattice in the bulk silicon near the channel stops, which due to the electric field present, electron-



(a) (b) (c)

Figure 4.12: Raw (left), corrected (middle) and thresholded (right) images of Vanilla, showing the stages used to correct for pattern noise. (b) and (c) represent the image data prior and post application of the thresholding. This shows a single frame from a ^{14}C exposure for one second integration time, with 3 β- events detected.

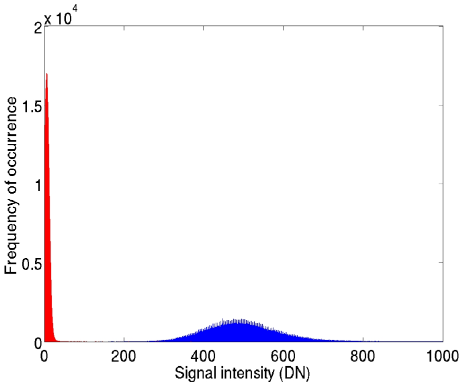


Figure 4.13: Intensity histograms of raw (blue) and corrected (red) images, Figures 4.12(a) and 4.12(b) respectively, acquired with the Vanilla sensor.

hole pairs are thermally generated. In CCD and CMOS imaging, this effect is highly influenced by temperature, being aggravated for high temperatures. The behaviour of these pixels can be classified in three different categories:

1. Pixels that consistently show high digital value.

2. Pixels that normally exhibit a leakage current similar to their neighbours, but eventually the leakage current jumps to a higher/lower level. This presents a bimodal intensity histogram. Pixel 4 in Figure 4.6(d) is an example of this category.

3. Pixels that sporadically exhibit different rates of leakage current. This case also presents a bimodal distribution, but these pixels are more difficult to detect automatically. This effect has been observed within this work only in the CMOS detector used, but not in the CCD detector.

The practical effect is that these pixels are usually bright as if an event has been detected causing false events. If the number of hot pixels is significant this could significantly affect Region-of-Interest (ROI) analysis or any other quantitative analysis. These pixels are discarded in the following way described below.

Based on the low activity contained in the samples used here, it is assumed that the corresponding probability of two events impacting in the same pixel in two consecutive frames is negligible. Hence, if a pixel is bright in two consecutive frames, the *apparent* event data of both events and all future events is discarded until this pixel returns to a value consistent with its neighbours. This method only fixes the problem if a hot pixel is most of the time bright and eventually turns off (category 2). If the pixel flickers (category 3), as has been observed in the CMOS detector, this problem has to be solved in a different way. These *flicker* pixels are removed by applying the FPN correction (Section 4.3.1) to all the images, and investigating the final composite image, i.e. the sum of all the thresholded images. The final composite image will exhibit pixels with low DN value, corresponding to noisy pixels that *detect* false events, pixels with a certain DN value corresponding to true events which impacted in the same pixel, and finally pixels with a much higher DN value corresponding to the aforementioned flickering hot pixels. A post-acquisition method used in practise to tackle this problem is to discard those pixels, after the composite image has been obtained, above a given N DN value. Two images are then obtained, one with the remaining pixels and that with those removed hot pixels. If the latter shows any visible structure, this means that the value used to discard the hot pixels, N, is too low and a higher value must be used. Although manual inspection has been used, the method could easily have been semi-automated by calculating the correlation of the residual image with the summed image, and then use a selected correlation threshold to ensure a correct N was applied. This process is manually repeated until no structure is visible in the image containing the hot pixels.

4.6 A Novel Region of Interest Post-Analysis

The aforementioned strategy was followed in this study to correct for dark current noise and FPN, in order to detect true β events present in each frame, obtaining a low rate of false events. The approach taken so far has largely followed the algorithm described in Section 4.3.1. In this section some further correction methods are described. The approach detailed in Section 4.3.1 thresholds the image, y, locally by considering the variance of each pixel $\sigma_{i,j}$, and globally by the coefficient k, producing a binary image, g. It has been empirically observed how a relative high k value was necessary to obtain an acceptable level of noise under room temperature acquisition. This has two concomitant negative effects: first it discards low energy β events and secondly it sub-samples the actual charge deposited by higher energy events in a cluster of pixels.

Using the methodology described above, it has been observed how some events that deposit charge in more than one pixel, are partially sampled. This situation is common in CMOS technology where no electric field is present in the active area, and charge may diffuse for several microns depending on the kinetic energy of the β electron and the thickness of the active volume. For a given cluster of pixels, only those pixels with higher digital values than the threshold defined by equation 4.2 are considered, therefore discarding other pixels that exhibit lower digital values but actually contain part of the deposited charge.

On the other hand, given the discrete nature of pixelated detectors, the spatial resolution is significantly affected by the detector pixel size among other design parameters. If the impinging β particle has enough kinetic energy to affect more than one pixel, and the deposited charge is accurately measured the hit event may be reconstructed in order to obtain spatial information on how the electron behaved in the silicon at sub-pixel level. Therefore it is of high importance to obtain accurate measurement of the deposited charge in each detected hit. To tackle this problem a dual thresholding method is detailed below in order to improve the measurement of charge deposition in pixelated detectors.

This approach is based first on the detection of β events using the methodology described above. It will be shown later that in this first step the sensitivity of the detector is imposed, given that this first step detects the total number of events present in the current frame. After some events have been detected a ROI, R_λ, is defined in the corrected image (Figure 4.12(b)) around each event detected in the previous step. The initial size of R_λ is 9x9 pixels, given that 5 pixels correspond to 125 μm, close to the expected maximum range of β- emitted by ^{14}C or ^{35}S (see Figure 3.1). Two different approaches are studied in this section. The first of them establishes a fixed threshold within R_λ, lower than the initial T defined in Section 4.3.1. The second approach establishes an adaptive threshold based on the statistics within R_λ.

4.6.1 Fixed Dual Thresholding

It was shown in equation 4.2 how the parameter k sets the threshold globally. It has already been stated that this approach produces deposited charge under-measurement, therefore poor sampling information.

In this approach the original parameter k is renamed to k_1, and a second value k_2 is used to establish a second threshold T_2 only within each R_λ. Using this approach the number of detected events in a single corrected frame, is imposed by k_1. This ultimately controls the sensitivity and the background noise level of the system. Low values of k_1 will lead to a large amount of events detected, representing true events (high sensitivity) but also false events, i.e. noise (high noise). On the other hand, high values of k_1 will lead to a few amount of events, reducing the sensitivity and the noise at the same time.

After the events being detected in the first thresholding step, each event is independently analysed, by thresholding again using k_2, being $k_2 < k_1$. This provides more accurate sampling information, in order to obtain sub-pixel resolution. This may also include some noise in the measurement, but it will be demonstrated below how this effect is not significant compared to the addition of pixels corresponding to partially sampled events.

4.6.2 Adaptive Dual Thresholding

In this second approach, as part of the ROI analysis, the pixel with the highest digital value in each local region R_λ, p_i, is compared with a threshold N, in order to differentiate between noisy events and true events. This threshold N is obtained from the histogram of the corrected image, defined in equation 4.1 by y, $h(y)$ (Figure 4.14) in the following way: let us define two simple concepts related with the occupancy of the histogram $h(y)$: (i) An isolated bin within the histogram $h(y)$ is a bin $h(z)$ with 1 DN height and empty bins in the contiguous positions. It must fulfil the three conditions shown in equation 4.3:

$$Isolated\ \ bin\ \ h(z)_{iso} \begin{cases} h(z) = 1\ \ \& \\ h(z-1) = 0\ \ \& \\ h(z+1) = 0. \end{cases} \tag{4.3}$$

(ii) A non-isolated bin is defined as a bin *h(z)* with occupancy >1, or with anterior or posterior contiguous bins with non-zero values. It must fulfil at least one of the conditions shown in equation 4.4:

$$Non-isolated\ \ bin\ \ h(z)_{non-iso} \begin{cases} h(z) > 1\ \ \| \\ h(z-1) \neq 0\ \ \| \\ h(z+1) \neq 0. \end{cases} \tag{4.4}$$

It is assumed here that the residual noise (residual dark current and FPN) is compactly concentrated in the lower intensity bins of the histogram *h(y)*, as shown in the red component of the histogram in Figure 4.13. True β- events, that deposit a significant part of their charge in one single pixel, if present in the image, will reside in isolated bins at higher digital values than the residual noise, whereas non-isolated bins are assumed to correspond to noisy pixels. These assumptions are made under the low probability of having two β events, in the same frame (one second), with exactly the same deposited charge in electrons, subsequently digitised (DN) and rounded. Under this improbable situation these two events would represent a non-isolated bin in $h(y)$ and would be classified as noise.

This framework assumes that low activity sources/samples are used in these experiments and short integration times are used. Under these assumptions, and using the previous definitions shown in equations 4.3 and 4.4, the value N is set at the intensity value corresponding to the last isolated bin after the lowest DN contiguous bins (considered to represent noise). This is graphically shown in Figure 4.14. In this specific real example the inset in Figure 4.14 shows the histogram detail representing the lowest values of the histogram. This figure shows the bins of the histogram cluster where the last non-isolated bins are located at positions 96 DN and 97 DN, which means that the last isolated bin is at position 101 DN. Therefore N has a value of 101 DN for this particular corrected image.

Once the threshold value N is obtained for a specific corrected image, each event is independently analysed, similarly to the previous thresholding method. A 9x9 pixels

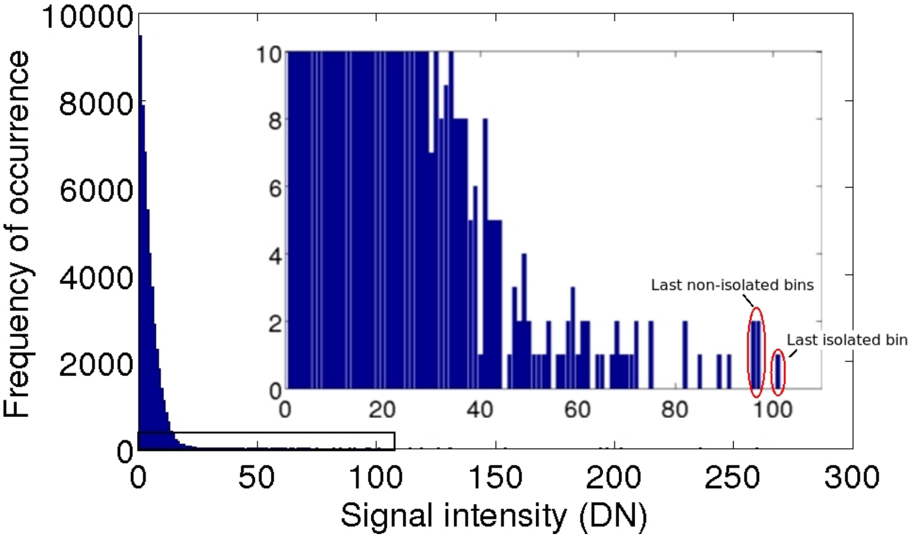
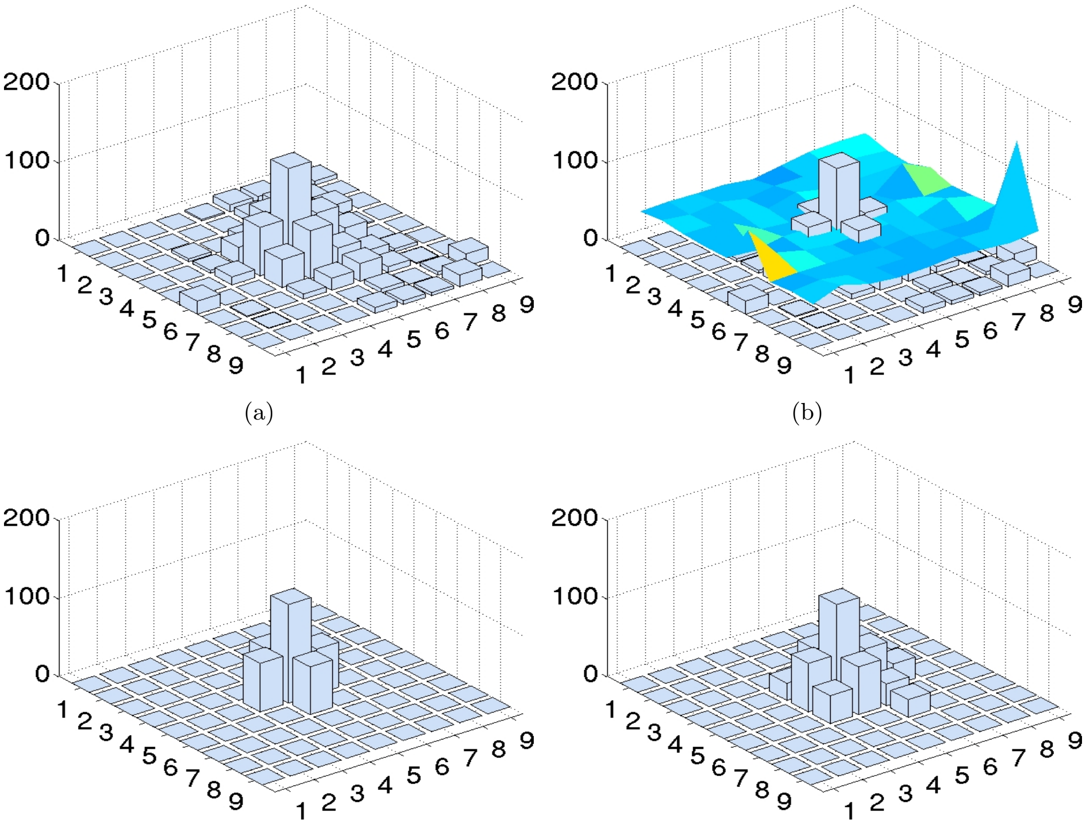


Figure 4.14: Exemplar intensity histogram $h(y)$ of a corrected image y with inset showing the lower intensity bins of the histogram. Note the two last non-isolated bins and the last isolated bin, that corresponds with the threshold chosen to distinguish between noisy pixels and true β- events. It is assumed that higher intensities correspond to higher probability of occupancy by genuine events.

R_λ around the highest intensity pixel of each detected event is analysed. If this pixel value is below N the event is considered noise, if it is above N then it is subject to be analysed. In this case, instead of using a fixed threshold within the R_λ as before, the statistical mode m_{R_λ} and the standard deviation σ_{R_λ} of R_λ are obtained, and R_λ is subsequently thresholded using $m_{R_\lambda} - \sigma_{R_\lambda}$ as threshold. After thresholding R_λ only those pixels spatially connected to the cluster hit where the central pixel resides, are considered as part of the true β- event containing deposited charge.

An example of adaptive dual thresholding process is shown in Figure 4.15. A 9x9 ROI placed on a detected event, and centred in the pixel with the highest digital value of R_λ, is shown as a 3D plot in Figure 4.15(a). Each bin represents the digital value read from each pixel of R_λ. It is observed that this cluster consists of a central pixel, where the particle deposited most of its energy, then some adjacent pixels contain part of the energy deposited by either the primary β particle, by secondary electrons generated in the silicon or as a result of dark current, and finally other pixels with a certain digital value corresponding to noise surrounding the event cluster under study. Figure 4.15(b) shows Figure 4.15(a) with the threshold defined independently for each pixel, as described by equation 4.2, plot on top. Only five pixels exhibit higher digital values than the threshold, as it is shown in Figure 4.15(b), therefore only these five pixels are considered as part of the hit cluster using the initial correction method described in Section 4.3.1 (Figure 4.15(c)). This example shows how no pixels are classified as false events, but this also shows how some adjacent pixels to the central pixel are missed, and thus are wrongly classified as noise. After applying the adaptive thresholding method explained above, adjacent pixels to the central cluster, that are initially missclassified as noise, are now correctly classified as pixels containing deposited charge, therefore considered as part of the true event and added to the hit cluster, as shown in Figure 4.15(d). By correctly classifying more pixels as part of the cluster hit this method obtains enlarged clusters. This is assumed to contain better representation of the true charge deposited. This claim will be further demonstrated in Section 6.3.1 by using experimentally validated Monte Carlo simulations.



(a) (b)

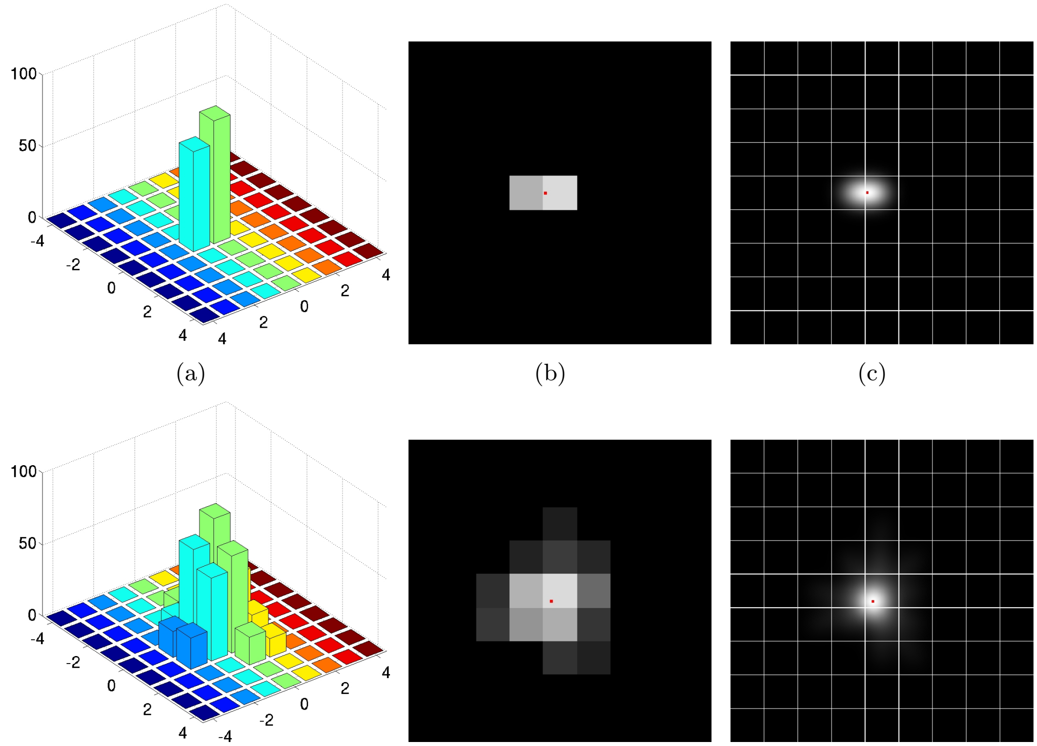
(c) (d)

Figure 4.15: 9x9 ROI cropped from a corrected image acquired with Vanilla shown as a 3D plot (a), same plot as (a) but with the threshold calculated with equation 4.2 for each pixel added on top (b), graph (b) thresholded using the original method (c) (some pixels with deposited charge have been removed) and resulting ROI obtained after implementing the improvement described above (d). This was assumed to contain a higher proportion of the true deposited charge, plus some additional residual dark current signal.

4.7 Cluster Hit Reconstruction

The methodology explained above, where more information is extracted from each cluster hit, has the potential of obtaining sub-pixel resolution. It has been observed how including more pixels in each cluster hit decreases the spatial resolution (see Table 5.3). On the other hand this can help to obtain a better reconstruction of the cluster hit, by obtaining a more accurate initial position where the β electron impinged, compared to the situation where less information is available.

An example of a real β- event is shown in Figure 4.16 to demonstrate the impact of this dual thresholding method. An exemplar 3D plot of deposited charge in each pixel is shown in Figures 4.16(a) and 4.16(d), using the initial correction method (described in Section 4.3.1) and the adaptive dual thresholding method respectively, using as a global thresholding parameter k=8 in both cases. The grey scale images of such 3D plots are shown in Figures 4.16(b) and 4.16(e) respectively, showing how with the standard method the cluster hit is comprised of only 2 pixels, while with the adaptive dual thresholding method the cluster hit is comprised of 14 pixels. The computed centre of gravity (COG) is indicated with a red point in both cases.



(a) (b) (c)

(d) (e) (f)

Figure 4.16: 3D plot of the deposited charge in each pixel using the traditional correction method (a) and the fixed dual thresholding method (d). Greyscale images of corresponding 3D plots showing the COG with a red point (b) and (e) respectively, and corresponding reconstructed cluster hits (c) and (f) respectively.

To compute the COG in Figures 4.16(b) and 4.16(e), the images were up-sampled by a factor of x100 (using nearest-neighbour interpolation), being the distance between COGs in Figures 4.16(b) and 4.16(e) 9.5 μm. There are other cases where this distance is longer (or shorter) depending on the energy and angle of incidence in the detector of the β electron. Monte Carlo (MC) simulations, where the real COGs of noiseless events are known, have also been undertaken to demonstrate that the new dual thresholding method obtains more accurate charge deposition information. The distance between the real COG, and the resulting COG after applying the dual thresholding method, for several simulated β- events was computed, showing lower errors than the distance between the real COG and the resulting COG after applying the initial correction method. This error could vary within a range of 1-10 μm.

It is known that β- particles describe tortuous paths when they interact in silicon, depositing energy at certain points in their trajectory, and also generating secondary electrons, when impacting with outer electrons with sufficient energy. From charge diffusion theory it is usually assumed that deposited charge diffuses in silicon and, if projected on the top surface of the active layer, a Gaussian-like distribution is obtained [78]. Moreover, it is also assumed here, and confirmed later in the MC simulations, that the pixel with the highest intensity value indicates where the β electron impacted initially in the silicon, and then started its erratic trajectory depositing charge in the neighbouring pixels. Based on these assumptions, and the information present in the pixelated images (Figures 4.16(b) and 4.16(e)), the cluster hits are reconstructed by creating straight lines between the COG of the cluster and the centre of each pixel within the cluster. Then a succession of Gaussians along each straight line joining each pixel with the COG is created as shown in Figures 4.16(c) and 4.16(f) for these cases shown in Figures 4.16(b) and 4.16(e) respectively. This succession of Gaussians emulates the charge diffusion process observed in each point where the particle deposited charge. As the depth of interaction is not known in each point the σ of the Gaussians has been arbitrarily set at 7.5 μm, observing satisfactory results visually in the final reconstructed clusters in Figures 4.16(c) and 4.16(f).

The effect of this cluster hit reconstruction in calibrated ^{14}C microscales and ^{3}H labelled brain tissue imaging will be presented in Chapter 5.

Chapter 5

Experimental Evaluation

To be able to compare quantitatively and qualitatively the performance of each detector and, more important, to highlight the differences between the CCD and the CMOS detectors used in autoradiography, the results for each single experiment are going to be presented simultaneously for the two detectors.

The performance, expressed as the sensitivity and background noise level, has been measured using calibrated microscales of ^{14}C, and ^{3}H in the case of the backthinned devices [58]. The microscales consist of known amounts of radioactivity homogeneously distributed in a plastic tissue equivalent polymer of 50(^{3}H)-120(^{14}C) μm thickness. Each microscale has 8 cells of varying activity; for ^{3}H the range is 0.11-4.04 kBq/mg and for ^{14}C, this is 1.11-31.89 kBq/g.

To demonstrate practical application of these detectors, tissue sections labelled with ^{3}H are also presented below. Tissue sections labelled with ^{14}C are not presented in this work because this radioligand is only used for spinal cord studies in our laboratory, and moreover these experiments do not represent a challenge in terms of spatial resolution.

The experiments were set up placing each sensor in a light proof box and placing the radioactive sample, either the microscale or the tissue sample, in direct contact with the surface of the sensor. For comparison purposes the samples were also exposed to conventional autoradiography film. When using film, the protocol used was the same as the typical protocol followed by experts in a life sciences laboratory. This was placing the samples in direct contact with film within a light proof cassette, making firm contact on the samples against the film, and after a sensible exposure time for each radioisotope, subsequently digitising the developed sheets of film.

5.1 Sensitivity

The sensitivity of a detector is a measure of how many β- events per unit area, emitted by the sample under study, are detected in a certain period of time given a certain amount of activity.

While introduced as the definition of sensitivity in this work, it is still not clear what is the standard measure of sensitivity in the scientific community. All the different

Table 5.1: Different measures used to measure sensitivity

Measure	Detector	Reference
counts	Betacamera	Puertolas *et al* [145]
e-/pixel/μCi	CCD+Scintillator	Karellas *et al* [81]
c/mm^2/kBq/g	CCD	Ott *et al* [130]
ADC/pixel/kBq/g	CCD	Kokkinou *et al* [88]
c/μg/μCi/g	MWPC	Petegnief *et al* [137]
cpm/kBq/g	MCP	Lees *et al* [99, 98, 96]
cps/kBq	MCP	Lees *et al* [100]
	Medipix I	Abate *et al* [1]
cps/mm^2/kBq/g	β-Imager	Tribollet *et al* [168]
	Medipix II	Mettivier *et al* [118, 120]
	CMOS MIMOSA V	Cappellini *et al* [32]

measures found in the literature have in common a calibrated microscale with known values of activity given by the manufacturer, typically of ^{14}C, being used, measuring detected events (counts), ADC units or electrons for each band of the microscale. The normalised number of counts per band is given for ^{14}C and ^{3}H in [145], exposed for 180 minutes and 150 minutes respectively, but no measure of sensitivity is given (it is assumed that the number of counts corresponds to the number of detected events, but this is not explicitly stated). The number of electrons per pixel against activity (μCi) has also been used [81] to assess the sensitivity of a scintillator intensified CCD. Sensitivity using direct detection with CCD detectors has also been measured under cooled conditions [130] by measuring counts per mm^2 against activity (kBq/g), and at room temperature [88] by measuring ADC units per pixel against activity. Counts per μg of standard tissue against the accumulated activity per mass unit is used to measure the efficiency of a multiwire proportional chamber in [137]. Another common measure is counts per minute (cpm) against activity (kBq/g) used in [96, 98, 99] to measure the sensitivity of microchannel plates (MCP) or in cpm against cpm/g given by a scintillator counter, which is derived from the actual activity of the source [13]. Counts per second (cps) against activity (kBq) is used to assess the sensitivity of MCPs applied to autoradiography in [100] and of Medipix I in [1]. Finally cps/mm^2 against activity (kBq/g) is used to measure the sensitivity of the β-Imager [168], the hybrid detector Medipix II [118, 120] and the MAPS CMOS detector MIMOSA V [32]. Given that this measure gives more information (spatial and temporal information) than the other aforementioned measures this is the one used in this work. All these measures are summarised in Table 5.1.

An exemplar composite image of calibrated ^{14}C microscales exposed for 100 minutes to the CCD sensor is shown in Figure 5.1 . Figure 5.2 demonstrates the corresponding image obtained from the CMOS Vanilla sensor with ^{14}C for the same exposure period. The initial dark current noise and FPN correction method detailed in Section 4.3.1 has been applied to both images, using a value of k=8 as global parameter (equation 4.2) for these two examples. Figure 5.3 shows the corresponding image obtained with autoradiographic film with the same reference source and for the same exposure period. Clearly, the level of contrast against background fogging makes image quantification with film impossible in such a short exposure time.

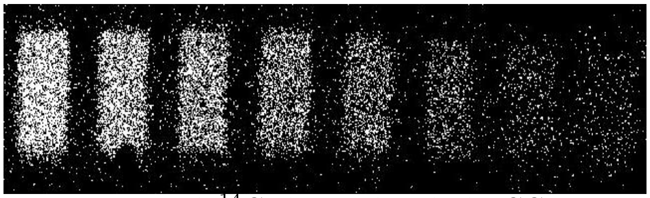


Figure 5.1: Composite image with ^{14}C obtained with the CCD sensor after 100 minutes exposure

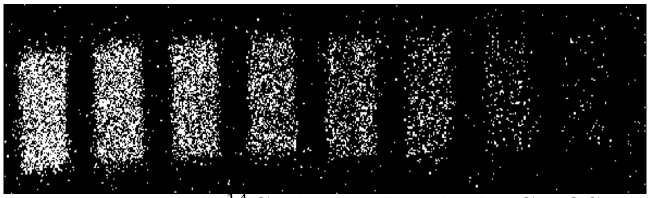


Figure 5.2: Composite image with ^{14}C obtained with the CMOS sensor after 100 minutes exposure

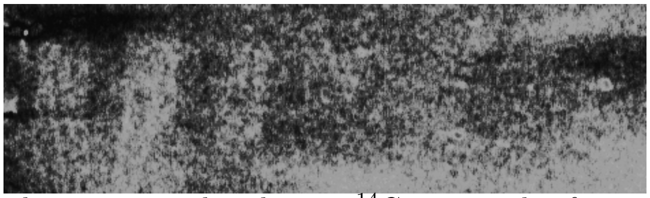


Figure 5.3: Exemplar image produced using ^{14}C microscale after exposure to Biomax MR film for 100 minutes.

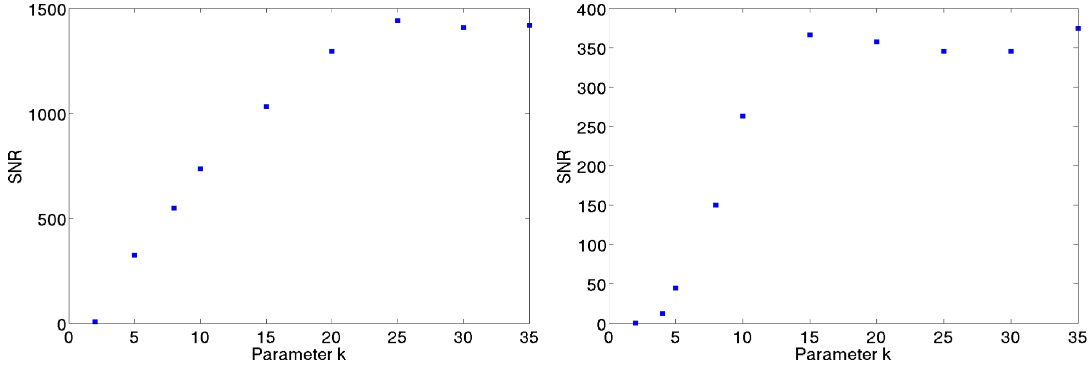
To analyse the sensitivity and the observed background noise level of the CCD and the CMOS sensor from the composite images (Figures 5.1 and 5.2), ROI analysis was undertaken to determine the number of counts per second per band (cps/mm^2) against the known specific activity per band (kBq/g obtained from the microscale manufacturer's data sheet [58]) to calculate the sensitivity, and the number of counts per second per mm^2 in a manually drawn ROI in the area corresponding to the background to calculate the background noise level.

If the cps/mm^2 measured in each band is plotted in the ordinate axis, and the activity (kBq/g) is plotted in the abscissa axis, a linear response of the detector is observed, and secondly, if a straight line is fitted to the data points, the sensitivity can be extracted from the slope of the fitting line.

As has been shown in the methodology section, the measure of the sensitivity and the background noise level is significantly affected by the threshold (equation 4.2) imposed on the corrected frames. As observed in equation 4.2, the threshold is ultimately dependant on the global parameter k. To establish an objective value for this coefficient, the relation of the parameter k with a measure of Signal-to-Noise ratio (SNR), defined by equation 5.1, has been calculated for the two detectors under the same conditions: same source, same post-processing and same length experiment.

$$SNR = \frac{I_M - I_B}{I_B} \tag{5.1}$$

where I_M is the number of counts per second per mm^2 measured in a given band and similarly I_B measured in the background. The evolution of the SNR with coefficient k for the CCD detector and the CMOS detector are shown in Figure 5.4.



(a)

(b)

Figure 5.4: Evolution of SNR with coefficient k for a ^{14}C source exposed to the CCD detector (left) and the CMOS detector (right).

The parameter k clearly controls the overall sensitivity of the detector, as it also controls the observed background noise level. This noise level controls ultimately the minimum detectable activity. Given this dependence the sensitivity and background noise level have also been calculated for the CCD detector and the CMOS detector for a range of k values from 2 up to 35, shown in Figures 5.5 and 5.6 respectively. It is observed in Figure 5.5 that the sensitivity of the CCD detector is ∼1.5 times better than the sensitivity of the CMOS detector. This can be explained by a higher fill factor exhibited by CCD technology, by the thinner dead layers located on the surface of the CCD detector and by the noisier pixel response (higher $\sigma_{i,j}$) observed in the CMOS detector compared to the CCD detector, as shown in Figures 4.9(d) and 4.4(d) respectively. It can also be seen how there exists a drop of sensitivity at low values of k, due to exacerbated noise that masks the signal. A linear evolution of the sensitivity is obtained in the k parameter range of 4-25 for both detectors. The negative slope within this linear range has been obtained, resulting in a rate of -1.4 10^{-4} cps mm^{-2} kBq g^{-1} k^{-1} for the CCD detector and -2 10^{-4} cps mm^{-2} kBq g^{-1} k^{-1} for the CMOS detector (although parameter k is unitless here it is presented as if k was its unit). This is translated in a slightly higher sensitivity of the CMOS detector to the k parameter. It is also observed in Figure 5.6 how the ordinate axis is in logarithmic scale to better distinguish the difference of background noise between both detectors. If shown in linear scale, for the CCD detector there is a plateau at k=8 which corresponds with a background noise of ∼3.7 10^{-4} cps mm^{-2}. Similarly for the CMOS detector the plateau starts at k=6 corresponding with a background noise of ∼13 10^{-4} cps mm^{-2}.

To see how the parameter k affects to the composite image, exemplar composite images, showing only the 5 bands with higher activity, obtained after 100 minutes of exposure

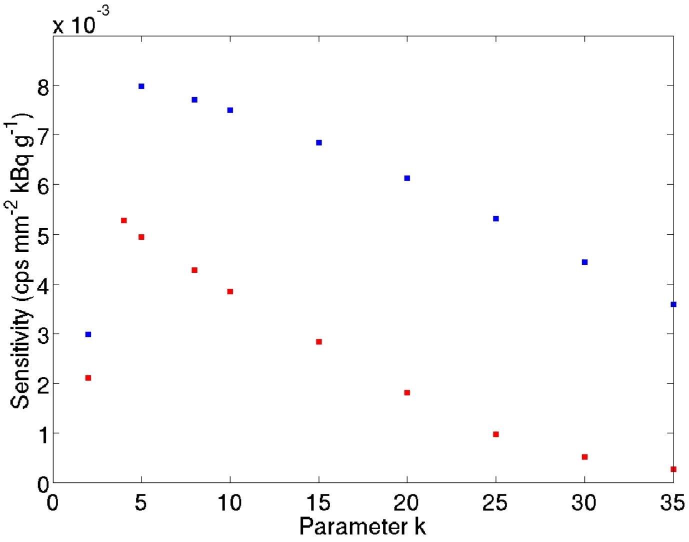


Figure 5.5: Evolution of sensitivity (cps mm^{-2} kBq g^{-1}) with parameter k for a ^{14}C source exposed to the CCD sensor (blue) and the CMOS detector Vanilla (red). In the k parameter range of 4-25, where the response can be approximated as linear, the drop rate is -1.4 10^{-4} cps mm^{-2} kBq g^{-1} k^{-1} for the CCD detector and -2 10^{-4} cps mm^{-2} kBq g^{-1} k^{-1} for the CMOS detector.

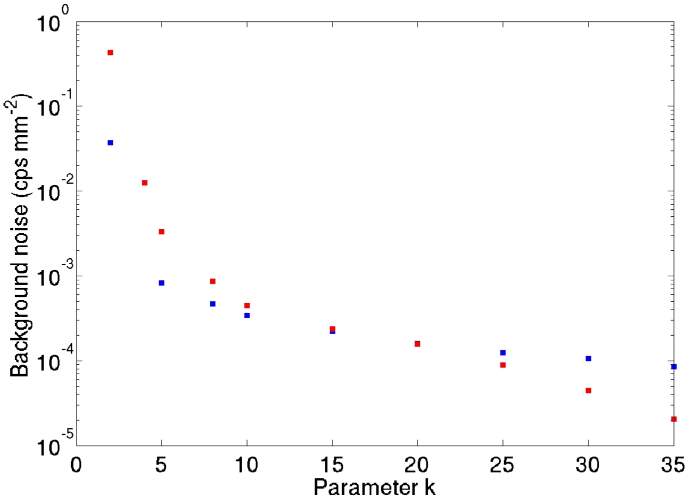
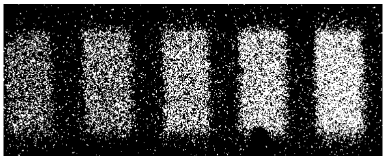


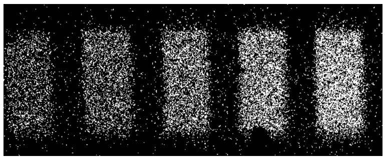
Figure 5.6: Evolution of background noise (cps mm^{-2}), shown in logarithmic scale, with parameter k for a ^{14}C source exposed to the CCD sensor (blue) and the CMOS detector Vanilla (red).

time are shown in Figure 5.7 for the CCD detector and in Figure 5.8 for the CMOS detector, for several values of k.

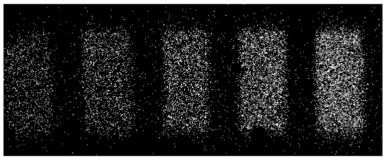
To put the sensitivity and background noise measurements in perspective, Table 5.2 compares the performance of the CCD detector and the CMOS detector with other detectors published in the literature.



(a) k=5

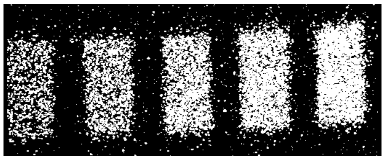


(b) k=15

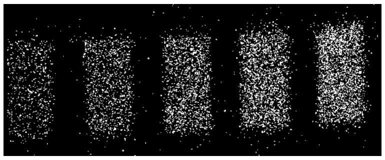


(c) k=30

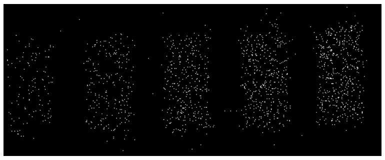
Figure 5.7: Composite images of a ^{14}C microscale exposed to the CCD sensor for 100 minutes for different values of the k parameter.



(a) k=5



(b) k=15



(c) k=30

Figure 5.8: Composite images of a ^{14}C microscale exposed to the Vanilla sensor for 100 minutes for different values of the k parameter.

Table 5.2: Sensitivity and background noise measured in silicon-based detectors.

Detector	Radioisotope	Sensitivity (cps mm^{-2} Bq mg^{-1})	Background noise (cps mm^{-2})	Experiment length
E2V CCD47-20	^{3}[H]	1.2 10^{-5}	2.9 10^{-4}	48 hours
E2V CCD47-20	^{14}[C]	24.6 10^{-3}	2.0 10^{-3}	12 hours
CMOS Vanilla	^{3}[H]	0.9 10^{-5}	4.7 10^{-4}	15 hours
CMOS Vanilla	^{14}[C]	11.4 10^{-3}	3.79 10^{-3}	100 minutes
CMOS MIMOSAV	^{3}[H]	2.0 10^{-5}	6.6 10^{-4}	2 hours
Hybrid MedipixII	^{3}[H]	1.0 10^{-5}	1.4 10^{-4}	1 hour

To obtain an optimum value for the global parameter k the SNR (Figure 5.4), sensitivity (Figure 5.5) and background noise (Figure 5.6) have been studied for both detectors.

Figure 5.4, where the evolution of the SNR with the control parameter k is shown, suggests that high values of k (k$\simeq$20 for both detectors), obtain the highest SNR that both technologies can achieve. On the other hand Figure 5.5 suggests that the highest sensitivity is obtained at very low k values (k$\simeq$5 for the CCD detector and k$\simeq$4 for the CMOS detector), compromising the number of detected false events, i.e. background noise. Finally the background noise (Figure 5.6), if shown in linear scale, shows how both detectors present a plateau at k=8 for the CCD detector and k=6 for the CMOS detector.

Considering the suggested k values by each graph for each detector, the optimum value taken in this work is that suggested by the background noise level. Too low k values, as the sensitivity graph suggests, would produce too noisy images so the SNR would be compromised. Too high values, as the SNR graph suggests, would produce very *clean* images, but the sensitivity would be extremely compromised, making necessary to extend the experiments for too long. The background noise graph suggests a middle point between two extremes. Therefore k=8 for the CCD detector and k=6 for the CMOS detector are the values used in the forthcoming experiments.

5.2 Calibration of the CCD and CMOS Detectors

In order to compare sensor performance it is necessary to calibrate the energy response of the CCD and the CMOS imagers, so that a comparative event threshold in calibrated keV units can be used between detectors. Both sensors were exposed to a point source of ^{241}Am (activity of 41 kBq) situated approximately 30 mm from the surface of the detector with a layer of insulating tape to absorb the associated α particles. The geometry was set up such that the chances of obtaining overlapping event clusters within a single image frame was considered negligible.

For the CCD detector, peaks corresponding to the ^{241}Am photon energies of 13.9 keV, 17.5 keV, 22 keV and 26 keV were observed in the reconstructed energy spectrum, as shown in Figure 5.9. Exposing the sensors to a ^{14}C source also provided a reference energy of 49 keV from the associated mean energy in the beta deposited energy spectrum.

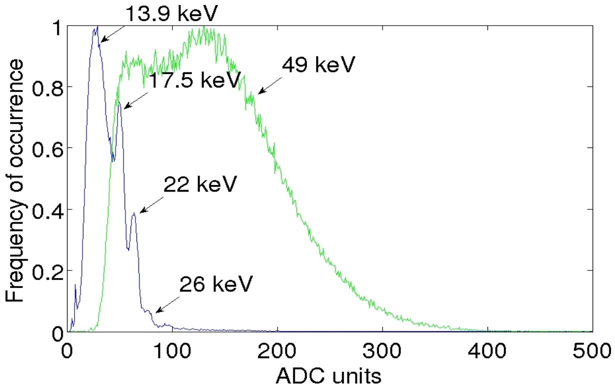


Figure 5.9: Energy spectra of ^{241}Am (blue) and ^{14}C (green) measured in the CCD detector, with the most significant energies indicated.

From these data obtained with the CCD detector, a fitted calibration line is shown in Figure 5.10 with a slope of 3.61 DN/keV (0.27 keV/DN) and an intersection point of -16 DN units. Assuming 3.6 eV is required to liberate one electron in silicon, then this yields a slope of 75 e-/DN units and an intersection point of 1234 electrons.

For the CMOS detector, peaks corresponding to the ^{241}Am photon energies of 13.9keV, 17.5 keV, 22 keV and 26 keV were observed in the reconstructed energy spectrum, as shown in Figure 5.11.

From these data obtained with the CMOS imager, a fitted calibration line is shown in Figure 5.12 with a slope of 24.1 DN/keV (0.041 keV/DN) and an intersection point

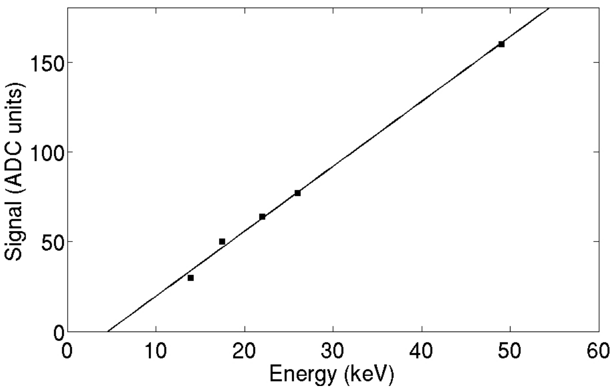


Figure 5.10: Energy calibration curve for CCD sensor. X axis represents the signal in keV and Y axis represents the signal in ADC units.

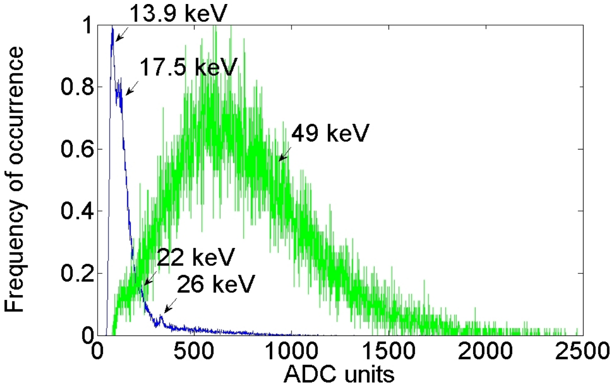


Figure 5.11: Energy spectra of ^{241}Am (blue) and ^{14}C (green) measured in the CMOS detector Vanilla, with the most significant energies indicated.

of -295.7 DN. Assuming 3.6 eV is required to liberate 1 electron in silicon, then this yields a slope of 11.5 e-/DN units and an intersection of 3400 electrons. Compared with the corresponding values presented for the CCD detector, this demonstrates first the higher amplification of the electrons read out from a pixel in to DN shown by the CMOS detector, and secondly the higher noise exhibited by this CMOS detector.

Having calibrated the CCD and the CMOS sensors energy response, every DN value can be quantified in keV.

5.3 Temporal Stability of the Vanilla Detector

Given that the experiments typically undertaken in autoradiography take for relative long periods of time (several hours) it is important to measure the variability (or stability) of the detector performance. The performance is measured here as the detector sensitivity (cps mm^{-2} kBq g^{-1}) and the background noise (cps mm^{-2}).

Several experiments have been consecutively undertaken, exposing the ^{14}C microscale to the CMOS Vanilla sensor, for different periods of time: 1, 2, 6, 12, 18, 36 and 48

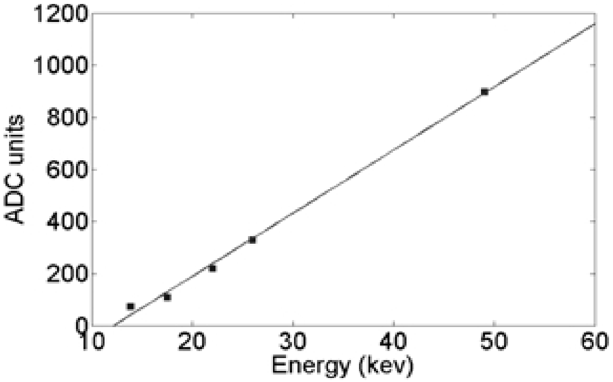
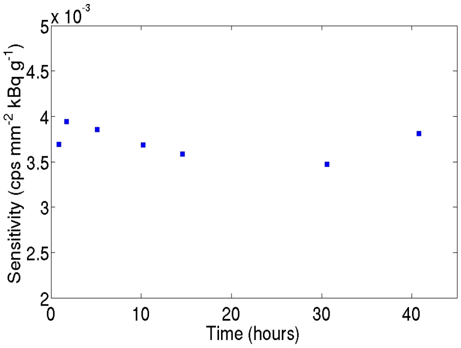


Figure 5.12: Energy calibration curve for the CMOS Vanilla sensor. X axis represents the energy in keV and Y axis represents the signal in DN units.

hours. The sensitivity and the background noise for each experiment are shown in Figure 5.13 and Figure 5.14 respectively. The experiment length shown in the abscissa axis represents the true experiment length discarding corrupted frames.



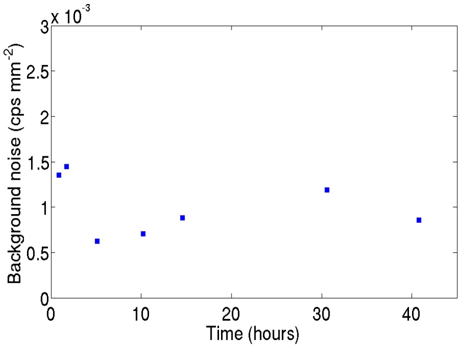


Figure 5.13: Evolution of sensitivity (cps mm^{-2} kBq g^{-1}) with time of the Vanilla detector for a ^{14}C microscale.

Figure 5.14: Evolution of background noise (cps mm^{-2}) with time of the Vanilla detector for a ^{14}C microscale.

These two figures show good stability of the CMOS detector, specially the sensitivity study, showing a maximum variation between the mean sensitivity (3.7 10^{-3} cps mm^{-2} kBq g^{-1}) and the most deviated measure of 5.06 %.

5.4 Cluster Distribution

The cluster distribution represents the histogram of the different event sizes detected in number of pixels. This distribution may be considered as a surrogate measure of Point Spread Function (PSF) of the system.

Depending on the methodology explained in Section 4.3.1 and the post-processing analysis explained in Section 4.6, the cluster distributions obtained are different in each case. The new thresholding approaches (Section 4.6) append more pixels to the cluster hit when a β- event is detected, compared to the traditional approach (Section 4.3.1). This will have an effect in the cluster distribution by showing higher occupancy in those bins corresponding with larger cluster sizes.

The choice of detector also makes a difference, given that deposited charge in silicon behaves differently in each technology. It has been described how deposited charge mainly diffuses in the field free region in the entire active volume of the CMOS sensor and in the field free region of the CCD sensor, but the CCD also benefits from a much larger depletion region, wherein deposited charge is quickly collected. It has been detailed (Section 4.1) that the field free region in this specific CCD detector is 13 μm, while the active volume of this CMOS detector is 20 μm thick completely field free (Section 4.2). This represents a larger charge diffusion process in the CMOS detector than in the CCD detector, producing larger cluster hits.

To show how the different approaches exposed here affect the spatial resolution of the system, given the absence of a point or a line source, a method described in [148] has been used to obtain an approximation to the true PSF. This method is based on the alignment of every detected event, centring every event in the pixel with highest intensity value. It is assumed in this work that this method gives a good approximation of the PSF of the system, making the measure independent on the distance between the original position where the particle was generated and the pixels where the particle interacted in the active area. This method assumes that the pixel with highest intensity value corresponds with the pixel where the β electron interacted first when impinged in the detector, as will be later demonstrated in the MC simulations.

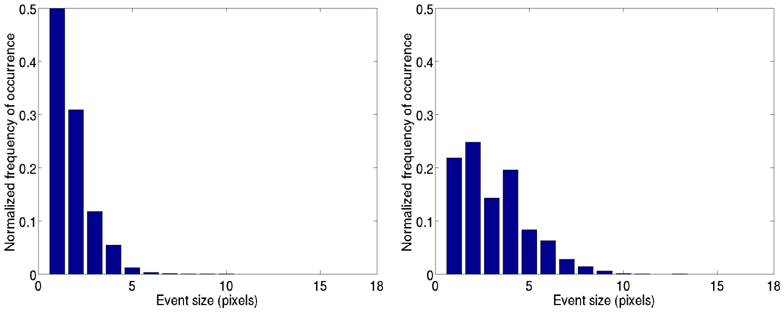
The results presented in this section are later confirmed by Monte Carlo simulations presented in Section 6.3.1 as part of a forward looking design process, using experience learnt from the sensors available for experimental work. The processes described here are reproduced and the final results compared in Section 6.3.1.

5.4.1 Cluster Distribution with no ROI Analysis

The normalised cluster distributions obtained with a ^{14}C source exposed for the 100 minutes on the CCD and CMOS detectors, where no post-processing is applied, for a value of k=8 as a global thresholding parameter for both sensors are presented in Figure 5.15.

In order to obtain an accurate cluster distribution of the genuine β- events, to compute these histograms shown in Figure 5.15, only those events beneath the radioactive bands of the microscale are considered here. Moreover, to also consider the false events that are beneath the radioactive bands of the microscale, another cluster distribution comprised of only false events, enclosed in an ROI with the same size as the sum of all the radioactive bands, is placed in the background, where no radioactive source is placed. This resulting cluster distribution comprised of false events, is then subtracted from the cluster distribution comprised of true events. This has been observed to give a more accurate cluster distribution if compared with MC simulations.

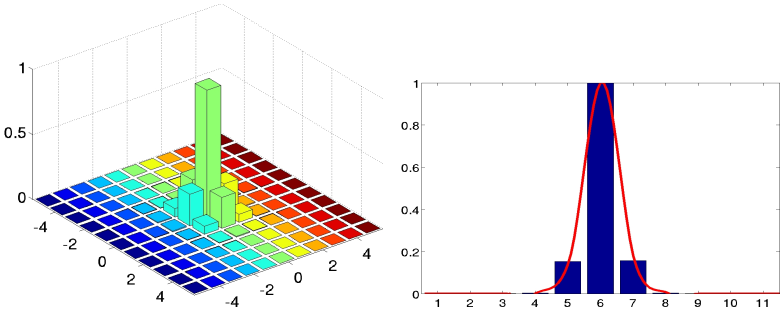
It is observed how for the CCD detector $\sim$80 % of the events are 1-2 pixels size, whilst for the CMOS detector most of the events are in the range of 1-4 pixels size. This



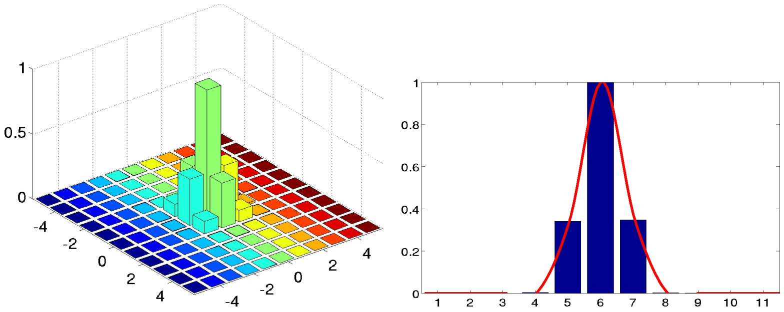
(a) (b)

Figure 5.15: Normalised cluster distribution for the CCD (left) and CMOS (right) detectors, exposed to a ^{14}C microscale for 100 minutes on both detector.

demonstrates an expected phenomenon, showing a larger charge diffusion effect in the CMOS detector, given that the field free layer is thicker (20 μm) compared to the CCD detector (13 μm). The PSF obtained, as has been explained above, is presented in Figure 5.16(a) and Figure 5.16(c) for the CCD detector and the CMOS detector respectively.



(a) (b)



(c) (d)

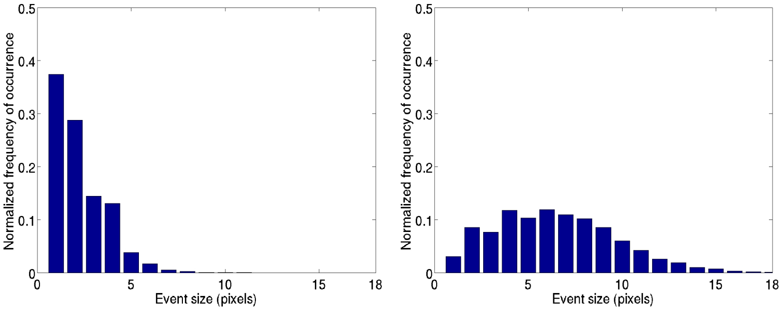
Figure 5.16: PSF for the CCD (a) and CMOS (c) detectors, obtained from the cluster distributions shown in Figure 5.15. Their profiles with interpolated PSFs are shown in (b) and (d) respectively.

The Full-Width-Half-at-Maximum (FWHM) measured from the PSFs shown in Figure 5.16(b) and Figure 5.16(d) are 1.24 and 1.54 times the pixel size respectively. This corresponds to 27.9 μm for the CCD detector (22.5 μm pixel size) and 38.6 μm for the

CMOS detector (25 μm pixel size).

5.4.2 Cluster Distribution with Fixed Dual Thresholding

By applying the algorithm with fixed dual thresholding detailed in Section 4.6.1, a fixed secondary threshold $k_2 < k_1$ is applied in every R_λ where an event has been previously detected. Following this, here is shown how pixels assumed to be misclassified using the traditional method are now classified as part of the cluster hit, therefore resulting in larger clusters. The cluster distributions measured for each detector with fixed dual thresholding applied, are shown in Figure 5.17. In these examples k_1=8 and k_2=4 for both detectors, with a ^{14}C microscale exposed for 100 minutes on the CCD detector and the CMOS detector.



(a) (b)

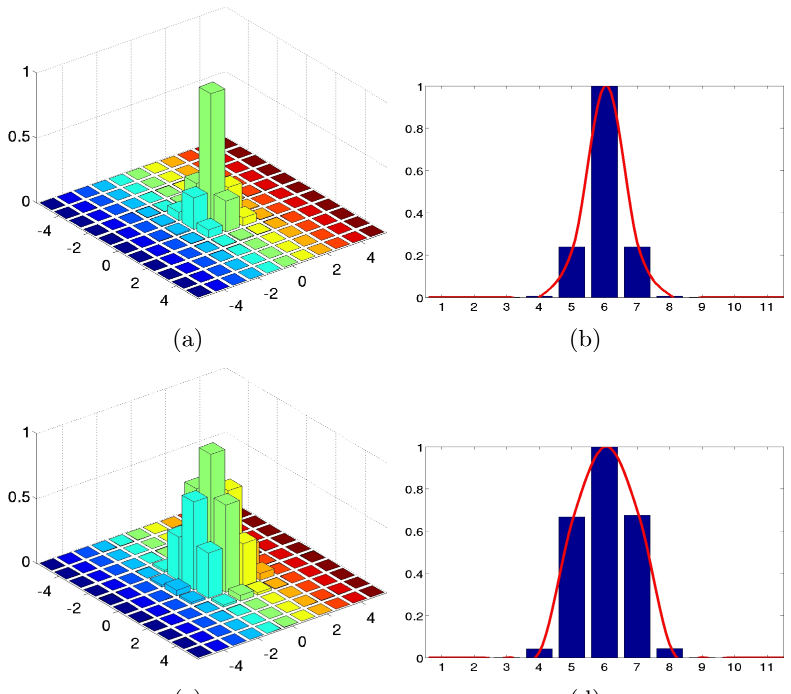
Figure 5.17: Normalised cluster distribution for the CCD (left) and CMOS (right) detectors, exposed to a ^{14}C microscale for 100 minutes, using fixed dual thresholding ROI analysis with k_1=8 and k_2=4 for both detectors.

In this situation it is observed how more pixels are affected by deposited charge compared to the previous scenario. For the CCD detector approximately 66 % of the events are 1-2 pixels large, and for the CMOS detector now events from 1-15 pixels size can be found. Similarly to the previous case the PSF is presented in Figure 5.18.

The FWHM measured from the PSFs shown in Figure 5.18 are 1.35 and 2.38 times the pixel size respectively. This corresponds with 30.3 μm for the CCD detector and 59.5 μm for the CMOS detector.

5.4.3 Cluster Distribution with Adaptive Dual Thresholding

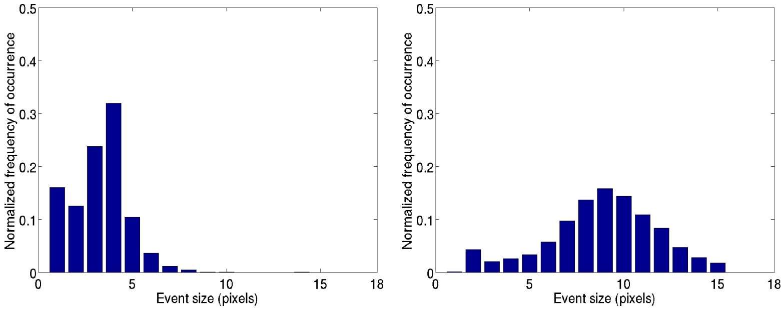
Similarly to the previous section, by applying the adaptive secondary threshold in every R_λ, where an event has been previously detected, as detailed in Section 4.6.2, misclassified pixels using the traditional method are now correctly classified as part of the genuine β event. The difference with the previous method, where a fixed dual threshold is used, is that in this case no fixed parameter is used, and the local threshold used in each R_λ is only dependant in its own statistics. The cluster distributions measured for the CCD detector and the CMOS detector are respectively shown in



(a) (b)

(c) (d)

Figure 5.18: PSF for the CCD (a) and CMOS (c) detectors, obtained from the cluster distributions shown in Figure 5.17. Their profiles with interpolated PSFs are shown in (b) and (d) respectively.



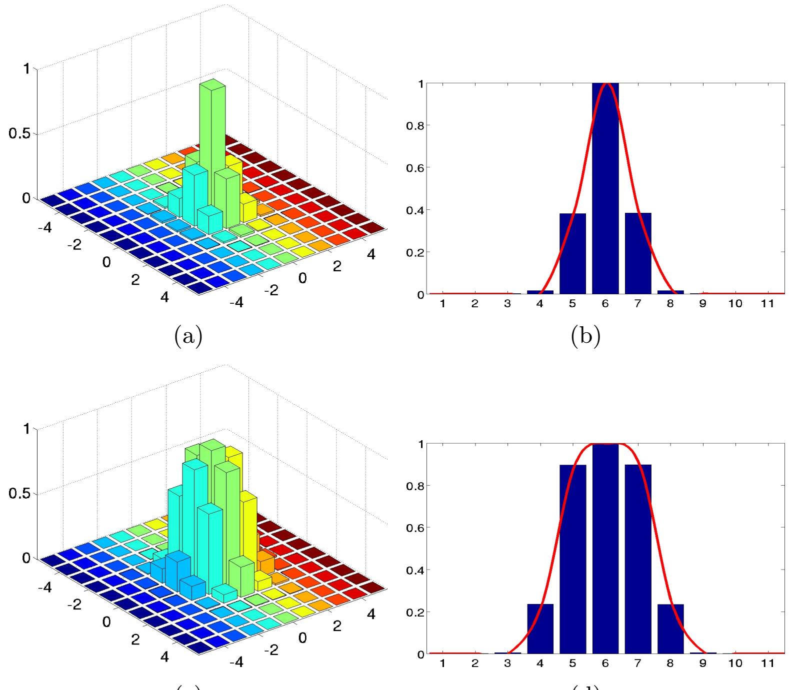
(a) (b)

Figure 5.19: Normalised cluster distribution for the CCD (left) and CMOS (right) detectors, exposed to a ^{14}C microscale for 100 minutes, using adaptive threshold for both detectors.

Figure 5.19. k=8 is used for both detectors in this example, with a ^{14}C microscale exposed for 100 minutes on the CCD detector and the CMOS detector.

There exist more events larger than 2 pixels in this situation. Now only approximately 28 % of the events are 1-2 pixels large for the CCD detector, most of the events (∼85 %) being 1-4 pixels size, while for the CMOS detector most of the events (∼45 %) are 7-11 pixels size. Similarly to the previous case the PSF is presented in Figure 5.20.

The FWHM measured from the PSFs shown in Figure 5.20 are 1.62 and 3.18 times the



(a) (b)

(c) (d)

Figure 5.20: PSF for the CCD (a) and CMOS (c) detectors, obtained from the cluster distributions shown in Figure 5.19. Their profiles with interpolated PSFs are shown in (b) and (d) respectively.

pixel size respectively. This corresponds with 36.4 μm for the CCD detector and 79.5 μm for the CMOS detector.

The spatial resolution measured using the three different approaches described above for the CCD detector and the CMOS detector, are summarised in Table 5.3, where the units are given in pixels, and also in μm depending on the pixel size of each detector.

5.5 Tissue Imaging Using ^{3}H

To demonstrate the potential of digital technology for direct detection with ^{3}H in tissue samples, several ex-vivo brain tissue sections labelled with the selective D_1 dopamine receptor radioligand ^{3}[H]SCH-23390[1] (4 nM) were prepared. These sections were exposed to the back-thinned CCD sensor for 11 hours (Figure 5.21) and to the back-thinned CMOS Vanilla sensor for 48 hours (Figure 5.22(b)), in comparison to the typical length of these experiments (∼5 weeks) when using film. These represent the first demonstrations of ^{3}H tissue imaging with room temperature CCD and CMOS detectors successfully reproducing the main target areas of uptake in a fraction of the usual 5 week exposure time required with conventional autoradiographic film [30].

After the two images shown in Figure 5.22 were obtained, a quantitative analysis was

---

[1] Schering Plough 23990 binding bound with ^{3}H

Table 5.3: FWHM measured for the CCD and the CMOS sensors using the three different approaches

	FWHM in pixel units and μm		
	no-ROI analysis	Fixed dual thresholding	Adaptive dual thresholding
CCD (22.5 μm pixel size)	1.24 (27.9 μm)	1.35 (30.3 μm)	1.62 (36.4 μm)
CMOS (25 μm pixel size)	1.54 (38.6 μm)	2.38 (59.5 μm)	3.18 (79.5 μm)

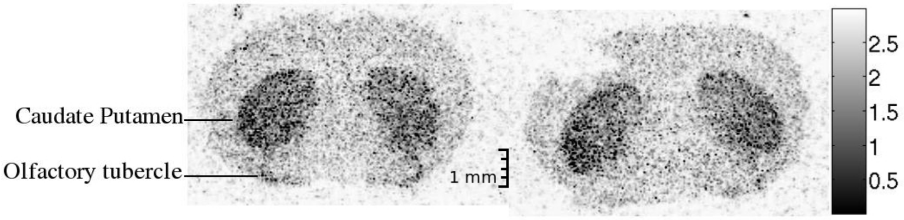
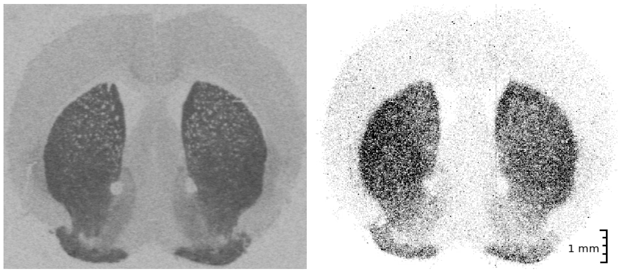


Figure 5.21: ^{3}H tissue image obtained with a CCD at room temperature, for 11 hours, demonstrating tritiated ligand binding to D1 receptors in a coronal mouse Striatum section, from the level of the Caudate, Bregma 0.62 mm (left) and 0.02mm (right).



(a) Film-based image (b) CMOS-based image

Figure 5.22: ^{3}H tissue image obtained with traditional film after 5 weeks (a) and with back-thinned CMOS sensor (b) at room temperature, for 37 hours, demonstrating tritiated ligand binding to D1 receptors in a coronal mouse Striatum section, from the level of the Caudate, bregma 0.86 mm.

manually undertaken. Several ROIs were drawn by a life science expert manually in both sections. These sections, indicated in Figure 5.23, are the Caudate Putamen (CPu), Cingulate Cortex (CgCx), Olfactory Tubercle (Tu), Accumbens Nucleus Shell (AcbSh) and Accumbens Nucleus Core (AcbC). The digital values of such regions were then converted to femtomoles per gram of ligand concentration, ^{3}[H]SCH-23390 in this specific case, and compared with the data extracted from the film autoradiogram (Figure 5.22(a)). In the case of the experiment undertaken with conventional film, six sections of contiguous bregmas were analysed to obtain an experimental mean and standard deviation of concentration in the ROIs under study. The values measured with the CMOS sensor and the conventional film are shown in Figure 5.24. The values of ligand density measured with the CMOS detector were observed to be close to the mean value measured with traditional film and within the range imposed by $\pm\sigma$ (also shown in Figure 5.24).

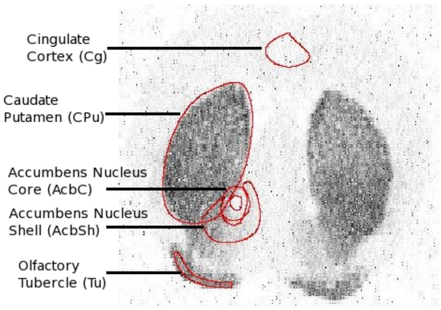


Figure 5.23: Fig. 5.22(b) with the regions under study labelled.

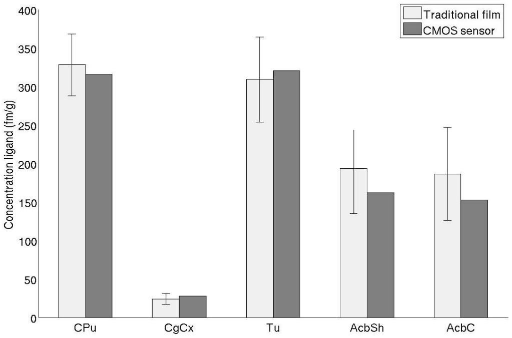


Figure 5.24: Comparison of ligand concentration in fm/g of the ROIs indicated in Figure 5.23 with traditional film after 5 weeks and with the CMOS sensor after 37 hours. The error bar in the measurements shown in the bars of film refer to 1 σ of the mean from the ligand concentration measured in 6 different sections of close bregmas. For the CMOS sensor only 1 of the sections was analysed.

Subsequently the raw images acquired with the CMOS sensor were re-processed to measure the ligand concentration as a function of time within the five aforementioned regions, thus a time-evolution graph of the measured ligand concentration was obtained. This indicated where the ligand concentration measured by the CMOS sensor in each region is quantitatively similar to the traditional film-based image. This graph is shown in Figure 5.25 only for the Caudate Putamen and the Cingulate Cortex as an example. Note the arrows in this graph (Figure 5.25) showing when the measured ligand concentrations of the analysed ROIs in the CMOS sensor reach the mean value minus one standard deviation. The 5 ROIs under study reached a concentration close to the final concentration, and within the range imposed by $\mu\pm1\sigma$ by the experiment undertaken with traditional film, in approximately 13 hours. This represents a reduction of a factor ∼50 in exposure time needed for CMOS direct detection compared to film in order to achieve the similar levels of qualitative accuracy.

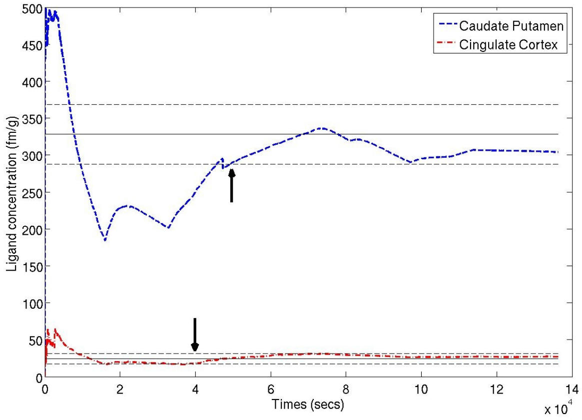


Figure 5.25: Evolution of the radioligand concentration in the Caudate Putamen (blue) and Cingulate Cortex (red) with time. The arrows indicate where the concentrations measured in the CMOS-based image are similar to those measured with traditional film.

In Table 5.4 the time at which the radioligand measured in the CMOS-based autoradiogram reaches the mean concentration measured in film exposed for 5 weeks (*t1*) and the mean minus one standard deviation (*t2*) are shown. The average of these two measurements is shown in *t3*.

Table 5.4: Time necessary by the Vanilla CMOS sensor to obtain comparable measurements of radioligand concentration.

ROI	Name	t1 (hrs)	t2 (hrs)	t3 (hrs)
1	Caudate Putamen (CPu)	19.0	13.7	16.3
2	Cingulate Cortex (CgCx)	12.9	10.4	11.6
3	Olfactory Tubercle (Tu)	14.7	11.2	12.9
4	Accumbens Nucleus Shell (AcbSh)	20.4	12.5	16.4
5	Accumbens Nucleus Core (AcbC)	20.0	12.5	16.3

5.6 Image Quality

It has been mentioned in this work, that a very important limiting factor of a digital detector, for the spatial resolution is the pixel size. To mitigate this limitation, and with the aim of obtaining sub-pixel resolution, two variations of a dual thresholding approach, where more information is extracted from detected cluster hits, have been presented in Section 4.6. Using this information, it has been shown in Section 4.7 how obtaining such extra information in each cluster hit, the estimated β electron trajectories are computed to have a better idea of where the β- particles impinged in the detector, and what trajectory occurred within the active volume of the detector.

To demonstrate the impact of this dual thresholding method in tissue imaging, this has been applied to ^{14}C calibrated microscales and ^{3}H labelled brain tissue sections. In both cases the result demonstrates a smoother film-like appearance without losing visible resolution.

A cropped region of a composite image obtained after 18 hours of a ^{14}C source exposed to the CMOS detector, corrected using the initial dark current noise and FPN correction method and the new fixed dual thresholding method (k_1=8,k_2=4), are shown in Figure 5.26. In Figure 5.26(a) the typical pixelated appearance is observed, whilst in Figure 5.26(b) the same image with event hits reconstructed as explained above is shown, exhibiting a more similar film-like appearance. To study the effect of this process on the spatial resolution 400 rows in each image have been averaged to obtain a 2.7 mm long cross-profile of the edges, shown with a red box in Figures 5.26(a) and 5.26(b). The resulting superimposed profiles are shown in Figure 5.26(c), where the blue plot represents the profile corresponding to the pixelated image (Figure 5.26(a)), and the red plot represents the profile corresponding to the processed image (Figure 5.26(b)). No degradation of the edges is observed in the processed profile shown in red.

In the case of ^{3}H brain tissue imaging the resultant image after ROI analysis has been applied is shown in Figure 5.27(c), alongside the pixelated image (Figure 5.27(b)) and the conventional film image obtained after 5 weeks of exposure time (Figure 5.27(a)).

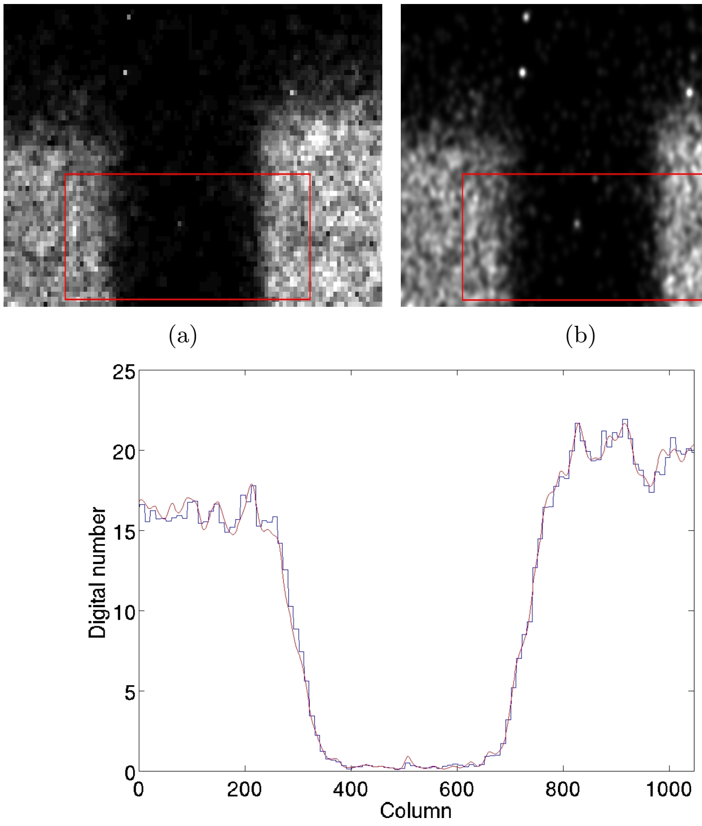
To better observe in more detail the improvement introduced by the hit reconstruction explained above, a zoom image of two different key functional regions, corresponding to the top part of the Caudate Putamen and to the Olfactory Tubercle (see Figure 5.23), present in the section shown in Figure 5.27 are shown in Figures 5.28 and 5.29. In these two images it can be observed how the image where hit reconstruction has been applied (c) is more similar to the film image (a), and show some structures that are better resolved compared to the pixelated image (b).

5.7 Conclusions

In this chapter three methods have been detailed to detect β- event from raw images acquired with a CCD and a CMOS sensor. These images have been shown to be corrupted mainly with dark current noise and FPN.

The main difference between the three methods is the way these subsample each event, having the sensitivity fixed by the thresholding parameter k used in equation 4.2. Depending on the method used to subsample each event will give more detailed information of the deposited charge in a cluster hit. If more pixels affected by the deposited charge are available, then more accurate centroiding techniques can be used to achieve sub-pixel spatial resolution.

This technique is applied to ^{14}C microscales autoradiographic data, showing a cross profile of the raw composite image, and the processed composite image using the fixed



(a) (b)

(c)

Figure 5.26: Composite image obtained after 18 hours of a ^{14}C source exposed to the CMOS detector, applying no ROI analysis (a) and using the fixed dual thresholding method (k_1=8,k_2=4) (b). A more film-like appearance is observed in (b) compared to (a) where the typical pixelated appearance is exhibited. Cross-profiles of 400 averaged rows (shown in a red box in (a) and (b)) are shown in (c), to demonstrate the absence of edge degradation. These cross-profiles correspond with a real length of 2.7 mm.

dual thresholding method and subsequent hit reconstruction. This demonstrates that applying this hit reconstruction techniques no spatial resolution is lost, giving the composite image a more film-like appearance.

This technique has also been applied to thin tissue autoradiographic data, showing better resolved small structures in the images where hit reconstruction techniques have been used.

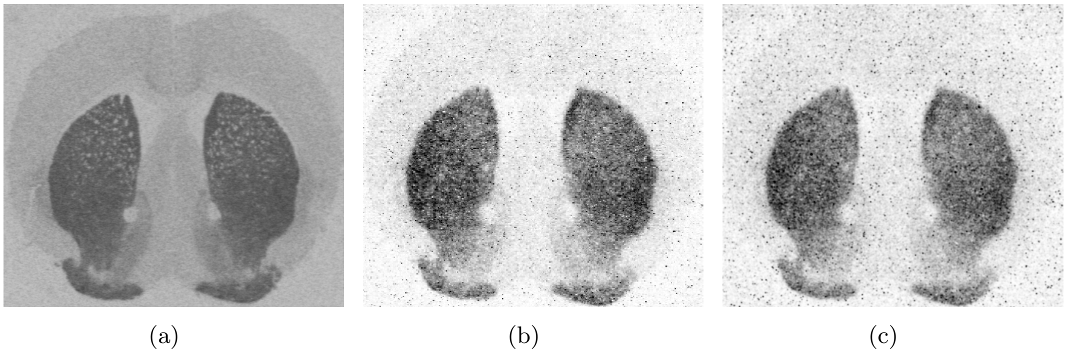


Figure 5.27: Composite image of a ^{3}H labelled brain tissue section obtained after 5 weeks exposed to conventional film (a), and after ∼37 hours exposed to the CMOS detector, using the fixed dual thresholding method (k_1=4,k_2=2) (b) and corresponding image after hit reconstruction has been applied (c). A more film-like appearance is observed in (c) compared to (b) where the typical pixelated appearance is exhibited.

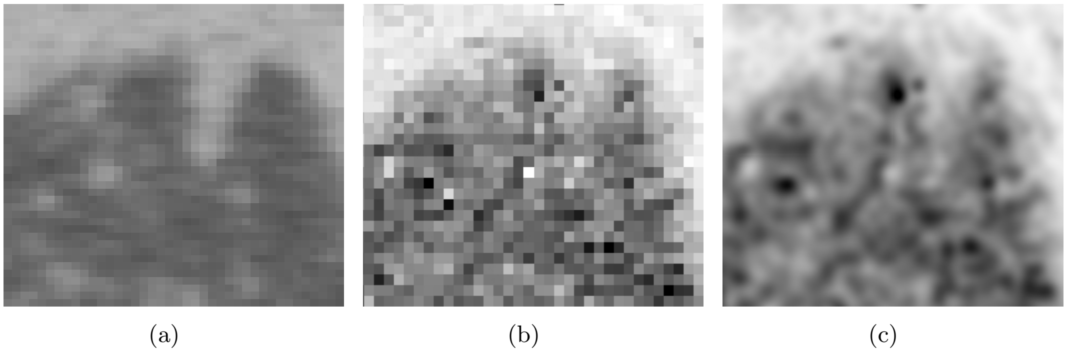


Figure 5.28: Film image (a), raw image (b) and reconstructed image (c) of the top part of the Caudate Putamen region, corresponding to Figures 5.27(b) and 5.27(c) respectively.

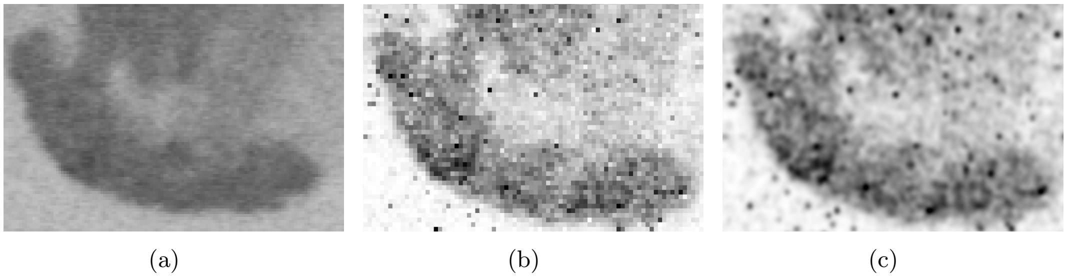


Figure 5.29: Film image (a), raw image (b) and reconstructed image (c) of the Olfactory Tubercle region, corresponding to Figures 5.27(b) and 5.27(c) respectively.

Chapter 6

Monte Carlo Simulations

6.1 Previous Work

Monte Carlo methods have been a very important tool for modelling complex systems characterised by a stochastic nature. These systems usually have many variables so that an analytical solution can not be easily obtained. Monte Carlo methods are therefore a type of numerical integration where many trials are calculated and a gross number derived to describe the outcome.

For physics-based simulation, there are several Monte Carlo toolkits available: Geant4 [38], MCNPX [27], Penelope [11] and EGSnrc [149] are among the most important implementations. For this study Geant4 has been used, together with its low energy package, because it has reliable accuracy in modelling interaction physics with electrons down to 250 eV. This is particularly relevant for studying ^{3}H imaging (average energy 5.9 keV and maximum energy 18 keV), along side higher energy emitters.

Despite the large body of work completed using Monte Carlo methods in medical imaging, specially in PET and SPECT, the use of this in autoradiography is relatively scant. Examples include Geant4 to simulate the detection of positrons emitted by a sample of ^{18}F-FDG with a storage phosphor detector [157]. MCNP5 has been used to simulate the heterogeneous dose distributions of ^{90}Y, ^{177}Lu and ^{111}In in rat kidneys autoradiograms [89]. The charge collection process of the MIMOSA CMOS architecture is investigated [44] using the commercial Monte Carlo toolkit ISE-TCAD [2]. This same sensor has also been studied with MCNPX analysing the efficiency of the sensor interacting with α-particles [128]. Low energy electron interaction in silicon has also been studied with the hybrid CMOS sensor Medipix2 [117] for electron imaging and with the MIMOSA CMOS sensor [43] for autoradiography using the CASINO Monte Carlo package, showing very superficial results as this Monte Carlo package has very limited functionality. The effect of backscattering in electron microscopy has been studied using customised Monte Carlo simulations, and compared to experimental data [121]. A system based on a double sided silicon strip detector with a synthetic collimator has been simulated using EGSnrc [138], to understand the interactions of low-energy gamma photons (^{125}I and ^{123}I) with the collimator and the detector in small animal imaging. There is however little evidence of Monte Carlo simulation work being used to understand the imaging trade-offs of particular imaging geometries. This is the starting point for this work.

6.2 Geant 4

Geant4 is a complex Monte Carlo package developed in C++, structuring all the different modules interacting in the simulation in C++ classes. Different modules, always present in the simulations undertaken in this work, can be enumerated as shown below:

- *Detector Construction*: This module defines the architecture and materials of the different parts, also known as volumes, of the detector. This module also defines the sensitive volumes. This means that when the simulated particles interact with the volumes of the detector, no measurement is taken unless the volume of the current interaction is defined as sensitive volume.

- *Physics List*: This module defines the particles that can be simulated and the physics processes associated to each particle. When the simulation is running, and the particles are interacting with matter, the measurements are taken in discrete form, hence after each certain distance the particles are stopped and polled regarding its status (type of particle, energy deposited, energy left, location, etc). In this module a threshold can be set to limit the number of secondary particles generated due to ionisation of e.g. silicon atoms. This threshold can be set by limiting the minimum distance that a particle has to travel, or the minimum energy that it has to contain to be tracked. Therefore if a β electron or a secondary electron generates more secondary electrons, these have to be able to travel a minimum distance or contain a minimum energy to be tracked. The purpose of this mechanism is to limit the number of generated secondaries to control the computing load of a simulation, or to avoid the infrared divergence effect, where thousands of particles with very low energy might be generated. On the other hand this mechanism has a significant impact on the spatial resolution and energy deposition of every particle, limiting the accuracy of the simulations.

- *Stacking Action*: In this module different particles can be prioritised or even *killed* at a certain step if necessary. This module is of great usefulness, because there may be particles not important for the simulation, that are generated and can make the simulation heavier and slower, thus these particles can be *killed*. The three possible classes in which a particle can be classified, if not *killed*, are *urgent*, *waiting* and *postpone-to-next-event*.

- *Analysis manager*: This module is the core of the simulation. When an event physically interacts with a sensitive volume a snapshot of this event, also called hit, is taken so position, time, momentum, energy, deposited energy and geometrical information can be measured. These hits are collected in different lists for each event for further analysis. Memory for this information is allocated in the heap. Therefore the following has to be taken into account: the number of particles to simulate, the step length and the information to be saved for each hit

> not to fill the heap. Each event has an *Initialise* method, a *ProcessHits* method and an *EndOfEvent* method. The initial properties of a given particle or manually user defined settings can be set in the *Initialise* method. All the different measurements that can be taken from a particle while it traverses matter are taken in the *ProcessHits* method. Finally, the final properties of a particle can be measured in the *EndOfEvent* method.

Once the simulation is coded with all the necessary modules compiled and linked the simulations are controlled via macro files. These files are simple scripts that define the shape, position and nature of the radioactive source, the visualisation driver to be used (if any) and the number of particles to be generated. Note that by using these macro files it is possible to define different architectures and generate particles for each architecture in a single simulation run. This allows results to be compared between different detector architectures in a single simulation (increasing computational load but reducing execution time). Macro files can be called from other macro files allowing increased complexity with higher hierarchy. Different levels of verbosity can also be used to control the amount of information that is presented on the screen at run time, which can be very useful for debugging purposes.

Geant4 allows one to use different random number engines. The default random number generator is the original universal random number generator Ranmar (proposed by Marsaglia and Zaman), but the random number generator used in this work is the Ranecu engine [73], due to its easiness to handle the internal status of the engine to repeat a certain event or experiment. The initialisation is carried out using a Multiplicative Congruential generator, with seeds taken from a seed table given an index. The period of the Ranecu engine is 10^{18}, shorter in comparison with the extended Mersenne/Twister engine (10^{6001}) or the Ranmar engine (10^{43}), all of them well below the longest experiment presented in this work where $3.2\ 10\ ^{6}$ random particles were generated. The speed of each random number generator, which is of great importance when generating millions of random numbers as in this work, measured in millions of generated uniform random numbers per second in a Pentium Pro 200 Mhz is 1.5, compared to 2.5 the Mersenne/Twister engine and 1.6 the Ranmar engine [71].

6.2.1 General Particle Source

The radioactive sources, in the aforementioned macro files, are generated using the General Particle Source (GPS) [59].

This package allows the user to define different energy spectra (linear, exponential, Gaussian), different angular distributions (directions in which the particles emanate from the source), different spatial samplings (2D or 3D surfaces) and multiple independent sources. This is very useful for example to simulate two different sources with a gap in between, and observe whether the gap is visible depending on the particle, energy, thickness of the sources, etc. Using this package a wide variety of source shapes can be created such as spheres, discs, boxes and many more, the point source being the default configuration. The size and position of this is also defined with this package.

The main advantage of using this package is that Geant4's particle generator is suitable for simple studies, but with GPS very complex setups can be created, and even typical continuous β spectra or user defined spectra can be used.

6.2.2 Visualisation Drivers

One of the main parts of any Monte Carlo package is the visualisation of the detector geometry for visual inspection of the simulation set-up. Geant4 is able to work with several different visualisation drivers. Figures 6.1, 6.2, 6.3 and 6.4 represent some exemplar geometries with some of the most common drivers.

- OpenGL. OpenGL is a standard 2D/3D graphics interface. This is a standard system component in Linux and Windows.
- RayTracer. This driver performs ray-tracing visualisation using the tracking routines of Geant4. The rendering can be aborted at any time before it finishes, to retry rendering from a different point of view. This can save a lot of time when debugging complex geometries.
- HepRep/Wired3. HepRep creates a HepRep XML file in HepRep1 format, prepared to be viewed with the Wired3 HepRep browser. This browser shows only certain parts of the simulations such as only certain particles or certain volumes of the detector. This can be very useful when very complex simulations are being visually inspected.
- VRML. This format is a standard format for representing interactive 3D vector graphics. Any VRML browser, such as *octagaplayer*, can be used.
- DAWN. This is a renderer which reads 3D geometrical data and visualise them. It is able to show vectorised graphics in high detail. A couple of useful tools can be used with this driver. DAWNCUT can perform a planar cut on a geometry to better observe its inner structures. DAVID is a tool that automatically detects overlapping volumes and highlights them to make the user aware of the error.

6.3 Comparison with Experimental Data

The validation of any Monte Carlo package is a critical issue to assure reliable results are produced. A thorough validation of the electromagnetic model applied to photons, electrons, protons and alpha particles in Geant4 is presented in [6]. In the aforementioned article the standard package, the low energy package and an alternative low energy package based on the Penelope analytical approach [11] are considered. The conclusion for electrons, the particle of interest in this work, is that the three packages show an excellent agreement in the energy range 1 keV to 100 TeV when compared to the NIST reference data [126].

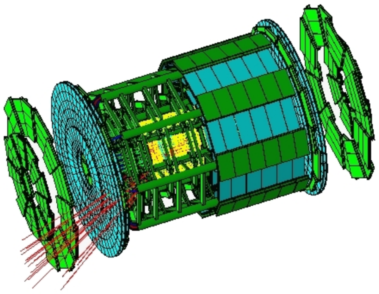


Figure 6.1: DAWN example

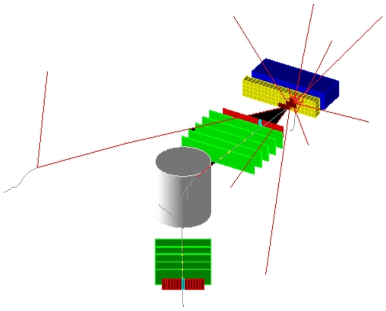


Figure 6.2: OpenGL example

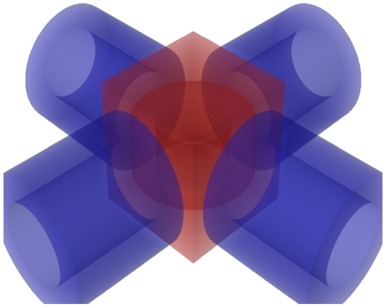


Figure 6.3: RayTracer example

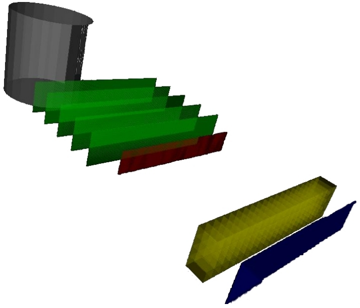


Figure 6.4: VRML example

To validate the results obtained from the experimental setups shown here, the electron stopping power and continuously slowing down approximation (CSDA) ranges have been measured, and these have been compared with those tabulated in the ICRU Report 37 [72]. In this work only the CSDA range of electrons from 10 keV up to 500 keV is considered, as this corresponds to the range of energies of the typical isotopes used in β- autoradiography. The measured range for each energy (blue squares) with its corresponding error margin (standard deviation of the measurement) is compared to the range obtained from ESTAR[1] (red squares) in Figure 6.5. The error measured is very low for low energies (∼1 %) but this increases somewhat for higher energies. The error bars shown in Figure 6.5 represent the standard deviation of the electron ranges measured with Geant4.

This experiment was undertaken with multiple scattering and the energy loss fluctuations that electrons suffer while they traverse the detector disabled, in agreement with ESTAR [6, 126]. To neglect the energy loss fluctuations is the same as to assume that the rate of energy loss at every point along the track is equal to the same as the total stopping power.

6.3.1 Validation with Experimental Data

A variety of detector architectures shown in this work, have been simulated to demonstrate that the architectures and the physical processes involved in the detection of β

---

[1] Stopping-power and range for electrons [126]

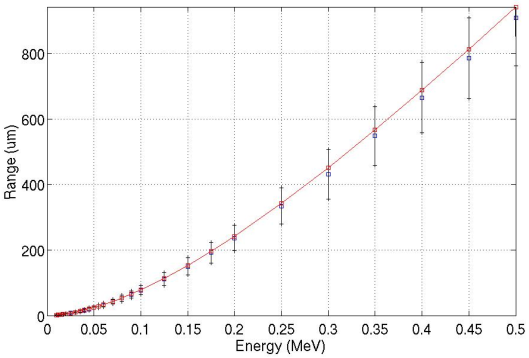


Figure 6.5: Measured electron range (blue squares) using Geant4 and NIST electron range (red squares) in silicon as a function of the electron incident energy.

particles are well understood. The detectors used in these simulations are typically comprised of (from bottom to top) a substrate, epitaxial layer (sensitive layer) and overlying structures. On top of these layers it is included an air gap between the top surface and the sample, the tissue sample itself and finally the glass microscope slide which is used to carry the tissue sample, as shown in Figure 6.6.

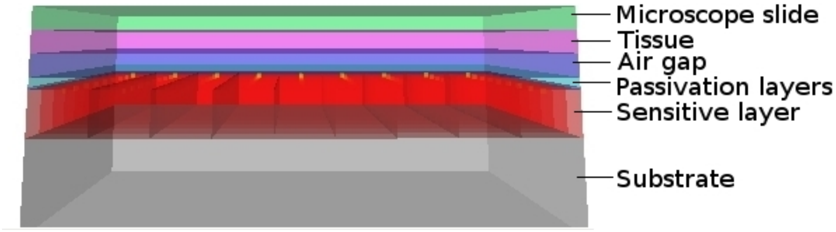


Figure 6.6: 3D cutaway of a digital detector used in these Monte Carlo simulations.

The tissue thickness can take two different values, either 120 μm (microscale thickness), for validation purposes, or 20 μm which is the typical tissue thickness. When a point source is used it is placed at the bottom surface of the tissue layer, to consider the best situation.

The dimensions of the common structures in CCD and CMOS detectors are shown in Table 6.1.

Table 6.1: Common dimensions used in the CCD and CMOS architectures

Layer	Glass slide	Tissue	Air gap	Sensitive layer	Substrate
Thickness	1.0 mm	120.0 μm 20 μm	variable	20.0 μm	500.0 μm

The structures where CCD and CMOS detectors differ are the pixel size and the structure and thickness of the overlaying layers. The dimensions are shown in Table 6.2.

A special mention has to be made for back-thinned devices (see Figure 3.2). In this case the typical top overlaying layers of a front-illuminated device, are located between

Table 6.2: Dimensions used in the CCD (left) and CMOS (right) architectures

Structure	Dimension
Pixel size	22.5 μm
Poly-Si	0.5 μm
SiO_2	0.1 μm
Si_3N_4	0.1 μm

Structure	Dimension
Pixel size	25 μm
SiO_2	4 μm
Si_3N_4	1 μm

the sensitive layer and the substrate. An extra passivation ~100 nm layer has to be added on top of the active layer, representing the residual SiO_2 layer remaining on the epitaxial layer after etching the substrate in the back-thinning process.

Several processes have been implemented in the simulations undertaken in this work, to validate the experimental data, to obtain as much realistic results as possible. The processes included in the simulations are detailed below, and also depicted in a work flow in Figure 6.7.

- Initial Monte Carlo Geant4 experiments, simulating the charge deposition process in silicon.
- Diffusion is added to the simulated data.
- After the charge diffusion is implemented, statistical noise, as measured from experimental data, is added to the entire noiseless image, resulting in a realistic simulated image.
- The Mean Image and the Standard Deviation Image used to corrupt the simulated data are directly measured on the detector. The Mean Image is corrected based on the difference between the mode of the Mean Image and the mode of the realistic simulated image, as described in Section 4.3.1. The realistic simulated image is corrected with the corrected mean image, resulting in a corrected image. This is subsequently thresholded. The threshold is directly obtained from k times the Standard Deviation Image.

Initial Monte Carlo Simulation

First the Monte Carlo simulation is undertaken using Geant4, implementing the corresponding structures of the detector. A 225x225x120 μm^3 piece of tissue of ^{14}C, covering 10 (225 μm/22.5 μm per pixel) and 9 (225 μm/25 μm per pixel) pixels for the CCD and the CMOS respectively, is placed on top of the *air gap* layer. A total of 10^6 particles have been simulated.

When a β electron or an electron generated by ionisation interacts with the silicon in the active layer some information is obtained and saved to file from the particle at the i*th* step: kinetic energy, deposited energy and position in euclidean coordinates within the detector. An example of the deposited energy by a β electron with an initial kinetic energy of 66 keV, in a cluster of pixels, is shown in Figure 6.13(a).

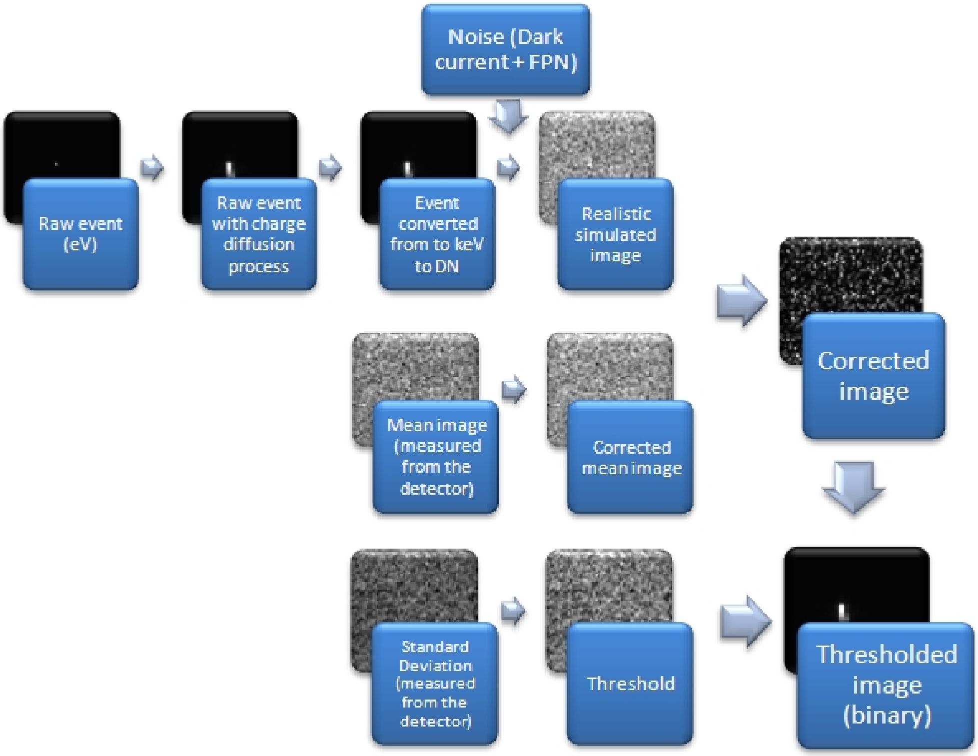


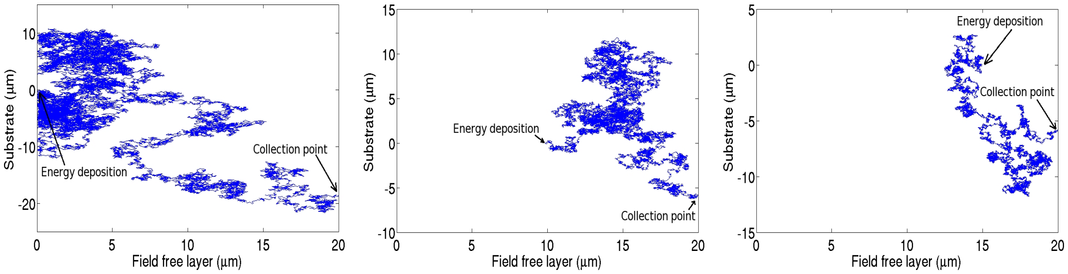
Figure 6.7: Image chain process of the noise addition and further noise correction of simulated data.

Charge Diffusion Implementation

As Geant4 does not implement the charge diffusion process that deposited charge suffers in silicon (see Section 3.1), this is implemented as a second step using a tailored script written in Matlab. The initial approach was to compute σ_{tot} (equation 3.9) for each step i based on the coordinates within the detector. This was obtained by computing first σ_i (equation 3.1), σ_d (equation 3.4) for particles in the depleted layer in the CCD detector and σ_{ff} (equation 3.8) for particles in the field free layer of the CCD detector and the epitaxial layer of the CMOS detector.

From the deposited energy at the *ith* step, the number of e-/hole pairs was computed (3.6 eV/e-h pair). Assuming that N e-h pairs were created, a Gaussian distribution with standard deviation σ_{tot} (equation 3.9) was randomly sampled extracting N samples. An example of the diffusion process is shown in Figure 6.13(b), where the deposited charge shown in Figure 6.13(a) has diffused following the diffusion theory detailed in Section 3.1.

It has been observed in this work how σ_{ff} is usually the most significant factor in equation 3.9, being sometimes σ_i (for low energies) and σ_d negligible (see Signal Generation in Silicon-based detectors in Section 3.1). To demonstrate the validity of equation 3.8 the experiment set up by Janesick [78] has been reproduced in this work. In this validation test 2000 electrons are simulated in a semi-infinite slab representing a field free layer. These electrons are collected in one of the boundaries, simulating the effect of the depleted layer in a CCD pixel, and reflected in the opposite boundary, simulating the effect of the substrate. The resulting charge distribution in the depleted layer boundary is approximated as Gaussian with an absorption depth dependant on the standard deviation σ_{ff}. The resulting σ_{ff}, after simulating a random walk for electrons originated in a semi-infinite silicon slab at different absorption depths, for different field free layer thicknesses, has been obtained. Three examples of one electron placed in a semi-infinite slab with a 20 μm thick field free layer at 20, 10 and 5 μm absorption depths are shown in Figure 6.8. In this figure the y axis on the left represents the boundary with the substrate, which reflects all the charge that reaches this point, and the y axis on the right represents the boundary with the depleted layer (CCD technology) or the charge collection diode (CMOS technology).



(a) Absorption depth 20 μm (b) Absorption depth 10 μm (c) Absorption depth 5 μm

Figure 6.8: Monte Carlo simulations of one electron showing a random walk in a 20 μm thick field free layer at 20 μm, 10 μm and 5 μm absorption depths.

The pseudo-random numbers that are generated to give direction to each electron are in the $[-1, 1]$ range, this has to be divided by a certain step size. It was observed how to obtain accurate results the step size at which each electron moves in the silicon was a critical parameter. A step size of $\frac{z_a}{500}$, being z_a the absorption depth, was observed to exhibit good accuracy, i.e. each step size has a maximum length of $\frac{z_a}{500}$ μm. This results in a longer step size for those events that deposit energy far from the charge collection points (Figure 6.8(a)$\Rightarrow$0.04 μm step size) than those that deposit energy close to the charge collection points (Figure 6.8(c)$\Rightarrow$0.01 μm step size).

An example of 10^5 electrons deposited in a 0.05 μm thick field free layer at an absorption depth of 0.05 μm, representing the worst possible situation where more diffusion occurs, is shown in Figure 6.9. The standard deviation measured from the data is 0.025 μm, half the field free layer, as equation 3.8 indicates. A Gaussian distribution with 0.025 μm standard deviation is plotted on top. It can be observed how the simulated bell-shaped charge distribution does not fit on the ideal Gaussian distribution very accurately, showing maximum discrepancy at the tails.

Similarly, 10^5 electrons have been simulated in a 20 μm thick field free layer at an absorption depth of 20 μm, as shown in Figure 6.10. The standard deviation measured from the simulated deposited charge distribution is 13.42 μm. In this particular case, this value is slightly higher than the expected half field free layer thickness, 10 μm. A Gaussian distribution with the aforementioned standard deviation is plotted on top in Figure 6.10.

In Figures 6.9 and 6.10 certain dissimilarity between the obtained simulated charge distribution and the assumed Gaussian distribution is observed. This dissimilarity,

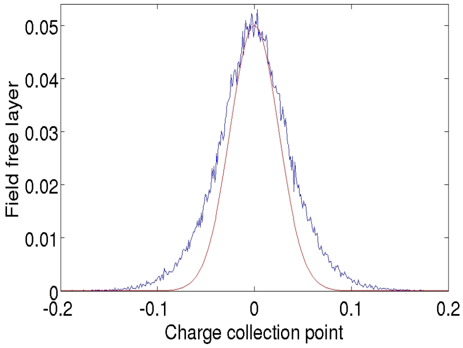


Figure 6.9: Example of 10^5 electrons projected on the charge collection point in a 0.05 μm thick field free layer at an absorption depth of 0.05 μm. A Gaussian distribution with 0.025 μm standard deviation is plotted on top, with a width as prescribed by Janesick [78].

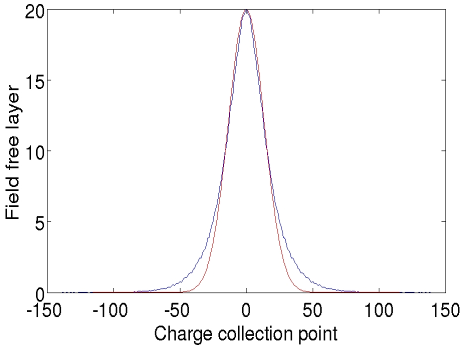


Figure 6.10: Example of 10^5 electrons projected on the charge collection point in a 20 μm thick field free layer at an absorption depth of 20 μm. A Gaussian distribution with 13.42 μm standard deviation is plotted on top, as predicted from [78].

showing a slightly more peaked charge distribution and with wider tails, has been previously noted by other authors, such as [25, 70, 134, 143]. It has to be noted how the standard deviation obtained for the case of a relatively thick detector (20 μm field free layer) is higher than the expected standard deviation. This implies that, for thick detectors, equation 3.8 is not as accurate as it is for thin detectors.

Besides the Monte Carlo simulation approach used by Janesick [78], detailed analytical approaches to study the charge diffusion in field free silicon (applied to X-ray radiation in CCD technology) can be found in [70, 134]. Very little work has been undertaken to validate these analytical approaches though. Relatively recent is the work of Prigozhin *et al* [143] where the analytical solution detailed by Pavlov *et al* [134] is successfully validated, using a cooled CCD detector (all sources of noise except read-out noise are removed) exposed to a ^{55}Fe source. By isolating the signal deposited by X-ray photons absorbed in the field free layer at different depths of interaction, the charge deposition shape of such events is studied and compared with the analytical distribution proposed by [134]. Another conclusion of Prigozhin's work is that the charge diffusion suffered in the depleted layer in CCD detectors can be neglected without greatly distorting the final charge distribution. The resulting charge distribution observed by Prigozhin *et al*, assuming recombination in the field free layer is negligible (diffusion length is much higher than the thickness of the field free layer) and the substrate works as a reflecting body (field free layer is p doped and substrate is p+ doped), follows equation 6.1:

$$q_{ff_1}(x) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \frac{z_a - 2mz_{ff}}{p_m(x)^3} \tag{6.1}$$

where z_a is the absorption depth, z_{ff} is the thickness of the field free layer and p_m is defined by equation 6.2:

$$p_m(x) = \sqrt{x^2 + (z_a - 2mz_{ff})^2} \tag{6.2}$$

This charge distribution has been compared with Hopkinson's *et al* work [70] obtaining very similar results. The charge distribution as defined by Hopkinson's *et al* follows equation 6.3:

$$q_{ff_2}(x) = \frac{z_a}{2\pi u(x)^3}(1 + \frac{u(x)}{\frac{1.15z_a}{\frac{z_a}{x_{ff}}+0.13}})exp(-\frac{u(x)}{\frac{1.15z_a}{\frac{z_a}{x_{ff}}+0.13}}) \tag{6.3}$$

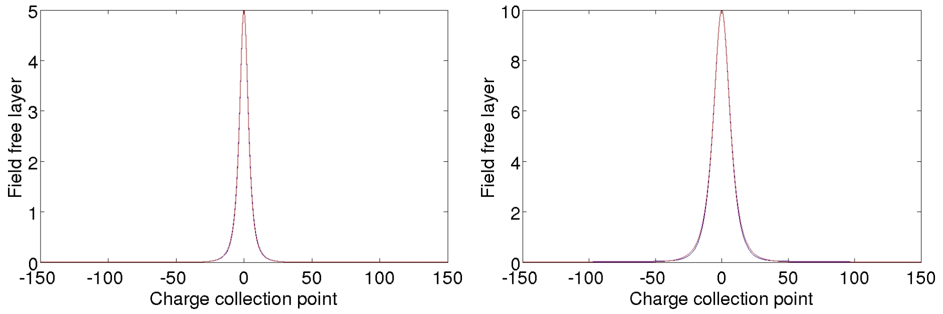
where $u(x)$ is described by equation 6.4:

$$u(x) = \sqrt{(x^2 + z_a^2)} \tag{6.4}$$

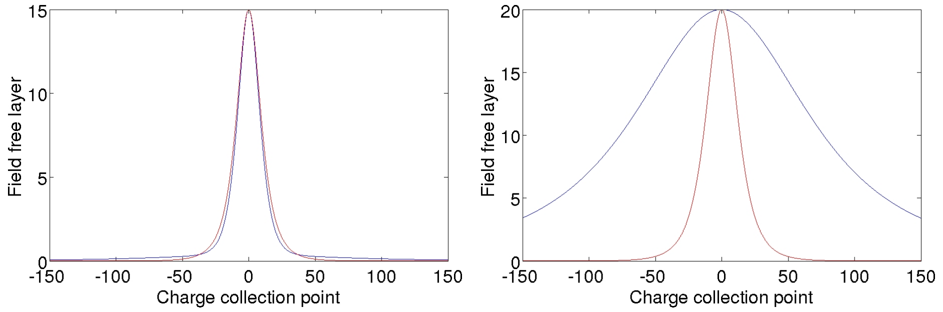
If both charge distributions are plotted together, as shown in Figure 6.11, it is observed how the results obtained by both equations are very similar in almost the whole field free layer thickness, but as the absorption depth approximates to the boundary with the substrate dissimilarities appear, as shown in Figure 6.11(d).

For comparison purposes with the approximated charge distribution proposed by Janesick [78] the same studies are shown in Figure 6.12 with the simplified Gaussian charge distribution plotted on top. Figure 6.12 demonstrates that the Gaussian approximation approximately fits Prigozhin's and Hopkinson's charge distributions for short and medium absorption depths, but differs greatly for deep absorption depths. It can be observed how the evolution of the analytical charge distributions is mainly based on a more peaked distribution and wider tails as the absorption depth increases.

After this study it has been observed how Janesick's model can be considered as a good approximation for thin detectors. For thick detectors, as the charge is deposited further from the charge collection element, the differences between Janesick's model and the real charge distribution measured by Prigozhin have higher differences. Hopkinson's model also seems to underestimate the charge distribution in the tails compared to Prigozhin's model. Finally Prigozhin's model has been implemented in this work given that it is the only work found in the literature that compares the analytically developed model with real data obtaining a good level of agreement.



(a) 5 μm (b) 10 μm



(c) 15 μm (d) 20 μm

Figure 6.11: Analytical charge distributions of an X-ray absorbed at different absorption depths in a 20 μm thick field free layer, obtained with Prigozhin's equation 6.1 (q_{ff_1}) shown in blue, and obtained with Hopkinson's equation 6.3 (q_{ff_2}) shown in red. Excellent agreement is shown at 5 μm and 10 μm, with minor divergence at 15 μm. At 20 μm there are major differences between the two models.

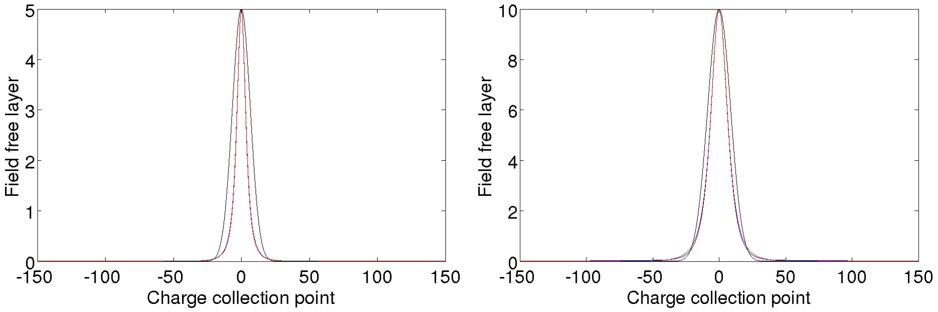
Detector Noise Implementation

Having validated the simulated charge deposition and diffusion processes, it was also necessary to consider adding realistic noise to the synthetic noiseless event data, as this was a significant source of potential corruption.

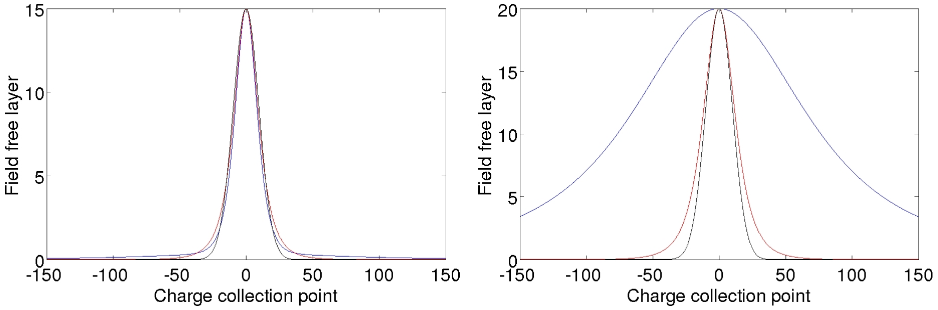
As explained in Section 4.3.1, prior to the correction of the noisy images acquired with the CCD or the CMOS detector, a reference set of blank frames is first necessary. The mean ($\mu_{i,j}$) and standard deviation ($\sigma_{i,j}$) of each pixel is then obtained from this data set.

For the simulations, this utilises *mean* and *standard deviation* images directly from the experimental data. The noiseless image has initially every pixel value in eV, but as the noise is added in digital numbers (DN), every pixel value is transformed to DN. For this transformation the ratio between the number of electrons per DN (e-/DN), i.e. gain, is necessary. In order to obtain this, the detector was calibrated as explained in Section 5.2 with an ^{241}Am point source. The resulting gain for the CCD detector is 250 eV/DN, and for the CMOS detector is 38.8 eV/DN.

After the image has been converted to DN, the noise $n_{i,j}$ is then independently obtained for each pixel, sampling a random number from a Gaussian distribution with pixel specific $\mu_{i,j}$ and $\sigma_{i,j}$ as parameters. For a given pixel $x_{i,j}$, the noise $n_{i,j}$ is added to obtain an observed pixel value $y_{i,j}$, given by equation 6.5.



(a) 5 μm (b) 10 μm



(c) 15 μm (d) 20 μm

Figure 6.12: Analytical charge distributions of an X-ray absorbed at different absorption depths in a 20 μm thick field free layer, obtained with Prigozhin's equation 6.1 (q_{ff_1}) shown in blue, obtained with Hopkinson's equation 6.3 (q_{ff_2}) shown in red, and approximated Gaussian distribution proposed by Janesick shown in black.

$$y_{i,j} = x_{i,j} + n_{i,j} \tag{6.5}$$

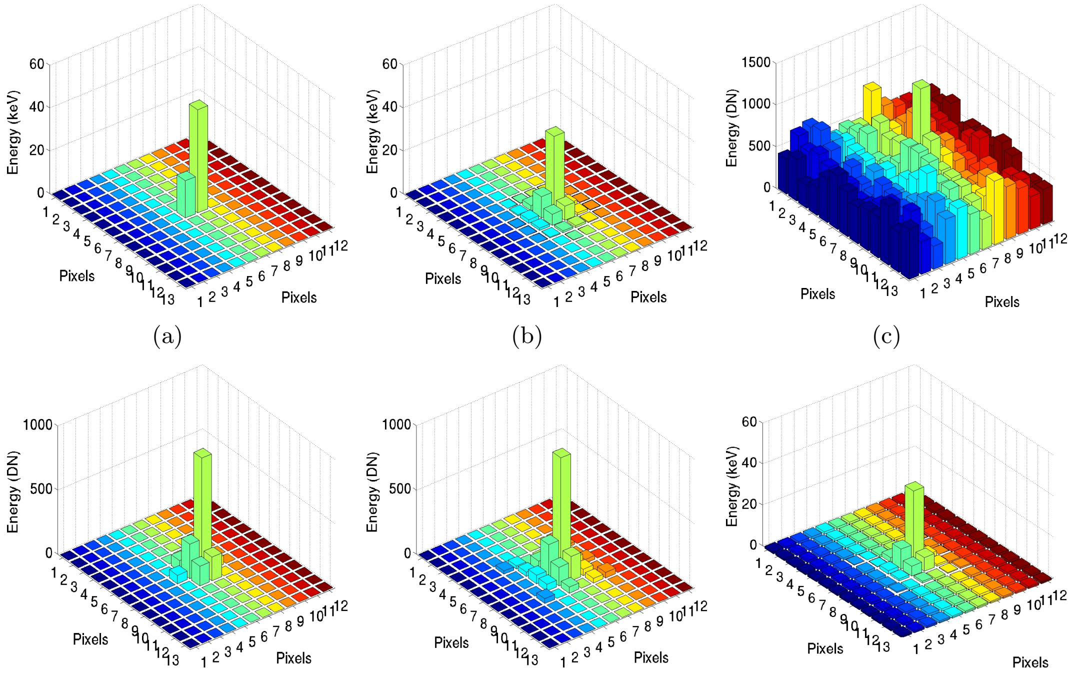
where $n_{i,j}$ is a random number sampled from a Gaussian distribution with mean and standard deviation $\mu_{i,j}$ and $\sigma_{i,j}$ respectively, obtained from experimental data. An example of the resulting noisy image y is shown in Figure 6.13(c).

Image Correction

Three different correction methods have been detailed in Section 4.6. These methods have been similarly implemented in order to fully validate the Monte Carlo simulations, with the standard correction method explained in Section 4.3.1 (Figure 6.13(d)) and the two novel corrections explained in Section 4.6. The result of generating synthetic noisy images, and then applying the dual thresholding method to Figure 6.13(d), using a fixed secondary threshold of k_2=4, is shown in Figure 6.13(e).

After the correction has been applied the image is converted from relative DN values back to energy (in eV) using the same calibration as before (Figure 6.13(f)). Finally the energy deposited by each β electron and the size of each hit (in pixels) are computed to obtain the final energy spectrum and cluster distribution, to be compared with experimental data.

Note the change of unit in Figure 6.13 (b) to (c) where the pixels have been converted from keV to DN to add the noise. Note also the more accurate sampling in (e), after post processing has been applied, compared with (d), where less pixels are included in the hit. In other words, (e) is more similar to (b), which is considered as the ground truth, compared to (d).



(a) (b) (c)

(d) (e) (f)

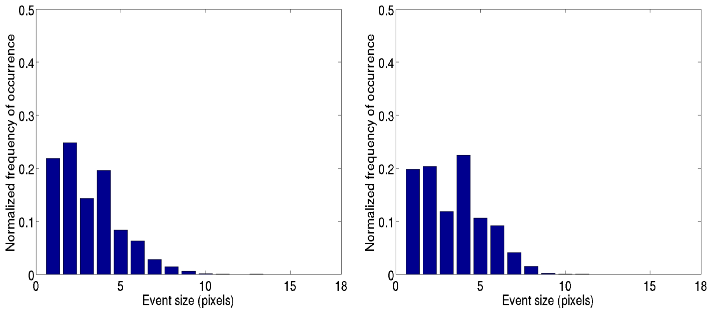
Figure 6.13: In Figure (a) a 66 keV β electron without any diffusion entering the detector from a non orthogonal direction is observed. The implementation of the diffusion process is shown in (b), noise is added in (c), initial correction is observed in (d), correction after post processing has been applied is shown in (e) and this is finally transformed to keV in (f).

Results of the Validation

The approach used in this work to evaluate the similarity of the simulated data compared to experimental data, is by comparing the cluster distribution obtained from the CMOS detector and the corresponding Monte Carlo simulations, for each different algorithm. The CCD is deliberately avoided in this comparison because of its lower diffusion and superior deposited charge confinement. The purpose of this validation is also to demonstrate that the correction method shown in Section 4.6 is based on the correct assumptions.

The experimental data taken with the CMOS detector is comprised of a 6 hours acquisition experiment, resulting in a total of 18360 images acquired. The number of events that were detected is 6570 events. The number of simulated events was 10^6 as mentioned above.

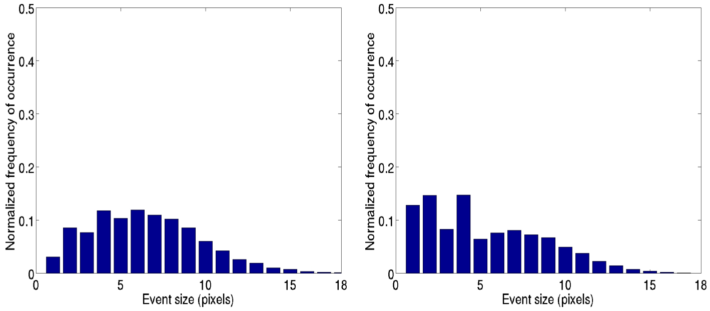
In Figure 6.14 the cluster distribution obtained from the acquired images from Vanilla and the cluster distribution obtained from the simulated setup, applying no post-processing and using k=8, are shown. Both distributions show very similar results, noticing how most of the events are concentrated in the first 4 bins, which means that most of the events are 1-4 pixels in size in both experimental and simulated results.



(a) (b)

Figure 6.14: Experimental normalised cluster distribution for the CMOS detector (left) and Monte Carlo simulated (right) using no ROI analysis.

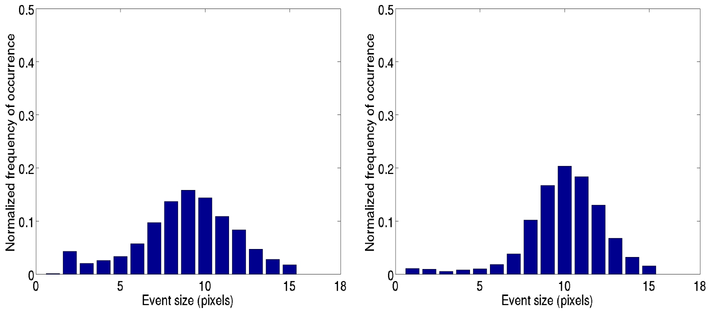
Similarly, in Figure 6.15, the cluster distribution obtained from the acquired images from Vanilla (Figure 6.15(a)) and the cluster distribution obtained from the simulated setup (Figure 6.15(b)), applying the fixed dual thresholding with k_1=8 and k_2=4 in the ROI analysis are shown. In this case, both distributions also show comparable results. In particular, note here how most events are now 5-10 pixels in size. Besides the random behaviour of the dark current in the pixels of the detector, which has been modelled, it was also observed how the experimental noise exhibits a time-dependant random component. This random component was basically comprised of an apparently random change of the global bias in the image (random change of dark current) and also in the random increase of noise that some pixels exhibit in the detector (random change of FPN). This effect has not been modelled as no reliable pattern of behaviour has been found. This may result in a higher disparity between cluster distributions compared to the previous case. The correction method explained in Section 4.3.1 tries to correct these effects, but this relies ultimately in certain models. If detector behaviour (or pixel behaviour) does not follow those models the correction method fails to correct the dark current noise and FPN.



(a) (b)

Figure 6.15: Experimental normalised cluster distribution for the CMOS detector (left) and Monte Carlo simulated (right) using fixed dual thresholding with k_1=8 and k_2=4.

Thirdly, in Figure 6.16 the experimental and simulated cluster distributions, resulting from the post-processing using adaptive thresholding in the ROI analysis, are shown. The value of thresholding parameter k used in equation 4.2 to initially detect the β-events is 8, to maintain consistency with the rest of the experiments shown above. In this case both distributions show a good match, showing in both cases how most of the 1-5 pixels events have now disappeared, and concentrating most of the events in the 7-12 pixels big events. In this case it is thought that the match is better because the method to adaptively threshold each ROI, is able to discard the noise (false events) better than in the previous situation. Therefore a slight mismatch between the noise shown in the experimental scenario and in that simulated does not make a significant difference for the results between experiment and simulation.



(a) (b)

Figure 6.16: Experimental normalised cluster distribution for the CMOS detector (left) and Monte Carlo simulated (right) using adaptive dual thresholding with k=8.

After these results of the Monte Carlo simulations of the β- particles interactions in silicon and the implemented charge diffusion process, it is considered that the simulations are sufficiently close to the experimental case, making possible a plausible set of predictive and/or analytical investigations in detector geometries and their effect on the spatial resolution and charge deposition.

6.4 Detector Considerations

Some simulations have been undertaken to understand the various effects of the different structures found typically in a CCD or a CMOS detector, and how these structures affect imaging performance for β- autoradiography. These effects are self-absorption effect, absorption depth and spatial resolution.

The self-absorption effect is due to the short electron range resultant of low and medium energy radioisotopes in comparison with the thickness of the radioactive source. If the source is thicker than the range of a certain β electron, this will be absorbed within the source. This effect means that a only fraction of the β electrons emitted by a radioactive source actually escape from the active volume.

The absorption depth is a critical feature as it affects the intrinsic spatial resolution and the capability of the detector of stopping the β particles.

The spatial resolution is an important section where the dependency of the main design parameters of the detector (pixel size and epitaxial layer thickness) are analysed to obtain the optimum spatial resolution depending on the radioisotope. In this section the system noise and the different correction methods presented in Sections 4.3.1 and 4.6 are considered to assess their effects on the final spatial resolution.

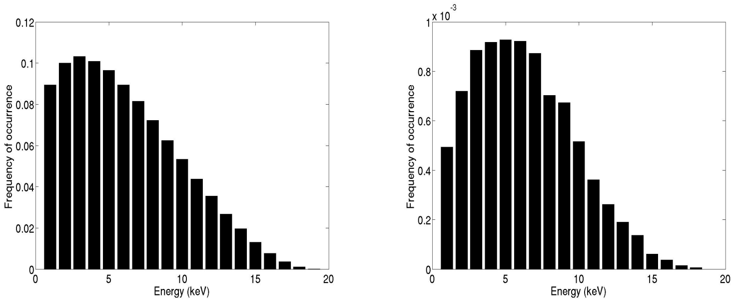
6.4.1 Self-absorption Effect

It has been shown in the past that accounting for β self absorption is very important to accurately measure the performance of a system [118, 145]. In both of these cases an estimation of the effective activity, i.e. fraction of β-particles escaping from a source and reaching the detector, is undertaken based on the range of ^{3}H and ^{14}C in the polymer-based microscale and in the geometry of the system.

The experiment presented in this section estimates the percentage of particles that reach the surface of the detector using a slab of tissue-equivalent polymer of ^{3}H and ^{14}C (density of ~1.1 g/cm^{3} [2]), using the entire range of energies of the corresponding β-spectra.

The experimental setup is simply a uniform radioactive slab of tissue (50 μm thick for ^{3}H and 120 μm thick for ^{14}C) on top of a slab of silicon. Then the intrinsic energy spectrum of the radioisotope under study and the deposited energy spectrum in the sensor are compared. Comparing the area under each distribution gives the fraction of particles that reach the detector (block of silicon). This allows to obtain the energy loss within the polymer microscale.

In Figure 6.17 the intrinsic energy spectrum (Figure 6.17(a)) and the deposited energy spectrum (Figure 6.17(b)) of a ^{3}H source are shown. From these graphs the fraction of ^{3}H particles that reach the detector is estimated at ~0.8 %. The same study is repeated for ^{14}C shown in Figure 6.18. From this simulation the estimated fraction of ^{14}C particles that reach detector is ~13.4 %, very close to the 14 % estimated analytically in [118].



(a) (b)

Figure 6.17: Intrinsic energy spectrum (a) and deposited energy spectrum (b) of ^{3}H within the slab of silicon. Note change in shape and dramatic change of the abscissa.

---

[2] GE Healthcare UK Ltd, Amersham Place, Little Chalfont, Buckinghamshire HP7 9NA, England

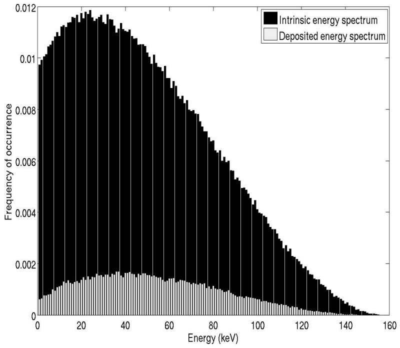


Figure 6.18: Intrinsic (black) and deposited (white) energy spectra of ^{14}C.

It is observed in both Figures 6.17 and 6.18 a general energy loss for the whole energy range, due to energy deposition within the source, but the energy spectra are not simply shifted to lower values, the shape of the energy spectra actually changes in both cases. This is due to a higher self-absorption effect of the low energy β electrons (short range) compared to the higher energy β electrons.

6.4.2 Absorption Depth

The absorption depth, also known as range, is where a specific particle with a specific kinetic energy is absorbed in matter. In the particular case of β- electrons in digital autoradiography, these interact with silicon depositing charge in each interaction, hence generating electron-hole pairs. These β- electrons slow down as their kinetic energy decreases, getting to a stopping point where no more energy is contained in the particle (other than thermal energy). As has been mentioned in Chapter 3, the trajectory of these β- electrons in silicon is tortuous, in contrast to the typical straight line trajectories obtained with X-ray photons. Electrons can undergo sharp deflections or maybe stopped completely in a single interaction. Electron ranges are highly variable even for electrons with the same energy [36]. The range is defined as the straight-line distance between the coordinate where the β particle initially interacted with the silicon and the coordinate where it was finally absorbed.

A tabulated measurement of absorption depth is found in the NIST website [126]. This information is also published in the International Commission on Radiation Units and Measurements (ICRU) report 37, which defines the *Stopping Powers for Electrons and Positrons*. As has been previously mentioned, this measure neglects energy-loss fluctuations and multiple scattering. In the case of β electrons, scattering is a significant process. Therefore the ranges provided by NIST may show high differences from what those particle ranges observed in the Monte Carlo simulations obtained here.

To obtain a better approximation of the absorption depth of β electrons, a point source of monoenergetic electrons placed in the centre of a slab of silicon has been simulated, for a wide range of energies (5-80 keV). The range obtained from these simulations neglecting scattering and taking it into account are shown in Figure 6.19, also with the range provided by NIST.

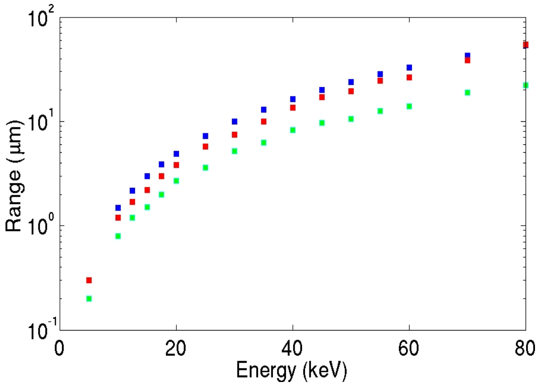


Figure 6.19: Electron range in silicon provided by NIST (blue), obtained with the Monte Carlo simulations without scattering (red) and with scattering (green).

It can be observed that NIST does not provide any information for energies below 10 keV. It can also be observed how the ranges provided by NIST are longer than those obtained in the simulations due to the absence of models for energy-loss fluctuations and multiple scattering in the NIST data. The difference shown by these sets of data (with and without scattering) becomes higher as the energy increases due to a higher multiple scattering effect.

6.4.3 Spatial Resolution

The range of a β- particle is controlled by its kinetic energy but this is less definite than for other particles. This is due to the fact that the trajectory of a β- particle follows a tortuous path due to its comparable mass to that of the outer electrons in the atoms environment. The particle will interact with these orbital electrons and the nuclei, changing direction abruptly and generating electron/hole pairs, thus depositing energy. This results in β-electrons with total path length greater than the penetration distance (straight line between initial interaction point and where the particle was finally absorbed).

It has been previously stated that the four intrinsic factors in a silicon-based detector that control the spatial resolution of an interaction are the pixel size, the electric field depth within the sensitive layer of the sensor, the resistivity of the silicon wafer and the thickness of the sensor [148]. Although [148] only considered X-ray interactions, here it is assumed that for β imaging, these factors are also the most relevant. It would be more appropriate to define the range of a β electron as the distance between the position where the β electron is generated, somewhere within the tissue sample, and the position where the particle is stopped. All the interactions of the particle in the middle of these two points will affect the range and concomitantly the spatial resolution. This is why any potential air gap is also an important parameter to consider in the observed spatial resolution for β imaging.

The electric field depth in the sensitive layer of the sensor will depend on the operational voltage of the sensor: being inherently low for CMOS sensors the electric field depth will also be low. This parameter is imposed by the technology and the feature size. For instance, for 0.18 μm CMOS processes, the voltage at which the sensor is operated is 1.8 V [77]. This electric field depth is non-uniform along the sensitive layer of the sensor and weakly extended in the sensitive volume [177], or even neglected [176]. The resistivity of the silicon wafer (typically $< 10\ \Omega$/cm) is also imposed by the CMOS production foundry and is lower than that for CCD detectors, to avoid the latch-up problem associated with CMOS electronic circuitry [77].

From an ionising radiation detection and imaging perspective, the only design parameters that can be imposed prior to fabrication (within reasonable ranges) are the pixel size and the thickness of the epitaxial layer. This is why these simulations have been focused in the study of these two parameters. The thickness of the layers used in the detector architecture for this set of experiments for ^{14}C and ^{3}H are detailed in tables 6.3 and 6.4 respectively. For the case of ^{3}H, a back-thinned architecture is used. This is why there is a top passivation layer of 0.1 μm, the residual product of the etching process on the sensor, and the bottom passivation layer of 5 μm, the same passivation layer on top of the epitaxial layer in the front-illuminated architecture (further details of the back-thinning process appear in Section 3.1). The range of thicknesses of the epitaxial (sensitive) layer used is up to 30 μm for ^{14}C and 20 μm for ^{3}H.

Table 6.3: Thickness of layers used for spatial resolution for ^{14}C

Layer	Glass	Tissue	Passive	Substrate
Thickness	1.0 mm	20.0 μm	5.0 μm	500.0 μm

Table 6.4: Thickness of layers used for spatial resolution for ^{3}H

Layer	Glass	Tissue	Passive (top)	Passive (bottom)	Substrate
Thickness	1.0 mm	20.0 μm	0.1 μm	5.0 μm	500.0 μm

In order to investigate these phenomenon further, a Monte Carlo simulation was set up as follows. A point source was placed in the centre of a central pixel from a 29x29 pixel array in direct contact with the surface of the detector (no air gap). The spatial resolution was measured as the Full-Width-Half-at-Maximum (FWHM) from the resulting Point Spread Function (PSF) in each experiment. For a ^{14}C point source 5 10^4 particles were simulated for each detector geometry, the pixel size range being 5, 10, 15 and 25 μm, and the epitaxial layer thickness range 5, 15, 20 and 30 μm, resulting in a spatial resolution map where the effects of the pixel size and the epitaxial layer thickness on the spatial resolution are studied. A total of 8 10^5 particles were simulated. For a ^{3}H point source 2 10^5 particles were simulated for each detector geometry, the pixel size range being 1, 2, 5 and 10 μm, and the epitaxial layer thickness range 1, 2.5, 5, 10 and 20 μm. A total of 4 10^6 particles were simulated. After all these combinations were simulated using Geant4, the charge diffusion process, system noise and noise correction, as described in Section 6.3.1, were added to obtain realistic results.

The number of particles simulated for each geometry for the ^{14}C point source (5 10^4 particles) was confirmed to be statistically sufficient by simulating 10^6 particles for a subset of the aforementioned geometries and obtaining identical results.

The number of particles simulated for the ^{3}H point source was four times higher compared to ^{14}C given that more particles are absorbed in the passivated layers, and the deposited energy is significantly lower. Therefore more statistics were necessary in order to compute the FWHM from the PSF reliably. After repeating an initial simulation several times using 2 10^5 particles, similar results were obtained, demonstrating statistical validity with this set of samples.

Initial Spatial Resolution Map

The spatial resolution map resulting from a simulation of the variation of the spatial resolution with the thickness of the epitaxial layer (5, 15, 20 and 30 μm) and the pixel size (5, 10, 15 and 25 μm) for a point source of ^{14}C with no air gap is shown in Figure 6.20. Similarly, a spatial resolution map resulting from a simulation of the variation of the spatial resolution with the thickness of the epitaxial layer (1, 2.5, 5, 10 and 20 μm) and the pixel size (1, 2, 5 and 10 μm) for a point source of ^{3}H with no air gap is shown in Figure 6.21.

Figures 6.20 and 6.21 do not include charge diffusion, noise or any kind of correction; these just represent the intrinsic spatial resolution obtained from each geometry for each radioisotope.

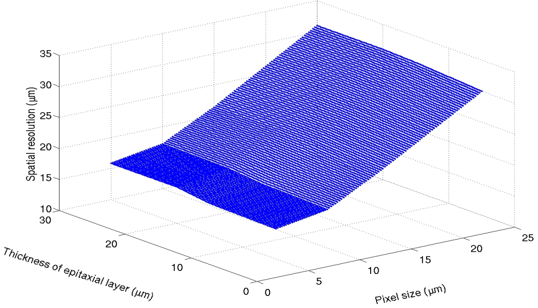


Figure 6.20: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{14}C with no air gap between the surface of the sensor and the source.

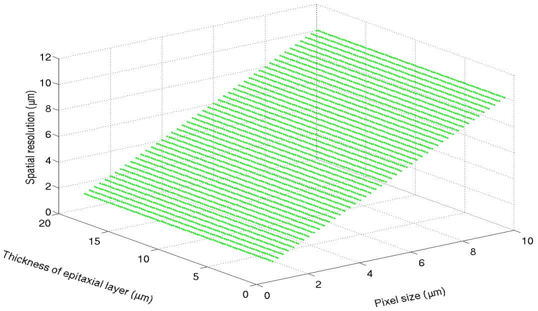


Figure 6.21: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{3}H with no air gap between the surface of the sensor and the source.

It is observed how in both examples, the data point with the smallest pixel size and thinnest epitaxial layer, 5 μm and 5 μm respectively for ^{14}C and 1 μm and 1 μm respectively for ^{3}H, shows the best spatial resolution, ∼15 μm for ^{14}C and ∼1 μm for ^{3}H. It can be observed how the dependence of the spatial resolution on the thickness of the epitaxial layer is very low, as oppose to the dependence with the pixel size in both cases. This high dependence is not shown in Figure 6.20 for ^{14}C below a pixel size of ∼10 μm, which corresponds approximately with the average range of ^{14}C β electrons.

Spatial Resolution Map with Charge Diffusion

Those results shown in Figures 6.20 and 6.21, are obtained using the data generated by the Geant4 Monte Carlo toolkit, where no charge diffusion, system noise or noise correction are included in the model. Nonetheless, these represent a baseline level of performance, against which the effect of further performance degrading processes can be measured. To obtain a more accurate study of the resulting observed spatial resolution based on the variation of the detector geometry, the charge diffusion process has been implemented as explained above (Section 6.3.1). The resulting spatial resolution variation, used for the same geometries used in Figures 6.20 and 6.21 is shown in Figures 6.22 and 6.23 for ^{14}C and ^{3}H respectively.

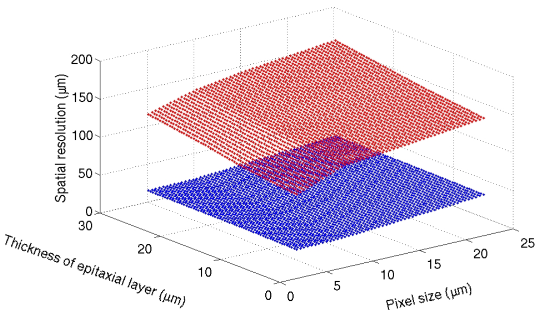


Figure 6.22: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{14}C with no air gap between the surface of the sensor and the source, with charge diffusion process (red) and without charge diffusion process (blue). Note that the data point corresponding to the smallest spatial resolution corresponds to a 5 μm thick epitaxial layer and a 5 μm pixel.

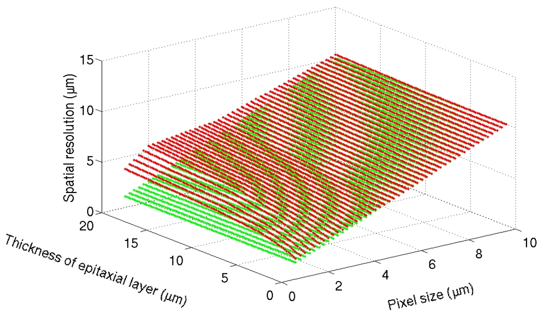


Figure 6.23: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{3}H with no air gap between the surface of the sensor and the source, with charge diffusion process (red) and without charge diffusion process (green). Note that the data point corresponding to the smallest spatial resolution corresponds to a 1 μm thick epitaxial layer and a 1 μm pixel.

It can be observed in Figures 6.20 and 6.21 how, without charge diffusion, the variation in epitaxial thickness does not provide a significant impact on the spatial resolution. If the charge diffusion process is considered, it is now observed in Figures 6.22 and 6.23 how the epitaxial layer thickness represents an important parameter in the detector design, having a negative impact when a thicker epitaxial layer is used.

Both Figures 6.22 and 6.23 show a significant degradation of the spatial resolution as both the pixel size and the epitaxial layer increase in size due to the charge diffusion process, being this process more critical for ^{14}C due to the higher amount of deposited charge (higher ionising energy).

In the specific case of ^{14}C (Figure 6.22), it is observed how the rate of spatial resolution degradation as the pixel size increases is higher for a pixel size $<\sim$10 μm along the entire range 5-25 μm. This is due to the higher impact of the charge diffusion process on the spatial resolution for a pixel size below the average ^{14}C mean energy. On the other hand, the rate of spatial resolution degradation holds a linear relationship with the epitaxial layer thickness.

The specific case of ^{3}H (Figure 6.23) shows some differences compared to the previous case. In this case the charge diffusion process does not show such high degradation of the spatial resolution compared to the previous scenario. The epitaxial layer thickness shows to have a higher effect on the spatial resolution for small pixel sizes (<∼4 μm) than for larger pixel sizes.

Spatial Resolution Map with System Noise and Noise Correction

This study is further analysed by including the effect of dark current noise and FPN, and subsequent noise correction. The three different correction approaches, described in Sections 4.3.1 and 4.6, are implemented to observe how these affect the spatial resolution in each different scenario. It has to be noted that the noise model applied to the raw images, obtained from the Monte Carlo simulations, was previously measured from the available detector, having 25 μm pixel size and 20 μm thick epitaxial layer. This is expected to change for other geometries but the absence of other available detector geometries limited the implementation of geometry dependant noise models. However the FPN is not expected to change significantly by changing the dimensions of the pixel size and the epitaxial layer thickness. Dark current noise, on the other hand, is expected to increase with size given that this is proportional to the pixel area A_p and the constant specified by the manufacturer R, representing the dark current noise at 300K, as shown in equation 3.11. In this respect, the noise implementation here may be somewhat overestimated for pixels <25 μm and underestimated for pixels >25 μm.

In Figures 6.24 and 6.25 the spatial resolution map for ^{14}C and ^{3}H are shown respectively, obtained after including the detector noise in the model and applying the FPN correction method described in Section 4.3.1. Figure 6.24 represents the intrinsic spatial resolution map for ^{14}C shown in blue, and the spatial resolution map obtained after applying the correction method shown in red. Similarly, Figure 6.25 represents the intrinsic spatial resolution map for ^{3}H in green, and the spatial resolution map obtained after including detector noise and applying the correction method in red.

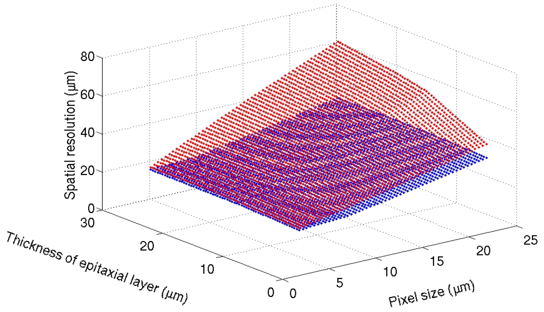


Figure 6.24: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{14}C with no air gap between the surface of the sensor and the source, without charge diffusion process (blue) and with charge diffusion process, noise addition and the FPN correction method described in Section 4.3.1 (red).

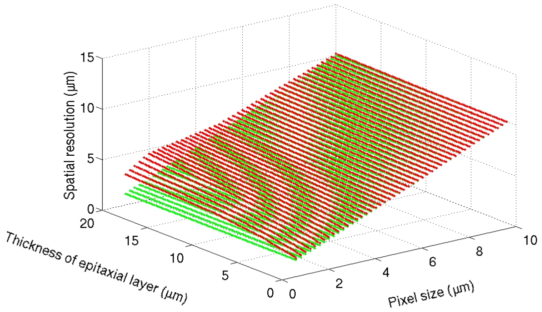


Figure 6.25: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{3}H with no air gap between the surface of the sensor and the source, without charge diffusion process (green) and with charge diffusion process, noise addition and the FPN correction method described in Section 4.3.1 (red).

The same graphs have been obtained applying the fixed dual thresholding method described in Section 4.6.1 for the two cases presented above, in Figures 6.26 for ^{14}C and 6.27 for ^{3}H. The two thresholds applied in the fixed dual thresholding method for ^{14}C are k_1=8 and k_2=4, and for ^{3}H these thresholds are k_1=4 and k_2=2.

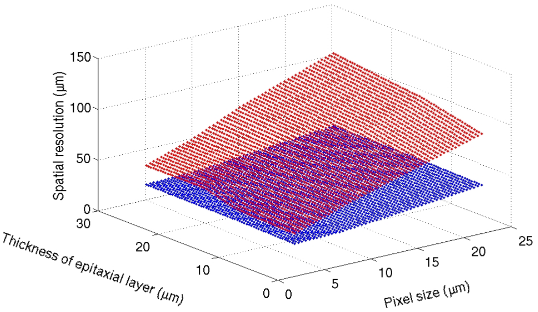


Figure 6.26: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{14}C with no air gap between the surface of the sensor and the source, without charge diffusion process (blue) and with charge diffusion process, noise addition and the fixed dual thresholding method as noise correction method (red).

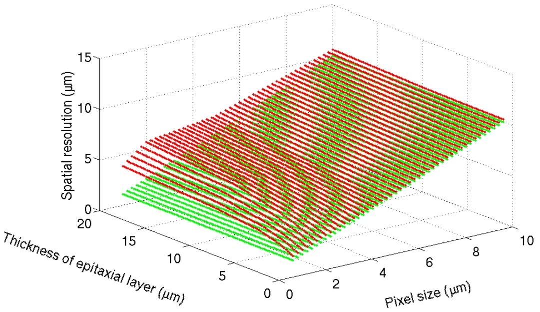
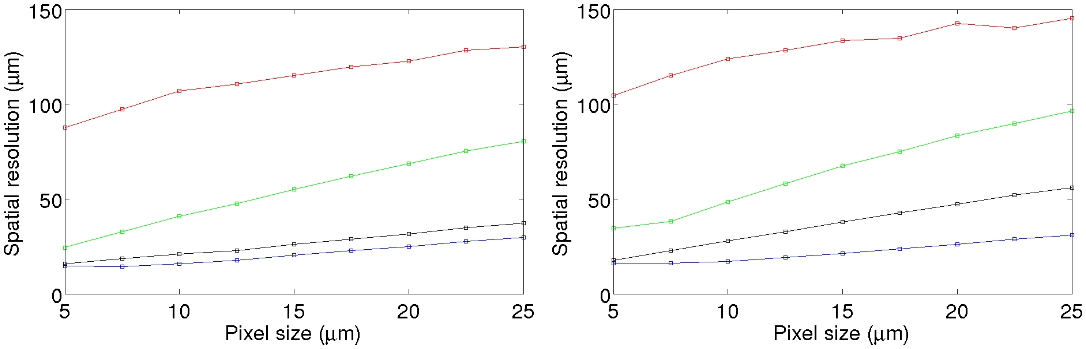


Figure 6.27: Simulated spatial resolution map of the digital sensor exposed to a point source of ^{3}H with no air gap between the surface of the sensor and the source, without charge diffusion process (green) and with charge diffusion process, noise addition and the fixed dual thresholding method as noise correction method (red).

In order to further study the spatial resolution maps presented above, due to the vast amount of data processed to produce each spatial resolution map (8 10^5 particles for ^{14}C and 4 10^6 for ^{3}H), two highly sampled profiles have been simulated. One profile is taken at a fixed 5 μm epitaxial layer and the other one at 20 μm epitaxial. The pixel size values simulated in these cases are 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5 and 25 μm for ^{14}C, and 0.5, 1, 2, 3, 4, 5, 7.5 and 10 for ^{3}H.

The resulting spatial resolution for each epitaxial layer for ^{14}C is shown in Figure 6.28 for four different cases. (1) for the raw data without charge diffusion, system noise nor noise correction (shown in blue), (2) for the data where charge diffusion has been included in the model (shown in red), (3) for the data where charge diffusion, system noise and the standard correction method (Section 4.3.1) have been included (shown in black), and finally (4) for the data where charge diffusion, system noise and the fixed dual thresholding correction method (Section 4.6, k_1=8 and k_2=4) have been included (shown in green).

From Figure 6.28 some conclusions can be extracted with respect to ^{14}C imaging. First this demonstrates that the up sampling introduced in the spatial resolution maps between data points shown in Figures 6.20, 6.22, 6.24 and 6.26 does not introduce artifacts. It demonstrates how the spatial resolution, measured from the raw data in both cases (blue graph), does not show an important dependency on the pixel size in the range of 5-10 μm. This approximately corresponds with the mean range of β electrons at 49 keV, the mean energy of ^{14}C. For a pixel size >10 μm the spatial resolution increases linearly with the pixel size. When the charge diffusion process, system noise



(a) Epitaxial layer 5 μm.

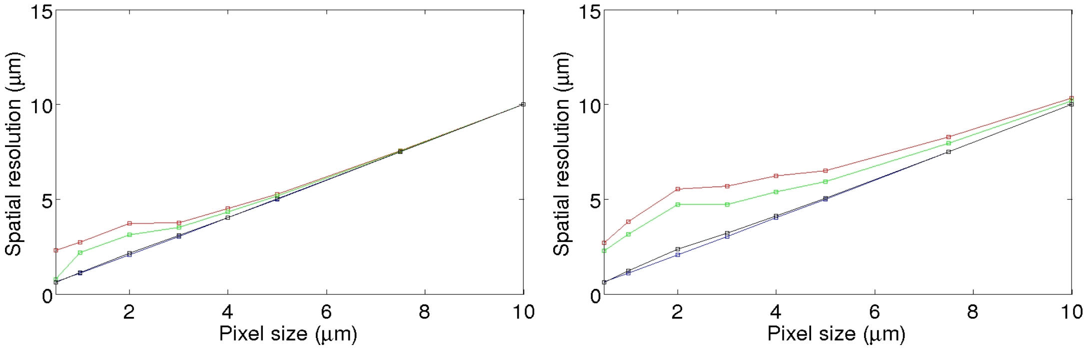
(b) Epitaxial layer 20 μm.

Figure 6.28: Spatial resolution obtained for ^{14}C for an epitaxial layer of 5 μm (a) and 20 μm (b) for a range of 5-25 μm pixel sizes. The blue graph represents the spatial resolution for the raw data, the red graph includes the effect of the charge diffusion, the black line includes the charge diffusion, system noise and correction method described in Section 4.3.1, and the green graph includes the charge diffusion, system noise and dual thresholding correction method.

and noise correction are included in the simulations, then the spatial resolution shows a higher dependence on the pixel size. In the particular case where the FPN correction method described in Section 4.3.1 is applied (black graph), the spatial resolution shows a linear dependence with the pixel size over the whole range of values studied here. This linear dependence is mostly observed in the other case, where the fixed dual thresholding correction method is applied (green graph).

Similar to the study shown above, the resulting spatial resolution for ^{3}H is shown in Figure 6.29 for the same four different cases as before. In this particular case the parameters used for the fixed dual thresholding method are k_1=4 and k_2=2. The epitaxial layer values used in these experiments are at 5 and 20 μm, and the pixel size values used here are 0.5, 1, 2, 3, 4, 5, 7.5 and 10 μm, as mentioned above.

From Figure 6.29 some conclusions can be extracted. Similarly with the previous case, for ^{14}C, this demonstrates that the up sampling introduced in the spatial resolution maps shown in Figures 6.21, 6.23, 6.25 and 6.27 does not introduce artifacts. Figure 6.29 demonstrates how the spatial resolution, measured from the raw data in both cases (blue graph), is strongly correlated with the pixel size. The effect of charge diffusion in this case is strongly limited by the epitaxial layer thickness. Note, in Figure 6.29(a), how the spatial resolution is unaffected by charge diffusion for pixel sizes larger than the epitaxial layer thickness, set at 5 μm. Figure 6.29(b) shows how for a thicker detector charge diffusion has a more significant effect, particularly at smaller pixel dimensions, resulting in a degraded spatial resolution. When charge diffusion, system noise and noise correction are included in the simulations, in the particular case where the FPN correction method described in Section 4.3.1 is applied (black graph), the spatial resolution shows a linear dependence with the pixel size, being very close to the intrinsic spatial resolution (blue graph) in the whole pixel size range studied here. In the case where the fixed dual thresholding correction method is applied (green graph), the spatial resolution graph seems to fall between that obtained with and without the



(a) Epitaxial layer 5 μm.

(b) Epitaxial layer 20 μm.

Figure 6.29: Spatial resolution obtained for ^{3}H for an epitaxial layer of 5 μm (a) and 20 μm (b) for a range of 0.5-10 μm pixel sizes. The blue graph represents the spatial resolution for the raw data, the red graph includes the effect of the charge diffusion, the black line includes the charge diffusion, system noise and correction method described in Section 4.3.1, and the green graph includes the charge diffusion, system noise and dual thresholding correction method.

charge diffusion process (red graph).

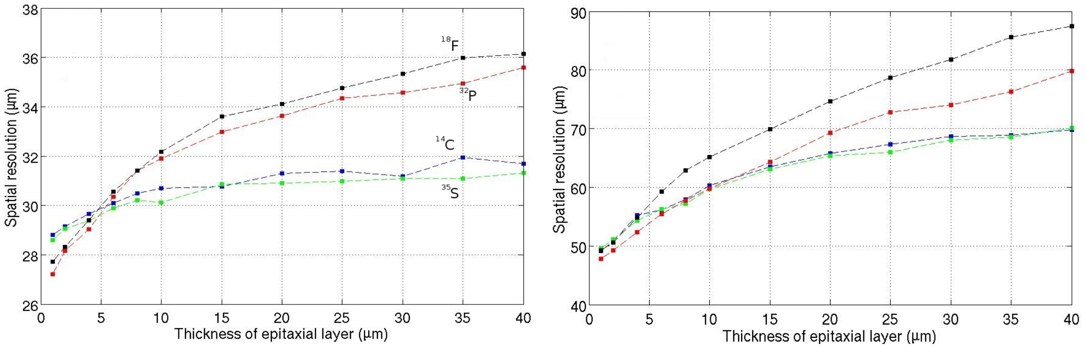
Overall, these simulations demonstrate that the best spatial resolution is obtained when pixel size is minimised, and that once FPN correction processes have been applied, there is only very weak dependence on epitaxial thickness.

6.4.4 Air Gap Effect on Spatial Resolution

In order to investigate the effect of the air gap between the surface of the sensor and the source on the spatial resolution a set of experiments has been undertaken using a fixed pixel size at 25 μm (the pixel size of the sensor used in the experimental part of the study) using four different radioisotopes; ^{14}C, ^{35}S, ^{32}P and ^{18}F. Different thicknesses have been considered to study the degradation of the FWHM with this effect. In this case, the aim is to study the intrinsic spatial resolution degradation with the air gap between the sample and the surface of the detector, therefore no charge diffusion, system noise or noise correction method is implemented to obtain these results.

The FWHM degradation for four different radioisotopes, ^{14}C, ^{35}S, ^{32}P and ^{18}F, has been studied for two different values of air gap: 0 μm shown in Figure 6.30(a) and 20 μm shown in Figure 6.30(b). Figure 6.30 shows how the air gap degrades the spatial resolution dramatically.

The same experiment has been undertaken for ^{3}H using a fixed pixel size of 1 μm for a range of epitaxial thicknesses varying from 1 μm up to 25 μm with no air gap (blue) and for a 5 μm air gap (red). It is shown in Figure 6.31 how the epitaxial thickness does not show any difference due to the short range of ^{3}H β electrons.



(a) Spatial resolution vs thickness of epitaxial layer with no air gap.

(b) Spatial resolution vs thickness of epitaxial layer with an air gap of 20 μm.

Figure 6.30: Spatial resolution vs thickness of epitaxial layer with different values of air gap for ^{14}C (blue), ^{35}S (green), ^{32}P (red) and ^{18}F (black) using a pixel size of 25 μm.

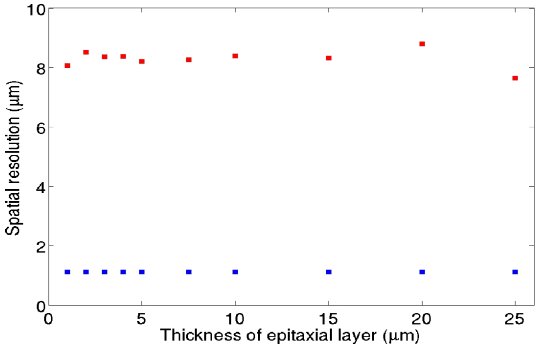


Figure 6.31: Spatial resolution vs thickness of epitaxial layer with no air gap (blue) and with a 5 μm air gap (reg) for ^{3}H using a pixel size of 1 μm.

6.4.5 Charge Collection

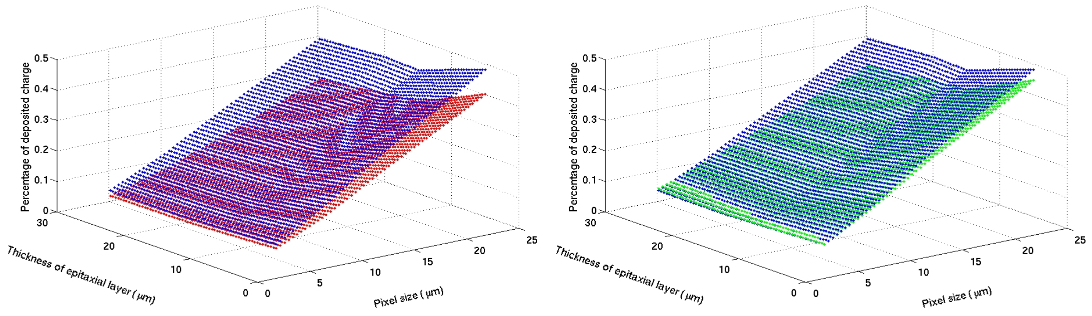
It has been described above how the detector architecture geometry, epitaxial layer thickness and pixel size, are of critical importance for the spatial resolution. Another parameter of significant influence for the final image quality is the proportion of charge (signal) collected by the sensitive volume of the detector. A thick detector with large pixels will collect more signal than a thin detector with small pixels, thus producing higher SNR images.

To study this effect in this work a point source in the centre of an array of 29x29 pixels is simulated. The proportion of charge deposited by each β electron in the central pixel, where the point source is centred, compared to the entire amount of deposited charge is analysed.

A similar comparison to the spatial resolution study, where the intrinsic deposited signal, the effect of charge diffusion and the resulting deposited signal after system noise and noise correction are implemented, are compared to see their respective effects.

Figure 6.32 shows the proportion of charge deposited in the central pixel of the β

electrons emitted by a ^{14}C point source (centred in the central pixel) for different architecture geometries, where the pixel size range goes from 5 μm up to 25 μm, and the epitaxial thickness goes from 5 μm up to 30 μm. Each data point represents a specific architecture where 5 10^4 β electrons have been simulated. Figure 6.32(a) shows the intrinsic proportion of charge deposited (no charge diffusion and no noise) in blue, compared to the proportion of charge deposited with charge diffusion process included in the model in red. Figure 6.32(b) shows the comparison between the intrinsic proportion of charge deposited in blue (the same as in Figure 6.32(a)) and the proportion of deposited charge with charge diffusion, system noise and the noise correction method described in Section 4.3.1 implemented in green.



(a) (b)

Figure 6.32: Proportion of charge deposited in a pixel by a ^{14}C point source for different detector architectures. Figure (a) shows the intrinsic proportion of deposited charge in blue with the proportion of deposited charge with the charge diffusion process added to the model in red. Figure (b) shows the same intrinsic proportion of deposited charge in blue with the proportion of deposited charge with charge diffusion, system noise and the noise correction method described in Section 4.3.1 implemented in green.

The intrinsic proportion of deposited charge shown in Figure 6.32 demonstrates how increasing the pixel size has a positive impact on the amount of deposited charge in a pixel, concomitantly improving the SNR of the final image, and therefore improving the image quality.

It is also observed in Figure 6.32 how decreasing the epitaxial thickness improves the proportion of charge deposited in a single pixel. This can be confusing as it is well known that a detector with a thick active volume collects more charge than a thin active volume. The reason of this apparently contradictory result is because a thicker detector collects more charge than a thin detector, but this charge is shared with more pixels, concomitantly showing lower proportion of deposited charge in one single pixel.

Figure 6.32(a) shows how the charge diffusion process produces a degradation of the proportion of deposited charge in one pixel, having higher impact for thicker active volumes. For instance, for a pixel size of 25 μm and a 5 μm epitaxial thick, it can be observed how the proportion of deposited charge drops from 0.5 (48.2%) down to 0.4 (40.1%) (x1.2 degradation rate), while for a 30 μm epitaxial thick the degradation of proportion of deposited charge goes from 0.4 (39.3%) down to 0.25 (25.8%) (x1.5 degradation rate).

As has been observed in the spatial resolution study (Section 6.4.3) the different correction methods tend to alleviate the effect of charge diffusion process, by ignoring those pixels with the lowest level of signal, because they are masked by dark current and FPN. In this example, where the correction method described in Section 4.3.1 is implemented, Figure 6.32(b) exhibits a less significant impact of the charge diffusion process. The proportion of deposited charge shown in all the different architectures studied in this section is closer to the intrinsic proportion of deposited charge, compared to the previous example (Figure 6.32(a)), but globally the same conclusions obtained in the previous case still hold. The charge diffusion process degrades the proportion of deposited charge in one pixel, this effect being more significant for thicker architectures. It can be observed in Figure 6.32(b) how the intrinsic proportion of deposited charge (blue) for small pixel sizes is slightly lower than that obtained with charge diffusion, system noise and noise correction implemented (green). This artifact is attributed to a deficient noise correction that is unable to remove all the noise in each pixel, therefore the signal measured in each pixel is overestimated due to the level of noise. This effect is only observed for small pixels, where the level of deposited signal is lower than for large pixels, because the level of noise assumed for all the detector architectures remains constant.

6.5 Conclusions

First two important issues, such as the self-absorption effect and the range of β- particles in silicon, largely discussed in the literature have been investigated. Self-absorption has been shown to have a significant impact in the efficiency of the detector, reducing the β- particles that reach to the surface of the detector to $\sim$0.8 % and $\sim$14 % for ^{3}H and ^{14}C respectively, as shown in Figures 6.17 and 6.18. Electron range (Figure 6.19) has also been shown to have a significant impact in the spatial resolution as largely explained in Section 6.4.3.

Secondly the two thresholding methods presented in this thesis work have been validated, demonstrating successful implementation of the diffusion process suffered by ionised charge in silicon, and correct modelling of the noise observed in the CMOS digital detector. The experimental setup is reproduced obtaining similar cluster distributions, showing a good match between both scenarios.

A thorough study on the spatial resolution of the detector has been presented. Given that the Monte Carlo simulations have been previously validated, the geometry of the detector has been modified to obtain an optimum detector geometry for each typical radioisotope used in β autoradiography.

From this set of experiments an optimum detector geometry can be chosen depending on the radioisotope used. Observing the intrinsic spatial resolution map (Figures 6.20 and 6.21) it seems clear that for ^{14}C a pixel size <10 μm does not improve the spatial resolution. From this value up to larger pixels, the spatial resolution then linearly scales with the pixel size. A similar effect has been observed for ^{3}H albeit with a smaller range of pixel sizes. This indicates that spatial resolution is strongly related with the mean range of the electrons emitted by each radioisotope, over and above pixel dimension.

From Figures 6.22 and 6.23 it is observed how the charge diffusion effect greatly degrades the spatial resolution. Several models found in the literature were implemented, resulting in very similar results for thin detectors. Dissimilarities were found when using thicker detectors as those used in this thesis. The only validated charge diffusion model with experimental data was implemented in this work for inclusion in the Monte Carlo data previously produced. This model has been shown to produce the largest dissimilarities compared with the rest of the models for deep energy deposition interactions in thick detectors (∼20 μm).

Charge diffusion could be reduced by applying an electric field to the active volume to collect the deposited charge, or by using a thinner detector. It has been observed that the outer pixels affected by the deposited charge in an arbitrary event only contain few electrons. These few electrons will be easily masked by dark current noise.

The correction methods applied to the noisy images demonstrate capability of detecting true events, and reliably rejecting false events, by obtaining similar cluster distributions to those obtained with Monte Carlo simulations representing a realistic environment. These methods can be separated in two depending on the goal of the method.

Using the FPN correction method described in Section 4.3.1, a single threshold, high enough to produce a low rate of false events, results in a spatial resolution map similar to the intrinsic spatial resolution map for low energies (^{3}H) as shown in Figure 6.25 and in more detail in Figure 6.29. For medium energies (^{14}C) some differences arise as the effects of the epitaxial layer thickness on spatial resolution increases as shown in Figure 6.24 and in more detail in Figure 6.28, due to higher levels of multiple scattering.

Where dual thresholding is applied it is observed how the spatial resolution is degraded with regard to the previous method, but as explained in the previous chapter this higher level of information can be made available for event reconstruction algorithms to improve the spatial resolution.

In all cases shown above, when charge diffusion, system noise and noise correction are included in the final model, it has been shown how the best spatial resolution always corresponds to the smallest range of pixel sizes. In the case of ^{14}C, this represents pixels up to 10-15 μm as being optimal. Above this range, the degrading effects of pixel dimension and epitaxial thickness somewhat degrade spatial resolution performance. Below this size, one might expect to see greater significance in the per pixel statistical uncertainty produced due to finer sampling of the total charge. For ^{3}H, the charge diffusion effect appears to be proportional to epitaxial thickness and pixel size. In this case, the optimal pixel dimension appear to be ∼1 μm.

It has been further demonstrated how the spatial resolution of the system is dramatically affected by the air gap between the surface of the sensor and the source. As can be observed in Figure 6.30, a 20 μm air gap produces ∼x2 increase in spatial resolution compared to no air gap when imaging ^{14}C or ^{35}S. For instance for an epitaxial thickness of 20 μm the intrinsic spatial resolution goes from ∼31 μm up to ∼65 μm. A similar degradation rate can be observed for higher energy radioisotopes such as ^{32}P and ^{18}F, showing a degradation from ∼34 μm for both radioisotopes without air gap up to 70 μm and 75 μm respectively with a 20 μm air gap.

These results demonstrate how an air gap between the detecting medium and the sample is a matter of great importance, and further explains the reason why, in traditional autoradiography using conventional film, the samples are placed in direct contact with film inside a cassette where the lid makes pressure on the samples to reduce the possibility of existence of a gap between the samples and the sheet of film.

It has also been shown the dependence of the epitaxial thickness and pixel size on the deposited charge in one single pixel. A large pixel shows higher proportion of deposited charge in one pixel, while a thicker epitaxial layer reduces the proportion of deposited charge as it is shared between neighbouring pixels. On the other hand it is known that thick active volumes collect more charge than thin active volumes, despite the fact that it is shared between neighbouring pixels.

Depending on the application and the main aim of the design, a tradeoff between resolution and deposited charge has to be found in any direct detection silicon imaging device. The choice of one parameter over the other is going to deprecate the performance in a certain way.

Following all this work, for a system needing ^{3}H imaging capability, it needs to have $\sim$1 μm pixel size or below and 1 μm epitaxial thickness. If this is not important, then this can be relaxed to $\sim$10 μm pixel size and 10-15 μm epitaxial thickness. The pixel size necessary for autoradiography is well below 25 μm, the pixel size of the detectors used in this thesis, although the epitaxial thickness is close to that one used in this work (20 μm).

After studying β electron interactions in silicon imaging by using Monte Carlo to understand the best way to design a silicon imaging sensor, this work studies in the next two chapters the second bottle-neck in the imaging chain post-acquisition analysis.

Chapter 7

Segmentation of Low Contrast to Noise Ratio Autoradiographic Data

7.1 Introduction

Image segmentation is one of the most fundamental aspects of biomedical image processing. It is widely used in a variety of applications such as delineation of anatomical structures, functional structures, tumours or extraction of features among some significant examples. There is even a wider variety of methods to segment a certain structure, depending on the specific problem.

Image segmentation of autoradiographic data plays an important role, as part of the final analysis and assessment of the autoradiograms obtained in a typical autoradiography experiment. The usual protocol followed in these type of experiments, involves manual segmentation of those regions that exhibit higher radioligand concentrations. In the specific case of this thesis work, where brain autoradiography is studied, those regions within the brain that exhibit higher radioligand concentration are manually delineated. These segmented regions represent areas in the brain that contain high levels of the neuroreceptor under study, hence a high level of radioligand uptake is observed. However, manual segmentation is very time consuming and repetitive. As the aim of this thesis addresses high throughput autoradiography, then having addressed the image acquisition phase, the next *bottle neck* in autoradiography throughput is the large amount of time needed for analysis, involving segmentation of key structures.

These segmented functional regions will be used as biological landmarks for further alignment with an atlas. Thus, it is assumed that these segmented landmarks have counterparts with the corresponding atlas bregma[1]. The reference point Bregma is shown indicated in a typical mouse skull in Figure 7.1, and also in a sagital mouse brain section in Figure 7.2 [135]. There exists a wide variety of types of landmarks. Those used in this work are segmented anatomical landmarks corresponding to key

---

[1] Coordinate system which defines the exact position of a certain slice in the brain

functional structures of the brain. These structures correspond to the regions in the autoradiogram with highest and lowest levels of uptake of the radioligand; i.e. these will be the brightest and the darkest regions respectively.

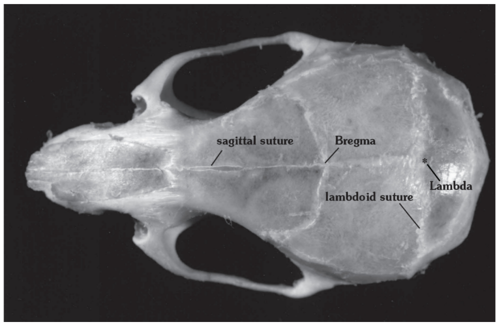


Figure 7.1: Mouse skull diagram showing the horizontal plane reference points, bregma and lambda [135].

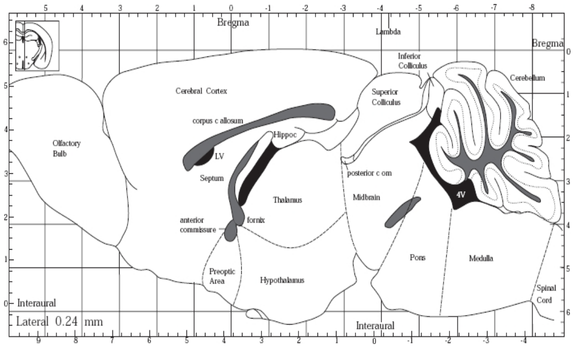


Figure 7.2: Sagital section of the mouse brain. The position of the skull marks (bregma, lambda and interaural) are indicated [135].

To put in proper perspective the real difficulty of this problem, for comparison with other imaging modalities, typical Contrast to Noise Ratio (CNR) for CT is around ∼200 and for MR is ∼300 for a 3T scanner and ∼100 for a 1.5T scanner. On the other hand typical values of CNR in functional imaging are: PET ∼10 and ∼1-4 for SPECT, similar values to those obtained in autoradiography. This is already suggestive that most of the segmentation approaches developed for CT and/or MR are most likely to misclassify the typical blurred edges in autoradiography.

7.2 Review on Segmentation Techniques

Image segmentation is one of the traditional branches of image analysis playing a key role in the biomedical arena. These methods are represented by a multiplicity of different approaches, as shown in different surveys [64, 132, 139]. Conventional segmentation approaches are based on histogram thresholding, edge detection (active contours), clustering, Bayesian methods and region extraction, or a combination of these.

Thresholding techniques are often discarded in medical imaging because these only consider histogram information, ignoring spatial information, and because of susceptibility to noise.

Edge detection based approaches have typically demonstrated good performance with high contrast images such as those found in Computed Tomography (CT) or Magnetic Resonance (MR), where boundaries are generally well defined, but for modalities with low CNR, as is often the case in functional imaging, boundaries appear diffuse and the level of noise is higher than that compared to CT and MR, due to low spatial resolution, partial volume effects and noise resulting from the statistical nature of radioactive decay [65]. As an example of edge detection, semi-automatic segmentation using a tailored model of active contours for brain autoradiographic data, is presented in [122]. The resulting segmentation data is visually assessed providing reproducible results, but user interaction is often required. Edge detection methods usually work in conjunction with a different approach to make the overall algorithm more robust. Some hybrid approaches using region growing, followed by active contours to refine the edge detection process, have been presented for PET tumour delineation [102, 184].

Attempts to simultaneously segment volumes of interest (VOIs), and to obtain a tracer kinetic model from the tissue time-activity curve (TTAC) is a very active area of research in image segmentation. A combination of clustering and region growing [83] and Markov Random Fields (MRF) [34] are some of the different approaches used in dynamic PET studies. Clustering time-activity curves using K-means or Maximum a Posteriori MRF (MAP-MRF) [152] are also some alternatives presented in the literature. Fuzzy C-means (FCM) and Fuzzy Hidden Markov Chains [65] have been other recent approaches used in segmentation of PET studies [187]. A comparison between Expectation-Maximisation (EM), Fuzzy C-Means and Independent Component Analysis (ICA) is presented in [87], applied to simulated and real dynamic PET brain images. The overall results are not satisfactory, but the authors conclude that ICA offers better results compared to the others, but it requires initial manual thresholding of the 3D images.

In the case of single slices or static images, where temporal information is not available, region growing has been shown to be an effective approach in 2D [111], where automatic delineation of fMRI structures using region growing outperforms clustering methods, and in 3D [170], where segmentation is applied for automatic extraction of blood vessels from Magnetic Resonance Angiography (MRA) data. This approach is widely used for biomedical image segmentation, due in part to its capability of segmenting noisy images and because it combines spatial connectivity and intensity information simultaneously. This technique relies on the idea that neighbouring pixels within the same region have some similarity, most often based on intensity values or statistics. As a result, a multitude of approaches based on region growing have been presented in the literature for segmenting a wide variety of imaging modalities, such as planar mammography [95], functional MRI [111], MR Angiography [170, 183], CT, MR, tumour segmentation [69, 114, 142] and PET [63].

Two novel measures are presented in [69], such as the gradient and the contrast that successfully delineate regions in MR images. These discontinuity measures have demonstrated good performance with mammogram image data [95], CT and MR images [69] but are too sensitive to noise when segmenting images with low CNR. Region growing relies on what is called the homogeneity criterion, where a pixel is appended to the region being grown if its intensity value is within some user-defined limits [142, 183, 111, 114], this aspect representing the most important feature of the algorithm. In [142] this margin is obtained from a previous learning step. Noticeably [183] used competitive region growing by growing several seed regions at the same time assuming that all the seeds are in direct contact. In [111] every voxel is considered a seed voxel at the initial point. Then the homogeneity criterion used to grow regions is based on the Pearson correlation coefficient between the time series of the voxel and the averaged time series of the voxels in a pre-merged region, and then compared to a pre-defined fixed threshold. The resulting regions are the largest segmented regions. This work also compares the performance of region growing, by analysing the receiver operating characteristic (ROC), with other methods typically used in fMRI, such as fuzzy c-mean clustering analysis (FCA), concluding that region growing outperforms compared to the mentioned method. A dynamic homogeneity criterion is also used in some approaches, where the user-defined margins change as the region grower evolves [63, 114]. A Bayesian approach is used by [133] to model the bimodal histogram generated when a region grower operates between two different tissues, using fuzzy c-means to estimate the parameters that describe each distribution. However, this relies on reasonably good CNR. Region growing has been shown to be highly dependent on the location and order of the original seeds chosen to grow the regions. The skeleton of, previously thresholded, blood vessels is used as seed regions to automate the initialisation of the region growing approach presented in [170].

In this work the segmentation issue of low CNR images (as often found in functional biomedical image data) is addressed using a region based approach. This automatic region grower, that automatically locates the seed regions for each structure, tries to detect significant variations in the statistics in the region under segmentation, or significant increments of the region size being segmented is demonstrated using simulated and pre-clinical imagery, demonstrating the suitability of this algorithm to low CNR images. This region grower segments the complete image until all the structures are labelled, i.e. included in a segmented region.

The methodology used in this work, for recursively segmenting objects represented in a low CNR image, results in a fully labelled scene. This involves first simple removal of an assumed background component, followed by initialisation and execution of the region grower, finally halting using some termination criteria, discussed in detail in section 7.3.4. Once initial regions have been grown, then these are discarded from the original image, and the region grower is then recursively executed repeatedly until all pixels have been assigned to a segmented region.

7.3 Methodology

Classically, segmentation is defined as the partitioning of an image I into n non-overlapping regions R_i, $I = \cup_{i=1}^{n} R_i$, which are homogeneous with respect to some characteristic. The distinctiveness or non-overlapping condition, i.e. $R_k \cap R_j = \emptyset$ for $k \neq j$ is then fulfilled, wherein each R_k being connected between them. Some segmentation approaches assume a priori knowledge of the final number of subregions n. However in the case of the algorithm described here, n is unknown.

The methodology now described is used for recursively segmenting objects represented in a low CNR image, resulting in a fully labelled scene, without knowing the number of subsets a priori, n, in contrast with other approaches [114, 183]. A flowchart of the entire segmentation process is shown in Figure 7.3. The steps that comprise this algorithm involve first a pre-processing step where the simple removal of an assumed background component and anisotropic filtering to reduce statistical noise are first applied (Section 7.3.1). Secondly the region grower is initialised and subsequently it is executed following a dynamic criterion, where the growing parameters are adjusted depending on the statistics of the image as the region grower evolves (Section 7.3.2). Finally the region grower halts using some termination criteria (discussed in more detail in Section 7.3.4). Once initial regions have been grown, then these are discarded from the original image, and the region grower is then recursively executed until all pixels have been assigned to a segmented region.

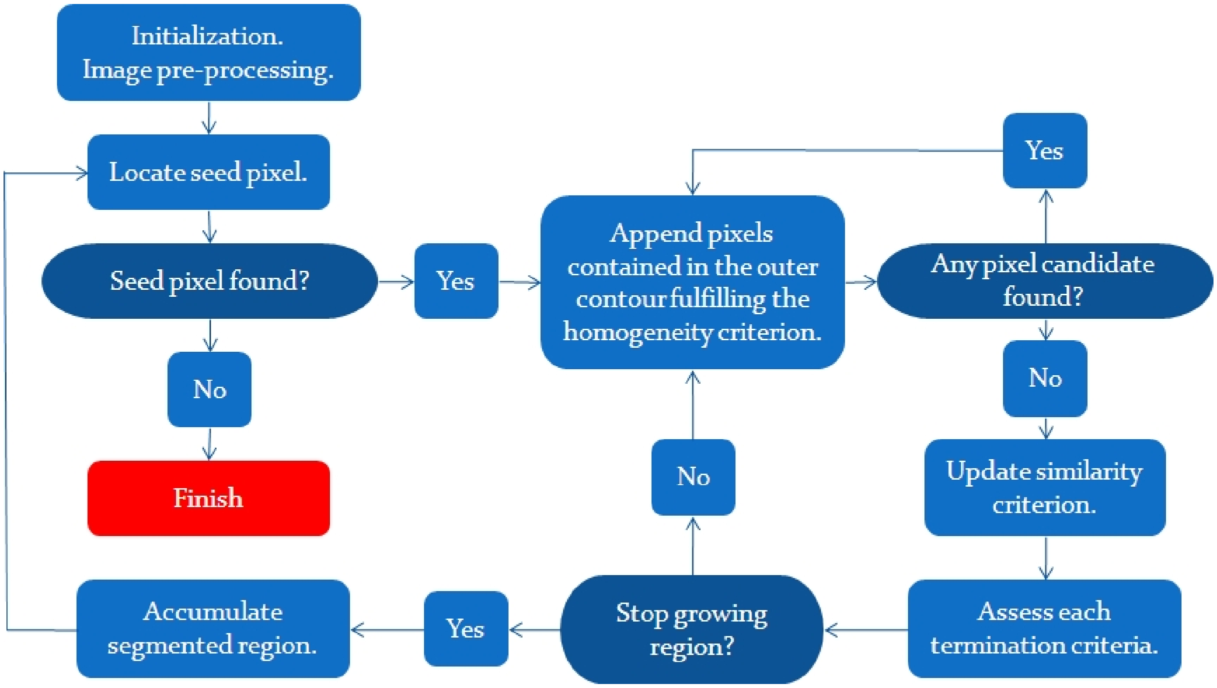


Figure 7.3: Flow chart showing the entire segmentation process.

It has been observed that after the autoradiogram has been segmented, is some cases there still exist some regions with low levels of ligand uptake (low intensity values) that have not been segmented. Therefore, once the region grower has segmented all the possible regions, the entire process is applied to an intensity-inverted version of the same autoradiogram. Depending on the specificity of the radioligand used in the original tissue section and the abundance of the neuroreceptors targeted in a specific experiment this dual-segmentation process may result in an over-segmented labelling.

If the radioligand used is not very specific this will bind to many different neuroreceptors, with the addition of the targeted one, resulting in a homogeneous image with abundant bright regions. If, on the other hand, the radioligand is highly specific it will only bind to those regions where a specific neuroreceptor is present resulting in an image with some very specific bright regions.

If the targeted neuroreceptor is very abundant the radioligand will be present in many regions of the tissue section, resulting in a corresponding image with abundant bright regions. This is the case of nicotinic neuroreceptors radiolabelled with ^{125}[I]-epibatidine. If the neuroreceptor can be located only in very specific regions of the brain this will result therefore in very specific bright regions. Clearly these issues alter the resulting

CNR of the image to be segmented. This variability in the target region might be addressed by using a region growing approach.

Due to the final over-segmentation produced in some cases where the radioligand is not very specific, the targeted neuroreceptor is abundant or also due to high statistical noise present in some images a region merging is undertaken at the end (Section 7.3.5).

The result of undertaking all these steps provide a final image with more regions segmented, compared to applying the region grower only once to the autoradiogram, easing the overall analysis of the autoradiograms.

7.3.1 Pre-processing

Prior to applying a region grower on the autoradiograms, the raw image data is first pre-processed. This pre-processing is comprised of two steps: first a thresholding step, where the background is removed, and secondly a filtering step, where an anisotropic filter is applied to the autoradiogram, to reduce statistical noise of the autoradiogram that may produce over-segmentation of certain regions. These two steps are described below.

Thresholding:

Initially images are thresholded before being segmented to remove the background noise. This is undertaken by computing the histogram of the image, and assuming this comprises of a set of n Gaussian-distributed components: all the pixels belonging to the first Gaussian distribution of the histogram are then set to zero. Low intensity values are assumed to correspond to absence of radioactivity, and presence of radioactivity is coded with high intensity values.

Due to sampling problems and presence of statistical noise, the histogram is smoothed by using a kernel density estimator, before finding the optimal setting for background thresholding. Given an intensity histogram f_x, with bins occupancy $\{x_1, x_2, ...x_N \mid \forall x_i > 0\}$, the equation of the statistical density estimator applied to the intensity histogram is shown in equation 7.1 [159]:

$$\hat{f}_x(h) = \frac{1}{Nh} \sum_{i=1}^{N} K(\frac{x - x_i}{h}), \tag{7.1}$$

where $\hat{f}_x(h)$ is the resulting smoothed intensity histogram for a certain bandwidth h, N is the total number of occupied bins in the intensity histogram, K is the kernel applied to the data operating on x, the occupancy for the bins in the original histogram f_x and x_i being a particular bin occupancy. A Gaussian function with zero mean and variance one is usually used for K. The kernel density estimator is then fully defined by equation 7.2:

$$\hat{f}_x(h) = \frac{1}{Nh}\sum_{i=1}^{N}\frac{1}{2\pi}e^{-\frac{1}{2}(\frac{x-x_i}{h})^2} \tag{7.2}$$

The main problem related with kernel density estimators is usually to choose a suitable value for the bandwidth h. AMISE (Asymptotic Mean Integrated Squared Error) is an algorithm used to calculate the optimal bandwidth, but due to the heavy computing cost traditionally necessary to run this algorithm a much easier option has been chosen in this thesis work. This is the popular Silverman's *rule of thumb* bandwidth [159] obtained with equation 7.3:

$$h = 1.06\sigma_x(N_xN_y)^{-0.2} \tag{7.3}$$

where σ_x is the standard deviation of the image, and N_xN_y is the total number of pixels in the image. Silverman's *rule of thumb* has been qualitatively validated obtaining usually satisfactory results by previous works.

In Figures 7.4(a) and 7.4(b) a raw autoradiogram and its corresponding histogram are shown. In Figure 7.4(c) the resulting smoothed intensity histogram is shown, presenting better defined peaks, smoother valleys and removing *false peaks* due to noise. As can be observed in Figure 7.4(c) the red line delimits the valley between the Gaussian corresponding to the background noise and the residual histogram.

This threshold is automatically obtained by finding the zero-crossing points of the second derivative of the smoothed histogram $\hat{f}_x(h)$, and finding the corresponding zero-crossing to the first valley between Gaussians. All the pixels below that intensity bin are set to zero and those above the threshold are considered as part of the autoradiogram.

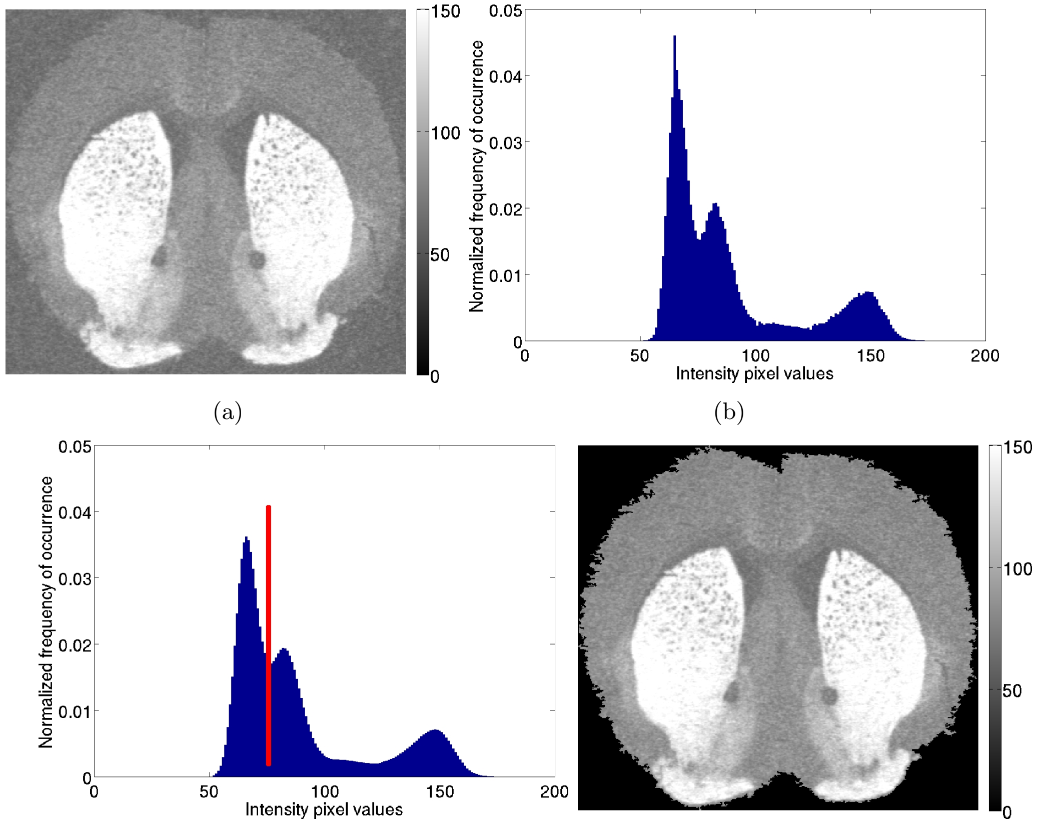
Few pixels of the background with higher value than the threshold will still remain, and some other pixels in the autoradiogram with lower values than the threshold might be set to zero because the Gaussians are slightly overlapped. To avoid this problem the thresholded image is morphologically processed. There will be some *holes* in the autoradiogram that will be filled and, outside the autoradiogram, there will be small blobs belonging to background noise. These will be removed by only considering the largest blob of the thresholded autoradiogram.

At the end of the pre-processing stage the image has the background set to zero and the autoradiogram has its original values without holes as shown in Figure 7.4(d).

In the data used for development of the method, there were some images where the background and the autoradiogram were not possible to separate by using this approach, given that the main body of the autoradiogram had very similar intensity values, compared to the background noise. In such cases the thresholding step was simply ignored.

Anisotropic Filtering:

Given the high statistical noise the autoradiographic data usually exhibits, an anisotropic



(a) (b)

(c) (d)

Figure 7.4: Original autoradiogram (a), corresponding intensity histogram (b), smoothed intensity histogram using a Gaussian-based kernel density estimator (c) and thresholded autoradiogram (d).

filter is first applied to the autoradiograms. An anisotropic filter is a common method to enhance the image quality in computer graphics, first introduced by Perona and Malik [136]. It is anisotropic because it is space variant, i.e. it is not homogeneous in all directions. It smooths homogeneous regions and enhance sharp edges.

This filtering process involves filtering an initial image $I(i,j)$ iteratively using the filter described by equation 7.4:

$$I(i,j)^{t+1} = I(i,j)^t + \lambda(c_U^t \nabla_U^t I_{i,j}^t + c_D^t \nabla_D^t I_{i,j}^t + c_L^t \nabla_L^t I_{i,j}^t + c_R^t \nabla_R^t I_{i,j}^t), \tag{7.4}$$

where $I(i,j)^t$ is the pixel value of the image at position (i,j) at iteration t, $I(i,j)^{t+1}$ is the pixel value at position (i,j) at the next iteration, the coefficient λ has a fixed value of 0.25, coefficients c_θ^t are the so called conduction coefficients calculated at the iteration t, and the terms ∇ are the finite differences calculated for 4-connectivity (i.e. pixels $(i,j-1)$, $(i,j+1)$, $(i-1,j)$ and $(i+1,j)$) as shown below:

$$\begin{cases} \nabla_U^t I_{i,j}^t = I_{i,j-1}^t - I_{i,j}^t \\ \nabla_D^t I_{i,j}^t = I_{i,j+1}^t - I_{i,j}^t \\ \nabla_L^t I_{i,j}^t = I_{i-1,j}^t - I_{i,j}^t \\ \nabla_R^t I_{i,j}^t = I_{i+1,j}^t - I_{i,j}^t \end{cases} \tag{7.5}$$

The conduction coefficients c^t are obtained with equation 7.6 [133]:

$$c^t = \frac{1}{1 + (\frac{\|\nabla^t I_(i,j)\|}{k})^2} \tag{7.6}$$

where $\|\nabla^t I_(i,j)\|$ is the norm of the finite difference for each case (U, D, L and R). k is a tuning value empirically obtained from the intensity histogram $f(x)$ defined by equation 7.7 [133]:

$$k = 0.6 \sum_{i=1}^{N} f(x) \tag{7.7}$$

The visual result on the autoradiograms after applying this filter is subtle and difficult to observe, but it has a positive impact on the subsequent processing of the image data, smoothing homogeneous regions and thus avoiding over segmentation in such regions, without affecting the edges contiguity.

7.3.2 Region Grower

The performance of most region growing techniques is highly dependant on the selection of the location of the initial seed pixels. Some approaches assume manual location of these, making this process user dependant and time consuming. Other approaches automatically locate the seed pixels by using different methods prior to the segmentation, such as image thresholding or morphological operations [170]. Other approaches use competitive region growing where all the pixels are considered initially as seed pixels, regions are separately grown from the initial seed pixels by adding pixels that satisfy the homogeneity criterion. In a second step some regions are selected from the obtained regions based on some predefined criterion [111]. In this work the seed pixel locations are chosen automatically as follows. Two assumptions are made in the initialisation:

- The initial region centred on a seed pixel (i.e. seed region) is NxN pixels assuming that the smallest region to be segmented is $2N$x$2N$ pixels. This assumption makes sure that the smallest anatomical region present in a tissue section is not missed due to the size of the initial seed region.
- The minimum distance between the centres of a segmented region and a new seed pixel is M pixels. This mitigates the effect of growing a new region which is actually part of an existing segmented structure, that has been undersegmented.

The first seed region is centred on a randomly located pixel that occupies the highest intensity value in the image. This is then enclosed by an NxN seed region, N being dependant on the size of the image. Some statistics, comprised of the mean, mode and standard deviation, and the highest pixel value in the seed region are calculated. For the specific case of the autoradiograms used in this work a region of 9x9 pixels was used. Therefore, prior to the segmentation, the autoradiogram is resized manually to make the smallest structure to be segmented (usually the hypothalamus) at least 18x18 pixels.

To make this resizing step user-independent for brain autoradiography image data, a typical autoradiogram where the hypothalamus is observed, has been taken, and the percentage of the the area of the hypothalamus compared to the whole section has been manually measured. By doing this a standard size (in pixels) for the autoradiograms to be segmented is set, prior to segmentation. This should be undertaken with an image containing the hypothalamus before a whole set of tissue sections is segmented, therefore the resizing parameters are first obtained and subsequently applied to all the autoradiograms under analysis.

7.3.3 Dynamic Similarity Criterion

A region grower is based on the similarity criterion, a criterion that candidate pixels have to fulfil to be appended to the region. This is with no doubt the most important feature of the region grower. Once the localised seed region statistics are computed, then a one pixel wide border, defined by T (equation 7.8), is considered, and each pixel of the border is evaluated.

$$T = \{x_{ij} \notin \bigcup_{i=1}^{n} R_i : N(x_{ij}) \bigcap_{i=1}^{n} R_i \neq \emptyset\} \tag{7.8}$$

where $N(x_{ij})$ is the set of immediate neighbours of pixel x_{ij}. The considered intensity pixel, x_{ij}, will be appended to the region $\{R_i\}$ if it fulfils the homogeneity criterion, i.e. if it is within the margins defined by equation 7.9:

$$x_{ij} \begin{cases} \in R_i & \text{if } m_o - k\sigma_{1,s} < \hat{x}_{ij} < m_o + k\sigma_{2,s} \\ \notin R_i & otherwise \end{cases} \tag{7.9}$$

where $\hat{x}_{ij}$ is the mean of the 3x3 ROI centred in x_{ij}, m_o (see Figure 7.5) is the mode of the seed region, $\sigma_{1,s}$ is the standard deviation of the pixels in the seed region below m_o, $\sigma_{2,s}$ is the standard deviation of the pixels in the seed region above m_o and k is a parameter that controls the marginal step size. This parameter k gives this region grower the capability of adapting the homogeneity criterion to the different difficulties exhibited in typical autoradiographic data with low CNR, such as blurred edges and high statistical noise.

The reference pixel x_{ij} is usually used in equation 7.9, instead of $\hat{x}_{ij}$, but due to the high statistical noise present in the image this represents an improvement to avoid a premature halt of the region growing. An initial value of k is set manually but this is later increased dynamically as in [114] (see Figure 7.6).

The use of two different standard deviations is based on the typical case of a textured region where the histogram of the original seed area is modelled with a Gaussian mixture [142]; thus the mode of the histogram approximately sets the middle of the mixture, one standard deviation ($\sigma_{1,s}$) represents the lower values of the mixture describing, in an approximate way, the variation of width of an assumed Gaussian like (but contaminated) distribution, and the second ($\sigma_{2,s}$) is a similar descriptor for the higher values of the mixture (see Figure 7.5). If the seed region is placed on a very homogeneous region and the histogram can be modelled with a single Gaussian, i.e. $\sigma_{1,s}$ and $\sigma_{2,s}$ are very similar, this does not represent a problem for the correct modeling of the statistics of the seed region.

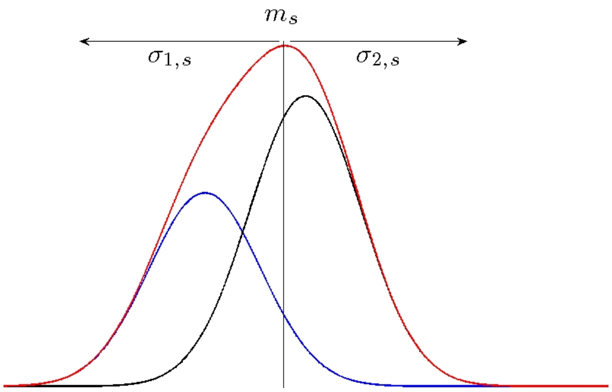


Figure 7.5: Model of a low CNR Gaussian mixture where two different standard deviations, $\sigma_{1,s}$ for lower values and $\sigma_{2,s}$ for higher values, are computed to better model the statistics of the original seed region.

Some parameters are measured from the intensity histogram at each iteration t to be used by the termination criterion based on the intensity histogram (explained below):

- mode of the current region being segmented (m_t).
- the difference between the mode intensity of the region at the current iteration t (m_t) and the mode of the original seed region (m_o), denoted as Δ_{m1}.
- the difference of the mode intensity of the region at iteration t (m_t) and the previous mode at iteration *t-1* (m_{t-1}), denoted as Δ_{m2}.

Every time the contour T (equation 7.8) is extracted, all the pixels contained in the one pixel wide T_t (t denoting that the contour T corresponds to the t$_{th}$ iteration) are candidates to be appended to the region $R_{i,t}$. After appending those pixels that fulfil equation 7.9 a new region is obtained $R_{i,t+1}$. Using the same parameter k, the contour T_{t+1} is extracted again in the next iteration $(t+1)$ from the new region $R_{i,t+1}$, to append those pixels that still fulfil equation 7.9.

Several iterations of the region grower will occur, using the initial parameter setting in equation 7.9, until no further pixels are appended to the region. Parameter k is then adaptively changed by a certain step size ($k = k + \triangle k$) in response to the CNR of the image; the step size $\triangle k$ used in the experiments shown here has a value of 1. This parameter can be thought as a way of controlling the sensitivity of the region grower to edges. If this parameter $\triangle k$ is large (>1), every time it is updated, the margins defined by equation 7.9 will allow to append more pixels, possibly missing some edges. On the other hand, if $\triangle k$ is small (<1), the algorithm could be too sensitive to noise and halt prematurely. To give $\triangle k$ a specific value is therefore a matter of controlling the sensitivity the algorithm given the noise present in the image data. For images with high CNR (>20) the step size could be higher (2-3) and for images with lower CNR (<5) the step size could be lower (0.2-0.5). By dynamically updating k, this parameter relaxes equation 7.9 so the number of pixels appended to the region being grown is increased. Figure 7.6 shows the evolution of the boundary for different values of k, from the initial *k=3* to the final *k=17* for this example.

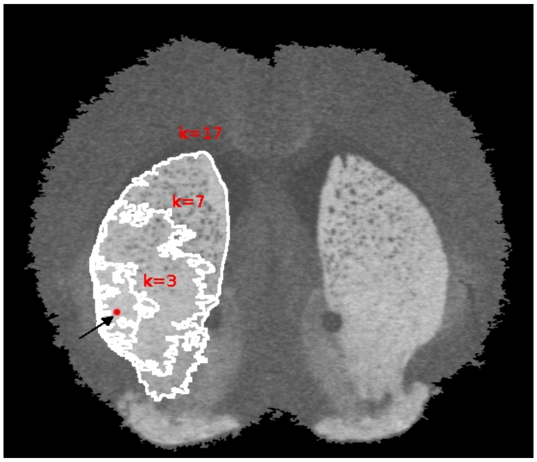


Figure 7.6: Evolution of the boundary as k automatically increases, relaxing equation 7.9, for *k=3*, *k=7* and *k=17*. The initial seed region, painted in red, is indicated with an arrow.

It is worth to make the difference at this point of the two different indices mentioned so far. The first of them, t, increases every time a new contour $T_{i,t}$ is extracted and subsequently the candidates in the contour that fulfil the homogeneity criterion are appended to $R_{i,t}$. This process is repeated until no more pixels can be appended to $R_{i,t}$, which is when the other index k increases, updating the similarity criterion, and t is initialised.

Every time t or k iterate the termination criteria are evaluated to find out if the region grower stopped because it found an edge, or because, due to statistical noise or inhomogeneities in the region, it found a false edge. The termination criteria can be divided in two groups: termination criteria evaluated when t is increased and termination criteria evaluated when k is increased.

The termination criterion evaluated when t is updated is briefly explained below and further detailed in Section 7.3.4:

1. A significant change of mode in the intensity histogram of $R_{i,t}$ while t iterates is considered to indicate an over-growing situation. This termination criterion is further explained below.

The termination criteria evaluated when k is updated are briefly explained below and further detailed in Section 7.3.4:

1. The number of iterations t that the region grower iterates appending pixels, until it can not append any more and k is updated is also considered. The physical explanation of this is that if k is updated (increased $\triangle$k), and few iterations t occur (with the new k value) until the region grower halts again, can be an indication of a very inhomogeneous region. Given this situation the region grower considers that it is leaking in to a different region (different texture).
2. The difference of areas between regions at k and $k-1$ is calculated. A significant jump between region sizes at consecutive iterations may indicate that the region grower is over-growing. This is futher explained in Section 7.3.4 below.
3. The mean of the gradient of the outer contour $T_{i,t}$ is defined as C^k. If a significant drop between C^k and C^{k-1} is observed the region grower considers that $T_{i,t}$ was placed on an edge in the previous iteration.

If none of the termination criteria determines that a discontinuity between textures has been found the parameter k is subsequently updated to continue appending pixels (see equation 7.9). In this case the region grower applies a penalty scheme to the outer contour $T_{i,t}$ detailed below.

Penalisation scheme:

The intensity values of those pixels that are contained in the outer contour $T_{i,t}$ of the segmented region at the current iteration $R_{i,t}$ are penalized. This penalty decreases the intensity value of only those pixels contained in $T_{i,t}$ by a certain value, depending on the relative overlap between the intensity histograms of $R_{i,t}$ and its corresponding outer contour $T_{i,t}$, defined as p_R and p_T respectively. The penalty is defined by equation 7.10:

$$T_{i,t} = T_{i,t} - w(\mu_s - \mu_c) \tag{7.10}$$

where w is a weighting factor (explained below) and μ_s and μ_c are the mean intensity value of $R_{i,t}$ and $T_{i,t}$ respectively. Note that μ_s and μ_c will always fulfil $\mu_s > \mu_c$, i.e. the mean of the pixel values in the region being segmented $R_{i,t}$ will always be higher than the mean of the pixel values in the contour $T_{i,t}$.

The physical phenomenon that explains this penalty is because some times the region grower stops appending pixels due to high statistical noise present in the region being segmented. If the intensity value of the pixels contained in $T_{i,t}$ are similar enough to those pixels in $R_{i,t}$ (which means the region grower stopped due to the presence of significant statistical noise), after updating parameter k these pixels in $T_{i,t}$ will be appended to the region being grown. On the other hand, if the intensity value of the pixels in $T_{i,t}$ are very dissimilar compared to those pixels in $R_{i,t}$, the intensity penalty will increase such dissimilarity of intensity value, keeping these pixels out of the region grown.

The weighting factor w is obtained from p_R and p_T, the instensity histograms of those pixels contained in $R_{i,t}$ and $T_{i,t}$ as mentioned above. These intensity histograms will sometimes be very low populated, i.e. constructed with very few samples when $R_{i,t}$ is small. Therefore some considerations have to be taken to reduce the effect of poor sampling and produce undesirable results.

To obtain w first a set of parameters $N(m, \sigma_u, \sigma_l)$ from each $R_{i,t}$ and $T_{i,t}$ are obtained, m being the mode, and σ_u and σ_l the standard deviation of those pixels below and above m respectively, similarly to that described above. Then p_R and p_T are modelled as a finite mixture model using the aforementioned set of parameters $N(m, \sigma_u, \sigma_l)$ for each distribution. An idealised plot of p_R and p_T is shown in Figure 7.7, where p_R is plot in blue and p_T in red, as p_R corresponds to $R_{i,t}$ which contains by definition higher intensity values than $T_{i,t}$.

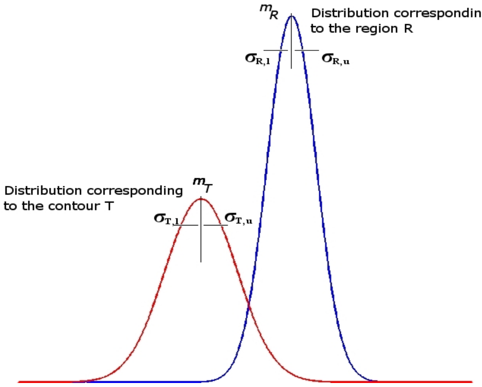


Figure 7.7: Intensity histogram of $R_{i,t}$, p_R, shown in blue and of $T_{i,t}$, p_T, shown in red. The parameters measured from each distribution are also shown.

After being obtained p_R and p_T a measure of the overlapping between these two is computed. In order to do this two parameters are obtained. First $p_R \cap p_T$, representing the measure of overlapping between both distributions, and secondly $p_R \cup p_T$, used to normalise the overlapping between both distributions $p_R \cap p_T$. The overlapping measure O is defined by equation 7.11, being $O = [0, 1)$.

$$O = \frac{\sum p_{p_R \cap p_T}}{\sum p_{p_R \cup p_T}} \tag{7.11}$$

If $O = 0$ means that there exists no overlapping between p_R and p_T (completely different tissues). The situation $O = 1$ never arises as this would imply both distributions are identical, a situation which is impossible because at this point the region grower is stopped because the intensity pixel of those pixels in $T_{i,t}$ can not be appended to $R_{i,t}$ due to intensity disparity.

After the model of each distribution, $N_R(m_R, \sigma_{R,u}, \sigma_{R,l})$ and $N_T(m_T, \sigma_{T,u}, \sigma_{T,l})$, and the overlapping O have been computed, w is then obtained following equation 7.12 to solve equation 7.10:

$$w = \begin{cases} 0 \ (No \ penalty) & \text{if } \sigma_{T,u} > 2\sigma_{T,l} \ \& \ \sigma_{R,l} > 2\sigma_{R,u}; \\ 1 \ (Maximum \ penalty) & \text{if O=0} \quad (p_R \cap p_T = \emptyset); \\ 1 - O \ (Medium \ penalty) & \text{otherwise} \quad (p_R \cap p_T > 1). \end{cases} \tag{7.12}$$

The first condition shown in equation 7.12 implies that there exists a substantial amount of overlapping, as shown in an idealised situation in Figure 7.8, where $\sigma_{T,u} > 2\sigma_{T,l}$ and $\sigma_{R,l} > 2\sigma_{R,u}$. In that case $w = 0$, which means that those pixels in $T_{i,t}$ are not penalized. The second condition establishes that if there exists no overlapping between p_R and p_T (i.e. $R_{i,t}$ and $T_{i,t}$ are very dissimilar) w is set to one, which means that those pixels in $T_{i,t}$ suffer maximum penalty, $\mu_s - \mu_c$. The last condition lies between the two extreme cases described above. The amount of penalty of the pixels in $T_{i,t}$ is related to the relative overlapping between the intensity histograms of $R_{i,t}$ and $T_{i,t}$, p_R and p_T respectively. The higher the overlapping (O↑) the lower the penalty ($w\downarrow$).

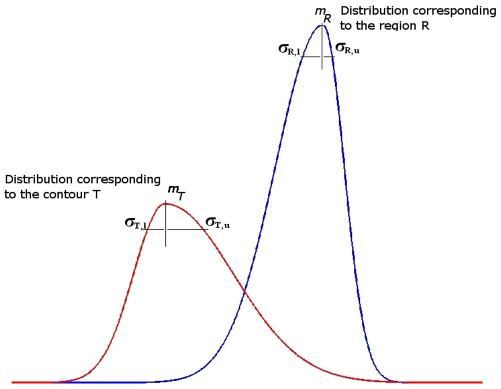
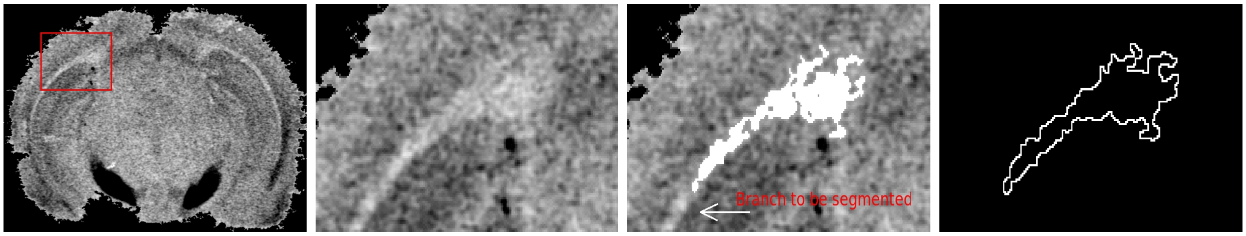


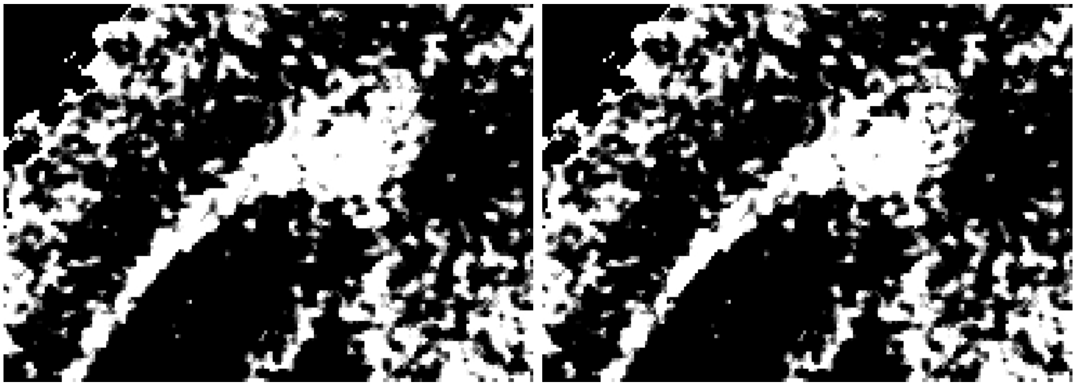
Figure 7.8: Intensity histogram of $R_{i,t}$, p_R, shown in blue and $T_{i,t}$, p_T, shown in red. The parameters measured from each distribution are also shown. This case shows a situation where $\sigma_{T,u} > 2\sigma_{T,l}$ and $\sigma_{R,l} > 2\sigma_{R,u}$.

The effect of the inclusion of the weighting factor w is observed in a practical case in Figure 7.9. Figure 7.9(a) shows the original autoradiogram, where the region under study is bounded by a red box. Figure 7.9(b) shows a close up of such region. The resulting segmentation for an initial k value $R_{i,t}$, superimposed on the autoradiogram, is shown in Figure 7.9(c). At this point the region grower can not append more pixels, so the weighting factor is evaluated. For this purpose the contour of the segmented region $T_{i,t}$ is extracted to obtain p_T, as shown in Figure 7.9(d). Figure 7.9(e) is the same image as Figure 7.9(b) but with a different intensity scale, where the contrast between the parts of the tissue being segmented and the tissue in the background are clearly differentiated. Figure 7.9(f) represents Figure 7.9(e) but with the weighting factor applied to those pixels in the contour (figure 7.9(d)), thus both images present the same pixel intensity in every pixel except in those pixels in the contour of the region being segmented where these have been penalized.

To understand this process note the bright region observed in Figure 7.9(b). In Figure 7.9(c) this bright region has been properly segmented but it stops appending pixels due to the high statistical noise present in the image. If only k is updated those pixels in the bright branch will be appended to the region being segmented, but some pixels located at the sides of the elongated segmented region (indicated in Figure 7.9(c)) will also be appended. By penalizing all the pixels in the contour as explained above, those pixels located at both sides of the elongated segmented region will not be appended, while those pixels in the mentioned bright branch will be appended.



(a) Original autoradiogram. (b) Zoom of the region being segmented. (c) Region being segmented indicated in white. (d) Contour of region being segmented.



(e) Original image. (f) Image with pixels in the contour after applying (7.12).

Figure 7.9: Process of the effect of weighting the pixels in the contour.

7.3.4 Termination Criteria

Termination criteria are those conditions that have to be fulfilled to halt the region growing process. The criteria used in this work have been developed based on the empirical observation of the properties observed in low CNR images such as digitised autoradiograms. These criteria are based on the points enumerated below:

- Intensity histogram of the region under segmentation.
- Too few iterations of t until k is updated.
- Difference between region sizes each time the parameter k is updated.
- Edge detection based on image gradient.

Termination criterion based on intensity histogram:

One of the conditions to stop the region grower is based on analysing the histogram of the segmented region at each iteration t. When segmenting low CNR imagery, it is observed that the object being segmented is usually described by a Gaussian-like distribution, but as the region grower appends more pixels to the region, a second Gaussian distribution emerges initially as a long tail. At a certain point the histogram of the region will develop into a bimodal but frequently asymmetric Gaussian mixture (see Figure 7.10).

As the histogram of the region grown develops into a bimodal histogram, the difference between the mode intensity at iteration t and the mode intensity computed from the initial seed region, defined by equation 7.13, is going to be significantly increased.

$$\Delta_{m1} = m_t - m_o \tag{7.13}$$

where m_t is the mode of the histogram at the current iteration t and m_o is the mode computed from the original seed region.

In some cases the mixture distribution can be highly asymmetric, making it difficult to distinguish these two components (see Figure 7.11). This effect might happen also when the histogram results in a Landau-like distribution. One way of detecting which situation has occurred is by studying the evolution of the mode of the histogram at each iteration, defined by equation 7.14.

$$\Delta_{m2} = m_t - m_{t-1} \tag{7.14}$$

where m_t is the mode of the histogram at the current iteration and m_{t-1} is the mode of the histogram at the previous iteration.

If the evolution of the histogram from the region being segmented results in a bimodal histogram, Δ_{m1} and Δ_{m2} will increase significantly at the iteration when the mode of the second distribution is higher than that of the first original distribution (indicated in Figure 7.10). In this case the region being grown has flooded across different tissues. A hard threshold (shown in Figure 7.10) is applied here as the lowest (non zero) histogram bin between the two distributions. Under this situation a portion of the object being segmented is incorrectly discarded (area A in Figure 7.10), and a portion of the segmented component is incorrectly assigned to the grown region (area B in Figure 7.10). After binarising the segmented region, it is observed that pixels corresponding to area B, are not spatially connected to the grown region being segmented. These pixels are therefore easily removed using binary morphological operations.

In the case, where a highly asymmetric long tailed Gaussian histogram shape is observed, neither Δ_{m1} nor Δ_{m2} will exhibit the step change seen in the previous case. Therefore region growing may also be halted by detecting any potential over-segmentation from observing the change in region size as well as described below.

In practise the intensity histogram of the region being segmented is not smooth, due to a mixture of high noise (variation) and low statistics. To distinguish between the two

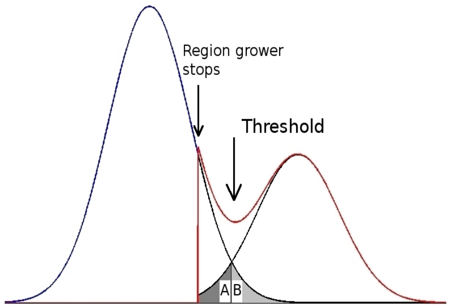


Figure 7.10: Idealised example of segmenting two different tissues resulting in a clearly distinguished bimodal distribution.

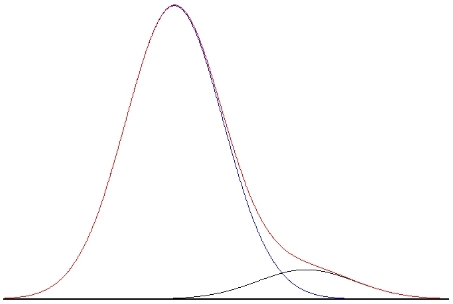


Figure 7.11: Idealised example of segmenting two different tissues with a highly asymmetric mixture distribution, making component separation difficult.

different situations explained above, and to avoid spurious halts of the region growing, then the two conditions shown in equation 7.15 have to be fulfilled to halt the region growing process:

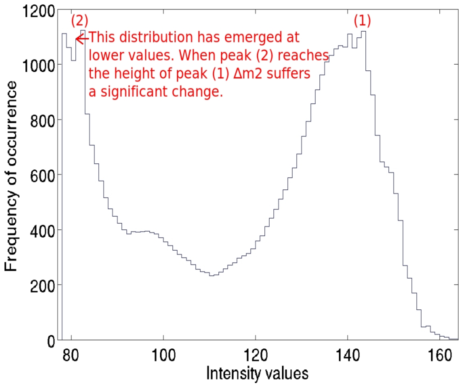
$$\begin{cases} max\{\Delta_{m1}\} - \sigma\{\Delta_{m1}\} > max\{\Delta_{m1acc}\} + \sigma\{\Delta_{m1acc}\} \\ max\{\Delta_{m2}\} - \sigma\{\Delta_{m2}\} > max\{\Delta_{m2acc}\} + \sigma\{\Delta_{m2acc}\} \end{cases} \tag{7.15}$$

where Δ_{m1acc} is a vector whose components represent the Δ_{m1} (7.13) accumulated values used each time k is incremented, and Δ_{m2acc} is similarly defined for Δ_{m2} (7.14). A real intensity histogram example is shown in Figure 7.12 with its corresponding mode (blue plot), Δ_{m2acc} (red plot) and Δ_{m2} (green plot) evolution, shown in Figure 7.13. The red arrows indicate every time the parameter k is increased. As mentioned previously index t iterates for an initial k value (k_{init}). When no more pixels are appended for that k value, this is increased by $k = k + 1$. After n iterations (k+n in Figure 7.13) a significant change of Δ_{m2} is detected, resulting in a halt of the region grower.

It can be observed how the mode flickers at the beginning when the histogram has low statistics, and then it stabilises until a large change is detected. This initial flickering might be discarded by checking whether there actually exists a valley between the distributions in the intensity histogram.

Termination criterion based on how many times t iterates until k is updated:

The second termination criterion evaluated every time k is updated is based on the number of times that t has been increased between consecutive iterations of k. The physical phenomenon in which this criterion is based on the following observation.



This distribution has emerged at lower values. When peak (2) reaches the height of peak (1) Δm2 suffers a significant change.

Figure 7.12: Example of intensity histogram where Δ_{m2} suffers a significant change.

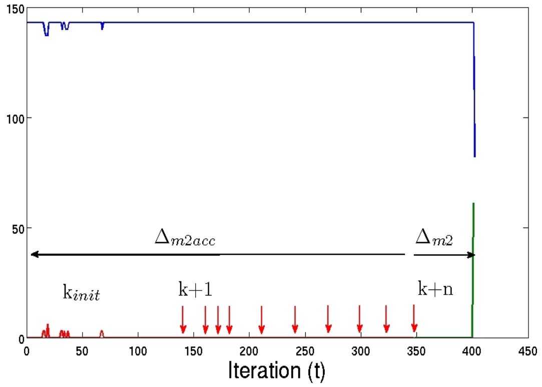


Figure 7.13: Evolution of the histogram mode (blue), Δ_{m2acc} (red) and Δ_{m2} (green) of intensity histogram shown in Figure 7.12. The increments of the parameter k are marked with red arrows.

While the region grower is growing on a relatively homogeneous region, the iterator t increases as more pixels are appended to $R_{i,t}$. Eventually the intensity value of the pixels in $T_{i,t}$ drop as the region grower gets close to the boundary of the functional region being segmented, preventing the region grower of appending more pixels. When this situation is given, k is increased to continue appending pixels. If the region grower halts after few iterations t (0-4 iterations) this situation suggests that the region grower is growing on a very inhomogeneous regions, i.e. a boundary. Therefore the region grower halts.

Termination criterion based on difference between region sizes:

Another of the termination criteria observes the difference in size of $R_{i,t}$ for consecutive values of k. The region size difference is measured as shown in equation 7.16.

$$I_k = A_k - A_{k-1} \tag{7.16}$$

where A_k is the region size of $R_{i,t}$ at iteration k and A_{k-1} is the region size of $R_{i,t}$ at iteration *k-1*. The region grower considers that region being segmented is leaking in to another region, and therefore over-growing, if $I_k > 5I_{k-1}$.

Termination criterion based on edge detection:

Before the region grower is applied to the autoradiogram, $G(i,j)$, the gradient of the image, is first computed as described in equation 7.17:

$$G(i,j) = \sqrt{\nabla_h I(i,j)^2 + \nabla_v I(i,j)^2 + \nabla_{d1} I(i,j)^2 + \nabla_{d2} I(i,j)^2} \tag{7.17}$$

where the terms $\nabla I(i,j)$ are the finite differences if pixel (i,j) with its 8 neighbours, as shown in equation 7.18:

$$\begin{cases} \nabla_h I(i,j) = \sqrt{(x(i,j)-x(i,j-1))^2+(x(i,j)-x(i,j+1))^2} \\ \nabla_v I(i,j) = \sqrt{(x(i,j)-x(i-1,j))^2+(x(i,j)-x(i+1,j))^2} \\ \nabla_{d1} I(i,j) = \sqrt{(x(i,j)-x(i-1,j-1))^2+(x(i,j)-x(i+1,j+1))^2} \\ \nabla_{d2} I(i,j) = \sqrt{(x(i,j)-x(i+1,j-1))^2+(x(i,j)-x(i-1,j+1))^2} \end{cases} \tag{7.18}$$

Having computed the gradient $G(i,j)$ of the entire image, every time the parameter k is updated, a binary template of $T(i,j)$, defined as $T_b(i,j)$, is extracted, and the gradient of only those pixels contained in $T(i,j)$ are summed up and normalised, dividing by the cardinality of the contour $(\sum T_b(i,j)^k)$, resulting in an updated gradient scalar measure defined by C^k:

$$C^k = \frac{\sum G(i,j)T_b(i,j)^k}{\sum T_b(i,j)^k} \tag{7.19}$$

If $C^k < C^{k-1}$ the region grower considers that it has missed an edge in the previous iteration $k-1$, therefore the region grower halts and considers the region segmented at iteration $k-1$ as valid.

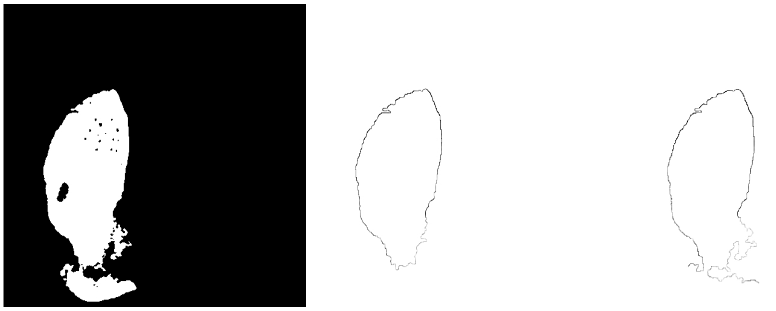
In Figure 7.14 this process is observed. Figure 7.14(a) represents the original image, Figures 7.14(c) and 7.14(d) represent the segmented region at iterations k-1 and k respectively. The gradient of the image $G(i,j)$ shown only where the contour of the segmented region is present $T(i,j)$ is shown in Figures 7.14(e) and 7.14(f) for k-1 and k respectively. For this specific example, at iteration k, C^k has a value of 0.38, while for the previous iteration C^{k-1} has a value of 0.53. Given this situation the region grower stops and takes the previous segmented region at k-1 as valid. Observing the original autoradiogram in Figure 7.14(a) it is observed that the segmented region actually corresponds with the left side of the known functional structure Caudate Putamen.

In order to avoid premature halting of the region grower, a second condition has to be fulfilled by C^{k-1} to stop the region grower. As has been mentioned, the level of statistical noise in the typical autoradiograms used here is significant, so there may be some iterations, typically early on in the region growing process, where C^k is slightly lower than C^{k-1}, sufficient to stop the region grower, but does not actually corresponds with a functional region. This is why a second noise check condition to be fulfilled by C^{k-1} has been introduced: C^{k-1} has to be above a certain threshold to actually be considered. This threshold is $\sigma_g + m_g$, where σ_g and m_g are the standard deviation and the mode of the gradient image $G(i,j)$ respectively (considering in $G(i,j)$ only those pixels above zero).

Combination of termination criteria:

As has been mentioned, the termination criterion based on significant jumps on the intensity histogram mode is evaluated every time the iteration index t is updated. The rest of termination criteria are all evaluated one after the other when k is updated.

(a) Original autoradiogram. (b) Gradient of original autoradiogram, $G(i, j)$. The image is grey scale inverted for illustration purposes. (c) Segmented region at k-1.



(d) Segmented region at k. (e) Gradient in the contour at k-1, C^{k-1}. The image is Gray scale inverted for illustration purposes. (f) Gradient in the contour at k, C^{k}. The image is Gray scale inverted for illustration purposes.

Figure 7.14: Process of edge detection to halt the region grower.

The first one to be evaluated is the one that considers how many iterations of t have happened between consecutive increases of k. Secondly the difference between consecutive region sizes is evaluated. Thirdly the edges of consecutive regions are evaluated as explained above.

If only one of them considers that the region growing process has to stop, then it stops. There is no priority established between them, thus the order in which they are evaluated is not critical.

In order to segment the rest of the objects in the image, consecutive seed pixels are located using the criterion of the next highest pixel value in the image excluding those already segmented. This pixel needs to be a minimum distance of M pixels (a value of 30 was used here) from the previously segmented regions to avoid over-segmentation by growing regions too close. These regions consecutively segmented comprise $R_a = \{R_{a1}, R_{a2}, ...R_{an}\}$, R_a being the final fully labelled image, n the total number of segmented regions and $R_{a\theta}$ each of the segmented regions. As has been mentioned, once R_a is obtained the autoradiogram is inverted in intensity and the process is repeated exactly in the same way, producing now $R_b = \{R_{b1}, R_{b2}, ...R_{bm}\}$, R_b being the final fully labelled image of the intensity inverted autoradiogram, m the total number of segmented regions and $R_{b\theta}$ each of the segmented regions.

7.3.5 Region Merging

Image segmentation algorithms often suffer from undersegmentation or oversegmentation because segmentation algorithms rely on relatively homogeneous textures. This effect has also been observed in the studies undertaken in this work. Therefore a post-processing step based on region merging is presented to obtain more robust image segmentation results.

First the labelled images obtained after the segmentation of the original autoradiogram (R_a) and the segmentation of the intensity-inverted autoradiogram (R_b) are combined.

The way the combination of R_a and R_b is undertaken is by considering the intersection between regions in R_a and R_b. Denoting the final fully segmented image as R_l the fusion initialises by setting $R_l = R_b$. Then the overlapping of each region of R_a with R_l is studied. The following considerations are then taken:

1. If $R_{a\theta} \cap R_l = \emptyset$ then $R_l = R_l \cup R_{a\theta}$.
2. If $R_{a\theta} \cap R_l \neq \emptyset$ and $N(R_{a\theta} \cap R_l) = 1$, $N(\cdot)$ being the number of different regions in the intersection, then $R_l = R_l \cup R_{a\theta}$.
3. If $R_{a\theta} \cap R_l \neq \emptyset$ and $N(R_{a\theta} \cap R_l) > 1$, then the proportion of overlapping between both regions, I_{ab} ($I_{ab} = R_{a\theta} \cap R_{b\theta}$) with $R_{a\theta}$ and with $R_{b\theta}$, defined as p_1 and p_2 respectively, is computed as defined in equation 7.20:

$$p_1 = \frac{\sum\limits_{i \in I_{ab}} x_i}{\sum\limits_{i \in R_{a\theta}} x_i} \qquad p_2 = \frac{\sum\limits_{i \in I_{ab}} x_i}{\sum\limits_{i \in R_{b\theta}} x_i} \tag{7.20}$$

where $\sum_{i \in I_{ab}} x_i$ represents the sum of the pixel values in I_{ab}, $\sum_{i \in R_{a\theta}} x_i$ represents the sum of the pixel values in $R_{a\theta}$ and similarly with $R_{b\theta}$. If $p_1 < 1\%$ & $p_2 < 1\%$ then I_{ab} is ignored. Otherwise I_{ab} is merged to $R_{a\theta}$ or $R_{b\theta}$ following equation 7.21:

$$\begin{cases} R_{a\theta} = R_{a\theta} \cup I_{ab} \quad if \quad \mu_{R_a} - \mu_{I_{ab}} < \mu_{R_b} - \mu_{I_{ab}} \Rightarrow R_l = R_l \cup R_{a\theta} \\ R_{b\theta} = R_{b\theta} \cup I_{ab} \quad if \quad \mu_{R_a} - \mu_{I_{ab}} > \mu_{R_b} - \mu_{I_{ab}} \Rightarrow R_l = R_l \cup R_{b\theta} \end{cases} \tag{7.21}$$

where $\mu_{I_{ab}}$ is the mean of the pixels in the overlapped area I_{ab}, μ_{R_a} is the mean of $R_{a\theta}$ and μ_{R_b} is the mean of $R_{b\theta}$.

An example, where R_a intersects with two regions, R_{b1} and R_{b21} ($N(R_{a\theta} \cap R_l) = 2$) is shown in Figure 7.15. The intersection of R_a with R_{b1} is shown in purple and denoted by I_{ab1} in Figure 7.15(d). Similarly, the intersection of R_a with R_{b2} is

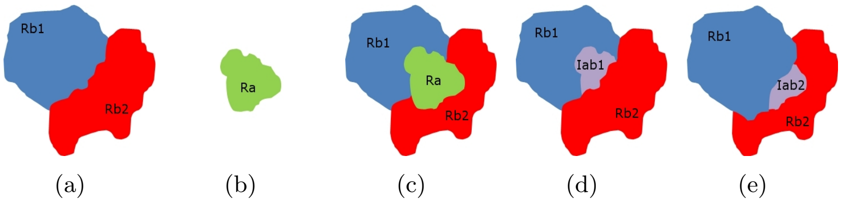
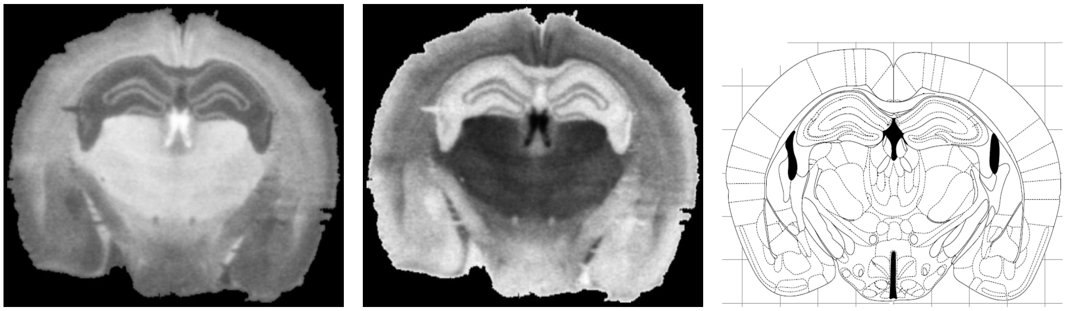


Figure 7.15: $R_b = \{R_{b1}, R_{b2}\}$ (a), $R_a = \{R_a\}$ (b), intersection of R_a and R_b (c), intersection of R_a with R_{b1} (I_{ab1}) shown in purple (d) and intersection of R_a with R_{b2} (I_{ab2}) shown in purple (e).

> shown in purple and denoted by I_{ab2} in Figure 7.15(e). Both regions, I_{ab1} and I_{ab2}, are independently evaluated as explained above.

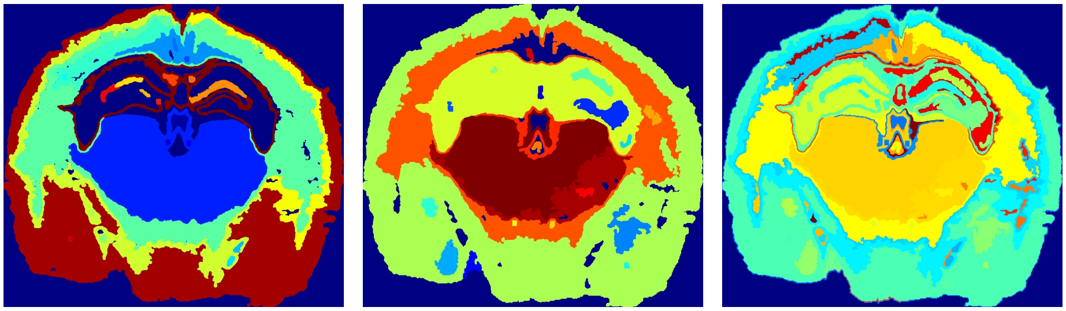
A real example of a fusion of labelled (segmented) images is shown in Figure 7.16. Figures 7.16(a) and 7.16(b) show the original and the intensity-inverted autoradiograms respectively. Figure 7.16(c) shows the corresponding atlas image (Bregma -1.82 mm) considered as the ground truth. Figures 7.16(d) and 7.16(e) show the corresponding labelled images and Figure 7.16(f) shows the final fusioned labelled map.



(a) Original autoradiogram.

(b) Intensity-inverted autoradiogram.

(c) Corresponding atlas image (Bregma -1.82 mm).



(d) Labelled map of original autoradiogram.

(e) Labelled map of intensity-inverted autoradiogram.

(f) Fusion of labelled maps.

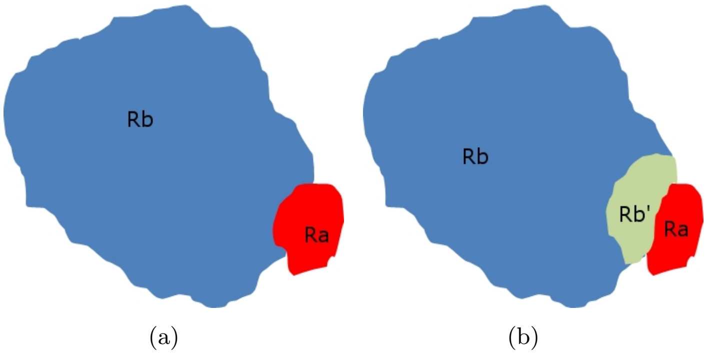
Figure 7.16: The fusion of labelled maps (d) and (e) produces a more detailed labelled map (segmentation) of the original autoradiogram (a).

This fusion results in an over-segmented autoradiogram. To obtain a more accurate segmentation some of these regions have to be merged. This has been undertaken by evaluating the similarity between intensity histograms of adjacent regions in the fusioned labelled image. The metric used in this work is the Dice similarity index [47], based on the proportion of overlapping of intensity histograms. The percentage of overlapping for two arbitrary regions R_{a1} and R_{b1}, being their intensity histograms h_1 and h_2 respectively, is defined in equation 7.22:

$$p_{ovl}(\%) = 2\frac{\sum h_1 \bigcap h_2}{\sum h_1 + h_2} 100\% \tag{7.22}$$

This measurement will result in 1 for a perfect overlapping and 0 for a total absence of overlapping.

In some cases where $|R_{a\theta}| << |R_{b\theta}|$, $|\cdot|$ being the size of the corresponding region, a direct comparison of h_1 and h_2 may not be appropriated due to the much higher amount of samples in one histogram compared to the other. This problem has been tackled in this work by considering only those pixels that comprise a sub-region of $R_{b\theta}$, denominated $R'_{b\theta}$, in close proximity of $R_{a\theta}$ when this situation is given. In this way, not only the histograms integrals are equal, but also it is assumed that those pixels from the subregion of $R'_{b\theta}$ show a more accurate similarity or dissimilarity measure with $R_{a\theta}$. Figure 7.17 shows an example where $|R_{a\theta}| << |R_{b\theta}|$ and therefore a sub-region of $R_{b\theta}$ shown in green ($R'_{b\theta}$) is used to calculate a more representative measure of $p_{ovl}(\%)$ (equation 7.22).



(a) (b)

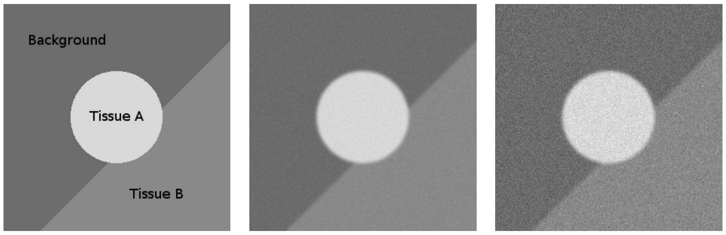
Figure 7.17: $R_{a\theta}$ and $R_{b\theta}$ where $|R_{a\theta}| << |R_{b\theta}|$ (a) and sub-region of $R_{b\theta}$, $R'_{b\theta}$, with similar size compared to $R_{a\theta}$ (b).

After $p_{ovl}(\%)$ has been computed for all those regions in contact in the merged labelled map (Figure 7.16(f)), if $p_{ovl}(\%) > p$ the corresponding regions will be merged, otherwise they will remain separate, p being manually set in each experiment.

7.4 Validation

The performance of the segmentation algorithm is assessed measuring how often the algorithm results in a correct decision, in order to predict how this will perform when used routinely. The issue here is that counting pixel regions as correct or incorrectly segmented regions is too complex due to lack of absolute ground truth in a functional biological image. We have used therefore a set of simple test simulated images where the ground-truth and noise are known. The metrics used to evaluate the segmentation technique presented here is based on measuring over-segmentation, under-segmentation and missing regions depending on the noise added to the simulated image [26].

An initial 300x300 pixels test simulated image, shown in Figure 7.18(a), is comprised of two different tissues (A and B) and a background. Tissue A is a 60 pixels radius circumference centred in the test image, while tissue B represents a piece of tissue with lower intensity value but larger in extension. The test image has been deliberately corrupted adding Gaussian noise with different values of variance, representing the typical statistical noise present in these kinds of imaging applications, and by blurring the edges filtering the ideal image with a Gaussian spatial filter using different variance values. Figures 7.18(b) and 7.18(c) represent the ideal image blurred with a σ=5 pixels PSF and with additive Gaussian noise with μ=0 and σ=2 and σ=6 respectively. These σ values are typically found in autoradiographic data.



(a) (b) (c)

Figure 7.18: Simulated noiseless image (a), simulated image blurred with a σ=5 PSF and additive Gaussian noise (μ=0 and σ=2) (b)and simulated image blurred with a σ=5 PSF and additive Gaussian noise (μ=0 and σ=6) (c).

These variance values used in the blurring process correspond to the system impulse response, defined by the Point Spread Function (PSF), that are observed in the different imaging modalities. In the case of autoradiography this blurring effect is mainly due to the β particle mean range (~0.3 μm for ^{3}H, ~13 μm for ^{14}C and ^{35}S) and by the deposited charge diffusion process in the active layer of a digital imaging device. Based on experimental work two different scenarios are considered here. For low energy radioisotopes, such as ^{3}H, a σ=2 pixels has been used, and for medium energy radioisotopes such as ^{14}C and ^{35}S, a σ=5 pixels has been used.

The assessment of this image segmentation algorithm is based on varying the Contrast-to-Noise-Ratio (CNR) of the test image. Mean intensity values and noise values have been directly measured in a variety of autoradiograms labelled with different radioisotopes to use real values in the assessment of the algorithm. The measure of the assessment is based on the percentage of over-segmented and under-segmented pixels of each tissue for different CNR values. The typical definition of CNR for an image with uniform noise in the background, as is this case, is shown in 7.23. This equation is not applicable in a real situation, where different values of noise (σ) are measured in every different tissue (texture) and background, but this is not considered in the test image for simplicity of the validation.

$$CNR = \frac{S_{AB}}{\sigma_{bck}} \tag{7.23}$$

where S_{AB} is the difference μ_A-μ_B, μ_A and μ_B being the mean intensities of ROIs in tissues A and B respectively, and σ_{bck} is the standard deviation of the background, considered as the noise. The Gaussian noise added to the test image has the same variance for the three structures (tissue A, tissue B and background), therefore the standard deviation measured in any of the regions is the same.

It is observed from equation 7.23 that the CNR depends on the difference of intensities between the two tissues under study and the statistical noise present. In the approach used in this work the difference of intensities between tissues is fixed to a value measured directly from autoradiographic data, but the statistical noise is increased.

After manually measuring the intensity in some of our autoradiograms, the values used for the assessment here are S_{AB}=40 (tissues A and B) and S_{Bbck}=35 (tissues B and background). The standard deviation measured in different regions of real autoradiographic data varied from 8 down to 3 in some cases. In this work a range of 0.5 up to 10 has been analysed, resulting in a CNR range of 81.0 (σ=0.5) down to 4.0 (σ=10), as observed in Figure 7.19. In Figure 7.20 the percentage of over-segmented and under-segmented pixels are shown when segmenting tissue A and tissue B. Each combination (PSF and σ) has been repeated over 30 times with different statistical noise realisations, therefore a statistical mean and $\pm\sigma$ for each combination has been obtained, as shown in Figure 7.20 with error bars. Those error bars that show high $\pm\sigma$ for CNR<6 in Tissue B (shown in black in Figure 7.20(a) and in red in Figure 7.20(b)) represent situations where either the region was barely segmented due to too much noise (Figure 7.20(a)), or where the segmentation completely leaked in the background tissue.

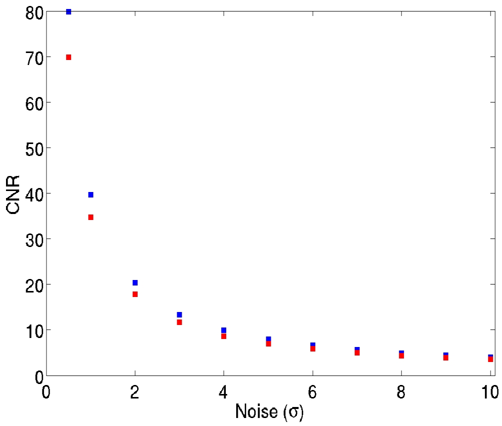
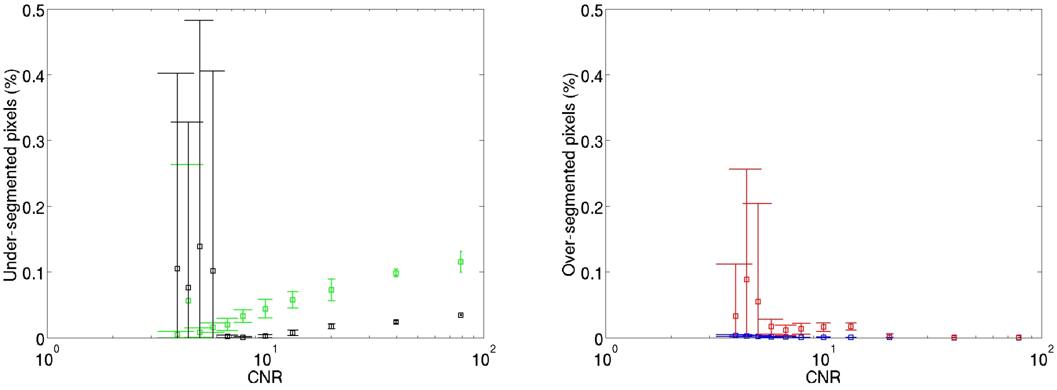


Figure 7.19: CNR measured between Tissue A and B (blue) and B and background (red) with statistical noise σ.

In Figure 7.20(a) it is observed how Tissue A slightly suffers under-segmentation <15% for low values of σ resulting from the blurring effect in the edges. As the noise increases the blurring affect takes less importance, therefore the number of under-segmented



(a) (b)

Figure 7.20: Percentage of under-segmented (left) and over-segmented (right) pixels of segmentation of tissue A (green and blue) and tissue B (black and red). Note the logarithmic scale in the abscissa axis.

pixels decreases. Tissue B suffers in some cases of under-segmentation for high values of noise due to premature halts. In Figure 7.20(b) Tissue A never suffers of over-segmentation, while Tissue B suffers of over-segmentation in some cases when σ >7. In general this demonstrates how this approach is able to segment reliably data with CNR$\geq$6.

7.5 Results

Several autoradiograms have been used to assess the performance of this segmentation algorithm. Only a few examples using different radioisotopes are presented here.

7.5.1 Iodine-125 labelled sections

Iodine-125 (^{125}I) is a radioisotope of Iodine used in autoradiography to bind to nicotinic ligands, such as Epibatidine or α-Bungarotoxin, to study nicotinic acetylcholine receptors (nAChRs) in brain. This radioisotope has not been mentioned previously in this thesis work because it is not a β- emitter, it is a gamma emitter with a maximum energy of 35 keV, some of which are internally converted to x-rays. Due to the large density of these nAChRs receptors in brain, resulting autoradiographic data usually shows a significant number of regions with high levels of uptake.

To demonstrate the ability of the region grower described in this chapter to segment low CNR autoradiographic data the CNR of some manually drawn ROIs have been measured. As mentioned before, the CNR described in equation 7.23 is not applicable in this situation where the noise (σ) is different for different tissues. A widely used alternative definition of CNR for cases where non-uniform noise is present in the image is then used to measure the CNR between regions [160]. This definition is shown in equation 7.24:

$$CNR = \frac{S_{AB}}{\sqrt{\frac{\sigma_A^2+\sigma_B^2}{2}}} \tag{7.24}$$

where σ_A and σ_B are the standard deviation (i.e. noise) measured in each tissue A and B respectively.

An example of a ^{125}I-labelled autoradiogram, where the studied ROIs are indicated in white boxes, is shown in Figure 7.21. The mean and standard deviation of those ROIs are shown in the table next to the autoradiogram.

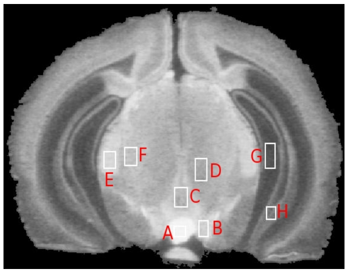


Figure 7.21: Autoradiogram of ^{125}[I]-epibatidine labelled brain section with ROIs under analysis indicated in white boxes.

Region	Mean	Standard Deviation
A	243	1.0
B	209	5.1
C	215	3.0
D	193	5.0
E	184	8.0
F	168	6.6
G	84	5.7
H	124	7.6

From those mean and standard deviation values shown in the above table, the resulting CNR measured between some significant adjacent regions, measured following equation 7.24, are shown in Table 7.1.

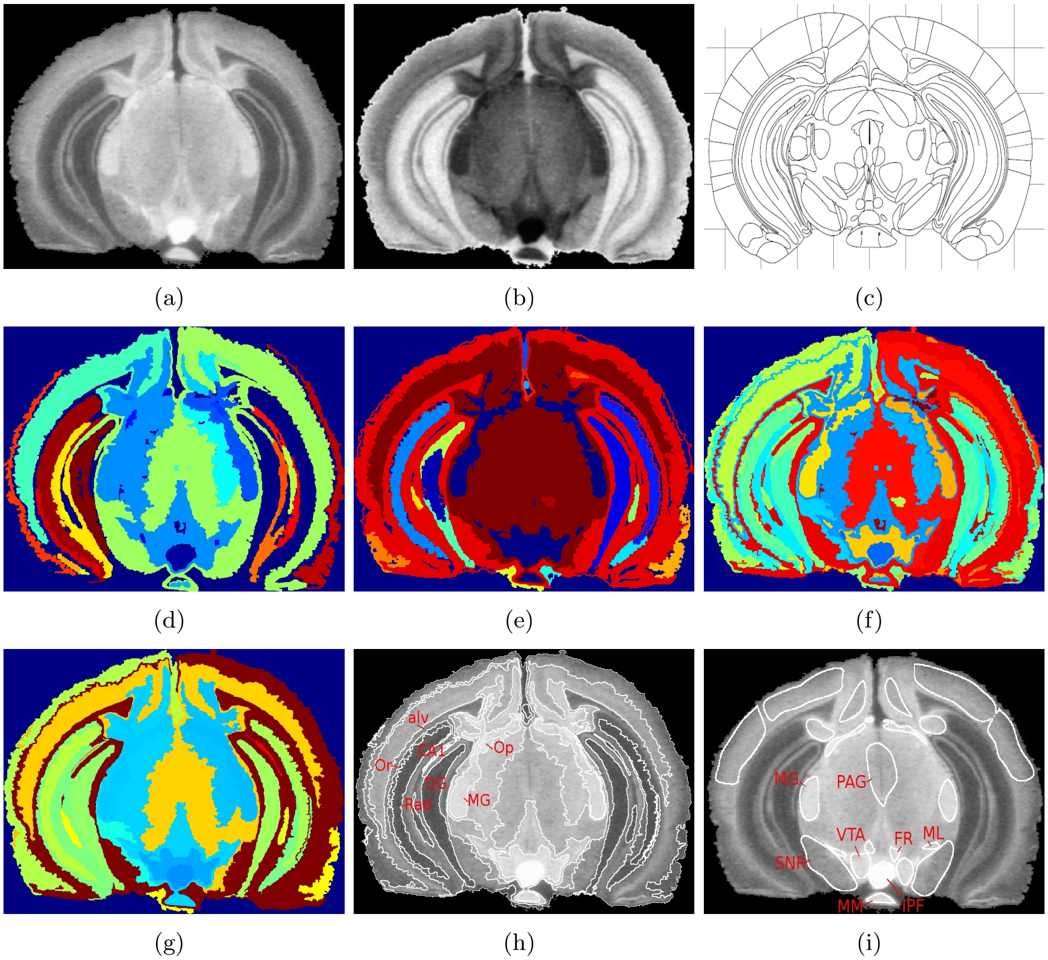
Table 7.1: CNR values for the regions shown in Figure 7.21

CNR_{AB}	9.2	CNR_{AC}	12.5	CNR_{CD}	5.3	CNR_{EF}	2.1	CNR_{GH}	5.9

Figure 7.22 represents an example of an autoradiogram of ^{125}[I]-epibatidine density binding to $\alpha_4\beta_2$ heteromeric nAChRs in a mouse brain section. This example shows the resulting segmentation of the original autoradiogram and its intensity inverted version. Observing the original autoradiogram (Figure 7.22(a)) it can be noted how some structures have a counterpart in the atlas image (Figure 7.22(c)), and some other structures do not, the reason being that atlases are created based on large data sets of images to obtain a generic atlas. This means that in practise images will not fit perfectly in to atlases.

Observing Figures 7.22(a) and 7.22(h) it can be identified how most of the brightest structures in the autoradiogram are correctly segmented. It can also be observed that those regions, where the CNR was measured in Figure 7.21, are successfully segmented.

This same autoradiogram was also segmented by an expert biologist following the same criteria as if it was an usual experiment, using the atlas book as reference. The resulting segmentation is shown in Figure 7.22(i). The name of those regions identified in Figure 7.22(i) are Periaqueductal Grey (PAG), Medial Geniculate Nucleus (MG), Interpeduncular Fossa (IPF), Medial Mammillary Nucleus (MM), Substantia Nigra (SNR), Fasciculus Retrofexlus (FR), Ventral Tegmental Area (VTA) and Medial Lemniscus (ML).



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 7.22: (a) Original autoradiogram of $\alpha_4\beta_2$ nAChRs binding in a mouse brain section (20 μm thick) labelled with ^{125}[I]-epibatidine (0.1 nM), cut at the level of Bregma -3.16 mm (b) original autoradiogram inverted, (c) corresponding atlas section, (d) resulting segmented autoradiogram, (e) resulting segmented inverted autoradiogram, (f) fusioned segmented autoradiograms (d) and (e), (g) fusioned segmented autoradiograms after merging regions with higher similarity, (h) original autoradiogram with those regions segmented delimited and (i) autoradiogram manually segmented by an expert biologist.

Some observations can be extracted by comparing the resulting fully segmented autoradiogram of this segmentation process, shown in Figure 7.22(h), and that same autoradiogram segmented by an expert biologist, shown in Figure 7.22(i).

First there are regions shown in Figure 7.22(i) clearly delineated based on the experience of the biologist that are not correctly identified by the segmentation process presented in this chapter, such as those regions observed in the cortical area. The region denominated as PAG does not show a distinct level of density ligand uptake compared to its background, thus being impossible to segment with the region grower. The region denominated as FR is delineated but it is not possible to distinguish from the VTA. The same happens with the region denominated as MG, which is correctly delineated but again, it merges with the Optic Nerve Layer (Op) region indicated in Figure 7.22(h). Those regions with clearly different levels of ligand density compared to adjacent regions are perfectly delineated, such as the SNR, MM and IPF. Some regions with low level of density ligand are also segmented such as the Dentate Gyrus (DG) or other small regions regions such as the Oriens Layer of the Hippocampus (Or), Alveus of the Hippocampus (alv), Stratum Radiatum of the Hippocampus (Rad) or Field CA1 of Hippocampus (CA1). Although these regions are not important for biologists they will help in the registration process detailed in next chapter to obtain a more accurate atlas warped on the autoradiogram.

In order to compare the manual segmentation and the automatic segmentation quantitatively the Dice similarity index between those regions that intersect has been computed. In the specific case of the example shown in Figure 7.22, those regions denominated as MM, IPF and VTA+FR have been evaluated. The contour of the mentioned regions manually segmented and the contour of the corresponding regions automatically segmented by the region grower are shown in Figure 7.23.

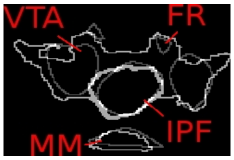


Figure 7.23: Contour of manually segmented regions and regions automatically segmented using the region grower with names of the regions indicated.

The Dice index, defined in equation 7.25, measured between those regions indicated in Figure 7.23 is shown in table 7.2.

$$D = 2\frac{A_1 \cap A_2}{A_1 + A_2} \tag{7.25}$$

Table 7.2: CNR values for the regions shown in Figure 7.23

Region	Dice index
IPF	0.89
MM	0.71
VTA+FR	0.61

In this case only three regions were compared. The MM and IPF regions were independently delineated by the expert and by the region grower, showing both a high Dice index. VTA and FR were delineated independently by the expert but the ligand density was equally distributed on that area, thus being segmented as one single functional area by the region grower. That is why the Dice index observed in Table 7.2 for the region VTA+FR is lower than in the other two regions.

7.5.2 Hydrogen-3 labelled sections

Selective Tritiated ligands are typically used to label receptors or transporters in the brain; among others selective ^{3}H ligands have been used to label various components of the dopaminergic system in the brain (e.g. D1, D2 and Dopamine Transporters). There is higher density of D1 receptors in Striatum regions of the brain compared to D2 receptors, therefore D1-labelled brain autoradiograms show brighter Striatum regions than those D2-labelled autoradiograms, and concomitantly higher CNR.

Similarly to the previous example the mean and standard deviation of several ROIs are also analysed, by manually placing these on significant regions of the original autoradiogram as shown in Figure 7.24.

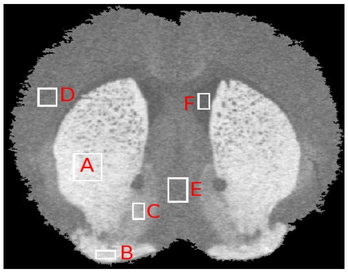


Figure 7.24: Autoradiogram of ^{3}H-labelled brain section with ROIs under analysis indicated in white boxes.

Region	Mean	Standard Deviation
A	153	6.7
B	141	5.9
C	116	6.3
D	82	5.0
E	85	5.0
F	70	4.3

Similar to the previous example, the CNR between some of the segmented regions shown in Figure 7.24 are shown in Table 7.3:

Table 7.3: CNR values for the regions shown in Figure 7.24

CNR_{AB}	1.9	CNR_{AC}	5.6	CNR_{AD}	12.0	CNR_{CE}	5.4	CNR_{EF}	3.2

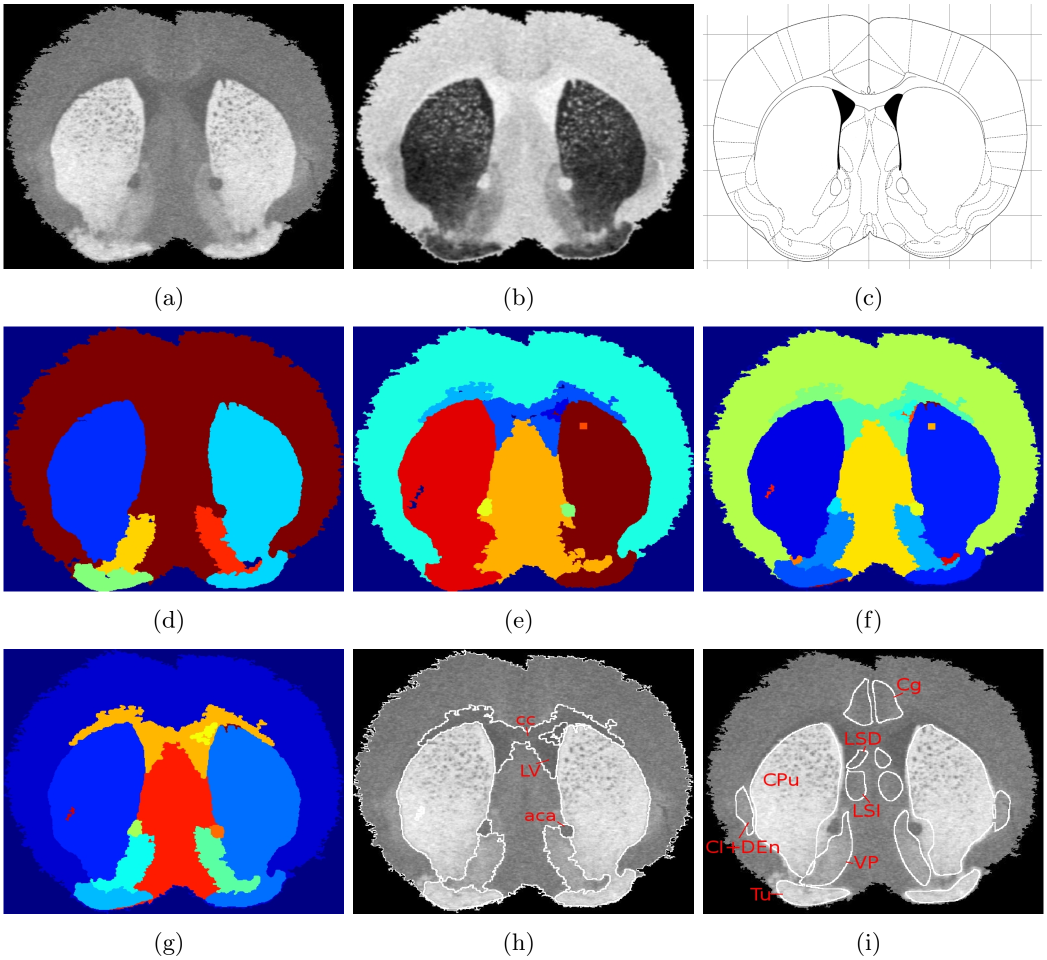
One example of an autoradiogram of mouse brain section labelled with ^{3}[H]SCH[2]-23390 to D1 receptors is shown in Figure 7.25. This section was cut at the level of the Striatum (Bregma +0.62 mm) where high density of ligand binding to dopamine D1 receptors is observed, specifically in the regions of the Caudate Putamen and the Olfactory Tubercle. Figure 7.25(i) shows the resulting manual delineation by an expert biologist, where the most important functional areas are indicated. These are the Caudate Putamen (CPu), Olfactory Tubercle (Tu), Ventral Pallidum (VP), Cingulate Cortex (Cg), Lateral Septal Nucleus (LSI and LSV) and Caudal Interstitial (CI) and Dorsal Endopiriform Nucleus (DEn).

Comparing Figure 7.25(h) and 7.25(i) it is observed how regions Cg, LSI and LSD are drawn based on experience, but they do not exhibit any change on ligand density, making impossible for the region grower to segment. On the other hand, regions CPu, Tu and VP are correctly delineated. The region denominated as CI+DEn is not delineated although a slight level of ligand density can be observed. The CNR measured between

---

[2] Schering Plough 23990 binding bound with ^{3}H

CI+DEn and the background tissue with lower ligand density is 3.5. In this case this CNR is too low to be segmented properly. It can also be observed how the Corpus Callosum (cc), Lateral Ventricle (LV) and the anterior part of the Anterior Commissure (aca), indicated in Figure 7.25(h), are segmented by the region grower. These regions are not segmented by biologists because they lack of importance for the purposes of their studies, but for the approach taken here all the regions have importance, for use in the registration process explained in next chapter.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 7.25: (a) Original autoradiogram of a mouse brain section (20 μm thick) cut at the level of the Striatum (Bregma +0.62 mm), labelled with ^{3}[H]SCH-23390 (4 nM) to D1 receptors, (b) original autoradiogram inverted, (c) corresponding atlas section, (d) resulting segmented autoradiogram, (e) resulting segmented inverted autoradiogram, (f) fusioned segmented autoradiograms (d) and (e), (g) fusioned segmented autoradiograms after merging regions with higher similarity, (h) original autoradiogram with those regions segmented delimited and (i) autoradiogram manually segmented by an expert biologist.

A similar quantitative comparison, compared to the previous example, has been undertaken. In the specific case of the example shown in Figure 7.25, those regions denominated as CPu, VP and Tu in both hemispheres have been evaluated. The contour of the mentioned regions manually segmented and the contour of the corresponding regions automatically segmented by the region grower are shown in Figure 7.26.

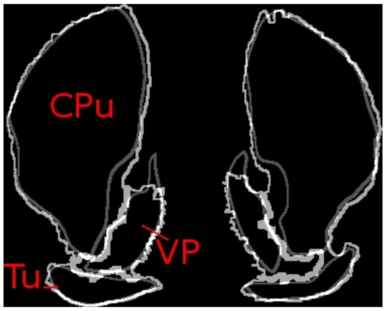


Figure 7.26: Contour of manually segmented regions and regions automatically segmented using the region grower with names of the regions indicated.

The Dice index measured between those regions indicated in Figure 7.26 is shown in table 7.4. In this case these Dice indices show higher values compared to the previous experiment. It is worth to mention that the region grower segmented the right CPu and Tu regions as one single region, due to the equally distributed ligand over those regions, showing no difference between regions. On the other hand a biologist is aware that both regions correspond to different anatomical regions but same functional regions, therefore they are delineated separately. This typical situation does not represent a problem for the later registration process, as both regions will be considered as one, and the warped atlas will define where is the limit between both anatomical regions.

Table 7.4: CNR values for the regions shown in Figure 7.26

Region	Dice index
CPu (left)	0.93
VP (left)	0.83
Tu (left)	0.90
CPu+Tu (right)	0.92
VP (right)	0.73

7.5.3 Sulphur-35 labelled sections

Sulphur-35 (^{35}S) is an isotope typically used to label, among others, the nucleotide ligand guanosine 5'-O-[γ-thio]triphosphate (GTPγS) resulting in the radioligand [^{35}S]GTPγS. This radioligand can be used to label the intracellular second messenger signal transduction G protein. The high energy of ^{35}S, compared to ^{3}H, (average energy ∼49 keV and maximum energy 167 keV) makes typical ^{35}S-labelled autoradiographic data to have intrinsically low spatial resolution, due to the high β electron range longer than the other radioisotopes used in this work (10 μm for a 50 keV β electron, see Figure 6.19). An additional issue with this type of autoradiographic data is the fact that G proteins show a widespread distribution of the protein in the brain, resulting in images with little difference of contrast between functional regions, being concomitantly more difficult to segment.

As in previous experiments, several ROIs are manually drawn and analysed, as shown in Figure 7.27. The mean and standard deviation of each indicated ROI are shown in the Table next to the autoradiogram.

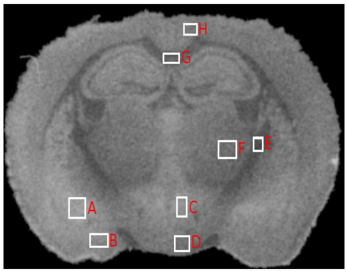


Figure 7.27: Autoradiogram of ^{35}S-labelled brain section with ROIs under analysis indicated in white boxes.

Region	Mean	Standard Deviation
A	156	4.9
B	137	7.3
C	143	4.9
D	114	6.0
E	81	5.4
F	110	5.5
G	96	7.0
H	121	4.6

From those mean and standard deviation values the CNR is obtained as shown in Table 7.5:

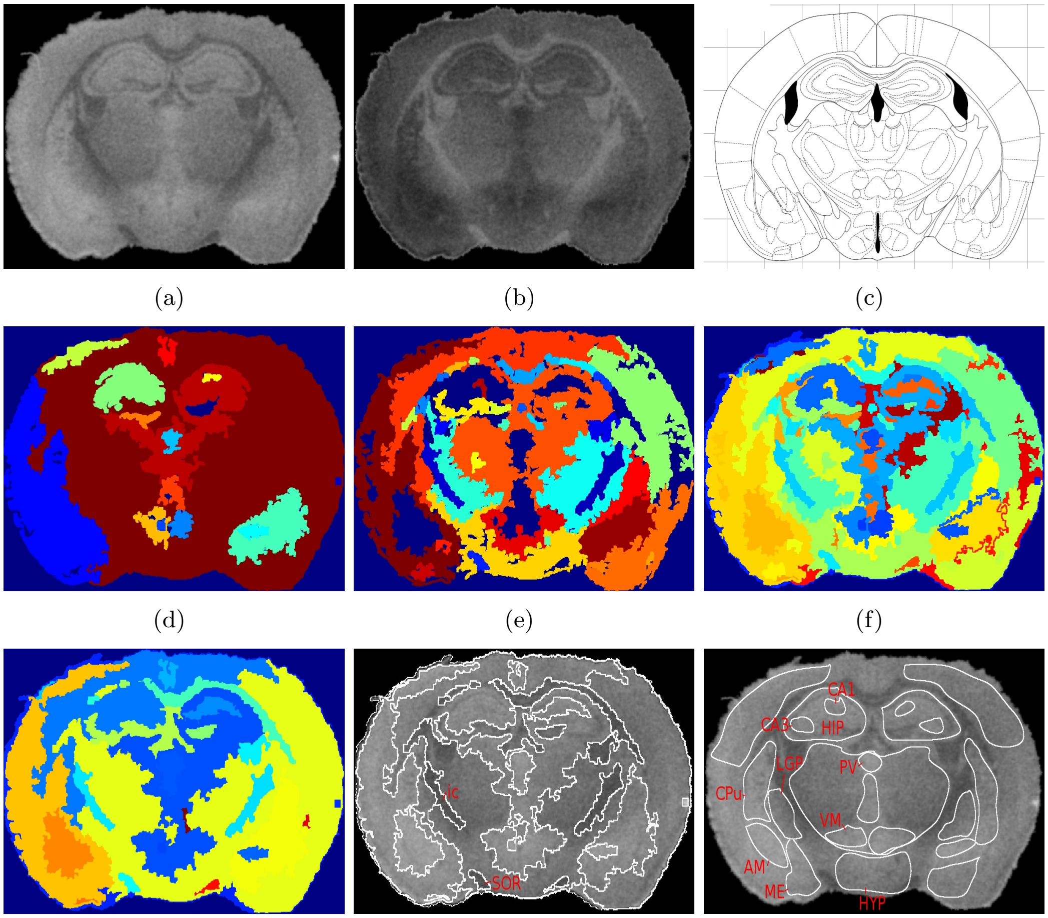
Table 7.5: CNR values for the regions shown in Figure 7.27

CNR_{AB}	3.0	CNR_{CD}	5.4	CNR_{EF}	5.2	CNR_{CF}	6.3	CNR_{GH}	4.0

One example of ^{35}[S]GTPγS labelled mouse brain section (20 μm thick) autoradiogram is shown in Figure 7.28. This example corresponds to an autoradiogram from the level of the Hippocampus (Bregma -1.46 mm). High density of G proteins are observed in the amygdaloid area and the hippocampal area. Figure 7.25(i) shows the resulting manual delineation by an expert biologist, where the most important functional areas are indicated. These are the Caudate Putamen (CPu), Hypothalamus (HYP), Amygdalae (AM), Medial Amygdaloid Nucleus (ME), Lateral Globus Pallidus (LGP), Hippocampus (HIP) and fields CA1 and CA3 of Hippocampus (CA1 and CA3), Ventromedial Thalamic Nucleus (VM) and Paraventricular Thalamic Nucleus (PV).

Comparing Figure 7.28(h) and 7.28(i) it is observed how multitutde of regions, such as CA1, CA3, HYP and ME among some examples, are drawn based on experience, but they do not exhibit any change on ligand density, making impossible for the region grower to segment. On the other hand, regions PV and AM in both hemispheres are correctly delineated. In addition it can also be observed how the regions Internal Capsule (ic) and the Supraoptic Nucleus (SOR), indicated in Figure 7.28(h), are segmented by the region grower. Similarly to the previous case, these regions are not segmented by biologists because they lack of importance for the purposes of their studies, but for the approach taken here all the regions have importance, as the further registration explained in next chapter is more accurate.

A similar quantitative comparison, compared to the previous examples, has been undertaken. In the specific case of the example shown in Figure 7.28, those regions



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 7.28: (a) Original autoradiogram of a mouse brain section (20 μm thick) cut at the level of the Hippocampus (Bregma -1.46 mm), labelled with [^{35}S]GTPγS to G proteins, (b) original autoradiogram inverted, (c) corresponding atlas section, (d) resulting segmented autoradiogram, (e) resulting segmented inverted autoradiogram, (f) fusioned segmented autoradiograms (d) and (e), (g) fusioned segmented autoradiograms after merging regions with higher similarity, (h) original autoradiogram with those regions segmented delimited and (i) autoradiogram manually segmented by an expert biologist.

denominated as PV and AM in both hemispheres have been evaluated. The contour of the mentioned regions manually segmented and the contour of the corresponding regions automatically segmented by the region grower are shown in Figure 7.29.

The Dice index measured between those regions indicated in Figure 7.29 is shown in table 7.6. In this case these Dice indices show lower values compared to the two previous experiments. In this example it was specially difficult to find regions similarly segmented by the expert and by the region grower. It has been observed how ^{35}S-labelled ligands have lower specificity and worse spatial resolution (due to the longer β-range) compared to ^{3}H and ^{125}I, resulting in specially low CNR autoradiograms. This

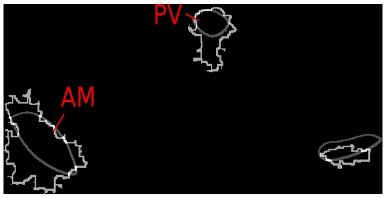


Figure 7.29: Contour of manually segmented regions and regions automatically segmented using the region grower with names of the regions indicated.

leads to higher disagreements with experts, as shown in Table 7.6.

Table 7.6: CNR values for the regions shown in Figure 7.29

Region	Dice index
PV	0.59
AM (left)	0.63
AM (right)	0.49

7.6 Conclusions

A segmentation algorithm, based on region growing, devised to be applied to functional data has been detailed in this chapter. The two main parts of this algorithm, the similarity criterion and the termination criteria, have been widely described and discussed.

It has been demonstrated how the similarity criterion is able to cope with high levels of statistical noise, and concomitantly low levels of CNR between adjacent tissues, due to its capability of dynamically adapting the region growing similarity criterion to the amount of noise of the section. It has been demonstrated how using a test image the region grower is able to segment robustly with levels of CNR down to ~6.

A combination of termination criteria has been used to be able to halt the region grower when a change of tissue texture or intensity, or a leaking situation are detected. Due to the high statistical noise and the high amount of partial volumes this task was specially difficult.

It can be observed from Tables 7.1 (^{125}I), 7.3 (^{3}H) and 7.5 (^{35}S), that show the CNR measured between some of the most significant functional structures in each experiment, how the autoradiogram labelled with ^{35}S shows the lowest values of CNR. This is related with the lower specificity of ^{35}S radioligands and with the longer β- range emitted by ^{35}S. This effect degrades the spatial resolution of the autoradiogram.

The algorithm has been validated with a set of test images based on a phantom. By modifying the statistical Gaussian noise the CNR was also modified in a controlled manner. The edges were also blurred by using a Gaussian mask to simulate the edge degradation resulting from the particle range. The result of the validation using the phantom is that this algorithm is able to segment reliably the different regions of the phantom with a maximum statistical noise $\sigma < 7$, corresponding to a CNR$\simeq$6.

To assess the region grower quantitatively, an expert has delineated manually the examples shown here, and these manually delineated autoradiograms have been compared with the resulting autoradiograms of the region grower. The Dice index between counterpart regions has been computed in all the overlapping regions. These Dice indices have shown high values for ^{3}H and ^{125}I, demonstrating high level of agreement between the manual and the automatic segmentation. The Dice indices obtained with ^{35}S were lower compared to the other two cases, representing a higher level of disagreement between both methods. It was observed in all the cases that an expert usually appears to under-segment the regions of interest to avoid incorrect measurements due to adjacent regions with lower or higher levels of ligand density. This effect does not happen with the region grower as it is able to detect blurry edges halting the region growing process before it leaks in to an adjacent region.

Chapter 8

Image Registration

8.1 Introduction

The final aim of autoradiography studies is to measure the absolute ligand concentration in internal brain structures. This is also known as Quantitative Autoradiography (QAR).

QAR involves ROI analysis of the autoradiogram using an implied correspondence with particular anatomical structures using a paper-based anatomical atlas [135]. Absolute image quantification is achieved by simultaneously exposing the autoradiography film to a set of calibrated microscales, and then expressing the digital values in each identified region in femtomoles per milligramme of ligand, using a video-based densitometry system. This analysis can then be used in life science applications such as genetic studies comparing knock out mice[1] with wild mice[2]. This labour intensive process is prone to systematic and random errors introduced by the user, because the delineation of ROIs of the key anatomical structures of the brain, typically of complex shapes, is undertaken be free-hand drawing. The aim of this study is to provide a method to automatically register an autoradiogram with an atlas, thus speeding up the labour-intensive process, avoiding quantisation labelling errors from manual segmentation and making the final quantitative results more accurate.

8.2 Review on Image registration

Image registration is a very active area of research in medical imaging, combining images from the same subject acquired with different imaging modalities, each modality showing information of a different nature such as functional and structural. The most active sub-area in image registration is based on the idea of aligning two different acquisitions from the same subject using different modalities, mainly combining functional and structural modalities [112, 124, 163, 166, 165, 181]. The aim of this is to be able

---

[1] A knock out mouse is a genetically modified mouse that has suffered the subtraction of a gen to see how this gen affects.

[2] A wild mouse is a mouse that has not suffered any genetic modification.

to obtain an accurate location in the anatomy of an active functional area. This application has broad application in diagnosis, pre-surgery planning or even during surgery.

Sometimes the images are acquired using the same modality from the same subject so the aim of image registration in this case is the correction of motion artifacts between the two images, because they were taken at different times (study of the evolution of a lesion), or because the patient just moved (correction of breathing and heart motion is a very active application) [21, 84, 169]. If these effects are not corrected they have a significant impact on the final resolution of the images, having the organ or tumour under study poorly delineated contours. In the case of Image Guided Radio Therapy (IMRT), where the tumour contours have to be as accurate as possible to reduce the dose delivered in healthy tissue, this is a critical problem [110, 146].

Some other times the images are obtained from different subjects with the same modality to create an atlas to be used as reference for future studies. There are many other combinations, but what all of them have in common is the necessity of aligning (registering) those images.

There are many different registration approaches and many different classifications such as [28, 174, 112, 68, 151]. The most extended classification criteria are presented here:

- **Dimensionality**. This criterion relies on the the dimensions (2D or 3D) of the images to register. It can be included here the registration of more than two images in time series. An example of 2D and 3D registration studies is shown in Figure 8.1.

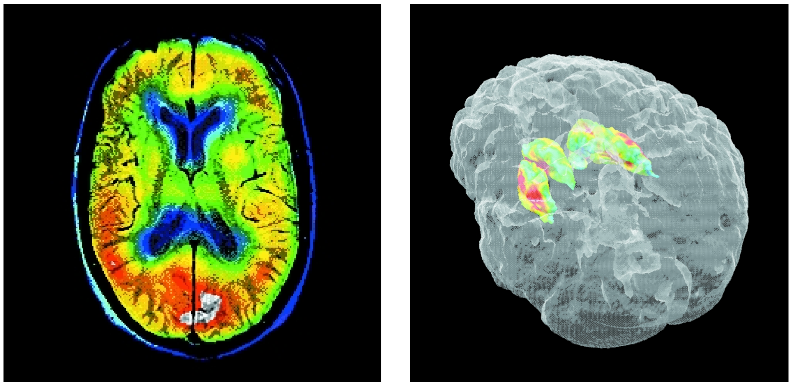
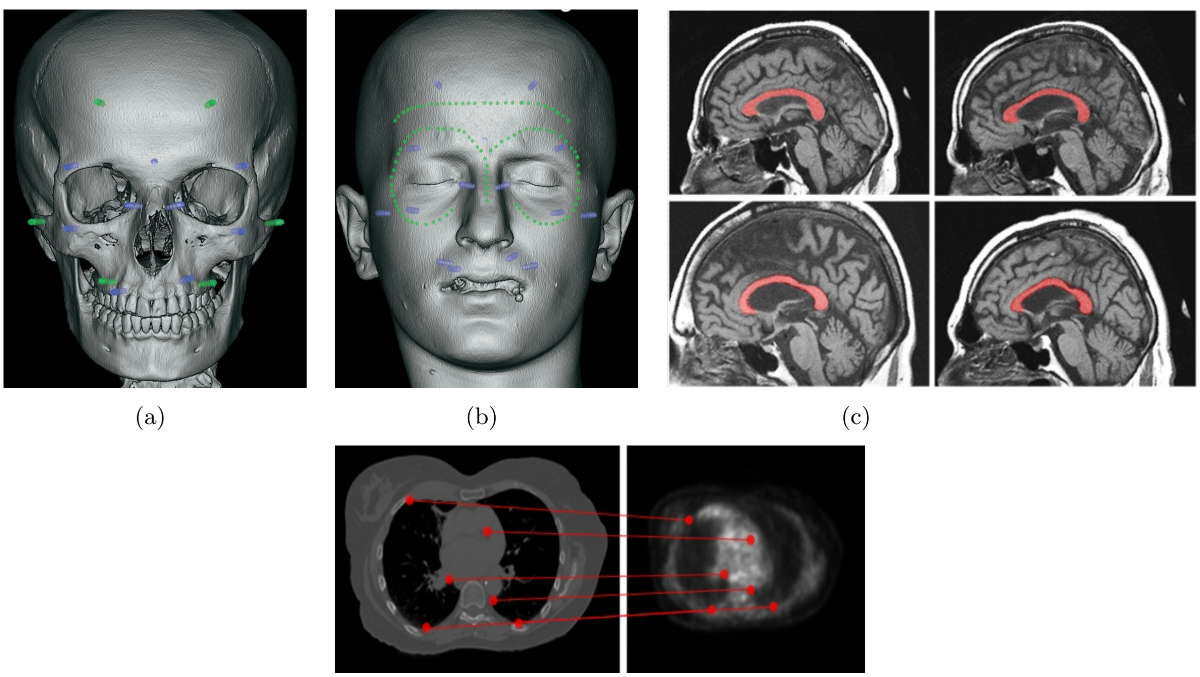

(a) (b)

Figure 8.1: Example of co-registered 2D SPECT-MRI images (a) and example of co-registered 3D image of the brain (MR) and basal ganglia (PET-^{18}F) (b) [85].

- **Nature of registration**. It can be extrinsic, with invasive (Figure 8.2(a)) or non-invasive markers (Figure 8.2(b)), or intrinsic. Intrinsic registration can be (1) based on anatomical or geometrical landmarks (Figure 8.2(d)), (2) based on segmentation of rigid or deformable models (Figure 8.2(c)) or (3) based on pixel/voxel properties.



(a) (b) (c)

(d)

Figure 8.2: Examples of different landmarks used in medical image registration [8]. Screws are used as invasive markers as shown in (a). Cutaneous landmarks can also be used as non-invasive markers (b). Anatomical structures, present in the source and target images, can be also used as landmarks, given a previous segmentation of such structures (c) as shown in this MR example. User defined or geometrical landmarks are also used (d) as in this CT-PET registration example.

- **Modalities involved**. Registration can be undertaken with images obtained from the same imaging modality (PET, SPECT, CT, MR, Autoradiography, Ultrasounds, Fluorescence, etc), i.e. monomodal, from different modalities, i.e. multimodal or intermodal, modality to model, or patient to model.

- **Subject**. The registration can be undertaken with images obtained from the same subject, i.e. intrasubject, from different subjects, i.e. intersubject or from patient to atlas.

- **Nature of transformation**. Here four different categories can be distinguished as shown in Figure 8.3. Rigid (the relative distances between points remains constant), affine (parallel lines remain parallel but the distance between points might change), projective (straight lines remain straight but the parallelism can be lost) and curved transformation (straight lines can be mapped onto curves). A lot of surveys classify these categories also as rigid and non-rigid techniques.

- **Parameter determination**. The transformation parameters can be obtained using a direct method or search-oriented. A direct method computes the aforementioned parameters based on corresponding points in the source and target

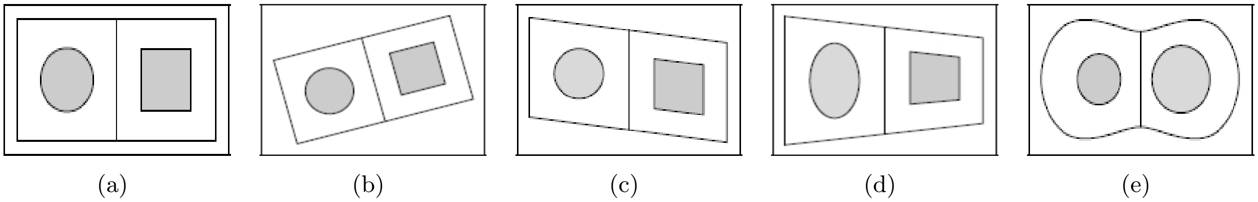


Figure 8.3: Examples of 2D transformations [112]. (a) Original image, (b) rigid transformation, (c) affine transformation, (d) projective transformation and (e) curved transformation.

> images, by interpolation or approximation, assuming that the problem is simple enough to allow a straightforward calculation. A search-oriented method starts from an initial guess, and based on a goodness-of-match or an energy function, refines the transformation parameters using an optimisation algorithm.

Usually the dimensionality, the modality and the subject are imposed by the application. The nature of registration and nature of transformation are the two main subjects to study to solve a specific problem. The formal mathematical problem consists of finding a transformation T to map one image (source image) to another (target image). This can be a transformation of the position x_A to x_B or it can consider also the intensity value of such position, i.e. $A(x_A)$ to $B(x_B)$ [68].

Historically image registration has often been considered as a rigid problem. Rigid transformations only allow rotations and translations. This method is very simple but it has very limited use. Usually this is only used for registration of intrasubject images with anatomical regions with few degrees of freedom such as bones or brain (usually considered as a rigid body because of the skull). Rigid registration allows for 3 degrees of freedom in 2D (2 for translation and 1 for rotation) or 6 degrees of freedom for 3D (3 for translation and 3 for rotation). Due to the lack of accuracy using only rigid registration, affine registration appeared to add scaling and shearing as transformations, so differences of scaling across scanners or large differences of scale between subjects could be corrected. For transformations involving 3D image data the number of degrees of freedom is therefore increased to 12 (3 for translation, 3 for rotation, 3 for scaling and 3 for shearing).

Biological tissues often have in general an important elastic component, therefore rigid or affine approaches may provide limited accuracy when aligned. The challenging work carried out at the moment by many groups involves non-rigid, also known as curved or elastic, registration techniques [41]. The main advantage of this technique is its capability of matching local characteristics while maintaining continuity and smoothness [68, 104].

Another component in a registration algorithm is the similarity measure used to assess how well two images match. These can be decomposed into geometric approaches and intensity approaches [41]. Intensity approaches match intensity patterns using mathematical or statistical criteria, so they are based on the existence of an intensity relationship between the two images. Typical measures are squared differences, correlation coefficient, measures based on optical flow and based on information theory, such as the well know mutual information or the intensity independent normalised mutual information. Mutual information was first defined by Shanon in the information theory field (Telecommunications), but it was demonstrated later by Wells and Viola [179] how mutual information has a lot of potential to be used in the medical imaging area.

Geometric approaches identify intrinsic or extrinsic landmarks in the two images to be matched. After the corresponding landmarks have been identified interpolation methods are used to warp the source image in a way that the landmarks match in the source and the target images consistently. Radial basis functions in general, and thin-plate splines and B-splines in particular, have to be mentioned due to their widespread use as interpolators in medical imaging. The main difference of these two families of splines is that thin-plate splines have infinite support (every basis function has global effect), and B-splines local support (every basis function has local effect). This is usually seen as an advantage in favour of the B-splines due to the better registration of local deformations. Cubic B-splines are usually the choice when this approach is used, due to low computational cost.

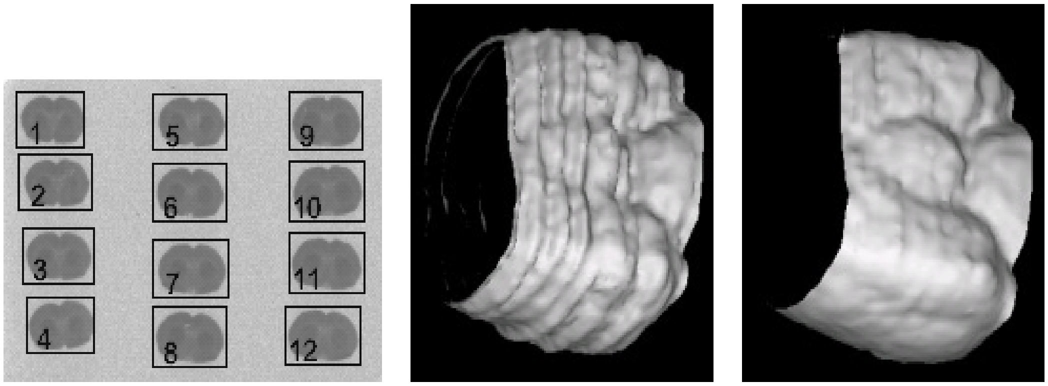
The needs of non-rigid registration in image-guided surgery is discussed in [66]. This discussion leads to a conclusion where, due to the computational cost of traditional elastic registration methods, the number of degrees of freedom has to be decreased. This is undertaken by using motion modelling, statistical shape modelling or biomechanical modelling. Motion modelling is based on the a-priori knowledge of the cyclic cardiac and respiratory motions. Using this knowledge the deformations in the organs caused by these sources of motion can be potentially corrected on real time [66].

Statistical shape modelling searches for a compact description of shape and its variations of organs across a population, using principal component analysis to study the most important shape variations [39].

Biomechanical modelling is based on the transformation of an underlying structure of the physical model. Elastic models treat the image as a linear elastic solid. This way of addressing the problem is only valid for small deformations. To face large deformations the elastic model is replaced by viscous fluid models. Physically-based numerical methods are based on the interpretation of the source image as an elastic solid or a viscous fluid where the deformations are controlled by applied forces [37]. Using this approach, the image registration is obtained as the solution of partial differential equations of continuum mechanics [60]. Due to the high level of elasticity provided by viscous fluids and the limited elasticity of biological tissues, these models do not always make biological sense, producing *unnatural-looking* results, particularly when large deformations are suffered by the object under study.

8.2.1 Image registration in Autoradiography

Most of the image registration applications with autoradiographic data presented in the literature are aimed at registering closely spaced two dimensional sections of the same subject to form a three dimensional model [186, 92]. This provides 3D information of activation areas in brain autoradiography studies. An example obtained from [92] is shown in Figure 8.4.



(a) (b) (c)

Figure 8.4: (a) Original autoradiograms used for 3D brain surface reconstruction. 3D reconstruction before registration (b) and after registration (c).

A first approach to register autoradiographic images is to compute the centre of mass and the minor and major axis. This approach based on the principal axes method was first used by [67]. An improvement over this method was developed by [186] where asymmetric, damaged and tilted autoradiographic sections are considered. This method has two problems: one is that the selection of the control points at the boundaries to be registered is manual; the second is that the alternative proposed for situations in which the shape is not correctly detected (damaged sections) is intensity-based, which makes it of limited application to intra-subject registration or between modalities with an intensity linear relationship. Rigid body transformation was applied in [92] using principal axes matching, and mutual information as similarity criterion, to create a three dimensional reconstruction of a rat brain from serial autoradiographic sections. To overcome the previous limitations, elastic registration was applied in [82] using mutual information as part of a global cost function, with thin-plate splines to register 2D autoradiographic images with video block face images[3]. In [113] a registration of autoradiographic images and fMR images is implemented to compare the different activation regions in symian brains. In this case a block matching algorithm (intensity based) is used obtaining a 3-D reconstruction of serial 2-D autoradiographic slices and a registration of this with an fMR brain volume in one single process.

All the published work so far to the best of the author's knowledge, using registration in autoradiography images has considered only intra-subject or inter-subject registration. Atlas registration is well documented in other modalities such as PET [4, 49] and MR [62, 45], but hitherto remains completely unexplored in autoradiography. Existing atlases used in other modalities are typically created registering images from different subjects, acquired with a given modality, sufficient to consider that a high percentage of the population is well represented. Then a statistical atlas is created by averaging the position of each organ or biological structure, thus intensity information is usually

---

[3] Reference video images

available. This fact offers the opportunity of using a wide variety of different registration approaches. This is not the case in autoradiography, where the aforementioned available atlas [135] lacks intensity information, and contains only boundary information, as observed in Figure 8.5. This limits the number of alternatives that can be used for the purpose of this application.

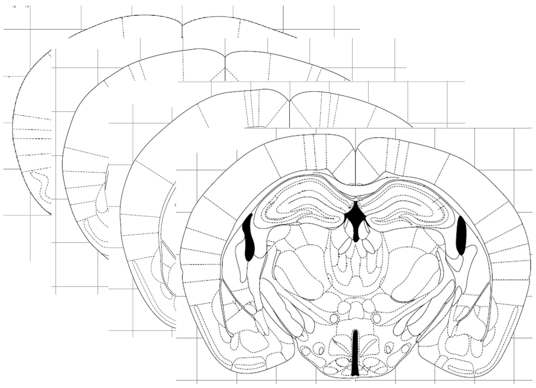


Figure 8.5: Stack with several exemplar atlas sections.

8.3 Pre-registration: Affine registration

As mentioned in Section 8.1, the aim of this work in this chapter is to provide a method to enhance throughput in autoradiography by developing an automated atlas-based registration. The first step in this registration process, as in most of the approaches found in the literature, is a global affine registration where global descriptors, such as centre of mass and principal axes, are aligned. A principal axes transformation [5], also considering scaling, is applied here. In this work five degrees of freedom are considered: rotation (1), translation (2) and scaling (2). The shearing effect is not considered here because visually it is seen to be negligible.

The centroid, the major and minor axes and the angle of rotation, defined as the angle between the major axis and the horizontal axis, of both the autoradiogram and the template are first obtained.

Then the transformation presented in matrix form, defined by equation 8.1, is applied to the atlas section, as this is the source image.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} s_x cos(\alpha) & -s_x sin(\alpha) \\ s_y sin(\alpha) & s_y cos(\alpha) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{8.1}$$

where x and y are the initial coordinates, s_x and s_y are the scaling factors in the x and y directions respectively, α is the rotation angle, t_x and t_y are the translational factors in the x and y directions respectively and x' and y' are the final coordinates. An example of this first step applied to an autoradiogram (target image) and its corresponding atlas image (source image) is shown in Figure 8.6.

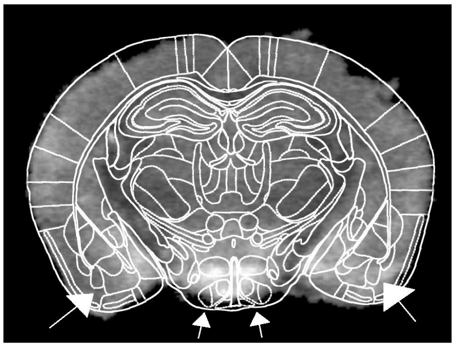


Figure 8.6: Affine registration of the autoradiogram and the slice of the atlas corresponding to that brain section (Bregma -1.70 mm). Some areas of poor correspondence are indicated by the arrows.

Non-linear deformations observed in the autoradiograms are produced during the cutting process of the tissue sections during sample preparation. This causes non-linear deformations and even loss of cortical regions. Moreover, as the source image is a synthetic atlas image, it is quite likely that key structures of the brain will fail to align with counterpart regions in the atlas following global rigid-body registration. For these reasons brain imagery from autoradiography has to be considered as soft (i.e. deformable) tissue. This is why affine registration has proved to yield limited precision.

8.4 Refined registration: Non-rigid registration

The second approach that has been considered to refine the registration is based in a special class of non-rigid elastic transformations, which allows the surfaces to be locally deformed, with smoothness constraints to maintain continuity.

The approach used in this work to obtain a similarity measure is the geometric approach, given that the atlas used (see Figure 8.5) does not provide any kind of intensity information thus ruling out an intensity based similarity measure. The intrinsic landmarks used here are comprised of a set of automatically selected geometrical landmarks and a set of anatomical regions present in the source and target images.

This problem has therefore been tackled by interpolation of landmarks. The whole registration problem thus can be divided in three main steps [150].

- Extraction of landmarks. The landmarks, in this work, are subdivided in two different sets. First geometrical landmarks are selected that represent strategically well suited control points in the external contour of the image. Secondly, a set of anatomical landmarks are those internal key structures that are identifiable in the autoradiogram as well as in the corresponding atlas image. These key structures are automatically segmented as those areas in the autoradiogram with the highest and lowest levels of uptake, corresponding with the brightest and darkest regions respectively (assuming that high intensity values in the image represent high level of radioactivity and vice versa). This information is taken from the previous segmentation produced in Chapter 7.

- Establishing correspondence between landmarks. The segmented landmarks correspond to certain structures in the brain architecture. This correspondence is undertaken manually by the user, but could be undertaken automatically looking for combinations of local structures in the area of the segmented region. This assumes that the initial affine registration brought the segmented region and its combination of structures in the atlas into some level of close proximity.
- Computing the transformation. Once the correspondence between the two datasets has been found a transformation is applied to the atlas to warp the atlas onto the autoradiogram.

8.4.1 Selection of geometrical landmarks

The contours of the autoradiogram and atlas present typically significant differences, the atlas having a smooth boundary and the autoradiogram boundary being noisy and highly tortuous. This difference is partly caused by the stresses induced by slicing the sections.

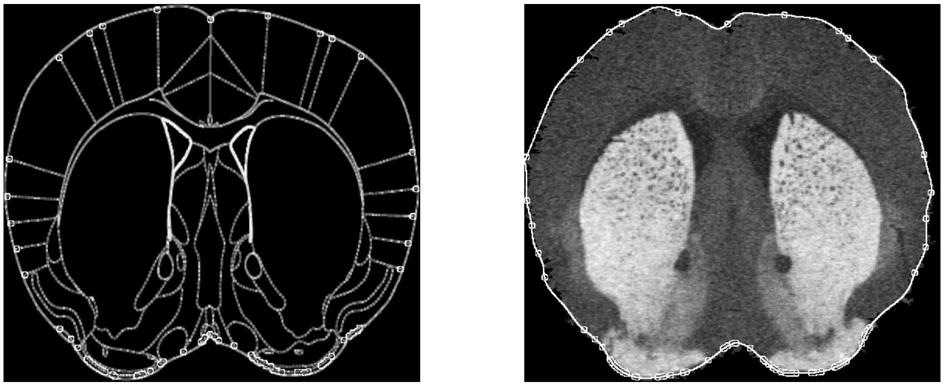
Contour smoothing: To address this problem, before selecting the geometrical landmarks, the segmented autoradiogram boundary is smoothed using cubic splines [103]. For this purpose M equidistant points are automatically selected from the contour to interpolate with the cubic splines. Depending on the distance between points, the boundary will be more or less smoothed, obtaining a smoother contour when further control points are selected, as can be observed in Figure 8.7. Here a smoother boundary is observed in Figure 8.7(c) with a contour distance (length along the contour between two consecutive points) between control points of 80 pixels, compared to that observed in Figure 8.7(b) with a contour distance between control points of 40 pixels.

Location of geometrical landmarks: After smoothing the contour of the autoradiogram the geometrical landmarks are automatically selected in the atlas image. These control points correspond to the junctions where the outer boundary of the atlas has physical contact with the delimiters of the regions in the cortex, as shown in Figure 8.8(a) with circles in the outer boundary. These control points are selected based on the assumption that after applying the elastic registration, the continuity of these contours of the structures in the cortex with the outer boundary will be preserved. Once these N control points on the atlas contour have been automatically located, their corresponding control points in the autoradiogram are automatically selected. These are selected as the physically closest points from the autoradiogram contour, as shown in Figure 8.8(b) with circles in the autoradiogram contour, to the previously located control points in the atlas contour. These two sets of control points will help to match

(a) Original boundary (b) Smoothed boundary using M=40 (c) Smoothed boundary using M=80

Figure 8.7: Original boundary (a), smoothed boundary using a distance of 40 pixels between points (b) and smoothed boundary using a distance of 80 pixels between points (c).

the internal structures of the atlas onto the autoradiogram considering the deformed external contour.



(a) (b)

Figure 8.8: Atlas (a) and autoradiogram (b) with the points selected to be used as geometrical landmarks marked with circles.

The result of registering both contours, using the geometrical landmarks shown in Figures 8.8(a) and 8.8(b), using linear radial functions (detailed in Section 8.4.5), is shown in Figure 8.9(b). For comparison purposes the atlas and autoradiogram registered using only an affine registration is shown in Figure 8.9(a). Comparing Figures 8.9(a) and 8.9(b) it can be observed how the atlas contour fits more accurately with the autoradiogram contour after the elastic registration of the atlas contour.

8.4.2 Selection of biological landmarks

The biological landmarks used in this work are derived from using the previous scene-based segmentation of key functional structures of the autoradiograms. The segmen-

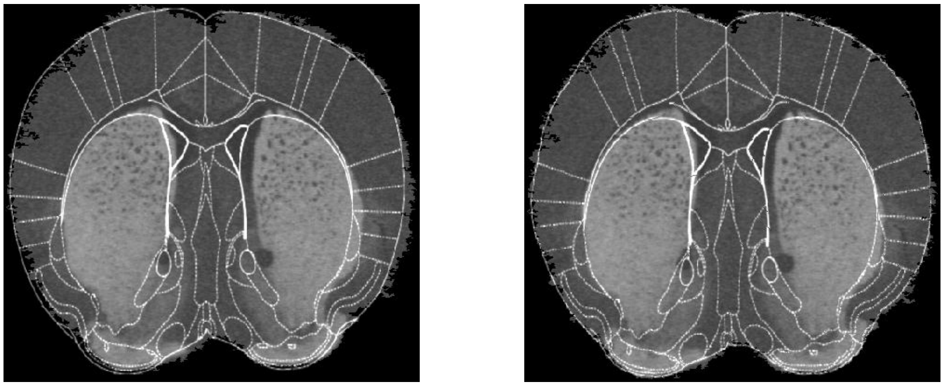

(a) (b)

Figure 8.9: Atlas and autoradiogram registered after affine registration (a) and after elastic deformation of the contour (b). Note how only the atlas contour fits more accurately on the autoradiogram contour, but significant registration errors are still observed in the internal structures.

tation process has been thoroughly described in Chapter 7.

As a first approach, the landmarks used for the registration of the internal structures of the autoradiogram were the centres of mass of each segmented region. Therefore the interpolation was undertaken using n landmarks, n being the total number of segmented regions. This approach resulted in a better local alignment of the internal structures of the autoradiogram with the atlas compared to just an affine registration. Nevertheless, due to the high elasticity of the brain tissue sections and the concomitant significant distortion produced by the cutting process, the boundary of the segmented internal regions used as landmarks did not perfectly match the boundary of the corresponding regions in the atlas.

To tackle this problem a second approach was considered. This second approach uses a set of m control points in the boundary of each segmented internal region, and the corresponding m control points in the atlas, as sets of landmarks to obtain a more accurate registration.

Once a set of regions A (equation 8.2) has been automatically segmented in the autoradiogram, and the corresponding regions have been manually identified in the atlas counterpart B (equation 8.3), the contour of each region is extracted.

$$A = \{A_1, A_2, \dots A_n\} \qquad A_i = \{a_{i1}, a_{i2}, a_{i3}, \dots a_{im}\} \tag{8.2}$$

$$B = \{B_1, B_2, \dots B_n\} \qquad B_i = \{b_{i1}, b_{i2}, b_{i3}, \dots b_{im}\} \tag{8.3}$$

A set of control nodes a_i is extracted from each contour in each segmented region of the autoradiogram A_i (equation 8.2), i being the region being processed and m the number of control nodes in the contour. An identical procedure is followed for the corresponding region of the atlas B_i (equation 8.3), producing a second set of control points b_i. The number of control nodes selected for each contour m, identical for both sets A_i and B_i, is proportional to the length of the contour. The heuristic criterion is to set a point from the contour as control point every 20 pixels along the contour. If the total number of control points extracted from a certain contour is less than 6 control points because it is too short, then the distance between control points is reduced to 10 pixels instead of 20 pixels.

To obtain a better description of the contour shape, some points where maximum curvature of the contour is measured are also used as control points. The curvature is measured as shown in equation 8.4:

$$c = \sqrt{\frac{t_x^2 \;\; t_y^2}{s}} \tag{8.4}$$

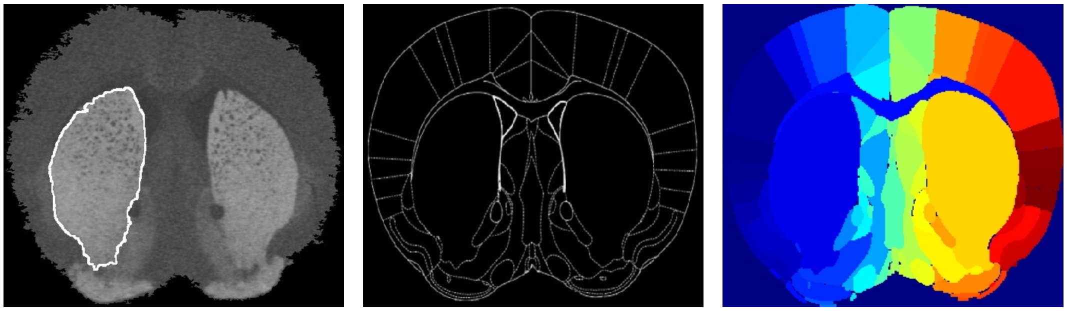
where t_x and t_y are defined by equation 8.5, x' and y' are the gradient of the x coordinates and the y coordinates respectively, and parameter s is defined by equation 8.6:

$$t_x = \frac{x'}{s} \qquad t_y = \frac{y'}{s} \tag{8.5}$$

$$s = \sqrt{x'^2 + y'^2} \tag{8.6}$$

The curvature of a point in a curve is usually measured as how its coordinates change related to its two adjacent points. Contour shapes in this work do not present acute angles. Therefore the way of measuring the curvature has been modified to obtain a more accurate detection of points with high curvature in the contour shape. Instead of considering the two adjacent points of each point, the two equidistant points at five pixels of distance by each side has been used. Using this method the 20 pixels with largest curvature of the contour are extracted and included in the set of control points A_i.

Figure 8.10 shows an example of an original autoradiogram with one of the segmented regions highlighted in Figure 8.10(a), the wireframe image of the corresponding bregma in Figure 8.10(b) and the labelled wireframe atlas image in Figure 8.10(c).



(a) (b) (c)

Figure 8.10: Original autoradiogram with one region under study delineated (a), wireframe image of corresponding bregma -1.82 mm (b) and labelled wireframe image (c).

After extracting A_i for a specific autoradiogram region, the way the control nodes are extracted from the contour in the corresponding atlas region, defined by B_i, is very important, as it has to keep maximum proximity between pairs of control points (a_{i1} with b_{i1}, a_{i2} with b_{i2} and so on), it has to maintain the one-to-one property of the warping transformation and at the same time it has to describe the contour as accurate as possible. If after the initial affine registration there still exists a significant missalignment between the internal structures in the autoradiogram under study and the corresponding regions in the atlas, the control nodes will not properly describe the contour of the atlas region. Besides there will be situations where the distance between consecutive control points along the contour is too large to describe properly the atlas contour.

Another problem that may occur is too close control points corresponding to two different structures. Given two different structures in the autoradiogram A_1 and A_2, and their corresponding set of control points, $\{a_{11}, a_{12}, a_{13} \ldots, a_{1m}\}$ and $\{a_{21}, a_{22}, a_{23} \ldots, a_{2m}\}$ respectively, if a given a_{1t} (belonging to A_1) is too close to another a_{2t} (belonging to A_2) this situation could result in an unexpected behaviour in the data interpolator.

To solve these two problems each set of control points is post-processed to obtain an accurate representation of the contour of each region while avoiding over-crowding of control points along the contour. This is achieved by limiting the maximum contour distance between adjacent control points along the region contour a_{it} and a_{it+1}, and also by limiting the minimum straight distance between control points from adjacent functional structures A_1 and A_2.

1. The first condition establishes that if between two consecutive control points along the contour the distance exceeds the maximum distance allowed, h consecutive control points in each side of the *gap* will be displaced in order to reduce that *gap*. In other words, the contour will be *stretched* to better describe the original contour of the region under study. The maximum distance along the contour allowed between consecutive control points is the mean contour distance between consecutive control points in the entire contour plus one standard deviation. The number of control points displaced when this situation is given (h), has been chosen arbitrarily as three for each side (total of six control points) for all the experiments shown in this work.

2. The second condition establishes that if the straight distance between two control points, a_{1t} and a_{2t}, belonging to two different functional regions, A_1 and A_2, is too close, then these control points are ignored by the subsequent interpolation process. This minimum distance has been set to five pixels for all the experiments shown in this work.

An example of the first problem described above using real data is shown in Figure 8.11. Figures 8.11(a) and 8.11(b) show the contour of the segmented region, corresponding to the Corpus Callosum of the mouse brain, and the corresponding set of 128 equidistant control points respectively. Figure 8.11(c) represents the contour of the corresponding region in the atlas. The resulting control points from the atlas contour, extracted with maximum spatial proximity to the control nodes of the autoradiogram (Figure 8.11(b)), are shown in Figure 8.11(d). Large *gaps* of distance between consecutive control points are indicated with red circles. The result of post-processing the set of control points from the atlas contour, reducing significant *gaps* between adjacent control points, is indicated with arrows in the resulting Figure 8.11(e), where these *gaps* are reduced by stretching the adjacent six control points per *gap* (3 control points per side).

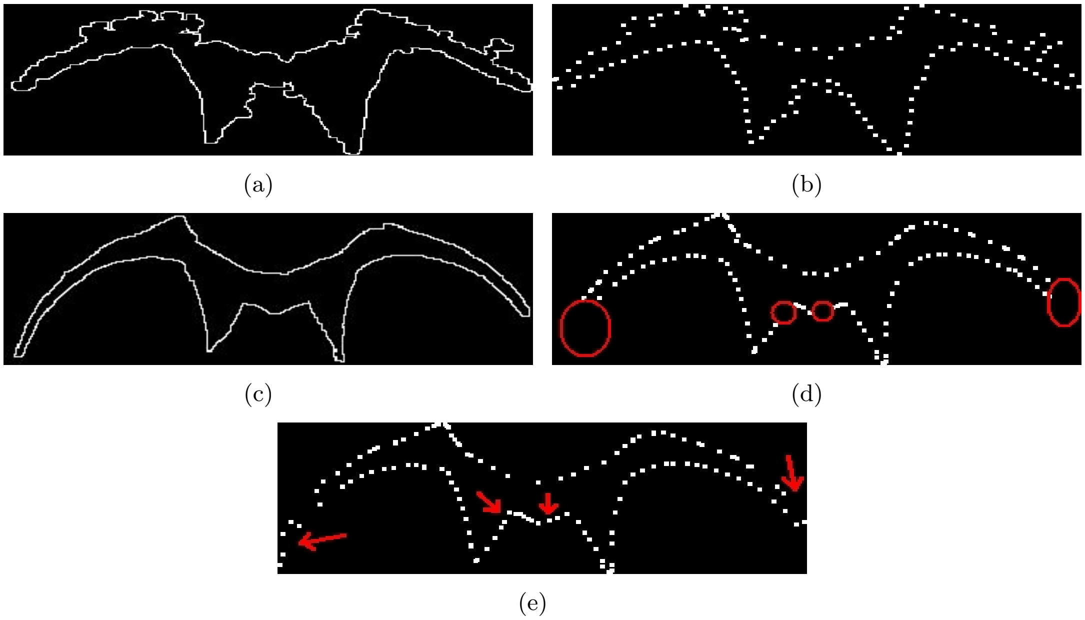

(a) (b) (c) (d) (e)

Figure 8.11: Contour of segmented region corresponding to the Corpus Callosum of the mouse brain (a), corresponding set of 128 equidistant control points (b), corresponding atlas region (c) and corresponding set of 128 control points obtained from the atlas, showing large *gaps* between consecutive control points indicated with red circles (d). Figure (e) shows the final set of control points from the atlas after correcting significant *gaps* between consecutive control points indicated with arrows (e).

This process is repeated for every segmented region in the autoradiogram. The control nodes of each region are accumulated to be used in the final interpolation of all the control nodes as described in the Section 8.4.3.

8.4.3 Scattered Data Interpolators

The problem of registration is viewed as a problem of interpolation in this work. This can be used to predict values at locations where there is no recorded observation on a surface or to deform surfaces. The latter is the case used in this study.

Having a set of control points in an image defined by $\mathbf{x}$, with coordinates in 2D (x, y), and their associated desired final coordinates (x', y'), it is necessary to find a transformation function $f(\mathbf{x})$ which translates all the points, that are not control points, to a new position $\mathbf{x'}$ (x', y'). These functions, better known as interpolators, have to fulfil some constraints to obtain satisfactory results. Among the most important constrains are the fact that the control points have to be exactly moved to their final coordinates, in other case they would be called approximators instead of interpolators, and the final result must be smooth providing continuity in the surfaces.

In this work three different approaches are considered: The Inverse Distance Weighted algorithm (IDW), Radial Basis Functions based algorithms (RBFs) and B-Splines functions.

8.4.4 Inverse Distance Weighted

The Inverse Distance Weighted method is based on the assumption that the interpolating curve is more influenced by nearby control points than by more distant points. The predicted value of a point is computed from the value of its neighbours. The classical formula of the IDW interpolator, sometimes called Shephard's method, is shown in equation 8.7 [156].

$$f(x) = \sum_{i=1}^{n} w_i(x) x_i \tag{8.7}$$

where n is the number of scattered points available to calculate the deformation field to apply to the source image, $w_i(x)$ is the weight dependent on the distance between the interpolation point to calculate x and the scatter point x_i as shown in equation 8.8. The classical form of the weight function is shown in equation 8.8, where $\sigma_i(x)$ is defined by equation 8.9.

$$w_i(x) = \frac{\sigma_i(x)}{\sum_{j=1}^{n} \sigma_j(x)} \tag{8.8}$$

$$\sigma_i(x) = \frac{1}{(d_i(x))^{\mu}} \tag{8.9}$$

being $d_i(x)$ the euclidean distance between the interpolation point to estimate x and the scatter point x_i. The parameter μ is the significance of the surrounding points around the interpolated value and controls the smoothness. The parameter $\mu > 1$ assures the continuity of the first derivative (which keeps the continuity of the transformation function $f(\mathbf{x})$), a value of 1.5 is visually valid for us. In this special case the differences between boundaries have to be considered.

Applying such deformations in the inner regions of the atlas, a combination of forces induced by the aforementioned landmarks and the points from the contour, produces a better match to the external contour and the internal structures. Ensuring this correspondence occurs robustly is a key goal of this work to ensure that reliable analysis can be undertaken using subsequent anatomically defined ROIs.

8.4.5 Radial Basis Functions interpolators

Another approach to scattered data interpolation is to create the interpolator, having n sites with initial coordinates $\{t_n\} : t_i = (x_i, y_i)$ and final coordinates $\{f_n\} : f_i = (x'_i, y'_i)$

for two dimensions, as a linear combination of radial functions *f(x)*. Then the resulting interpolator function is described by equation 8.10 [104].

$$f(x) = p_m(x) + \sum_{i=1}^{n} \lambda_i \phi(||x - x_i||) \tag{8.10}$$

where p_m is a low degree polynomial, λ_i is a set of coefficients to be computed, $|| \cdot ||$ denotes the Euclidean norm and $\phi(x)$ is the radial function (called radial because the argument is the distance between two points x and x_i). Some popular choices of this function are given below:

$$\left.\begin{array}{ll}
\phi(x) = x & (Linear) \\
\phi(x) = x^2 log(x) & (Thin-plate\ spline) \\
\phi(x) = e^{-x^2/2\sigma^2} & (Gaussian) \\
\phi(x) = (x^2 + c^2)^{1/2} & (Multiquadratic) \\
\phi(x) = I^k(1-x)^{\lfloor d/2 \rfloor + k + 1}(x) & (Wendland)
\end{array}\right\} \tag{8.11}$$

where σ is the standard deviation of the Gaussian function, c in the multiquadratic function is a positive constant and the Wendland function is described below.

The choice of RBF determines the characteristics of the transformation $f(x)$. The main properties of the RBFs are locality, solvability and efficiency [55]. Locality defines if the function has local or global support, this is if the transformation for a certain point only affects those surrounding points or all the points in the image respectively. Solvability defines if the linear equation system that needs to be solved (see below) can be solved, i.e. the matrices involved in the matrix system that has to be solved are non-singular (the inverse can be computed). Efficiency depends on the dimension (2D or 3D) and the size of the data set that has to be handled. It is also an important feature if the matrix system that has to be solved is comprised of sparse or dense matrices.

The first four RBFs shown in equation 8.11 have infinite support, i.e. the addition or modification of one single pair of control points will affect the entire transformation function $f(x)$. As opposed to functions with infinite support, the Wendland function has local support, as it is defined only inside a user defined radius a, as shown in equation 8.12:

$$\phi(x) = \begin{cases} p(x) & \text{0≤x≤a;} \\ 0 & \text{x>a} \end{cases} \tag{8.12}$$

where $p(x)$, also written as $p_{d,k}(x)$, is a univariate function that depends on the space dimension d (2 for 2D and 3 for 3D) and on a smoothness factor k. The form of $p_{d,k}(x)$ is defined in equation 8.13:

$$p_{d,k}(x) = I^k(1-x)_+^{\lfloor d/2 \rfloor + k + 1}(x) \tag{8.13}$$

where I^k is the integral operator applied k times and the subscript + means that the function is only defined inside the circle of radius a as mentioned previously, as shown in equation 8.14:

$$(1-x)_+^{\nu} = \begin{cases} (1-x)^{\nu} & 0 \leq \mathrm{x} \leq \mathrm{a}; \\ 0 & \mathrm{x} > \mathrm{a} \end{cases} \tag{8.14}$$

The function $p_{d,k}(x)$ for different values of smoothness k is shown in Table 8.1 [180]:

Table 8.1: Explicit formula of Wendland's functions

$p_{d,0}(x) = (1-x)_+^d$	k=0
$p_{d,1}(x) = (1-x)_+^{(d+1)}[(d+1)x+1]$	k=1
$p_{d,2}(x) = (1-x)_+^{(d+2)}[(d^2+4d+3)x^2+(3d+6)x+3]$	k=2

The Wendland function is not differentiable at $x = 0$ for $k = 0$, therefore it is not smooth. To keep the function $p_{d,k}(x)$ with small degree, not to deprecate the efficiency of the algorithm, $k = 1$ is chosen as smoothness degree. The specific case of this work is 2D, therefore $d = 2$. The function $p_{d,k}(x)$ used in this work is $p_{2,1}(x) = (1-x)^3(3x+1)$, therefore $\phi(x)$ is defined as shown in equation 8.15:

$$\phi(x) = \begin{cases} (1-x)^3(3x+1) & 0 \leq \mathrm{x} \leq \mathrm{a}; \\ 0 & \mathrm{x} > \mathrm{a} \end{cases} \tag{8.15}$$

From the general equation for RBFs (equation 8.10), a general solution for the low degree polynomial, p_m, is a linear combination of monomials defined by equation 8.16 [104].

$$p_m(x) = \sum_{i=1}^{M} a_i g_i(x) \tag{8.16}$$

where M is the dimension of the space where the monomials are defined [150], defined by equation 8.17, and $g_i(x)$ is described by equation 8.18 [24].

$$M = \frac{(d+m-1)!}{(d!(m-1)!)} \tag{8.17}$$

where m is the order of the derivative function to be minimised to obtain the resulting interpolator funtion 8.10 [150]. In this work m has been set to 2 for conveniency.

$$g_1(x) = 1 \; g_2(x) = x \; g_3(x) = y \tag{8.18}$$

being x and y the initial coordinates of the n known sites. For 2D images $M=3$. This imposes the minimum number of landmarks to use, hence at least 3 non colinear landmarks have to be supplied.

The coefficients λ_i from equation 8.10 are obtained from the constraint that the values of the function at the control points, *f(x$_i$)*, are known so n linear equations can be used as described by the linear system in equation 8.19.

$$\begin{pmatrix} A & Q \\ Q' & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ a \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix} \tag{8.19}$$

where A is a matrix comprised of the radial basis function between each couple of control points as shown in equation 8.20:

$$A = \begin{pmatrix} 0 & \phi(||x_1 - x_2||) & \phi(||x_1 - x_3||) & \cdots & \phi(||x_1 - x_n||) \\ \phi(||x_2 - x_1||) & 0 & \phi(||x_2 - x_3||) & \cdots & \phi(||x_2 - x_n||) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi(||x_n - x_1||) & \phi(||x_n - x_2||) & \phi(||x_n - x_3||) & \cdots & 0 \end{pmatrix} \tag{8.20}$$

Q is a matrix with the initial coordinates, λ and a are the coefficients of the radial functions and the monomials respectively and f is a matrix with the final coordinates. These are defined by 8.21, 8.22 and 8.23 respectively.

$$Q = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{pmatrix} \tag{8.21}$$

$$\lambda = \begin{pmatrix} \lambda_1, & \lambda_2, & \cdots, & \lambda_n \end{pmatrix}^T \tag{8.22}$$

$$a = \begin{pmatrix} a_1, & a_2, & \cdots, & a_M \end{pmatrix}^T \tag{8.23}$$

The dimensions of the system are as shown in 8.24:

$$\begin{pmatrix} nxn & nxM \\ Mxn & MxM \end{pmatrix} \begin{pmatrix} nx1 \\ Mx1 \end{pmatrix} = \begin{pmatrix} nxM - 1 \\ MxM - 1 \end{pmatrix} \tag{8.24}$$

The final linear system of equations is fully described for a 2D case (M=3) as follows:

$$\begin{pmatrix} 0 & \phi(||x_1 - x_2||) & \phi(||x_1 - x_3||) & \cdots & \phi(||x_1 - x_n||) & 1 & x_1 & y_1 \\ \phi(||x_2 - x_1||) & 0 & \phi(||x_2 - x_3||) & \cdots & \phi(||x_2 - x_n||) & 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \phi(||x_n - x_1||) & \phi(||x_n - x_2||) & \phi(||x_n - x_3||) & \cdots & 0 & 1 & x_n & y_n \\ 1 & 1 & 1 & \ldots & 1 & 0 & 0 & 0 \\ x_1 & x_2 & x_3 & \ldots & x_n & 0 & 0 & 0 \\ y_1 & y_2 & y_3 & \ldots & y_n & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} f(x_1, y_1) \\ f(x_2, y_2) \\ \vdots \\ f(x_n, y_n) \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{8.25}$$

After obtaining the λ and a coefficients, by inverting the matrix on the left hand side in equation 8.19, all the points of the image are evaluated with the obtained *f(x)* function obtaining the transformed image. To solve the typical system of linear equations in *Matlab*, defined in matrix form by $Ax = b$, is typically using the command $x = A \setminus b$. This approach to invert the system and obtain x employs LU factorisation, also known as Gaussian elimination.

8.4.6 B-Splines interpolators

B-Splines approximation techniques are a very popular approach due to its local support and the high efficiency of these functions. The approach in this work is to use B-Splines functions as interpolators. The most popular B-Spline is the cubic B-Spline because it gives a well accepted level of smoothness keeping the function simple, to obtain high computational efficiency.

To clarify the terminology, for the different interpolators detailed above, control points refer to the initial scattered data set of points with known initial and final position. B-Splines theory refers as control points to those points that comprise an overlying uniform lattice $\boldsymbol{\phi}$. Those points with known initial and final positions are simply defined as initial data set of scattered data points $\mathbf{x}$.

The basic idea is, given a set of scattered data points $\mathbf{x}$, to find a transformation function $f(\mathbf{x})$, as mentioned above. While for the previous interpolators the interpolating function is continuous, i.e. it is defined in every point of the space, the main difference with B-Splines is that the transformation function is only defined in some specific control points that cover the entire space forming an overlying lattice defined by $\boldsymbol{\phi}$ [93].

Let ϕ_{ij} be the value of the ijth control point on the lattice, located at (i, j). The interpolation function in an arbitrary point (x_w, y_z), at euclidean coordinates (w, z), is defined by equation 8.26:

$$f(x_w, y_z) = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(s) B_l(t) \phi_{(i+k)(j+l)} \tag{8.26}$$

where $i = \lfloor x \rfloor - 1$, $j = \lfloor y \rfloor - 1$, $s = x - \lfloor x \rfloor$, $t = y - \lfloor y \rfloor$, B_k and B_l are uniform cubic B-Splines, defined by equation 8.27, and shown in Figure 8.12, $\phi_{(i+k)(j+l)}$ are the 16 control points surrounding (x_w, y_z) as shown in Figure 8.13 and $\lfloor\rfloor$ represents the lower integer after rounding. This represents the locality of this B-Spline, and can be extended by increasing the degree of the B-Spline, i.e. considering more control points around (x_w, y_z).

$$\left.\begin{array}{ll} B_0(t) = (1-t)^3/6 & -2 \le t < -1 \\ B_1(t) = (3t^3 - 6t^2 + 4)/6 & -1 \le t \le 0 \\ B_2(t) = (-3t^3 + 3t^2 + 3t + 1)/6 & 0 \le t \le 1 \\ B_3(t) = t^3/6 & 1 < t \le 2 \end{array}\right\} \tag{8.27}$$

B_k and B_l weight the contribution of each control point to $f(x,y)$ depending on the distance t.

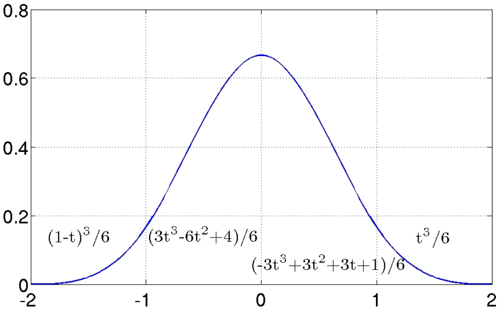


Figure 8.12: Cubic B-Spline defined by equation 8.27.

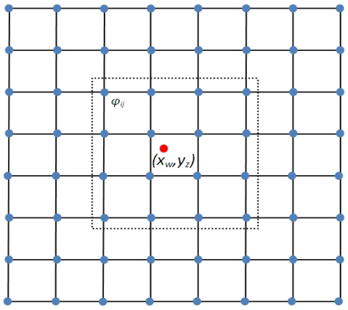


Figure 8.13: Lattice of control points $\boldsymbol{\phi}$. A scattered point (x_w, y_z) is shown in red with its corresponding 16 neighbouring control points surrounded by a dashed red square.

With this formulation the problem is reduced to solve the value of the control points that comprise the lattice $\boldsymbol{\phi}$. To do this it is known that a landmark from the original frame (x_w, y_z), belonging to the initial set of scattered landmarks $\mathbf{x}$, has to be transformed to the known new location (x'_w, y'_z) using the B-Spline interpolator shown in equations 8.28 and 8.29 for x_w and y_z respectively.

$$x'_w = \sum_{k=0}^{3}\sum_{l=0}^{3} B_{kl}\phi_{kl} \tag{8.28}$$

$$y'_z = \sum_{k=0}^{3}\sum_{l=0}^{3} B_{kl}\phi_{kl} \tag{8.29}$$

where B_{kl} is $(B_k(s)B_l(t))$ the B-Spline function corresponding to the local control point at location (k,l) on the lattice, out of the 16 neighbour control points of a specific landmark (x_w, y_z). The problem of solving the value for the control points on the lattice $\boldsymbol{\phi}$ is undertaken for the x and y coordinates independently, therefore there are two lattices implemented, ϕ_x and ϕ_y. The values for ϕ_{kl_x} and ϕ_{kl_y} are obtained by minimising $\sum_{k=0}^{3}\sum_{l=0}^{3}\phi_{kl}^2$ in equations 8.28 and 8.29 respectively, resulting in equations 8.30 and 8.31 respectively [93]:

$$\phi_{kl_x} = \frac{B_{kl}x'_w}{\sum_{a=0}^{3}\sum_{b=0}^{3}B_{ab}^2} \tag{8.30}$$

$$\phi_{kl_y} = \frac{B_{kl}y'_z}{\sum_{a=0}^{3}\sum_{b=0}^{3}B_{ab}^2} \tag{8.31}$$

B_{kl} will be different for each coordinate, x and y, because it depends on the distance t between the point (x_w, y_z) and the klth control point on the lattice in each direction independently.

If the points belonging to the initial set of scattered points $\mathbf{x}$ are too close it is very likely that there will be cases where one single control point on the lattice ϕ_{kl} is affected by more than one scattered point in $\mathbf{x}$. In other words, their neighbourhoods comprised of 16 control points are overlapped, as shown in Figure 8.14.

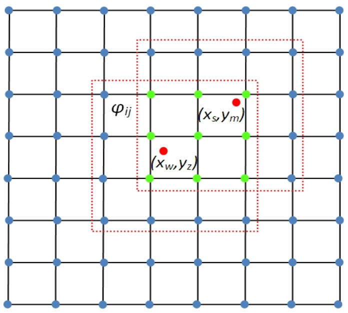


Figure 8.14: Lattice of control points $\boldsymbol{\phi}$ with two scattered points (x_w, y_z) and (x_s, y_m) showing overlapped neighbouring regions. The overlapping control points on the lattice are shown in green.

This situation leads to an error in ϕ_{ij}. This error can be defined as the difference between the final value, after considering the contribution of all the surrounding scattered data points (ϕ_{ij}), and the expected value if only one scattered data point is considered (ϕ_c), shown in equation 8.32:

$$e(\phi_{ij}) = \sum_c (\omega_c \phi_{ij} - \omega_c \phi_c)^2 \tag{8.32}$$

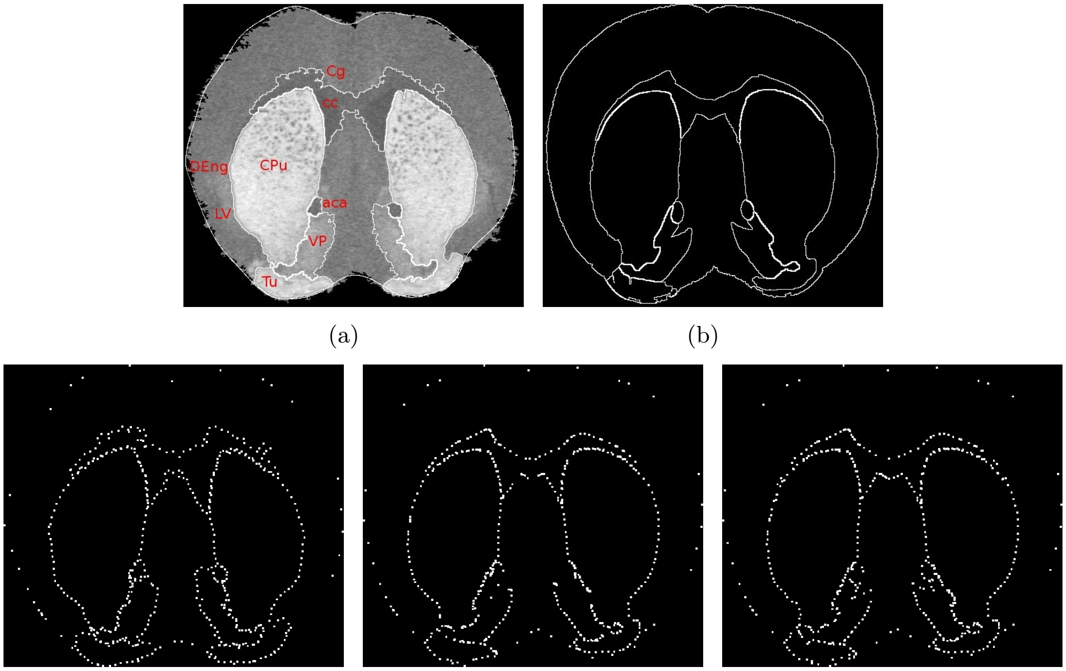
To minimise this error $e(\phi_{ij})$ equation 8.32 is differentiated with respect to ϕ_{ij}, resulting in equation 8.33.

$$\phi_{ij} = \frac{\sum_c \omega_c^2 \phi_c}{\sum_c \omega_c^2} \tag{8.33}$$

After computing the values of all the control points from the entire lattice $\boldsymbol{\phi}$, equation 8.26 is applied to all the scattered data points $\mathbf{x}$. Due to the aforementioned error $e(\phi_{ij})$ the final positions of the initial scattered data points will not coincide with the known final positions. Therefore, in this specific situation where scattered data points are too close and surrounding areas of control points overlap, B-Splines work as an approximator instead of an interpolator. This error, and therefore reduced registration accuracy, increases when increasing the distance between control points in the lattice, as it is more likely that areas of control points overlap.

8.5 Results

To be able to observe the differences between different interpolators a mouse brain section with relatively simple internal structure has been chosen. This is a mouse brain section labelled with ^{3}[H]SCH-23390 cut at the level of the Striatum that shows high levels of ligand uptake in the Caudate Putamen (CPu) and Olfactory Tubercle (Tu), medium levels of ligand uptake in the Ventral Pallidum (VP), in the Dorsal Endopiriform (DEng) and in the Cingulate Cortex (Cg) and low levels in the Corpus Callosum (cc), the Anterior Commissure (aca) and Lateral Ventricle (LV) (see Figure 8.15(a)).



(a) (b)

(c) (d) (e)

Figure 8.15: Original autoradiogram of mouse brain section with regions with high, medium and low level of radioligand uptake delineated after segmentation (a), corresponding atlas regions (b), control points of the segmented autoradiogram regions (c), initial control points of the corresponding atlas regions (d) and post-processed control points of the atlas regions.

The regions that have been successfully segmented are the CPu, Tu, VP aca, and cc as shown in Figure 8.15(a). The corresponding regions in the atlas are indicated in Figure 8.15(b). Figure 8.15(c) represents the control points extracted from all the autoradiogram contours from the regions indicated in Figure 8.15(a). Similarly the initial unprocessed set control points of the corresponding atlas regions is shown in Figure 8.15(d), and in Figure 8.15(e) after processing the set of control points.

The results of registering the internal regions shown in Figure 8.15(a) with their corresponding regions in the atlas shown in Figure 8.15(b) with the different scattered data interpolators are shown below. From these results only visual qualitative assessment can be taken. By showing these results on their own there is no way to compare qualitatively these scattered data interpolators. For this purpose a leave-one-out evaluation is presented in Section 8.6.

8.5.1 Results with the Inverse Distance Weighted interpolator

The result of using a scattered data interpolator based on Inverse Distance Weighted interpolator is shown in Figure 8.16. Figure 8.16(a) shows the autoradiogram with the corresponding warped atlas section superimposed on top after the elastic transformation function applied to the atlas, Figure 8.16(b) shows a control grid that shows the warping function used in this example and Figure 8.16(c) shows the displacement vectors over the surface after applying the transformation function.

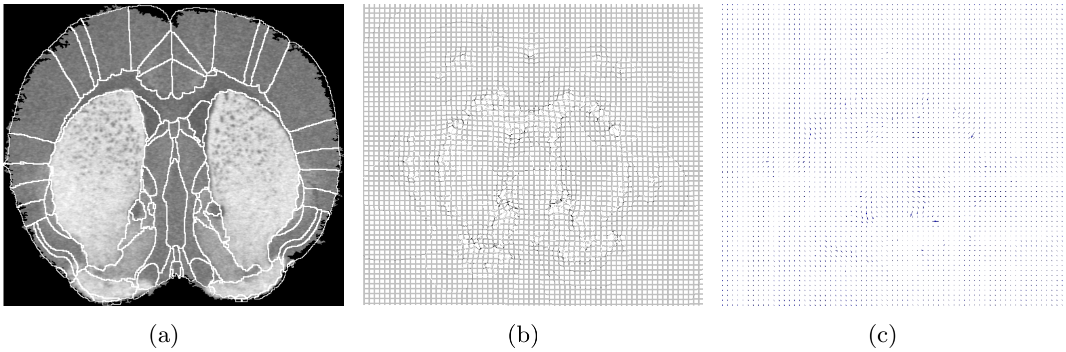

(a) (b) (c)

Figure 8.16: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying an Inverse Distance Weighted interpolator (a). The image in the middle shows a control grid showing the warping function (b). The right image shows the displacement vectors corresponding to the warping function (c).

8.5.2 Results with Radial Basis Functions

The result of using a scattered data interpolator based on linear Radial Basis Functions is shown in Figure 8.17. Figure 8.17(a) shows the autoradiogram with the corresponding warped atlas section superimposed on top after the elastic transformation function applied to the atlas, Figure 8.17(b) shows a control grid that shows the warping function used in this example and Figure 8.17(c) shows the displacement vectors over the surface after applying the transformation function.

The result of using the more extended RBF based on thin-plate splines is shown in Figure 8.18. Figure 8.18(a) shows the autoradiogram with the corresponding warped atlas section superimposed on top after the elastic transformation function applied to the atlas, Figure 8.18(b) shows a control grid that shows the warping function used in this example and Figure 8.18(c) shows the displacement vectors over the surface after applying the transformation function.

RBFs based on multiquadratic functions demonstrated an important dependence on the parameter c shown in equation 8.11. The higher c is the transformation has less effect on the atlas. This is clearly observed in Figure 8.19, where three values of c have been used, c=1 in Figure 8.19(a), c=10 in Figure 8.19(b) and c=100 in Figure 8.19(c). The corresponding control grids are shown in Figures 8.19(d), 8.19(e) and 8.19(f) respectively, and the corresponding displacement vectors are shown in Figures

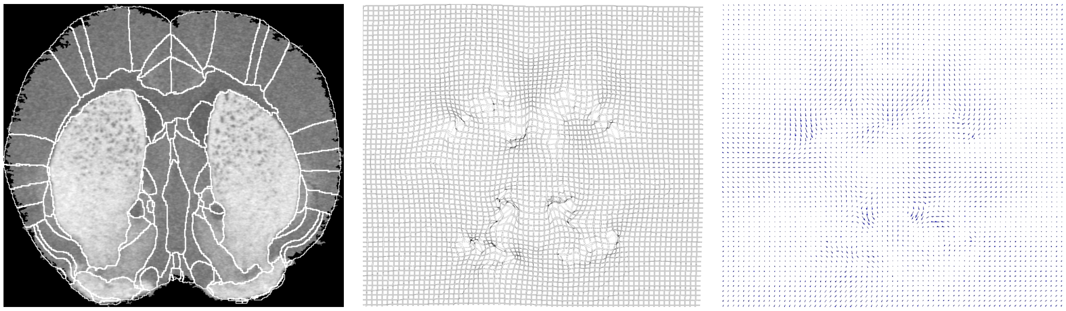

(a) (b) (c)

Figure 8.17: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying a scattered data interpolator based on linear RBFs (a). The image in the middle shows a control grid showing the warping function (b). The right image shows the displacement vectors corresponding to the warping function (c).

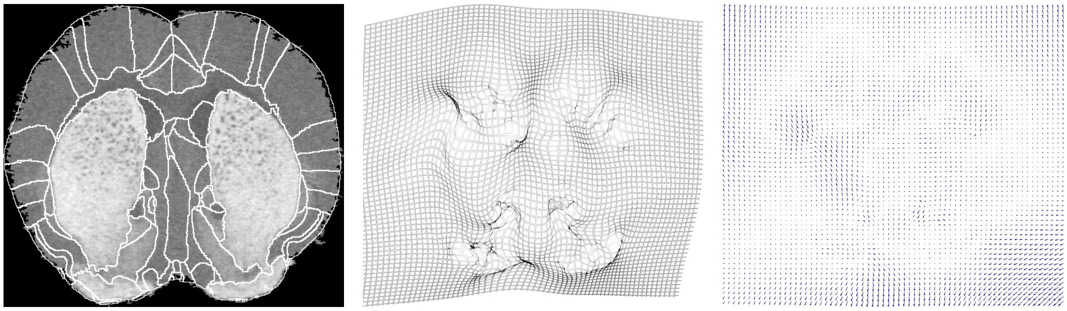
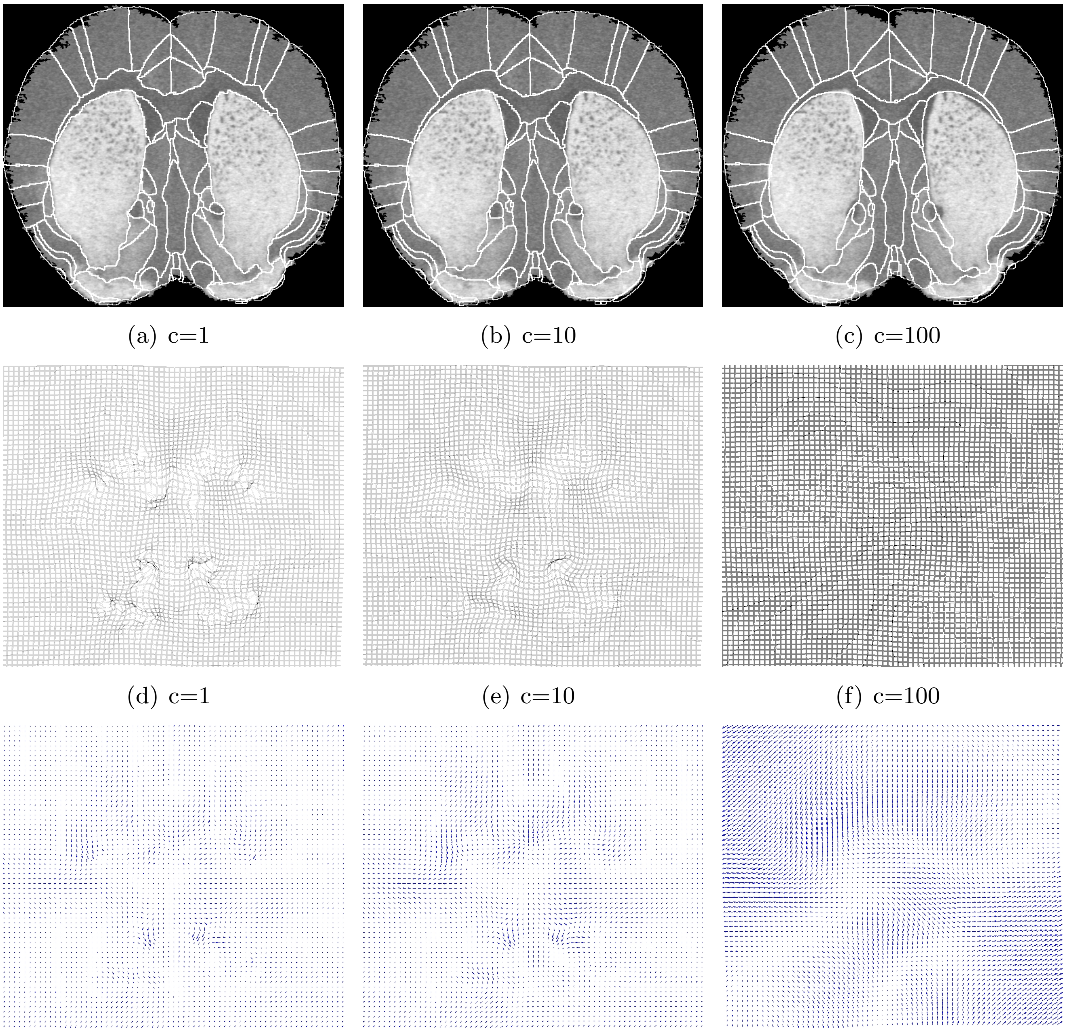

(a) (b) (c)

Figure 8.18: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying a scattered data interpolator based on thin-plate splines RBFs (a). The image in the middle shows a control grid showing the warping function (b). The right image shows the displacement vectors corresponding to the warping function (c).

8.19(g), 8.19(h) and 8.19(i) respectively. It can be observed from the control grids and the displacement vector fields how the warping function effect diminish as parameter c increases.

The specific case of RBFs based on the Wendland function is shown in Figure 8.20. This function has local support as opposite to the other types of RBFs shown above. The local support is controlled by imposing a maximum radius that limits the effect of the Wendland function. The effect of varying the radius is shown in three examples in Figure 8.20, for values a=10 (Figure 8.20(a)), a=30 (Figure 8.20(b)) and a=100 (Figure 8.20(c)). The control grids are shown in Figures 8.20(d), 8.20(e) and 8.20(f) respectively and the displacement vectors are shown in Figures 8.20(g), 8.20(h) and 8.20(i) respectively.

It is observed in Figure 8.20(a) that a ripple effect appears in the borders of those regions containing control points for a small radius. This suggests that the radius should be


(a) c=1 (b) c=10 (c) c=100

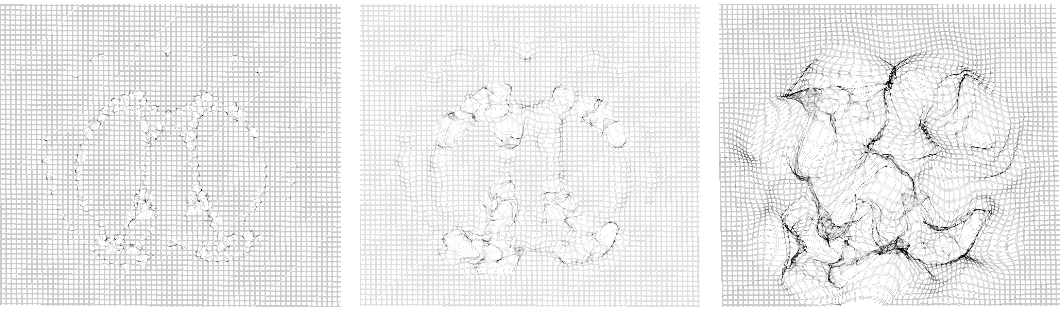
(d) c=1 (e) c=10 (f) c=100

(g) c=1 (h) c=10 (i) c=100

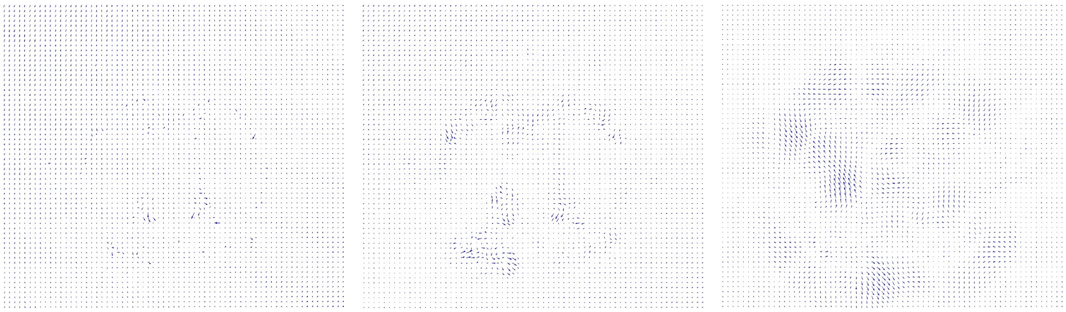
Figure 8.19: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying RBFs based on multiquadratic functions for c=1 (a), c=10 (b) and c=100 (c). The three images in the middle row (d), (e) and (f) show the respective control grids showing the warping functions. The three images in the bottom row (g), (h) and (i) show the displacement vectors corresponding to each warping function.

at least half of the distance between consecutive control points in the contour. This ripple effect disappears for larger radii as observed in Figure 8.20(b). For large radii the interpolator then behaves more as a global support function, as shown above, than as a local support function. The local support effect is observed from the control grids for small radii in Figures 8.20(d) and 8.20(e).

(a) a=10 (b) a=30 (c) a=100



(d) a=10 (e) a=30 (f) a=100

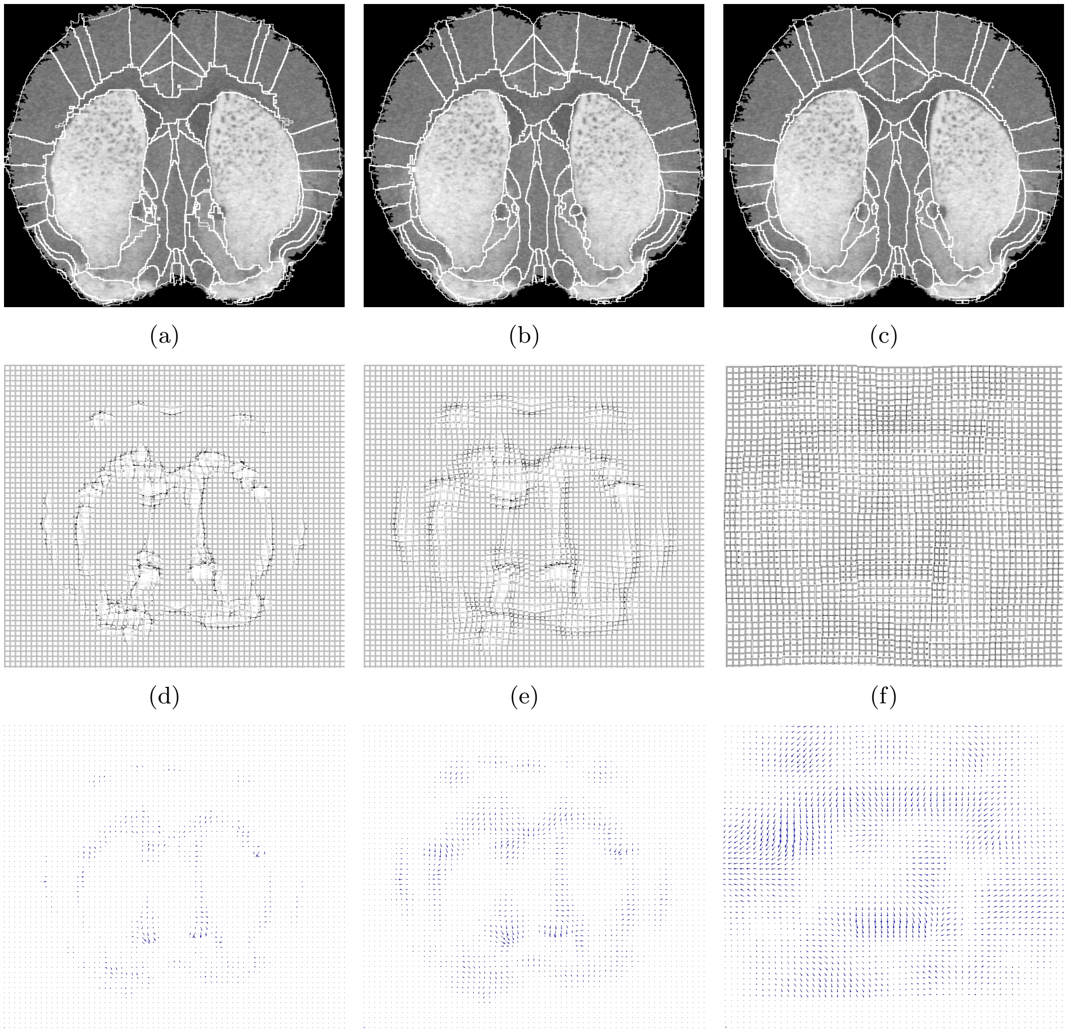


(g) a=10 (h) a=30 (i) a=100

Figure 8.20: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying RBFs based on the Wendland function for a=10 (a), a=30 (b) and a=100 (c). The three images in the middle row (d), (e) and (f) show the respective control grids showing the warping functions. The three images in the bottom row (g), (h) and (i) show the displacement vectors corresponding to each warping function.

8.5.3 Results with Cubic B-Splines

The result of using cubic B-Splines is shown in Figure 8.21. Different lattice interspaces have been used to see how as the lattice interspace increases a smoother warping function is applied, generating artifacts in some cases, but the registration is less accurate. The lattice interspace used is 5 pixels (Figure 8.21(a)), 10 pixels (Figure 8.21(b)) and 30 pixels (Figure 8.21(c)). The control grids are shown in Figures 8.21(d), 8.21(e) and 8.21(f) respectively and the displacement vectors are shown in Figures 8.21(g), 8.21(h) and 8.21(i) respectively.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 8.21: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying cubic B-Splines with a lattice interspace of 5 pixels (a), 10 pixels (b) and 30 pixels (c). The three images in the middle row (d), (e) and (f) show the respective control grids showing the warping functions. The three images in the bottom row (g), (h) and (i) show the displacement vectors corresponding to each warping function.

The compact support of this function is clearly observed in the control grids where only the surrounding areas to the initial scattered data points are transformed, keeping intact the remaining areas of the atlas. It is also observed how, while the lattice control points interspace increases, then the registration accuracy reduces. This approach does not then work as an interpolator but more as an approximator, as described in Section 8.4.6.

8.6 Validation

Given all the qualitative results shown above, differences in terms of smoothness and accuracy are observed between the different scattered data interpolators. These different interpolators are assessed in this section by obtaining a quantitative comparison between methods. A leave-one-out approach has been used.

It was shown in Chapter 7 an automatic algorithm for image segmentation in autoradiography. From the autoradiogram labelled with ^{3}H shown in Section 7.5.2 eight regions were segmented, and a set of control points extracted from the contour of each region. These regions are the Caudate Putamen (CPu), Olfactory Tubercle (Tu), Accumbens Nucleus (Acb), Ventral Pallidum (VP), Corpus Callosum (cc), Anterior Commissure (aca) and Lateral Ventricle (LV).

The leave-one-out approach has been implemented by considering 7 out of the 8 regions and leaving one out, and after warping the atlas with a specific interpolator, a measure of overlapping between the left-out region in the atlas and the corresponding region in the autoradiogram is measured. Some of these regions are observed to be merged, because the segmentation process failed to segment them independently. This failure is mainly due to the absence of difference or almost no difference of intensity between regions, i.e. no intensity edge , and as the segmentation algorithm is based on region grower, these kind of situations are not detected. The reason why these regions do not show an intensity edge is explained by the very similar ligand uptake of these regions, and concomitantly the similar density of D1 neuroreceptors. These regions with absence of intensity edge are the Accumbens Nucleus with the Ventral Pallidum (Acb+VP), the Corpus Callosum with the Anterior Commissure (cc+LV) and, in the case of the right hemisphere, there is no edge between the Caudate Putamen and the Olfactory Tubercle. These situation do not represent a problem for this algorithm as long as the corresponding regions are correctly identified in the atlas. The regions that have been analysed for comparison between algorithms are the cc+LV, CPu (left side), Acb+VP (left side), aca (left side) and Tu (left side).

The percentage of overlapping for each region (cc, CPu, VP, aca and Tu), as defined in equation 8.34, after applying the Inverse Distance Weighted interpolator, linear RBFs, thin-plate splines RBFs, multiquadratic RBFs (a=10), Wendland function-based RBFs (a=30) and Cubic B-Splines (for a lattice interspacing of 10 pixels) is shown in Figure 8.22.

$$O = 2\frac{\mid Rad \cap Atl \mid}{\mid Rad \mid + \mid Atl \mid} \tag{8.34}$$

where Atl represents the left-out region under study in the atlas, Rad represents the corresponding region in the autoradiogram and $\mid \cdot \mid$ represents the size of the region.

The four bars on the left shown in Figure 8.22 for each region represent the global support interpolators, while the two bars on the right represent the local support interpolators, the Wendland function-based RBF and the Cubic B-Splines.

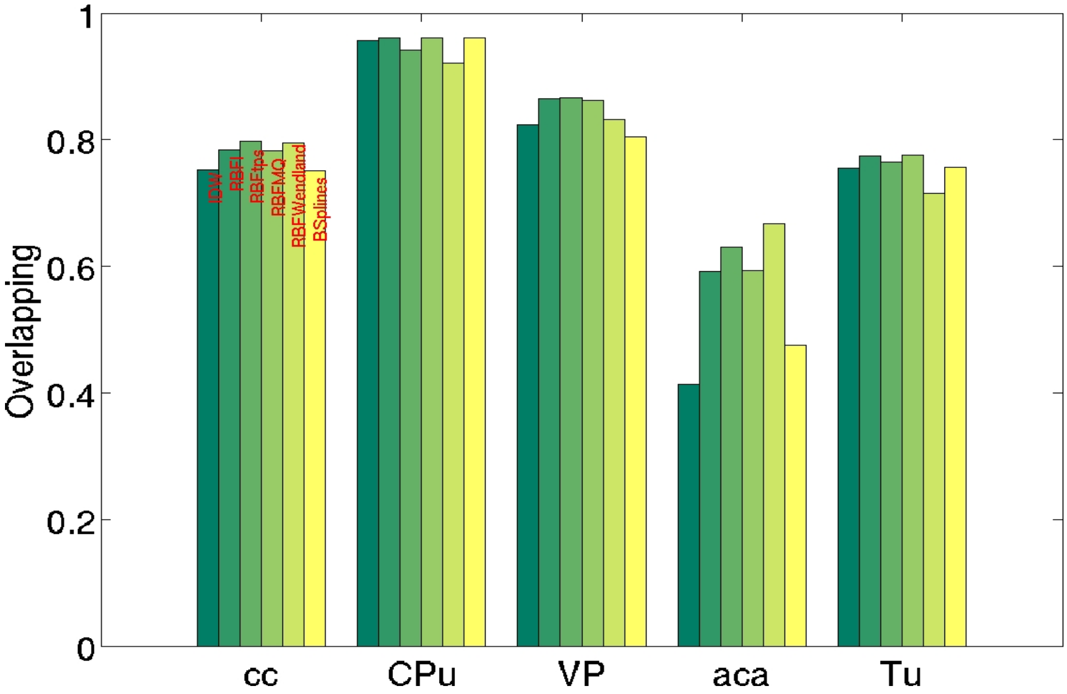


Figure 8.22: Percentage of overlapping for each of the 5 different analysed regions for the IDW, linear RBFs, thin-plate splines RBFs, multiquadratic RBFs and B-Splines respectively.

8.7 Conclusions

It has been observed that B-Splines with closer distance between control points in the lattice ϕ performs better showing higher overlap between the left-out-region and the corresponding autoradiogram region. It has also been observed how for multiquadratic RBFs, the higher the parameter a is the lesser accuracy is obtained.

The Caudate Putamen (CPu) shows a very similar, and overall high overlapping percentage between all the algorithms, explained by its large size and concomitantly easiness to register. This makes this region a not very good region to compare the different interpolators.

It is noticed how the Anterior Commissure region (aca), on the other hand, shows high variation of overlapping between algorithms. This is due to the small size of this region, making this specific region difficult to register compared to the other analysed regions.

The Olfactory Tubercle (Tu) also shows relative variation of overlapping, mainly due to the large difference of size between the original autoradiogram and atlas. This demonstrates that when a specific region in an autoradiogram largely differs from its counterpart in the atlas image, if this is not considered in the registration process as part of the biological landmark data set, the registration error will be significant.

Looking at the global support interpolators in detail it is observed a slight better performance of the thin-plate splines RBFs compared to the linear RBFs and the multiquadratic RBFs, being the performance of these two very similar in general. It can also be observed that the Inverse Distance Weighted interpolator performs slightly worse than the RBFs-based interpolators for all the analysed regions.

Regarding the local support interpolators, Wendland functions-based RBFs performs slightly better than Cubic B-Splines but this is not consistent along all the analysed regions.

In general none of the registration algorithms shows a clear higher performance over the others. Local support interpolators show slightly better performance for smaller regions, while very little difference is observed for larger regions.

Chapter 9

Conclusions and Future Work

9.1 Concluding Remarks

The aim of this work was to present a new perspective on conventional autoradiography by reducing the typical times spent in a typical autoadiography experiment. This was focused on the two main stages where most time is spent: sample exposure (to traditional film) and post-acquistion image analysis.

9.1.1 Acquisition

Sample exposure to film is largely dependant on the radioisotope being used and this in turn depends on the marker-neuroreceptor under study. The chemical nature of a specific marker will show certain preference to bind to a certain radioisotope, i.e. a marker based on Sulphur atoms will be easier to bind to ^{35}S than any other typical β-radioisotope. Samples labelled with ^{35}S and ^{125}I are usually exposed to traditional film for $\sim$4 days, samples labelled with ^{14}C are usually exposed for $\sim$1 week and samples labelled with ^{3}H are usually exposed for $\sim$5 weeks. In the case of ^{3}H, sample exposure requires long periods of time because the low energy of the β electrons emitted by ^{3}H, which causes that most β- are absorbed in the protective layers or the gel that comprises conventional film. Only a small proportion of the emitted β- ionise silver atoms in film. ^{35}S and ^{14}C require only a few days of exposure time because of the higher energy compared to ^{3}H. ^{125}I requires also few days because it is an X-ray emitter, and X-rays are not so easily absorbed in the protective layers or gel compared to β-. Therefore a higher proportion of X-rays are able to ionise the silver atoms in film.

After exposing the sample and developing the film through delicate manual procedures, the film is subsequently manually analysed to extract the radiomarker density of those regions showing high radioligand binding density. This process is prone to error and is susceptible to inter- and intra variability of life scientists analysing the images. This makes the process unreliable if the region under study shows medium or low density of ligand binding. Moreover, this process can take approximately an hour on average per brain section. This analysis process usually occupies in total approximately three days of an expert life scientist's time.

Several alternatives have been considered in the past to replace conventional film. The only two alternatives considered by life scientists as relatively successful are phosphor plates and the commercial Biospace: the β- and μ-Imagers. The former fails to obtain competitive spatial resolution and has been largely discarded as an option by most research groups, and the latter claims to produce a spatial resolution of 15 μm and 50 μm for ^{3}H for the μ-Imager and the β-Imager respectively. However their high cost and very small Field of View make these systems inaccessible and/or unattractive to most research groups.

As a first part of this work two silicon solid-state alternatives, such as CCD and CMOS technologies, have been thoroughly analysed and compared. CCD and CMOS technologies are two competing technologies with some points in common but also some important differences. Digital technologies offer the possibility of on-line monitoring a specific sample at the same time as the digital image is being formed. Experiments have demonstrated that single events can be observed on the screen as they deposit energy in the silicon. This has a great potential as life scientists can discard or continue experimenting with new radiomarkers or new drugs, and do not need to wait for a long period of time to see a final autoradiogram. This will save a lot of time in preliminary studies.

CCD and CMOS technologies have demonstrated higher sensitivity compared to conventional film, as well as their increased dynamic range. It has been shown the necessity of back-thinning to be able to detect ^{3}H β electrons in both technologies. In the specific case of CMOS technology a quantitative comparison has been undertaken by measuring the ligand density in some predefined regions of interest for 36 hours. By comparing the measured density ligand concentration in the film autoradiogram, that had been exposed for 4 weeks, similar density ligand concentrations were obtained with the back-thinned CMOS detector in $\sim$13 hours for the 5 regions under study. This represents an increased capture speed by a factor of 51 times in sensitivity compared to conventional autoadiography film. To make this comparison more accurate the same section could be exposed to film for different periods of time and subsequent image analysis would compare different ligand density concentrations along time, but due to the high cost of autoradiography film and the lengthy development of film and manual analysis renders this approach is prohibitive.

Another important feature that these technologies have to show to be a viable alternative to conventional film is their spatial resolution. Autoradiography in general requires high spatial resolution on the order of 1-10 μm depending on the radioisotope used.

This is an even more important requirement for brain autoradiography, the main focus of this work, where a spatial resolution of 1 μm is often desired by life scientists. It is important to note that due to multiple scattering inside the gel and the digitisation step conventional film does not produce such high spatial resolution. In the specific case of our Biosciences laboratory, using a digitization system with an 11 μm pixel size CCD camera, the spatial resolution for ^{14}C has been measured resulting in approximately 30-40 μm.

The pixel size and the thickness of the sensitive layer have been identified, through Monte Carlo simulations, to be of significant importance in both technologies to obtain high spatial resolution. These Monte Carlo simulations have been undertaken, including dark current modelling and fixed pattern noise simulations derived from the detector, and an experimentally validated charge diffusion model in order to obtain realistic results.

Monte Carlo simulations have demonstrated that spatial resolution scales largely with the pixel size, and the higher the energy of the radioisotope the higher the dependency on the sensitive layer thickness. For ^{3}H a spatial resolution of 1 μm can theoretically be achieved using a detector with $\sim$1 μm pixel pitch. For higher energy radioisotopes, such as ^{14}C and ^{35}S, the thickness of the sensitive layer plays a far more significant role, degrading the spatial resolution as it becomes affected by higher multiple scattering.

To reduce this effect, using a thinner sensitive layer is a possibility, resulting in reduced sensitivity as less charge is deposited. Another alternative is to use CCD technology, where the first few microns of the sensitive layer are affected by an electric field. This electric field makes the deposited charge to drift towards the charge collection points, thus minimising the charge diffusion process and reducing the multiple scattering effect.

9.1.2 Image analysis

The second part of this work is focused on the posterior to acquisition image analysis process. As a life scientist, the main goal of autoradiography is to study the different levels of ligand density present in a certain tissue, after the subject under study has been treated with an experimental drug. In order to do so, after exposing the tissue section to conventional film, this has to be developed and digitised. Then, an expert life scientist manually delineates all the significant regions in the tissue section taking a paper version of an atlas as a reference. Sometimes these regions are clearly visible due to a high or low level of ligand density in comparison with the surrounding regions. However, there are times when the life scientist delineates regions that do not show different levels of ligand concentration and there is no information in the image that represents the different tissues. In these cases the life scientist's expertise is playing a major role largely compromising the delineation accuracy.

In this work an approach to semi-automatically delineate all the regions of the tissue section under study has been presented. This was divided into two stages: firstly, regions with high and low levels of ligand density are automatically segmented, and secondly, after extracting a set of control points from those segmented regions, the autoradiogram is registered with its corresponding atlas image.

Autoradiography presents the typical problem present in functional data, such as high statistical noise and low definition in the edges of the different functional regions. As has been mentioned, the radioisotope used depends on the neuroreceptor under study. Depending on the radioisotope the edges will be more or less defined. A low energy radioisotope will produce better defined edges than a higher energy radioisotope, making the segmentation process easier.

Region growing has been chosen in this work because it is a very robust option given that it uses intensity and spatial information simultaneously. The two main components of a region grower are the similarity criterion followed to append pixels to a growing region and the stopping criteria needed to know when to stop appending pixels.

Due to the high statistical noise typically present in autoradiographic data, a dynamic similarity criterion has been used. This criterion automatically updates while a region grows considering the amount of statistical noise present in the image. This avoids a premature halt in the algorithm.

Given the complexity of the data being analysed in this work, the termination criteria are comprised of a set of conditions that contemplate different possible situations that have been observed. Detection of leakage in to a background region and edge detection are the two conditions that have been used to stop growing a region. This combination of criteria have been validated with synthetic data, by modelling the level of statistical noise and measuring the level of over- and under- segmentation. This has demonstrated that the region grower presented here is able to properly segment autoradiographic data under the presence of a high level of statistical noise, with the added difficulty of blurred edges between different tissues.

As synthetic data does not fully consider all the problems that can appear in real data, this region grower has also been tested with different film autoradiographic sections using different radioisotopes. The CNR of different regions being segmented has been manually measured showing low levels of CNR, representing the difficulty of this scenario. These results were quantitatively compared with manual segmentation by expert life scientists, measuring the Dice index between corresponding manually and automatically segmented regions, and resulting in high levels of agreement. It was also observed how the region grower was able to segment certain regions with low levels of ligand density and low importance for life scientists, but of high importance to increase the accuracy of the subsequent registration process. A significant dependence of the final success of the segmentation process has been observed on the quality of the anatomy of the tissue section, as well as how well the radioligand was distributed on the section.

Subsequent to the segmentation, manual correspondence between segmented regions and counterpart regions in the atlas was established. From these, a set of control points were extracted. These control points were refined to better represent the shape of each region. Finally, some of the most popular interpolators have been used to compute the transformation function for all the pixels in the image. These interpolators have been divided between different families: interpolators with global and local support, showing in general a similar performance.

This process is successfully based on two assumptions: first that the internal anatomy of the section is approximately well described by its corresponding atlas section, and secondly, that the segmentation process has been successful at correctly identifying those regions with higher and lower density of ligand concentration. If these two assumptions are fulfilled, the registration process produces a warped version of the atlas image, superimposed on the autoradiogram, thus automatically delineating all the anatomical regions that are present in the atlas image. Thus it is possible to extract the density of ligand concentration in the autoradiogram automatically.

9.2 Future Work

Comparing tritiated brain autoradiographic samples acquired with the back-thinned Vanilla sensor and with a commercial back-thinned CCD, under temperature controlled conditions, the changes of temperature were shown to have a significant impact on the background noise of the final results. From the technological part of this thesis, a natural extension of this work is to use the back-thinned CMOS sensor at room temperature but under temperature controlled conditions, by using a thermoelectric temperature controller (TED).

The correspondence between segmented regions and atlas regions is undertaken manually in this work. To have a fully automatic registration algorithm this process should be automated. An approach based on spatial closeness between counterpart regions and comparing region features, such as size, orientation, minor and major axes, was initially tried, but problems arose for small regions that are not close enough to establish the right correspondence.

A common problem that life scientists claim when they have to segment manually autoradiographic data, is that the cutting process is not perfect and the thin sections that they cut are not parallel to the coronal axis. There is usually a slight tilting between the coronal plane and the blade. This results in the reference that they use [135] to delineate the different regions under study being not exact. Therefore in a single section, there may be more than one bregma needed as reference. This issue makes the delineating process inaccurate depending on the tilting angle between the coronal plane and the blade.

The algorithm presented here to automatically delineate the entire autoradiogram does not consider this problem, as this is entirely undertaken in 2D. A possible extension of this algorithm would be to develop the algorithm for a 3D scenario.

First a 3D model of the atlas brain should be built based on the reference [135]. Secondly the correct tilting angle should be searched to find the right correspondence between the 3D atlas and the different functional structures of the autoradiogram.

In order to add information to this algorithm to make the whole process more accurate intensity information could be added to the atlas. There exists atlases in the internet with intensity information made of histological sections, i.e. high resolution pictures of brain sections. If some kind of relation or correspondence is found between the different structural regions from the atlas and the histological sections, and then it is used to establish a correspondence with the functional structures of the autoradiogram, the process could be undertaken completely automatic and with higher information.

Appendix A

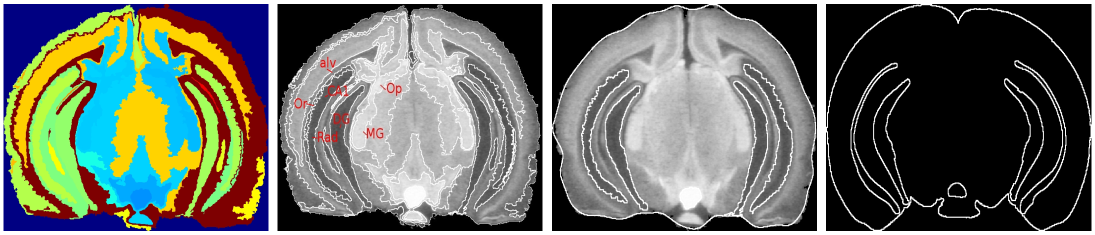
Extra results of Registration

The automatic registration algorithm explained in Chapter 8 has been applied to those autoradiograms segmented shown in Section 7.5.1, where a ^{125}I labeled autoradiogram is segmented, and Section 7.5.3, where a ^{35}S labeled autoradiogram is segmented.

Not all the data interpolators have been tried in these autoradiograms given the significant similarity of the quantitative results obtained after the validation in Section 8.6.

A.1 Results with Iodine-125

The segmentation of this section has been shown in Section 7.5.1. The resulting labeled and segmented autoradiograms are shown in Figures A.1(a) and A.1(b) respectively. Those regions, identified in the autoradiogram and atlas image, used in the registration as source and target sets of landmarks, are shown in Figures A.1(c) and A.1(d) respectively.



(a) (b) (c) (d)

Figure A.1: Labeled ^{125}I autoradiogram (a), resulting segmented autoradiogram with some functional regions indicated (b), autoradiogram with regions used in the registration delineated (c) and corresponding delineated regions in the atlas image (d).

The result of after applying affine registration between the atlas (source image) and the autoradiogram (target image) is shown in Figure A.2. Internal functional structures show significant missmatches between the atlas and the autoradiogram, making necessary elastic registration.

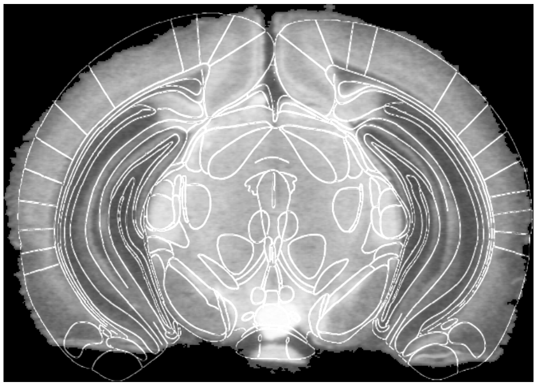


Figure A.2: Affine registration of the autoradiogram and the slice of the atlas corresponding to that brain section (Bregma -3.16 mm). In general poor correspondence between the internal functional regions can be observed.

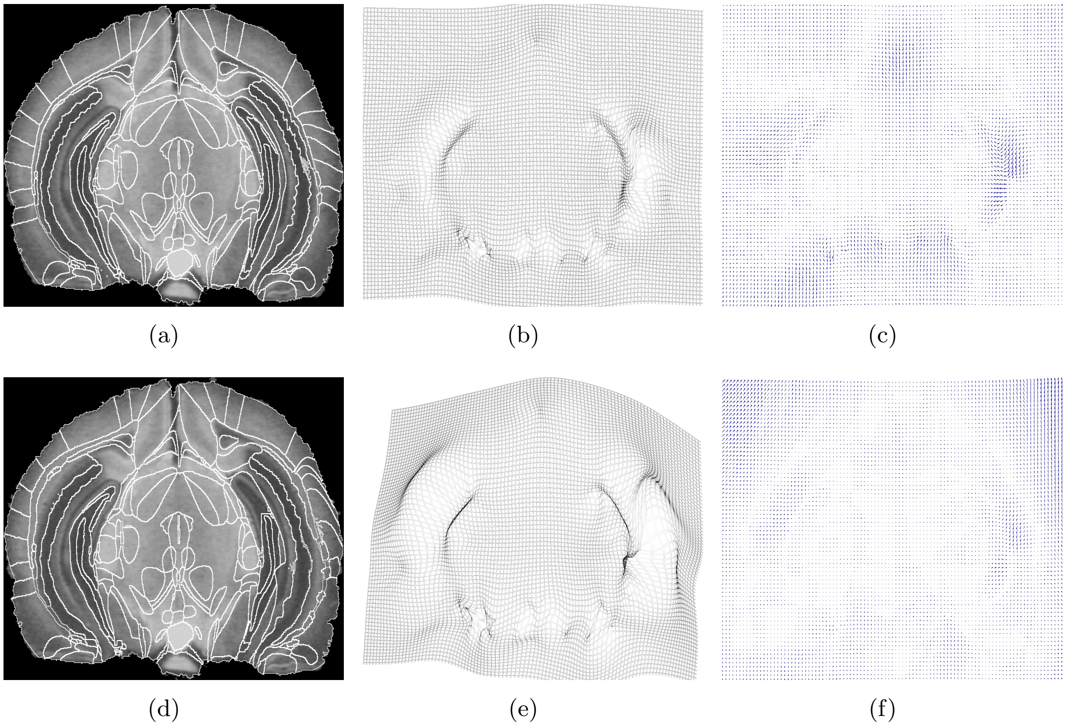
The resulting registered autoradiogram after applying a scattered data interpolator based on linear RBFs and based on thin-plate splines RBFs, as representative global support interpolators, are shown in Figures A.3(a) and A.3(d) respectively. The resulting control grids and vector fields, are shown in Figures A.3(b) and A.3(c) respectively for the linear RBFs, and in Figures A.3(e) and A.3(f) respectively for the thin-plate splines RBFs.

The resulting registered autoradiogram after applying a scattered data interpolator based on Wendland RBFs with a local effect of 30 pixels (radius of Wendland function), and the same autoradiogram after applying a data interpolator based on B-splines with a lattice interspace of 10 pixels, are shown in Figures A.4(a) and A.4(d) respectively. The resulting control grids and the vector fields, are shown in Figures A.4(b) and A.4(c) respectively for the Wendland function, and in Figures A.4(e) and A.4(f) respectively for the B-splines.

A.2 Results with Sulphur-35

The segmentation of this section has been shown in Section 7.5.3. The resulting labeled and segmented autoradiograms are shown in Figures A.5(a) and A.5(b) respectively. Those regions, identified in the autoradiogram and atlas image, used in the registration as source and target sets of landmarks, are shown in Figures A.5(c) and A.5(d) respectively.

The result of after applying affine registration between the atlas (source image) and the autoradiogram (target image) is shown in Figure A.6. Internal functional structures show significant missmatches between the atlas and the autoradiogram, making necessary elastic registration.


(a) (b) (c)

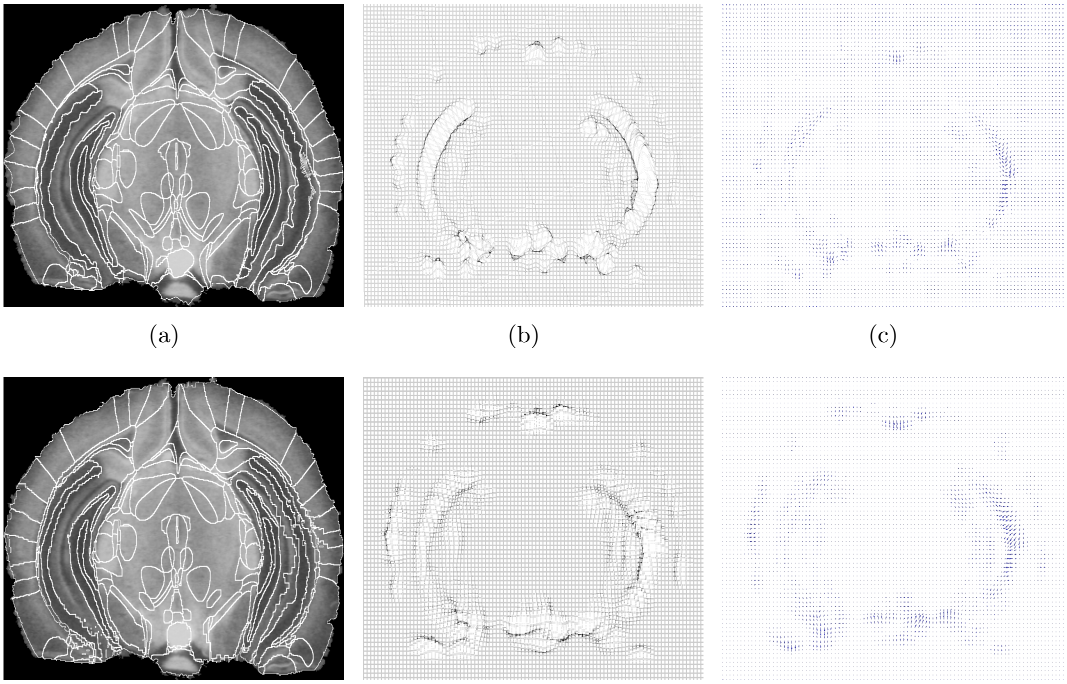
(d) (e) (f)

Figure A.3: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top after applying a scattered data interpolator based on linear RBFs (a). The image in (b) shows the resulting control grid showing the warping function, and the image in (c) shows the displacement vectors corresponding to the warping function. The same autoradiogram, superimposed on top of the atlas image after applying a scattered data interpolator based on thin-plate splines RBFs is shown in (d). The image in (e) shows the resulting control grid showing the warping function, and the image in (f) shows the displacement vectors corresponding to the warping function.

The resulting registered autoradiogram after applying a scattered data interpolator based on linear RBFs and based on thin-plate splines RBFs, as representative global support interpolators, are shown in Figures A.7(a) and A.7(d) respectively. The resulting control grids and vector fields, are shown in Figures A.7(b) and A.7(c) respectively for the linear RBFs, and in Figures A.7(e) and A.7(f) respectively for the thin-plate splines RBFs.

The resulting registered autoradiogram after applying a scattered data interpolator based on Wendland RBFs with a local effect of 30 pixels (radius of Wendland function), and the same autoradiogram after applying a data interpolator based on B-splines with a lattice interspace of 10 pixels, are shown in Figures A.8(a) and A.8(d) respectively. The resulting control grids and the vector fields, are shown in Figures A.8(b) and A.8(c) respectively for the Wendland function, and in Figures A.8(e) and A.8(f) respectively for the B-splines.

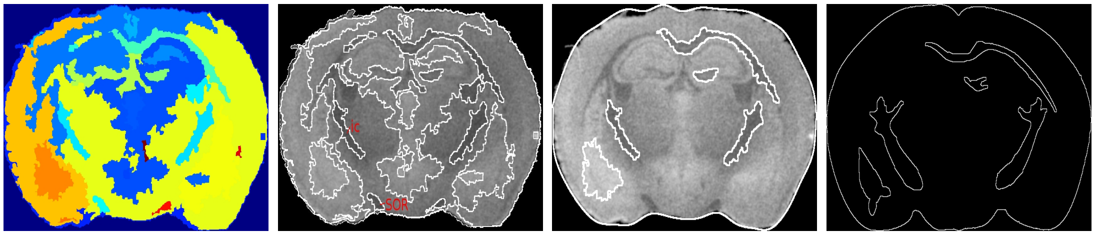
This case is a good example where the radiolabeled ligand shows low specificity, which makes the segmentation process a difficult task. Therefore, few functional structures can be identified in the autoradiogram and in the atlas. The resulting interpolation



(a) (b) (c)

(d) (e) (f)

Figure A.4: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying a scattered data interpolator based on Wendland RBFs (a). The image in (b) shows a control grid showing the warping function and the image in (c) shows the displacement vectors corresponding to the warping function. The same autoradiogram superimposed on top of the atlas image, after applying a scattered data interpolator based on B-splines with a lattice interspace of 10 pixels is shown in (d). The image in (e) shows a control grid showing the warping function and the image in (c) shows the displacement vectors corresponding to the warping function (f).

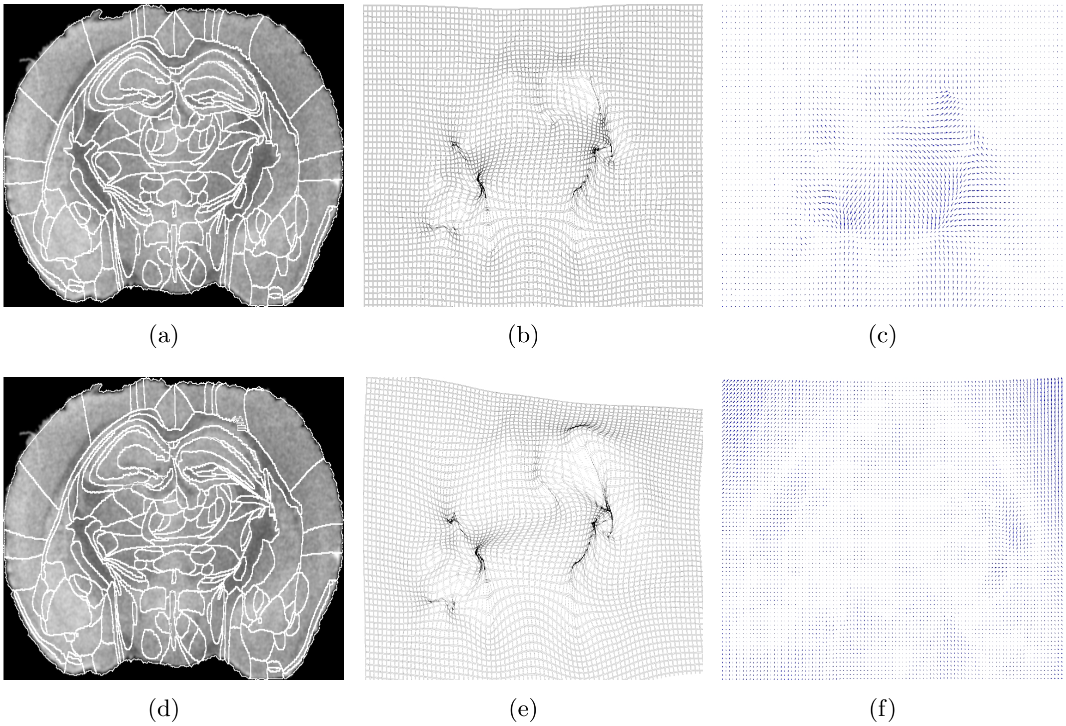


(a) (b) (c) (d)

Figure A.5: Labeled ^{35}S autoradiogram (a), resulting segmented autoradiogram with some functional regions indicated (b), autoradiogram with regions used in the registration delineated (c) and corresponding delineated regions in the atlas image (d).

process shows similar results with all the interpolators as expected, shown overall good results. In this case, even though few regions were identified and used as input for the

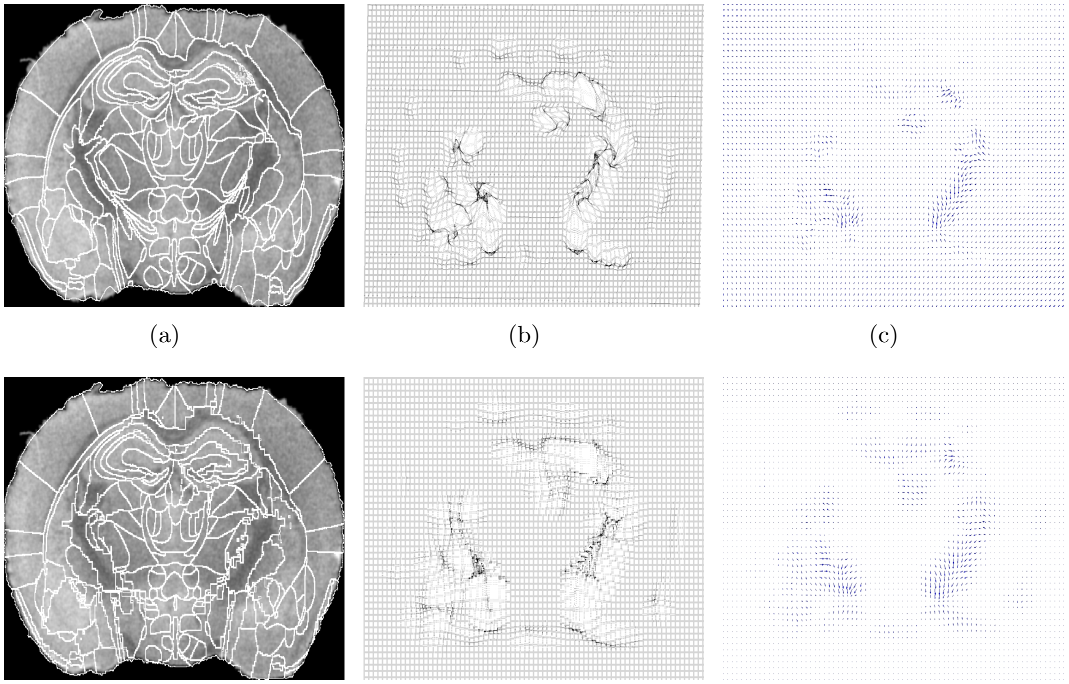
Figure A.6: Affine registration of the autoradiogram and the slice of the atlas corresponding to that brain section (Bregma -1.46 mm). In general poor correspondence between the internal functional regions can be observed.



(a) (b) (c)

(d) (e) (f)

Figure A.7: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top after applying a scattered data interpolator based on linear RBFs (a). The image in (b) shows the resulting control grid showing the warping function, and the image in (c) shows the displacement vectors corresponding to the warping function. The same autoradiogram, superimposed on top of the atlas image after applying a scattered data interpolator based on thin-plate splines RBFs is shown in (d). The image in (e) shows the resulting control grid showing the warping function, and the image in (f) shows the displacement vectors corresponding to the warping function.



(a) (b) (c)

(d) (e) (f)

Figure A.8: Autoradiogram of mouse brain section with corresponding atlas section superimposed on top, after applying a scattered data interpolator based on Wendland RBFs (a). The image in (b) shows a control grid showing the warping function and the image in (c) shows the displacement vectors corresponding to the warping function. The same autoradiogram superimposed on top of the atlas image, after applying a scattered data interpolator based on B-splines with a lattice interspace of 10 pixels is shown in (d). The image in (e) shows a control grid showing the warping function and the image in (c) shows the displacement vectors corresponding to the warping function (f).

interpolation algorithm, the registration of both images makes the quantitative analysis more accurate and less error-prone than free-hand delineation.

Appendix B

List of Publications:

The following publications have resulted from the work documented in this thesis:

Journal publications:

1. J. Cabello, A. Bailey, I. Kitchen, M. L. Prydderch, A. T. Clark, R. Turchetta and K. Wells. Digital Autoradiography Using Room Temperature CCD and CMOS Imaging Technology. Physics in Medicine and Biology, 2007; 52:4993-5011.

Conference publications:

1. J. Cabello, A. Bailey, I. Kitchen, M. Guy and K.Wells. Segmentation of low contrast-to- noise ratio images applied to functional imaging using adaptive region growing. In SPIE Medical Imaging, 2009., volume 7259, pages 40-52, Feb 2009.

2. J. Cabello, A. Holland, K. Holland, A. Bailey, I. Kitchen and K. Wells. Betacam: A Commercial Approach to B- Autoradiography. In SPIE Medical Imaging, 2009., volume 7258, pages 3P-10 Feb 2009.

3. J. Cabello, A. Bailey, I. Kitchen, R. Turchetta and K. Wells. A Dual Threshold Method to Independently Control Spatial Resolution and Sensitivity in β Imaging. In Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE, pages 1-7, Oct 2008.

4. J. Cabello and K. Wells. A Monte Carlo Investigation into the Fundamental Limitations of Digital β-Autoradiography: Considerations for Detector Design. In Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE, volume 5, pages 3625-30, Oct 2007.

5. J. Cabello, K. Wells, A. Bailey and I. Kitchen. Semi-automatic elastic registration applied to a β-Autoradiography brain atlas. In Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE, volume 6, pages 4303-7, Oct 2007.

6. J. Cabello, K. Wells, A. Metaxas, A. Bailey, I. Kitchen, A. Clark, M. Prydderch, and R. Turchetta. Digital autoradiography imaging using CMOS technology: First tritium autoradiography with a back-thinned CMOS detector and comparison of CMOS imaging performance with autoradiography film. In Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE, volume 5, pages 3743-6, Oct 2007.
7. J. Cabello, A. Bailey, I. Kitchen, A. T. Clark, J. P. Crooks, R. Halsall, M. L. Key-Charriere, S. Martin, M. L. Prydderch, R. Turchetta and K. Wells. Digital Autoradiography Using CCD and CMOS Imaging Technology. In Nuclear Science Symposium Conference Record, 2006. NSS '06. IEEE, volume 4, pages 2607-12, Oct 2006.

Bibliography

[1] L. Abate, E. Bertolucci, M. Conti, G. Mettivier, M. C. Montesi, and P. Russo. GaAs pixel arrays for β imaging in medicine and biology. *Nuclear Instruments and Methods in Physics Research Section A*, 460:97–106, 2001.

[2] ISE-TCAD ISE Integrated Systems Engineering AG. Zurich/CH, Software Release 5.0 and 6.0 Users Manuals.

[3] R.K. Ahrenkiel, W.K. Metzger, M. Page, R. Reedy, J. Luther, and J. Dashdorj. Relationship of recombination lifetime to dark current in silicon p-n junctions. In *Photovoltaic Specialists Conference, 2005. Conference Record of the Thirty-first IEEE*, pages 895–8, Jan 2005.

[4] K. R. Alper, E. R. John, J. Brodie, W. Gunther, R. Daruwala, and L. S. Prichep. Correlation of PET and qEEG in normal subjects. *Psychiatry Research: Neuroimaging*, 146:271–82, 2006.

[5] N. M. Alpert, J. F. Bradshaw, D. Kennedy, and J. A. Correia. The Principal Axes Transformation- A method for image registration. *Journal of Nuclear Medicine*, 31:1717–1722, 1990.

[6] K. Amako, S. Guatelli, V. Ivanchenko, M. Maire, B. Mascialino, K. Murakami, L. Pandola, S. Parlati, M. G. Pia, M. Piergentili, T. Sasaki, and L. Urban. Validation of Geant4 electromagnetic physics versus protocol data. *Nuclear Science Symposium Conference Record, 2004 IEEE*, 4:2115–9, 2004.

[7] Aptina Imaging, Inc (former Micron Imaging, Inc) http://www.aptina.com/.

[8] ARIS*ER Consortium. http://www.ariser.info/.

[9] C. D. Arvanitis, S. E. Bohndiek, G. Royle, A. Blue, H. X. Liang, A. Clark, M. Prydderch, R. Turchetta, and R. Speller. Empirical electro-optical and x-ray performance evaluation of CMOS active pixels sensor for low dose, high resolution x-ray medical imaging. *Medical Physics*, 34:4612–25, 2007.

[10] S. U. Ay and E. R. Fossum. A 76 x 77mm^2, 16.85 Million Pixel CMOS APS Image Sensor. In *Symposium on VLSI Circuits, 2006. Digest of Technical Papers. 2006*, pages 19–20, 2006.

[11] J. Baro, J. Sempau, J. M. Fernandez-Varea, and F. Salvat. PENELOPE: an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons

and positrons in matter. *Nuclear Instruments and Methods in Physics Research Section A*, 100:31–46, 1995.

[12] N. Barthe, K. Chatti, P. Coulon, S. Maitrejean, and B. Basse-Cathalinat. Recent technologic developments on high resolution beta imaging for quantitative autoradiography and double labeling applications. *Nuclear Instruments and Methods in Physics Research Section Av*, 527:41–5, 2004.

[13] N. Barthe, P. Coulon, C. Hennion, D. Ducassou, B. Basse-Cathalinat, and G. Charpak. Optimization of a new scintillation gas detector used to localize electrons emitted by ^{99m}Tc. *The J. Nucl. Med.*, 40:868–75, 1999.

[14] J. E. Bateman. Quantitative autoradiographic imaging using gas-counter technology. *Electrophoresis*, 11(5):367–75, 1990.

[15] M. Battaglia, B. Hooberman, D. Contarato, L. Greiner, T. Kim, H. Wieman, and R. Foglia. Back-thinning of CMOS Monolithic Active Pixel Sensors for the ILC at LBNL. In *International Symposium on Detector Development for Particle, Astroparticle and Synchrotron Radiation Experiments*, Apr 2006.

[16] M. A. Baysal and E. Toker. A new CMOS-based digital imaging detector for applications in mammography. *Proc. SPIE*, 5923:109–19, 2005.

[17] E. Belau, R. Klanner, G. Lutz, E. Neugebauer, H. J. Seebrunner, A. Wylie, T. Bohringer, L. Hubbeling, P. Weilhammer, J. Kemmer, U. Kotz, and M. Riebesell. Charge collection in silicon strip detectors. *Nuclear Instruments and Methods in Physics Research*, 214(2-3):253–60, Sep 1983.

[18] E. Bertolucci, M. Conti, G. Grossi, G. Madonna, E. Mancini, P. Russo, M. Caria, P. Randaccio, del Guerra A., M. Gambaccini, R. Marchesini, M. Marziani, A. Taibi, R. Beccherle, M. G. Bisogni, U. Bottigli, M. E. Fantacci, V. Rosso, A. Stefanini, R. Tripiccione, and S. R. Amendolia. Autoradiography with silicon strip detectors. *Nuclear Instruments and Methods in Physics Research Section A*, 381:527–530, 1996.

[19] E. Bertolucci, M. Conti, G. Mettivier, M. C. Montesi, and P. Russo. Betaview: a digital β-imaging system for dynamic studies of biological phenomena. *Nuclear Instruments and Methods in Physics Research Section A*, 478:109–113, Feb 2002.

[20] M. Bigas, E. Cabruja, and M. Lozano. Bonding techniques for hybrid active pixel sensors (haps). *Nuclear Instruments and Methods in Physics Research Section A*, 574:392–400, May 2006.

[21] P. M. Bloomfield, T. J. Spinks, J. Reed, L. Schnorr, A. M. Westrip, L. Livieratos, R. Fulton, and T. Jones. The design and implementation of a motion correction scheme for neurological PET. *Physics in Medicine and Biology*, 48:959–78, 2003.

[22] R. K. Bock and A. Vasilescu. *The Particle Detector BriefBook*. Springer, 1998.

[23] S. E. Bohndiek, E. J. Cook, C. D. Arvanitis, A. Olivo, G. J. Royle, A. T. Clark, M. L. Prydderch, R. Turchetta, and R. D. Speller. A CMOS active pixel sensor system for laboratory-based x-ray diffraction of biological tissue. *Physics in Medicine and Biology*, 53:655–72, 2008.

[24] F. L. Bookstein. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(6):567–85, 1989.

[25] T. M. V. Bootsma, E. J. van Zwet, A. C. Brinkman, J. W. den Herder, L. de Jong, P. de Korte, and S. M. Olsthoorn. Synchrotron calibration and response modelling of back-iiluminated XMM-RGS CCDs. *Nuclear Instruments and Methods in Physics Research Section A*, 439:575–81, 2000.

[26] K. Bowyer. Validation of medical image analysis techniques. In *Handbook of Medical Imaging, Vol. 2: Medical Image Processing and Analysis.* New York: SPIE, 2000.

[27] J. Briesmeister. *MCNP a general Monte Carlo N-Particle Transport Code.* MCNP a general Monte Carlo N-Particle Transport Code, 2000.

[28] L. G. Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376, 1992.

[29] A. Brunettia and R. Cesareo. X-ray tomography using a CMOS area detector. *Nuclear Instruments and Methods in Physics Research*, A258:485–9, 2007.

[30] J. Cabello, A. Bailey, I. Kitchen, A. Clark, J. Crooks, R. Halsall, M. Key-Charriere, S. Martin, M. Prydderch, R. Turchetta, and K. Wells. Digital autoradiography using CCD and CMOS imaging technology. In *Nuclear Science Symposium Conference Record, 2006 IEEE*, volume 4, pages 2607–12, Oct. 29 2006-Nov. 1 2006.

[31] J. Cabello, K. Wells, A. Metaxas, A. Bailey, I. Kitchen, A. Clark, M. Prydderch, and R. Turchetta. Digital autoradiography imaging using CMOS technology: First tritium autoradiography with a back-thinned CMOS detector and comparison of CMOS imaging performance with autoradiography film. In *Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE*, volume 5, pages 3743–6, Nov 2007.

[32] C. Cappellini, A. Bulgheroni, M. Caccia, V. Chmill, M. Jastrzab, F. Risigo, and P. Scopelliti. Imaging of biological samples with silicon pixel detectors. *Nuclear Instruments and Methods in Physics Research Section A*, 591:34–7, 2008.

[33] G. Charpak, R. Bouclier, T. Bressani, J. Favier, and C. Zupancic. The use of multiwire proportional counters to select and localise charged particles. *Nuclear Instruments and Methods in Physics Research Section A*, 62:265–268, 1968.

[34] J. L. Chen, S. R. Gunn, M. S. Nixon, and R. N. Gunn. Markov random field models for segmentation of PET images. In *IPMI '01: Proceedings of the 17th International Conference on Information Processing in Medical Imaging*, pages 468–74, London, UK, 2001. Springer-Verlag.

[35] L. Chen, L. S. Gobar, N. G. Knowles, Z. Liu, A. F. Gmitro, and H. H. Barrett. Direct Imaging of Radionuclide-Produced Electrons and Positrons with an Ultrathin Phosphor. *The Journal of Nuclear Medicine*, 49:1141–5, 2008.

[36] S. R. Cherry, J. A. Sorenson, and M. E. Phelps. *Physics in Nuclear Medicine*. Saunders, 2003.

[37] G. E. Christensen. *Deformable shape models for anatomy*. Phd thesis, Washington University, August 1994.

[38] GEANT4 Collaboration. Geant4: A simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A*, A506:250–303, 2002.

[39] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *COMPUTER VISION AND IMAGE UNDERSTANDING*, 61:38–59, 1995.

[40] J.P. Crooks, J.A. Ballin, P.D. Dauncey, A.-M. Magnan, Y. Mikami, O. Miller, M. Noy, V. Rajovic, M. Stanitzki, K.D. Stefanov, R. Turchetta, M. Tyndel, E.G. Villani, N.K. Watson, and J.A. Wilson. A novel CMOS monolithic active pixel sensor with analog signal processing and 100% fill factor. In *Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE*, volume 2, pages 931–5, Nov 2007.

[41] W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: theory and practice. *The British Journal of Radiology*, 77:140–153, 2004.

[42] M. W. Davidson and M. Abramowitz. Molecular Expressions, Concepts in digital imaging technology, 2008. http://micro.magnet.fsu.edu/.

[43] G. Deptuch. Tritium autoradiography with thinned and back-side illuminated monolithic active pixel sensor device. *Nuclear Instruments and Methods in Physics Research Section A*, 543:537–548, Jan 2005.

[44] G. Deptuch, M. Winter, W. Dulinski, D. Husson, R. Turchetta, and J. L. Riester. Simulation and measurements of charge collection in monolithic active pixel sensors. *Nuclear Instruments and Methods in Physics Research Section A*, 465:92–100, 2001.

[45] P. F. D'Haese, V. Duay, R. Li, A. de B. d'Aische, T. E. Merchant, A. J. Cmelak, E. F. Donnely, K. J. Niermann, B. Macq, and B. M. Dawant. Automatic segmentation of brain structures for radiation therapy planning. *Proceedings of the SPIE*, 5032:517–26, May 2003.

[46] S. Dhawan. Introduction to Microchannel Plate Photomultipliers. *IEEE Transactions on Nuclear Science*, 28(1):672–6, Feb 1981.

[47] L. R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, Jul. 1945.

[48] B. Dierickx, G. Meynants, and D. Scheffer. Near-100% fill factor standard CMOS active pixel. In *CCD & AIS workshop*, page P1, 1991.

[49] B. Dogdas, D. Stout, A. F. Chatziioannou, and R. M. Leahy. Digimouse: A 3D whole body mouse atlas from CT and cryosection data. *Phys. Med. Biol.*, 52:577–87, 2007.

[50] E2V Technologies, http://e2v.com/.

[51] T. E. Everhart and P. H. Hoff. Determination of Kilovolt Electron Energy Dissipation vs Penetration Distance in Solid Materials. *Journal of Applied Physics*, 42:5837–46, 1971.

[52] B. F. Fandrich, R. Kohler, and F. Jenichen. *CCD-Cameras for X-Ray Investigations*. Fa. Proscan elektronische Systeme GmbH, Weberstrasse 6, 2003.

[53] T. J. Fellers and M. W. Davidson. Hamamatsu, Review Articles, Digital camera readout and frame rates. http://digitalimagingu.com.

[54] H. Föll. Semiconductors I, 2000. http://www.tf.uni-kiel.de.

[55] M. Fornefett, K. Rohr, and H. S. Stiehl. Elastic Registration of Medical Images Using Radial Basis Functions with Compact Support. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 402–7, 1999.

[56] B. A. Fowler, M. Godfrey, J. Balicki, and J. Canfield. Low Noise Readout using Active Reset for CMOS APS. In *SPIE Proc. Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, volume 3965, pages 126–35, 2000.

[57] Fujifilm Life Sciences, West Avenue Stamford, CT 06902 USA.

[58] GE Healthcare Life Sciences, UK, http://www.gelifesciences.com.

[59] Geant4 General Particle Source, http://reat.space.qinetiq.com/gps/.

[60] E. Gladilin, V. Pekar, K. Rohr, and H. S. Stiehl. A comparison between bem and fem for elastic registration of medical images. *Image and Vision Computing*, 24(4):375–379, April 2006.

[61] J. A. Goyette, G. D. Lapin, M. G. Kang, and A. K. Katsaggelos. Regularized iterative image restoration algorithms applied to autoradiography. *Proc. IEEE EMBS*, 15:490–1, 1993.

[62] P. J. Greer, V. L. Villemagne, J. Ruszkiewicz, A. K. Graves, C. C. Meltzer, C. A. Mathis, and J. C. Price. MR atlas of the baboon brain for functional neuroimaging. *Brain Research Bulletin*, 58:429–38, 2002.

[63] T. Grenier, C. Revol-Muller, N. Costes, M. Janier, and G. Gimenez. 3D Robust Adaptive Region Growing for segmenting [^{18}F]fluoride ion PET images. In *Nuclear Science Symposium Conference Record, 2006 IEEE*, volume 5, pages 2644–8, 2006.

[64] R. M. Haralick and L. G. Shapiro. Image Segmentation Techniques. *Computer Vision, Graphics, and Image Processing*, 29:100–32, 1985.

[65] M. Hatt, F. Lamare, N. Boussion, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson, C. Cheze-Le Rest, and D. Visvikis. Fuzzy hidden markov chains segmentation for volume determination and quantitation in PET. *Physics in Medicine and Biology*, 52(12):3467–3491, 2007.

[66] D. J. Hawkes, D. Barratt, J. M. Blackall, C. Chan, P. J. Edwards, K. Rhode, G. P. Penney, J. McClelland, and D. L. G. Hill. Tissue deformation and shape models in image-guided interventions: a discussion paper. *Medical Image Analysis*, 9:163–75, 2005.

[67] L. S. Hibbard, J. S. McGlone, D. W. Davis, and R. A. Hawkins. Three dimensional representation and analysis of brain energy metabolism. *Science*, 236:1641–2, 1987.

[68] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46:1–45, 2001.

[69] S. A. Hojjatoleslami and J. Kittler. Region growing: A new approach. *IEEE Transactions on Image Processing*, 7:1079–84, 1998.

[70] G. R. Hopkinson. Analytic modeling of charge diffusion in charge-coupled-device imagers. *Optical Engineering*, 26(8):766–72, 1987.

[71] W. Hoschek. Colt Project, http://acs.lbl.gov/∼hoschek/colt/.

[72] ICRU. Stopping Powers for Electrons and Positrons. Technical Report 37, International Commision on Radiation Units and Measurements, 1984.

[73] F. James. A review of pseudorandom number generators. *Computer Physics Communications*, 60:329–44, 1990.

[74] J. Janesick. *Scientific Charge Coupled Devices*, volume PM83. SPIE PRESS, Jan 2001.

[75] J. Janesick. Dueling detectors. *Proc. SPIE*, 4669A(45):30–33, February 2002.

[76] J. Janesick. Charge coupled CMOS and hybrid detector arrays. In *Proc. SPIE*, volume 5167, 2003.

[77] J. Janesick. Charge-coupled CMOS and hybrid detector arrays. *SPIE*, 5167:1–18, 2004.

[78] J. Janesick, T. Elliot, S. Collins, T. Daud, D. Campbell, and A. Dingzian. CCD advances for X-ray scientific measurements in 1985. In *SPIE Proc. X-ray Instrumentation in Astronomy*, volume 597, pages 364–80, 1985.

[79] J. Janesick and G. Putnam. Developments and Applications of High-Performance CCD and CMOS Imaging Arrays. *Annual Review of Nuclear and Particle Science*, 53:263–300, Dec 2003.

[80] S. Kanekal, A. Sahai, R. E. Jones, and D. Brown. Storage-phosphor autoradiography: A rapid and highly sensitive method for spatial imaging and quantitation of radioisotopes. *Journal of Pharmacological and Toxicological Methods*, 33:171–178, 1995.

[81] A. Karellas, H. Liu, C. Reindhart, L. J. Harris, and A. B. Brill. Imaging of a radionuclide emissions with a low-noise charge-coupled device. *IEEE Trans. Nuclear Science*, 40:979–82, 1993.

[82] B. Kim, J. L. Boes, K. A. Frey, and C. R. Meyer. Mutual information for automated unwarping of rat brain autoradiographs. *Neuroimage*, 5:31–40, 1997.

[83] J. Kim, W. Cai, D. Feng, and S. Eberl. Segmentation of VOI from multidimensional dynamic PET images by integrating spatial and temporal features. *IEEE Transactions on Information Technology in Biomedicine*, 10:637–46, 2006.

[84] A. P. King, R. Boubertakh, K. L. Ng, Y. L. Ma, P. Chinchapatnam, G. Gao, T. Schaeffter, D. J. Hawkes, R. Razavi, and K. S. Rhode. A technique for respiratory motion correction in image guided cardiac catheterisation procedures. *Medical Imaging 2008: Visualization, Image-guided Procedures, and Modeling*, 6918(1), 2008.

[85] K. Kneöaurek, M. Ivanovic, J. Machac, and D. A. Weber. Medical image registration. *Europhysics News*, 31:4, 2000.

[86] M. Kohno, K. Imanishi, A. Bamba, M. Tsujimoto, H. Murakami, T. G. Tsuru, and K. Koyama. Charge diffusion and loss as a function of absorption depth in X-ray CCD. In *SPIE Proceed. X-Ray and Gamma-Ray Instrumentation for Astronomy XII*, volume 4497, pages 149–57, 2002.

[87] H. Koivistoinen, J. Tohka, and U. Ruotsalainen. Comparison of Pattern Classification Methods in Segmentation of Synamic PET Brain Images. In *Proceedings of the 6th Nordic Signal Processing Symposium (NORSIG 2004)*, pages 73–6, Jun. 9 2004-Jun. 11 2004.

[88] E. Kokkinou, K. Wells, M. Petrou, and A. Bailey. Digital autoradiography imaging using direct irradiation of a CCD between 278-309 K. *IEEE Transactions on Nuclear Science*, 50:1702–7, 2003.

[89] M. Konijnenberg, M. Melis, R. Valkema, E. Krenning, and M. de Jong. Radiation dose distribution in human kidneys by octreotides in peptide receptor radionuclide therapy. *J. Nucl. Med.*, 48:134–42, 2007.

[90] Y. Kvinnsland and A. Skretting. Methods for separation of contributions from two radionuclides in autoradiography with a silicon strip detector. *Physics in Medicine and Biology*, 45:1183–1193, 2000.

[91] A. Lacassagne and J. Lattes. Repartition du polonium (injecte sous la peau) dans l'organisme de rats de greffes cancereuses. *C. r. Seanc Soc. Biol.*, 90:352–3, 1924.

[92] J. S. Lee, J. S. Kim, D. S. Lee, and K. S. Park. Three dimensional reconstruction of rat brain from autoradiographic images using image registration by maximization of mutual information. *Nuclear Science Symposium Conference Record, 2003 IEEE*, 3:1552–3, 2002.

[93] S. Lee, G. Wolberg, and S. Y. Shin. Scattered Data Interpolation with Multilevel B-Splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–44, 1997.

[94] S. C. Lee, H. K. Kim, I. K. Chun, S. Y. Cho, M. H. Lee, and M. H. Cho. A flat-panel detector based micro-CT system: performance evaluation for small-animal imaging. *Physics in Medicine and Biology*, 48:4173–85, 2003.

[95] Y. J. Lee, J. M. Park, and H. W. Park. Mammographic mass detection by adaptive thresholding and region growing. *International Journal of Imaging Systems and Technology*, 11:340–6, 2001.

[96] J. E. Lees and G. W. Fraser. Efficiency enhancements for mcp-based beta autoradiography imaging. *Nuclear Instruments and Methods in Physics Research Section A*, 477:239–243, 2002.

[97] J. E. Lees, G. W. Fraser, and P. Carthew. Large microchannel plate detectors for ^{14}C autoradiography. *Nuclear Instruments and Methods in Physics Research Section A*, 45:349–353, 1998.

[98] J. E. Lees, G. W. Fraser, and P. Carthew. Microchannel plate detectors for ^{14}C autoradiography. *IEEE Transactions on Nuclear Science*, 45(Issue 3, Part 2):1288–92, 1998.

[99] J. E. Lees, G. W. Fraser, and D. Disdale. Direct beta autoradiography using MCP detectors. *Nuclear Instruments and Methods in Physics Research Section A*, 392:349–353, 1997.

[100] J. E. Lees, A. Murray, A. C. Perkins, and G. W. Fraser. Autoradiography of High-Energy radionuclides using a microchannel plate detector. *IEEE Transactions on Nuclear Science*, 49:153–155, 2002.

[101] J. E. Lees, J. F. Pearson, G. W. Fraser, J. M. Hales, and P. G. Richards. An MCP-based system for beta autoradiography. *IEEE Transactions on Nuclear Science*, 46:636–638, 1999.

[102] H. Li, W.L. Thorstad, K.J. Biehl, R. Laforest, Y. Su, K.I. Shoghi, E.D. Donnelly, D.A. Low, and W. Lu. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. *Medical Physics*, 35(5):3711–21, Aug 2008.

[103] N. Liepins and J. Sekino. Application of cubic splines to contour plotting. *Mathematical Magazine*, 63:343–345, Dec 1990.

[104] J. A. Little, D. L. G. Hill, and D. J. Hawkes. Deformations incorporating rigid structures. *Computer Vision and Image Understanding*, 66:223–232, 1997.

[105] X. V. Liu, B. A. Fowler, S. K. Onishi, David P. V., D. Wen, H. Do, and S. Horna. CCD / CMOS Hybrid FPA for Low Light Level Imaging. In *SPIE Proc. Infrared and photoelectronic imagers and detector devices*, volume 5881, pages 58840C.1–58810C.9, 2005.

[106] K. Ljunggren and S. E. Strand. Beta camera for static and dynamic imaging of charged-particle emitting radionuclides in biologic samples. *Journal of Nuclear Medicine*, 31:2058–2063, 1990.

[107] K. Ljunggren and S. E. Strand. Reduction of noise in the beta camera for low activity applications. *IEEE Transactions on Nuclear Science*, 41:1666–1669, 1994.

[108] S. J. Lockett, J. E. Bateman, A. R. Bradwell, and D.B. Ramsden. The enhancement of radiographic images from a multiwire camera using a maximum entropy algorithm. *Physics in Medicine and Biology*, 33(12):1373–83, 1988.

[109] S. J. Lockett, A. R. Bradwell, J. E. Bateman, M. R. Hawkesworth, J. R. Taylor, S. Maudsley, C. A. Hanham, and D. B. Ramsden. Preliminary assessment of a multiwire camera for quantitative autoradiography of tritium-labelled substances. *Physics in Medicine and Biology*, 31(5):535–42, 1986.

[110] W. Lu and T. R. Mackie. Tomographic motion detection and correction directly in sinogram space. *Physics in Medicine and Biology*, 47:1267–84, 2002.

[111] Y. Lu, T. Jiang, and Y. Zang. Region growing method for the analysis of functional MRI data. *Neuroimage*, 20:455–65, 2003.

[112] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2:1–37, 1998.

[113] G. Malandain, E. Bardinet, K. Nelissen, and W. Vanduffel. Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations. *Neuroimage*, 23:111–27, 2004.

[114] M. Mancas, B. Gosselin, and B. Macq. Segmentation using a region-growing thresholding. In E. R. Dougherty, J. T. Astola, and K. O. Egiazarian, editors, *Image Processing: Algorithms and Systems IV. Edited by Dougherty, Edward R.; Astola, Jaakko T.; Egiazarian, Karen O. Proceedings of the SPIE, Volume 5672, pp. 388-398 (2005).*, volume 5672 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 388–398, March 2005.

[115] K. J. McCarthy, A. Owens, A. D. Holland, and A. A. Wells. Modelling the X-ray response of charge coupled devices. *Nuclear Instruments and Methods in Physics Research Section A*, 362:538–46, 1995.

[116] C. McFee. UCL Department of Space & Climate Physics, Detector Physics Group, Noise sources in a CCD. http://www.mssl.ucl.ac.uk/.

[117] G. McMullan, D. M. Cattermole, S. Chen, R. Henderson, X. Llopart, C. Summerfield, L. Tlustos, and A. R. Faruqi. Electron imaging with medipix2 hybrid pixel detector. *Ultramicroscopy*, 107:401–13, 2007.

[118] G. Mettivier, M. C. Montesi, and P. Russo. First images of a digital autoradiography system based on a a medipix2 hybrid silicon pixel detector. *Physics in Medicine and Biology*, 48:173–81, 2003.

[119] G. Mettivier, M. C. Montesi, and P. Russo. Tritium Digital Autoradiography with a Medipix2 Hybrid Silicon Pixel Detector. *Nuclear Instruments and Methods in Physics Research Section A*, 516:554–563, 2004.

[120] G. Mettivier, M. C. Montesi, and P. Russo. Digital Autoradiography with a Medipix2 Hybrid Silicon Pixel Detector. *IEEE Transactions on Nuclear Science*, 52:46–50, 2005.

[121] A. C. Milazzo, P. Leblanc, F. Duttweiler, L. Jin, J. C. Bouwer, S. Peltier, M. Ellisman, F. Bieser, H. S. Matis, H. Wieman, P. Denes, S. Kleinfelder, and N. H. Xuonga. Active pixel sensor array as a detector for electron microscopy. *Ultramicroscopy*, 104:152–9, 2005.

[122] J. Mohr, A. Hess, M. Scholza, and K. Obermayer. A method for the automatic segmentation of autoradiographic image stacks and spatial normalization of functional cortical activity patterns. *Journal of Neuroscience Methods*, 134:45–58, 2004.

[123] Multidimensional Integrated Intelligent Imaging Consortium, http://mi3.shef.ac.uk/.

[124] R. Myers. The application of PETMR image registration in the brain. *British Journal of Radiology*, 75:S31–S35, 2002.

[125] W. Neeser, M. Böcker, P. Buchholz, P. Fischer, P. Holl, J. Kemmer, P. Klein, H. Koch, M. Löcker, G. Lutz, H. Matthäy, L. Strüder, M. Trimpl, J. Ulrici, and N. Wermes. The DEPFET pixel BIOSCOPE. *IEEE Transactions on Nuclear Science*, 47:1246–1250, 2000.

[126] NIST. National Institute of Standards and Technology.

[127] Nobel Prize Web, http://www.nobelprize.org.

[128] A. Nourreddine, A. Nachab, D. Husson, and S. Higueret. MCNPX simulation for α-particle detection by CMOS active pixel sensor. *Radiation Measurements*, 40(2-6):275–8, 2005.

[129] A. Orbom, M. Dahlbom, T. Olafsen, A. M. Wu, and S. Strand. Serial Digital Autoradiography with a Silicon Strip Detector as a High Resolution Imaging Modality for TRT Dosimetry. In *Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE*, volume 6, pages 4054–6, Nov 2007.

[130] R. J. Ott, J. MacDonald, and K. Wells. The performance of a CCD digital autoradiography imaging system. *Physics in Medicine and Biology*, 45(7):2011–2027, Jul 2000.

[131] M. Overdick, A. Czermak, P. Fischer, V. Herzog, A. Kjensmo, T. Kugelmeier, K. Ljunggren, E. Nygård, C. Pietrzik, T. Schwan, S. E. Strand, J. Straver, P. Weilhammer, N. Wermes, and K. Yoshioka. A bioscope system using double-sided silicon strip detectors and self-triggering read-out chips. *Nuclear Instruments and Methods in Physics Research Section A*, A392:173–177, 1997.

[132] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26:1277–94, 1993.

[133] Z. Pan and J. Lu. A bayes-based region-growing algorithm for medical image segmentation. *Computing in Science and Engineering*, 9:32–8, 2007.

[134] G. G. Pavlov and J. A. Nousek. Charge diffusion in CCD X-ray detectors. *Nuclear Instruments and Methods in Physics Research Section A*, 428:348–66, 1999.

[135] G. Paxinos and K. B. J. Franklin. *The mouse brain in stereotaxic coordinates.* Academic Press, 2001.

[136] P. Perona and J. Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–39, 1990.

[137] Y. Petegnief, A. Petiet, M. C. Peker, F. Bonnin, A. Meulemans, and D. Guludec. Quantitative autoradiography using a radioimager based on a multiwire proportional chamber. *Physics in Medicine and Biology*, 43:3629–3638, 1998.

[138] T. E. Peterson, D. W. Wilson, and Barrett H. H. Application of silicon strip detectors to small-animal imaging. *Nuclear Instruments and Methods in Physics Research Section A*, 505:608–11, 2001.

[139] D. L. Pham, X. Chenyiang, and L. P. Jerry. A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.

[140] Andor Technology PLC. CCD Sensor Architectures, 2008. http://www.andor.com/learn/digitalcameras.

[141] J. Plusquellic and C. Patel. Principles of VLSI Design/VLSI Systems, 2000. http://www.cs.umbc.edu/CSEE/index.html.

[142] R. Pohle and K. D. Toennies. Segmentation of medical images using adaptive region growing. *Proc. SPIE Medical Imaging*, 4322:1337–46, 2001.

[143] G. Prigozhin, N. R. Butler, S. E. Kissel, and G. R. Ricker. An Experimental Study of Charge Diffusion in the Undepleted Silicon of X-ray CCDs. *IEEE Transactions on Electron Devices*, 50(1):246–53, Jan 2003.

[144] M. L. Prydderch, N. J. Waltham, R. Turchetta, M. J. French, R. Holt, A. Marshall, D. Burt, Bell. R., P. Pool, Eyles. C., and H. Mapson-Menard. A 512x512 CMOS Monolithic Active Pixel Sensor with integrated ADCs for space science. *Nuclear Instruments and Methods in Physics Research Section A*, 512(1-2):358–67, 2003.

[145] D. Puertolas, H. Piedigrossi, D. Leutz, T. Gys, and C. D'Ambrosio. An ISPA camera for beta autoradiography. *IEEE Transactions on Nuclear Science*, 43:2477–2487, 1996.

[146] D. Rangaraj, G. Palaniswaamy, and L. Papiez. DMLC IMRT delivery to targets moving in 2D in Beam's eye view. *Medical Physics*, 35(8):3765–78, 2008.

[147] L. Reuen, R. Kohrs, J. J. Velthuis, L. Andricek, P. Fischer, F. Giesen, H. Krüger, G. Lutz, M. Mathes, H. G. Moser, I. Peric, R. H. Richter, C. Sandow, E. von Törne, M. Trimpl, J. Treis, and N. Wermes. Performance of a DEPFET Prototype Module for the ILC Vertex Detector. *IEEE Transactions on Nuclear Science*, 53(3):1719–25, Jun 2006.

[148] S. A. Rodney and J. L. Tonry. Characterizing Charge Diffusion in CCDs with X-Rays. *The Publications of the Astronomical Society of the Pacific*, 118:866–73, 2006.

[149] D. W. O. Rogers and A. F. Bielajew. The Use of EGS for Monte Carlo Calculations in Medical Physics. *Nuclear Instruments and Methods in Physics Research Section A*, 100:31–46, 1995.

[150] K. Rohr, H. S. Stiehl, R. Sprengel, J. Buzug, T. M. Weese, and M. H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging*, 20:526–34, 2001.

[151] W. Rui and L. Minglu. An overview of medical image registration. In *ICCIMA '03: Proceedings of the 5th International Conference on Computational Intelligence and Multimedia Applications*, page 385, Washington, DC, USA, 2003. IEEE Computer Society.

[152] A. Saad, B. Smith, G. Hamarneh, and T. Moller. Simultaneous segmentation, kinetic parameter estimation, and uncertainty visualization of dynamic PET images. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 10(Pt 2):726–33, 2007.

[153] D. Scheffer. A Wafer scale Active Pixel CMOS Image Sensor for generic X-ray radiology. In Jiang Hsieh and Michael J. Flynn, editors, *Medical Imaging 2007: Physics of Medical Imaging*, volume 6510, page 65100O. SPIE, 2007.

[154] D.K. Schroder, B.D. Choi, S.G. Kang, W. Ohashi, K. Kitahara, G. Opposits, T. Pavelka, and J. Benton. Silicon epitaxial layer recombination and generation lifetime characterization. *IEEE Transactions on Electron Devices*, 50(4):906–12, Apr 2003.

[155] B. Shangera and R. J. Ott. Preliminary studies using silicon strip detectors in digital autoradiography. *IEEE Transactions on Nuclear Science*, 40(4):992–4, 1993.

[156] D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, pages 517–524, New York, NY, USA, 1968. ACM.

[157] P. M. Shikhaliev, T. Xu, J. L. Ducote, B. Easwaramoorthy, J. Mukherjee, and S. Molloi. Positron autoradiography for intravascular imaging: feasibility evaluation. *Phys. Med. Biol.*, 51:963–79, 2006.

[158] Silicon Strip Detectors in the ATLAS Experiment. http://hepwww.rl.ac.uk/OpenDays98/Detectors/silicon.htm.

[159] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26. Chapman & Hall/CRC, 1986.

[160] X. Song, B. W. Pogue, S. Jiang, M. M. Doyley, H. Dehghani, T. D. Tosteson, and K. D. Paulsen. Automated region detection based on the contrast-to-noise ratio in near-infrared tomography. *Applied Optics*, 43(5):1053–62, Feb. 2004.

[161] M. Sonoda, M. Takano, J. Miyahara, and H. Kato. Report on the development of a CR system using an imaging plate. *Radiology*, 148:833, 1983.

[162] R. J. Stover, M. Wei, Y. Lee, D. K. Gilmore, S. E. Holland, D. E. Groom, W. W. Moses, S. Perlmutter, G. Goldhaber, C. Pennypacker, N. W. Wang, and N. Palaio. Characterization of a Fully Depleted CCD on High Resistivity Silicon. In *Proceedings of The SPIE Conference on Solid State Sensor Arrays: Development and Applications*, volume 3019, pages 183–8, 1997.

[163] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):25–35, 1997.

[164] Walter E. Stumpf. *Drug Localization in Tissues and Cells. Receptor Microscopic Autoradiography.* IDDC Press, Chapel Hill, NC, 2003.

[165] D. W. Townsend. Dual-Modality Imaging: Combining Anatomy and Function. *Journal of Nuclear Medicine*, 49(6):938–55, 2008.

[166] D. W. Townsend and T. Beyer. A combined PET/CT scanner: the path to true image fusion. *The British Journal of Radiology*, 75:S24–S30, 2002.

[167] L. K. Townsley, P. S. Broos, G. Chartas, E. Moskalenko, J. A. Nousek, and G. G. Pavlov. Simulating CCDs for the Chandra Advanced CCD Imaging Spectrometer. *Nuclear Instruments and Methods in Physics Research A*, 486:716–50, Nov 2002.

[168] E. Tribollet, J. J. Dreiffus, G. Charpak, W. Dominik, and N. Zaganidis. Localization and quantitation of tritiated compounds in tissue sections with a gaseous detector of β particles: Comparison with film autoradiography. *Proc. Natl. Acad. Sci.*, 88:1466–8, 1991.

[169] B. M. W. Tsui, W. P. Segars, and D. S. Lalush. Effects of Upward Creep and Respiratory Motion in Myocardial SPECT. *IEEE Trans. Nuclear Science*, 47(3):1192–5, 2000.

[170] Y. Tuduki, K. Murase, M. Izumida, H. Miki, K. Kikuchi, K. Murakami, and J. Ikezoe. Automated Seeded Region Growing Algorithm for Extraction of Cerebral Blood Vessels from Magnetic Resonance Angiographic Data. In *Annual - Proceedings of International Conference of the IEEE Engineering in Medicine and Biology*, volume 3, pages 1756–9, 2000.

[171] R. Turchetta, J. D. Berst, B. Casadei, G. Claus, C. Colledani, W. Dulinski, Y. Hu, D. Husson, J. P. Le Normand, J. L. Riester, G. Deptuch, U. Goerlach, S. Hiwueret, and M. Winter. A monolithic active pixel sensor for charged particle tracking and imaging using standard VLSI CMOS technology. *Nucl. Instr. and Meth. A*, 458:677–89, 2001.

[172] R. Turchetta, M. French, S. Manolopoulos, M. Tyndel, P. Allport, R. Bates, V. O'shea, G. Hall, and M. Raymond. Monolithic active pixel sensors (MAPS) in a VLSI CMOS technology. *Nucl. Instr. and Meth. A*, A501:251–9, 2003.

[173] J. Ulrici, P. Fischer, P. Klein, G. Lutz, W. Neeser, L. Richter, R. Strüder, M. Trimpl, and N. Wermes. Imaging performance of a DEPFET pixel Boiscope system in tritium autoradiography. *Nuclear Instruments and Methods in Physics Research Section A*, 547:424–36, 2005.

[174] P. A. van den Elsen, E. J. D. Pol, and M. A. Viergever. Medical image matching-A review with classification. *IEEE Engineering in medicine and biology*, 12:26–39, 1993.

[175] B. Van Zeghbroeck and M. Silver. *Principles of Semiconductor Devices and Heterojunctions.* Prentice-Hall, 2007.

[176] E. G. Villani. Charge diffusion in undepleteed regions of silicon particle detectors: analysis and simulation. *Nuclear Instruments and Methods in Physics Research A*, 539:125–31, 2005.

[177] E. G. Villani, R. Turchetta, and M. Tyndel. Analysis and simulation of charge collection in. Monolithic Active Pixel Sensors (MAPS). *Nuclear Physics B*, 125:184–8, 2003.

[178] K. Wells and E. Kokkinou. UK patent GB2397959: Imaging system for use in Autoradiography, 2006.

[179] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Med Image Anal*, 1(1):35–51, Mar 1996.

[180] S. M. Wong, Y. C. Hon, and M. A. Golberg. Compactly supported radial basis functions for the shallow water equations. *Appl. Math. Comput*, 127(1):79–101, 2002.

[181] R. P. Woods, J. C. Mazziotta, and Cherry S. R. MRI-PET registration with automated algorithm. *Journal of computer assisted tomography*, 17:536–46, 1993.

[182] Y. Yamane, N. Ishide, Y. Kagaya, D. Takeyama, N. Shiba, M. Chida, Y. Sekiguchi, T. Nozaki, T. Ido, and K. Shirato. Quantitative double-tracer autoradiography with tritium and carbon-14 using imaging plates: Application to myocardial metabolic studies in rats. *The Journal of Nuclear Medicine*, 36:518–24, 1995.

[183] J. Yi and J. B. Ra. A locally adaptive region growing algorithm for vascular segmentation. *International Journal of Imaging Systems and Technology*, 13:208–14, 2003.

[184] L. Yu-Bu, S. Soo-Min, L. Jae-Sung, and K. Myoung-Hee. Tumor segmentation from small animal PET using region growing based on gradient magnitude. In *IEEE International Workshop on Enterprise Networking and Computing in Healthcare Industry (Healthcom 2005)*, pages 243–7, Jun 2005.

[185] Q. Zhang, Y. Li, B. Steele, X. Wu, W. R. Chen, J. Rong, and H. Liu. Comparison of a CMOS-based and a CCD-based digital x-ray imaging system: Observer studies. *Journal of Electronic Imaging*, 14:3002–8, 2005.

[186] W. Zhao, T. Y. Young, and M. D. Ginsberg. Registration and three-dimensional reconstruction of autoradiographic images by the disparity analysis methods. *IEEE Trans. Medical Imaging*, 12:782–91, 1993.

[187] W. Zhu and T. Jiang. Automation segmentation of PET image for brain tumors. *Nuclear Science Symposium Conference Record, 2003 IEEE*, 4:2627–2629 Vol.4, 19-25 Oct. 2003.

[188] J. M. Zuo. Electron Detection Characteristics of a Slow-Scan CCD Camera, Imaging Plates and Film, and Electron Image Restoration. *Microscopy Research and Technique*, 49:245–68, 2000.