DISSERTATION
SUBMITTED TO THE
COMBINED FACULTIES FOR THE NATURAL SCIENCES AND FOR MATHEMATICS
OF THE RUPERTO-CAROLA UNIVERSITY OF HEIDELBERG, GERMANY
FOR THE DEGREE OF
DOCTOR OF NATURAL SCIENCES

Put forward by
M.Sc. Aerospace Eng. - Ilaria Sacco
Born in: Genova, Italy
Oral examination: 9th November, 2016

Development of highly integrated PET/MR detector modules

Referees: Prof. Dr. Peter Fischer
Prof. Dr. Joao Seco

ABSTRACT

In the last 20 years, PET imaging has become a very valuable technique for clinical diagnosis. It gives functional information on glucose metabolism and cancerous cell growth and, if coupled with precise anatomical information from MRT scans, can lead to very accurate results.

The aim of the present work is the design of highly performing PET detector modules, to be used for both clinical and pre-clinical applications and compatible with operations inside MR scanner. A single-layer ceramic module, hosting both sensors and readout electronics, efficiently thermally decoupled thanks to an integrated liquid cooling channel, has been designed, fabricated and tested. Novel single-ended front-end architecture, studied for low input impedance and good timing resolution, has been designed and integrated on PETA ASIC for SiPM readout. SiPM dies for maximum light collection and identification of sub-millimeter scintillating crystals have been designed and characterized. The assembly is very compact, with an area of $32\,mm \times 32\,mm$, about $8\,mm$ thick without crystals, with up to 144 readout channels. It has proved excellent ToF resolution of about $230\,ps$ CRT at FWHM for detectors in coincidence. The same module has been able to identify sub-millimeter scintillating crystals.

ZUSAMMENFASSUNG

Im Verlauf der letzten 20 Jahre hat sich die Bildgebung mittels PET zu einem sehr wertvollen Instrument der klinischen Diagnose entwickelt. Sie liefert funktionelle Informationen über den Glukose-Stoffwechsel und das Wachstum krebsartiger Zellen und kann, wenn sie mit präzisen anatomischen Informationen von MRT-Bildern verknüpft wird, zu sehr genauen Ergebnissen führen.

Das Ziel dieser Arbeit ist der Entwurf hochperformanter Detektormodule für PET, die sowohl in klinischen als auch in vor-klinischen Anwendungen eingesetzt werden können, und kompatibel mit dem Einsatz in MR-Scannern sind. Ein Keramik-Modul, das sowohl Sensoren als auch Ausleseelektronik enthält, die dank integrierter Kanäle für FlÃijssigkühlung effektiv thermisch entkoppelt sind, wurde entworfen, gefertigt und getestet. Eine neuartige single-ended Architektur für das Frontend, optimiert für eine niedrige Eingangsimpedanz und eine gute Zeitauflösung, wurde entworfen und in den PETA-ASIC für SiPM-Auslese integriert. SiPM-Chips mit höchstmöglicher Lichtempfindlichkeit und die Erkennung von Szintillatorkristallen kleiner einem Millimeter wurden entworfen und charakterisiert. Der Aufbau ist sehr kompakt, hat eine Fläche von $32\,mm \times 32\,mm$, ist ohne Kristalle ungefähr $8\,mm$ hoch, und arbeitet mit bis zu 144 Auslesekanälen. Die gezeigte Time-of-Flight Zeitauflösung ist ungefähr $230\,ps$ (FWHM) für zwei Detektoren in Koinzidenz. Das gleiche Modul konnte Szintillator-Kristalle kleiner einem Millimeter unterscheiden.

ACKNOWLEDGMENTS

> *"Choose a good mentor who asks original questions. Be patient, do not give up: work hard and preserve. Be passionate and excited about what you are doing: think of your scientific profession as if it was your hobby. Luck is important too, but remember, very often luck is not blind: it hits those who are ready."*
>
> - Aaron Ciechanover,
> Nobel Laureate in Chemistry 2004

First of all, my sincere thanks goes to my PhD advisor Peter Fischer, for having been the good mentor I was looking for. With his questions and answers, his help, the opportunities he gave me, the trust and his constant enthusiasm, he made my journey to this goal more exciting and interesting than I ever expected.

Thanks to Christian Kreidl and Michael Ritzert, and all my colleagues at the Computer Science Institute: I cannot even count how many times they helped me in so many different ways, with interesting discussions and always useful suggestions.

Thanks to my parents and Alessia, once more. They have been there, from far away, with always the same encouraging smiles and unbound trust. At last, thanks to Elisa: I will never be grateful enough for her patience and support.

CONTENTS

I INTRODUCTION AND BACKGROUND 1

1 INTRODUCTION 3

2 POSITRON EMISSION TOMOGRAPHY 5

- 2.1 Physical principle and medical purpose 5
- 2.2 Detector technology 7
  - 2.2.1 Scintillating crystals 8
  - 2.2.2 Optical sensors 10
  - 2.2.3 Read-out electronics 11
- 2.3 Challenges in detector design 12
  - 2.3.1 Physical limits 12
  - 2.3.2 Spatial resolution 13
  - 2.3.3 Timing resolution and Time-of-Flight 15
  - 2.3.4 Detection sensitivity 17
- 2.4 Scanner design 18
  - 2.4.1 Small-animal PET scanners 18
  - 2.4.2 Whole-body PET scanners 19
- 2.5 Data acquisition and corrections 19
  - 2.5.1 Attenuation correction 20
  - 2.5.2 Random and scatter correction 21
  - 2.5.3 Dead time correction 22
- 2.6 Combined scanners 22
  - 2.6.1 State of art: PET/CT scanners 23
  - 2.6.2 The challenge of PET/MR scanners 24

3 SIPM: A NEW PHOTON SENSOR FOR PET 29

- 3.1 The avalanche photo-diode 29
- 3.2 SiPM modeling and readout 31
- 3.3 Proprieties 33
  - 3.3.1 Gain and signal linearity 34
  - 3.3.2 Photon detection efficiency 35
  - 3.3.3 Noise sources 36

4 THE SUBLIMA PROJECT 39

- 4.1 The starting point: HYPERImage 40
- 4.2 Towards integrated PET/MR 42
  - 4.2.1 System analysis and simulation 42
  - 4.2.2 MR-transparent detector cubes: alternative designs 43
  - 4.2.3 Novel reconstruction algorithms 44

5 SURVEY ON OTHER PROJECTS 45

- 5.1 ASIC Designs for SiPM Readout 45
- 5.2 Position-sensitive SiPMs 47
- 5.3 Research projects on novel PET detectors 50

        5.3.1 Novel ToF-PET detector systems 50
        5.3.2 High-spatial resolution detectors 52
    5.4 Commercial Whole-Body ToF-PET/CT scanners 53
    5.5 Commercial Whole-Body PET/MR scanners 55

II DEVELOPMENTS 57
6 READOUT ELECTRONICS DESIGN 59
    6.1 Low input impedance single-ended front-end 59
        6.1.1 Input stage 60
        6.1.2 Discriminator logic 68
        6.1.3 Charge integrator 70
    6.2 ASIC design 74
        6.2.1 Channel architecture 74
        6.2.2 Phase-locked loop and TDC design 76
        6.2.3 Readout logic 79
        6.2.4 Chip versions 80
7 NOVEL PHOTON DETECTOR DESIGNS 83
    7.1 Rectangular SiPM for maximum light collection 83
    7.2 Interpolating SiPM 84
        7.2.1 ISiPM Topology 85
        7.2.2 Reconstruction Errors 87
        7.2.3 Prototype Design 89
    7.3 A Digital Photon Counter 91
        7.3.1 Pixel Architecture 91
        7.3.2 Trigger and Multiplicity Generation 93
        7.3.3 Control and Readout Architecture 95
8 DETECTOR MODULE 99
    8.1 Design and Features 99
    8.2 Internal liquid cooling system 101
    8.3 Assembly in clinical and pre-clinical configuration 103
9 TEST SETUPS 107
    9.1 Back-to-back coincidence setup 107
        9.1.1 FPGA readout architecture 108
    9.2 Modified setup for flexible SiPM readout 109

III RESULTS 111
10 TIMING RESOLUTION 113
    10.1 Single channels coincidence resolution 113
        10.1.1 Considerations on energy resolution 114
    10.2 Module coincidence resolution 116
    10.3 Effects of temperature on timing resolution 119
11 SMALL CRYSTAL IDENTIFICATION 121
    11.1 Light-guide-based Module 121
    11.2 ISiPM-based Module 123
    11.3 Event reconstruction with IDP 127
12 PERFORMANCE AT HIGH DATA RATE 129

IV CONCLUSION AND OUTLOOK 133
13 CONCLUSION 135
14 OUTLOOK 137

V APPENDIX 139
A PETA CHARACTERIZATION 141
B PERSONAL CONTRIBUTION 147
Bibliography 151

LIST OF FIGURES

Figure 2.1 Schematic PET ring 7
Figure 2.2 Positron decay and annihilation scheme 12
Figure 2.3 DOI effect on spatial resolution 14
Figure 2.4 Various coincidence error sources 15
Figure 2.5 Time of Flight effect on spatial resolution 17
Figure 3.1 APD cross section 29
Figure 3.2 SiPM array schematic and photo 30
Figure 3.3 SiPM Model 31
Figure 3.4 Effects of input and quench resistance on SiPM output voltage 32
Figure 3.5 Effects of the amplifier design parameters on output signal 34
Figure 3.6 SiPM signal saturation 35
Figure 3.7 PDE dependance on over voltage 36
Figure 3.8 Effects of temperature and bias voltage on the DCR and the crosstalk 37
Figure 4.1 Sublima PET/MR whole-body scanner 39
Figure 4.2 HYPERImage PET module 41
Figure 4.3 HYPERImage PET Insert 42
Figure 5.1 Position Encoded SiPMs 49
Figure 5.2 Linearly-Graded SiPMs 50
Figure 5.3 Hyperion pre-clinical scanner 52
Figure 5.4 DigiPET pre-clinical scanner 53
Figure 6.1 Channel simplified schematic 60
Figure 6.2 Low-impedance input stage schematic 63
Figure 6.3 Gain stage schematic 64
Figure 6.4 Input impedance dependance on the input stage bias currents 66
Figure 6.5 Input DAC schematic and output 67
Figure 6.6 Input impedance and DC level for different reference DAC settings 67
Figure 6.7 Discriminator logic circuit 68
Figure 6.8 Simulated bandpass frequency range at the discriminator input 69
Figure 6.9 Circuit for threshold definition 69
Figure 6.10 Circuit for charge integration 70
Figure 6.11 Effect of different waiting times on energy measurement 72
Figure 6.12 Integrator transient response on feedback current 73
Figure 6.13 PETA: Picture and block architecture 74

Figure 6.14 PETA Channel schematic 75
Figure 6.15 PLL block schematic 77
Figure 6.16 Coarse counter selection principle 79
Figure 6.17 Scheme of channel readout architecture 80
Figure 7.1 Picture of a rectangular SiPM die with 12 sensors 84
Figure 7.2 ISiPM cell assignment examples 86
Figure 7.3 Simulated systematic error for devices with different cell density and cluster size 87
Figure 7.4 Map of systematic errors in position reconstruction 88
Figure 7.5 Reconstruction error due to photon detection statistics and counting noise 89
Figure 7.6 ISiPM micro pictures 90
Figure 7.7 Rectangular ISiPM die 90
Figure 7.8 Pictures of IDP chip 92
Figure 7.9 IDP pixel architecture 92
Figure 7.10 Hit-Or generation in IDP 93
Figure 7.11 Multiplicity tree logic 94
Figure 7.12 Delay propagation on multiplicity tree 95
Figure 7.13 IDP1 Readout architecture 96
Figure 8.1 Comparison between Sublima and HYPERImage detector modules 99
Figure 8.2 Single layer detector views 100
Figure 8.3 Scheme of the LTCC cross-section 101
Figure 8.4 Infra-red images of the ceramic substrate cooled with internal water channel 102
Figure 8.5 Substrate temperature dependence on internal water temperature 104
Figure 8.6 Picture of a fully assembled module 104
Figure 8.7 Assembled clinical detection module 105
Figure 8.8 Picture of an ISiPM-based module 106
Figure 9.1 Back-to-back coincidence setup 108
Figure 9.2 Assembled clinical detection module 110
Figure 10.1 CRT resolution between single SiPMs 114
Figure 10.2 Energy spectra for single SiPM dies 115
Figure 10.3 CRT resolution between two channels on two modules 116
Figure 10.4 Average CRT between detectors 118
Figure 10.5 Internal cooling setup configuration 119
Figure 10.6 CRT dependance on temperature 120
Figure 11.1 Reconstructed flood-map with light-guide-based module 122
Figure 11.2 Reconstructed flood-map with an ISiPM-based module 124
Figure 11.3 Profile plot with ISiPM 125

Figure 11.4 Energy resolution per crystal with ISiPM 125
Figure 11.5 Effects of photon cluster size on ISiPM position reconstruction 126
Figure 11.6 DOI estimation with ISiPM 127
Figure 11.7 Array reconstruction with IDP1 128
Figure 11.8 Full event frames acquired with IPD1 128
Figure 12.1 Detected and expected rate per channel 130
Figure 12.2 SiPM current versus activity 131
Figure 12.3 Timing and energy resolution dependance on data rate 132
Figure A.1 Input resistance estimation circuit 141
Figure A.2 Measured input DC voltage 142
Figure A.3 Noise level per channel 142
Figure A.4 Characterization of the energy integrator 143
Figure A.5 Effects of discharge current on ADC 144
Figure A.6 Channel dead-time estimation 145
Figure A.7 Intrinsic CRT of PETA channels 146

LIST OF TABLES

Table 2.1 Proprieties of inorganic scintillators 9
Table 2.2 Proprieties of common photon sensors 11
Table 3.1 SiPM parameters used in simulation 33
Table 5.1 ASICs for SiPM readout 48
Table 5.2 Commercial ToF-PET/CT scanners 54
Table 5.3 Commercial PET/MR scanners features 55
Table 6.1 Input stage options in test chip ISIS1 61
Table 14.1 Features of PETA6 137

ACRONYMS

ADC	Analog to Digital Converter
APD	Avalanche Photo Diode
BGO	Bismuth Germanium Oxide
CP	Charge Pump
CRT	Coincidence Resolving Time
CT	Computed Tomography
DCR	Dark Count Rate
DOI	Depth of Interaction
dSiPM	Digital SiPM
ESR	Enhanced Specular Reflector
ETH	Eidgenössiche Technische Hochschule
FBK	Fondazione Bruno Kessler
FDG	Fludeoxy Glucose
FOV	Field of View
FPGA	Field Programmable Gate Array
GSO	Gadolinium Oxyorthosilicate
IDP	Interpolating Digital Photo-sensor
IMS	Institute for Microelectronics Circuits and Systems
ISiPM	Interpolating Silicon Photo-Multiplier
LaBr$_3$	Lanthanum(III) bromide
LFSR	Linear Feedback Shift Register
LOR	Lino of Response
LSB	Least Significant Bit
LSO	Lutetium Oxyorthosilicate
LTCC	Low Temperature Co-fired Ceramic
MAB	Module Adapter Board

MR	Magnetic Resonance
MSE	Micro System Engineering GmbH
NaI	Sodium Iodide
PCB	Printed Circuit Board
PDE	Photon Detection Efficiency
PDPC	Philips Digital Photon Counting
PET	Positron Emission Tomography
PETA	Position Energy Timing ASIC
PFD	Phase-Frequency Detector
PLL	Phase Locked Loop
PMT	Photo Multiplier Tube
RF	Radio Frequency
SAFIR	Small Animal Fast Insert for mRi
SAR-ADC	Successive Approximation Register ADC
SiPM	Silicon Photo-Multiplier
SPAD	Single Photon Avalanche Diode
SPECT	Single-Photon Emission Computed Tomography
SPU	Singles Processing Unit
SSEP	Sublima Stack Evaluation Platform
STiC	SiPM Timing Chip
Sublima	SUB-nanosecond Leverage in PET/MR ImAging
TDC	Time to Digital Converter
ToF	Time of Flight
ToT	Time over Threshold
VCO	Voltage-Controlled Oscillator
WLS	Wavelength Shifter

Part I

INTRODUCTION AND BACKGROUND

This first part is an overview of goals and applications of the present work. The motivations behind this research are reported in Chapter 1, while Chapter 2 describes the physics principles on which PET imaging is based and technologies commonly used in detector design. Due to their importance for novel and MR compatible PET systems, Chapter 3 is dedicated to Silicon Photo-Multipliers technology and to their proprieties. Details on Sublima project, aimed to demonstrate the feasibility of a PET/MR combined system (Chapter 4) and a brief overview on the state-of-art in this research field (Chapter 5) are also presented.

1

INTRODUCTION

Positron emission tomography (PET) is an extremely useful tool in oncology, cardiology and neurology diagnosis. It allows the detection of cancerous cells, the evaluation of vascular diseases and atherosclerosis, the study of Alzheimer's and other brain diseases. It can also be used to assess the efficiency of cancer treatments on patients. PET has proved great potential and is now a world-wide popular imaging technique.

The resulting image quality is affected by many factors, one of them being the detector intrinsic resolution. To give an idea of commercial scanner performance, spatial resolution achievable with presently available systems is in the order of $4 - 5\,mm$ at the scanner center, and degrades for structures placed at the edges of the field of view. In addition, it has to be noted that for acquiring enough data, the scanning procedure can take as long as half an hour, with a reduction of comfort for the patient. A long imaging sequence can create problems with non-cooperative patients or children, and even in standard situations any non volunteer movement during data acquisition could cause major artifacts in the final image if not compensated for correctly.

This work aims to optimize the intrinsic resolution of PET detectors, for both human and small-animal imaging applications. Moreover, since PET imaging does not offer clear identification of anatomic body structures, the proposed detector has been designed to be magnetic insensitive, allowing its use in an integrated architecture with magnetic resonance. Nowadays MRi can be considered the best available option for clear distinction of different tissues, with the additional advantage of not requiring any radioactive dose to the patient.

The challenge of designing an optimized detector has been addressed from many perspectives and all components have been taken into account, in order to propose a scalable, high-performing, dense and compact solution to be eventually integrated within MRi systems. Front-end electronics and photon sensors have been designed for fast readout and maximum light collection. Innovative solutions for outstanding spatial resolution have been tested as well, specifically indicated for small-animal applications, where structures to be identified are in the order of mm or less.

The designed detector module, adjustable with minimum changes for clinical and pre-clinical PET imaging, is efficiently cooled and requires minimal infrastructure to operate. It reaches excellent performance, in terms of both intrinsic spatial and timing resolution. It could be used to build a highly-efficient, ultra compact and MR compatible PET ring with time-of-flight capability.

This thesis is divided into four parts. Part I describes the physics principles and proprieties of PET imaging systems, with detailed description of technologies available today. A survey on proposals and results by other research projects for improving PET image resolution is reported as well. Part II describes the technologies developed for pushing the limits in gamma detection resolution: fast and precise front-end electronics, position-sensitive SiPMs and very compact assembly on a single-layer module. The results obtained with dedicated test setups, in different rate conditions and detector configurations, are described in Part III. Finally, the fourth part collects conclusions and comments on the achieved results and a discussion of future work. Candidate's detailed contribution to the presented work is reported in the Appendix B.

This work has been developed within the Sublima project [1].

---

[1] Sublima is a project supported by the European Union, with the FP7 frame, under the grand No.241711.

2

POSITRON EMISSION TOMOGRAPHY

The origin of nuclear medicine can be dated with the discovery of radioactivity, by Bequerel and Curie during the later part of the 19^{th} century. It has offered great opportunities to create images of the inner body and find diagnosis, without the need of any surgery.

Positron Emission Tomography (PET) is one of the most recent imaging techniques: it benefited from developments in tomographic reconstruction, directly related to the increase in computing power available nowadays. Since its first appearance during the 1980's, it has shown such good potential in medical applications to be now considered a world-wide standard technique [1].

2.1 PHYSICAL PRINCIPLE AND MEDICAL PURPOSE

PET relies on the detection of coincidence products coming from the annihilation of a positron with an electron and uses them for reconstructing the position of the source event. During a β^+ process, a proton decays into a neutron, releasing a positron and a neutrino, following the formula:

$$p^+ \rightarrow n + e^+ + \nu + E \tag{2.1}$$

where E is the energy released in the process. While neutrinos travel through matter without interacting, positrons are unstable in nature and travel for a certain path, depending on their kinetic energy, until they interact with one of the many electrons available. The annihilation of the two particles takes place, following the equation:

$$e^+e^- \rightarrow \gamma + \gamma \tag{2.2}$$

where the two gamma particles are high energy photons. Assuming ideally that the positron kinetic energy is zero when the annihilation takes place, it can be deduced that the two gammas travel in perfectly opposite directions, with a fixed energy that is equal to the positron and electron rest mass $E_0 = 511\,keV$.

PET detectors are placed around the point of interaction and must identify coincident impacting particles, with fixed energy. In an ideal model, the annihilation always happens along the line connecting the two triggered detectors: if a large number of lines of interaction are detected, the source position can be reconstructed with high level of precision.

Compared with other imaging systems, such as SPECT (Single Photon Emission Computed Tomography) and CT (Computed Tomography), the possibility to detect coincident particles is a great advantage. It prevents the introduction of a collimator in front of the detector, usually necessary for selecting only particles coming from a known direction. Collimators reduce data rate and, being heavy and dense materials within the field of view, cause a large fraction of scattered particles.

In medical applications, positron emitters must be injected into the patient's body. It is important that the emitter gets accumulated in cells of interest, in order to give information on their position within the body. The injected compound is named "radio-tracer": it is a molecule where an atom has been replaced with a radioactive positron emitter. Its chemical and biological proprieties depend on the specific target of the application.

Usually, the molecule used for PET imaging is glucose-based: cancerous cells, but also brain and heart tissues, consume glucose at higher rate when compared with other tissues or neighbor "healthy" cells. In such a way, once the radio-tracer has reached the target, the pharmaceutical gets accumulated into the structures of interest and the decay rate detection can start. The resulting image obtained with a PET system shows in fact the different uptake rates of the tracer within the imaged region.

Nowadays, the most common pharmaceutical available is fludeoxy glucose (FDG). FDG was first discovered by Al Wolf and his group at the Brookhaven National Laboratory, and injected in human volunteers in 1976. It is still the most efficacious and easy to use radio-tracer for PET in oncology, neurology and cardiology diagnosis [2, 3]. The glucose molecule is labeled with the positron emitter Fluorine-18 (^{18}F), which is currently the most common option. ^{18}F is a pure positron emitter and is so small not to interfere with the biological proprieties of the body cells. It has a half-life of about 110 minutes, which makes it suitable for clinical applications, where no residual radioactivity must remain in the body 20 hours after injection. It is a good tracer for cancerous cells and, due to the high activity normally detected, allows the identification of even very small lesions (down to a few mm in size). Still, the relatively short half-life introduces some complexity for its production in a cyclotron and especially its transport to the clinical structures, which must be carefully planned [4]. Another pure positron emitter sometimes used in PET imaging with specific compounds is Carbon-11, which has the main limitation of very short half-life, only about 20 minutes.

Depending on the radio-tracer, different tissues or cells and specific biologic processes can be targeted. PET can be used for studying glucose absorption, cell metabolism, blood flow or oxygen use. It does not give any information about the internal body structures, which are usually identified by their different physical characteristics. For this reason, it is defined as a functional imaging technique.

2.2 DETECTOR TECHNOLOGY

PET scanners are typically built in a ring configuration, so that coincidence events coming from the body in the field of view can be detected simultaneously. A schematic view of a gamma detector ring is shown in Fig.2.1. The figure shows also a portion of the volume within the ring diameter, considered the useful field of view (FOV), where the body must be placed in order to avoid large artifacts.

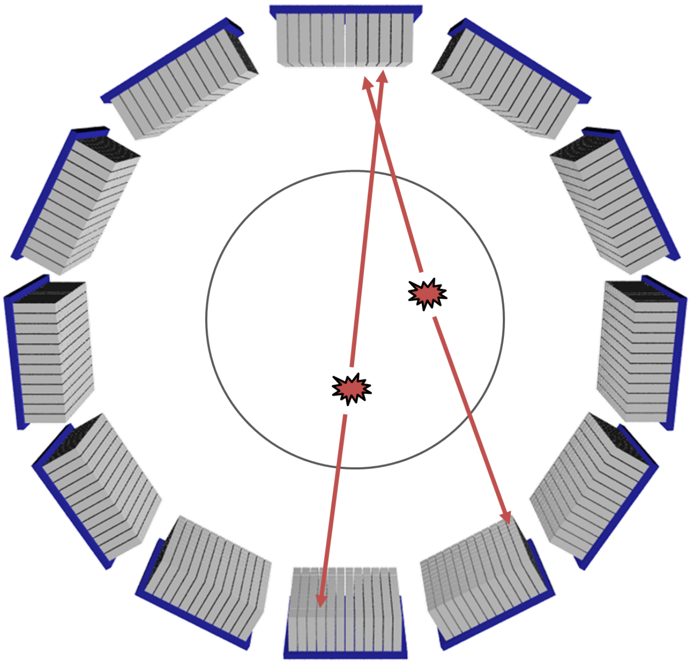


Figure 2.1.: Schematic PET scanner with one ring, showing two annihilation events and relative coincident gamma traveling in opposite directions. The field of view (black circle) is about half the size of the ring for whole-body scanners.

Block detectors are typically used in the ring design, each of them formed by an array of pixels. The line connecting two pixels which have triggered opposite gamma particles from the same annihilation event is called “line of response” (LOR). In the ideal case, the annihilation event happens at some point along the LOR and a large number of LORs allow the reconstruction of the source position.

Detector characteristics play a crucial role for the final image resolution. Each module is typically composed of scintillating crystals, which absorb gamma particles and produce visible light. The number of photons is proportional to the absorbed energy. Emitted light is detected by optical sensors and converted into an electrical signal, which can be read out and eventually stored by the front-end electronics. The full detection chain is described, taking into account different technologies, in the following sections.

2.2.1 *Scintillating crystals*

Scintillating inorganic materials are used in PET detectors for conversion of high energy gamma particles into optical photons. Due to their atomic structure, an electron in an atom can be excited from the valence band to the conduction band, when the absorbed energy is high enough. In such an unstable state, the electron travels around the reticule till it de-excites and recombines again via photon emission. This process is inefficient: the emission and absorption wavelength are the same, few photons are emitted and not in the visible range.

For this reason, inorganic materials are doped with impurities: they modify the atomic energy levels, allowing intermediate gaps. When an electron gets excited into the conduction band, the free hole ionizes an impurity atom. The electron is then trapped within it, leaving the impurity, also called the activator, in an unstable state. The de-excitation of the electron from such a level is fast (from tens to hundreds ns, depending on the material [5]) and causes emitted light. In this case, the scintillator emits a higher number of photons, at different wavelength and then not affected by self absorption and with lower energy, usually in the visible spectrum range.

The number of emitted photons is proportional to the absorbed energy and the absorption can take place through different physical processes. For gamma particles at $511\,keV$, the only possible interactions with the atoms are the photo-electric and the Compton scattering effects. Photo-electric effect is the mechanism from which an electron is excited at enough energy to break its atomic bounds and get released in the reticule. This is the most preferable interaction in gamma detection, because it does not foresee energy losses. Compton effect happens when a gamma particle and an electron interact via inelastic scattering. The results are a scattered electron and a gamma at lower energy deviated by a certain angle. This situation is generally undesired in detectors.

The number of photons produced by photo-electric effect is proportional to ρZ^4, where ρ is the scintillator density and Z its atomic number. On the other side, Compton scattering probability is also proportional to ρ. Compton scattered gammas could get absorbed again after the first interaction with an electron: in this case, the total absorbed energy is close to the original $511\,keV$, but due to the double process, leads to poor resolution. Alternatively, the two interactions could be detected separately, by two adjacent scintillators. In this case the detected energy is reasonably smaller than expected, not impacting the energy resolution but reducing the overall coincidence rate. For these reasons, the requirements for scintillating materials include a high atomic number, together with a reasonably high density, for good stopping power of incoming radiation. Moreover, photon emissions must be fast and with a large light yield, for good timing resolution and signal quality. Usually the refraction index of the scintillating material is also taken into account during detector design. Since the scintillator must be coupled with optical sensors using transparent glue, a good matching of the refraction indexes is desirable, in order to avoid internal total reflection and light losses. Finally, scintillators should be easy to fabricate, which is often a challenge and dependent on the targeted size.

Common materials used in PET detectors are listed in Table 2.1, together with their most important proprieties [1, 6].

	NaI	BGO	GSO:Ce	LSO:Ce	$LaBr_3$:Ce
Density $[g/cm^3]$	3.67	7.13	6.71	7.40	5.29
Atomic number	51	75	59	66	47
Decay time $[ns]$	230	300	30-60	35-45	18
Light output $[h\nu/keV]$	43	8.2	12.5	27	70
Peak emission $[nm]$	415	480	430	420	356
Refraction index	1.85	2.15	1.85	1.82	1.88
Attenuation $[cm^{-1}]$	0.34	0.92	0.62	0.87	0.47
Photo-electric effect [%]	17	40	25	32	13

Table 2.1.: Proprieties of most common inorganic scintillators used in PET applications.

Sodium iodide (NaI), usually Tallium doped, is amongst the very first materials used in medical applications, but has lost popularity, due to its low density and consequent poor stopping power. Bismuth germanium oxide (BGO) is a pure inorganic scintillator, which was very often used in PET applications in the past (for example, in [7, 8, 9]), but with a relative slow response, when compared with other materials (up to ten times slower) and low light yield. Cerium-doped gadolinium oxyorthosilicate (GSO:Ce) has acceptable proprieties, which strongly depend on the doping ratio. Stopping power is not optimal, when compared with other materials, and crystals are fragile, making the fabrication and assembly challenging. More recently, cerium-doped lutetium oxyorthosilicate (LSO:Ce) has gained popularity for its favorable performance. It has short decay time and good light yield. It is nowadays a well established material for medical imaging (for example [10, 11]) and can be produced with stable performance. With the addition of small quantities of yttrium, cheaper crystals with comparable proprieties can be grown (LYSO): they have similar density, with a sufficiently high atomic number useful for detection power and photo-electric interaction. They are an established cost effective alternative for large detector systems [12, 13]. Finally, recent studies are demonstrating the great potential of Lanthanum(III) bromide ($LaBr_3$): the shorter decay time (less than 20 ns) makes it a good candidate for applications where extremely accurate timing resolution is needed [14]. At the moment fabrication costs are still high, making prohibitive the construction of a full ring detector with such material.

2.2.2 *Optical sensors*

Photons emitted by scintillating crystals are converted into electrical signals by optical sensors. In PET history, different devices have been used in experimental and commercial systems.

Photon sensors must have a high detection efficiency, with good match to the wavelength range emitted by the scintillators. High gain and fast response improve the signal quality and timing performance. Generally also mechanical considerations should be taken into account: complexity of the final system depends on the size and on the operation point of every component (cooling strategy, power supply, assembly). Recently, with the ongoing research on combined PET/MR scanners (see Section 2.6.2), insensitivity of the sensor to magnetic fields is also considered a mandatory requirement.

The first photon sensors used in PET scanners, very popular since the sixties, are photomultiplier tubes (PMT) [15]. PMTs are composed of a vacuum tube, with cathode and anode at opposite surfaces. When a photon impacts at the cathode, electrons are produced via photoelectric effect. Electrons travel from a dynode (a plate used as an electrode) to the next one, biased at higher voltage, and get multiplied at every stage. The main advantages of PMTs are their large amplification gain, in the order of $10^6 - 10^7$ electrons produced per detected photon (dependent on the number of dynodes) and relatively low cost, which make them a good option for large systems. On the other side, they are sensitive to magnetic fields, which deviate the electrons from their path along the tube. Additionally, their size is normally pretty large, compared to other devices (up to some cm long in some cases), adding complexity to system assembly. PMTs have been used in many experimental scanners at high resolution [7, 16, 17]

Starting in the nineties, Avalanche Photo-Diodes (APDs) were introduced on the scenes. APDs are formed by doped p-n junction, typically in silicon, where, due to the opposite carriers charge, a neutral depletion region appears. The width of this region can be enlarged with the application of an external reverse voltage, in the order of few hundreds of volts (details on APD junctions can be found in Section 3.1). When a photon gets absorbed it creates an electron-hole pair through photo-electric effect, which moves under the effect of the internal electric field, generating an avalanche of other electron-hole pairs, resulting in an electrical signal. APDs have very good sensitivity and are able to detect a large fraction of the total number of photons impacting the surface (up to 85% of detection efficiency in best cases). Their gain is low, around $500-1000$ pairs per photon, so often additional amplifiers are necessary. APDs are typically employed in small-animal PET scanners, where timing constraints are not strict and where compactness is preferred [18, 19]. Performance fluctuations depending on temperature force the use of an efficient cooling system for precise applications.

Finally, the latest photon sensors considered, available since the nineties, are the Silicon Photo-Multipliers (SiPMs). SiPMs are composed as arrays of APDs built on a common substrate, working in Geiger mode, with a typical active area of some mm^2. The diodes work in parallel, with good sensitivity and fast response. As with the APDs, they are insensitive to magnetic fields, but also have very good timing performance and low noise. SiPM design, characteristics and proprieties are more extensively described in Chapter 3.

Proprieties of the described photon sensors are summarized in Table 2.2.

	PMT	APD	SiPM
Gain $[e/\gamma]$	$\approx 10^6-10^7$	$100-200$	$\approx 10^6$
Quantum efficiency [%]	≤ 35	≤ 80	$20-50$
Power voltage $[V]$	500-3000	100-500	20-40
Threshold sensitivity $[ph.e]$	1	≈ 10	1
Compactness	no	yes	yes
Magnetic compliance	no	yes	yes

Table 2.2.: Proprieties of most common photon sensors used in PET applications..

2.2.3 *Read-out electronics*

Front-end electronics is the last step of a detection module: it records the signal arrival time, its energy and its position within the array. Depending on the photon sensor and also on the scintillator proprieties, characteristics and requirements for the read-out architecture can be different.

As a general consideration, resolution can be optimized with good signal-to-noise ratio, so that the electronic contribution to possible fluctuations in the signal can be neglected. Typically, thermal and intrinsic noise within scintillators (such emission fluctuations or fluorescence activity) and within optical sensors (dark count noise or optical crosstalk) are in fact the most limiting factors for system performance. Other important requirements are fast response and data processing, for reducing dead time and improving sensitivity. Good energy estimation helps in distinguishing annihilation events from other events at similar energy which could have happened within the field of view. A brief description of the read-out front-end requirements is provided only for a SiPM-based detector, in Chapter.3.

2.3 CHALLENGES IN DETECTOR DESIGN

The design and the performance of the detection module obviously affect its image reconstruction capability, in terms of spatial resolution of lesions and image contrast. Many characteristics contribute to additional errors in the image: some of them can be optimized for better performance, while some others depend on physical processes related to the PET detection principle and determine the achievable intrinsic limit.

2.3.1 *Physical limits*

Image resolution depends on the ability of the system to identify the exact position of the radio-tracer within the body and its decay rate. Such measurements are limited by physical errors, which are difficult to prevent and impossible to measure.

First of all, the statistical behavior of the source emitter introduces uncertainty in the results. Imaging results depend on the estimated radio-tracer uptake rate in different cells and the fluctuations in the number of decays could lead to wrong results. The effective dose injected into the patient, which depends on the radio-tracer specific activity but also on eventual contaminants in the compound, must be taken into account. Half-life of the radioactive component must be reasonable long, so that the pharmaceutical can be produced, injected into the patient and imaged once accumulated in targeted cells, but short enough not to be detectable in the patient after more than a day.

In addition to the limits of the radio-tracer precision, the annihilation process generating the signals is also a source of error for the final resolution. Two factors introduce not-measurable errors, as schematically drawn in Fig.2.2. First of all, the positron emitted by the nucleus travels for a certain distance Δd, before releasing enough kinetic energy to be able to interact with an electron. This distance depends on positron energy, which is a characteristic for each radionuclide, with statistical fluctuations. In medical applications, common radionuclides emit at energies in the order of 0.5 to 5 MeV, for a path of 0.1 to 2 cm. Such values are estimated for straight paths, which is usually

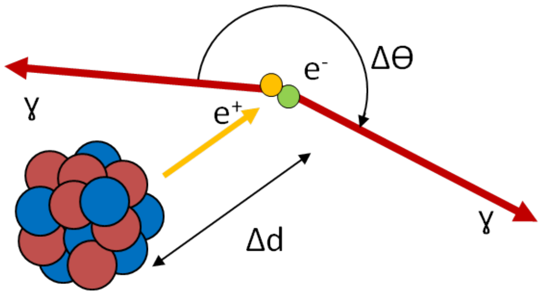


Figure 2.2.: Positron decay and annihilation scheme.

not the case, since positrons interact and get deviated by other particles. The interesting value is the perpendicular distance of the original molecule from the line of annihilation: its value is much smaller than the positron average path, but still in the order of $0.5 - 3\,mm\,rms$ for $^{18}F - FDG$, causing a blurring effect in the final image for which there is not correction.

Additionally, the emitted gamma rays are never exactly opposite. The positron does not annihilate when its energy is zero, so that the small amount of residual momentum is conserved in a noncolinearity effect of the products trajectories. The results is, also in this case, a blurring shade in the final image of few mm.

These errors originate from physical processes within the body (decay statistics, interactions of positron with electrons) and are not measurable and not correctable for. Further study on the radio tracers could lead to some improvements but nevertheless they represent an intrinsic limitation for the image quality.

2.3.2 *Spatial resolution*

Detector spatial resolution is defined as the minimum distance at which two different events are still recognizable. Since, in the present case, we are taking into account only detectors with pixelated scintillators, the spatial resolution depends on the size of single crystals.

Together with the pixel size, crystal shape and ring configuration determine the volume included between two scintillating detectors. Requirements in detector design are clearly different for humans or small-animals scanners. In case of pre-clinical PET, where ring diameter is in the order of $20\,mm$ or less and lesions to be detected are a few mm wide, pixel size must be as small as possible, typically $1\,mm$ or possibly less. For human imaging applications, where ring radius is about $40\,cm$, additional scattering in the body and parallax effects have a large impact on the image resolution, so that a small pixel size contributes to the complexity of the system without offering a real benefit to the final result.

The uncertainty on the absorption position along the crystal thickness can degrade the final image resolution, especially when long elements are used. Such measurement is called depth of interaction (DOI). Effects of DOI uncertainty, simplified for a bi-dimensional ring, are shown in Fig.2.3. Scintillators have lateral size d and thickness h. In the ideal case, gamma absorption always happens at the crystal top surface for both detectors and LOR is as wide as the scintillators d. In real situations, gamma particle penetrates few mm within the crystal before getting absorbed. The absorption position is statistically represented by the attenuation coefficient of the scintillator, which is a measure of its stopping power and is defined as the percentage of gamma particles absorbed when traveling along one centimeter of material. The

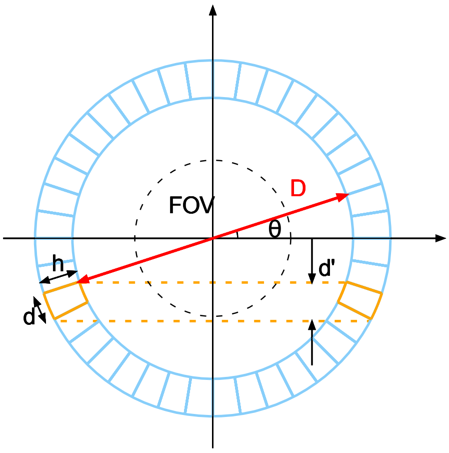


Figure 2.3.: Depth of interaction effect on spatial resolution: for a pair of coincidence detectors placed at the borders of the field of view, the width for possible annihilation is larger (d') and depends on the crystal thickness, on the detectors relative position and on the ring diameter.

actual volume containing the annihilation position is then larger than expected (orange dashed volume in the figure). Most severe effects are related to coincidences detected at the edges of the FOV.

In Fig.2.3, two detectors are taken into account: they are placed within the ring in such a way to detect coincidences originated at the borders of the targeted FOV, representing the worst possible case. In these conditions, the width of the volume of interest is expressed as:

$$d' = d \cos\theta + h \sin\theta \tag{2.3}$$

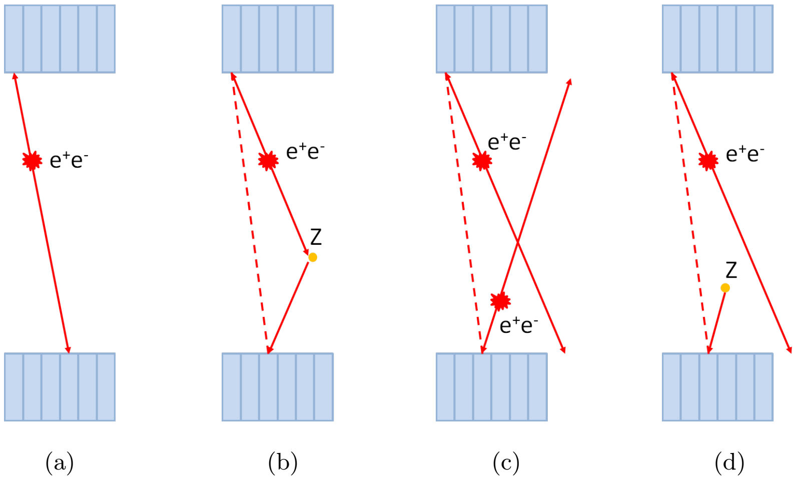
When the FOV is small, compared to the diameter of the system (D in the figure), the angle θ is also small and the difference between d' and d is less important. Such simple geometrical considerations are important during scanner design and justify why often PET rings are much larger than the size required to accommodate a patient (together with considerations on patient comfort). As an example, for a human-body scanner with a radius of $40\,cm$ and detector crystal size of $4\,mm \times 4\,mm \times 20\,mm$, DOI effect degrades the spatial resolution down to $40\,\%$, for a source placed only $10\,cm$ away from the center of the field of view [20].

Many strategies have been investigated for improving DOI resolution. One option is double-side light readout, on both scintillator faces: the absorption position can be estimated by geometrical consideration [21] or, with more recent and precise technologies, using timing measurements at the two sides [22, 23]. Another solution is controlling light sharing between crystals along their thickness, in order to measure the absorption point using different signal amplitudes from neighbor scintillators [24]. Probably the simplest alternative uses two or more layers of crystals. They are assembled opportunely shifted in the lateral plane or optically decoupled to produce a shifted signal at the photon sensor surface [25]. More sophisticated designs take advantage of the different characteristics of two scintillating materials, also assembled in two layers, but recognizable by their different time response using single-side readout [26, 27].

2.3.3 *Timing resolution and Time-of-Flight*

Since the precision of the detector components cannot be infinite, a certain time window where events are accepted as simultaneous must be defined during data acquisition. The coincidence window finite width allows the detection of false coincidences. If the fraction of false events is too high, the final image resolution could be badly degraded.

Different physical effects produce false coincidences, the most important being Compton scattering. Its effects are schematically represented in Fig.2.4, together with other less common events which could cause false coincidence detections.



(a) (b) (c) (d)

Figure 2.4.: Various sources of coincident events in opposite detectors: (a) true coincidence from annihilation event, (b) false coincidence detection from a scattered gamma ray after interaction with an atom Z: the reconstructed event is placed along the dashed LOR, (c) coincidence products from two different annihilation events within the coincidence window, (d) false coincidence detection, where one or both signals are actually coming from a random non annihilation event at atom Z within the FOV. Figure from [1].

The fraction of scattered gammas can be reduced with good energy resolution detectors. Still, gammas that release insufficient energy during inelastic impact with other particles, and are then deviated by small angles, are included in the $511\,keV$ peak and in the reconstruction process. For large FOV, Compton scattering is an important source of errors. Moreover, within the body there are large volumes of dense material which interact with gamma rays and deviate their trajectories.

Random coincidences are originated by two annihilation events, happening simultaneously: if only one product per event is detected, the reconstructed LOR does not correspond to any actual decay position. Such a situation is more probable for high dose scanning techniques. In a similar way, false coincidences can be detected if a non annihilation decay event happen within the field of view, together with a partially detected annihilation. Even in this case, the missing detection of one of the annihilation products allows the reconstruction of a false LOR.

An obvious way to limit false coincidence detection is to reduce at minimum the coincidence time window. This solution has an effect on random coincidence detection, however Compton scattered events, which appears almost simultaneously, would still be included. Clearly, a time window must be defined as a compromise between reduction of uninteresting events and sensitivity to good coincidences. It is usually in the order of $4 - 5\,ns$ for commercial human-body scanners.

In the case where the arrival time can be measured with enough precision, the source position along the LOR could be estimated. Such measurement, known as Time-of-Flight (ToF), allows higher signal-to-noise ratio and ideally a smaller number of events for identifying a lesion with the same precision of a standard system. The source position, with respect to the ring center, is given by:

$$\Delta D = \frac{\Delta t \cdot c}{2} \tag{2.4}$$

where c is the speed of light and Δt is the travel time difference measured by the opposite detectors. It has to be noted that for measuring a distance of few mm, the detectors must be able to detect a time difference in the order of only some tens of ps.

Fig.2.5 illustrates the concept of TOF measurements [20]. When reconstructing the annihilation position, LOR is usually divided into volume units, named voxels. Each voxel within the FOV is assigned to a certain probability that an event occurred within it. When no ToF information is available, the probability distribution is uniform along each LOR (Fig.2.5(a)) and all the voxels contribute to the reconstruction procedure. A large number of LORs, coming from the same source and with different orientations, must be detected to have a sufficient signal-to-noise ratio and good image resolution. If TOF precision can be reached, the source position along the LOR is instead estimated with a Gaussian-shaped probability function: the voxels far from the events do not contribute and a larger signal-to-noise ratio is possible. The Gaussian function width is determined by the coincidence resolving time (CRT) of the instruments, which is defined as the minimum time distance at which the system is able to identify different events.

ToF-PET scanners are presently becoming the target for clinical applications, thanks to the great potential they offer in improving image quality. Systems with CRT of about $600\,ps$ at FWHM are already commercialized (see Section 5.5): in such a case the Gaussian distribution

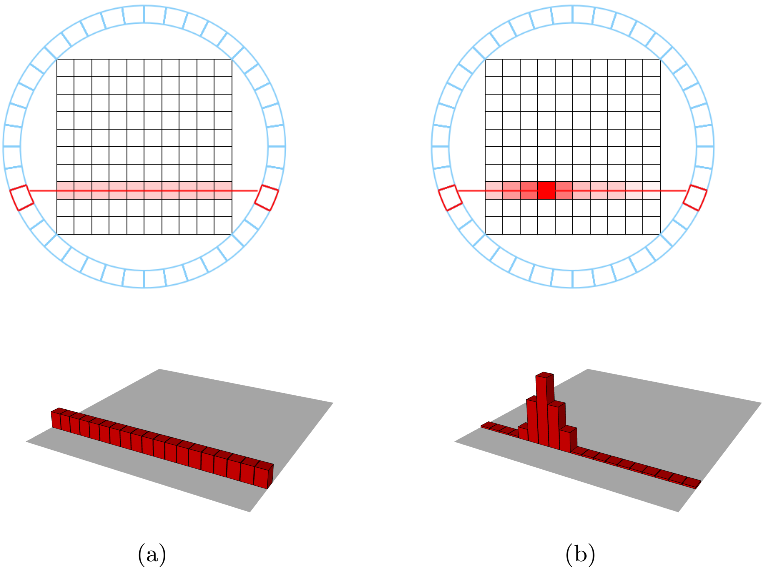


Figure 2.5.: Effect of Time-of-Flight resolution on the source position resolution: (a) if no ToF is available, the annihilation position probability is distributed uniformly for all the voxels along the LOR; (b) with ToF, the probability distribution is a Gaussian curve along the LOR, its width depending on the timing resolution of the system. The probability functions are shown also in 3D plots below.

probability along the LOR would be about $9\,cm$ wide. These values are already good for human applications, even if not optimal. Many research groups are presently targeting a coincidence detection precision within a range of only a few cm, corresponding to about $200\,ps$ CRT at FHWM between detectors.

2.3.4 *Detection sensitivity*

Being able to detect the largest possible number of coincidences is crucial for the final signal-to-noise ratio of the reconstructed image, but must be compromised with the limited amount of radioactivity that can be injected into the patient and the overall system costs and complexity.

Factors determining the overall sensitivity are both related to single detector module design and to the whole scanner geometry. Detection efficiency of each component is clearly very important, such as scintillators stopping power, sensor efficiency and their optical coupling. Their geometry and arrangement in a ring determines gaps and areas insensitive to gamma particles. Solid angle is also an important factor in scanner design: it is defined as the angle projected by a source onto the PET system and it should be as large as possible, to improve the overall sensitivity. It depends on the ring diameter, on its axial length and on the distance between the source and the detectors. A large diameter has the effect of reducing the distance between the source and the detector and then reduce the solid angle and the geometrical efficiency. Still, a ring that is too small is not comfortable for the patient and also introduces more severe parallax problems coming from DOI uncertainty. A long axial dimension improves the efficiency, but can lead to prohibitive costs. Usually, the detector ring is some tens cm long in the axial direction. Finally, the dead time of each component is a limitation in the maximum detectable counts: it comes from both sensors and readout electronics. During the recharging phase of a sensor or the readout and data processing, for example, components are "blind" to new events.

In order to give an idea of the system complexity, it must consider that the sensitivity of a 3-D PET scanner is in the order of $2 - 10\%$, already good if compared with the respective SPECT result which is only $0.01 - 0.03\%$, due to the presence of the collimator in the FOV.

2.4 SCANNER DESIGN

In clinical practice, a whole-body ring is used for complete scanning of patients, while more specialized and higher resolution machines are available on the market for specific body regions, such as for breast and brain imaging. New radio-pharmaceuticals, diagnosis techniques and reconstruction procedures are normally tested on dedicated small-animals scanners, which have different requirements and constraints. The most common design and constraints for whole-body and small-animal PET scanners are briefly described here.

2.4.1 *Small-animal PET scanners*

Pre-clinical scanners target mice, rat and rabbit imaging. Small rings necessary for imaging them allows the use of few detector modules, with a large reduction of cost and complexity. Additionally, some small-animal scanners can fit inside already existing machines, such as MR, and offer a good opportunity to test detector compatibility with other technologies.

These rings typically have a diameter varying from 12 to 20 mm. The most challenging requirement is the spatial resolution, which must be good enough to identify small lesions in the animal body. It is usually achieved with small pixelated scintillators: in common designs their lateral size is the order of 1 to 2 mm. Crystals thickness is typically between 10 to 15 mm, for reducing DOI effects. With these design parameters, blurring effects due to noncolinearity and DOI uncertainty have a limited impact on the final image, in the order of few tenths of mm [20]. Even if using small crystals, the sensitivity of the overall system is reasonably high, thanks to good geometrical efficiency achievable for such small animals, with a limited number of detector modules.

For the concerns of timing resolution, requirements are not too strict. ToF measurements do not introduce a great benefit: for mice dimensions, spatial resolution of $3\,cm$ along the LOR for a CRT at FHWM between detectors of about $200\,ps$ (which is already very challenging to achieve) is not helpful. Furthermore, a detector with small pixelated crystals is more affected by light losses at the boundaries, with a consequent degradation of timing performance. For these reasons, in pre-clinical applications detected coincidence pairs with CRT of about $1\,ns$ at FHWM are considered sufficient. Spatial resolution at the FOV center typically reaches values around 1 to $2\,mm$.

2.4.2 *Whole-body PET scanners*

Clinical scanners are more challenging to design, first of all for the simple fact that they need to be large enough to image a person. A large number of detector modules are necessary for obtaining a decent geometrical sensitivity, with consequent high costs. Volume inside the scanner must be large enough to fit most of the possible patients and the scanning procedure should be short and smooth, to avoid involuntary movements and related artifacts. Since axial dimension of the ring is usually not sufficient for a complete scanning of the body, the patient's bed moves within the ring, acquiring data in different positions and recombining them later during offline data processing.

The ring diameter is typicall about $80\,cm$, while the interesting FOV is in the order of $40\,cm$. ToF resolution allows improvement of the final image quality and is nowadays the target of most research groups (see Chapter 5 for details). Scintillators pixel size is kept on the order of few mm ($4\,mm \times 4\,mm$ commonly). Very small pixels require a large number of readout channels, with the related complexity. Crystals can be as long as to $25\,mm$, with DOI related problems. Clinical scanners have a spatial resolution of about $4\,mm$ at the FOV center, which degrades up to 50% moving at the FOV borders. The coincidence time window is kept to a few ns.

2.5 DATA ACQUISITION AND CORRECTIONS

Coincidences are detected and recorded by the system in the defined time window. The acquisition software looks for coincidences only between planar detectors, in a 2-D acquisition, or between all detectors in the ring, for a 3-D readout.

In order to acquire purely 2-D data, septa are placed between detectors so that undesired particles get absorbed and only coplanar rays are detected. This approach reduces random and single rate per detector, so that the fraction of wrong coincidences and the dead time of the sensor are lower, but has large impact on the overall instrument sensitivity. Almost one order of magnitude in sensitivity can be gained with a 3-D acquisition mode. It is more sensitive to scattered and random particles, for which it is necessary to apply appropriate correction algorithms, described briefly in the following paragraphs.

Source position is reconstructed with iterative models or with Fourier-based algorithms [28]. It is then clear that complexity and relative processing time for a 3-D volume is a drawback of using a full acquisition mode, but it is small if compared with the gained sensitivity and the shortest scanning time required.

Nowadays, almost all whole-body scanners available for clinical applications work in 3-D acquisition mode, as well as all the small-animals and brain imaging systems.

2.5.1 *Attenuation correction*

Particles originating within the body are attenuated by the body itself. Attenuation is the most important effect in PET and causes large signal losses, between 50 to 95% of emitted particles, high for large patients.

The probability that a particle reaches the detector is inversely proportional to the path it has to travel within the body and it also depends on the tissue it travels through. Taking into account a bi-dimensional scanner and two detectors, the length of the body intersecting the LOR is named L. The positron source is placed at a distance x from one of the two edges, along the LOR. In such conditions, and since both coincident gamma particles must be detected, the probability to record an event can be written as:

$$P_{event} = e^{-\mu x} \cdot e^{-\mu(L-x)} = e^{-\mu L} \tag{2.5}$$

where μ is the linear attenuation coefficient (assumed constant in the formula), which depends on the tissue characteristics along the path. It is about $0.095\,cm^{-1}$ for soft tissues (brain, muscles, etc.), $0.12 - 0.0.14\,cm^{-1}$ for bones and $0.03 - 0.04\,cm^{-1}$ for lungs, which are very porous and filled with air [20].

Attenuation correction is necessary to avoid major artifacts in the final image and incorrect uptake estimation from certain regions. More specifically, effects on the final image can be distorted structures, due to the different behavior of gammas for different directions and fake high activity coming from boundary regions or from areas with low attenuation correction factors.

Correction algorithms take into account attenuation effects of different tissues within the body and “recount” the uptake activity for the whole body. Ideally, they add missing counts to highly attenuated areas and subtract counts from areas where attenuation is minimal, such as at skin and border regions. Usually, both the images, attenuated and not-attenuated, are then used in clinical practice, because a comparison is useful for identifying missing lesions or confirming the diagnosis.

Few techniques are common in clinical applications for the estimation of attenuation correction factors to be used during data analysis. Transmission scans through the patient body are used for this purpose. A source is placed along the LOR connected to each pair of detectors: data is acquired without and with the patient, so that the ratio between counted events can be used as an estimation. The scanning procedure of the empty ring must be performed every day, while the transmission scan of the patient must be taken before injection of the radio-pharmaceutical. It is also very important that the patient does not move between scans, to avoid large artifacts in the final image. The most common source used for transmission scan is ^{68}Ge, which generates the positron-emitter ^{68}Ga by electron capture.

Alternatively, a transmission scan can be performed immediately after the PET procedure: in this way, there are fewer chances that the patient moves on the bed, since the measurement is much faster, without waiting periods between scans. On the other hand, it requires the ability to distinguish between transmitted and emitted data.

The last strategy is to acquire transmitted and emitted data simultaneosly in one scan. From a certain point of view, this is surely the optimal solution: it minimizes problems related to patient movements and reduces scanning time. There is still the challenge of distinguishing PET data from the transmitted particles, which in this case also contribute to additional scattered and random events.

Since nowadays almost all PET scanners used in clinical applications are combined with CT apparatus, additional transmission scans are often not necessary. CT-based attenuation correction factors are extracted and applied to PET results (see Paragraph 2.6.1 for details).

2.5.2 *Random and scatter correction*

False counts must be subtracted from the final result, to reduce noise and improve image resolution. Random coincidences are not related to any particular position or region and add uniform background to the image, degrading the contrast. Two methods are normally used for the estimation of random coincidence rates. The first one is a so-called "delayed coincidence window" and can be performed online: coincidences between two detectors are counted on a time window delayed enough so that true and scattered coincidences cannot be taken into account. The remaining rate of coincidences correspond to random events.

Alternatively, random coincidences rate can be estimated starting from single random rate per detector and taking into account the coincidence time window. Such technique requires that each detector continuously monitors its own single detection rate.

Scattered events can have a very large impact on the final resolution, especially for 3-D scans of the abdomen. In some cases, fractions of scattered events can be as high as 70% of the total. Clearly, for being affected by Compton scattering is enough that only one of the two gamma particles gets scattered along its path. The detector usually does not distinguish between particles scattered in the body and those scattered within the scintillating crystals. Two approaches are usually followed in this case.

The first one takes into account an emitted and a transmitted scan separately. It is assumed that all attenuation seen with the transmission scan at $511\,keV$ is actually due to Compton scattering. With geometrical considerations and computer modeled physics, scatter distribution is estimated for each detector and subtracted from emission data. A different way to estimate scattered events is to consider only the portion of the FOV outside the patient body. Once random events have been subtracted, the only source originated in that volume is the mispositioned event due to scatter. The behavior inside the body can be deduced by extrapolating the Compton profiles with mathematical models and subtracting them from the emission data, in order to obtain the final uptake rate. This solution is applicable only if the whole FOV is not occupied by the subject, so that a large volume can be analyzed for the estimation of Compton profiles. If the available volume is too small, the reconstruction algorithm can produce considerable artifacts in the image.

2.5.3 *Dead time correction*

Dead time depends on the single data rate of each detector, which is typically much larger than coincidence rate. If not properly corrected for, it can lead to underestimation of the activity in high dose regions. The correction algorithm uses models for detector pairs or for the whole system. Common models follow paralyzable and non-paralyzable dead time assumptions. Non-paralyzable models are used when, during dead time, the detector is completely blind to any incoming event. Alternatively, if an event occurs and dead time increases, the model is assumed to be paralyzable.

Dead time correction is more important for applications at high dose, where single count rates becomes an issue for the detector.

2.6 COMBINED SCANNERS

PET imaging systems give information on specific cellular functions, but are not able to clearly distinguish between different types of tissues. In clinical practice, having combined information of the body's structures and on specific biological functions in the same image is of great help for the diagnosis.

Combined PET/CT systems are nowadays common in clinical applications. A new research field, which is gaining popularity for its potential, is the integration of PET scanner within MR imaging systems: better contrast, especially in soft tissues, obtained with MRi has to be paid for with a higher complexity of the integrated system.

2.6.1 *State of art: PET/CT scanners*

PET/CT scanners were developed starting in the 1990's and can be nowadays considered as a standard imaging tool [29]. Computerized Tomography (CT) is a diagnosis technique which measures the fraction of high energy X-rays transmitted through the body from different angles. In most common scanners, a fan beam rotates around the patient and moves along the axial direction. On the opposite side of the field of view a matrix of detectors registers the transmitted radiation. With moveable beam and detectors, the system is able to scan a large 3-D field of view, giving volumetric images of the body. It gives excellent anatomical information, with good contrast amongst different types of tissues, in relatively short scanning times. The most important drawback is that patients are exposed to a high radiation dose, of about $20\,mSv$[1], depending on clinical requirements and on the local standard regulations [30].

Combination of PET and CT has been so successful that nowadays almost all PET scanners available for clinical procedures are working in tandem configuration. The primary reason of the success of such combined technologies is the fact that scanning sequences are actually performed sequentially, with an evident simplification of the design. A motorized bed moves from one ring to the next one in the same room. Despite its simplicity, this solution is clearly not optimal: the patient could move between sequences, with related artifacts in the final image. An important advantage of combining the two techniques is the possibility to use transmission data from the CT scan for extracting attenuation correction coefficients necessary for the PET image reconstruction. Since CT transmitted energy is in the order of $40 - 140\,keV$, models have been developed over the years for adjusting the attenuation correction factors for more energetic gamma particles, showing very good results [31].

Typically, PET/CT is successful for whole-body imaging, while it gives worse results in brain applications, due to the poor contrast between soft tissues.

---

[1] Sievert (Sv) is the unit in the International System of Units of the equivalent dose. It measures the health effect of ionizing radiation on human body.

2.6.2 *The challenge of PET/MR scanners*

Combined PET/MR scanners could solve some problems encountered with PET/CT systems. MR imaging offers in fact an higher contrast in soft tissues, when compared with CT, and does not expose patients to any radiation dose. It was invented in 1970 and has seen a rapid development in clinical applications since the 1980's.

It benefits from the magnetic proprieties of hydrogen atoms, which compose the largest part of human body. Protons in atoms, positively charged, spin around their own axis and behave like tiny magnets. When atoms are subjected to a strong magnetic field, named B_0 and directed along the longitudinal axis, they tend to align with it, resulting in a net magnetization along the same direction. The combination of spinning rotation and stable magnetic field causes the protons to precess at a characteristic frequency f_L, defined with the Larmor equation:

$$f_L = g \cdot B_0 \tag{2.6}$$

where g is a characteristic constant of each atom, called gyromagnetic constant, and B_0 is the applied magnetic field magnitude ($3\,T$ in clinical applications).

Larmor frequency is used for transmitting energy to protons, taking advantage of the resonance effect. Radio-frequency (RF) transmitting coils are placed around the field of view and produce pulsed electric field, at the hydrogen Larmor frequency. Atoms aligned with the main field get tilted so that the net magnetization of the protons lays on the transverse plane. They drift back to their initial orientation in a certain time, called T_1-relaxation time, which depends on the tissue and on its hydrogen content. The signal is received with similar RF coils, where the flowing current changes depending on the net magnetization of the atoms.

The net magnetization vector stays aligned on the transverse plane, for the first instants after the RF pulse. Protons start to immediately de-phase in other directions on the transverse plane. This second time constant, called T_2-relaxation time, or spin-to-spin relaxation, is also detected by receiver coils. T_2 is usually much faster than T_1.

Standard RF sequences are applied in order to excite protons with a specific time pattern. In the same way, the signal is sampled at defined intervals, where the net magnetization for the different tissues is most different, so that it is easier to recognize features in the image with high contrast. Depending on the diagnosis requirements, the image can be weighted on T_1 or T_2 parameter [1]. For example, fat tissues have very fast T_1-relaxation time, when compared with water, so that they appear bright in T_1-weighted images, while water is generally dark. On the other side, edemas or infections are white in T_2-weighted images, while fibrous and calcified tissues appear black, giving information on some types of functional and physiological processes as well.

In order to include spatial information in the output signal, a magnetic field gradient is applied along the longitudinal direction. It varies linearly in space: in this way, specific slides can be selected for being imaged and a 3-D image is produced, combining them.

MR tomography offers very high quality imaging, especially of soft tissues, such as brain or heart. The main disadvantage is that a full body scan still requires about 30 minutes, when compared with the few minutes necessary for a complete CT scan.

Its integration with PET is a challenging subject for many reasons. First of all, MR imaging does not give any information about attenuation correction coefficient of different tissues. Since it is based on a completely different physical principle, there is no direct way to estimate the behavior of ionizing radiation within the body. Moreover, any interference between PET detectors and RF transmitting and receiving coils must be avoided, because they could cause pickup at pulsing frequency ($128\,MHz$) in the electronics. Magnetic field gradients can induce eddy currents if circuit loops are present on the detector. On the other side, interferences of PET detectors on magnetic field must be avoided as well, for not distorting it or causing additional noise.

The larger effort needed for a complete integration is justified by the better results expected in some clinical fields, such as pediatric diagnosis, neurology and cardiology. Compared studies have been carried out to identify the advantages and drawbacks of both PET/CT and PET/MR systems [32, 33].

For practical integration of PET and MR technologies for simultaneous imaging, three alternatives are generally considered [34, 35]:

SEQUENTIAL SCANNING

A sequential design is the simplest way to combine the two technologies and is similar to the solution adopted for PET/CT scanner: rings are placed in the same room, at some distance from each other. A motorized bed moves between them so that measurements are performed in sequence. Scanners must still be properly shielded from each other and separated by adequate distance. This solution has the great advantage not to require large modifications to the original scanners: it is not expensive, but does not allow truly simultaneous imaging. Both scanning sequences can take some time (up to 40 minutes), and this could be stressful for patients. Movements during data acquisition are hard to recognize and can introduce artifacts. Moreover, motorized bed and the two systems need quite some space, which could be prohibitive in practice for clinics and research institutes who want to replace their instrumentation.

REMOVABLE PET INSERT

A more challenging solution is a small PET ring inserted in an MR ring. In such case, detectors must be newly designed, in order to be insensitive to magnetic fields and gradients and RF transparent. All components must be carefully chosen. The MR would not need large modifications, but the limited inner volume puts a constraint on the maximum PET ring diameter. Such alternative could be useful in brain imaging or small-animal applications. It allows truly simultaneous scanning with the two technologies. Mutual interference with different designs have been studied in [36, 37, 38, 39]. Great potential of truly simultaneous PET/MR scanning for brain imaging have already been demonstrated in [40].

INTEGRATED SYSTEM

A full integration of PET and MR systems requires substantial design modifications to both of them and is the most challenging and expensive solution. It is also the only design which guarantees truly simultaneous imaging of the whole body. MR instrumentation needs modification in order to allow the insertion of permanent PET ring and at the moment few designs have been proposed. A first option (targeted for the moment to pre-clinical applications) uses a split superconductive magnet [41], with a gap for the PET ring, placed behind the RF coil. The bore magnet generates a field of only $1\,T$ to reduce its impact on optical sensors. Light is transmitted via fibers to PMTs located outside the magnetic field, with large losses. A design by Siemens, commercialized only in the last few years uses a radio-frequency coil of reduced size for allowing enough space for photo-detectors and scintillators [42]. The result is a degradation of the MR performance, which depends on the coil diameter. Signa PET/MR by General Electric Healthcare includes a shaped RF shied and body coil, for the accommodation of the PET ring, which is placed within the gradient coil (see Section 5.5 for details on performance of already commercialized integrated scanners). The alternative proposed by Sublima collaboration (see Section 4) is based on a split-gradient coil, with a completely novel design, and very compact PET detectors, which can be placed in the gap between coils.

Research is ongoing in parallel with studies of integrated designs, in order to find reliable solutions for the estimation of the attenuation correction coefficients, without the need of additional transmission scans. In all proposed designs, there is in fact no space for the introduction of a rotating transmission detector. MR images distinguish tissues by their proton density and relaxation times and these quantities must be related to the ionizing radiation behavior with matter at $511\,keV$. This effect is tough to identify for some structures, like bones and cavities, which have very different attenuation coefficients but similar signal intensities in the MR. MR-based attenuation correction coefficients should be derived from special MR sequences, T_1-weighted slices or approximation models [43, 44]. An approach uses template maps, generated with transmission scans and MR, of a sample of healthy subjects. The correlation between the templates is then use to adjust the patient MR map and extrapolate an estimated transmission scan attenuation map [45]. Alternatively, multi-step registration sequences have been proposed, in order to adjust a CT representative map so that it matches the MR scan of the patient [46]. Special sequences have also been used to identify tissues in the body and assign them to an attenuation correction coefficient, based on their density and composition and their expected interaction with radiation. This is the most challenging approach, because the assignment of a proper coefficient is tough for certain areas, but has the advantage of taking into account real measurements from the scanned patient [47, 48].

Truly simultaneous PET/MR imaging techniques also offer the possibility to perform online motion correction, which is not available with any other scanner. Even if PET scanner can achieve spatial resolution in the order of few mm, resulting images, especially of the whole body, are still affected by movements of the patient. It is important to note that respiratory movements can be up to 5 times larger then the intrinsic spatial resolution of the machine. In brain studies, the situation is less critical, because rigid movements of the head can be monitored with dedicated sensors and corrected for. Using emission PET data for monitoring patient movements is very inefficient and produces very noisy signals which do not improve the final image resolution. In addition, some radio-tracers have very fast kinetics and their signal depends on the patient response to them, making it very difficult to track them online. The areas at the borders of the field of view also have lower probability of being recorded properly and are usually affected by artifacts. CT sequences are performed in very short time so that cannot be used for movement tracking. In addition, the movements affect the transmission scan results and the relative attenuation correction coefficients.

Alternatively, MR offers the opportunity to record online patient movements and re-assign proper attenuation correction coefficients to a specific time frame [49, 50, 51, 52].

3

SIPM: A NEW PHOTON SENSOR FOR PET

Silicon Photo Multipliers are on the scene since only fifteen years but already widely employed in many applications, from medical imaging systems to calorimetry in high-energy physics experiments. They offer many advantages, when compared with other well established photon-sensors, such as APDs and PMTs, and they are presently the object of intense research.

A SiPM is built as a matrix of many APD cells: the common silicon substrate is biased for them all at the same voltage, while the output of each diode is connected to a necessary quench resistor, usually in poly-silicon. All diodes are then connected in parallel to a common output. The produced signal can be considered as the sum of signals from the single APDs.

3.1 THE AVALANCHE PHOTO-DIODE

To better understand the behavior of a SiPM, it is necessary to analyze how a single avalanche diode operates. Its cross section is shown in Fig.3.1. The figure shows an n-on-p implantation, but clearly both polarities are acceptable. The diode is composed of two highly doped regions in contact, n^+ and p^+ in the figure, with excess of electrons and holes respectively, built within a weakly doped substrate p^-. A quench resistor in poly silicon is also represented in the figure. The common biased substrate is highly positive doped as well.

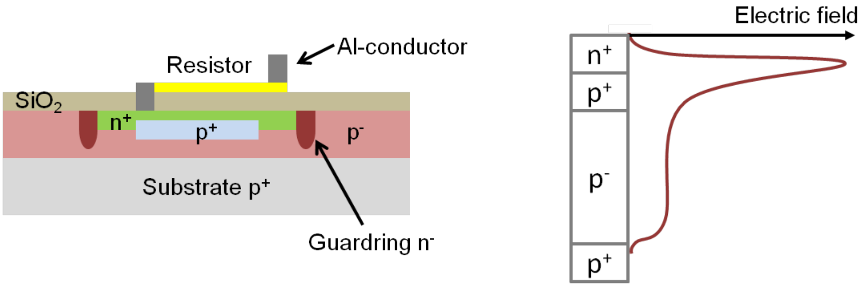


Figure 3.1.: APD cross section and electric field within the substrate.

When no bias is applied, free carriers diffuse at the boundaries between differently doped regions, leaving behind an excess of charge.

Charge concentration in the proximity of the borders creates an electric field which counteracts the diffusion process, until the equilibrium is reached. At this point, a depletion zone is created, where no free carriers are available anymore and whose depth depends on the doping level of the semi-conductor regions. The application of an external reverse bias has the effect of enlarging the build-in depletion region even more. Especially, for the avalanche diode in the figure, the electric field at the highly doped borders becomes very strong and the p^- substrate gets fully depleted, for reasonably high bias voltages. Electric field value along the cross section of an APD is reported in Fig.3.1, on the right. For avoiding edge effects at the sharp corners of the doped regions, a guard ring is typically included in the design.

A photon traveling through the diode releases its energy in the thick depleted region, producing an electron-hole pair. Due to the electric field, carriers drift in opposite directions and one of them gets accelerated by the very high electric field at the highly doped regions border. The value of the electric field in that area defines the ADP operation mode. If electric field is below a certain threshold, named “breakdown”, APD works in linear region. The carrier moving at high speed generates another pair, with a cascode of subsequent interactions, as an avalanche. The gain of an APD working in linear mode is about $10^1 - 10^3$ and the output signal is proportional to the number of photons absorbed within the substrate.

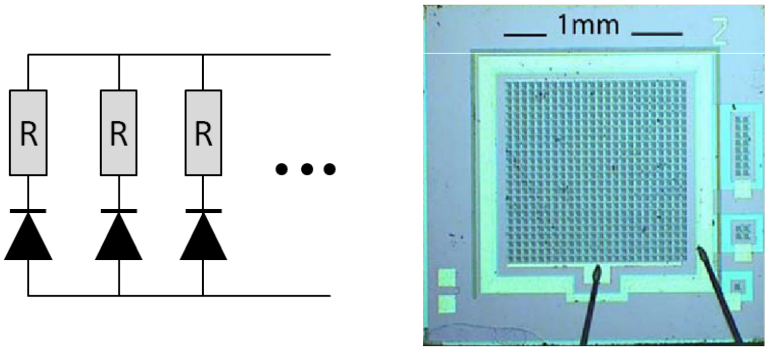


Figure 3.2.: SiPM array schematic (left), with diodes connected in parallel with individual resistors, and picture (right) of a $1\,mm \times 1\,mm$ device from [53].

If the external applied bias is high enough, the diode operates behind the linear region, in the so-called “Geiger mode”. In such conditions, the electric field at the high-doped junction is strong enough to generate an ideally never ending signal. The generated carrier has sufficient energy to generate more pairs through a self-sustaining mechanism and the silicon junction become conductive. The avalanche does not extinguish until the internal bias at the junction drops below the breakdown value. Any proportionality between the number of impacting photons and signal amplitude gets lost: each diode generates the same output for one absorbed photon or more and is insensitive until the avalanche stops. For this reason, an additional quench resistor per diode is necessary: current flowing through it induces a voltage drop at the junction which interrupts the signal generation as soon as the breakdown value is reached. The gain for APDs operating in Geiger mode is about 10^6.

SiPMs work with APDs operating in Geiger mode. Proportionality between signal output and detected photons is obtained by connecting many diodes in parallel: each of them produces the same "unit" signal if an avalanche is triggered and contributes to the final amplitude. A schematic view of the diodes connected in a SiPM matrix, together with a picture of a real device is shown in Fig.3.2.

3.2 SIPM MODELING AND READOUT

In a SiPM model, each diode can be represented by a capacitor C_d, connected with the quench resistor R_q. The quench includes a parasitic capacitance C_q, connected in parallel to R_q in the model. Both capacitors are typically on the order of tens fF, with C_d a few times larger than C_q. Another parasitic capacitance C_g has to be added in parallel to all diodes and corresponds to the metal grid spanning the entire detector area. For a SiPM of few mm^2 size, it is usually around tens to hundreds pF.

A schematic circuit of a SiPM with N diodes, where only one cell triggers an avalanche, is in Fig.3.3 [54]. The avalanche is simulated as a current spike source, in parallel with the diode capacitance. The schematic also includes the input impedance R_s of the readout front-end. The idle diodes are connected in parallel, acting as a load. The voltage drop at their quench resistors can be considered reduced of $(N-1)$ and the total load can be expressed as a parallel connection of $(N-1)C_d$ and $(N-1)C_q$, so that the total capacitance seen by a single triggered diode is equal to:

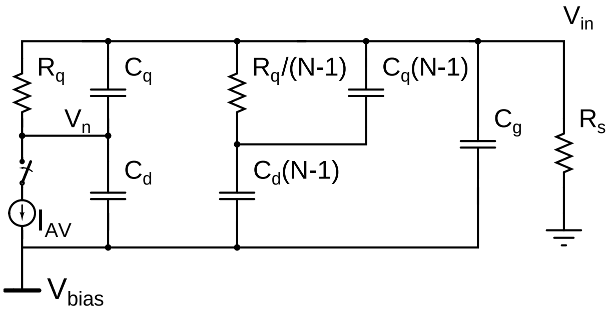


Figure 3.3.: SiPM model: in the schematic, only one diode is triggered by the avalanche spike current (I_{AV}) in a matrix of N pixels.

$$C_{tot} = \left((N-1)\,C_q || \,(N-1)\,C_q\right) ||C_g \tag{3.1}$$

Bias voltage V_{bias} must be higher than the breakdown point V_{br} and the over-voltage is defined as $OV = V_{bias} - V_{br}$.

When a photon triggers an avalanche, the switch in the model closes and a current spike discharges the capacitors, releasing a charge Q equal to:

$$Q = (C_d + C_q) \cdot OV \tag{3.2}$$

Voltage at node V_n drops to V_{bias} immediately after the spike and the current generated by the avalanche flows within R_q, until V_n is again slightly below the breakdown and the pairs production stops.

The output signal rising edge depends on the diode resistance R_d, not shown in the schematic model. Its fast time constant can be approximated with:

$$\tau_{rise} = R_d \cdot (C_d + C_q) \tag{3.3}$$

The falling edge associated with the recharge phase can be divided into a fast and a slow component. The fast component depends on the parasitic capacitive load formed by the not triggered cells. In the model approximation, and assuming that $R_s \ll R_q/(N-1)$, it can be written as:

$$\tau_{fall_1} = R_s \cdot C_{tot} \tag{3.4}$$

with R_s being the input resistance of the current amplifier and C_{tot} defined in 3.1. The second slow time constant mainly depends on the quench resistor size:

$$\tau_{fall_2} = R_q \cdot (C_d + C_q) \tag{3.5}$$

The value of R_q is temperature dependent, leading to slightly different time responses of the device depending on the temperature conditions.

Effects of different input and quench resistances have been simulated and plotted in Fig.3.4. SiPM parameters used for the simulations are reported in Table 3.1: values have been calculated for an IRST SiPM in [54].

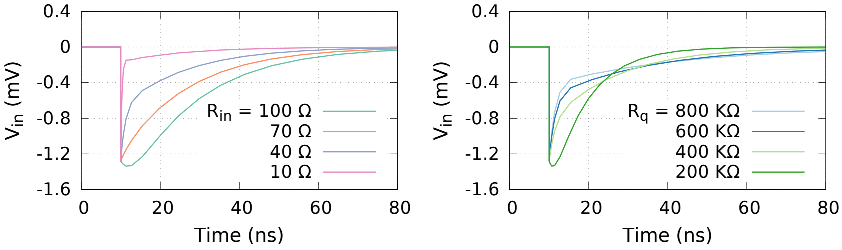


Figure 3.4.: Effects of input (left) and quench (right) resistance on SiPM voltage signal V_{in}.

The time constants can be used for approximating the signal pulse shape, as proposed in [55]:

$$V_{in}(t) \cong \frac{Q}{C_d + C_q} \left(\frac{Cq}{C_{TOT}} e^{-\left(\frac{t}{\tau_{fall_1}}\right)} - \frac{R_s}{R_q} \frac{C_d}{C_q + C_d} e^{-\left(\frac{t}{\tau_{fall_2}}\right)} \right) \tag{3.6}$$

	SiPM IRST N = 625 $V_{bias} = 35$ V
$R_q\,(k\Omega)$	393.75
$V_{br}\,(V)$	31.2
$Q\,(fC)$	148.5
$C_d\,(fC)$	34.13
$C_q\,(fC)$	4.95
$C_g\,(pF)$	27.34

Table 3.1.: SiPM parameters used in simulation. Values by [54].

The charge immediately released at the beginning of the avalanche depends on the the parasitic capacitance and is equal to :

$$Q_{in} = Q\frac{C_q}{C_d + C_q} \tag{3.7}$$

From a comparison between fast and slow components of the pulse shape, it is possible to estimate which parameters affect at most the timing performance of the device. The peak height ratio can be approximated with:

$$\frac{V_{slow}^{max}}{V_{fast}^{max}} \cong \frac{R_s C_d C_{TOT}}{R_q C_q^2} \tag{3.8}$$

The best timing performance corresponds to a fast spike and a slow recovery tail, meaning that the ratio must be as small as possible. For what concerns the SiPM parameters, the total capacitance has to be kept at minimum, while the ratio C_q/C_d should be maximized, also for enlarging the prompt charge in Eq.3.7. The readout input impedance contributes as well to the timing resolution: the ratio R_s/R_q must in fact be minimized for a faster pulse shape.

The design of an amplifier optimized for SiPM readout should take into account the previous model characteristics. The input impedance must be minimized and its bandwidth, which is typically limited, must be large enough to properly match the SiPM characteristics. Effects of the combinations of the two parameters on the amplifier output are shown Fig.3.5, for the same device in Table 3.1, taking into account the signal produced by 10 cells triggered simultaneously. The current amplifier used for the simulation is presented in Paragraph 6.1.1 and is not described in detail here. The amplifier bandwidth contributes to an additional delay on the signal rising slope, while the input resistance has large effects on the pulse shape. The fast and slow components are easily recognizable in the figure for impedance of only 10 Ω.

3.3 PROPRIETIES

SiPMs have some clear advantages when compared with other photon detectors. First of all, they are very compact sensors and are particu-

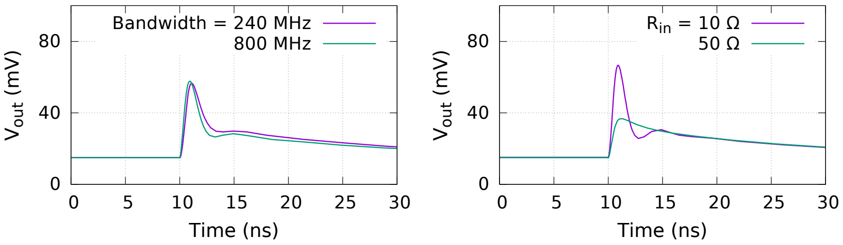


Figure 3.5.: Effects of bandwidth and input resistance of the readout amplifier on the signal shape. The output for an amplifier with an input impedance of 20Ω at different cut-off frequencies is shown on the left. On the right, the effects of the input impedance, for an amplifier with a cut off frequency of $400\,MHz$ is reported.

larly indicated for applications with mechanical constraints. They are powered at relatively low bias voltages, on the order of $60\,V$ or less, with very large gain. Moreover, they are insensitive to magnetic fields. Their most important proprieties and limitations are described here more in details.

3.3.1 *Gain and signal linearity*

The total number of electron-hole pairs produced during an avalanche is dependent on the over voltage value and on the diode characteristics. The SiPM gain G can be defined as the ratio between the charge released by a triggered diode and the unit carrier charge. Taking into account Eq.3.2, G can be written as:

$$G = \frac{(C_q + C_d)OV}{e} \tag{3.9}$$

where e is the elementary charge and is equal to about $1.60 \times 10^{-19}\,C$. While the dependance on the over voltage is clear, it has to be pointed out that SiPM gain can also shift with temperature, because the APD breakdown voltage varies with it. Deriving equation 3.9 for temperature, and with the necessary substitutions, the following formula is obtained:

$$\frac{1}{G}\left(\frac{\partial G}{\partial T}\right) = -\frac{1}{OV}\left(\frac{\partial V_{br}}{\partial T}\right) \tag{3.10}$$

The dependance of the breakdown voltage on the temperature is positive, leading to a reduction of the gain for higher temperature. The effect is almost linear and in the order of tens of mV per degree, depending on the technology. Also for this reason, SiPMs are typically stabilized at a certain temperature during operations.

When a photon triggers the internal avalanche, the diode is insensitive until its recharge is almost completed. This effect is not evident when only few photons get detected, but can become relevant for signals corresponding to many photons in short time. The final effect is the saturation of the output signal amplitude, which depends on the number of diodes in the array. Considering the incoming photons as Poisson distributed, the number of fired cells can be expressed as:

$$N_{fired} = N_{pixel} \cdot \left(1 - e^{-\frac{N_{photon}}{N_{pixel}}}\right) \tag{3.11}$$

where N_{fired} is the number of triggered diodes, N_{pixel} is the total number of sensitive pixels in the SiPM array and N_{photon} are the photons in the impacting cluster. As example, Fig.3.6 shows the number of triggered diodes as a function of the incoming photons, as foreseen from Eq.3.11, for three devices with different total number of pixels. Taking into account scintillators light output (reported in Table 2.1 for most common scintillating materials used in PET applications) it is easy to understand that SiPM non-linear behavior can have a large impact on the output signal, especially within the energy range useful in PET. In Fig.3.6, blue lines show expected photons output for 511 keV absorption in three scintillating materials (NaI, LSO:Ce and LABr3:Ce), assuming, as example, that for all of them only 25% of the incoming photons interact with the sensor. Even if the light output is large and forces the SiPM to work in saturation condition, the intrinsic non-linearity effect of the device can be deducted by its characteristics and corrected for.

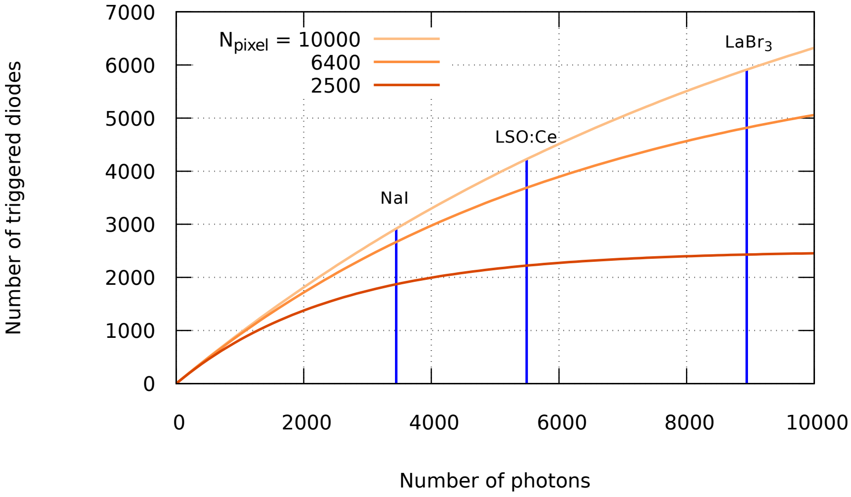


Figure 3.6.: SiPM signal saturation and its dependance on the number of pixels in the array. Ideal output for three scintillating material is plotted as well. The number of expected photons has been estimated knowing light output for 511 keV absorbed energy and assuming a detection efficiency of 25% in all three cases.

3.3.2 *Photon detection efficiency*

Photon detection efficiency (PDE) defines the capability of the sensor to produce an avalanche for every photon impacting on its surface. It is defined as the ratio between the number of triggered diodes and the total number of incoming photons. Its value depends on the wavelength of emitted light and on the bias voltage of the device and is equal to:

$$PDE(\lambda, OV) = QE(\lambda) \cdot \epsilon(OV) \cdot GE \tag{3.12}$$

$QE(\lambda)$ is the quantum efficiency of the diode and represents the probability of absorbing a photon. It is a function of photon wavelength and depends on fabrication technology. $\epsilon(OV)$ is named “Geiger efficiency” and indicates the probability of generating an electron-hole pair after the photon absorption. It is obviously less than 1 and its value increases with larger over-voltage. Finally, since not the entire SiPM area is photon-sensitive, the last factor GE represents the geometrical efficiency of the array. It takes into account the real fraction of sensitive area on the device and the reflection effects at the borders. An example of PDE dependance on the bias voltage and on the input light wavelength is shown in Fig.3.7, for two different devices (data published in [56, 57, 58]). Increasing the bias voltage usually has a positive effect on the PDE, but also affects the dark count noise (see next paragraph).

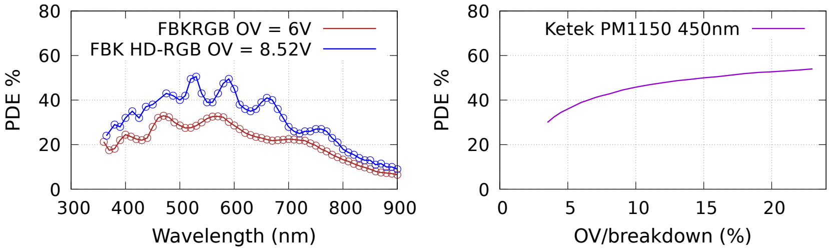


Figure 3.7.: PDE dependance on light wavelength and over voltage, for two commercial SiPM technologies. Data published in [56, 57, 58]

3.3.3 *Noise sources*

The signal from a diode triggered by noise is not distinguishable from a “real” signal and can affect the overall performance. Different effects usually contribute to the noise level within a SiPM.

The first important contribution is given by the dark count rate (DRC), which is the rate of thermally-generated electron-hole pairs in the darkness. The thermal energy is converted into vibrations of the material molecules, which can be enough to create carrier pairs, especially if impurities are trapped in the lattice, adding more intrinsic energy band levels for the electrons, and making easier for them to escape their stable state. The probability of carriers generation, depending on temperature, is well-known and expressed as:

$$p(T) = C \cdot T^{\frac{3}{2}} e^{-\frac{E_g}{2k_b T}} \tag{3.13}$$

where C is a constant depending on the material and the fabrication process, k_B is the Boltzmann constant (equal to $1.38 \times 10^{-23}\, m^2 Kg\, s^{-2} K^{-1}$),

E_g is the energy gap typical of the material and T is the temperature. Additional carriers not related to actual photon absorption come also by diffusion from the quasi neutral region in the diode, also contributing to the overall noise. In order to reduce the DCR, the first obvious solution is to operate at low temperatures. Since carriers generated by noise behave exactly as pairs generated by a photon absorption, at higher bias voltage, which translates into a higher gain, the probability to trigger an avalanche in the diode is larger, and therefore has a positive effect on the signal-to-noise ratio. Obviously, as stated in Eq.3.13, the fabrication technology and doping strategies are also important for reducing the rate of noisy-generated pairs.

Crosstalk between diodes is another source of false triggers within the array. When an avalanche is triggered, hot carriers can induce IR photons, which then are good candidates for triggering neighbor pixels [59]. This process is independent of how the first pair of carriers has been generated, and can originate from both dark count events and real signals. A crosstalk-induced event is recognizable because it is almost simultaneous with the original event. It can be measured as the ratio between double and single dark counts. The crosstalk probability between neighbor pixels is higher for higher gain. An useful technique for its reduction, which is now quite common in fabrication technology, is to introduce isolation barriers between diodes, called trenches [60]. Even if trenches have shown good efficiency in shielding the diode from the neighbor activity, part of the produced photons comes from reflection at the common substrate, and they are then not affected by the trenches.

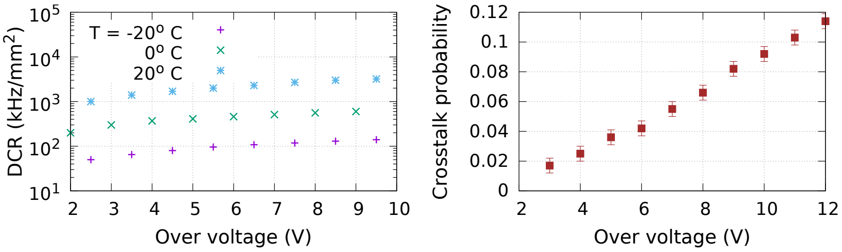


Figure 3.8.: Effects of temperature and bias voltage on the DCR and the crosstalk for FBK SiPMs in HD-RGB technology [57].

Fig.3.8 shows an example of DCR and crosstalk probability, as a function of the over voltage, for the FBK device presented in [57], fabricated in their HD-RGB technology. DRC variations with temperature are also plotted.

One more effect that can degrade SiPM signal is the so-called afterpulsing. In ideal conditions, when an avalanche is triggered, all carrier pairs are created in very short time and contribute to the fast rising edge of the signal. If there are impurities in the substrate, some carriers can get trapped in middle band energy levels. As this is a metastable state, they are released later, contributing to the signal with a large delay and possibly re-triggering another avalanche. The same diode can then trigger twice during a single event detection. A larger over voltage corresponds to a larger number of carriers produced, and also to larger impact of the after-pulsing on the final performance. For similar reasons, the temperature also affects the after-pulsing shape, inducing thermally generated carriers which can trigger successive avalanches in the diode.

As a general statement, high bias voltage usually has a good influence on sensor performance: it allows a larger gain and improves detection efficiency of the sensor. On the other hand, it simultaneously affects the noise rate, which must be kept to minimum levels to avoid false triggers which are indistinguishable from real absorption signals. Except for typical precautions during fabrication, the most straight forward way to keep noise under control is a reduction of operating temperature, which is not always easy to achieve and can put strong constraints on the assembly of a SiPM-based system. Recent developments in SiPM fabrication technology have shown very promising improvements in noise rate, for conditions close to room temperature.

4

THE SUBLIMA PROJECT

Sublima (SUB-nanosecond Leverage in PET/MR ImAging) is an European project, founded within the 7th Framework Program (2010-2015) and aimed to the design of new detection solutions for fully integrated PET/MR imaging scanner with outstanding performances [61]. It was based on the results obtained by the previous HYPERImage project.

The project target was a complete integration of a PET system with ToF capability within an MR scanner, in such a way that a truly simultaneous acquisition can take place. For reaching this goal, timing, spatial and energy resolution must be pushed to their limit, together with an optimal photon detection efficiency. Sublima aimed to detect up to 40% of gamma photons, which is a very challenging fraction, compared with the 10% normally detected by conventional PET scanners. In Fig.4.1, a sketch of the proposed system architecture is shown.

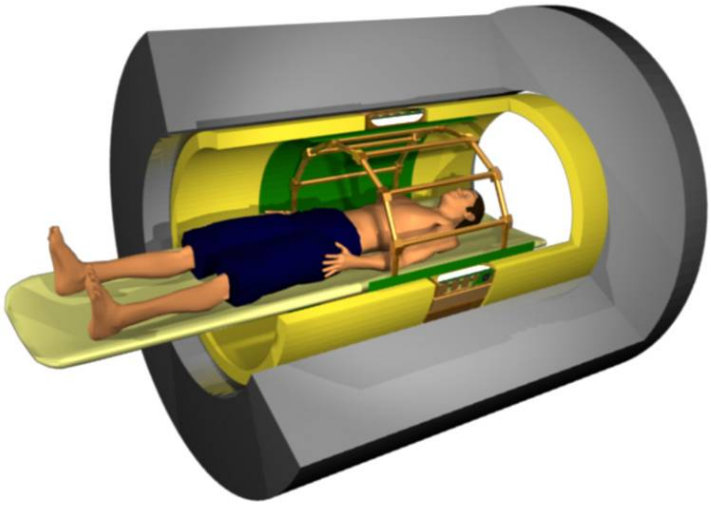


Figure 4.1.: Sublima PET/MR whole-body scanner.

The MR system had to be redesigned, for integrating a PET ring about $20\,cm$ long in axial direction. The proposal was to place the ring in between two split gradient coils, within the magnetic bore and behind the RF coils, in a design proposed already within the HYPERImage project. Each detector component, from scintillators to optical sensors and readout electronics, must be proved to be insensitive to magnetic field and gradients. The block detectors must be seamless, for reducing possible scattering and improving efficiency, and very compact, in order to fit into the tiny space available. The split gradient coils have been redesigned, for ensuring performance comparable with standard architectures. The targeted demonstrator is designed for operation with bore at $7\,T$ in pre-clinical imaging, and at $3\,T$ in whole-body applications.

To fully exploit the chances offered by simultaneous scanning of the body with the two technologies, new algorithms for MR-based motion correction and attenuation coefficients correction have been studied.

4.1 THE STARTING POINT: HYPERIMAGE

HYPERImage was a project granted by the European Union's 7th Framework Program, active in the period 2008-2011 [62]. Its goal was the development of technologies necessary for an integrated PET/MR demonstrator, with ToF resolution in clinical applications.

The outcome has been a compact detection module with good timing resolution, used for populating a small-animal ring demonstrator which has been tested in $3\,T$ bore MR environment for proving the concept. The PET insert can be placed on patient table. The scalability of the module to clinical applications has been investigated on a whole-body demonstrator, where PET detectors have been placed in the gap between split gradient coils.

A picture of the proposed detector is shown in Fig.4.2 [63]. It is composed of three PCB layers, with lateral size of about $32\,mm \times 32\,mm$. The first board hosts a matrix of 8×8 SiPMs. The sensors are produced in monolithic dies, each of them with 2×2 SiPMs. The area of each SiPM is $4\,mm \times 4\,mm$, fabricated in FBK technology described in [64]. SiPMs are coupled with an array of 22×22 LYSO crystals, with an overall pitch of $1.3\,mm$ and $10\,mm$ thickness, in pre-clinical configuration. Between sensors and scintillators, a $1.3\,mm$ thick glass plate is glued as light guide. The sensor board communicates with a front-end readout board: it includes two PETA3 ASICs in the last version, each of them with 36 differential-ended channels. Chips are wire-bonded on the top surface. Integrated TDC assigns time-stamp information per event, with a bin-width of $50.2\,ps$. The energy value is digitized internally using an 11-bits ADC. Data are sent to a pre-processing stage: a Xilinx Spartan3A FPGA on the third board takes care of data cleaning and readjusting, before communication with the main readout board.

Modules are plugged on a Singles-Processing-Unit (SPU), shown in Fig.4.3-left. Each unit can host up to 6 detectors, in a 2×3 configuration. It processes data from the modules and is connected to the control PC via an optical gigabit Ethernet link. The stacks are cooled using copper pipes, thermally coupled with the ASICs board backside and the FPGA, placed between the processing and the analogue readout boards. The PET insert, designed for hosting 10 SPU modules is

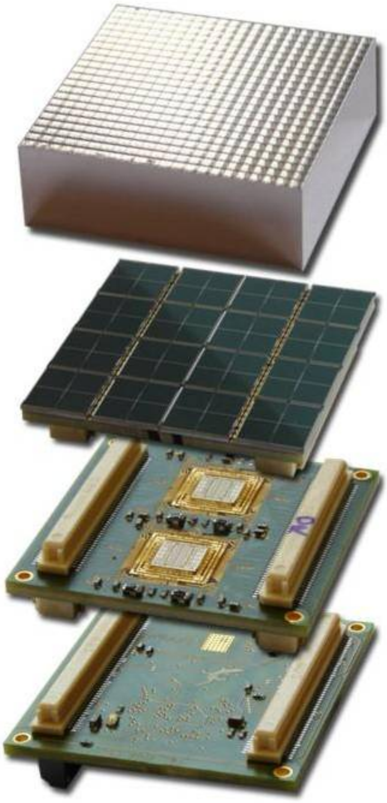


Figure 4.2.: HYPERImage PET detection module: it is composed of three PCB layers (data-processing, front-end analogue readout and SiPM array) and coupled with a LYSO array 10 *mm* thick with 1.3 *mm* pitch.

shown in Fig.4.3-right. It has a FOV of 99 *mm* in axial direction, when fully populated and diameter of 160 *mm* [65].

Detector performances have shown no degradation in MR environment with bore field at 3 *T*: spatial resolution of rods down to 1.6 *mm* in a phantom has been proved, comparable with state-of-art results in pre-clinical imaging. Also, no interference of the PET technology on RF coil noise or on B_0 field homogeneity has been registered [66][67].

Within the HYPERImage project, another detector block has been developed in parallel with the analogue design. It is based on digital SiPMs, produced at Philips Digital Photon Counting (PDPC) [68] and has same lateral dimensions of the analogue stack, using and array of 8×8 sensors. Its design is described in detail in [69].

For proving the scalability of the technology to clinical ring dimensions, the necessary modifications of the MR scanner have been tested. Split gradient coils and PET transparent RF coils have been tested in 3 *T* modified Achieva MR scanner, together with two PET modules. In terms of timing performance, HYPERImage detector has reached ToF resolution of about 449 *ps* CRT at FWHM, during MR EPI acquisition [70].

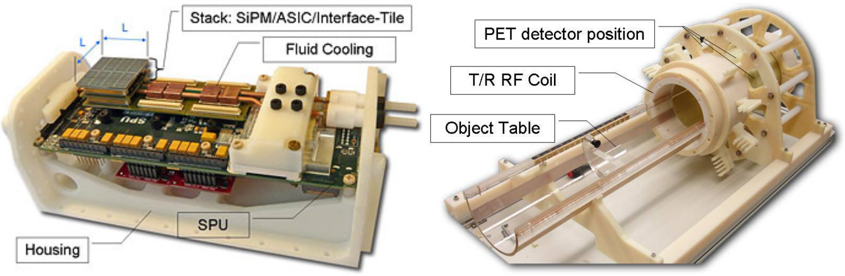


Figure 4.3.: HYPERImage SPU control unit with one detector module (left) and PET insert backbone (right). Pictures by Philips.

Strong effort has been put also for estimating the attenuation correction maps for the PET signal using MR data. Since such research has gone forward during Sublima project, the latest results on this topic are discussed in the next paragraph.

HYPERImage has succeeded in proving the feasibility of a PET detector, with a technology that does not interfere with the MR scanner, reaching good performance with still room for improvement.

4.2 TOWARDS INTEGRATED PET/MR

Sublima takes advantage of developments and results of HYPERImage and pushes their performance even forward. The optimization of every detection and reconstruction step has been addressed, in order to reach the best resolution at the state of art for both clinical and a pre-clinical systems.

4.2.1 *System analysis and simulation*

Effects of the detector characteristics on the reconstructed image quality have been extensively simulated for understanding the most critical requirements for a system. Since PET/MR scanners are not yet conventional, it is not clear at the moment which are the most important parameters to address during optimization. The most interesting results for the present work, which have been taken into account during detector design (Section 7.1) are reported in [71]. Monte Carlo simulations have been used for studying the effects of ToF and pixel size on the image quality, for a whole-body scanner with inner diameter of $60\,cm$ and without any DOI estimation. Results have shown that, for a system with ToF resolution of $600\,ps$ CRT at FWHM, lesions detectability does not benefit from pixel size smaller than $2.6\,mm \times 2.6\,mm$. This is especially true for positions close to the FOV center, where DOI effects are less relevant. On the other side, simulations have shown that for pixel size of $2.6\,mm \times 2.6\,mm$ a large improvement is obtained pushing timing resolution between detectors down to $200\,ps$ CRT at

FWHM. Even in this configuration, benefits are less evident for areas more affected by DOI uncertainty.

Simulations on a whole-body scanner have been studied also in [72], where effects of DOI estimation have been also taken into account. For simulations, detectors with 2 layers of scintillating crystals have been considered. Results have shown how small pixel size and improvements in timing resolution have larger impact if compared to DOI estimation on two layers. DOI helps especially in giving more homogenous performance in larger areas of the FOV, improving lesion detectability at the edges.

4.2.2 *MR-transparent detector cubes: alternative designs*

Few design options have been investigated for choosing the best alternative for the final system. The detector cube based on analogue SiPM tile and front-end ASIC is the central subject of this work and its design is detailed described in Part II.

An alternative option is based on digital photon counters, where sensitive area and readout logic are fabricated on the same silicon substrate. The detection module is based on digital tiles from PDPC [68]. Three detector versions have been developed within Sublima. They offer the advantage not to require separate readout front-end development and can ideally be used with very low threshold, equivalent to the first photon in the best case. On the other side, the technology is affected by dark count, so that its performance is very sensitive to operation temperature.

Each tile is composed of 8×8 sensors, about $4\,mm \times 4\,mm$, for an overall sensitive area of about $32\,mm \times 32\,mm$. Pre-clinical modules with pixelated LYSO crystals, with pitch of $1\,mm$ and $12\,mm$ thickness have been assembled, with an interposed light guide. They have shown energy resolution of 12.8% at FWHM, with 20% of cells disabled for dark count rate reduction. The system has been operated at 5^oC, reaching timing resolution of 265 or $450\,ps$ CRT at FWHM, depending on the trigger level [73]. When coupled in clinical configuration, with LYSO $22\,mm$ thick, the detectors reach 220 and $300\,ps$ CRT at FWHM for measurements at 10^oC.

Digital modules have been also assembled with monolithic scintillator blocks. The monolithic block requires precise calibration strategy, but allows DOI estimation. The gamma absorption point can be deduced from the size of the detected cluster. Moreover, monolithic blocks maximize the module detection efficiency, avoiding the losses at crystals borders, are cheap and simplify the assembly. The designed modules have shown very good timing and energy resolution, reaching less than $200\,ps$ CRT at FWHM with different sizes of monolithic blocks, operating the system at -25^oC. The energy resolution has been measured to be 12.6% at FWHM [74]. Absorption position in the transverse monolithic plane can be reconstructed with spatial resolution of about $1\,mm$ at FWHM, in average on the all area [75].

Novel scintillating materials have also been investigated. $\mathrm{LaBr_3{:}Ce^3}$, which has fast response and is indicated for applications aiming to good timing resolution, has been characterized with different Ce^3 concentration and has shown great potential [14]. As previously discussed (Paragraph 2.2.1), production costs are still too high to make it seriously competitive for clinical applications. LSO:Ce and LYSO crystals have also been taken into account and compared with Ca doped versions [76].

4.2.3 *Novel reconstruction algorithms*

Reconstruction and data correction methods have been improved, for achieving better lesion detectability but also for being compatible with PET/MR [77]. A short overview on the developed techniques are reported here for completeness.

Attenuation correction algorithms represent one of the most challenging issue for the application. A design of an annulus-shaped source to be placed at the boarders of the FOV and used for transmission scan has been proposed. Transmission data can be recognize from ToF information in this configuration [78]. In such approach, no MR data are used for estimation of the attenuation map and an additional dose to the patient is necessary, but direct measurements can be taken simulataneosly with emission registration. For reducing effects of scattering and random coincidences, a similar approach but using line sources placed around the FOV have been investigated as well and evaluated on a real PET/MR system [79].

An alternative approach using MR data has been investigated as well. A fast simulation method for validating the theory has been proposed, based on real MR data set [80]. Tissues are segmented and assigned to an attenuation correction coefficient, depending on their typically known behavior in emission scans. Additional MR derived deformation fields and projection noise are added to data, for more realistic results. Based on such simulation strategy, both attenuation and motion correction algorithms have been developed. The large impact of respiratory motion correction for lesions detectability has been proved for the abdominal region, where only MR-based attenuation correction are usually not precise enough [81, 82]. Tracking motion position allows the assignment of attenuation correction map for each frames, with error reduction of about 50% in simulations.

5

SURVEY ON OTHER PROJECTS

The present overview is an attempt to describe the most significant developments in the broad field of ToF-PET and combined PET/MR. The focus is kept on solid state PET detectors based on SiPM sensors. For this reason, Section 5.1 collects information on ASIC designs dedicated to SiPM readout and their results.

Novel SiPM topologies for very good spatial resolution of small photon clusters are instead presented in Section 5.2. Both designs have actually been developed within the Sublima project, but are here described as comparison with the position-sensitive SiPM design presented as a part of this thesis.

A survey on developed detectors and results in research applications, for both small-animal and whole-body scanners, is presented in Section 5.3. Only research projects developed within the last five years and with outstanding performances in coincidence timing or spatial resolution have been taken into account for the review.

Additional information on already commercialized scanners in whole-body imaging are reported as well: Section 5.4 refers to PET scanners with TOF capabilities, which are nowadays commonly coupled with CT scans. Section 5.5 presents instead the very first scanners on the market integrated with MR technology.

5.1 ASIC DESIGNS FOR SIPM READOUT

Thanks to the good proprieties of SiPMs, many research groups are nowadays involved in the design of dedicated ASICs with optimal performance in SiPM readout. Few designs, aiming to exceptional coincidence timing resolution for ToF applications, are taken into account in this review, which does not aim to be exhaustive. All reported data and results are cited as available on published sources at the time of writing. For each design, results obtained with the latest chip version, when available, have been taken into account.

STiC (SiPM Timing Chip), developed by University of Heidelberg for the Endo ToF-PET US project [83], is designed in UMC $180\,nm$ technology and includes 64 channels in its latest design. The sensor can be read out with a selectable differential or single-ended front-end. The input stage is designed as a grounded gate stage with current feedback path, for ensuring large bandwidth and fast operation. Energy and time information are encoded in timestamps and have separate programmable threshold. The chip uses linearized time-over-threshold (ToT) measurement for the estimation of the signal energy. TDC design is the same of PETA ASIC, described in more details in section 6.2.2, with bin-width of $50\,ps$. It includes a 64-word FIFO for storing data, read out every $6.4\,\mu s$ through LVDS serial link at $160\,Mbits/s$ with an $8-10$ bits encoding.

Within the same project, a second ASIC, named TOFPET, has been designed by a collaboration between INFN and University of Torino (Italy), and INESC-ID, LIP and IST UTL in Lisbon (Portugal) and fabricated in UMC $110\,nm$ technology [84]. The input stage is designed as single-ended front-end, with an adjustable impedance in a range of $10-60\,\Omega$. It is based on a regulated-gate cascode: the injected current is mirrored on two branches, for timing and energy measurement respectively. Signal energy is estimated through ToT linearization. It includes a TDC with selectable $50-25\,ps$ bin-width. Triggered data are read out with $8-10$ bits encoding with LVDS lines, at $160\,Mbits/s$. Characterization measurements of the second chip version (TOFPET2) have been published recently: smaller TDC bin-width of $20\,ps$ and higher affordable data rate per channel (from 100 to $600\,kHz$) have been implemented [85].

FlexToT is ideally designed by University of Barcelona and CIEMAT Madrid (Spain) [86]. The ASIC hosts 16 channels and is designed in AMS $350\,nm$ technology. It also makes use of ToT measurement for avoiding integration and digital conversion of the signal amplitude. It does not include any TDC and ADC: the fast rising edge signal is sampled externally by an FPGA which also takes care of data processing. The pulse width is proportional to the injected signal amplitude. The signal including time and energy information can be few tens ns long. Clearly, timing resolution depends on the external TDC design as well. A suggested application is in performing DOI measurements with different scintillator materials, taking advantage of their different time constant and signal shape.

BASIC is an ASIC born from the collaboration between Universities of Pisa and Bologna (Italy), to be used in PET/MR systems. The latest version (BASIC2) hosts 32 channels, each of them based on a very low input impedance current amplifier. In the most recent design, it includes an ADC for digital conversion of the energy measurement. Energy is measured as integrated charge delivered by the input signal: the voltage peak reached at the end of the integration is measured and digitized internally. The hit signals are OR-ed on a fast path within the chip and sent out on a dedicated pad. The signal is used for starting the readout and gets sampled by an external precise TDC [87].

SPIROC2c is a design developed by Omega group in Orsay (France) [88]. Each channel (36 in total) can tune the DC input level in a range of

$5\,V$ and includes two amplifiers, with different gains, for handling large and small signals. A slow shaper adjusts the signal, which gets stored on an internal memory and used for energy estimation. A fast shaped signal gets instead compared with a configurable threshold and used on an internal TDC ramp for measuring the arrival time. Both energy and timing values gets converted using an internal 12-bits Wilkinson ADC. The same group designed and commercialized other ASIC for SiPMs readout (PETIROC, EASIROC) with similar features, not described here.

The last design taken into account for this review, is the ASIC NINO, designed from LAA collaboration at CERN. It is not specifically targeted to SiPM readout, since it has been developed for the readout of Multigap Resistive Plate Chamber [89]. It has been used with good performances in SiPM readout and for this reason is included in this overview. Channel design is fully differential and consists of high precision amplification stages and a discriminator. After a first pre-amplifier, four stages with low gain and large bandwidth amplify the signal. Input stages are kept at controlled offset with a low frequency feedback loop. The output signal is sent to a pulse stretcher, which shapes the waveform depending on the energy information before sending it out via LVDS pads. Output pulses can be from 0.5 to few ns long, with a data rate in the range of $100 - 400\,MHz$ maximum , in standard operation mode. External TDC/ADC designs are necessary.

The most important characteristics of the ASICs described in this section are reported in Table 5.1.

5.2 POSITION-SENSITIVE SIPMS

Standard SiPMs give good estimation on the number of impacting photons, but does not include any information on the detection position. The signal of many diodes in the array is read out in parallel on the same output channel. Few designs have been proposed in recent years, for overcoming this limit. They are mainly targeted to small-animal PET scanners, where required high spatial resolution is impossible to obtained with a 1:1 coupling configuration between SiPMs and crystals, due to the enormous number of readout channels necessary for such assembly. Signal is usually spread on many sensors, whose detected amplitudes are used for mathematically reconstructing the center of gravity of the original event. An intrinsically position sensitive SiPM design allows to read out a single sensor per event, without the need of any neighbor information, saving readout bandwidth.

The design proposed within this thesis is described in details in Section 7.2. Other two devices are taken into account in this review.

The first one has been proposed in [90]. Each avalanche diode in the array is connected to one of the available outputs. Light produced within a scintillating crystal is read out by cells in a certain position. If

	STiC	TOFPET	FlexToT	BASIC32	SPIROC2c	NINO
Technology	UMC 0.18 μm	UMC 0.11 μm	AMS 0.35 μm SiGe BiCMOS	AMS 0.35 μm CMOS	AMS 0.35 μm SiGe	IBM 0.25μm CMOS
Area (mm^2)	25	25	10	22	32	8
Channels	64	64	16	32	36	8
Power/ch (mW)	20	8	10	> 2.65	2.5	40
Front-end	SE/DE	SE	SE	SE	SE	DE
Input range (V)	0.7	0.5	-	1	5	–
Input Impedance (Ω)	< 50	$10 - 60$	30	17	AC	20
TDC bin (ps)	50	50(25)	–	–	300	–
Jitter (ps)	< 30	< 25	< 18	< 625	100	< 25

Table 5.1.: ASICs for SiPM readout and their characteristics.

the cells assignment to the output channels is properly encoded, then crystals can be recognized, mathematically weighting the output signal amplitudes. Two versions have been designed, using one and two-dimensions cell distributions. In the 1-D layout (visible in Fig.5.1-left), a square device is subdivided into four strips, each of them read out by two output channels. The assignment of each channel to the cells follows a linear pattern, with a decrease/increase density of the output distributions. In case of a 2-D device, the configuration designed is shown in Fig.5.1-right. The sensor is read out using only four output channels for the whole matrix. Cells assignment follows a linear distribution, with the same decrease/increase patter of the 1-D device, in both horizontal and vertical directions. In general, such devices offer a reduction in the overall number of readout channels and have shown good spatial resolution, clearly resolving a scintillator array with crystals of $0.93\,mm \times 0.93\,mm \times 10\,mm$, coupled with a sensitive area of $4\,mm \times 4\,mm$. The main limit of such design is that resolution is dependent on the sensor size and subdivision (in strips or in pixels) and relays on a precise crystal placement.

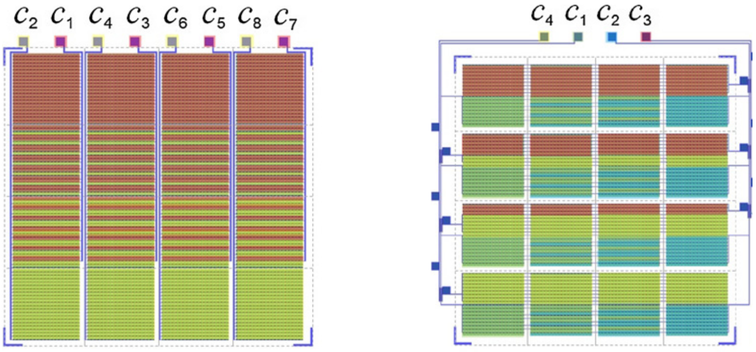


Figure 5.1.: Position Encoded SiPMs by [90]. On the left, example of 1-D encoding using two readout channels for vertical strip. On the right, 2-D encoding topology with 4 readout channels.

An alternative design with linear encoding of the detection position within the matrix has been first proposed in [91]. Devices are known as linearly-graded SiPMs (LG-SiPMs). In such architecture, position is encoded using resistive current dividers between diodes. In a 1-D case, the signal from one column gets spread onto two output channels so that the current read out from them varies linearly depending on the pixel position. In a 2-D architecture, two channels decode the column position and other two the row position. The schematic architecture of a 2-D LG-SiPM is shown in Fig.5.2. LG-SiPMs performance in crystals identification are not dependent on the scintillator position and the same device can be coupled with arrays with different pitches. An additional advantage is that many sensors can be connected together, so that four channels are sufficient for a large sensitive area. On the other side, a very low input impedance front-end architecture is necessary for avoiding signal degradation and then also resolution capability. Iden-

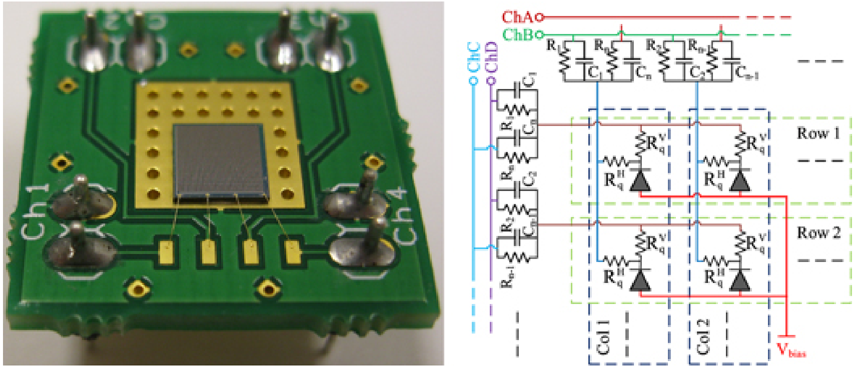


Figure 5.2.: Linearly-Graded SiPMs: sensor picture (left) and architecture of a 2-D encoded resistive current divider used for reading out the array (right) by [92].

tification of crystals in an array with $0.53\,mm$ pitch has been achieved [92].

5.3 RESEARCH PROJECTS ON NOVEL PET DETECTORS

Extensive research is going on for improving the performances of clinical and pre-clinical PET detectors. Most part of recent projects is also oriented versus MR compatible technologies.

Detectors are typically optimized for either timing or spatial resolution, in order to address the most important requirements for clinical and pre-clinical scanners. For this review, focus is kept on novel detectors and technologies with great potential in one or both applications.

5.3.1 *Novel ToF-PET detector systems*

Many research projects aim to ToF performances, with CRT in the order of few hundreds of ps.

Most collaborations are in the initial phase of investigating the optimal design of ToF-PET detectors, with ideally good compatibility with MR applications. Only solutions which have already proved results at system level, or are very close to, and with coincidence timing resolution less than $300\,ps$ (FWHM) are considered.

A group from Jagiellonian University in Cracow (Poland) proposed organic scintillating detectors for ToF-PET, instead of the more traditional inorganic materials [93]. Such choice is justified by strong reduction of the overall cost for whole-body solution, which normally contains thousands crystals. Polymer scintillators emit very fast signals, with decay time in the order of only few ns (compared with tens of ns of inorganic scintillators) with an advantage for timing resolution that can reach CRT values below $100\,ps$. On the other side, since photoelectric effect during absorption is almost negligible, emitted photons can be considered as totally produced by Compton scattering, with degradation of energy resolution. Since gamma particles, scattering within the FOV release different energy in the scintillator, an energy cut can be still applied at lower energies. Losses in the signal can then be compensated by longer FOV, thanks do the more affordable price of materials. A prototyped solution built in a ring of $30\,cm$ long plastic scintillators, oriented in the axial direction and read out at the edges by two matrix of SiPMs, has proved CRT for annihilation events of $266\,ps$, using a collimated ^{68}Ge source [94]. Energy resolution $\sigma(E)/E$ at the Compton edge is 7.5%.

Using a similar design concept design, a collaboration from ETH University in Zurich (Switzerland) has designed and built a prototype ToF-PET detector with long LYSO scintillator ($3\,mm \times 3\,mm \times 100\,mm$) oriented along axial direction. The design is similar to the previously developed detector AX-PET [95]; ETH group has introduced a digital-based readout, aimed to TOF performances. Light is detected with dSiPMs at the scintillator edges. In addition, thin wavelength shifter (WLS) are placed between crystals, in orthogonal direction, and read out by one dSiPM array, for estimation of the absorption axial position. Two prototype blocks have already been tested: they include only 4 long scintillators, for allowing a fast readout of the dSiPM (dead time increases when reading out full 8×8 matrix of sensors), and 8 WLS strips, placed in two rows in the central region. Temperature during operation is kept at about $16^{o}C$, for reducing dark count rate in the sensors. CRT between two blocks, measured only for events within the photo-electric peak and detected by the WLS system, has reached about $270\,ps$ at FWHM.

An uncommon configuration has been proposed within the project Endo ToF-PET-US [83]. Its goal is the design of a multimodal imaging system for endoscopic exams. The image obtained with ToF-PET scanner is combined with ultra-sound results. A miniaturized PET detector is integrated on the scope and placed in the proximity of the target. Coincidences are acquired together with a large detector plate, placed around the patient. The system asymmetry and the small detector size on the probe are challenging constraints for the design. Scope detector is designed with LYSO:Ce crystals (size $0.71\,mm \times 0.71\,mm \times 10\,mm$) read out with digital SiPMs. The external plate is composed of 256 block detectors, each one with 4×4 LYSO:Ce crystals (size $3.5\,mm \times 3.5\,mm \times 15\,mm$) coupled with analogue SiPMs. Both STiC and ToF-PET ASICs have been designed within the project for SiPM readout (Section 5.1). Measurements on dedicated test setup with MPPCs from Hamamatsu and LYSO crystals of $3.1\,mm \times 3.1\,mm \times 15\,mm$ have achieved ToF resolution. STiC has reached $220\,ps$ CRT at FWHM, when connected in differential-ended front-end configuration in the latest chip version and energy resolution 12% at FWHM [96, 97]. TOFPET reached $270\,ps$ CRT at FWHM and 17% energy resolution at FWHM at the photo peak, defined as preliminary result in [97]. Results for whole system are presently not available.

Stanford University (California, USA) has developed a ToF-PET detection module with very good timing performances [98]. Detector is based on LYSO and SiPM sensors and uses multichannel low noise RF amplifiers for readout. Waveforms are digitized with commercial modules , with 12 bits resolution and $500\,MHz$ input bandwidth, for offline data processing. This configuration has achieved very good timing resolution: with short crystals of $3\,mm \times 3\,mm \times 5\,mm$ CRT is measured to be $206\,ps$ at FWHM, with energy resolution of about 11% at FWHM (without considering SiPM non linearity effect).

5.3.2 *High-spatial resolution detectors*

Recent developments in high resolution detectors dedicated to small-animal imaging are presented here. The focus is kept on the design of full systems.

Philips presented in 2012 its first digital insert for small-animal imaging, named Hyperion-II^D and compatible with MR applications [69]. The project has been partially granted within Sublima. Each detector module is composed of scintillating crystals $12\,mm$ long, with an overall pitch of $1\,mm$, coupled through a trenched light guide with a 4×4 quad dSiPMs (64 sensors in total for a block detector of about $30\,mm \times 30\,mm$) produced by PDPC [68]. Modules are mounted in a 3×2 configuration on a common board (black box in Fig.5.3), which is assembled on the gantry. The gantry FOV is $210\,mm$ in the transverse diameter and $96\,mm$ in the axial direction. The insert has already been operated in MR, studying the interferences of technologies on each other, showing good uniformity in B_0 field distribution and stable performances of PET detectors. Single crystals reach energy resolution of about 14% at FWHM, inside and outside the MR bore.

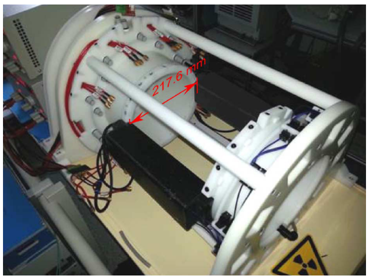


Figure 5.3.: Hyperion-II^D gantry, picture from [69].

Seoul National University (South Korea) developed a small-animal insert, compatible with MR operations at $7\,T$. It is based on analogue SiPM tiles, produced by Hamamatsu. Sensors are coupled with LYSO array through optical grease, without any light guide in between. Each crystal is $1.2\,mm \times 1.2\,mm \times 10\,mm$. A matrix of 9×9 crystals is read out by 4×4 SiPMs: signals are multiplexed through resistive charge division network, for position encoding. Commercial amplifiers and TDC components extract timing and energy information from signals [99]. The detector achieved energy resolution of 14.2% at FWHM in a wide range of temperatures. PET insert is shielded with carbon fiber and copper film, with an assembly that minimizes eddy currents. It must be placed between the transceiver and the gradient coil. It reaches sub-millimeter resolution in the order of $0.5\,mm$ at FWHM for reconstruction of point sources placed at the center of the FOV [100].

DigiPET is a scanner with very large solid angle, designed by Ghent University (Belgium) and partially supported by Sublima project [101]. Its target is mice brain imaging and is smaller than a standard ring. A prototype with four block detector has been assembled in a square arrangement with a distance between opposite detectors of only $34.5\,mm$. A picture of the prototype is visible in Fig.5.4. Each block is composed of a monolithic LYSO of $32\,mm \times 32\,mm \times 2\,mm$, read out with a matrix of digital SiPMs from PDPC. Reduced sensitivity due to thin scintillators is compensate by large solid angle. From first measurements, the system has reached energy resolution of 18% at FWHM and CRT of $680\,ps$ at FWHM, with intrinsic spatial resolution within the detector of only $0.54\,mm$.

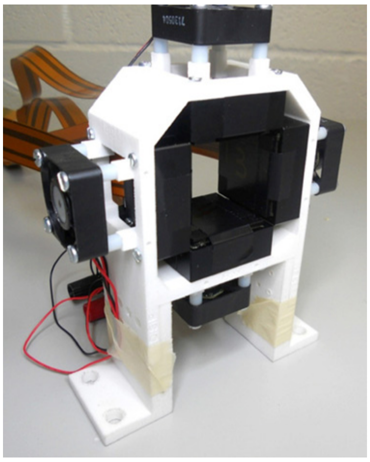


Figure 5.4.: DigiPET square scanner [101].

Results obtained with the most novel scanners are showing very promising progresses in the field. Intrinsic spatial resolution of *mm* or less, in some cases, has been achieved and allows spatial resolution at the center of the FOV of less than $1\,mm$. More established solutions, such commercialized system Inveon micro-PET by Siemens, available on the market since around ten years, demonstrate spatial resolution in the FOV in a range between 1 and $2\,mm$ [102].

5.4 COMMERCIAL WHOLE-BODY TOF-PET/CT SCANNERS

A short overview on the commercialized ToF-PET systems is here briefly reported, for understanding the established results and the impressive progresses already done by research in the field. All clinical scanners here described are built in tandem configuration with CT systems. The specification of PET detectors are reported for knowledge in Table 5.2; some values are presently not available [103].

Philips Ingenuity TF is on the market since 2006 and has been the first commercialized PET/CT scanner [104]. It reaches coincidence timing resolution of $502\,ps$ at FWHM. Discovery-690 [105] and Biograph mCT [106], by General Electric and Siemens, have been commercialized later and are based on similar technologies. Biograph mCT has

	Philips Ingenuity TF	Siemens Biograph mCT	GE Discovery 690	Philips Vereos Dig	Toshiba Celesteion
Technology	PMT + LYSO	PMT + LSO	PMT + LYSO	dSiPM + LYSO	PMT + LYSO
Crystal Size (mm^3)	$4 \times 4 \times 22$	$4 \times 4 \times 20$	$4.2 \times 6.3 \times 25$	$4 \times 4 \times 19$	$4 \times 4 \times 12$
Energy resolution (%)	11.1	11.5	12.4	11.1	NA
CRT at FWHM (ps)	502	527.5	544.3	316	≈ 410
Coinc. window (ns)	4.5	4.1	4.9	4	NA
Axial FOV (mm)	180	218	157	164	196
Resolution (mm)					
Transaxial at $1-10\,cm$	$4.8-5.1$	$4.4-4.95$	$4.7-5.06$	$4.1-4.5$	$5.1-5.1$
Axial $1-10\,cm$	$4.73-5.23$	$4.4-5.9$	$4.74-5.55$	$3.96-4.3$	$5.0-5.4$
Sensitivity (cps/kBq)	7.3	9.7	7.4	5.7	NA

Table 5.2.: Proprieties of the commercial ToF-PET/CT scanners available nowadays, produced by Siemens, Philips, General Electrics and Toshiba.

slightly worst timing resolution, if compared with Ingenuity TF, but offers larger sensitivity and longer FOV in axial direction. Toshiba has produced its PET/CT scanner with ToF capability [107], available on the US market since 2014. It reaches better timing resolution (down to about $410\,ps$ at FWHM) if compared with the previous ones, but with worst spatial resolution within the filed of view. These four scanners are all based on PMT sensors. Philips Vereos Digital [108] is even more recent and is the first scanner based on digital SiPM sensor by PDPC [68]. Crystals are coupled in a 1:1 configuration with SiPM arrays, leading to very high count-rate and good timing resolution, down to only $316\,ps$ at FWHM. Spatial resolution in both axial and transaxial planes improves of almost $1\,mm$ compared with other scanners.

5.5 COMMERCIAL WHOLE-BODY PET/MR SCANNERS

PET/MR first scanners are today available on the market. Three producers offer combined PET/MR technologies: Siemens Healthcare, with the scanner Biograph-mMR [42]; Philips with Ingenuity TF [109] and General Electrics Healthcare with Signa PET-MR scanner[110]. The scanners have pretty different designs and performance [111]. The most important features and characteristics are summarized in Table 5.3.

	Biograph-mMR	Ingenuity TF	Signa
Integration	Simultaneous	Sequential	Simultaneous
	PET detector		
Technology	APD + LYSO	PMT + LYSO	SiPM + LBS
Crystal Size (mm^3)	$4 \times 4 \times 20$	$4 \times 4 \times 22$	$3.95 \times 5.3 \times 25$
Timing Resolution	no TOF	TOF	TOF
CRT at FWHM (ps)	2900	525	385
Axial FOV (mm)	258	180	250
Sensitivity (cps/kBq)	13.2	> 6	21
	MR system		
Field B_0 (T)	3.0	3.0	3.0
Bore size (cm)	60	60	60

Table 5.3.: Proprieties of the commercial PET/MR scanners available nowadays, produced by Siemens, Philips and General Electrics.

Philips Ingenuity TF has been the first PET/MR commercialized system. It does not offer simultaneous imaging, but is a sequential scanner with two coaxial rings, with FOV at $4.2\,m$ distance properly shielded. It reaches TOF resolution with CRT of about $525\,ps$ at FWHM and axial and transverse spatial resolution at $1\,cm$ from the FOV center of $4.7\,mm$ [112].

Biograph-mMR has been the first truly simultaneous PET/MR on the market. It is composed of 8 detector rings, each one with 56 modules for totally more than 4000 channels. It does not reach ToF resolution, with a coincidence resolution of only $2.93\,ns$. It achieves average spatial resolution of about $4.3\,mm$ at FWHM at $1\,cm$ distance from the center of the FOV, with scatter fraction of about 40%. MR has not substantial impact on PET detectors performances; it has been measured good homogeneity of the static magnetic field and no RF interferences [113].

More recently, General Electrics has presented its simultaneous TOF PET/MR scanner Signa. PET detector reaches CRT at FWHM of $385\,ps$ outside the MR, compared with $394\,ps$ in simultaneous operations [114]. MR signal to noise ratio shows small degradation when operating with PET, but static magnetic field distribution is still homogenous. It has achieved transverse spatial resolution of $4.2\,mm$ and higher sensitivity at the center of the FOV, if compared with other systems (see Table 5.3).

Part II

DEVELOPMENTS

This part describes the designs implemented for improving the detector resolution. The goal is to reach both outstanding timing and spatial resolution, in combination with high sensitivity and compatibility with MR technology.

Chapter 6 is dedicated to the readout electronics developed for SiPM readout. A novel single-ended front-end architecture has been included in an ASIC design with TDC/ADC conversion.

Chapter 7 describes novel optical sensor designs. SiPM layout has been optimized for maximizing light collection. In addition, an innovative SiPM topology, named Interpolating SiPM, has been developed targeting very good intrinsic spatial resolution for small-animal PET detectors. The design of a SPAD array with integrated readout logic in CMOS technology is also described, useful for studying light distribution in many applications.

Finally, Chapter 8 is dedicated to the detector module, its cooling strategy, features and assembly configuration for clinical and pre-clinical scanners.

6

READOUT ELECTRONICS DESIGN

Dedicated front-end electronics for SiPM readout has been designed and characterized. The architecture aims to very low input impedance and wide bandwidth, for excellent timing resolution. Few front-end design options have been taken into account during development phase: all them have been characterized on a test chip. The selected one has been integrated on latest versions of PETA ASIC (Position Energy Timing ASIC).

6.1 LOW INPUT IMPEDANCE SINGLE-ENDED FRONT-END

A single-ended front-end channel appears to be the best solution for a compact detector module: it minimizes the number of connecting lines and does not require additional external decoupling components on the board assembly. Compared with differential-ended architectures, it has the disadvantage of larger sensitivity to noise.

Focus has been kept on very fast architecture, with extremely low input impedance and contained power consumption. The discriminator logic has been realized with fully differential circuitry, so that the overall noise is kept within reasonable range. Signal energy is measured with charge integration and digitized with a simple and fast architecture, for keeping at minimum channel dead time, without losing too much in resolution.

For design and optimization of the front-end architecture, the SiPM model described in Section 3.2 has been considered. Since the SiPM signal is mainly a current pulse, the input stage is designed as a current buffer. A proper dynamic range is mandatory for maintaining linear performance in many conditions. The amplified current is converted into voltage and compared with a programmable threshold for trigger definition. Mirrored current is injected in a parallel branch for energy estimation.

The input stage is directly connected to the SiPM output and input DC voltage can be adjusted for varying sensor bias voltage and gain. Fine adjustment of the over-voltage from the readout side can be used for compensating small inhomogeneities in the breakdown voltage between sensors, normally resulting after fabrication.

A simplified view of the full single-ended channel is shown in Fig.6.1: it includes all circuit blocks that are described in details in the following paragraphs.

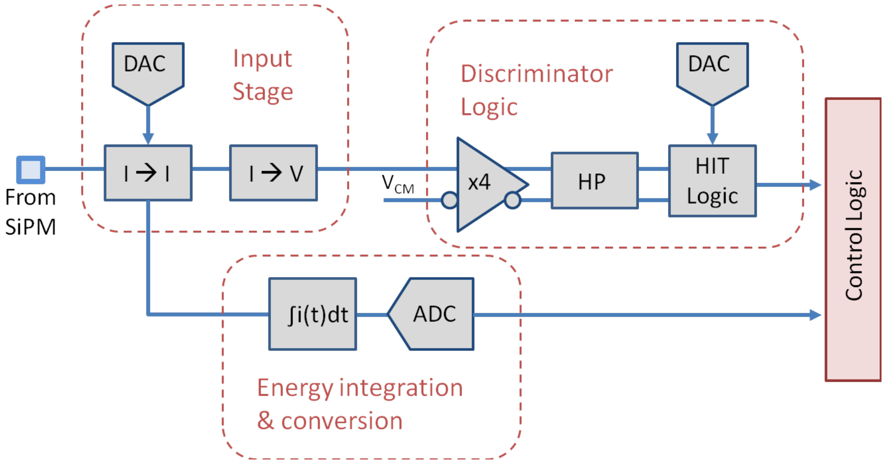


Figure 6.1.: Single-ended channel schematic view: SiPM signal is amplified and converted into a voltage pulse by the input stage. The pulse is used by the differential discriminator logic for defining the trigger. The amplified current signal is integrated onto a capacitor and digitized.

The proposed architecture is designed for reading signals from $n-on-p$ diodes, with negative input pulse. This design decision is justified by the choice of reading out SiPMs produced with RGB technology by FBK [56].

6.1.1 *Input stage*

The front-end architectures taken into account during development phase have been integrated on the test chip ISIS1 [115]. The input stages are all based on the same amplifier concept and have been optimized and simulated for fast and precise hit generation. Measurements on a dedicated test setup have been used for the selection of the most performing ssolution, to be integrated on PETA ASIC.

A brief overview on the considered architectures is reported for completeness in the next paragraph. Details on all design options are not included here: only the chosen architecture is fully described in this section. More information on ISIS1 chip concept can be found in Section 6.2.4.

DESIGN OPTIONS: OVERVIEW Three designs have been investigated. Their simplified schematic and proprieties are reported in Table 6.1. The architectures are based on a grounded-gate stage: additional features (green parts in the figures) have been added to the basic schematic, in order to match the requirements of the application. Results reported in the table are obtained with simulated design, using the same SiPM model for all of them, with a total load capacitance of about 40 pC.

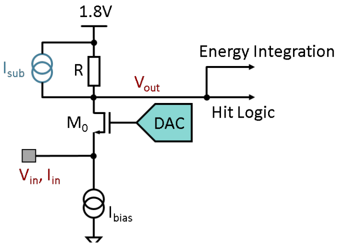
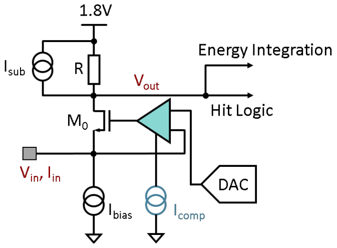
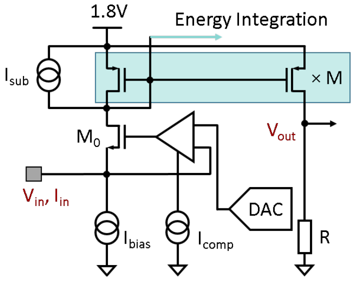
Design	Proprieties	Comments
Common-gate Amplifier  1.8V I_{sub} R Energy Integration V_{out} Hit Logic M_0 DAC V_{in}, I_{in} I_{bias}	$R_{in} \approx 25-40\,\Omega$ $I_{tot} = 3.5\,mA$ $\omega_{LP} \approx 150-250\,MHz$ $G_{I\to V} = 100$	V_{out} depends on I_{sub}; V_{in} set with M_0 gate voltage; Additional $V \to I$ conversion to energy path
Regulated Amplifier  1.8V Energy Integration I_{sub} R V_{out} Hit Logic M_0 V_{in}, I_{in} DAC I_{bias} I_{comp}	$R_{in} \approx 3\,\Omega$ $I_{tot} = 6-7\,mA$ $\omega_{LP} \approx 650\,MHz$ $G_{I\to V} = 100$	V_{in} directly set with DAC; one more current source
Regulated Amplifier with Mirror  1.8V Energy Integration I_{sub} × M M_0 V_{out} V_{in}, I_{in} DAC I_{bias} I_{comp} R	$R_{in} \approx 3\,\Omega$ $I_{tot} = 7-8\,mA$ $\omega_{LP} \approx 650\,MHz$ $G_{I\to V} = M \times 100$	Larger dynamic range than previous cases; current injected into energy path

Table 6.1.: Input stage options designed in test chip ISIS1. Results obtained with simulated circuit: R_{in} is the input impedance, I_{tot} the total current consumption for the stage, ω_{LP} the low-pass cutoff frequency, and G the gain of the current to voltage conversion.

For a common grounded-gate stage, the input impedance R_{in} is limited to $\approx 1/g_m$, where g_m is the transconductance of the *nmos* cascode M_0 and is defined as:

$$g_m = \frac{\partial I_D}{\partial V_{gs}} \tag{6.1}$$

I_D is the current at the drain node and V_{gs} is the gate-source voltage difference. Assuming that M_0 is working in saturation (true for $V_{gs} > V_{th}$ and $V_{ds} \geq \left(V_{gs} - V_{th}\right)$, where V_{th} is the threshold voltage and V_{ds} the drain-source voltage) and without considering the modulation length effect in first approximation, I_D is equal to:

$$I_D \approx \frac{\mu C_{ox}}{2} \frac{W}{L} \left(V_{gs} - V_{th}\right)^2 \tag{6.2}$$

where μ is the mobility of the main carriers, C_{ox} is the gate capacitance per unit area, W is the gate width and L its length. Using the previous equations, it is easy to find the dependance of g_m on the drain current:

$$g_m \approx \mu C_{ox} \frac{W}{L} \left(V_{gs} - V_{th}\right) \approx \frac{2I_D}{\left(V_{gs} - V_{th}\right)} \tag{6.3}$$

In order to achieve the lowest possible input impedance for such architecture, a wide gate transistor is necessary, combined with large drain current and relative high power consumption. The fundamental limit for this stage is the behavior of the *nmos* in weak inversion area, where equation 6.2 is not valid anymore and g_m decreases.

As reported in Table 6.1, the designed cascode amplifier can achieve minimum impedance of about $25 - 40\,\Omega$, with a bias current in the order of $3 - 4\,mA$. The cascode M_0 is $400\,\mu m$ wide, which has been considered to be a good compromise between low input impedance, parasitic capacitance in the branch and layout constraints. In order to define the DC input voltage, directly connected to the SiPM cathode, the cascode gate is controlled by an internal DAC, whose design is described in details in the next paragraph. An additional current source, named in the figure as I_{sub} is used for the regulation of the output voltage DC level. This first design has a low bandpass cut off frequency, of about $150 - 200\,MHz$ and produces a voltage output which must be used for both the discriminator logic and the energy integration branch. The gain of the voltage conversion depends on the load resistor size (R in the figure) , which is $100\,\Omega$ in the submitted design. The energy is measured, in all architectures, via charge integration and therefore the signal must be converted into current again in order to be evaluated.

The other two designs use a differential amplifier which regulates the cascode gate voltage, so that the input DC voltage is kept at a reference value. The reference is produced by the same DAC used in the first architecture. The advantages of the regulation feedback, in terms of bandwidth and input impedance, are described in detail in the next paragraph.

In addition, the last proposed design uses a *pmos* load, instead of simple resistor, as termination of the input branch. This option offers two advantages: since the *pmos* has larger resistance, it allows a larger dynamic range. Additionally, the current signal can be mirrored on a parallel branch for the voltage conversion, whose gain can be adjusted with the mirror ratio M and the load resistor size ($100\,\Omega$ in the design). A similar mirror is used for the energy injection path, without the need of any additional conversion of voltage into current.

Regulated circuits reach in simulation very low input impedance, with a still reasonable power consumption. The small signal model showing the effects of the feedback loop is described in the next paragraph. The higher power consumption of the last proposed design takes into account the additional current flowing in the parallel branch used for voltage conversion. The regulated amplifier has larger bandwidth if compared with the simple grounded-gate stage, reaching in the optimized design a cut off frequency of about $650\,MHz$ in simulation, coupled with a SiPM with total capacitance of $40\,pF$.

Thanks to its good performances, obtained with relatively low current consumption, the last design has been selected for integration on the ASIC architecture. The circuitry blocks included in the channel are described in more details in the following sections. Characterization measurements of the front-end can be found in Appendix A.

REGULATED CASCODE AMPLIFIER WITH CURRENT MIRROR

The schematic of the selected input stage is shown in Fig.6.2. The figure reports also the corresponding small signal model used for estimating the circuit input impedance.

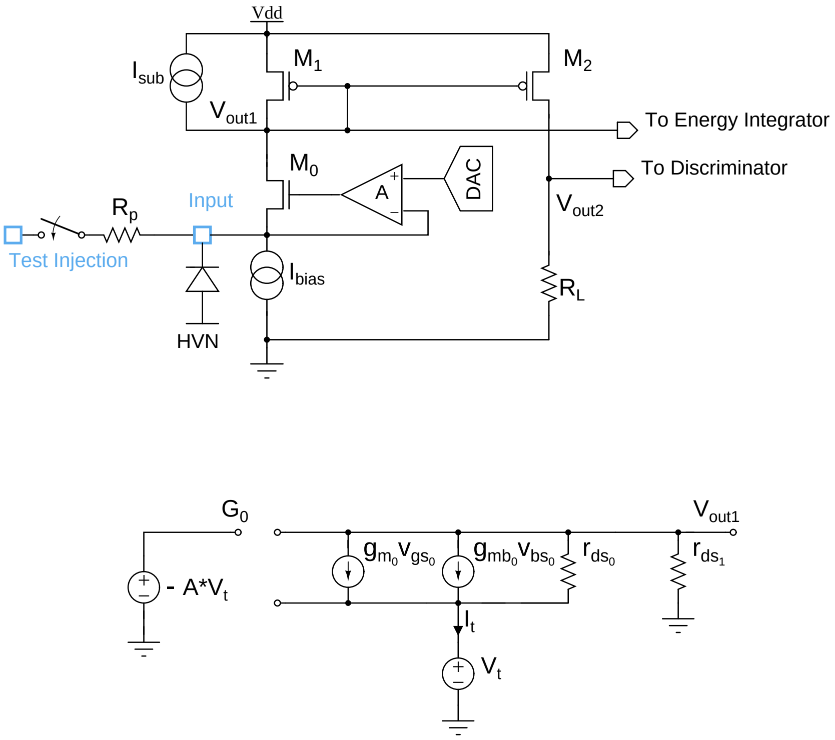

Vdd
I_{sub}
M_1
M_2
V_{out1}
To Energy Integrator
M_0
To Discriminator
A
DAC
R_p
Input
V_{out2}
Test Injection
I_{bias}
R_L
HVN
G_0
V_{out1}
$g_{m_0}V_{gs_0}$
$g_{mb_0}V_{bs_0}$
r_{ds_0}
r_{ds_1}
- A*V_t
I_t
V_t

Figure 6.2.: Schematic of the channel input stage and relative small signal model used for estimating the input impedance.

For the small signal circuit, the *pmos* M_1 has been considered as a simple resistor r_{ds_1}, used as a load at the output node. The voltage at M_0 gate, G_0 in the figure, is equal to $(-A \times V_t)$, where A is the amplifier gain and V_t is the voltage pulse at the input node, producing a current I_t [116] . Using Kirchoff's laws at the input node and considering M_0 bulk connected to ground, the following equations can be written:

$$\begin{cases} I_t = -g_{m_0} V_{gs_0} - g_{mb_0} V_{bs_0} + \frac{V_t}{r_{ds_0}} \\ V_{gs_0} = V_{G_0} - V_{S_0} = -AV_t - V_t \\ V_{bs_0} = V_{B_0} - V_{S_0} = 0 - V_t \end{cases} \tag{6.4}$$

Since the input impedance r_{in} can be written as $\frac{V_t}{I_t}$, the previous equation leads to:

$$r_{in} = \frac{1}{\left[\frac{1}{r_{ds_0}} + (g_{m_0}(1+A) + g_{mb_0})\right]} \approx \frac{1}{g_{m_0}(1+A)} \tag{6.5}$$

where it has been assumed that $r_{ds_0} \gg 1$ and $g_{mb_0} \ll g_{m_0}$. The introduction of the feedback loop has the great advantage to reduce the input impedance of a factor $\approx A$, for values of $A \gg 1$. From the simulated results, the input impedance is reduced of a factor about 10, compared with the respective value obtained with non-regulated cascode, paid with almost double current consumption.

The regulation amplifier is a standard gain stage with differential inputs and single ended output: the schematic is shown in Fig.6.3. The two inputs are connected to the DC input node and to the reference voltage defined with the internal DAC. In the design, M_3 is identical to M_5 (working in saturation) and M_4 matches M_6. In idle condition, the equation $\Delta V_{in} = 0$ is true and $I_{D_3} = I_{D_4} = I_{D_5} = I_{D_6} = \frac{I_{ampl}}{2}$. When a small change v_{in} is added to the differential input nodes ΔV_{in}, the drain currents at M_3 and M_5 change in opposite directions around the

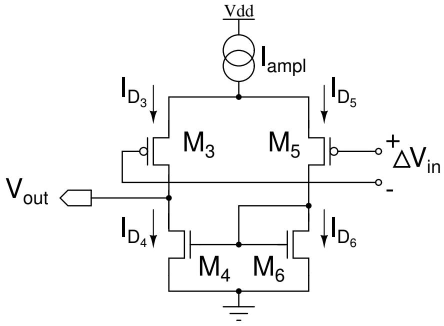


Figure 6.3.: Gain stage schematic with differential inputs and single-ended output.

idle value. Assuming a perfect match between M_3 and M_5, it can be considered that $g_{m_3} = g_{m_5}$ and that:

$$\begin{cases} I_{D_3} = \frac{I_{ampl}}{2} + i_{d_3} = \frac{I_{ampl}}{2} + g_{m_3}\frac{v_{in}}{2} \\ I_{D_5} = \frac{I_{ampl}}{2} + i_{d_5} = \frac{I_{ampl}}{2} - g_{m_3}\frac{v_{in}}{2} \end{cases} \tag{6.6}$$

On the right branch of the circuit, it is still true that $I_{D_5} = I_{D_6}$ and, thanks to the current mirror, $I_{D_4} = I_{D_6}$. At the output node, the current mismatch produces a voltage change, which can be written as a function of v_{in} using Eq.6.6:

$$v_{out} = (i_{d_3} - i_{d_5})r_{out} = g_{m_3} \cdot v_{in} \cdot r_{out} \tag{6.7}$$

where r_{out} is the output impedance of the circuit, assumed to be purely resistive, and is given by the drain-source resistance of M_3 and M_4. The amplifier gain is then equal to:

$$A = \frac{v_{out}}{v_{in}} = g_{m_3} \cdot r_{out} = g_{m3}(r_{ds_3} \| r_{ds_4}) \tag{6.8}$$

Since the drain-source resistance is defined as $r_{ds} = \frac{\partial v_{ds}}{\partial i_{ds}} = \frac{1}{g_{ds}}$, where g_{ds} is the common source output conductance, the previous equation can also be written as:

$$A = \frac{g_{m3}}{g_{ds_3} + g_{ds_4}} \tag{6.9}$$

Assuming that both M_3 and M_4 are in saturation, terms in equation 6.9 can be replaced with their large signal approximation:

$$A = \frac{\sqrt{2\mu C_{ox}\frac{W_3}{L_3}\frac{I_{ampl}}{2}}}{(\lambda_3 + \lambda_4)\frac{I_{ampl}}{2}} \propto \frac{1}{\sqrt{I_{ampl}}} \tag{6.10}$$

For g_{m_3} the equation 6.3 has been used, while g_{ds} has been replaced by its approximation $(\lambda \cdot I_D)$, where λ is the channel length modulation parameter. Once the transistor size has been defined, the contribution of the current goes actually versus a reduction of gain. Combining equations 6.5 and 6.9, the input impedance of the whole regulated stage can be written as:

$$r_{in} \approx \frac{1}{g_{m_0}\left(1 + \frac{g_{m_3}}{g_{ds_3} + g_{ds_4}}\right)} \tag{6.11}$$

The amplifier current affects at the same time the frequency response of the gain stage. In particular, the low pass cut off frequency is equal to:

$$\omega_{LP} = \frac{1}{r_{out}C_{out}} \approx \frac{g_{ds_3} + g_{ds_4}}{C_{load} + C_p} \tag{6.12}$$

where C_{load} is the capacitive load at the output node. C_p is the total parasitic capacitance seen by the circuit at the output node and is given by $\left(c_{bd_3} + c_{bd_4} + c_{gd_2}\right)$ where $c_{bd_{3,4}}$ are the bulk-drain capacitance for $M3$ and $M4$ and c_{gd_2} is the gate-drain capacitance for $M2$. As reported previously, the values of $g_{ds_{3,4}}$ are linearly dependent on the amplifier bias current.

The effects of the input stage bias current and of the amplifier current (I_{bias} and I_{ampl} respectively) are reported in Fig.6.4. Simulations show well how the input stage bias current has larger effects on the input impedance, while the amplifier bias affects at most the cut-off frequency, with small effects on the impedance value at low frequency. A proper compromise must be found during operation, so that low input impedance, sufficient bandwidth and reasonable current consumption can be achieved.

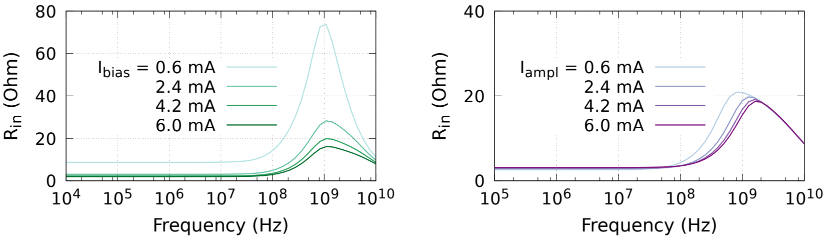


Figure 6.4.: Input impedance frequency behavior simulated for different bias currents (I_{bias} and I_{ampl} as defined in Fig.6.2 and 6.3). On the left, $I_{ampl} = 4.8\,mA$; on the right $I_{bias} = 1.5\,mA$.

The current is converted on a mirror branch into voltage on a resistor: the conversion gain depends on the ratio (W_2L_1/L_2W_1) and on the load resistor R_L, as shown in Fig.6.2. Using the SiPM model described in section 3.2, the conversion gain has been chosen so that signal coming from few cells operating at $4\,V$ over breakdown could trigger the discriminator logic with a differential threshold set to about $20\,mV$ and with only 2 stages of amplification (the discriminator circuit is described in 6.1.2). A gain of $200\,V/A$, obtained with a ratio of 2 and a $100\,\Omega$ resistor has been designed.

The DAC employed for the definition of the input reference voltage has been implemented with an output range of about $1.6\,V$, for being used in both the non-regulated and the regulated input stages within the test chip (see Table 6.1). The schematic of the circuit is reported in Fig.6.5. On a multiple-channel design, the left branch defines the gate bias values and can be shared between many channels. The right branch is included in each channel: a parallel connection of 32 current sources, biased with the common reference voltages, is used for inducing a voltage drop onto a $100\,k\Omega$ resistor (R_{Load} in the figure). The unit current source is dimensioned to be $500\,nA$, so that a bin width of $50\,mV$ is achieved. In order to maximize the output resistance, a cascode design has been implemented. A sufficiently high bias at the gate voltage of cascode M_9 ensures the saturation of the branch. The DAC output varies from about $200\,mV$ up to the bias voltage $1.8\,V$.

Simulated results for the output voltage are reported in Fig.6.5-right, for all possible DAC settings, showing the good linearity of the design.

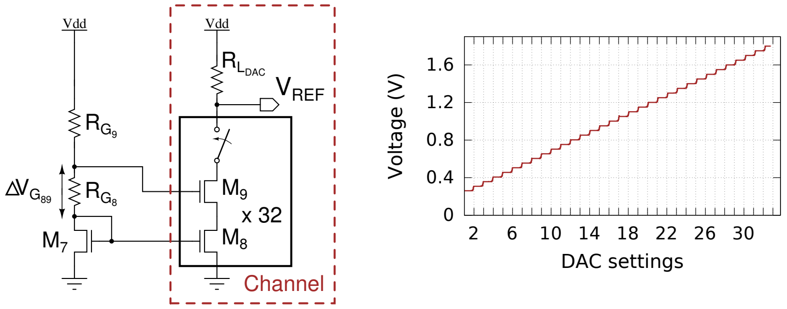


Figure 6.5.: Schematic view of the DAC defining the input stage DC point and simulated results of the output voltage for all possible settings.

The reference value, used within the regulation loop for setting the input stage DC level, must be chosen properly so that cascode M_0 works in saturation region, where its transconductance is at maximum. For too high values of the DC input, the transistor reaches its weak inversion region and the input impedance of the whole stage gets larger. The design has been optimized for a maximum voltage swing at the input node of about $0.8\,V$, keeping the input impedance lower than $20\,\Omega$, with a minimum value of about $3\,\Omega$. The simulated results for the input resistance and the DC voltage, for different values of the DAC output voltage, are shown in Fig.6.6. Considering that the technology operates at only $1.8\,V$, a swing of about half the supply range is a very good compromise and can be used for adjusting SiPM gain on a per channel basis, without degradation of the readout performance.

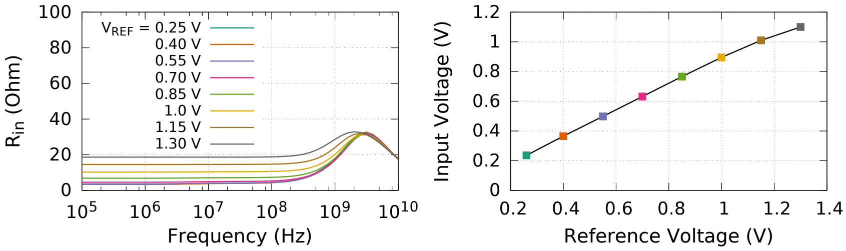


Figure 6.6.: Simulated input impedance and DC level for different reference DAC settings.

For testing purposes, each channel includes also a current injection path, useful for characterization measurements. Only one pad is typically available for many channels for the analogue injection. Each channel is connected to it via a $1\,k\Omega$ resistor and the path can be enabled with an *nmos* switch with about $40\,\Omega$ resistance. A voltage pulse injected into the pad introduces a current signal, with amplitude of about $V_{pulse}/1k\Omega$. Since the DC input voltage of the front-end is kept at a fixed value by the regulation loop, the pulse is DC-decoupled with a capacitor placed on the readout PCB. In each designed chip, the input pads are always connected with protection diodes, which do not allow the injection of voltage pulses larger than about $700\,mV$ below/above negative/positive power supply limits respectively. Naming V_{in} the DC input level of a channel, the maximum current that can be linearly injected through the test path is:

$$I_{max} = \frac{V_{in} + 700\,mV}{1\,K\Omega} \tag{6.13}$$

With the present design, maximum injected current of $1\,mA$ has been considered sufficient during testing phase.

6.1.2 *Discriminator logic*

The discrimination logic generates a hit flag if the input signal is above a certain threshold. This part of the channel is entirely realized with differential logic, for minimizing noise contribution. Its schematic circuit is shown in Fig.6.7.

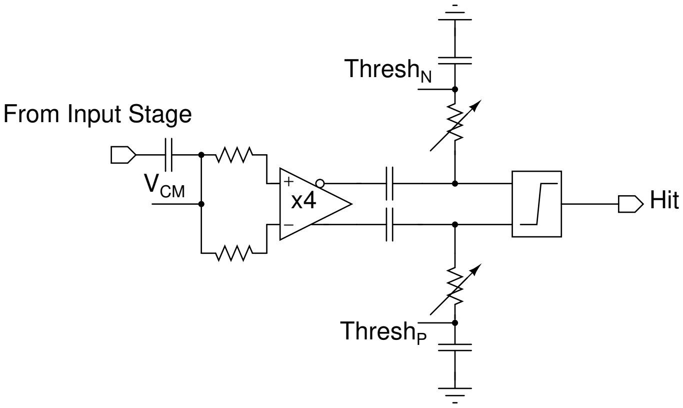


Figure 6.7.: Discriminator logic circuit: input signal is mirrored on a differential logic, amplified and filtered. The output is a pulse as wide as the crossing time of the opposite pulses.

The positive voltage pulse from the input front-end gets at first decoupled with a series capacitor. DC input level is defined with a voltage common mode, provided usually from an input bias pad, set at $1\,V$ and used as the negative input of the first amplifier stage. The amplification chain is composed of two identical modules, each of them with gain of 2. No low-frequency feedback loop is necessary for baseline correction.

The amplified pulses are decoupled with series capacitors: the DC voltages at the discriminator inputs are set to the threshold values ($Thresh_N$ and $Thresh_P$ in the figure). The discriminator logic produces a full CMOS pulse, when the differential signals cross. Its width is equivalent to the crossing time of the differential pulses. The signal is fed into a fast hit logic latch (not shown in the figure), whose output is used for triggering the timing information with the TDC and for starting the energy integration procedure.

A high-pass programmable filter is introduced for reducing the baseline contribution in the discriminator. Four programmable values are selectable: the cutoff frequency spans from about 15 to 80 MHz. Simulated effects on the bandpass frequency at the discriminator input are reported in Fig.6.8, for the positive signal of the differential pair. In the figure, together with filter effects, it is reported the original transfer function of the input stage with no SiPM connected.

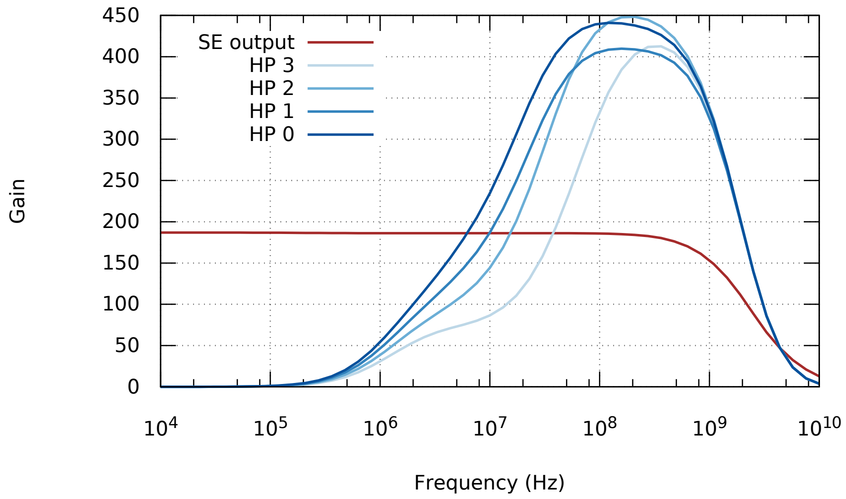


Figure 6.8.: Simulated frequency response of the positive signal at the discriminator input: effects of the programmable high-pass filter are shown, together with the transfer function of the single-ended input stage.

Threshold values used within the discriminator are generated with the circuit drawn in Fig.6.9. Such circuit is included in each channel and uses a current DAC for threshold definition. The voltage reference values visible in the figure (Ref_N and Ref_P) are common for many channels and in PETA are defined internally using a simple resistor divider. The difference between references defined internally is 100 mV,

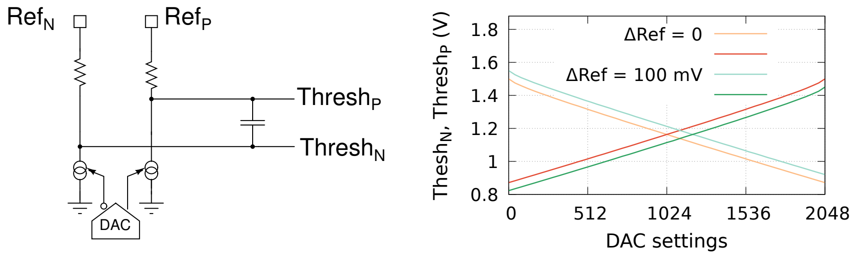


Figure 6.9.: Schematic of the circuit for threshold definition: a DAC defines complementary currents in the two branches which cause voltage drop at the resistors. Simulated results, for two values of reference difference ΔRef and for the entire DAC range are shown on the right.

but it can be over-written via two input pads in all PETA versions. Two complementary currents, defined by an internal 11-bit DAC integrated in each channel, flow onto the $100\,k\Omega$ resistors and generate a voltage drop at the circuit outputs. The maximum available current on both output together is $6\,\mu A$, with a LSB of about $3\,nA$. This leads to a maximum voltage sweep at the circuit outputs of about $600\,mV$, to be added to the external differential threshold.

The estimated voltage threshold is defined as:

$$\Delta Thresh = 2 \cdot LSB \cdot DAC \cdot 100\,k\Omega + \Delta Ref \qquad (6.14)$$

Simulated results for $Thresh_N$ and $Thresh_P$ are reported in Fig.6.9-right, for two different settings of the reference voltages and for the whole DAC range.

6.1.3 *Charge integrator*

Energy is measured as integrated charge delivered by the sensor in a certain time window. The amplified signal from the input stage is mirrored for injection into the circuit and integrated on a capacitor, with a standard amplifier with capacitive feedback. The charge stored is digitized measuring the time necessary for completely discharging the capacitor with a fixed current.

A schematic view of the designed circuit is shown in Fig.6.10. The input signal is mirrored into the integration path through parallel connection of 32 *pmos* transistors (together named M_1 in the figure). The ratio can be selected independently for every channel on a chip, between 1/640 and 1/20. Such wide range has been chosen for being compatible with detector operation in many configurations.

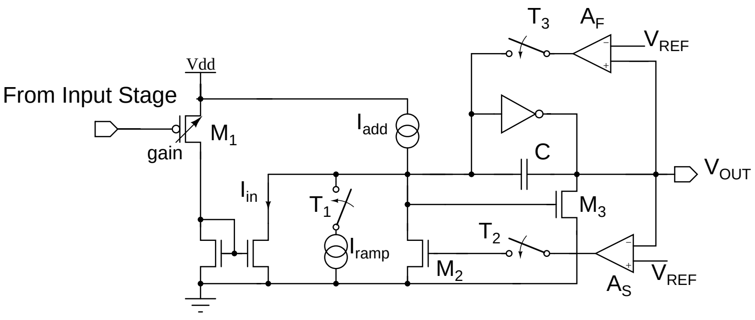


Figure 6.10.: Simplified view of the energy integrator circuit: switches state does not correspond to any operational condition.

Two feedback loops are necessary for operating the circuit: a fast differential operational amplifier A_F is used for defining the voltage at the capacitor input node, in such a way that the output voltage is kept fixed at a reference value. The reference voltage V_{REF} is defined on PETA with an internal 8-bits DAC, in a range between 100 and 600 mV. The second feedback loop controls, with the same input signals, the gate voltage of a wide *nmos* (M_2 in the figure) used for current compensation. The slow control is necessary for filtering low frequency fluctuations and keeps the circuit stable. The *nmos* M_2 is wide enough to compensate part of the current flowing from the source I_{add}.

The charge integration is divided into three phases, which can be summarized as follows (switch names referred to Fig.6.10):

- idle state (T_1, T_2, T_3 closed): the capacitor output voltage is kept at fix value by the feedback loops. The current generated from source I_{add} flows in sink I_{ramp} and through M_2. It also compensates for baseline current mirrored from input stage;

- hit integration (T_1 closed, T_2, T_3 open): when the hit logic detects a hit, the feedback loops are disabled. Current through M_2 is constant, because its gate is maintained for long time at the original bias thanks to its large channel length. Equilibrium between currents through M_2, I_{add} and I_{ramp} is still valid and the input signal gets integrated on the capacitor. The output voltage rises until the signal is over;

- capacitor discharge (T_1, T_2, T_3 open): after a programmable waiting time, switch T_1 is opened by the control logic. Source I_{add} produces now an excess of current not compensated by the circuit and equivalent to I_{ramp}, which discharges the capacitor. The value of I_{ramp} is programmable within the chip. Once the output voltage reaches again the reference value, feedback loops are enabled and the integrator circuit is back to its idle state.

When the waiting time is over, the integrator output voltage depends on the input charge following the equation:

$$V_Q(T_{wait}) = V_{REF} + \int_{T_{hit}}^{T_{wait}} i_C(t) C \, dt \tag{6.15}$$

where T_{hit} is the trigger time defined by the hit logic and T_{wait} is the programmable waiting time before starting with the discharge. The equation 6.15 is true only if compensation and discharge current sources (respectively I_{add} and I_{ramp} in the figure) are properly dimensioned. In that case, current i_C is proportional to the sensor signal. Wrong circuit setting results in an integrated baseline, which is constantly added to the input signal when feedback loops are disabled. Even if baseline can be corrected for and does not affect the energy resolution, it adds dead time to the channel because it contributes to longer discharge times.

Time necessary to the output voltage for reaching again the reference value is used as measure of the integrated charge. With constant discharge current I_{ramp} and assuming that the pulse integration on the capacitor is over, it can be written that:

$$V_Q(T_{wait}) - I_{ramp} T_{discharge} C = V_{REF} \tag{6.16}$$

Using equations 6.15 and 6.16, the discharge time, depending on input and discharge current is deduced:

$$T_{discharge} = \frac{\int_{T_{hit}}^{T_{wait}} i_C(t)}{I_{ramp}} \tag{6.17}$$

The programmable value T_{wait} can be set in PETA from 0 up to about $200\,ns$. This range has been chosen because it is large enough to cover typical scintillator signals. Depending on the selected waiting time, circuit can work in two conditions, which are shown in Fig.6.11 for a step current pulse. For step input current, as in the figure, different waiting times have no effect on final energy estimation. If discharge starts when the pulse is still active (Fig.6.11-left), the integrated current $i_C(t)$ is equal to the input current i_{in} only until the discharge starts. After that, the current injected onto the capacitor is equal to $i_{in} - I_{ramp}$ (it works for reasonably small values of I_{ramp}), until the input pulse is over and discharge follows as normal. If waiting time is larger than the pulse width (Fig.6.11-right), once the input signal is over, the capacitor output voltage stays constant till the discharge starts. De-integration time is identical in both case. For real SiPM signals the two conditions are not identical anymore, because the current signal is not constant. Starting the discharge once the input charge has been completely collected on the capacitor ensures correct measurements, but it requires waiting time of about 3 times the scintillator decay constant. In case of LYSO crystals, with decay time of about $40\,ns$, a reasonable setting is waiting time of about $150\,ns$. A strategy for reducing channel dead time consists in using shorter waiting time, in the order of decay time or slightly more: an error is introduced in Eq.6.17, because I_{ramp} has to be summed with the input current. The error is more important

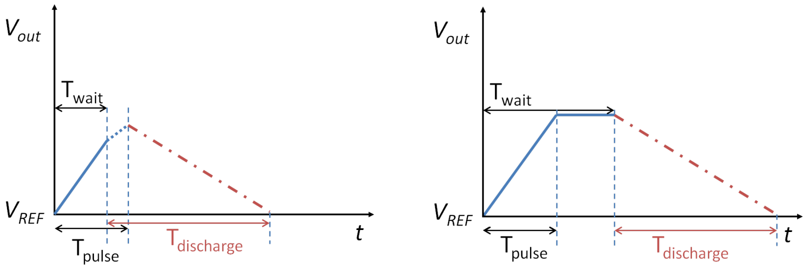


Figure 6.11.: Effects of different waiting times before discharge starts for a step current input pulse. On the left, discharge starts while the signal is still getting integrated on the capacitor; on the right, well after the integration is over. The discharge time is the same for both cases, thanks to the linearity of the signal and of the capacitor.

for low energy events, which in most applications, such 1:1 coupling between crystals and sensors, are considered not interesting.

The fast feedback loop generates a full CMOS signal, result of the comparison between the capacitor output and the reference voltage. Such value is used by the digital logic in the channel for controlling the switches in the circuit and defined when the discharge is over. The effects of the feedback on the discharge time are shown in Fig.6.12, for different feedback bias currents. A fast loop ensures an earlier detection of the capacitor condition.

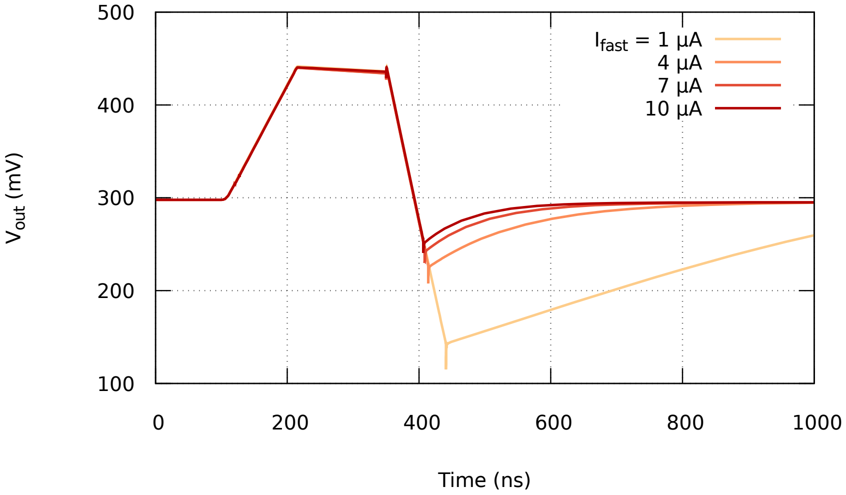


Figure 6.12.: Transient response for different values of the fast amplifier bias current. Effects on the discharge time are visible.

The whole energy integration circuit needs about $200\,\mu A$ for operating. The discharge time can be adjusted changing the discharge current I_{ramp} and the variable gain at the injection mirror. In all PETA versions, time is estimated with a 9-bit counter with bin-width of $3.2\,ns$. Assuming that energy resolution dE/E_{511} is at best 10% at FWHM for the $511\,keV$ photo-peak, standard deviation at the same energy is equal to:

$$\sigma_{511} = \frac{FWHM_{511}}{\sqrt{8\ln 2}} \tag{6.18}$$

The intrinsic binning noise of the ADC is defined as $\sigma_{ADC} = LSB/\sqrt{12}$. Requiring that $\sigma_{511} = 12 \cdot \sigma_{ADC}$, LSB can be written as a fraction of the measured photo-peak energy:

$$LSB = \frac{\sqrt{12} \cdot FWHM_{511}}{12 \cdot \sqrt{8\ln 2}} \approx 0.12 \cdot FWHM_{511} = 1.2\% \tag{6.19}$$

From Eq.6.19, ADC setting should be adjusted so that the photo-peak is detected at about $(1/0.012)\,LSB \approx 70\,LSB$. In this case, the most common events within the photo-peak would require about $220\,ns$ for completing the digitization of the energy measurement.

6.2 ASIC DESIGN

PETA (Position Energy Timing ASIC) is a chip dedicated to SiPM readout, whose first version is available since 2006. All chip versions starting from PETA4 integrate the single-ended design and are the only ones taken into account in this dissertation, when not explicitly declared.

The ASIC is designed in UMC $180\,nm$ technology with six metal layers, on an area of $5\,mm \times 5\,mm$ and with solder ball connection. A picture of PETA4 is shown in Fig.6.13-left. Pitch between bumps is about $272\,\mu m$. The chip is structured in two identical halves: each of them includes 18 readout channels (see Fig.6.13-right). The halves are independent from each other, with separate phase-locked loop (PLL) circuits and readout logic. The central area of the layout is reserved to the common bias based on 11-bits current DACs. The main blocks shown in the figure are described in more details in the following paragraphs.

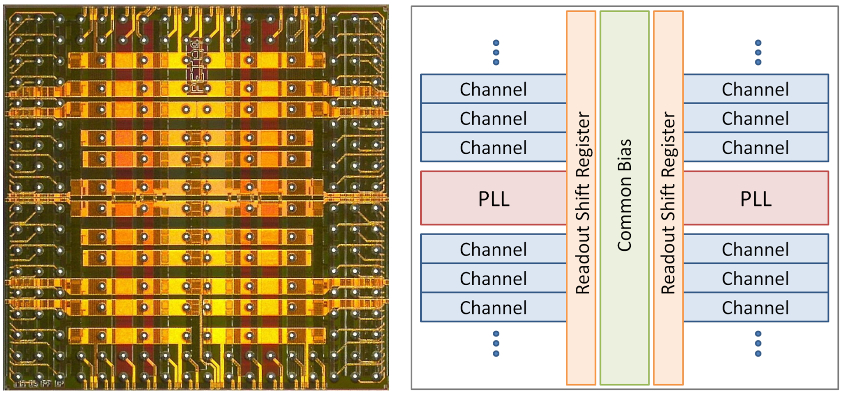


Figure 6.13.: Picture of PETA4 chip with bumps (left) and block architecture scheme (left).

The chip layout has been carefully studied for avoiding any loop in traces, which would cause noise pickup from the MR gradient field.

6.2.1 *Channel architecture*

Each channel is composed of two front-end options. The novel single-ended front-end is integrated in parallel with a differential-ended stage, which has been also included in previous PETA versions, where it has achieved ToF performance [117]. The two designs share part of the bias lines. The three input pads necessary per channel cannot be connected simultaneously on the designed substrate, due to the extremely small pitch. For this reason, two versions of the module, for differential and single-ended readout, have been designed during the project.

The analogue front-end, in both cases, takes care of a first amplification of the input signal (current amplifier in the single-ended, voltage amplifier for the differential-ended). The amplified signal is used in both channel by the discriminator logic and within the integrator. The block architecture of a whole channel is shown in Fig.6.14.

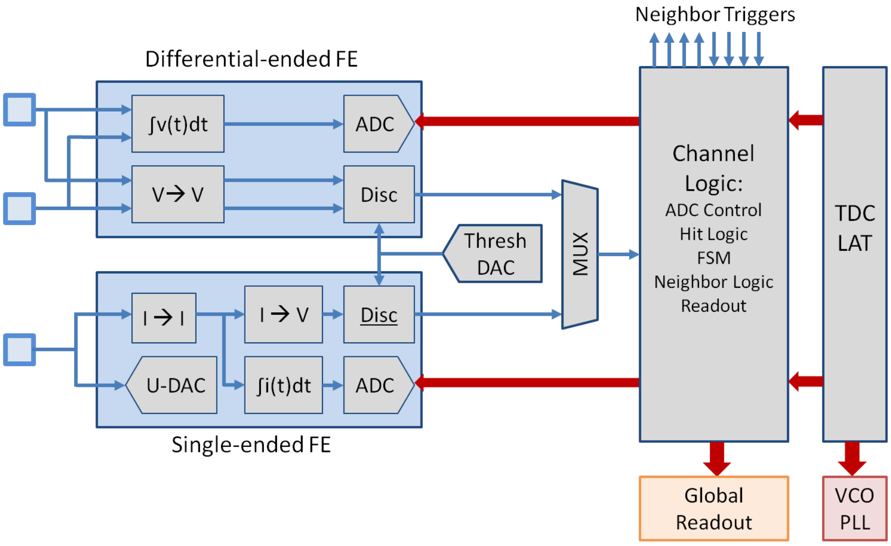


Figure 6.14.: Block architecture of PETA channel.

The differential-ended channel needs decoupling components on the substrate board: the SiPM signal is connected to one of the inputs, depending on the diode polarity. The opposite input is usually tied to ground, if SiPMs cannot be read out fully differentially. The voltage signal is amplified with a 5-stage amplifier, each block with a gain of 2. A feedback loop is included for offset regulation. The amplified signal is used for the generation of the hit pulse with a circuit identical to the one described in 6.1.2 for the single-ended channel. The signal integration is performed also in this case by an amplifier with capacitive feedback, enabled only when a hit is detected by the discriminator logic. A second amplifier calculates the difference of the integrator voltage before and after integration and converts it into current. Such current is then used in a comparator, for the digital conversion of the signal with a SAR-ADC (Successive Approximation Register ADC), with programmable accuracy from 9 to 11 bits. Once conversion is over, the integrator is reset. The whole channel design is fully differential in this case: more details on its architecture can be found in [118].

The hit signal is used for latching the event timestamp. Each channel includes a synthesized logic block, clocked at 312 MHz, which takes care of hit definition and controls veto mechanism and readout. In typical applications in 1:1 configuration between crystals and SiPMs, the trigger threshold is set at very low values for the best timing resolution, but low energy events are also not interesting for the analysis and should be discarded during readout in order to reduce dead time. The channel logic takes into account different constraints for the final decision on the event:

- Energy threshold: after digitization of the integrated energy, the result is compared with a custom defined minimum value for acceptance of the event. This threshold mechanism is included for both front-end options. In the single-ended channel, it can be used only at the end of the integration and conversion procedure, so that no dead time in channel is saved but only readout bandwidth. In the differential front-end, it is included the possibility to enable an early comparison, which happens before the conversion and avoids useless dead time.
- Time over threshold (ToT): in order to avoid the analogue-to-digital conversion for low energy signals, a ToT veto mechanism is available. It is integrated only as an option for the differential-ended channel, where the SAR-ADC long conversion time can have a large impact on channel dead time. For the single-ended version it would give no contribution, since the charge integrator must be discharged in any case and short discharge time for low energy signals is not considered as a main limitation.
- Neighbor logic: some applications require that events with low energy get read out and included in data analysis, even if they do not trigger the hit logic. In PETA, channels assigned to the same neighbor group are all read out if at least one of them is defined as hit, thanks to an additional mechanism in the logic. Neighbor logic is fully programmable: maximum four groups of neighbors can be assigned per half chip.

Channel power consumption is in the order of $35/8\,mW$ when operating the single-ended front-end, for analogue/digital blocks respectively. Differential analogue front-end consumes about $16/8\,mW$ instead.

6.2.2 *Phase-locked loop and TDC design*

Two Phase-Locked Loop circuits (PLL) are integrated within the chip, one per side. They are necessary for stabilizing the oscillating frequency of a 16-stage voltage-controlled oscillator (VCO). The fully differential design is a standard second order PLL, optimized for a reference frequency of $625\,MHz$.

A simplified schematic view of the circuit, not showing the differential architecture, is in Fig.6.15

The VCO is composed of 16 differential inverters and the design has been optimized for a delay at each stage of about $2 \times 16/625\,MHz = 50\,ps$. It can be adjusted changing the inverter bias current: the delay time is defined as the time necessary to charge the buffer output capacitance at the voltage output and is inversely proportional to the bias current. The oscillation can be obtained with an even number of stages simply switching the differential lines in one position. More details on the VCO implementation can be found in [118].

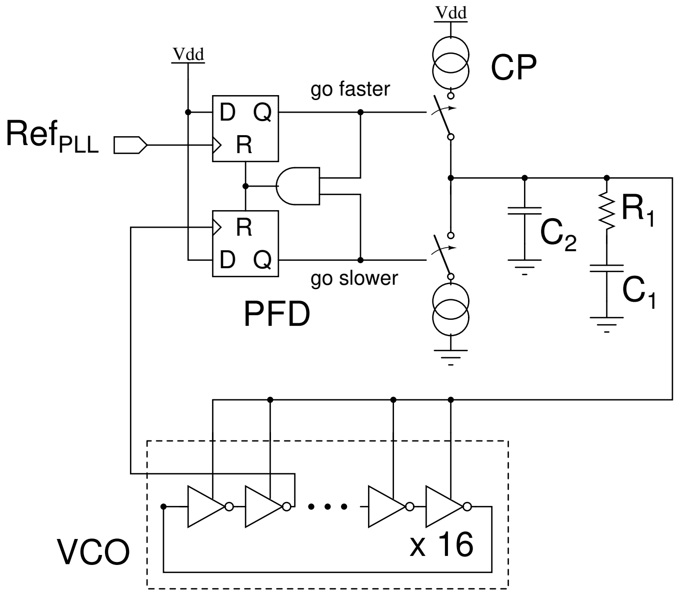


Figure 6.15.: Schematic view of the PLL blocks.

Within the PLL circuit, the VCO oscillating frequency is compared with the input reference in a phase-frequency detector block (PFD). The PFD detects the phase difference between its two inputs and generates two outputs, corresponding to a positive or negative difference respectively. The pulses are proportional to the input shift and control the charge pump (CP) injection. CP switches are complementary and only one of them can be closed at time. The current generated within CP increases/decreases the output voltage of the C_2 capacitor, which controls the VCO ring delay. Together with C_2, the RC filter $(C_1, R_1$ in the figure) defines the frequency behavior of the system.

PLL and VCO designs have already proved good performance in previous PETA versions. In order to reduce the number of components necessary on the assembled board, the RC loop filter, which was external for PETA3, has been integrated within the design. Its effects on the loop frequency behavior are here briefly taken into account.

Considering the parallel connection of C_2 and $(C_1 + R_1)$, the filter transfer function $F(s)$ can be written as:

$$F(s) = \frac{1 + R_1 C_1 s}{s\,(C_1 + C_2 + sR_1C_1)} \tag{6.20}$$

Assuming that the contribution of VCO to the transfer function can be approximated as an integrator with gain K_{VCO}, the open loop transfer function is:

$$G(s) = K_{VCO}\, I_{CP} \left(\frac{1 + R_1 C_1}{s^2\,(C_1 + C_2 + sR_1C_1C_2)} \right) \tag{6.21}$$

The open loop has one stabilizing zero at $s_z = 1/\,(R_1C_1)$ and three poles, two of them at zero and the highest one at $s_{p_3} = 1/\,(R_1C_2)$, where it has been assumed that $C_2 \gg C_1$. They depend on the filter dimension and must be accurately chosen in order to operate the close loop in a stable region. Integrating the RC filter on the chip avoids to place up to 8 SMD components on the substrate (2 components per chip side and 2 per each differential signal). For being able to modify the zero and pole coordinates and ensures stability, the size of R_1 is programmable with an internal DAC, while C_1 and C_2 are kept at fixed values.

The stabilized frequency is used by the internal digital logic, which runs at half the locked frequency (about $312\,MHz$), and for timestamp assignment. Each event is assigned to a coarse and a fine timestamp: together, they fully characterize the event arrival time. Coarse and fine counters are defined per half chip: timestamps values are sent to all channels, so that when a hit is generated by the hit logic, the channel stores the time information in low power latches.

The time within the VCO is defined by the only buffer in the ring which is changing its status, while all the others are stable. For this reason, all the valid fine timestamps are encoded as a train of 0 and 1, where the transition position defines the arrival time. The decoding into a 5-bit value is performed on the chip: when all the states are 0 (or 1), it means that the last (or first) stage is about to change, corresponding to fine timestamps 0 (or 31).

The output from the eighth oscillator stage is buffered and used as main clock for the coarse counter, which determines the coarse timestamp, with bin-width of $1.6\,ns$, corresponding to a full VCO oscillation. For ensuring stable timestamp registration for any arrival time, two counters are actually included on the chip, switching their values on the positive (master counter) and on the negative clock edge (slave counter) respectively. In such a way, at least one of them is surely stable when the hit signal is triggered. The selection of the stable coarse timestamp to be read out is performed directly on-chip. It requires a short calibration, while both data are read out (debug readout mode) and compared with the distribution of the fine timestamp value. An example showing coarse and fine counter waveforms on a chip is reported in Fig.6.16. Due to the fact that the coarse counter is clocked on a buffered VCO output, and also to internal delays in the design, the fine timestamp period is not necessarily synchronized with the coarse counter clock. It can be divided into zones where master or slave counters are surely stable ($SelMaster$ and $SelSlave$ zones in the Fig.6.16). The calibration is used for estimating the shift between coarse and fine counter periods. The information is stored in a configuration bit and used during normal readout operation, for choosing the stable coarse timestamp depending on the fine one.

The coarse timestamp is encoded in 15 bits linear-feedback shift register (LFSR), for a total number of $\left(2^{15}-1\right)$ available states. A full coarse period takes then $\left(2^{15}-1\right)\cdot 1.6\,ns \approx 52.4\,\mu s$.

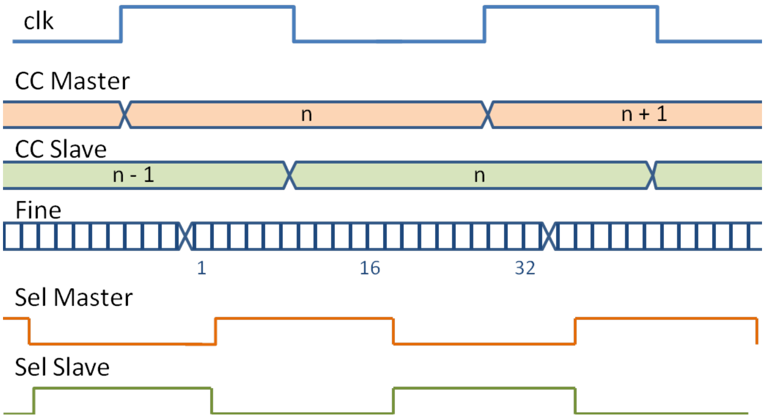


Figure 6.16.: Coarse counter selection principle: depending on the VCO oscillation (fine timestamp fo the detected event), the chip selects which coarse timestamp has a stable value at event arrival time.

6.2.3 *Readout logic*

Data are shifted out at the maximum design frequency of $160\,MHz$ ($320\,MHz$ in the last PETA6 version) through two registers, one per chip side, via LVDS lines. An additional shift register, used for debug data, can be read out via dedicated LVDS pads, or can be connected in chain to the main register and read out from the same line, in a so-called “debug readout mode”. The chip also generates an LVDS hit signal, which is OR-ed between all channels and can be used for triggering the readout procedure.

Each channel output register is composed of 37 (34 for PETA6) bits, both in main and in debug register. The main register contains, as first bit, the hit flag for the channel. The other 36 bits transmits event energy and timestamps, channel address and additional information on the hit type (such as “test hit” or “master hit”).

During normal readout operation, the full register per channel is shifted out. PETA includes also a zero-suppression (or “bypass”) feature, for larger readout bandwidth. A schematic view of the shift register logic for one channel is shown in Fig.6.17. In bypass mode, the hit bit is used to determine whether the full register must be read out or not. In such case, only the hit bit is loaded onto the shift register and the time necessary for complete readout of half chip is:

$$T_{bypass} = \frac{(N_{hits} \times 36 + 18 \times 1)}{f_{clk}} \tag{6.22}$$

where N_{hits} is the total number of hits detected to be read out and f_{clk} is the shift clock frequency. In normal readout mode, the time necessary for reading the shift register does not depends on N_{hits} and is equal to:

$$T_{NObypass} = \frac{(37 \times 18)}{f_{clk}} \tag{6.23}$$

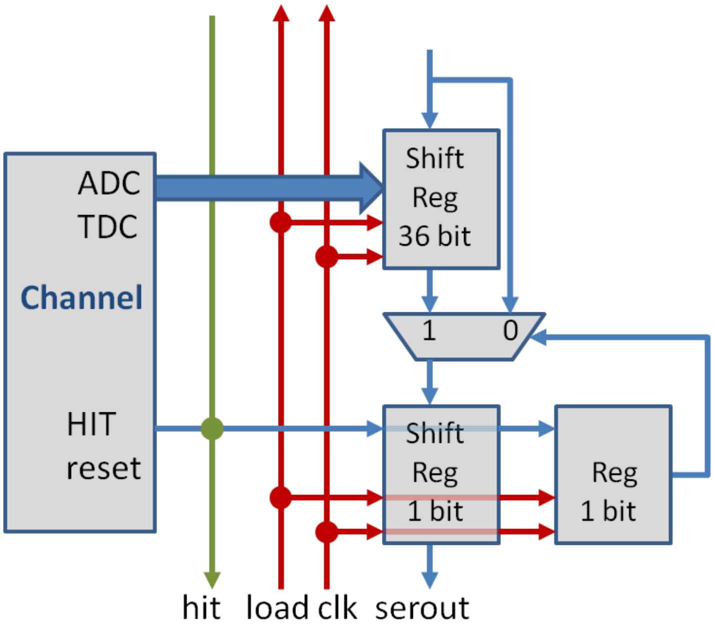


Figure 6.17.: Scheme of the channel readout architecture: control lines are red, common output green and local lines are blue. If working in bypass mode, the 36 bits of the main shift register are included in the shift chain only if an hit has been detected.

Only 5 LVDS signals are necessary for reading out data from the chip and they are all shared between left and right registers. Except for the load signal and the shift clock, a common reset signal for the internal state machine is necessary. Once data have been loaded onto the register, the hit logic is cleared with a dedicated signal, so that the channel is available again for the next event detection. Every chip version includes also an “inhibit” feature, which permanently suspends event detection if necessary.

6.2.4 *Chip versions*

Single-ended front-end design has been developed into a dedicated test chip and then integrated within three PETA versions until now. Except for minor bugs adjusted from one version to the other, the most significant modifications introduced in successive PETA ASICs are listed here for completeness. All results reported in the next chapters have been obtained with PETA5, when not explicitly specified.

ISIS1 (submitted in 2011) ISIS test chip has been used for the characterization of different front-end options, designed in UMC 180 nm technology, on a die of about $1.5\,mm \times 3\,mm$. Details on the front-end design strategy are reported in paragraph 6.1.1. A LC oscillator generates the internal clock. The output is encoded per channel with a single-wire protocol: the first fast rising edge is used for sampling timing information with an external TDC, while the pulse width encodes the integrated charge. Additional information, such as debug bits and channel address are sent out on the same line. The design includes 5 pairs of channels for coincidence measurements.

PETA4/5 (submitted respectively in 2012 and 2014) These two chip versions are very similar. PETA4 presents an error in the synthesized logic, preventing the use of the differential-ended channel, which cannot latch the TDC timestamp when triggering a hit. PETA5 main blocks have been described in the previous sections. It cannot operate bypass readout option when using the single-ended front-end.

PETA6 (submitted in 2015) This version solves some issues discovered in the previous design. The readout architecture has been changed for better performance at high data rate and reduced dead time. PETA6 has been designed in two versions, for opposite polarities of the single-ended front-end, in order to read out both n-on-p and p-on-n SiPMs. An additional version with fewer bumps and larger pitch ($500\mu m$) has also been submitted, for being used in dedicated test setups. A detailed comparison with PETA5 can be found in Chapter 14.

7

NOVEL PHOTON DETECTOR DESIGNS

SiPM performance are strongly dependent on their fabrication technology, especially in terms of noise, detection efficiency and temperature stability. Anyway, a careful design and adequate geometry can help to improve the light detection in a module.

The work has been focused on optimizing the SiPM die geometry, for making it compatible with the detector module design. Sensor based on a novel topology aimed to exceptional spatial resolution have been fabricated and tested. This chapter includes also a brief description on the design of a CMOS chip, with integrated SPAD array on the same silicon substrate, which stores the hit information per pixel and represents a powerful tool for studying light distribution of single events.

7.1 RECTANGULAR SIPM FOR MAXIMUM LIGHT COLLECTION

The detection module designed within Sublima had to be compatible with the HYPERImage backbone structure, where each stack was about $32\,mm \times 32\,mm$ in lateral dimensions. SiPM dies are typically built on an area of few mm^2, for facilitating fabrication process and keeping the overall capacitance within acceptable values. The placement of many dies on the module substrate is a tough challenge, because the area between sensors is insensitive to incoming photons and degrades the global efficiency. Gaps for wire-bonding connection must be large (in the order of hundreds μm) to allow the machine to reach the substrate.

The design of rectangular SiPM dies partly solves the problem: the sensor array can be very homogeneous with no wire-bonding gaps in between the active region, with optimal light collection.

For defining the SiPM size, simulation results by Thoen et.al [71] (described in Paragraph 4.2.1) have been taken into account. They have shown that the optimal pixel for whole-body PET rings, in case no DOI measurement is available but with ToF resolution, is about $2.6\,mm \times 2.6\,mm$.

For this reason, SiPM sensors with $2.5\,mm$ pitch have been organized on rectangular die, in a 6×2 configuration. The active area per SiPM is $2.25\,mm \times 2.25\,mm$ and the total die size is about $15\,mm \times 5\,mm$. The 12 output signals per die are available at the wire-bonding pads which are placed along one of the short edges. Picture of a rectangular die with 12 sensors is shown in Fig.7.1-left: wire-bonding pads are visible on the left edge. Tracks layout is modeled for reducing the resistivity of longer lines, so that all sensors on the same die have similar performance. A micro picture of the variable shaped tracks is reported in Fig.7.1-right.

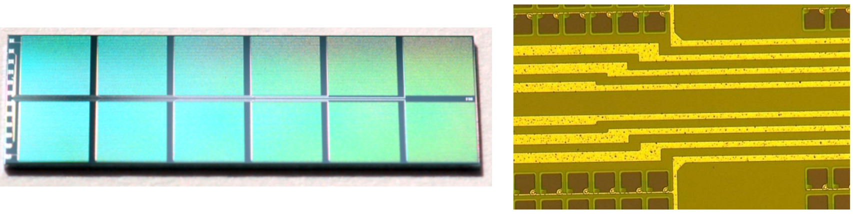


Figure 7.1.: Picture of a rectangular SiPM die with 12 sensors with pads at a short edge (left); micro picture of tracks with variable width (right), used for reducing the resistivity mismatch for sensors at different distance from the pads.

A detector module is fully populated with 12 rectangular dies, placed in two columns of 6 dies each. Bonding wires are only at two edges, with no large gap necessary in between the dies. Placing them at about $100\,\mu m$ distance, an homogenous pitch of $2.5\,mm$ on the whole array is achievable. More details on the final detector configuration can be found in Chapter 8.

SiPMs prototypes have been fabricated in FBK HD-RGB technology [57], with cell size of $25\,\mu m \times 25\,\mu m$.

7.2 INTERPOLATING SIPM

Interpolating SiPMs (ISiPMs) are devices aimed to reconstructing the center of gravity of small photon clusters with high accuracy. This characteristics is especially useful for identifying very small scintillating crystals coupled with the sensitive area and then ideally optimal for pre-clinical applications targeting good spatial resolution.

In the most common detector configuration for high spatial resolution, an array of small scintillating crystals is coupled with a sensor matrix through an interposed light guide. Light from a scintillator is spread on many sensors and the measured amplitudes are mathematically interpolated for reconstructing the center of gravity of the detected cluster. The main disadvantage of this strategy is the reduced readout bandwidth, due to the large number of channels that has to be used for the reconstruction, and the poor timing resolution resulting from spreading the signal on many devices.

ISiPM topology aims to solve these issues [119]. The common output signal, usually available in any SiPM, is in this case replaced with multiple lines. Each avalanche diode within the array is connected to one of the available outputs, so that a proper weight function of the detected amplitudes can be used for estimating the center of gravity of the triggered cells. The goal is to obtain information on the photon cluster position without regard to its shape or orientation. The device is intrinsically position-sensitive and does not require any information from neighbor sensors for the reconstruction of the crystal position. On the other side, a single device needs multiple readout channels.

The designed prototypes have been fabricated by FBK [120]. The first two square prototyped versions, with a sensitive area of about $7.4\,mm \times 7.4\,mm$ have been realized in RGB technology [56], while the last sensor, to be assembled onto the detection module (described in Chapter 8) has been fabricated in HD-RGB technology [57].

7.2.1 *ISiPM Topology*

The prototypes have been designed with four output channels, each of them ideally corresponding to one of the corners of a rectangular/square sensor. Each diode within the array is connected to only one of the available channels, depending on its position within the array.

As stating point, let's consider an ideal square device where each point on the surface can be associated to four arbitrary values. Four continuous density functions $d_{ij}(x,y)$ are defined: the indexes i,j refer to the corners of the rectangular/square device, where each of them is ideally associated to an output channel. For the ISiPM square design, linear normalized distributions have been chosen:

$$\begin{aligned} d_{ij}(x,y) &= \frac{(i \cdot x + 1)(j \cdot y + 1)}{4} \\ &\text{for} \qquad i,j = \pm 1 \end{aligned} \tag{7.1}$$

with the axis origin placed at the center of the device and $x,y \in [-0.5\,;\,0.5]$. Every density function has its maximum in a corner and the minimum in the opposite one. At the center, the four densities are equal and normalized on the area.

In the considered case of a continuous device, any position could be reconstructed from the values of the four density distributions, using the appropriate weight function:

$$\begin{aligned} \vec{r_{rec}} &= \sum_{i,j} d_{i,j}(x,y)\vec{C_{ij}} \\ \text{for} \qquad \vec{C_{ij}} &= \begin{pmatrix} i \\ j \end{pmatrix} \end{aligned} \tag{7.2}$$

Ideally, minimum three density distributions are sufficient to uniquely reconstruct the position on a 2-D surface from an inverted weight function. It has been decided to keep the ISiPM design simple and symmetrical, using linear distributions referred to each sensor corner.

Since SiPMS are instead discrete devices, it is not possible to use the exact linear distributions in Eq.7.1, but instead each cell can be assigned to a single output channel. The assignment algorithm must ensure that the mean local density on an arbitrary large area is as close as possible to the ideal continuous value. The algorithm uses an iteration routine for defining the connection of each diode. At first, it takes into account square clusters of arbitrary size N and considers all cells not yet assigned to any output. It processes the full area starting from a corner, moving in steps of N cells, and performs the following commands:

- for each of the four density distributions, calculates the sum of the ideal density values $I_{i,j}$. The theoretical values are chosen taking into account the position of all diodes in the considered $N \times N$ area;
- calculates the sum of the ideal density values for the only diodes already assigned $F_{i,j}$ in the same area;
- if $I_{i,j} > \left(F_{i,f} + 1\right)$, surely at least an assignment to corner (i,j) is missing in the considered area. A random free position gets assigned to the (i,j) corner;
- it moves to the next $N \times N$ cluster and repeats the previous steps on the whole device area;
- the cluster size N is doubled and the procedure repeats;
- the iteration ends when all the cells have been assigned to an output.

An example of cell assignment for two devices with 40×40 and 160×160 diodes for initial cluster size $N = 2$ is shown in Fig.7.2. In the figure, different colors correspond to the available channels and each pixel is a diode in the SiPM array.

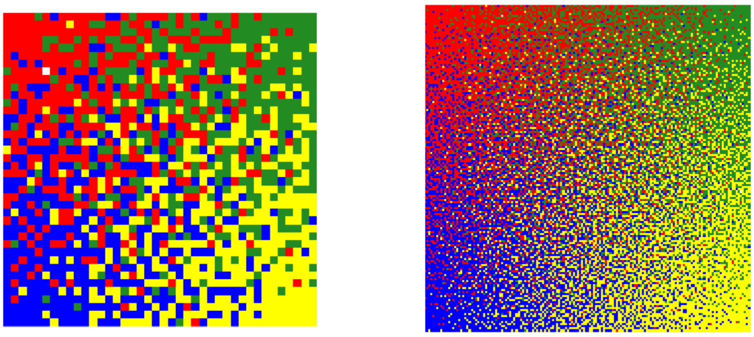


Figure 7.2.: Examples of cell assignment for ISiPM with 40×40 (left) and 160×160 (right) cells: each color corresponds to an output channel and each pixel in the image is a diode in the array.

In the discrete case, the center of gravity position can be reconstructed taking into account a reasonably large number of pixels. If the number of triggered diodes per channel is expressed as $T_{i,j}$, for an arbitrary shaped group of pixels, the center of gravity position $\vec{r}_{cog}$ is given by:

$$\vec{r}_{cog} = \frac{\sum T_{i,j}\vec{C}_{ij}}{\sum T_{i,j}} \tag{7.3}$$

The ISiPM reconstruction principle is not dependent on the orientation or on the shape of the detected cluster $T_{i,j}$.

7.2.2 *Reconstruction Errors*

The reconstruction of photon cluster center of gravity is affected by error, which depends on the both statistical effects and topology limitation.

Systematic error is introduced due to the discretization of the ideal local density function. In the extreme case of a single diode triggering the event, the topology offers no chance of reconstructing the detection position. The accuracy of the reconstruction method depends on the number of diodes contributing to the result: it improves for larger cluster and for denser arrays. An estimation of ideal reconstruction error is proposed in Fig.7.3 for devices with different numbers of pixels and photon cluster sizes. For the simulation, square clusters have been considered. It has been assumed that all diodes within the cluster trigger the event simultaneously, contributing to the reconstruction algorithm. At this stage, no assumption has been made on the real analogue signal or on the device saturation. The triggered cells belonging to the four distribution groups have been counted and summed to be used in Eq.7.3.

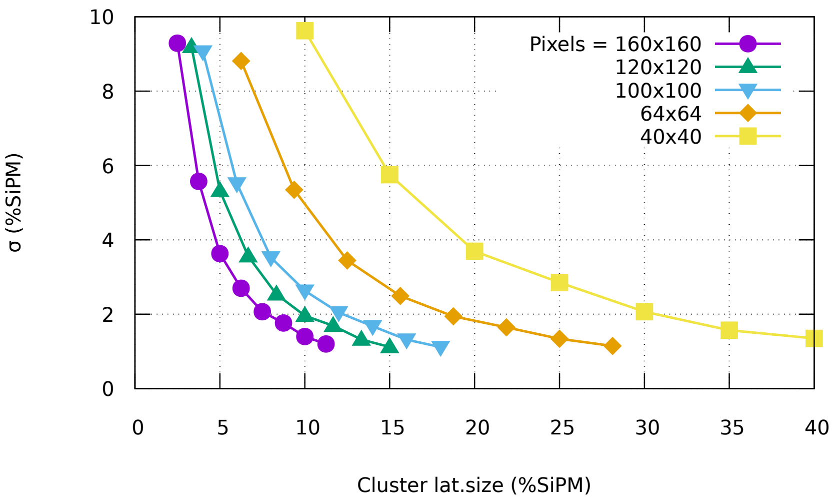
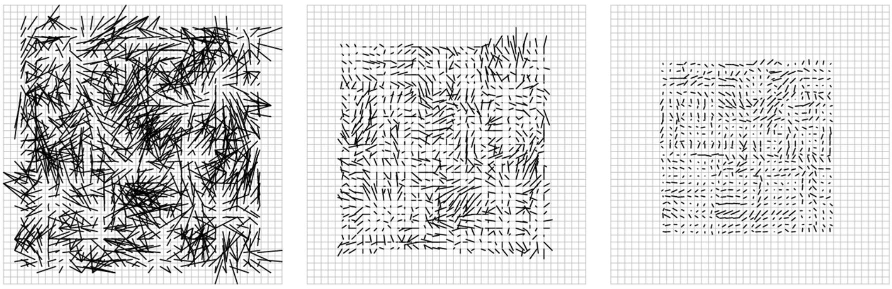


Figure 7.3.: Simulated systematic error for devices with different cell density and square cluster size. Both the error (standard deviation) and the lateral cluster size are represented as a fraction of the ISiPM lateral dimension.

The systematic error also depends on the diode assignment and on the detected cluster position. It could ideally be corrected for, knowing the size and the position of the impacting photon cluster (assuming for example that photons coming from scintillating crystals, placed in fixed positions). Fig.7.4 shows the systematic error in position reconstruction for an ISiPM with 40×40 cells. Square clusters of different lateral size $D = 5$, 10 and 15 have been used in simulation. The vectors on the area represent the distance between the real center of gravity for a cluster and its reconstructed position.



a) b) c)

Figure 7.4.: Systematic error dependance on the detected cluster size. Each line in the plots connects the real center of gravity of a cluster (size $D \times D$) to its reconstructed position. The errors are calculated as standard deviation and expressed as percentage of device lateral size. a) $D = 5$, $\sigma_x = 5.5\%$, $\sigma_y = 6.0\%$. b) $D = 10$, $\sigma_x = 2.0\%$, $\sigma_y = 2.1\%$. c) $D = 15$, $\sigma_x = 1.2\%$, $\sigma_y = 1.0\%$.

The plots offer a more intuitive representation of the effects of the cluster size on the center of gravity reconstruction. For a very large clusters of 15×15 cells (about 14% of the total SiPM area), the reconstruction error, expressed as standard deviation, is almost not detectable, reaching about 1.1% of SiPM size at FWHM in both x and y directions. It has to be noted that the simulation assumes that all the emitted photons hit the sensitive area, so that no consideration is made at this point on border or light spreading effects.

Another factor that affects the reconstruction accuracy is the overall detection efficiency. Only a fraction of the diodes within the cluster area actually trigger the event and contribute to the reconstruction algorithm, leading to additional degradation of the final result. The detection efficiency depends on many factors, such as the SiPM technology and layout, the optical coupling and wavelength matching with the scintillating crystals, and the bias voltage (for details on SiPM proprieties see Paragraph.3). Simulated reconstruction error, for different cluster sizes and fraction of detected photons per event, is shown in Fig.7.5-left. The simulation refers to an ISiPM with 160×160 diodes. Fig.7.5-right shows the effect of noise on the reconstructed position. In this case, it has been assumed that 80% of the diodes within the square cluster trigger the event. Noise has been added to the four output and

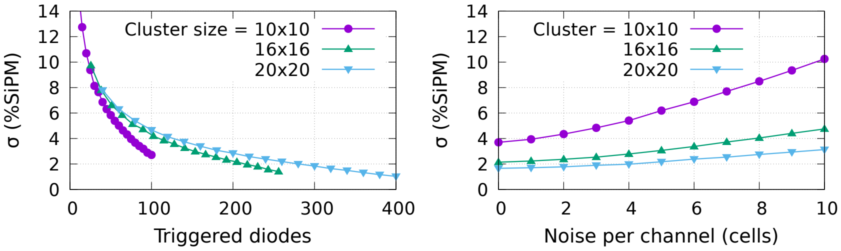


Figure 7.5.: Reconstruction error (standard deviation) due to photon detection statistics (left) and counting noise (right). The results are simulated on a device with 160 × 160 diodes. In the first case, not all the diodes in the cluster trigger the events, in the second case noise per channel is added in cells unit in a simulated case with 80% of detected photons in the cluster.

plot in cells unit. Its effect is less severe for large clusters, which have benefit from larger signal-to-noise ratio.

In all the previous simulated results, the error is always reported as the standard deviation of the reconstructed position, expressed as a percent ratio of the ISiPM lateral size. The size refers to the number of diodes and not to the actual dimension of the die. For this reason, it has to be kept in mind that the resolution here reported strongly depends on the cells density of the sensor. With recent technology, producers can achieve cell size down to even $15\,\mu m \times 15\,\mu m$.

As additional remark, the position reconstruction accuracy is not the only requirement for a position-sensitive SiPM, but has to be combined with good energy resolution. For any detected cluster, if all diodes within it trigger the event, the reconstructed position is ideally the best possible, but the energy information is completely lost. In practice, photon clusters are not so precisely defined and light sharing between adjacent crystals, or within the optical glue, has also an effect on their size. However, system requirements must be considered in order to find the best compromise between spatial and energy resolution.

7.2.3 *Prototype Design*

Two prototype versions have been fabricated at FBK with different technologies, in order to first prove the ISiPM concept. They both have shown very promising results [121, 122]. Based on their design, rectangular die with ISiPM topology have been designed and fabricated, to be used on the detection module described in Chapter 8, and whose results have been collected in Section 11.2.

The prototypes have both sensitive area of about $7.4\,mm \times 7.4\,mm$. The first production has been realized using a single metal layer for the signal routing in between the array. It has 100 × 100 avalanche diodes, each with size of $74.5\,\mu m \times 73\,\mu m$. A micro picture of the first device is shown in Fig.7.6-left. Since only one metal layer is available, two

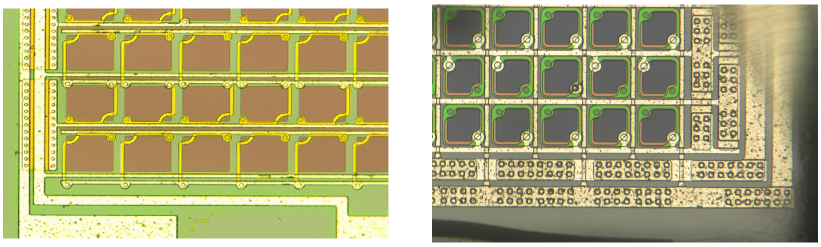


Figure 7.6.: Micro pictures of one (left) and two-metal (right) layer technology ISiPMs. The pictures refers to devices with cell size of $74.5\,\mu m \times 73\,\mu m$ and $60\,\mu m \times 60\,\mu m$ respectively.

parallel tracks are placed between each row of diodes, as visible in the picture. In a double row, the lines are connected to output 0 and 1, while in the next one output 2 and 3 are available, with a pattern that is repeated on the whole matrix. In this way, each diode is connected either directly with the closest track (top or bottom) with a metal connection, or to the second one using a poly-silicon bridge. The main disadvantages of such layout are the high track resistivity and large capacitance, combined with a high risk of electrical crosstalk between neighbor parallel tracks and possible shortcuts.

The second prototyped version has been fabricated with two-metal layer technology (micro picture in Fig.7.6-right). Two versions with different cell sizes and densities have been designed and tested: 120×120 cells of $60\,\mu m \times 60\,\mu m$ and 160×160 cells of $45\,\mu m \times 45\,\mu m$, for a total sensitive area of $7.2\,mm \times 7.2\,mm$ in both cases. As visible in the picture, the tracks are placed both horizontally and vertically between rows and columns of diodes. Each pixel is "surrounded" by the four signals, so that an easier layout is possible. Such configuration prevents some of the disadvantage of the single-metal layer layout: tracks are still long and with high resistivity, but crosstalk between signals is reduced, together with the risk of shortcuts.

Figure 7.7.: Rectangular ISiPM die, with two ISiPM sensors, with total area of $15\,mm \times 5\,mm$.

Rectangular dies have been designed and fabricated for being assembled on the detector module. As reported in Section 7.1, rectangular dies with wire-bonding pads at one edge allow to populate the module maximizing the light collection, since no gap for wire-bonding is required in between the sensitive area. The silicon die has a total area of about $15\,mm \times 5\,mm$, with two ISiPMs per die. The cell assignment algorithm is still based on linear density distribution of the four outputs in Eq.7.1, with adjusted boundaries for rectangular geometry. The rectangular sensors have been fabricated with two-metal layer HD-RGB technology, with cell size of $25\,\mu m \times 25\,\mu m$.

A picture of a rectangular ISiPM die is shown in Fig.7.7.

7.3 A DIGITAL PHOTON COUNTER

The most classical approach in photon detection consists of using analogue sensors and reading them out via specialized front-end electronics.

An interesting alternative is represented by digital SiPMs: each diode (or groups of diodes) is combined on the same silicon substrate with dedicated readout logic, which stores the hit information until readout starts. In such a way, the number of detected photons can ideally be counted for each event.

A prototype of digital SiPM has been developed and tested. The chip is here named Interpolating Digital Photo-sensor (IPD): it is fabricated with $0.35 \mu m$ CMOS technology by Fraunhofer Institute for Microelectronics Circuits and Systems [123]. The technology offers the possibility to integrate SPAD cells within CMOS chip and has shown low dark count rate of less than $50\,cps$ per SPAD with diameter of $30 \mu m$ at room temperature [124].

The die has an active area of $5\,mm \times 5\,mm$ and contains 88×88 pixels, each of them $56.44\,\mu m \times 56.44\,\mu m$. Two versions have been designed and fabricated, with similar pixel logic but different readout architecture.

The first chip (IDP1) has pixels with active area of $25.76\,\mu m \times 47.48\,\mu m$, for a total fill factor of about 37.7%. It integrates a programmable RAM, used for defining local density distributions of maximum four weight functions, with the same principle of ISiPM topology (described in Section 7.2) so that it allows center of gravity reconstruction on-chip.

The second prototype (IDP2) has pixels with larger active area for a fill factor of about 55%. It is designed for fast readout and does not include the RAM design anymore. Eight parallel output lines allow the readout of a maximum of $400\,k$ full frames per second [125]. A picture of the second chip version is shown in Fig.7.8. Details on the matrix layout are visible as well: active pixels are placed with asymmetric layout so that two columns share control lines, routed vertically (micro picture is taken on IDP1 matrix).

Characterization measurements for IPD2 can be found in [126].

7.3.1 *Pixel Architecture*

Each SPAD is connected at the anode to a quench resistor of $500\,k\Omega$ in both chip versions. A capacitor takes care of the DC-decoupling between the diode and its readout logic. The simplified schematic of the pixel logic is drawn in Fig.7.9.

The positive pulse produced by the triggered avalanche is fed into a discriminator, which converts the signal into full CMOS range. The

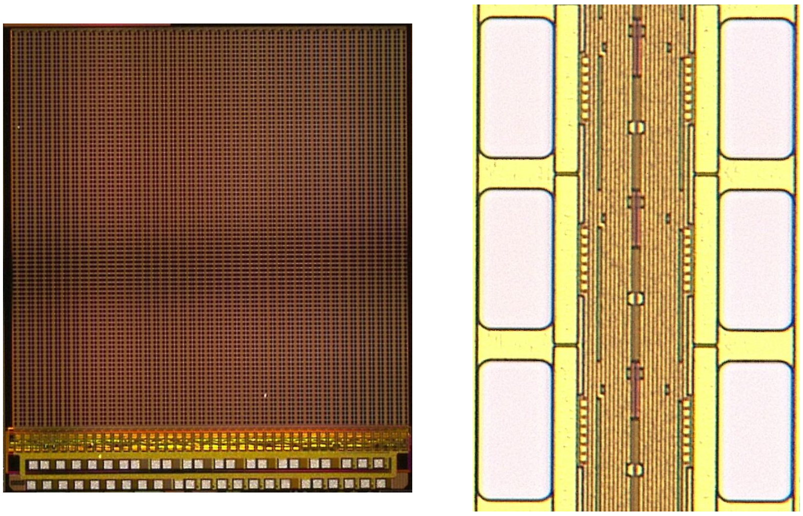


Figure 7.8.: Picture on the left shows the whole IPD2 chip: wire-bonding pads are placed at the lower edge. Micro picture on the right shows details on the matrix of IDP1, on an area of 2 × 3 pixels.

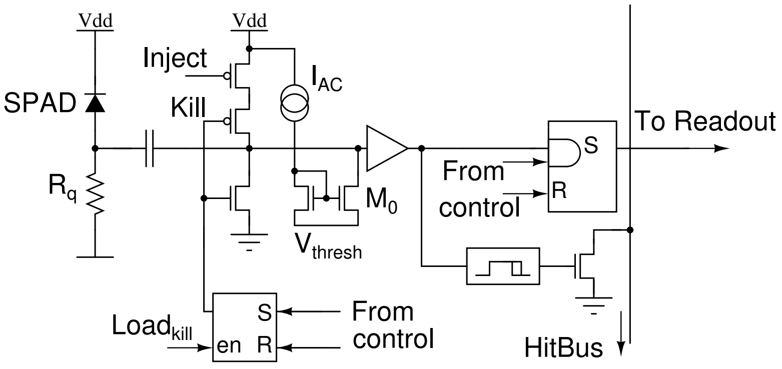


Figure 7.9.: Simplified schematic of the pixel logic within IPD1.

DC input level of the discriminator is kept at the voltage V_{thresh} (or at ground, in the second chip version), using a current mirror. The conductance of M_0, working in linear region, can be changed with the current source I_{AC}. The pixel can be disabled using a "killing" feature, which ties the discriminator input to ground. This option can be very useful for the "hot pixels", which are affected by higher noise rate and degrade device performance. A test pulse can be injected in all pixels not killed, using a dedicated *pmos*. The injection path is common for the entire matrix.

The hit signal produced by the discriminator is used by a mono-flop circuit, which generates a pulse with variable programmable width (in the order of about $10 - 20\,ns$) that drives the gate of a *nmos*, connected to a hit bus. The hit bus in each column is pulled high with a bias source. If one *nmos* in the column detects a hit, the hit bus is pulled to ground. The active-low signal is used within the trigger and multiplicity logic at the end-of-column logic block.

When the external control logic receives the first hit signal, it enables a memory latch per diode for a programmable accumulation time, typically on the order of 100 *ns*. In this time window, all hits coming from the same photon cluster are stored on the chip.

Once the accumulation time is over and the events are loaded into the respective shift register, the latches are reset and the pixel is available for the next event.

7.3.2 *Trigger and Multiplicity Generation*

The trigger logic defines if at least one hit has been detected within the whole matrix. It uses the hit signal generated per column on the hit bus, which is active-low, and generates a pulse with fast negative edge that is detected and sampled by an external TDC. Multiple hits within a column are not detectable by the logic.

The OR-hit between all pixels on the matrix is generated with two different circuits in both chip versions. In the first case, the hit bus of each column is received by an inverter, which drives another *nmos* gate. The pull-down *nmos* at the end of each column are connected in parallel to the same bus. In the first chip version, the signal is directly connected to the output pad and the setup needs an external pull-up mechanism for operating. The simplified schematic view of the circuit for only four column as example is shown in Fig.7.10-left. In IDP2, the pull-up is integrated on chip, the signal is inverted and produced a full CMOS pulse available at the output pad.

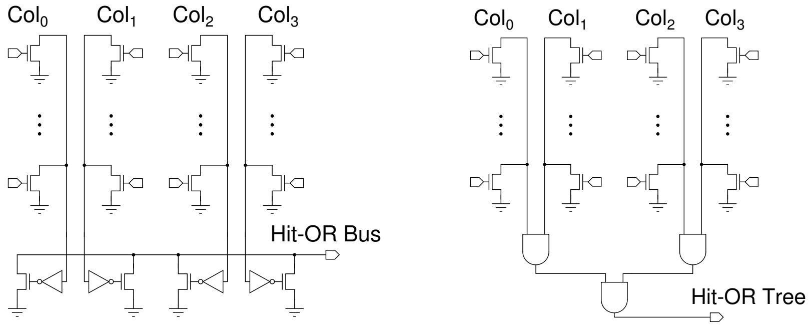


Figure 7.10.: Hit-Or generation (left) with fast floating bus. Hit Tree (right) using AND ports. Both circuit are schematically represented for only four pixel columns.

A second alternative for the trigger generation is shown in Fig.7.10-right: a tree of AND ports is used for transmitting a fast negative edge signal when at least one hit is detected. The design is optimized for timing and layout has been implemented ensuring constant delay propagation for every column. Simulations show that delay within the whole tree is less than 1 *ns*. It has also to be considered that signal propagation along the hit bus is dependent on the pixel position within the column. For very accurate timing resolution, delay of single pixels can be calibrated using the available test injection path.

A counter for the trigger multiplicity is necessary for distinguishing photons emitted from scintillator (or other light sources) from noise signals. The counter is designed as a “multiplicity tree”: the hit busses per column are evaluated and four outputs are available on pads, for both chip versions. The outputs correspond to conditions: $N_{trigg} \geq 1$, ≥ 2, ≥ 3, ≥ 4, where N_{trigg} is the number of triggered columns. The tree is based on the simple idea that, taken two inputs A and B, for the output Q, defined as the sum of them, the following equations are valid:

$$\begin{cases} Q \geq 1 \; if \;\; (A \geq 1 \wedge 1) \vee (1 \wedge B \geq 1) \\ Q \geq 2 \; if \;\; (A \geq 2 \wedge 1) \vee (A \geq 1 \wedge B \geq 1) \vee (1 \wedge B \geq 2) \\ Q \geq 3 \; if \;\; (A \geq 3 \wedge 1) \vee (A \geq 2 \times B \geq 1) \vee (A \geq 1 \wedge B \geq 2) \vee \\ \qquad\qquad (1 \wedge B \geq 3) \\ Q \geq 4 \; if \;\; (A \geq 4 \wedge 1) \vee (A \geq 3 \wedge B \geq 1) \vee (A \geq 2 \wedge B \geq 2) \vee \\ \qquad\qquad (A \geq 1 \wedge B \geq 3) \vee (1 \wedge B \geq 4) \end{cases} \tag{7.4}$$

For a matrix of 88 columns, a multiplicity of four simultaneous events in four different columns have been considered sufficient for the present application.

The tree is composed of identical OR-NAND-INV logic blocks. Each block is designed for 5 pairs of input signals, which is the minimum number of inputs required for defining multiplicity ≥ 4 (see last condition in Eq.7.4). Only the first tree level (corresponding to $N_{trigg} \geq 1$ and ≥ 2) has been designed with smaller blocks, with only 3 pairs of input signals, for keeping the tree layout more compact. The basic 5-inputs block, connected for multiplicity ≥ 2, is shown in Fig.7.11.

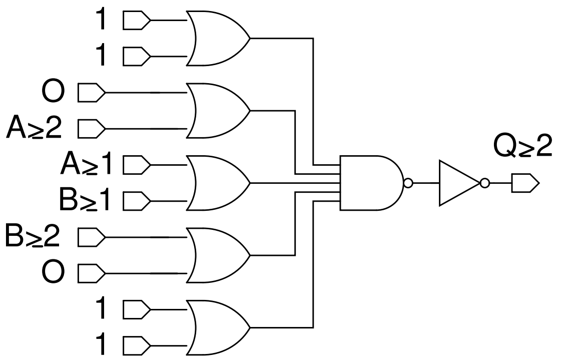


Figure 7.11.: Multiplicity tree logic: the base OR-NAND-INV block is connected with signal for multiplicity ≥ 2 as example.

Four unit blocks with proper input connections produce the necessary four outputs and they compose the multiplicity tree unit. The choice of using the same block for all the multiplicity outputs ensures a homogenous delay for all of them.

Each logic block is optimized for timing resolution, with low jitter and fast slope at the signal negative edge and is implemented with symmetrical design, for reducing the skew between input signals. Wide *nmos* transistors have been used for fast signal propagation and design has been carefully simulated. Input signals have been injected in random columns and constant delay on the four multiplicity outputs has been proved, in the order of only $4\,ns$ in the second IPD version. An example of simulated results is shown in Fig.7.12.

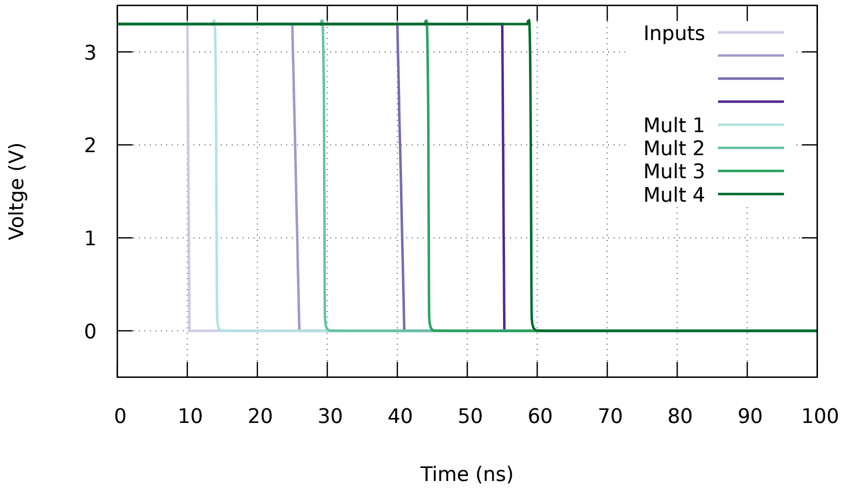


Figure 7.12.: Delay propagation within multiplicity tree for the four available outputs, after injection of pulses into four different pixel columns.

7.3.3 *Control and Readout Architecture*

In order to maximize the fill factor, control lines are kept at minimum and, when possible, shared between adjacent columns of pixels. For what concerns control and readout architecture, the two versions are different and are here treated separately.

IPD1: INTEGRATED CENTER OF GRAVITY RECONSTRUCTION IPD1 offers the opportunity to configure pixels singularly using two vertical and one horizontal lines per column. The combination of the two vertical signals, together with the horizontal signal enabling the selected row, allows the use of the different features. In particular, except for the kill-pixel feature present also on IPD2, it offers the possibility to exclude the pixel from the common hit production at the hit bus. This option could be useful for hot pixels, which could not be used for the trigger but still get accumulated in the final count. Since the columns are placed in pairs, each block of control lines includes actually four vertical signals. Due to a bug on the prototype, all even and odd pixels are actually selected in parallel, so that it is not possible to configure pixels singularly. IDP1 has also a so-called "freeze"-option: after detection of an event, the content of the hit latches can be immediately stored on a parasitic capacitor, so that the information on the first photon position for each event can be included in the readout, without affecting the normal hit collection.

The pixel is read out with a double buffering shift register: as soon as the accumulation window is over, the pixel is cleared and available for the next event.

Data from each double column are shifted until the end-of-column logic. Each column of pixels is read out with two output lines, shifting data from even/odd pixels respectively, for emptying the whole column at twice the speed. In slow-mode readout, the full 2-D hit pattern can be read out through an horizontal shift register.

In fast readout mode, the readout is based on a concept similar to the ISiPM topology (Section 7.2 for details). Each pixel is assigned to one of the four available colors and the mapping configuration is stored on the on-chip RAM. The assignment is done in such a way that the center of gravity of any cluster of triggered pixels can be reconstructed using a simple weight function of the four distribution sums. Since this topology has shown very promising results on analogue devices, it has been implemented on a digital sensor; also different density distributions of the four (or less) colors could be tested. A schematic view of the implemented design is shown in Fig.7.13. The number of triggered pixels per color is summed up on-chip. The four serial outputs per dou-

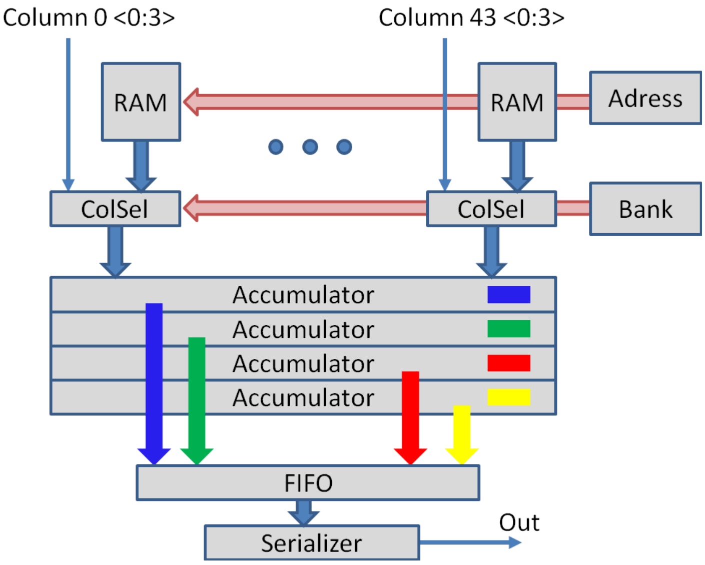


Figure 7.13.: IDP1 Readout architecture option with "color accumulator" feature, to be used for a readout based on the ISiPM concept.

ble column are compared with the RAM content of the corresponding row, for all the available colors. Four accumulator synthesized blocks (one per color) take care of counting the fired pixels, and store the result on a FIFO, which is read out serially. The readout is implemented with JTAG interface.

IPD2: FULL FRAME FAST READOUT The second chip version aims to faster readout and has solved some of the problems encountered with the first design. The matrix can be read out with up to 8 serial LVDS links, using double data rate. The number of control lines has been reduced as well, in order to improve the fill factor.

8

DETECTOR MODULE

A single-layer module has been designed on a Low Temperature Co-fired Ceramic (LTCC) substrate. Design and fabrication have been finalized in collaboration with Micro System Engineering GmbH [127]. The substrate hosts both the optical sensors and the read-out electronics, thermally decoupled by an internal channel within the ceramic, used for liquid cooling. All components on the assembly are not ferromagnetic, for ensuring MR compatibility.

The final detector height is only $8\,mm$, taking into account the connector size and not considering the scintillator array thickness, for a total area of about $32\,mm \times 32\,mm$.

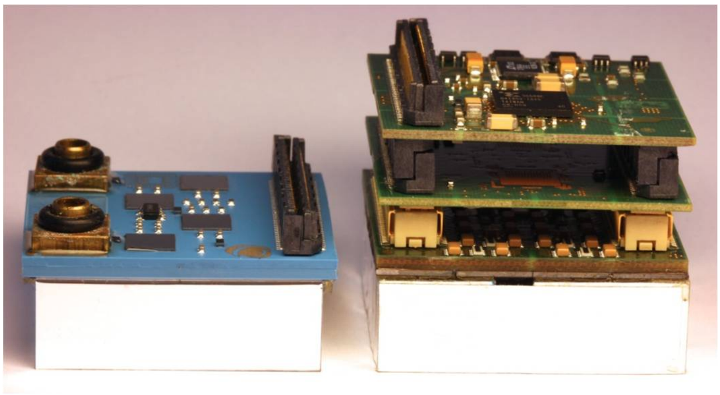


Figure 8.1.: Comparison between Sublima (left) and HYPERImage (right) detector modules, from a side view. The modules are both coupled with $1\,cm$ thick scintillator array. HYPERImage module is composed of three PCB layers (sensors, ASICs and data pre-processing board), while Sublima module hosts sensors and readout electronics on the same substrate.

A picture of the detector with $1\,mm$ thick crystal array is in Fig.8.1: it shows a comparison with the three-layers detector developed within HYPERImage project.

8.1 DESIGN AND FEATURES

The module is designed for hosting an array of 12×12 SiPMs, for 144 output signals in total, read out with four flip-chip mounted PETA5 ASICs. The SiPM dies have rectangular shape, as detailed described in

Paragraph 7.1. The pitch of $2.5\,mm$ between SiPMs is homogeneous on the whole matrix. A picture of the top and bottom surfaces is shown in Fig.8.2. The golden wire-bonding pads for the connection of the SiPM signals are placed at the external edges. The longest tracks on the module are about $30\,mm$ long and correspond to signals from SiPMs placed at the center of the array. No current loops are present, for ensuring MR compatibility.

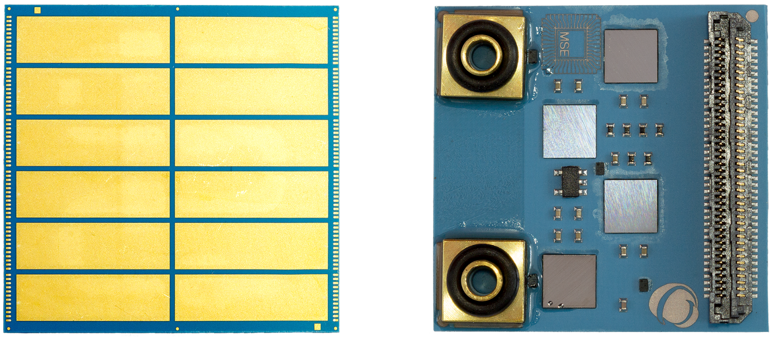


Figure 8.2.: Top and bottom views of the LTCC substrate. The top side (left) hosts the SiPM array, while the bottom side (right) is used for the read-out electronics and connection to the main read-out board.

The bottom surface (in Fig.8.2-right) is reserved to the read-out electronics. Four PETA5 can be operated in parallel with very few lines. Control signals are transmitted through fast LVDS lines, while slow CMOS signals are used for storing the configuration. The chip address is hard-coded on the substrate. All chips on the module use the same reference frequency for the internal PLL circuit and are synchronized.

An 80-pins connector is sufficient for communication with the main readout board, without any data pre-processing stage. Since most of the necessary logic is integrated on chip, the assembled design needs very few external component for stable operation. Except for few decoupling capacitors necessary between ground and power lines, the LVDS termination resistors are soldered on board, as close as possible to the ASICs. It has to be noted that PETA6 design integrates the termination resistors as well, with additional space saved on the assembly. As visible in the right picture in Fig.8.2, four small I^2C temperature sensors ($TMP\,108$ from Texas Instruments) are placed in strategic positions on the surface: close to the cooling inlet and outlet, between two ASICs (hottest spot on board) and at the connector (coldest spot on board). The sensors “alert” output pin is connected in parallel to the power enable input of all chips, so that if any of the sensor detects a temperature above the programmable threshold, the alert output shuts down the whole module instantaneously, preventing damages to the chips. An additional EEPROM for storing the configuration has been included on the assembly.

The inlet and outlet connectors for the cooling system are also visible in Fig.8.2: the picture shows a first connection version, with brass sockets glued on the surface. Classical O-rings prevent leakages. Solutions with a single-piece connector would help avoiding misalignment between the parts and subsequent leakages.

The power consumption of the whole module, taking into account both analogue and digital power, in normal readout operation can reach about $6\,W$, without considering the SiPM contribution.

8.2 INTERNAL LIQUID COOLING SYSTEM

Due to the high-compactness of the module, with sensors and electronics on the same substrate, and the large number of readout channels on the design, a very efficient cooling system is mandatory.

Ceramic substrates are a very valuable solution for this challenge, due to their high thermal conductivity, which is in the order of $2-3\,W/(m\,K)$ [127]. Such value is about ten times higher than the respective value for a standard multi-layer PCBs, where copper, with high thermal conductivity, and dielectric (usually FR-4) with very low conductivity, are stacked on top of each other in the design.

Moreover, the fabrication process of the LTCC substrate allows the integration of internal empty structures, few mm wide, where cold liquid can flow through for efficiently thermally decoupling top and bottom surfaces [128].

A side view of the ceramic internal layer design is schematically reproduced in Fig.8.3.

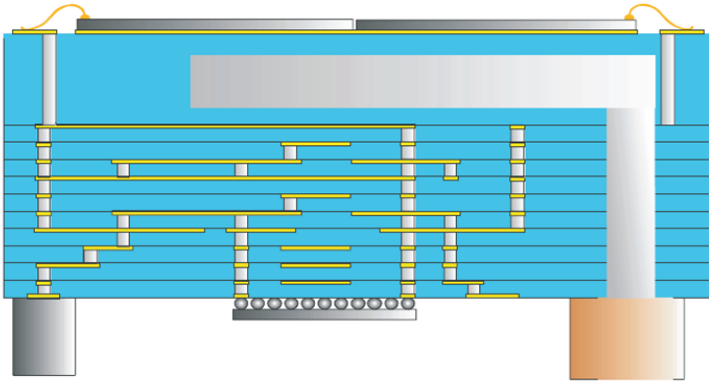


Figure 8.3.: Schematic of the LTCC cross-section, not in scale: 11 layers are used for signal routing and about 40 layers for mechanical purposes (source: Micro System Engineering GmbH, 2014).

11 ceramic layers, closer to the readout electronics, have been used for signal routing. Particular attention has been dedicated for routing fast LVDS signals and including ground planes in the design. Since, within the assembly, SiPMs are the temperature-sensitive devices and are placed on the top surface, it has been chosen to integrate the internal channel within the upper layers (about 40 of them have been necessary). The internal channel is about $2\,mm$ wide, following a dou-

ble “S” path, so that cooling liquid flows above all chips for maximizing the cooling efficiency.

Characterization measurements have been taken in order to prove the efficiency and the stability of the cooling concept. Water at different temperatures has been pumped into the substrate with a stable flow of about $0.35\, cm^3/s$. The temperature has been monitored using both the I^2C sensors available on the module and an infra-red camera pointed at the top surface. Double measurements have been particularly useful for estimating the mismatch between the temperature at the top surface (which affects the SiPMs performances) and the value measured by the sensors placed on the bottom side, which can be read out during data acquisition. It has been proved, in different conditions, that the maximum deviation between the two measurement techniques, for the hottest spot on board, is always less than $1^o\, C$.

Power dissipated by the system through water can be simplified as:

$$W_d = C_w \Phi_w \Delta T_w \tag{8.1}$$

where C_w is the water specific heat capacity (about $4.2\, J/(K\, cm^3)$), Φ_w is the liquid flow rate and ΔT_w is the increase of the water temperature. Considering that measurements have been taken on a module with only three fully operating chips, the total power dissipated by the electronics is in the order of $4.4\, W$. Using Eq.8.1, the theoretical water temperature increment, for a very good thermal coupling, is about $3^o\, C$. Experimental results are shown in Fig.8.4: four infra-red pictures of the substrate top surface have been taken using water at different temperatures. Pictures correspond to flowing water at 21, 16, 12 and $8^o\, C$ at the inlet. Water temperature at inlet and outlet has also been monitored

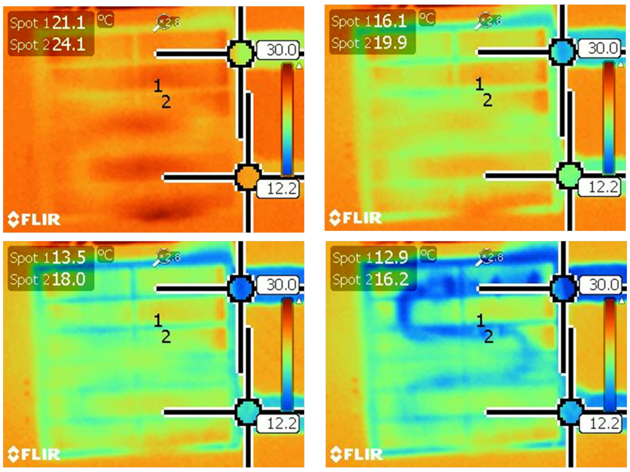


Figure 8.4.: Infra-red images of the ceramic substrate cooled with water flowing through the internal channel for four different liquid temperatures. From top-left to right-bottom the measured water temperature at the inlet is 21, 16, 12 and $8^o\, C$.

at fixed positions using the infra-red camera. The difference between them (reference points are visible in the figure) is about $3.8 \pm 0.8^o\, C$ in average, very close to the theoretical value, showing the efficient thermal coupling of the assembly. Discrepancies from the theoretical value are related to large error in the estimation of the water flow for the used test setup. In addition, it has been measured that the module reaches stable conditions in about ten seconds, four times faster than the value obtained for cooling the same assembly with external copper pipes in thermal contact with the chips backside.

Taking into account the border surface between water and ceramic, the convective thermal exchange, can be expressed as:

$$W_d = h\, D_{ch} \int_0^{L_{ch}} (T_{LTCC}(x) - T_w(\,x))dx \tag{8.2}$$

where W_d is the total exchanged power, h is the heat transfer coefficient at the boundary, D_{ch} and L_{ch} are the channel width and length, $T_{LTCC}(x)$ and $T_w(x)$ are the temperatures of substrate and water respectively, as function of the position along the channel. Deriving the substrate temperature from Eq.8.2, it is easy to obtain:

$$T_{LTCC}(x) = \frac{dW_d}{dx} \cdot \frac{1}{h\, D_{ch}} + T_w(x) \tag{8.3}$$

Since on the setup there is not chance for measuring the substrate temperature distribution along the channel, the highest temperature on the surface has been considered. For different power consumption conditions and different water temperature at the inlet, the hottest detected position is always in the same point X_{HOT}. The dependance of $T_w\ (X_{HOT})$ on the inlet temperature is linear and can be deduced from Eq.8.1, where heat sources are assumed discrete and corresponding to the chips position along the channel. The expected linear curves of $T_{LTCC}(X_{HOT})$ as function of the water temperature at the inlet are reported in Fig.8.5. Data have been acquired for two different values of W_d, enabling one or three ASICs at time.

8.3 ASSEMBLY IN CLINICAL AND PRE-CLINICAL CONFIGURATION

The single-layer module must be scalable, for application in both clinical and pre-clinical scanners. Since the two systems aim to pretty different performance, the module has been adjusted to fit the requirements in both cases.

Rectangular SiPMs described in Section 7.1 have been used for populating the ceramic top surface. The sensor placement has been realized by Micro System Engineering GmbH [127], with a distance between dies of about $100\,\mu m$ only. A picture of a fully populated module, together with a detailed view of the bonding wires at the array edge is shown in Fig.8.6.

For achieving the best timing resolution for ToF in clinical applications, the array has been coupled with a matrix of LYSO crystals with

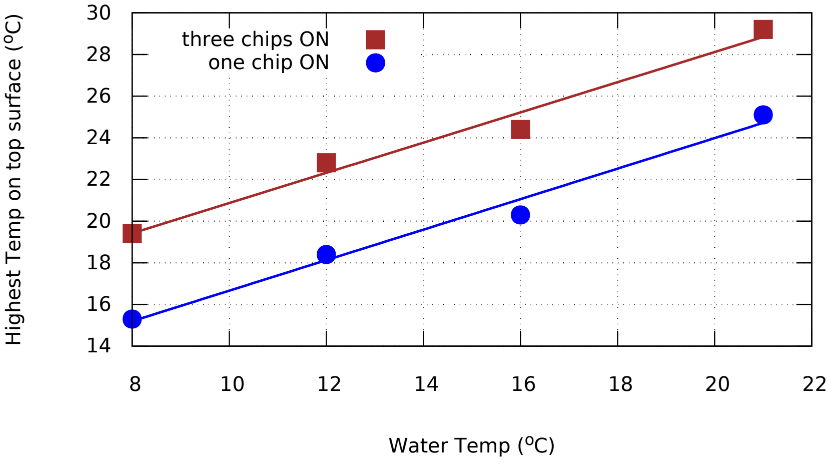


Figure 8.5.: Substrate temperature dependence on water temperature at the inlet: the hottest measured point on top surface has been taken as reference for the plot. Both temperatures have been measured with infra-red camera.

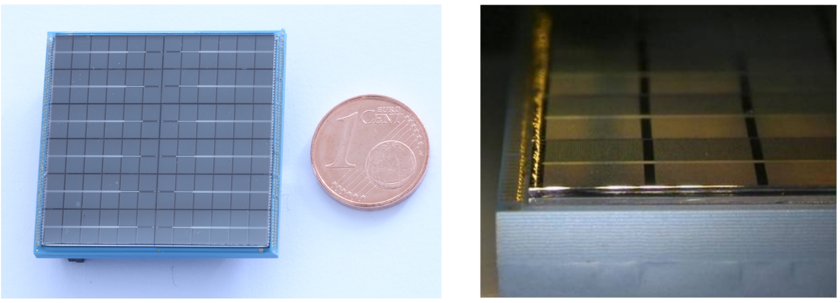


Figure 8.6.: Picture of a fully assembled module with 144 sensors and detailed view of the bonding wires at one edge.

a regular pitch of $2.5\,mm$. LYSO has very good performance in terms of light yield and response time and is available at reasonable price (proprieties listed for LSO:Ce in Table 2.1 are valid for LYSO as well). The crystals are only $10\,mm$ thick: this can be considered an unusual design solution for whole-body rings, where commonly crystal thickness is $20-25\,mm$. It has been chosen as a good compromise between stopping power, and then detection sensitivity, and design compactness, taking into account the mechanical constraint of the Sublima ring.

In order to minimize light sharing between adjacent scintillators, the array has been built with three layers of reflective material between crystals: a central $0.011\,\mu m$ aluminum foil, plus two $65\,\mu m$ enhanced specular reflector (ESR [129]) layers glued on crystals with adhesive. ESR is commonly used in PET imaging systems for its high reflection coefficient, reaching about 98.5% of the total impacting light. The value slightly depends on the incoming angle and is optimized for wavelength above $400\,nm$ [130]. Light transmitted through ESR layer is then reflected or absorbed by the aluminum foil. Aluminum reflects 87% of impacting light and absorbs the other 13%, with no transmitted fraction.

Fig.8.7 shows a picture of a fully assembled module in clinical configuration. The crystal block has been mounted at a flip-chip mounting machine at the Computer Science Institute, using a customized holder on the machine rotating arm, large enough to grab the array. Alignment between crystals and sensors is ensured thanks to a careful placement at the microscope. Transparent glue with refraction index of about 1.5 is used for both optical coupling and for protecting the bonding-wires [131].

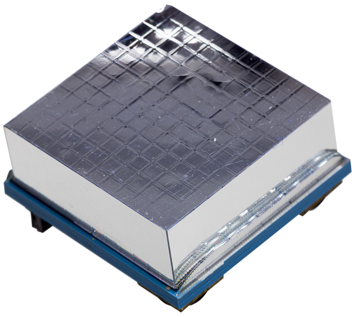


Figure 8.7.: Clinical gamma detection module.

For what concerns the pre-clinical detection module, two versions have been assembled. The first one is based on the same SiPM array shown in Fig.8.6. A 1 *mm* thick glass plate has been mounted on the array with the same transparent glue used for the crystal array assembly. It is used as light guide for spreading the optical photons on a bunch on neighbor sensors. For small-animal applications, it has been chosen to employ an LYSO array with a pitch of 0.833 *mm*, with pixel size of $0.755\,mm \times 0.755\,mm \times 10\,mm$ and ESR reflector between crystals. Such configuration requires to read out energy information from a reasonable large number of neighbor sensors. Neighbor logic within the chip allows readout of groups of channels.

A second design uses an array of ISiPM devices described in Section 7.2. In such a case, since the sensors is intrinsically position-sensitive, no information from neighbor pixels is needed and no light guide has to be mounted on the array. The first prototype has been built on a dedicated PCB, connected to the ceramic substrate through an adapter board (see pictures of the prototype module in Fig.8.8). This solution allows to test multiple ISiPM arrays with the same ceramic module and has been considered more practical in the very initial test phase. The array has been coupled with the same LYSO array used in the standard pre-clinical module version, with an overall pitch of 0.833 *mm*. In this setup, ISiPMs do not benefit from internal liquid cooling, but at this testing stage it has been considered a minor issue, compared with the advantage in flexibility.

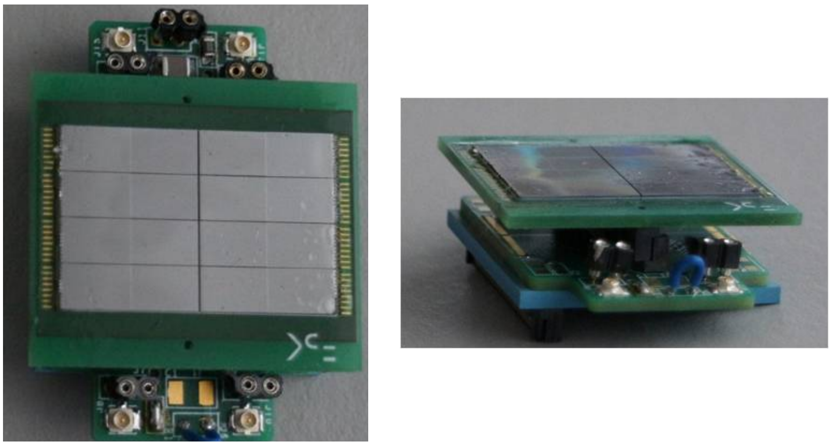


Figure 8.8.: Picture of an ISiPM-based prototype module.

9

TEST SETUPS

All the developed designs have been characterized and tested on dedicated setups. The description of each of them is here not presented, with the exception of a large coincidence setup used for reading out two modules in coincidence, emulating ideally a small diameter ring.

The main disadvantage of the ceramic substrate is its limited flexibility, preventing the use of different SiPM devices on the assembly. For overcoming at this problem, an adapter board has been fabricated.

The chips are controlled in all cases with an FPGA. Control firmware (in Verilog) and software (in C++) have been adjusted from previous available versions implemented by Michael Ritzert.

9.1 BACK-TO-BACK COINCIDENCE SETUP

A coincidence setup has been designed for the detection of back-to-back gamma radiation from a positron decay with two detector modules. This configuration emulates the geometry of a real PET ring and ensures realistic measurements at system level.

The setup is composed of an horizontal board, named here SSEP (Sublima Stack Evaluation Platform). It hosts a Xilinx Spartan6 FPGA for reading out and controlling up to 8 PETA5 in parallel, through USB2.0 interface. PC-based software, implemented under Linux, is used for programming the FPGA.

A fast quartz oscillator at $622.08 MHz$ with LVDS output is used as reference clock. A clock divider with low jitter multiplies the signal at the same frequency on two LVDS outputs with no phase shift. These signals are used as PLL reference clocks on the modules. A third output is divided on board with selectable ratio and used as reference clock within the FPGA and for the ASICs readout. Frequency of 104, 124 or $156\,MHz$ can be selected.

At the sides, the SSEP board can be connected with two vertical PCBs, named MABs (Module Adapter Board). Each MAB hosts one detection module: different versions have been designed for reading out all the prototyped stacks available. The vertical boards are placed at about $150\,mm$ distance: this geometry has been considered as a good compromise between very long traces on the PCB, stiffness of the setup and realistic small-animal ring dimensions. In addition, SSEP includes a free slot at the center compatible with a rotational stage, in order to emulate phantom reconstruction with just two detector modules. A picture of the setup, together with two MAB boards with modules in clinical configuration is visible in Fig.9.1.

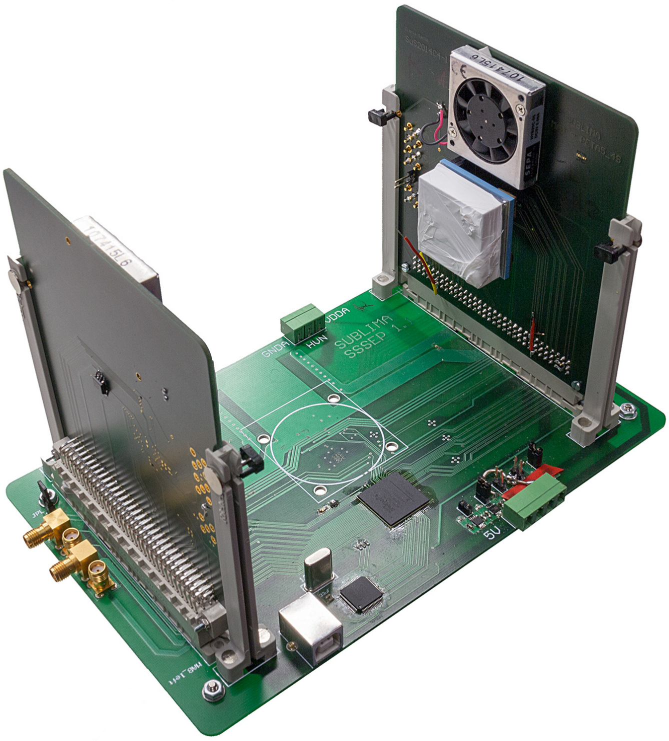


Figure 9.1.: Picture of the back-to-back coincidence setup. Two modules, with 8 PETA5 ASICS in total, are read-out in parallel with Xilinx Spartan6 FPGA via USB 2.0 interface.

The system is not optimized for internal liquid cooling. Since it operates only two modules at time, a simple series connection with plastic flexible pipes between them has shown to be a good solution.

9.1.1 *FPGA readout architecture*

PETA readout internal architecture is described in Paragraph 6.2.3. Data per channel are loaded onto a shift register on each chip half and the channel is cleared, to be immediately ready for the next event. Control lines are shared between four chips on a substrate and FPGA is programmed for synchronous readout of the two modules as well. This means that 8 PETAs on the setup are always read out simulataneosly.

The readout sequence can be based on two approaches:

- Hit-based readout: data are loaded onto the shift register only when at least one hit in any of the chips has been detected. The OR-ed LVDS hit signals are used by the FPGA for starting the readout procedure;

- Continuous readout: data are loaded and read out at fixed time intervals, even if no hit is stored on chips. FPGA selects the interesting data to be saved and discards useless information.

Each frame of data read out from the chips during a full shift is assigned to a 13-bit timestamp, generated within the FPGA. It increments every time the chips get emptied: in continuous readout, the frame timestamp has constant width and can be used for estimating the detection rate of the chips.

For the hit-based readout strategy, the counter increments only when receiving a hit and is uncorrelated to real time. Frame timestamps have variable bin-width depending on the hit arrival times. It can be useful when bypass option is active, where only triggered channels are shifted out. In this case, the continuous readout strategy could have effects on the overall dead time, since fixed frame timestamp width is typically calibrated on the unlikely case that all channels trigger an event. For few triggered channels read out in bypass mode, hit-based readout seems then a more efficient alternative.

The continuous readout option has been used for all the presented measurements, with shift clock at $124.4\,MHz$. With such configuration, taking into account Eq.6.23, the time necessary for emptying the shift register is about $5.35\,\mu s$. The chip is read out at intervals of $6\,\mu s$.

FPGA includes a FIFO for each half chip, used for storing event data to be sent out via USB2.0 interface. Stored data are 8 bytes long and includes:

- FPGA frame counter timestamp
- PETA timestamps (fine and coarse counter)
- integral energy (9 bits for single-ended front-end, selectable 9 to 11 for differential-ended front-end)
- channel and chip ID
- hit type (test, neighbor logic master or slave, hit with invalid timestamp)

If debug data are read out from the second available shift register, another 8 bytes are stored in the FIFO for each event.

FIFOs are read out sequentially through an FTDI interface working at $60\,MHz$. If at least one FIFO gets full during data acquisition, the readout process stops until all FIFOs are empty again. This step is clearly the bottle neck of the readout chain and is not optimized for high data rate measurements.

9.2 MODIFIED SETUP FOR FLEXIBLE SIPM READOUT

The single layer module offers a great advantage in terms of compactness and channel density, but it has the main drawback of requiring dedicated SiPMs assembled on the top surface in order to be operated.

For allowing more flexibility, a modified setup has been designed. An adapter PCB (designed by FBK) is conductive glued on the ceramic top surface, using the SiPM large bias pads. The board is slightly shorter than the LTCC in one direction, in order to allow wire-bonding between the two layers, using the SiPM input pads. Chip input signals are then available on a 80 pins connector and on few other sockets on the PCB.

The adapted module can be used for reading out SiPMs with different geometry or fabricated with different technology. It offers the necessary flexibility for using the developed PETA readout in many conditions, but is not optimal. It has been very useful for estimating limit in timing resolution with very efficient SiPMs (Paragraph 10.1) and testing new ISiPM prototypes (Paragraph 7.2).

Clearly, longer tracks from SiPM outputs to the chip pads can have an influence on the results, especially when timing resolution is pushed to the limit.

An example of the setup limitations is visible in Fig.9.2. In the picture, a full tile is read out using the adapter connector and a single die is connected through cable about $10\,cm$ long to one of the sockets available on the board. In case of coincidence timing measurements, the cable connection is clearly not optimal and can have a large impact on the final results. Additional comments on this can be found together with results in Paragraph 10.1).

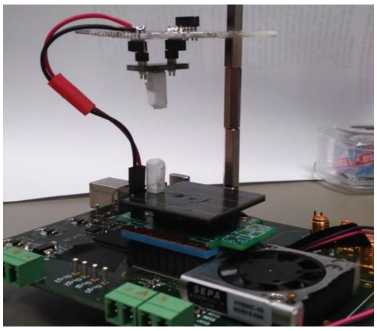


Figure 9.2.: Adapted module in coincidence with single die.

Part III

RESULTS

In this part, results obtained with the designed detection modules are collected. Measurements acquired with 1:1 coupling between sensors and crystals, for ToF resolution applications, are described in 10. Chapter 11 is dedicated to the results of pre-clinical modules, assembled for very high spatial resolution. Finally, the setup has been operated at high data rate and chip performance, in terms of both timing and energy resolution, has been evaluated and are presented in Chapter 12.

10

TIMING RESOLUTION

Timing performance presented in this chapter has been evaluated for modules in 1:1 configuration between sensors and scintillators, in order to maximize light collection. For all measurements, LYSO crystals have been used, with variable thickness depending on the application. SiPMs of two different technologies, both produced by FBK, have been employed in the measurements. Results are shown for the entire detection module and also for a reference setup with only two readout channels.

10.1 SINGLE CHANNELS COINCIDENCE RESOLUTION

Single SiPM dies have been used as reference for studying the achievable limit in coincidence timing resolution with PETA5. They are not affected by optical crosstalk from neighbor devices and can ideally achieve better results. In order to read them out, the modified setup described in Section 9.2 has been employed.

Two dies fabricated in FBK HD-NUV technology have been chosen for the measurements. HD-NUV SiPMs have large PDE in the blue and near ultra-violet range, up to $53\,\%$ for $400\,nm$ wavelength at $9\,V$, and dark count rate of only $200\,kHz$ at the same voltage, which is not much affected by temperature variations [132]. The cell size is only $25\,\mu m \times 25\,\mu m$, for a total fill factor of about $72\,\%$. Devices of $4\,mm \times 4\,mm$ have been optically glued with two LYSO crystals of $3\,mm \times 3\,mm \times 5\,mm$. Coincidence events from a point-form ^{22}Na placed between crystal top surfaces have been acquired. Reference measurements by FBK, using discrete electronics on a dedicated PCB, reports CRT between channels of about $100\,ps$ at FWHM [132].

Measurements have been acquired at stable temperature of $20^o\,C$ and results are shown in Fig.10.1. Corrections for the TDC not linear bin-width and for time walk effect are also included in the figure. The time walk effect correction takes care of the delay in the arrival time for signals with lower energy, which have a slower rising edge and triggers later the discriminator logic. On the other hand, high energy signals behave oppositely and trigger an early timestamp. The best CRT achieved is about $182\,ps$ at FWHM at $11\,V$ over the breakdown, which is amongst the best values at the state of art.

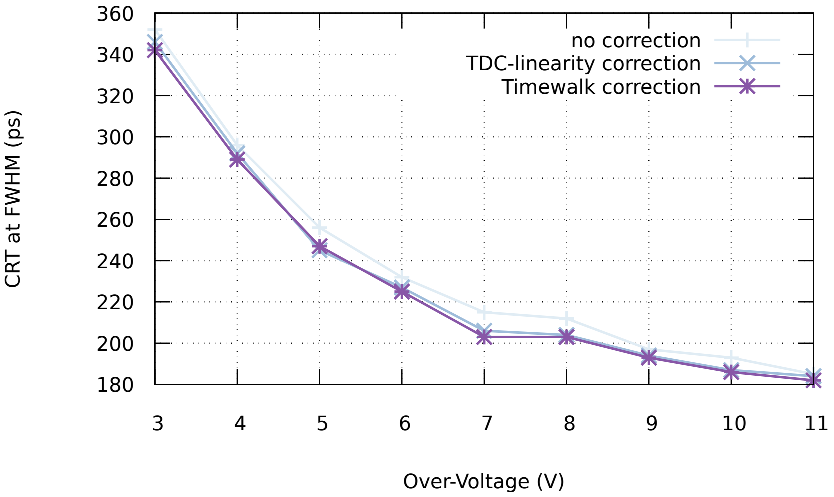


Figure 10.1.: Dependance of the CRT resolution on the over-voltage applied to SiPMs ($4\,mm \times 4\,mm$ sensors couple with $3\,mm \times 3\,mm \times 5\,mm$ LYSO crystals). Corrections for the TDC bin width non linearity and for time walk effect are included in the plot.

The achieved resolution is still far from the theoretical limit of the design (with injected analogue pulses the coincidence timing resolution is about $77\,ps$ at FWHM, as reported in Appendix A). SiPM limit resolution for the technology, measured by FBK, is also on the same order. The setup limitations have probably the largest effect on the achievable resolution: signals are brought from SiPMs to chip input pads using in this case about $10\,cm$ long cables for both devices, adding large jitter to the measurements and explaining, at least partially, the results. A more compact design could allow lower CRT values.

10.1.1 *Considerations on energy resolution*

It has been noted that high bias voltage affects the charge integrator performance, leading to a non-linearity effect in addition to the SiPM non linear behavior. It is due to under-estimation of the dynamic range within the front-end.

As an example of this effect, energy spectra detected by one channel for two values of the SiPM bias are shown in Fig.10.2. For a rough estimation of the electronic contribution to the saturation effect, the SiPM behavior is first investigated.

The selected SiPMs have $1600\,cells/mm^2$ from datasheet, for totally 14400 cells in the area below the crystal. Typically, light output from LYSO is to about $27\,\gamma/keV$ (see Table 2.1). The actual number of photons depends on crystal surface treatment and on optical coupling, but for a single crystal wrapped in Teflon tape, it is assumed that all emitted photons impact the sensor surface. Taking into account the characterization measurements of the SiPM reported in [132], it is

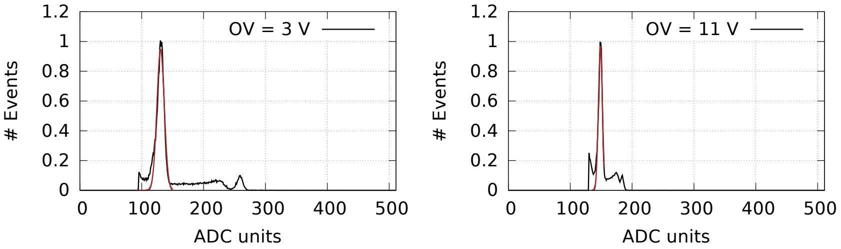


Figure 10.2.: Detected energy spectra measured by the same channel for two values of over-voltage.

possible to estimate a PDE of about 40 and 55% at 400 nm respectively for the considered over voltage of 3 and 11 V. Eq.3.11 is used for calculating the number of triggered diodes for sodium peak energies of 511 keV and 1.27 MeV (positron emission peak and single gamma emission peak). In a completely linear setup, the ratio between peaks is about 2.48.

From previous assumptions and considering only SiPM saturation effect, the expected peak ratio, calculated for the two bias voltages, is about 2.06 and 1.79 at 3 and 11 V over breakdown respectively. Spectra measured with PETA5, and shown as example in Fig10.2, correspond to peak ratio of 1.98 and 1.34 in the two cases. While for high over voltage a saturation effect in the chip is very probable, for low bias voltage it can be assumed that the integration procedure works in linear range, with an energy resolution of about 11.1% at FWHM before linearization. The energy spectrum has been linearized, considering that the number of generated photons in the scintillator is proportional to the absorbed gamma energy through a constant β and that the ADC response is linear with the total number of triggered cells with a gain α. Eq.3.11 can then be re-written as:

$$ADC_u = \alpha \cdot N_{fired} = \alpha \cdot N_{pixel} \cdot \left(1 - e^{-\frac{\beta \cdot En_{keV}}{N_{pixel}}}\right) \tag{10.1}$$

where En_{keV} is the detected energy in keV and ADC_u is the integrator measurement result. From such equation it is easy to deduce the dependance of the energy on the ADC units:

$$En_{KeV} = -\frac{N_{pixel}}{\beta} \cdot \log\left(1 - \frac{ADC_u}{\alpha N_{pixel}}\right) \tag{10.2}$$

Unknown α and β can be found from the two known peak positions. After linearization the energy resolution degrades to about 12.4% at FWHM. This value is in the same order of the best values reported in literature, with ASIC readout.

Energy spectra measured with discrete electronics have shown very linear behavior, in the the order of $9-10\%$ in a wide range of bias voltages [132]. It is reasonable to conclude that a proper correction of the integration linear range in next front-end design should lead to energy resolution of about $12-13\%$ at FWHM for wide range of bias settings. Electronic calibration could help in estimating the energy resolution in non-linear regions, but the maximum injected current is not enough for realistic comparison with SiPM signal at high over voltage in the present design. The very quick progresses in SiPM fabrication technology have make possible to operate devices at very high bias which were not foreseen at the time of design. The next PETA version has already been adjusted for this purpose.

10.2 MODULE COINCIDENCE RESOLUTION

Two modules assembled in 1:1 coupling configuration have been used on the coincidence setup described in Section 9.1 with a point-form ^{22}Na source.

The first results, presented in Fig.10.3, show the timing coincidence resolution variation of two opposite channels, one per module, for different bias voltage [133]. Only the two selected channels have been read out during data acquisition, but the whole modules have been bias at high voltage, so that crosstalk effects between sensors are present. Measurements have been acquired at different temperatures and for over-voltage up to $9\,V$: the energy saturation effect within electronic front-end and SiPMs, discussed in the previous section, prevents the use of higher bias settings in this configuration. In the same figure, results obtained with a different test setup from FBK, with the same SiPM

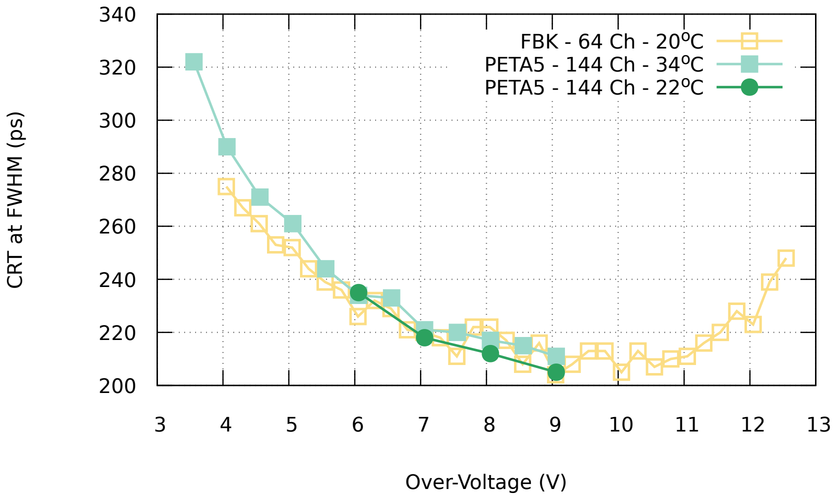


Figure 10.3.: CRT resolution between two channels on opposite modules. Plots show results obtained with two test setups, using the same SiPM technology: modules read out with PETA5 (array of 144 sensors, $2.25\,mm \times 2.25\,mm$ SiPMs coupled with $2.25\,mm \times 2.25\,mm \times 10\,mm$ LYSO crystals), SiPM tile read out with discrete electronics at FBK (array of 64 sensors, $4\,mm \times 4\,mm$ SiPMs coupled with $3.9\,mm \times 3.9\,mm \times 22\,mm$ LYSO crystals); results by [134].

technology, are presented as well. Data are published in [135, 134]. In such case, coincidences have been acquired with arrays of 8×8 SiPMs, with active area per sensor of about $4\,mm \times 4\,mm$. Each SiPM has been coupled with a LYSO crystal of $3.9\,mm \times 3.9\,mm \times 22\,mm$ and data have been acquired at 20^oC. Since setups, SiPM active area, crystal thickness and temperature conditions are different, the presented plot does not mean to be an actual comparison between setups and performance. Nevertheless, it gives anyway an impression on the limit resolution achievable with the technology in different configurations and conditions.

The best results obtained with both setups are in the order of $200\,ps$ CRT at FWHM. In case of PETA5 readout, the best coincidence resolution is about $205\,ps$, when the whole array is biased at $9\,V$ over breakdown and the temperature is about 22^oC.

Coincidences have been acquired for a large sensitive area, in order to estimate the timing resolution for the entire system, with an overvoltage of $8.3\,V$ and stable temperature at 20^oC. An area of 6×3 SiPMs on each module has been considered: the point-form ^{22}Na source has been placed in different positions within the field of view for detecting coincidences along the largest possible number of LOR between modules. Results are combined together on the map shown in Fig.10.4-top. Map axis represent the channels read out for each modules: missing results on the map correspond to missing LORs (and then missing source positions) during data acquisition. One channel (number 13 on module 2) was not operational during measurements.

It is clear from the map that some channels have slightly worst performance (as example channel 17 on module 2 and channel 11 on module 1), due to their higher noise rate and trigger threshold during acquisition. No correlation has been noted between sensor position on the array and achieved timing resolution, showing that short connection lines on the compact module and shaped tracks on the dies are effective.

The coincidence values are represented as a frequency histogram on the plot shown in the lowest part of Fig.10.4: the mean value for all the considered pairs is only $232\,ps$ at FHWM. Such result can be considered a reliable estimation of the system performance: measurements have been taken at room temperature, powering the whole SiPM arrays and a large sensitive area has been taken into account. For all these reasons, this result clearly demonstrates that the module can achieve outstanding timing resolution and can be considered the present state-of-art for system performance.

For estimating the energy resolution achievable with the module, assumptions have been made as in Section 10.1.1 and the charge integration has been considered linear for measurements at low bias voltage.

For an over voltage of $3\,V$, the modules have measured a peak ratio in sodium spectrum, averaged on 72 channels (2 chips fully read out per

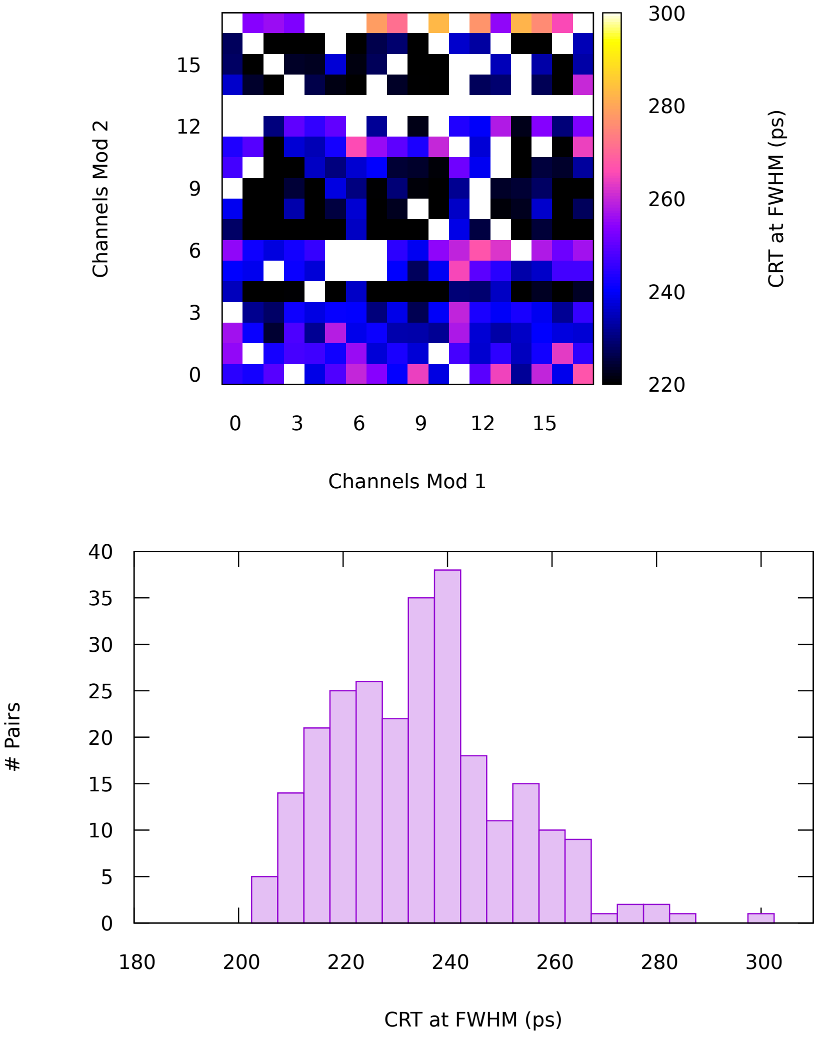


Figure 10.4.: Map and relative frequency histogram of the measured CRT values at FWHM for pairs of opposite channels on an area of 6 × 3 sensors per detector. 18 channels per module have been read out, with the whole array biased. Missing data in the map correspond to missing LOR during data acquisition. One not-functional channel is included.

module), of about 1.9 and a non-linearized energy resolution of 13.1% at FWHM. Each SiPM has an active area of $2.25\,mm \times 2.25\,mm$, with 8100 diodes in total. Assuming a priori that the entire saturation effect is due to SiPM limited number of cells, the measured peak ratio corresponds to detection efficiency of the assembly of about 24%. This value is very reasonable, taking into account that the SiPM PDE in the same condition is about 30% ([57]) and that part of the light can be absorbed by the aluminum foil between crystals, or spread on neighbor sensors within the optical coupling. For this reason, it has been considered that the approximation of ADC working in linear region for over-voltage of $3\,V$ makes sense. After linearization, the resolution reaches about 15.0% at FWHM, averaged on all readout channels.

10.3 EFFECTS OF TEMPERATURE ON TIMING RESOLUTION

HD-RGB SiPMs on the detector module have stable dependance on the temperature and dark count rate less than $2 \times 10^6\,cps$ at 20^oC [57].

The efficient internal cooling system described in Paragraph 8.2 has been used with water at different temperatures. The inlet and outlet sockets has been adjusted in order to use flexible plastic pipes for connecting two modules in series on the coincidence test setup. 3-D printed sockets have been glued on the substrate with water resistant tape, as shown in the picture in Fig.10.5.

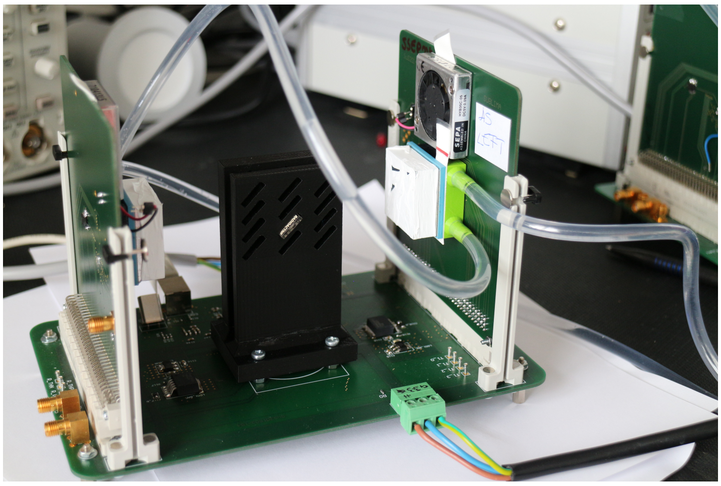


Figure 10.5.: Setup configuration for internal liquid cooling: the liquid flows within one module (on the right) and then into the opposite one. The temperature difference between modules measured at the inlets is about 4^oC.

Since modules have been connected serially, only one chip per stack is powered during measurements for limiting the water temperature difference between detectors. In these conditions, temperature difference between tiles is about 4^oC (also due to the length of the pipes). The dependance of the CRT resolution on the water temperature is given in Fig.10.6. Five pairs of channels on opposite modules have been considered for the analysis. Water is cooled down at the minimum achievable value, with the present setup, of 8^oC at the inlet. This corresponds to temperature at the substrate top surface of about 15 and 19^oC for the first and the second module respectively. The correlation between water temperature at inlet connection and the temperature of the SiPM sensors has been described in Section8.2.

The results are well in agreement with the expectations: operative temperature at the top surface in the order of about 20^oC, which corresponds to water temperature below 16^oC in the plot in Fig.10.6, contributes to improve the timing performance. All considered channel pairs achieve CRT resolution less than $225\,ps$ at FWHM. Lower temperatures do not give large contribution to the average timing resolution:

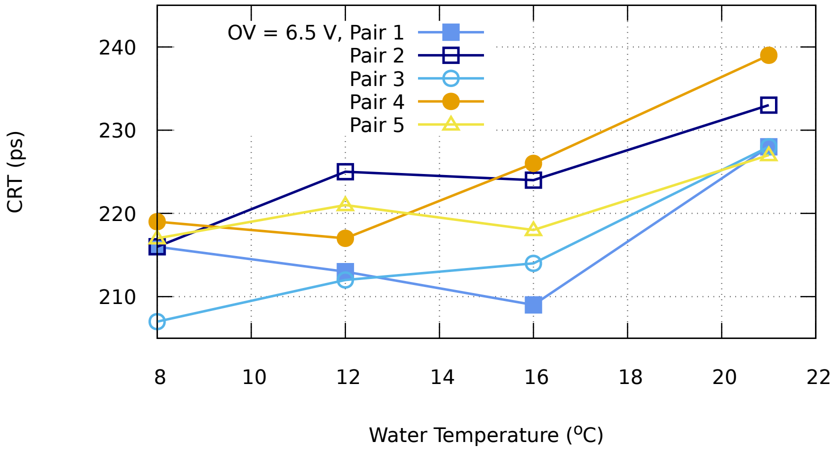


Figure 10.6.: Coincidence timing resolution at FWHM between five pairs of channels, for over voltage = $6.5\,V$, at different internal water temperature.

this is due to the temperature difference between modules and to the small temperature range analyzed with the setup. On the other side, measurements prove that the system can reach ToF performance using water at only $15^{o}C$, simplifying the design of a full ring system.

11

SMALL CRYSTAL IDENTIFICATION

The two detector solutions described in Section 8.3 for pre-clinical applications have been tested with the same scintillating array, for a better comparison of results. An array of 36 × 36 LYSO crystals has been coupled with SiPM matrix using transparent grease (refraction index about 1.5) in both cases.

11.1 LIGHT-GUIDE-BASED MODULE

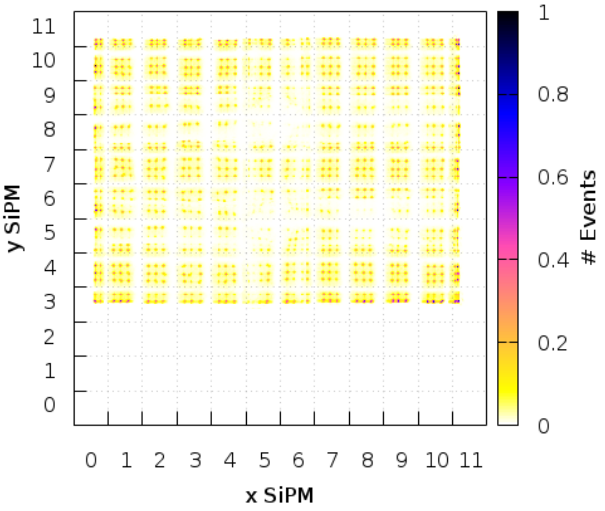
The module based on classical light-guide is composed of 12 × 12 SiPMs. Data have been acquired enabling neighbor groups of 18 channels per half chip. Such configuration corresponds to area of 6 × 3 SiPMs read out in parallel, if at least one channel in the group detects an event. For crystal array of $0.833\,mm$ pitch and light spreader $1\,mm$ thick, an area of 3 × 3 SiPMs around the master event has been taken into account for mathematical reconstruction of the original source position. If more than one master channel is detected, the event with larger integrated charge is considered as center of the interesting area.

The measured energy have been adjusted for baseline and for gain mismatch between channels, using a previous set of calibration data, obtained in a 1:1 coupling configuration with LYSO array with $2.5\,mm$ pitch. The weight function used for estimating the source position is based on Anger logic [136]:

$$\begin{cases} x_r = \frac{\sum_S X \cdot en_{XY}}{\sum_S en_{XY}} \\ \\ y_r = \frac{\sum_S Y \cdot en_{XY}}{\sum_S en_{XY}} \end{cases} \tag{11.1}$$

where S is the group of neighbor SiPMs taken into account for the reconstruction and en_{XY} is the energy detected by each of them, weighted on the respective position X and Y within the array. This reconstruction strategy shows problems at the edges of the array, where information is partial, due to missing neighbors.

The reconstructed flood-map is in Fig.11.1, together with the profile plot of one row. Totally, 36 × 36 LYSO crystals are placed on a 12 × 12 SiPM array. Ideally, and in case of a precise array placement, 3 × 3 scintillators should be reconstructed on each SiPM area. The detector used for the measurements includes one PETA not fully functional, resulting in missing positions in the lower part of the sensitive area.



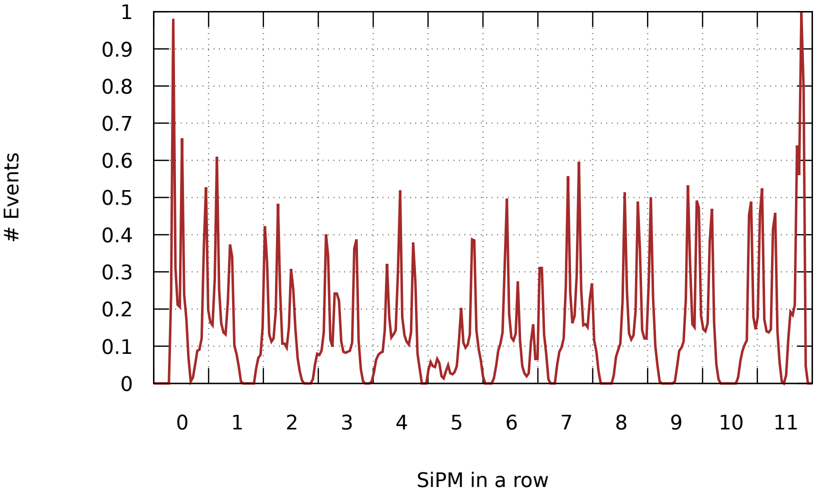


Figure 11.1.: Reconstructed flood-map and profile plot of one central row, using a light-guide-based module coupled with a LYSO array with 0.833 mm pitch. One PETA5 ASIC on the module is not fully functional, explaining the missing positions in the lower part of the map. No cut at the photo-peak has been applied to the data set.

From the profile plot it is easy to deduce that the reconstruction algorithm fails to identify crystals at the array edges: the two elements at the border are not distinguishable in the reconstructed data. Additional artifacts can be seen at the edges between left and right halves of the matrix, which correspond to left and right side of the same PETA. Chip halves are read out separately and do not share neighbor logic information within the chip. The light output is not always enough for triggering channels on both chip halves, and for this reason the same crystal at the border gets reconstructed in different positions, because the group of neighbors taken into account for the reconstruction is variable. In these positions, the algorithm generate more than 1 peak per crystal.

Similar effects appear at the borders between different readout chips: each PETA5 ASIC takes care of the read-out of 3 rows of SIPMs. From the flood-map, it is visible that, in most cases, light output is large enough to trigger two groups of neighbors, one per chip, and allows a proper reconstruction. In the corner between groups, when more than two triggered events are necessary for enabling the neighbor logic in all the interesting areas, spatial resolution degrades and most events get discarded on energy base. Trigger could be forced with lower discriminator threshold in the corner areas. All artifacts based on energy mismatches could be corrected for with better gain calibration per channel.

Even including all these artifacts, crystals in the array are clearly identified, showing the good spatial resolution of the module. Moreover, flood-map has been reconstructed from data within the full energy spectra. Proper calibration per channel is necessary in order to estimate the photo-peak position of each crystal and select only annihilation events. Calibration data obtained coupling the SiPM array with a pixelated LYSO array in a 1:1 configuration have shown to be not precise enough for this purpose.

The module has already achieved very promising spatial resolution, but there is still room for improvement in the next future.

11.2 ISIPM-BASED MODULE

The ISiPM-based module has been coupled with the same array of 36×36 LYSO crystals with $0.833\,mm$ pitch. Neighbor logic is programmed per sensor: four channels are connected to the same group but do not share trigger information with neighbor ISiPMs. No light guide is necessary and the array is directly coupled with the sensors with optical grease. Bonding wires at the array edges in the prototyped module are not protected and the scintillator block has been placed manually on the sensor matrix. For this reason, small misalignment between ISiPMs and crystals have been considered acceptable and have not been corrected for during data acquisition, in order not to damage the prototype.

The reconstructed flood-map is shown in Fig.11.2. Each ISiPM should be ideally coupled with 9×6 crystals. Along the horizontal direction, 9 peaks per sensor are easy identifiable in the flood-map, while along the vertical axis one crystal per ISiPM is missing. This effect is simply due to a bad alignment, that causes light from a crystal to be spread on two ISiPMs, which are read out separately in the presented example. It can be completely ignored for the present purpose, since it does not depend on an ISiPM limitation in resolution and could be easily solved with a more precise assembly. In the array, it is also visible

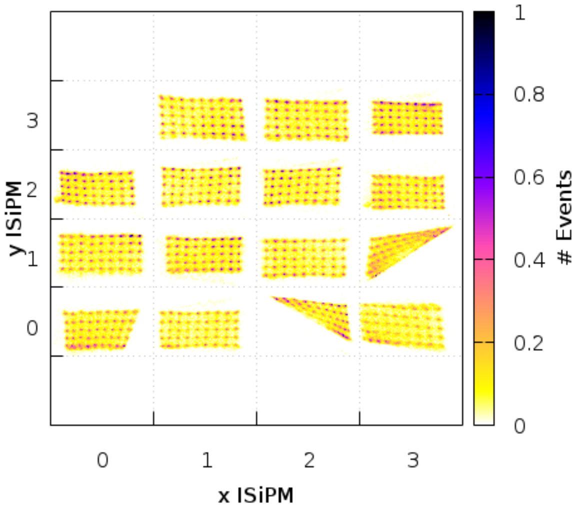


Figure 11.2.: Reconstructed flood-map using an ISiPM-based module coupled with LYSO array with 0.833 *mm* pitch between crystals.

that an ISiPM is not operative and that two channels, for two different sensors, are not read out, resulting in deformed reconstruction.

In order to evaluate the reconstruction quality for the ISiPM topology, a single crystal row is analyzed as example. The profile plot is shown in Fig.11.3: data are projected on the horizontal plane and fitted with a sum of Gaussian functions, one per crystal. The peak-to valley ratio of the fitted curve for the selected row is about 13 considering only events included in the photo-peak. The contribution of low energy events degrades the value down to about 4. The fraction of bad reconstructed positions is estimated from the fitted curves: the intersection points between the Gaussian shapes are used for defining the crystal borders. For each fitted curve, the fraction of integral area belonging to neighbor crystals is considered to be associated to false reconstructed positions. In such conditions, for the whole row, the amount of false reconstructed events is $< 1.5\%$ ($< 3.5\%$ if no energy cut is applied on data) and the error is always at maximum corresponding to one crystal position. Considering that such good results are obtained for already a sub-millimeter crystal pitch, it is clear that the device has great potential for high spatial resolution applications.

Since readout scheme is programmed not to share trigger information between neighbor sensors, light losses are expected for events happening at the borders of the sensitive area. In these cases, part of the emitted light is detected by the neighbor ISiPM, but not included in the data processing. The largest effect of such inaccuracy is a shifted energy spectrum for crystals placed at the borders. The map reported in Fig.11.4-left shows the detected position of the 511 keV peak for all crystals on an ISiPM (9×5 positions), as a fraction of the central

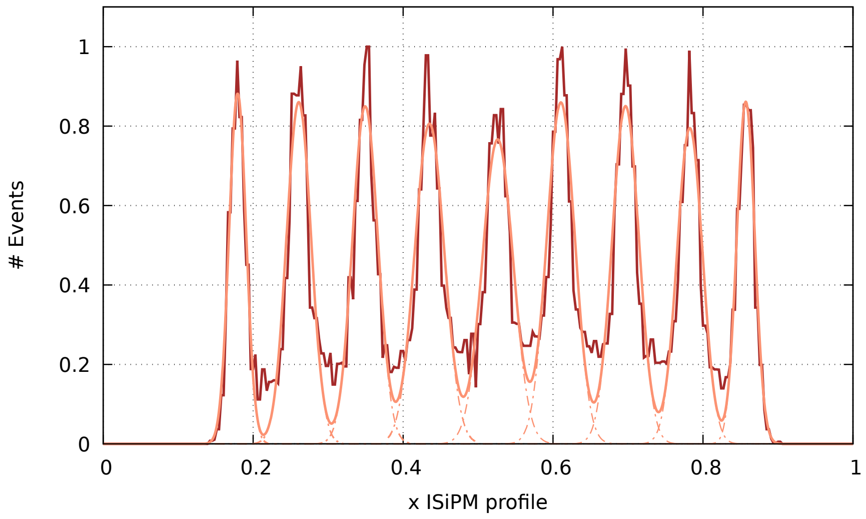


Figure 11.3.: Profile plot of one reconstructed row of crystals. The profile is fitted as a sum of Gaussian curves.

crystal data, used as reference. Crystals have been identified fitting a matrix of Gaussian peaks with fix pitch in x and y direction. Elements at the edges are clearly associated with lower detected energy. Such effect is less visible for crystals placed along the horizontal borders: it can be again explained with the misalignment of the array in that direction, where crystals are not placed exactly at the limit of the sensitive area and are less affected by light losses. In the same figure, on the right, the distribution of the energy resolution per channel is also mapped as a function of the crystal position within the array. In this case differences are negligible, showing that photon losses do not affect the final resolution and can be neglected. Overall energy resolution dE/E of about 16.7% at FWHM has been achieved in average on an ISiPM, which is already a good result considering the small size of the scintillating crystals used for the measurements. Due to the very large number of cells in a sensor (up to about 60000) and the low overvoltage applied for the measurements (only 3 V), both the SiPM and the readout electronics are reasonably operating within their linear range.

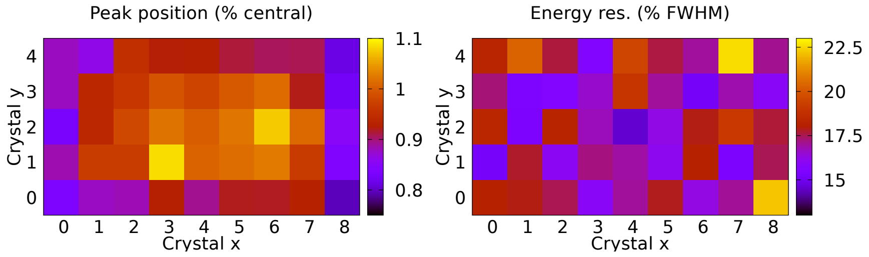


Figure 11.4.: Photo-peak position per crystal on an ISiPM (9 × 5 reconstructed elements), referred to the central crystal (left). Energy resolution per crystal (right).

Energy is estimated summing up the four amplitudes from an ISiPM for each event. Mismatches between channel gains lead to degradation of the energy (and spatial) resolution. A preliminary calibration of the amplifier gain for all channels in a sensor could then lead to an additional improvement in the energy resolution.

An additional consideration can be done for understanding the relation between spatial and energy resolution for position sensitive SiPMs. Performance of the device can be adjusted allowing, or preventing, light spreading on the sensitive area. For improving energy accuracy, the highest number of cells should trigger the emission event. Spreading light from a crystal on many diodes affects the position reconstruction as well, when the cluster is not fully detected by the sensor: the asymmetry at the borders introduces an error, resulting in a reconstructed position closer to the sensor center and an overall degradation in the detected peak-to-valley ratio. As example, two flood-maps obtained with the first square ISiPM prototype are reported in Fig.11.5: measurements have been acquired in both case with a 6×6 LYSO array with crystal size of $1\,mm \times 1\,mm \times 10\,mm$ each, optically decoupled with ESR layer and with an overall pitch of about $1.13\,mm$. The sensitive area is $7.4\,mm \times 7.4\,mm$ and light emission has been manipulated using different materials on the array top surface: aluminum foil for maximizing reflection in one case, and black tape for reducing it in the second one. The detected pitches, in both x and y direction, is $p_x = 0.52 \pm 0.09\,mm$ and $p_y = 0.74 \pm 0.06\,mm$, in case of maximal light collection with aluminum foil on top. In case part of the photons get absorbed by the black top layer, the squeezing effect toward the center is less evident, with detected pitches of $p_x = 0.64 \pm 0.10\,mm$ and $p_y = 0.82 \pm 0.04\,mm$. On the other side, energy resolution is clearly better if more photons are emitted per event: in the first case, the resolution at FWHM is about 17.9%, while it degrades to about 20.1% in case of partially absorbed cluster. Some artifacts in the reconstructed flood-map are related to defects in the LYSO array, such as reflector

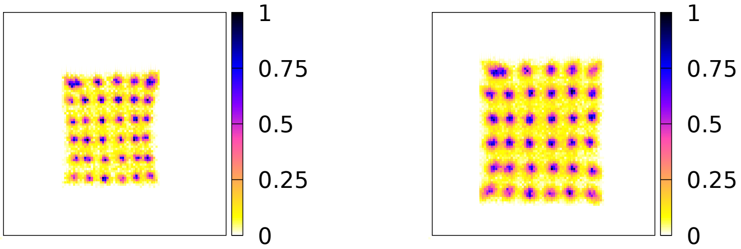


Figure 11.5.: Cluster size effects on ISiPM position reconstruction: the flood-maps are obtained with the same LYSO array (6×6 crystals, $1.13\,mm$ pitch). The lateral surfaces are optically decoupled with ESR reflector. In the left case, the top side is covered with aluminum foil, in the right one is covered with black tape for avoiding light reflection.

layer missing between two crystals, or bad positioned reflective layers, and do not depend on the ISiPM technology.

Thanks to the good results in crystal identification, the module has been used for testing the DOI capability of the system. The most obvious strategy for introducing information on the radial gamma absorption position is to assembly two layers of scintillating crystals on top of each other, with shifted centers of gravity. The reconstructed position directly encodes in this way also the DOI information.

Data have been acquired with two LYSO arrays of 5×5 crystals, with an overall pitch of $0.8775\,mm$. The array, with total area of $5\,mm \times 5\,mm$ is coupled with one ISiPM sensor in the matrix through optical grease. This configuration maximizes the light compression on the sensitive area, with the previously described effects on spatial resolution. The reconstructed flood-maps, using a single array of crystals and the combination of two layers, are shown in Fig.11.6 Without taking into account the predictable effects at the array edges, the ISiPM is accurate enough to distinguish crystals closed to the array center. In such areas, three rows of crystals are well separated, showing that the device can ideally identify small scintillating crystals placed at very short distance. It can be assumed that, even if manually assembled, the center of gravity of crystals in the two layers are placed at about $0.8755/\sqrt{2} \approx 0.62\,mm$, proving once more the very promising spatial resolution offered by the ISiPM topology.

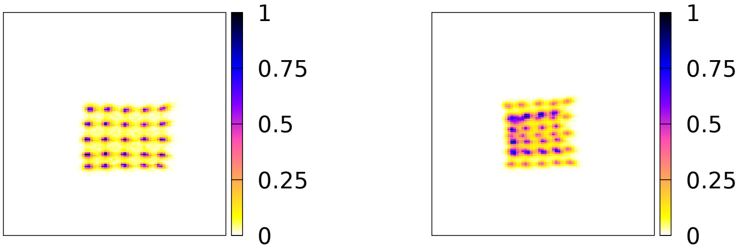


Figure 11.6.: Flood-maps obtained with one and two layers of LYSO arrays of 5×5 crystals with $0.8775\,mm$ pitch. A part from expected boundary effects, ISiPM-based module is able to identified crystals assembled on two shifted layers, with a distance between centers of gravity of about $0.65\,mm$.

11.3 EVENT RECONSTRUCTION WITH IDP

IDP chip has been coupled with LYSO arrays for proving its ability in reconstruction of emission events. Measurements here presented have been acquired by Roberto Blanco with IPD1, in slow readout operation mode: the full frame has been read out per event. In order to discard noisy events, only frames with multiplicity ≥ 4 have been taken into account. The accumulation time is 200 ns. Flood-maps resulting by summing all output frames are shown in Fig.11.7: they have been acquired with two LYSO arrays with 0.8775 mm and 0.4775 mm pitch. Crystals are optically decoupled with an ESR layer. The higher photon count in between crystals is related to the fact that diodes at the boundary detect events emitted from both elements.

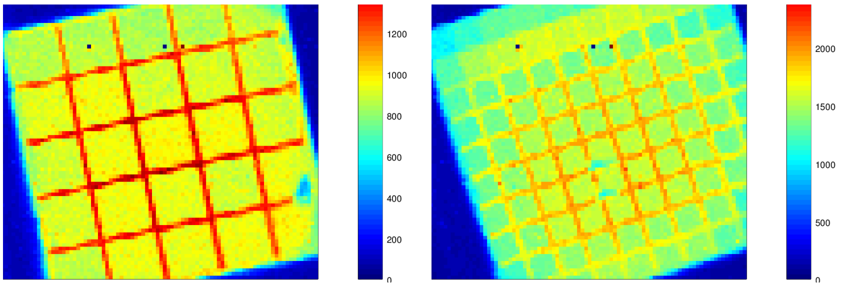


Figure 11.7.: Flood-maps of two LYSO arrays reconstructed with IDP1: full event frames are read out in slow readout mode. Array pitches are 0.8775 mm (left) and 0.4775 mm (right). Hot pixels have been discarded from the flood-map.

The great potential of the device can be even better deduced looking at single event frames, reported in Fig.11.8, for the array with smaller pitch. The light distribution for every emission event can be studied, since position of all triggered diodes in the array is read out during full frame readout mode. Three cases are reported as example in the figure: a 511 keV absorption from annihilation event from ^{22}Na, high energy absorbed event from gamma decay in ^{22}Na (peak at 1.27 MeV) and a Compton scattered gamma re-absorbed at lower energy in a different crystal. It is also visible how, for such small crystals, the photo-peak events saturates the SiPM area below the scintillator, corresponding to only about 100 avalanche diodes. Event energy can still be estimated using the light spread on the whole sensitive area.

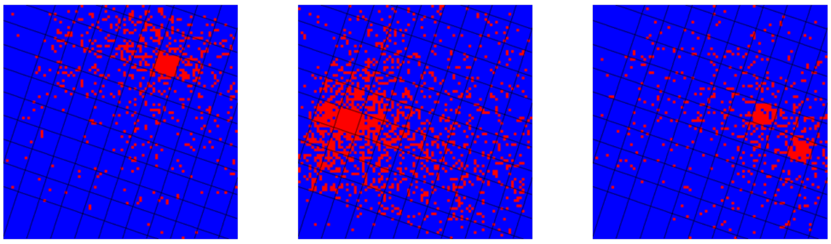


Figure 11.8.: Full event frames acquired with IPD1: the three frames correspond to a 511 keV absorption (left), high-energy absorption (center), and a Compton scattered gamma, re-absorbed at lower energy in a different crystal (right).

12

PERFORMANCE AT HIGH DATA RATE

The modules have been tested in high data rate environment in collaboration with ETH Zurich, Switzerland, as a possible option of the SAFIR project [137]. SAFIR (Small Animal Fast Insert for mRi) goal is the design of a PET insert for small-animal imaging, MR-transparent and compatible with existing MR scanners. The project targets good spatial resolution, high sensitivity and exceptional timing resolution, in the order of $300\,ps$ CRT at FWHM. The operations will be at about $500\,MBq$, ten times higher than the pre-clinical practice, and aiming to resolve temporal evolution of the tracer concentration.

Measurements with modules in 1:1 configuration, mounted on the back-to-back coincidence setup, has been taken. The pixel size is comparable to the SAFIR requirement, where pixels will be $2\,mm \times 2\,mm$ in sensitive area. A vial long about $30\,mm$ and filled with $^{18}F - FDG$ solution, at initial activity of $567\,MBq$, has been placed in the field of view between two modules. Data have been acquired at fixed temperature of about 35^oC, because the internal liquid cooling infrastructure was not available at the facility during test phase. For such a reason, only one chip per module has been biased during measurements, reading out a total sensitive area of about $30\,mm \times 7.5\,mm$. Data sets used in the following analysis have been acquired at random time intervals, for a maximum of 15 minutes each one, so that the detection rate can be considered constant for each set (^{18}F has half life of about 110 minutes). Modules are placed at about $55\,mm$ distance from the vial, which lays horizontally in front of the sensitive area.

Expected rate per channel for such configuration has been calculated with Monte Carlo simulation. An attenuation coefficient for LYSO crystals of $0.87\,cm^{-1}$ for $511\,keV$ gamma particles has been considered. The vial has been approximated with a one-dimensional linear source with uniform activity density. Only events included in the photo-peak ($\pm 2\sigma$) have been taken into account for the analysis. It has been assumed that, on a whole sodium spectrum, photo-peak events represent 30% of the all emission events.

Data have been acquired in continuous readout mode, as described in Section 9.1.1, at fixed time intervals of $6\,\mu s$. Results are summarized in Fig.12.1. One line shows the approximated theoretical event rate obtained with Monte Carlo simulation, based on geometrical considerations on the setup and averaged along the central crystal row.

Measured rate at difference source activities is averaged on all active channels and is shown as well in the plot. The ideal rate per channel, based on a simplified geometrical model, is used only as proof of principle and shows good agreement with the measured rate. Data measured for activity less than $150\,MBq$ have been used for fitting the actual linear rate $r_l(A)$, depending on the source activity A. For higher activity, effects due to the readout dead time are expected in the acquired data. For this analysis, the readout system has been considered to be affected by a non-paralyzable dead time, meaning that events happening during dead time get lost and do not add additional dead time. Assuming that the source decay rate follows a Poisson distribution and knowing that frames are read out every $\tau = 6\mu\,s$, the non-paralyzable rate at high activity can be approximated using the formula:

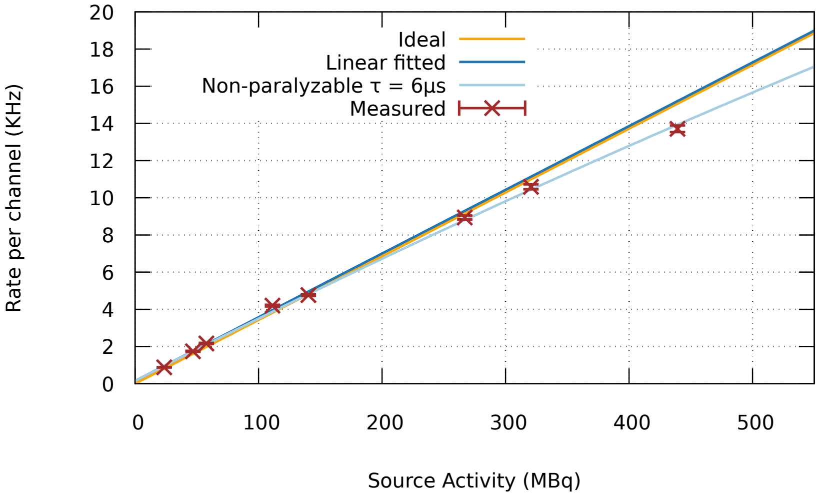


Figure 12.1.: Expected and measured rate per channel. Three expected curves are represented in the plot: a theoretical rate simulated from geometrical considerations on the setup, a linear rate fitted using only the data acquired with activity $\leq 150\,MBq$ and the expected rate for non-paralyzable read-out with dead time equal to $6\,\mu s$. The measured detection rate is averaged on all active channels.

$$r_{np}(A) = \frac{r_l(A)}{1 + r_l(A)\tau} \tag{12.1}$$

Both the linear rate and the non-paralyzable rate are plotted as well in Fig.12.1. The non-paralyzable function fits well the measured data: taking into account the residual values, their standard deviation is about $180\,Hz$, very good for the approximations considered for the analysis. The fit demonstrates that the readout dead time is the most limiting factor for the system sensitivity. With the present geometry, a source of about $500\,MBq$ at the center of the field of view generates more than $17\,kHz$ event rate per channel and PETA5 readout can afford up to about $15.5\,kHz$, with event loss of about 9%. This result is already good but has to be considered as a very conservative prediction: chip design is in fact optimized for readout at $160\,MHz$. Moreover, reading out the shift register with zero suppression mode (not possible with PETA5 due to a design bug) would allow additional loss reduction of about $4-5$ times.

It has also to be pointed out that, even if PETA5 can cope with reasonably high data rate, the USB2.0 interface is not optimized for large bandwidth and drastically limits the acquisition capability. PETA6 has already been designed for coping with much higher data rate, for ideally no loss at the same conditions.

Another interesting consideration can be done on the sensor conditions when detecting high data rate. Current flowing through SiPM tiles has been monitored with a multimeter during acquisition. The registered value is referred to two detector modules, with totally 288 sensors biased in parallel and operating at $8\,V$ over breakdown. Measured current values are reported, depending on the source activity, in Fig.12.2. It is interesting to note that current per module can reach up to $9\,mA$ for a source at $500\,MBq$: considering that SiPMs are biased at about $36\,V$, a SiPM tile dissipates up to $350\,mW$, which is still a pretty low value for a whole module.

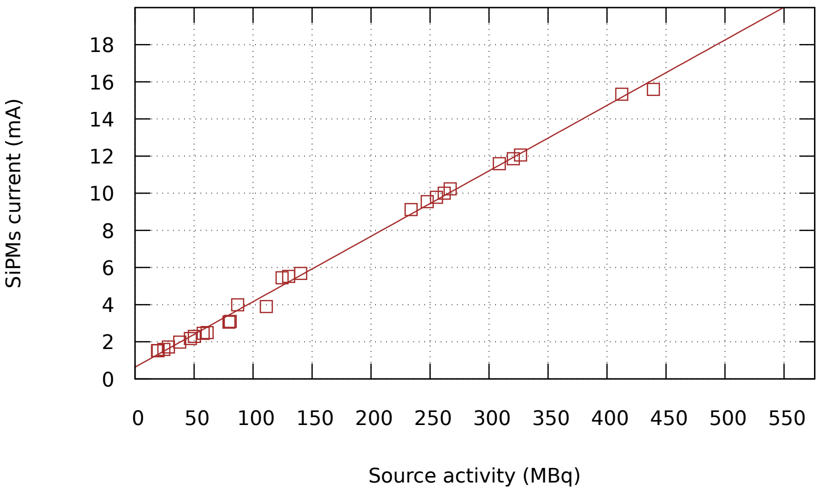


Figure 12.2.: Dependance on the SiPM current (two modules operating in parallel) on the source activity.

The timing and energy resolutions have also been monitored to prove the stability of the front-end at any data rate. Results are reported in Fig.12.3. The plot on the left shows the CRT at FWHM between a pair of channels on opposite modules, at different source activity, for corresponding data rate up to about $14\,kHz$ per channel. The standard deviation for different rate conditions is about $10\,ps$: such value depends on changes in temperature, which has not been controlled during data acquisition, and on different energy cut around the photo-peak. As visible from the plot, no clear dependance is noticeable on the data rate. The CRT mean value on the all range is about 263 ps at FWHM, which is worst than the detector performance shown in Section 10.2, but is affected by the high operation temperature. The right plot in Fig.12.3 shows the energy resolution for each detector, averaged on all working channels. The difference between the two modules is probably due to imperfections in the assembly in one case. Anyway, the resolution is constant for different activities, in the order of 10% at FWHM (not linearized), showing that the high rate does not affect the front-end performance.

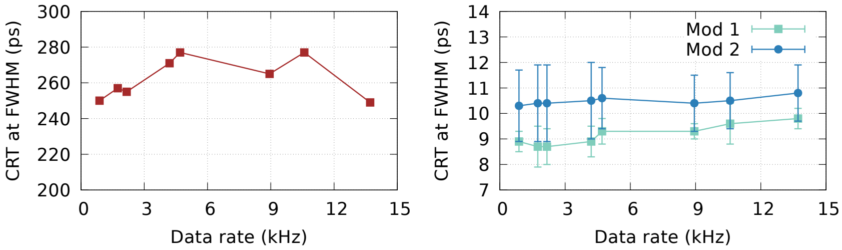


Figure 12.3.: Coincidence timing (left) and energy resolution (right) dependance on the data rate. Timing plot is referred to a pair of opposite channels. Energy resolution (dE/E at FWHM) is averaged on all working channels on two detector modules.

Part IV

CONCLUSION AND OUTLOOK

Chapter 13 summarizes project goals and measurement results. Finally, an outlook on the next steps already planned for operating the modules on larger systems and for further improvements is reported in Chapter 14.

13

CONCLUSION

The goal of this work was to design and test compact PET detector modules, with excellent timing and spatial resolution, compatible with operation in MR. Each detector component has been carefully optimized and modules in clinical and pre-clinical configuration have been fabricated and characterized.

The module is composed of a single LTCC substrate, with readout electronics and SiPMs on opposite surfaces and an internal liquid cooling channel for ensuring maximal thermal decoupling. All components on the assembly are not magnetic; signal routing on the board and within the ASIC has been studied for avoiding pick up noise from magnetic field gradients. The assembled module is extremely compact, with total height less than $1\,cm$ without scintillators. It can read out up to 144 channels, with a novel single-ended front-end architecture with low input impedance, in the order of 10Ω or less.

Modules for clinical applications have been coupled with $10\,mm$ thick LYSO array with pitch of only $2.5\,mm$. Coincidence timing resolution between modules is $232\,ps$ at FWHM at 20^oC. This value has been obtained using many lines of response between pairs of channels, proving the performance at system level. Similar resolution on whole modules in realistic configuration for operation within a scanner were not yet been proved at the time of writing, so it can be stated that the design reaches state-of-art performance. Comparable results with SiPM arrays have been achieved by other research groups, but only using discrete electronics, thin crystals and a single reference for coincidences ($206\,ps$ in [98]) or at extremely low temperature ($200\,ps$ at -25^oC in [74]).

Using reference SiPMs in HD-NUV technology by FBK, the measured coincidence timing resolution reached $182\,ps$ at FWHM, with still room for improvement in an optimized setup.

The choice of using scintillators only $10\,mm$ thick for a clinical module was justified by the required compactness of the module, in order to allow the insertion in the tiny gap between split gradient coils, proposed by Sublima project.

The dynamic range of the charge integrator included in the single-ended front-end has shown to be too small for very high SiPM over voltages (not predictable during design) and has already been adjusted in the next chip version.

Two options for pre-clinical applications with very good spatial resolution have been tested as well: both are able to identify crystals within an array with only $0.833\,mm$ pitch. Position sensitive ISiPMs have also the additional advantage of requiring only four output channels per event and no information from neighbor sensors, with possibly larger readout bandwidth. The topology is able to distinguish crystals at about $0.62\,mm$ distance.

Even in this case, the module has shown great potential and results are amongst the best achieved with similar configuration. Pixelated scintillators with $0.53\,mm$ pitch have been identified by [92], but using discrete electronics readout with low input impedance.

Chip power consumption (about $43\,mW$ per channel) is reasonable, but proper cooling infrastructure is necessary. The channel integrated within LTCC is a very efficient solution, but needs careful design on a large system in order to avoid leakage and ensure sufficient flow per module. The next PETA version has already been designed with lower power consumption (about $10\,mW$ less per channel).

MR compatibility has been considered during design, but performance stability must be proved in magnetic field environment.

The designed module offers a compact alternative with outstanding resolution for both clinical and pre-clinical applications. Thanks to the excellent performance demonstrated with the detectors, PETA is being considering in preliminary phase by three research groups at the moment, for the design of MR-compatible PET inserts (small-animal PET scanner within project SAFIR [137], brain PET scanner at Stanford University in Craig Levin's group [138], breast PET scanner at RWTH Aachen University in Volkmar Schulz's group).

14

OUTLOOK

Single-layer substrate has proved to be a very valid option and, for this reason, next module version will be based on the same design.

A new PETA version (PETA6) has already been designed and fabricated. First characterization measurements are presently in the preliminary phase and have not been included in this thesis. Bump bonds layout is compatible with previous versions, so that new chips can be assembled on ceramic substrate without modifications of the module design. The main features of the new ASIC, compared with the previous PETA5, are summarized in Table 14.1.

	PETA5	PETA6
SE Polarity	Negative	Negative/Positive
SE power Consumption (mW/ch)	43	33
Readout Frequency (MHz)	160	320
Bypass readout	Yes (bug)	Yes
Channel data (bit)	37	34
Double buffer	No	Yes
LVDS outputs	2	2 (+2)
Temperature Sensor	No	Floating diode

Table 14.1.: New features in PETA6, compared with the previous PETA5 version.

PETA6 has been designed aiming to high data rate operations. The channel data have been shortened to 34 bits and the readout shift registers can be emptied at twice the frequency compared with PETA5. A second optional LVDS serial output is available per half chip, in case the number of routed lines is not an issue on the assembly. The channel includes an additional internal buffer for storing one event on-chip and reducing dead time. It has been estimated that the front-end needs about $300 - 500\,ns$ for completing the decision process for an event, depending on the energy integrator settings. Simulation results show that, in these conditions, the new features can allow data rate, in best cases, of about $200\,kHz$ per channel with losses lower than 5%.

In addition, a new circuit for detecting the PLL operation condition (locked/not locked) has been designed by Michael Ritzert and integrated on chip. Termination resistors for LVDS signals have been integrated, for reducing the number of external components necessary on board. A floating diode can be used for measuring chip temperature.

The single-ended dynamic range has been adjusted for larger current injected into the channel. A more uniform power distribution, with a reduction of about $10\,mW$ per channel, has also been achieved.

PETA6 has been designed in two versions, for reading out both SiPM polarities with the single-ended front-end stage. In this way, the technology is ideally compatible with any SiPM available nowadays on the market. Single-ended front-end with inverted polarity is based on the same architecture, but needs to be characterized for fully proving its performance.

An important step to address in the next future is the characterization of the detection modules in MR environment, to fully prove their compatibility.

In order to operate the modules in ring configuration, the design of a processing unit board hosting many blocks is needed. A preliminary architecture has already been investigated. An unit with fast Ethernet interface would host 6 modules, placed in a 2×3 configuration. Each pair of neighbor detectors is read out with a dedicated FPGA. Precise synchronization between units through optical link is an option. Infrastructure for parallel cooling of many substrate has to be integrated on the system.

Such setup would be a big step forward the assembly of a full prototype scanner, compatible for both clinical and pre-clinical applications, and would allow to investigate the improvements in image reconstruction using the optimized modules.

Part V

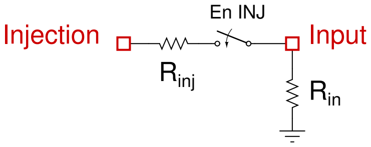
APPENDIX

A

PETA CHARACTERIZATION

Chip features and performance have been tested on dedicated setups for all versions, using analogue pulses injected through the available test path. Results presented here have been obtained with PETA5 (where not differently specified) and aim to a complete characterization of the single-ended front-end architecture.

For what concern the input impedance evaluation, measurements have been taken with the test chip ISIS1, where the resistivity from the test point to the chip pad is minimal, if compared with the full assembled setup available for PETA chips. Considering one channel, a voltage pulse ΔV_{inj} is injected via the test path enabling the corresponding low impedance switch. A schematic of the input stage and resistances is in Fig.A.1, where the red squares represent the available test points on the PCB.



$$R_{in} = R_{INJ} \cdot \frac{\Delta V_{in}}{(\Delta V_{in} + \Delta V_{inj})} \tag{A.1}$$

Figure A.1.: Simplified circuit for the estimation of the input resistance. Red pads are accessible from the PCB.

Measurement of the voltage drop at the input pad ΔV_{in} for a known voltage pulse at the injection ΔV_{inj}, allows the estimation of the input resistance of the channel, as expressed in Eq.A.1. With such configuration, the input resistance for the regulated cascode has been measured to be about $7 - 8\,\Omega$, very close to the simulated value of $3 - 4\,\Omega$.

The input DC voltage, defined by the reference DAC, has been measured at the test injection pad, enabling a single channel at time. A full range measurement is reported in Fig.A.2: it shows, consistently with the simulation results, that the DC input can be changed within a range of about $800\,mV$, before reaching saturation regions for both high and low DAC settings. The two plateau correspond to conditions which must be avoided during operation: in case of very low input volt-

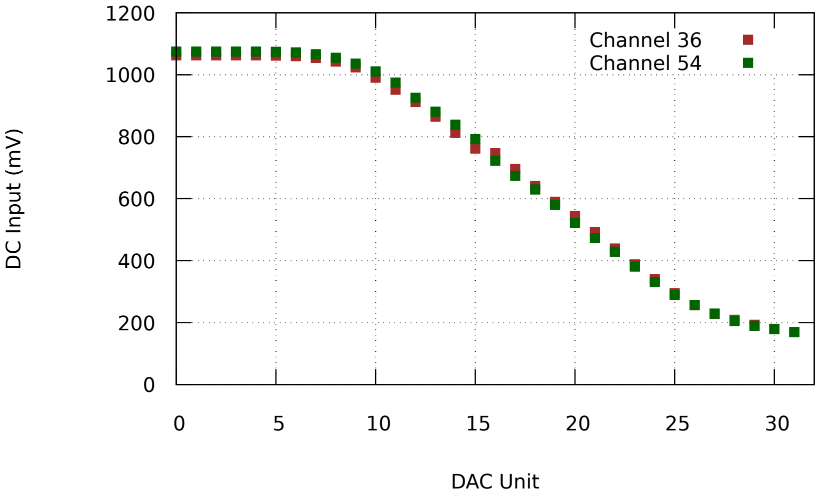


Figure A.2.: Measured input DC voltage for different settings of the input DAC.

age, the *nmos* providing the bias current (I_{bias} in Fig.6.2) is no more in saturation region. In case of high input voltage instead, the regulation amplifier does not operate in its designed range.

The discriminator noise is studied per channel: common differential reference is kept at the fixed value defined by the internal resistor divider, while the fine threshold per channel is trimmed using the current DAC. The discriminator result has been studied for differential threshold in a range of $-100/+100\,mV$ and reported in Fig.A.3. In the plot, blue strips represent regions where the discriminator has triggered a noisy hit. The y-axis reported the theoretical differential threshold for each channel, calculated from the DAC setting. It is clear from the

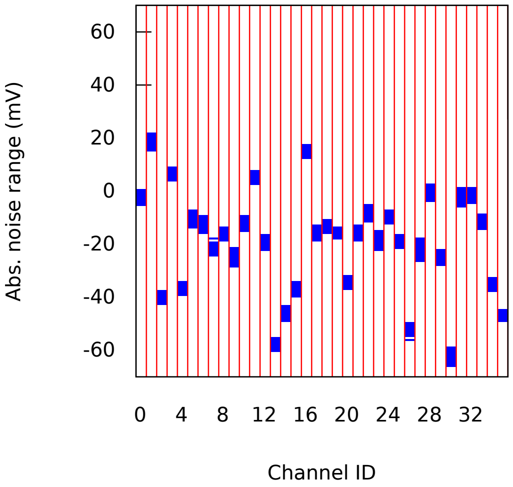


Figure A.3.: Absolute noise level per channel at the discriminator. Blue strips are areas where the discriminator triggers a noisy hit.

plot that the common mode level is different for every channel and does not correspond to the theoretical zero-level, with a maximum offset of about $60\,mV$. For this reason, a proper threshold calibration per channel is necessary in order to use homogenous threshold settings during data acquisition. The minimum differential threshold per channel is homogenous on the chip and is about $10\,mV$.

The energy integrator linearity has been characterized with voltage pulses injected through the test path. As stated in equation 6.13, due to the protection diodes at the pad, the maximum current that can be injected into the channel is about $1\,mA$ (depending on the DC input level of the channel). The integrator measures injected charge, which depends on the voltage pulse as follow:

$$Q_{inj} = \frac{\Delta V_{inj}}{R_{inj}} \cdot \Delta T_{inj} \tag{A.2}$$

where ΔV_{inj} is the height of the input pulse, ΔT_{inj} its width and R_{inj} the resistance of the injection path (about $1\,k\Omega$ by design). Fig.A.4-left shows the integrator results for different injected charge. Both width (curve ΔT) and height (curves ΔV) of the pulse have been varied for the measurements. The saturation due to the protection diodes is visible when the injected current is too high. For higher DC input of the channel (v_{in} in the figure) such effect appears later. The integrator shows linear behavior for very wide signal and large charge: the plot is obtained with pulse up to $1\,\mu s$ long.

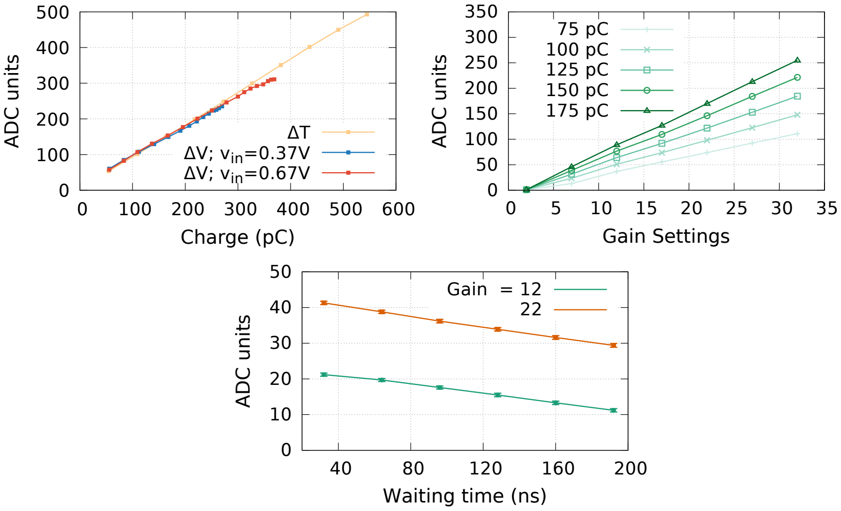


Figure A.4.: Characterization of the energy integrator and digital conversion. On the left, measurements with injected voltage pulses at different width and height. On the right, examples of fix charge signal injected and integrated with different injection gain. Below, effects of different waiting times before discharge starts for an injected charge of about $50\,pC$.

The right plot shows instead the effect of the ADC gain at the input mirror. Fixed charged signals are injected into the channel and integrated with different gain settings, proving the linearity of the injection circuit. In Fig.A.4- below, the effects of different waiting times on the digitized value are reported, for two gain settings and a injected charge of about 50 pC, pulse width of 50 ns. Leakage current affects the voltage at the feedback capacitor, once the pulse is over, causing a drift of about 10 bits in the ADC value during the whole available range, between 0 and 200 ns. As visible from the plot, the loss is constant for different gain settings and the final effect is a small shift which is considered acceptable.

The effects of discharge current (equal to I_{ramp} in the circuit in Fig.6.10) have also been taken into account. Larger currents allow faster de-integration times and are therefore preferable for shortening the channel dead time. Examples of results obtained for different injected charges and discharge currents, are reported in Fig.A.5. The relative error is also shown in Fig.A.5-right, proving that the ADC configuration can be adjusted with stable performance for a large range of injected charge and discharge times.

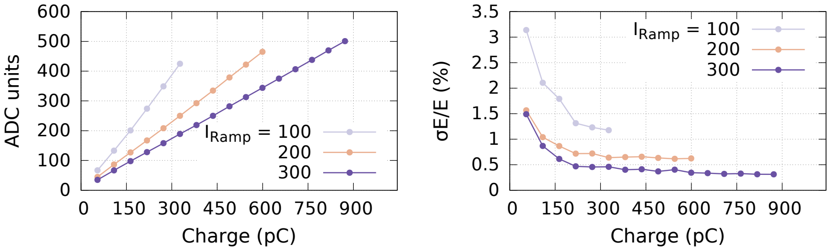


Figure A.5.: Effects of discharge current on measured energy, for different injected charge. The relative error is constant over a large range, except for very small integrated charge, and gets better for larger discharge currents. I_{ramp} is expressed as DAC units.

Consideration on the channel dead time are also necessary when configuring the integrator parameters. In all measurements, the integrator parameters has been configured for detecting the photo-peak within the lowest 150 bins, for a discharge time of maximum about 450 ns.

In order to estimate overall channel dead time (hit detection, integration waiting time, capacitor discharge and final state machine decision), multiple pulses have been injected at short time distances in a channel. Channel dead time can be estimated in this way: a low energy event is pulsed within the channel and energy threshold is configured for being just above the injected value. After a short delay, a second pulse with higher energy, which would not be discarded by the ADC threshold, is injected. If the two pulses are separated well enough and channel dead time is over, the higher energy event is detected. Fig.A.6 shows an example of double pulses injected per channel: the plots represent the percentage of detection of the second pulse, when the first pulse in the channel gets discarded by the energy threshold condition. Three sets of data for different discharge times, obtained with variable injection gain at the integrator input, are represented in the plots. The left plot shows the result depending on the absolute time difference between

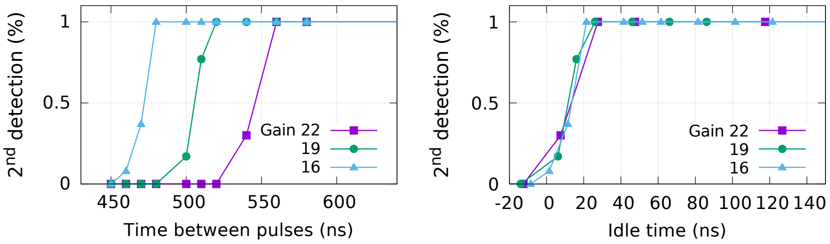


Figure A.6.: Channel dead-time estimation using double pulses in a channel with different integrated charge. The first pulse is discarded with energy threshold veto mechanism and the second one is read out. y-axis in both plots represent the percentage of detection for the second event: on the left is dependent on absolute time difference between input pulses, on the right is dependent on idle time.

pulses. The channel dead time depends on different factors: the waiting time (programmable) between the hit detection and the discharge start, the de-integration time and the reset time (programmable). The rejected pulse is always corresponding to input charge of $120\,pC$, while the detected one is $300\,pC$. The processing time for the first event can be estimated, knowing the programmable waiting and reset times, the energy digitized value and the decision process time necessary within the internal state machine. Such approximated value is subtracted from the time difference between the pulses and defined as “idle time”. If the idle time is larger than zero, the channel should always be able to detect and store the second pulse, while for idle time lower than zero, the second hit is received during channel theoretical dead time. Right plot shows again the detection percentage of the second hit, as a function of the estimated idle time within the channel. Measurements prove that the channel is ready almost immediately after the event suppression and that dead time can be reduced with a proper channel configuration.

Finally, timing performance of single-ended front-end have been studied. TDC design sets a limit to the best possible coincidence resolution achievable by the system, which is expressed as:

$$CRT_{TDC} = \frac{b}{\sqrt{12}} \cdot \sqrt{2} \cdot 2.35 \approx 0.96 \cdot b \tag{A.3}$$

where b is the TDC bin width, in PETA equal to $50.2\,ps$. This leads to a minimum CRT value of about $48ps$ at FWHM.

Intrinsic coincidence resolution for a pair of channels is shown in Fig.A.7. Channels read out for these measurements belong to different chip halves, meaning different VCO and PLL circuits. No appreciable difference has ever been noticed between halves in the same chip and between different ASICs. Bin-width correction algorithm is used for linearization of the VCO delays between stages, leading to CRT of about $77\,ps$ at FWHM in the best case. Results have been achieved with internal differential threshold set at about $30\,mV$ above the discriminator electronic noise. The measured CRT is larger than the theoretical

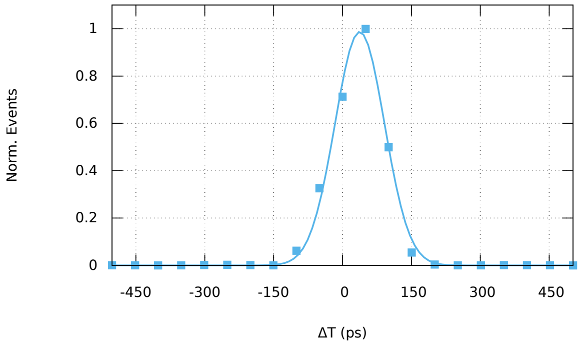


Figure A.7.: Coincidence timing resolution between two PETA channels using external voltage pulses. CRT is 125 *ps* before bin-width correction and 77 *ps* at FWHM after.

value, meaning that improvements are still possible. A different injection circuit could possibly lead to better results.

B

PERSONAL CONTRIBUTION

My contribution to the presented work is here listed in detail. Studies, designs and results not included in the list are to be intended as someone else's work.

DESIGNS AND MEASUREMENTS

- Test chip ISIS1
    - Design, simulation and optimization of different single ended front end architectures and integration in test chip ISIS1;
    - Design and assembly of a dedicated PCB for characterization of ISIS1;
    - Software and firmware update for operating the test setup. Original version already available by Michael Ritzert;
    - Design of a TDC integrated on Xilinx Spartan6 FPGA, with bin-width of about 50 *ps* and intrinsic resolution of about 50 *ps* CRT at FWHM. External TDC is necessary for measuring timing resolution from the one-wire encoded information read out from ISIS1;
    - Characterization of ISIS1 and measurements with SiPMs coupled with LYSO crystals in coincidence configuration;

- PETA4 and differential module
    - Selection of the most performing front-end from ISIS1, optimization of it and integration on the final design PETA4;
    - Collaboration with MSE GmbH, for designing a LTCC substrate for two PETA4 flip-chip mounted and connected in differential-ended operation mode;
    - Support for the design of a PCB dedicated to the differential module readout (design by Roberto Blanco);
    - Characterization measurements of PETA4 with differential and single-ended front-ends;
    - Collaboration with MSE GmbH, for modifying the existing LTCC substrate for differential-ended PETA4 and making it compatible with single-ended operation mode. Design of an adapter board for testing the modifications and coincidence measurements with SiPMs and LYSO crystals;

  - Design of an adapter board for reading out ISiPM prototypes with differential-ended front-end PETA4 and related measurements;
  - Software and firmware update for operating new test setups (original version by Michael Ritzert);
  - Implementation of software for crystal position reconstruction using ISiPM dies read out with PETA4;

- ISiPM characterization
  - Reconstruction of LYSO arrays with ISiPM prototypes (version 1 and 2) on a dedicated test setup (setup designed by Peter Fischer)
  - Studies on light sharing between scintillators and its effect on center of gravity reconstruction using ISiPM prototypes. DOI measurements with two layers of scintillating crystals and investigations on achievable limit in spatial resolution;
  - Implementation of software for data analysis and evaluation of position reconstruction and energy resolution per crystal;
  - Implementation of simulation algorithm for light transmission within a monolithic crystal coupled with ISiPM;

- PETA5 and high-density module
  - Optimization of single-ended front-end and integration in PETA5 design;
  - Collaboration with MSE GmbH, for designing of LTCC with 4 PETA5 flip-chip mounted connected in single-ended operation mode;
  - Design of test PCB for one LTCC module with four PETA5;
  - Software and firmware update for reading out four PETA5 (original version by Michael Ritzert);
  - Characterization measurements of PETA5;
  - Studies on temperature effects using internal water cooling; assembly of the cooling infrastructure with help by Christian Kreidl;

- Measurements with clinical and pre-clinical modules
  - Communication with Proteus Inc. for different LYSO arrays fabrication; arrays mounted on the module by Christian Kreidl;
  - Design of composite test setup for coincidence measurements between two detector modules; software and firmware update for parallel readout of 8 PETA5;
  - Coincidence measurements between detector modules;
  - Timing measurements with FBK tiles, with adapter board on LTCC substrate, designed by Alberto Gola, FBK;
  - Studies on achievable limit in timing resolution with SiPM read out with single-ended front-end in PETA5;

  - Studies on position reconstruction of LYSO arrays: measurements performed with standard detector module with light spreader. Studies on the effects of the neighbor logic assignment and of different reflective layers between crystals on the position reconstruction. Position reconstruction algorithm adjusted from an original version by Michael Ritzert;
  - Studies on position reconstruction of LYSO arrays with a matrix of ISiPMs, read out with PETA5 via adapter board and implementation of dedicated position reconstruction algorithm;
  - Studies on position reconstruction of LYSO arrays using LG-SiPMs by FBK and implementation of position reconstruction algorithm;
  - Implementation of Monte Carlo simulation for the estimation of data rate per channel in coincidence configuration, for a linear source placed at the center of the FOV;
  - Supervision of bachelor thesis on the characterization of PLL circuit within PETA5;

- Design of IDP1/IDP2
  - Design, simulation and optimization of multiplicity tree logic in IDP1 and IDP2;

- PETA6 and dedicated test setup
  - Optimization of single-ended front-end for power reduction and larger dynamic range; integration in PETA6- Version A design;
  - Design and simulation of single-ended front-end for readout of signals with inverted (positive) polarity. Integration of the design in PETA6-Version B;
  - Supervision of master thesis on characterization measurements of PETA6 and dedicated test setup;
  - Supervision of internship work on PETA6 probing with needles on wire-bonding pads at the probe station.

MEASUREMENTS IN COLLABORATION WITH OTHER RESEARCH GROUPS

- Reconstruction of photon clusters center of gravity using a monolithic scintillator coupled with ISiPM prototype, with collimated Na^{22} source - collaboration with Radoslaw Marcinkowski, Stefan Vanderberghe (Ghent University), Ghent, Belgium (01-2013);
- Studies on internal liquid cooling efficiency, performed with a LTCC module with two flip-chip mounted PETA4 chips - collaboration with Rainer Dohle (MSE Engineering GmbH), Berg, Germany and Roberto Blanco (07-2013);
- Readout at high data rate: measurements with coincidence test setup with two detection modules performed with $^{18}F - FDG$

source with original activity of about $500\,MBq$ - collaboration with Jannis Fischer, Werner Lustermann (ETH Zurich), Zurich, Switzerland (06-2015);

- Measurements with a detection module assembled with 8×8 SiPMs by FBK, read out with two PETA5 chips and coupled with $5\,mm$ thick monolithic LYSO block, with collimated source - collaboration with Pieter Mollet (Ghent University), Mannheim, Germany (07- 2015);
- Investigations on limit in timing performance for the single-ended front-end in PETA5. Measurements taken with HD-NUV SiPMs by FBK read out with PETA5; test on LYSO center of gravity reconstruction with rectangular ISiPMs and LG-SiPMs by FBK, read out with PETA5 - collaboration with Alberto Gola, Alessandro Ferri, Claudio Piemonte (FBK), Trento, Italy (10/11 - 2015);

CONFERENCE ATTENDANCE

- Oral presentation at IWORID 2012 - “A low power front-end architecture for SiPM readout with integrated ADC and multiplexed readout”
- Poster presentation at PSMR 2013 - “A new position sensitive Silicon Photo-Multiplier with sub-millimeter spatial resolution for photon clusters identification in PET applications”
- Oral presentation at TWEPP 2013 - “PETA4: a multi-channel TDC/ADC ASIC for SiPM readout”
- Oral presentation at IEEE NSS-MIC 2013 - “Interpolating Silicon Photo-Multiplier: a Novel Position-Sensitive Device for Photon Cluster Identification with sub-millimeter Spatial Resolution”
- Oral presentation at IEEE Sensors 2013 - “A new position sensitive silicon photomultiplier with sub-millimeter spatial resolution for photon-cluster identification”
- Oral presentation at IEEE NSS-MIC 2014 - “A Compact, Water Cooled, 144-Channel Photo Sensor Module for gamma Detection in PET”
- Oral presentation at Elba Advanced Detector Conference 2015 - “A Compact, High-Density Detection Module for Time-of-Flight Measurements in PET Applications”
- Poster presentation at Elba Advanced Detector Conference 2015 - “SPAD Array Chips with Cluster Reconstruction and Fast Full Frame Readout”
- Oral presentation at IWORID 2016 - “A sub-millimeter resolution detector module for small-animal PET applications”

BIBLIOGRAPHY

[1] Victor I.Mikla and Victor V.Mikla. *Medical Imaging Technology.* Elsevier Insights, 2014.

[2] M. Reivich et al. The [18F] fluorodeoxyglucose method for the measurement of local cerebral glucose utilization in man. *Circulation Research*, 44:127–137, 1978.

[3] P. Som et al. A fluorinatedglucose analog, 2-fluoro-2-deoxy-d-glucose(f-18): Nontoxic tracer for rapidtumor detection. *J Nucl Med*, 21:670–675, 1980.

[4] Richard Zimmermann. *Nuclear Medicine , Radioactivity for Diagnosis and Therapy.* EDP Science, 2006.

[5] C. L. Melcher. Scintillation crystals for PET. *J Nucl Med*, 41:1051–1055, 2000.

[6] C. W. E. van Eijk. Inorganic scintillators in medical imaging. *Phys Med Biol*, 40:R85–R106, 2002.

[7] S.R. Cherry et al. Micropet: A high resolution PET scanner for imaging small animals. *IEEE Trans Nucl Sci*, 44(3):1161–1166, 1997.

[8] M. Watanabe et al. A high resolution PET for animal studies. *IEEE Trans Med Imag*, 11(4):577–580, 1992.

[9] W. M. Digby et al. Detector, shielding and geometry design factors for high-resolution PET system. *IEEE Trans Nucl Sci*, 37(2):664–670, 1990.

[10] M. Conti et al. First experimental results of time-of-flight reconstruction on an LSO PET scanner. *Phys Med Biol*, 50(19):4507–4526, 2005.

[11] J.L. Humm et al. From PET detectors to PET scanners. *Eur J Nucl Med Mol Imaging*, 30(11):1574–1597, 2003.

[12] J. Chen et al. Large size LSO and LYSO crystals for future high energy physics experiments. *IEEE Trans Nucl Sci*, 54(3):718–724, 2007.

[13] T. Kimble et al. Scintillation properties of LYSO crystals. *IEEE Nucl Sci Conf*, 3:1434–1437, 2002.

[14] H. T. van Dam et al. Optical absorption length, scattering length, and refractive index of $LaBr_3 : Ce^{3+}$. *IEEE Trans Nucl Sci*, 59(3):656–664, 2012.

[15] H. Iams and B. Salzberg. The secondary emission phototube. *Proceedings of the Institute of Radio Engineers*, 23(1):55–64, 1935.

[16] M. Watanabe et al. A high resolution animal PET scanner using compact PS-PMT detectors. *IEEE Nucl Sci Conf*, 2:1330–1334, 1996.

[17] K. Ziemonsa et al. The clearPET project: development of a 2nd generation high-performance small animal PET scanner. *Nucl Instr Meth Phys Res A*, 537(1-2):307–311, 2005.

[18] 0. Fries et al. A small animal PET prototype based on LSO crystals read out by avalanche photodiodes. *Nucl Instr Meth Phys Res A*, 387(1-2):220–224, 1997.

[19] K. S. Shah et al. Position sensitive APDs for small animal PET imaging. *IEEE Trans Nucl Sci*, 3:1411–1415, 2002.

[20] S. R. Cherry et al. *Physics in Nuclear Medicine*. Elsevier Saunders, 2014.

[21] J. S. Hubert et al. An LSO scintillator array for a PET detector module with depth of interaction measurement. *IEEE Nucl Sci Conf*, 48(3):684–688, 2002.

[22] S. Seifert and D. R. Schaart. Improving the time resolution of TOF-PET detectors by double-sided readout. *IEEE Trans Nucl Sci*, 62(1):3–11, 2015.

[23] Y. Yang et al. Depth of interaction resolution measurements for a high resolution PET detector using position sensitive avalanche photodiodes. *Phys Med Biol*, 51(9):2131–2142, 2006.

[24] R.S. Miyaoka et al. Design of a depth of interaction (DOI) PET detector module. *IEEE Trans Nucl Scie*, 45(3):1069–1073, 1998.

[25] T. Miyaoka et al. A four-layer depth of interaction detector block for small animal PET. *IEEE Trans Nucl Sci*, 51(5):2537–2542, 2004.

[26] N. Inadama et al. A depth of interaction detector for PET with GSO crystals doped with different amounts of Ce. *IEEE Trans Nucl Sci*, 49(3):629–633, 2002.

[27] M. Streun et al. Pulse shape discrimination of LSO and LuYAP scintillators for depth of interaction detection in PET. *IEEE Nucl Sci Conf*, 3:1636–1639, 2002.

[28] F. H. Fahey et al. Data acquisition in PET imaging. *J Nucl Med Technol*, 30(2):39–49, 2002.

[29] T. Beyer et al. A combined PET/CT scanner for clinical oncology. *J Nucl. Med.*, 41(8):1369–1379, 2000.

[30] D. J. Brenner and E. J. Hall. Computed tomography - an increasing source of radiation exposure. *N Engl J Med*, 357:2277–2284, 2007.

[31] P. E. Kinahan et al. Attenuation correction for a combined 3D PET/CT scanner. *Med Phys*, 25(10):2046–2053, 1998.

[32] A. Drzezga et al. First clinical experience with integrated whole-body PET/MR: Comparison to PET/CT in patients with oncologic diagnoses. *J Nucl Med*, 53(6):845–855, 2012.

[33] G. K. von Schulthess et al. A look ahead: PET/MR versus PET/CT. *Eur J Nucl Med Mol Imaging*, 36(1):S3–S9, 2009.

[34] G. Delso et al. PET/MRI system design. *Eur J Nucl Med Mol Imaging*, 36(1):S86–S92, 2009.

[35] S. I. Ziegler et al. Technical and methodological aspects of PET/MR. *Clin Transl Imaging*, 1(1):11–16, 2013.

[36] H. F. Wehrl et al. Assessment of MR compatibility of a PET insert developed for simultaneous multi-parametric PET/MR imaging on an animal system operating at 7 T. *Magn Reson Med*, 65(1):269–279, 2011.

[37] J. Wehner et al. PET/MRI insert using digital SiPMs: Investigation of MR-compatibility. *Nucl Instr Meth Phys Res A*, 734(B):116–121, 2014.

[38] H. W. Schlemmer et al. Simultaneous MR/PET imaging of the human brain: Feasibility study. *RSNA*, 248(3):1028–1035, 2008.

[39] C. Catana et al. Simultaneous in vivo positron emission tomography and magnetic resonance imaging. *PNAS*, 105(10):3705–3710, 2008.

[40] H. F. Wehrl et al. Simultaneous PET-MRI reveals brain function in activated and resting state on metabolic, hemodynamic and multiple temporal scales. *Nat Med*, 19:1184–1189, 2013.

[41] A. J. Lucas et al. Development of a combined microPET-MR system. *Technol Cancer Res Treat*, 5(4):337–341, 2006.

[42] Siemens Healthcare. Biograph-mmr. `http://www.healthcare.siemens.com/magnetic-resonance-imaging/mr-pet-scanner/biograph-mmr`. Accessed: 2016-09-05.

[43] M. Hofmann et al. Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques. *Eur J Nucl Med Mol Imaging*, 36(1):S93–104, 2009.

[44] G. Wagenknecht et al. MRI for attenuation correction in PET: methods and challenges. *MAGMA*, 26(1):99–113, 2013.

[45] E. R. Kops et al. Template-based attenuation correction for PET in MR-PET scanners. *IEEE Nucl Sci Conf*, pages 3786–3789, 2008.

[46] E. Schreibmann et al. MR-based attenuation correction for hybrid PET-MR brain imaging systems using deformable image registration. *Med Phys*, 37(5):2101–2109, 2010.

[47] H. Zaidi et al. Magnetic resonance imaging-guided attenuation and scatter corrections in three-dimensional brain positron emission tomography. *Med Phys*, 30(5):937–948, 2003.

[48] R. R. Kops et al. Attenuation correction in MR-PET scanners with segmented T1-weighted MR images. *IEEE Nucl Sci Conf*, pages 2530–2533, 2009.

[49] J. Ouyang et al. MR-based motion correction for PET imaging. *Semin. Nucl. Med.*, 43(1):60–67, 2013.

[50] S. Fuerst et al. Motion correction strategies for integrated PET/MR. *J Nucl Med*, 56(2):261–269, 2015.

[51] C. Tsoumpas et al. Simultaneous PET/MR acquisition and MR-derived motion fields for correction of non-rigid motion in PET. *Ann Nucl Med*, 24(10):745–750, 2010.

[52] C. Wuerslin et al. Respiratory motion correction in oncologic PET using T1-weighted MR imaging on a simultaneous whole-body PET/MR system. *J Nucl Med*, 54(3):464–471, 2013.

[53] R. G. Wagner et al. The next generation of photo-detectors for particle astrophysics. *Argonne National Laboratory (ANL), Report No. ANL-HEP-TR-09-27*, 2009.

[54] F. Corsi et al. Modelling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design. *Nucl Instr Meth Phys Res A*, 572(1):416–418, 2007.

[55] G. Collazuol et al. Silicon photo-multipliers: status and perspectives. *MEDAMI*, 2014.

[56] N. Serra et al. Characterization of new FBK SiPM technology for visible light detection. *Jinst*, 8:P03019, 2013.

[57] C. Piemonte et al. Characterization of the first FBK high-density cell silicon photomultiplier technology. *IEEE Trans Electr Dev*, 60(8):2567–2573, 2013.

[58] Ketek. Silicon detectors for x-ray and optical spectroscopy. `http://http://www.ketek.net/`. Accessed: 2016-08-06.

[59] J. Bude et al. Hot-carrier luminescence in Si. *Phys. Rev. B*, 45:5848, 1992.

[60] W. J. Kindt et al. Optical cross talk in geiger mode avalanche photodiode arrays: Modeling, prevention and measurement. *Solid-State Device Res Conf*, pages 192–195, 1998.

[61] Sublima. SUB nanosecond Leverage in PET/MR ImAging. `http://www.sublima-pet-mr.eu/`. Accessed: 2016-06-05.

[62] HYPERImage. Hybrid PET/MR. `http://www.hybrid-pet-mr.eu/`. Accessed: 2015-09-05.

[63] T. Solf et al. Solid-state detector stack for ToF-PET/MR. *IEEE Nucl Sci Conf*, pages 2798–2799, 2009.

[64] N. Zorzi et al. Development of large-area silicon photomultiplier detectors for PET applications at FBK. *Nucl Instr Meth Phys Res A*, 636(1):S208–S213, 2011.

[65] V. Schulz et al. A preclinical PET/MR insert for a human 3T MR scanner. *IEEE Nucl Sci Conf*, pages 2577–2579, 2009.

[66] V. Schulz et al. SiPM-based preclinical PET/MR insert for a human 3T MR: first imaging experiments. *IEEE Nucl Sci Conf*, pages 4467–4469, 2011.

[67] C. Lerche et al. Calibration and stability of a SiPM-based simultaneous PET/MR insert. *Nucl Instr Meth Phys Res A*, 702:50–53, 2013.

[68] Philips digital photon counting. `http://www.digitalphotoncounting.com/`. Accessed: 2016-04-01.

[69] B. Weissler et al. Design concept of world's first preclinical PET/MR insert with fully digital silicon photomultiplier technology. *IEEE Nucl Sci Conf*, pages 2113–2116, 2012.

[70] HYPERIMAGE Collaboration. Hybrid pet-mr system for concurrent ultra-sensitive imaging. *Final Report*, 2013. Accessed: 2016-08-08.

[71] H. Thoen et al. Influence of detector pixel size, TOF resolution and DOI on image quality in mr-compatible whole-body PET. *Phys. Med. Biol.*, 58(18):6459–6479, 2013.

[72] S. Surti et al. Design study of a whole-body PET scanner with improved spatial and timing resolution. *IEEE Trans Nucl Sci*, 60(5):3220–3226, 2013.

[73] P. M. Dueppenbecker et al. Development of an MRI compatible digital SiPM based PET detector stack for simultaneous preclinical PET/MRI. *IEEE Nucl Sci Conf*, pages 3481–3483, 2012.

[74] H. T. van Dam et al. Sub-200 ps CRT in monolithic scintillator PET detectors using digital SiPM arrays and maximum likelihood interaction time estimation. *Phys. Med. Biol.*, 58(10):3243– 3257, 2013.

[75] S. Seifert et al. First characterization of a digital SiPM based time-of-flight PET detector with 1 mm spatial resolution. *Phys. Med. Biol.*, 58(9):3061– 3074, 2013.

[76] D. N. ter Weele et al. Scintillation properties of Ca co-doped L(Y)SO:Ce between 193 K and 373 K for TOF-PET/MRI. *EJNMMI Phys*, 1(1):A10, 2014.

[77] V. Keereman et al. Challenges and current methods for attenuation corrections in PET/MR. *MAGMA*, 26(1):81–98, 2013.

[78] P. Mollet et al. Simultaneous MRcompatible emission and transmission imaging for PET using time-of-flight information. *IEEE Trans Med Imaging*, 31(9):1734–1742, 2012.

[79] P. Mollet et al. Attenuation correction for TOF-PET with a limited number of stationary coincidence line-sources. *4th Jülich MR-PET Workshop, Abstracts*, 2013.

[80] C. Tsoumpas et al. Fast generation of 4D PET-MR data from real dynamic MR aquisitions. *Phys Med Biol*, 56(20):6597–6613, 2011.

[81] I. Polycarpou et al. Impact of respiratory motion correction and spatial resolution on lesion detection in PET: a simulation study based on real MR dynamic data. *Phys. Med. Biol.*, 59(3):697–713, 2014.

[82] C. Buerger et al. Investigation of MR-based attenuation correction and motion compensation for hybrid PET/MR. *IEEE Trans Sci*, 59(5):1967–1976, 2012.

[83] T. C. Meyer et al. Endo-TOFPET-US: A multimodal ultrasonic probe featuring time of flight PET in diagnostic and therapeutic endoscopy. *Nucl Instr Meth Phys Res A*, 718:121–125, 2013.

[84] M. D. Rolo et al. TOFPET ASIC for PET applications. *J Instr*, 8(02):C02050, 2013.

[85] A. Di Francesco et al. TOFPET2: a high-performance ASIC for time and amplitude measurements of SiPM signals in time-of-flight applications. *J Instr*, 11:C03042, 2016.

[86] A. Comerma et al. Flextot - current mode ASIC for readout of common cathode SiPM arrays. *IEEE Nucl Sci Conf*, pages 1–2, 2013.

[87] F. Corsi et al. BASIC: an 8-channel front-end ASIC for silicon photomultiplier detectors. *IEEE Nucl Sci Conf*, pages 1082–1087, 2009.

[88] S. Conforti Di Lorenzo et al. SPIROC: design and performances of a dedicated very front-end electronics for an ILC Analog Hadronic CALorimeter (AHCAL) prototype with SiPM read-out. *J Instr*, 8:C01027, 2013.

[89] F. Anghinolfi et al. NINO: an ultra-fast and low-power front-end amplifier/discriminator ASIC designed fro the multigap resistive plate chamber. *Nucl Instr Meth Phys Res A*, 553(1-2):183–187, 2004.

[90] V. Schulz et al. Sensitivity encoded silicon photomultiplierâĂŤa new sensor for high-resolution PET-MRI. *Phys Med Biol*, 58(14):4733–4748, 2013.

[91] A. Gola et al. A novel approach to position-sensitive silicon photomultipliers: First results. *IEEE Nucl Sci Conf*, pages 1–4, 2013.

[92] J. Du et al. Evaluation of linearly-graded SiPMs for high resolution small-animal PET. *Biomed Phys Eng Express*, 1(4):045008, 2015.

[93] P. Moskal et al. TOF-PET detector concept based on organic scintillators. *Nucl Med Rev*, 15(C):C81–C84, 2012.

[94] P. Moskal et al. Time resolution of the plastic scintillator strips with matrix photomultiplier readout for J-PET tomograph. *Phys Med Biol*, 61:2025, 2016.

[95] AX-PET. A demonstrator for an axial PET camera with WLS strips and G-APD readout. `https://twiki.cern.ch/twiki/bin/view/AXIALPET/WebHome`. Accessed: 2016-08-06.

[96] T. Harion et al. STiC - a mixed mode silicon photomultiplier readout ASIC for time-of-flight applications. *J Instr*, 9:CO2003, 2014.

[97] M. Zvolsky et al. EndoTOFPET-US - a miniaturised calorimeter for endoscopic time-of-flight positron emission tomography. *J Phys Conf Ser*, 587:012068, 2015.

[98] J. Y. Yeom et al. Readout electronics and data acquisition of a positron emission tomography time-of-flight detector module with waveform digitizer. *IEEE Trans Nucl Sci*, 60 (5):3735–3741, 2013.

[99] G. B. Ko et al. Development of a front-end analog circuit for multi-channel sipm readout and performance verification for varios PET detector designs. *Nucl Instr Meth Phys Res A*, 703:38–44, 2013.

[100] G. B. Ko et al. Simultaneous multi-paramteric PET/MRI with silicon photomultiplier PET and ultra high-filed MRI for small animal imaging. *J Nucl Med*, 57(8):1309–1315, 2016.

[101] J. Y. Yeom et al. DigiPET: sub-millimeter spatial resolution small-animal PET imaging using thin monolithic scintillators. *Phys Med Biol*, 59(13):3405–3420, 2014.

[102] E. P. Visser et al. Spatial resolution and sensitivity of the inveon small-animal pet scanner. *J Nucl Med*, 50(1):139–147, 2009.

[103] S. Vandenberghe et al. Recent developments in time-of-flight PET. *EJNMMI Phys*, 3(1), 2016.

[104] J. A. Kolthammer et al. Performance evaluation of the Ingenuity TF PET/CT scanner with a focus on high count-rate conditions. *Phis Med Biol*, 59(14):3843–3859, 2014.

[105] V. Bettinardi et al. Physical performance of the new hybrid PET/CT Discovery-690. *Med Phys*, 38(10):5394–5411, 2011.

[106] A. W. Jacobi et al. Physical and clinical performance of the mCT time-of-flight PET/CT scanner. *Phys Med Biol*, 56(8):2375, 2011.

[107] K. C. Burr et al. A new modular and scalable detector for a time-of-flight pet scanner. *IEEE Nucl Sci Conf*, pages 2830–2834, 2012.

[108] M. Miller et al. Characterization of the Vereos Digital Photon Counting PET system. *J Nucl Med*, 56(3):434, 2015.

[109] Philips. Ingenuity tf. `http://www.usa.philips.com/healthcare/product/HC882380/ingenuity-tf-petmr-system`. Accessed: 2016-29-05.

[110] General Electrics Healthcare. Signa pet/mr. `http://www3.gehealthcare.com/en/products/categories/magnetic_resonance_imaging/signa_pet-mr`. Accessed: 2016-29-05.

[111] S. Vandenberghe et al. PET-MRI: a review of challenges and solutions in the development of integrated multimodality imaging. *Phys Med Biol*, 60(4):R115–R154, 2015.

[112] H. Zaidi et al. Design and performance evaluation of a whole-body ingenuity TF PET-MRI system. *J Nucl Med*, 56:3091–3106, 2011.

[113] G. Delso et al. Performance measurements of the siemens mMR integrated whole-body PET/MR scanner. *J Nucl Med*, 52(12):1914–1922, 2011.

[114] C. S. Levin et al. Design features and mutual compatibility studies of the time-of-flight PET capable GE SIGNA PET/MR system. *IEEE Trans Med Imaging*, 35(8):1907–1914, 2016.

[115] I. Sacco et al. A low power front-end architecture for SiPM readout with integrated ADC and multiplexed readout. *Jinst*, 8:CO1023, 2013.

[116] A. Rivetti. *CMOS Front-end Electronics for Radiation Sensors.* CRC Press, Taylor and Francis Group, 2015.

[117] C. Piemonte et al. Performance of FBK SiPMs coupled to PETA3 read-out ASIC for PET application. *Nucl Instr Meth Phys Res A*, 718:345–346, 2013.

[118] M.Ritzert. Development and test of a high performance multi channel readout system on a chip with application in PET/MR. *PhD. Thesis, Heidelberg University*, 2014.

[119] P. Fischer and C. Piemonte. Interpolating silicon photomultipliers. *Nucl Instr Meth Phys Res A*, 718:320–322, 2013.

[120] Fondazione Bruno Kessler. Silicon radiation sensors. `https://srs.fbk.eu/`. Accessed: 2016-08-06.

[121] I. Sacco et al. A new position-sensitive silicon photomultiplier with submillimeter spatial resolution for photon-cluster identification. *IEEE Sens Conf*, pages 1–4, 2013.

[122] I.Sacco et al. Interpolating silicon photo-multiplier: a novel position sensitive device with submillimeter spatial resolution and depth of interaction capability. *IEEE Nucl Sci Conf*, pages 1–3, 2013.

[123] Fraunhofer. Ims. `http://www.ims.fraunhofer.de/en/homepage.html`. Accessed: 2016-06-04.

[124] D. Bronzi et al. Low-noise and large-area CMOS SPADs with timing response free from slow tails. *Proc Eur Sol St Dev Res Conf (ESSDERC)*, pages 230–233, 2012.

[125] P. Fischer et al. SPAD array chips with full frame readout for crystal characterization. *EJNMMI Phys*, 2:A3, 2015.

[126] M. Schork. Characterization of CMOS SiPM IDP2. *Master Thesis, Heidelberg University*, 2016.

[127] Micro System Engineering. an MST company. `http://www.mst.com/msegmbh/`. Accessed: 2016-08-06.

[128] R. Dohle et al. Very compact, water-cooled SiPM module for PET/MRT applications. *EMPC*, pages 1–6, 2015.

[129] 3M Optical Systems. Vikuiti enhanced specular reflector (esr). `http://www.apioptics.com/pdf/ESR.pdf`. Accessed: 2015-10-10.

[130] E. Lorincz et al. Modeling and optimization of scintillator arrays for PET detectors. *IEEE Trans Nucl Sci*, 57(1):48–54, 2010.

[131] I. Sacco et al. A compact, water-cooled, 144-channel photo sensor module for gamma detection in PET. *IEEE Nucl Sci Conf*, pages 1–4, 2014.

[132] C. Piemonte et al. Performance of NUV-HD silicon photomultiplier technology. *IEEE Trans Electr Dev*, 63(3):1111–1116, 2016.

[133] I. Sacco et al. A compact, high-density gamma-detection module for time-of-flight measurements in PET applications. *Nucl Instr Meth Phys Res A*, 824:233–236, 2016.

[134] A. Ferri et al. Performance of a 64-channel, 3.2x3.2cm2 SiPM tile for TOF-PET application. *Nucl Instr Meth Phys Res A*, 824:196–197, 2016.

[135] A. Ferri et al. First results with SiPM tiles for TOF PET based on FBK RGB-HD technology. *EJNMMI Physics*, 2(Suppl 1):A86, 2015.

[136] H. O. Anger. Scintillation camera with multichannel collimators. *Journal of Medicine*, 5:515–531, 1964.

[137] SAFIR. Small Animal Fast Insert for mRi. `https://safir.ethz.ch/home/`. Accessed: 2016-09-05.

[138] C.M. Chang et al. Preliminary PET performance evaluation of an RF field-penetrable brain-sized PET insert for simultaneous PET/MR imaging. *J Nucl Med*, 56(3):99, 2015.