CMOS Active Pixel Sensors in Bio-Medical Imaging

Michela Esposito

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

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

May 2015

© Michela Esposito 2015
Summary

Bio-medical imaging is a large umbrella term which covers a number of different imaging modalities used in healthcare today, spanning pre-clinical imaging, to diagnostic imaging and imaging to assist and plan patient treatment. This field of research is pivotal to driving advances in healthcare. This is underpinned by advances in new detector technologies which have the potential to reduce image acquisition time and dose, improving image quality and offer more accurate tools for diagnosis and treatment.

Large area CMOS Active Pixel Sensors (APSs) have the potential to deliver these advances in such demanding and continuously evolving field; large imaging area, together with low noise, low cost, fast readout, high dynamic range and potential for in-pixel intelligence have made this technology an ideal candidate to displace currently used imaging technologies in this field.

This thesis represents the first investigation into the capabilities of large area CMOS APSs to be used across a number of different imaging modalities in bio-medical science, spanning protein imaging to proton Computed Tomography (CT), using both ionising and non-ionising radiation sources. A novel characterisation of the detector performance has been carried out and set into context of commonly used detectors for bio-medical imaging. Considering the performance parameters assessed for this detector, in comparison with digital detectors commonly used in the clinical practise, this demonstrates how such large area sensor technology may be successfully employed in bio-medical imaging.

The novel large area CMOS APS, studied in this work, is proposed as a multi-modality imaging platform for use in pre-clinical science. For the first time direct “contact print” imaging of radioactive and optical labeled biological samples on a large imaging area has been demonstrated, showing its potential application to a broad range of ionising and non-ionising imaging probes. The protein detection capability of this detector has been compared with both film emulsion and commercially available digital systems, demonstrating a higher resolution in protein detection than either film emulsion or a commonly used commercial CCD-based western blotting detection system. Also, when detection capabilities of this imaging system are compared with the state-of-the art devices for tissue autoradiography, this detector system exhibits a sensitivity comparable to that reported for its competitors, whilst offering the largest imaging area. Both these proof of concepts pave the way for large area CMOS APSs to be used as a multi-modality imaging platform in life science.

The radiation hardness of a novel large area CMOS APS, designed for medical applications and hardened-by-design, is presented. The radiation damage, produced in this sensor by X-ray and proton irradiation, has been studied as function of total ionising dose and displacement damage dose. The damage contributions from ionising and non-ionising energy deposition have been separated for the proton field and proved independent from proton energy providing a further verification of the Non Ionising Energy Loss (NIEL) scaling hypothesis. The lifetime of this detector for routine use in
clinical practice has been evaluated as high as 4 years when used in a typical MegaVolt-age radiotherapy environment, demonstrating how such large area sensor technology may be successfully employed in X-ray and proton based imaging applications.

The feasibility of using CMOS APSs as energy-range detectors in proton CT has been demonstrated. Capability of single proton counting, together with potential of energy deposition measurements, have been demonstrated for CMOS APSs. Furthermore, experimental work, based on a simple stack of two CMOS sensors, as well as simulation work has been carried out to prove the capability of such a detection system for proton tracking. Novel algorithms have been developed to perform proton tracking in a CMOS energy-range telescope designed to perform proton CT, paving the way for a new generation of imaging devices to be used in this application.

Key words: CMOS APS, Wafer Scale Sensors, Bio-Medical Imaging, Autoradiography, Western Blotting, Chemiluminescence, Radiology, Mammography, Radiation Hardness, Monte Carlo Simulations, Charge Transport, Proton Therapy, Proton CT

Email: m.esposito@surrey.ac.uk - m.esposito@physics.org
WWW: http://www.eps.surrey.ac.uk/
Acknowledgements

First of all, I would like to thank my supervisor Dr Kevin Wells for his continued support, advice and profitable discussions over the course of this work. I would also like to thank my second supervisor Prof. Philip M. Evans for his guidance on proton CT, and Prof. Nigel M. Allinson (University of Lincoln) for his advice on many occasions. Many thanks to Dr Thalis Anaxagoras (ISDI Ltd) for introducing me to CMOS sensors and for his continuous advice.

In addition, I would like to thank the members of the MI3+ and PRaVDA consortia, especially Dr Gavin Poludniowski (Karolinska University Hospital), Dr Tony Price (University of Birmingham), Dr Jon Taylor (University of Liverpool), Mr Chris Waltham (University of Lincoln) and Dr Grainne Riley (University of Lincoln) for their support and collaboration. Also, many thanks to present and past members of the CVSSP Medical Imaging Group for all the useful discussions, especially to Dr. Oliver Diaz and Dr. Mary Yip.

I would like to thank Dr Alexis Bailey and Dr Jane Newcombe (University of Surrey) for providing the biological samples used in this work. My thanks also extend to aSpect Systems GmbH and ISDI Limited for their support and development of the PRaVDA system, and to the operators and physicists at the iThemba LABS and at the University of Birmingham for producing and maintaining the proton beams that were used for the measurements presented in this work.

Finally, a personal thanks to my husband Roberto for his continuos support and encouragement.
Contents

List of Abbreviations 1

1 Introduction 5
  1.1 Imaging in biology and medicine 7
  1.2 Challenges in bio-medical imaging and motivations 10
  1.3 Structure of the thesis 13
  1.4 Achievements and major contribution 15

2 The DynAMITe CMOS Active Pixel Sensor: design overview and proof of concept demonstration 17
  2.1 CMOS APSs fundamentals 18
     2.1.1 Readout modalities 20
     2.1.2 Large area CMOS APSs 21
  2.2 The DynAMITe detector 23
     2.2.1 Pixel design 25
     2.2.2 Readout 25
     2.2.3 Stitching 27
  2.3 Proof of concept 29
     2.3.1 Charge collection 29
     2.3.2 Non Destructive Readout 33
  2.4 Discussion 34

3 Characterisation of the DynAMITe CMOS APS 35
  3.1 Motivation 35
  3.2 Methodology for detector performance evaluation 37
### 3.2 Photon Transfer curve
- 3.2.1 Photon Transfer curve .............................................. 37
- 3.2.2 Per-pixel Photon Transfer Curve ................................. 41
- 3.2.3 Fixed Pattern Noise ................................................. 41
- 3.2.4 Contrast-to-Noise Ratio ............................................ 42
- 3.2.5 Detective Quantum Efficiency ..................................... 43

### 3.3 Detector performance
- 3.3.1 Photon Transfer Curve .............................................. 45
- 3.3.2 Per-pixel Photon Transfer Curve ................................. 47
- 3.3.3 Photon Transfer Curve in stitching blocks ......................... 49
- 3.3.4 Fixed Pattern Noise ................................................. 50
- 3.3.5 Contrast-to-Noise Ratio ............................................ 55
- 3.3.6 Detective Quantum Efficiency ..................................... 59

### 3.4 Discussion .................................................................. 61

### 4 Large area CMOS APSs for Western Blotting ................. 63

#### 4.1 Western Blotting ................................................. 63
- 4.1.1 Protein sequencing and immuno-detection ........................ 64
- 4.1.2 ECL detection in western blotting ................................. 66

#### 4.2 Motivations .......................................................... 68

#### 4.3 Proof of principle .................................................... 69
- 4.3.1 Biological samples and experimental setup ....................... 69
- 4.3.2 Decay curve .......................................................... 70
- 4.3.3 ECL imaging .......................................................... 71
- 4.3.4 Linear range, sensitivity and SNR ................................. 72
- 4.3.5 Comparison with film-based images ............................... 73

#### 4.4 Chemiluminescence detection using a large area CMOS APS 75
- 4.4.1 Linearity and Quantum Efficiency ................................. 76
- 4.4.2 Biological samples and experimental set-up ....................... 77
- 4.4.3 Chemiluminescence detection ..................................... 78

#### 4.5 First images of protein sequences ................................ 81
- 4.5.1 Biological samples and experimental set-up ....................... 81
- 4.5.2 Imaging of immuno-detected proteins ............................. 82
## 4.6 Comparative performance assessment

4.6.1 Biological samples and experimental set-up

4.6.2 Detection capability

## 4.7 Discussion

## 5 Radiation Hardness assessment

5.1 Motivation

5.2 Radiation damage in Silicon sensors

5.2.1 Ionising damage

5.2.2 Non-ionising damage

5.2.3 NIEL scaling hypothesis

5.2.4 The radiation hardness design of the DynAMITe detector

5.3 Irradiation experiments

5.3.1 X-ray irradiation

5.3.2 Proton irradiation

5.3.3 Dosimetry

5.4 X-ray irradiation

5.4.1 Comparison with a commercial CMOS imager

5.4.2 Dark current

5.4.3 Fluorescence

5.5 Proton irradiation

5.5.1 Electro-optical parameters

5.6 Ionisation versus displacement damage effects

5.6.1 Dark current

5.6.2 Dark current non-uniformity

5.6.3 Conversion gain

5.7 Detector lifetime

5.8 Discussion
6 Large area CMOS APSs for proton Computed Tomography 123

6.1 Imaging in Proton Therapy ........................................... 123

6.1.1 Proton CT ............................................................ 126

6.1.2 The PRaVDA pCT ...................................................... 130

6.2 Motivations ............................................................... 130

6.3 Materials and Methods .................................................. 131

6.3.1 Proton facilities ....................................................... 131

6.3.2 Double DynAMITe detector ....................................... 132

6.3.3 Experimental parameters ......................................... 135

6.3.4 Image processing ..................................................... 138

6.4 Proton counting .......................................................... 138

6.4.1 Proton counting at UoB ............................................. 139

6.4.2 Proton counting at iThemba ...................................... 142

6.4.3 Time structure ......................................................... 144

6.5 Correlation ............................................................... 145

6.5.1 Correlation in uncollimated beam experiments ................. 148

6.5.2 Correlation in collimated beam experiments .................. 151

6.5.3 Background radiation and SNR ................................... 153

6.5.4 Correlation in the collimated 30 mm range beam experiments 156

6.6 Proton tracking .......................................................... 158

6.6.1 Tracking algorithm for the Double DynAMITe experiments 161

6.6.2 Simulations of proton displacement ............................ 162

6.6.3 MDA applied to simulated data ................................. 164

Effects of noise and detector misalignment on the reconstruction of simulated data ................................................. 166

6.6.4 MDA applied to the experimental data .......................... 168

6.7 Proton tracking in the PRaVDa Range Telescope ................... 171

6.7.1 Tracking: the reconstruction algorithm ........................ 172

6.7.2 Monte Carlo simulations ............................................ 173

6.7.3 Tracking efficiency .................................................. 175

6.8 Discussion ................................................................ 176
## 7 Conclusions and Further Work

7.1 Conclusions ........................................ 180

7.2 Further work ...................................... 184

### A Digital autoradiography with a large area CMOS APS

A.1 Dark correction .................................... 189

A.2 Signal analysis and clustering techniques ............... 191

A.3 Linearity, sensitivity and Minimum Detectable Activity ... 192
  
  A.3.1 Sensitivity, background and Minimum Detectable Activity . . . 192

A.4 Tissue imaging ..................................... 196
  
  A.4.1 Imaging of $^{14}\text{C}$-sulphur labelled tissue sections .... 196

A.4.2 Imaging of $^{125}\text{I}$Epibatidine labeled brain sections .... 199

A.5 Discussion ......................................... 200

### B Multiple Coulomb Scattering

B.1 Rossi-Greisen equation ................................ 204

B.2 Molière theory ...................................... 207

### C GEANT4 simulations

C.1 Cross section validation ............................. 211
  
  C.1.1 X-rays ........................................... 212

  C.1.2 Protons .......................................... 214

C.2 Simulations of detector response ....................... 214
  
  C.2.1 Charge diffusion ................................ 216

  C.2.2 Charge collection ................................ 218

  C.2.3 Charge sharing and digitalisation ............... 220

  C.2.4 Validation ....................................... 221

### D List of publications

References ............................................ 229
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS</td>
<td>Active Pixel Sensor</td>
</tr>
<tr>
<td>AR</td>
<td>Autoradiography</td>
</tr>
<tr>
<td>C-C FPN</td>
<td>Column-to-Column Fixed Pattern Noise</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge-Couple Device</td>
</tr>
<tr>
<td>CDS</td>
<td>Correlated Double Sampling</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal-Oxide-Semiconductor</td>
</tr>
<tr>
<td>CNR</td>
<td>Contrast-to-Noise ratio</td>
</tr>
<tr>
<td>COV</td>
<td>Coefficient of Variation</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>DD</td>
<td>Double DynAMITe</td>
</tr>
<tr>
<td>DDD</td>
<td>Displacement Damage Dose</td>
</tr>
<tr>
<td>DQE</td>
<td>Detective Quantum Efficiency</td>
</tr>
<tr>
<td>ECL</td>
<td>Enhanced Chemiluminescence</td>
</tr>
<tr>
<td>ELG</td>
<td>Enclosed Layout Geometry</td>
</tr>
<tr>
<td>EPID</td>
<td>Electron Portal Imaging Detector</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>FOP</td>
<td>Fiber Optic Plate</td>
</tr>
<tr>
<td>FP</td>
<td>Frenkel Pair</td>
</tr>
<tr>
<td>FPI</td>
<td>Flat Panel Imager</td>
</tr>
<tr>
<td>FPN</td>
<td>Fixed Pattern Noise</td>
</tr>
<tr>
<td>FWC</td>
<td>Full Well Capacity</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full Width at Half Maximum</td>
</tr>
<tr>
<td>HBD</td>
<td>Hardness-By-Design</td>
</tr>
<tr>
<td>HRP</td>
<td>HorseRadish Peroxidase</td>
</tr>
<tr>
<td>INL</td>
<td>Integral Non-Linearity</td>
</tr>
<tr>
<td>KERMA</td>
<td>Kinetic Energy Released to the Matter</td>
</tr>
<tr>
<td>LET</td>
<td>Linear Energy Transfer</td>
</tr>
<tr>
<td>MCP</td>
<td>Micro-Channel Plates</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple Coulomb Scattering</td>
</tr>
<tr>
<td>MDA</td>
<td>Minimum Detectable Activity</td>
</tr>
<tr>
<td>MDA</td>
<td>Minimum Displacement Algorithm</td>
</tr>
<tr>
<td>MV</td>
<td>MegaVolatge</td>
</tr>
<tr>
<td>MWPC</td>
<td>Multi-wire Proportional Chamber</td>
</tr>
<tr>
<td>NDR</td>
<td>Non Destructive Readout</td>
</tr>
<tr>
<td>NIEL</td>
<td>Non Ionising Energy Loss</td>
</tr>
<tr>
<td>P</td>
<td>Pixel</td>
</tr>
<tr>
<td>P-P FPN</td>
<td>Pixel-to-Pixel Fixed Pattern Noise</td>
</tr>
<tr>
<td>pCT</td>
<td>proton Computed Tomography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
</tr>
<tr>
<td>PKA</td>
<td>Primary knock-on atom</td>
</tr>
<tr>
<td>PTC</td>
<td>Photo Transfer Curve</td>
</tr>
<tr>
<td>QA</td>
<td>Quality Assurance</td>
</tr>
<tr>
<td>QE</td>
<td>Quantum Efficiency</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>RT</td>
<td>Range Telescope</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium Dodecyl Sulfate-Polyacrilamide Gel Electrophoresis</td>
</tr>
<tr>
<td>SiPM</td>
<td>Silicon Photo-Multiplier</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SP</td>
<td>Sub-Pixel</td>
</tr>
<tr>
<td>SSD</td>
<td>Silicon Strip Detector</td>
</tr>
<tr>
<td>STI</td>
<td>Shallow Trench Isolation</td>
</tr>
<tr>
<td>TID</td>
<td>Total Ionising Dose</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Bio-medical imaging is a large umbrella term which covers a number of different imaging modalities spanning pre-clinical imaging, to diagnostic imaging and imaging to assist and plan patient treatment. It is usually not simple to define a precise moment in time when a new branch of science and technology is born, as this is commonly the outcome of a number of concatenated discoveries and events. However, this is not the case for bio-medical imaging.

It was November the 8th 1895 and the German physicist Wilhelm Conrad Röntgen, Professor and Head of the Department of Physics at the Julius-Maximillan University in Würzburg, was experimenting with cathode rays using evacuated glass bulbs. Professor Röntgen noted that when a current was passed across the bulb, a barium platinocyanide screen was seen to fluoresce. Simultaneously he noted the effect of the new phenomenon on photographic plates. Three days before Christmas of that year he brought his wife into his laboratory and they emerged with a photograph of the bones in her hand and the ring on her finger (see Figure 1.1a). A few days later, Röntgen submitted his findings to the Würzburg Physical Medical Institute [1], paper which later appeared in an English translation on Nature with the title “On a new kind of rays” [2].

The importance of this discovery was immediately recognised by contemporary scientists as well as the general public, with public lectures and demonstrations, as shown
Chapter 1. Introduction

Figure 1.1: a) First radiograph produced in 1895: the hand of Frau Röntgen [2]. b) A flyer inviting to a public demonstration of X-ray photography distributed at an 1896 exhibition at Crystal Palace, London [3].

in the flyer of Figure 1.1b, distributed at an 1896 exhibition in Crystal Palace, London. Professor Röntgen was then awarded the very first Nobel Prize in Physics in 1901. Cathode rays, or X-rays as Röntgen named them “for the sake of brevity” [2], were discovered and bio-medical imaging and the specialist area of radiology was born.

However, more than a century has passed from these early experiments for bio-medical imaging to become the science that we know today. A significant breakthrough in bio-medical imaging was represented by advances in electronics and computational resources available in the second half of the 20th century. It followed from the important developments in experimental nuclear physics of the 20th century, made possible by a number of advances in electronics such as the invention of electronic valves (by the Philips company in 1917), transistors (developed by the Americans John Bardeen and Walter H. Brattain in 1948) as well as integrated circuits (developed by a group of researchers at Intel under the direction of Ted Hoff in 1969), together with the theoretical work of Alan Turing on a programmable calculating machine [4], which led to the invention of the first computer. With such experimental and computational resources
now available, bio-medical imaging evolved into the myriad of different activities that we know today.

1.1 Imaging in biology and medicine

Modern bio-medical imaging is a large branch of science spanning the breadth of the entire healthcare pathway (Figure 1.2). This has inevitably led to further specialism with different imaging modalities for pre-clinical and clinical imaging.

Pre-clinical science utilizes structural, functional, metabolic and molecular probes for a wide range of applications including characterising disease phenotypes and molecular mechanisms, evaluating drug efficacy and identifying potential bio-markers, in order to drive advances in healthcare and fundamental bioscience. Such activity relies on a wide range of techniques requiring different probe detection methods, such as $\beta^-$ or $\gamma$ autoradiography of specifically labeled tissue, *ex-vivo* positron autoradiography of small radio-labeled animals tissue sections for comparative analysis with *in-vivo* imaging studies, optical detection of immuno-detected proteins, DNA and RNA labeled with chemiluminescent or fluorescence probes and UV shadow imaging for DNA analysis. Some exemplar images produced with these imaging modalities are shown in Figure 1.2.

Once a new drug, a new diagnostic or treatment technique has been proved successful at the pre-clinical stage, i.e. on small animals or on *ex-vivo* samples, then this has to demonstrate efficacy in humans via clinical trials before reaching the general public. It can then be safely used in medicine.

The following stage for bio-medical imaging in the healthcare path involves patient imaging for diagnosis and treatment purposes. Diagnostic imaging is based on creating visual representations of the interior of the body for clinical analysis and medical intervention, relying on both analysis of patient morphology and functional studies on patient organs. Some examples of diagnostic imaging in medicine are provided in Figure 1.2, showing a chest radiograph, a functional study of the brain activity through
Positron Emission Tomography (PET) and tomographic images of the brain anatomy in Computed Tomography (CT) slices.

Diagnosis, however, is not the arrival point for imaging in medicine. In fact, imaging is also a valuable tool for assisting patient treatment. Figure 1.2 shows some examples of medical imaging modalities to assist patient treatment. This includes dose planning for external beam radiotherapy, where dose estimation is registered over a CT scan, and sentinel lymph node images acquired in an surgical setting using an intra-operative gamma camera.

Figure 1.2 (preceding page): Imaging in healthcare. A schematic representation of the role which imaging plays in pre-clinical and clinical sciences. Pre-clinical imaging is often usually carried out on *ex-vivo* animal tissues or extracted proteins, DNA and RNA. Exemplar images are reported for tissue autoradiography (*image courtesy of Dr Alexis Bailey, University of Surrey, UK*), western blotting of proteins via immuno-detection (*image courtesy of Midwest Scientific, St. Louis, US*), and for DNA shadow imaging (*image courtesy of the University of Southampton, UK*). Clinical imaging is considered for use on the general public, after such techniques have proved themselves successful and safe in clinical trials. Diagnostic imaging involves a number of techniques for *in-vivo* anatomic and functional investigations, including general radiology (*image courtesy of Dr Nevit Dilmen, University of Istanbul, Turkey*), Positron Emission Tomography (PET) (*image courtesy of Dr Jens Maus, Helmholtz-Zentrum Dresden-Rossendorf, Germany*), Computed Tomography (CT) (*image courtesy of Dr Mikael Häggström, Uppsala University Hospital, Sweden*). Imaging in medicine can also assists patient treatment as in dose planning for external beam radiotherapy (*image courtesy of Varian, Palo Alto, US*) and for sentinel lymph node imaging in an intra-operative setting (*image from P. Olcott at al., Phys. Med. 30(3):340-5, 2014*).
1.2 Challenges in bio-medical imaging and motivations

The first radiograph ever produced, Frau Röntgen’s hand (see Figure 1.1a), was made using glass photographic plates, where an X-ray sensitive emulsion of silver salts was applied to glass plates. Film plates, subsequently substituted for photographic film in 1918, represented the only detection and recording medium available for bio-medical imaging until the 1950s, where new imaging detector technologies started being developed to improve image quality, reduce dose to the patient and the required exposure time, using intensifier screens in radiology first, and subsequently moving to digital detectors. Even today, over 100 years later, film emulsion is still used in many applications for pre-clinical sciences, e.g. in autoradiography or western blotting.

Imaging with film retains several advantages, which still renders it as a detector of choice in some fields today. Due to the fine granularity of silver crystals in the emulsions, film offers an unbeatable spatial resolution. It is also inexpensive and suitable for long exposures, due to the low dark signal which can be generated in this medium. However, its drawbacks are certainly significant: film delivers low sensitivity, thus requiring potentially long exposures, and a limited linear response over a small range of signal amplitudes and intensities (about two decades); it ages with alterations of its chemical properties and thus its sensitivity, and film images are strongly dependant on the exact type of development procedure adopted. Also, chemical development of the virtual image is a long “off-line ” procedure which requires availability of dedicating facilities (dark rooms) and toxic chemicals, thus adding additional costs, time (and risks) to standard medical investigations. Additionally, in a time where a transition to digital medicine is taking place, film images require scanning techniques to be digitised, as the optical density (film darkening) response is not linear with dose.

The last few decades of the 20th century have seen a large effort to develop modern technologies for bio-medical imaging, to replace the archetypal victorian film emulsion. Most of the work in this field has been driven by the important developments achieved in radiation detector technologies for basic physics research, and effort has
been made to translate such leading edge technologies into bio-medical imaging. Numerous attempts have been made, including gaseous, scintillator-based or solid state detectors[5]. However, few of these research efforts ever reached the point of being routinely used in pre-clinical science laboratories or medical facilities. In fact, in order for such prototypes to be successfully commercially exploited, there is a need to deliver specific practical requirements such as cost and ease of operation, besides providing the needed operational performance (noise, imaging area, image quality).

For these reasons, digital detectors which have actually become a viable commercial alternative to film emulsion stand apart from detectors developed for basic physics research; these are in fact the scientific development of technologies developed for large scale production for the consumer market. Flat Panel Imagers (FPIs) and Charge-Coupled Devices (CCDs) are the digital detector technologies of choice for many applications in clinical and pre-clinical imaging, respectively. Both technologies are based on the development of an already available commercial process, arising from consumer-based flat-panel display technology for the former technology, and general imaging cameras for the latter. Following on from a mass production technology, these two kinds of imaging sensor technology can offer relatively low production costs to be then routinely used in laboratories and medical facilities, together with the advantage of being mature technologies leading to ease of use and maintenance.

Nevertheless, FPIs and CCDs offer some significant drawbacks when used for biomedical imaging. Despite reaching very large imaging areas (50×50 cm$^2$), FPIs present a high noise floor, relatively large pixel pitch, low frame rate and image lag [6, 7]. In fact, FPIs present a high readout noise (>1000 e$^-$), due to the inherently high pixel noise combined with line R/C noise sources, which tends to increase for large arrays [8]. Pixel pitch is also physically limited by resolution and tolerance in the photolithographic techniques used, as well as by the need to have a small resistance value in the ON state for fast readout [7]. Moreover these devices suffer from a low frame rate, mainly due to the low electron mobility in amorphous silicon ($\sim$ 1 cm$^2$ V$^{-1}$ s$^{-1}$) compared to crystalline silicon (typically 1400 cm$^2$ V$^{-1}$ s$^{-1}$), and image lag, due to the
abundance of charge traps in the a-Si lattice. This produces a decay time comparable to the readout speed, which needs to be compensated for, algorithmically [9].

Similarly, CCDs, although offering a very low noise performance ($\approx e^{-}$) suffer low frame rate, due to the sequential readout access and require operation at low temperature [10].

Complementary Metal-Oxide-Semiconductor (CMOS) Active Pixel Sensors (APSs) represent the next stage of detector technology development in bio-medical imaging. In the last decade, those detectors have become a major category of high-volume semiconductor production, driven by the large consumer-led demand for multimedia applications. In fact, CMOS APSs are now capable of offering low noise at room temperature (60-150 e) [11, 12, 13], as each pixel contains an active circuit [7], and a pixel pitch in the order of 25-50 $\mu$m, deriving from higher resolution lithographic processes and higher levels of integration reached in the CMOS manufacturing processing, technically and economically driven by integrated circuit applications [14]. APSs also offer a higher frame rate, based on the higher charge mobility of crystalline silicon compared to a-Si in FPIs, true random access via column parallel readout and a relative absence of image artifacts (e.g. image lag). These advantages, together with low power consumption ($\leq$ 1 W), decreasing proportionally to the minimum feature size in the specific technology used [15] and potential for a low cost and fast scaling technology based on standard consumer-based CMOS fabrication techniques, have made CMOS APSs, a valuable alternative in the bio-medical imaging field.

Furthermore, APSs are based on the use of in-pixel active circuits, which can then be designed to program operations at the pixel level. On-chip intelligence for APSs can be developed down to the pixel level to implement counting functions, rudimentary image processing, selective readout of Region-Of-Interest (ROI) and selection of ROIs triggered by an external event. On-chip intelligence also has the potential to implement a number of readout modalities which can improve imaging tasks, such as on-line dose sensing [16] to minimise patient dose, and dynamic tracking of specific imaging feature...
Recently, developments in photolithography have made available the realisation of large area devices integrated onto the complete width of a silicon wafer to create a contiguous sensor array, so that sensors of arbitrary size can be manufactured, limited only by wafer size. CMOS APSs can now meet the demand in imaging of most clinical and pre-clinical imaging applications.

1.3 Structure of the thesis

The previous Section describes, in principle, the potential advantages that CMOS APS technology may offer to the field of biomedical imaging. Therefore, the thesis is focused on the investigation of CMOS APSs for a range of biomedical imaging applications, spanning preclinical bio-imaging to proton CT therapeutic imaging. This will demonstrate the breadth of potential for this technology, capitalising on novel performance assessment methodologies developed for wafer scale sensor technology.

In order to set the scene on APS technology, basic working principles and design of such detectors are described in Chapter 2, together with a detailed description of the large area CMOS APS used in this work.

The next conceptual step to demonstrate the use of CMOS APSs in bio-medical imaging is assessing the detector’s intrinsic imaging performance. Chapter 3 reports on the detector characterisation in terms of noise, gain, quantum efficiency and Contrast-to-Noise ratio. Particular emphasis is placed into assessing these figures on a regional basis across the detector area, and on a per-pixel basis. In fact, large area sensors can suffer from performance non-uniformity, thus leading to limited imaging performance and reduction on real dynamic range, when used in routine applications. Detector performance is also compared with state-of-the-art technology for radiology, by means of Detective Quantum Efficiency calculated with a cascaded linear model of the detector response.
Chapter 1. Introduction

The suitability of CMOS APSs in pre-clinical science is then demonstrated using a “direct contact printing approach” in western blotting imaging in Chapter 4. Such imaging technology has never been used in this field, and thus a first proof of principle of chemiluminescence detection with a small area CMOS APS is reported. The chemiluminescence signal is then studied for a large area CMOS APS and detection parameters such as linearity, sensitivity and quantum efficiency are assessed. The first images of protein sequences with a CMOS APS are presented and protein detection performance is compared with the “gold standard” film emulsion and with a commercially available digital bio-imaging system.

For imaging applications which use high doses of ionising radiations, e.g. radiology or proton Computed Tomography, assessing the radiation hardness of such a detector is fundamental. In fact, in order for such technology to be used in routine radiological applications, it must provide a suitable life time before radiation damage becomes significant. Chapter 5 reports a study on the radiation hardness of a large CMOS APS. Radiation damage mechanisms are analysed for both ionising and non-ionising energy deposition and the response of the detector to X-rays (i.e. purely ionising) and to protons (i.e. producing ionising and non-ionising energy deposition) fields are studied. Particular emphasis in this Chapter is given in separation of ionising and non-ionising contributions, so that the response of the detector to an arbitrary field of radiation can be predicted. Furthermore, the expected life time of this detector when routinely used in Mega Voltage radiotherapy and proton CT is discussed.

Chapter 6 discusses the feasibility of using large area CMOS APSs as energy-range detectors for proton CT. As such technology has never been used in this application, a first proof of concept is provided. This Chapter is based on verifying the fundamental requirements for a detector to be used for proton CT. It provides demonstration of single proton imaging, and verification of the capability of CMOS APSs to track protons as they traverse a stack of such detectors. An analysis in terms of correlation is reported for a simple stack of only two sensors, and an algorithm for proton tracking
is introduced and applied to simulated and experimental data. The proton tracking algorithm is then extended to a simulated full size energy-range telescope, based on a large number of CMOS detectors.

Two further aspects of this work are represented by a proof of principle of large area CMOS detector for tissue autoradiography and by detailed GEANT4-based Monte Carlo simulations, used in several places within the thesis (Chapters 5 and 6). Appendix A reports on the use of a large area CMOS APS for tissue autoradiography, in a comparative approach with state-of-the-art technology in the field. The GEANT4 simulation framework, used for this work, is described in Appendix C, together with new charge transport classes (developed by the author), to simulate charge diffusion and collection in the detector.

1.4 Achievements and major contribution

This thesis represents the first investigation into the capabilities of large area CMOS APSs to be used across a number of different imaging modalities in bio-medical science, spanning protein imaging to proton CT, using both ionising and non-ionising radiation sources.

The main achievements, including contributions to the field, contained in this thesis may be summarised as follows:

- A novel methodology for a per-pixel analysis of the electro-optical performance of a wafer scale CMOS APS has been developed, to address characterisation of inhomogeneity issues arising from the stitching techniques used to manufacture wafer scale sensors, one of the main limitations of this innovative technology.

- For the first time a large area CMOS APS has been used for chemiluminescence detection in proteomics. This has produced better performance compared to state-of-the-art imaging systems in the field, paving the way for CMOS sensors to be used in proteomics research.
• For the first time a large area CMOS sensor, the first solid state device which can offer an imaging area comparable to film emulsion, has been used to produce tissue autoradiography images.

• A novel procedure for noise suppression, based on spectroscopic analysis of the incoming signal, has been developed to improve imaging quality in autoradiography, an application where a very low noise is required.

• For the first time the radiation hardness of a wafer-scale imaging detector, specifically designed for bio-medical imaging has been assessed and an estimation of its lifetime for routine clinical applications provided. In addition, the ionising and non-ionising radiation damage has been studied to isolate these two contributions in a mixed radiation field and used to provide a further verification of the Non Ionising Energy Loss (NIEL) scaling hypothesis.

• A first verification of the proton counting capabilities of CMOS APSs have been provided in the context of proton CT.

• Novel algorithms have been developed to perform proton tracking in a CMOS energy-range telescope designed to perform proton CT, paving the way for a new generation of imaging devices to be used in this application.

Finally, a list of journal and conference publications resulting from this work, are reported in Appendix D.
Chapter 2

The DynAMITe CMOS Active Pixel Sensor: design overview and proof of concept demonstration

Image sensors have become a major category of high-volume semiconductor production, driven by the large consumer-led demand for multimedia applications, such as still and video digital cameras for general imagery, as well as industrial machine vision and military applications. Even when developed for consumer based applications, it appears evident how such sensors can be translated for scientific applications, mainly space or medical related, where requirements for low noise, high responsivity, large dynamic range and high spatial resolution are met.

In the 1990s, monolithic Complementary Metal-Oxide Silicon (CMOS) image sensors started emerging as a serious alternative to charge-coupled devices (CCDs) [18] which had dominated both consumer and scientific imaging application market for decades, as a natural result of advances in CMOS technology for processors and DRAMs. This new technology provides potential for integrating all imaging functions onto a single
chip, effectively reducing size, cost and power consumptions for image sensors [10, 19]. Standard CMOS technology allows for manufacture of monolithic integrated devices, where all the functions for timing, exposure control and ADC can be implement on a single piece of silicon, leading to the production of so called camera-on-a-chip [20].

CMOS image sensors can be classified in two groups: Passive Pixel Sensor (PPS) and Active Pixel Sensor (APS). The former consists of pixels provided with a photosensing element [21], i.e. a photodiode, and a MOSFET switch for readout, while the latter also includes a source follower transistor, used as active amplifier. Although PPSs offer a high fill factor, since each pixel is provided with a single transistor, those sensors also suffer from a higher noise and a lower frame rate due to higher capacitive loads, compared to APSs, and do not allow to scale to large array sizes. For this reason, scientific interest is now focused on APSs.

2.1 CMOS APSs fundamentals

APSs were described for the first time by Noble in 1968 [22], and the design of such sensors has greatly evolved in the 1990s with Yadid-Petch’s work on random pixel access and electronic shuttering [23]. A basic APS pixel architecture is shown in Figure 2.1a. This pixel employs a photodiode, as a light sensing element, and a circuit consisting of three transistors: a reset transistor (Reset) to control the integration time, a source-follower transistor (SF) acting as a buffer amplifier, and a row-select transistor (RS) responsible to connect the pixel to the column buses for readout. This design is known as a 3T pixel, comprising three transistors, and can be considered the simplest circuit for APSs.

Pixel operation can be classified in three main stages:

- The reset stage. The gate of the Reset transistor is pulsed ON for a certain amount of time to remove information from the pixel at the end of each integration time period. The reset can either be “soft” or “hard”, depending on the voltage
2.1. CMOS APSs fundamentals

\( V_{\text{RESET}} \) used [24, 25]. The soft reset corresponds to the condition where the gate-to-drain voltage of the reset transistor is lower than the threshold voltage of the transistor, so that charge from the sense node can thermally cross the reset gate barrier to the \( V_{\text{DD}} \) (see Figure 2.1a) drain region. On the other hand, in hard reset operation mode the gate-to-drain voltage exceeds the transistor threshold, thus allowing a steady state operation. Soft reset results in a high saturation level and low read noise at the cost of image lag and low-illumination nonlinearity, while hard reset shows no image lag and greater linearity, at the expense of increased noise and reduced saturation level, and thus resulting in a decrease in dynamic range.

- **The phototransconductance stage.** Following pixel reset, the photodiode capacitor is discharged through a constant integration time at a rate proportional to the incident illumination. The silicon photodiode is sensitive to light from UV to near IR (20-1200 nm). It usually consists (see Figure 2.1b) of a PN junction, i.e. a \( \text{P}^+ \) epitaxial layer and an \( \text{N}^+ \) well reversely biased, thus creating a depletion region. During this stage, a fraction of charge generated in the PN junction of the photodiode, both thermally (\( n_{\text{DARK}} \)) and photogenerated (\( n_{\text{PE}} \)), is collected at the \( \text{n}^+ \) photodiode contact. This charge packet is accumulated on the photodiode intrinsic capacitance and at all other parasitic capacitances connected to the photodiode node, known cumulatively as integration node capacitance, which provides charge-to-voltage conversion. The voltage generated by the collected charge at the integration node capacitance \( C_{\text{NODE}} \) can be written as:

\[
V_{\text{SIGNAL}} = \frac{Q}{C_{\text{NODE}}} = \frac{(n_{\text{PE}} + n_{\text{DARK}})q}{C_{\text{NODE}}}
\]  

(2.1)

- **The readout stage.** At the end of each integration period the voltage of Equation 2.1 is presented to a column bus using a row select transistor. The select transistors connect all the pixels to the column buses, and at a given time all the pixels on a given horizontal row are placed onto their corresponding column buses. Following this process, the pixel is reset by pulsing the reset gate and the entire cycle is then repeated.
Figure 2.1: A basic 3T APS pixel (a) connected to a n+ sense node represented by a photodiode. (b) A cross section of a photodiode with a PN junction [10].

The working principle described here is relevant to the 3T pixel, the simplest design for APS, but could be easily extended to more complex architectures featuring 4, 5 or 6 in-pixel transistors, namely 4T, 5T and 6T respectively. The use of such additional transistors, with respect to the standard 3T design, allows for greater array functionality and, in most cases, improved performance, although, at the expenses of fill factor and pixel size.

2.1.1 Readout modalities

In a 3T pixel array the readout of all pixels cannot be performed simultaneously, but a rolling readout technique has to be applied. In fact all the pixels in a row of a 3T pixel array are readout simultaneously, while consecutive rows are readout sequentially. This process is called Rolling Shutter [10, 19]. Figure 2.2 shows the time dependence of the rolling shutter technique. The red region in the schematics, labelled $T_R$, represents the time needed to read and reset a row, while the green area represents the exposure time, the actual time frame when the specific row is sensitive to light ($T_E$).

For a sensor working in rolling shutter mode, the minimum exposure time allowed for each row corresponds to the time needed to read and reset all the other rows in the pixel array. Thus, for an array of $N$ rows the minimum exposure time ($T_E$) can be calculated as a function of $T_R$, $T_E = N \times T_R$, i.e. the minimum time required to read and reset $N$ rows. As can be seen from Figure 2.2, the readout of each row is shifted with respect to the previous/following one, as the readout of a row does not start before
the previous row has been readout. At any given time instant, different rows will be in a different phase of their exposure/readout cycle. Considering an event happening at a given time $T$ in Figure 2.2, this will be recorded in Frame # 3 for rows at the top of the array, in Frame # 2 for rows at the bottom and might fall in between two frames for some other row, thus without being detected. For this reason rolling shutter generates image artefacts, when there is relative motion between the sensor and the scene to be imaged.

An alternative approach to the rolling shutter is called *Snapshot mode* [23]. This methodology makes use of a storage element inside the pixel, which can work as a mechanical shutter. All the rows can be exposed simultaneously, and then the exposure is stopped and each pixel value is stored inside the in-pixel storage element, ready to be sequentially readout. It is worth noting that the sequential readout, used for snapshot mode, affects only the time at which pixels are readout (i.e. sequentially with row numbers), while all pixels in the array are exposed simultaneously. Although the snapshot mode solves the problem of the rolling artefacts, it also has some drawbacks. In fact, one of the main concerns with this modality is the pixel sensitivity and charge lost during storage time, together with reduced capacity for signal storage which can often result in signal loss. Also the frame rate achievable with this technique is lower than with rolling shutter, limiting its advantages to those applications where high frame rates are not required.

### 2.1.2 Large area CMOS APSs

Recently, developments in photolithography have made available the realisation of large area devices integrated onto a silicon wafer to create a contiguous sensor array. The associated technique, referred to as a stitching process [33, 34], is based on the use of a mask reticle, comprising several functional blocks or a larger circuit (e.g. pixels and readout electronics), which is stepped and repeated, in whole or in part, across a silicon wafer to create modularly different sectors of a large circuit, so that sensors of arbitrary size can be manufactured, limited only by wafer size. As a result of this, a number of large area CMOS APSs is currently available and listed in Table 2.1, together with
<table>
<thead>
<tr>
<th></th>
<th>Tile Area (cm²)</th>
<th>Side Buttable size (µm)</th>
<th>Pixel Frame rate (fps)</th>
<th>Noise floor (e⁻)</th>
<th>Conversion gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>RedEye1 [26]</td>
<td>2.46×4.92</td>
<td>3</td>
<td>48</td>
<td>4.5</td>
<td>150</td>
</tr>
<tr>
<td>C97732 DK-11 [27]</td>
<td>12×12</td>
<td>–</td>
<td>50</td>
<td>1</td>
<td>1250</td>
</tr>
<tr>
<td>CMOS APS [16]</td>
<td>7.73×14.5</td>
<td>3</td>
<td>33.55</td>
<td>8.7</td>
<td>175</td>
</tr>
<tr>
<td>VLA CMOS [28]</td>
<td>4.9×9.8</td>
<td>3</td>
<td>96</td>
<td>1.3</td>
<td>250</td>
</tr>
<tr>
<td>CMOS APS [29]</td>
<td>7.37×7.75</td>
<td>–</td>
<td>18</td>
<td>1.25</td>
<td>240</td>
</tr>
<tr>
<td>LAS [12]</td>
<td>5.6×5.6</td>
<td>–</td>
<td>40</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>CMOS APS [30]</td>
<td>12×15</td>
<td>3</td>
<td>150</td>
<td>30</td>
<td>–</td>
</tr>
<tr>
<td>CMOS APS 2923 [31]</td>
<td>11.5×14.5</td>
<td>2</td>
<td>75</td>
<td>26</td>
<td>361.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>164.9</td>
</tr>
</tbody>
</table>


\(^a\)High Full Well \(^b\)Low Full Well

Table 2.1: Summary of the design specifications and optical performance for CMOS large area detectors.
2.2. The DynAMITe detector

A large area CMOS APS is the detector investigated in this work. The following Sections will provide details on the design of this APS and proof of concept demonstrations of some key aspects of its novelty.

2.2 The DynAMITe detector

The CMOS APS used in this work was developed by the EPSRC funded MI3+ collaboration, for a range of bio-medical applications including life science imaging, diffraction breast imaging and radiotherapy portal imaging. This sensor is referred to as Dynamic Range Adjustable for Medical Imaging Technology or DynAMITe. DynAMITe is fabricated in a standard 0.18 μm CMOS technology using reticle stitching [33], covering a total active area of 12.8×13.1 cm$^2$. It is also designed to be two-side buttable so that the imaging area can be further increased up to 25.6×26.2 cm$^2$. A picture of the silicon wafer from which the DynAMITe sensor has been diced is shown in Figure 2.3a, while the first deployment of this detector is shown in Figure 2.3b.

Figure 2.2: Schematic representation of the rolling shutter readout mode for a pixel array of $N$ rows ($P - Q$). The read and reset time is reported in red ($T_R$), while the exposure time, when pixels are sensitive to radiation, is shown in green ($T_E$).
Figure 2.3:  
(a) The DynAMITe chip wafer. The boundaries of the chip region are visible with the ruler denoting a 12.8 cm edge dimension. 
(b) The DynAMITe detector in its first deployment.

Figure 2.4:  
(a) A schematic representation of the array arrangement. Red circles represent the Pixels, placed at 100 μm pitch, while cyan circles are Sub-Pixels, placed at 50 μm pitch. 
(b) A schematic representation of the diodes arrangement for the DynAMITe detector. Sub-pixels show a deeper depletion width compared to Pixels.
2.2. The DynAMITe detector

2.2.1 Pixel design

The novel concept underlying the development of this detector lies in simultaneously using high and low well capacity diodes in the same pixel array. High well capacity diodes are able to provide high dynamic range, whereas low well capacity diodes can deliver low noise, offering an extended dynamic range when these are combined. This sensor thus consists of two grids of different well capacity diodes, which are geometrically super-imposed. Large well capacity diodes, referred to as Pixel (P) diodes are placed on a 1280×1312 pixel matrix with 100 μm pitch, whereas low well capacity diodes, referred to as Sub-Pixel (SP) diodes are arranged on a 2560×2624 pixel grid with 50 μm pitch (Figure 2.4a). SP diodes are reset at a higher reset voltage than P diodes, leading to the creation of different depletion widths (Figure 2.4b). The deeper depletion width of SP type diodes allows, as radiation interacts with the sensor, the first generated charge to be collected by such low noise pixels. Hence low intensity signals are detected with an intrinsic low noise, resulting in in a good Signal-to-Noise Ratio (SNR) performance. As Sub-Pixels reach near saturation, then P type diodes start collecting providing an extended dynamic range. The intrinsic higher noise of those pixels does not degrade the SNR performance of the sensor, as the higher noise level has to be compared with a higher intensity signal.

A proof of principle of this concept is provided in Section 2.3.1.

2.2.2 Readout

The readout architecture of the DynAMITe sensor is based on a dual readout chain: each pixel is connected to two reset, select and output lines (Figure 2.5). Doubling the readout electronics allows either to increase the readout speed or to readout different sensor regions or pixel types in combination.

An increase in readout speed is achieved when the whole pixel array or a Region of Interest (ROI) is readout by using both readout chains. In fact, the two separate electronic chains can be synchronised in a ping-pong arrangement where two rows are readout simultaneously, one by each chain. This arrangement effectively reduces the
number of rows/pixels readout by each chain to a half, thus increasing the readout speed of a factor of two.

By using the ping-pong architecture, the maximum readout frame rate\(^1\) for the 1280×1312 P type matrix is designed for 90 fps whereas this figure is 30 fps for the 2560×2624 SP type matrix, while the use of ROIs allows for an even faster readout, with frame rate proportional to the ROI size. As the dual readout chain can be used either in the ping-pong architecture or operated independently, several readout modalities are implemented for this sensor:

- Pixel full matrix (90 fps);

\(^1\)The maximum frame rate is the reciprocal of the minimum integration time of the sensor, i.e. the time needed to readout all the pixels. This is defined as

\[
T = (N_{rows} \times T_{row}) + \left( \frac{N_{pixels}}{N_{amp}} \times T_{clock} \right)
\]

where \(N_{rows}\) is the number of rows to readout, \(T_{row}\) is the time required to sample a row, \(N_{pixels}\) is the number of pixels in the array, \(N_{amp}\) is the number of amplifiers and \(T_{clock}\) is the time required to clock a pixel voltage towards the amplifiers.
2.2. The DynAMITe detector

- Sub-Pixel full matrix (30 fps);
- Pixel full matrix (45 fps) readout simultaneously with Sub-Pixel full matrix (15 fps);
- Pixel full matrix (45 fps) readout simultaneously with Pixel ROI (714 fps for $100 \times 100^2$ pixel ROI);
- Sub-Pixel full matrix (15 fps) readout simultaneously with Sub-Pixel ROI (714 fps for $100 \times 100$ pixel ROI);
- Pixel full matrix (45 fps) readout simultaneously with Sub-Pixel ROI (714 fps for $100 \times 100$ pixel ROI);
- Sub-Pixel full matrix (15 fps) readout simultaneously with Pixel ROI (714 fps for $100 \times 100$ pixel ROI).

Additionally by using the above readout modalities it is possible to perform Non Destructive Readout (see Section 2.3), where the full frame of the entire pixel matrix can be readout at a given frame rate using one electronic chain, while an ROI is simultaneously readout at a higher frame rate using the second chain. As the ROI is not reset, during readout, collected charge is accumulated until the full matrix is readout. Non Destructive Readout allows for Correlated Double Sampling (CDS) [35] and online dose sensing. A proof of principle of this concept is provided in Section 2.3.2.

The DynAMITe sensor has also been designed according to the radiation hardness-by-design methodology [36]. A more detailed discussion on the radiation hardness-by-design of this detector is reported in Chapter 5.

2.2.3 Stitching

Advances in photolithography have made available the realisation of large area devices integrated onto a silicon wafer to create a contiguous sensor array by using the so

---

$^2$A $100 \times 100$ ROI is presented in this example. However ROI size can be chosen arbitrary, leading to different frame rates proportional to the ROI size.
Figure 2.6: Schematic of the stitching block process used for the DynAMITe sensor. Eight types of sub-block mask are reported: blocks E, B, C, F which contribute to the sensor active area, blocks A and D for row addressing, block G for the Grey code counter, blocks H and I for the outputs.

called *stitching block process*. A schematic of the stitching block process used for the DynAMITe sensor is reported in Figure 2.6, showing the 8 types (labelled as A-I regions) of sub-block mask used for manufacture. The sensor imaging area has been constructed by stepping a sub-block mask of 18.0 mm × 25.6 mm across the wafer (E block), creating a 180×256 pixel block and a 360×512 pixel block for Pixels and Sub-Pixels respectively. Additional blocks are placed at the edges of the sensor:

- 1 A-type block and 5 D-type blocks for row addressing;
- 1 G-type block for the Grey Code Counter which drives the row addressing;
- 7 H-type blocks and 1 I-type block for the outputs;
- 7 B-type (18 mm × 3.2 mm), 1 C-type (2 mm × 3.2 mm) and 5 F-type blocks (2 mm × 25.6 mm) as further imaging area to reduce dead space at the edge in the 2-side buttable configuration.
E blocks are repeated five times in the Y-axis direction and seven times in the X-axis direction, realizing an imaging area of $1260 \times 1280$ P-type pixels and $2520 \times 2560$ SP-type pixels which increases up to $1280 \times 1312$ and of $2560 \times 2624$ pixels respectively when B,C and F-type blocks are taken into account.

### 2.3 Proof of concept

The novel concepts exploited in the development of the DynAMITe sensor rely on a double level of duality, as described in Section 2.2. The first level of duality, i.e. the use of two different full well capacity diodes in each pixel, should allow the use of each diode type independently, thus offering either a low noise and low full well capacity on a small pixel pitch or a higher noise and higher full well capacity on a larger pixel pitch. Alternatively, using small and large well capacity diodes in combination, the sensor should be able to provide an extended dynamic range, based on the low noise of Sub-Pixels and high full well capacity of Pixels.

The second level of duality of the DynAMITe sensor consists in the use of two independent electronic chains for readout. Such architecture is meant to offer an increase in readout speed, when a single pixel camera is readout (either Sub-Pixels or Pixels), or to provide simultaneous readout of Pixels and Sub-Pixels when those are combined together. The use of two electronic chains for readout also allows implementation of a Region-of-Interest based readout for Non Destructive Readout (NDR).

The following sections investigate the novel concepts described above, thus representing a proof of principle for this novel CMOS APS.

#### 2.3.1 Charge collection

The charge collection process for each diode type of the DynAMITe detector has been studied, in order to verify the sequential collection by Sub-Pixels first and then Pixels. The sensor was exposed to a uniform illumination field provided by an LED array at
Figure 2.7:  
a) Output signal for the Sub-Pixels (red circles) and Pixels (black squares) as a function of the illumination level.  
b) Output signal for the Sub-Pixels operating alone (red circles) or synchronously with Pixels operating synchronously (black squares) as a function of the illumination level.
2.3. Proof of concept

Figure 2.8: (a) Schematic representation of the NDR process: secondary camera (blue symbols) is readout with an exposure time $T_2$. The primary camera (red symbols) is readout with an exposure time $T_1 = 4 \times T_2$ until reset occurs. Mean signal (b) and noise (c) from NDR is represented as function of time for the ROI of the secondary camera readout at 26.3 fps (black squares), together with the full frame of the primary camera readout at 3.6 fps (red circles). The full frame of the primary camera operated alone is readout at 3.6 fps (blue triangles) is also shown.
523 nm with illumination level increasing from 0 to 350 nW/cm$^2$. The signal amplitude from both the Sub-Pixel and Pixel cameras operating synchronously, with a 190 ms integration time, was measured.

A further test was performed in order to evaluate any potential lack of charge collection in the Sub-Pixel camera, due to the Pixel camera collecting at the same time. The sensor was exposed to a uniform illumination field with the illumination level increasing from 0 to 50 nW/cm$^2$ with 190 ms integration time. The output signal of the Sub-Pixel camera was measured when it was operated alone and also when operated synchronously with the Pixel camera.

Figure 2.7a shows the sensor output versus illumination level for both Sub-Pixel and Pixel cameras. The Sub-Pixel camera collects with higher efficiency than the Pixel camera, until near saturation. Sub-Pixels reach saturation at about 40 nW/cm$^2$ when the corresponding Pixels are at the 14% of their dynamic range, given the 190 ms integration time. After this point Pixels collect all the high intensity signal with a linear behaviour, over the complete range investigated (0-350 nW/cm$^2$), and reach saturation at the end of this interval.

The output signal for the Sub-Pixel camera is shown in figure 2.7b when the Sub-Pixels collect alone (i.e. Pixels are disabled), and in the case that Pixels and Sub-Pixels are operated synchronously. There is no appreciable difference in the Sub-Pixels output in the two different conditions, meaning that synchronous operation of both diode matrices does not affect the charge collection process.

Thus, one of the main concepts underlying the development of the DynAMITe detector, i.e. the sequential charge collection of the in-pixel diodes types based on their different reset voltages (see Section 2.2), has been demonstrated together with the possibility of extending the dynamic range, when the two full well capacity diodes are operated simultaneously.
2.3.2 Non Destructive Readout

The dual readout circuits of both large and small diode matrices facilitate NDR. The full frame of the entire pixel matrix (referred to as primary camera) can be readout at a given frame rate, while a Region-Of-Interest (ROI), referred to as secondary camera, is simultaneously readout at a higher frame rate. One of the two readout chains is used for the primary camera, whereas the other chain addresses the secondary camera. If the reset signal is disabled for the secondary camera, this camera is reset only via the reset signal for the primary camera allowing charge to be integrated in the ROI for an exposure time equal to the one at which the whole matrix is readout. Figure 2.8a schematically shows this functionality.

The NDR architecture was tested by evaluating signal and noise in both primary and secondary cameras and compared with the performance of the primary camera operating alone.

The DynAMITe sensor was exposed to a 523 nm LED with a brightness of 20 nW/cm$^2$. A 512 × 2513 pixel ROI was defined and readout at 26.3 fps (primary camera), whereas the whole matrix was readout at 3.6 fps (secondary camera). For comparison purposes, a similar data set, with the same illumination conditions and exposure time, was acquired for the primary camera while the secondary camera was disabled. Mean signal and noise were analyzed for each of these conditions, where noise was calculated evaluating the standard deviation of the difference of two consecutive frames$^3$.

The mean signal (Figure 2.8b) or noise rms (Figure 2.8c) is shown as function of the integration time for the ROI of the secondary camera readout at 26.3 fps (black squares), for the full frame of the primary camera readout at 3.6 fps, simultaneously with the secondary camera (red circles), and for the full frame of the primary camera readout at 3.6 fps alone (blue triangles).

Signal and noise for the primary camera operating with and without the secondary camera are comparable. Thus when operating simultaneously, both readout schemes can operate independently without compromising sensor performance. The signal mea-

$^3$A more detailed discussion on noise sources and analysis methodologies is reported in Chapter 3.
Chapter 2. The DynAMITe CMOS APS

The signal measured in the secondary camera follows the pattern expected for NDR (see Figure 2.8a), linearly increasing till reset occurs in the primary camera. The maximum signal measured in the secondary camera is comparable, but slightly lower than that of the primary camera. This can be explained by accounting for the fact that both cameras are read-out by two different electronic chains, thus potentially suffering variations in gain.

The noise measured in the ROI of the secondary camera is 35% lower noise compared with the primary camera (both with and without ROI selection). This can be explained by considering that noise calculation by subtraction of consecutive frames in the NDR mode (i.e. when camera is not reset frame-by-frame) effectively represents performing Correlated Double Sampling (CDS [35])\(^4\), thus leading to suppression of constant noise components and reduction in the total noise measured.

Thus, one of the novel concepts in the development of the DynAMITe detector, i.e. the use of a double electronic chain for readout, has been demonstrated successfully working in performing NDR and CDS.

2.4 Discussion

The DynAMITe detector has been presented as a novel, large area APS imaging sensor capable of two inherently different resolutions each with different noise and saturation performance in the same pixel array. This novel design has great potential for use in a variety of biomedical imaging applications fulfilling the requirements of large imaging area with high dynamic range and frame rate.

\(^4\)CDS is based on the use of two samples during the pixel read-out cycle, one soon after reset and the second before the next reset occurs. The two samples are then subtracted proving a measurement devoid of constant components, since the common offset of the two samples is canceled by subtraction.
Chapter 3

Characterisation of the DynAMITe CMOS APS

3.1 Motivation

Accurate characterisation of device performance is critical in scientific imaging, particularly for medical and biological imaging, where it is required to ensure device performance is suitable for the intended application. Device characterisation in terms of parameters such as conversion gain, noise, quantum efficiency, full well capacity, Contrast-to-Noise ratio (CNR) and Detective Quantum Efficiency (DQE) are also needed for optimisation of system performance.

However, recent developments in photolithography have made available the production of large wafer scale CMOS APSs, based on the use of the stitching process. This motivates the need for a regional characterisation of detector performance and assessment of the uniformity of this performance over the whole sensor area.

In fact, despite CMOS stitched sensors being monolithic, sources of non-uniformity of response and regional variations can affect wafer scale sensor performance, representing a significant challenge for these detectors. Non-uniformity of stitched sensors can arise from a number of factors related to the manufacturing process, including lack of refocussing of the stepper at each reticle position, variations of the amplification, variation
between readout components, wafer defects and variations across the wafer, such as threshold voltage, leakage current, substrate doping, mobility, trans-conductance, etc.

Non-uniformity in large area sensors can represent a significant drawback as regional differences in noise and conversion gain can generate local changes in imaging performance, such as Signal and Contrast-to-Noise Ratio, introducing anisotropy in image resolution. Such factors also produce uncertainty when the sensor is calibrated using parameters derived globally over the entire sensor area, ultimately limiting the sensor resolution. Finally, large variations in noise and conversion gain can produce a decrease in dynamic range.

Non-uniformity and regional variations in large area stitched sensors have been investigated by some groups recently. Zin et al. [37] investigated spatial non-uniformity of a Large Area Sensor (LAS) [12], [30]. Sensor non-uniformity was evaluated in each of the stitched regions of the sensor resulting in variations between 3.85-5.67 and 3.79-7.02% for read noise and gain respectively.

In this chapter, an investigation into the spatial non-uniformity of a wafer scale stitched CMOS APS is reported. Key metrics for sensor performance, such as read noise, conversion gain and full well capacity have been optically characterised for the whole sensor by means of the Photon Transfer Curve (PTC) technique. Nevertheless when optical characterisation involves wafer scale sensors, where manufacturing techniques are based on regional processes (i.e. reticle stitching), evaluating averaged parameters in specific regions of the sensor can be misleading. To assess inhomogeneity issues arising from the stitching techniques used to manufacture wafer scale sensors, a novel per-pixel characterisation of a large area sensor has been developed. A per-pixel PTC analysis of the sensor was performed, and results have been grouped on a per pixel basis and at a stitching block level, providing a more detailed understanding of the uniformity of response of the sensor. Furthermore the Fixed Pattern Noise (FPN) has been evaluated at different grouping levels across the entire sensors array and for each stitched region. Finally uniformity and regional variations of this large area CMOS sensor have been
3.2 Methodology for detector performance evaluation

3.2.1 Photon Transfer curve

Assessment of the sensor performance in terms of image noise and signal is an essential step to optimise imaging sensors and evaluate their reliability in specific applications. Noise sources in APSs can be classified into two different types: temporal noise, i.e. having temporal dependence, and spatial noise, due to noise sources which vary across the detector matrix. Read and shot noise belong in the first category, the first being due to signal-independent noise sources (e.g. reset noise, off-chip and on-chip amplifier noise, quantisation noise, dark current shot noise), the latter depends on the detected
Figure 3.2: a) An ideal PTC [38]. b) Schematic comparison between a Photon Transfer Curve generated by averaging noise and signal in a region of interest (left) and by evaluating it as a per-pixel function (right). Index \( i \) refers to row, index \( j \) to columns and \( k \) to the progressively increasing illumination levels.

signal (Poisson statistics of interacting radiation quanta). FPN is considered a spatial related source of noise, as it is ascribable to non-uniformities in the manufacturing process and to differences in pixel and column voltages as well as variations in column amplifiers.
3.2. Methodology for detector performance evaluation

The Photon Transfer Curve (PTC)[38, 39, 40, 41] represents the standard for evaluating performance parameters, such as read noise, conversion gain and full well capacity, and provides means to isolate and quantify noise components in the sensor response.

An ideal PTC for a CMOS sensor is illustrated in Figure 3.2a where noise is plotted against signal, at different illumination levels, on a log-log scale. Three different noise regimes can be identified in the plot. A first region, at low signal levels, is dominated by the read noise which is independent of the signal level (corresponding to a slope equal to zero in the plot). The second regime is dominated by the shot noise, i.e. by the noise associated with the Poisson statistics of incoming photons, which is related to the square root of the signal. Since the PTC in represented in logarithmic coordinate, the slope of the shot noise region is $1/2$. The third region is dominated by the FPN and is proportional to signal, with a slope of 1 in the plot. Finally, saturation occurs at some illumination within the FPN region. Charge spreads between pixels, smoothing and lowering the noise components.

The signal generated in a pixel by $P$ incident photons, expressed in digital numbers (DN)$^1$, can be written as

$$S(DN) = P \cdot QE \cdot \eta_i \cdot S_v \cdot A_{INT} \cdot A_{EXT} \cdot A_{ADC}$$ (3.1)

where $QE$ is the interacting efficiency (interacting photons/incident photons), $\eta_i$ is the quantum yield (number of $e^-$ generated per interacting photon), $S_v$ is the sensitivity of the sense node (V/e$^-$), $A_{INT}$ is the gain of the in-pixel amplifier (V/V), $A_{EXT}$ is the gain of the external amplifier (V/V) and $A_{ADC}$ is the gain of the ADC (DN/V). A schematic representation of the signal generation process is shown in Figure 3.1. The conversion gain can be defined as

$$K(e^-/DN) = \frac{1}{S_v \cdot A_{INT} \cdot A_{EXT} \cdot A_{ADC}}$$ (3.2)

For incident photon of wavelength $\lambda > 400$ nm, a single electron-hole pair is generated per interacting photons, i.e. $\eta_i = 1$. Using this and Equation 3.2, Equation 3.1 can be

---

$^1$This is the pixel voltage converted into a digital value by sensor ADCs.
Chapter 3. Characterisation of the DynAMITe CMOS APS

re-written as

\[
S(DN) = \frac{P_I}{K(e^{-}/DN)}
\]  

(3.3)

where \(P_I\) is the number of interacting photons, i.e. \(P_I = P \times QE\) The conversion gain \(K\) can then be evaluated by relating it to the signal variance \(\sigma_S^2\). The variance of Equation 3.3 can be found by applying the propagation of the errors formula, and by adding the read noise \(\sigma_R^2(DN)\) in quadrature

\[
\sigma_S^2(DN) = \left(\frac{\partial S}{\partial P_I}\right)^2 \sigma_{P_I}^2 + \left(\frac{\partial S}{\partial K}\right)^2 \sigma_K^2 + \sigma_R^2(DN)
\]  

(3.4)

Using the assumption of a Poisson distribution for the interacting photons (\(\sigma_{P_I}^2 = P_I\)) and of a negligible variance for the conversion gain (\(\sigma_K^2 \sim 0\)), Equation 3.4 becomes

\[
\sigma_S^2(DN) = \frac{P_I}{K^2} + \sigma_R^2(DN)
\]  

(3.5)

Substituting Equation 3.3 into Equation 3.5, the conversion gain can be expressed as follows

\[
K = \frac{S(DN)}{\sigma_S^2(DN) - \sigma_R^2(DN)}
\]  

(3.6)

where \(\sqrt{\sigma_S^2(DN) - \sigma_R^2(DN)}\) is the signal shot noise \(\sigma_{shot}\). Plotting logarithmically the signal noise \(\sigma_S\) of the sensor as a function of the mean signal \(S\) gives the PTC. In order to calculate the conversion gain \(K(e^{-}/DN)\), it is necessary to extract the signal shot noise from the total sensor noise. Temporal invariant components can be removed from the signal noise subtracting two consecutive frames which will yield a FPN suppression. The read noise can be suppressed by subtracting the variance of a differenced dark frame (\(P_I = 0\)) from the signal variance \(\sigma_S^2\). The PTC data are generated exposing the sensor to different illumination levels from dark to saturation, provided by an LED array centred at \(\lambda = 523\) nm (bandwidth 35 nm) coupled to a lens and a single neutral density filter, to achieve uniform illumination. A calibrated photodiode was also used to estimate the photon flux at the sensor position. The rms sensor noise was then evaluated as a function of the sensor signal. The gain of the system is determined from the noise measurements, under assumption of Poisson distributed input signal noise. Conventionally, both noise and signal are calculated as an average in an ROI (512 \times 512 Sub-Pixel and 256 \times 256 Pixel ROI in this work). Hence this procedure provides an
average information on the behaviour of the sensor.

### 3.2.2 Per-pixel Photon Transfer Curve

In order to gain information on the uniformity of response and regional variations of the sensor, the Photon Transfer Curve is evaluated on a per pixel basis from generalising Equation 3.6 into:

\[
K(i, j) = \frac{S(DN)(i, j)}{\sigma_S^2(DN)(i, j) - \sigma_R^2(DN)(i, j)}
\]

(3.7)

representing a relation between the pixel signal \(S(i, j)\) at location \((i,j)\) and the associated pixel signal noise \(\sigma_S(i, j)\), resulting in the evaluation of a per-pixel conversion gain \(K(i, j)\) and a read noise \(\sigma_R(i, j)\) matrix. A schematic to highlight the differences between PTC resulting from ROI average and the per-pixel procedure is shown in Figure 3.2b).

### 3.2.3 Fixed Pattern Noise

Fixed Pattern Noise (FPN) represents the spatial variation of the output image of a sensor due to different gains and offsets in pixel transistors and to column amplifiers. FPN can be analysed considering separately the effect due to pixels and columns [42]:

\[
F_{i,j} = X_{i,j} + Y_j
\]

(3.8)

where \(X_{i,j}\) is the pixel-to-pixel(P-P) FPN due to gain and offset variation in pixel transistors and \(Y_j\) is the column-to-column (C-C) FPN due to variation in column amplifiers of the column parallel readout. The two terms contributing to the FPN can be calculated by evaluating the difference between each pixel signal \(S_{i,j}\), under uniform illumination, averaged over \(N\) frames, and the overall average signal \(\bar{S}\) given by

\[
\bar{S} = \frac{\sum_{i,j,m} S_{i,j,m}}{LMN}
\]

(3.9)
where \(i, j, m\) are rows, columns and frames indices and \(L, M, N\) are their maximum values. The components of the FPN (\(Y_j\) and \(X_{i,j}\)) and their variances (\(\sigma_Y^2\) and \(\sigma_X^2\)) are calculated as follows

\[
Y_j = \frac{1}{L} \sum_{i=1}^{L} F_{i,j}
\]

\[
X_{i,j} = F_{i,j} - Y_j
\]

\[
\sigma_Y^2 = \frac{1}{M-1} \sum_{j=1}^{M} Y_j^2
\]

\[
\sigma_X^2 = \frac{1}{M(L-1)} \sum_{i=1}^{L} \sum_{j=1}^{M} X_{i,j}^2
\]

Column-to-column (\(\sigma_Y^2\)) and pixel-to-pixel FPN (\(\sigma_X^2\)) metrics are conventionally expressed in terms of the percentage of the average signal \(S\) for a given illumination level, which is 50% of saturation in this work.

### 3.2.4 Contrast-to-Noise Ratio

The uniformity of the image quality performance, deriving from the electro-optical properties listed in the previous sections, have been investigated to facilitate understanding of detector performance in clinical routine applications.

X-ray measurements in a common mammography set-up were performed to evaluate the uniformity of the CNR across the pixel array.

A 25 kVp X-ray source with Mo anode was used together with a 30 µm Mo filtration. A CsI scintillator, 150 µm thick, was chosen to allow photon conversion before detection and this was coupled to a 3 mm thick Fiber Optic Plate (FOP). A test object of 45 mm × 25 mm × 120 mm, made of Perspex, was used to assess the CNR in the
3.2. Methodology for detector performance evaluation

stitching blocks. The object was placed above the detector with the 45 mm side parallel to the beam axis, in order to deliver a breast-equivalent thickness of 53 mm [43]. Since the object only covers 20% of the detector area (its projection on the detector area is 25\text{mm} \times 120\text{mm}), it was stepped across the detector reticle and images acquired at each step.

Signal within the test object (\(S_o\)) and within the background (\(S_b\)) were evaluated in ROIs corresponding to the detector stitching block, together with their standard deviation (\(\sigma_o\) and \(\sigma_b\) respectively), to derive CNR:

\[
\text{CNR} = \frac{S_o - S_b}{\sqrt{\sigma_o^2 + \sigma_b^2}} \tag{3.14}
\]

In order to assess the uniformity of the CNR across the detector area, a figure of merit (\(\text{CNR}_{\text{uni}}\)) was evaluate following [44]:

\[
\text{CNR}_{\text{uni}} = \frac{N(\Delta_{15\%}-\text{CNR}_{\text{mean}})}{N_{\text{total}}} \% \tag{3.15}
\]

as the ratio between the number of ROIs where CNR is within 15% of the mean CNR (\(\text{CNR}_{\text{mean}}\)) and the total number of ROIs used for the study (\(N_{\text{total}}\)).

3.2.5 Detective Quantum Efficiency

Detective Quantum Efficiency (DQE) is commonly used method to characterise imaging performance, describing the Signal-to-Noise ratio through the imaging chain [45]. DQE is a frequency-dependent figure, however many of the signal and noise transfer properties can be studied at zero-frequency (DQE(0)).

A cascaded linear system can be used to model signal and noise properties of imaging detectors [46]. This approach is based on describing the imaging system as a series of discrete stages, each of which represents a quantum gain or blurring process. For a gain stage, the mean fluence output quanta \(\bar{q}_i\) can be related to the input quanta \(\bar{q}_{i-1}\) by means of the mean gain \(\bar{g}_i\) at that stage \(i\):

\[
\bar{q}_i = \bar{g}_i \bar{q}_{i-1} \tag{3.16}
\]
For an imaging system based on indirect conversion, the first stage of the cascaded linear system is represented by the interaction of X-ray incident quanta in the converter. The gain at this stage is given by

\[
\bar{g}_1 = \frac{E_{\text{MAX}}}{\int_0^{E_{\text{MAX}}} q_0(E)(1 - e^{-\mu(E)T})dE} \int_0^{E_{\text{MAX}}} q_0(E)dE
\]

(3.17)

where \( q_0(E) \) is the incident X-ray spectrum and \( \mu(E) \) is the attenuation coefficient at the energy \( E \) and \( T \) is the thickness of the converter. The generation and emission of optical quanta in the converter are represented by the gain \( \bar{g}_2 \), comprising two terms \( \bar{g}_{2a} \) and \( \bar{g}_{2b} \). The first term represents the average number of optical photons generated per interacting X-ray:

\[
\bar{g}_{2a} = \frac{E_{\text{MAX}}}{\int_0^{E_{\text{MAX}}} q_1(E)\bar{g}_{2a}(E)dE} \int_0^{E_{\text{MAX}}} q_1(E)dE
\]

(3.18)

where \( q_1(E) \) is the spectrum of the interacting photons and \( \bar{g}_{2a}(E) \) is the mean number of optical photons generated per X-ray of energy \( E \). The second term contributing to the gain at the converter stage, \( \bar{g}_{2b} \), is the fraction of generated optical photons which will exit the converter. The latter term (\( \bar{g}_{2b} \)) can be expressed in terms of its variance, or Poisson excess \( \epsilon_{g_2} \) [47], related to the Swank noise [48]. Values of the Swank noise reported in [49] are used in this work. The final gain stage is represented by the fraction of optical quanta detected. Following [46], this factor (\( \bar{g}_3 \)) is assumed to be equivalent to the fill factor.

This cascaded model allows prediction of the theoretical DQE(0) for imaging systems, once properties of the scintillator system and detector parameters (i.e. read noise, pixel pitch and fill factor) are known. The DQE(0) can be expressed as

\[
\text{DQE}(0) = \frac{\bar{g}_1 \bar{g}_2 \bar{g}_3}{1 + \bar{g}_3(\bar{g}_2 + \epsilon_{g_2}) + \frac{\sigma_{\text{pixel}}^2}{\bar{q}_0 \bar{q}_1 \bar{g}_2 \bar{g}_3}}
\]

(3.19)

Theoretical prediction of DQE(0) is of great importance for this work, as it allows comparison with detection performance of the detector under study with conventional FPIs, routinely used in X-ray imaging applications.
3.3 Detector performance

3.3.1 Photon Transfer Curve

Figure 3.3 displays the PTCs for Sub-Pixels and Pixels calculated from averaged noise and signal in an ROI (see Figure 3.2b). The noise components are decomposed into “read and shot noise” and “shot noise” after FPN subtraction. The curves for signal shot noise, plotted logarithmically, as a function of the mean output signal, provide the read noise and Full Well Capacity (FWC), i.e. the signal level corresponding to the maximum variance in the PTC. The read noise and FWC can then be converted from relative digital numbers to absolute units of electrons after deriving the conversion gain of both pixel arrays, as described in Section 3.2.1. The conversion gain is then calculated using the slope of these curves in their respective linear region. Dynamic range can be calculated from these parameters as $DR = 20 \log_{10} \frac{FWC}{\sigma_R}$, while Integral Non-Linearity (INL) is the difference between the data points and the linear regression fit ($\Delta$) in the linearity plot (Figure 3.3b), computed as $INL = (\Delta_{\text{max}} - \Delta_{\text{min}})/ADC_{\text{fullscale}} \times 100$.

The derived electro-optical performance parameters calculated from the PTC for Sub-Pixel and Pixel arrays are reported in Table 3.1. It can be seen in Table 3.1 how Sub-Pixels offer a lower noise and lower FWC compared to Pixels, and consequently lower conversion gain. This is due to the choice of different diodes to be used for Pixels and Sub-Pixels. In fact, as discussed in Section 2.2 and 3.3, the concept underlying the development of the DynAMiTe detector was to use diodes with low noise and low FWC for the Sub-Pixels and diodes with higher noise and higher FWC for Pixels, so that a combination of these two can lead to an increased dynamic range.

The linearity curve for both pixel arrays, displayed in Figure 3.3b, shows the result of using the conversion gain to calculate the number of signal electrons as a function of the number of incident photons. The slope of these curves yields a quantum efficiency $\eta$ of 45% for the Sub-Pixel array and 38% for the Pixel array ($\lambda=523$ nm). This can be explained when considering the diode area inside both pixel arrays. In fact, as discussed in Section 2.2 and 3.3, each Sub-Pixel pixel is fitted with 4 parallel-connect 0.6 $\mu$m diameter diodes, with a diode-area-to-pixel-area-ratio of 0.002; whilst, Pixels are
Figure 3.3: \(a\) Read and shot noise PTCs displayed with the decomposed shot noise component evaluated in a region of interest for Sub-Pixels and Pixels. \(b\) Linearity curves for Sub-Pixels and Pixels calculated in an ROI. The number of electrons produced in the pixel for a given number of incident photons is plotted as function of the photon flux. The slope of these curves in the linear region represents the sensor Quantum Efficiency (\(\eta\)).
3.3. Detector performance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pixels</th>
<th>Sub-Pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversion gain ($e^−/DN$)</td>
<td>293.3 ± 0.8</td>
<td>50.0 ± 0.2</td>
</tr>
<tr>
<td>Read noise ($e^−$)</td>
<td>780 ± 1</td>
<td>149.9 ± 0.7</td>
</tr>
<tr>
<td>Full well capacity ($e^−$)</td>
<td>1.9 ± 0.7 × 10^6</td>
<td>0.3 ± 0.2 × 10^6</td>
</tr>
<tr>
<td>Quantum efficiency $\eta$ (%)</td>
<td>38</td>
<td>45</td>
</tr>
<tr>
<td>Dynamic range (dB)</td>
<td>68</td>
<td>66</td>
</tr>
<tr>
<td>INL (%)</td>
<td>2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of the performance parameters of the Pixel and Sub-Pixel arrays calculated for a region of interest.

provided with a single diode of 1 μm diameter, with a diode-area-to-pixel-area-ratio of 0.0003, implying a lower collection efficiency for the Pixels.

3.3.2 Per-pixel Photon Transfer Curve

Conversion gain and read noise have been calculated as a per-pixel function, using Equation 3.7. A conversion gain and read noise value of zero has been assigned to Pixels and Sub-Pixels corresponding to a determination coefficient ($R^2$) lower than 0.5, i.e. poor fit quality, in the linear fit of Equation 3.7, (less than 1% of the pixel array). The conversion gain matrices for Pixels and Sub-Pixels are shown in Figure 3.4a and 3.5a respectively, together with the relevant histograms (Figure 3.4b), whereas the read noise of Pixels and Sub-Pixels is displayed in Figure 3.4c and 3.5c respectively together with the relevant histograms (Figure 3.4d).

The colour scales represent the conversion gain per-pixel in $e^−/DN$ for Figure 3.4a and 3.5a, while the read noise expressed in $e^−$ (Figure 3.4c and 3.5c). Matrices for conversion gain and read noise (Figures 3.4 and 3.5) are devoid of specific patterns of variation across the entire imaging area, suggesting the stitching procedure in sensor fabrication did not introduce variations at reticle level or at the reticle edges.

Additionally data reported for the Sub-Pixel array present a group of pixels with sig-
significant deviation from the average values between column 83 and column 96, distinguishable by a vertical dashed line in Figure 3.5a and c. This is due to readout issues at the amplifier level for the specific sensor under characterisation.

From the Central Limit theorem it can be shown that a random variable might be modelled as log-normal if it can be written as the multiplicative product of many independent random variables each of which is positive [50]. In this case, both conversion gain and read noise can be considered quantities which arise from a series of different amplification and conversion stages, i.e. multiplicative stages. In fact, as radiation interacts with the sensor, charge, generated inside the depletion region or reaching this region due to thermal diffusion, is collected. Collected charge is then multiplicatively amplified by a series of amplifiers both inside the pixel and on the sensor periphery, before being sampled and digitalised. This process is summarised in Equation 3.1 and schematically represented in Figure 3.1.

All the multiplicative factors of Equation 3.1 can be considered random variables assumed to be normally distributed for each pixel, and the total distribution across the entire matrix is still normal, due to the Central Limit theorem. Thus is possible to schematise the signal generation process as a multiplicative product of many normally distributed random variables. Hence, conversion gain and read noise, which are proportional to signal, can be modelled as global log-normal distributions.

Histograms within Figures 3.4 and 3.5 have been fitted with a log-normal probability density function $f(x)$

$$f(x) = \frac{1}{x \cdot \sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (\log(x) - \mu)^2 \right\}, \ x > 0 \quad (3.20)$$

where $\mu$ and $\sigma$ represent the mean and the standard deviation of the log-normal distributed variable. The log-normal fit for all distributions of Figures 3.5 and 3.4 gave a coefficient of determination $R^2 > 0.99$. Mean ($\mu$) and standard deviation ($\sigma$) have also been evaluated and results are reported in Table 3.2 with the fitting errors. The read noise distributions present a mean value of $262 \pm 1 e^-$ and $887 \pm 1 e^-$ with a standard deviation of $168.14 \pm 0.01 e^-$ (64% of the mean) and $451.75 \pm 0.07 e^-$ (50% of the mean)
3.3. Detector performance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Symbol</th>
<th>Unit</th>
<th>Pixels</th>
<th>Sub-Pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>Gain</td>
<td>$K$</td>
<td>e$^{-}$/DN</td>
<td>296 ± 1</td>
<td>59 ± 1</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Gain</td>
<td>$K$</td>
<td>e$^{-}$/DN</td>
<td>90.14 ± 0.08</td>
<td>20.94 ± 0.02</td>
</tr>
<tr>
<td>$\sigma/\mu \times 100$</td>
<td>Gain</td>
<td>$K$</td>
<td>%</td>
<td>50</td>
<td>64</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Noise</td>
<td>$\sigma_R$</td>
<td>e$^{-}$</td>
<td>887 ± 1</td>
<td>262 ± 1</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Noise</td>
<td>$\sigma_R$</td>
<td>e$^{-}$</td>
<td>451.75 ± 0.07</td>
<td>168.14 ± 0.01</td>
</tr>
<tr>
<td>$\sigma/\mu \times 100$</td>
<td>Noise</td>
<td>$\sigma_R$</td>
<td>%</td>
<td>30</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 3.2: Results of the log-normal fit for read noise and conversion gain distributions in term of mean $\mu$ and standard deviation $\sigma$ for Sub-Pixel and Pixel arrays respectively. The conversion gain distributions feature a mean value of $59 \pm 1e^{-}$ and $296 \pm 1e^{-}$ with a standard deviation of $20.94 \pm 0.02e^{-}$ (35% of the mean) and $90.14 \pm 0.08e^{-}$ (30% of the mean) for Sub-Pixel and Pixel arrays respectively. When the values for conversion gain calculated on a per-pixel basis (see Table 3.2) are compared with the average conversion gain of Table 3.1, these appear comparable between the two methodologies. However, the per-pixel analysis lead to larger values of read noise (see Table 3.2) for both pixel types. Such a behaviour can be related to the fact that for the per-pixel PTC all pixels are accounted for, including the ones at the periphery of the sensor, where large variations in the process or in the Si wafer can be present.

Comparing the fit parameters in Table 3.2, it can be observed that the relative width ($\sigma/\mu$) of the two distributions, namely conversion gain and read noise, are comparable for the two pixel types. In fact the relative width for conversion gain is 50% and 60%, while these figures are 30% ad 35% for read noise for Pixels and Sub-Pixel respectively. This is expected since the width of these distributions is due to the same manufacturing variations for both pixel types.

3.3.3 Photon Transfer Curve in stitching blocks

The conversion gain and read noise matrices have been averaged in each of the 35 E-type stitching blocks constituting the sensitive regions of the detector (see Figure
2.6). The percentages of variation per block with respect to the average value on the entire matrix are displayed in Figure 3.6 for both Pixel and Sub-Pixel arrays. The average percentage of variation is 1.31% for the conversion gain of Pixels and 1.18% for Sub-Pixels, whereas it is 1.32% for the read noise of Pixels and 1.52% for Sub-Pixels, with a coefficient of variation (COV) \( \leq 1.9\% \) for all of them. Per block variation in terms of read noise and conversion gain are comparable for both pixel types as expected.

This latest result is comparable with the analysis on non-uniformity reported by [37], who showed that gain and read noise of the LAS sensor, calculated per stitching block, exhibit a COV of the order of 3.79 - 7.02% and 3.85 - 5.67% respectively. When these COV values are compared with a value \( \leq 1.9\% \) measured for DynAMITe, this demonstrates a higher level of uniformity in terms of optical performance.

### 3.3.4 Fixed Pattern Noise

The level of FPN is shown in Figure 3.7 for Pixels (a) and Sub-Pixels (b) at half saturation. Column-to-Column (CC) and Pixel-to-Pixel (PP) FPN values have been calculated in different regions of the sensor: “whole array” refers to the FPN calculated on the full imaging area, “column output” refers to the FPN calculated in one of the seven column outputs of the sensor, “row sub-stitch” refers to the FPN calculated in a region of the column outputs limited in one of the five row stitching areas. The C-C FPN has been measured as 1.3% for Pixels and 2.0% for Sub-Pixels, whereas the P-P FPN results in 2.2% for Pixels and 4.2% for Sub-Pixels on the whole array.

FPN, calculated on the whole array, is globally higher for the Sub-Pixel than for the Pixel arrays. This is due to the total diode node capacitance of Sub-Pixels being dominated by the diode itself which suffers higher variations leading to more mismatches among pixels. Conversely, the large capacitance offered by the Pixel diodes works as a low-pass filter which can filter out diode-to-diode variations. Limiting the region for measurement to a column output or row sub-stitch leads to a decrease in FPN. C-C FPN is reduced by a factor 0.4 in the column output and by a factor of 0.6 in a row sub-stitch for the Sub-Pixels, whereas these values are 0.5 and 0.7 respectively for the
3.3. Detector performance

Figure 3.4: Conversion gain and read noise matrices for Pixels, resulting from the per-pixel photon transfer curve. Colour scales represent the conversion gain value per-pixel expressed in $e^-/DN$ (a)) and the read noise per-pixel in expressed $e^-$ (c). The relevant histogram for each of the matrices is reported (b) and (d).
Figure 3.5: Conversion gain and read noise matrices for SubPixels, resulting from the per-pixel photon transfer curve. Colour scales represent the conversion gain value per-pixel expressed in $e^-/DN$ (a) and the read noise per-pixel in expressed $e^-$ (c). The relevant histogram for each of the matrices is reported $b)$ and $d)$. 

(a) Gain-SP

(b) Gain-SP

(c) Noise-SP

(d) Noise-SP
3.3. Detector performance

Figure 3.6: Percentage of variation of conversion gain and read noise per stitching block with respect to the average value across the whole matrix for Pixels \( a \) and \( d \) and for the Sub-Pixels \( b \) and \( d \) respectively. Grey scales represent the percentage of variation with respect to the average value on the whole matrix expressed in \( \% \).
Figure 3.7: Percentage of FPN, both C-C and P-P, measured in different regions of the Dynamite detector for Pixels (a) and Sub-Pixels (b).
3.3. Detector performance

Pixels. P-P FPN is reduced by a factor of 0.5 in the column output and by a factor of 0.6 in the row sub-stitch for the Sub-Pixels. However, Pixels do not show any significant decrease of P-P FPN confining calculation in limited regions (column output or row sub-stitch), as P-P FPN for Pixels is dominated by the low-pass filter mentioned above.

Moreover, FPN, when evaluated for each of the 35 sensor stitching blocks (Figure 3.8), results in higher values at the edges of the sensor, e.g. stitching block 1 and 35, attributed to higher manufacturing variation in the CMOS processes at the edge of the wafer. C-C FPN shows a similar trend of variation among stitching blocks for Sub-Pixels and Pixels (Figure 3.8a). This can be explained taking into account the column parallel read-out and the fact that column amplifiers of both type of pixels are placed close to each other. Hence variations at amplifier level for one diode type could result in similar variations for the amplifiers of the other diode type, thus leading to a similar behavior in the columnar spatial variations, i.e. C-C FPN. Moreover both C-C and P-P FPNs show a periodic variability with a period of 5 stitching blocks, corresponding to the number of row stitching block per column. Thus a higher FPN is related to those stitching blocks which lay at the bottom of the sensor (blocks 5, 10, 15, 20, 25, 30, 35 in Figure 2.6) which may be representative of a region of higher process variation being nearer the edge of the original wafer.

3.3.5 Contrast-to-Noise Ratio

A further stage in this investigation led to test this system in a mammographic configuration, in order to assess to what degree detector non-uniformities affect image quality in a typical radiographic application. Figure 3.9a displays a composite image, obtained by averaging single frames, showing the Perspex test object being imaged in different positions to cover all the stitching blocks of this sensor. Object and background signals and standard deviations are then calculated to assess the CNR across the stitching blocks, after correcting for gain variations [51]. The top seven stitching blocks of the sensor showed some device-related sensitivity, and have been therefore discarded from
Figure 3.8: Percentage of Fixed Pattern Noise (FPN), both column-to-column (C-C) (a) and pixel-to-pixel (P-P) (b), measured in each of the 35 stitching blocks of the Dynamite detector for Pixels and Sub-Pixels
analysis. CNR per stitching block is reported in Figure 3.9b for Pixels and Sub-Pixels.

On average CNR results were derived to be higher for Pixels (average value of 34.4) compared to Sub-Pixels (average value of 5.9), with a COV of 13%, compared to 26% COV for Sub-Pixels across all the 35 stitching blocks. This difference can be explained by the higher FPN Sub-Pixels feature, due to the total diode node capacitance being dominated by the diode itself, compared to Pixels. The uniformity of the CNR across both pixel arrays ($CNR_{uni}$) has been evaluated according to Equation 3.15, resulting in 79% uniformity for Pixels and 37% for Sub-Pixels.

The results obtained in terms of CNR and uniformity of CNR have been compared with the state-of-the-art large area sensors for mammography. A comparative analysis for CNR and uniformity is reported in [44] for a number of mammography system used in BreastCheck, the National Breast Screening Program of Ireland. The mammography systems investigated in the referenced work are: GE Senographe DS and GE Senographe Essential [52], both a-Si flat panels coupled to a CsI scintillator, Lorad Selenia[53], an a-Se flat panel used for direct detection, and Sectra MDM-L30 and M-40 [54], both based on the use of strip sensors in direct detection. Comparative data are shown in Table 3.3.5 together with some of the experimental settings. The thickness of the Perspex test object (45 mm) is not reported in the table, as it is the same for all the systems used in this comparison.

The Pixels of the DynAMITe sensor show the highest mean CNR when compared with the other detection systems of Table 3.3.5 (34.4). $CNR_{mean}$ is lower for Sub-Pixels (5.9) than that of FPIs (GE DS, GE Essential and Lorad Selenia), but results higher than that of the Sectra detectors. The CNR uniformity for Pixels is comparable with the highest among the $CNR_{uni}$ values for the systems included in Table 3.3.5 (GE DS, GE Essential and Lorad Selenia). On the other side, the CNR uniformity for Sub-Pixels is comparable with that achieved by the strip detectors (Sectra MDM-L30 and M-40).
Figure 3.9: (a) Composite image displaying the Perspex test object used for CNR measurements stepped across the pixel array. Areas where signal and background are calculated in a single stitching block are displayed in blue and yellow respectively. (c) CNR per stitching block for both Pixels (P) and Sub-Pixels (SP).
3.3 Detector performance

<table>
<thead>
<tr>
<th>Detector</th>
<th>Type</th>
<th>T/F</th>
<th>kVp</th>
<th>(CNR_{\text{mean}})</th>
<th>(CNR_{\text{uni}}(%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE DS</td>
<td>a-Si/CsI</td>
<td>Rh/Rh</td>
<td>29</td>
<td>8.14</td>
<td>96.2</td>
</tr>
<tr>
<td>GE Essential</td>
<td>a-Si/CsI</td>
<td>Rh/Rh</td>
<td>29</td>
<td>12.83</td>
<td>88.9</td>
</tr>
<tr>
<td>Lorad Selenia</td>
<td>a-Se/Direct</td>
<td>Mo/Mo</td>
<td>28</td>
<td>11.09</td>
<td>89.3</td>
</tr>
<tr>
<td>Sectra MDM D40</td>
<td>Si SS/Direct</td>
<td>W/Al</td>
<td>29</td>
<td>2.47</td>
<td>36.1</td>
</tr>
<tr>
<td>Sectra MDM L30</td>
<td>Si SS/Direct</td>
<td>W/Al</td>
<td>35</td>
<td>3.65</td>
<td>45.6</td>
</tr>
<tr>
<td>DynAMITe Pixels</td>
<td>APS/CsI</td>
<td>Mo/Mo</td>
<td>25</td>
<td>34.4</td>
<td>79</td>
</tr>
<tr>
<td>DynAMITe Sub-Pixels</td>
<td>APS/CsI</td>
<td>Mo/Mo</td>
<td>25</td>
<td>5.9</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 3.3: Comparative data for CNR and uniformity in CNR reported for a number of large area detectors used in mammography and for the DynAMITe CMOS APS. All systems are tested with a 45 mm thick Perspex test object. T/F stands for Target/Filtration materials of the X-ray tube.

3.3.6 Detective Quantum Efficiency

Evaluation of detector performance is typically carried out using different methodology for APSs and FPIs. The former are usually studied by means of their photon transfer properties for evaluation of read noise, conversion gain and FPN [38, 41], as in this work. The latter are investigated using DQE, Noise Power Spectrum and Modulation Transfer Function [45], in radiological application-like conditions. For this reason a direct comparison between performance parameters measured for the APS used in this work and reported in literature for FPIs might not be directly achievable. To address this point a theoretical calculation of the DQE at zero-frequency, based on detector performance parameters such as read noise, fill factor and pixel pitch, has been undertaken comparatively for the DynAMITe detector and a number of FPIs.

Calculations of DQE(0) have been performed, following Equation 3.19, for the DynAMITe detector and for a number of FPIs, under radiology imaging conditions. Detectors involved in the comparison are a-Si flat panels reported in [55, 46, 56]. Parameters used for the DQE(0) calculation, such as pixel pitch, read noise and fill factor are reported
### Table 3.4: Summary of imaging conditions and detector parameters used for the \( \text{DQE}(0) \) calculations.

<table>
<thead>
<tr>
<th>Application</th>
<th>Energy (kVp)</th>
<th>Exposure (mR)</th>
<th>Anode</th>
<th>Filtration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiology</td>
<td>110</td>
<td>0.03-3</td>
<td>W</td>
<td>2.75 mm Al</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Detector Technology</th>
<th>Pixel pitch (( \mu )m)</th>
<th>( \sigma_R ) (e(^-))</th>
<th>Fill factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>DynAMITe SP</td>
<td>APS</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>DynAMITe P</td>
<td>APS</td>
<td>100</td>
<td>780</td>
</tr>
<tr>
<td>Jee et al. 2003 [55]</td>
<td>a-Si FP</td>
<td>97</td>
<td>2000</td>
</tr>
<tr>
<td>Siewerdsen et al. 1997 [46]</td>
<td>a-Si FP</td>
<td>127</td>
<td>5000</td>
</tr>
<tr>
<td>Jaffray et al. 2001 [56]</td>
<td>a-Si FP</td>
<td>400</td>
<td>4000</td>
</tr>
</tbody>
</table>

In Table 3.4 together with the X-ray settings used. The same scintillator (150\( \mu \)m thick CsI) was used for calculation of \( \text{DQE}(0) \) for all the detectors involved in the comparison. In fact, as \( \text{DQE}(0) \) (Equation 3.19) depends on both scintillator-related (\( \bar{g}_1, \bar{g}_2 \) and \( \epsilon_{g_2} \)) and detector-related parameters (\( \sigma_R, \bar{g}_3 \) and \( a_{\text{pixel}} \)), the use of the same conversion system permits exclusion of the interaction and conversion stages from the comparison, which effectively becomes a comparison of the effects of detector-related parameters.

The dependence of the \( \text{DQE}(0) \) with exposure for all the detectors involved in the comparison is shown in Figure 3.10. For low exposure values (\( \bar{q}_0 \) in Equation 3.19), the differences in read noise dominate the \( \text{DQE} \) and curves in Figure 3.10 show a significant difference between each other. At higher exposure values those differences tend to flatten out as the number of X-ray quanta dominate the additive term in the denominator of Equation 3.19, and the scintillator efficiency and quantum gain represent the detection limit of the system.

Pixels and Sub-Pixels of the DynAMITe detector show the highest detection efficiency among the detectors used in this comparison, due to the inherently lower noise of APS compared to FPIs. Sub-Pixels have the highest efficiency, because of their noise level
which is the lowest among the detectors used in this comparison (see Table 3.4). The DQE(0) for Pixels and for the flat panel reported in [56] are comparable, even if subject to a very different noise level (100 and 4000 $e^-$ respectively). This results from the effect of a larger pixel pitch (400 $\mu$m) balancing off the effect of a higher noise in Equation 3.19. Flat panels presented by [55] and [46] have a comparable DQE, the lowest in the comparison, since it is dominated by a high noise (2000 and 5000 $e^-$ respectively) on a relatively small pixel pitch (97 and 127 $\mu$m respectively).

Figure 3.10: Calculated DQE(0) versus exposure for the DynAMITe Pixels and Sub-Pixels and for FPIs under radiological imaging conditions. The typical exposure range of a specific application is highlighted in figure.

3.4 Discussion

An investigation into the uniformity of response of this sensor and regional variations has been carried on by means of electro-optical performance assessment and X-ray imaging.
For the first time a per-pixel analysis of the electro-optical performance of a wafer scale CMOS APS has been carried out, to address inhomogeneity issues arising from the stitching techniques used to manufacture wafer scale sensors. A complete model spanning the signal generation in the pixel array, from charge generation inside the pixel sensitive volume to the digitisation of the generated signal, has been provided and proved capable of accounting for noise and gain measured distributions. Evaluation of uniformity in terms of noise and conversion gain highlighted a similar degree of homogeneity for both pixel types, being due to the same lithographic processes, while Fixed Pattern Noise showed an higher degree of variation, being ascribed to the electronic design of the pixels.

When tested in a conventional X-ray radiology set-up (i.e. mammography), the DynAMITe detector showed a uniformity in terms of CNR among the highest when compared with mammography detectors commonly used in clinical practise.

Finally a theoretical calculation of DQE at zero-frequency allowed comparison of the detection performance of the DynAMITe detector with FPIs, used in the medical imaging field. Evaluation of DQE showed how the intrinsic lower read noise of APS, compared to FPIs, results in higher DQE performance.

Considering the performance parameters assessed for this detector, in comparison with digital detectors commonly used in the clinical practise, demonstrates how such large area sensor technology may be successfully employed in medical imaging.
Chapter 4

Large area CMOS APSs for Western Blotting

4.1 Western Blotting

Many applications in pre-clinical sciences rely on the localisation within a specimen of a label (such as an optical or radioactive tag) bound to a specific structure (such as a gene or a drug receptor), by placing the specimen against some particular type of sensitive imaging sensor. A number of different labels (both ionising and non-ionising) that bind to specific biomolecules are available, to fulfil particular needs in terms of chemical binding and detection performance, for a range of different applications.

Western Blotting electrophoresis sequencing is an analytical technique widely used in the field of Functional Proteomics to detect, recognise and quantify specific labeled target molecules in biological samples. Chemiluminescence labelling represents the common form of labelling in this field. It requires neither external illumination nor filtering optics and does not produce an inherent label-related false-positive background, compared to other available techniques (e.g. fluorescence). A suitable choice for chemiluminescence labelling is to use Enhanced Chemiluminescence (ECL) reagents which produces optical emission at 425 nm. Target biomolecules (proteins), follow an immuno-detection procedure which ends in emission of a 425 nm photon. Hence de-
detecting such photons corresponds to gathering information on the spatial distribution of the biomolecule of interest.

### 4.1.1 Protein sequencing and immuno-detection

In order for protein to be immuno-detected in western blotting, these have to follow a process known as sequencing, which is schematically shown in Figure 4.1a. Proteins, extracted from a living organism (Escherichia Coli bacteria in Figure 4.1a, are denaturated by heating at a high temperature and loaded onto a support medium, usually a Sodium Dodecyl Sulfate-Polyacrilamide Gel Electrophoresis (SDS-PAGE), to undergo electrophoresis. SDS-PAGE gels are used to separate proteins, or more generally biomolecules, based on their mass. The majority of proteins, which are charged polypeptides, can bind to a fixed amount of SDS, resulting in masking the intrinsic charge of the polypeptide chains, such that the net charge per unit mass becomes constant. With a constant net charge for all the proteins bound to the gel, those proteins can migrate subject to the same force per unit mass in an electric field. However, as the gel is constituted by a porous material which works as a sieve, the distance that these proteins are able to travel, known as electrophoretic separation, will depend on the ratio between protein molecular radius, and thus mass, and the gel pore size. In this way the electrophoretic separation, obtained by applying an electric field to the gel, depends only on the protein molecular radius, related to the molecular mass, by means of the gel sieving effects. By using calibration markers of known molecular radius and knowing the gel pore size, it is possible to establish a relation between protein molecular radius and electrophoretic separation. A schematic of this process is shown in Figure 4.1b.

Proteins separated by electrophoresis have then to be transferred to a suitable medium, usually a membrane, for immuno-detection (see Figure 4.1c). Immuno-detection of proteins in western blotting is a multi-step process, schematically represented in Figure 4.1c. In fact, in order to be detected, proteins need to be chemically bound in several stages, until the last stage where light is emitted for detection. Proteins bind a specific

---

1 Polypeptides are chains of amino acids, constituting protein molecules.
4.1. Western Blotting

Figure 4.1: a) Block diagram of the work-flow required for proteins western blotting sequencing. b) Schematic representation of the electrophoretic process. A $t=0$, three different proteins $P_1, P_2, P_3$ with molecular radius $R_1 > R_2 > R_3$ are placed in an electric field $E$. Proteins migrate under the effect of the Coulombian force $F = Q \times E$, with $Q$ their resulting charge after protein binding to SDS. After a time $t=T$ proteins are separated, according to the ratio between their molecular radius and the radius of the gel pores. c) Schematic representation of protein detection in Western Blotting by means of chemiluminescent emission.
primary antibody, which then binds a secondary specific antibody. The secondary antibodies are bound to $\text{H}_2\text{O}_2$ molecules in order to emit light when exposed to a Luminol substrate ($C_8H_7N_3O_2$), according to the following chemical reaction:

$$
\text{Luminol} \rightarrow 3 - \text{APA}[\circ] \rightarrow 3 - \text{APA} + \gamma(425\text{nm}) \quad (4.1)
$$

where 3-APA is the product 3-aminophtalate and $3 - \text{APA}[\circ]$ is its excited state. As the excited state relaxes to the ground state, the excess energy is liberated as visible blue light photons of 425 nm wavelength.

### 4.1.2 ECL detection in western blotting

The most commonly used detection medium in western blotting is film emulsion, which is placed in a film cassette and in close contact with the sample, once the chemiluminescence reaction has started. Film emulsion offers an unbeatable spatial resolution.
due to its fine granularity, although for this particular application a spatial resolution of \(\approx 100 \, \mu \text{m}\) is adequate to resolve the finest details in the gel. However, film also presents severe drawbacks (limited dynamic range, non-linear response and low sensitivity) which strongly limits the imaging performance often entailing lengthy exposures: a significant bottleneck in the routine experimental work-flow. Moreover, in those applications, where it is not possible to assess \textit{a priori} the uptake of ligands in the biological sample of interest, an expensive and time consuming procedure of multiple exposures is required.

To overcome these limitations, a small number of western blotting digital imaging systems have being developed in the last two decades, mainly based on the use of CCDs [59, 58, 60, 57] and single avalanche diodes [61]. However, these systems exhibit serious shortcomings in imaging performance, such as a low frame rate due to the inherent sequential read-out of CCDs, and require operation at low temperature to achieve a reasonable noise level. Furthermore, since those CCD based system use optics to focus the sample image on their small imaging area, sensitivity is compromised, due to poor geometrical efficiency.

CMOS APSs have the potential to overcome such issues. In fact, such technology has the capability of offering a relatively low noise at room temperature, thus removing the need for bulky and cumbersome cooling systems. Furthermore, CMOS APSs can be manufactured with large areas (several hundreds of square centimetres), comparable with the largest samples in this application, and \textit{direct contact} imaging can be performed. Thus, samples can be placed directly on the detector surface, so maximising the geometrical efficiency. These advantages, together with potential for low cost, fast scaling technology and fast readout, suggest CMOS APSs as a viable alternative to film emulsion and CCD based imaging system for western blotting.
4.2 Motivations

The potential advantages of using a large area CMOS APS for imaging in pre-clinical science, and more specifically in western blotting have been discussed in the previous Section. However, a more detailed assessment of the detection capability of CMOS APSs in western blotting, as well as a comparative analysis with currently used detection systems, are needed.

This chapter reports on the viability of using a large area CMOS APS for western blotting imaging. Signal characterisation, imaging capabilities and performance comparison with detection systems routinely used in this field are studied.

As discussed in the previous Section protein detection in western blotting is a multi-step process which terminates with secondary antibody being activated by Luminol and emitting Enhanced Chemiluminescence Light (ECL). Thus, detection of ECL is the mean used in this application to gather information on the biological molecules object of study.

For the initial studies reported in the chapter, only this last stage, namely chemiluminescence emission by secondary antibody, has been studied to characterise the response of the sensor to the source of interest. The final part of the study involves the use of proteins and the whole western blotting flow (from denaturation to chemiluminescence activation, see Section 4.1.1) is reproduced to relate the results back to biological molecules to be imaged and quantified, i.e. proteins. The majority of biological samples used in this section have been prepared by the author at the Microbial and Cellular Sciences laboratories (Faculty of Health and Medical Science).

The viability of using a CMOS APS in western blotting has been first tested with a small area APS (Section 4.3), whose response to chemiluminescence light has been studied and compared to conventional detection techniques, i.e. film emulsion.

A detailed analysis of the DynAMITe detection performance to chemiluminescence light is reported for a number of different samples and compared with film emulsion
4.3. Proof of principle

(Section 4.4).

Finally, imaging capabilities of the DynAMITe detector for protein detection in western blotting are presented (Section 4.5), together with a comparative analysis with film emulsion and a commercially available CCD-based digital imaging system (Section 4.6).

4.3 Proof of principle

Preliminary studies on the suitability of CMOS APS for chemiluminescence detection in life science were first carried out using the Vanilla APS developed by the MI3 consortium [62]. This sensor features a relatively small area \( (12.8 \times 12.8 \text{ mm}^2) \), a pixel pitch of 25 \( \mu \text{m} \) and a readout noise in the order of 50 e\(^-\) [30], and has been extensively tested for digital autoradiography [63]. However, the limited imaging area of this sensor, used here only for preliminary demonstration of proof of concept, requires the creation of smaller samples compared to those routinely used in western blotting, with the constraint of both dimensions being smaller than 1 cm. For comparison, the smallest protein gels routinely available for western blotting are \( 9 \times 6 \text{ cm}^2 \).

4.3.1 Biological samples and experimental setup

Chemiluminescent samples were prepared by the author using Enhanced Chemiluminescence (ECL) labeled secondary antibody. A 70 \( \mu \text{l} \) drop of anti-mouse IgG HorseRadish Peroxidase (HRP) antibody (A8924, Sigma-Aldrich) in a 0.6 mg/ml concentration was diluted (1:5000) in a Tris Buffered saline solution. The resulting solution was spotted on a Hybond - LEP membrane (Amersham Biosciences). The membrane was then washed in a Lumi-Light Western Blotting Substrate (Roche Applied Science) in order to activate the ECL emission by means of the Luminol contained.

A second sample was prepared using anti-human IgG HRP antibody (A8667, Sigma-Aldrich) ranging from \( 117 \times 10^{-5} \) to \( 41 \times 10^{-2} \mu \text{g} \). The resulting solutions were spotted on a Hybond - LEP membrane (Amersham Biosciences). The same procedure as described before was used to activate ECL emission. The experiment was set up by
placing the sensor in a light proof box, with the sample in direct contact with the
detector surface. Images were acquired at a frame rate of 1 frame per second. In order
to correct for the dark offset and detector noise (see Chapter 2), all the acquired data
have been thresholded with respect to the per-pixel threshold $T_{i,j}$:

$$T_{i,j} = \mu_{i,j}^{ref} + k[\sigma_{i,j}^{ref} + (M^{ref} - M^{curr})]$$

(4.2)

where $\mu_{i,j}^{ref}$ is the mean value of the pixel $i,j$ in a reference dark acquisition, $\sigma_{i,j}^{ref}$ is the
standard deviation of the pixel $i,j$ in a reference dark acquisition, $M^{curr}$ is the mode
of the current frame and $M^{ref}$ is the mode of a reference dark acquisition. The term
$(M^{ref} - M^{curr})$ takes into account drifts in $\mu_{i,j}$ due to the thermal variation of the
dark current and the constant $k$ was chosen equal to 2 in order to ensure a confidence
interval of 95%.

### 4.3.2 Decay curve

A 70µl drop of anti-mouse HRP antibody was imaged for 8 hours at 1 frame per
second. Signal generated by ECL light, after thresholding according to equation 4.2,
is reported as a function of time in Figure 4.3. Analysis of the data indicates an
exponential decay law for the chemiluminescence process. The exponential decay of
this process can be explained similarly to the behaviour of radioactive decays. In fact,
given $N$ ECL-activated proteins in a sample, each of which is bound to a Luminol
molecule, the decay of a Luminol molecule to its ground state with subsequent light
emission (see Equation 4.1) is a stochastic process at the level of a single protein and
the probability for the bound Luminol molecule to relax to the ground state does not
depend on time. Thus, the number of relaxation events $(-dN)$ expected to happen in a
time interval $dt$ is proportional to the number of Luminol molecules $N (-dN/dt \propto N)$.
The expected relaxation events $-dN/N$ is proportional to an increment of time $dt$ by
means of the relaxation rate, or lifetime, $\lambda (-dN/N = \lambda dt)$. The solution to this first-
order differential equation is $N(t) = N_0 \exp^{-\lambda t}$, with $N_0$ number of Luminol molecules
at $t = 0$. From exponential fit of data in Figure 4.3, a lifetime of 34.80 ± 0.03 min
has been measured. Reference data, to compare this figure to, have not been found
available in literature.
4.3. Proof of principle

Figure 4.3: Decay curve of the chemiluminescent emission: signal generated in the whole imaging array versus elapsed time, after spotting a solution droplet of 70 µl of anti-mouse IgG HRP antibody and activating. The exponential fit is also shown, indicating an experimental life time of 34.80 ± 0.03 min.

4.3.3 ECL imaging

In figure 4.4 the first chemiluminescence images obtained with a CMOS sensor at room temperature are shown.

Insets a-d show respectively the chemiluminescent sample (anti-mouse IgG HRP antibody) at a time from sample activation corresponding to 0.05, 0.5, 1 and 2 lifetimes, and thus to a residual antibody mass of 95%, 60%, 37% and 13% respectively. The
Figure 4.4: A drop of anti-mouse IgG HRP antibody activated with Lumi-Light western blotting substrate exposed to the Vanilla sensor for 1 s for a inferred residual sample mass of 95% (a), 60% (b), 37% (c) 13% (d). Grey scale represents signal generated in each pixel expressed in Digital Number (DN).

residual antibody mass was inferred by means of the elapsed time from sample activation with Lumi-Light, the lifetime estimated in Section 4.3.2 and the initial mass. The grey scale represents the signal generated in each pixel, expressed in arbitrary Digital Number (DN).

4.3.4 Linear range, sensitivity and SNR

The linear range and the sensitivity of the Vanilla sensor to ECL light have been evaluated, by analysing the detector response over 5 hours as a function of the residual anti-mouse IgG HRP antibody inferred mass, calculated from known initial mass and the estimated lifetime (see Section 4.3.2) in a separate experiment. Figure 4.5 shows the detected signal per time unit and area unit in a Region of Interest (ROI) plotted against the residual anti-mouse IgG HRP antibody inferred mass.

Starting from a high mass level the sensitivity curve shows a linear behaviour (linear range), then approaches a saturation level defined by the detector noise floor. From
the data of Figure 4.5, the linear range of the system extends down to a lower limit, which corresponds to a Minimum Detectable Mass (MDM) of 0.5 ng. The higher limit of the linear region is limited by the amount of antibody used at the beginning of the experiment (6.4 ng). Thus, the antibody mass at which the curve of Figure 4.5 is expected to plateau, that is, the maximum value of the linear portion of the sensitivity plot or saturation limit, could not be determined due to the limited activity range of the sample used. The noise level, shown in figure 4.5 with a dotted line, has been measured independently as 22153 DN s$^{-1}$mm$^{-2}$. The slope of the linear region of the curve in figure 4.5 represents the system sensitivity measured as $(723.70 \pm 3.14) \times 10^4$ DN s$^{-1}$ ng$^{-1}$ mm$^{-2}$.

Figure 4.6 shows the Signal-to-Noise Ratio (SNR) as a function of the exposure time (in the range 3-120 s) measured at different fractions of lifetime of the chemiluminescent sample (from 0.05 to 5 lifetime), corresponding to a residual sample mass reported as a percentage in figure. It has been observed that the SNR, at each fixed exposure time, decreases exponentially with the residual mass.

4.3.5 Comparison with film-based images

A comparative analysis between western blotting images acquired with the Vanilla CMOS APS and conventional film-based images has been undertaken, in order to evaluate and quantify the improvement in the detection process.

A set of 32 blots of anti-human IgG HRP antibody ranging from 1.17 ng to 0.39 µg was exposed to the Vanilla detector after activation with Lumi-Light. This sample was imaged with conventional film emulsion for 5 s (see Figure 4.7a) and 1 min (see Figure 4.7b) respectively. Selected blots of this samples were also imaged with the Vanilla sensor in order to compare the response. Blot $I_2$ (2.73 ng of anti-human IgG HRP antibody) and $N_1$ (2.73 µg of anti-human IgG HRP antibody) were imaged with the Vanilla system for the same exposure time used for film images (see Figure 4.7c and d). Blot $I_2$ is not visible with film emulsion exposed for 5 s (Figure 4.7a), whereas it is visible when the blot is imaged with the Vanilla detector with the same exposure time.
Figure 4.5: The relative signal in the ROI after thresholding divided by the exposure time and the area are plotted versus the anti-mouse IgG HRP inferred antibody mass, inferred from the initial mass after exponential decay. A linear fit is shown with a straight line in the linear region. The noise level is also indicated with a dotted line.

Figure 4.7c), due to the a higher sensitivity of this APS than that of conventional film emulsion.

Exposing film emulsion for a longer time (1 min in Figure 4.7b) allows detection of all the activity values blotted on the sample. Nevertheless saturation occurs. In fact blot N1 is saturated in the film image exposed for 1 min (Figure 4.7 (b)). The same blot is not saturated when exposed to the Vanilla sensor for the same exposure time (Figure 4.7d), implying a larger dynamic range for the digital detector.

Furthermore, the two blots under investigation were imaged for 1 s with the Vanilla system (see Figure 4.7e and f) demonstrating that the system is capable of clearly detect both above the noise level, with a dramatic reduction in exposure time compared
Chemiluminescence detection using a large area CMOS APS

A proof of principle, as well as an evaluation of detection capabilities of CMOS APSs in western blotting has been presented in the previous section. However, the limited imaging area of the Vanilla detector makes this detection system unpractical to be used in the routine laboratory practice in life science. A larger imaging area is then required in order for this technology to impact upon this field. The DynAMITe CMOS APS, featuring an imaging area of 12.8×13.1 cm², increasable to 25.6×26.2 cm² when four sensors are tiled together (see Section 2.2), can offer the large imaging area required in...
Chapter 4. Large area CMOS APSs for Western Blotting

This section reports on the evaluation of the detection capabilities of the large area DynAMITe detector in western blotting.

4.4.1 Linearity and Quantum Efficiency

Before exposing the DynAMITe detector to ECL activated biological samples, the response of the detector to a light source with a wavelength close to that produced by chemiluminescence (425 nm) was tested to assess linearity, sensitivity and Quantum Efficiency (QE).

The DynAMITe detector was exposed to a calibrated Light Emitting Diode (LED) array with wavelength centred at 468 nm (bandwidth 35 nm).

Figure 4.8 shows the linearity plot for the DynAMITe detector exposed to blue light. The specific signal rate, after signal calibration in units of e\(^-\) (see Section 3.2.1), is plotted versus the source light irradiance. A linear fit has been performed for this data and is shown as a dotted line in figure 4.8. The DynAMITe detector presents a linear
response to blue light in the range $0.1-2 \times 10^5$ photons s$^{-1}$ pixel$^{-1}$ and a sensitivity of $(0.38\pm0.1)$ e$^-$/photon, calculated as the slope of the linear region in the response curve, corresponding to a quantum efficiency of 38% at 468 nm.

4.4.2 Biological samples and experimental set-up

Secondary anti-human IgG HRP antibody (A8667, Sigma-Aldrich), ranging from 20 to 0.1 nl diluted in Tris Buffered Saline, have been spotted on a Hybond - LEP membrane (Amersham Biosciences) in a dot blot manifold. Secondary antibody concentration is reported in Figure 4.9. The membrane was then washed in a Lumi-Light Western Blotting Substrate (Roche Applied Science) in order to activated the ECL emission.
Chapter 4. Large area CMOS APSs for Western Blotting

Figure 4.9: A map distribution of the sample concentration. Concentration of anti-human secondary antibody are expressed in nl.

The sample was prepared in duplicate to be exposed to the DynAMITe digital detector and film emulsion. The ECL activated membrane was then placed in close contact with the detector surface and imaged in a light-tight box. Multiple CMOS images were acquired at 1-second frame rate and dark corrected following Equation 4.2. The duplicated sample was exposed to film emulsion. Standard procedures for film development were adopted.

4.4.3 Chemiluminescence detection

Figure 4.10 shows a comparison between images of ECL activated antibody spots, obtained with film emulsion (left column) and with the DynAMITe detector (right column). The reference to the dot position is reported for all the images according to the distribution map of Figure 4.9. Images are shown for a 1-second (Figure 4.10 a, f), 3-second (Figure 4.10 b, g) 5-second (Figure 4.10 c, h), 30-second (Figure 4.10 d, i) and 1 minute (Figure 4.10 e, j) exposure time.

Lowest activity blots in Figure 4.10 (C2-C10) are not visible in the film image with 1 second exposure time (Figure 4.10 a). On the other side the same blots are clearly visible above the background with the same exposure time for the DynAMITe detector (Figure 4.10 f), thus showing a higher sensitivity of the digital detector compared to film. An horizontal profile of row C exposed for 1 second to film emulsion (a) and to DynAMITe (b) is reported in Figure 4.11. In 1-second film image only the two highest
Figure 4.10: Secondary anti-human antibody activated with Luminol imaged with film emulsion (left) and with the DynAMiTe detector (right). Images taken at several exposure times are reported: 1 s (a, f), 3 s (b, g), 5 s (c, h), 30 s (d, i) and 1 min (e, j). The reference to the dot position is reported for all the images according to the distribution map of Figure 4.9.
Figure 4.11: Horizontal profile of blots of row C (see Figure 4.9) imaged with film emulsion a) and the DynAMiTe detector b) for 1 second. It is worth noting that the profile for film emulsion is opposite to the one measured for the digital detector, as signal detection in film emulsion corresponds to a darkening of the emulsion.
activity blots (C12 and C11) are distinguishable as peaks above the background (4.11 (a)), whereas all the spotted blots are distinguishable in (4.11 (b)).

Film images at 5 s, 30 s and 1 min (Figure 4.10 b-e) show saturation for all the blots in rows A-B. This leads to a spread of the signal on the film resulting in an incapability to distinguish different blots. On the other side saturation is absent in all the images reported for DynAMITe (Figure 4.10 f-j) because of a larger dynamic range.

4.5 First images of protein sequences

In this later stage of the work, after characterisation of ECL light signal and related detector imaging performance, experiments of protein imaging are discussed.

4.5.1 Biological samples and experimental set-up

*Escherichia coli* (*E. Coli*) total proteins were denaturated by heating at high temperature and subsequently loaded onto a SDS-PAGE 12% Acrylamide gel in amount of 5, 10 and 15 µl. Protein gel was subjected to electrophoresis for 1 hour by an electric potential difference of 200 V. The electric field resulting from the potential difference allowed proteins to migrate through the gel, which behaves like a sieving medium, with a velocity depending upon their molecular weight.

Separated proteins were then transferred to a Polyvinylidene Fluoride (PVDF) membrane (Hybond - LEP membrane, Amersham Biosciences) by overnight electroblotting in order to be immuno-detected.

Following the procedure described in Section 4.1.1 and schematically represented in Figure 4.1, samples were incubated with Bovine Serum Albumin, as a blocking reagent to prevent non-specific binding and then washed with human serum containing primary antibodies and then with secondary antibodies (anti-human IgG HRP antibody A8667, Sigma-Aldrich). In order to allow chemiluminescence emission, the membrane was washed with Lumi-Light Western Blotting Substrate (Roche Applied Science). The sample was prepared in duplicate to be exposed to the DynAMITe digital detector and film emulsion. The ECL activated membrane was then placed in close contact with the
detector surface and imaged in a light-tight box. Multiple images were acquired at a 1-second frame rate and dark corrected following Equation 4.2. A duplicated sample was exposed to film emulsion. Standard procedures for film development were adopted.

4.5.2 Imaging of immuno-detected proteins

In Figure 4.12 the first images of western blotting of proteins via immuno-detection with a room temperature CMOS APS are presented. Membrane sections containing 5, 10 and 15 µl of proteins, together with a calibration marker, were imaged with both the DynAMITe detector and film emulsion, for a cumulative integration time of 30 s, 1 and 3 min. Film images, although windowed between minimum and maximum signal, are strongly affected by saturation in the longer exposure images, whereas the DynAMITe detector remains unsaturated due to a larger dynamic range. An intensity profile across the lines containing 5 µl of proteins is plotted in Figure 4.13 for both detection systems, obtained with 1 s exposure time for DynAMITe and 1 min exposure time for film emulsion. Detected peaks, numbered in figure, are comparable in number for both systems with an exposure time ratio of 60.

4.6 Comparative performance assessment

A further experiment on detection of western blotting of proteins was designed to effectively compare the detection performance of the DynAMITe detector with the detection systems routinely used in the field: film emulsion and a CCD based system (FluorChem Q, see Section 4.1.2 ).

4.6.1 Biological samples and experimental set-up

Denaturated E. Coli proteins were loaded onto a SDS-PAGE 12% Acrylamide gel in dilutions of 1:500, 1:1000, 1:1500 and 1:2000 to evaluate proteins detectability at different concentrations.
Figure 4.12: First images of optical labeled protein sequence imaged using the multi-modality imaging system. Samples were imaged with both the DynAMITe detector and film emulsion for 30 s, 1 min and 3 min exposure time. Imaging for both detection systems are displayed on a false colour scale and are windowed between maximum and minimum signal detected at the longest exposure time. Colour scales are expressed in Digital Numbers for both set of images.
Figure 4.13: Intensity profile across the lines containing 5 µl of *E. coli* proteins (see Figure 4.12) displayed for the DynAMITe detector (red line) using 1 s exposure time and for film emulsion (blue line) with 1 min exposure time. Detected peaks are numbered.

A similar procedure to that described in Section 4.5.1 was used to separate, transfer and activate the samples. A calibration marker was used to assess the molecular weight of detected proteins. Individual protein columns were excised from the original membrane, to separate protein columns by their concentration and expose a range of concentrations to each of the detectors under investigation. Figure 4.14 schematically shows this process. The CCD-based system (FluorChemQ) was exposed to 6 protein columns: 3 columns were for a 1:500 dilution and 3 for 1:1000. The DynAMITe detector was exposed to a total of 9 protein columns: 3 from the 1:500 dilution batch, 2 from 1:1000, 2 from 1:1500 and 2 from 1:2000. Film emulsion was exposed to the same dilutions as FluorChemQ, but only two columns were used for each dilution. A number of nine ECL activated protein columns (see Figure 4.14) were then placed in close contact with the detector surface and imaged in a light-tight box. Multiple images were acquired at a 1-second frame rate and dark corrected following Equation 4.2. A second fraction of the samples (four protein columns, see Figure 4.14) was exposed to film emulsion and standard procedures for film development were adopted. A third fraction of the samples (six columns, see Figure 4.14) was imaged with the FluorChemQ, using
4.6. Comparative performance assessment

Figure 4.14: A schematic representation of the samples exposed to the various detectors (FluorChemQ, DynAMiTe and film emulsion) with their respective dilutions and relative position in the imaging area.

the high sensitivity mode, i.e. with the shortest distance possible between samples and CCD detector. The same exposure time was used for all the three systems in order to compare sensitivity and dynamic range.

4.6.2 Detection capability

Protein samples were imaged with the three different detection systems under study (DynAMiTe detector, film emulsion and FluoChem Q) for different values of the exposure time: 30 s, 1 min an 5 min. Images acquired with the three systems are displayed in figure 4.15 on a logarithmic colour scale.

For the shortest exposure time (30 s) the first protein band (lowest molecular weight in the column) at the bottom of the column is visible for all the detection system used and is highlighted with a red oval. This corresponds to the red line at approximately 20 kDa in the gel image (left column of Figure 4.15), as calculated with a calibration marker. However, a second band of proteins is visible for the DynAMiTe detector at least at the lowest dilution (1:500). This is highlighted with a cyan arrow and corresponds to the 37 kDa line in the gel image.

At the following exposure time (1 min) this second protein band is still not visible for
Figure 4.15: Images obtained by exposing the protein samples to film emulsion, FluorChemQ and DynAMITe detector for 30 s, 1 min and 5 min. A logarithmic colour scale for the digital images in reported. On the left of the picture a schematic representation of a gel with molecular weights (coloured lines) reported as calculated with a calibrated marker. A red oval shows the protein bands of the lowest molecular weight (20 kDa) detected by all the systems. A cyan arrow shows the protein band at 37 kDa detected by the DynAMITe detector, while the yellow arrow shows the protein band at 50 kDa. Intensity profiles across the two grey dashed lines are shown in Figure 4.16.
4.6. Comparative performance assessment

Figure 4.16: Intensity profiles across the two dashed grey lines in Figure 4.15 for the FluorChemQ system (a) and for the DynAMiTe detector (b) for the lowest dilution (1:500) and for 5 min exposure time. Three peaks are clearly visible in the DynAMiTe system, compared to only one distinct peak in the FluorChemQ system.
Figure 4.17: Intensity profiles for the FluorChemQ system and the DynAMITe detector for the two dilutions: 1:500 (a) and 1:1000 (b). A linear fit is for each dataset is also reported, whose a gradient represents the system sensitivity.
the FluorChemQ system and for film emulsion. With 5 min exposure time a further protein band is detected by the DynAMITe system. This is highlighted with a yellow arrow and corresponds to the 50 kDa line in the gel image. At the same exposure time, film emulsion and the CCD-based system are still capable of detecting only the highest intensity protein band.

Profiles along the lowest dilution columns (1:500) and with the highest exposure time (5 min) for the DynAMITe detector (across the two dashed grey lines in Figure 4.15) and the FLuorChemQ system were drawn for a quantitative comparison. Intensity profiles for the two detection systems are shown in Figure 4.16. As seen previously, at the longest exposure time (5 min) the DynAMITe detector is capable of imaging 3 protein bands while the CCD-based system can just detect the highest intensity one. This can be explained as due to the lower sensitivity of the FluorChemQ system when compared with the DynAMITe detector. The former employs a small area (few cm$^2$) sensor and focusing optics placed at a large distance from the image sensor, thus suffering from low geometric efficiency, whilst the latter is based on the direct contact of samples with the detector surface, which facilitate much higher geometric efficiency for light collection.

The sensitivity curves for both detection systems are displayed in Figure 4.17 for two dilutions (1:500 and 1:1000), reporting signal measured in the highest intensity band as function of the exposure time. Both systems show a good linearity ($R^2=0.99$ for the linear fit of the these data), but sensitivity, which can be measured as the gradient of these curves, is significantly lower for the CCD-based system. In fact, at the lowest dilution (1:500), linear fit of the data of figure 4.17 results in a sensitivity of 12.5±0.6 DN/s for FluorChemQ and 30.64±2 DN/s for the DynAMITe detector. Sensitivity at the higher dilution (1:1000) is 1.39±0.04 DN/s and 6.8±0.6 DN/s, respectively.

4.7 Discussion

The viability of using a large area CMOS APS for protein detection in western blotting has been studied in this chapter. A proof of principle of using a CMOS APS for ECL
light detection has been provided using the small area Vanilla APS. Signal characterisation, as well as imaging performance such as linearity, dynamic range and SNR, have been presented in a comparative approach with the conventionally used film emulsion. From this first analysis APSs demonstrated potential as a suitable detection technology for this application.

However, western blotting requires a larger imaging area and for this reason further tests were conducted with the large area DynAMITe detector. Imaging experiments with ECL light showed a higher sensitivity and a larger dynamic large for this detector when compared to film emulsion. Finally, a comparative test showed that the DynAMITe detector is able to detect more protein bands compared to using standard film emulsion or when compared to a commonly used commercial camera-based western blotting detection system. Thus, CMOS APSs may be considered a viable alternative to commonly used detection system in western blotting, offering higher intrinsic performance.
Chapter 5

Radiation Hardness assessment

5.1 Motivation

In Chapter 3 an investigation into the imaging performance of the DynAMITe detector has been reported, with a particular focus on medical imaging applications. The imaging performance of this detector have been comparatively studied with respect to Flat Panel Imagers (FPIs), routinely used in clinical practice. From comparative data, the APS object of this study demonstrated a number of advantages compared to FPIs, including a lower noise floor and small pixel pitch resulting in a better DQE performance (section 3.3.6) and a comparable uniformity in terms of imaging capabilities over the whole imaging area (section 3.3.5). These advantages, together with the possibility for CMOS APSs to be designed to program operations at the pixel level, for developing on-chip intelligence, make APSs suitable for potentially replacing FPIs in clinical applications.

While CMOS APSs are indeed starting to be a commercial alternative to FPIs in radiological applications [26, 27, 32], the imaging needs in the medical field are also changing. In fact particle therapy, and more often proton therapy (see Chapter 6), are recently starting to spread worldwide with a significant increase in the number of new facilities being planned [64]. Thus, devices for treatment monitoring as well as for new imaging modalities (e.g. proton Computed Tomography (CT) [65]) are needed.
Requirements for CMOS APSs, in order to impact upon these broader medical imaging needs, become even more demanding with the requirements of a significant radiation tolerance to both X-rays and particle radiation fields. Several investigations have been carried out to assess the radiation tolerance of CMOS devices for X-rays and γ irradiation [66, 67] and for particle radiation fields[68, 69], as well as to propose new design techniques to enhance radiation hardness [36]. However none of these detectors is specifically designed for medical imaging, and as, such, application-specific requirements in terms of noise, resolution, imaging area are not fulfilled.

In this chapter the radiation hardness-by-design [36] of the DynAMITe detector is presented and its radiation tolerance studied when irradiated with X-rays and protons. From the point of view of radiation damage, while X-rays produce only ionisation damage in the detector, proton fields can be considered as a mixed field. In fact, as protons travel through the detector volume, ionisation charge is produced resulting in the build-up of charge trapped in the oxide and interstate charge at the interface between the silicon and silicon dioxide [36]. In addition to ionisation, protons also interact by means of Non Ionising Energy Loss (NIEL) with the detector crystalline structure, producing displacement of a recoil atom from its lattice position which results in the generation of new energy levels in the band-gap of the semiconductor leading to alterations of its intrinsic electrical properties [70].

Thus, radiation damage produced by ionisation and by non ionising processes is studied in this Chapter to assess the DynAMITe sensor for use in highly ionising bio-medical applications. Particular emphasis is placed upon the separation of ionising and non-ionising contributions to the radiation damage, in order to predict the sensor response when exposed to a different radiation field.

The rest of this chapter is organised as follows: Section 5.2 provides an overview of the damage mechanisms in Silicon detector produced by ionising and non ionising radiation together with details on the radiation hardness design of the DynAMITe detector;
in Section 5.3 the experimental procedure used in the study is reported. Section 5.4 reports on the radiation damage produced by X-rays and the performance compared with a commercial CMOS APS for radiology applications. Damage produced by proton irradiation is reported in Section 5.5, while comparison between ionising and non ionising radiation damage and separation of the two contributions in a mixed radiation field is discussed in Section 5.6. An estimation of the operational life of the this detector for MegaVoltage radiotherapy and proton CT is discussed Section 5.7, followed by conclusion in Section 5.8.

5.2 Radiation damage in Silicon sensors

Particles and photons, traveling through a medium, can deposited energy by means of collisional energy loss or atomic displacement [70]. When the energy loss results in the excitation or emission of an atomic electron, the energy loss process is referred to as Energy Loss by ionisation. Conversely energy loss processes are called Non-ionising Energy Loss or NIEL when these involve atomic displacements or collisions where the knock-on atom does not move from its lattice, dissipating the imparted energy as lattice vibrations.

5.2.1 Ionising damage

The ionising radiation effects on CMOS devices are mainly related to the build-up of charge trapped in the oxides and interstate charge at in the interface between silicon and silicon dioxide (Si/SiO$_2$) [36]. The process of radiation damage by ionising energy deposition can be described as a series of four consecutive steps [71], schematically represented in Figure 5.1.

Firstly ionising particles, passing through the CMOS device, generate electron-hole (e-h) pairs along their track. Immediately after e-h pair creation, a fraction of these recombine (Figure 5.1c). This phenomenon happens in a very short time window (in the order of 0.1 ps), limited by the time needed to remove electrons from the gate
Figure 5.1: The charge distribution in a gate oxide at three time points following exposure to a pulse of irradiation \((t = 0)\) for a thick gate oxide biased positively at the gate electrode. (b) Initially after irradiation electron-hole pairs are generated throughout the oxide. (c) After a few picoseconds a fraction of the electron-hole pairs recombine. A fraction of the holes tunnel out of the oxide and the electrons are swept out of the oxide. (d) The remaining holes drift to the SiO2/Si interface, where they are either captured in deep traps, interact to form interface states, or tunnel out of the oxide. Courtesy of [71].

Figure 5.2: A schematic diagram showing the creation of Frankel pairs in the Si crystalline structure.
5.2. Radiation damage in Silicon sensors

oxide under the gate electric field, given their higher mobility compared to that of holes. The amount of recombination happening in this narrow time window is a function of the density of ionisation (Linear Energy Transfer or LET) and gate electric field.

After recombination has taken place, in step two, the only charge carriers left in the gate oxide are holes, which drift towards the Si/SiO₂ interface under the gate electric field, on a time scale several decades longer compared to that of recombination given their lower mobility compared to that of electrons. As a result of the first two steps, a fraction of holes remain trapped at the Si/SiO₂ interface (oxide trapped charge shown in Figure 5.1c). This trapped positive charge, can be neutralised by electrons tunnelling out from silicon or by thermal emission of electrons from the oxide valence band (step three). The possibility of oxide trapped charge neutralisation suggests that ionisation radiation damage can be mitigated by annealing at either room or at higher temperature. Finally in the fourth step, formation of inter-state traps can occur, being related to dangling bonds between silicon and silicon dioxide [72].

Generation of charge density in oxides affects CMOS device operations at different levels: trapped and inter-state charges produce degradation of transistor performance if localised in the gate oxide, resulting in an increase of dark current, voltage threshold and sub-threshold voltage swing shifts, and a decrease in Maximum Output Voltage Swing. Trapped charge localised in the oxide, which provides transistor insulation, result in the generation of inter-device and intra-device parasitic currents, contributing to pixel dark current increase [36, 71].

5.2.2 Non-ionising damage

Non-ionising or displacement damage occurs when particle interaction with the detector crystalline structure results in the displacement of a recoil atom from its lattice position, leading to the generation of new energy levels in the band-gap of the semiconductor [70]. Primary knock-on atoms (PKAs) create in the silicon lattice a so called

\[ \text{Mobility in crystalline Si in typically } \leq 1400 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \text{ for electrons, while for holes this value is } \leq 450 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \]
Chapter 5. Radiation Hardness assessment

**Frenkel pair (FP):** a pair of defects consisting of a vacancy and interstitial in the crystal lattice. A schematic representation of this process is provided in Figure 5.2. FPs can migrate through the lattice to either form a stable point defect or annihilate if interstitials and vacancies recombine.

Displacement damage is a threshold process: creation of a FP can occur only if the deposited energy is higher than the displacement threshold energy, estimated in the order of 13-33 eV for silicon [70]. If energy imparted to the crystal lattice is much higher than the damage threshold, then a cascading-displacement occurs with creation of closely spaced groups of defects in a small spatial region (**defect clusters**).

Defects induced in the crystal lattice periodicity give rise to new energy levels in the bandgap of the irradiated semiconductor, thus altering its intrinsic and electrical properties [73]. Major effects in semiconductor devices affected by displacement damage are listed below:

1. Thermal generation of e-h pairs through a level near midgap, contributing to carrier generation and thus to leakage current increase.
2. Enhancement of e-h recombination. A free carrier of one sign can be captured at a defect centre and then followed by capture of a carrier of opposite polarity. This process reduces charge carrier's lifetime, leading to gain reduction.
3. Defects can act as scattering centres, reducing charge mobility.
4. Compensation of acceptors/donors in the irradiated semiconductor results in alterations of the majority carrier concentration, inducing type inversion in the semiconductor depleted region (e.g. from *p* to *n* type).

**5.2.3 NIEL scaling hypothesis**

The NIEL scaling hypothesis is based on the assumption that displacement damage scales linearly with energy imparted in collisions, regardless of spatial distribution of defects and specific annealing sequence taking place. This assumption represents a
powerful tool for studying radiation damage, as it allows evaluation of damage independently from particle type and energy.

Prior literature (see [70] and references therein) suggests that the damage effect due to NIEL is expressed by the damage function \( D(E) \) in units of MeVmb:

\[
D(E) = \sum_k \sigma_k(E) \int f_k(E, E_R) P_k(E_R) dE_R
\]

(5.1)

where \( E \) is the incoming particle energy, \( \sigma_k(E) \) is the cross-section for the k-th reaction occurring in the cascade, \( f_k(E, E_R + dE_R) \) is the probability that a recoil atom is generated with kinetic energy between \( E_R \) and \( E_R + dE_R \), and \( P_k(E_R) \) is the partition energy function for the recoil nuclei following the Linhard screened-potential scattering theory.

The damage function can be directly related to NIEL by the displacement KERMA (Kinetic Energy Released to the Matter)²:

\[
D(E) = \frac{A}{N_A} \frac{dE}{dx}(E)|_{\text{non-ion}}.
\]

(5.2)

where \( A \) is the atomic weight of Si and \( N_A \) is the number of atoms per cm\(^3\) in the bulk Si. Using \( D(E) \) it is possible to define a hardness factor \( k \) to compare damage efficiency of different radiation sources of different energy:

\[
k = \frac{\int D(E) \phi(E) dE}{D_n(1\text{MeV}) \int \phi(E) dE}
\]

(5.3)

where \( \phi(E) \) is the radiation incoming energy spectra and \( D_n(1\text{MeV}) = 95\text{MeVmb} \) is the damage value for 1 MeV neutrons, assumed as in reference [74]. By means of the factor \( k \) it is possible to define a 1 MeV neutron equivalent fluence \( \phi_{eq} \) for any radiation source of energy spectra \( \phi(E) \):

\[
\phi_{eq} = k \int \phi(E) dE.
\]

(5.4)

This figure can be interpreted as the equivalent 1 MeV neutron fluence required to produce the same displacement damage produced by a certain fluence of a certain

²For silicon with \( A=28.086 \text{ g/mol} \), the relation between NIEL and D is the following: \( 100\text{MeVmb} = 2.144\text{keVcm}^2\text{g}^{-1} \)
radiation source.

In this work the radiation tolerance to particle fields, evaluated for the sensors under investigation, will be expressed in $\phi_{eq}$.

### 5.2.4 The radiation hardness design of the DynAMITe detector

CMOS design for scientific and especially space applications, where a significant radiation tolerance is required, have usually relied on dedicated foundry processes to deliver radiation hardness. However, the inherent complexity of these processes, together with the relatively low volume of production, has made them lag behind in terms of minimum feature size scaling [75], compared to consumer based integrated circuit applications [14].

For this reason new techniques for radiation hardness, namely radiation hardness-by-design (HBD) and based on standard foundry processes, have started becoming popular in the last decade [36]. HBD is a CMOS design approach to mitigate the effect of radiation damage using specific layout techniques at the transistor, component and device level through standard foundry processes.

The DynAMIte detector has been designed according to the HBD methodology. In fact all the in-pixel transistors have been designed with source and drain physically enclosed using an Enclosed Layout Geometry (ELG) [36, 76] in order to reduce the edge-leakage, which is generated in the transition area between the gate oxide and the isolation oxide (Shallow Trench Isolation), used to produce transistor-by-transistor insulation, after exposure to radiation. P$^+$ doped guard rings have been added in around the diode to prevent radiation induced inter-device leakage current.

Moreover, the Sub-Pixel photo-diode has been split in four interconnected diodes, placed at pixel corners, to mitigate the reduction in charge collection efficiency, an effect of the displacement damage, by shortening the charge carrier’s path to the collection node [68].
5.3 Irradiation experiments

5.3.1 X-ray irradiation

The DynAMITe detector has been exposed to an X-ray field (W anode, 160 kVp, 0.5 mm Cu filtration) up to a cumulative dose of 94 kGy(Si). An average of 15 hours of exposure per day was performed while the detector was biased, reset and readout in order to recreate actual operational conditions.

The detector was mounted with a scintillator (140 µm thick Gd$_2$O$_2$S:Tb) and a 3 mm Fiber Optic Plate (FOP). Performance parameters such as dark current, offset, gain and dynamic range have been evaluated at each exposure step for two separate region of interest (ROIs) of approximately 1 cm$^2$ area each. The rest of the detector was covered with a 2 mm thick Tungsten shield.

The dark current was measured by subtraction of two dark images, one with a long integration time and one with a short integration time, in order to correct for the remaining Fixed Pattern Noise. The dark current measured with this procedure was converted into dark current density using the conversion gain of the sensor calculated at each exposure step.

For comparison, a commercial APS designed for bio-medical applications, has been irradiated in the same test experiment up to a cumulative dose of 0.8 kGy (Air Kerma), when it showed loss of light sensitivity and pixel resolution. The maximum delivered dose in this experiment (94 kGy(Si)) represents the highest deliverable dose within the time allocated for this experiment.

5.3.2 Proton irradiation

The DynAMITe detector has been directly exposed, i.e. without scintillator and FOP, to proton beams at the MC40 cyclotron at the University of Birmingham. A proton beam with an average energy of 29.4 MeV, collimated to an area of 1 cm$^2$, was used to
Table 5.1: Total ionising doses (dose to water \(D_{W}\) and to silicon), and proton fluences delivered in the proton irradiation experiments.

<table>
<thead>
<tr>
<th>(D_{W}) (kGy)</th>
<th>Dose (kGy(Si))</th>
<th>Fluence (p cm(^{-2}))</th>
<th>(D_{W}) (kGy)</th>
<th>Dose (kGy(Si))</th>
<th>Fluence (p cm(^{-2}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.3</td>
<td>(3.3 \times 10^{10})</td>
<td>0.1</td>
<td>0.7</td>
<td>(6.8 \times 10^{8})</td>
</tr>
<tr>
<td>0.3</td>
<td>0.8</td>
<td>(9.8 \times 10^{10})</td>
<td>0.3</td>
<td>2.2</td>
<td>(2.0 \times 10^{9})</td>
</tr>
<tr>
<td>0.6</td>
<td>1.6</td>
<td>(1.9 \times 10^{11})</td>
<td>0.6</td>
<td>4.4</td>
<td>(4.0 \times 10^{9})</td>
</tr>
<tr>
<td>1.0</td>
<td>2.6</td>
<td>(3.2 \times 10^{11})</td>
<td>1.0</td>
<td>7.3</td>
<td>(6.7 \times 10^{9})</td>
</tr>
<tr>
<td>1.9</td>
<td>5.3</td>
<td>(6.5 \times 10^{11})</td>
<td>3.0</td>
<td>22.1</td>
<td>(2.0 \times 10^{10})</td>
</tr>
<tr>
<td>3.0</td>
<td>8.0</td>
<td>(9.8 \times 10^{11})</td>
<td>5.0</td>
<td>36.7</td>
<td>(3.4 \times 10^{10})</td>
</tr>
<tr>
<td>5.0</td>
<td>13.2</td>
<td>(1.6 \times 10^{11})</td>
<td>15.0</td>
<td>110.0</td>
<td>(1.0 \times 10^{11})</td>
</tr>
</tbody>
</table>

The DynAMITe sensor was also exposed to a lower energy beam (1.9 MeV), obtained by degrading the unattenuated 29.4 MeV beam with a 6.6 mm PMMA absorber, up to an integral dose to water of 15 kGy. A summary of dose delivered in these experiments, both dose to water \(D_{W}\) and to silicon, together with the relevant proton fluences is reported in Table 5.1.

After each irradiation step, the sensor response was characterised in terms of electro-optical performance, via integration time sweep under constant illumination, to estimate conversion gain, full well capacity and quantum efficiency (see Section 3.2.1), and in dark conditions, to assess dark current.

### 5.3.3 Dosimetry

Dosimetry for X-rays and protons was carried out at both facilities by means of a transmission ionisation and Markus chamber respectively, providing readings of ionising dose in air and in water at each irradiation step.
5.4. X-ray irradiation

In order to correlate measurements of dose to air/water to dose deposited in the detector or in components of interest (dose to Silicon or dose to SiO₂), Monte Carlo simulations were performed using the GEANT4 toolkit[77].

Simulation were run to calculate air kerma \( (AK^{\text{sim}}) \) for X-rays irradiation or dose to water for proton irradiation \( (WD^{\text{sim}}) \), and energy deposition in the \( i \)-th detector layer (e.g. SiO₂, epitaxial Si etc.) as function of simulated air kerma and dose to water \( (E_{\text{dep}}^{i}(AK^{\text{sim}}) \text{ or } E_{\text{dep}}^{i}(WD^{\text{sim}})) \). This allows to measured air kerma \( (AK^{\text{meas}}) \) and dose to water \( (WD^{\text{meas}}) \) to be related to deposited energy, and thus to the total ionising dose deposited in the detector volume (TID(Si)) or to specific detector components (e.g. TID(SiO₂)).

Moreover, for the X-ray irradiation experiment, secondary radiation generated as fluorescence X-ray in the shield, scintillator and FOP was taken into account and separately analyzed per generating chemical element.

The Displacement Damage Dose (DDD) was calculated for both proton irradiation experiments, by means of the NIEL scaling hypothesis, for a proton fluence \( \phi_p \) of energy \( E \) as:

\[
DDD = \phi_p(E) \cdot KERMA(E)|_{\text{non-ion}}
\]

(5.5)

where the damage KERMA was derived from the damage function \( D(E) \) in equation 5.1 [70].

5.4 X-ray irradiation

Following X-ray irradiation as described in Section 5.3.1, results are presented below.

5.4.1 Comparison with a commercial CMOS imager

Comparative data for dark current and dynamic range are shown in Figure 5.3 and 5.4 for both the DynAMiTE detector and a CMOS APS device commercially available for
radiology applications.

Figure 5.3 demonstrated exponential increase in dark current for the commercial device (*Comm. dev.*) up to a dose of 700 Gy, where loss of light sensitivity and pixel resolution occur (dotted line in Figure 5.3 and 5.4). At this dose the relative increase in dark current for the DynAMITe detector is lower than 150% (100 times lower than the increase observed in the commercial device) and is almost constant up to the maximum delivered dose shown in figure 5.3 (1.8 kGy (Si)).

The residual dynamic range\(^3\), calculated as a fraction of the initial dynamic range, is shown in Figure 5.4. The DynAMITe detector exhibits a residual dynamic range of about 99% of that seen at pre-irradiation, up to 1.2 kGy Air Kerma (1.8 kGy(Si)), whereas the commercial device falls below the failure limit of 50% at 650 Gy. The DynAMITe detector has been further irradiated up to a dose of 94 kGy(Si) without showing any functionality issues with a residual dynamic range of 88% of pre-irradiated performance.

### 5.4.2 Dark current

Dark current density has been measured at each irradiation step to evaluate radiation-induced damage. The pre-irradiation dark current density of the DynAMITe detector is 59 pA/cm\(^2\), a figure which increases up to 293 pA/cm\(^2\) at the maximum delivered dose of 94 kGy(Si).

Figure 5.5 shows the dark current density increase as a function of the total ionising dose to silicon TID(Si) on a semi-logarithmic scale. The dark current increase shows a threshold behaviour with TID: it is constant at the first doses delivered and then starts increasing. This mechanisms would seem contradictory compared with TID induced damage mechanisms, where the number of trapped charge in the oxide and thus the dark current should increase continuously with TID, and with some literature [78, 79, 80] . However, Bogaerts et al. reported a similar threshold mechanism in

\[^3\] Dynamic range is defined as \(20 \log(FWC/\sigma_r)\), where \(FWC\) is the Full Well Capacity and \(\sigma_r\) is the read noise (see Section 3.2.1)
5.4. X-ray irradiation

Figure 5.3: Relative increase in dark current reported for the DynAMITe detector and a commercial device (Comm. dev.) in both the exposed ROIs.

Figure 5.4: Residual dynamic range reported for the DynAMITe detector and a commercial device (Comm. dev.) in both the exposed ROIs.
A physical explanation to this phenomenon could be related to the presence of an additional threshold mechanism, other than build-up of trapped charge in the oxides, which suddenly enhances dark current, e.g. depletion of an interface. It could also be the case of a parasitic dark current in the pre-irradiated device which shadows the genuine increase in dark current with TID up to a point (threshold) where the dark current increases with TID is dominant.

The dark current increase due to the total ionising dose $TID$ ($\Delta C(TID)$) has been fitted, over an assumed linear range of values, with the empirical formula [67]:

$$\Delta C(TID) = \max\left\{0, K \log_{10} \frac{TID}{TID_{thre}}\right\}$$  \hspace{1cm} (5.6)

where $TID_{thre}$ is the total ionising dose threshold, necessary to observe a dark current density increase, and $K$ is the dark current increase per decade. Fitting of data of Figure 5.5 has resulted in evaluation of a radiation damage threshold $TID_{thre}=204$ Gy(Si) and a dark current density rate $K=96 \pm 5$ pA/cm$^2$/decade.

Recent studies on TID induced dark current [69, 81] have shown a monotonic increase of dark current with TID, suggesting that the cause of degradation remains constant with TID. Two main contributions to the dark current can be highlighted: one arising from the junction perimeter and the other from the junction area. In those recent studies [82, 81], as well as for older technologies [83, 84], it has been shown how the TID induced dark current is proportional to the perimeter junction, being this contribution the dominant one. The dark current current increase with TID for the device under study has then to be considered generated in the peripheral depleted Shallow Trench Isolation (STI) interface, where the depletion region touches the STI interface. For this reason, the use of a pinned photodiode, where the depleted region is not in contact with the surrounding oxides, could represent an improvement in TID induced dark current.

Dark current density histograms are displayed in Figure 5.6 for six delivered doses (0, 0.6, 1, 10, 50, 94 kGy(Si)). Two different phenomena are observable in the plot: a shift of the dark current density distribution towards higher current values, and a broadening of the distributions. It is also to note that the dark current density dis-
5.4. X-ray irradiation

5.4. X-ray irradiation

The distribution at 1 kGy(Si) (green curve) is shifted to the left of the distribution at the previous irradiation step (0.6 kGy(Si), red curve). This behaviour can be related with self-annealing occurring in early device life with heat generation resulting in emptying deep traps formed in the oxides during device fabrication. After a Gaussian fit to these distributions, the coefficient of variations, namely the ratio between the standard deviation and the mean of the corresponding Gaussian distributions, is plotted in Figure 5.7 as a function of TID. The highest coefficient of variation occurs for the pre-irradiation distribution (18%), whereas this value vary randomly between 8 and 11% for all other dose levels implying no significant spatial non-uniformity arising from radiation damage produced up to 94 kGy(Si), in agreement with [69].

![Figure 5.5: Dark current density increase as a function of the total ionising dose to silicon (semi-logarithmic scale).](image)

5.4.3 Fluorescence

For the X-ray irradiation experiment, the detector was not directly exposed to the radiation source but a number of additional elements were added in order to use a realistic experimental setup (e.g. FOP, scintillator) or to protect unexposed regions of the sensor (e.g. shields). However, the introduction of this additional elements between source and detector might contribute to the total dose delivered to the sensor, through
Figure 5.6: Dark current density histograms at six ionising doses.

Figure 5.7: Coefficient of variation of the dark current density distributions evaluated in Figure 5.6.
secondary radiation production. For this reason the contribution of fluorescence X-rays to the total energy deposited in the detector has been studied.

Energy absorption in the detector layers, as calculated from Monte Carlo simulations described in Sec 5.3.3, is shown in Figure 5.8 as percentage of the integral energy emitted by the source. The maximum of energy deposition occurs for shield (18%), scintillator (13%) and FOP (31.5%) as expected because of their thickness and atomic number, compared to the other detector layers. The percentage of energy absorbed in silicon (including silicon nitrite and oxide) is 0.23%.

The relative contribution to the total energy absorbed in silicon from the X-ray fluorescence component is shown in Figure 5.9, displaying the percentage of energy deposited in each of the detector layers, due to fluorescence generated in the shield, scintillator and FOP. The main contribution to fluorescence is represented by the FOP (between 25 - 18%), because of the relatively high Z elements of which it is made (Pb, La, Zr, Ba), whereas the scintillator and shield contributes for a fraction evaluated in the range 0.1-0.7%.

Figure 5.10 shows the contribution to the energy absorbed in the detection layer due to fluorescence X-rays generated in the FOP per chemical element. Lead represents the first source of fluorescence in the FOP (71-65%) because of its relatively high abundance (30%), followed by Lanthanum (20-33%) present in the FOP in a percentage of 12%. Barium and Zirconium offer a smaller contribution (2-0.4%) depending on their smaller abundance in the FOP. Even so L-shell fluorescence X-rays due to these two elements are in the range 2-6 keV, implying a high cross section for such photons.

Although a non negligible dose to the detector is due to fluorescence X-rays generated in the FOP, the choice of using such a FOP arises from the need to reproduce an experimental set-up as close as possible to a conventional set-up in X-ray imaging.

5.5 Proton irradiation

Following proton irradiation as described in Section 5.3.2, results are presented below.
Figure 5.8: Simulated energy absorption in the detector layers. The logarithmic color scale represents the energy absorbed as percentage of the integral energy emitted by the source. Figure is not to scale.

Figure 5.9: Simulated relative contribution to the total energy absorbed in silicon from fluorescence X-ray produced in shield, scintillator and FOP.
5.5. Proton irradiation

Figure 5.10: Simulated relative contribution to the total energy absorbed in silicon from fluorescence X-ray generated in the FOP per chemical element.

5.5.1 Electro-optical parameters

Mean-variance curves [41] are shown in Figure 5.11 for the Dynamite sensor exposed to 29.4 MeV protons, for each of the irradiation levels reported in Table 5.1. Two different phenomena can be observed in Figure 5.11: a decrease in conversion gain in $e^-/DN$ (reciprocal of the slope in the linear region of the mean-variance curve) and a decrease in full well capacity (maxima of the mean-variance curve). The former can be explained as a change in the photodiode capacitance, related to the depletion of the isolation oxide and generation of an inversion channel, as discussed in [78]. The latter can be due to radiation-induced changes in and Maximum Output Voltage Swing (MOVVS) [36], as reported in [79, 80]. However, a validation of these hypotheses would require direct capacity measurements as well as direct analog voltage measurements on the sensor output. Such measurements can not be performed on the sensor under study.

Conversion gain, extrapolated from the curves of fig. 5.11, was used to calculate the linearity response for this sensor. Exposing the sensor to a uniform white light field, signal level has been recorded as a function of the exposure time, which is proportional to the number of optical photons impinging the sensor (see fig. 5.12). The mean signal
Chapter 5. Radiation Hardness assessment

Figure 5.11: Mean-variance curve for the DynAMITe detector exposed to 29.4 MeV protons for each irradiation level of Table 5.1.

Table

at the shortest exposure time (153 ms) goes from $4 \times 10^4 e^-$ in the unexposed detector (blue symbols in fig. 5.12) to $7 \times 10^2 e^-$ at the highest dose delivered (black symbols). The slope of this curve, when the signal was calibrated in units of $e^-$, represents the detector Quantum Efficiency (QE) [41]. Fig. 5.12 shows how QE decreases as radiation dose increases. This can be explained as a sum of two independent effects: decrease in conversion gain, due to changes in the photo-diode capacity [78], and decrease in collection efficiency. This effect can be either due to a decrease in charge carrier’s lifetime due to the bulk damage generated in the sensitive volume [68, 85], although marginal at this fluencies, or to a change in surface recombination velocity due to an increase in the dielectric/Si interface state density increase, as concluded in [80, 86].
5.5. Proton irradiation

Figure 5.12: Linearity curve for the DynAMITe detector exposed to 29.4 MeV protons for each irradiation level of Table 5.1.

Figure 5.13: Dark current density increase for the DynAMITe detector when irradiated with 160 kVp Xrays and 29.4 MeV protons.
5.6 Ionisation versus displacement damage effects

Radiation damage effects due to proton irradiation, reported in the previous section, are the result of two different damage mechanisms: ionisation and displacement damage. In order to discriminate these two contributions to sensor damage, proton irradiation experiments were compared with X-ray irradiation.

5.6.1 Dark current

Dark current increase as a function of the Total ionising Dose to Silicon (TID(Si)) is shown in Figure 5.13 for the DynAMITe sensor exposed to 160 kVp X-rays and 29.4 MeV protons. The dark current increase in this figure is reported in units of pA cm$^{-2}$, thus including the changes in conversion gain. As shown previously, the dark current increase for the DynAMITe sensor irradiated with X-ray (blue symbols in Figure 5.13) presents a damage threshold at $TID_{thre} = 204kGy(Si)$. After this point the increase in dark current is linear with $log_{10}(\frac{TID}{TID_{thre}})$ with a slope evaluated as $K_1 = 96 \pm 5pA/cm^2/\text{decade}$.

The increase in dark current, measured after protons irradiation (red symbols in Figure 5.13) shows a behaviour comparable with the one observed for X-ray irradiation. The increase in dark current due to proton irradiation is compatible with a damage threshold mechanics. However an independent evaluation of the damage threshold dose for 29.4 MeV protons is not possible, since the lowest dose delivered (268 Gy(Si)) falls already within the linear region. Thus the same damage dose threshold is assumed for both radiation fields.

The dark current increase curve for proton irradiation consists of two different regimes: first, a region (up to 1.5 kGy(Si)) where the dark current increase is linear with $log_{10}(\frac{TID}{TID_{thre}})$ and a second region where the dark current falls off. A possible explanation for this reduction in dark current could be the generation of a parasitic negative current in the PMOS reset transistor, which balances off the increase in dark current at diode level. A linear fit of $K_2log_{10}(\frac{TID}{TID_{thre}})$ is performed in the linear region,
resulting in a dark current increase rate of $K_2 = 40 \pm 5 \text{pA/cm}^2/\text{decade}$.

The dark current increase rate $K_2$ for the DynAMITe detector irradiated with 29.4 MeV protons, is significantly lower (by a factor 1.7), when compared to that produced by X-ray irradiation. This difference can be ascribed to the generation of a parasitic negative current in the PMOS reset transistor, which compensated the increase in dark current due to TID. 5.6.3.

![Figure 5.14: Dark current density distributions for the DynAMITe detector when irradiated with 29.4 MeV protons and 160 kVp X-rays for an integral TID of 300 Gy(Si).](image)

### 5.6.2 Dark current non-uniformity

Non ionising energy deposition in CMOS sensors is known to produce two effects related to the device dark current: increase of average value of dark current and introduction of non-uniformities, appreciable as dark current spikes [85, 87].

The dark current density distributions for the DynAMITe sensor exposed to 29.4 MeV protons and 160 kVp X-rays for an integral TID of 300 Gy(Si) are displayed in Figure 5.14. As seen previously (Figure 5.6) at this dose, and also over the whole range of
Figure 5.15: Dark current density distributions for the DynAMITe detector when irradiated with 29.4 MeV protons for several TID.

TID investigated, dark current density distributions after X-ray irradiations do not show any significant radiation-induced non-uniformity. However, as result of proton irradiation, the dark current density distribution for 300 Gy(Si) (Figure 5.14) shows significant non-uniformity. In fact, while the dark current of the majority of pixels is distributed according to a normal distribution comparable to that of X-ray irradiated pixels, a fraction of these corresponds to a tail with a quasi-exponential decrease.

Figure 5.15 shows the dark current density distributions after irradiation with 29.4 MeV protons for several values of proton fluences. The unexposed dark current density (continuous line in Figure 5.15) is normally distributed, suggesting a uniform dark current distribution of the non-irradiated device. As TID increases (and so proton fluence), the ratio between peak area and tail area decreases, showing an incremental non-uniformity increase with proton fluence. However at proton fluence larger then $2 \times 10^{11} \text{pcm}^{-2}$ (TID=1.6 kGy(Si)), the generation of a parasitic negative current in the reset transistor leads to an inversion in trend for the dark current density distribution as also seen for the average dark current in Figure 5.13.
The quasi-exponential behaviour of the dark current density tails can be related to the concept of probability of generation of electro-active defects in the pixel array [87, 85]. In fact dark current density measurements of Figure 5.15 show that the number of dark current spikes increases with proton fluence, and thus with the number of defects generated in the diode depleted volume.

### 5.6.3 Conversion gain

![Conversion gain graph](image)

Figure 5.16: Relative gain variation for the DynAMITe detector irradiated with 160 kVp X-rays, 29.4 MeV and 1.9 MeV protons.

Figure 5.16 shows the conversion gain as a function of the TID for the SP pixels of the DynAMITe sensor irradiated with 160 kVp X-rays, 29.4 MeV protons and 1.9 MeV protons. The ratio between measured gain at a certain TID level and unexposed gain shows a different trend for proton and X-ray irradiation.

When pixels are exposed to X-ray radiation, the decrease in gain is slower and linear, decreasing with a $\log_{10}(TID)$ decay. Data for proton irradiation suggest a threshold mechanism and a fast drop in gain (with a higher slope compared to X-ray irradiation).

In the case of X-ray irradiation, only ionising effects, i.e. the build-up of charge trapped
Figure 5.17: Relative gain variation for the DynMITe detector, as a function of DDD, when irradiated with 29.4 and 1.9 MeV protons.

in the dioxide and the interface-state charge, affect the gain variations. For proton irradiation these ionising effects have to be summed up to non-ionising energy deposition and generation of displacement damage in the bulk Si.

The conversion gain can then be expressed as function of both TID(SiO$_2$) and DDD:

$$G(TID(SiO_2), DDD) = G_0 - \Delta G(TID(SiO_2)) - \Delta G(DDD)$$

where $G_0$ is the gain of the unexposed sensor, $\Delta G(TID(SiO_2))$ is the variation in gain produced by ionising energy deposition in the dioxide top layers and $\Delta G(DDD)$ is the change in gain due to displacement damage in bulk Si.

From the experimental data for X-rays of Figure 5.16, $\Delta G(TID(SiO_2))$ is evaluated by means of a linear fit of the function

$$\frac{\Delta G(TID(SiO_2))}{G_0} = a + b \cdot \log_{10}(TID(SiO_2))$$

giving a linear coefficient $b = -0.056 \pm 0.003 \text{dec}^{-1}$. 
5.7 Detector lifetime

Once the gain variation due to ionisation is known as a function of dose deposited to the dioxide top layers, it is possible to isolate the contribution of DDD ($\Delta G(DDD)$) in the proton irradiated sensor by means of eq. 5.7.

Figure 5.17 shows the relative change in conversion due to displacement damage ($\Delta G(DDD)/G_0$) for the DynAMITe detector exposed to 29.4 and 1.9 MeV protons. Values of $\Delta G(DDD)/G_0$ for 1.9 MeV protons show a threshold damage mechanism and then a linear behaviour with the logarithm of DDD.

The threshold level is not evident for the high-energy irradiation, since the lowest dose delivered falls in the linear region. Displacement damage dose threshold has been estimated for the low energy data as $DDD_{thre} = 66.9 TeVg^{-1}$ and is assumed to be independent from proton energy since DDD is calculated in the NIEL scaling hypothesis (Equation 5.5), thus independent from particle type and energy in the energy range considered (up to 30 MeV).

A linear fit of the following relation was performed for both data series:

$$\frac{\Delta G(DDD)}{G_0} = a + K\log_{10}\left(\frac{DDD}{DDD_{thre}}\right)$$  \hspace{1cm} (5.9)

giving a linear coefficient $K(29.4 MeV) = 0.6 \pm 0.2 dec^{-1}$ for higher energy protons and $K(1.9 MeV) = 0.5 \pm 0.2 dec^{-1}$ for lower energy ones. The fit parameters are consistent for the two energies data series within their errors, providing a further verification to the NIEL scaling hypothesis in the range of energies investigated, and demonstrating the efficacy of the separation method applied for ionisation and displacement damage.

5.7 Detector lifetime

Finally, it is worth evaluating the expected lifetime of the DynAMITe CMOS APS in routine clinical usage.

Routine use of CMOS sensors in clinical practice ranges from diagnostic imaging (radiology and Computed Tomography (CT)) to electron portal imaging in MegaVoltage
(MV) radiotherapy. While the former involves delivering a relatively low dose per exam (in the order of 5 mSv [88]), the latter involves a much higher radiation dose being delivered and as such represents a significant challenge for imaging detectors in terms of operational life.

Imaging for patient positioning in MV radiotherapy usually corresponds to 0.2 Gy delivered to the patient [89] per fraction. However Electron Portal Imaging Detectors (EPIDs) are usually not removed after patient positioning, but are subject to the whole treatment dose (ca. 2 Gy per fraction [90]). Thus assuming a cumulative dose of 2.2 Gy per fraction delivered to the patient, and 20 patients treated per day over 250 working days per year, the cumulative dose to water, delivered at the patient position per year, can be estimated as 16.5 kGy.

In order to compare this value with the relevant dose delivered to the detector, Monte Carlo simulations\(^4\) were performed using a standard MV EPID arrangement\(^5\)[89]. From simulated energy deposition, it was calculated that the yearly dose of 16.5 kGy to patients equates to a dose to the detector of 23.8 kGy(Si). In order to predict the lifetime for the DynAMITe detector in MV radiotherapy, a nominal failure condition for the detector was chosen as a reduction of 50% in conversion gain. A 50% decrease in conversion gain corresponds to a dose of 94 kGy(Si) in the X-ray irradiation studies (see Figure 5.16). This yields a predicted lifetime of 3.9 years for this detector under MV radiotherapy conditions.

For proton radiography and pCT, although these techniques are still matter of research and not yet established in the clinical practise, a similar calculation can be performed. Proton imaging is based on the use of a relatively low flux, for protons to be imaged individually [65], with a maximum energy of typically 65 MeV and a to-

---

\(^4\)Details on the implementation of these simulations are provided in Appendix C

\(^5\)A therapeutic 6 MV beam was simulated incident on a 21 cm diameter spherical water phantom, whose centre is placed at 1 m from the source position. The DynAMITe detector, together with a Gadox conversion screen and a 1 mm thick Copper plate, was placed at 1.6 m from the source. Dose deposited in the detector was scored as a function of the dose deposited in the water phantom.
The chosen failure threshold of 50% reduction in conversion gain corresponds to a proton fluence of $9.8 \times 10^{10} \text{cm}^{-2}$ for a 29.4 MeV proton field. These values can be converted in a 1 MeV neutron equivalent fluence, using the NIEL scaling hypothesis (section 5.2.3), in order to provide a figure independent of proton energy.

However, using the NIEL scaling hypothesis, which has been been verified for this detector in the energy range 2-30 MeV, is subject to large limitations. In fact, going from the range of energies where this hypothesis has been verified to energies of interest for the proton CT (65 MeV), cross sections for interactions change notably. Nevertheless, even with such limitations, it has to be considered a useful tool for comparing most particles and energy dependencies of the damage observed in silicon particle detectors.

The resulting 1 MeV neutron equivalent fluence at the failure condition for the DynA-MITe detector is $\phi_{eq}=2.3 \times 10^{11} \text{n(1 MeV)cm}^{-2}$, while the neutron equivalent fluence for a pCT scan is $\phi_{eq}(\text{CTscan})=4.4 \times 10^{7} \text{n(1 MeV)cm}^{-2}$. Thus a lifetime corresponding to 3300 pCT scans can be envisaged for this CMOS APS. If the device is used in a single room to provide planning CT scans for a 3-room proton facility treating 750 patients per year, this is a lifetime of over four years. However if the clinical need is for daily imaging for guiding treatment, the lifetime is then of the order of three months.

However, it is important to note that those estimated figures of lifetime derive from accelerated radiation tests, as performed in this study. The detector under investigation has been exposed to much higher dose rates, compared to what would be expected in the clinical use, and often irradiation has not been followed by a time of storage, as would happen in the practical applications. Those two factors contribute to an underestimate of the detector lifetime, since in a realistic use lower dose rates and storage time can lead to a significant recovery of the radiation damage. Thus, lifetime calculated in this section have to be considered the lower limit of the expected detector lifetime.
5.8 Discussion

The radiation hardness-by-design of the DynAMITe detector has been presented. The radiation damage, produced in this sensor by X-ray and proton irradiation, has been studied as function of total ionising dose and displacement damage dose.

X-ray irradiation has resulted in a decrease in conversion gain and increase in dark current. When this performance is compared with a commercial CMOS APS for radiology applications, the DynAMITe sensor shows a radiation tolerance 100 times higher. At the maximum dose delivered (94 kGy(Si)), the detector remained fully functional. A similar investigation has been carried out for proton irradiation studies up to proton fluence of $2 \times 10^{11}$, resulting in an increase in dark current and decrease in conversion gain and full well capacity.

The dark current increase produced by proton irradiation, due to ionisation and displacement damage, has been compared with the dark current increase produced by ionisation only, in the X-ray irradiation study. The proton irradiated detector has shown non-uniformities in dark current distribution, e.g. exponential tails or spikes, characteristic of displacement damage and their relative frequency has been linked to the proton fluence.

Changes in conversion gain produced by proton irradiation have been studied as function of both ionisation and displacement damage. After subtraction of ionisation contribution, due to total ionising dose deposited in Si dioxide, the resulting gain variations have been modelled as function of the displacement damage for both proton beams used in this study (29.4 and 1.9 MeV). When expressed as function of displacement damage dose, gain variations have been proved independent from proton energy providing a further verification of the NIEL scaling hypothesis.

Finally, the utility of this radiation-hardened CMOS APS in the clinical environment has been considered. For example, a single treatment fraction and relevant imaging in
MV radiotherapy has an absorbed of about 2.2 Gy, which equates to an absorbed dose at the sensor of 3.2 Gy. Hence, CMOS APS devices based on an equivalent technology and design as employed in DyNAMITe would have a working life of about 30000 treatment fractions or nearly four years when used in a typical MV radiotherapy environment. When this figure is calculated for pCT a lifetime of 3300 scans, or four years when used for treatment planning, can be envisaged, as lower limit of the detector lifetime. This level of radiation tolerance suggests that this detector design may be an excellent candidate for routine use in clinical practice, employed in X-ray and proton based imaging applications.
Chapter 6

Large area CMOS APSs for proton Computed Tomography

6.1 Imaging in Proton Therapy

In 1919 Professor Sir E. Rutherford demonstrated that the hydrogen nucleus could be extracted from the nuclei of nitrogen atoms by collision [92], postulating the existence of a fundamental nuclear constituent, for which he later coined the term proton [93]. However, it took until the 1955 before the first patients were treated with proton therapy at the Lawrence Berkeley Laboratory (LBL) in California [94]. For this goal to be achieved two steps were fundamental: Lawrence building the first cyclotron in 1930 [95], paving the way for the production of therapeutic proton beams, and the intuition of R. R. Wilson in 1946, who first saw the advantage of using protons for radiotherapy. Wilson stated: “It will be easy to produce well collimated narrow beams of fast protons, and since the range of the beam is easily controllable, precision exposure of well defined small volumes within the body will soon be feasible” [96].

The depth dose distribution in muscular tissue for mono-energetic protons of clinically relevant energies is shown in Figure 6.1a. This is the main motivation for using protons for cancer treatment, instead of X-rays, a common oncologic treatment option since the end of the 19th century [97].
Protons, compared to X-rays or photons, have the potential of delivering the planned dose over a small depth range, defined and controlled by the proton energy, while relatively sparing surrounding healthy tissues [98], thus reducing the so called therapeutic ratio, i.e. the ratio between dose delivered to healthy tissues and dose delivered to the planned target volume. In fact, the most prominent difference between protons, and more generally heavy charged particles, and photons is the finite range of a proton beam. As photons penetrate in a medium, they show a small build-up region followed by an exponential decay. In contrast, the energy transferred to a medium by protons is inversely proportional to the proton kinetic energy, as protons lose energy mainly because of electro-magnetic interactions with orbital electrons of the medium they interact with. The outcome of this interaction is that the more protons slow down, the higher the energy they transfer per track length, resulting in a maximum energy deposition at a certain depth in the medium transversed. A schematic representation of this process is shown in Figure 6.1b. Relative depth dose, proportional to the energy deposited per length unit in a transversed tissue volume, is reported for 15 MV photons (with a spectrum end-point at 15 MeV) and for a proton beam with a spread-out Bragg peak. The tumor volume is highlighted in red with an ideal dose distribution. Notably, proton dose depth curve is much more similar to the ideal dose distribution that the one for photons. It is also of note that the proton curves stops abruptly distal to the target volume.

From the first experimental patients’ treatment at LBL in 1950s, it took a few more decades for this technique to be further developed and become a truly clinical option: in 1990, the first hospital-based proton therapy facility was opened at the Loma Linda University Medical Center (LLUMC) in California [100]. At the end 2013 there were 43 active proton therapy facilities worldwide, having treated more than 100000 patients to that date [101]. These large numbers, together with a significant increase in the number

---

1 A strictly monoenergetic proton beam is often unsuitable for cancer treatment, as the monoenergetic Bragg peak is too narrow compared to the tumor volume (see Figure 6.1a). It is then necessary to spread out the Bragg peak to provide uniform dose within the target volume, by providing a suitably weighted energy distribution of the incident beam. Spread-out Bragg peaks are often achieved by using a range modulator wheel.
6.1. Imaging in Proton Therapy

Figure 6.1: 

(a) Depth dose curve for protons of clinically relevant energies in muscular tissue [99]. The narrow peak for this mono-energetic protons is shown. 

(b) Comparison between depth dose curves for 15 MV photons and a proton spread-out Bragg peak. The target volume is highlighted in red, together with an ideal dose distributions. Notably, proton dose depth curve is much closer to the ideal dose distribution than for photons. It is also of note that the proton curves stops abruptly distal to the target volume [98].
of new facilities being planned [64], make proton therapy pivotal in the medical physics research.

6.1.1 Proton CT

The use of proton therapy for cancer treatment makes the need for developing new and more accurate imaging modalities for treatment planning, based on direct measurements of tissue stopping power instead of tissue density, as in conventional X-ray Computed Tomography (CT), to reduce the error in converting the latter quantity into the former [102].

This need appeared evident since the early days of proton therapy. The Nobel laureate who won the Nobel Prize for Medicine in 1979 with Hounsfield for the invention of the X-ray CT, A. M. Cormack, said in his seminal paper on CT reconstruction [103]:

*The next application of the solution [for Computed Tomography]... concerns the recent use of the peak in the Bragg curve for the ionisation caused by protons, to produce small regions of high ionisation in tissue. The radiotherapist is confronted with the problem of determining the energy of the incident protons necessary to produce the high ionisation at just the right place, and this requires knowing the variable specific ionisation of the tissue through which the protons must pass.*

Since the first proof of concept of pCT were developed in the '70s by Hanson [104] and, not surprisingly Cormack [106](see Figure 6.2a), a number of technologies has been proposed to perform pCT, as means of mapping tissues stopping power for an accurate treatment planning. Although pCT has benefited from many developments, in terms of detector technology and image reconstruction, it is conceptually very similar to the first alpha CT scanner developed 1975 and, surprisingly, trialled on humans (see Figure 6.2b).

The basic requirements for pCT lie in measuring position and direction of individual protons and assessing their residual energy in an energy discriminating detector, in
6.1. Imaging in Proton Therapy

Figure 6.2:  
(a) The first tomographic reconstruction of a plastic phantom with pCT [104]. and  
(b) the first alpha scanner trialled on humans [105].  
(c) A schematic representation of an instrument for pCT. A patient is placed between two planes of tracker detectors (Tracker 1 and Tracker 2), so that entrance and exit proton trajectories can be measured in order to infer proton most-likely paths in the patient. Residual proton energy is then measured in a calorimeter, and energy loss by protons on a specific path inside the patient is reconstructed.
order to infer the most likely path in the patient of each proton and the energy deposited, thus the stopping power, along the inferred path. A schematic representation of a detector apparatus for pCT is shown in Figure 6.2c.

Since 2000 a number of research group around the world started working on developing clinically usable pCT system, triggered by the growing interest for proton therapy in oncology. An overview on the main pCT systems developed in the last decade is provided next.

A research group based on the collaboration of Loma Linda University (LLU) and Santa Cruz Institute for Particle Physics, is working on several aspects on developing a pCT system, both technological and theoretical. This group has deployed a first pCT prototype in 2003, based on the use of Silicon Strip Detectors (SSDs) as trackers and a segmented CsI calorimeter coupled to photodiodes \(^{107}\) (see Figure 6.3a,b). One of the main limitations of this system is, however, the slow decay time of the CsI scintillator \((\approx 1 \mu s)\), strongly affecting the event-rate capability of this system \(^{108}\). For this reason, a new system is currently under development making use of a segmented plastic scintillator, to allow for a faster read-out, coupled to Silicon Photo-Multiplier (SiPM) \(^{109}\).

An Italian project for a PRoton IMaging device (PRIMA), based at the INFN-LNS laboratories in Catania, has recently developed two pCT prototype systems (see Figure 6.3c,d). Both based on the use of SSDs as trackers and segmented YAG:Ce calorimeter coupled to photodiodes \(^{110}\), while improvements from the first to the second prototype are mainly represented by an increase of field of view and event-rate capabilities.

The Massachusett General Hospital (MGH) recently developed a system using only a matrix array diode based on time-resolved dose measurements and without performing proton tracking \(^{112}\). Although this system is simple and low cost, spatial resolution is limited by the lack of proton tracking, which results in reconstructed images suffering from halo effects, as can be seen in Figure 6.3e.
Figure 6.3: a) The LLU pCT system and b) a pCT slice of a contrast phantom (diameter 14 cm) [107]. c) The PRIMA pCT system [110] and d) a pCT slice of a PMMA phantom (2 cm diameter) [111]. e) Two pCT slices of a PMMA phantom image by using the MGH pCT system (12 cm diameter)[112].
6.1.2 The PRaVDA pCT

Previous prototypes of energy-range detectors for pCT are chiefly based on the use of scintillator-based calorimeters, to measure proton residual energy. However, such an approach is limited by the need for only a single proton passing through the energy-range detector per read-out cycle. A novel approach to this problem is the use of pixelated detectors, where the independent read-out of each pixel allows the simultaneous measurement of the residual energy of a number of protons in the same read-out cycle, facilitating a faster and more efficient pCT scan.

The Proton Radiotherapy Verification and Dosimetry Application (PRaVDA, Wellcome) [113] consortium is developing a pCT system based on the use of Silicon Strip Detectors (SSDs) as trackers [114], to provide the protons’ most likely path through the patient, and CMOS Active Pixel Sensors (APSs) used in an energy-range telescope, to infer residual proton energy by measuring the position where the proton stopped in the telescope. A schematic representation of the proposed PRaVDA system is shown in Figure 6.4.

6.2 Motivations

This Chapter investigates the major aspects related to the proof of concept of using CMOS APSs for pCT. In order for CMOS detectors to be successfully used in an
energy-range telescope for pCT, these have to demonstrate capability to distinguish individual protons and to track protons as they go through a stack of sensors. This means answering the following questions:

- Can a CMOS APS count individual protons?
- If so, what is the range of linearity and what trade-off can be found between noise floor and counting efficiency?
- Can a simple two-layer range telescope demonstrate imaging of a single proton passing through the two pixellated detectors?
- Can the tracks of individual protons be reconstructed through a two-layer range telescope?

The following questions will be addressed in this Chapter by analysing experimental data acquired at two proton facilities, spanning a range of proton energies, and by using a rudimentary range telescope consisting of two DynAMITe detectors stacked together and synchronised. Finally, an algorithm for proton tracking, together with an assessment of its efficiency, will be provided for the PRaVDA range telescope.

6.3 Materials and Methods

6.3.1 Proton facilities

Two proton sources were used for the experiments described in this chapter: the University of Birmingham (UoB) MC40 cyclotron (Birmingham, UK) and the iThemba radiotherapy facility at the iThemba Laboratories (Cape Town, SA).

The MC40 cyclotron is a non-therapeutic proton facility with a maximum beam energy of 36.6 MeV and a maximum beam size of 5 cm, obtained by magnetically defocusing the proton beam. The beam produced at this facility has a narrow energy distribution with a Full Width at Half Maximum (FWHM) ≈ 0.1 MeV. For all the experiments described below the beam current was measured with a transmission chamber at the exit of the nozzle. The conversion factor from measured beam current to proton fluence
has been measured to be 1/160 protons per nA. The range of beam currents available for this beam varies from fractions of pA to tenths of nA.

The iThemba beam is actively used as a therapeutic facility providing a much higher energy (191 MeV). A schematic diagram of the iThemba beam line is shown in Figure 6.5. The maximum beam range achievable at the patient position, or iso-centre, is 240±0.4 mm range with a FWHM of 25±1.0 mm (measured as 50% of maximum dose on the distal side of the Bragg peak in water). The large area beam (10 cm diameter) is achieved by using a system of passive scattering components and collimators, while the beam energy can be degraded by graphite attenuators. In the proton therapy beam set-up, the current range achievable at this facility is in the range 0.1-100 nA, however, for the purposes of the experiments reported in this Chapter, a different set-up was used making available lower currents. The physics research injector cyclotron (SPC2 in Figure 6.5 a) was used with an external ion source, allowing copper filters, featuring evenly spaced holes, to further reducing the beam current. Those filters, designed for experiments of a Japanese research group, produce a reduction in beam current of a nominal factor of 10\(^{-2}\) or 10\(^{-4}\).

The beam current at iThemba was measured using an air monitor, placed before the final lead collimator (shown in Figure 6.5b). The output of the air monitor was expressed in uncalibrated units, and calibration data were not available to convert its readings to proton flux. Moreover, low value of currents, achievable with the filters described above, fall below the sensitivity threshold of the air monitor, so that current measurements with this instrument were not available for all the experiments performed.

### 6.3.2 Double DynAMITe detector

The DynAMITe CMOS APS was used for the experiments described in this Chapter. Part of this work, where the experimental goal was to prove the capability of CMOS APSs to tracks individual protons, has been performed by using a stack of two DynAMITe detectors placed one after the other, in order to provide a two-layer range
6.3. Materials and Methods
telescope as proof of principle of the PRaVDA range telescope. This configuration, using the two detectors stacked, is referred in this work as *Double DynAMITe* (DD). A picture showing the DD set-up at iThemba is shown in Figure 6.6. A minimum detector-to-detector distance of 10 mm was achievable, given the size of the electronic boards of the two detectors, and each detector features a 5 mm thick Al back-plate (see Figure 6.6). It is of note that the thickness of the Al plate after the first detector, represent a challenge for proton tracking, as this will provide additional material for protons to scatter in. Also, due to geometrical constraints, the two detectors are flipped (180 degree rotation in the image plane) one respect to the other. Furthermore, precise alignment as not possible for the two detectors, so that calibration of the relative misalignment had to be performed, by using high current beams and aligning the beam profiles in both detectors.

The readout of the two detectors was synchronised in order to allow for proton tracking, so that images were acquired simultaneously for both detectors. The same master clock, which is the basic timing signal for detector operations such as exposure, reset and readout (see Chapter 2), was provided to both detectors guaranteeing synchronous operations. The synchronisation of the two detectors was tested by flashing an LED.
placed in the enclosure of the two detectors, with an ON time comparable with the detector integration time and with an OFF time several times longer. However, it has to be noted that the ON and OFF cycle of the LED could not be synchronised with the detector readout, so that there is a finite probability that the ON time of the LED falls in between frames for the two detectors. Figure 6.6c shows the output of the two detectors, named *Master* and *Slave*, as a function of the frame number as the LED is turned ON and OFF. Both curves follow each other over time without any appreciable drift, so that the synchronous readout of the DD stack was demonstrated.

### 6.3.3 Experimental parameters

Several experimental parameters, such as beam energy, current, size, exposure time, were varied for the experiments described in this Chapter. Before introducing them tabularly, it is useful to discuss some guidelines which have been adopted in choosing those parameters.

One of the main requirements for proving proton counting is that the beam current has to be low enough so that event pile-up is negligible within the shortest exposure time allowed by the detector. As the detector used for these experiments was not designed for particle counting but for integrated imaging, the shortest exposure time achievable with a full frame readout of the SP camera (see Section 2.2) was 153 ms. However, this exposure time was too long for the minimum currents allowed in several experiments (reported in Table 6.1), so that a trade-off had to be found between the detector area to readout and the necessity to image individual protons. For this reason most of the experiments were performed using ROIs in the detector area, as small as 10 rows (or 0.5 mm), to achieve a suitably short exposure time. On the other hand, when the experimental task was to track individual protons, a larger area was needed so that protons scattered in the first detector had a high probability of interacting in the ROI of the second detector and be imaged. A further trade-off between occupancy, i.e. the fraction of detector pixels which sees a proton, and exposure time had to be found.

The parameters used for the experiments of this chapter are summarised in Table 6.1.
Figure 6.6: a) The DD set-up during experiments at iThemba. An optical collimation tool is used to project the beam area on the detectors. The stack of detectors is placed at iso-centre on a treatment coach. b) A schematic showing the configuration of the DD set-up. c) The signal measured in the two sensors (named Master and Slave) as a function of the frame number, when an LED is cyclically turned ON and OFF in the enclosure. Both curves follow each other over time without any appreciable drift.
### UoB

<table>
<thead>
<tr>
<th>#</th>
<th>Energy</th>
<th>Current</th>
<th>Filter</th>
<th>Beam size</th>
<th>Camera</th>
<th>Rows</th>
<th>Exp time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>36 MeV</td>
<td>1.5-500</td>
<td>–</td>
<td>5</td>
<td>SP</td>
<td>10</td>
<td>0.76</td>
</tr>
<tr>
<td>b</td>
<td>36 MeV</td>
<td>10-500</td>
<td>–</td>
<td>5</td>
<td>P</td>
<td>10</td>
<td>0.76</td>
</tr>
</tbody>
</table>

### iThemba

<table>
<thead>
<tr>
<th>#</th>
<th>Energy</th>
<th>Current</th>
<th>Filter</th>
<th>Beam size</th>
<th>Camera</th>
<th>Rows</th>
<th>Exp time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>191 MeV</td>
<td>10-100</td>
<td>$10^{-2}$</td>
<td>10</td>
<td>SP</td>
<td>10</td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>60 (30 mm)</td>
<td>20-100</td>
<td>$10^{-2}$</td>
<td>10</td>
<td>SP</td>
<td>10</td>
<td>0.76</td>
</tr>
<tr>
<td>5</td>
<td>191 MeV</td>
<td>10, 50, 100</td>
<td>$10^{-4}$</td>
<td>5</td>
<td>SP</td>
<td>2560</td>
<td>153</td>
</tr>
<tr>
<td>6</td>
<td>60 (30 mm)</td>
<td>0.1, 0.5</td>
<td>$10^{-2}$</td>
<td>5</td>
<td>SP</td>
<td>2560</td>
<td>153</td>
</tr>
<tr>
<td>9</td>
<td>191 MeV</td>
<td>1, 10</td>
<td>$10^{-4}$</td>
<td>0.5</td>
<td>SP</td>
<td>400</td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>60 (30 mm)</td>
<td>0.1, 0.5</td>
<td>$10^{-2}$</td>
<td>0.5</td>
<td>SP</td>
<td>400</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 6.1: Experimental parameters used at UoB and iThemba. The ROI used is specified here only as a number of rows, as the number of columns to readout was fixed (2520 columns). For the iThemba experiments the value of current reported was the beam current before applying the $10^{-2}$ or $10^{-4}$ filter. Also, when beam energy is expressed as a length, this refers to the beam range measured as 50% of maximum dose on the distal side of the Bragg peak in water.
6.3.4 Image processing

Images acquired for these experiments were dark corrected by subtraction of the average of a number of dark frames, taken soon before each experiment, to reflect changes in ROI size, exposure time, noise and temperature. Further to dark correction, images were thresholded with respect to a reference value chosen as three times the temporal noise, and equal to 19 DN. A clustering algorithm\(^2\) was used to account single hit events spread over multiple pixels. The number of connected events (counts), size of events (cluster size), integral signal deposited in each event, as sum of the signal generated in each pixel forming a clustered event (signal), were used in the analysis of these data.

6.4 Proton counting

The hypothesis to be verified in the proton counting experiments was that the DynAMITe detector is capable of imaging individual protons. This will offer a proof of concept on the use of CMOS detectors in an energy-range telescope for pCT, as well as informing on parameter optimisation for the design of an ad-hoc device.

A way to verify that events measured in the detector are genuinely due to individual protons being imaged is to observe their dependence on beam current. The number of detected events should be proportional to beam current, in the case of single proton being imaged, until occurrence of saturation, due to the pile-up of events. Conversely, the occurrence of pile-up can be verified by analysing the relationship between event size and beam current, which should be constant at a certain level until a point where pile-up starts occurring. Other features of detected events, such as signal generated, which can be related to energy deposited, can be used to support the hypothesis of proton counting.

Some exemplar images of individual protons being imaged by the DynAMITe detector are shown in Figure 6.7. Figure 6.7a shows a few bright clusters resulting from the

\(^2\)The cluster algorithm is provided by the *bwconncomp* Matlab function.
6.4. Proton counting

Figure 6.7: a) Bright clusters produced by 36 MeV protons interacting in the Master sensor of the DD stack. b) Superimposed images of 36 MeV protons for Master (coloured in red) and Slave (coloured in green) sensors.

interaction of 36 MeV protons in the Master sensor of the DD stack. Figure 6.7b shows two superimposed images for Master (coloured in red) and Slave (coloured in green) sensors when the stack is exposed to 36 MeV protons.

6.4.1 Proton counting at UoB

The DD stack was exposed at the UoB cyclotron to 36 MeV protons, from an uncollimated beam (5 cm diameter). A small ROI (10×2520 pixels) was used to allow an exposure time short enough to be able to image individual events (see Table 6.1).

The number of events detected as a function of the beam current is shown in Figure 6.8, for the DynAMITe P camera (Figure 6.8a) and SP camera (Figure 6.8b). Both detector cameras show a range linearity of counts with beam current over a range. Counts for the P camera were linear in the whole range investigated (10-500 pA), while the SP camera shows a linear behaviour up to \( \approx 100 \) pA and then exhibits saturation.
The reason for saturation of the SP camera above 100 pA current can be explained in terms of pile-up occurring in this camera. This can be verified in Figure 6.8c, which shows the cluster size measured for both cameras as a function of the beam current. For the P camera, cluster size is almost constant over the range of current investigated and approximately equal to 1 pixel (with a 100 µm pitch). Conversely, for the SP camera, the cluster size is between 2.5 and 3 pixel (with a 50 µm pitch) below 100 pA, and then this parameters increases linearly with current up to ≈ 5 pixels. Since the energy deposition related to each event has to be considered constant with current at any given energy, the occurrence of an increase in cluster size has to be explained with more events recorded in adjacent pixels, forming a multi-event single cluster (i.e. pile-up).

A second main difference in the behaviour of P and SP cameras to note in Figure 6.8 (a-b) is the different sensitivity of the two cameras in detecting single events. Sensitivity for proton counting can be defined as the slope of the linear region of Figure 6.8 (a-b). The sensitivity for the SP camera has been measured as $3.3 \times 10^{-1}$ particle pA$^{-1}$, while this value becomes as low as $9.6 \times 10^{-3}$ particle pA$^{-1}$ for the P camera. This large difference in proton counting has to be related to the different design of the two cameras. In fact, as discussed in Section 2.2 and 3.3, the P and SP diodes are designed to offer different noise floors, are placed on the detector matrix with different geometries covering different areas and thus offering different collection efficiency. The noise floor for the SP camera was measured to be as low as 149.9±0.7 e$^-$, while this value for the P camera is 780±1 e$^-$ (see Table 3.1). Also SP and P pixels are intrinsically and geometrically different. In fact each SP pixel is fitted with 4 parallel-connect 0.6 µm diameter diodes, with a diode-area-to-pixel-area-ratio of 0.002; whilst, P pixels are provided with a single diode of 1 µm diameter, with a diode-area-to-pixel-area-ratio of 0.0003, implying a lower collection efficiency for the P pixels.

Given the higher sensitivity of the SP camera, only this camera has been used in the remainder of this work.
6.4. Proton counting

Figure 6.8: Number of events detected for the DynAMiTe P camera (a), SP camera (b) and cluster size c) as a function of the beam current, when exposed to 36 MeV protons at UoB. Linear fits are also shown for a) and b) and obtained by choosing the data range to fit by maximising the fit determination coefficient $R^2$. By adopting this procedure, the saturated regions have not been used for linear fit of the data as shown in b).
6.4.2 Proton counting at iThemba

A similar experiment was performed at iThemba to prove the proton counting capability of the DynAMITe detector at clinically relevant energies. The two energies used are 191 MeV, the pristine undegraded energy of the iThemba beam which might be useful for beam Quality Assurance (QA) purposes, and the 30 mm range beam, corresponding to the average energy expected after the patient (considering a 240 mm beam degraded by a 210 mm of water), which corresponds to a mean energy of 60 MeV and is the energy of interest for the energy-range telescope in pCT. A 10×2520 ROI, of the same size as at UoB, was used (see Table 6.1 for Experiment 1 and 2).

Figure 6.9a shows the number of events detected for the SP camera as a function of the beam current, measured in uncalibrated air monitor units, for the two beam energies investigated. At both energies the dependence on the beam current is linear over the whole range of currents investigated, excluding the occurrence of pile-up as can be seen from Figure 6.9b. In fact Figure 6.9b shows the cluster size as a function of the beam current, which, although affected by some random fluctuations, has no dependence on the beam current. Moreover, it can be seen how the cluster size for the degraded beam is larger (2.4 pixel) than that for the pristine beam (1.7 pixel), due to the higher energy deposition (dE/dx) at lower energy.

Performing a linear fit on the two curves of Figure 6.9a provides the sensitivity of the DynAMITe detector at the two energies used. At the highest energy the sensitivity has been measured as 1.1±0.1 particle per air monitor count, while this value is 0.7±0.2 particle per air monitor count for the 30 mm range beam. At the low energy the sensitivity is lower than at the higher one. This can be explained considering that the 30 mm range beam is a degraded beam, obtained by letting the pristine beam passing through a thickness of 124 mm graphite. In fact, for a degraded beam a larger angular dispersion is expected, resulting in a larger beam size, compared to the smaller pristine beam. Since the air monitor is placed just before the final collimator, as explained in Section 6.3.1 and Figure 6.5, a fraction of the beam measured by the air monitor is blocked by the final collimator and can not reach the detector. Furthermore, when a
Figure 6.9: Number of events $a)$ and mean cluster size $b)$ for the DynAMITe SP camera for the iThemba pristine 191 MeV (blue symbols) and for the degraded 30 mm range beam (red symbols) as a function of the beam current expressed in uncalibrated air monitor units. Linear fits are also shown for $a)$. 

A larger beam (as a degraded beam) is readout with a small ROI, as in this experiment (see Table 6.1), the fraction of beam detected in the ROI is smaller compared to a pristine beam. All these factors lead a decreased sensitivity for the degraded beam compared to the pristine one, explaining the experimental evidence of Figure 6.9 $a)$. 

A further step to demonstrate proton counting is related to verify the energy deposition measured (signal collected in a clustered event) against the theoretical expected value. For this reason, GEANT4 Monte Carlo simulations were performed, with a geometry reflecting the experimental set-up used. Ad hoc classes to simulate charge diffusion and collection in the detector were written by the author. Theoretical aspects and practical implementation of the charge diffusion and collection models used can be found Appendix C. 

Simulated and measured energy spectra (in units of DN) are shown in Figure 6.10.
Figure 6.10: Measured and simulated energy spectra (in units of DN) for the pristine (a) and 30 mm range (b) beam.

for the pristine energy (a) and for the degraded beam (b). Both simulated and measured histograms resemble the shape of a spread-out Landau distribution, as expected for energy loss of charged particles in thin absorbers [115]. Measurements and simulations show a good agreement at both energies, suggesting once again that the signal generated in the detector is genuinely due to individual protons being imaged. Some discrepancies can be found at higher signal values, towards the tail of the measured distribution, where the measured spectrum is higher. This fact can be explained accounting for some pile-up events, which although rare might occur in the detector, generating clustered events of higher energy, while in idealised simulations protons are always treated individually, removing the chance of pile-up.

6.4.3 Time structure

When experimental data described in the previous Section were analysed as a function of time, i.e. frame number, some regular fluctuations in the number of counts with time could be seen. Figure 6.11a shows the number of counts detected in the front and back detector of the DD stack (namely Master and Slave) as a function of the frame number, for a fixed beam current (40 nA with a $10^{-2}$ filter), extracted from the dataset
of Experiment 1 (see Table 6.1). For both detectors the number of counts regularly fluctuates between higher and a lower count region. This behaviour can be successfully fit with a square wave, as shown in Figure 6.11b. Additionally, spikes between high and low count regions are present. The frequency and spectral components of this periodicity can be studied by using the Fast Fourier Transform (FFT). The waveform period so calculated, expressed as a number of frames, was plotted as a function of the nominal current used for each experiment. For most values of current the calculated period was 26.3 frames (≈ 20 ms) for both Master and Slave. For a few current values (0.6-0.8 nA) this values was as low as 9.1 frames (≈ 7 ms). For the majority of data points and considering an exposure time of 0.76 ms (see Table 6.1), the observed time structure corresponds to a frequency of 50.0 Hz. The frequency of the time structure present in this experiment would suggest a possible coupling or interference from the mains, either for the beam line or for the detectors. The absence of this time dependence in dark frames acquired in the same room, would exclude the problem being related to the detector. Other experiments at iThemba (Experiment 5, 6, 9 and 10 of Table 6.1) were not affected by this phenomenon, since their exposure time was significatively longer than this characteristic period. Effects of this time structure, especially for proton tracking, will be further discussed in the next Section.

6.5 Correlation

The aim of proton tracking is to estimate the probability that two particular events in the two layers are due to the same proton (i.e. to say which proton is which). Capability of proton tracking represents an important requirement in order for CMOS detectors to be used as RT for pCT. However, before effectively trying to track protons, i.e. to identify which proton is which across the DD stack, it is useful to perform a preliminary study of this problem in terms of correlation. The fact that the number of events seen in the two layer shows a good degree of correlation is evidence that the same proton is detected in both layers. This a requirement for protons to be tracked across the two layers. It has to be noted that for the purpose of verifying correlation of number of events detected in two DD layers, timing is of relatively low importance as long as
Figure 6.11:  
(a) Number of events detected for front and back sensor (Master and Slave) as a function of the frame number for a given current (40 nA with $10^{-2}$ Filter).  
(b) The number of events for the Master detector is fitted with a square wave.  
(c) The period of the time structure calculated by using the FFT for both detectors.
event pile-up is negligible. However, for proton tracking, timing or more specifically readout speed and frame occupancy, will have to be accounted for. In fact, in order for proton tracks to be reconstructed across the DD stack, or the full RT, events will have to be sufficiently separated in space to reduce ambiguities and so maintain a sufficient tracking efficiency. The parameter used in this work to assess correlation between the two detectors is the Pearson’s correlation coefficient. Given two random variables $X$ and $Y$, with expected values $\mu_X$ and $\mu_Y$ and standard deviation $\sigma_x$ and $\sigma_y$ respectively, the Pearson’s correlation coefficient for these two variables $\rho_{X,Y}$ is given by

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$  \hspace{1cm} (6.1)$$

where $E[\ ]$ is the expected value operator and $\text{cov}(\ )$ is the covariance. A Pearson’s correlation coefficient of +1 (-1) indicates a perfect direct (inverse) linear relationship. The more $\rho_{X,Y}$ approaches to zero, the more the data are uncorrelated. An advantage of using Pearson’s correlation coefficient relies in this coefficient being unbiased (i.e. is not sensitive to different offsets in the two samples), which is of particular interest for analysis of the DD experiments where data might be correlated even if not lying on the $x = y$ bisector when one variable is plotted against the other, i.e. with intercept different from zero (due, for example, to higher noise in one detector) and slope different from one (due, for example, to a geometrical efficiency lower than one between the two detectors).

The interpretation to give to the correlation coefficient will depend on the experimental parameters of Table 6.1. As beam and detector geometry, i.e. ROI size, changes (see Figure 6.12) the correlation coefficient can be interpreted differently. In fact, for Experiments 1 and 2 a very small ROI was readout (10 rows or 0.5 mm) when exposed to a large beam (10 cm diameter). Given the small ROI size of this experiment and the large beam area, the ROI readout can be considered as an equilibrium region within the larger beam area, where the number of protons seen in the first detector ROI, and scattered outside the second detector ROI, equates the number of protons falling outside the first detector ROI and then scattered inside the second detector ROI. In this case a high correlation coefficient supports the evidence that the number of protons
which cross both detector ROIs for each time unit is the same, or proportional depending on the possible different noise levels and efficiencies of the two detectors. A high correlation coefficient in this experiment will not demonstrate that the same protons are detected in both detectors, as the proton beam spread out from one detector to the other, the probability that the same proton will cross two geometrically corresponding ROIs in the two detectors is low. Thus, for Experiments 1 and 2 data, although they can be provide useful information in terms of correlation and detector synchronisation, will not ultimately be suitable to prove proton tracking.

The geometry of Experiments 5, 6, 9, 10 are different (see Figure 6.12). In these collimated beam experiments, the ROI size is larger than the beam area allowing for protons scattered outside the first detector to still fall within the readout ROI of the second detector. In this case the correlation coefficient can be interpreted differently, related to the probability of seeing the same protons in the two detectors.

### 6.5.1 Correlation in uncollimated beam experiments

Data from Experiment 1 are shown in Figure 6.13a, where the number of events in the slave detector is plotted against number of events in the master one, for three different values of current (10 nA, 50 nA, 100 nA with a $10^{-2}$ filter). Although each data group for the various currents span across a large range of values, because of the time
6.5. Correlation

structure discussed before, a good linear relation is shown between events in the two
detectors. The data points of Figure 6.13a have an intercept consistent with the x=y
bisector, while the slope is lower than one. The slope of this plot can be considered
as being the efficiency of the DD stack, in fact it represents the ratio between number
of counts in the slave detector and number of counts in the master. The efficiency
for this set of experiments is reported in Figure 6.13b over the range of currents used.
Efficiency is between 0.88 and 1.02, randomly varying with current without showing
any evidence of dependence on the beam current, as expected in the assumption that
pile-up is not occurring (hypothesis verified in Figure 6.9a.

Figure 6.13c shows the Pearson’s correlation coefficient over the whole range of cur-
rents investigated. The data of Experiment 1 show a good degree of correlation with a
Pearson’s coefficient always greater than 0.8. Given the geometry of this experiment,
such a high correlation can be interpreted as a high probability that the number of
protons scattered outside the ROI of the second detector is compensated by a same
number of protons scattered inside this ROI. Thus the ROI is in equilibrium.

Although a very high correlation coefficient is found for data of Experiment 1, the
finding of a time structure in these data sets makes these measurements weaker. In
fact, as the correlation coefficient depends mainly on the covariance of the two datasets
(Equation 6.1), the presence of periodic regions of high and low counts can artificially
increase the correlation, since counts in the master and slave follow each on the pe-
riodic waveform. For this reason, a high and low counts regions were defined in the
periodic data of Experiment 1. For each dataset a threshold was chosen corresponding
to the average value of the periodic waveform, and then counts above this threshold
are considered as high and those below are considered as low. This process is shown in
Figure 6.14a. High and low counts regions are then analysed separately, reducing the
possibility of a high correlation coefficient being due to those fluctuations. Number of
events in the slave plotted against number of events in the master for the high and low
counts region are shown in Figure 6.14b. Similarly as in Figure 6.13a, data points have
a linear behaviour and lie on the x=y bisector, indicating a high level of correlation.
Figure 6.13: a) Number of events in the slave for each frame as a function of number of events in the master, for some values of the beam current. The $x = y$ bisector is also shown. b) Efficiency of the DD set-up, i.e. slave-to-master count ratio, over the range of current investigated. c) Pearson’s correlation coefficient for data of Experiment 1.
6.5. Correlation

The Pearson’s correlation coefficient for the high and low counts region is shown in Figure 6.14c and compared with the same parameter calculated on the original data (raw as in Figure 6.13c). Although the separation in high and low count regions reduces the value of the Pearson’s coefficient, especially at the lowest current value, this value still remains sufficiently high to prove a good correlation between proton counts in the DD stack.

6.5.2 Correlation in collimated beam experiments

The same correlation analysis, described in the previous Section, was performed on the collimated beam experiments (Experiments 5, 6, 9 and 10 from Table 6.1) to test correlations in experiments where the beam area is fully contained in the ROIs, so that the detection of same protons across the stack can be demonstrated. The correlation plot, i.e. counts in the slave versus counts in the master detector is shown in Figure 6.15 for data of Experiment 5 a) and Experiment 9 b), whose experimental parameters described in Table 6.1.

For both experiments shown in Figure 6.15, although counts in both detectors increase with beam current, suggesting the evidence of proton counting, data points are circularly distributed for each current value suggesting a poor degree of correlation. For comparison, Pearson’s correlation coefficient and efficiency is reported in Table 6.2 for all the experiments at iThemba, together with a summary of the experimental parameters used. From data of Table 6.2, it appears that only Experiment 1, i.e. a small ROI in an uncollimated beam, shows a correlation coefficient high enough to support correlation. For all the other collimated beam experiments, regardless of the energy, the Pearson’s correlation coefficient is often close to zero and the values of efficiency are largely different from experiment to experiment, up to a paradoxical point reached for 5 cm beam experiments where efficiency is greater than 1, showing more counts in the back detector than in the front one.

A direct comparison of the outcomes of the several experiments in Table 6.2 is not possible, since many parameters vary from experiment to experiment including the relative beam/ROI geometry, beam current, filtration, exposure time. However, since the efficiency seems to vary substantially across the various experiments, a further
Figure 6.14:  

(a) The process to separate high and low counts regions in the periodic data of Experiment 1.  

(b) Number of events in the slave as a function of number of events in the master, for some values of the beam current, for high and low counts regions. The $x = y$ bisector is also shown.  

(c) Pearson’s correlation coefficient for data of Experiment 1, separated in high and low counts regions and for the original data (raw).
### 6.5. Correlation

<table>
<thead>
<tr>
<th>#</th>
<th>Energy</th>
<th>Current</th>
<th>Filter</th>
<th>Beam size</th>
<th>Rows</th>
<th>Exp time</th>
<th>Pearson’s</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>191 MeV</td>
<td>10</td>
<td>10^{-2}</td>
<td>10</td>
<td>10</td>
<td>0.76</td>
<td>&gt;0.83</td>
<td>&gt;0.9</td>
</tr>
<tr>
<td>5</td>
<td>191</td>
<td>10</td>
<td>10^{-4}</td>
<td>5</td>
<td>2560</td>
<td>153</td>
<td>0.72</td>
<td>1.4</td>
</tr>
<tr>
<td>5</td>
<td>191</td>
<td>50</td>
<td>10^{-4}</td>
<td>5</td>
<td>2560</td>
<td>153</td>
<td>0.05</td>
<td>1.2</td>
</tr>
<tr>
<td>5</td>
<td>191</td>
<td>100</td>
<td>10^{-4}</td>
<td>5</td>
<td>2560</td>
<td>153</td>
<td>0.18</td>
<td>1.1</td>
</tr>
<tr>
<td>6</td>
<td>60 (30 mm)</td>
<td>0.1</td>
<td>10^{-2}</td>
<td>5</td>
<td>2560</td>
<td>153</td>
<td>-0.12</td>
<td>1.9</td>
</tr>
<tr>
<td>6</td>
<td>60 (30 mm)</td>
<td>0.5</td>
<td>10^{-2}</td>
<td>5</td>
<td>2560</td>
<td>153</td>
<td>-0.79</td>
<td>1.5</td>
</tr>
<tr>
<td>9</td>
<td>191</td>
<td>1</td>
<td>10^{-4}</td>
<td>0.5</td>
<td>400</td>
<td>24</td>
<td>-0.02</td>
<td>0.4</td>
</tr>
<tr>
<td>9</td>
<td>191</td>
<td>10</td>
<td>10^{-4}</td>
<td>0.5</td>
<td>400</td>
<td>24</td>
<td>0.20</td>
<td>0.4</td>
</tr>
<tr>
<td>10</td>
<td>60 (30 mm)</td>
<td>0.1</td>
<td>10^{-2}</td>
<td>0.5</td>
<td>400</td>
<td>24</td>
<td>-0.05</td>
<td>0.4</td>
</tr>
<tr>
<td>10</td>
<td>60 (30 mm)</td>
<td>0.5</td>
<td>10^{-2}</td>
<td>0.5</td>
<td>400</td>
<td>24</td>
<td>-0.13</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 6.2: Pearson’s correlation coefficient and efficient, namely the ratio between counts in the slave and counts in the master, for all the experiments performed at iThemba together with a summary of the experimental parameters used. Also, when beam energy is expressed as a length, this refers to the beam range measured as 50% of maximum dose on the distal side of the Bragg peak in water.

An investigation into the noise property of the detector was needed.

#### 6.5.3 Background radiation and SNR

For all the experiments described in the previous Section, a detector threshold $T_1$ ($T_1=19$ DN) has been used, equal to three times the noise floor, as measured in dark conditions in the treatment room. However, poorly correlated data in the collimated beam experiments together with a highly variable efficiency across the stack (see Figure 6.13b) resulted in the need to further investigate the noise level in the detectors.

Noise calculated as described above, is mainly due to detector intrinsic electronic noise and some background radiation in the treatment room due to relatively long lived isotopes. However, during the actual experiments some additional sources of noise are present, such as secondaries generated in the beam line, in the collimator, in the detector, in the detector Al back-plates, or secondaries scattered back from the treatment
Figure 6.15: Number of counts in the slave detector plotted against the number of counts in the master one for data of Experiment 5 (a) and 9 (b). The $x = y$ bisector is also shown.

Figure 6.16: a) SNR calculated for Experiments 1, 5 and 9 with the beam OFF is considered ($N_1$) or with the beam ON and the beam stopper in place ($N_2$). b) Pearson’s correlation coefficient, obtained by numerical simulated data, as a function of SNR.
6.5. Correlation

In order to evaluate this contribution to the detector noise and background radiation, a further dataset was acquired with the so called beam stopper in place, a 5 cm thick brass plate to completely stop 191 MeV protons. The number of events measured in dark condition (beam OFF) and with the beam stopper (beam ON + beam stopper) are reported in Table 6.3.

Given the same imaging area and the same exposure time for the two experiments of Table 6.3, the dataset with the beam ON and the beam stopper in place shows a higher number of counts compared to the same data acquired with beam OFF. Also, the efficiency across the stack of two sensors is different for the two experiments: higher for the beam stopper data. These two facts suggest the presence of secondary radiation in the beam line and in the treatment room, and also its anisotropy along the beam direction. This secondary radiation has to be added up to the primary signal when the beam is not stopped, and might be responsible for the poor correlation and non-constant efficiency observed in the collimated beam experiments.

In order to assess the contribution of the secondary radiation to the correlation of the experiments reported in the previous Section, an analysis of the Signal-to-Noise Ratio (SNR) in terms of noise floor has been performed. The SNR for some of the experiments of Table 6.1 is plotted as function of the nominal beam current in Figure 6.16a. Two different noise levels are chosen to calculate the SNR, either the noise level measured with the beam OFF (N₁) or the one measured with the beam ON and the beam stopper in place (N₂). A line corresponding to SNR=1 is also plotted for reference. For Experiment 1 data, the SNR is much higher than the limit SNR=1 for both the noise values considered (ranging from 10³ to 10⁴). However, for collimated beam experiments (Experiment 5 and 9) at low proton fluences values, most data points are barely above the SNR=1 when the noise floor N₁ is used, and below this line using N₂ as threshold. It can then be deduced from Figure 6.16a how the higher noise level when the beam is ON, due to secondary radiation, leads to a SNR often less than the unity for most of the experiments. This very low figure explains the lack of correlation in the collimated experiments.
To have a measure of the effect of SNR on Pearson’s correlation coefficient, Figure 6.16b shows the value of this coefficient for numerical simulated data with a SNR ranging from $10^{-1}$ to $10^4$. As expected the correlation coefficient increases with SNR, starting from $\rho_{X,Y}=0.1$ for the lowest SNR value, passing through $\rho_{X,Y}=0.5$ at SNR=1 (where there is a 50-50% chance that a generic hit belongs either to the signal or noise distributions), then up to a saturation value of 1 for SNR greater than 70.

From data of Figure 6.16a, it appears that collimated beam experiments (Experiment 5 and 9) are affected by a low SNR. To try to improve this ratio, it is useful to compare the noise and signal energy spectra to set detection threshold which could further reject noise counts.

The spectra for the beam stopper data (beam ON), dark signal (beam OFF) and Experiment 1 are shown in Figure 6.17a. It is visible how the three spectra are largely overlapped, and any attempt to discriminate noise from proton signal based on the signal amplitude, is ineffective. However, the lower energy data (30 mm range beam) might offer this possibility since the energy deposition ($dE/dx$) is higher, helping to distinguish the genuine proton signal from noise and background. For comparison, energy spectra measure in Experiment 1 and 6 are shown in Figure 6.17b. Comparison of these two spectra suggests that using a higher threshold $T_2=30$ DN for experiments with the 30 mm range beam (Experiment 6 and 10) can successfully suppress part of the noise, improving the SNR and thus the correlation.

### 6.5.4 Correlation in the collimated 30 mm range beam experiments

The 30 mm range beam data appear then to be the most suitable of the collimated beam data (see Table 6.1) to prove correlation, since a higher threshold can improve the SNR. Figure 6.18a shows the Pearson’s correlation coefficient for data of Experiment

---

3 Numerical simulations were performed by assigning a number of counts to the master sensor, extracted from a Poisson distribution with mean $S$. The number of counts in the slave is assumed to be 90% of the number of counts in the master, to allow for a realistic efficiency. Such data sets result perfectly correlated ($\rho_{X,Y}=1$). The noise for each SNR value was calculated in order to give the required SNR, where the signal is the sum of the previously calculated signal and noise.
6.5. Correlation

<table>
<thead>
<tr>
<th>ROI</th>
<th>Current</th>
<th>Master</th>
<th>Slave</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam OFF</td>
<td>2560×2520</td>
<td>0</td>
<td>4500</td>
<td>1930</td>
</tr>
<tr>
<td>Beam Stopper</td>
<td>2560×2520</td>
<td>0.4 nA</td>
<td>6450</td>
<td>9030</td>
</tr>
</tbody>
</table>

Table 6.3: Counts in master and slave and their ratio, efficiency, for full frame ROI images acquired with the beam OFF or with the beam ON and stopped with the brass beam stopper.

Figure 6.17: a) Energy spectra measured for a reference dark set (beam OFF), with the beam stopper in place (and beam ON) and with 191 MeV protons (Experiment 1). b) Energy spectra measured at 191 MeV (Experiment 1) and for the 30 mm range beam (Experiment 6).
6 and 10, at the two values of current used, when a threshold based on the detector
noise is chosen \((T_1=19\text{ DN})\) or when a higher threshold, based on the sum of detector
noise and background radiation is used instead \((T_2=30\text{ DN})\). For data of Experiment 6
the use of the high threshold significantly improves the correlation, going from 0.17, at
the lowest current using the low threshold, to 0.68 with a high threshold. Similarly for
the high current values, \(\rho_{X,Y}\) improves from 0.5 to 0.75. However, such improvement
for the second experiment (Experiment 10) is not visible. This can be explained by
looking at the same quantity (Pearson’s correlation coefficient), when this is plotted
against the SNR estimated for the two experiments (see Figure 6.18b). Even when the
high threshold is used, Figure 6.18b shows how the SNR for Experiment 10 is around
1, resulting in a low correlation, while this values is 2.5 or 7 for Experiment 6, for the
two currents used respectively, resulting in a higher correlation.

In this way a good degree of correlation has been proven also for the collimated beam
experiments, when this exhibits a sufficiently high SNR (Experiment 6). Given the ge-
ometry of this experiment (see Figure 6.12 and Table 6.1), the high correlation between
number of protons seen in the first detector with number of protons seen in the second
one, suggest that the same protons are seen in each frame in both detectors.

6.6 Proton tracking

Proton tracking represents a key task for the PRaVDA instrument. As discussed in
Section 6.1.2, the stack of CMOS sensors will be used to infer proton energy from their
range, measured from the last layer where a proton generates signal.

However, in order for the energy of individual protons to be measured, the track of
each proton has to be reconstructed through the whole RT and related to the direc-
tional information of the proton track, provided by the tracker detectors upstream (see
Figure 6.4). Since the beam current has to be high enough to allow a reasonable ac-
quision time for pCT, proton tracking through the RT will involve reconstruction of
multiple proton tracks per frame. Higher occupancy values, i.e. the fraction of detec-
6.6. Proton tracking

Figure 6.18: a) Pearson’s correlation coefficient for Experiment 6 and Experiment 10 when a threshold based on the detector noise is used ($T_1=19$ DN), or when a threshold based on the detector noise and background radiation is used ($T_2=30$ DN). b) Pearson’s correlation coefficient for Experiment 6 and 10 plotted against SNR, with a threshold of 30 DN.

or pixels which sees a proton, will limit the capability of correctly tracked protons, as the level of ambiguity increases when the distance between adjacent hits decreases. Together with occupancy, another factor which will limit proton tracking capabilities is related to Multiple Coulomb Scattering (MCS), i.e. to protons deflecting from their original trajectory due to Coulomb interaction with nuclei of the material they transverse. Details of the physics of MCS interaction will be provided in Appendix B.

Although the tracking algorithm to be used for the PRaVDA RT will be based on the use of strip detectors for information on the incoming angle of protons and will have to resolve proton tracks through a stack of CMOS sensors, some interesting proof of principle for particle tracking with CMOS APSs can be found in the DD experiment.

For the purposes of proton tracking, it is important to discuss the physical quantities which can be measured in the DD experiments and how these are related to the theory of MCS, whose prediction of the expected scattering angle will be used to assess the validity and efficiency of the tracking algorithm. Figure 6.19 shows schematically a proton track crossing two detectors. The particle impinges the first detector with
Figure 6.19: Schematic representation of a proton track crossing two detectors, placed at a distance $d$. The particle impinges the first detector with an angle $\alpha$ and, after being scattered at an angle $\theta$, is detected in the second detector at an angle $\alpha + \theta$.

and angle $\alpha$, then crosses the first detector where undergoes a number of small angle scattering due to MCS, exiting this detector with an angle $\theta + \alpha$. The first angle ($\alpha$) reported in Figure 6.19 is known as incidence angle, while the second one ($\theta$) is the scattering angle predicted by MCS models. The difference between the position of the particle in the first and second detector is defined as displacement. It is a function of the incident angle $\alpha$ and the scattering angle $\theta$, while the linear quantity, related to scattering angle $\theta$ only, is defined as deflection.

For the DD experiment the quantity which can actually be measured is displacement, as the difference between the position of a hit in the first and second sensor, while the incidence angle is unknown. In the assumption that the incident angle is much smaller than the scattering angle, $\alpha << \theta$, the displacement can be approximated as just a function of the scattering angle $f(\theta + \alpha) \approx f(\theta)$, and displacement can be approximated with deflection. Thus the measured displacement can be verified against the theoretical deflection predicted by the MCS model.
6.6. Proton tracking

6.6.1 Tracking algorithm for the Double DynAMITe experiments

From DD experiments described in Section 6.3, experiments at a degraded beam energy (30 mm range beam, equivalent to 60 MeV) appear to be the most suitable for demonstrating proton tracking, since they are performed at an energy close to the energy expected for proton to enter the RT. Amongst the 30 mm range beam experiments, the data taken by using the 5 cm diameter collimator (Experiment 6) have been chosen to test the tracking procedure, as they show a good degree of correlation (see Figure 6.18). For analysis of the DD experimental data a simple algorithm based only on the use of two detectors has been developed, thus without any prior information on proton velocity, as will be available for the PRaVDA RT. This algorithm is based on the assumption that hits which show a minimum layer-to-layer displacement are assumed to belong to the same proton. This assumption is motivated by the quasi-Gaussian distribution of the scattering angle, as described in Appendix B. Considering the \( m^{th} \) of \( N \) hits in the layer \( n \) of the stack, at position \( \vec{r}_{m,n} \), the most likely match in the \( n+1 \) layer will be the \( j^{th} \) hit at position \( \vec{r}_{j,n+1} \) such that

\[
\| \vec{r}_{m,n} - \vec{r}_{j,n+1} \| = \min_{1 \leq j \leq N} \| \vec{r}_{m,n} - \vec{r}_{j,n+1} \| \quad (6.2)
\]

In order to implement this simplified version of the tracking algorithm, a rectangular matrix (of dimensions \( N \times M \), where \( N \) is the number of hits in the first detector and \( M \) is the number of hits in the second detector) is built for each pair of acquired images reporting the distance between each hit of the first detector and every hit of the second one. The absolute minimum of this matrix is found, corresponding to a match for the relevant hits in the first and second layer. The relevant row and column corresponding to the minimum are then deleted from the matrix, as a match for the two corresponding hits has been found, and the algorithm is applied recursively \( P \) times, with \( P \) minimum of \( M \) and \( N \).

This algorithm will be referred to as Minimum Displacement Algorithm (MDA) in the remainder of this work.
6.6.2 Simulations of proton displacement

GEANT4 Monte Carlo simulations were performed to test the reconstruction algorithm based on the minimum displacement. The geometry used in Experiment 6 was simulated and the main simulations parameters are listed below:

1. Beam
   (a) Gaussian energy distribution $G(\mu, \sigma)$, $\mu = 190.8$ MeV and $\sigma = 1.53$ MeV
   (b) Gaussian angular distribution $G(\mu, \sigma)$, $\mu = 0$ deg and $\sigma = 0.57$ deg
   (c) Circular beam of 5 cm radius
   (d) Beam energy is degraded with a 124 mm graphite absorber to deliver a 30 mm range beam
   (e) Beam is collimated with a brass tube of 2.5 cm inner radius and 5 cm outer radius

2. CMOS RT
   (a) A stack of two CMOS sensors is placed at 20 cm from the beam nozzle
   (b) Each sensor consists of 14 $\mu$m thick epitaxial (sensitive volume) and 753$\mu$m substrate
   (c) $10 \times 10 cm^2$ active area
   (d) 50$\mu$m pixel pitch
   (e) 10mm sensor-to-sensor distance
   (f) 5mm thick Aluminium plate is placed behind each sensor

The displacement distribution (as defined in Figure 6.19) calculated from simulated data for both $x$ and $y$ coordinates is shown in Figure 6.20a. Displacement distributions of figure 6.20a agree with the expected Gaussian distribution for small scattering angles, and thus displacement values (see Appendix B). However for larger angles, or displacement values, the distributions deviate significantly from a Gaussian behaviour.

\[^4\text{Details on the implementation of these simulations are provided in Appendix C}\]
6.6. Proton tracking

According to the Molière theory for MCS described in Appendix B (see Figure B.1). Both distributions, fitted with a Gaussian function in Figure 6.20a, are comparable and centred in the origin with $\mu_x = 0.2 \pm 0.2$ pixel and $\mu_y = 0.08 \pm 0.07$ pixel for the $x$ and $y$ coordinates, respectively. Similarly, the standard deviation of the distributions is comparable and equal to $\sigma_x = \sigma_y = 16.7 \pm 0.1$ pixel, which corresponds to a scattering angle $\theta_{rms} = 0.0833 \pm 0.0005\,rad$, given a pixel size of 50 $\mu$m and a detector-to-detector distance of 10 mm.

Figure 6.20: a) Displacement distributions in the $x$ and $y$ coordinated for simulated data. A Gaussian fit is shown for both distributions, with $x$ and $y$ distributions appearing to be comparable, as expected given the circular symmetry of MCS. b) Tracking efficiency, i.e. ratio of correctly tracked proton tracks to the total number of proton tracks, plotted versus occupancy level. A rational fit is also shown.

This value should be compared with $\theta_{th}^{rms} = 0.0515\,rad$, obtained from the Rossi-Greisen equation (Equation B.15). Simulated and theoretical scattering angles are not comparable within their errors. The theoretical value for the scattering angle underestimates the simulated one. This can be explained by accounting for the assumptions used for this comparison. Firstly, as discussed at the beginning of this Section, deflection can be approximated by displacement on the assumption of a small incidence angle ($\alpha$ in Figure 6.19). However, for the degraded beam simulated here (attenuated by 124 mm
of graphite), the incoming angle might be comparable with the scattering angle and effectively the comparison run here is between $\theta$ (from the theoretical model) and $\theta + \alpha$ (from simulations). Also, it has to be considered that the theoretical calculation does not account for the energy distribution of incoming protons as well as for their angular distribution. In fact, protons whose energy is degraded by both interaction with the graphite degrader and the collimator will exhibit some high tails in their energy spectrum, resulting in a large scattering angle, from Equation B.14. Similarly, incoming protons angular distribution results in a large angular spread, due to multiple scattering in the degrader and collimator. Protons entering the detector at a large angle will transverse a larger effective thickness than the one seen by a proton impinging the detector perpendicularly, corresponding to the parameter $L$ in equation B.14. An effective larger material thickness will then lead to larger scattering angles.

6.6.3 MDA applied to simulated data

The MDA has been applied to simulated data binned to give several values of occupancy (100, 250, 750, 1000, 2000, 3000, 5000, 10000 protons per frame corresponding to an occupancy range between 0.004 and 0.4% given a detector matrix of 500 × 500 pixels). As the number of protons per frame increases, the capability of correctly tracking proton tracks decreases, as shown in figure 6.20b. Figure 6.20b reports the fraction of proton tracks correctly reconstructed across the detector stack, namely tracking efficiency, as a function of the occupancy level. Tracking efficiency decreases as a rational\(^5\) function (see fit of Figure 6.21b), going from nearly one for the lowest occupancy level to 0.6 for an occupancy of 0.03% (750 protons per frame), and then this value is close to 0.1 for the highest occupancy level (10000 protons per frame or 0.4% occupancy).

The level of occupancy also affects the displacement distributions obtained by applying the MDA to the simulated data, as shown in figure 6.21a. For increasing values

\(^5\)A rational function is any function which can be defined by a rational fraction, i.e. an algebraic fraction such that both the numerator and the denominator are polynomials. In the function used for fitting purposes here, both numerator and denominator of first degree polynomial: $y(x) = \frac{p_1 x + p_2}{q_1 + x}$. 

of occupancy, the displacement distributions become narrower and further divert from the Gaussian behaviour, i.e. presenting higher tails. The rms deviation of these distri-

Figure 6.21: a) Displacement distributions for one projected coordinate obtained by applying the proton tracking algorithm to simulated data. Distributions are displayed for occupancy values of 0.004, 0.04 and 0.2 %. b) Displacement rms plotted against the occupancy level and fitted with a rational function. The simulated displacement is also reported as a dashed line.

To quantify this feature, the MDA was applied to uniformly displaced data. Numerical simulations were performed assigning uniformly distributed positions to hits in the first sensor. The relevant position in the second detector was then calculated by shifting the original position in the first detector by a value \( d \) extracted from a uniform distribution \( d \in U(-30 \text{ pixel}, 30 \text{ pixel}) \). Hits in the two detectors were then subsequently
Chapter 6. Large area CMOS APSs for proton Computed Tomography

Figure 6.22: Reconstructed displacement distribution for numerically simulated uniformly displaced hits. The dashed line shows the simulated distribution, while the solid curves show the displacement distributions obtained by applying the MDA to those data for several values of occupancy or particles per frame (ppf).

reconstructed by using the MDA. The simulated displacement distribution is shown as a dashed line in Figure 6.22, together with the reconstructed distributions for several value of occupancy, ranging from 10 to 2000 hits per frame. Reconstructed displacements distributions of Figure 6.22 narrow down progressively, as occupancy increases, going from the simulated box-like distribution to a bell function.

Effects of noise and detector misalignment on the reconstruction of simulated data

The MDA has been applied so far to simulated data, where only the physics and geometry of the experiment have been accounted for. However, in order to use these simulated results to predict experimental results, some more realistic effects have to be considered, such as noise (both detector noise and background radiation) and detector misalignment.

Noise was assumed to be uniformly distributed across the detector matrix, with a count rate of 1, 5, 10, 20, 50 and 80% of the signal count rate, corresponding to a SNR of 101, 21, 11, 6, 3 and 2.25 respectively. When those values are compared with the
6.6. Proton tracking

Figure 6.23: a) Displacement distributions for one projected coordinate obtained by applying the proton tracking algorithm to simulated data for several values of SNR. The simulated displacement is shown with a dashed line. b) Tracking efficiency for several values of SNR, compared with noiseless simulation (red symbols and curve).

SNR measured for Experiment 6 (see Section 6.5.3), measurements stand at the lower end of this range. In fact for the lowest current used in the Experiment 6, the estimated SNR is \( \simeq 2.5 \), while this value is relatively higher for the higher current used \( \simeq 7 \).

The measured displacement, as obtained by applying the MDA to simulated data is shown in Figure 6.23a for several values of SNR. At relatively high SNR \( (SNR > 10) \), the reconstructed displacement rms is weakly affected by SNR. However, for lower SNR values \( (SNR < 10) \), the effect of the added noise results in an increase in reconstructed displacement rms, going from the simulated values of 16.7 pixel (dashed line in Figure 6.23a), to \( \simeq 22 \) pixel for the lowest SNR at the lowest occupancy. For higher occupancies the discrepancy between high and low SNR decreases. The tracking efficiency of the MDA applied to simulated data is shown in Figure 6.23b for several values of SNR and compared with the tracking efficiency of noiseless simulations. As expected, the tracking efficiency decreases with SNR: going from an efficiency of \( \simeq 1 \), for the lowest occupancy when noise is not accounted for, to \( \simeq 0.5 \) at the lowest SNR. In addition
to noise, experimental data might also suffer from misalignment between the two detectors. As described in Section 6.3, the two detectors were physically flipped and no precision alignment tools were available at the time of the experiment. For this reason, acquired images had to be flipped and subsequently aligned by fitting a beam profile at high current for the two detectors.

A certain amount of 1D-misalignment has been added to simulated data (corresponding to an SNR=3), by shifting in one direction the simulated hit position in the second detector of an amount of 1, 5, 10 and 25 pixels. The displacement rms calculated for these misaligned data is shown in Figure 6.24a. For low values of misalignment (5 pixels) the difference with perfectly aligned detectors is negligible. However, for increasing values of misalignment the reconstructed displacement becomes larger (31 pixels for a 25 pixel misalignment at the lowest occupancy). A further effect due to detector misalignment is visible in Figure 6.24b, where the displacement distributions are plotted for several values of misalignment. As misalignment increases the centre of these distributions is shifted towards negative values. In fact, as the algorithm is based on the use of the minimum distance, the artificial shift in the hit position in the second detector due to misalignment is compensated by moving the 1D displacement distribution towards negative values.

### 6.6.4 MDA applied to the experimental data

Finally, the tracking algorithm was applied to the DD data of Experiment 6. The reconstructed displacement distributions for the $x$ and $y$ coordinates are shown in Figure 6.25a and b for data at 0.1nA and 0.5 nA (with the $10^{-2}$ filter), respectively. Both distributions show a Gaussian core, with higher tails in agreement with the Molière theory (see Figure B.1). The displacement rms calculated for the low current data (Figure 6.25a) in the Gaussian region is $\sigma_1 = 20.01 \pm 0.04$ pixels, while at the higher current (Figure 6.25b) this value becomes $\sigma_2 = 11.8 \pm 0.4$ pixels, which has to be compared with the simulated displacement rms $\sigma_{sim} = 16.7 \pm 0.1$ pixels.

The low current experiment provides an overestimate of the simulated displacement, while the high current one an underestimate. This can be explained when considering
Figure 6.24: a) Displacement distributions for one projected coordinate obtained by applying the proton tracking algorithm to simulated data with a SNR of 3 and several values of misalignment. b) Displacement distributions for several misalignment values at a SNR=3.

Figure 6.25: Displacement distribution reconstructed in the X and Y coordinate for the low current (a) and high current experiments (b). Gaussian fits are also shown.
Figure 6.26: Displacement distribution reconstructed for several binning values at the low current (a) and high current experiments (b). Gaussian fits are also shown

the effect of occupancy, noise and misalignment observed in simulations. For Experiment 6 data at the low current a SNR of 2.5 has been estimated in Section 6.5.3, which considering an occupancy of 0.04% gives a displacement rms of $\approx 18$ pixels (0.9 mm), as extrapolated from the plot of Figure 6.23 a. This value is expected to be higher when some misalignment is accounted for (see Figure 6.24 a explaining the measured value of $\sigma_1$). Also, when looking at the data at higher current, a smaller displacement is observed. This is easily explained by accounting for the dependence of the displacement on the occupancy (see Figure 6.21 b).

Data of Experiment 6 have then been binned to reproduce measurements at higher occupancy values. The $x$ and $y$ displacement distributions for several binning values at the two currents are shown in Figure 6.26 a and b. Similarly to that observed in Figure 6.21, the reconstructed displacement distributions keep their Gaussian core with high tails, and narrow down as the bin size, and so the occupancy, increases. The displacement rms calculated for these measured distributions is then compared with data obtained by applying the MDA to simulated data in Figure 6.27. Displacement rms
6.7 Proton tracking in the PRaVDa Range Telescope

Figure 6.27: Displacement rms is shown for data of Experiment 6 for several binning values and compared with the simulated displacement (dashed line), noiseless simulations, simulations with SNR=3 and with simulations with SNR=3 and misalignment of 10 pixels.

obtained for the two currents of Experiments 6 (red and black symbols) is compared with the simulated displacement (dashed line), noiseless simulations, simulations with SNR=3 and with simulations with SNR=3 and misalignment of 10 pixels. Experimental data seem to be relatively consistent with simulated data. This demonstrated that individual protons can be tracked in the DD stack with a certain efficiency depending on the occupancy, noise and misalignment.

6.7 Proton tracking in the PRaVDa Range Telescope

An investigation into the proton tracking capability of CMOS APSs has been provided in the previous Section. It is now useful to extend the tracking algorithm as assessing its efficiency over the whole RT.

The proton CT modality of the PRaVDA system requires proton tracks to be reconstructed across the whole detection system. Hits in the four sets of silicon strip detectors
Chapter 6. Large area CMOS APSs for proton Computed Tomography

(SSDs), as well as in the CMOS range telescope (RT), will need to be disentangled to reconstruct proton tracks. A very short exposure time, comparable to the cyclotron beam clock (37 ns), is envisaged for the SSDs, resulting in an extremely low level of occupancy and thus in a high tracking efficiency. Conversely, the exposure time of the CMOS RT is expected to be several order of magnitude higher (1 ms) than that of SSDs, requiring a trade-off between experiment length and tracking efficiency.

In order to provide information for an optimum trade-off in terms of tracking efficiency, simulated data are presented here to test the efficiency of a tracking algorithm developed for the CMOS RT. This algorithm takes advantage of directional information provided by SDs for protons impinging the RT, and it is based on a layer-to-layer adaptive calculation of proton trajectory to estimate the expected proton coordinates in the following RT layer.

This algorithm uses the assumption of ideal detectors. Detector specific effects such as noise (i.e. false positives), detection threshold, (i.e. missing hits below threshold), and detector rolling shutter are not accounted for. Thus the reconstruction efficiency estimated here has to be considered as an idealised case, and as such, this figure is expected to deteriorate when those detector specific phenomena are factored in.

6.7.1 Tracking: the reconstruction algorithm

Compared to the DD experiment, the PRaVDA system will benefit from a more complex set-up where proton tracks can be characterised both in terms of proton hits in each CMOS layer and velocity unity vector, i.e. trajectory direction. In fact, from the point of view of the RT, incoming proton trajectories are completely known by means of their velocity vectors measured by the last set of SDs. Also velocity unit vectors (or directions) can be calculated \textit{a posteriori} across the RT, by accounting for proton displacement every pair of layers. The proposed tracking algorithm for the PRaVDA system works as follows:

Given a number of $N$ incident protons per frame crossing $M$ detection layers (where layer $n = 0$ correspond to the last set of SDs and $n = M − 1$ is the last CMOS sensor in the RT), the position vector and velocity unit vector of the $i^{th}$ of $N$ protons at the
layer \( n = 0 \) are measured as \( \vec{r}_{\text{meas}}(i, n) \) and \( \hat{v}_{\text{meas}}(i, n) \) respectively. The expected position of the proton \( i \) in the layer \( n + 1 \) (first CMOS layer) \( \vec{r}_{\text{exp}}^{(i, n+1)} \) is calculated by means of the following parametric set of equations:

\[
\begin{align*}
\vec{r}_{\text{exp}}^{(i, n+1)} &= \vec{r}_{\text{meas}}^{(i, n)} + \hat{v}_{\text{meas}}^{(i, n)} \cdot t \\
t_x^{(i, n+1)} &= d
\end{align*}
\]  

(6.3)

where \( t \) is the equation parameter and \( d \) is the sensor-to-sensor distance. The measured position in the layer \( n \) will then correspond to that of \( j^{th} \) of the \( N \) detected hits in the layer \( n \) such that:

\[
j: \| \vec{r}_{\text{exp}}^{(i, n+1)} - \vec{r}_{\text{meas}}^{(j, n+1)} \| = \\
\min_{1 \leq j \leq N} \| \vec{r}_{\text{exp}}^{(i, n+1)} - \vec{r}_{\text{meas}}^{(j, n+1)} \|
\]

(6.4)

The position of the \( j^{th} \) hit is then assigned as position of the proton \( i \) in the layer \( n + 1 \):

\[
\vec{r}_{\text{meas}}^{(i, n+1)} = \vec{r}_{\text{meas}}^{(j, n+1)}
\]

(6.5)

Known the position of the proton at the layers \( n \) and \( n + 1 \), a new velocity unit vector can adaptively be calculated:

\[
\hat{v}_{\text{meas}}^{(i, n+1)} = \frac{\vec{r}_{\text{meas}}^{(i, n+1)} - \vec{r}_{\text{meas}}^{(i, n)}}{\| \vec{r}_{\text{meas}}^{(i, n+1)} - \vec{r}_{\text{meas}}^{(i, n)} \|}
\]

(6.6)

This process has then to be iterated for all the \( M - 1 \) layers of the RT and for each proton track.

### 6.7.2 Monte Carlo simulations

The tracking algorithm previously described has been applied to simulated data. GEANT 4 simulations\(^6\) have been performed for the complete PRaVDA system, according to the specification below:

1. Beam

\(^6\)Details on the implementation of these simulations are provided in Appendix C.
(a) Gaussian energy distribution $G(\mu, \sigma)$, $\mu = 190.8$ MeV and $\sigma = 1.53$ MeV
(b) Gaussian angular distribution $G(\mu, \sigma)$, $\mu = 0$ deg and $\sigma = 0.57$ deg
(c) Circular beam of 5 cm radius

2. Four set of SDs

(a) Three strips per set
(b) Silicon thickness 150 $\mu$m
(c) $10 \times 10$ cm$^2$ active area
(d) 100 $\mu$m pixel pitch
(e) 5 mm strip-to-strip distance

3. Phantom

(a) Sphere of 21 cm diameter
(b) Phantom material is water

4. CMOS RT

(a) RT made of 16 CMOS sensors
(b) Each sensor consists of 14 $\mu$m thick epitaxial (sensitive volume) and 753 $\mu$m substrate
(c) $10 \times 10$ cm$^2$ active area
(d) 200 $\mu$m pixel pitch
(e) 1 cm sensor-to-sensor distance
(f) 1 mm thick PMMA attenuator placed in front of each sensor

Position and velocity unit vectors of impinging protons have been scored in the last strip module and in all the layers of the CMOS RT.
6.7.3 Tracking efficiency

The tracking algorithm described previously has been applied to simulated data to evaluate the tracking efficiency for the PRaVDA RT, namely the ratio between number of tracks correctly reconstructed and the number of simulated tracks. Figure 6.28 shows the tracking efficiency as a function of frame occupancy, for the first (# 1), middle (# 5) and last layer (# 16) of the RT. The value at the last layer represents the total tracking efficiency of the RT. Tracking efficiency appears to decrease linearly with occupancy, and it also has a dependency on the layer number, decreasing towards the end of the range telescope. For occupancy values below 0.4% or 100 protons per frame, tracking efficiency is comparable for the three different layers considered and is close to 99%. For higher occupancy values, the efficiency drops sharply and larger differences are appreciable for the different layers. For comparison, tracking efficiency for the DD stack calculated using the MDA from Figure 6.21, is also reported in Figure 6.28. The advantage of using directional information for proton tracking is clear, with a significant gain in tracking efficiency for the RT compared the DD. For the DD stack, at the highest occupancy considered (0.4%) the tracking efficiency is 11%, when at the same
occupancy this value is 89% for the last layer of the RT.

Even with a significant gain in efficiency for tracking in the RT, a very low occupancy level would be desirable, as mis-reconstructed proton tracks will deteriorate resolution of the reconstructed CT images. However, the occupancy level is inversely proportional to the experimental length and thus to dose, since a fixed number of proton tracks is needed for CT reconstruction. A trade-off then has to be found between the efficiency in reconstructing proton tracks and the necessity of having an experiment realisable in terms of experimental length.

Table 6.4 reports tracking efficiency\(^7\) in the first (\(\eta_{FIRST}\)) and last (\(\eta_{LAST}\)) layer of the RT for several occupancy values, ranging from 0.002 % to 2%. Each of those values are compared with the number of frames required per angle \(^8\), the exposure time required at each angle and the total exposure time for acquiring the projections over 180°. From data of Table 6.4, it appears that an occupancy level of 0.4% can guarantee a reasonably high efficiency (89% at the last layer), with an experimental length of 50 minutes for a pCT scan to be performed. Although such an acquisition time is still acceptable for the purposes of the development of a new technology, it might result too long for a clinical application. Conventional X-ray CT scans take between 5 and 10 minutes to be performed, depending on the part of the body to be imaged. A duration equal or lower than 10 minutes represents an acceptable limit to which pCT scans should be aiming to. Nevertheless, these results are promising proof of concept which might lead to a clinically usable system with upgrades of the detector technology, mainly in terms of readout speed, further improving those figures.

6.8 Discussion

Some of the proof of concept, needed for CMOS APSs to be used in an energy-range telescope for pCT, have been discussed in this Chapter.

\(^7\)Fraction of total proton tracks which can be correctly reconstructed.

\(^8\)In the assumption that \(3 \times 10^9\) protons over 180° are required for CT reconstruction.
### Table 6.4: Tracking efficiency in the first ($\eta_{FIRST}$) and last ($\eta_{LAST}$) layer, expected number of frames per angle, expected exposure time per angle and expected total exposure time.

<table>
<thead>
<tr>
<th>N protons per frame</th>
<th>Occupancy$^a$ (%)</th>
<th>$\eta_{FIRST}$ (%)</th>
<th>$\eta_{LAST}$ (%)</th>
<th>N Frames$^b$ per angle</th>
<th>Exp.Time$^{c,d}$ (s)</th>
<th>Total exp.$^d$ (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.002</td>
<td>99.9</td>
<td>99.8</td>
<td>$3.3 \times 10^6$</td>
<td>$3.3 \times 10^3$</td>
<td>10000</td>
</tr>
<tr>
<td>10</td>
<td>0.004</td>
<td>99.9</td>
<td>99.8</td>
<td>$1.7 \times 10^6$</td>
<td>$1.7 \times 10^3$</td>
<td>5000</td>
</tr>
<tr>
<td>50</td>
<td>0.02</td>
<td>99.7</td>
<td>99.1</td>
<td>$3.3 \times 10^5$</td>
<td>333.3</td>
<td>1000</td>
</tr>
<tr>
<td>100</td>
<td>0.04</td>
<td>99.4</td>
<td>98.4</td>
<td>$1.7 \times 10^5$</td>
<td>166.7</td>
<td>500</td>
</tr>
<tr>
<td>500</td>
<td>0.2</td>
<td>97.8</td>
<td>93.7</td>
<td>$3.3 \times 10^4$</td>
<td>33.3</td>
<td>100</td>
</tr>
<tr>
<td>1000</td>
<td>0.4</td>
<td>96.0</td>
<td>89.1</td>
<td>$1.7 \times 10^4$</td>
<td>16.7</td>
<td>50</td>
</tr>
<tr>
<td>5000</td>
<td>2</td>
<td>87.0</td>
<td>60.1</td>
<td>$3.3 \times 10^3$</td>
<td>3.3</td>
<td>10</td>
</tr>
</tbody>
</table>

$^a$ Based on a 500 × 500 pixel active area.

$^b$ Based on the requirement of $3 \times 10^9$ proton over 180°.

$^c$ This is the detector live time, any dead time resulting from data transmission etc. is not accounted for.

$^d$ Based on a readout speed of 1000 fps.
The capability of proton counting for CMOS APSs has been demonstrated through a wide range of energies (36-191 MeV) and also at the expected energy for pCT (60 MeV). The signal measured in the proton counting experiment has been compared with simulated data, showing good agreement which further supports the evidence that we can observe proton counting with this technology. The effect of the detector noise floor and collection efficiency have also been studied with respect to proton counting, providing figures for the design of the PRaVDA RT.

Prior to investigating proton tracking, an analysis in terms of correlation between events detected in the two stacked CMOS detectors has been performed. For those experiments where a high SNR was achieved, event correlation has been proved, suggesting the evidence that individual protons are imaged in the two detectors.

Finally, the capability of proton tracking has been demonstrated for CMOS APSs by using a reconstruction algorithm based on minimising the layer-to-layer lateral displacement. Such an algorithm has been applied to simulated data and the effect of occupancy, noise and detector misalignment have been accounted for. Applying the same algorithm to experimental data has given results comparable with simulations, demonstrating the potential capability of proton tracking for CMOS APSs. A further analysis on the track reconstruction task has been carried out for the proposed PRaVDA RT, by using simulated data, in order to assess the efficiency of the complete reconstruction algorithm, and this has been linked to other important experimental parameters, such as beam current and exposure time, demonstrating, for the first time, that proton tracking for pCT in a clinical situation is viable using CMOS APSs.
As discussed in Chapter 1, bio-medical imaging is a large term umbrella ranging from pre-clinical imaging, to diagnostic imaging and imaging to assist and plan patient treatment. This field of research is pivotal to both driving advances in healthcare and delivering patient care. It can benefit from new detector technologies which have the potential to reduce image acquisition time, dose and improve image quality.

Large area CMOS Active Pixel Sensors (APSs) have the potential to deliver these advances in such a demanding and continuously evolving field. A large imaging area, together with a low noise, low cost, fast readout, high dynamic range and potential for in-pixel intelligence have made this technology an idea candidate to replace currently used imaging technologies in biological and medical research and practice. This thesis represents the first investigation into the capabilities of such large area CMOS APSs to be used across a number of different imaging modalities in bio-medical science, spanning protein imaging to proton CT, using both ionising and non-ionising radiation sources.

The main conclusions of this work are drawn together and presented in this Chapter. These are followed by a summary of proposed further work based on the conclusions of this thesis.
7.1 Conclusions

A detailed characterisation of a large area CMOS APS, namely DynAMITe, has been presented and set into context of commonly used detectors for medical imaging (Chapter 3). Such a characterisation largely relies on a novel methodology, based on a per-pixel analysis of the detector performance, which has been developed to assess inhomogeneity issues arising from the stitching techniques used to manufacture wafer scale sensors.

The per-pixel characterisation of the DynAMITe sensor showed noise and conversion gain being log-normally distributed across the pixel matrix. A theoretical model for signal generation in the pixel array has been provided and proved capable of accounting for noise and gain measured distributions. The relative width for the pixel distribution of conversion gain is 50% and 60%, while these figures are 30% and 35% for read noise for Pixels and Sub-Pixel respectively. Those figures of merit are largely comparable between the two pixel types, providing the evidence that such inhomogeneities arise from the manufacture processes, which is the same for both pixel types, and is not related to the specific pixel design. Furthermore, when the performance evaluation for this sensor is carried on at stitching blocks level, the coefficient of variation for conversion gain and noise for both pixel types is relatively low ($\leq 1.9\%$). This figure is considerably lower of that calculated by Zin et al. [37] for the LAS sensor [12], a large area APS designed for bio-medical applications, reported to be in the range 3.79 - 7.02%, demonstrating a higher level of uniformity for the DynAMITe detector in terms of optical performance.

A further investigation has focused on assessing performance metrics commonly used as benchmark for medical imaging detectors, so that a direct comparison between this new technology and detectors routinely used in clinical practice could be made in order to determine the suitability of large area CMOS APSs in medical imaging. Contrast-to-Noise ratio (CNR) has been measured, in mammographic conditions, as high as 34.4 for the DynAMITe Pixels with a uniformity factor of 79%. The figure for CNR is the highest among a number of digital detectors commonly in mammography, as reported by [44], whilst the uniformity factor is higher than that reported for a Silicon Strip De-
tector (SSD) based mammography system, but lower compared to Flat Panel Imagers (FPDs).

Detective Quantum Efficiency at zero-frequency (DQE(0)) has also been theoretically evaluated by using a cascaded linear system model, and compared with FPDs commonly used for radiology imaging. DQE(0) result was higher (0.2 - 0.4) for the Sub-Pixels than that of any other FPI involved in the comparison. This analysis provides evidence of how the intrinsic lower read noise of APSs, compared to FPDs, results in higher DQE performance.

Considering the performance parameters assessed for this detector, in comparison with digital detectors commonly used in the clinical practice, demonstrates how such large area sensor technology may be successfully employed in medical imaging.

The novel large area CMOS APS, studied in this work, has been proposed as a multimodality imaging platform for use in pre-clinical science (Chapter 4 and Appendix A), and compared with the state-of-the-art detector technologies for western blotting and tissue autoradiography. For the first time direct “contact print” imaging of radioactive and optical labeled biological samples on a large imaging area have been demonstrated, showing its potential application to a broad range of ionising and non-ionising imaging probes, from western blotting in proteomics to tissue autoradiography.

When the protein detection capability of the DynAMITe detector in western blotting was compared with both film emulsion and commercially available digital systems, DynAMITe demonstrated sensitivity to more protein bands than either film emulsion or a commonly used CCD-based western blotting detection system. Denatured *E. Coli* proteins, after electrophoretic separation and chemiluminescence activation, have been imaged with the DynAMITe detector, film emulsion and a commercially available CCD-based imaging system. The DynAMITe imaging system proved itself capable of detecting protein bands corresponding approximately to a molecular weight of 20, 37 and 50 kDa when exposed for 5 minutes, while the other two systems were only able
to detect the more abundant, and thus more active, band at 20 kDa. The higher sensitivity of the DynAMITe system for this protein sample has been evaluated as $6.8 \pm 0.6$ DN/s, while the same figure is as low as $1.39 \pm 0.04$ DN/s for the CCD-based system. Thus CMOS APSs, largely benefiting from an increased sensitivity due to direct “contact print” imaging, can be considered a viable alternative to commonly used imaging systems in western blotting.

Detection capabilities of this large area imaging system have been compared with the state-of-the-art devices for tissue autoradiography, mainly based on the use of hybrid CMOS sensors [116, 117], small area CMOS and CCD sensors [63] and Micro-Channel Plates [118, 119]. The DynAMITe sensor exhibits a sensitivity as high as $3.6 \times 10^{-3}$ cps mm$^{-2}$ kBq$^{-1}$ g and a Minimum Detectable Activity of 0.06 Bq; a level of performance comparable with that reported for its competitors. Moreover, the large imaging area available for this APS ($\approx 164$ cm$^2$) represents a significant improvement in meeting the imaging needs of most autoradiography applications, compared to an imaging area of a few squared-centimetres offered by the other silicon-based direct detection competitors. However, the DynAMITe detector used in this study has been developed for a large range of bio-medical applications and thus some of its design specifications, such as the intrinsic pixel size of 50 µm, is inadequate for some autoradiography application (e.g. brain tissue imaging, where features of interest are a few microns wide), but might be adequate for others, such as as whole-body autoradiography of small animals. Thus, the imaging potential for large area CMOS APSs in autoradiography have been demonstrated, but, in order for this technology to truly impact in this field, a smaller pixel pitch and higher spatial resolution are needed.

Whilst CMOS APSs have started to become a commercial alternative to Flat Panel Imagers in radiological applications [26, 27, 32], the imaging needs in the medical field are also changing, with particle therapy, and more often proton therapy, becoming pivotal in cancer treatment. Requirements for CMOS APSs, in order to impact upon these broader medical imaging needs, become even more demanding with the requirements of a significant radiation tolerance to both X-rays and particle radiation fields. The radi-

---

**Chapter 7. Conclusions and Further Work**

Chapter 7. Conclusions and Further Work
ation hardness of a novel large area CMOS APS, designed for medical applications and hardened-by-design, has been studied (Chapter 5). The radiation damage, produced in this sensor by X-ray and proton irradiation, has been studied as function of total ionising dose and displacement damage dose. The damage contribution from ionising and non-ionising energy deposition has been separated for the proton field and been demonstrated to be independent from proton energy providing a further verification of the Non Ionising Energy Loss (NIEL) scaling hypothesis.

The utility of this radiation-hardened CMOS APS in the clinical environment has been considered by estimating its lifetime in routine applications. For example, for this detector to be used in Mega-Voltage radiotherapy a lifetime of 30000 treatment fractions or nearly four years can be envisaged, largely comparable with current technologies.

The radiation tolerance assessed for this detector, for routine use in clinical practice, demonstrates how such large area sensor technology may be successfully employed in X-ray and proton based imaging applications.

In Chapter 6, potential advantages in using CMOS APSs as an energy-range telescope for proton-CT (pCT) have been discussed, offering a pixellated readout where a larger number of events can be detected in parallel per each readout cycle, compared to conventionally used scintillator based calorimeters. In order for CMOS APSs to be used in proton CT, this detector technology has to demonstrate capability of single proton imaging and being able to track protons as they pass through a stack of such detectors, for proton range to be measured and ultimately for proton energy to be reconstructed.

Capability for individual proton counting, together with the potential for energy deposition measurements for CMOS APSs has been demonstrated for the first time over the large range of energies of interest for this applications. Furthermore, experimental work, based on a simple stack of two CMOS sensors, as well as simulation work has been carried out to prove the capability of such a detection system for proton tracking.
Chapter 7. Conclusions and Further Work

The feasibility of using CMOS APSs as energy-range detectors in pCT has been demonstrated, by studying event correlation in two stacked detectors. An analysis in terms of correlation between events detected in the two stacked CMOS detectors has been performed. For those experiments where a high SNR was achieved (SNR\(\geq\)2), significant correlation was confirmed (\(\rho \geq 0.7\)), suggesting the evidence of same protons been imaged by the two detectors, a condition necessary to perform proton tracking.

Novel algorithms have been developed for proton tracking in a CMOS-based Range Telescope (RT), showing a tracking efficiency of 89% at 0.4% occupancy with a readout of 1000 fps, resulting in an acquisition time of 50 minutes for a full pCT scan to be performed. Although such an acquisition time is still acceptable for the purposes of the development of a new technology, it may be too long for clinical application. Nevertheless, these results are a promising proof of concept which might lead to a clinically usable system with upgrades of the detector technology, mainly in terms of readout speed, further improving those figures.

7.2 Further work

Several interesting aspects of this work, arising from the results produced in this work, may be further explored.

The imaging performance of the DynAMITe detector have shown the potential for large area CMOS APSs to be routinely used in clinical imaging. Although a quantitative assessment of the detector performance in diagnostic imaging has been mainly focused on general radiology and mammography, the potentialities of CMOS detector technology in this field can be further exploited and extended to other diagnostic applications, e.g. fluoroscopy, CT, nuclear medicine imaging. Besides further work on, for example, optimisation (in terms of readout speed and imaging performance), choice of the scintillator material, the potential for in-pixel intelligence in CMOS sensor (arising from the monolithic nature of this technology) can really represent a breakthrough in this field. Adaptive imaging correction, photon counting, spectroscopic capabilities,
on-line dose sensing and optimisation are all interesting features which can arise from the in-pixel intelligence of CMOS sensors and that can represent a significant improvement in terms of imaging quality and dose reduction in diagnostic imaging.

Detection capabilities of the DynAMiTe detector in proteomics, exceeding performance of both film emulsion and CCD-based systems, have demonstrated the potential for large CMOS APSs to become the routine detector of choice in bio-imaging, paving the way for the development of a single platform for multi-modality imaging in pre-clinical science. Further steps in this direction will have to involve extending this platform to other imaging modalities used in this field. One such example is fluorescent imaging, especially considering the recent trend in substituting radioactive biological probes with non-radioactive ones (i.e. light emitting probes) due to health and safety concerns. However, in contrast to chemiluminescence and radioactive probes in this work, fluorescent imaging needs external illumination for the fluorescence reaction to be activated as well as optical filters to separate excitation light from fluorescence emission, based on wavelength. Such an experimental setting needs careful choice in terms of illumination and wavelength filtering, in order not to overshadow the advantages in terms of detection and geometrical efficiency arising from “contact print” imaging presented in this work. Flat field illumination and wavelength filters directly printed on a thin support medium are options to be looked at for fluorescence imaging with CMOS sensors, so that a single imaging system, featuring a compact size and with high efficiency, can be used for a variety of imaging application in today’s life science laboratories.

Moving to imaging for proton therapy, the full potential of large area CMOS APSs have yet to be fully exploited. While a first proof of concept of the capabilities of such detector technology for patient imaging prior to treatment has been provided in this thesis, systematic work to realise an energy-range telescope for proton CT is needed. The design of CMOS detectors with a ultra-fast readout represents the next step of this process, so that individual proton counting can be performed while using the whole sensor area. Furthermore, algorithms for proton tracking and residual energy measurements will need to be further developed, possibly by using an approach
based on machine learning. The residual energy measurements from the energy-range telescope could also greatly benefit from energy deposition measurements ($dE/dx$) in each of the CMOS layers of the energy-range telescope described in this work. Such approach could improve the telescope energy resolution and ultimately the resolution in mapping the patient tissue stopping power, thus leading to better treatment planning.
Appendix A

Digital autoradiography with a large area CMOS APS

Autoradiography (AR) is a well established technique for structural and metabolic analysis of cells and tissues. It is aimed at the localization, within a specimen, of a label (such as radioactive tag) bound to a specific structure (such as genes, specific morphological structures or a drug receptors), by placing the specimen against some type of sensitive imaging sensor.

AR samples are labeled with radioactive $\beta$ emitters (e.g. $^{14}$C, $^{3}$H, $^{35}$S, $^{32}$P), bound to specific biomolecules, which can be taken up by living organisms or can chemically bind other biomolecules of interest. Detecting the position and the activity of radiolabels can provide information on both the metabolism of the binding biomolecule used and on the binding sites itself. The principle of $\beta$-labelled autoradiographic process is that radioactive decays taking place within the specimen produce emission of particles ($\beta$) which, after detection over a suitable exposure time, can produce an image of the activity map in the specimen.

Film emulsion is the detection medium of choice for AR. It offers an unbeatable spatial resolution (1-5µm for low energy beta-emitters) due to its fine granularity, but on the other side it offers significant drawbacks which limit its imaging performance: film emulsion suffers from a limited dynamic range (2-3 decades) and non-linear re-
Appendix A. Digital autoradiography with a large area CMOS APS

response, which obscures high and low activity regions within the same sample to be simultaneously resolved. Moreover, film emulsion is characterised by low sensitivity which requires lengthy exposures, up to several months for low energy emitters ($^3$H): a significant bottleneck in the routine experimental work-flow.

A range of alternative digital technologies have been proposed to address these limitations, mirroring the general transition from analogue-based to digital-based systems, generally observed in medical imaging. Indirect detection in AR has been investigated through storage phosphors and scintillators, offering a 10 to 100 fold increase in sensitivity compared to film emulsion [120] together with a relative high spatial resolution (20 $\mu$m [121]). Multi-wire proportional chambers (MWPCs) have been used for this application exhibiting a relatively low spatial resolution (400 $\mu$m [122]), whereas a better resolution is offered by Micro-Channel Plates (MCPs): 26 $\mu$m for small area devices and 60 $\mu$m for larger areas [118, 119]. Alternative approaches have been attempted with micro-strip detectors [123, 124] and gaseous detectors [125].

Silicon based pixel detectors have been studied at length, as a suitable digital alternative to film emulsion, ranging from Charge Coupled Devices (CCDs) [63], CMOS Active Pixel Sensors (APSs) [63] to CMOS Hybrid pixel sensors [116, 117]. Even where these silicon based systems have demonstrated suitability for AR applications, because of a good sensitivity, low background and acceptable resolution, they are not yet used routinely because of their of their modest active area (a few square centimetres) compared with the typical sample sizes (between 10 and 100 $cm^2$). Their limited active area places an upper bound on throughput, off-setting the benefits of any performance gain, and so limiting application beyond proof-of-concept.

Recent advances in photolithographic techniques [33] have made available reticle-stitching process to scale up CMOS APSs up to wafer scale (13 cm × 13 cm) (see Chapter 4). Wafer scale CMOS APSs can therefore present a valuable alternative to overcome the performance limitation of film emulsion whilst offering a sensitive area that meets the needs of many AR imaging applications. First exemplar results of imaging beta-labelled ex-vivo tissue sections with the DynAMITe detector will be presented in this Appendix,
A.1 Dark correction

\[ \beta \] AR is an application characterised by low activity samples. Signals generated in CMOS sensors exposed to radiation can be normally considered as the sum of the detector dark current, due to thermally generated electrons, detector noise and the charge generated by the incoming radiation. A conventional procedure for dark current correction in CMOS involves subtracting a dark reference frame from the image of interest \([38]\). However, dark current is not constant over time but represents stochastic processes within each pixel, so that dark subtraction will result in a residual dark current related noise in the image of interest, due to dark current fluctuations. When these residual dark current fluctuations have to be compared with low intensity signals, they may appear of similar magnitude to the desired signal, particularly where long exposure times are used. Hence imaging applications which generate very low levels of signal,
such as AR, could then be effectively swamped by the sensor noise, thus obscuring useful image information.

The dark correction procedure proposed in this work is based on modelling the individual pixel response and evaluating the dark current fluctuations over time. This is reliant on the following set of assumptions:

1. The dark signal of a single pixel can be regarded as a Gaussian distributed random variable over time, as shown in Figure A.1;

2. The probability of two $\beta$ particles impinging the detector in the same position in the same frame is negligible (i.e. no overlapped events) and also the likelihood of the same pixel being fired in two consecutive frames in negligible.

Dark current is thus represented by per pixel average and standard deviation calculated over a 1 hour long dark reference set. Corrected images $\text{Corr}(i,j)$ are obtained from raw images $\text{Raw}(i,j)$ as follows:

$$\text{Corr}(i,j) = \text{Raw}(i,j) - [\text{Mean}_{\text{dark}}(i,j) + k \cdot \text{Mean}_{\text{dark}}(i,j) + q \cdot \text{StDev}_{\text{dark}}(i,j)]$$ (A.1)

where $\text{Mean}_{\text{dark}}(i,j)$ represents the average calculated over a dark reference set and corrects for the dark current offset, $k = \frac{\text{Raw}(i,j) - \text{Mean}_{\text{dark}}(i,j)}{\text{Mean}_{\text{dark}}(i,j)}$ is an adaptive weighting factor which takes into account temperature shifts of the sensor resulting in dark current shift, and $q \cdot \text{StDev}_{\text{dark}}(i,j)$ is the pixel-by-pixel standard deviation of a dark reference set multiplied by a weighting factor $q$ which corrects for dark current fluctuations. An application of the dark correction procedure is displayed in Figure A.2. Figure A.2 'Raw image' shows a detector ROI, wherein a $\beta$ interacted. However any specific particle tracks due to the $\beta$ signal is swamped by dark signal. Figure A.2 'Dark correction' shows the previous image after applying conventional dark subtraction procedure [38]. A large cluster of hit pixels is visible together with a number of single pixel clusters, which are false events (residual dark signal). In Figure A.2 'Dark correction' the raw image is corrected following the procedure described in this section. In this image the large cluster (true event) is still visible while the residual dark signal is suppressed, increasing the false event rejection rate, without compromising event sensitivity.
A.2 Signal analysis and clustering techniques

Figure A.2: Application of the dark correction procedure. *Raw image*: signal detected in a ROI where a β interacted is displayed as raw data. *Dark subtraction*: the previous image is corrected by conventional dark subtraction [38]. *Dark correction*: raw image corrected following the dark correction procedure. A magnified view of the β track is also displayed.

A.2 Signal analysis and clustering techniques

A β particle impinging the sensor with sufficient energy can potentially generate signal in more than one pixel, i.e. forming clusters of connected pixels, due to its range in the detector active volume, and, due to possible charge sharing across neighbouring pixels. In order to further suppress any residual signal unrelated to the source used, i.e. residual dark current fluctuations, cosmic and environmental radiation, an event analysis procedure has been developed, taking advantage of multi-pixel events to discriminate between source related events and background events.

Event analysis allows identification of a suitable range, in terms of cluster size and energy deposited, where the probability of signal generated by the source of interest is higher. A standard $^{14}$C source, has been used for this purpose. The cluster size distribution produced by a $^{14}$C source placed in close contact with the detector surface has been measured, resulting in an average cluster size of 2.5 pixel.

Energy summation of $^{14}$C events has been performed using a source covering a limited area of the detector surface. Two region of interests (ROIs) were set: the first one
corresponding to the source and the other one to a region where no sources are present. Signal recorded in the two ROIs, after dark correction, was summed on the pixels forming each cluster to reconstruct the signal produced by single events. Hence two different spectral distributions are recorded: a spectrum due to residual dark current fluctuations, environmental and cosmic radiation (ROI w/o source), and a spectrum due to the aforementioned phenomena and source signal (ROI w/ source). Cluster spectra for signals generated in the two ROIs are shown in Figure A.3 for different cluster size events (1-6 pixel clusters).

Selecting spectral regions of the source signal which are not overlapped to the background spectrum (Figure A.3) allows selection of regions where the probability to find a true event is higher. This post-processing procedure, allows further correction of the acquired images, increasing the rejection rate for false events.

**A.3 Linearity, sensitivity and Minimum Detectable Activity**

**A.3.1 Sensitivity, background and Minimum Detectable Activity**

Calibration microscales\(^1\) (shown in Figure A.4 _left_) were used to quantitatively assess the performance of the Dynamite detector for \(^{14}\)C AR in terms of sensitivity, background and minimum detectable activity (MDA). Linearity and sensitivity of the detector are derived by imaging various bands of different activity. Sensitivity is defined as the rate of detected events per unit area divided by the source activity. A ROI of 30 ×70 pixels was delineated for each band in the microscale, and the total number of events (number of beta-particle rays detected) in each band was measured. The total particle count rate per unit area of each band plotted versus its activity concentration represents the sensitivity curve, whereas its slope is the system sensitivity.

\(^1\)AR calibration microscales (Amersham biosciences) are in the form of polymer strips (21 mm × 3.5 mm × 120 µm containing \(^{14}\)C radioactivity uniformly incorporated into the polymer and arranged in eight bands of increasing activity concentration in the range 1.13 - 32.9 kBq g\(^{-1}\).
A.3. Linearity, sensitivity and Minimum Detectable Activity

Figure A.3: Spectral distributions of signal generated in a ROI covered with a $^{14}$C source (blue symbols) and without source (red symbols) for different event size (1-6 pixel clusters) after dark correction.
Figure A.4:  

(a) $^{14}$C calibration microscale displayed with the specific activity of each band.  

(b) Sensitivity curve: specific count rate for each band is plotted versus band specific activity. A linear fit of these data, whose slope represents the system sensitivity, is shown. Horizontal dotted line represents the background level, while the vertical dotted line is the MDA.

The resulting sensitivity curve is shown in figure A.4 right with an exposure time of 2 h. The system sensitivity for $^{14}$C beta-particles is $3.6 \times 10^{-3}$ cps mm$^{-2}$ kBq$^{-1}$ g and the background level was measured as low as $2.19 \times 10^{-4}$ cps mm$^{-2}$. Combining these figures and considering the microscale mass and geometric efficiency [116], the Minimum Detectable Activity (MDA) of the system can be calculated as the interception point between the sensitivity curve and the background level. MDA results for DynAMITe to be as low as 0.06 Bq.

The performance of the DynAMITe system for $^{14}$C AR is next compared to that of several state-of-the-art systems (Table A.1). The comparative data of Table A.1 indicate how sensitivity, background level, MDA of the DynAMITe system are comparable with those reported for its competitors, while on the other side it offers the largest imaging area available.
### A.3. Linearity, sensitivity and Minimum Detectable Activity

Table A.1: Detection performance comparison of $^{14}$C AR systems

<table>
<thead>
<tr>
<th>Detector</th>
<th>Area (cm$^2$)</th>
<th>Sensitivity (cps mm$^{-3}$ kBq$^{-1}$ g$^{-1}$)</th>
<th>Exposure time (h)</th>
<th>Background (mBq)</th>
<th>MDA (Bq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DynAMTe [This work]</td>
<td>12.8×13.1</td>
<td>3.6×10$^{-3}$</td>
<td>2</td>
<td>2.18×10$^{-4}$</td>
<td>0.06 Bq</td>
</tr>
<tr>
<td>Timepix[117]</td>
<td>0.14×0.14</td>
<td>4.11×10$^{-3}$</td>
<td>3.5</td>
<td>3.5×10$^{-3}$</td>
<td>N.A.</td>
</tr>
<tr>
<td>Timepix[117]</td>
<td>0.11×0.11</td>
<td>4.7×10$^{-3}$</td>
<td>10</td>
<td>18×10$^{-3}$</td>
<td>N.A.</td>
</tr>
<tr>
<td>Medipix2[116]</td>
<td>0.11×0.11</td>
<td>4.7×10$^{-3}$</td>
<td>10</td>
<td>2.1×10$^{-3}$</td>
<td>0.012 Bq</td>
</tr>
<tr>
<td>E2V CCD47-20 [63]</td>
<td>0.17×0.26</td>
<td>24.6×10$^{-3}$</td>
<td>12</td>
<td>3.8×10$^{-3}$</td>
<td>0.025 Bq</td>
</tr>
<tr>
<td>CMOS Vanilla[63]</td>
<td>0.13×0.13</td>
<td>11.4×10$^{-3}$</td>
<td>3.3</td>
<td>6.7×10$^{-3}$</td>
<td>N.A.</td>
</tr>
<tr>
<td>Cooled CCD[126]</td>
<td>0.17×0.26</td>
<td>N.A.</td>
<td>6.6×10$^{-3}$</td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>µImager[121]</td>
<td>0.93×0.93</td>
<td>20</td>
<td>0.04 cpm mm$^{-2}$</td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>MCP[119]</td>
<td>0.93×0.93</td>
<td>0.031 Bq</td>
<td>N.A.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A.4 Tissue imaging

To demonstrate the potential of CMOS APS technology for tissue autoradiography, images of several radio-labelled samples were produced and compared with film emulsion.

A.4.1 Imaging of $^{14}$C-sulphur labelled tissue sections

$^{14}$C tissue samples used in this work were generated as part of a series of in vivo experiments investigating the role of showering on skin decontamination following a radiological incident (ORCHIDS, U.K. Health Protection Agency). In vivo experiments on porcine tissues were designed to compare a putative optimized showering protocol against the current UK protocol employed by the Fire and Rescue Service (FRS protocol), as well as compare the penetration of $^{14}$C-sulphur mustard at the snout to the flank in an animal model.

Figure A.5 shows the first images of $^{14}$C labeled tissue sections imaged with a wafer scale CMOS APS. Images of two tissue sections, dissected from the bridge of snout (sample A) and planum rostrale (sample B), are displayed in Figure A.5. Images labeled as $a$, $b$ and $c$ represent the digital images obtained with 3, 6 and 8 h integration time respectively. Images $d$ show an image obtained when the same sample is exposed to film emulsion for one week.

Figure A.5 demonstrates visibility of all the details contained in the film image $d$) with only 3 h exposure time (sample A, Figure A.5$a$) and in 6 h (sample B, Figure A.5$b$). From this comparison it emerged that the DynAMITe imaging system is able to offer similar quality images as film emulsion with an improvement in exposure time of several decades (between 56 and 38 fold), depending on the sample activity.
Figure A.5: $^{14}$C labeled skin section (sample A a) and sample B b) imaged for 3 h (a), 6 h (b), 8 h (c) represented on the same false colour scale shown in counts/pixel. The same sample imaged with film emulsion after 1 week exposure (d). The red box highlights the section of the original sample exposed to the digital detector.
Figure A.6: $^{125}$I[Epibatidine brain tissue sections imaged with the DynAMITe detector with 1 h a, 5 h b and 9 h c exposure time. Image a) is windowed in [0,1]. A colour scale calibrated in DN pixel$^{-1}$ is also displayed for images b) and c).
A.4. Tissue imaging

Figure A.7: Signal plotted versus exposure time for each sample (A-E). A linear fit for each of the samples is also displayed together with the determination coefficient $R^2$. The background level for this experiment is represented with a dotted line.

A.4.2 Imaging of $^{125}$I Epibatidine labeled brain sections

Mice brain slices were labeled with $^{125}$I Epibatidine to undertake addiction studies on the nicotine brain receptors $\alpha_4\beta_2$-nAChR. In fact $^{125}$I Epibatidine binds with high affinity several nicotine receptors. Brain sections were incubated for 2h with $^{125}$I Epibatidine, with a solution of $^{125}$I Epibatidine and Cytosine to evaluate competitive ligands and with nicotine salts for evaluation of non specific binding.

Samples were imaged after 16 days from the sample preparation with a residual activity of 83% following routine film exposure, given the $^{125}$I half life $T = 59.4$ d. The first images of $^{125}$I label tissue section acquired with a large area CMOS APS are displayed in Figure A.6 a. Images of five slides labeled A-E, containing 12 tissue slices each, are displayed for 1 h, 5 h and 9h exposure in a),b),c) of fig A.6 respectively, together with a colour scale calibrated in Counts pixel$^{-1}$.

$^{2}$The radioisotope $^{125}$I decays by electron capture into $^{125}$Te with emission of a 35.5 keV gamma. Together with these gammas, low energy conversion electrons and Auger electrons are emitted in the energy range 3-30 keV, which, given their short range, can offer a good spatial resolution.
Figure A.7 shows the linearity plot for the imaged samples. For each slide (A-E) signal detected in one of the tissue slices has been plotted as function of the exposure time. The background level for this experiment is represented with a dotted line in figure. For samples C, D and E, which present a higher activity, signal shows a linear behaviour ($R^2 > 0.99$) in all the time range investigated. Samples A and B, characterised by a lower activity, have a signal level comparable with the background level after 1 hour exposure. The linear fit for lower activity samples (slides A and B) has been then performed in the range 2-9 h exposure time, resulting in a coefficient of determination $R^2 > 0.99$. Hence all the samples investigated in this work present a linear behaviour with time, above the background level.

A.5 Discussion

Although the DynAMITe detector is not specifically designed for AR, some tests have been carried out to assess its suitability for this application. Detection capabilities of the DynAMITe imaging detector have been assessed for AR applications and compared with the “gold standard” film emulsion and with state-of-the art digital devices in this field. The detection capabilities of this detector are comparable to those reported for its competitors. First images of tissue samples with a large area CMOS APSs have been presented, demonstrating the potential of this technology for digital AR. Moreover, the large imaging area available for this APS ($\approx 164 \text{ cm}^2$) represents a significant improvement in meeting the imaging needs of most autoradiography applications, compared to an imaging area of a few squared-centimetres offered by the other competing digital technologies.

However, the detector used in this study has been developed for a large range of biomedical applications and thus some of its design specifications, such a pixel size of 50 $\mu$m, result inadequate for some autoradiography application (e.g. brain tissue imaging, where features of interest are a few microns wide), but might be adequate for others, such as as whole-body AR of small animals. Thus, imaging potentials for large area
CMOS APSs in autoradiography have been demonstrated, but, in order for this technology to truly impact in this field, a smaller pixel pitch and higher spatial resolution are needed.
Appendix A. Digital autoradiography with a large area CMOS APS
Appendix B

Multiple Coulomb Scattering

When a charged particle passes through matter in the neighbourhood of a nucleus, its trajectory is subjected to deflection due to the interaction with the Coulomb field generated by the nucleus. This process is known as Multiple Coulomb Scattering (MCS), which, to a first approximation, can be considered an elastic process.

As protons traverse a material, they undergo a large number of collisions with nuclei, each of which is almost negligible. Therefore the observable related phenomenon is the statistical outcome, i.e. a random walk in angle, of numerous tiny deflections. The MCS angular distribution can be considered Gaussian in first approximation from the Central Limit Theorem, since it results from a large number of many small angle deflections. However, the occurrence of large scattering angle, although rare, makes the Central Limit theorem not really applicable to this process and the complete angular distribution for MCS can be seen as a Gaussian core with a single scattering tails.

Several theories were published in the 1930s and 40s to predict the the exact shape and width of the MCS angular distributions, but the most widely accepted is that proposed by Molière [127] in 1947 with a paper in German, and then improved by Bethe [128]. Molière’s theory of MCS is algebraically quite complicated, so before discussing it for a simple case, it is worth presenting first a simplified formula, known as Rossi-Greisen equation for the mean square of the scattering angle, which is in most cases sufficient for the most radiotherapy purposes.
## Appendix B. Multiple Coulomb Scattering

### B.1 Rossi-Greisen equation

Following [115], for the proton-nucleus interaction to be elastic, total momentum conservation is required. This can be expressed as requiring that the incident particle, of charge $ze$, acquires an equal and opposite transverse momentum with respect to that acquired by the interacting nucleus of charge $Ze$. Since this transverse momentum can be considered very small compared to the momentum $p$ of the incident particle, this particle is scattered at angle $\theta$, given by the ratio of the transverse momentum to the total momentum $p$ as a function of the impact parameter $b$:

$$\theta \approx \frac{2Zze^2}{bv}p^{-1} = \frac{2Zze^2}{bvp}$$  \hspace{1cm} (B.1)

where $v$ is the particle velocity. From equation B.1, the absolute value of the deflection $d\theta$ at the angle $\theta$ can be written in terms of the differential impact parameter variation $db$ at $b$:

$$d\theta = \frac{2Zze^2}{b^2vp}db = \frac{\theta^2vp}{2Zze^2}db.$$  \hspace{1cm} (B.2)

Thus the probability of collision $dP_{el}$ for a particle transversing a thickness $dx$ with an impact parameter between $b$ and $b + db$

$$dP_{el} = 2n_A\pi bdbdx = 2\frac{N\rho\pi}{A} bdbdx$$  \hspace{1cm} (B.3)

Substituting $b$ from Equation B.1 and $db$ from Equation B.2, Equation B.3 becomes

$$dP_{el} = 2\frac{N\rho\pi}{A} \left[ \left( \frac{2Zze^2}{vp} \right) \theta^{-1} \right] \left[ \left( \frac{\theta^2vp}{2Zze^2} \right)^{-1} d\theta \right] dx$$

$$= 2\frac{N\rho\pi}{A} \left( \frac{2Zze^2}{vp} \right)^2 \frac{d\theta}{\theta^5} dx$$  \hspace{1cm} (B.4)

$$= \frac{N\rho}{A} \left( \frac{2Zze^2}{vp} \right)^2 \frac{d\Omega}{\theta^4} dx$$

where the last step derives from assuming that the solid angle $d\Omega$ can be written as $d\Omega \approx 2\pi \theta d\theta$ for small scattering angles.

The probability of collision can then be re-expressed as

$$dP_{el} = \Xi(\theta) d\Omega d\chi$$  \hspace{1cm} (B.5)
where \( d\chi = \rho \, dx \) is the thickness of the material transversed, expressed in g cm\(^{-2}\) and \( \Xi(\theta) \), the differential scattering probability, can be written as

\[
\Xi(\theta) \, d\Omega = \frac{N \rho}{A} \left( \frac{2Zze^2}{vp} \right)^2 \frac{d\Omega}{\theta^4} \\
= 4N \frac{Z^2}{A} r_e^2 \left( \frac{zmc}{\beta p} \right)^2 \frac{d\Omega}{\theta^4} \left[ g^{-1} \text{cm}^2 \right]
\]

Equation B.6 is obtained introducing the classical electron radius \( r_e = e^2/mc^2 \) and is known as the Rutherford scattering formula and although valid in many cases of practical use it does present some limitations. The theoretical expression for Equation B.6 depends on the spin of the incident particle for large deflections, however, for small deflections this dependence can be neglected and one can use this expression to a first approximation [129]. Moreover, the finite size of the nucleus and the screening of the nucleus field by outer electrons will also reduce the validity of Equation B.6 for very small and very large deflection angles. When the finite size of a nucleus is accounted for by assuming that the nucleus is not contained in a point but spread across a sphere of radius \( r_n \approx 0.5 \, r_e A^{1/3} \), it can be shown that the value of \( \Xi(\theta) \), calculated by using Equation B.6, is not affected by the finite size of the nucleus for \( \theta < \lambda/(2 \pi r_n) \) with \( \lambda = h/p \) is the de Broglie wavelength of the incoming particle, while \( \Xi(\theta) \) approaches rapidly to zero for \( \theta > \lambda/(2 \pi r_n) \) [129]. The finite size of the nucleus can be accounted by considering the maximum deflected angle for Equation B.6 to be valid

\[
\theta_{\text{max}} = \frac{\lambda}{2 \pi r_n} \approx 2 \frac{h}{p r_e A^{1/3}} = 2 \frac{mc}{p \alpha} A^{-1/3}
\]

Also it can be shown that, taking as effective nuclear radius the Thomas-Fermi radius \( a_Z = \frac{\alpha_0}{Z^{1/3}} \), in which \( \alpha_0 \) is the Bohr radius, the nuclear field screening by the outer electrons does not affected the scattering probability \( \Xi(\theta) \) for \( \theta > \lambda/(2 \pi a_Z) \). A minimum deflection angle for Equation B.6 to be valid can then be calculated as

\[
\theta_{\text{min}} = \frac{\lambda}{2 \pi a_Z} = \frac{h}{pa_0 Z^{-1/3}} = \frac{mca}{p} Z^{1/3}
\]

Using values of Equation B.7 and B.8, the mean square of the scattering angle \( \langle \theta^2 \rangle \) at a depth \( \chi + d\chi \) in the transversed material can be calculated by its value at \( \chi \) in addition to the scattering angle in the thickness \( d\chi \)

\[
d\langle \theta^2 \rangle = d\chi \int_0^{2\pi} \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} \theta^2 \Xi(\theta) \, d\Omega
\]
which, using Equation B.6, becomes

\[
d\langle \theta^2 \rangle = d\chi d\phi \int_0^{\theta_{\text{max}}} \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} \frac{\theta^2}{2} 4 N \frac{Z^2}{A} r_e^2 \left( \frac{zmc}{\beta p} \right)^2 \frac{d\Omega}{\theta^4}
\]

then considering \( d\Omega \simeq \theta d\phi d\theta \) and using Equations B.7 and B.8, \( d\langle \theta^2 \rangle \) can be written as

\[
d\langle \theta^2 \rangle = 4 N \frac{Z^2}{A} r_e^2 \left( \frac{zmc}{\beta p} \right)^2 \int_0^{\theta_{\text{max}}} \frac{2\pi \theta d\theta}{\theta^2} \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} \theta d\theta
\]

\[
= 8 \pi N \frac{Z^2}{A} r_e^2 \left( \frac{zmc}{\beta p} \right)^2 \ln \left( \frac{\theta_{\text{max}}}{\theta_{\text{min}}} \right) d\chi
\]

\[
= 8 \pi N \frac{Z^2}{A} r_e^2 \left( \frac{zmc}{\beta p} \right)^2 \ln \left( \frac{2 \frac{mc}{\beta p} A^{-1/3}}{173 Z^{1/3}} \right) d\chi
\]

\[
= 8 \pi N \frac{Z^2}{A} r_e^2 \left( \frac{zmc}{\beta p} \right)^2 \ln \left( \frac{173 Z^{1/3}}{Z^{1/3}} \right) dx
\]

where the last step arises from \( A \approx 2Z \) and \( d\chi = \rho dx \). Introducing the radiation length as \( X_0 = 4 \frac{N e}{A} \alpha Z^2 r_e^2 \ln(173/Z^{1/3}) \), one obtains

\[
d\langle \theta^2 \rangle \simeq 4 \frac{\pi}{\alpha X_0} \left( \frac{zmc}{\beta p} \right)^2 dx
\]

Integrating Equation B.12 over a thickness \( L \) (expressed in cm) of material transversed, the so-called \textit{Rossi-Greisen equation for the mean square of the scattering angle} can be obtained

\[
d\langle \theta^2 \rangle = \frac{L}{4} \frac{\pi}{\alpha X_0} \left( \frac{zmc}{\beta p} \right)^2
\]

\[
= \frac{4 \pi}{\alpha} L \frac{1}{X_0} \left( \frac{zmc}{\beta p} \right)^2 \left( \frac{z}{vp} \right)^2
\]

where \( E_s^2 = [4(mc^2)\pi]/\alpha = 21.2 \text{MeV} \). The root mean square (rms) of the scattering angle can be written as

\[
\theta^{\text{rms}} = \sqrt{d\langle \theta^2 \rangle} = E_s \left( \frac{z}{vp} \right) \sqrt{\frac{L}{X_0}}
\]

whose projection onto a plane containing the initial particle trajectory is

\[
\theta^{\text{rms}}_{\text{proj}} = \frac{1}{\sqrt{2}} \theta^{\text{rms}}
\]
B.2 Molière theory

A more detailed analysis of the MCS is provided by Molière and Bethe, where the screening effect of the outer electrons is accounted for [127, 128]. Due to the algebraically complicated formulation, a simplified case will be discussed here. Assume that the scattering target consists of a single element (atomic weight $A$, atomic number $Z$), and that it is so thin (thickness $t$) that energy lost by protons transversing the target is negligible ($t <<$ proton range). Also it has to be assumed that $Z$ is large enough so that scattering by atomic electrons is negligible too. Following [130], the characteristic scattering angle is defined as:

$$\chi_c^2 = \frac{c_3 t}{(p v)^2} \tag{B.16}$$

with $c_3$ defined as

$$c_3 = 4 \pi N_A \left( \frac{e^2}{\hbar c} \right)^2 (\hbar c)^2 \frac{z^2 Z^2}{A} \tag{B.17}$$

where $N_A$ is the Avogadro’s number and $\hbar c$ is the fine structure constant. The physical interpretation of $\chi_c$ is that, on average, a proton will undergo a single scatter larger than $\chi_c$, transversing the target. Defined the constants $c_1$ and $c_2$

$$c_1 = \left[ \left( \frac{e^2}{\hbar c} \right) z Z \right]^2 \tag{B.18}$$

and

$$c_2 = \left[ \frac{1}{0.855} \left( \frac{e^2}{\hbar c} \right) (m_e c^2) Z^{1/3} \right] \tag{B.19}$$

a screening angle $\chi_a$ can be calculated

$$\chi_a^2 = \chi_0^2 (1.13 + 3.76 \alpha^2) \tag{B.20}$$

with $\alpha$ being the Born parameter give by $\alpha^2 = c_1/\beta^2$. The physical interpretation of $\chi_a$ is that of the angle at which the single scattering cross section departs from the Rutherford approximation $1/\theta^4$ (see Equation B.6), because of atomic electrons screening the nuclear charge. Next, a quantity being the natural logarithm of the effective number of collisions in the target can be calculated

$$b = ln \left( \frac{\chi_c^2}{1.167 \chi_a^2} \right) \tag{B.21}$$
which is related to the reduced target thickness $B$, defined as the root of the Equation

$$B = \ln b - b.$$  \hfill (B.22)

Finally, the Molière’s characteristic scattering angle can be defined

$$\theta_M = \frac{1}{\sqrt{2}} \left( \chi_c \sqrt{B} \right)$$  \hfill (B.23)

which is the equivalent of $\theta^{\text{rms}}$ in Equation B.14 calculated in the Rossi-Greisen approximation and it is typically 6% larger than this [130]. In order to calculate the distribution of the scattering angle $\theta$, Molière uses a reduced angle

$$\theta' = \frac{\theta}{\chi_c \sqrt{B}}$$  \hfill (B.24)

to approximate the scattering distribution $f(\theta)$ in $1/B$

$$f(\theta) = \frac{1}{2 \pi \theta_M^2} \frac{1}{2} \left[ f^{(0)}(\theta') + \frac{f^{(1)}(\theta')}{B} + \frac{f^{(2)}(\theta')}{B^2} \right]$$  \hfill (B.25)

The first order term is Gaussian

$$f^{(0)}(\theta') = 2 e^{-\theta'^2},$$  \hfill (B.26)

whilst terms of higher order are reported explicitly in [127].

The first proton measurements to verify Molière theory were performed by Bichsel in 1958 [131], who bombarded targets of Al, Ni, Ag and Au with low energy proton (< 5 MeV) accelerated in a Van de Graaff using nuclear emulsions as detectors. Experimental data are reported in Figure B.1 and compared with the full Molière’s theory (solid line) and with the Gaussian approximation (dashed line). Apart from verifying the validity of the Molière model, data of Figure B.1 shows that, for small scattering angles, the angular distribution is approximately Gaussian and, by integrating data of Figure B.1, it can be shown that 98% of the protons fall in the Gaussian region. For this reason, for many practical applications, easier models, such as the Rossi-Greisen, can be adequate in predicting the MCS.
Figure B.1: Scattering angle distribution ($\theta^2$) for protons transversing a photographic emulsion (dots). Experimental data are compared with the complete Molière’s model (solid line) and with the first order of this model, i.e. Gaussian approximation (dashed line) [131].
Appendix C

GEANT4 simulations

GEANT4 [77, 132] is an object oriented toolkit for the simulation of particle interactions with matter. It provides advanced functionality for all the domain related to detector simulations: geometry and material specifications, definition of particles, physics processes, tracking, event and run management, user interface and visualisation. Although originally developed for High Energy Physics experiments at CERN, GEANT4 now finds application in many domains of experimental physics, thanks to its object oriented approach. In fact, the object oriented approach of GEANT4 allows for polymorphism so that each object of the simulation toolkit can be handled transparently by the GEANT4 kernel. This means, for instance, that the GEANT4 physics library can handle any physics process transparently, irrespective of its specific features. This powerful technological feature is at the ground of the wide set of models and physics processes available for GEANT4, which make this simulation toolkit one of the most ubiquitous in the experimental physics research.

C.1 Cross section validation

Although GEANT4 has been widely validated in the literature for X-rays [133, 134] and protons [135, 136] (the two sources of interest for this work) and in very thin absorbers [137], such as the sensitive volume of CMOS sensors, an in-house validation against
reference data has been performed in order to provide confidence in the subsequent results.

C.1.1 X-rays

The first validation exercise was to compare X-rays interaction cross sections of GEANT4 and NIST-XCOM [138]. Simulations have been performed with mono-energetic X-ray beams in the energy range of interest for the simulations of Chapter 5 (20-200 keV) and the attenuation of such beams measured in a 1 mm thick slab of Silicon, representing the detector thickness used in this work. X-ray attenuation was studied in terms of probability of interaction or quantum detection efficiency \( A_Q \)

\[
A_Q = \frac{N_{int}}{N_0} = 1 - e^{-\mu(E,Z)T}
\]  

(C.1)

where \( N_0 \) is the number of photons incident on the detector, \( N_{int} \) is the number of photons interacting transversing an absorber of thickness \( T \), \( \mu(E,Z) \) is called attenuation coefficient and is function of the photon energy and material atomic number. Results were scored by separating contributions to \( A_Q \) due to the different interaction processes, i.e. photo-electric and Compton interactions. Rayleigh scattering was not studied individually since this is not of interest for the simulations used in this thesis. The total quantum detection efficiency was also studied, accounting for all the possible interaction processes (i.e., including Rayleigh scattering and pair production, although the latter is below threshold at the energies considered). Figure C.1 shows a comparison in terms of quantum detection efficiency for the GEANT4 simulations with the same quantity, as calculated from the attenuation coefficients from the NIST database, for photo-electric effect (Figure C.1a), Compton scattering (Figure C.1b) and for the sum of all possible interactions (Figure C.1c).

The agreement between measured and tabulated values from Figure C.1 results good over the whole range of energies investigated (20-200 keV), proving that the simulation toolkit used in this study can reliably reproduce the most common physics interactions in the energy range of interest with sufficient accuracy.
Figure C.1: Quantum detection efficiency ($A_Q$), defined in Equation C.1, as measured from GEANT4 Monte Carlo simulations and from NIST attenuation coefficients for photo-electric interactions (a), Compton scattering (b) and for the sum of all processes (c), i.e. photo-electric effect, Compton and Rayleigh scattering and pair production.
C.1.2 Protons

A cross section validation, similar to that reported in the previous Section, was performed for proton beams, as this source of particles has been widely used in Chapter 6. Mono-energetic proton beams in the range 30-230 MeV, which is the range of interest for proton CT as discussed in Chapter 6, were simulated and the energy deposition across a 25 cm Silicon slab measured. The dose depth curves for these simulations are shown in Figure C.2a for a number of mono-energetic proton beams in Silicon. The Bragg peak can be seen shifting deeper in Silicon, as the range increases with energy. The range of these proton beams has been measured from the curves of Figure C.2a, by considering the depth at which the dose in in Silicon falls to 50% of the maximum dose (on the distal side of the Bragg peak). These values of range measured from simulations were then compared with the projected range as reported in the NIST-PSTAR database[139] and reported in Figure C.2b. The agreement between simulated and tabulated values of range show a good agreement over the whole energy range investigated, proving that the simulation toolkit used in this study can reliably reproduce the most common proton physics interactions in the energy range of interest for proton CT with sufficient accuracy.

C.2 Simulations of detector response

GEANT4 represents an important tool for simulation of the physics interaction of particle and radiation with matter. This tool allows a snapshot of the state of a particle at each time instant, while travelling through a specified geometry, and, particularly to score the energy deposition along the particle path. However, in most practical cases, simulations are performed to predict the response of a detector in particular experiments. For this reason the detector response to ionisation was simulated, by adding to the GEANT4 standard library new classes for charge transport, created by the author.

In order to reproduce the detector response to the interaction with a specified kind of particle, a charge transport model has been developed, consisting of the following steps:
C.2. Simulations of detector response

Figure C.2: a) Normalised depth dose curves for mono-energetic proton beams in the range 30-230 MeV in Silicon. b) Proton range as measured from simulations (red symbols) and as tabulated in the NIST PSTAR database (blue curve) as a function of proton energy.

1. The energy deposition in the sensitive volume of the detector, a volume where charge generated can be collected, is scored by using the standard GEANT4 libraries.

2. The energy deposited in the sensitive volume is then converted into number of $e^-/\hbar$ pairs, assuming an electron-hole pair creation energy of Silicon ($E_{eh} = 3.6$ eV/pair) [140].

3. Charge diffusion in the detector field-free sensitive region is accounted for, leading to the definition of an electronic cloud.

4. Collection of this electronic cloud is then performed by accounting for collection efficiency as function of the depth of interaction.

5. Charge sharing amongst adjacent pixels is reproduced and number of collected electrons are sampled across the detector pixel matrix.

6. Detector electronics is then included by using the measured detector conversion gain and noise.
Figure C.3: a) Width of the charge cloud due to diffusion plotted as a function of the depth of interaction for a 12 µm thick field-free volume. b) Charge collection efficiency across the DynAMITe CMOS sensor. c) Schematic representation of charge sharing process in the DynAMITe detector.

In the following Sections, each of these steps above will be analysed in detail and a comparison between simulated and measured detector response will be shown, to prove the validity of this model.

**C.2.1 Charge diffusion**

From GEANT4 standard libraries the energy deposition at a point in space \((x_0, y_0, z_0)\), where \(x, y\) defines the detector plane and \(z\) is the orthogonal coordinates to these two, or the beam direction, can be recorded. If the impinging particle has sufficient kinetic
energy, it may produce ionisation, generating charge carriers at a mean rate of 3.6 eV/electron, which corresponds to three times the energy band gap of silicon [140]. Electrons liberated from ionised Silicon atoms generate a spherical electron cloud with an initial radius \( \sigma_i \) given by:

\[
\sigma_i[\mu m] = k E[keV]^\alpha
\]  

(C.2)

with \( k=0.0062 \, \mu m/keV \), \( \alpha = 1.75 \) and \( E[keV] \) being the energy deposited at location \((x_0, y_0, z_0)\). Equation C.2 follows directly from the well know energy-range relation \( R = kE^\alpha \), where both constants depend on the material [141].

Such generated charge will then undergo diffusion in the CMOS field-free sensitive layer, until it eventually recombines or reaches the collection point. The generated charge cloud diffuses over \( 4\pi \), and it is possible to approximate the projection of this cloud on the collection point as a Gaussian [142]. In Janesick’s seminal paper [142], the width of the Gaussian distribution resulting from charge diffusion in a field-free volume \( (\sigma_{ff}) \) is described by the following Equation:

\[
\sigma_{ff} = \frac{z_{ff}}{2} \sqrt{1 - \left( \frac{z_a}{z_{ff}} \right)^2}
\]  

(C.3)

where \( z_{ff} \) is the thickness of the field-free region and \( z_a = z_0 - z_{ff} \) is the difference between the depth of the interaction \( z_0 \) and the thickness of the field-free region. Such relation, derived by simulations in [142], has been later verified by other authors [143, 144, 145] for field-free regions of CCDs. The dependence of \( \sigma_{ff} \) with the depth of interaction \( z_0 \) is shown in Figure C.3a for a 12\( \mu \)m thick field-free detector. The worst case in terms of charge diffusion is represented when energy deposition occurs at the bottom of the field-free region \( (z_0 = 0) \), leading to largest possible cloud \( (\sigma_{ff} = 6\mu m) \), whereas the best case scenario occurs for charge generated just below the collection area \( (z_0 = z_{ff}) \), where the charge does not suffer charge diffusion \( (\sigma_{ff} = 0) \).

The width of the final Gaussian distribution, \( \sigma_{total} \), is then the sum of the two contributions mentioned above, initial electron cloud generation and charge diffusion, and can be expressed by the following Equation:

\[
\sigma_{total} = \sqrt{\sigma_i^2 + \sigma_{ff}^2}.
\]  

(C.4)
Appendix C. GEANT4 simulations

C.2.2 Charge collection

Charge generated in the detector volume undergoes diffusion processes as explained in the previous Section. Using Equations C.2, C.3 and C.4, it is possible to build a quasi-continuous spatial distribution of the ionisation charge for each energy deposition event. However, not all of this charge can be collected by the pixel diodes. In fact, charge diffusing through the detector volume can undergo recombination processes.

The probability for a charge recombination or collection is dependent on the specific region of the detector volume where charge is generated. In order to understand differences among different detector regions, it is useful to recall some of the concept of CMOS sensors architecture discussed in Chapter 2.

In standard CMOS technology a thin lightly p-type Silicon epitaxial layer is grown on a heavily doped p-type substrate (p+ sub). Within the epitaxial layer, n+ wells (n+ well) structures are formed. The p-type epitaxial layer (p epi) represents the detector sensitive volume, while the n+ well/p epi diode junction acts as a charge collection element. The detector is only partially depleted across the n well/p epi junction (≃ 1 µm), so the charge is collected mainly through a thermal diffusion mechanism, as discussed in the previous Section. Also, because of the particular doping profile realised across the sensor (p epi/p+ sub), the junction between epitaxial layer and substrate represents a potential barrier, limiting the diffusion of charge generated in the epitaxial layer towards the substrate. Therefore, three different collection regions can be identified in a CMOS sensor:

1. A heavily doped n+ regions placed at the top of the detector volume, where the high doping concentration and the physical presence of p+ wells, for the realisation of the in-pixel transistors, reduce the carrier lifetime and collection is only partial;

2. A lightly doped epitaxial layer, where the low doping concentration lead to a charge carrier lifetime (τn), the time available for charge collection before recombination occurs, which is much smaller than the diode collection time (Tc). The collection efficiency in this region is almost 100%.
C.2. Simulations of detector response

3. A heavily doped p substrate, where the collection efficiency shows an exponential decay due to the short lifetime of charge carriers. Charge collection efficiency decreases for deeper generation points, as recombination is more likely to occur before this charge can reach the collection diode.

The exponential decay in charge collection efficiency for the substrate can be explained by using the extension of Ramos theorem as applied to induced charge in semiconductor detectors [146]. In fact the charge transport equation for the excess of minority carrier generated in the p+ substrate can be written as:

\[
\frac{\partial \Delta n}{\partial t} = D_n \frac{\partial^2 D_n}{\partial^2 z} - \frac{D_n}{\tau_n} \tag{C.5}
\]

where \( D_n \) is the electron diffusion coefficient and \( \tau_n \) is the charge carrier lifetime. By using appropriate boundary conditions for Equation C.5, the solution of this equation at the collection electrode, as a function of the charge generation point \( z_0 \), can be expressed as [147]¹:

\[
q(z_0) = N \exp \left( -\frac{z_0 - w_2}{L_n} \right) \tag{C.6}
\]

where \( q(z_0) \) is the charge reaching the collection electrode, \( N \) is the charge generated at \( t = 0 \) in \( z_0 \), \( w_2 \) represents the edge of the epitaxial layer, and \( L_n \) is called electron diffusion length (\( L_n = D_n \tau_n \)).

The charge collection profile used for the simulations described in this Appendix is shown in Figure C.3b. A charge collection efficiency of 80% is assumed for the n+\(_{\text{well}}\) region (\( \simeq 1\mu m \)), followed by a full collection efficiency (100%) in the epitaxial layer [148]. Charge collection in the substrate follows the exponential decay of Equation C.6, using a carrier lifetime \( \tau_n \) of \( 3.5 \times 10^{-8} \) s resulting from the doping concentration of the CMOS sensor used in this work. The charge collection profile of Figure C.3b is consistent with experimental measurements carried out on similar CMOS technologies [148].

¹As boundary conditions for Equation C.5, it is assumed that \( \Delta_n = 0 \) at the edge of the depletion region \( z = w_2 \), and at the back of the collection electrode \( (z = D) \). Also, it assumed that the all the charge entering the depletion layer is collected.
C.2.3 Charge sharing and digitalisation

Charge diffusion and collection in CMOS APSs have been described in the previous two Sections, allowing to build a quasi-continuous spatial distribution of the ionisation charge collected for each energy deposition event. A further step concerning charge sharing across adjacent pixels and signal digitalisation is needed to fully simulate the detector response.

As charge diffuses in the detector volume, the electron cloud width increases as discussed in Section C.2.1. This lateral spread, especially when energy deposition events are close to the pixel edge, can result in charge being collected by diodes of several neighbour pixels. This phenomenon, known as *charge sharing*, is one of the main limitations of pixellated detectors with relatively small pitch and, although design solutions exist to mitigate this effect (e.g. by using guard rings around the pixels), it can affect imaging and spectroscopic performance.

For the DynAMITe detector, the subject of this work, charge sharing is a significant effect. In fact, as discussed in Chapter 2 and schematically shown in Figure 2.4, the 50 $\mu$m pixels (Sub-Pixels) feature four diodes placed at the four corners of the pixel, so that diodes of adjacent pixels are placed close to each other. In order to model the charge sharing phenomenon in this detector, the quasi-continuous distributions of charge collected for each ionising event are sampled on an artificial matrix with pixels corresponding to half of the pixel size of the detector (25 $\mu$m). In this way each detector pixel (50 $\mu$m) is split in four artificial quadrants (25 $\mu$m), each of these containing one of the four in-pixel diodes. This process is schematically shown in Figure C.3c, where the actual detector matrix is represented by a continuous line, while the artificial sampling grid is shown as a dotted line for the central pixel. Each energy deposition event is then associated to a pixel quadrant, given the calculated spatial distribution of the charge cloud after diffusion and collection (the energy deposition event is represented by a red dot in Figure C.3c). Such charge is then equally split among the four adjacent diodes, around the relevant pixel quadrant (pink dot in Figure C.3c).
C.2. Simulations of detector response

The charge collected per ionising event in each pixel quadrant is then re-sampled over the actual detector matrix, i.e. summing up the charge collected for each of the four pixel diodes. The charge collected per pixel has then to be converted into detector Digital Number (DN), by accounting for digitalisation processes described in Figure 3.1 of Chapter 2, and noise has to be added. Signal conversion, from unit of e\(^-\) to unit of DN, can be performed by using the conversion gain distribution measured in Section 3.3.2, i.e. conversion gain for each detector pixel can be randomly extracted from log-normal distributions with mean and standard deviations reported in Table 3.2. Similarly, detector noise can be added to each detector pixel, randomly extracted from log-normal distributions with mean and standard deviations reported in Table 3.2.

C.2.4 Validation

The model for charge transport and detector response simulations, described in the previous Sections, has been validated against experimental measurements acquired with the DynAMITe detector. The experimental data-set consists of individual protons measurements acquired at the iThemba Lab cyclotron (see Section 6.3.1) with a degraded 30 mm range beam, corresponding to a mean energy of 60 MeV. Such a degraded beam was then further degraded by using PMMA attenuators of a thickness of 14, 16, 18 and 22 mm, resulting into a mean proton energy in the range 38-18 MeV. Monte Carlo simulations of the experimental set-up were performed using the standard GEANT4 libraries and, in addition to this, custom classes for charge transport using the aforementioned modelling framework, created by the author.

Detected energy spectra, resulting from signal generated in the detector by individual protons, are shown in Figure C.4 for the experimental measurements and simulations, for the five PMMA attenuators used. At each proton energy considered, the signal distributions is suggestive of a a spread-out Landau energy-loss curve [115], as expected for the energy-loss of charge particles in thin absorbers. Signal distributions, in units of DN, result largely comparable, in terms of width and peak position, for measurements and simulations, providing a qualitative verification of the agreement of the charge transport model with experiments.
Figure C.4: Measured (red curve) and simulated (blue curve) spectra for signal generated by individual protons in the DynAMITe detector, when exposed to a 60 MeV beam degrade by a thickness of 14 a), 16 b), 18 c), 22 d) mm of PMMA.
In order to provide a quantitative verification of the validity of the charge transport model, distributions of Figure C.4 have been fit with a Landau distribution [149], to calculated the most probable value (peak position), corresponding to the mean value of Bethe-Bloch energy loss. Figure C.5 shows the most probable signal, resulting from a fit of the Landau distributions of Figure C.4, for measurements (red symbols) and simulations (blue curve) as a function of the attenuator thickness, i.e. proton energy. The agreement is satisfactory in the whole range investigated, and, even for the data point at 20 mm (see Figure C.4d), where some discrepancies are observable in the spectra, measurements and simulations are in agreement within their errors.

The agreement between simulated and measured data show a good agreement over the whole energy range investigated, proving that the charge transport model used in this study can reliably reproduce the detector response in the energy range of interest for proton CT with sufficient accuracy.
Appendix D

List of publications

The following publications have resulted from the work documented in this thesis:

Journal publications


technolog”, *Physics in Medicine and Biology* 05/2014; 59(11):2569. DOI:10.1088/0031-9155/59/11/2569

**Conference Publications**


References


[41] EMVA. European machine vision association standard 1288 standard for characterization and presentation of specification data for image sensors and cameras, 2005.


[101] PTCOG Secretary. Particle therapy patient statistics (per end of 2013), June 2014.


References


[121] N. Barthe, K. Chatti, P. Coulon, S. Maitrejean, and B. Basse-Cathalinat. Recent technologic developments on high-resolution beta imaging systems for quantitative
References


