University of Trento

Department of Physics
Ph.D. in Physics
XXXVI cycle

Design and implementation of the Hybrid Detector for Microdosimetry (HDM): Challenges in readout architecture and experimental results

FINAL THESIS

Supervisor:
Prof. Chiara La Tessa

Candidate:
Enrico Pierobon

Contents

"Curiosity is the most important."

– Tasuku Honjo

Part I

Standard microdosimetry

Introduction

Approximately 50% of all patients with localized malignant tumors undergo treatment with ionizing radiation. This is usually combined with tumor resection and/or chemotherapy [Delaney et al., 2005]. High energy photons (up to 25 MeV) are by far the most commonly used type of radiation in this field, but in the last decades the use of accelerated particles, in particular protons and carbon ions, has begun to gain popularity in treating cancer.
The primary rationale for employing charged particles in radiotherapy is their advantageous depth-dose profile, commonly referred to as the "Bragg curve". After many years from Bragg's studies, Robert R. Wilson became the first to propose the use of particle beams for cancer treatment when he published an article in the medical journal Radiology titled "Radiological Use of Fast Protons" [Wilson, 1946]. Since then, ion-based radiotherapy has made remarkable strides, developing into an established cancer treatment strategy. By the end of 2020, over 290,000 patients worldwide had received treatment through particle therapy, with roughly 250,000 undergoing proton therapy, 40,000 receiving carbon ion treatment, and around 3,500 being treated with helium, pions, or other ions [PTCOG, 2020]. Nevertheless, the same attention has not been given to all ions suitable for this purpose. Scientists primarily concentrated on establishing, refining, and perfecting treatment procedures exclusively for protons and carbon beams, overshadowing the use of other particles. After the medical community gained expertise and trust in this approach, they began to embrace new challenges and once more ventured into exploring the potential of using various ions in radiotherapy. Based on physical and biological experimental data, ^{4}He and ^{16}O ions emerged as the most promising candidates. Consequently, several heavy ion centers were created or improved to provide the capability to deliver these particles. [Rovituso and La Tessa, 2017].

While ion therapy continues to advance in precision and effectiveness, numerous uncertainties in treatment persist as a significant barrier to fully harnessing this form of treatment. Indeed, a specialized beam model is required for every ion type, necessitating its development and integration into the treatment planning system. The latter is the software responsible for calculating doses and managing beam delivery for each treatment. To accomplish this objective, it is essential to comprehensively characterize the relevant physical processes occurring between the beam and the patient's body, including the generation of secondary fragments from nuclear reactions, as well as the energy dissipation and lateral dispersion of primary and secondary ions. These interactions lead to alterations in the composition, energy, and direction of the radiation field, thereby influencing both the depth dose and lateral profile. One of the primary uncertainties in ion therapy relates to our understanding of the energy released within the radiation field, which serves as the basis for estimating the dose deposition and the induced radiobiological effects. The pioneering research conducted by Raymond E. Zirkle (1902 - 1988) and subsequent studies have underscored that the biological effectiveness of ionizing radiation, in conjunction with the amount of absorbed energy, hinges on the spatial distribution of energy deposition [Rossi, 1959]. This distribution is a fundamental component of the concept known as "radiation quality".

Radiation field quality

While it is widely acknowledged that characterizing the radiation field is of paramount importance in ion therapy, no universally accepted, quantitative definition for the concept of "radiation quality" exists. The International Commission on Radiological Protection (ICRP) in the *Publication 60 recommendations of ICRP* (1991) state that: "the probability of stochastic effects is found to depend, not only on the absorbed dose, but also on the type and energy of the radiation causing the dose. This is taken into account by weighting the absorbed dose by a factor related to the quality of the radiation". However, even the definition of "quality of the radiation" lacks a precise and universal interpretation.

When attempting to delineate the concept of radiation quality, a valuable initial reference point can be found in Dr. Dudley Goodhead's discussion [Goodhead, 2013], where he addressed the question, "What is radiation quality?" He initiated the discussion by highlighting that the biological consequences of ionizing radiation hinge upon three key factors: i) the amount of radiation (dose, fluence, ...), ii) the temporal distribution (dose rate, fractionation), and iii) the spatial arrangement (track structure). It is worth noting that only the spatial arrangement is directly connected to the radiation quality, and the assortment of radiation tracks at a given location is stochastically determined by the types and energies of charged particles present at that specific site. Radiation quality is subsequently characterized by the fluence spectrum of radiation particles at the specified locations within the target material. This spectrum provides information on the relative number of particles categorized by type and energy. It encompasses both charged and neutral components and is influenced by the source's characteristics and the material it traverses.
Finally, variations in radiation quality can lead to variations in biological effectiveness. This is true even if the energy deposited by the radiation is the same (same dose). The investigation of radiation quality also serves as a valuable instrument for exploring the fundamental mechanisms of radiobiological effects. Nonetheless, comprehending the intricacy of this subject requires a concise overview of the diverse electromagnetic and nuclear interactions governing the spatial distribution of energy deposition by radiation in matter. Furthermore, to appreciate the connection between radiation quality and biological harm, a fundamental insights into the biological effects of radiation will be provided.

Physical and biological background

When charged particles traverse a medium, two primary categories of physical interactions must be taken into account [Schardt, 2016]:

- *Electromagnetic interactions* which rule the energy deposition in matter. Coulomb forces between the positively charged ion and the negatively charged orbital electrons of the medium atoms are established. These interactions also concur in the lateral spread of particles.
- *Nuclear interactions*, which cause a loss of primary ions and the production of secondary radiation. In contrast to electromagnetic interactions, nuclear interactions are responsible for a more widely lateral spread of particles.

Energy deposition in matter

Ions interact inelastically with the shell electrons of the medium nuclei, resulting in an energy loss dE/dx well described by the Bethe-Bloch equation [Bethe, 1953]

$$S := -\frac{dE}{dx} = \left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \rho 2\pi N_a r_e^2 m_e c^2 \frac{z^2}{\beta^2}\frac{Z}{A}\left(\ln\left(\frac{2m_e c^2\beta^2}{\bar{I}^2\left(1-\beta^2\right)} - 2\beta^2 - 2\frac{C}{Z} - \delta\right)\right), \quad (1)$$

where:

- z and Z are the nuclear charge of the projectile and the target, respectively;
- m_e and e are the mass and charge of the electron, respectively;
- $\bar{I}$ is the mean ionization energy of the target atom or molecule. Mean ionization energy cannot be calculated to sufficient accuracy from first principles so it is effectively an adjustable parameter of the theory. For liquid water, accurate Bragg curve measurements for protons and different heavier ions show a values of 78 eV [Steidl et al., 2008];

- β is the speed the projectile relative to the speed of light;
- ρ is target density;
- ϵ_0 is the vacuum permittivity;
- N_a is the Avogadro's Number;
- A is the target mass number.

Another adjustment to the Bethe-Bloch equation considers how the effective charge of the projectile varies with its velocity. At high velocities, the atomic electrons are fully stripped away, rendering the ion's effective charge identical to its atomic charge number, denoted as z. As the ion slow down, the mean charge state decreases due to the interplay of ionization and recombination processes and z has to be replaced by the effective charge z_{eff} described by the Barka's empirical formula [Barkas and Evans, 1963]:

$$z_{\text{eff}} = z\left(1 - \exp\left(125\beta z^{\frac{2}{3}}\right)\right).$$

The maximum energy transfer is reached when the projectile velocity is:

$$v_p \simeq z^{-\frac{2}{3}} v_0\,, \tag{2}$$

where $v_0 = e^2\hbar = 1/137$ is the Bohr velocity.

An illustration of the energy loss curve, referred to as the "Bragg curve" for ions, in relation to penetration depth is depicted in Figure 1. This graph provides a comparative view of the distinct behaviors exhibited by carbon ions in contrast to γ-rays (from a ^{60}Co radioactive source) and high energy photons. In radiotherapy with high energy photons, the greatest doses are delivered at shallow depths, typically just beneath the skin. To safely treat deeply seated tumors, multiple angles are employed by rotating a LINear ACcelerator (LINAC) to focus the beams on the tumor, as outlined in [Fiorino et al., 2020]. Unlike photons, ions lose only a minimal fraction of their kinetic energy through ionization as they traverse the shallow depths often referred to as the entrance channel or plateau region. However, as they approach the end of their trajectory, the energy transfer undergoes a significant increase, culminating in a maximum value called the Bragg peak, followed by a rapid decline. The location of the Bragg peak can be precisely tailored to the desired tissue depth by modifying the kinetic energy of the incident ion.

Mean range and energy straggling

Charged particles gradually dissipate their energy along their paths through matter, ultimately coming to a halt. Consequently, charged particles possess a finite and relatively well-defined range. To facilitate discussions, the concept of the *Continuous-Slowing-Down Approximation* (CSDA) range, denoted as r_0, is defined as follows:

$$r_0 := \int_0^{E_0} \frac{1}{S(E)} dE.$$

This value represents the average distance covered by a charged particle with initial kinetic energy E_0 as it comes to a stop.

Equation 1 suggests that plotting the energy loss of a single ion as a function of penetration depth would yield an extremely sharp peak near its stopping point. However, energy loss

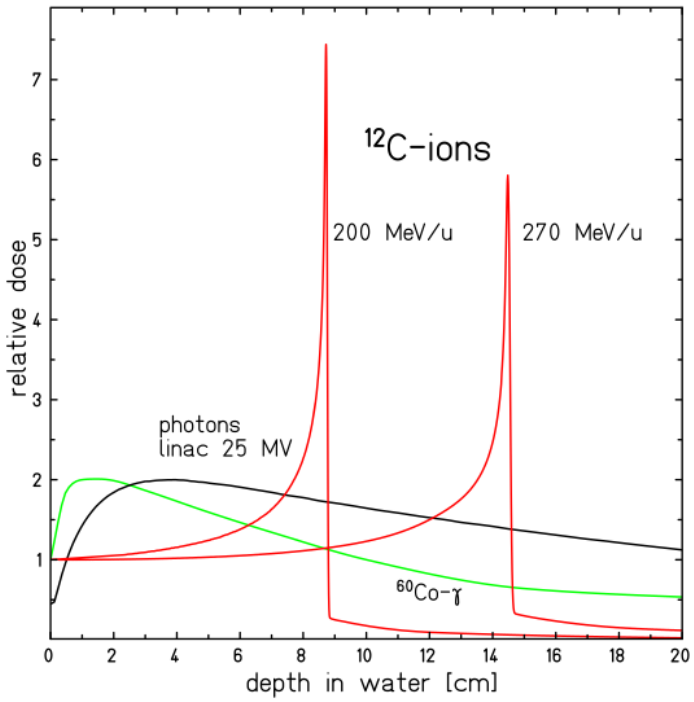


Figure 1: Depth-dose profiles of X-rays, megavolt photon beams, and ^{12}C ions in water. Figure taken from [Schardt et al., 2010]

follows a stochastic process, and thus, statistical fluctuations introduce broadening of the Bragg peak for a monoenergetic beam [Comfort et al., 1966]. These variations lead to the phenomenon known as *energy straggling*, which is effectively characterized by the Landau distribution [Wilkinson, 1996] for extremely thin absorbers and the Vavilov distribution for thinner absorbers [Bichsel, 1968]. In situations involving a substantial number of collisions or within thick absorbers, the Vavilov distribution approaches a Gaussian form [Ahlen, 1980], with a standard deviation (σ) represented as:

$$\sigma_E = 4\pi z_{\text{eff}} Z e^4 N \Delta x \left[\frac{1-\beta^2/2}{1-\beta^2}\right] .$$

The range straggling variance σ_R^2 is related to the energy-loss straggling variance σ_E^2, and the ratio between σ_R^2 and the mean range R is proportional to $\frac{1}{\sqrt{M}}$, with M being the ion mass. Thus, the range straggling gets smaller for heavier ions, being maximum for protons and, for example, a factor 3.5 larger compared to the ^{12}C ions value [Schardt et al., 2010].

Lateral beam spread

The lateral spreading of charged particles as they pass through matter is primarily attributed to their elastic Coulomb interactions with the target nuclei, a phenomenon well-described by the Molière Theory [Moliere, 1948]. For small angles, where higher-order terms in Molière's solution can be neglected, the angular distribution can be approximated by a Gaussian function, with the standard deviation determined by the Highland approximation:

$$\sigma_\theta[rad] = \frac{14.1\,MeV}{\beta\,p\,c} Z_p \sqrt{\frac{d}{L_{rad}}} \left[1 + \frac{1}{9}\log\left(\frac{d}{L_{rad}}\right)\right] . \quad (3)$$

Here, Z_p and p represent the charge and momentum of the projectile, respectively. The absorber material is characterized by its thickness, denoted as d, and radiation length, indicated as L_{rad}. As a consequence, for a given thickness, materials containing heavier elements cause a greater angular spread. When calculated with Equation 3, the lateral spread results in a smaller value for heavy ions compared to protons at a specific depth, due to the influence of the factor βpc.

Nuclear interactions

In comparison to electromagnetic interactions with atomic electrons within the medium, the likelihood of ions undergoing nuclear reactions is significantly lower (for example, a 209 MeV proton has a 24% probability of nuclear interaction in water before stopping [Paganetti, 2018]). However, as penetration depths increase, these interactions yield critical effects. At energies in the several hundred MeV/u range, violent nuclear spallation reactions can lead to the partial or complete disintegration of both the projectile and target nuclei (e.g., in head-on collision). Peripheral collisions, on the other hand, are, for purely geometrical reasons, the most common nuclear reactions that occur along the trajectory of the ion. These peripheral collisions are less severe than head-on collisions because the beam particle loses one or a few nucleons without completely disintegrating. These reactions can be effectively described as a two-step process (see Figure 2) through the abrasion-ablation model [Serber, 1947].

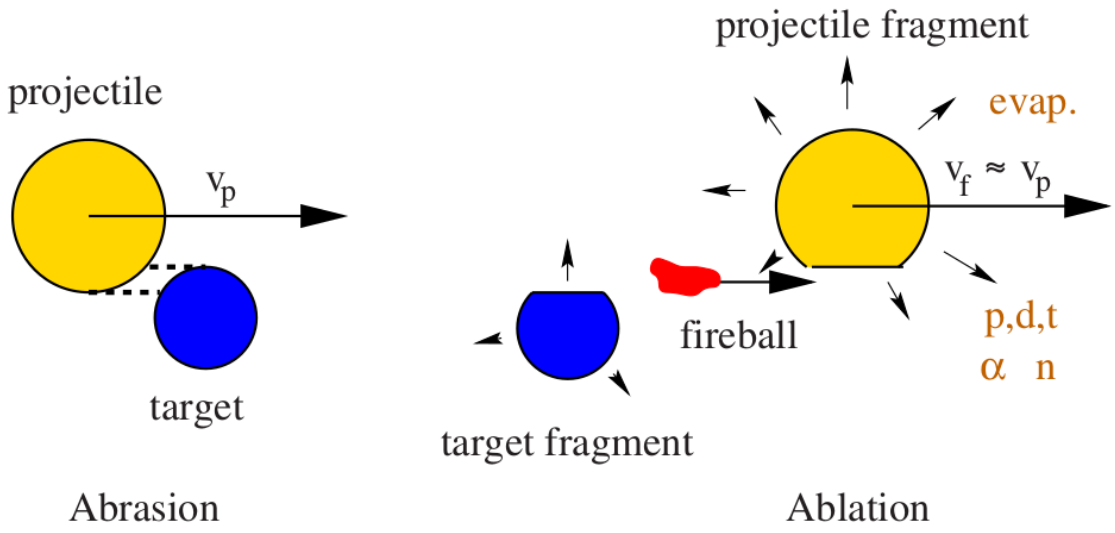


Figure 2: Illustration of the abrasion-ablation model of peripheral collisions at high energies, according to [Serber, 1947]. Figure from [Gunzert-Marx et al., 2008]

In the initial stage, called *abrasion*, nucleons within the overlapping reaction zone, often referred to as *fireball*, undergo abrasion, while the outer nucleons, called *spectators*, experience minimal effects. Subsequently, the remaining fragments from both the projectile and target undergo de-excitation in the second step, known as *ablation*, through the evaporation of nucleons or clusters.
The study of fragmentation reactions has been a focal point in nuclear physics [Goldhaber and Heckman, 1978, Hüfner et al., 1975], and extensive experimental data are accessible for numerous combinations of projectiles and targets, spanning a wide range of beam energies [Otten, 1989, Norbury et al., 2012, Horst, 2020, Luoni et al., 2021].

The phenomenon of nuclear fragmentation yields several significant effects pertinent to

radiotherapy using ion beams:

- nuclear reactions result in the depletion of primary beam particles and an accumulation of lower-Z fragments. These effects gain increasing significance as penetration depth increases.

- The secondary (or higher order) projectile-like fragments move at about the same velocity as the primary ions. As a result, they generally have a greater range than the primary beam, reaching depths beyond the Bragg peak of primaries particles.

The impact of nuclear fragmentation on the Bragg curve is shown in Figure 3. With increasing penetration depth, the peak-to-entrance dose ratio becomes gradually smaller, mainly caused by the exponentially diminishing flux of primary ions. The buildup of lower-Z fragments is clearly visible in the dose tail behind the Bragg peak at larger depths.

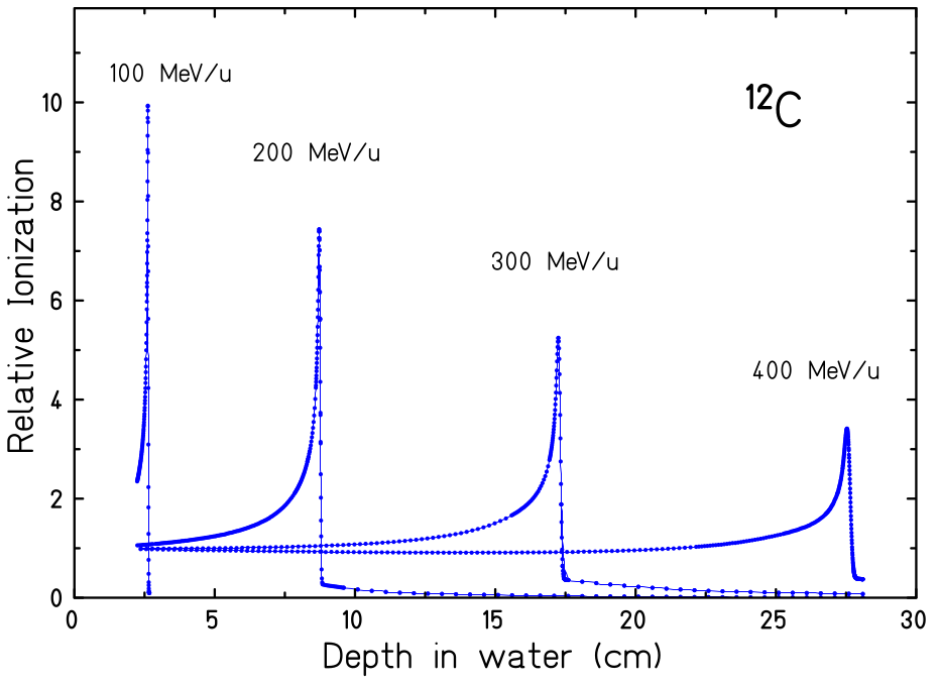


Figure 3: Measured Bragg curves of ^{12}C ions stopping in water. The dose tail behind the Bragg peak increases at larger depths due to nuclear fragmentation. Figure taken from [Schardt et al., 2008].

Fragmentation processes exert influence not only on the physical dose distribution but also on the biological effects of ions. Various ions generate distinct fragment species, each marked by its own unique biological impact. When working with protons, it is possible, as a reasonable approximation, to rely entirely on the physical dose released within the Bragg peak. This dose is easy to measure as it is primarily caused by the same type of radiation. However, for heavier ions, this approximation becomes invalid, because the physical dose is insufficient to differentiate the contributions of various ion species.

Linear Energy Transfer - (LET)

The quality of the radiation field is characterized through either a macroscopic or microscopic approach. The widely adopted macroscopic description relies on *linear energy transfer* (LET),

defined as the energy released by radiation per unit length:

$$LET = \frac{dE}{dl} \,. \tag{4}$$

It should be noted that LET is defined only for charged particle. LET helps in the description of the radiation quality differentiating in, low- and high-LET charged particles. This difference is due to the spatial structures of the ionization.

In the case of low LET, typical for the electron and high energy photons, the ionization density and consequently the energy released along the particle track is relatively small compared to high LET particles. Typically, ions are classified as high LET particles. To give a quantitative comparison, the absorbed dose in a sphere with radius $\simeq 7\,\mu\mathrm{m}$ of $1\,\mathrm{Gy}$ can be achieved with about 1000 electron tracks. The same dose can be achieved with only 4 α particles in the same volume [Mayles et al., 2007, Goodhead, 1988]. This shows how much more energy is transferred per unit distance along the high LET α particle tracks than that transferred by the low LET electrons. The difference in how the energy is released, and consequently the LET, has consequences for the biological effect.

Biological effect of radiation

The primary goal of curative cancer radiotherapy is to render all malignant cells inactive. At the same time, the incidence of radiation-induced normal tissue complications must be kept at an acceptable level. To achieve this goal, tumor cells must be incapacitated to the point where they are unable to pass on their genetic abnormalities to their progeny. One effective way to achieve this result is to induce damage to the deoxyribonucleic acid (DNA) that contains all genetic information. All cells have multiple mechanisms for repairing their DNA. Assuming that only unrepaired or improperly repaired DNA damage leads to cell death, the repair capacity of the cell can be quantified by the survival probability, denoted S, after exposure to a radiation dose D. This probability S is typically described by a linear-quadratic (LQ) model [McMahon, 2018]:

$$S(D) = \exp\left[-\alpha D - \beta D^2\right] \,. \tag{5}$$

The parameters α and β are factors specific to the type of cells. The repair capacity, or radiosensitivity, of different tissues is manifested in the quadratic term β and is often expressed through the ratio $\frac{\alpha}{\beta}$. Consequently, cells with a high α/β ratio exhibit limited repair capabilities, while the reverse holds true [Fowler, 1989].

Depending on the classification based on the LET, high- or low- LET, the biological effect is different. High LET, typical of heavier particles, produces a higher ionization density along the particle track. As depicted in Figure 4, the dissimilar spatial patterns of energy deposition between high energy photons and charged particles (carbon ions) are evident, along with the impact of these depositions on DNA. Low LET radiation tracks produce a random spatial distribution of energy deposition events within the cell volume. In contrast, ions (high LET) deposit their energy through the emission of secondary electrons, leading to highly localized energy deposition along the path of the primary particles [Scholz, 2006].

In terms of quantifying the efficacy of ions in inducing biological damage, the most widely used metric is *Relative Biological Effectiveness* (RBE). RBE is defined as the ratio of the physical absorbed dose delivered by a reference radiation (250 keV photons) to the dose administered by ions that produce an equivalent biological effect:

$$RBE \equiv \left.\frac{D_{ref}}{D_{ion}}\right|_{\text{isoeffect}}$$

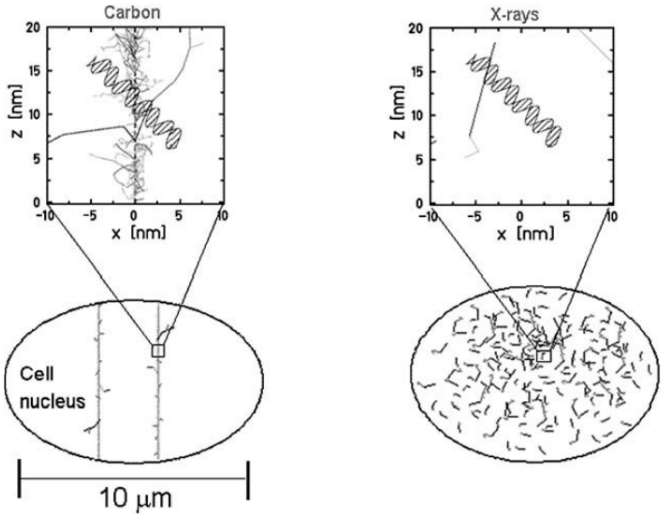


Figure 4: (Left) Carbon track structure in cell nucleus: the electrons density is highly localized around the ion path producing a large number of ionization to the DNA.
(Right) Uniform energy depositions of X-rays in the cell nucleus [Scholz, 2006].

RBE is of great importance in heavy ion therapy, serving as a key parameter to characterize both cell eradication and normal tissue complications [Dieter Schardt, 2010].

To comprehend how radiation quality influences biological damage, it is valuable to examine the behavior of the RBE in relation to the LET of the radiation. The connection between LET and RBE serves as a crucial link between the physical properties of radiation and their biological effects. In this context, variations in LET constitute the primary source of variability in RBE. Consequently, an accurate assessment of RBE necessitates a comprehensive understanding of the LET spectrum. During treatment, the LET value of the radiation will change as the radiation travels from the entry channel to the target. This is especially true if the tumor is deep-seated. This phenomenon arises from the nuclear and electromagnetic interactions of primary ions with the patient's body, resulting in changes in the energy of the primaries and the generation of secondary fragments emitted at various angles and with a wide range of energies. The resultant mixed radiation field leads to spatial variations in RBE, both within and outside the tumor. Figure 5 illustrates the LET dependence on the RBE for different ion types. The RBE increases with increasing LET up to a maximum value that is ion-specific. Particles with high LET exhibit a greater ionization density, leading to more pronounced damage and, consequently, an elevated RBE. However, empirical evidence suggests that the RBE decreases beyond a certain LET threshold. This phenomenon is often attributed to the "overkill" or "saturation" effects. At high LET, the induced damage in the cell is exceptionally high, exceeding the value required to deactivate the cell. Therefore, any excess damage leads to the same biological result, culminating in cell death. The net effect is a reduction in the *effectiveness* of deactivating the cell, and therefore a reduction in the relative biological *effectiveness*. While the LET is the standard and more common method for correlating radiation quality with its biological impact, it is measured at a macroscopic scale, significantly larger than the cellular level where the actual damage occurs. This macroscopic approach overlooks the stochastic nature of energy deposition, which leads to variations from one cell to another, and it assumes LET to remain constant throughout the material. Additionally, this approach involves several simplifications, such as the omission of track structure considerations and the neglect of the distribution of delta ray energy and its connection to

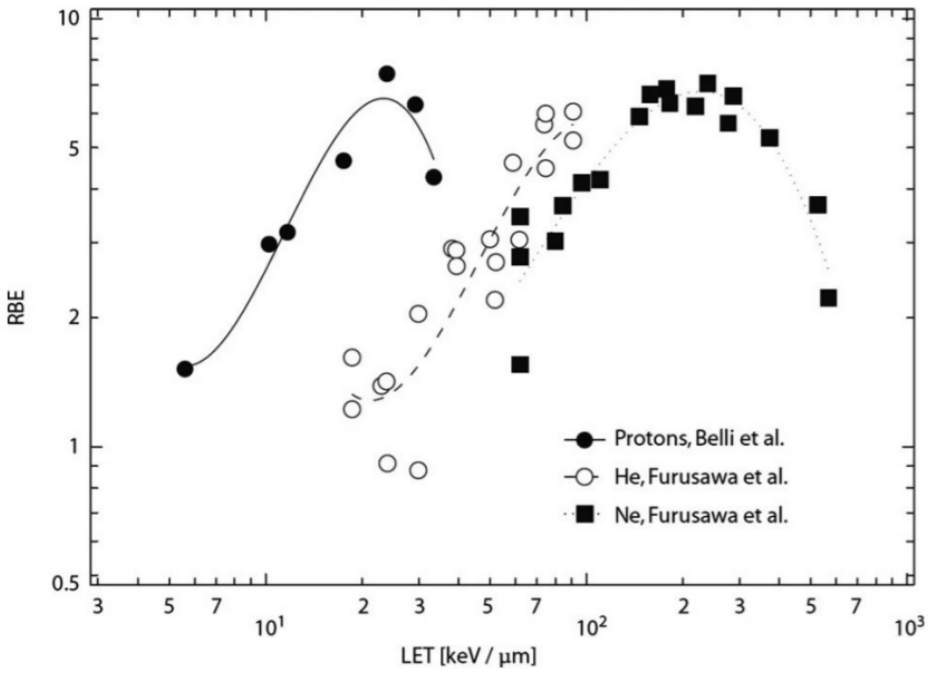


Figure 5: Dependence of the RBE on LET for protons, Carbon and Neon ions for experiments with V79 hamster cells. Figure taken from [Scholz, 2003].

spatial dose.
Microdosimetry, contrary to LET where a single value is measured, provides an entire distributions of variables related to energy depositions in tissue regions of micrometer-scale dimensions. It delves into the effects of radiation on cellular structures, acknowledging the stochastic nature of energy deposition within these regions. This microscopic approach considers entire distributions of variables related to energy depositions in tissue regions of micrometer-scale dimensions.
Microdosimetry measured quantity is the so called *lineal energy* y defined as:

$$y = \frac{\epsilon}{\bar{l}}, \tag{6}$$

the energy released ϵ, divided by the *mean chord length* $\bar{l}$. Despite the tempting similarity to Equation 4, the y describe a completely different domain of interaction. The y is defined in the typical size of the cell and consequently inherits a stochastic behavior from the stochastic nature of the energy deposition ϵ. While ϵ can be measured directly, the value of $\bar{l}$ is calculated as the average distance traveled by the particle inside the detector using the mean chord length approximation. The latter depends on both the geometry of the detector and certain assumptions about the radiation field, which is typically assumed isotropic and uniform. As the main topic of this thesis, a new Hybrid Detector for Microdosimetry (HDM), capable of estimating the track length (l) inside the microdosimeter, is proposed. In this way, it is possible to drop the mean chord length approximation in favor of a new quantity:

$$y_T = \frac{\epsilon}{l}. \tag{7}$$

This new quantity can improve the description of the radiation quality, consolidating the link with the biological effect. Two extreme scenarios for y_T can be considered, where the same energy $\epsilon = 2\,\mathrm{keV}$ is deposited over a small track $l_0 = 0.1\,\mu\mathrm{m}$ or a longer track $l_1 = 2\,\mu\mathrm{m}$. In

the case of l_0, y_T would be 20 keV/µm, while in the case of l_1, y_T would be 1 keV/µm. A different biological effect is likely to be observed from the different way the energy is released. However, when using the mean chord length approximation and the classical definition y, these distinctions are not captured as the ys are assumed to be identical.
In addition, using the estimate track l, the geometric dependence of the detector is inherently considered. This enhancement ensures better comparability of microdosimetric measurements obtained with detectors of different geometries.

The contribution of the present work to the field

This thesis is divided in two main parts. The first part falls within the scope of conventional microdosimetry. The fundamental characteristics and key parameters of microdosimetry are provided in chapter 1.
The next section, section 2.2, is dedicated to the data analysis process used to build microdosimetric spectra from acquired experimental data.
Next, a section, section 2.3, is dedicated to the development of a graphical user interface (GUI) to assist users in the analysis of microdosimetric data. This software, written in C#, enables users to make real-time adjustments of the main parameters required for data analysis and visualize the changes. The implementation and functionality of such a program are discussed.
After that, section 2.4 addresses the issue of pileup in microdosimetric spectra. Pileup occurs when multiple events are collected and processed simultaneously due to their close timing, distorting the spectral data. To address this problem, a novel method for estimating pileup probability and quantifying associated errors is proposed. This method provides a quantitative understanding of the impact of pileup on microdosimetric spectra. The study involves microdosimetric measurements conducted at various particle rate intensities, and Monte Carlo simulations using Geant4 [Agostinelli et al., 2003, Allison et al., 2006, Allison et al., 2016] to estimate pileup contributions to the experimental data.

This part is based on the following original work:

- Pierobon E., et al. "A novel methodology to estimate pileup effects and induced error in microdosimetric spectra" *Medical Physics* (2023 - submitted)

The second part of this thesis focuses on the development and characterization of the novel Hybrid Detector for Microdosimetry (HDM). To evaluate the accuracy of the mean chord length approximation, Monte Carlo simulations using the Geant4 toolkit were conducted. Analysis of track length distributions for all particles passing through the Tissue Equivalent Proportional Counter (TEPC) revealed that the mean chord length was not always a reliable representation for the entire population.

In order to obtain accurate track length information experimentally, a novel hybrid 2-stage microdosimeter (HDM) is designed. It consists of a spherical TEPC followed by four layers of low gain avalanche detectors (LGADs).

HDM delivers lineal energy spectra in a tissue-equivalent medium, providing an event-by-event measurement of the path length and submillimetric spatial resolution. To evaluate the detector's performance, a feasibility study was conducted, as reported in chapter 3. Various configurations, including the distance between detectors and the number of strips in a single LGAD, were tested, and an assessment of the HDM response to protons and carbon ions at different water depths was performed. Naturally, the realization of the detector requires a dedicated, custom, readout system.
The primary objective of my PhD is to design a readout system for HDM. This work is presented in two distinct chapters: the first chapter focuses on the development of the readout

system for LGADs (chapter 4), while the second chapter concentrates on the readout system for TEPC (chapter 5). In both chapters, bottom-up approach is adopted, starting with the raw signals from the detectors and ending with how the data is presented to the end-user. Throughout these chapters, I share the challenges encountered during signal processing. In addition, both chapters provide information on the results achieved with the implemented acquisition system and present a comprehensive analysis of performance. Regarding the LGADs, an experimental campaign was conducted to validate the sensors' detection capabilities and to assess the functionality of the complete readout system. For the TEPC, a mew data analysis was necessary due to the different information provided by the new acquisition. This new data analysis process is detailed in section 5.3. Additionally, an experimental campaign specific to the TEPC readout was carried out to assess its capability in producing microdosimetric spectra.

This part is based on the following original works:

- Missiaggia M., Pierobon E., et al. "A novel hybrid microdosimeter for radiation field characterization based on TEPC detector and LGADs tracker: a feasibility study." *Frontiers in Physics* (2021) [Missiaggia et al., 2021]

Chapter 1

Microdosimetry

1.1 Quantities and distributions

The formalism of the measured and derived quantities in microdosimetry can be found in detail in [Booz et al., 1983] and in [Zaider et al., 1996]. In this section, I present an overview of the key microdosimetry variables of interest for this work.

1.1.1 Energy deposited and imparted

Energy deposited

The quantity ϵ_i is the *energy deposits* in a single interaction i:

$$\epsilon_i = T_{in} + T_{out} + Q_{\Delta m} \,,$$

where:

- T_{in} represents the kinetic energy of the incident ionizing particle (exclusive of the rest mass);
- T_{out} represents the sum of the kinetic energies of all ionizing particles leaving the interaction (exclusive of the rest mass);
- $Q_{\Delta m}$ represents the changes of the rest mass energy of the atoms and all particles involved in the interaction ($Q_{\Delta m} > 0$ implies a rest mass decrease , $Q_{\Delta m} < 0$ implies a rest mass increase).

It is important to note that:

- ϵ_i is a stochastic quantity;
- ϵ_i is energy deposited at the point of interaction (which is also called transfer point);

Energy imparted

The energy imparted by radiation to matter in a given volume is ϵ:

$$\epsilon = \sum_i \epsilon_i \,,$$

where the summation is performed over all energy deposits ϵ_i in that volume. It can be noticed that:

- ϵ is a stochastic quantity;
- the energy deposits over which the summation is performed may be related to one or more events, i.e. due to one or more statistically independent particle tracks.

1.1.2 Specific energy

The specific energy, z, is the quotient of ϵ by m, where ϵ is the energy imparted by ionizing radiation to matter of mass m:

$$z = \frac{\epsilon}{m}\,,$$

Further, it is worth noticing that:

- z is a stochastic quantity. The value of the distribution function $F(z)$ is the probability that the specific energy is equal to or less than z. The probability density $f(z)$ is the derivative of $F(z)$ with respect to z:

  $$f(z) = \frac{dF(z)}{dz}\,,$$

  The probability density $f(z)$ includes a discrete component (a Dirac delta function) at $z = 0$ for the probability of no energy deposits.
  The expectation value:

  $$\bar{z} = \int_0^\infty z f(z) dz\,,$$

  is called mean specific energy and is a non-stochastic quantity.

- The specific energy may be caused by one or more energy deposits events. The distribution function of the specific energy deposits in a single event, $F_1(z)$, is the conditional probability that a specific energy less than or equal to z is deposited if one event has occurred. The probability density, $f_1(z)$, is the derivative of $F_1(z)$ with respect to z:

  $$f_1(z) = \frac{dF_1(z)}{dz}\,,$$

  and it is also called the single event distribution of z. The expectation value:

  $$\bar{z_F} = \int_0^\infty z f_1(z) dz\,,$$

  is called *frequency-mean specific energy per event*, and is a non-stochastic quantity.

- It is also useful to consider the dose distribution of z per energy deposit event. Let $D_1(z)$ be the fraction of absorbed dose per event delivered by energy deposits events of specific energy less than or equal to z. The dose probability density, $d_1(z)$, is the derivative of $D_1(z)$ with respect to z:

  $$d_1(z) = \frac{dD_1(z)}{dz}\,.$$

  The expectation value is

$$\bar{z}_D = \int_0^\infty z d_1(z) dz \,,$$

it is called *dose-mean specific energy per event*, and is a non-stochastic quantity. The relations between $f_1(z)$ and $d_1(z)$, and between $\bar{z}_F$ and $\bar{z}_D$ are

$$d_1(z) = \frac{z}{\bar{z}_F} f_1(z) \,,$$
$$\bar{z}_D = \frac{1}{\bar{z}_F} \int_0^\infty z^2 f_1(z) dz \,.$$

1.1.3 Lineal energy and LET limitations

In the past few years, LET has emerged as a key parameter in characterizing the quality of radiation. However, the definition falls short when applied to electrons: there is no volume of interest small enough to ignore the finite range of electrons, yet large enough to overlook the energy loss and lateral escape of δ rays. In the case of heavy ions, there are certain combinations of site sizes and particle energies for which the LET concept can predict energy deposition reasonably well. Nonetheless, even in such cases, the LET has limited utility as it does not offer information about the energy distribution within these sites. Rossi proposed that the use of linear energy transfer as a measure of radiation quality be replaced by the energy deposited by charged particles and their secondaries in volumes of specified size.

The *lineal energy*, y, is defined as the ratio between the energy imparted to the matter in a volume by a single energy-deposit event (ϵ) and the mean chord length in that volume ($\bar{l}$)

$$y = \frac{\epsilon}{\bar{l}} \,.$$

With this definition, y has no ambiguity. unlike LET.
It is important to notice that:

- the mean chord length in a volume is the mean length of randomly oriented chords. Under the assumptions of an hisotropic and uniform field, for a convex body, the mean chord length can be written as:
$$\bar{l} = \frac{4V}{S} \,,$$
where V is the volume and S is the surface area of this body. It is worth stressing produced clinical beams do not satisfly the mean chord lenght hypotesis as they fail to satisfly the isotropicity hypotesis. The mean chord length of different convex volume is given in Table 1.1 using the following parameters:
  - r is the radius of sphere or cylinder;
  - h is the cylinder length;

  The lineal energy is a stochastic quantity and is useful to consider its distribution function, $F(y)$, which represents the probability that the lineal energy is equal to or less than y.
  The probability density, $f(y)$, is the derivative of $F(y)$ with respect to y:

	Sphere	**Cylinder**
$\bar{l}$	$\frac{4}{3}\,r$	$\frac{2rh}{r+h}$

Table 1.1: Mean chord lengths of the most used active volumes of gas based microdosimeters.

$$f(y) = \frac{dF(y)}{dy}\,, \tag{1.1}$$

and it is also called lineal energy distribution. It is important to keep in mind that y is defined for single energy-deposit events only.

- The lineal energy distribution, $f(y)$, is independent of the absorbed dose or dose rate.
- The expectation value:
$$\bar{y}_F = \int_0^\infty y f(y) dy\,, \tag{1.2}$$
is called *frequency-mean lineal energy*, and it is a non stochastic quantity.
- It is also useful to consider the *dose distribution* of y. Let $D(y)$ be the fraction of absorbed dose delivered with lineal energy less than or equal to y. The *dose probability density*, $d(y)$, of y is the derivative of $D(y)$ with respect to y:
$$d(y) = \frac{dD(y)}{dy}\,. \tag{1.3}$$
- The distribution $d(y)$ is independent of the absorbed dose or dose rate.
- The expectation value:
$$\bar{y}_D = \int_0^\infty y d(y) dy\,,$$
is called *dose-mean lineal energy*, and it is a non stochastic quantity.
- The relations between $d(y)$ and $f(y)$ and between $\bar{y}_D$ and $\bar{y}_F$ are:
$$d(y) = \frac{y}{\bar{y}_F} f(y)\,,$$
and
$$\bar{y}_D = \frac{1}{\bar{y}_F} \int_0^\infty y^2 f(y) dy\,. \tag{1.4}$$

Since microdosimetric spectra are usually plotted on an x-log scale, it is preferable to plot $yf(f)$ or $yd(y)$ instead of $f(y)$ or $d(y)$ to estimate the fraction of events with lineal energy values in a given range of interest. The basis of this is the fact that:

$$\int_{y_1}^{y_2} f(y) d(y) = \int_{y_1}^{y_2} (y f(y))\, d\log y\,.$$

An additional quantity used in biological applications that accounts for the *saturation* effect and is parameterized by y_0 based on the method introduced by [Powers et al., 1968] is defined as follows:

$$y^* = \frac{y_0^2 \int_0^\infty \left[1 - \exp\left(-y^2/y_0^2\right)\right] f(y)\ \mathrm{d}y}{\int_0^\infty y f(y)\ \mathrm{d}y} \tag{1.5}$$

Quantity	Unit of measurement
ϵ_i	keV
ϵ	keV
y	keV/μm
z	Gy
V	μm^3
S	μm^2
$\bar{l}$	μm
d	g/cm^3

Table 1.2: Units of measurement of the principal microdosimetric parameters and of the parameters used in calculations.

1.1.4 Relationships between microdosimetric quantities

There are several relationships between microdosimetric quantities and their macroscopic analogues. Some of these are of a general nature and follow directly from the definition of the quantity, or can be derived mathematically. Others are restricted to a particular geometry or make use of specific approximations.

Relationships between ϵ, y and z

The relation between the quantities ϵ, y and z is:

$$\epsilon = y\bar{l} = zm\,.$$

The same relationships hold for the frequency and dose means. Numerical constants in the following formulas are chosen for the set of units reported in Table 1.2.

Then

- $z = ky$;
- $z = \frac{k}{\bar{l}}\epsilon$;
- $z = \frac{\epsilon}{m}$.

With some calculations:

$$\begin{aligned} z\,[Gy] = z\left[\frac{J}{kg}\right] &= k\frac{\epsilon}{\bar{l}}\left[\frac{keV}{\mu\,m}\right] = \\ &= k\frac{\epsilon}{\bar{l}}\left[\frac{J}{\mu\,m}\right]\frac{1}{\rho}\left[\frac{cm^3}{g}\right] = \\ &= k\frac{1.6\,10^{-16}}{\bar{l}}\left[\frac{J}{\mu\,m}\right]10^{15}\left[\frac{\mu\,m^3}{kg}\right]\,, \end{aligned}$$

from which we obtain that $k = 0.16\,\frac{\bar{l}}{V}$.
For the important special case of a spherical volume:

$$k = 0.16\frac{\bar{l}}{V} = 0.16\frac{\frac{4}{3}\frac{d}{2}}{\frac{4}{2}\pi\left(\frac{d}{2}\right)^3} = \frac{0.204}{d^2}\,.$$

Then, since in this work it has been always used a spherical microdosimeter, the absorbed dose in Gy is simply given by the following relationship

$$z[Gy] = \frac{0.204}{d^2\,[\mu\,\mathrm{m}^2]}y\,[\mathrm{keV}/\mu\,\mathrm{m}]\,. \tag{1.6}$$

Chapter 2

Experimental microdosimetry

Microdosimeters utilize either gas counters or solid-state detectors to capture the energy depositions of individual events within their active volume. The first type of detector built for microdosimetry were gas proportional counters made of tissue-equivalent plastic and filled with tissue-equivalent gas mixtures. Due to these features, they were referred to as "tissue-equivalent proportional counters" (TEPCs). In the early 1950s, H.H. Rossi and W. Rosenzweig pioneered the development of a spherical TEPC designed to assess the LET in complex radiation environments. Since one of the limitations of the LET is that it is defined as an average quantity in a well defined location, Rossi and Rosenzweig suggested operating the proportional counter at low pressure. This choice ensured that the product of gas density and detector diameter was on the order of 10^{-4} g/cm^2, equivalent to approximately 1 micrometer of material with unit density [Rossi and Rosenzweig, 1955]. Measurements of energy deposited by neutron irradiation, when compared with a sphere at standard pressure, showed much greater variation in the distribution of recoil proton path lengths. This experiment provided compelling evidence for the stochastic nature of energy deposited charged particles. While this phenomenon had been well-established theoretically, its direct observation had been limited to qualitative variations between traces left by identical ionizing particles in cloud chambers. Rossi recognized that the range of energy deposition observed in the simulated micrometer volume closely resembled the energy deposited in biological cells. This was later highlighted as a relevant parameter in the description of biological damage [Braby, 2015].

Manufactured from tissue-equivalent materials, TEPCs are microdosimeters with the optimal tissue equivalence property, providing accurate responses over a wide range of energy depositions for both primary and secondary charged particles. Furthermore, their high detection efficiency is attributable to electron multiplication within the gas filling. Due to these factors, various types of TEPCs have been constructed and employed for radiation field characterizations. Miniaturized TEPCs (referred to as mini-TEPCs), featuring sensitive volumes smaller than 1 mm, have been developed to function at therapeutic beam intensities [De Nardo et al., 2004]. These mini-TEPCs operate within the range of approximately 0.3 to 2 µm in equivalent tissue diameter. Various TEPC prototypes designed for use in ion-beam therapy have been researched and created. Examples include the low-pressure avalanche-confinement TEPC [Mazzucconi et al., 2019] and the gas electron multiplier (GEM) TEPC [Farahmand et al., 2004].

Another class of microdosimeters that do not rely on gas and instead operate in a semiconductor are the solid-state microdosimeters. The exploration of semiconductor detectors for microdosimetry begun in 1980 [Dicello et al., 1980]. More recently, a silicon microdosimeter, comprising an array of microscopic p–n junctions based on silicon-on-insulator (SOI) technology, was manufactured and subjected to testing with various radiation fields relevant to particle therapy [Rosenfeld et al., 2000]. Notably, [Agosteo et al., 2006] has proposed a silicon

telescope composed of ΔE and an E stage detectors (1.9 µm and 500 µm thick, respectively). Silicon detectors offer micrometer-scale sensitive volumes, making them highly valuable for microdosimetry. When coupled with tissue-equivalent converters, they are suitable for assessing the quality of radiation therapy beams. Utilizing micrometer-scale volumes helps mitigate the influence of wall effects on the recorded spectra. Wall effects are a significant concern when employing TEPCs, as events may deposit their energy in the TEPC's sensitive volume, including charged particles generated in the TEPC walls, which could only be partially captured at a micrometer-scale tissue site.
Additionally, in TEPCs, the electrodes require biasing at several hundred volts, necessitate precise gas pressure control, and face limitations in high-intensity radiation environments because the minimum geometrical cross-sectional area of the sensitive volume cannot be reduced below 1 mm^3. In contrast, solid-state microdosimeters operate at low voltage, and the thickness of their sensitive volume can be as small as 1 µm, making them suitable for measurements even in extremely intense radiation fields. Furthermore, pixelated arrays of detectors can be assembled, enabling simultaneous two-dimensional mapping of the radiation field quality [Colautti et al., 2020]. Compactness, low cost, portability, low power consumption and low sensitivity to vibration are other advantages of solid-state microdosimeters. However, the following problems need to be solved when using a silicon device for microdosimetry:

i lower signal-to-noise ration, limiting the minimum detectable energy.

ii Corrections for tissue equivalence should be made for the silicon sensitive zone.

iii The sensitive volume must be limited to a region of known dimensions.

iv Due to their geometry they are very sensitive to the direction of the incident particles.

v Since the sensitive zone of a silicon device can be approximated to a parallelepiped, a shape equivalence correction should be made when referring to a spherical site.

vi Arrays of detectors should be considered because the efficiency of a single detector of micrometric dimensions is very low [Agosteo and Pola, 2011]. The total area of the should be on the order of 2.5 mm × 2.5 mm, corresponding to a few hundred detectors [Lim et al., 2009].

2.1 Tissue equivalent proportional counter LET-1/2

In this work, a spherical TEPC was employed for measuring all microdosimetric spectra. The LET-1/2 TEPC model by Far West Technology is a proportional counter utilizing a spherical chamber with tissue-equivalent walls and a tissue-equivalent filling gas. The spherical design ensures that its response remains independent of the radiation source direction.

The detector consists of a spherical cavity with an internal diameter of 12.7 mm. It is housed in a vacuum-sealed 1.8 mm thick aluminum shell attached to an aluminum stem. This configuration allows the sphere to be waterproof, enabling the possibility of conducting measurements in water. The gas-filled sphere within the TEPC has an internal diameter of 12.7 mm and a wall thickness of 1.27 mm. Table 2.1 summarize the geometrical specifications of TEPC LET-1/2.

To simulate biological tissue, both the filling gas and the sphere wall are required to be tissue-equivalent.
The gases most commonly used are methane-based or propane-based. Since it has been demonstrated the equivalence of pure propane gas to fill TEPC, with the TE-propane, our detector was filled with pure propane gas [Chiriotti et al., 2015].

Aluminium stem	
Length	254 mm
Diameter	6 mm
Aluminium shell	
Length	6.35 cm
Wall thickness	1.8 mm
Diameter	2 cm
Sphere	
Wall thickness	1.27 mm
Internal diameter	12.7 mm

Table 2.1: Geometrical specifications of TEPC LET-1/2.

H	**C**	**N**	**O**	**F**	**Na**	**Mg**	**Si**	**P**	**S**	**K**	**Ca**
ICRU tissue, muscle [ICRU, 1964]											
10.2	12.3	3.5	72.9	-	0.08	0.02	-	0.2	0.5	0.3	0.007
Muscle-equivalent plastic A-150 [Smathers et al., 1977]											
10.1	77.6	3.5	5.2	1.7	-	-	-	-	-	-	1.8
Muscle-equivalent gas, with propane [Srdoč, 1970]											
10.3	56.9	3.5	29.3	-	-	-	-	-	-	-	-

Table 2.2: Elemental composition of compounds and mixtures in percentage by weight of the so-called *A-150* muscle-equivalent plastic and the muscle-equivalent gas propane-based, compared to the ICRU muscle definition.

A muscle equivalent plastic known as "A-150" is the primary microdosimetry material forming the wall of the TEPC LET-1/2 sphere. In Table 2.2, the elemental composition of muscle-equivalent compounds and mixtures is presented in weight percent, including the muscle-equivalent propane-based gas, and is compared to the ICRU 1964 muscle definition [ICRU, 1964].

A-150 is a mixture of polyethylene, nylon, carbon, and calcium fluoride. The elemental composition of *A-150* is very close to standard muscle tissue with respect to hydrogen and nitrogen. There is, however, a substitution of carbon for oxygen; in fact, no practically usable solid has an oxygen content equal to that of tissue. While trying to keep the composition as close to the muscle as possible, the A-150 is also designed to tightly seal the gas in the active area. [Smathers et al., 1977].

To simulate energy deposition within a microscopic volume of biological tissue, a larger cavity filled with a tissue-equivalent gas of much lower density is used. The simulation of this microscopic volume relies on two key variables: the physical dimensions of the detector (in the case of LET-1/2 the diameter d_g) and the density of the gas ρ_g. These variables together constitute a conjugate pair that governs the simulation of site size.
In order to be equivalent, it is required that the energy loss of the charged particles traversing the tissue sphere is the same as in the gas sphere, for equivalent trajectories. Let ΔE_t and ΔE_g be the mean energy losses of charged particles in tissue and gas, respectively

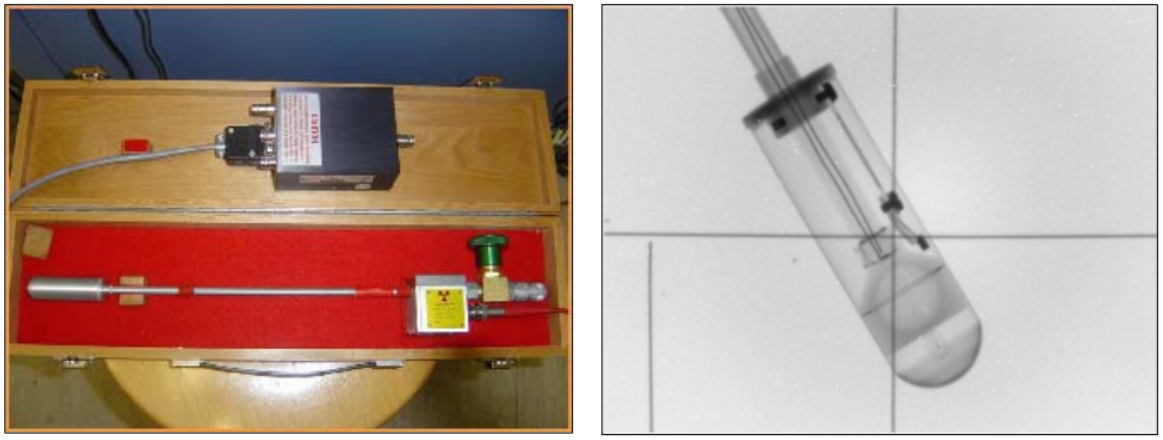


Figure 2.1: The TEPC LET-1/2 and its X-rays image

$$\Delta E_t = \left(\frac{S}{\rho}\right)_t \rho_t d_t \,, \tag{2.1}$$

$$\Delta E_g = \left(\frac{S}{\rho}\right)_g \rho_g d_g \,, \tag{2.2}$$

$$\tag{2.3}$$

where $\left(\frac{S}{\rho}\right)_t$ and $\left(\frac{S}{\rho}\right)_g$ are the mass stopping powers, ρ_t and ρ_g are densities and d_t and $d_g = d_t \kappa_{gt}$ are the diameters of the tissue and gas spheres, respectively.

So,

$$\Delta E_t = \Delta E_g \,. \tag{2.4}$$

By the definition of tissue equivalence, both the biological tissue and the tissue-equivalent gas have the same mass stopping power. Substituting Equation 2.1 and Equation 2.2 in Equation 2.4 and using $d_g = d_t \kappa_{gt}$:

$$\rho_g = \frac{\rho_t}{\kappa_{gt}} = \rho_t \frac{d_t}{d_g} \,. \tag{2.5}$$

Substituting ρ_t and d_g in Equation 2.5, it is possible to calculate the gas density for different site sizes. An approximation of the pressure required to simulate different site sizes can be obtained by using the gas density values, assuming the behavior of an ideal gas equation as:

$$PV = nRT \,,$$

or

$$PV = \frac{\rho}{M} RT \,,$$

where ρ is the gas density (in g/cm^3), M is the gas molecular weight, R is the universal gas constant, T is assumed to be the room temperature of 293.15 K.
In our case, the filled propane gas is at a pressure of 60 mbar to be equivalent to 2 μm of tissue. The corresponding density is:

$$\rho = \frac{P[atm] \cdot M[g/mol]}{R[\frac{m^3 \cdot atm}{K \cdot mol}] \cdot T[k]} = 1.08 \cdot 10^{-4} g/cm^3 \tag{2.6}$$

where $R = 8.206 \cdot 10^{-5} \frac{m^3 \cdot atm}{K \cdot mol}$, $M = 44.1 g/mol$.

2.2 Acquisition chain

To process and acquire signals from the LET-1/2 TEPC detector a dedicated acquisition composed by Nuclear Instrumentation Modules (NIM) is required. The internal design of the TEPC is carefully engineered to efficiently amplify and capture electrons generated by ionizing radiation within its active region. This is accomplished by using a static electric field strategically applied to create an avalanche effect on the electron released by ionizing radiation. Therefore, the main objectives of this design are to proportionally amplify the signal originating from the initial ionization event and to ensure the collection of all the electrons produced in the process. The magnitude of this electric field determines an initial level of signal amplification, allowing the user to adjust the detector gain. However, it is crucial not to exit the proportional region, because the output signals would cease to have a linear correlation with the initial number of released electrons. In all our applications, the TEPC LET-1/2 was operated at a voltage of +700 V provided by a high voltage power supply model NHQ 206L (Iseg).
The current signal generated by the collection of ionization events within the active area is extremely small (at most tens of µA [Rossi et al., 1996]), making direct processing impractical. As a result, the initial step in the acquisition chain involves the use of a charge-sensitive preamplifier, which in this work was the A422 model from CAEN [CAE, a]. In the acquisition chain, employing a preamplifier as the initial stage is a common and practical choice.

The primary role of the preamplifier is to transform the current signal into a more manageable voltage signal. This is accomplished by inverting the polarization of the TEPC collected current and by encoding it into a measurable voltage signal. Figure 2.2 shows the signal originated from radiation, captured after the preamplification stage. A zoom is presented in the figure to better appreciate the rapid variation of the signal with a slew rate of 1425 ± 23 V/ms. The rapid amplitude change occurs just within 0.2 µs.

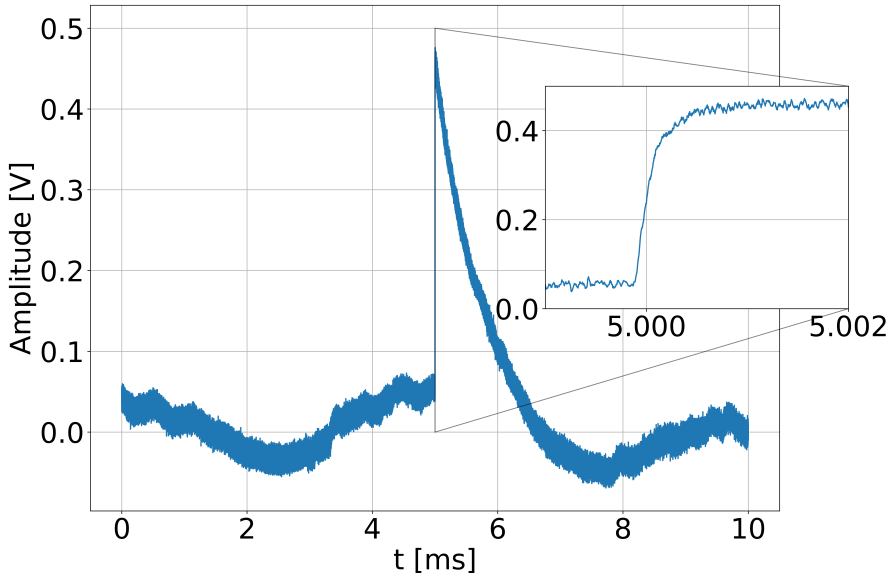


Figure 2.2: TEPC signal after the preamplifier. The box illustrates a zoom of the region containing the rapid increase of the signal with an estimate slew rate of 1425 ± 23 V/ms.

After the energy deposition signal has been processed by the preamplifier, it is fed simultaneously to three differently configured *shaping amplifiers*. The reason for this is the need for

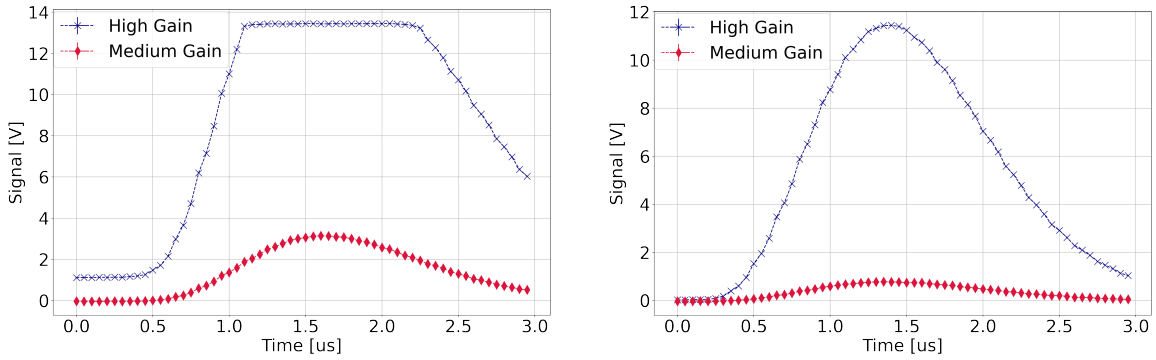
good resolution over a wide dynamic range in microdosimetry. The task of the amplifiers is to shape and amplying the input signal, efficiently transforming the rapid signal shown in Figure 2.2 in a semi-Gaussian waveform, although other waveforms are possible depending on the specific model. This process encodes the energy deposition data from the preamplifier into the amplitude of the resulting output signal.

Shaping amplifiers are characterized by a parameter known as the "shaping time constant", which determines how the signal is filtered and, consequently, the duration of the output signal. Ideally, the choice of this shaping constant should consider both parallel and serial noise contributions, in order to minimize the overall noise by selecting an appropriate shaping time [Letzter and Webster, 1970, Knoll, 2010]. However, it is important to note that not all experimental setups exhibit this characteristic and, in most cases, noise reduction cannot be achieved only through the adjustment of the shaping time constant. Therefore, regulating the shaping time constant involves finding a compromise between achieving sufficient integration time for accurate output while mitigating the risk of signal pileup. Longer shaping constants result in longer signals and are therefore more prone to overlap, as the tail of a signal may persist while a new signal is being initiated. This is true especially at a relatively high particles rate, where the temporal distribution of the signal becomes more dense. This combination of conditions increases the likelihood of signal overlapping, which can, in the worst scenario, irreparably corrupt the acquired data. On the other hand, shorter signals produced with smaller shaping constants are less prone to overlap, but tend to have shorter integration times. This shorter integration time can result in signals with higher levels of noise, reducing the signal quality and making it more challenging to distinguish meaningful events from background noise.
A shaping time of 2 µs was found to be the better compromise between signal duration and integration time.

Although the shaping constant is the same for all three shaping amplifiers, the gain is different. The three gain settings are roughly: ×1000, ×100, and ×10. Because of this choice, the higher gain is also called "high gain", the middle gain "medium gain", and the lower gain "low gain". In this work, the N968 module by CAEN [CAE, b] is used for the high- and medium-gain amplification, while the 7243E module by Intertechnique is used for the low gain. The manual of the N968 states that the maximum output amplitude is 10 V. With a signal of limited amplitude, to prevent signal saturation, e.g. reaching the maximum output voltage, different amplification stages must be adopted. In this way, signals that saturate in the higher gain stages are likely to be correctly processed by an amplifier with lower gain, independently from the signal amplitude after the preamplifier. This can be observed in Figure 2.3a, which shows the signals after high and medium gain amplification. The high gain signal displays a flat region at its peak amplitude, which is a characteristic indicator of amplifier saturation. On the other hand, the signal processed by a lower gain, which in this specific case is associated with the medium gain, is processed correctly.

A similar scenario is anticipated for the medium and low gain signals: the medium gain signal is saturating, while the low gain signal is being processed correctly.

To effectively utilize the information obtained from all three gain stages, it is essential to merge the spectra. Consequently, the amplifiers' gains are selected in a way that, for a specific range of input signals amplitude, the same signal is accurately processed by both the high and medium gain amplifiers, without saturation. This scenario is shown in Figure 2.3b where the same signal is correctly processed by both medium and high gain amplifiers. In the same way, for a given range of input pulses, the same signal can be processed correctly by both a medium gain amplifier and a low gain amplifier. These gain selections allow for the establishment of linear relationships between the signal amplitudes processed by different gains. Specifically, linear relations can be identified between the signal amplitudes processed by the high gain



(a) (b)

Figure 2.3: Signal acquired after the high gain (blue cross) and medium gain (red diamond). In Figure 2.3a the high gain signal is saturating while the medium gain signal is processed correctly. In Figure 2.3b the signal is processed correctly in both the medium and high gains.

and medium gain, as well as between the medium gain and low gain amplitudes.

Following the amplification stage, the signals are prepared for processing and storage. This function is carried out by MultiChannel Analyzers (MCAs), which are responsible for computing a histogram based on the amplitudes of the incoming signals. To accommodate for the three signal amplification levels, separate MCAs were employed. The high and medium gain signals were processed using an Ortec MCA Model 927 [ORT, b], while the low gain signal was handled by an Ortec Model 926. The Model 927 offers two inputs and has a higher resolution with 2^{16} channels, while the Model 926 has a resolution of 2^{10} channels. These MCAs are controlled using the dedicated software Maestro [ORT, a], which allows the user to control the acquisition and to store the acquired data in text files. In addition, Maestro provides some useful information about the acquisition: histogram total counts, *real time*, and *live time*. The real time is the actual acquisition time, while the live time takes into account the signal processing time required by the MCA. In fact, each time a signal is acquired, a time interval referred to as *dead time* (t_d) is required to process the signal. The value of t_d can be estimated as the difference between the real time (t_r) and the live time (t_l):

$$t_d = t_r - t_l \,. \tag{2.7}$$

While the MCA is in dead time, it cannot process any other signal, losing all energy deposition events arriving from the TEPC in that temporal interval. In the ideal scenario, the live time should match the real time, meaning that the MCA was always active and no events are lost. Dead time is usually expressed as a percentage of the total time:

$$t_d[\%] = \frac{t_d}{t_r} \cdot 100 = \left(1 - \frac{t_l}{t_r}\right) \cdot 100 \,. \tag{2.8}$$

Through experimental verification, I established that when the dead time exceeds approximately 20%, the reliability of Maestro in estimating the dead time estimate diminishes. Thus, it is advisable to stay within this threshold to ensure an accurate dead time estimation. As the dead time is closely related to the particle rate recorded by the TEPC detector, it is essential to carefully select the particle rate to prevent the acquisition from reaching saturation. Finally, Maestro allows the user to save the acquired histogram with the additional real time and live time information. These files, one for each amplification (high, medium and low), are the starting point for the data analysis.

The acquisition chain is schematized in Figure 2.4.

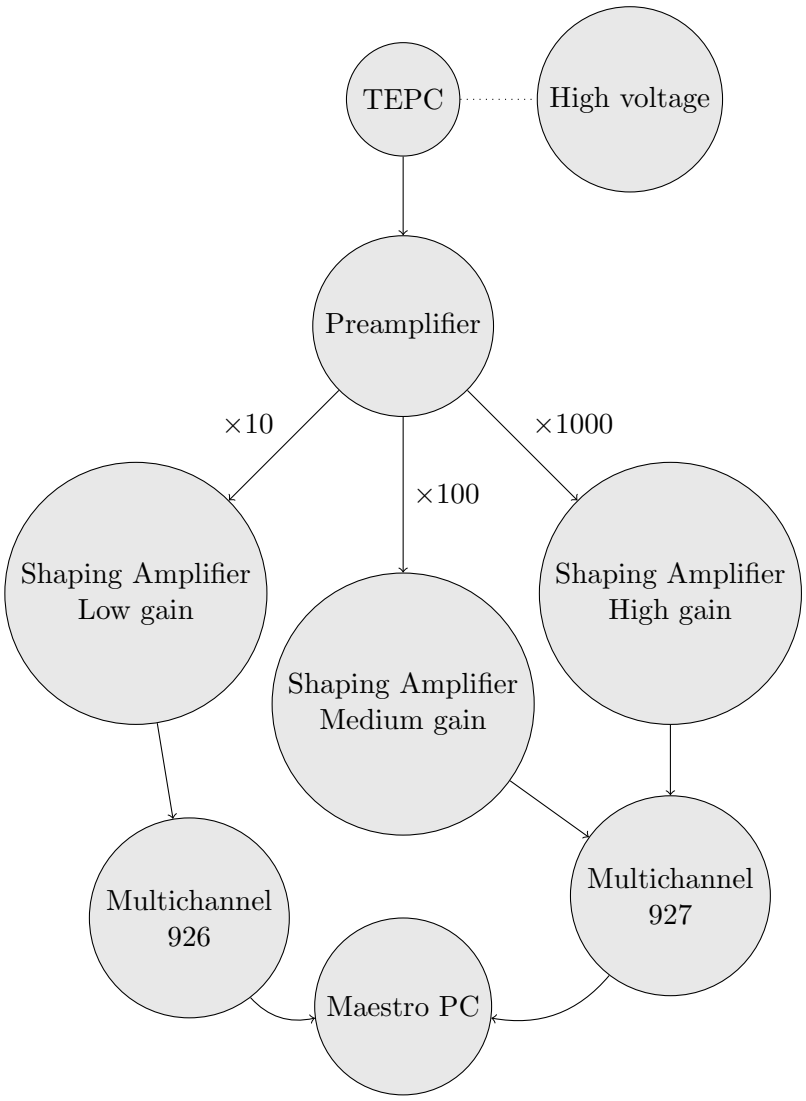


Figure 2.4: The TEPC receives a voltage input (HV), and its raw signal is initially directed to a preamplifier for amplification. Subsequently, the amplified signal is distributed into three amplifiers, each offering different gains (low, medium, and high). The outputs from these amplifiers are then routed to the Multichannel Analyzers, managed by the Maestro software.

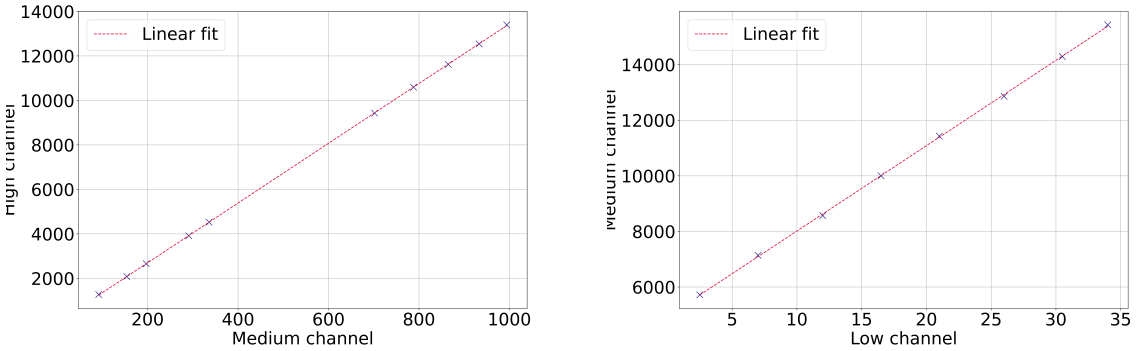
Calibration and intercalibration

Once the data has been acquired using the acquisition chain described above, the actual microdosimetric spectrum must be derived from the acquired data. Two main steps are required to build a microdosimetric spectrum from our initial data, namely *intercalibration* and *calibration*. Intercalibration is the process necessary to properly join the MCA data acquired from the three gain levels into a single one, while calibration is necessary to assign each channel to a lineal energy value (y). Once these steps have been performed, the microdosimetric spectra and its derived quantities can be calculated.

Intercalibration

Intercalibration is the process of determining the relationship between the amplitudes of signals processed by different amplifiers operating at varying gain levels. As mentioned in

section 2.2, the three gain levels are chosen to allow the same signal within a specific amplitude range to be coherently processed by two amplifiers, either high and medium, or medium and low. In this way, by using an arbitrary waveform generator to simulate the detector signals with a specific amplitude, it is possible to simultaneously populate certain channels of the MCAs for high and medium and iterate the process for the medium and low channels. Then the amplitude of the injected pulse can be adjusted to populate different regions of channels. By iterating this process and plotting the respective channels, it is possible to obtain Figure 2.5a. A linear fit of the type $y = mx + q$ was performed on the populated common channels. The same process is repeated also for medium-low channels as shown in Figure 2.5b.



(a) (b)

Figure 2.5: Intercalibration of the high-medium in Figure 2.5a, while Figure 2.5b shows the intercalibration for the medium-low. A linear fit was also performed for both datasets. For the high-medium the obtained fit coefficients are: $m_{\text{h-m}} = 13.434 \pm 0.014$ and $q_{\text{h-m}} = 12 \pm 10$. For the medium-low the estimated fit coefficients are $m_{\text{m-l}} = 306 \pm 2$ and $q_{\text{m-l}} = 4944 \pm 35$.

It is then possible to linearly transform the channel of the medium gain and low gain spectra acquired with the MCA to an high gain spectrum using the previously found linear coefficients:

$$\text{ch}_{\text{medium to high}} = \text{ch}_{\text{medium}} \cdot m_{\text{h-m}} + q_{\text{h-m}}, \ \text{ch}_{\text{low to medium}} = \text{ch}_{\text{low}} \cdot m_{\text{m-l}} + q_{\text{m-l}}, \tag{2.9}$$

In this way, all three spectra are related to the high gain amplification and it is possible to combine the channels coming from different gain levels. Furthermore, it has been found through experimental verification that the uncertainties due to the intercalibration process are negligible compared to the uncertainties due to the calibration step (more on this in the next section). The coefficients obtained are valid as long as the parameters of the medium and low shaping amplifiers are not modified. In this way, the same intercalibration can be reused for several acquisitions, as long as it remains valid. This intercalibration procedure automatically accounts for and corrects nonlinearity effects that may occur with the medium and low gain multichannel analyzers.

Calibration

Calibration is an essential step in the construction of a microdosimetric spectrum from raw data. Its purpose is to establish a correlation between the MCA channels and the lineal energy y. To achieve this, specific data points within the microdosimetric spectra are identified and linked to their corresponding y-values.

To obtain a robust calibration, multiple calibration points were considered. In particular, I have found that a reliable and easily identifiable point is the peak of the $f(y)$ spectra. The peak of $f(y)$ corresponds to the most probable y and consequently to the most probable energy deposition. In the case of protons as the radiation source, the most probable energy deposition is due to the primary protons.
As an additional reference point, the proton edge [Conte et al., 2013, Moro et al., 2015] is adopted in the calibration process. Physically, the proton edge corresponds to the maximum amount of energy that a proton can release inside the microdosimeter. Therefore, in the case of proton edge deposition, the geometry of the microdosimeter plays a crucial role in determining energy deposition. This is because, to achieve proton edge deposition, the microdosimeter's boundary must coincide with the end of the proton range.

An experimental campaign was conducted at the Proton Therapy Center in Trento with the primary objective of calibrating the TEPC detector. To optimize the separation of the signal from electronic and environmental noise and improve the visibility of the $f(y)$ peak, the lowest derivable proton energy of 70 MeV was used [Tommasino et al., 2017]. Lowering the primary proton energy results in a shift of the $f(y)$ peak towards higher values of y because slower protons deposit more energy within the microdosimeter, consequently leading to a higher y values.
To obtain multiple calibration points, I shifted the peak of the $f(y)$ spectra, lowering the proton energy even more by using RW3 water equivalent material and placing it along the beam line. In addition to that, 37 mm of RW3 were used in combination with a primary energy of 70 MeV to obtain proton edge depositions in the TEPC.
The experiments were replicated using Monte Carlo simulations with Geant4, simulating microdosimetric spectra for various thicknesses of RW3. It is anticipated that, especially within the range of clinical energies (from 0 up to around 230 MeV), Monte Carlo accurately models the transport of primary particles. As a result, the peak of $f(y)$ spectra has sufficient accuracy to be used as reference point in the calibration.
Regarding the proton edge, a dedicated simulation was performed with 37 mm RW3 material, replicating the experimental setup.
The shape of the proton edge follows a Fermi-like function [Conte et al., 2013]:

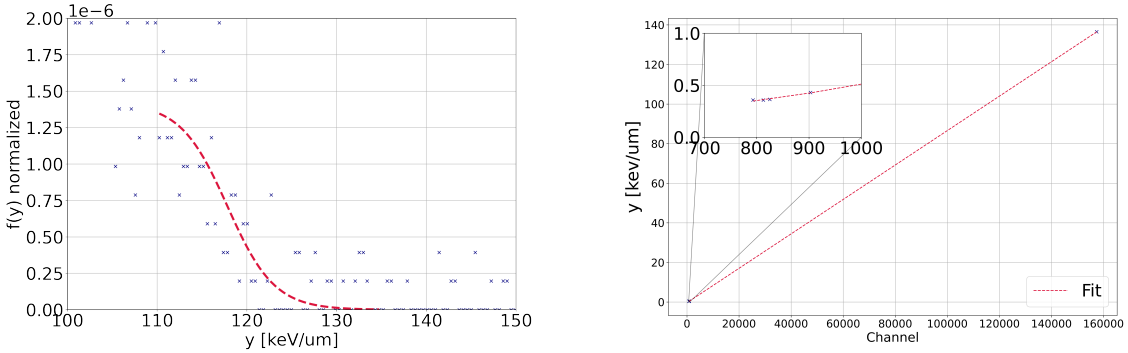
$$f(x) = \frac{A}{1 + \exp\left[B\left(x - C\right)\right]}\,. \tag{2.10}$$

Figure 2.6a shows the proton edge fit performed on simulation data with a primary beam of 70 MeV and 37 mm or RW3 material to lower the energy.
It has been established that the most stable and precise reference point can be determined using the Fermi-like function. This reference point corresponds to the intercept of the tangent line through the inflection point with the x-axis, as described in previous studies [Conte et al., 2013, Moro et al., 2015]. Mathematically, it can be expressed as:

$$x_{\mathrm{TC}} = \frac{2}{B} + C\,. \tag{2.11}$$

The same point can be identified by applying the same procedure for the experimentally acquired microdosimetric spectra.

The calibration process involves plotting the simulated values of y for the peaks of the $f(y)$ distribution as well as the x_{TC} against their corresponding channels from the experimentally acquired MCA data. This relationship is visually represented in Figure 2.6b, with an added zoomed-in view to better visualize the densely populated data points in the lower y region. To best fit the data of Figure 2.6b a quadratic relation of the form $y = a^2 \cdot \mathrm{ch}^2 + b \cdot \mathrm{ch} + c$ was



(a) (b)

Figure 2.6: Figure 2.6a shows the proton edge of simulated data where the Equation 2.10 was fitted. The estimated coefficients were: $A = (1.3 \pm 0.2) \times 10^{-6}$, $B = 0.03 \pm 0.1\,\mu m/keV$ and $C = 117 \pm 1\,keV/\mu m$. Figure 2.6b shows the calibration points and the quadratic fit performed on the data.

adopted, leading to the following calibration equation:

$$y[keV/\mu m] = \\ = (9.875 \pm 7) \cdot 10^{-10} \cdot \mathrm{ch}^2 + (7.1 \pm 1.1) \cdot 10^{-4} \cdot \mathrm{ch} - (2.2 \pm 0.9) \cdot 10^{-1}\,.$$

With this information, it is finally possible to associate a channel value with a value y, thus calibrating the spectra.

2.3 Data analysis with Mandarina

Once the calibration and intercalibration procedures are completed, the microdosimetric spectra can be calculated from the MCA data. To accomplish this task automatically, I developed a software called *Mandarina* written in C#. C# is a modern, high-level, object-oriented programming language developed by Microsoft. It was first introduced in 2000 as part of Microsoft's .NET [Hejlsberg et al., 2003, Wigley et al., 2009], and is often used to develop desktop applications primarily due to its easy-to-use graphical interface while retaining all the potential of an object-oriented programming language that uses a C++-like syntax. Overall, by relining to C# it is possible to create a Graphical User Interface (GUI) in a few steps, and thanks to its sintax and included libraries, data analysis can be implemented as well as porting code from existing implementations of C-like languages. While these advantages are notable, it is crucial for the programmer to also prioritize the robustness of the program. To ensure the reliability of the data analysis process, all input from the end user must undergo validation to prevent errors during the code execution.
A screenshot of Mandarina running is shown in Figure 2.7.

Mandarina requires Maestro files as input data with a specific patterns: fa_1a_2_*gain*.Spe where

- a_1a_2 are the index of the run, so that multiples file can be analyzed.
- *gain* is a string that defines if the files come from the "high", "medium" or "low" gain.

A filename of type f01_high.Spe is associated with f01_medium.Spe and f01_low.Spe, and all three are needed to construct the microdosimetric spectrum. If a different filename is specified,

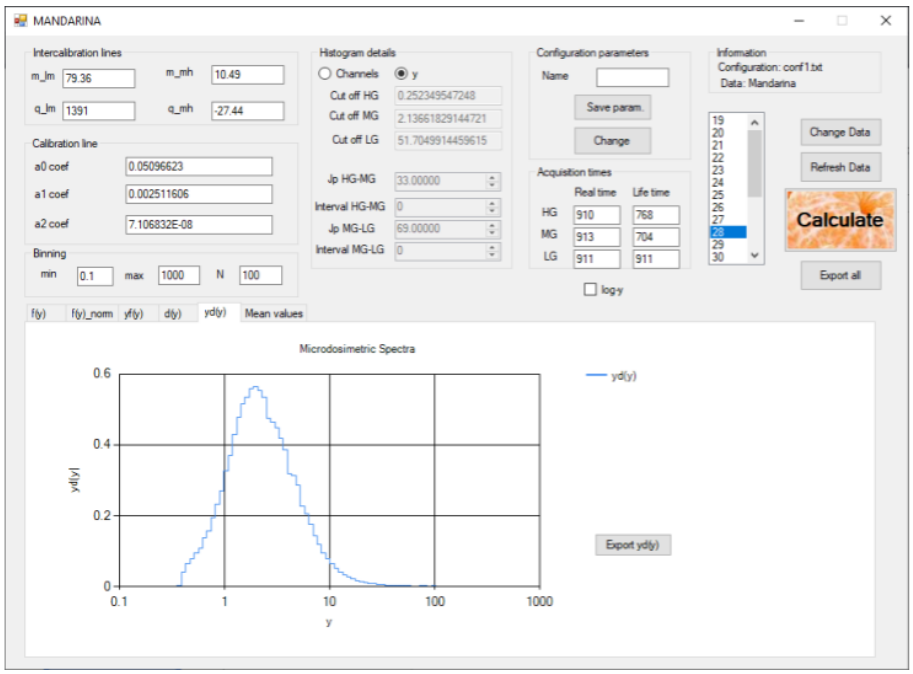


Figure 2.7: A screenshot of Mandarina graphical user interface.

or if not all files are found, the program simply ignores the file.
Another crucial step is to acquire a noise file. This file is obtained acquiring with Maestro without exposing the detector to radiation, typically for a period of around 5 minutes. It is important to perform this noise acquisition step before each experimental session. This allows for the assessment of the noise contribution and helps in identifying any unusual sources of noise. Mandarina requires specific naming for the noise file: noise_high.Spe, noise_medium.Spe and noise_low.Spe.

Mandarina incorporates an automatic noise correction process, which utilizes the provided noise files. This correction involves subtracting the noise-affected channels, with properly normalized noise. Given that the acquisition time of the noise differs from that of the microdosimetric measurement, a normalization step is executed to align the noise counts with the acquisition time. Mathematically, this adjustment involves scaling the counts in the i-th bin by the ratio of the real times of the noise (t_r^{noise}) and the real time (t_r) of the acquisition that needs correction:

$$n_i^{\text{noise, corrected}} = n_i^{\text{noise}} \frac{t_r}{t_r^{\text{noise}}} \,. \tag{2.12}$$

Once this correction is applied, the noise contribution can be subtracted from each bin of the spectrum. Usually, this correction predominantly affects the lower channels of the high gain because their relatively small amplitude compared to the noise makes them more susceptible to noise contributions. Indeed, the initial high gain channels (those with channel numbers below 650) and some of the medium gain channels (below 80) are primarily impacted by noise. In such cases, noise subtraction proves ineffective due to noise overwhelming the signals. Consequently, these channels are directly eliminated by the MCA, with their counts set to zero. Furthermore, if the noise contribution remains excessively high in certain channels, rendering noise subtraction ineffective, Mandarina provides the option to remove these channels if they fall below a specified channel value. With Mandarina, users can visualize these channel thresholds in terms of channel number or in y after applying the calibration procedure. This feature grants users greater control over the regions of the spectra they wish to discard.

Channel removal functionality is accessible for all individual data sets, including high, medium, and low gains. Typically, the high gain data set requires more attention due to its greater susceptibility to noise.

To improve accuracy, a further correction is performed for each bin of the acquired spectrum to correct for the data acquisition dead time. Each bin count is multiplied by the ratio of real time to live time:

$$n_i^{\text{adjusted time}} = n_i \frac{t_r}{t_l} \,. \tag{2.13}$$

The ratio between real time and live time should take into account the missed counts due to dead time. In the ideal scenario where $t_r = t_l$, no correction is made. Usually $t_l < t_r$, and therefore the ratio between the two compensates for the counts lost due to dead time. This correction affects all bins of the spectra, changing the amplitude but not the shape as in the case of the noise correction.

To summarize, each bin is adjusted by both the noise and the dead time correction:

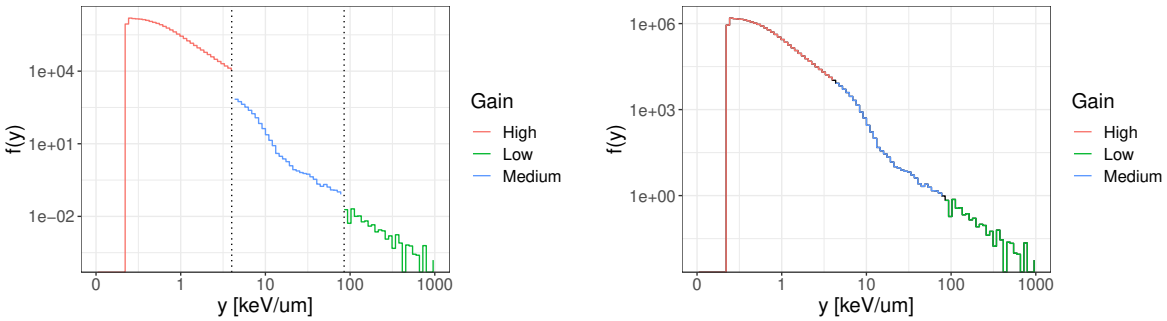
$$n_i^{\text{adjusted}} = n_i \frac{t_r}{t_l} - n_i^{\text{noise}} \frac{t_r}{t_r^{\text{noise}}} \,. \tag{2.14}$$

The implemented pileup rejection system available in the MultiChannel Analyzer is not reliable. Therefore, to avoid distortion of the spectra, it is preferable to implement a custom dead time correction. Once the histogram counts of all the amplifications are corrected, Mandarina applies the intercalibration. The user can provide the coefficients estimated from Figure 2.5 necessary to merge the three spectra into a single one. After the intercalibration, the program applies the calibration using the coefficient provided by the user. The result of these two steps is three calibrated spectra consisting of the high, medium, and low gain components, each spanning different channel ranges, ready to be merged. The next step is referred to as *re-binning*. Normally, microdosimetric spectra are built with a logarithmic binning from 0.1 keV/µm to 1000 keV/µm with 100 bins. As a consequence, the three spectra have to be adjusted to the new binning. However, Mandarina provides the ability to customize this range and the number of bins.

The next step is called *joining* and corresponds to the adjustment of the spectra height and merging of the high, medium, and low gain spectra into a single one. Although the three spectra are calibrated, intercalibrated, and have the same x-scale as a result of rebinning, they may not have compatible heights. This is true even after the correction for dead time and noise implemented in Equation 2.14. To adjust the heights, Mandarina simply takes the ratio of one or more overlapping bins and uses it to normalize the entire spectra. The user can select the bin to use as the joining point within the overlapping range between high-medium and medium-low. In addition, the user can select whether a single bin can be used, or instead use the average height ratio of multiple bins. In the joining process illustrated in Figure 2.8, one can observe the three separate spectra from high, medium, and low gains, each adjusted within its respective region as defined by the joining points. For the high-medium joining, the joining point is set at 4 keV/µm, while for the medium-low gain joining, it is positioned at 85 keV/µm. These spectra components are presented in more detail in Figure 2.8a. Figure 2.8b shows the final result after joining. Normally, the high gain is used as the reference, so the medium gain amplitude of the spectra is adjusted to match the high gain height in the bin selected for joining (4 keV/µm). The same process is repeated to join the medium and low gain, using the adjusted medium gain as a reference and correcting the low gain.

Once spectra are correctly joined, Mandarina offers the user the possibility of modifying in real time the joining point to check the robustness and see how changing joining point affects the final spectra.

Mandarina places a strong emphasis on user flexibility, enabling the modification of essential parameters such as calibration, intercalibration, binning, and joining points, while



(a) (b)

Figure 2.8: Joining process of the high, medium and low spectra. Figure 2.8a shows the single component adjusted to span their domain defined by the joining points: 4 keV/µm for the high-medium joining and 85 keV/µm for the medium-low marked as dotted vertical lines. Figure 2.8b shows the final result after the joining of the spectra using a single bin as overlap.

providing real-time visualization of the spectral changes. This allows users to promptly observe the impact of parameter adjustments on the spectra. In fact, Mandarina shows a preview of the $f(y)$, $yf(y)$, $d(y)$ and $yd(y)$ distributions (equations (1.1) and (1.3)) in different tabs that the user can select. It also calculated and displays the main average value typical of microdosimetry: y_F, y_D and y^* (equations (1.2), (1.4) and (1.5) respectively). Mandarina also offers the possibility to export all the calculated microdosimetric spectra in Comma Separated Valued (CSV) format, with appropriate names that keep the same file numbering a_1a_0 of the input files. In this way, a set of microdosimetric spectra with the same parameters can be calculated and stored in the internal memory of the running PC.
Mandarina can keep track of the parameters used to calculate the spectrum in a dedicated configuration file. This file can be stored in the PC and can be loaded as necessary allowing for the storing of different parameters and configuration used to calculate the microdosimetric spectra.

A screenshot of the Mandarina Graphical User Interface (GUI) is shown in Figure 2.9. The different region of the program are highlighted in different colors specifying their functionality. In summary, using Mandarina, the user can modify:

The intercalibration , highlighted in red in Figure 2.9, this comprehends two line equations coefficients: "m_mh" and "q_mh' for the high-medium, "m_lm" and "q_lm" for the medium-low.

The calibration , highlighted in orange in Figure 2.9, includes three parameters of a quadratic equation: "$a0$ coef", "$a1$ coef" and "$a2$ coef", corresponding to an equation of the form: $a0\ \text{coef}^2 c^2 + a1\ \text{coef}\, c + a2$ where c is the channel.

The binning of the microdosimetric spectrum, highlighted in green in Figure 2.9, defines the minimum ("min" in the GUI), the maximum ("max") and the number of bins ("N") used to build the microdosimetric spectrum.

The cut position to remove channels where the noise contribution is too high, highlighted in light blue in Figure 2.9, removes the channels below the selected threshold. Mandarina can convert the selected channel values to y to provide feedback to the user. Cut off HG" defines the cut for the high gain spectrum, while "Cut off MG" and "Cut off LG" define the cut for the medium and low gain spectra.

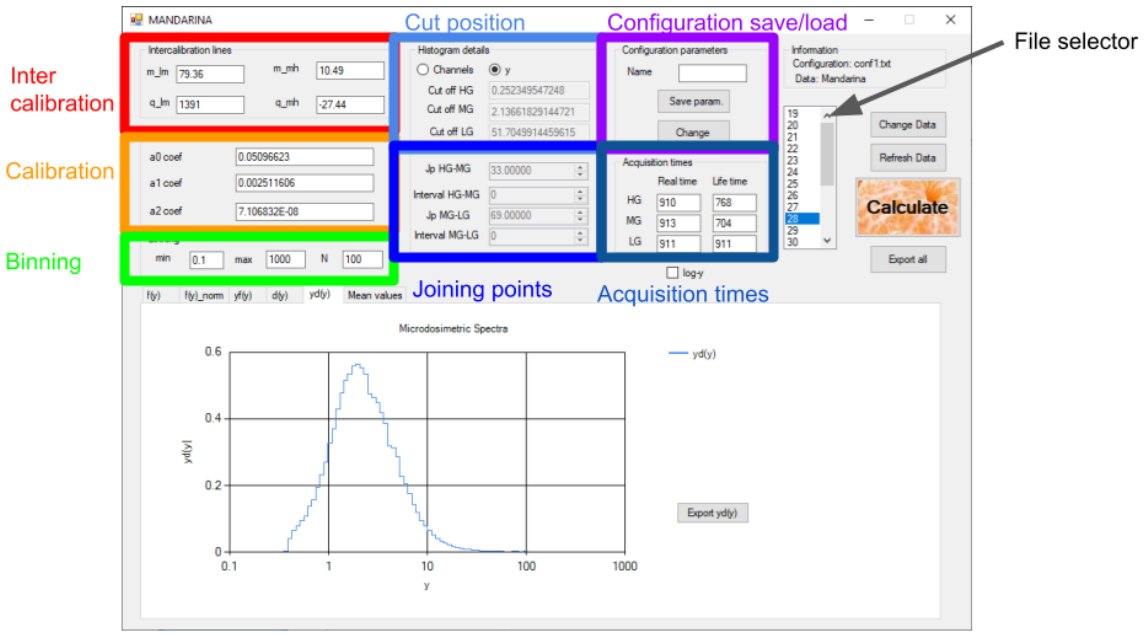


Figure 2.9: A screenshot of Mandarina graphical user interface. All the different user inputs are highlighted in different colors depending on their functionality.

The joining points , highlighted in dark blue in Figure 2.9, specifies which point is to be used as a joining point between high-medium ("Jp HG-MG") spectra and, in case the user wants to specify multiple points, "Interval HG-MG" selects how many. The same selection can be made for "Jp MG-LG" and "Interval MG-LG" for the medium-low joining. As in the case of the cut position, Mandarina can provide the values of joining points directly in y to help the user.

One can also load and save a configuration file containing all the aforementioned variables (purple square in Figure 2.9). This configuration file allows for the analysis of multiple spectra using the same settings, and the resulting microdosimetric spectra and data can be displayed and saved as needed.
Finally, in the blue square of Figure 2.9, Mandarina conveniently presents the acquisition times, both real time and live time, for all acquired spectra high, medium, and low.

2.4 Detector Pileup

2.4.1 Introduction

The Tissue Equivalent Proportional Counter (TEPC) is widely recognized as the standard reference detector in microdosimetry [ICR, 1983, Magrin et al., 2023, Missiaggia et al., 2020, Missiaggia et al., 2023, Colautti et al., 2020]. Depending on the active volume size, TEPCs can sustain different particle rates, ranging from low intensities in the order of 10^3 of particles per seconds for spherical geometries of $\sim 10\,\text{cm}$ radius [Farah et al., 2015], to clinical intensities above 10^6 particles per seconds for cylindrical active regions of $\sim 100\,\mu\text{m}$ of radius [Conte et al., 2020]. When the particle rate exceeds the detector capability, it leads to an effect known as pileup, which can influence the microdosimetric spectrum.

Pileup can originate both from the detector collection and the electronic acquisition chain. If the particle rate is sufficiently high, an event can traverse the active area while the elec-

trons of a previous ionization are still being collected. This second event can disrupt the electric field and consequently influence the collection of electrons, creating a distorted signal. When these ions are produced at a rate exceeding their recombination rate, they accumulate within the active region of the detector. This accumulation disrupts the electric field, leading to an irregular proportionality in the charge collection. Pileup can also occur in the electronic acquisition chain if the time between multiple events is too short and no pileup rejection is implemented [Langen et al., 2002]. The impact of electronic pileup on the spectra depends on the sequential components of the chain, but typically causes the overlap of two or more signals, resulting in the corruption or complete loss of data. The direct impact of pileup on the microdosimetric spectra is the broadening and shifting of the lineal energy $f(y)$-distribution, and, consequently, of all spectra derived from it. This, in turn, results in fluctuations in the frequency-mean lineal energy y_F and the dose-mean lineal energy y_D. These microdosimetric quantities are commonly used for characterizing radiation quality in microdosimetry-based radiobiological modeling [Hawkins, 1994, Bellinzona et al., 2021], and their inaccuracy potentially leads to limited predictions of the biological outcomes.

Acquiring high-quality microdosimetric data without pileup is challenging, especially in clinical facilities where the particle rate is limited by design and operation constraints. Microdosimetry is considered a superior tool compared to macroscopic Linear Energy Transfer (LET) when it comes to characterizing the quality of the radiation field. Due to this advantage, its potential application for daily Quality Assurance (QA) in clinical facilities is currently under investigation [Magrin et al., 2023, Rosenfeld, 2016, Bianchi et al., 2021, Colautti et al., 2020].

Many beam delivery companies, especially those in charge of clinical facilities, are reluctant or unable to adjust the beam current outside of the clinical setting. Any change in beam delivery could result in a different dose delivered to the patient and therefore any adjustment to the beam line is prudent. Consequently, medical facilities are faced with the challenge of being unable to lower particle rates to levels that would effectively mitigate the issue of pileup [doi, 2007].

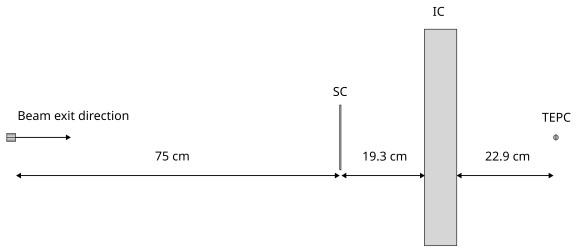
Pileup in TEPCs has only been investigated with neutron beams and in a limited range of particle rates (3 kHz to 55 kHz) [Langen et al., 2002]. In this study, the response of a commercial TPEC to clinical protons at various rates within the range of 10^3 to 10^6 particles per second was experimentally investigated. Combining the experimental findings and GEANT4 Monte Carlo simulations, we developed a methodology to accurately evaluate the pileup on the measured spectra and built a rate-pileup curve. This work yields two significant contributions: i) it offers a direct measurement of pileup and its impact on radiation quality in the context of clinical protons, and ii) it outlines a precise methodology for evaluating the degree of pileup from measured microdosimetric spectra in a wide range of experimental conditions.

2.4.2 Material and methods

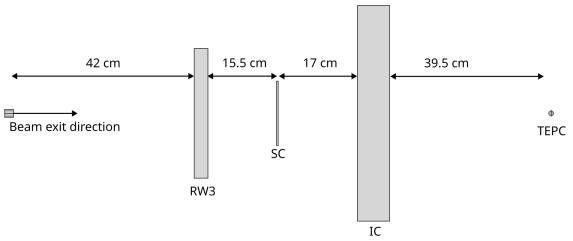
Experimental setup

All measurements were performed at the Proton Therapy Center in Trento (Italy) [Tommasino et al., 2017]. The microdosimetric spectra were acquired with a spherical TEPC model LET-1/2 from Far West Technology filled with propane gas, and equivalent to a tissue sphere of 2 µm in diameter. The irradiations were performed at two proton energies of 11 MeV and 70 MeV, with beam rates between 1.2 ± 0.2 kHz and 2.60 ± 0.11 MHz. These two energies were selected to explore pileup effects at two distinct beam qualities, and for lineal energy distributions fully above the TEPC readout noise limit of 0.2 keV/µm. Since the minimum active energy delivered by the facility is 70 MeV, we used 35 mm of RW3 (solid water) material to degrade the beam energy to 11 MeV. The scheme of the experimental setups used for the two proton energies

is illustrated in Figure 2.10. The beam rate was monitored with an Ionization Chamber (IC) placed upstream of the TEPC. The raw output provided by the IC integrates the energy deposited by all events during the acquisition time and requires calibration to be converted into a rate. For this purpose, we placed a 1 cm thick plastic scintillator (SC) in front of the IC, and used it to calibrate the IC count for both proton energies. To avoid degradation of the beam energy and saturation of the SC for rates above $\approx 5 \times 10^5$ Hz, the SC was removed during the actual acquisition of the microdosimetric spectra. The TEPC was placed at the isocenter, 125 mm away from the beam exit. In both experimental setups, the TEPC was fully immersed in the beam profile. The same is true for the IC and the SC.



(a) Experimental setup for the 70 MeV proton beam measurement.



(b) Experimental setup for the 11 MeV proton beam measurement.

Figure 2.10: Geometry for experimental setup using 70 MeV primary proton beam and, a modulated 70 MeV initial energy using 35 mm of RW3 to obtain $\simeq$ 11 ± 5 MeV average residual energy. For both experiments, the scintillator was removed prior to the acquisition of the microdosimetric spectra.

Electronic acquisition chain

The electronic acquisition chain for the TEPC is described in section 2.2.

The SC signal was connected to a discriminator module N841 by CAEN, and the number of protons detected was counted with a timer and counter module ORTEC-871. The IC was read out using specialized acquisition software that generated files, including counts and relative timestamps with a resolution of 200 ms.

Monte Carlo simulations

All simulations were performed using Geant4 Monte Carlo toolkit [Agostinelli et al., 2003] version 10.06.p01, using experimental geometry illustrated in Figure 2.10. However, to obtain the closest match of the simulated peak with the experimental data, an additional 2 mm of RW3 was required for a total of 37 mm of RW3. The single scattering mode was used to describe electromagnetic interactions, while hadronic interactions were described by the G4_QGSP_BIC_HP physic list.

2.4.3 Results

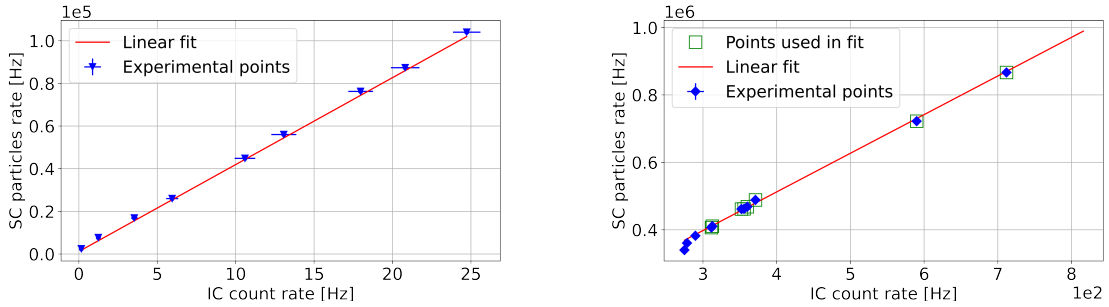
In this study, we explored the connection between particle rate and pileup for a spherical TEPC exposed to clinical proton beams of varying energies. To construct a rate-pileup calibration curve, we first had to determine the particle rate at the TEPC and assess the degree of pileup for the corresponding microdosimetric spectrum.

Particles rate

In principle, MAESTRO can estimate the particle rate at the TEPC. However, we observed experimentally that as the dead time exceeds $\sim 20\%$, MAESTRO measurements of the particle count become inaccurate. To overcome this issue, and correctly estimate the particle rate at the TEPC independently of the beam intensity, we developed a two-step process. First, we

used the plastic scintillator (SC) to convert the read-out signal from the ion chamber (IC) into a particle rate. As a second step, we conducted a calibration between the TEPC and the calibrated IC to precisely determine the TEPC rate. The reason for not directly calibrating the TEPC with the SC is that the plastic scintillator cannot cover the entire range of rates selected for this study, as it begins to saturate around 0.5 MHz. Following this approach, we could assess the rate at the TEPC without relying only on MAESTRO.

The SC versus IC read-outs are plotted in Figure 2.11 for both proton energies.



(a) 70 MeV proton energy.

(b) 11 MeV protons energy.

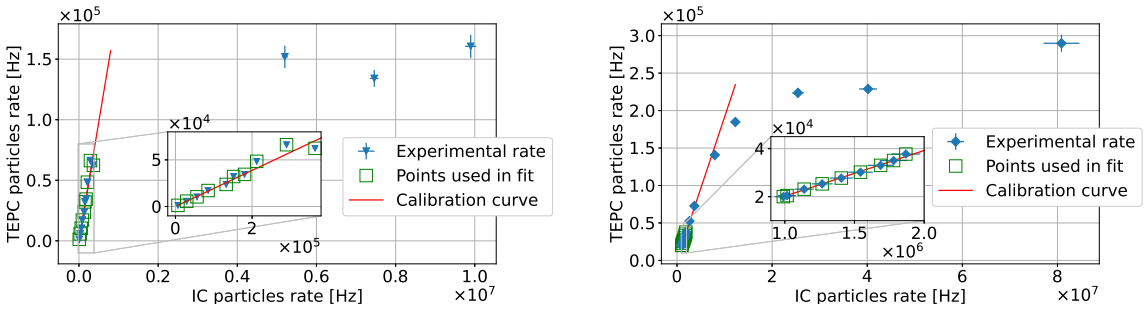
Figure 2.11: Particle rate measured with the scintillator (SC) versus counts acquired with the ionization chamber (IC) for the two proton energies of 70 and 11 MeV. The experimental points (blue markers) were fit with a linear function, represented by the red solid lines. While for the 70 MeV dataset, we used all available data for the fit, for the 11 MeV datasets, we only used those marked with a green square.

	m [Hz]	δm [Hz]	q [Hz]	δq [Hz]
11 MeV energy	1146	267	53901	9421
70 MeV energy	4076	29	2839	582

Table 2.3: Calibration coefficients obtained from linear regression on the rates obtained at the IC and the SC in Figure 2.11.

While the experimental dataset collected at 70 MeV exhibits a consistent linear behavior across the entire range, at 11 MeV, the IC begins to display a non-linear response when the SC rate drops below 4×10^5 Hz. We hypothesize that this deviation stems from the combination of two factors: i) slower protons release more energy in the IC, generating a larger signal, and ii) the absorber used to degrade the proton energy to 11 MeV causes the beam spot to broaden due to electromagnetic and nuclear scattering, leading more protons to be deflected at angles that are large enough to miss the IC. Although individual protons generate a higher signal at 11 MeV than at 70 MeV, this effect is out-weighted by the fact that fewer protons reach the IC at 11 MeV. As a result, the IC collects a lower signal at 11 MeV, approaching the non-linearity region close to the detection limit. For this reason, we decided to exclude the experimental points collected at 11 MeV with an SC rate $< 4 \times 10^5$ Hz from the calculation of the IC-SC calibration curve, discarding points with non-linear behavior. Using the function $y = mx + q$, we fit both the 11 MeV and 70 MeV datasets to obtain the linear curves plotted in Figure 2.11, and whose parameters are reported in Table 2.3.

The m coefficient, serving as a multiplicative factor applied to the SC rate, shows an increasing trend with proton energy. This behavior reflects the fact that low-energy protons deposit more energy, leading to a higher IC signal and, consequently a lower calibration slope. The intercepts of both curves are not consistent with zero within the error bar, suggesting that



(a) 70 MeV proton energy. (b) 11 MeV protons energy.

Figure 2.12: Relation between TEPC and calibrated ionization chamber (IC) particle rates measured for the two beam energies. A zoom of the low-rate region is shown for both graphs in the box. Points marked by the green square were fit with linear regression to obtain a calibration function (red line).

the SC collects a background signal when the beam is off. Although background subtraction is performed before each acquisition and is used to correct the SC output, the non-zero intercept indicates that this subtraction has limited accuracy and that an offset parameter in the fit must be included to enhance robustness.

Once the IC counts are converted into a rate with the parameters of Table 2.3, we can use this data to calibrate the TEPC and evaluate the rates when MAESTRO dead time exceeded $\sim 20\%$. The plot of the rate measured by the IC versus that estimated at the TEPC with MAESTRO is shown in Figure 2.12.

	m	δm	q	δq
11 MeV energy	0.0190	0.0004	1199	471
70 MeV energy	0.1963	0.0040	−48.5	26.9

Table 2.4: Calibration coefficients obtained from a linear regression of the rates obtained at the TEPC and at the IC in Figure 2.12.

At both proton energies, the rate measured by the TEPC is proportional to the IC below $\sim 5 \times 10^4$ Hz, while it saturates above. We selected all data in the proportionality region, and fitted them with the linear function $y = mx + q$, obtaining the calibration coefficients listed in Table 2.4. The m value obtained for the two beam energies differ by approximately one order of magnitude. This discrepancy indicates that a much larger number of particles will traverse the IC but not the TEPC at 11 MeV energy, because of lateral scattering along their trajectory. Also, in this case, the intercept of both curves is not zero, indicating that the background subtraction at the IC is not accurate.

Pileup estimation

We developed a stochastic algorithm to estimate the pileup probability starting from a microdosimetric spectrum. This algorithm relies on the GEANT4 Monte Carlo toolkit [Agostinelli et al., 2003] to replicate the microdosimetric spectra at various probabilities of pileup. We start by fixing the pileup probability $p \in [0, 1]$. Then, for each primary particle indexed as i, we generated a random number p_i from a uniform probability distribution in [0, 1]. If $p_i < p$, we marked the particle as *pileup-event* and summed its total energy deposition to the first non *pileup-event*. We reproduce the microdosimetric spectra running the previous algorithm over different values of pileup probability from 0.001 to 0.99. This procedure was adopted for both the 11 MeV and the 70 MeV proton energies. An example is shown in Figure 2.13 for the 70 MeV beam.

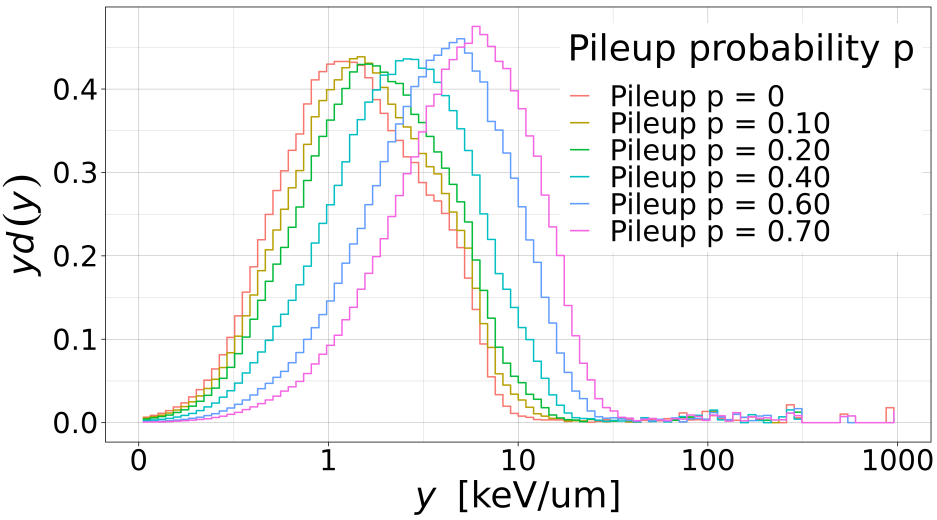


Figure 2.13: Microdosimetric $yd(y)$ spectra of 70 MeV protons with different pileup probability $p = 0,\ 0.10,\ 0.20,\ 0.40,\ 0.60,\ 0.70$. The spectra have been simulated with the GEANT4 Monte Carlo code.

As pileup increases, it becomes more likely that a recorded energy deposition in the spectrum did not result from a single event, but from the sum of two or more *pileup-events*. The direct effect is that higher energy depositions become more probable, and as a consequence, the microdosimetric spectrum shifts to larger y values.

The simulated microdosimetric spectra of varying pileup probabilities were used to evaluate the correct pileup on the experimental spectra. To measure the consistency between the simulated and experimental microdosimetric spectra, we utilized the Kolmogorov-Smirnov (K-S) test [Eadie et al., 1971]. This nonparametric test evaluates the likelihood that two probability density distributions satisfy the null hypothesis of being statistically equivalent. In our analysis, we considered a $\mathfrak{p}$ value above 5% to be considered statistically significant. We performed the K-S test for all simulated-experimental microdosimetric $d(y)$ spectra pairs. To analyze the data, we defined a matrix (SM) according to Equation 2.15, where the entry at position $(i,\ j)$ is the value of K-S test evaluated at the i-th experimental spectrum against the j-th simulated spectrum:

$$SM := \{D_{\text{K-S}}(d(y)_i,\ d(y)_j)\}_{ij}\ , i \in \{\text{experimental spectra}\},\ j \in \{\text{simulated spectra}\} \tag{2.15}$$

To enhance the test robustness, we focused on the portion of the spectrum within approximately ±30% of the mode, containing at least 80% of the spectrum counts. This region predominantly comprises primary particle events, whereas the tails of the spectrum are more likely to be populated by rare events. By setting a $\mathfrak{p}$ value threshold above 95%, it is possible that for a given experimental spectrum, several simulated spectra with different pileup probabilities are considered compatible according to the K-S test. In this case, we used the lowest and highest probabilities among the compatible simulations to establish the uncertainty interval for the pileup range associated with the specific experimental spectrum.

Figure 2.14 shows the how the K-S test has found a mathing between compatible experimental and simulated spectra. In particular, a pileup probability of 0.326 ± 0.0145 was found to be compatible with the experimental spectra at an estimated particles rate of 73.50 ± 5.99 kHz.

Using the K-S test, we could estimate the pileup single probability value, or range of values, associated with each experimental spectrum. Since each measurement was acquired

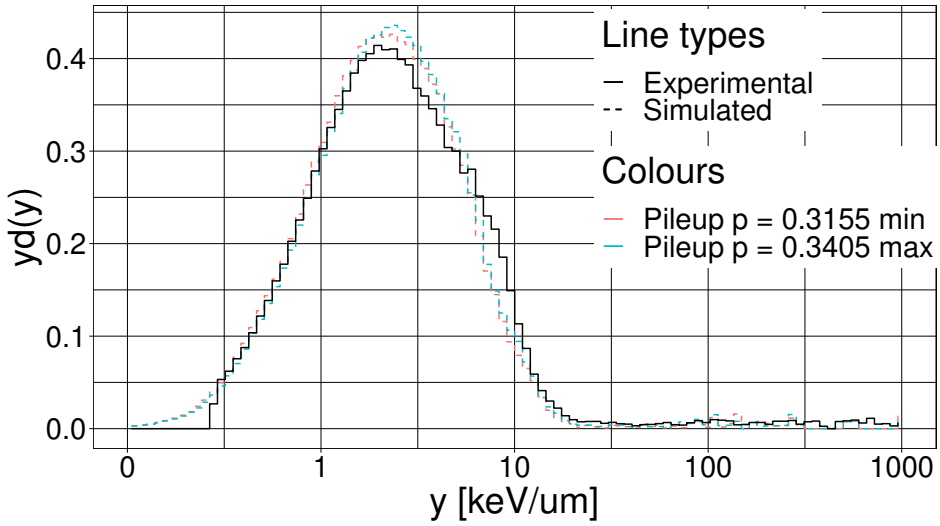


Figure 2.14: Result of the K-S test finding a range of simulated spectra with pileup probability [0.3155, 0.3455] with an experimental spectra aquired with an estimated particles rate of 73.50 ± 5.99 kHz.

with a specific particle rate, we built pileup-rate curves for both proton energies, reported in Figure 2.15.

The measurements can be accurately described by the Equation 2.16. The exponential of Equation 2.16 can be approximated to its linear term $\frac{x}{\mu}$ with a 1% uncertainty for rates below 136.0 kHz at 11 MeV, or 28.2 kHz for 70 MeV. As the beam intensity increases, both datasets deviate from the linear trend and eventually reach a saturation level. Equation 2.16 was used to fit the data, which is presented in Figure 2.15.

$$p(x) = a + b\left(1 - \exp\left[-\frac{x}{\mu}\right]\right) \tag{2.16}$$

The fit parameters are provided in Table 2.5.

11 MeV energy	0.022	0.018	0.75	0.02	348956	32846	1.1
	a	δa	b	δb	μ [Hz]	$\delta\mu$ [Hz]	χ^2_r
70 MeV energy	−0.005	0.007	0.81	0.01	136683	5724	0.64

Table 2.5: Parameters obtained from fitting the rate-pileup curves with Equation 2.16 for both proton energies. The curves are plotted in Figure 2.15.

Ideally, the zero rate point should correspond to a zero pileup. This is only true for the 70 MeV dataset, while the 11 MeV energy has a small offset of 0.022 ± 0.018. We hypothesize that this offset at low energy is due to the limited accuracy of GEANT4 in simulating the experimental setup. To correctly reproduce the shape and position of the measured 11 MeV microdosimetric spectrum, we added two extra 2 mm of RW3 in the simulation geometry. It is possible that although this correction resulted in an adequate reproduction of the microdosimetric spectrum, it still carried a discrepancy in the pileup evaluation. The 70 MeV curve consistently remains above the 11 MeV curve, indicating a higher occurrence of a pileup at 70 MeV for a given rate. This observation is supported by higher values of the μ coefficient, implying a smoother exponential curvature, indicating lower pileup effects (Table 2.5). At high particle rates, electronic signals generated by two distinct events will probably overlap.

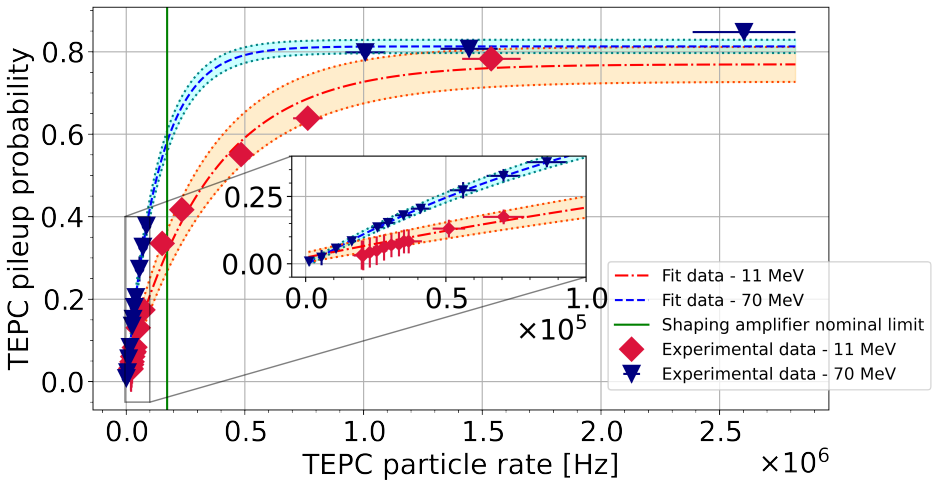


Figure 2.15: Particles rate at the TEPC versus the corresponding pileup assessed from the microdosimetric spectra for 11 MeV (red diamonds) and 70 MeV (blue triangles) energy protons. The dashed lines represent the fit calculated with Equation 2.16. The confidence intervals for both datasets are marked by the color regions delimited by the dotted lines. The shaping amplifier nominal limit of $2.9 \cdot t_{\text{shaping}}$ is shown as a vertical green line, and represents the threshold for the electronic pileup.

This is likely to occur on the element of the acquisition chain where signals are slowest (longest duration). In our acquisition chain, the slowest signal (longest duration) is produced after the shaping amplifier. The signal duration depends on the amplifier settings, which in this case were set to $2.9 \cdot t_{\text{shaping}} = 2.9 \cdot 2\,\mu\text{s} = (172\,\text{kHz})^{-1}$ with a shaping time of 2 µs. Thus, if two or more signals are separated by less than $(172\,\text{kHz})^{-1}$, they will overlap. This reference value, which is also depicted in Figure 2.15 and applicable to both energies, serves as a guideline for the occurrence of signal overlap. Finally, the reduced χ^2 in Table 2.5 for both energies suggests that the experimental errors are well estimated, and the overall trend is compatible with Equation 2.16 as $\chi^2_r \simeq 1$.

To further quantify the impact of pileup on microdosimetric quantities, we evaluated the experimental y_F values at different particle rates. We used the y_F measured without the influence of pileup as the reference "true value", and then calculated the percentage error on the y_F obtained at different pileup levels. The results are plotted in Figure 2.16.

The y_F values monotonically increase as the particle rate increases. Consequently, the percentage error also increases, reaching a maximum of 85% in the extreme case of pileup equal to 0.848 ± 0.008. The increase in the y_F value is a direct consequence of pileup as observed in Figure 2.13.

2.4.4 Discussion and conclusions

The acquisition of microdosimetric data at particle rates of clinical interest is always affected by pileup, regardless of the type of detector used. This effect distorts the spectra and their derived quantities, compromising the accuracy of the radiation field characterization. To overcome this issue, we have developed an innovative approach for evaluating pileup at different beam rates. In this work, we focused on two clinically relevant proton energies of 70 and 11 MeV, representative of the beam plateau and tumor region, respectively. The data indicate that pileup increases linearly with a rate up to approximately 28.2 kHz, and then starts to saturate (Figure 2.15). In addition, higher beam energies lead to a larger pileup value at a

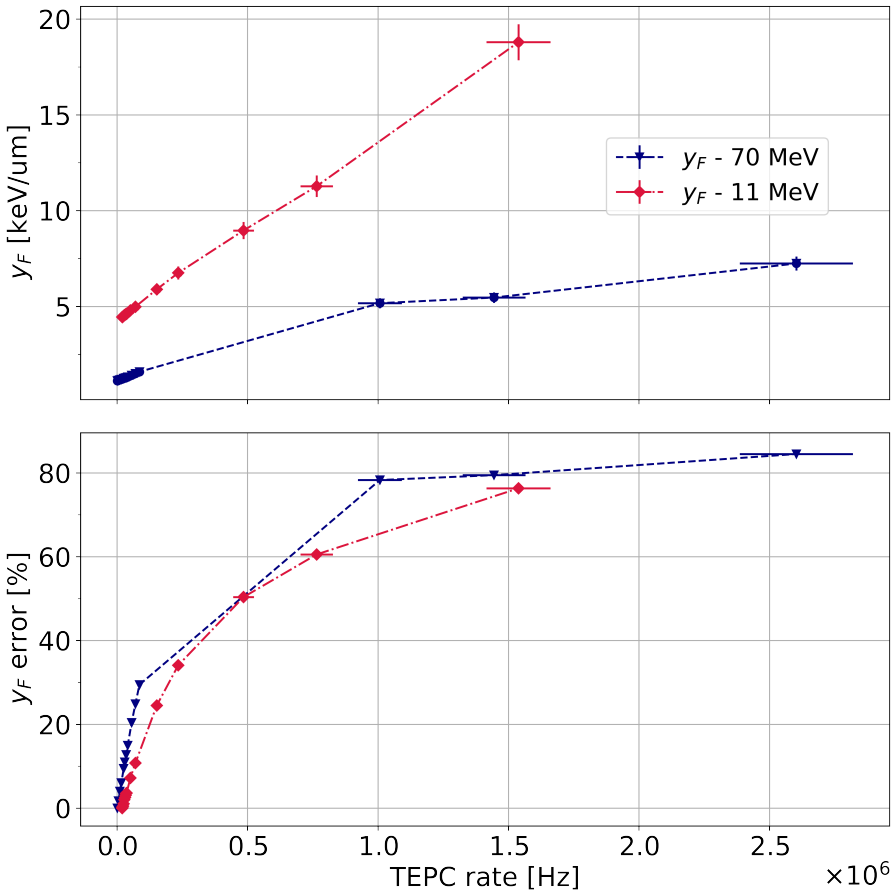


Figure 2.16: Upper plot: y_F values measured with the 70 MeV (blue triangles) and 11 MeV (red diamonds) proton energies. Lower plot: percentage errors on the y_F values calculated using the zero pileup y_F as a reference value. The dashed lines connecting the points are drawn to guide the reader's eye.

given particle rate. Although the curve parameters depend on the beam energy, the shape is identical for both datasets. For this reason, we hypothesize that all clinical protons will exhibit the same behavior. Therefore, based on this assumption, the measured curves can extrapolate the expected pileup for beams with energies close to those investigated. For energies significantly distant from those studied in this research or when dealing with different ion species, we cannot apply the pileup-rate curves of Figure 2.15 to predict experimental pileup. However, we can still utilize our methodology to evaluate the pileup level from the measured microdosimetry spectra in these cases.

With the methodology outlined in this paper, we estimated pileup-induced errors on the microdosimetric y_F. Our findings suggest that a rate of 10.9 ± 0.6 kHz corresponding to a pileup of 0.05 ± 0.02 results in an error of y_F below 5% for both proton energies. The curve of Figure 2.15 exhibits a linear region up to a pileup of 0.15 ± 0.03, corresponding to a y_F uncertainty of 11 %. If this degree of uncertainty is acceptable, then microdosimetry can be applied to measure radiation quality up to a beam rate of approximately 29.3 ± 1.3 kHz. Although this value is 2 orders of magnitude lower than the clinical range, it is more attainable for clinical facilities.

In conclusion, the method proposed in this work represents a new and generalized way

to estimate the pileup contribution to microdosimetric spectra, bringing the application of microdosimetry closer to clinical scenarios. Although we developed this methodology from experimental proton spectra measured with the TEPC, the algorithm can be applied to any microdosimeter, as we as for other ion species and energies.

Part II

Hybrid microdosimetry

Chapter 3

Hybrid Detector For Microdosimetry - HDM

3.1 Introduction

A limitation shared by all microdosimeters is that while ϵ is directly measured, the value of $\bar{l}$ has to be theoretically estimated as the mean path travelled by a particle inside the active area of the detector, and thus depends on the detector geometry. In addition, $\bar{l}$ values calculated for standard geometries can be used only if the microdosimeter is exposed to a homogeneous and isotropic field (also called uniform isotropic randomness) [Kellerer et al., 1985] and a different $\bar{l}$ value will be obtained under different irradiation conditions, i.e. for other types of randomness. Some attempts have been made to overcome the limited accuracy of the the mean chord length concept, e.g. in unidirectional particle field when the isotropicity assumption drops. An example is SOI microdosimeters [Bolst et al., 2017], which are composed of 3D sensitive volumes (SVs) arrays with a well defined thickness, and thus path length. Furthermore, few theoretical studies focused on finding a formula of the mean path length for both uniform [Cruz et al., 2001] and non uniform [Santa Cruz et al., 2001] radiation fields. So far, only the calculation for a uniform isotropic randomness could be successfully applied to experimental methodologies. Estimating the path length l is a critical parameter in microdosimetry that will influence the accuracy of the radiation field quality characterization [Abolfath et al., 2020]. In fact, for a given energy ϵ deposited in the detector, the resulting y value can assume a wide range of values depending on the l. For example, if ϵ=10 keV in a 2 µm diameter sphere made of tissue, y can varies from 5 keV/µm to 100 keV/µm just considering l values ranging from the sphere diameter to abritrary small l values, 0.1 µm in the case of this example.

For this reason, since the quantity y is traditionally intended as the ϵ over the mean chord length value $\bar{l}$, we introduce a new quantity y_T, defined as ϵ divided by the particle real track length l.

Hybrid Detector for Microdosimetry (HDM) is indeed designed to measure the y_T. This detector have been specifically intended for particle therapy application, where a knowledge of the y_T yields a more direct link to the biological damage.
HDM is composed of a spherical TEPC followed by four layers of Low Gain Avalanche Detectors (LGADs) [Pellegrini et al., 2014]. LGAD is a recent technology in silicon systems featuring detection of particles in a wide energy range with improved accuracy for timing and tracking measurements [Pellegrini et al., 2014, Sadrozinski et al., 2016, Vignati et al., 2022]. The LGAD application in particle therapy has been also recently investigated [Vignati et al., 2017, Vignati et al., 2022, Vignati et al., 2020, Sacchi et al., 2020, Sacchi et al., 2021]. In the proposed setup, the TEPC will provide the energy deposition ϵ directly in a tissue-equivalent

medium while the LGADs will offer information about particle spatial distribution with a precision of about 200 or 300 µm, depending on the chosen configuration.

Details of the detector components, geometrical configurations as well as read-out solutions are illustrated here. Using GEANT4 toolkit [Agostinelli et al., 2003, Allison et al., 2006, Allison et al., 2016], we investigated HDM performances when exposed to protons and carbon ions in the therapeutic energy range. The influence on all microdosimetric quantities when the real l is used instead of the mean track length approximation is discussed. Detection efficiency and tracking precision are also reported.

3.2 HDM feasibility study

Together with the TEPC that provides energy depositions, LGADs constitute the tracker stage of HDM. An introduction on these silicon detectors is provided here.

3.2.1 Low Gain Avalanche Detector

Low Gain Avalanche Detector, also called Low Gain Avalanche Diode (LGAD), is an innovative and rapidly evolving technology in silicon detection system. It was first fabricated at Institute of Microelectronics of Barcelona (CNM-IMB) [IMB,] clean room facilities [Pellegrini et al., 2014]. From then, it has been used for particle timing and tracking and, more recently, its application in radiotherapy has been explored [Vignati et al., 2017, Vignati et al., 2022, Vignati et al., 2020, Sacchi et al., 2020, Sacchi et al., 2021]. Unlike standard PIN diodes [Hazdra et al., 2018], the defining feature of LGADs is an additional p^+ region. Figure 3.1 shows a comparison between a traditional PIN and an LGAD where it is possible to observe the extra p^+ labeled in the picture as "p^+ gain layer". The addition

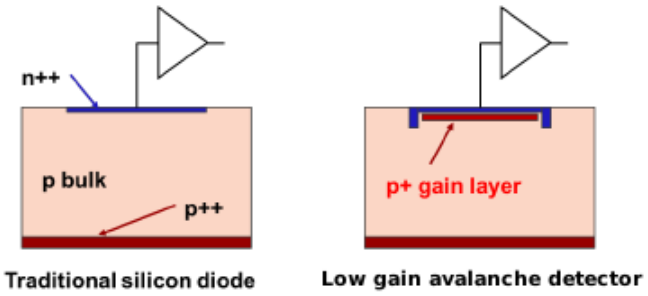


Figure 3.1: Section view of a standard PIN (left) and a Low Gain Avalanche Detector (right). Picture from [Ferrero et al., 2019].

of this extra p^+ layers add to the electric field profile a region featuring a high gradient. Figure 3.2 shows a representation not in scale of the electric field profile inside the LGADs. Taking advantage of this property, the increased gradient of the electric field can induce a controlled charge multiplication (at least $\sim 300\,\mathrm{kV/cm}$ [Sola et al., 2019b] are required), thus amplifying the magnitude of the initially deposited charge. Like PIN diodes, LGADs require reverse polarization. In this it is possible to create a depletion region (named "p-bulk" in Figure 3.2). If sufficient energy, typically exceeding 3.6 eV, is deposited into the p-bulk region, an electron-hole pair can be generated. The 3.6 eV energy threshold is crucial as it represents the average amount of energy required to produce an electron-hole pair in silicon [S.M. Sze, 2006]. After the creation of the electron-hole pair, due to the electric field, the electron will drift towards the top of Figure 3.2 while the hole will drift towards the bottom. Upon reaching the high-gradient electric field, the electron undergoes a controlled multiplication effect, resulting

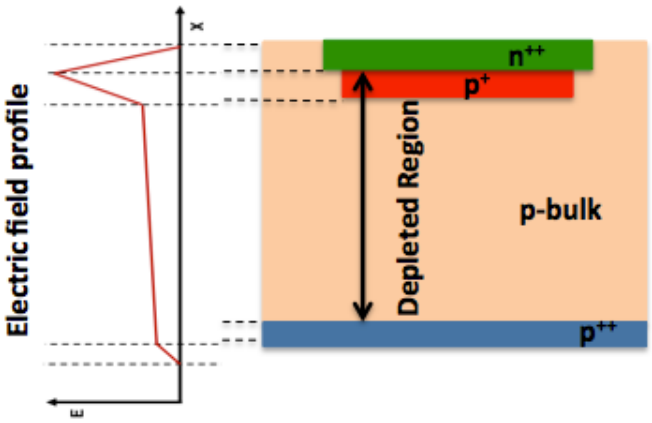


Figure 3.2: Section view of an LGAD with its electric field profile schematized. Picture from [Ferrero et al., 2017].

in the generation of multiple electron-hole pairs. As a consequence, the originally generated electron-hole pair can be amplified, depending on the gradient and the applied inverse polarization voltage, up to a factor ~ 100 [Sola et al., 2019c]. With this internal detector design, LGADs offer impressive features, including exceptional time resolution (< 40 ps), rapid signal duration (approximately a few nanoseconds), and outstanding radioresistance, capable of withstanding up to 2.5×10^{15} n_{eq} cm^{-2} [Marti Villarreal et al., 2023, Bisht et al., 2022]. These attributes make LGADs the foundational choice for the timing systems employed in the ATLAS [Casado et al., 2022] and CMS [Cms, 2019] experiments. Achieving these features relies significantly on the detector geometry, with particular emphasis on optimizing the LGAD thickness to align with timing, capacitance, and practical fabrication requirements. This optimization typically results in a thickness of the active area of approximately 50 µm, as reported in [Giacomini, 2023]. To attain this desired active area thickness, it becomes necessary to employ a carrier wafer, often referred to as a substrate, with a thickness on the order of a few hundred microns. This step is essential for making the manufacturing process viable.

The main features of the LGADs used for the HDM prototype can be found in [Sola et al., 2019a]. In particular, the active region is 50 µm thick, while the substrate is 300 µm and can be thinned down to 20 µm postproduction (70 µm of total thickness). A constraint on the detector geometry is that the optimal active area of one strip is ~ 2 mm^2, which corresponds to 5 pF of capacitance; in fact, the read-out chip has been designed accordingly. Two LGAD geometric configurations are available: one with 71 strips and another with 34 strips. For what concerns the 71-strips configuration, the area is already optimized, while the 34-strips configuration has a larger area. However, previous experiments with similar area have shown the feasibility of this capacitance also.

Furthermore, the dead area between two strips must be 66 µm wide independently of the strip width. Thus, narrower strips result into a higher spatial resolution but also a decreased detection efficiency due to a larger dead area and a resulting lower fill factor. In addition, to cover the same area more strips are needed, which translates into a larger number of channels to be read-out.

To find the optimal detector geometry for our application, we simulated three configurations: i) 34 strips, each 294 µm wide and 12.5 mm high (sensor height 13.8 mm and width 13.4 mm); ii) 71 strips, each 114 µm wide and 12.5 mm high (sensor width 14 mm and height 13.8 mm) and iii) 288 strips, each 114 µm wide and 50.22 mm high (sensor height 51.52 mm and width 51.84 mm). An image of the design project of this configuration of the complete sensor

is given in Figure 3.3 (left panel). While the first two configurations are now being produced, configuration (iii) is not currently feasible and was tested to investigate the tracking efficiency for a larger detector with the same spatial resolution of the 71 strips detector (ii).

3.2.2 HMD geometry

The LGAD position with respect to the TEPC determine the detector performances and the optimal configuration depends on the goal of the specific measurement. In this work, we investigated the configuration with the TEPC upstream of the 4 LGAD layers. This setup has been chosen because we wanted to characterize the radiation field with standard microdosimetric measurements, without possible artifacts due to the LGADs in front. The distances between the detectors can be found in Figure 3.3 (right panel). In particular, the first LGAD have been placed as close as possible to the TEPC to minimize lateral scattering and energy loss of particles exiting the microdosimeter.

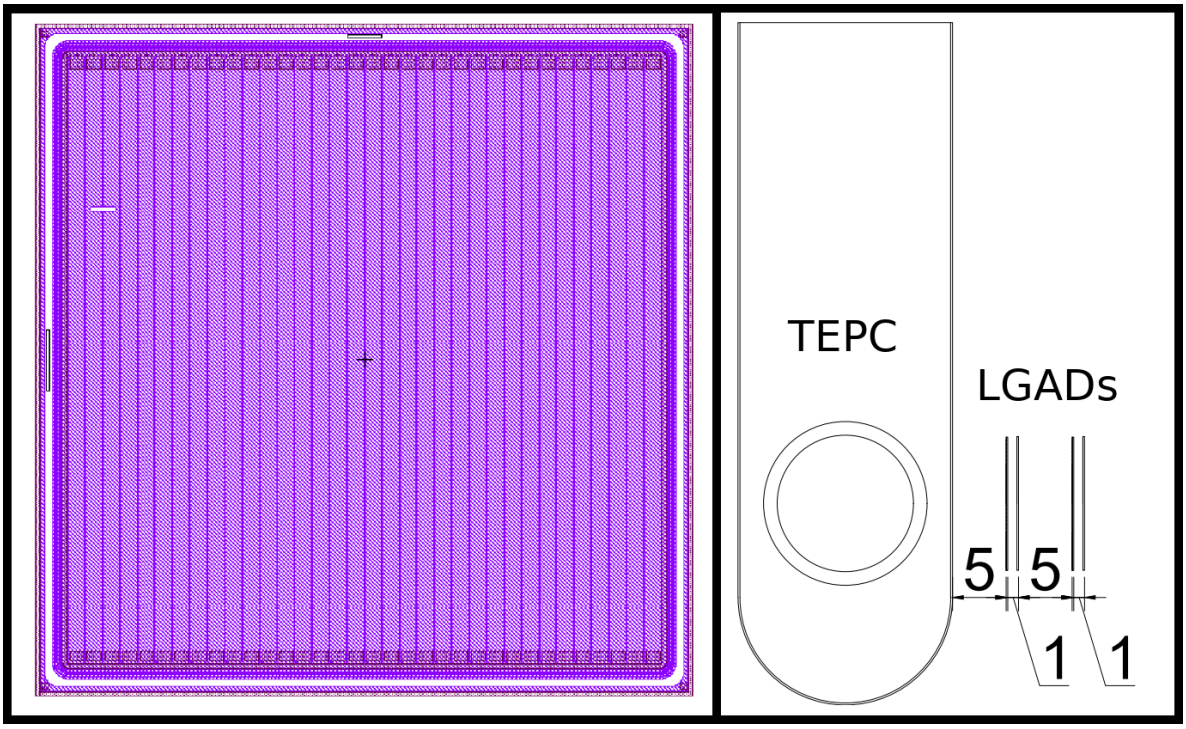


Figure 3.3: Left panel: design of one LGAD sensor with 34 active strips. Right panel: Scheme of the HDM setup, showing the TEPC followed by four LGAD layers. Distances between detectors are reported in millimeters.

3.2.3 Geant4 simulations of HDM

To investigate the detector performances, we run Monte Carlo calculations using Geant4 toolkit. As the HDM design is optimized for applications in particle therapy, we focused the study on the response to protons and carbon ions at therapeutic energies. All calculations were run to acquired a minimum of 10^6 events on the TEPC, which is considered an adequate statistics for experimental measurements [Missiaggia et al., 2020].

In addition, since microdosimetry deals with patterns of single energy deposition in tissue at the micrometer scale, we computed the energy deposition ϵ of a particle traversing the

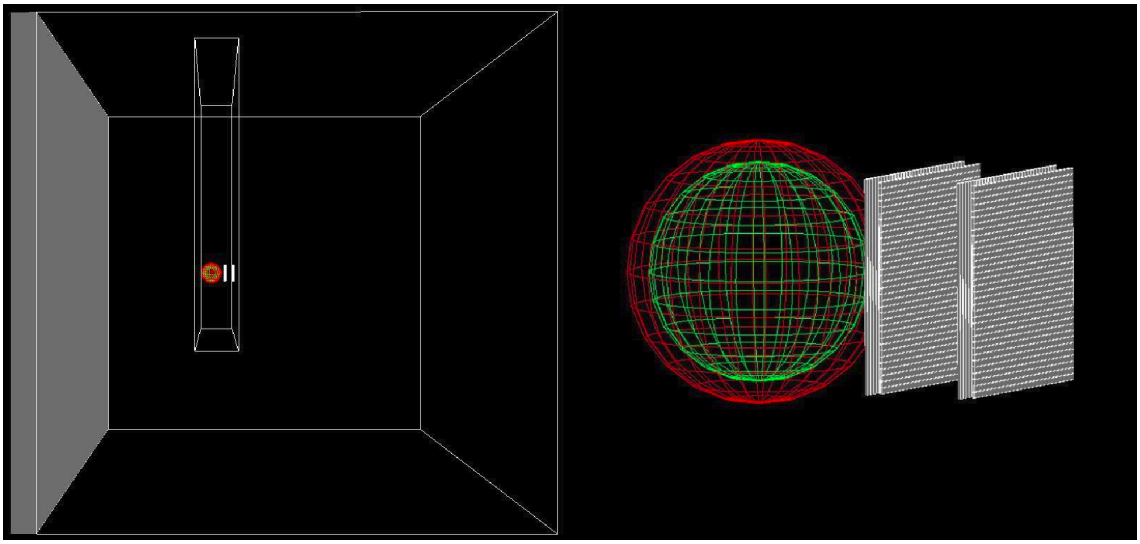


Figure 3.4: 3D scheme of the geometry used for all Geant4 simulations. Both the TEPC and the four 24-strips LGADs are contained in PMMA box filled with air. The box is placed inside a water phantom, whose walls are made of PMMA. A broader view is show in left panel, while a zoom on HDM is illustrated in right panel.

TEPC as the sum of the energy deposited by the primary event and all the related secondary particles that entered the detector.

The simulation geometry consisted of a water phantom with PMMA walls (1.74 cm water equivalent thickness) where the hybrid system was placed. To reproduce a realistic setup, HDM was contained in an additional air box 2.8×20×2.8 cm thick. A 3D view of the setup is shown in Figure 3.4.

The water phantom was irradiated with 290 MeV/u carbon ions and 150 MeV protons, which have the same range in water ($\sim$ 160 mm). The beam spots were circular with a 3 cm radius to ensure that the detectors were fully immersed in a homogeneous and isotropic radiation field. The detector box was placed at 10.74 cm in water along the beam direction. This depth represented a good compromise to assess HDM performances in a relatively mixed field in terms of particle species and energies, but upstream of the Bragg peak, where most particles have a low energy and thus might stop inside the TEPC.

3.2.4 Tracking

Tracking algorithm

To measure a particle track, the LGADs were positioned to have the strips in different directions, two horizontals (x plane) and two verticals (y plane). By coupling two sensors with different orientations, a spatial position for a particle can be measured. Thus, two pairs of sensors are the minimum requirement for reconstructing a particle track. To reproduce a realistic experimental scenario, in the simulation we scored only the position of the strip hit by the particle. Then, we used a lineal interpolation to reconstruct the particle path inside the TEPC, from which we could estimate the real track length.

Tracking efficiency

Using Geant4 simulations, we studied the HDM tracking efficiency. As a first step, we focused on identifying the lost events and divided them into three categories:

1. particles that reach all the detectors, but traverse an inter-strip dead zone in at least one of the LGADs;

2. particles that range out before reaching the fourth LGAD;

3. particles that undergo lateral scattering and are deflected outside the solid angle covered by all detectors.

Category 1 is related to the probability to hit a dead region and thus depends on the LGAD geometry. Assuming a uniform radiation field, the probability to reach an active strip is given by A_{act}/A_{tot}, where A_{act} is the total area covered by active strips and A_{tot} the total area of the sensor, including both active strips and dead inter-strips. As the probabilities of hitting the active region of two sensors are independent, the overall probability of the joint event is the product of the single probabilities. To test the validity of these assumptions, in the simulation we also scored the particles traversing the inter-strip regions.

For category 2, we investigated the minimum detectable kinetic energy for each ion type, i.e. the minimum energy that a particle must have to pass through all detectors. The values for all particle species of interest have been estimated with LISE++ toolkit version 10.0.6a [Tarasov and Bazin, 2008]. These kinetic energy cutoffs depends only electromagnetic interactions in the detector layers and do not take into account additional losses due to multiple Coulomb scattering (MCS). To estimate a realistic kinetic energy detection threshold, we performed simulations of HDM exposed to a given particle species and decreases the initial energy until we found the minimum value required to traverse all detectors. We then repeated the test for the ion types of most interest.

The percentage of particles deflected outside the solid angle covered by all detectors (category 3) dependent on the LGADs size. To assess this value and its dependence on the LAGDs geometry, we performed simulations for every configuration described in subsection 3.2.1.

For the events seen by the TEPC and by an active zone of each of the 4 silicon layers (i.e. the trackable particles), we investigated the tracking accuracy using the algorithm described in section 3.2.4. From the simulations, we could extract the real particle track and compare it to that reconstructed with the tracking algorithm, estimating a mean discrepancy between the predicted and actual values. The tests were repeated for all LGADs configurations taken into consideration.

3.2.5 Results

Radiation field characterization in the TEPC

The composition of the radiation field entering the TEPC was investigated at a depth of 10.74 cm in beam. The results include kinetic energy spectra of all particle species, track length distributions and microdosimetric spectra $yd(y)$ obtained with both the real track length and the mean chord length. The results are shown in Figure 3.5 and Figure 3.7 for protons and carbon ions, respectively. In detail: panels **A, B** of Figure 3.5 and panels **A, B, C, D** of Figure 3.7 illustrate the kinetic energy distributions of all particles entering the TEPC, with and without the contribution from the primary ions (in both cases the energy distributions of the single components are normalized to one); the track distributions of all the particles are plotted in panel **C** for protons and in panel **E** for carbons, with the mean chord length of 8.47 mm marked with a dashed red line; panels **D** for protons and **F** for carbons contain a comparison between the microdosimetric spectra calculated with the mean chord length approximation ($yd(y)$) or the real track length ($y_T d(y_T)$). Furthermore, the mean values and standard deviations of the track length distributions are also reported in Table 3.1 for both ions of interest.

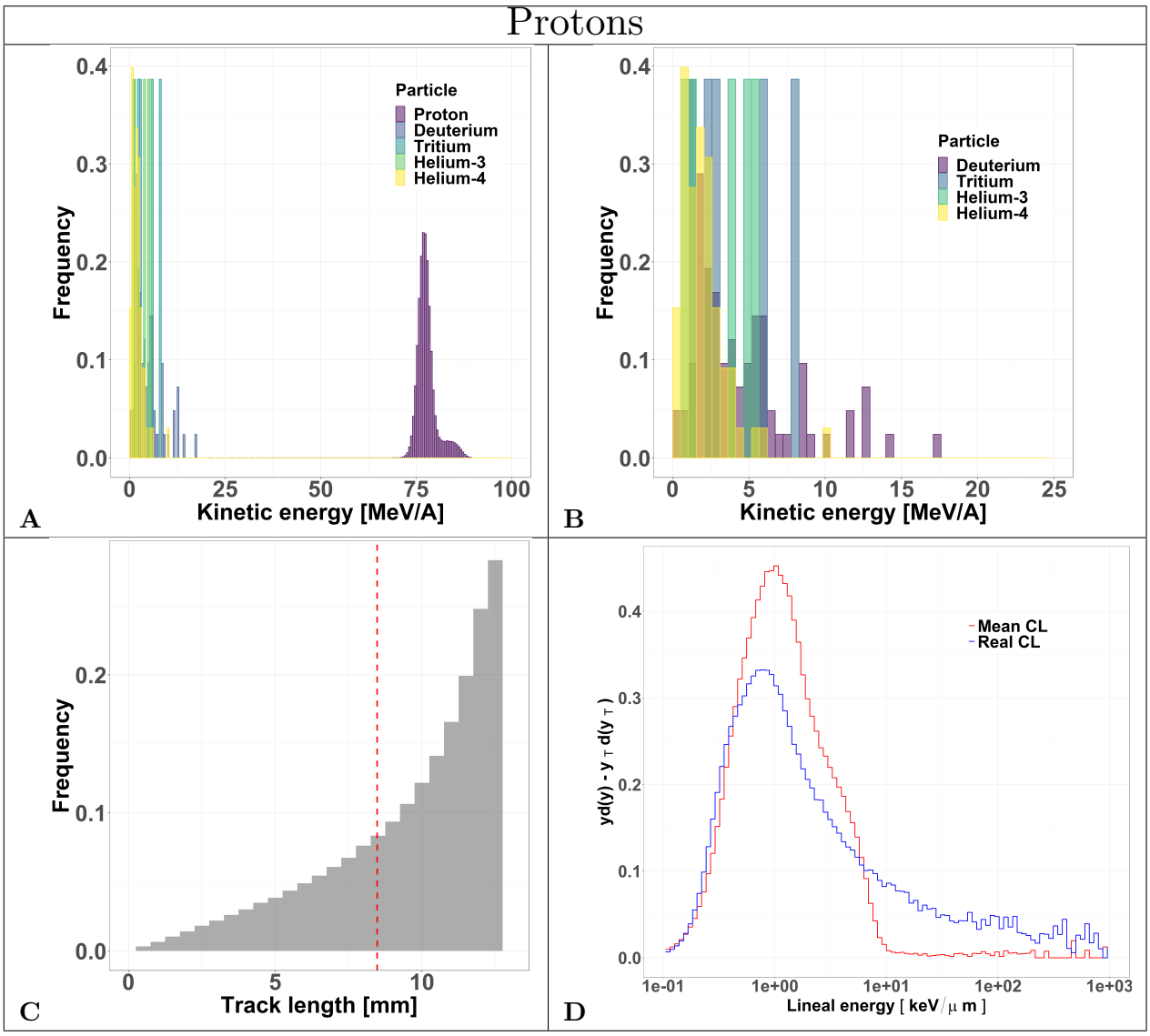


Figure 3.5: Characterization of the radiation field generated by 150 MeV protons after traversing 10.74 cm of water and seen by the TEPC. Panels **A** and **B**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **C**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **D**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line) .

Secondaries produced by protons, are mostly low-energy (below 10 MeV) and the distribution does not have a peak. For carbon ions, the energy of all fragments species peaks around 170 MeV/u, which is the residual primary beam energy (Figure 3.5, panel **A**). Protons can only generate fragments from the target nuclei, and thus their energy will be relatively low [Tommasino and Durante, 2015]. Carbon ions, instead, can produce both projectile and target fragments, whose kinetic energies have a much wider range, peaking at the same value as the primary ions [Mohamad et al., 2018, Tommasino et al., 2015].

The track length distributions of both protons and carbon ions are very broad and do not present a peak. Furthermore, the mean track length calculated for both protons and carbon

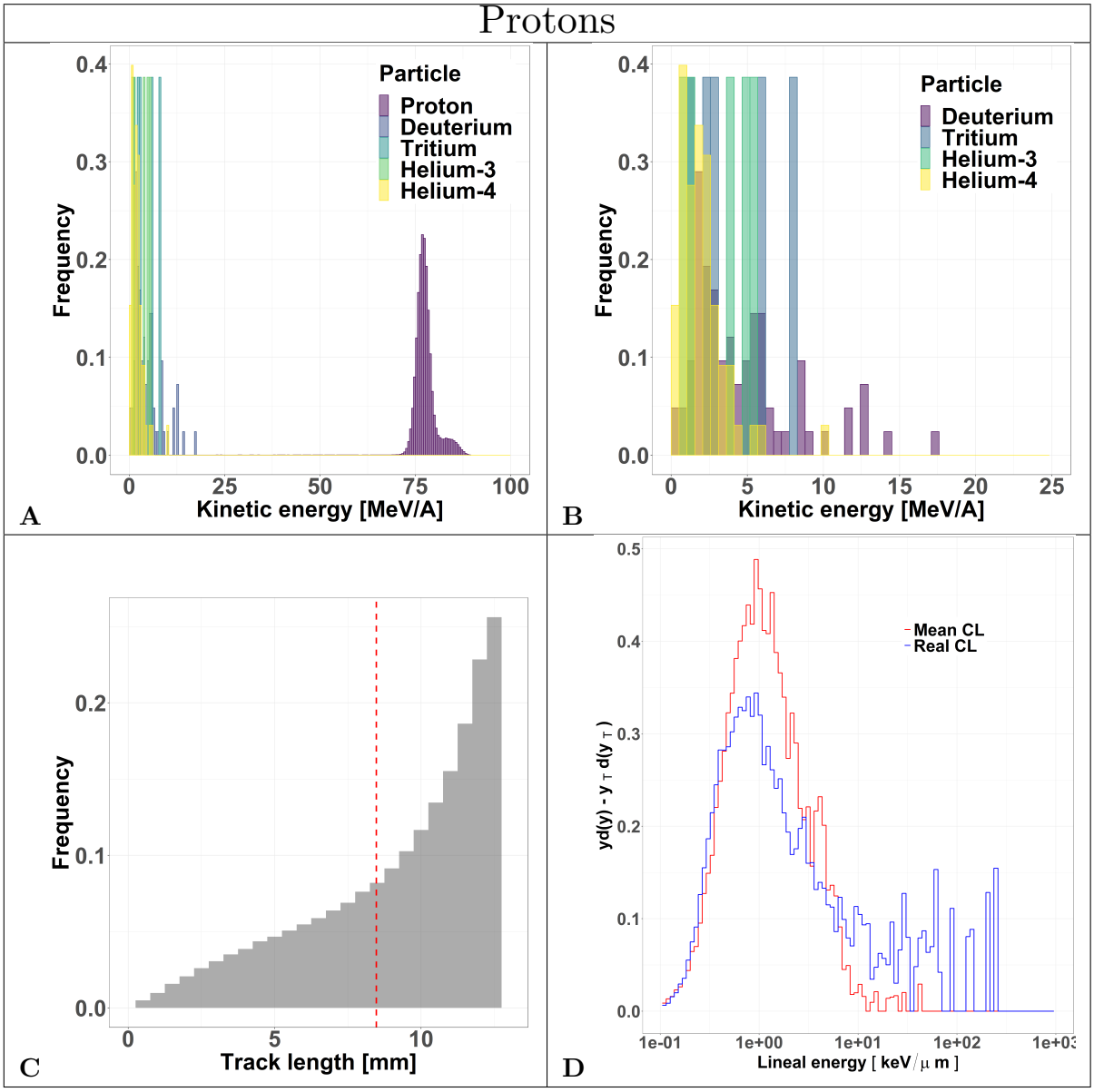


Figure 3.6: Characterization of the particles lost by HDM when irradiated with 150 MeV protons at a depth or 10.74 cm in water. Panels **A** and **B**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **C**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **D**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line).

ions is higher than the mean chord length, indicating that the latter does not provide an accurate description of the system. The limitation of the mean chord length approximation can be further investigated by comparing the standard microdosimetric $yd(y)$ spectra with those obtained with the real track length ($y_T d(y_T)$). The latter distributions show a non negligible contribution in the high y_T region. Those contributions are due to events that deposit energy along a small chord length and they are underestimated in the $yd(y)$ spectra where the mean chord length value is used. These events have a very high y_T and thus can be relevant for radiobiological effects, especially for high y-values.

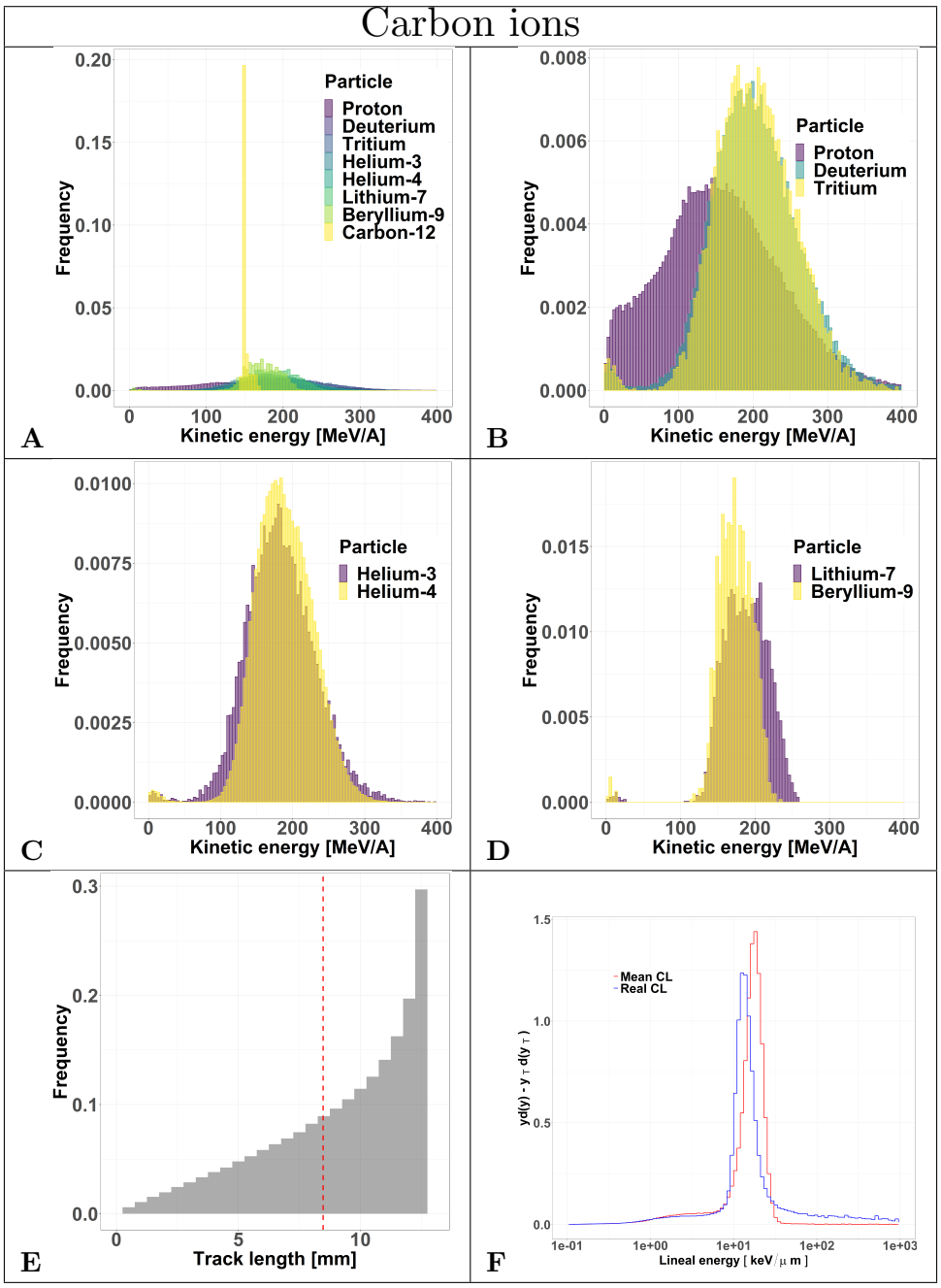


Figure 3.7: Characterization of the radiation field generated by 290 MeV/u carbon ions after traversing 10.74 cm of water and seen by the TEPC. Panels **A** to **D**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **E**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **F**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line).

Particles tracked by HDM

We investigated HDM tracking efficiency as well as the characteristics of the tracked events. Table 3.2 illustrates for carbon ions and protons the percentage of particles tracked by HDM, their mean track length values, their standard deviations and the average discrepancy between

Ion	**Mean track length [mm]**	**Standard deviation [mm]**
Carbon	9.17	3.03
Proton	9.53	2.8

Table 3.1: Mean track length values and standard deviations of protons and carbon ions traversing a spherical TEPC of 12.6 mm diameter. The mean chord length of this detector exposed to a uniform and isotropic radiation field is 8.47 mm.

Ion	**Configuration**	**Tracked particles [%]**	**Mean track length of tracked particles [mm]**		**Standard deviation [mm]**		**Mean absolute tracking error [mm]**
			Real	**Reconstructed**	**Real**	**Reconstructed**	
Carbon	*34 strips*	31.4	10.10	10.09	2.42	2.43	0.38
	71 strips	12.1	9.99	10.00	2.53	2.52	0.20
	288 strips	14.6	9.53	9.55	2.81	2.79	0.25
Proton	*34 strips*	45.8	10.06	9.89	2.37	2.50	0.91
	71 strips	15.3	9.91	9.91	2.46	2.47	0.24
	288 strips	16.6	9.63	9.64	2.68	2.67	0.28

Table 3.2: Percentage of particles tracked by HDM, including their mean track length, standard deviations and the absolute values of the mean tracking error of the algorithm with respect to the actual value. The results are reported for both protons and carbon ions and for three LGAD configurations (34, 71 and 288 strips).

the reconstructed and the real track length. The latter values are reported for the three sensor geometries (34, 71 and 288 strips) described in subsection 3.2.1.

The results show that, as expected, the 71 strips configuration collects the least amount of events because of the reduced fill factor. Increasing the sensor dimension while keeping the same fill factor increases the number of collected events (288 strips configuration). The mean track length and standard deviation obtained with the tracking algorithm are in good agreement with the real values obtained directly from the simulation. This is confirmed also by the small values of the mean absolute error, defined as the average absolute value of the difference between the real track length and the reconstructed one.

The accuracy of the reconstructed tracks in the three sensor configurations (34, 71 and 288 strips) was further studied in Figure 3.8 and Figure 3.9 for protons and carbon ions, respectively. We compared the track length distribution obtained directly from Geant4 with that reconstructed with the algorithm. The data are presented as density color plots in panels **A, C** and **E**; the green dotted line marks a perfect prediction of the algorithm, the red and blue colors represent regions of high and low events density, respectively. The distributions have a cone-like shape, implying a better accuracy of the reconstructed tracks of large lengths. This result is further supported by the presence of high density regions around the green line in the large track lengths zones.

To further assess the accuracy of the tracking algorithm, in panels **B, D** and **F** we com-

pared the track distributions of all particles traversing the TEPC with those detected by HDM and either obtained directly from the simulation or estimated with the tracking algorithm. Independently of the primary ion type, the 34 and 71 strips configurations systematically underestimate the distributions for small tracks. On the contrary, the 288 strips configurations provide a more accurate estimation of the whole track distributions, especially for protons.

The track distributions obtained with the three configurations were used to calculate microdosimetric $yd(y)$ and $y_Td(y_T)$ spectra for all particles tracked by HDM. The results are shown in Figure 3.10 and Figure 3.11 for protons and carbon ions respectively. Results show that the $yd(y)$ spectra differ from the $y_Td(y_T)$ ones, with a peak value shifted to the right in all cases. On the contrary, the $y_Td(y_T)$ distributions obtained with the real track length and with the reconstructed track length are similar mostly in the bell shape regions. It can be nonetheless seen that they differ in the tails due to the higher discrepancy between the real track length and the reconstructed ones for small track lengths. The accuracy between the two increases from the first sensor configuration (panels **A**) to the last one (panels **C**) under both radiation fields.

Particles lost by HDM

As discussed in section 3.2.4, we can group lost particles into three categories: i) particles with a kinetic energy under the minimum required to traverse all the detectors, ii) particles lost due to MCS and iii) particles that reach all detectors, but cross an inter-strip in at least one LGAD.

The minimum kinetic energies necessary to pass all detectors have been studied and are reported in Table 3.3 for all particles of interest. The values calculated with LISE++ are indicated for all particles while those obtained with Geant4 only for selected ions representative of the radiation field. The results obtained with the two methodologies agree very well for protons but have a a higher discrepancy for carbon ions.

Using Geant4 outputs, we characterized the particles lost in terms of kinetic energy when entering the TEPC and track length traversed inside the detector. The results are reported in Figure 3.6 and Figure 3.12 for protons and carbon ions, respectively. In panels **A** and **B** of Figure 3.6 and **A**, **B**, **C** and **D** of Figure 3.12 the kinetic energy spectra of all particle types are plotted with and without the contribution from the primaries. Independently of the fragment type, the energy spectra have the same shape of those reported in Figure 3.5 and Figure 3.7, where all events are considered. These result indicate that the probability for a particle to be lost is independent of the charge and energy (for energies above the minimum threshold reported in Table 3.3). Panels **C** of Figure 3.6 and **E** of Figure 3.12 illustrate the track distributions of lost particles, together with the mean chord length (red dotted line). The left side of the distribution appears to be more populated compared to the distribution of all events (Figure 3.5 and Figure 3.7), suggesting that there is a higher chance of loosing a particle if it has a small track length. Such events, in fact, traverse the TEPC edges and geometrically have a larger probability of missing the sensors, considering also MCS effects. In panels **D** for protons and **F** for carbons, the microdosimetric $yd(y)$ and $y_Td(y_T)$ spectra of particles that are not tracked by HDM are shown. Similarly to panels **D** of Figure 3.5 and **F** of Figure 3.7, where all particles are taken into account, the $y_Td(y_T)$ distribution peaks are shifted to the left for both protons and carbon ions radiation fields. Further, the high-y regions are significantly lower than the high-y_T regions; again, this is due to the real track lengths overestimation caused by using the mean chord length value.

Furthermore, the number of particles that reach at least one of the inter-strip passive regions with respect to the total number of events reaching the detectors (i.e. traversing either an active strip or an inter-strip region) has been estimated to be 63% for the 34 strips

	Proton [MeV/A]	**Deuterium** [MeV/A]	**Tritium** [MeV/A]	**Helium-3** [MeV/A]	**Helium-4** [MeV/A]	**Lithium-7** [MeV/A]	**Beryllium-9** [MeV/A]	**Boron-11** [MeV/A]	**Carbon-12** [MeV/A]
LISE++	17	11	8	20	17	20	24	28	34
GEANT4	17	12	9	-	17	-	-	-	37

Table 3.3: Minimum kinetic energies for several isotope types necessary to traverse all the detectors. The values have been calculated with LISE++ toolkit and, for the most representative of the radiation field, also with Geant4.

	y_D	$y_{D_T,TEPC}$	$y_{D_T,TRACKED}$
Carbon	16.53	40.49	40.79
Proton	3.75	23.73	21.08

Table 3.4: y_D and y_{D_T} values evaluated for both protons and carbon ions. y_{D_T} values have been calculated for all the particles that traverse the TEPC ($y_{D_T,TEPC}$) and for particles that HDM is able to track $y_{D_T,TRACKED}$ with the 288 strips configuration.

configuration and 81.5% for the 71 strips configuration. Increasing the number of strips in each sensor results in a significant improvement of the detection efficiency.

Finally, to asses the HDM capability to provide an accurate microdosimetric characterization of the radiation field, the following quantities have been calculated: standard y_D values considering all events traversing the TEPC; $y_{D_T,TEPC}$ values calculated from the $y_T d(y_T)$ spectra for all particles traversing the TEPC; $y_{D_T,TRACKED}$ values calculated from the $y_T d(y_T)$ spectra for all particles tracked by HDM. The results for both ions of interested are reported in Table 3.4.

The large discrepancy between y_D and $y_{D_T,TEPC}$ confirm the results shown in Figure 3.5, Figure 3.7, Figure 3.10 and Figure 3.11) and proves that the mean chord length approximation applied to the TEPC spectra does not provide an accurate description of the radiation field quality. The $y_{D_T,TRACKED}$, instead, is very close to $y_{D_T,TEPC}$, suggesting that the population of events tracked by HDM is representative of the actual field.

3.2.6 Discussion

An innovative design for a hybrid microdosimeter (HDM: hybrid detector for microdosimetry) is presented in this section. HDM is a two-stage detector composed by a TEPC and four layers of LGAD sensors. The combination of two different types of sensors (gas- and silicon-based) results in detection performances not offered by any existing microdosimeter. In fact, the TEPC gives a direct measurement of energy deposition in tissue while the LGADs provides particle tracking. The latter information has two main advantages: it improves the TEPC lateral spatial resolution to submillimetric precision and offer the real track length traverse by each particle in the TEPC. An improved lateral spatial resolution is helpful especially in a non homogeneous field, as the beam edges (both lateral and near the end-of-range). In these regions, the TEPC is only partially traversed by primary particles, whose path length might substantially deviate from the mean chord, depending on the detector position.

To assess the detector capability, we performed Monte Carlo simulations using Geant4 toolkit. As the primary application of HDM is particle therapy, we investigated its perfor-

mances exposed to protons and carbon ions at a certain water depth.

The limitations of the mean chord length for our geometry are evident by looking at the track length distributions of all particles traversing the TEPC (Figure 3.5, Figure 3.7 and Table 3.1). This approximation is based on the specific assumption that the TEPC is exposed to a uniform isotropic radiation field. In the cases considered here, although the beam generates such type of randomness, the water surrounding the TEPC causes the isotropy assumption to drop, with a direct consequence on the resulting mean track length. To further validate this, simulations without the water phantom has been performed and a mean track length value of 8.56 mm has been obtained for protons and 8.45 for carbon ions, both in accordance to the nominal mean chord length value.

However, even if a mean value of chord length based on more appropriate kind of randomness is used, the data reveal that a mean value is non representative of the whole track length distribution, since the standard deviations are rather large. This behavior is noticeable by the broadness of the track distributions in panels **C** of the Figure 3.5 and **E** of Figure 3.7.

Discrepancies between the mean chord and the real track length translate into difference between the standard $yd(y)$ and the alternative $y_T d(y_T)$ microdosimetry spectra (Figure 3.5, Figure 3.7), the more evident being in the high y_T regions. The majority of particles populating these areas have a track length substantially smaller than the mean chord, and thus their actual lineal energy is is systematically underestimated if using the mean chord approximation.

The detector efficiency is defined by the number of particles that traverse the LGADs active regions, i.e. those that are tracked. This number depends on the LAGD configuration, i.e. the number of detection strips contained in a sensor. As the dead interstrip area is the same independently of the configuration, for a given total area of the sensor, by lowering the number of strips the detection efficiency increases. However, a larger number of strips results in a superior spatial resolution. To optimize the detector design for our application, we investigated HDM performances using three different LGAD configurations: 34, 71 and 288 strips per sensor.

Detection and tracking efficiencies were assessed by studying the composition of the radiation field detected by HDM versus the radiation field incoming on the TEPC. We identified three categories of events: i) particles detected by the entire system (i.e tracked events); ii) particles lost (i.e. only traversing the active volume of some detectors); iii) particles non-trackable (i.e. those with not enough energy to reach the fourth LGAD).

For each category, we studied the kinetic energy spectra, track length distribution, real track versus track reconstructed with the tracking algorithm and microdosimetric spectra.

Independently of the primary ion and LGAD configuration, the mean track length of the tracked events is always higher that the value of all incoming particles. Events traversing the TEPC with a small track have a higher probability to miss the LGAD detectors. In fact, LGADs with 34 and 71 strips configurations have a total height and width comparable to the TEPC diameter, so if a particle reaches the TEPC with a given angle with respect to the primary beam direction, it is probable that its path will not cross all the LGADs. This hypothesis is confirmed by the fact that the 288 strips configuration collects a significantly higher portion of small-track particles (Figure 3.8 and Figure 3.9). Furthermore, for this configuration the mean track length of the tracked events is closer to the value of all particles (see Table 3.2). The mean tracks obtained when HDM is exposed to protons and carbon ions are similar for the 34 and 71 strips configurations. For the 288 strips configuration, HDM provides a more accurate track distribution for protons than for carbon ions. In fact, secondary fragments produced by protons reach, on average, smaller scattering angles compared to those generated by carbon ions [Rovituso and La Tessa, 2017].

However, those are the chords that suffer most from a high error on the tracking, as panels

A, **C** and **E** of Figure 3.8 and Figure 3.9 show for all the configurations. Furthermore, panels **A**, **D** and **G**, besides confirming the above mentioned fact that the bigger sensor takes better into account lower track lengths, they demonstrate also that the spatial resolution of the sensors, namely the widths of their strips, has a clear effect on the homogeneity of the track distribution. In fact, it can be noticed that the lower the spatial resolution is, the more the reconstructed tracks will have some preferential track lengths.

Finally, a comparison between panels **F** of Figure 3.8 and Figure 3.9 supports the hypothesis that, for protons, the 288 strips configuration is able to collect a track distribution which is very similar to the real one, while for carbon ions the distribution is still slightly underestimated for the small tracks.

Differences in the track length distributions for the LGAD configurations translate into different microdosimetric $y_T d(y_T)$ spectra (Figure 3.10 and Figure 3.11). A bigger sensor, like the 288 strips configuration, is able to collect more events with smaller TEPC tracks, which are the main contributors of the high y_T region.

The characterization of lost events indicates that the majority is caused to the LGADs fill-factor (interstrip regions). Thus, this issue can be resolved by increasing the measurement time to collect enough statistics.

For events that suffer MCS in the detectors, if the deviation angle is large enough they will be lost. In fact, even trying to enlarge LGADs or place them at a given angle with respect to the beam direction, the reconstructed track would be affected by errors too large to make the data of any value. The probability of loosing a particle because of MCS strongly depends on the HDM position in the radiation field. Depths in the Bragg peak regions as well as distal positions represent the worst cases because of the low kinetic energy of the particles populating these regions. This reflects on the fact that the proposed HDM configuration has limitations in those regions since it cannot operate close to and at distal part of the SOBP.

Finally, particles that do not have enough kinetic energy to reach all detectors are also a limit of HDM detection efficiency. Nonetheless, this issue can be partly solved by exploring the possibility of producing LGADs with thinner active layers or decreasing the substrate width. For instance, reducing the total LGAD thickness down to 70 µm is considered achievable in the near future.

3.3 HDM Readout system

The implementation of HDM readout is a challenging task: due to its hybrid nature, this detector must simultaneously capture two different types of information: the energy deposition recorded in tissue-equivalent material by the TEPC and the reconstruction of the particle tracks provided by the LGADs. As the main topic of this thesis, one chapter is dedicated to the development of the LGAD readout: chapter 4; while another chapter is dedicated to the TEPC readout: chapter 5.

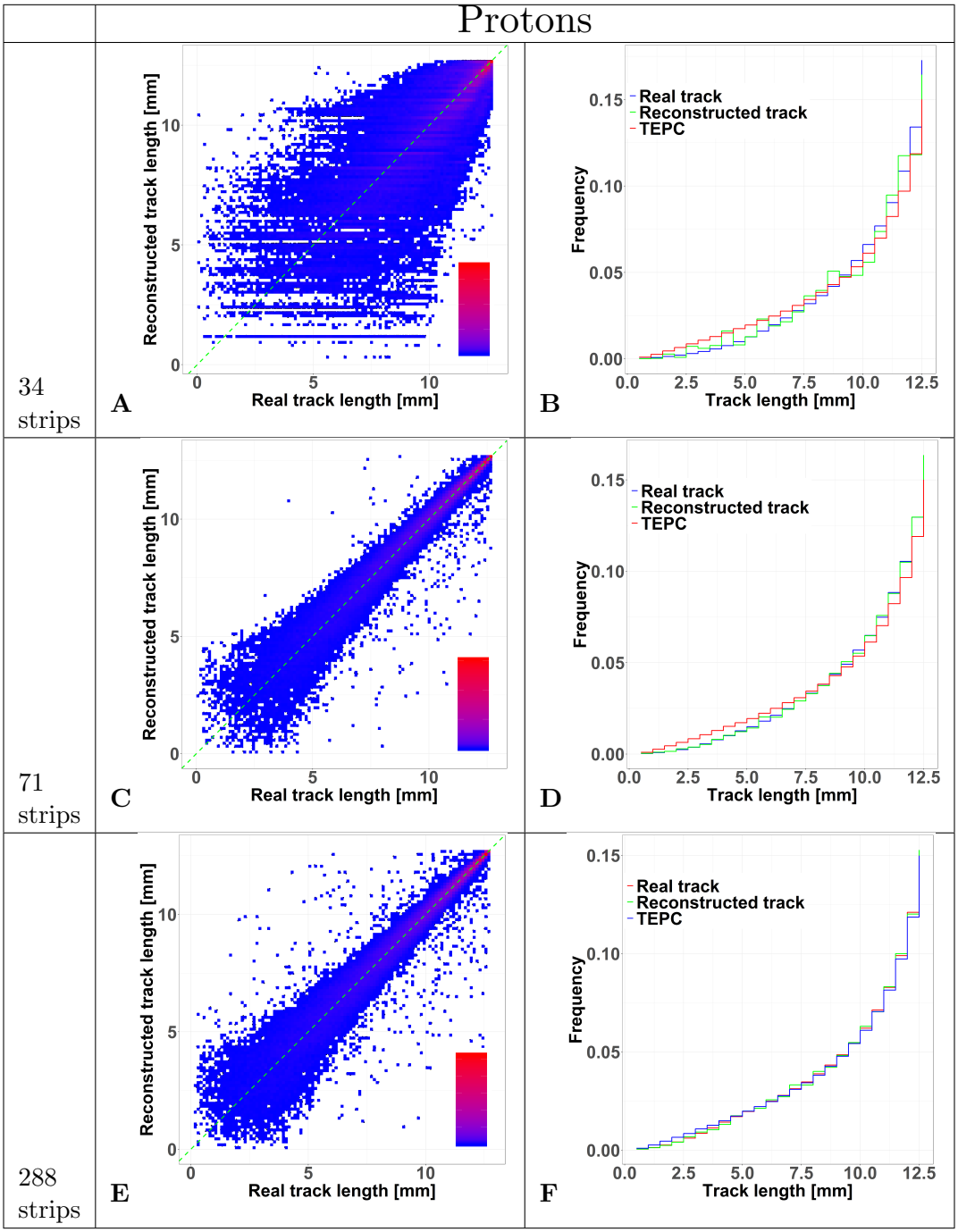


Figure 3.8: HDM performances when exposed to 150 MeV protons at 10.74 cm depth in water. The results are shown for 34, 71 and 288 strips LGAD configurations. Panels **A, C, E** shows 2D color plots of track length obtained with HDM versus real track length calculated directly with Geant4. The green dashed line at 45 degrees indicates the perfect agreement between the two datasets. The colors represent regions with a high (red) or low (blue) density of events. Panels **B, D, F** illustrate the comparison between the track length distributions of particles tracked by HDM considering the real track lengths calculated with Geant4 (blue line - TEPC) or that reconstructed with the tracking algorithm (green line - Reconstructed track). The distributions of the real track lengths obtained directly from the simulation, considering the whole particles population (tracked and not tracked by HDM) is also shown (red line - Real track).

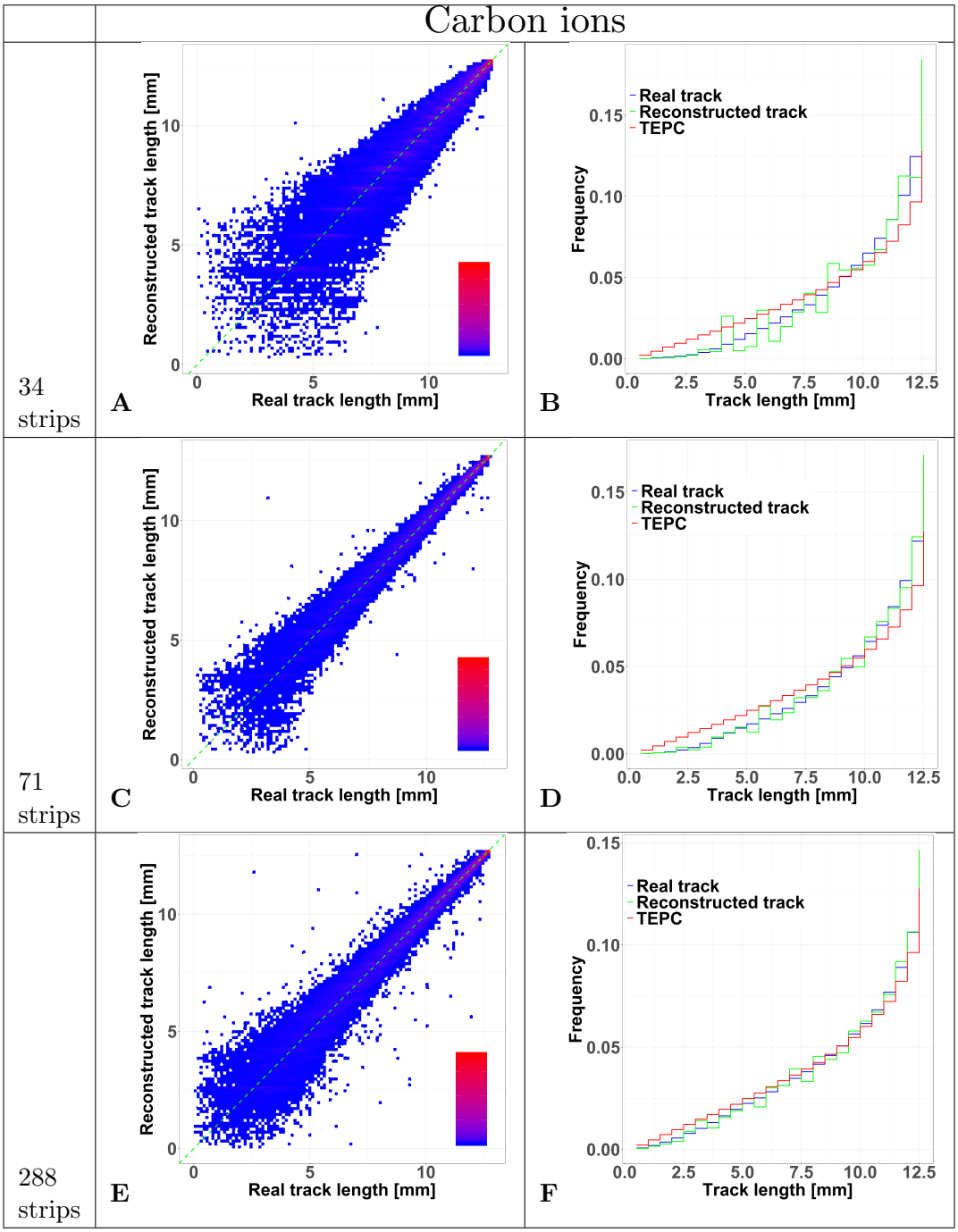


Figure 3.9: HDM performances when exposed to 290 MeV/u carbon ions at 10.74 cm depth in water. The results are shown for 34, 71 and 288 strips LGAD configurations. Panels **A, C, E** shows 2D color plots of track length obtained with HDM versus real track length calculated directly with Geant4. The green dashed line at 45 degrees indicates the perfect agreement between the two datasets. The colors represent regions with a high (red) or low (blue) density of events. Panels **B, D, F** illustrate the comparison between the track length distributions of particles tracked by HDM considering the real track lengths calculated with Geant4 (blue line - TEPC) or that reconstructed with the tracking algorithm (green line - Reconstructed track). The distributions of the real track lengths obtained directly from the simulation, considering the whole particles population (tracked and not tracked by HDM) is also shown (red line - Real track).

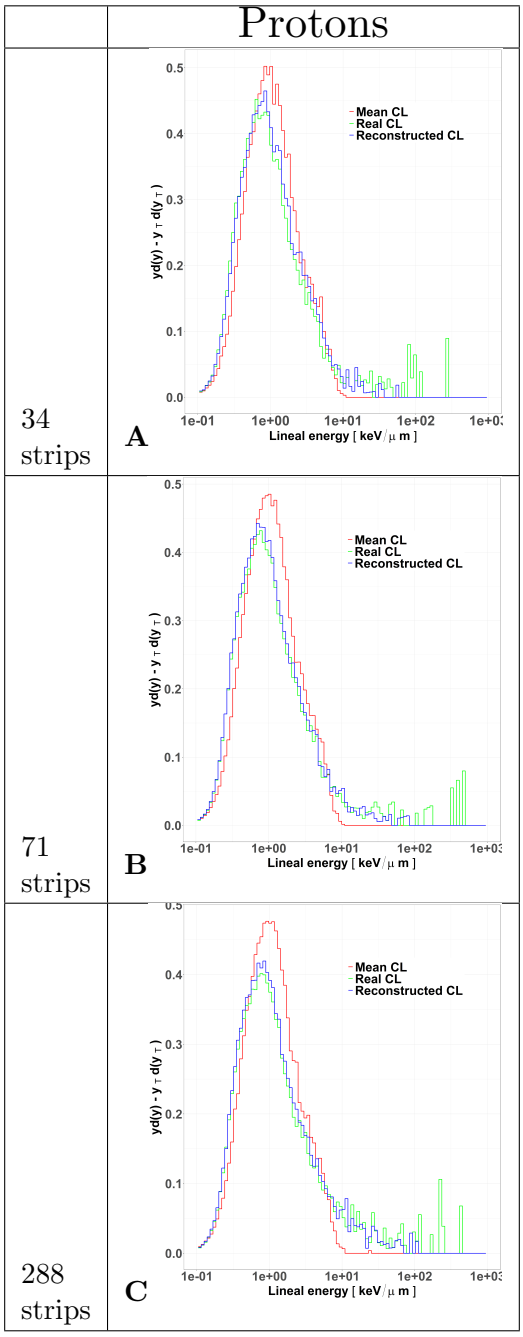


Figure 3.10: Microdosimetric spectra of all particles tracked by HDM when irradiated with 150 MeV protons at a depth or 10.74 cm in water. The distributions include the standard $yd(y)$ spectra calculated with the mean chord length (red line) and the $y_T d(y_T)$ spectra obtained either with the real track length (green line) or with the value estimated with the tracking algorithm (blue line). The distributions are shown for LGAD configurations with 34 (panel **A**), 71 (panel **B**) and 288 (panel **C**) strips..

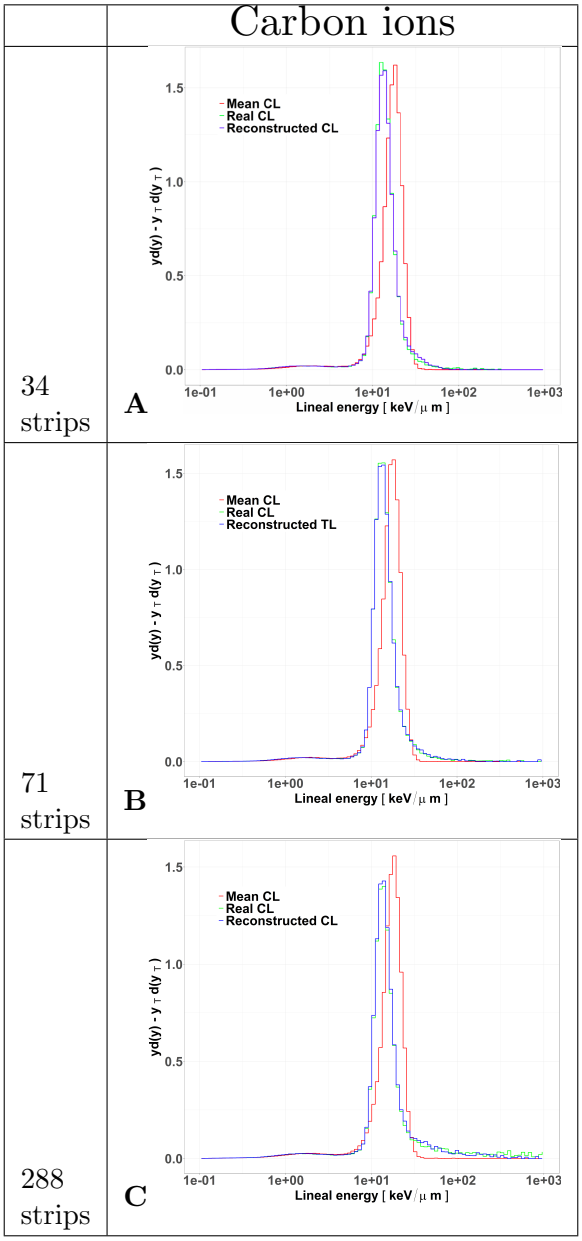


Figure 3.11: Microdosimetric spectra of all particles tracked by HDM when irradiated with 290 MeV/u carbon ions at a depth or 10.74 cm in water. The distributions include the standard $yd(y)$ spectra calculated with the mean chord length (red line) and the $y_T d(y_T)$ spectra obtained either with the real track length (green line) or with the value estimated with the tracking algorithm (blue line). The distributions are shown for LGAD configurations with 34 (panel **A**), 71 (panel **B**) and 288 (panel **C**) strips.

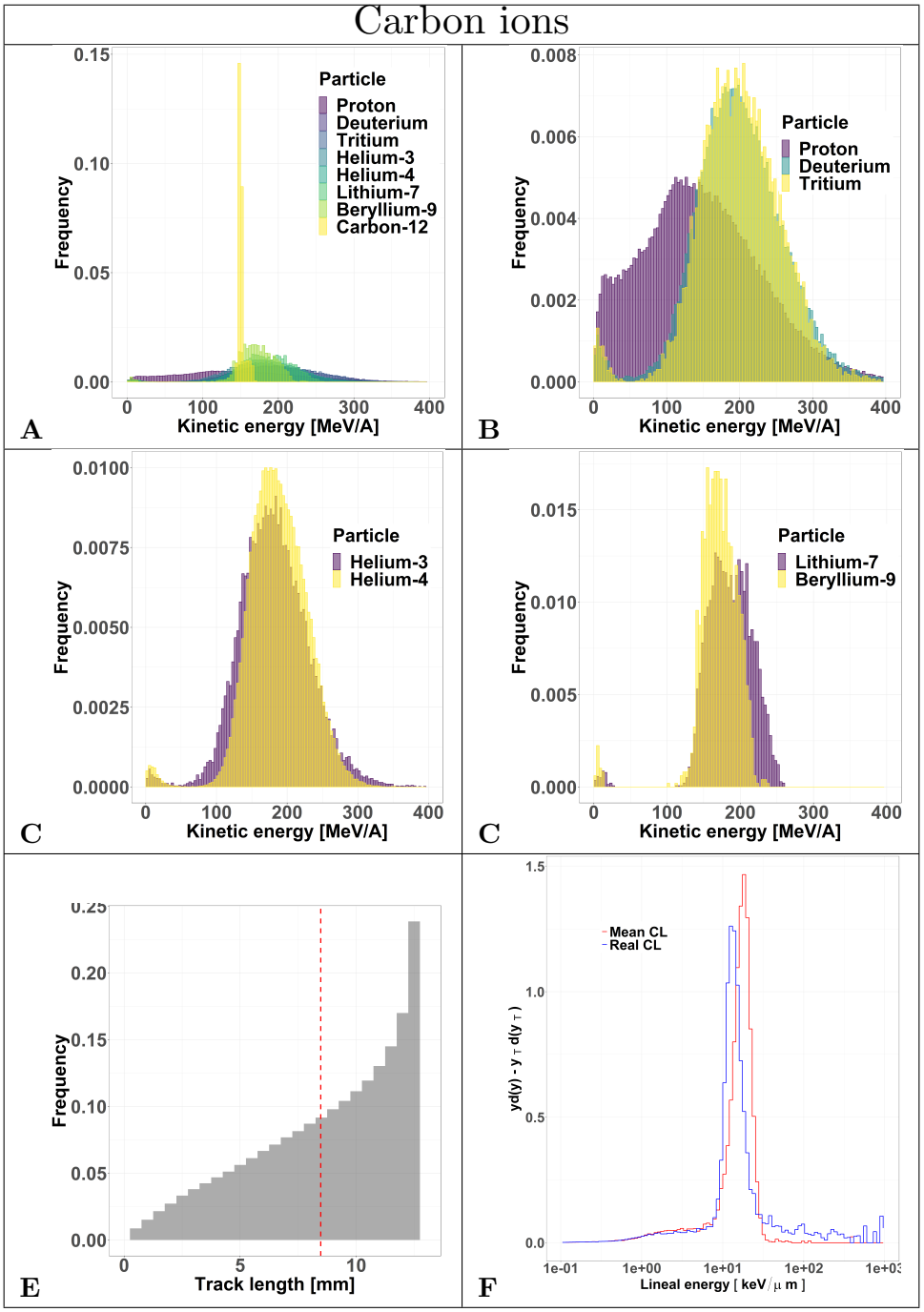


Figure 3.12: Characterization of the particles lost by HDM when irradiated with 290 MeV/u carbon ions at a depth or 10.74 cm in water. Panels **A** to **D**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **E**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **F**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line).

Chapter 4

Low Gain Avalanche Detector readout for HDM

4.1 Introduction

In this section, I present an overview of the existing hardware dedicated to handling the readout of Low Gain Avalanche Detectors (LGADs). I focus on the signal processing aspect, covering how signal is generated within the solid-state detector, electronic signal processing, and the available solutions for implementing the readout and displaying data to the end user, following a bottom-up approach.

4.1.1 Low Gain Avalanche Detector signal

A Low Gain Avalanche Detector (LGAD) is a semiconductor detector designed for the detection and measurement of ionizing radiation. More detailed information on the geometry of LGAD can be found in subsection 3.2.1. LGADs operate under a reverse bias voltage, which creates a powerful electric field to accelerate charge carriers generated by ionizing radiation. As a result, a characteristic signal is produced as a function of time, as depicted in Figure 4.1. The LGAD signal can occur when sufficient energy, namely 3.6 eV [S.M. Sze, 2006], is deposited inside the p-bulk region to create an electron-hole pair. Due to reverse bias, the strong electric field accelerates the electron toward the anode and the holes toward the cathode (1. *drift of electrons* in Figure 4.1). As they gain energy from the electric field, they can reach the critical state where additional electron-hole pairs can be created (2. *onset of multiplication* in Figure 4.1), leading to a controlled charge multiplication effect and consequent amplification of the signal. At the end of the multiplication, the signal has reached its maximum amplitude (3. *end of multiplication* in Figure 4.1). Finally, the electrons are collected, leaving only the hole drift (4. *drift of holes* in Figure 4.1), since the hole mobility is typically several times lower than that of the electrons [S.M. Sze, 2006]. The holes are then collected (5. *end of drift* in Figure 4.1), setting the signal to the baseline.

It is important to underline that charge collection time is heavily dependent on the thickness of the sensor (active reagion). In Figure 4.1 the thickness of the sensor is 300 µm leading to a total collection time of $\simeq$ 30 ns. For sensors with a thinner active region, the collection time decreases [Sadrozinski et al., 2016]. This capability is particularly valuable in beam monitoring applications that rely on counting individual protons in therapeutic beams. In such scenarios, the increased particle flux requires shorter signals to mitigate the counting inefficiency caused by the simultaneous occurrence of temporally close signals, a phenomenon known as pileup.

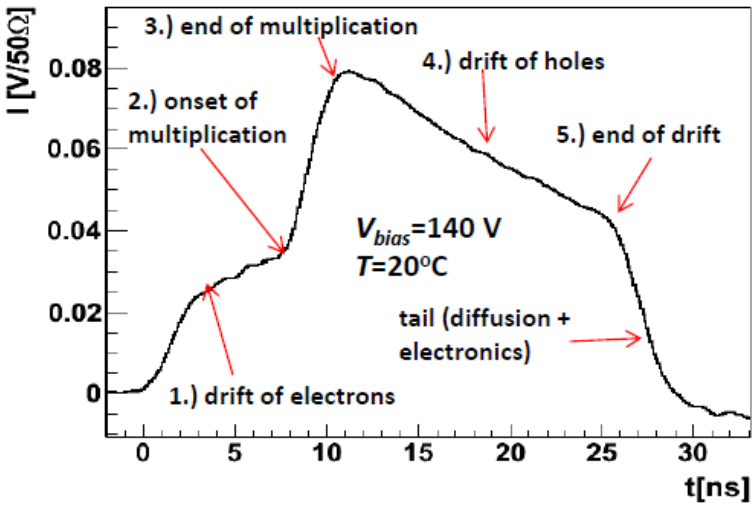


Figure 4.1: Induced signal shape of an LGAD from a short-pulse red laser illumination. V_{bias} = 140 V at a temperature of 20 °C. LGAD thickness is 300 µm. From [Kramberger et al., 2015].

Figure 4.2 shows a simulation of the LGAD pulse for different thicknesses of 50, 150, 300 µm. The observable effect is a reduction of the pulse width as the thickness decreases. LGADs sensors employed for HDM have a thickness of 50 µm, producing a signal of $\simeq$ 1.5 ns, which is challenging to handle due to its small amplitude and fast peaking time.

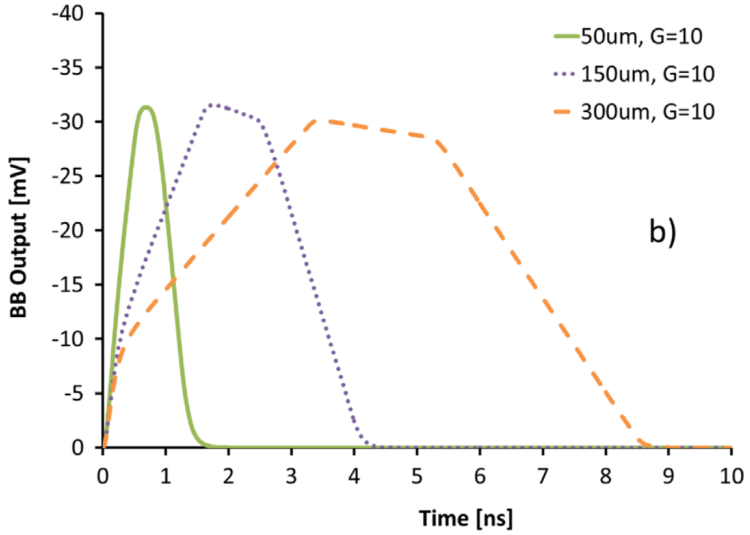


Figure 4.2: LGAD simulated signal as function of the thickness 50, 150, 300 µm. From [Sadrozinski et al., 2016].

To deal with very high particle rates in therapeutic applications, the readout must be fast enough to separate different signals close in time without introducing significant dead time. In conclusion, the readout faces a significant challenge due to the short pulse duration, approximately in the nanosecond range, and the small amplitude. To tackle this demanding task, the Istituto Nazionale di Fisica Nucleare (INFN) of Turin has developed a dedicated Application-Specific Integrated Circuit (ASIC).

4.1.2 ABACUS ASIC

As described in subsection 4.1.1, reading the pulse of the LGAD signal is a challenging task due to its small duration ($\simeq$ 1.5ns). To overcome this difficulty, a dedicated application-specific integrated circuit (ASIC) was developed. ASICs are specialized integrated circuits designed for specific applications or tasks. Unlike general-purpose integrated circuits, ASICs are customized and optimized for a specific function or set of functions. They are designed for high performance, low power consumption, and often to meet specific requirements such as speed, power efficiency, or area constraints. For the specific task of handling LGAD strips for high-speed particle counting, an ASIC named ABACUS (Asynchronous-logic-based analog counter for ultra-fast Silicon strips), shown in Figure 4.3, was designed. This ASIC is capable of handling 24 channels, expecting a minimum charge of 4 fC up to 150 fC [Fausti et al., 2021].

The main task of ABACUS is to discriminate single LGAD signals originating from each

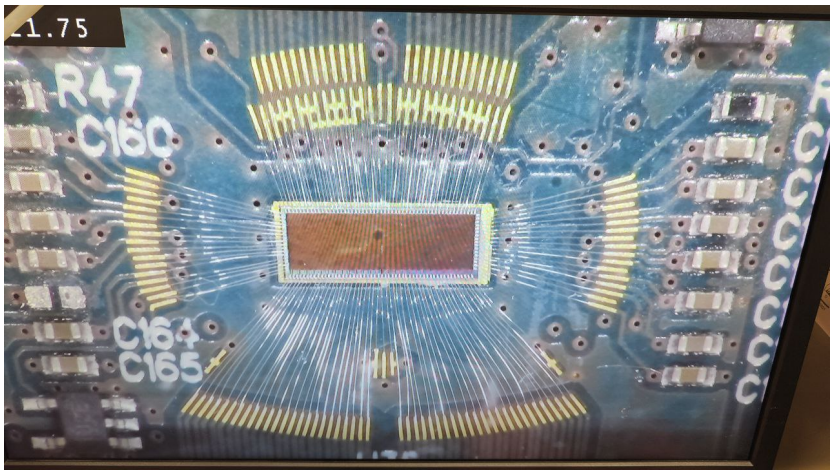


Figure 4.3: A photo of the ABACUS with bondings.

LGAD strip with a reference value that can be adjusted internally. The LGAD strip signal is first amplified and then compared to the reference threshold value. If the signal exceeds the threshold value, ABACUS generates a logical pulse. Figure 4.4 shows the block diagram of a single channel where it can be categorized into four distinct parts based on their respective functions: amplification (blocks 1 and 2), discrimination (blocks 3 and 4), driver (blocks 5 and 6), and feedback reset (blocks 7 and 8). To ensure signal shaping, standard digital buffers have been incorporated to achieve a precise reset signal for the amplifier feedback capacitor (block 9). Finally, a test pulse circuit (block 10) with a known input capacitance allows the user to inject a well-defined charge into the system for testing purposes.

The amplification design incorporates a Charge Sensitive Amplifier (CSA). After amplification, the signal is sent to a two-stage leading-edge discriminator. This discriminator compares the amplified signal against a threshold. Whenever the amplified signal exceeds this threshold, a pulse is generated, indicating the presence of a signal exceeding the desired level. For debugging purposes, there is an option to inject a known charge through TP_injection in Figure 4.4. This allows to simulate an LGAD strip signal caused by ionizing radiation without having the sensor connected. The pulse then triggers a feedback reset circuit designed to quickly return the input signal to the baseline level. This allows the ASIC to quickly accept new pulses, making the chip suitable for fast signal processing. Simultaneously, the output signal from the discriminator is transmitted off-chip using a Current Mode Logic (CML) differential driver. The CML driver is chosen for its ability to achieve high-speed and low-power signal transmission by employing a differential signaling scheme. With

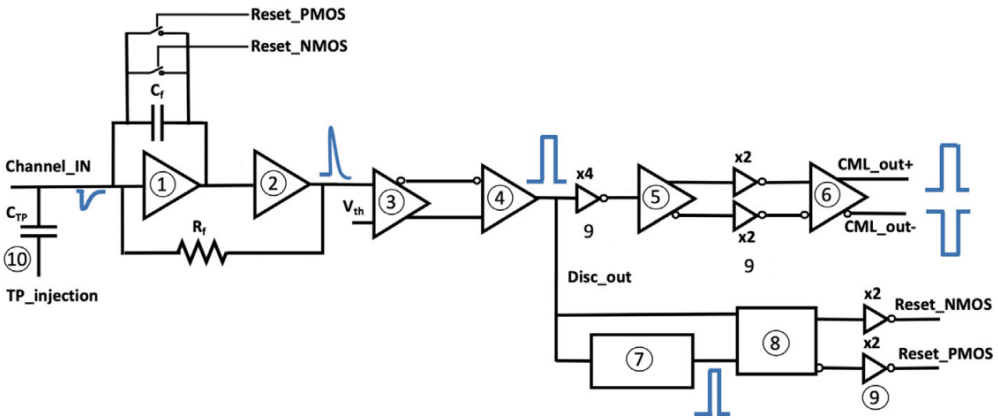


Figure 4.4: Block diagram of the ABACUS channel. Charge sensitive amplifier (1), low-pass filter (2), leading edge discriminator (3, 4), single ended to differential converter (5), current mode logic driver (6), pulser (7), recovery (8), buffers for signal shaping (9), test pulse circuit (10). From [Fausti et al., 2021]

an internal termination of 50 Ω, the driver ensures efficient signal transmission by matching the impedance and minimizing reflections. The use of differential signaling enhances noise immunity and reduces common-mode noise, contributing to reliable signal propagation. The ability to set a proper threshold level is of utmost importance, as it enables the discrimination of genuine signals from background noise. In most cases, the total signal comprises a combination of noise and the desired signal. A significant challenge lies in determining the appropriate threshold level for each channel to accurately detect the signal. An example depicting this process is illustrated in Figure 4.5, where a signal is represented by the blue line and surpasses a predefined threshold indicated by the red horizontal line. When this event occurs, a logic pulse is generated as indicated by the green curve. If the threshold is set too low, the system may generate false pulses in response to noise spikes. On the contrary, if the threshold is set too high, smaller signals are not detected. This threshold configuration process is fundamental to the operation of ABACUS, as it effectively discriminates signal pulses from the LGAD by generating logical pulses when the threshold levels are correctly configured. This process has to be done for all channels independently, as the signal from each LGAD strip can have different pedestals. To overcome this problem, the ASIC is able to adjust the discrimination threshold independently for each channel using the signal produced by a Digital to Analog Converter (DAC). Therefore, it is possible to set or read a threshold value using appropriate serial commands sent through dedicated Serial INput (SIN) data line and read through Serial OUTput (SOUT) data line.

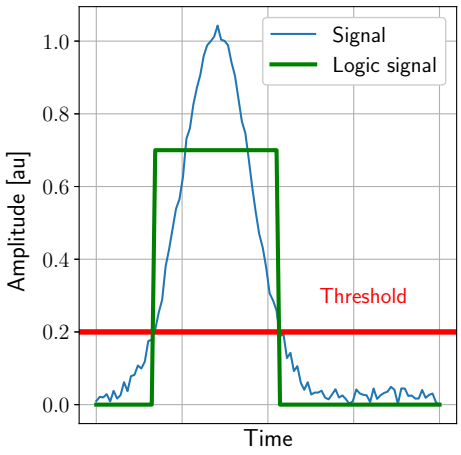


Figure 4.5: Example of signal detection by threshold. When the analog signal (shown in blue) exceeds the predefined threshold level (depicted in red), a logic signal is triggered and raised to a high level (represented in green).

The user has to provide a synchronous clock signal (CLK) that aligns with the serial streams. The inputs to the system are sampled precisely at the rising edge of the clock signal.

Therefore, it is essential to ensure that the provided clock signal is synchronous and properly synchronized with the serial data streams to ensure accurate sampling and reliable operation of the system. The typical sequence, composed of two sequences of 24 bits for a total of 48 bits, is shown in Table 4.1. Before any operation, a 16-bit initialization sequence (init) of 0xA5A5 (equivalent to 0b1010010110100101) is required. The next two bits in the sequence determine the type of operation: 0b11 indicates a write (set) operation for a threshold, while 0b10 indicates a read operation for a threshold value. The subsequent five bits encode the channel address, encoded from 0 to 23, with the most significant bit (MSB) denoted as a_4. A maximum of 32 channels can be selected with 5 bits, but as the ABACUS chip supports only 24 channels, any selected channel above 24 will result in no operation. Following the channel address, three bits of arbitrary value are included in the sequence. Historically, these bits served specific purposes in different ASIC versions but are currently unused. In the case of write operations, the next six bits define the actual DAC value, where v_5 represents the MSB. Lastly, to complete the command, an additional 16 bits of arbitrary values are required. This results in a sequence of 2×24 bit packets that form the complete command structure.

Init 0xA5A5 16 bits	RW selector 2 bits	DAC address 5 bits	NA 3 bits	DAC value 6 bits	Tail 16 bits
0b1010010110100101	0b11-write 0b10-read	$a_4a_3a_2a_1a_0$	don't care don't care	$v_5v_4v_3v_2v_1v_0$ don't care	don't care don't care

Table 4.1: ABACUS SIN command sequence to perform read or write (RW) operations related to the internal threshold configuration. An initialization sequence (init) of 16 bits, represented as 0xA5A5 = 0b1010010110100101, is required before any operation. The next 2 bits specify the operation type: 0b11 for writing or setting a threshold, and 0b10 for reading a threshold. The subsequent 5 bits encode the channel address of the DAC, with a_4 as the MSB. After that, 3 unused bits follow, and in the case of write operations, the subsequent 6 bits define the DAC value, with v_5 as the MSB. Finally, the command is completed with 16 arbitrary bits, resulting in a sequence of 2×24 bit packets.

During read operations, it is necessary for the user to provide a valid clock for an additional 24 cycles. This allows the ASIC to transmit the required information through the SOUT pin, which can be sampled on the rising edge of the CLK signal. It is important to note that the 6 bits sent as the DAC value are disregarded during this process. The read sequence sent by the ASIC via SOUT is encoded according to Table 4.2. The first two 0b11 bits acknowledge the correct interpretation of the read SIN command. If the bits are not 0b11, the read command has encountered an error. The next 5 bits encode the channel address as confirmation from the read command, in the same configuration as writing operations with a_4 as MSB. As with a SIN command, 3 additional "don't care" bits are sent. Finally, the 6 bits representing the actual DAC value are sent, again with v_5 as MSB.

Correct command 2 bits	DAC address 5 bits	NA 3 bits	DAC value 6 bits
0b11	$a_4a_3a_2a_1a_0$	don't care	$v_5v_4v_3v_2v_1v_0$

Table 4.2: ABACUS read command SOUT. The first two bits, 0b11, confirm the correct interpretation of the read SIN command. If these bits are not 0b11, it indicates an error. The next 5 bits encode the channel address with a_4 as the most significant bit (MSB). 3 additional unused ("don't care") bits are sent. Finally, the 6 bits representing the actual DAC value are transmitted, with v_5 as the MSB.

The current version of ABACUS threshold can span a maximum of $\sim 24\,\text{mV}$ by adjusting the DAC values from a minimum of 0 to a maximum of 63. This has proven to be insufficient and should be extended in the next production, which is currently underway.

Furthermore, it should be noted that Current Mode Logic (CML) standard, output format of the ASIC, it not commonly used in electronics. As a result, an additional step is required in the acquisition chain to convert the output signal to Low Voltage Differential Signaling (LVDS), which is a more widely adopted standard. This conversion process involves additional external components to enable seamless compatibility and ensure proper signal transmission and reception. While the inclusion of LVDS conversion adds complexity and cost to the acquisition chain, it is necessary to bridge the compatibility gap and facilitate efficient transfer of signals between the ASIC and the FPGA or other components in the system.

An example of an LVDS digital signal generated by the ABACUS chip, converted to LVDS and measured with a differential probe is shown in Figure 4.6. The signal was generated by an external pulser connected to TP_injection, with a synchronous trigger signal displayed in dark blue. The ABACUS signal is represented in magenta, with an amplitude of 9.53 mV and a width of 2.0 ns. It is important to note that detecting this type of fast signal presents a significant challenge. According to the Nyquist-Shannon sampling theorem [Shannon, 1949], sampling a signal with a duration of $(2.0\,\text{ns})^{-1} = 0.5\,\text{GHz}$ requires a sampling frequency of at least 1 GHz.

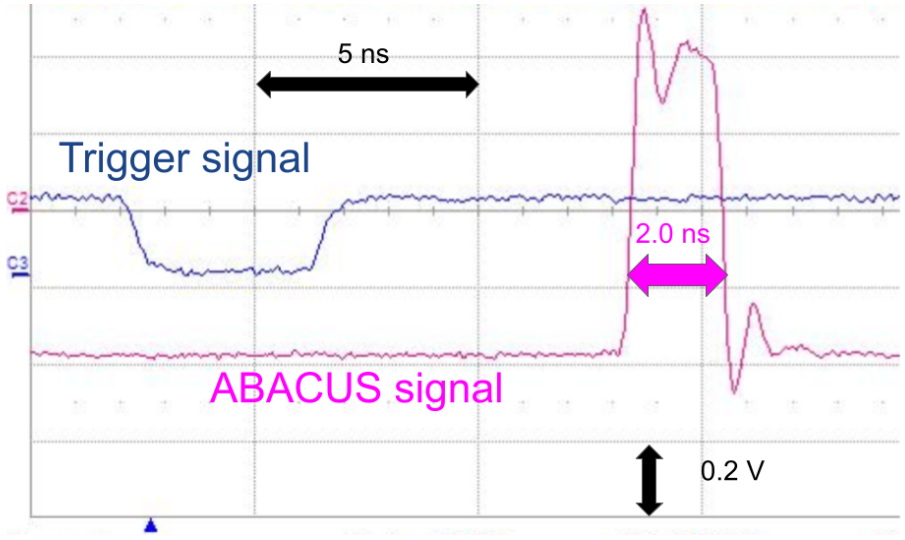


Figure 4.6: Digital output signal from an ABACUS chip after external pulsing via TP_injection. The CML signal is further converted to LVDS and measured with a differential probe (magenta line). The amplitude of the latter is 9.53 mV and the width is 2.0 ns. A trigger signal of the pulser, in sync with the excitation pulse, is also shown in dark blue.

In summary, to effectively process LGADs signals, a dedicated ASIC called ABACUS has been developed. ABACUS is designed to handle up to 24 LGAD signals and operates by comparing the input signal with an adjustable threshold. When the signal exceeds the threshold, ABACUS generates a fast CML signal. Additionally, it incorporates a specialized reset circuitry that facilitates swift recovery of the baseline, making it suitable for rapid counting operations.

4.1.3 ESA_ABACUS PCB

To host the whole system: ABACUS and LGAD, a special Printed Circuit Board (PCB), namely ESA_ABACUS, has been developed by the Istituto Nazionale di Fisica Nucleare (INFN) of Turin [Fausti et al., 2018]. PCBs are the backbone of modern electronic devices and systems. They serve as the foundation for interconnecting and supporting electronic components, providing a platform for the assembly, organization, and electrical interconnection of various circuit elements. The ESA_ABACUS design serves as a physical support structure to accommodate both the LGADs and up to six ABACUS chips. The ABACUS chips, designated with a number from 1 to 6, are glued and bonded to the ESA_ABACUS, which provides the necessary connectors, power supply, and voltage biasing. ESA_ABACUS plays a critical role in converting the digital signal from the ABACUS Current Mode Logic (CML) standard to the more widely supported Low Voltage Differential Signaling (LVDS) standard.

The board is symmetrically designed to have two independent regions. Each region requires its own power and can support up to 3 ABACUS. Since both analog and digital signals are present, the PCB is designed to have 3 different voltage sources that are kept separate to minimize noise: analog power, digital power and service power. Both analog and digital power require 3 V while service power requires 4 V. The voltages are further down-regulated to 1.2 V for analog and digital power, while service power is down-regulated to 3.3 V. In order to minimize the impact on other components, the reverse polarization for the LGAD is designed to be floating. This design choice ensures that any potential effects or disruptions caused by reverse polarization are minimized and do not adversely affect the performance or operation of the other components on the board.

For debugging and testing purposes, the board has dedicated test points for each ASIC, allowing voltage measurements to verify their expected nominal values. This helps to identify potential issues and ensure proper functionality. Deviations from the expected values require further investigation and troubleshooting. In addition, the board is equipped with two SubMiniature version A (SMA) connectors for each ASIC. These connectors allow the user to directly pulse the input channels of the ABACUS, specifically targeting the even or odd channels. By carefully choosing the shape of the pulse signal, selecting the amplitude, slew rate and duration, it is possible, knowing the capacitance C_{TP} in Figure 4.4, to inject a specific charge to mimic the energy deposition in the LGAD strip. This feature provides a convenient means for users to input test signals directly into the ABACUS system, facilitating thorough testing and evaluation of the readout system. Figure 4.7 displays the top region of the ESA_ABACUS board, where all the connectors are located. The power connectors for analog, digital and service can be found at the bottom of the photo. As mentioned earlier, the board follows a symmetrical design with independent left and right sides. Hence, there are a total of six connectors: three for the left side and three for the right side. The hole in the center of the board is designed for hosting the sensor and to limit the interaction of the beam with unnecessary material. On the top, the High Voltage (HV) connector responsible for supplying inverse polarization to the LGAD sensor can be seen. Along the region of the board, six SMA connectors are visible. Each pair of SMA connectors is designed for pulsing the even or odd channels of a specific ASIC. In the case of the board depicted in Figure 4.7 and Figure 4.8, it is configured to handle only three ABACUS chips, which explains the presence of only six (three for odd channels and three for even channels) SMA connectors instead of twelve. Finally, 5 flat cable connectors are present to handle the configuration and data transfer to the readout back-end.

The bottom side of the board is illustrated in Figure 4.8. At the top of this side, there are six voltage regulators, one dedicated to analog power, one for digital power, and one for service power, on each of the two regions of the board. The region intended for hosting the

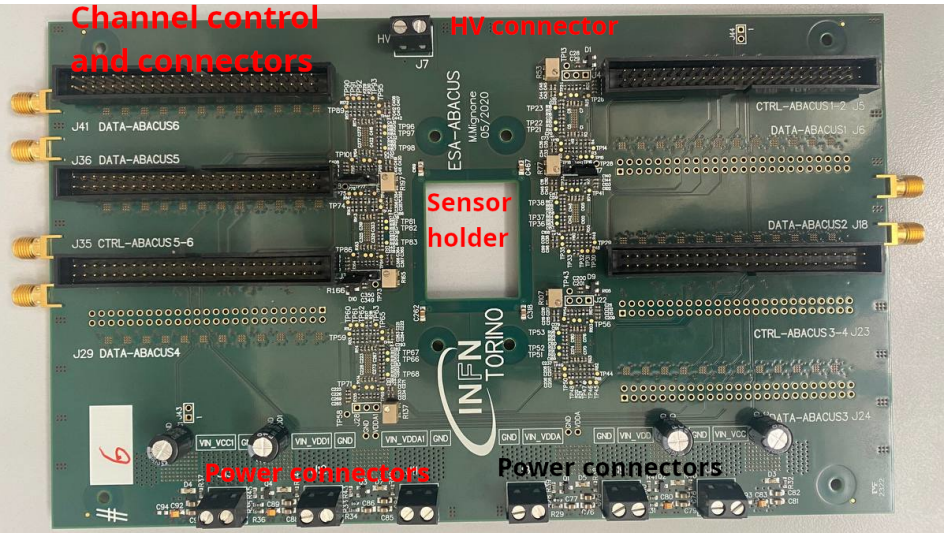


Figure 4.7: The top view of the ESA_ABACUS board. All the connectors, including the readout connectors, are situated on this side of the board. The reverse biasing for the LGAD is positioned on the top surface of the board. In the center of the board, channel connectors and control connectors are located. On the bottom side of the board, the power supply connectors can be found, providing the necessary power to the system.

ABACUS chips is highlighted in purple. As mentioned earlier, this specific board configuration is designed to accommodate only three ASICs. Hence, three lines of Current Mode Logic (CML) to Low Voltage Differential Signaling (LVDS) converters are highlighted in the figure; each line corresponds to the set of converters for the 24 signals of an ABACUS chip.

Each ESA_ABACUS module is equipped with three 16-bit Digital to Analog Converters (DACs) of type LTC2602. These DACs are specifically designed to establish a shared threshold value for all 24 channels within each ASIC. In this design, the thresholds are set in a two-step process: 1. LTC2602 is used to find and set the common threshold. 2. Using the ABACUS chip DAC to fine-tune individual thresholds. This enables precise adjustment of the threshold values to meet the specific requirements of each channel. When employing this combination of DACs, the ESA_ABACUS and ABACUS chips work in tandem to provide both a common threshold for all channels and the flexibility for channel-specific threshold calibration.

Each LTC2602 DAC is equipped with two voltage outputs, allowing for the simultaneous handling of two ABACUS chips. Considering that the maximum voltage threshold is 1.2 V, the minimum adjustment increment for the 16-bit DAC is determined by dividing the range of 1200 mV by 2^{16} (the total number of possible digital values for a 16-bit DAC). This yields a minimum adjustment of 0.018 mV. To configure the DAC value, a serial protocol is adopted. The signal carrying the configuration data is transmitted through the Serial Interface Data Input (SDI) pin and is sampled on each rising edge of the Serial Interface Clock (SCK). It is essential for the user to provide a valid SCK signal that is synchronized with the SDI signal. In addition to the SDI and SCK pins, there is another pin called the Chip Select/Load Input (CS/LD) that the user must take care of. This pin should be connected to a low logical level for the entire duration of the communication stream. The communication protocol, outlined in Table 4.3, consists of a 24-bit sequence. The first four bits are used to select the operation, with the focus primarily on the write command for our purposes. However, it is worth remarking that other commands are implemented to serve specific applications, such as putting the DAC in a dormant state to minimize power consumption. The subsequent four bits are used to select the address of the DAC or specify which ASIC to refer to. It is important

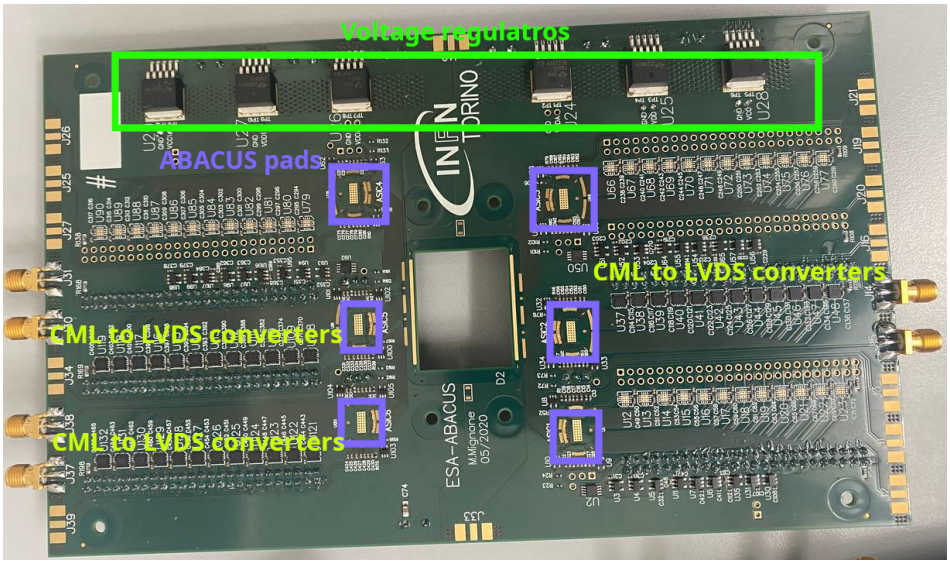


Figure 4.8: The bottom view of the ESA_ABACUS board showcases the voltage regulators positioned on the upper part. The board also features six dedicated pads for hosting the ABACUS chips. Furthermore, a substantial portion of the board is allocated for the CML to LVDS converters, essential for converting signals from the Current Mode Logic (CML) standard to the Low Voltage Differential Signaling (LVDS) standard

to highlight that the LTC2602 DAC has two outputs, each connected to a separate ABACUS (e.g., ABACUS 1 and ABACUS 2 are connected to one LTC2602). As a result, updating the odd address will modify the threshold for ABACUS 1, while updating the even address will modify the threshold for ABACUS 2. Lastly, the transmitted value of the threshold, with the most significant bit (MSB) represented by v_{15}, is conveyed to complete the communication sequence. It is important to note that precise timing constraints must be observed, particularly

Command 4 bits	DAC address 4 bits	DAC value 16 bits
0b0011-write	0b1111-all 0b0000-even 0b0001-odd	$v_{15}v_{14}\ldots v_1v_0$

Table 4.3: The 24-bit writing sequence sequence for LTC2602 equipped on the ESA_ABACUS. The first four bits determine the operation, primarily utilizing the write command. The subsequent four bits select the DAC address or the corresponding ASIC. The protocol takes into account that the LTC2602 DAC has two outputs connected to separate ABACUS chips, where odd addresses can update ABACUS 1, 3 or 5 and even addresses can update ABACUS 2, 4, or 6. The remaining bits transmit the threshold value, with the MSB denoted as v_{15}.

when setting the CS/LD pin. To ensure proper functionality, the logic level of the CS/LD pin must be lowered at least 7 ns prior to the initiation of the data transfer. Adhering to this timing requirement is crucial to guarantee accurate communication and reliable operation of the system. To facilitate the verification and debugging of the DAC threshold, the user can directly measure the threshold using a dedicated test point. This feature enables the user to easily check and read the threshold value. It is noteworthy that the LTC2602 DAC does not support read operations. Therefore, the availability of the dedicated test point provides a convenient means for the user to evaluate and monitor the threshold. ESA_ABACUS also

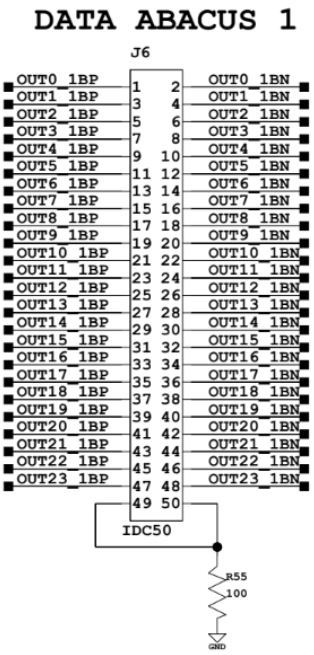
provides the option to bypass the threshold level set by the DAC and use a trimmer instead by setting a jumper on a specific pin. This alternative mode of operation eliminates the need to implement the specific transfer protocol associated with the DAC. However, it should be noted that using a trimmer for threshold adjustment requires manual intervention with a screwdriver. While this mode of operation provides flexibility, it lacks the precision achievable with DAC-based threshold adjustment, making it useful only during testing and debugging.

The connection between ESA_ABACUS and the readout system follows the LVDS standard, which relies on differential signaling. This standard requires the use of two transmission lines for each channel or data line. To perform physical connection, standard flat cable connectors with 25 pairs are utilized. However, the last pair of the connector is designated for grounding, resulting in a total of 24 usable pairs for data transmission. This allows the signals of the 24 channels managed by an ABACUS to be carried by a single flat cable. The detailed scheme is described in Figure 4.9, where each channel consists of two pairs ending with the "P" or "N", depending on the polarity: positive or negative, respectively. In Figure 4.9a the connection scheme for the ABACUS output signal is shown, where the 24 pairs are connected to the respective ASIC output, converted to LVDS. The thresholds can be programmed for both the ABACUS and the LTC2602, as mentioned above. The communication with these devices is done via the second type of connector, as shown in Figure 4.9b. This connector provides the signals necessary to control the LTC2602. These signals include DAC_SDI, DAC_SCK, and DAC_CS/LD, corresponding to SDI, SCK, and CS/LD, respectively. For the ABACUS chips, as described in subsection 4.1.2, SIN, SOUT and CLK are mapped, in the special case of ABACUS-1 to DAC_IN_1, DAC_OUT_1, DAC_CLK_1. Using the same connector, a second ABACUS chip can be controlled by mapping DAC_IN_2, DAC_OUT_2, DAC_CLK_2 to the same referenced port of ABACUS-2. Thresholds for 48 channels can be managed with this configuration: LTC2602 controlling the common thresholds of the even ASIC (24 channels) and the odd ASIC (24 channels), and two ABACUS chips controlling the thresholds of 2×24 individual channels. The two pairs labeled FPGA_1 and FPGA_2 are currently unused, but they can provide communication to the subsequent ESA_ABACUS control connector present on the board, managing the other 48 channels. This allows for the establishment of communication between different control blocks, creating a communication loop between control 1-2, control 3-4, and control 5-6, if needed.

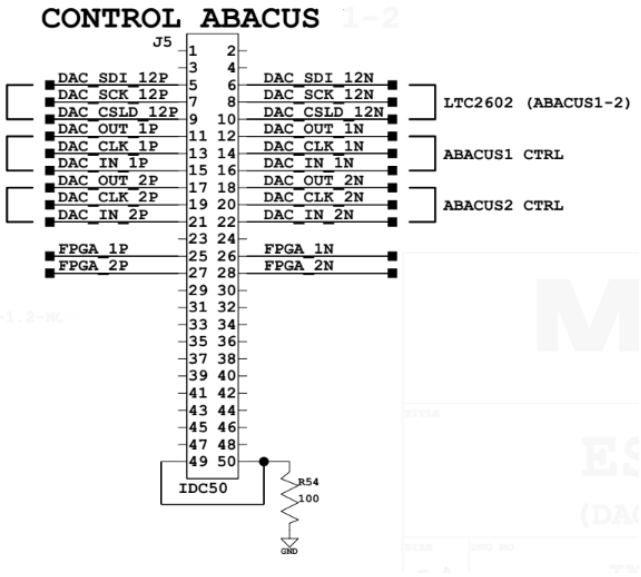
In summary, reading a total of 71 strips, as in the case of Hybrid Detector for Microdosimetry (HDM), requires the use of three ABACUS. This requirement is due to the fact that each individual ABACUS can control a maximum of 24 channels. In addition, when using adjacent ABACUS chips (e.g. ABACUS-1 and ABACUS-2), a single shared LTC2602 can be used. As a result, the total number of connections required to manage the data is simply the number of channels: 71 pairs for the data of ABACUS-1-2-3. While the required pairs to control the three ABACUS and the ESA_ABACUS thresholds, assuming two adjacent ASICs sharing one LTC2602, are 9 for ABACUS-1-2-3, 3 for LTC2602 controlling ABACUS 1-2 and 3 for LTC2602 controlling ABACUS-3. The total number of required pairs is 15.

4.1.4 Field Programmable Gate Array

The final stage of the readout is implemented and managed by Field Programmable Gate Array (FPGA). FPGAs are a type of integrated circuit that can be programmed or configured to perform various digital logic functions. Unlike traditional fixed-function integrated circuits, FPGAs can be reprogrammed to implement different logic circuits, making them highly flexible and adaptable. They consist of an array of configurable logic blocks and programmable interconnects that can be customized to meet specific application requirements. This programmability allows for rapid prototyping, design iteration, and even on-the-fly reconfigura-



(a) ABACUS data pin out connections. A total of 24 channels can be handled.



(b) ABACUS and ESA_ABACUS pin out control connections. According to subsection 4.1.2 three transmission lines are require for ABACUS: SIN, CLK and SOUT here named DAC_IN, DAC_CLK, and DAC_OUT respectively. In the case of LTC2602 programming, again three transmission lines are required DAC_SDI, DAC_SCK, and DAC_CS/LD for SDI, SCK, and CS/LD, respectively.

Figure 4.9: The flat connector pin assignment for ESA_ABACUS data ASIC-1 (a - left) and control 1-2 (b - right) in compliance with the LVDS communication protocol. As a differential protocol, it requires the use of two lines for signal transmission, forming a pair. In this pin assignment, the letter “P” or “N” at the end of each line label indicates the polarity of the respective transmission line, where “P” represents the positive line and “N” represents the negative line.

tion during operation. I relied on FPGA technology for its performance and versatility in implementing the final interface between hardware and end-users. Therefore, the FPGA task is to handle the digital data sent by the ESA_ABACUS and to configure the discrimination threshold, according to the protocol described in subsection 4.1.2 and subsection 4.1.3.

To accomplish this challenging task, the Zynq 7000 SoC ZC702 Evaluation Kit was used. This board is comprehensive of the FPGA, based on the Zynq architecture, mounted on top of a development board with predefined interfaced like ethernet port, USB port, serial port, SD card interface and two FPGA Mezzanine Card (FMC) Low Pin Count (LPC) connectors [Xil, b].

The Zynq 7000 SoC ZC702 Evaluation Kit features two FMC LPC interfaces that adhere to the VITA 57.1 specification [VIT,]. These interfaces require a specific adapter to connect the flat cable to the ESA_ABACUS. To address this adaptation requirement, the Istituto Nazionale di Fisica Nucleare (INFN) of Turin has developed a specialized PCB known as the “FMC breakout”. This board enables the conversion of each FMC connector into two 50-pin flat cables, namely data and control. However, it is important to note that not all FMC pairs are connected to the FPGA accordingly to VITA specifications [VIT,]. As a consequence, some pins are left unconnected or connected to static 3.3 V, making them unusable to send data. A diagram of the pins available for an FMC is shown in Figure 4.10, where all usable

pins are shown. The latter, thanks to the FMC breakout interface, are further divided into two 50 flat pin cables, one with 24 available pairs, the other with 12 available pairs. The other FMC port on the FPGA follows an analogous scheme. As a result, the total number

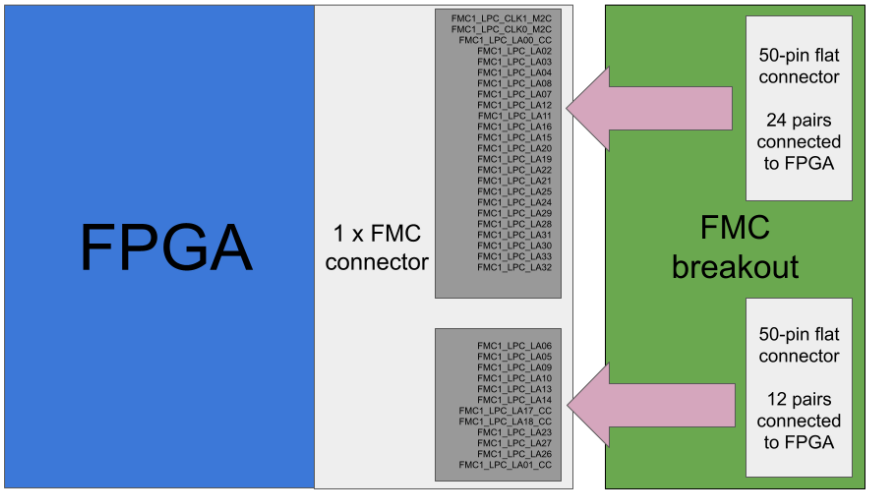


Figure 4.10: Block scheme of the FPGA FMC connector. In total, all the pairs of FMC usable to send data are 36. These are adapted to the ESA_ABACUS connectors thanks to the FMC breakout board by splitting them into a 50-pin flat cable carrying 24 usable pairs and a second one, carrying 12 usable pairs. The analogous scheme is valid for the second FMC port available in the ZC702.

of available pairs on a FMC LPC on the Evaluation Board is $24 + 12 = 36$. Since there are two connectors available on the board, the total number of differential transmission lines is $36 \times 2 = 72$. In the specific application of one layer of Hybrid Detector for Microdosimetry (HDM), where 71 channels need to be managed by the readout, according to the flat cable definition of the ESA_ABACUS described in subsection 4.1.3, a total of 71 pairs are required for readout in addition to 15 pairs for control. As a result, a single FPGA board cannot handle both data and control of the ESA_ABACUS equipped with a 71 strips sensor. To address this challenge, the use of multiple FPGAs is required, and custom flat cables need to be employed to fully utilize the available pair connections. In the specific scenario, assuming the utilization of ABACUS 1, 2, and 3, the data channels from ABACUS 1 can be handle by a 50-pair flat cable with 24 actual pairs. The same arrangement applies to ABACUS 2. However, for the remaining data connectors of ABACUS 3, a custom flat cable is needed to divide them into two sets of 12 pairs each, which can then be connected to the FMC breakout boards. A connection scheme illustrating this configuration is depicted in Figure 4.11. With one FPGA managing the data from the board, the other required FPGA can manage the thresholds by controlling the ABACUS DACs and the LTC2602s. This design is modular and scalable to read out the required 4 layers, so one FPGA can handle the data for one ESA_ABACUS, for a total of 4 FPGAs. Regarding threshold controls, it is possible to use a single dedicated FPGA, but with the disadvantage of having to create a custom 50-pin ribbon cable from the required $9 + 6 = 15$ pairs to connect to a 24-pair FMC connector.

ABACUS signal processing

Moving on to the detection of the digital logical pulse, as discussed in subsection 4.1.2 and illustrated in Figure 4.6, the capture of a signal with a duration of only 2 ns presents a sig-

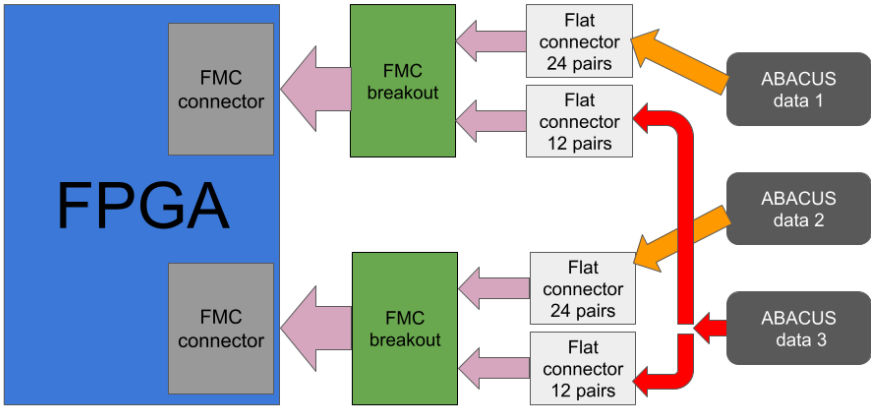


Figure 4.11: Block scheme of data FPGA connection to ESA_ABACUS in the specific case of 71 strip LGAD. Three ABACUS chip are required, in this case labeled 1, 2 and 3. While data from ABACUS 1 and 2 can be directly managed using a standard 50-pin flat cable for each ASIC (yellow arrows), data from ASIC 3 has to be divided with a specially designed connector on the remaining 12-pairs input (red arrows).

nificant challenge for a single FPGA. FPGAs have the ability to sample digital signals based on the selected clock frequency, where a higher clock frequency typically results in improved performance. Although increasing the clock speed might seem like a potential solution, it introduces various design challenges that can lead to inconsistent implementation, incorrect data capture, or even complete system failure due to timing violations. Operating at higher clock frequencies also requires careful consideration of signal integrity issues. High-speed signals are more susceptible to noise, crosstalk, and other forms of interference. Failure to properly manage these issues can lead to signal degradation, reduced reliability, and degraded system performance. Moreover, faster clock rates typically result in increased power consumption and heat generation within the FPGA. This can pose thermal management challenges, potentially leading to overheating and reduced overall system reliability. Lastly, designing for higher clock speeds often requires more complex and sophisticated design techniques. This can lead to increased development time, higher design complexity, and potentially increased cost. Therefore, while faster clocking can provide performance benefits, it is essential to carefully balance these advantages with the associated design challenges and ensure that the FPGA design can meet all timing, power, and signal integrity requirements. One solution for sampling high-speed digital signals is to rely on the Serializer/Deserializer (SERDES) primitive. SERDES are typically used when a single transmission line is available and a fast, high throughput of data is required. Its primary function is to convert a parallel digital input of data from a slow clock domain into a serial data stream along the transmission line (serialization or SER) with a high speed clock, typically a factor of 4 to 12 faster than the slow clock domain. Once the data has been sent, the deserialization (DES) process occurs by sampling the transmission line data with the high speed clock. The output is then presented in parallel format in the slow clock domain, where each output of the bus correspond to sampling points along the transmission line at specific time intervals. The size of output pins depends on how the SERDES is configured. The size of the output also defines how faster the high speed clock is with respect to the slow clock signal: assuming an output of dimension n, the fast clock is n times faster than the slow clock. The input signal is then sampled at

n equal spaced points by the fast clock. The total time covered by all n points corresponds to one slow clock cycle. These sampled points represent the digital data that was originally serialized. When the output is presented in parallel format in the slow clock domain, each output of the bus corresponds to a specific sampling point along the transmission line. This allows the reconstructed parallel data to be processed and further utilized in the slower clock domain. They are usually available by default on most of the development board's connection pins, and, they can be used to sample the digital pulse signal generated by the ABACUS chip each time a signal above threshold is detected. In fact, by using SERDES in the deserializer configuration, it is possible to sample an input line with a frequency that is several times higher than the main clock line speed, thus allowing the detection of fast signals. The process of deserialization in SERDES can be visualized using a diagram, as depicted in Figure 4.12. In this scheme, the input pulse to be detected is represented by a magenta square in the fast clock domain. The SERDES operates in a 1:10 configuration, where the fast clock domain is 10 times faster than the slow clock domain. Each time interval spanned by a slow clock iteration is sampled 10 times by the fast clock, capturing the input pulse at different, equally spaced intervals. These sampled points are then transmitted to the parallel output on the next clock cycle. Referring to the specific example shown in Figure 4.12, the input pulse is detected as a high level during the 3rd sample interval. Consequently, during the next clock cycle, the parallel output 3 reflects a high level signal.

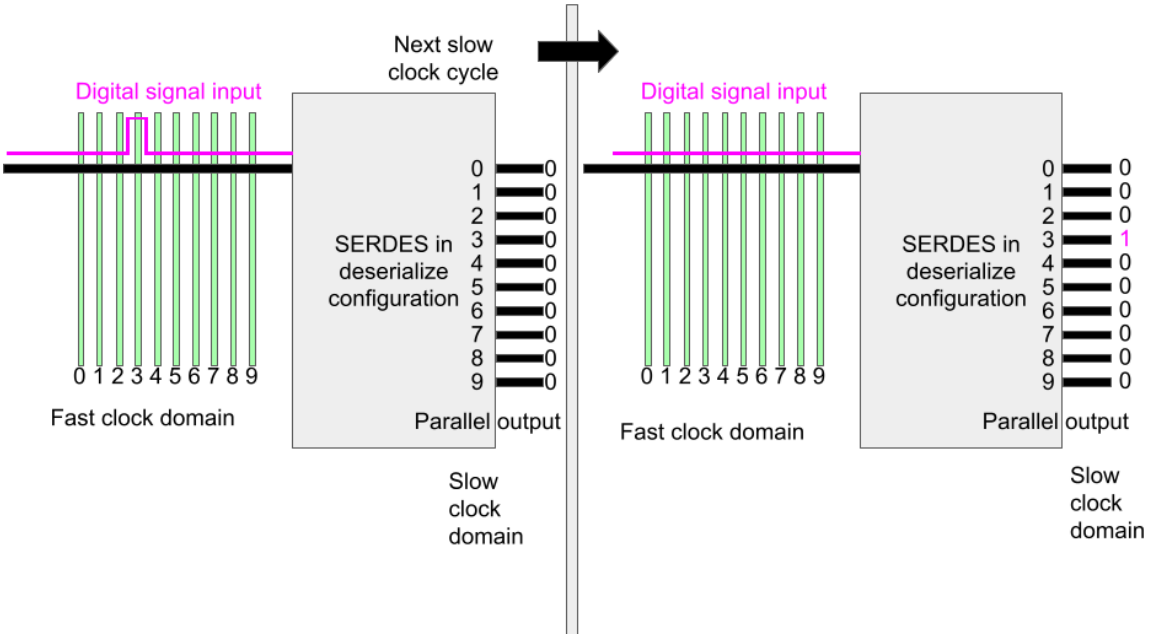


Figure 4.12: The operation of a SERDES in a deserializer 1:10 configuration. In the fast clock domain, a digital signal (depicted in magenta) is inputted into the SERDES. This signal is then sampled at 10 equally spaced points along the transmission line. Among these samples, only the third sample detects the signal as a high logical level. As a result, in the subsequent slow clock cycles, the parallel output of the SERDES will have the third element of the bus set to a high level.

With this method is possible to archieve faster sampling of digital signals. By design FPGAs logic can detect signals at positive (or negative) edge of the selected clocking frequency. Since a frequency of 100 MHz would be not enough to sample the LGAD signal using SERDES, it is possible to increase the sampling frequency by a factor 10, reaching 1 GHz.

With this sampling frequency, effective detection and processing of pulse signals similar to those illustrated in Figure 4.6, which are generated by the ABACUS chip, can be achieved. As explained in subsection 4.1.2, meeting the Nyquist-Shannon sampling theorem's requirements necessitates a minimum sampling frequency of 1 GHz. By utilizing SERDES in the aforementioned 1:10 configuration with a 100 MHz slow clock, these rapid pulses can be reliably detected.

Direct Memory Access - DMA

Once the data is collected and the acquisition is working properly, the most challenging task is to move the digital data from the FPGA to a safe storage location where it can be analysed by the end user. This problem is addressed by most FPGA vendors with the Zynq FPGA design. The Zynq architecture combines a Processing System (PS) and a Programmable Logic (PL) within a single chip. The PS is a hard processor system based on ARM Cortex-A9 cores, while the PL consists of reconfigurable logic fabric (FPGA) that can be programmed to implement custom digital circuits. The PS includes two ARM Cortex-A9 cores, along with various peripherals such as memory controllers, input/output interfaces, timers, and communication interfaces. The PS runs a full-fledged operating system and handles tasks that require high-level processing capabilities, such as running software applications, executing operating system functions, and handling external communication. On the other hand PL is a reconfigurable fabric that consists of configurable logic blocks, programmable interconnects, and various IP cores. The PL allows for the implementation of custom digital circuits tailored to specific application requirements. The PS and PL are connected through Advanced Extensible Interface (AXI) buses, enabling seamless communication and data transfer between the two domains. This allows for efficient integration of software and hardware components, enabling the system to leverage the processing capabilities of the PS and the flexibility of the PL to achieve optimal performance. Overall, the Zynq architecture with its PS and PL combination offers a powerful and versatile platform for implementing a wide range of applications, combining the flexibility of programmable logic with the processing capabilities of the ARM Cortex-A9 cores.

To serve the specific purpose of sending data from PL to PS and vice versa in a few clock cycles, I relayed in the Direct Memory Access or (DMA). Historically, DMA was designed to facilitate the transfer of data to specific memory locations without requiring the CPU to handle the entire process, thus preventing overloading and time-consuming operations. CPU cycles are only required to initiate the memory transfer by providing the DMA controller with the necessary instructions and parameters, such as the source and destination addresses and the amount of data to be transferred. The DMA controller then performs the data transfer independently, freeing the CPU for other tasks. Today, the DMA controller is an important part of a modern motherboard.

It is possible to use DMA to transfer data from the PL to the PS and vice versa. The requirements to make DMA work are different depending on the design architecture, namely whether the PS is running an operating system (OS) or not. In general, depending on the overall readout demands, the implementation may differ in running a Linux-based OS, or not running an OS and relying on the library provided by the board developer (Xilinx) to perform basic operations. As a more flexible choice, I decided to run a Linux OS on the PS. The disadvantage is that more effort is needed to get the OS to detect and control DMA. To ensure that the OS effectively manages the DMA in the PS section of a system, a kernel driver is implemented. This driver acts as an interface between the hardware and the OS, allowing the OS to control and manage data transfer operations. By implementing a kernel driver, the system achieves maximum flexibility, as all operations can be performed within

the so-called kernel space without restrictions. This means that the OS has direct access to hardware resources, including DMA, enabling efficient management and utilization of these resources. However, it is important to note that with such freedom comes the risk of potential hardware damage if incorrect operations are performed. The kernel driver must be carefully designed and implemented to ensure proper handling of the hardware and prevent any harmful operations. In Linux-based operating systems, user programs typically run in the so-called user space, which has restrictions in place to prevent unauthorized or dangerous operations that could potentially harm the hardware. This separation between the user space and the kernel space provides a layer of protection, ensuring that only authorized actions are allowed to interact with the hardware. By utilizing a kernel driver and enforcing restrictions in the user space, the system achieves a balance between flexibility and safety. The kernel driver enables efficient control and management of the DMA in the PL, while the user space restrictions prevent accidental or malicious operations that could lead to hardware damage.

To implement a working DMA from PL to PS, the user must: package the data according to the AXI protocol. Correctly send the data to a DMA IP core provided by Xilinx [Xil, a] and develop a kernel driver to handle the DMA operations. In this particular implementation, the DMA IP core was configured in scatter-gather (SG) mode in order to optimize and achieve the best performance. The scatter-gather DMA is an enhanced version of the traditional DMA technique, designed to efficiently handle data transfers involving non-contiguous memory buffers. In a traditional DMA transfer, data is transferred between a source and destination in a linear manner, where the DMA controller retrieves data from a single source address and stores it in a contiguous block of destination memory. This approach works well when the data is stored in a continuous memory region. However, in scenarios where data is scattered across multiple non-contiguous memory regions, traditional DMA becomes less efficient. This is where scatter-gather DMA comes into play. In scatter-gather DMA, the data transfer process is divided into smaller transactions, each involving a specific memory buffer. The scatter-gather DMA controller uses a descriptor table, often called the Scatter-Gather List (SGL), to specify the source and destination addresses for each transaction. The SGL consists of individual descriptors, each containing information about a specific memory buffer, such as its address and length. The descriptors are arranged in a linked list, allowing the DMA controller to traverse through them and perform the data transfers. When a scatter-gather DMA transfer is initiated, the DMA controller starts with the first descriptor in the list, retrieves data from the corresponding source address, and stores it in the destination memory. Once a transfer is complete, the DMA controller moves to the next descriptor in the list and repeats the process until all descriptors have been processed. This scatter-gather approach enables efficient data transfers involving non-contiguous memory regions, as the DMA controller can automatically handle the scattered data without requiring intervention from the CPU. It eliminates the need for software intervention to manage the individual memory buffers, resulting in reduced CPU overhead and improved data transfer efficiency.

Scatter-Gather-Direct Memory Access (SG-DMA) (hereafter referred to as DMA) requires that the input data follow the specific AXI Stream protocol. The AXI Stream protocol provides unidirectional data flow, where data moves only from a source (producer) to a destination (consumer). It does not support bi-directional data transfer. Data is transferred in bursts of fixed length without explicit addressing. The burst length is predefined and data is transferred continuously until the burst is complete. Unlike other AXI protocols, AXI Stream does not use explicit addressing. Instead, the data stream itself carries payload data and control signals:

- TVALID (Transmit VALID): Indicates that valid data is available from the source.
- TREADY (Transmit READY): Indicates that the destination is ready to accept the

data.

- TLAST (Transmit LAST): Indicates the end of a burst or packet of data.

The AXI Stream protocol operates on a data-ready basis. The source component asserts the TVALID signal when valid data is available, and the destination component asserts the TREADY signal when it is ready to receive the data. Data transfers occur when both TVALID and TREADY are asserted simultaneously. The TLAST signal is used to indicate the end of a burst or packet of data. It helps the receiver identify the boundaries of the data stream, allowing proper parsing and processing of the incoming data. A scheme of the such transfer is shown in Figure 4.13. Usually, the DMA, if correctly configured is ready to receive the data so the TREADY signal is raised to high logical level. Once the data is ready to be sent, the TVALID signal is also raised to the high level. Again, sampling of the signal occurs always at the positive edge of the ACLK and so the transfer occurs at the rising edge of T3 in Figure 4.13 when both TVALID and TREADY are asserted.

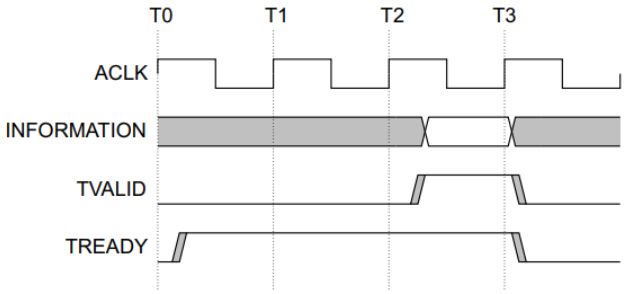


Figure 4.13: Example of when data can be transmitted according to the AXI stream protocol. Both TVALID and TREADY must be asserted simultaneously. If this condition is met, the data is transmitted on the positive edge of ACLK as on the T3 rising edge.

By employing a data stream following the AXI stream protocol, I can transmit data to the DMA, which can then directly write this data to memory. In the context of Linux running on the PS, users can access the memory sections containing the data transmitted by the PL.

4.2 Implementation

In this section, I present the solution I have successfully implemented on the ZC702 development board. I begin by focusing on the Programmable Logic (PL), where data processing and collection take place. Subsequently, I will go into the Processing System (PS), which operates with a Linux Operating System (OS) to efficiently manage the data from the PL.

4.2.1 Programmable Logic - PL

Programmable logic (PL) is where the user can implement custom operations. Because the code the user writes is actually converted into a logic circuit, timing and signal propagation are well known. This is one of the main reasons why PL hosts time-critical operations. The user can actually implement real-time logic while keeping the required time of these operations under strict constraints. In addition, the PL can be programmed to perform custom operations such as creating an AXI stream of data. All of this came to the coast that the programming language is typically at a very low level as the code has to be translated into an actual logic circuit.

As in the previous sections, I follow a bottom-up approach to explain the data processing flow. Firstly, I illustrate the data processing using the SERDES. Next, I discuss the data handling to the Processing System (PS) through Direct Memory Access (DMA). Finally, I explore the modules responsible for controlling the thresholds of the ABACUS and the LTC2602. All PL development and implementation have been performed with Vivado 2022.2.

Data processing

In the specific case of detecting the signal from ABACUS, to achieve a sampling frequency of at least 1 GHz, the SERDES primitive has to be carefully configured. The slow clocking frequency was selected at 100 MHz and SERDES used in $1:10$ configuration with Double Data Rate (DDR) option. Double Data Rate (DDR) is a memory technology that allows for the transfer of data on both the rising and falling edges of a clock signal, effectively doubling the data transfer rate compared to single data rate (SDR) memory. This is achieved by using a technique called "data strobing", where the data is aligned with the clock edges and captured on both transitions. Thus, in a DDR configuration, a fast clock of 0.5 GHz allows sampling at 1 GHz, the minimum sampling frequency required to detect ABACUS signals. A single SERDES primitive is not sufficient to achieve a 1:10 ratio. Two SERDES primitives in master-slave configuration are required, as a single primitive can cover [0:8] parallel output[1]. The slave SERDES in cascade configuration will handle the remaining [9:10] parallel output.

Assuming correct signal processing, a 10-bit bus is read for each channel on each slow clock cycle. Each ABACUS digital pulse is then recognizable by a $0 \rightarrow 1$ transition of the bus. In the example of Figure 4.12, the output bus is $0,0,0,1,0,0,0,0,0,0$ and there is only a $0 \rightarrow 1$ transition from the second to the third bit. In general it is required from the FPGA to detect such transition within one clock cycle, and, depending on the number of transition, add them up to a cumulative counter. This is implemented using a LookUp Table (LUT). LUT is a fundamental component in the context of FPGAs. It is a configurable memory element that allows for the implementation of logic functions. A lookup table essentially stores a truth table, where the inputs to the table correspond to specific memory addresses, and the outputs represent the logic function values associated with those addresses. When programming an FPGA, designers can define the desired logic function by configuring the values stored in the lookup table. By properly configuring the lookup table, complex logic functions can be realized within the FPGA, providing flexibility and versatility in designing digital circuits. To overcome the counting issue within a clock cycle, a special LUT capable of counting the number of $0 \rightarrow 1$ transactions has been implemented. Note that special care must be taken when dealing with transitions at the edge of the sequence. If we consider an arbitrary SERDES sequence with the first bits at high level ($a_n = \{1,0,\dots 0,0\}$) as output, it is not possible to determine where the first bit of the transition $0 \rightarrow 1$ comes from. Additional information consisting of the previous last bit is required as shown here where two successive output sequence of the SERDES are presented:

$$\{0,0,\dots \overbrace{0,1}^{\text{transition}}\}\{1,0,\dots 0,0\}\ 1 \text{ transition identified correctly} \tag{4.1}$$

$$\{0,0,\dots 0,0\}\{1,0,\dots 0,0\}\ 1 \text{ transition not identified correctly} \tag{4.2}$$

It is important to note that two or more consecutive high digital levels can be sampled by SERDES. This situation can occur if the length of the signal from the ABACUS is longer than

[1] The [n:m] notation is typical of the Verilog programming language and indicates a bus of $m-n$ elements, e.g. $[0:8]$ indicates a bus of 8 elements labeled with numbers from 0 to 7 with rightmost element as most significant bit.

the sampling period of the SERDES. One solution is to always store the last bit of the preceding sequence and extend the lookup tables to the 11 bits. With this method, all necessary information is always available, and signal transitions are processed correctly. Generating the LUT can become a tedious process since, being a truth table, all the possible outcome have to be addressed. One way to avoid manually implementing all $2^{11} = 2048$ possible combinations of 11 bits is to use metaprogramming techniques. Specifically, a programming language can be used to generate the necessary LUT code based on a set of rules or conditions. In this scenario, a Python3 script was developed to generate the required LUT in the Verilog language. The resulting code was then integrated into the FPGA project as part of the overall design. This approach of using a program to generate code is not limited to this specific case. It is a common practice to employ code generation tools or scripts to automate the creation of repetitive or similar code segments with minor variations. By utilizing metaprogramming techniques, developers can significantly reduce manual effort, improve code consistency, and improve overall productivity.

To count the number of transitions from the output of SERDES in the deserializer configuration, a counter register with a width of 32 bits is used. This architecture must be implemented for each FMC input dedicated to counting ABACUS data. To facilitate this process and to avoid redundant manual coding, a special Python3 script is used, again, making use of metaprogramming techniques.

DMA implementation

A Verilog module was created to facilitate efficient and orderly data transmission. This module is responsible for packaging the counter data according to the AXI stream protocol. After the data is properly packaged, it can be sent to the DMA, which in turn sends the data to the PS for further processing.

The functionality of this module can be effectively represented by a finite-state machine (FSM) diagram. To enhance the readability and maintainability of the code, two separate FSMs were implemented to manage the DMA data transfer process. The primary FSM is responsible for resetting, and initiating the overall process, while the secondary FSM is triggered by the primary FSM and oversees the actual data transfer. The primary FSM, which is responsible for controlling state transitions, is shown in Figure 4.14. Under normal conditions, the state machine remains in the idle loop, waiting for either a run request command or a reset command. This approach ensures that the data transfer process is properly controlled and synchronized, allowing efficient and reliable DMA operations. When a run request is sent, the FSM changes to `start` and triggers the second state machine shown in Figure 4.15. The latter FSM is responsible for the actual data transfer in compliance with the AXI stream protocol.

Before transmitting the data, I used multiple latches to ensure synchronous reading of all channels. A latch is a basic circuit element in digital electronics that stores a single bit of data. It operates on the principle of latching or capturing the input value and holding it until a subsequent signal is received. Latches are commonly used to synchronize and align data signals in digital systems. In this context, the latch ensures that the data from all channels is sampled at the same time, thus enabling a consistent and accurate data transfer.

When a reset command is received, all logic counters and latches are forced into the initialization state. When the reset is complete, the FSM automatically returns to the idle state, ready to accept a new request. Not all conditions are explicitly shown in Figure 4.15. In fact, the implementation of the code has been carefully designed to address potential issues that may arise during DMA transfers, such as failures or incomplete transfers. To mitigate these challenges, various measures have been taken. Firstly, a timer is incorporated into the

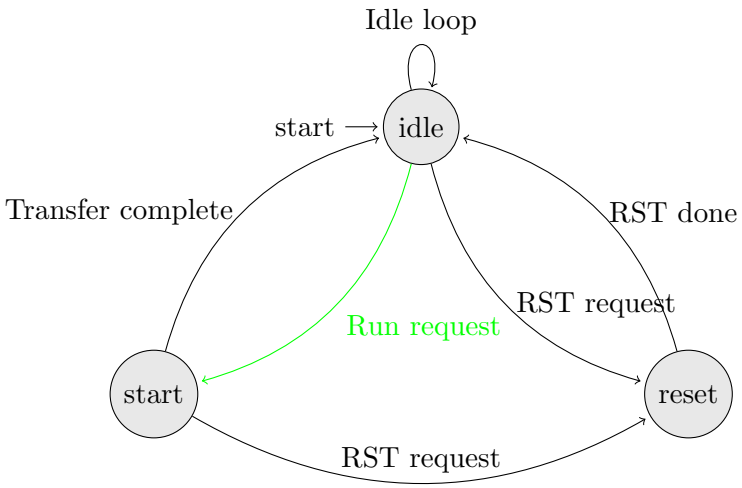


Figure 4.14: Primary FSM diagram designed and implemented in the FPGA logic. This state machine is responsible for the resetting and the initialization of the write process.

system to automatically reset the state machine in the event of an initialized and aborted transfer. This proactive approach helps prevent potential deadlocks and ensures the system remains responsive. Additionally, real-time monitoring is implemented to verify the readiness of the primary state machine for initializing data transfers. This real-time check helps optimize efficiency by avoiding unnecessary delays and maximizing overall throughput.

To maximize DMA data throughput, the most efficient configuration was achieved by sending data chunks of 128 bits at each DMA transfer. This implies that at each positive edge of ACLK, if TREADY and TVALID are asserted simultaneously, 4 counters of 32 bits each are sent to the PL using the following encoding format:

$$\overbrace{\underbrace{C1_{31}, C1_{30}, \dots C1_1, C1_0}_{\text{channel 1}}, \underbrace{C2_{31}, C2_{30}, \dots C2_1, C2_0}_{\text{channel 2}}, \underbrace{C3_{31}, C3_{30}, \dots C3_1, C3_0}_{\text{channel 3}}, \underbrace{C4_{31}, C4_{30}, \dots C4_1, C4_0}_{\text{channel 4}}}^{\text{128 bits}} . \tag{4.3}$$

CM_n is defined as the n^{th} bit of the counter associated with channel M. In the specific case of LGADs with 71 strips, the total number of transfer or ACLK clock cycles is 16. The counter registers are sent in order from 0 to 70, and in Equation 4.3 the first packet of 4 counters is represented; further 15 data transfers of the same encoding are required to send all counter values. To avoid wasting the last 32 bits of data stream of the 16th cycle that are not associated to any channel, a timestamp is sent as well. This timestamp is stored in a 32 bit counter register incremented at each ACLK clock cycle. The latter is treated in an analogous way of any other LGAD channel: by latching the value when data transfer is requested. With this additional information, it is possible, if the ACLK clock frequency is known, to have a time correspondence with reasonable accuracy: assuming a frequency of ACLK of 100 MHz, the time resolution is 10 ns. If such a resolution is not sufficient, more complex methods beyond the scope of this work are required, but for a specific Hybrid Detector for Microdosimetry (HDM) application it is considered sufficient. Running an incremental counter at this counting frequency with only 32 bits, on the other hand, will result in an overflow occurring after $2^{32} - 1 = 4294967295$ clock cycles or every $(2^{32} - 1)/100\,\text{MHz} \simeq 42.95\,\text{s}$. If handled correctly, overflows are acceptable and can be corrected in data analysis by identifying the occurrence and correcting the register value by adding 2^{32}. Naturally, the counter must be stored in a 64 bit unsigned integer variable.

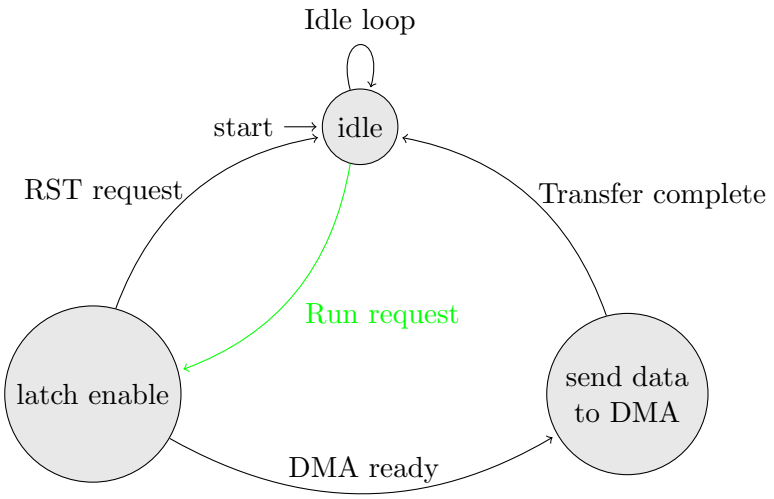


Figure 4.15: Secondary FMS diagram designed and implemented in the FPGA logic to handle the latching of the counters and the actual DMA data transfer.

On the PS side, a Linux kernel driver was developed as interface to the hardware device. The DMA was treated as a character device, which involved implementing basic system calls such as `open()`, `read()`, `write()`, and `close()` to provide the necessary functionality for user space applications. The implementation of the kernel driver follows a common approach to mapping device memory using structures or structs (`struct` types in C or C++). These structures provide a convenient way to describe and organize the properties and configuration of the hardware device within the driver code. Within this driver, three primary structs have been implemented to handle different aspects of the device:

- The first struct was responsible for managing the Scatter-Gather List (SGL), a technique used in DMA transfers to efficiently process data from multiple non-contiguous memory regions.

- The second struct contained essential information about the device, including its name and the number of DMA channels it supports. This struct also hosts the third struct.

- The third struct served as the main struct for the DMA functionality. It contained fields required by the kernel to correctly identify the device class, such as the physical address of the device, channel buffers for data transfers, the SGL, and the direction of the DMA transfers.

To control DMA in user space, a common approach is to provide a struct that is mapped into user space memory as a user-to-kernel interface. The following struct has been developed to accomplish this task.

```
struct channel_buffer {
    unsigned int buffer[N_TRANSFERS];
    enum proxy_status {
        PROXY_NO_ERROR = 0,
        PROXY_BUSY = 1,
        PROXY_TIMEOUT = 2,
        PROXY_ERROR = 3
    } status;
}
```

The array `buffer`, consisting of `unsigned int` of 32 bits, stores the actual data values sent by the PL component. Each element in the array represents the data from a specific channel, with the index corresponding to the n-th channel. In the case of reading 71 channels, the total number of transfers, `N_TRANSFERS`, is set to 72, where 71 transfers are for the channels and 1 transfer is for the timestamp. To facilitate the decoding of DMA status and identify potential errors, the `proxy_status` enumeration type is used. This enumeration provides a set of predefined values that can be used by the user to interpret the DMA status and handle any exceptional conditions that may occur during the data transfer process.

This struct is then contained in the following struct:

```
struct channel {
        struct channel_buffer *buf_ptr;
        int fd;
};
```

Where `fd` is the file descriptor pointing to the DMA device. A file descriptor is a unique identifier that is assigned to a file or device when it is opened by a process. It is an abstract representation of the opened file or device, allowing the process to interact with it using various system calls. The file descriptor can be opened in read-only mode (`O_RDWR`) with the following line of code:

```
struct channel rx_channel;
rx_channel.fd = open("/dev/dma_proxy_rx", O_RDWR);
```

`/dev/dma_proxy_rx` is the system path where the DMA device is created by the kernel driver once is loaded. By using this nested structs, multiples channels can be opened e.g., in the case of both a reciving and a transmitting channel. In addition to that, multiple channel buffer can be mapped, in the following example, the mapped memory span `RX_BUFFER_COUNT = 32` meaning that 32 independent copies of the struct `channel_buffer` are available.

```
rx_channel.buf_ptr =
    (struct channel_buffer *) // casting
        mmap(
                NULL,
                sizeof(struct channel_buffer) * RX_BUFFER_COUNT,
                PROT_READ | PROT_WRITE,
                MAP_SHARED,
                rx_channel.fd,
                0
            );
```

Increasing the buffer count in the shared memory between the PL and PS has the potential to enhance data transmission speed. With a larger buffer count, more data can be stored and transmitted concurrently, allowing for faster processing. However, it is important to note that when the buffers reach maximum capacity, additional clock cycles may be necessary to free up memory for new data. This can introduce some latency in the system. Thus, while a higher buffer count can improve overall throughput, it may also introduce additional overhead in managing the memory,

The user can command the DMA transfers with the `ioctl` (an abbreviation of the input/output control) function. This function, implemented in the kernel drivers, requires: the file descriptor, the command, and an integer to select which channel buffer is in use.

```
ioctl(rx_channel->fd, START_XFER,  &i); //start a DMA transfer
ioctl(rx_channel->fd, FINISH_XFER, &i); //wait trasnfer complete
```

The first instruction, which passes the argument `START_XFER`, initiates a DMA transfer. This instruction is non-blocking, meaning that the program continues executing without waiting for the transfer to complete. The responsibility of initiating the transfer lies with the DMA itself, which will start it once it is ready. To keep track of the execution and status of the DMA

transfer, a cookie can be used. The cookie serves as a unique identifier or token associated with the transfer. It allows the program to later query the DMA or check the status of the transfer, providing a way to synchronize or monitor the progress of the transfer if needed. The other instruction, which takes the argument `FINISH_XFER`, waits for the completion of a previously submitted DMA transfer. Unlike `START_XFER`, this instruction is blocking, meaning that the program will pause execution until the transfer is complete. By using the blocking `FINISH_XFER` instruction, the program ensures that it does not proceed further until the transfer is finished. However, to avoid potential deadlock situations, a timeout timer is implemented. If the transfer takes too long and the timeout timer expires, an error is triggered, allowing the program to handle the situation accordingly. For both function implementations, the variable `&i` is the index of the channel buffer, as mentioned above, multiple channel buffers can be run independently.

For the simplest DMA transfer that loop-back data in memory, the user should write a function like this:

```
int simple_transfer(
        struct channel * tx_channel,
        struct channel * rx_channel,
    ){
    // use one buffer only
    const int buffer_id_tx = 0;
    const int buffer_id_rx = 0;

    // set the value to send
    // on buffer 0
    tx_channel->buf_ptr->buffer[buffer_id_tx] = 42;

    const int i = 0;

    // non blocking
    ioctl(rx_channel->fd, START_XFER, &buffer_id_rx);

    // bloking
    ioctl(tx_channel->fd, FINISH_XFER, &buffer_id_tx);

    // print recived data
    std::cout << "recived␣=␣" <<
        rx_channel->buf_ptr[0].buffer[0] << std::endl;


    return 0;
};
```

By designing the kernel driver and the user space interface carefully, the process of performing a DMA transfer can be simplified to just a few lines of code. The complexity of managing the SGL and the DMA itself is abstracted away from the user, as it is handled by the kernel driver. This abstraction allows user space programs to easily initiate DMA transfers without having to worry about the intricacies of the underlying hardware.

The flow of data in the system is illustrated in Figure 4.16. The flow of data is depicted by the magenta arrows. Starting from the counters, the data is processed and serialized using the AXI streaming protocol by a custom Verilog module called the DMA controller. Once the data are in this format, they can be handled by the DMA IP core. The DMA operation also requires instructions from the PS side. These instructions include the configuration of the registers and the SGL. The Linux kernel plays a vital role in managing these instructions through the implementation of a kernel driver. When properly configured, the DMA can directly write data to the memory. The user has control over DMA functionality through the kernel, utilizing the provided `ioctl` commands. Additionally, the user can access the memory where the DMA data were transferred, again through the kernel. All commands

are represented by blue arrows. Overall, this system architecture allows for efficient data flow and control between the various components, enabling seamless integration of the DMA functionality into the overall design of the system.

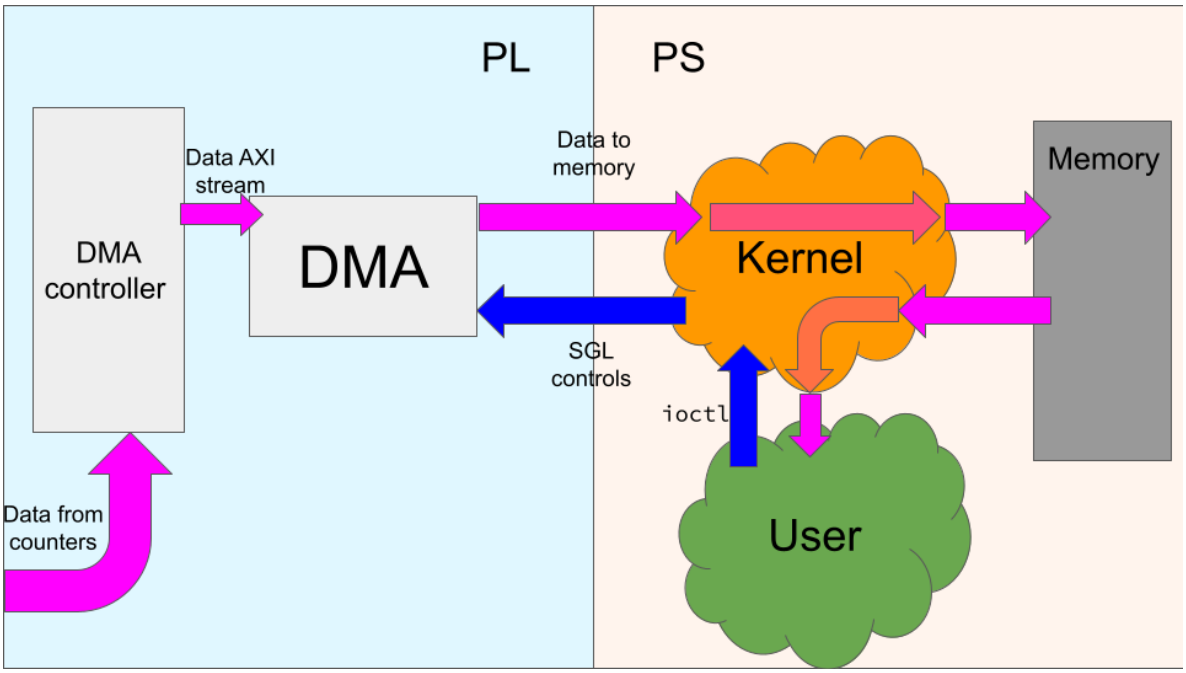


Figure 4.16: Block diagram showing the transfer of data from the meter register to the user. The data flow within the system is represented by the magenta arrows. Data from the counters are serialized to the AXI streaming protocol thanks to a specialized Verilog module called the DMA controller. The AXI streaming data are compatible with the DMA IP core. To operate the DMA IP, instructions are required from the PS side to configure registers and provide the SGL. The Linux kernel handles these instructions through a driver. Once the DMA is configured, it can write data directly to memory. Users can control the DMA through the kernel, using the implemented `ioctl` commands. Instructions are shown as blue arrows. Finally, the kernel provides the user with access to the memory where the DMA data were transferred.

It should be stressed that the whole process of implementing a working kernel driver, but in general producing programs on the PS of the ZC702, requires cross-compilation. Cross-compilation is required when developing software for a target platform that has a different architecture or operating system from the development machine. In the case of the ZC702 board and the Linux kernel, cross-compilation is necessary because the development machine typically runs on a different architecture (e.g., x86) than the Zynq SoC (ARM-based). Cross-compilation involves setting up a toolchain specifically designed for the target platform. This toolchain consists of a compiler, linker, and other necessary tools that can generate executable code compatible with the target platform architecture. The toolchain ensures that the compiled code is compatible with the target system's processor and operating system. By using a cross-compiler, developers can write and compile code on their development machine and generate executable binaries that can be run on the target platform. This allows efficient software development without the need to directly compile code on the target device. Furthermore, in these applications where an OS is utilized on top of an FPGA, the OS is often streamlined to maximize performance and minimize resource usage. This involves reducing unnecessary components and functionalities typically found in a general-purpose PC operating system.

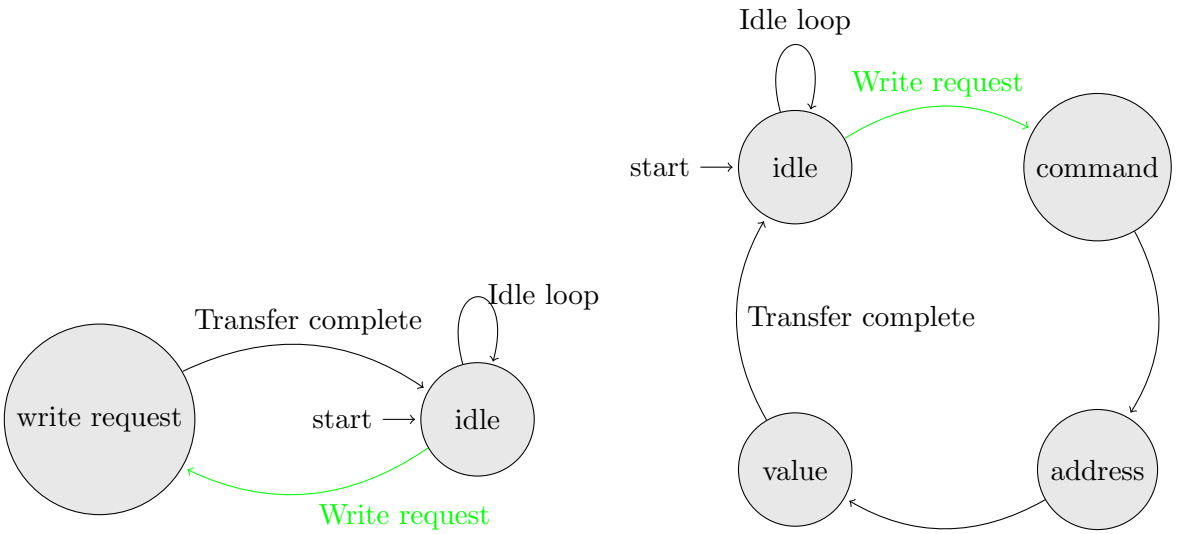
The first reason for keeping the OS lightweight is to optimize performance. By removing unnecessary components, the system can allocate more resources to the critical tasks and processes required for the specific application. This can lead to an improved overall performance and response of the system. The second reason is to minimize resource consumption. FPGA-based systems often have limited resources, such as memory and processing power. As a consequence, the OS resources are optimized for the specific task, but in general, if not specifically provided, lack general debugging tools and text editors such as the GNU Project debugger (GDB).

Threshold controls

In the previous subsections (subsection 4.1.2 and subsection 4.1.3), I explored the communication protocols necessary to manage the threshold discrimination levels of the LGADs. To address the protocol requirements, I developed two distinct Verilog modules. The first module, called "DAC_external_controller" (DEC), is designed to control the thresholds of the LTC2602. The second module, named "DAC_internal_controller" (DIC), is responsible for managing the thresholds of the ABACUS. These Verilog modules provide the necessary functionality to handle the threshold settings of both the LTC2602 and the ABACUS, enabling efficient control over LGAD threshold discrimination.

Again, FSM diagrams prove to be an effective way to represent the functionality of Verilog modules. For both the DEC and the DIC modules, the design philosophy is to implement two FSMs: a primary and a secondary. The primary FSM is responsible for managing user requests and triggering the secondary FSM when necessary. Secondary FSM is solely focused on handling the communication protocol. Each payload component required for serial transmission is represented as a state in the FSM. As each payload component is transmitted, the FSM seamlessly transitions to the next component. Furthermore, this design approach allows for targeted debugging and troubleshooting. If a bug is discovered in a specific state of the FSM, it is possible to address and correct that particular state without needing to modify the entirety of the program structure. This modular design greatly simplifies the debugging process and improves the overall maintainability of the codebase. Developers can focus on isolating and resolving issues in a specific state, minimizing the impact on the rest of the program's functionality. Another design constraint requires the modules to accept the minimum necessary input parameters in a parallel form (bus). The other option is to use a serial protocol, which is generally avoided in FPGA design. One of the few exceptions may be when communication resources within the modules are limited. A serial design architecture helps in reducing the required I/O pins for each module. In the specific application of Hybrid Detector for Microdosimetry (HDM), the selected FPGA has been proven capable of implementing and sustaining the required design, and therefore no such limitation needs to be considered in code development.

Both DEC and DIC share the same design methodology, resulting in a similar modular architecture. Starting with the DEC module, a typical SPI payload contains the elements shown in Table 4.3: the command, the DAC address, and the DAC value. The implementation of the primary and secondary FSMs is shown in Figure 4.17. Once a write request is received, the primary state machine goes from the `idle` loop to a `write request` state. The secondary state machine follows analogously, initializing the communication by preparing the command. The SC/LD output is indirectly controlled by the primary FSM since the timing diagram requirements of the LTC2602 require a minimum of 7 ns to elapse between the initialization of the SC/LD to low level and the first data transfer. To meet this requirement, the primary FSM ensures that the SC/LD is set to the required low level before the secondary FSM is allowed to advance from `idle` to any state. Once triggered, the secondary FSM remains



(a) Primary FMS in charge of user request ad secondary FMS activation.

(b) Secondary FMS in charge of communication payload

Figure 4.17: Primary and secondary FMS implemented in the DEC module to communicate with the LTC2602.

in the `command` state for the duration of 4 SPI Clock (SCK) cycles. During this time, the output values are updated on the negative edge of the serial clock as they are sampled by the LTC2602 on the rising edge. This gives the electronics enough time to set the value before sampling. Similarly, 4 SCK cycles are used to transmit the address in the `address` state. An additional 16 SCK cycles are then used to transmit the actual DAC value in the corresponding state. Once transmission is complete, both FSMs transition back to the `idle` loop state, ready for subsequent instructions or requests. An example of transmission is shown in Figure 4.18 where the three transmission states are highlighted in different colors. SCK, SDI and SC/LD are shown as well.

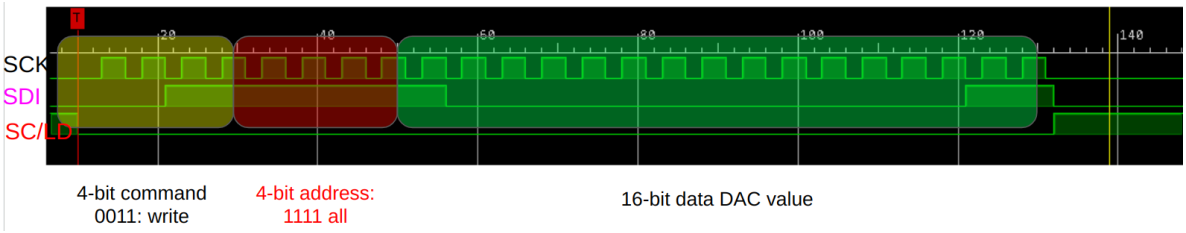


Figure 4.18: Actual DEC SPI communication from the FPGA to the LTC2602. The SCK, SDI and SD/LD lines are shown. Different communication payloads are highlighted in different colors to reflect their corresponding states in the FSM diagram Figure 4.17. The first 4 SCK cycles send a 0b0011 corresponding to a write command. The next 4 bits sent correspond to the address, selecting both outputs by sending 0b1111. The last 16 bits represent the value. In this case 0b1000000000000011 = 32771 corresponding to a threshold value of 0.6V. Note how the SDI value is updated on the falling edge of SCK.

DEC ports are shown in Figure 4.19. The module design was specifically chosen to ensure that all instructions are synchronized with a clock frequency of 100 MHz provided by the clk_in port. This is a commonly adopted design choice and helps to avoid potential synchronization

issues with other modules in the system. The user can customize the SPI communication speed by providing the desired SCK frequency to the module. The maximum value tolerated by LTC2602 is 50 MHz. However, since the threshold adjustment procedure does not have strict timing requirements and does not need to be performed at the highest possible speed, a more relaxed clock frequency of 100 kHz is provided. This choice is intended to support the integrity of the signal by reducing the likelihood of data communication errors. The desired SCK clock is supplied to the module via the clk_serial port. The write_now input serves as a trigger to initiate the write process. When a pulse is detected on this input, the module gathers the necessary inputs from the remaining ports. The [15:0] DAC_value bus represents the actual value to be transmitted, while the [0:3] command and [0:1] channel determine the command and address selection, respectively. For address selection, 0b00 denotes even output, 0b01 denotes odd output, and 0b11 denotes all output. The output ports spi_clk, spi_data and nspi_enable correspond to the SCK, SDI and SC/LD respectively. They can be connected directly to the desired LTC2602. A write_done flag is also exported. This allows the user to keep track of the module status.

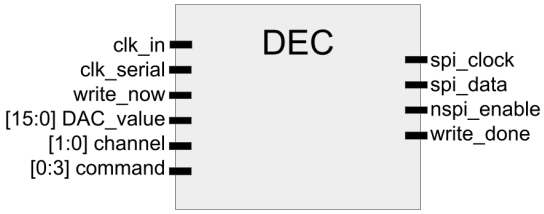
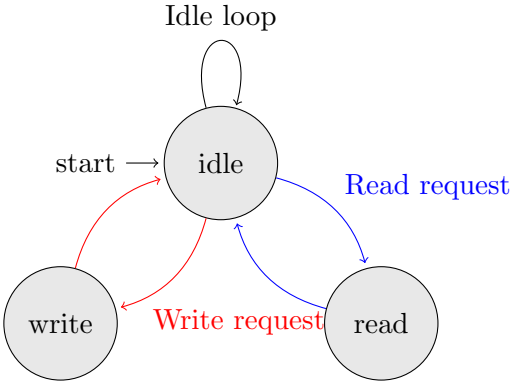


Figure 4.19: DEC input and output ports. The module takes in the desired SCK clock through clk_serial and initiates the write process when write_now is triggered. Inputs such as DAC_value, command, and address determine the transmitted data. The output ports spi_clk, spi_data, and nspi_enable correspond to SCK, SDI, and SC/LD, respectively, enabling direct connection to the LTC2602. Additionally, a write_done flag is provided to monitor the module's status.

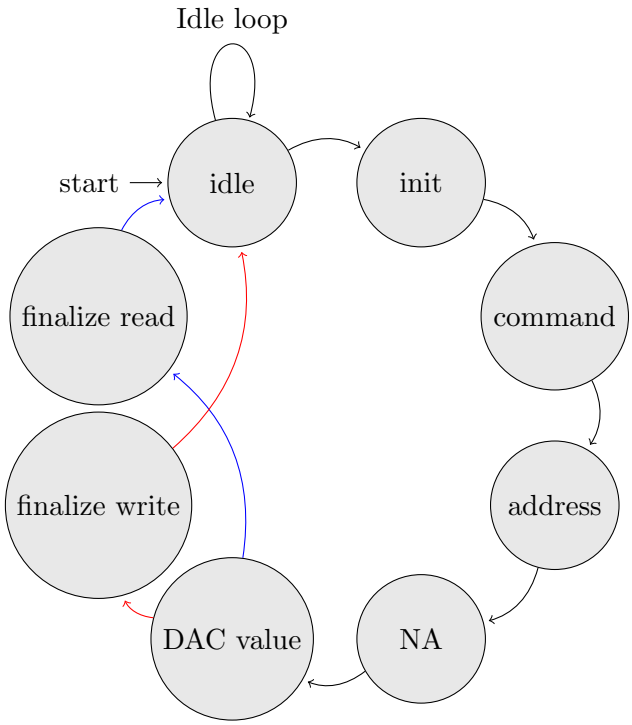
Overall, this module serves as a crucial interface between the user and the LTC2602. By appropriately configuring the input register of the DEC module, data can be transmitted in accordance with the specified serial format. This allows users to effectively control and adjust the threshold levels of the LTC2602.

The DIC communication is more complex as both read and write operations are implemented. In addition to that, the payload size is larger, as shown in Table 4.1. I start with the FSM diagram in Figure 4.20, showing the primary and secondary FSM. Normally, the primary FSM, is in the `idle` loop. The user can then request write or read operations. In the case of write operations, the FSM follows the red arrows into the `write` state and stays there until the operation is completed. Similarly, if a read operation is requested, the FSM will follow the blue arrows into a `read` state. Once the operation is complete, the machine will return to the `idle` loop. The secondary state machine diagram follows the same design philosophy as the one adopted for DEC: one state is implemented for each component of the payload. It is worth mentioning that since the ABACUS chip sequence has no specific time constraints, the output is always set to the desired level on the negative edge of the previous CLK cycle to avoid errors in the command interpretation. This will give the electronic circuitry enough time to set the value before the value is read. In the special case of communication initialization, which is handled in the `init` state, the secondary state machine ensures that the value is set at least half a CLK period in advance. This state is also responsible for

sending the initialization sequence 0x$A5A5$ =0b1010010110100101. Next, the `command` state, depending on whether a read or write request is submitted, sends the appropriate data: 0b01 for read while 0b11 for write. Next, the `address` state is responsible for sending the 5-bit data associated with the address of the channel to be written or read. After that, the `NA` state simply populates the unused payload sequence. Finally, the `DAC Value` state sends the



(a) Primary FMS in charge of user request ad secondary FMS activation.



(b) Secondary FMS in charge of communication payload.

Figure 4.20: Primary and secondary FMS implemented in the DIC module to communicate with ABACUS chip.

6 bits representing the new channel threshold. In the case of write operations (red arrow), the module enters the `finalize write` state where the 16-bit fill sequence is sent. While, in the case reading (blue arrow), the module enters the `finalize read` state where the entire payload of the response sequence, shown in Table 4.2, is stored in memory within the module. To complete the sequence, regardless of the required read or write task, the state machine returns to the `idle` loop. By default, the ABACUS chip discards DAC values during read operations. Therefore, the `DAC value` state can be used universally for both read and write operations, eliminating the need for a dedicated state. This design approach saves resources during module synthesis and ensures a consistent flow regardless of the operation that is being performed.

A sequence of successful writing sampled directly from the FPGA to the ABACUS is shown in Figure 4.21. Both the FPGA SIN and CLK are shown. The sequence is designed to program channel 0b00001 = 1 with the value of 0b000001 = 1 DAC units.

In the case of read operations, the sequence sent is shown in Figure 4.22. CLK, SIN and SOUT are shown. The sequence requests to read channel 0b01111 = 15. The ABACUS answers the request by sending two bits 0b11 according to the table Table 4.2 to confirm the correct interpretation of the command. Next, we have the confirmation of the address:

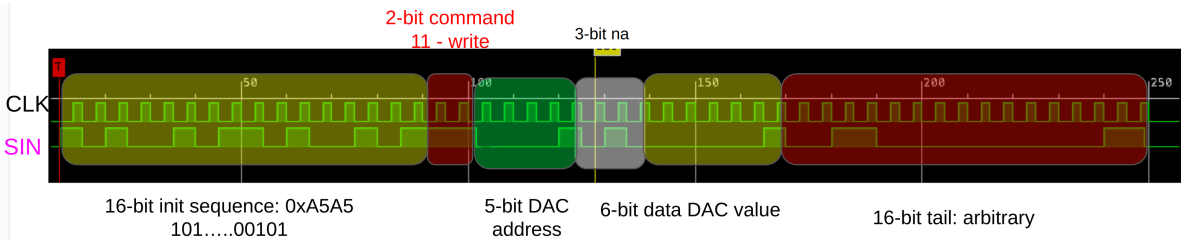


Figure 4.21: Write sequence sent from the FPGA using the DIC module. Different states of the FSM are highlighted in different colors. Both CLK and Serial INput (SIN) lines are shown. This sequence programs the channel 1 to a value of 1 DAC units. According to Table 4.1, the first initialization sequence is sent, followed by the write command. The channel address is then sent. The next 3 bits are unused. The commands end with 6 bits containing the actual DAC value and 16 bits of arbitrary values as a tail. The total sequence then consists of 48 bits. It should be noted that the SIN values are updated on the negative edge of CLK, since the ABACUS acquires the SIN values on the rising edge.

0b01111 = 15, matching the one sent. Finally, after 3 bits of padding, the actual value of the DAC is sent: 0b111111 = 63 DAC units, the maximum possible value. Then, to complete the sequence, another 16 bits of arbitrary values, not shown in the figure, are sent.

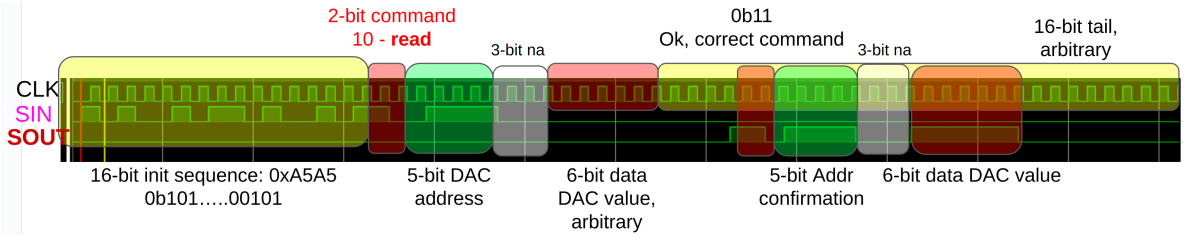


Figure 4.22: DIC read sequence according to Table 4.1 and Table 4.2. CLK, SIN and SOUT are shown from top to bottom. The sequence requests to read channel 0b01111 = 15. The ABACUS answers with two bits 0b11 to confirm the correctness of the command. The address confirmation follows: 0b01111 = 15, corresponding to the sent value. After three bits of padding, the actual DAC value is sent: 0b111111 = 63 DAC units.

The port scheme implemented in the DIC scheme is shown in Figure 4.23. As in the case of the DEC, the user has to supply the clock with a frequency of 100 MHz, which is used to synchronize the FSM via the clk port and the clock used in serial communication connected to the DAC_clk port. Again, the ABACUS has no explicit timing requirements, and to be conservative, I used a 100 kHz frequency clock as DAC_clk; in addition, the threshold setting of the ABACUS is not time constrained, so a relaxed clock frequency can be provided. This was found to be acceptable as Figure 4.21 and Figure 4.22 were obtained under these conditions. Since the DEC also needs a SCK, I connected the same 100 kHz clock line to optimize the overall design without having to create a new one. The port execute_command triggers the execution of the command when a signal pulse is detected. The module creates a payload using the values on [0:5] DAC_in_value as the actual DAC value and [0:4] DAC_channel as the channel address. A read or write command is sent depending on the value of bit_write_or_read: 0b0 to read, 0b1 to write. One DIC module can handle two ABACUS. This design was mainly adopted to optimize the use of each ESA_ABACUS control connector, since it is designed to connect two ABACUS chips. The user can choose to send the command to the even or odd output by setting the selector input accordingly: 0b0 to communicate with the even-labeled

ABACUS, 0b1 to communicate with the odd-labeled ABACUS. The ready output port is a flag the user can use to monitor the status of the module: if it is raised, the DIC is busy and any pulse sent through execute_command is discarded. Since the DIC can handle two ABACUS, two clock ports SDAC_clk_0 SDAC_clk_1, two data ports SDAC_data_0 SDAC_data_1 and two input ports SDAC_data_in_0 SDAC_data_in_1 are used to handle respectively CLK, SIN and SOUT of the ABACUS even (port name ending with 0) or odd (port name ending with 1). If the user requires a read command, the whole 16 bits output according to Table 4.2 is assigned to the DAC_out_value bus.

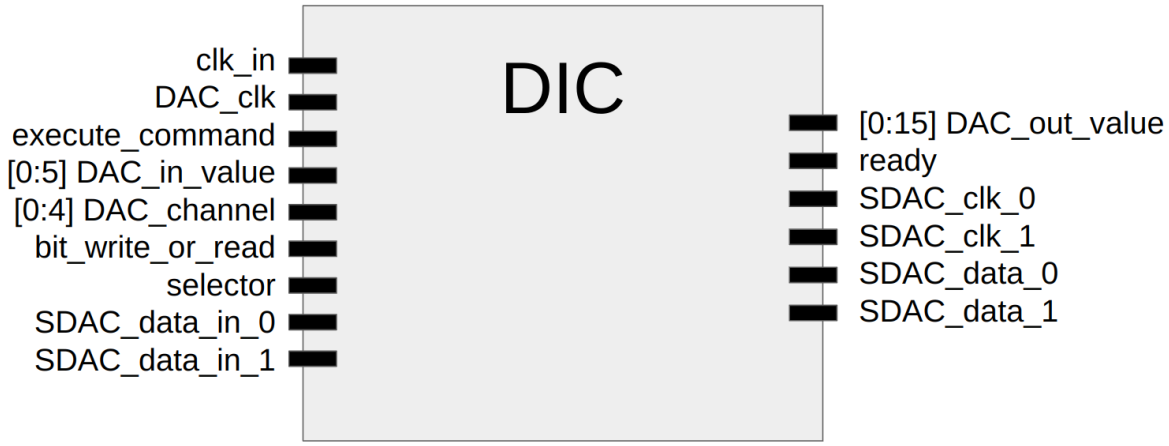


Figure 4.23: The DIC module input/output ports. Interfaces with the ABACUS chip and requires a 100 MHz clock via the clk port for FSM synchronization. The DAC_clk port, using a 100, kHz clock, handles the ABACUS serial CLK. The execute_command port initiates the command execution, generating a payload based on [0:5] DAC_in_value as the DAC value and [0:4] DAC_channel as the channel address. The bit_write_or_read input dictates whether to read (0b0) or write (0b1) data. One DIC module can manage two ABACUS chips, and the selector input determines whether to communicate with the even-labeled (0b0) or odd-labeled (0b1) ABACUS. The ready output flag allows users to monitor the module's status. For read commands, the full 16-bit output in the form of Table 4.2 is mapped to the DAC_out_value bus. Ports directly connected to the ABACUS are: two clock ports (SDAC_clk_0 and SDAC_clk_1), two data ports (SDAC_data_0 and SDAC_data_1), and two input ports (SDAC_data_in_0 and SDAC_data_in_1). These ports handle the CLK, SIN, and SOUT signals of the even (port name ending with 0) or odd (port name ending with 1) ABACUS, respectively.

PS-PL communication via AXI4-L

To simplify the transmission of command from the PS running Linux, I used the Advanced eXtensible Interface-Light 4 (AXI4-L) communication protocol. The AXI4-L communication protocol is a simplified version of the full Advanced eXtensible Interface protocol used for communication between different modules in an FPGA system. AXI4-L is designed for lightweight, simple, and low-latency communication between the PS and peripheral devices or IP cores in the PL of the FPGA. Unlike the full AXI4 protocol, which supports multiple channels, multiple transactions, and advanced features like burst transfers and data streaming, AXI4-L has a reduced feature set to make it more lightweight and easier to implement. It typically supports only a single channel with a limited number of signals for read and write transactions. AXI4-L transactions are simple and require fewer clock cycles compared to the full AXI4 protocol, making it well-suited for simple control and configuration tasks.

It is commonly used for communicating with control registers, status flags, or configuration registers in peripheral devices or custom IP cores. Since AXI4-L is simpler and more straightforward than the full AXI4 protocol, it is often preferred for communication with peripheral devices or custom IP blocks that do not require the advanced features provided by the full AXI4 protocol. It allows for efficient and direct communication between the PS and the FPGA's programmable logic, enabling seamless integration of custom hardware within the FPGA design.

This design choice enables the mapping of necessary registers into the memory of the running OS. Consequently, these registers can be configured from user space, facilitating communication between the PL and the PS. Although DMA could be an alternative, the primary goal is to control the Verilog module's state machine, which does not demand extremely tight time constraints. While reading the LGAD counter status demands fast and large data transfer, other functionalities in this context do not require the same urgency. Thus, the AXI4-L protocol effectively addresses communication requirements.

Four 32-bits registers were used to communicate with DEC, DIC, DMA and output respectively. The first three registers were assigned to a specific Verilog module, while the last one is only used to read the control registers. This design choice enhances code readability and reduces the likelihood of bugs. Although not all 32 bits are utilized in each register, the suboptimal partial usage of only 4 registers has no significant impact on PL implementation or the OS.
The registers are mapped in the kernel space starting from 0x43c0_0000. 0x43c0_0000 has been chosen as the base address of the DEC controller. 0x43c0_0004 is the base address of the DIC controller while 0x43c0_0008 is the base address of the DMA and other control registers. Finally, register 0x43c0_000c is dedicated to reading status flags. The register spacing is due to the fact that the ZC702 is based on a byte-addressable architecture, where each address refers to a single byte. As a result, each register is shifted by 4 units ($4 \times 8\,\text{bit} = 4\,\text{byte} = 32\,\text{bit}$). The reading and writing of these registers is an atomic operation, ensuring that it is executed as a single indivisible unit. This property prevents interruptions and guarantees that the PS cannot update the values concurrently while the PL is performing a read or write operation and vice versa.

The DEC modules registers are shown in Table 4.4. The first element at offset 0x0 is a flag that serves as a trigger for the DEC data transfer. To initiate a new operation, the user must reset this value to 0b0, since the DEC is triggered by transitions from 0 to 1. This ensures prevention of continuous command sending while the transfer flag is 0b1. The following 16 bits at offset 0x1 are dedicated to the actual value of the LTC2602 DAC. The remaining 2 bits at offset 0x11 select the output of the LTC2602: 0b00 selects even channels, 0b01 selects odd channels, and any other value selects all channels. Finally, the last 4 bits are reserved for the command. As described in subsection 4.1.3, the LTC2602 can support multiple commands. To set a threshold, the required command is 0b0011.

Address	[19:22]	[17:18]	[1:16]	0
Description	DAC command	DAC address	DAC value	Enable

Table 4.4: Register space at physical address 0x43c0_0000 used to control the DEC module.

The registers are designed to be integrated with the ports of the DEC controller shown in Figure 4.19. This direct integration allows them to be easily connected to the corresponding ports, ensuring the specific functionality required by the DEC controller. As an example of a

write command, if the user writes 0x1e3ffd, the register is mapped as follows:

$$0\text{x1e3ffd} = 0\text{b}\underbrace{000000000}_{\text{unused}}\ \underbrace{0011}_{\text{command}}\ \underbrace{11}_{\text{address}}\ \underbrace{0001111111111110}_{\text{value}}\ \underbrace{1}_{\text{enable}}\ . \tag{4.4}$$

This command selects all outputs and sets a new threshold value of 0b0001111111111110 = 8190 DAC units, which corresponds to a voltage of 0.14996V.

The DIC module registers are shown in Table 4.5. At offset 0x0, the first element is a flag that initiates the DIC data transfer. To start a new operation, the user must reset this value to 0b0, as the DIC is triggered by transitions from 0 to 1. This precaution prevents continuous sending of commands while the transfer flag is 0b1. The next 5 bits at offset 0x1 represent the DAC value to be sent. The next 4 bits at offset 0x7 represent the DAC address. The 12th bit defines whether the operation is a read operation (0b0) or a write operation (0b1). Finally, the DAC selector selects the ABACUS to be considered: 0b0 for even numbered ABACUS, 0b1 for odd numbered ABACUS. Again, as for the DEC design, the registers are defined to

Address	13	12	[7:11]	[1:6]	0
Description	DAC selector	Read/Write	DAC address	DAC value	Enable

Table 4.5: Register space at physical address 0x43c0_0004 used to control the DIC module.

be directly connected to the ports of Figure 4.23 having the functionality required by the DIC module. As an example, writing the register with the value 0x6263 correspond to the following:

$$0\text{x6263} = 0\text{b}\underbrace{00000000000000000}_{\text{unused}}\ \underbrace{1}_{\text{write}}\ \underbrace{10001}_{\text{address}}\ \underbrace{110001}_{\text{value}}\ \underbrace{1}_{\text{enable}} \tag{4.5}$$

Enables the transmission and performs a write command on channel 0b10001 = 17 to the value of 0b110001 = 49 DAC units.

The 32-bit register with physical address 0x43c0_0008 is used to map the basic functionality of the DMA and SERDES user to acquire and transfer data from the PS to the PL. The relative register function and offset are reported in Table 4.6. The first bit is connected to one of the available output LEDs on the development board. This is to facilitate debugging and to verify that the board is processing the commands correctly. The second bit, located at offset 0x1, is used to request a DMA transfer of the counter register values that are associated with the LGADs. The DMA interface will only initialize a transfer if it detects a $0 \rightarrow 1$ transition. Before another transfer is requested, the user must reset this bit to 0. The next 3 bits are used to reset the DMA, reset the counter register associated with the LGADs, and reset the SERDES respectively.

Address	4	3	2	1	0
Comment	SERDES reset	Counter reset	DMA reset	Request DMA transfer	LED

Table 4.6: Register space at physical address 0x43c_0008 used to control the DMA, SERDES and counter registers associated with each LGAD channel. In addition, one of the board's LED is controlled for debugging purposes.

Reset commands are generally required in electronic systems, including FPGA designs, to initialize and establish a known state. When a system powers up or undergoes a reset, it might be in an unpredictable state with undefined values in registers and memory. This uncertainty can lead to unintended behavior and potential errors in the system's operation. By including reset commands, designers can ensure that the system starts from a known, well-defined state, eliminating any ambiguity.

The last registers, offset 0x43c0_000c, are dedicated to reading, mapped in Table 4.7. The first bit, at offset 0x0, is set to high when the DEC module is in use. The second bit has the same functionality, but for the DIC module. In case of read operations, the 16 bit sent by the ABACUS according to Table 4.2 are mapped to the register from 2 to 17. Finally, the bit at offset 18 is set to high when the DMA is ready for operation. Thanks to the AXI4-L protocol,

Address	18	[2:17]	1	0
Description	DMA ready	DIC received value	DIC idle	DEC write done

Table 4.7: Register space at physical address 0x43c0_000c that controls Verilog module status and DIC read command response. These registers are used in read-only mode.

the user can easily manage modules by writing from the user space the selected address. This can be accomplished using a simple C++ program. As an example, the following code opens the required address, checking for possible errors:

```
int fd_axil;
const uint32_t axi4l_size = 4UL; //in bytes. 4UL = 16 bytes = 128 bit
const uint32_t axi4l_pbase = 0x43C00000;
volatile uint32_t* axi4l_ptr = NULL;

if ((fd_axil = open("/dev/mem", O_RDWR | O_SYNC)) < 1){
    std::cerr << "Falied␣to␣open␣/dev/mem" << std::endl;
    exit(EXIT_FAILURE);
}
axi4l_ptr = (volatile uint32_t *) mmap(
        NULL,
        axi4l_size,
        PROT_READ|PROT_WRITE,
        MAP_SHARED,
        fd_axil,
        axi4l_pbase
    );

if(axi4l_ptr == MAP_FAILED){
    std::cerr << "Map␣falied" << std::endl;
    exit(EXIT_FAILURE);
}
```

If the `mmap` is executed successfully, the user can access any of the spanned memory addresses as a vector. This is because by casting the `axi4l_ptr` as `uint32_t`, each element of the vector is expected to be contained in 4 bytes or 32 bits: the exact size of the registers on the PS. The following code resets the status of the DMA register to 0x43c_0008 and requests a reset of the LGAD counter registers. It then requests a transfer to the DEC module by setting the value 0x1e3ffd as in Equation 4.4.

```
#define AXI4L_REG_EXTERNAL_DAC_OFFS 0
#define AXI4L_STATE_MACHINE_CONTROLLER 2
#define AXI4L_IDLE 0x0
#define AXI4L_RESET_COUNTER 0x8

// LGAD counter register reset
axi_ptr[AXI4L_STATE_MACHINE_CONTROLLER] = AXI4L_IDLE;
axi_ptr[AXI4L_STATE_MACHINE_CONTROLLER] = AXI4L_RESET_COUNTER;

// DEC module update
axi4l_ptr[AXI4L_REG_EXTERNAL_DAC_OFFS] = AXI4L_IDLE;
axi4l_ptr[AXI4L_REG_EXTERNAL_DAC_OFFS] = 0x1e3ffd
```

Note the use of the `volatile` type in the declaration of `axi4l_ptr`. By using this definition, the compiler knows that the value at this address can be updated externally. This is necessary when reading the register, because the values can be updated by the PS. Specific functions

have been implemented to assist the user in generating the configuration address. These functions take as input the parameters to be set; in the case of the DEC, Enable, Value, Address, and Command. They then generate the value to be written to the appropriate memory address. Overall, this design choice simplifies the communication between PL and PS with a few lines of code. By having specific registers with predefined functionalities, the interaction between the two domains becomes straightforward and efficient.

4.2.2 Processing System - PS

The Processing System (PS) constitutes a pivotal element within the System on Chip (SoC) or FPGA setup, encompassing one or multiple processor cores, memory controllers, input/output interfaces, and other critical components essential for executing an operating system (OS) and higher-level software applications. The ZC702 development board is equipped with a Dual ARM Cortex-A9 processor configuration [Xil, b]. Depending on the nature of the required functions, the processor can operate in two distinct modes: the *bare metal* mode or the *Linux* mode. In bare metal applications, the PS is streamlined to perform the fundamental tasks necessary to achieve specific objectives, and no full-fledged Operating System (OS) is used. Consequently, users are limited to employing libraries offered by the hardware provider (Xilinx) to create their solutions. The *bare metal* approach is particularly advantageous when intricate control over hardware resources and optimized performance tailored to specific tasks are imperative. On the contrary, running a Linux system on the PS is also a viable option. This approach has the benefit of complete open source accessibility and, in turn, enjoys support from the broader Linux user community. Additionally, this environment offers greater versatility with access to a wealth of tools and libraries, rendering it well-suited for intricate applications demanding networking, file systems, and other high-level functionalities. However, it is worth noting that this route requires a degree of familiarity with Linux systems. In this implementation, I chose to implement a Linux system because it is more versatile and can provide similar performance to the bare metal implementation.

Embedded Linux system

To make use of the PS a customized embedded Linux system is required. However, this process is not straightforward, as it involves cross-compilation due to the target device's ARM architecture. Additionally, the booting process of an operating system, while well-defined, requires the proper configuration and coordination of various elements to ensure successful initialization. To build and boot a complete running Linux-based SO, running on the arm processor of the Zynq the following step are required:

1. build the First Stage Boot Loader (FSBL): the FSBL, also known as a bootstrap loader or simply boot loader, is a small program or piece of code that resides in the read-only memory (ROM) or on a separate chip within a computer system. Its primary function is to initiate the process of loading the OS into the computer's memory and starting its execution.

2. Build the Platform Management Unit Firmware (PMUFW): The PMUFW controls the power-up, reset, and monitoring of resources within the system, including inter-processor interrupts and power management registers.

3. Build U-Boot: U-Boot, or Universal Boot Loader, is executed by the FSBL and is responsible for loading the Linux kernel, selecting and loading the device tree, and handing it over to the kernel.

4. Build the Root File System (rootfs): the rootfs represents the typical hierarchical file tree structure of Linux systems. By default, pre-built rootfs images are available. However, if the user wishes to customize the rootfs by adding extra programs, specific tools such as Buildroot or Yocto can be used to create a custom rootfs.

5. Build the device tree blob: the device tree blob is a compiled version of the device tree that provides essential information to the Linux kernel about the system's devices and required drivers.

6. Build the Linux kernel: this step involves compiling the Linux kernel, including any necessary custom drivers such as those handling Direct Memory Access (DMA).

7. Prepare the boot image: the boot image is created by combining all the prepared components, including the FSBL, U-Boot, kernel and rootfs images, and the device tree blob.

One notable advantage of this workflow is its modular nature, where each step operates independently, leading to more flexible debugging capabilities. In the event that a particular step encounters an issue or failure, it can be addressed and corrected individually without affecting the other components. This modularity enables efficient troubleshooting and targeted fixes, allowing for quicker iterations and improvements in the overall system. The final step, which involves creating the boot image, is a crucial task that needs to be performed whenever changes are made. The boot image can be prepared for various destinations, such as an SD card or FLASH memory on the development board. This flexibility allows for the deployment of the boot image to different storage mediums depending on the specific requirements and preferences of the system. In this implementation, I opted to execute the boot from an external SD memory card. This requires specific configuration from U-Boot setting the expected filesystem and images paking method. Once the boot has been completed, the SD card is mounted on the rootfs and all the SD space usable to store eventual data.

One of the benefits that outweigh the effort of booting a Linux operating system on an FPGA is the availability of high-level functionality, such as file writing, multithreading and Internet Protocol (IP) capabilities. This allows users to easily access and utilize features like data transfer to an external PC. It has been observed that storing data in the SD card may not be the most efficient method due to potential bottlenecks in I/O operations. To enhance performance and ensure data integrity, I opted to use the Transmission Control Protocol over Internet Protocol (TCP/IP) [Forouzan, 2002] to transmit data via Gigabit Ethernet. This approach offers an additional layer of security by storing the data away from the FPGA. Despite taking precautions to physically position the board in low radiation zones, there is always the potential risk of data corruption. By promptly sending the data off the board, this risk can be mitigated, especially when utilizing the reliability and error-checking features of TCP/IP. Indeed, TCP includes built-in error-checking mechanisms to ensure the reliable delivery of data. In the event that a packet transmission fails or encounters errors, the protocol automatically attempts to resend the data to ensure successful delivery.

To further enhance performance, the implementation of multithreading can provide significant benefits. The Zynq 7000 devices feature dual-core ARM Cortex-A9 processors, allowing for efficient utilization of the available processing power. By spawning a new thread and binding it to an unused core, both cores can be fully utilized. With multithreading, maximum performance can be achieved as higher priority system calls, such as RAM refresh or network broadcasting, can be handled separately. This ensures that critical tasks are executed efficiently without impacting the performance of the main processing threads. By leveraging the power of multithreading, the system can effectively handle multiple tasks simultaneously, resulting in improved overall performance and responsiveness.

Readout software

By integrating all the above components, threshold setting and DMA data transfer, the so-called "threshold scan" is possible. Threshold scans are performed by triggering the readout of the LGAD counters at different values of the channel thresholds. The derived data can provide important information such as the minimum detectable energy for each LGAD strip and the optimal threshold level. In order to perform these threshold scans in an automated way, it is necessary to implement a dedicated program; this program should be able to effectively control both the readout and configuration processes. To solve this problem, I developed a C++ program that is able to perform threshold scans and is also designed to handle data safely. Using the capabilities of the operating system, my program ensures secure data storage by implementing a reliable server-client data transfer mechanism based on TCP/IP. The program running on the board acts as the server, while a client PC acts as the data receiver. Once the data is acquired from the FPGA board, it is immediately transferred to the client PC for storage. This approach minimizes the exposure of the data to hazardous radiation environments and reduces the risk of data corruption. In addition, offloading data storage responsibility to a separate client PC optimizes FPGA resources.

The following steps are performed by the implemented program:

1. initiate server-client TPC/IP connection.
2. Initialize and configure DMA buffers.
3. Initialize the control register via AXI4-L.
4. Set the ABACUS thresholds to a common level using the DIC controller: 0 DAC units.
5. Reset the counter values to 0.
6. Set the threshold level of the LTC2602 using the DEC controller to `N` DAC units.
7. Perform a DMA transfer .
8. Send the data via TCP/IP to the client.
9. Set a new threshold level by increasing `N` of specific amount: `N += m`. If `N < MAX` go to item 7. Else disconnect from the client, free the memory and terminate the execution.

The purpose of the first 5 points is to correctly configure the communication to both the PS and the server client. If all these points are successful, the program can proceed to set the first threshold on `N` DAC units. The counter register data of the LGADs is then sent from the PS to the PL via DMA. Once the data is received by the PL, it is sent back to the client PC via Ethernet. At this stage, `N` is incremented and depending on its value, the program can either terminate or continue to acquire a new threshold. If `N < MAX`, the program will set the new threshold value and will continue to acquire and send the data. In the other case, if `N` exceeds the maximum limit `MAX`, the threshold scan is finished: the connection to the client is terminated and the allocated memory is freed. Additionally, the user can configure the program to write the counter data to the SD memory where the OS is running. This is useful in the unfortunate situation where the Ethernet connection cannot be established, or for debugging purposes. This design requires also the client program to be implemented. The only task of the client is to accept the communication from the server and to decode the data stream. Once the data is available, the program can format and store it in a convenient format. To optimize readability, the program saves the data in Comma Separated Values (CSV) format. The first column contains the threshold value in DAC units. Next, each

N+1th column stores the Nth channel. The last column is reserved for the time stamp. With this formatting, each row corresponds to a different threshold.

This design uses only one DMA channel buffer. As described in section 4.1.4, the DMA driver is capable of handling multiple buffers. By using 32 buffers, it is possible to perform 32 transfers on each DMA request. This further optimization has recently been implemented in the code and has consistently improved DMA performance. As a result, at any value of threshold, 32 fast transfers are received by the client.

4.3 Results

In the next section, I present the results obtained using the data acquisition described above and implemented. Acquisition was tested with and without LGAD sensors. The test without sensor allows to evaluate the data transfer and the correct data processing of the whole system. Without a sensor, it is also possible to inject a known charge into even or odd channels using the test points described in subsection 4.1.2. However, tests performed with a radiation source are intended to verify the capability of the LGADs to detect ionizing radiation. From the threshold scan it was possible to derive the minimum detectable energy by the strip sensor. The following sections present the results of both test configurations: without and with LGADs.

4.3.1 DMA benchmark

One important test, conducted without connecting a sensor, is to verify and characterize the data flow through the DMA. One of the relevant parameters is the transmission speed. In fact, it is possible to quantify the transmission speed using the timestamp generated by the DMA controller. The user can rely on this timestamp if it has sufficient accuracy. By definition, as described in section 4.1.4, the value of the timestamp is obtained by counting each positive edge of the 100 MHz clock. Alternatively, the user can rely on a dedicated function from the C++ standard library that is commonly used for timing [STD,]. However, the precision depends on the hardware being used. By relying on the timestamp provided internally in PS, this can be avoided, having a well-defined time precision and, in addition to that, synchronous triggering with other PS operations.

By converting the clock counter register to a time, the time difference between two DMA transfers can be estimated. This way, I can evaluate performance and throughput. The test was carried out on an implementation capable of handling 24 channels. Two 50 pins ribbon were connected to the ESA_ABACUS, one to control the 24 channel of ABACUS 1, and one connected to the 12 control pins of ABACUS 1 and 2. All of that according to the Figure 4.10 scheme, using the FMC breakout mezzanine board. In this setup, the FPGA board is capable of reading up to 24 channels. Therefore, the DMA transfer was optimized to 7 chunks of 128 bits. The first 6 chunks store the actual counter values of the LGAD according to Equation 4.3. The remaining 128 bits were used to carry additional information, such as the 32-bit timestamp, a 32-bit encoding of the number of DMA packets sent for transfer integrity verification, and 64 zero bits for padding. To test the whole system, a threshold scan was performed from 30000 DAC units = 0.5493 V to 37998 DAC units = 0.6958 V since this is the typical range where the signal is expected. To speed up the process, the threshold was increased by 2 DAC units; therefore, a total of $(37998 - 30000)/2 = 3999$ DAC values were explored. To make use of the 32 DMA buffers and improve efficiency, 32 read commands have been sent for each threshold value. One read command for each DMA buffer. As a result, a total of $3999 \cdot 32 = 127968$ reads of all LGAD counter registers have been captured. Once the

data are available from the PS, it is sent via TCP-IP protocol to a client written in Python. The latter formats and writes the received data to disk in a CSV format.

The captured data are then analyzed: overflow correction is required because the total duration of the capture exceeds the $\simeq 42.94\,\mathrm{s}$ maximum time representable in a 32-bit register with a 100 MHz clock, as discussed in section 4.1.4. To have a more direct physical meaning, given the clock frequency, the timestamp was converted to a time. The register carrying information on the DMA transfer was checked to ensure that no packages are lost. All 127968 reads were preset, confirming that there was no loss. Regarding the DMA transfer speed, Figure 4.24 shows the time difference between successive readings of the 24 counters.

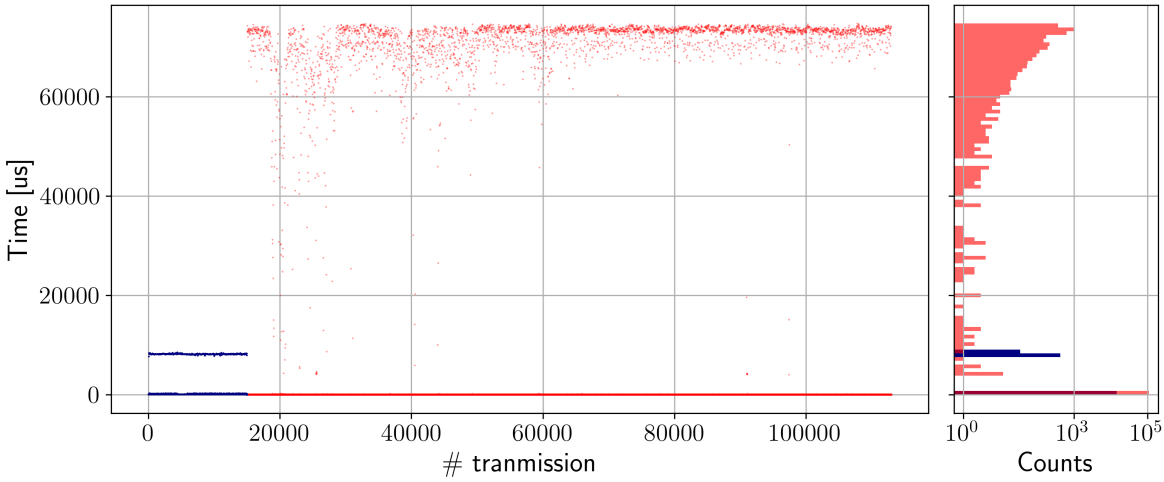


Figure 4.24: The left plot shows the time difference of consecutive reads using DMA. For each threshold, 32 reads are performed for a total of 127968 points. The right histogram shows the distribution of the time differences. Two regimes can be observed, a faster one shown in blue and a slower one shown in red.

From Figure 4.24, two distinct regimes can be identified, separated by the 15006^{th} reading, marked with different colors. The region before this threshold, shown in blue, represents the first population, while the region after it, shown in red, corresponds to the other, slower events population. The two identified populations can be further divided into two regimes. Within the first population, there is a faster regime with an average speed of $27.390 \pm 15.275\,\mu\mathrm{s}$ and a slower regime with an average speed of $8185 \pm 86\mu\mathrm{s}$. On the other hand, the second population exhibits a slower average transmission time. It also has a fast component with an average speed of $25.86 \pm 10.68\,\mu\mathrm{s}$ and a broader, slower component. The change in transmission behavior observed after the 15006^{th} reading could be attributed to buffer utilization. Both the Ethernet TCP/IP protocol and disk writing utilize buffers. When a buffer becomes full, it leads to a slowdown in the reading process. Once the system is saturated, this slowdown becomes irreversible. To recover from this situation, the transmission speed needs to be reduced by decreasing the transfer request, allowing the buffer to empty. However, no lost or corrupted data was detected.

Secondly, the structure inside each regime is attributed to the change in threshold. Figure 4.25 shows the the time difference of 64 successive reads. For a fixed threshold value, to make use of the DMA buffers, 32 reads are performed, one for each buffer. After 32 reads, the threshold value is changed. The peaks in transfer delay occur every 32 transfers, transfer #0 and transfer #32 in Figure 4.25, suggesting that setting the new threshold is time consuming. This behavior is common to all remaining data transfers not shown in the plot. In addition to that, the time difference peak of Figure 4.25 is of $\simeq 8311\,\mu\mathrm{s}$, which is consis-

tent with the estimated average speed of the slowest-first population of Figure 4.24. Further

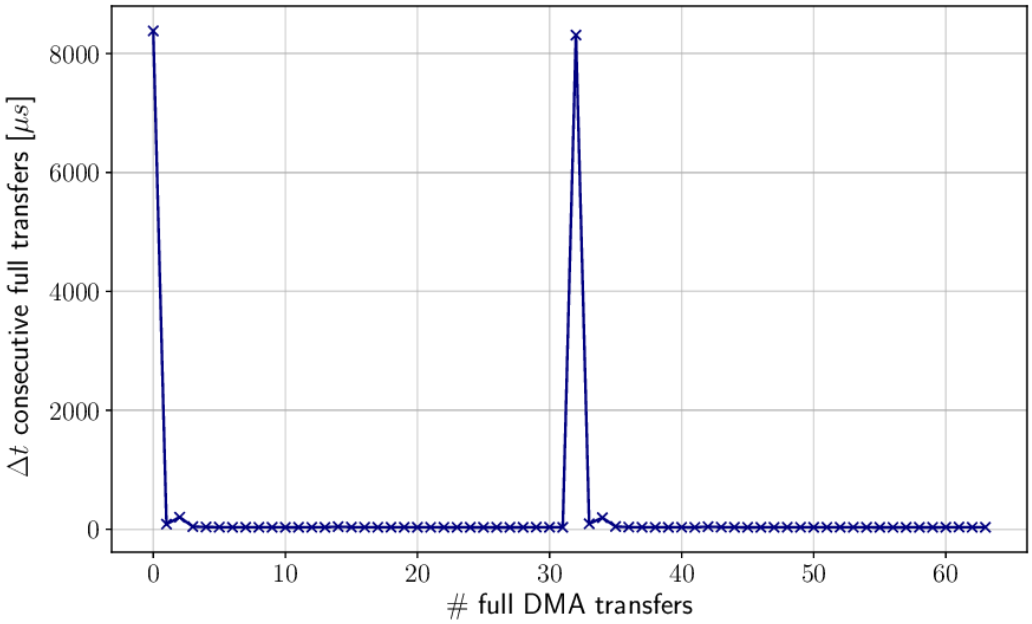


Figure 4.25: Time difference for all the performed full DMA transfers.

analysis conducted on the two population have shown that the total counts of slow-elements, for both of population 1 and 2 are exactly 3999. This number corresponds to the threshold changes, corroborating the hypothesis of slowing down the transfer speed when the threshold is changed. Investigating only the DMA transfer speed, I removed the slow-element from both populations. Using again the timestamp as time reference, the system can process data with an average speed of 31.20 Mbit/s, reaching a maximum of 49.53 Mbit/s. To improve system performance, an alternative approach could be to utilize the UDP/IP protocol instead of TCP/IP. Unlike TCP, which is connection-oriented and ensures reliable data transmission, the UDP protocol features reduced overhead and less error-checking mechanisms. Using UDP allows for faster data transfer. Packets are sent without establishing a connection or waiting for acknowledgments. However, it is essential to consider that UDP does not guarantee delivery or packet order, making it suitable for scenarios where occasional packet loss or out-of-order packets can be tolerated.

An additional improvement could be achieved by using multithreading techniques. The ARM Cortex-A9 version available on the PS is equipped with 2 cores, but the current program running on the PS utilizes only one core. To fully utilize the second core, modifications are required in the program to enable multithreading. By doing so, one thread can be dedicated solely to handling DMA, significantly increasing data transfer from PL to PS. Meanwhile, the main program can focus on configuration and data transmission. By effectively distributing tasks across multiple threads and cores, the overall efficiency of the system can be enhanced, leading to faster data processing. Utilizing both cores of the ARM Cortex-A9 processor will allow better resource utilization and maximize potential hardware capabilities.

It is important to note that in the specific application of threshold scanning, the time constraints are not severe. However, in other implementations, such as for HDM, the system response time becomes critical. In tracking applications, the readout must be able to distinguish between successive events. If the readout cannot keep up with the detection rate, it

may lead to overlapping events, and consequent loss of information.

4.3.2 Proton beam test

Proton beam tests were performed at the Trento Proton Therapy Center. This allowed the characterization of the entire readout system and the LGAD sensors. Several threshold scans were performed at different proton energies ranging from 70 MeV to 228 MeV. These scans allowed us to determine the minimum detectable energy in the sensors by focusing on the highest deliverable energy of 228 MeV, which poses the greatest challenge for detection due to the lower energy release in the LGADs. In addition, the test aimed to determine the optimal inverse polarization bias for the sensors. This bias is critical to depleting the p-bulk region of the sensors, as discussed in subsection 4.1.1. The choice of bias directly affects the signal-to-noise ratio, so it is essential to find the most appropriate bias setting to achieve optimal sensor performance. An optimal balance must be found between gain, and thus signal-to-noise separation, without increasing the leakage current of the sensor too much and without stressing the sensor too much with bias voltages close to breakdown.

Expeimental setup

The test was performed using an LGAD sensor with 48 strips. In this test, the acquisition system read data from 24 strips forming a single ABACUS, while the other 24 strips were intentionally left unconnected. The LGAD sensor was placed at the isocenter, defined at 125 cm from the beam exit. In this configuration, the beam profile covers the entire sensor. A picture of the experimental setup is shown in Figure 4.26 The high voltage was supplied by CAEN DT5519E, which also allows monitoring of the current absorbed by the sensor. Each

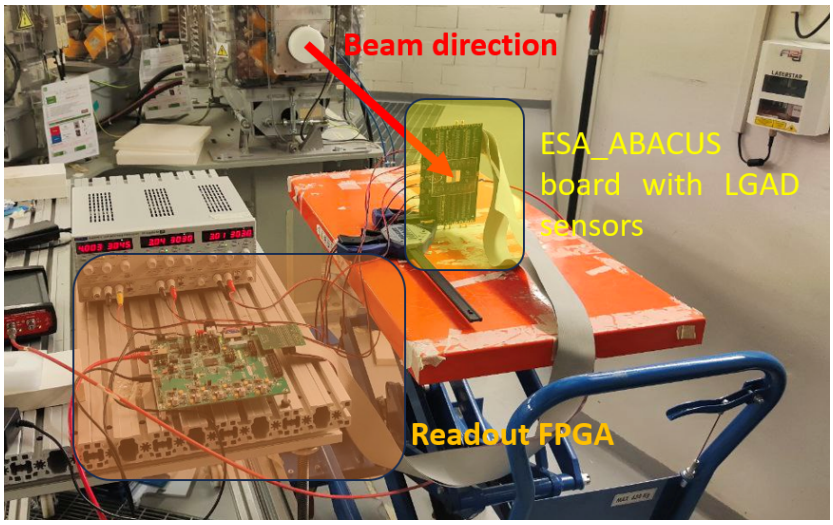
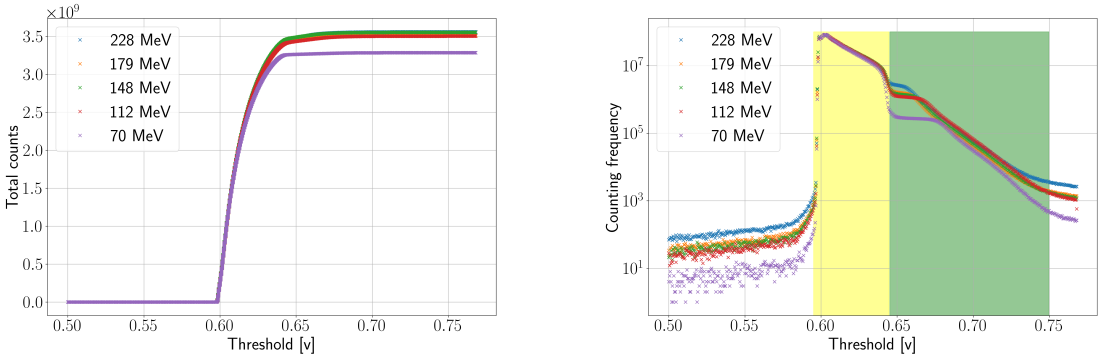


Figure 4.26: Experimental setup of the conducted test at the Trento Proton Therapy center.

power supply of the ESA_ABACUS was provided by a PL303QMT-P. The three outputs were connected to the power source required by the board. In conclusion, two 50-pin ribbon cables were connected to the ESA_ABACUS board. One cable was used to control the 24 channels of ABACUS 1, while the other was connected to the 12 control pins of both ABACUS 1 and ABACUS 2. This connection was made following the scheme shown in Figure 4.10, using the FMC breakout mezzanine board.

Threshold scans

The first part of the measurements, threshold scans, were performed with protons of 228, 179, 148, 112 and 70 MeV energy. The threshold scan range was selected from 0.5 V to 0.768 V with a step of 0.476 mV. To increase the number of counts collected at each threshold and thus reduce the statistical counting error, the program running in the PS was forced to stop at each threshold for at least 20 ms. All The results of the threshold scan for channel 0 are shown in Figure 4.27. Figure 4.27a shown the total counts at the selected threshold, while Figure 4.27b shows the counting frequency calculated utilizing the timestamp information from the PL. Since the range of the thresholds scans is kept constant for all the runs, it is expected that each threshold scan have the same duration.



(a) Total number of counts

(b) Counting frequency

Figure 4.27: Total counts (a - left) and counting frequency (b - right) for channel 0 from the threshold scan conducted using protons of varying energies: 228 MeV, 179 MeV, 148 MeV, 112 MeV, and 70 MeV. In Figure 4.27b the yellow region contains counts that are mainly influenced by noise. In the green region, the counting frequency is attributed to the presence of the actual proton beam.

In Figure 4.27a, the total counts gradually increase as the threshold is raised until it reaches around 0.6 V. At this point, the total counts experience a sharp rise. After this, the total counts remain stable until the threshold reaches its maximum value. This pattern is consistent across all energies. The observed behavior can be explained as follows. When the threshold is set too low, it does not intercept any signal and the counts remain relatively stable. However, as soon as the threshold intersects the noise level, the ABACUS begins counting with its maximum capacity. This results in a sharp increase in the number of counts recorded. Once the threshold is increased beyond the noise level, the total counts stabilize at their final values. From this kind of plot, it is difficult to distinguish the signal from the noise although the bream signal was present.

A more complete description can be obtained by looking at the right graph of Figure 4.27b. In this graph, two distinct regions can be identified, each representing different characteristics of the signal. The yellow region consists primarily of counting contributions from noise, whereas the green region corresponds to the signal generated by the actual proton beam. This happens because the proton beam signal is expected at higher thresholds outside the main noise contribution. To help the reader, in Figure 4.28 I show a comparison of a threshold scan performed without and with a proton beam of 70 MeV. The data exhibit a common contribution from noise, which starts at approximately the same threshold values in both cases. However, the noise's contribution to the counting frequency is less significant in the no-beam case. As the threshold is increased beyond the noise level, the counting frequency

drops to zero in the absence of the beam. In contrast, when the beam is present, a distinct signal is observable, indicating the presence of protons being counted. It is also possible to observe a slight increase in the threshold values attributed to the noise, suggesting that the presence of the beam alters the electronic noise of the board.

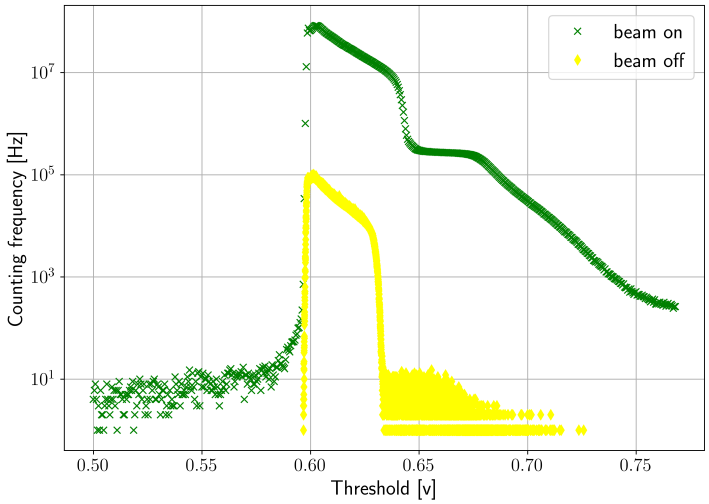


Figure 4.28: Comparison of threshold scan counting frequency without beam (yellow), and with proton beam of 70 MeV (green).

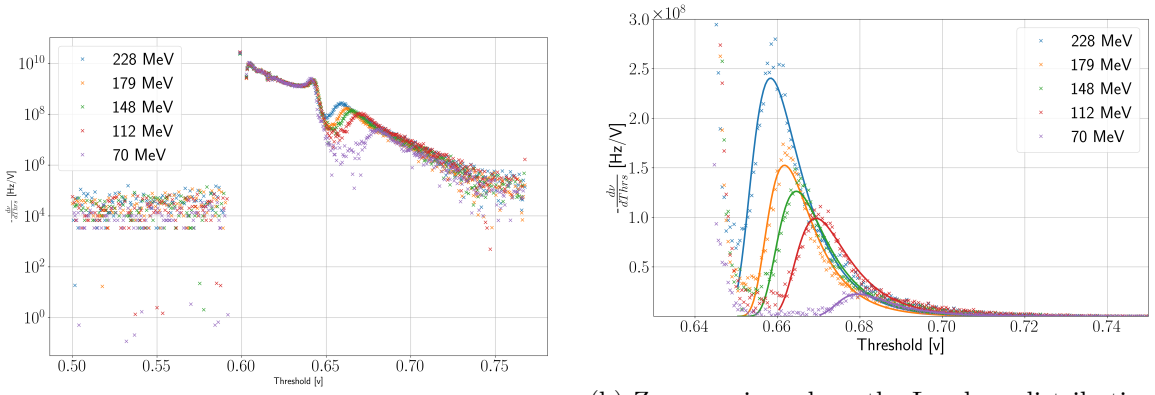
With further data processing it is possible to estimated the probability the of energy deposition in the sensors strips at a given threshold. I define $p(x)$ as the probability density function of the signal having amplitude x. I write as $f(x > x_{\text{threshold}})$ the counting frequency with signals of amplitude x, larger than $x_{\text{threshold}}$. $x_{\text{threshold}}$ can be configured, as previously described, by the user. Then I can relate the signal amplitude probability density function $p(x)$ to $f(x > x_{\text{threshold}})$) assuming a constant frequency of the particles f_0:

$$f(x > x_{\text{threshold}}) = f_0 \int_{x_{\text{threshold}}}^{\infty} p(x')dx' \,. \tag{4.6}$$

It is then possible to recover $p(x)$ since $f(x > x_{\text{threshold}})$ is measured when performing a threshold scan.

$$A\, p(x = x_{\text{threshold}}) = -\frac{df(x_{\text{threshold}})}{dx} \,, \tag{4.7}$$

where A a normalization constant that account for the f_0. By calculating the first derivative of the count frequency with respect to the threshold, it is possible to estimate the probability of having a signal at that threshold multiplied by a constant. Since multiple threshold values are collected, it is possible to recover the shape of such a probability distribution. The results are shown in Figure 4.29 for channel 0 at proton energies of 228, 179, 148, 112, 70 MeV. Again, there is a significant noise contribution before ~ 0.65 V in agreement with Figure 4.27b. At values of threshed > 0.65 V it is finally possible to identify the amplitude probability distributions. It is expected that this distribution is represented by a Landau distribution convolved with a Gaussian distribution [Knoll, 2010]. However, I have found that experimentally a Landau distribution without the Gaussian convolution represents the experimental data well and can therefore be fitted instead to simplify the process.



(a) Entire threshold scan region.

(b) Zoom region where the Landaus distributions are visible and fitted.

Figure 4.29: Amplitude probability distributions calculated using Equation 4.7 for channel 0, varying energies: 228 MeV, 179 MeV, 148 MeV, 112 MeV, and 70 MeV.

Figure 4.29b shown a zoom of Figure 4.29a where the Landau distributions are visible and fitted. More details on how the fit was performed are available in Appendix subsection 4.5.1. As the primary particle's energy increases, the peak of the Landau distribution shifts towards lower threshold voltages, approaching the region dominated by noise. This shift is a direct consequence of the lower energy being released on the sensors by higher primary particle energies, in accordance with the Bethe-Block equation. The decrease in energy deposition results in a reduction in the signal strength, leading to a shift in the peak of the distribution towards lower threshold values. It can also be observed from Figure 4.29b that the dispersion of the Landau distributions decreases as the primary particle energy increases. This effect can be observed in Figure 4.31 for all the channels where the Full Width Half Maximum (FWHM) of the Landau distribution is plotted agains the primary particles energies. A clear decrease of the FWHM value is visible as the primary particles energies increases. This reflects the fact that a passive modulator is used to obtain energies lower than 228 MeV. As a consequence, the lower the energy, the higher the interaction of the beam with the modulator, corresponding to a higher energy straggling and, therefore, a wider energy distribution. This behavior is common to all the channels. As the beam energy decreases, the amplitude of the probability distributions also decreases. This behavior is due to the reduced number of protons reaching the detector. As mentioned above, a passive modulator is used to reduce the energy, and an additional effect is the potential loss of the particle population due to scattering with the modulator, and transport if the same number of particles are extracted from the cyclotron. This is exactly the case, as all irradiations were performed with the same particle rate extracted from the cyclotron. Consequently, fewer protons reach the detector, leading to a reduction in the amplitude of the observed probability distributions.

Baseline estimation

By fitting each Landau distribution for the 24 channels, the Most Probable Value (MPV) can be quantified. The procedure on how the fit has been performed is reported in the appendix: subsection 4.5.1. The MPV represents the most probable energy loss of charged particles passing through the thin layers and is crucial in characterizing the detectors' energy response. When the experimental data are fitted by Landau distribution, the MPV can be determined, providing valuable insights into the interactions between charged particles and the detector material. The two main parameters of the Landau distribution are the MPV and

the "width" (sigma), which represent the peak and the spread of the energy loss, respectively. The physical interpretation of the MPV corresponds to the most probable energy loss of charged particles and provides an estimate of their average energy loss as they pass through the detector material. Figure 4.30 displays the MPV as a function of the primary proton energies. As anticipated, with increasing primary particle energy, less energy is released in the LGADs. The energy released by an incoming particle with energy E is well described by the Bethe-Block equation under the assumption of a purely electromagnetic interaction, which predicts in first approximation a $\propto 1/E$ relationship. This can be derived from Equation 1, assuming fixed projectile and target characteristic and neglecting effective charge $Z_{eff} \simeq Z$ and other correction therms inside the logarithm. The main dependece on the particle can then be expressed as $\frac{1}{\beta^2}$. We then used an equation of the form:

$$F(E) = b + \frac{a}{E}\,, \tag{4.8}$$

to fit each MPV energy dependence of Figure 4.30 with b and a constants. The baseline of can be estimated from Equation 4.8 in the limit as E approaches infinity. This idealization represents a particle with infinite energy; according to out parameterization of the Bethe-block equation, zero energy would be deposited in the detector. As a consequence, the parameter b is the baseline value. With the estimated baseline for each channel, the separation between noise and actual signals generated by charged particles interacting with the detector can be quantitatively evaluated. This information is crucial to correctly define the threshold and evaluate whether there is an optimal configuration with the ABACUS chip. If the baseline values fall outside the adjustable threshold range, this indicates that the channel cannot be used. In such cases, sacrificing the channel might be necessary to ensure the reliable data acquisition of the remaining channels.

Figure 4.32 shows the estimated values of the baselines for the 24 channels. The maximum value is $0.6716 \pm 0.0008\,\mathrm{V}$, while the minimum value is $0.6209 \pm 0.0005\,\mathrm{V}$. The average value is $0.6410\,\mathrm{V}$ with a standard deviation of $0.0128\,\mathrm{V}$.

It is possible to compare baseline values by performing the same analysis on the threshold scans at different LGAD voltage biases. By choosing and maintaining a fixed energy of the primary protons and by changing the voltage bias, it is possible, again performing the limit of such voltage to zero, to have another estimation of the baselines. We selected the 70 MeV energy as it is the easiest to detect, and performed threshold scans at 120, 100, 80, 60 and 40 V. It is expected that as the voltage bias is lowered, the strength of the electric field affecting the collection of electrons and the gain will decrease. The main consequence is a reduction in signal amplitude. Figure 4.33 shows for channel 0 the experimental amplitude probability distribution, as well as the fit performed. As the bias voltage is lowered a shift of the distribution towards lower voltage values is observed. To characterize this behavior, the signal probability distributions were fitted with Landau functions for each channel and for each bias voltage. The resulting behavior is shown in Figure 4.30. This time, the relation between bias voltage and MPV is linear and fitted with a linear function:

$$y(V) = mV + q\,, \tag{4.9}$$

with V voltage while keeping m and q constants. The baseline can be calculated by projecting the line equation when the bias is zero $V = 0$. As a consequence, q of Equation 4.9 is the baseline value. Figure 4.35 shows a comparison between the baseline values shown in Figure 4.32 and the baseline calculated using the LGAD voltage bias. A an average offset of $0.0222 \pm 0.0019\,\mathrm{V}$ is observed as the baseline calculated from the energy dependence of the MPV have systematically higher values. Despite this offset, both descriptions show the same behavior, except for the channel 11. When investigating this specific channel, no anomalies

or irregularities were detected in the data analysis. The origin of the offset could be due to the non-ideal condition in terms of polarization and uniformity of the electric field inside the LGADs sensors when the bias voltage is too low. This could affect the drift of the charges, which affects the overall sensor performance.
However, the spanned range of the baselines measured with the two methods agrees: 0.05070 ± 0.0009 V. This is a significant parameter, as the ABACUS must be able to accommodate a common threshold for the channels within this range. The current version of the ABACUS faces a limitation in terms of threshold adjustment, which allows a maximum of ± 24 mV for tuning thresholds. Consequently, the ABACUS internal DAC range does not allow one to fine-tune all required thresholds. This issue is expected to be addressed in the upcoming production of ABACUS, where an extended DAC range is expected, allowing the full threshold adjustment range.

Voltage bias selection

When considering the best value of the voltage to supply as a bias to the LGADs, two effects compete with each other. The first effect is the amplification of the signal as the bias voltage increases. Ideally, the highest possible voltage voltage should be chosen before the sensor reaches the breakdown. The second effect is due to the sensor temperature: as the temperature increases, the leakage current increases. As more current passes through the sensor, the temperature increases, leading to a positive feedback loop. To prevent this, we limited the high-voltage power supply to a maximum current. However, if the current exceeds the imposed limit, the power supply starts to decrease the voltage, leading to a variation in the baseline levels. Therefore, the ideal bias voltage is the minimum voltage, to prevent heating problems, with sufficient gain to detect the smaller amount of energy released inside the LGADs.

Figure 4.36 shows the charge distribution of channel 0 for a proton beam of the highest energy, 228 MeV, at different voltage bias of 260, 240, 220, 200, 180, 160, 140 V. Significantly, the signal remains discernible across all derivable energies, even with a bias as low as 140 V. Decreasing the voltage bias beyond this point leads to the signal disappearing.

As the voltage bias decreases, the energy distribution's spread also diminishes. I attribute this behavior to the reduction in the charge multiplication effect within the sensor. With fewer charge carriers, the velocity distribution becomes more focused, aligning closely with the average values of the ideal velocity distribution. In the context of HDM development, the charge distribution information is not crucial, as the requirement is for a digital signal from each strip when traversed by radiation. With this in mind, the voltage bias was set at 220 V, achieving a suitable balance between detection and preventing overheating.

Crosstalk and overheating

Only half of the board was used during beam time, as controlling 24 channels requires only one ABACUS chip. Consequently, only one side of the ESA_ABACUS needs to be powered and used to control and read the ASIC. By switching on the entire board with 2 additional ABACUS (ABACUS 5 and ABACUS 6) it is possible to observe a drift in the threshold scan profiles. Figure 4.37 shows the threshold scan for channel 0, with half and full ESA_ABACUS powered. This behavior is common to all the remaning channels. This offset towards higher threshold can be observed when the board is fully powered on. First, this shift was attributed to a thermal effect, as the full powering of the board significantly increased the temperature. In Figure 4.40 a thermal image of the front and back of the ESA_ABACUS with fully powered is shown. The highest temperatures up to 58° C can be observed where the ASICs are located.

Compared to thermal measurements, where only half of the board was powered on, the higher temperature is substantially lower with a maximum of 48° C.

To address concerns about temperature drift, additional studies were conducted at the Istituto di Fisica Nucleare of Turin using a thermostatic bath for thorough system response testing. In this setup, it was possible to place the ESA_ABACUS in an environment with a controlled temperature, thanks to the thermostatic bath. With the help of a fast arbitrary signal generator, it was possible to simulate the signal from an LGAD strip using the dedicated inputs. Real-time board temperature measurements were made using two PT100 thermometers [Bernstein, 2021] one placed near the sensor and the other near the ABACUS. Threshold scans were again performed in the same configuration as during beamtime, with half of the board powered up while maintaining controlled temperature conditions. Care was taken to ensure that the measured temperature remained stable before initiating a threshold scan. Figure 4.38 shows the noise region of the threshold scans at 40, 20, 15, 10 ° C. The figure shows only the noise region, but tests simulating the injection of a charge using an arbitrary waveform generator have shown that the offset in question is common to both signal and noise. Observing the noise or the signal as the temperature changes is therefore indifferent. In contrast to the observations made in Figure 4.37, where an increase in temperature led to a shift towards lower thresholds (right), the effect seen in Figure 4.38 is that a temperature rise causes a shift to the left. As a result, thermal effects were ruled out as a possible cause of the drift, as they would have produced the opposite behavior. Subsequent tests revealed an unexpected interaction between the two sides of the board. While they are supposed to work independently, as soon as a sensor is bonded and the board is fully powered up, they are connected by contact with the sensor. When the board is half powered, only ABACUS 2 is connected to the sensors mounted on one half. On the other hand, when the board is fully powered, ABACUS 5 and ABACUS 6 are active and connected to the sensors mounted on the other half of the board. My hypothesis is that if both sides of the board are connected to the sensors, there might be crosstalk that leads to an offset in the threshold scan. This hypothesis is not contradicted by the observed offset offset in Figure 4.37 at controlled temperature, therefore excluding the thermal contribution. It is important to note, however, that this may not be the only explanation, as other factors, such as leakage current from the LGAD affecting the ABACUS chip, could also be a viable explanation.

4.4 Conclusion

In conclusion, the work presented in this study has demonstrated the successful implementation and characterization of a readout system for Low Gain Avalanche Detectors (LGADs) using the ZC702 FPGA development board. By implementing appropriate controls for the ESA_ABACUS LTC2602 DAC, it was possible to configure the thresholds for the even or odd - 24 channels. Again, by implementing controls for the ABACUS ASIC, it is possible to achieve fine control over individual channel thresholds. Both controllers were implemented in the FPGA. Using the ARM processor integrated into the Zynq FPGA, I successfully created and executed a Linux distribution that boots from the SD card. This operating system environment provides users with a more robust platform for submitting commands and interacting with the FPGA system. In this specific case, I used the AXI4-L protocol to facilitate the control of both ABACUS and ESA_ABACUS thresholds. This was accomplished by writing to specific memory areas within the Linux user-space. In terms of signal processing from the LGADs, the pre-processing accomplished by the ABACUS yields a rapid digital signal that cannot be effectively sampled by the FPGA using conventional methods due to its short duration. This necessitated the utilization of SERDES (Serializer/Deserializer) primitives to sample the signal at a suitable speed, enabling the subsequent detection and counting of

LGADs strip activation. Given the high data throughput involved, specific data handling measures were required. In this implementation, the Direct Memory Access (DMA) mechanism was employed to facilitate the smooth transfer of data from the FPGA to the user in a convenient format. However, despite its advantageous speed, using the DMA entails certain drawbacks. To ensure its correct usage, the Linux system must be capable of effectively managing it. This entails the implementation of a kernel driver that properly maps the DMA registers and incorporates the DMA functionality (including elements like the Scatter Gather List). This allows the user to work on top of the driver, controlling the DMA with simple and concise instructions, such as asking the DMA to transfer from the FPGA to memory. Once the data becomes accessible within the Linux user space, it can be transmitted using the Transmission Control Protocol-Internet Protocol (TCP/IP) to an external PC situated beyond the experimental area. This distant PC offers a secure setting for the storage of the obtained data, as well as the capability to undertake pre-processing and initial analysis. This ensures real-time validation and verification of the acquired data.

To verify the correct implementation of the data flow, two main tests were performed. A "dry test", without a connected LGAD sensor and a proton beam test, where a 24 channel sensor was connected. The dry test was conducted to analyze the data flow characterizing the DMA and TCP-IP performances. Also, to verify that nothing has been lost during the transfer.
The proton beam test, performed at the Protontherapy center in Trento, was conducted to characterize the sensor response to clinical energies, ranging from 70 MeV to 228 MeV, the latter is the hardest to detect due to the smaller energy released in LGADs. The threshold scan methodology provided valuable insights into the behavior of the LGAD sensors at different proton energies. By analyzing the threshold scan data, it was possible to extract the probability distribution of the protons releasing energy in the LGAD strips. The shape of the latter should follow a Landau distribution convolved with a Gaussian distribution. By fitting these distributions, important parameters such as the Most Probable Value (MPV) can be estimated. By detecting the Landau distribution of protons at 228 MeV, I proved that the sensor is capable of working with clinical proton energies. Furthermore, by analyzing the threshold scan data, it was possible to define the baselines of the channels, which is crucial for distinguishing between signal and noise. This step is critical for accurate data interpretation and the successful application of the readout system. Baseline values were estimated in two different ways. At fixed voltage bias, finding the MPVs of the Landau distributions of different energies $E = 228, 179, 148, 112, 70$ MeV by fitting the Landaus on the experimental data. The behavior of the MPVs follows $\propto 1/E$ (Figure 4.30) and, taking the limit of $E \to \infty$, it is possible to estimate the zero energy released in the sensor or the baseline. The second way to estimate the baseline is to build, for a fixed delivered proton energy, an MPVs relation at different voltage biases V. Again, the MPVs were estimated by fitting the Landaus of energy released distribution in the LGADs. This follows a linear equation (Figure 4.34), and, ideally, taking the limit for $V \to 0$, it is possible to estimate the baseline. The baselines estimated using both methods exhibit a similar trend, in agreement with the observation in Figure 4.35. However, there exists an average offset of 0.0222 ± 0.0019 V between the two methods, whose source I hypnotize originates from the nonuniform polarization of the LGADs at low bias voltage. Nonetheless, the dispersion of the baselines is the crucial parameter, as each channel requires an accurate threshold configuration to effectively differentiate between noise and signal. The maximum obtained rage is 0.05070 ± 0.0009 V, in agreement with both of the methods mentioned above. Although this baseline compensation can theoretically be achieved using the ABACUS chip, the current version is limited in its ability to adjust individual channel thresholds within a range of ± 24 mV, rendering it insufficient for precise configuration of all channels. This issue is recognized by the designer, and the forthcoming

production is expected to rectify this limitation.

We also explored the minimum polarization bias to strike a balance between effective particle detection and preventing sensor overheating. While this decision hinges on factors such as environmental conditions (temperature, electronic noise, etc.), in our specific experimental setup at the Protonterapy center, I found that 220 V was the optimal choice. This bias level ensured successful signal detection while effectively mitigating the risk of sensor overheating.

Additional experiments were conducted to investigate the perplexing response of the board, which exhibited a threshold shift when operating at full power. Detailed tests conducted in a controlled thermal environment revealed that this threshold shift could be attributed to crosstalk between the two sides of the ESA_ABACUS. Although the initial design aimed for the two sides to remain separate and autonomous, the sensor wiring shared by ASIC2 on one side and ASIC5 and ASIC6 on the other connects these two sections together when the board was fully powered.

This chapter provides the backbone for the implementation of a complete single-layer hybrid detector for microdosimetry (HDM). Despite the fact that the ABACUS chip does not provide sufficient threshold range to specifically select the required threshold each individual channel, this design can be extended to a complete HDM layer composed of 71 LGAD strips sacrificing some channels. As mentioned above, since most of the implementation of Verilog code relies on metaprogramming techniques, a possible extension to 71 channels is relatively easy to implement. However, more than one FPGA is required according to the pin-out configuration underlined in subsection 4.1.4. Further implementation should aim at designing a *Master* FPGA responsible for controlling the threshold for all layers. This means that FPGAs handling different HDM layers need to be able to communicate with each other. However, since all of the PS are running under Linux and threshold adjustment operation is not time-constrained, this can be done by using TCP-IP communication. In addition, buffer stages are needed to avoid the saturation observed in subsection 4.3.1. Buffer states should be implemented in both the PL and PS.

HDM will require tracking capability. As a consequence, the triggering of multiple layers is required. This operation is time-constrained, it cannot be done from the *Master* FPGA over the Ethernet connection. A dedicated connection line is required between the layer FPGAs and the *Master* FPGA. It would be ideal to implement the trigger condition in the *Master* FPGA. Once this condition is met, the *Master* triggers a readout of the other FPGAs. By carefully timing and designing the buffer state in the PL, it is possible to recover the trigger event in the specific stage of the buffer. This method allows the acquisition to continue while the buffer of interest is sent to an acquisition PC, in this case via TCP-IP, as the operation is no longer strictly time constrained.

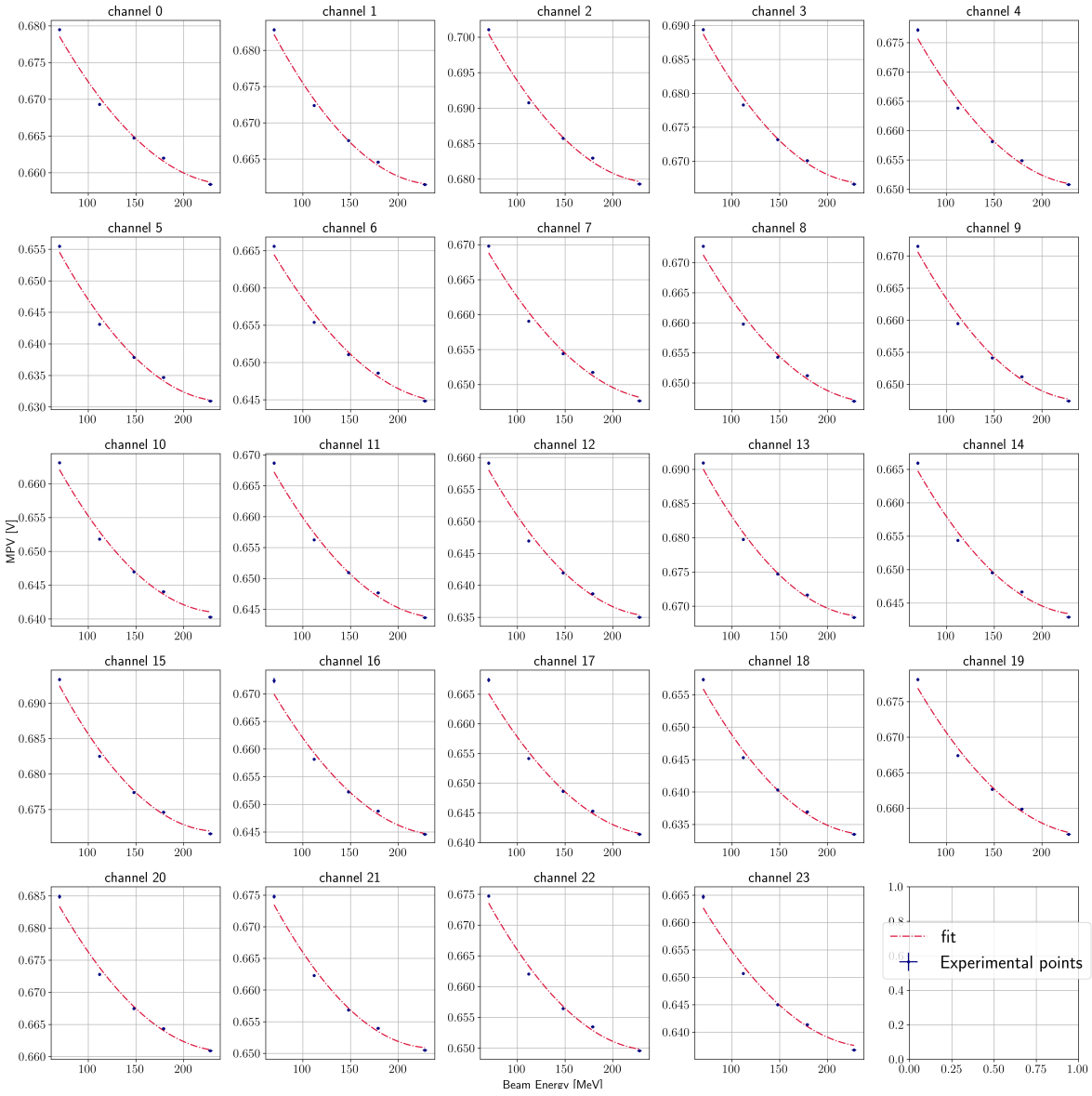


Figure 4.30: MPV calculated by fitting the Landau distributions for each channel, as function of the primary particle energy with Equation 4.8. LGAD bias of −220 V. The error bars are not visible.

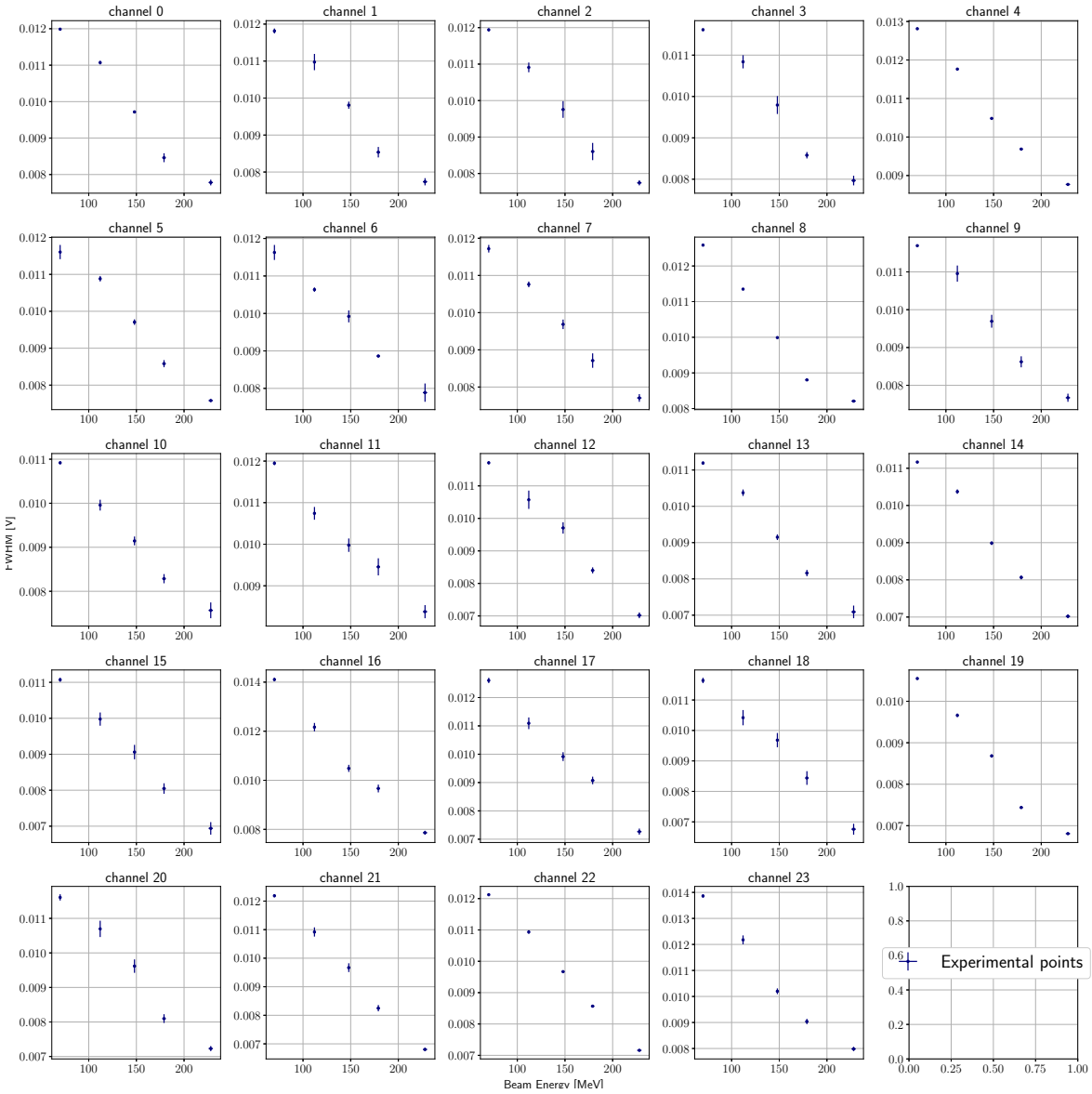


Figure 4.31: Full Width Half Maximum (FWHM) calculated by fitting the Landau distributions for each channel, as function of the primary particle energy with Equation 4.8. LGAD bias of -220 V.

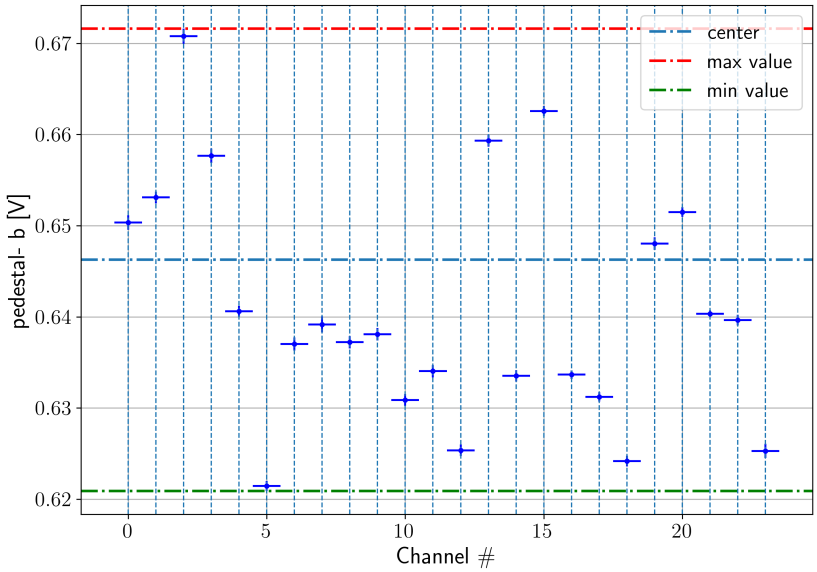


Figure 4.32: Calculated baseline for all the 24 channels using Equation 4.8 in the limit of $E \to \infty$ corresponding to no energy deposited.

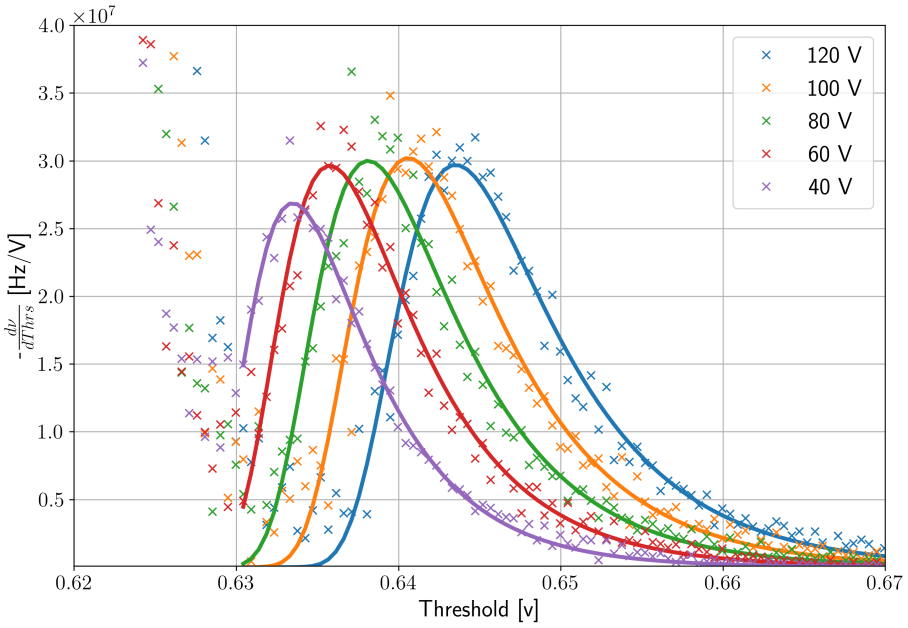


Figure 4.33: Amplitude probability distribution at different voltage bias of 120, 100, 80, 60 and 40 V for 70 MeV protons for channel 0. Figure shows also the fit performed on the landau distribution.

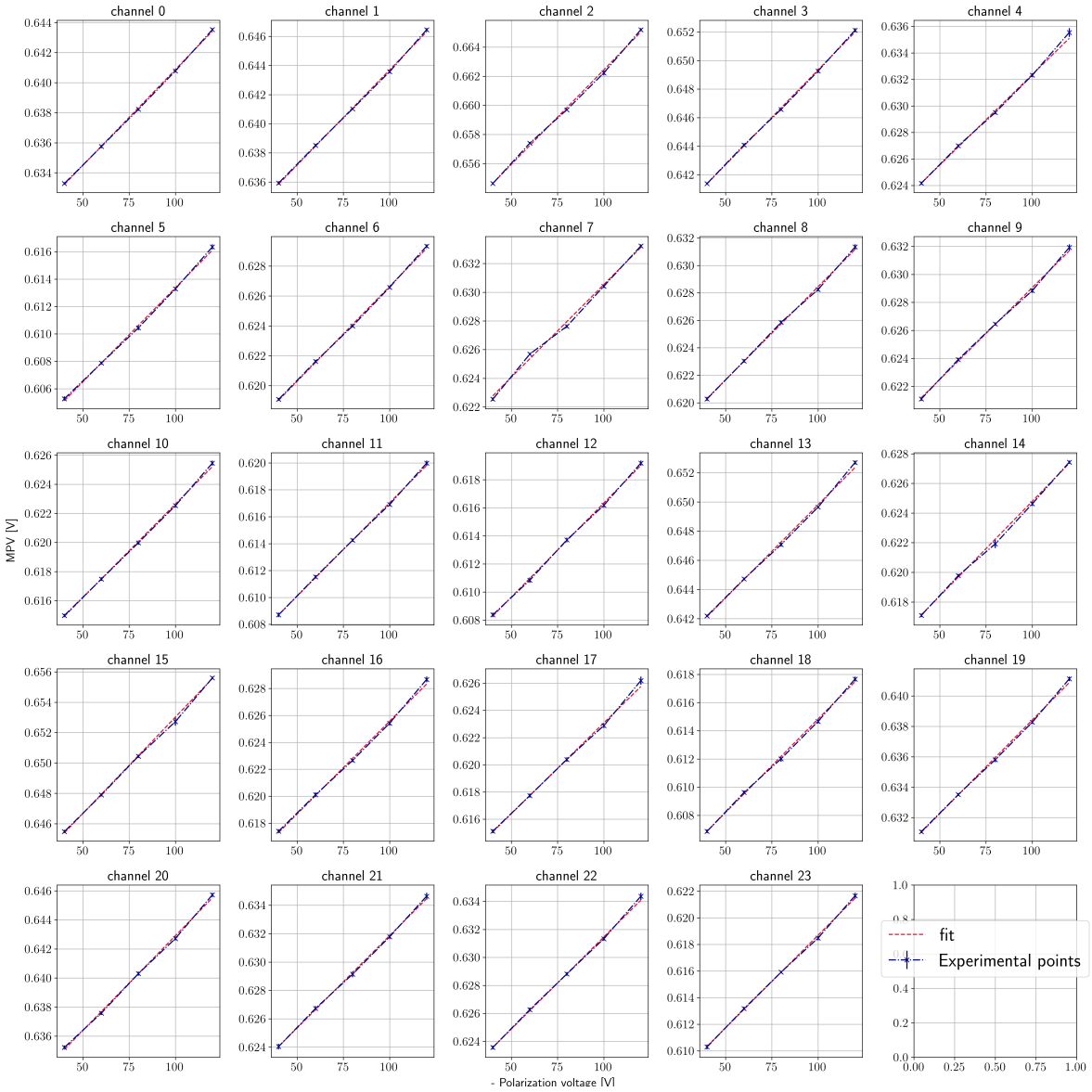


Figure 4.34: MPV calculated by fitting the Landaus distributions for each channel, as function of the LGAD voltage bias. Then a fit with Equation 4.9 is performed to estimate the bias voltage of each channel. The error bars are not visible. Proton energy at 70 MeV.

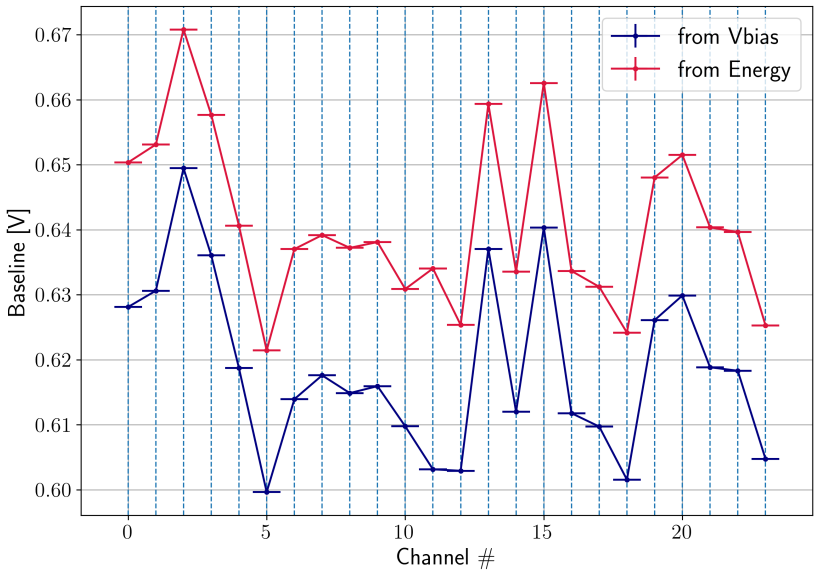


Figure 4.35: Comparison baseline calculated using the primary proton energy dependence and the LGAD voltage bias dependence.

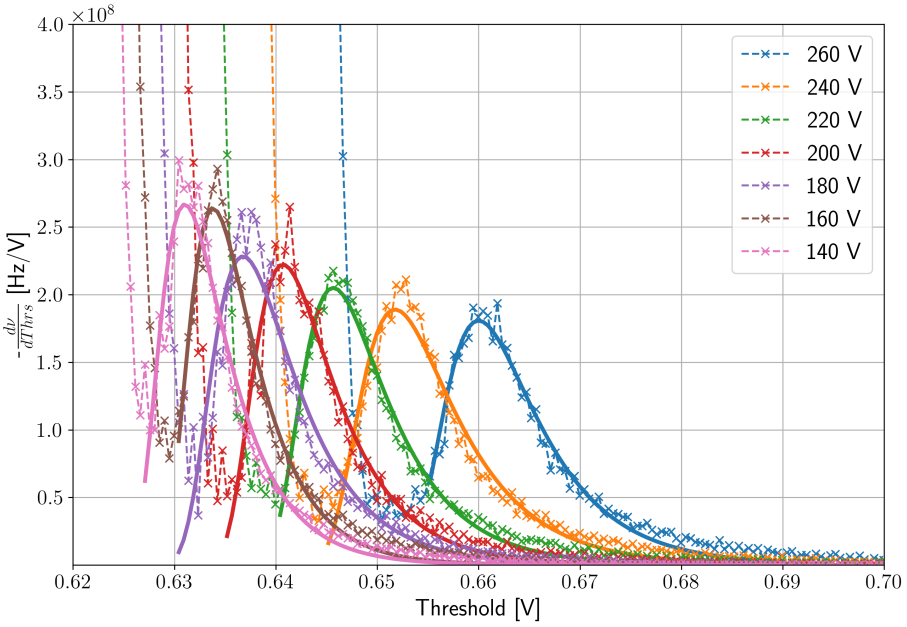


Figure 4.36: Charge probability distributions for protont at 228 MeV in different Bias voltage of the LGADs for channel 0.

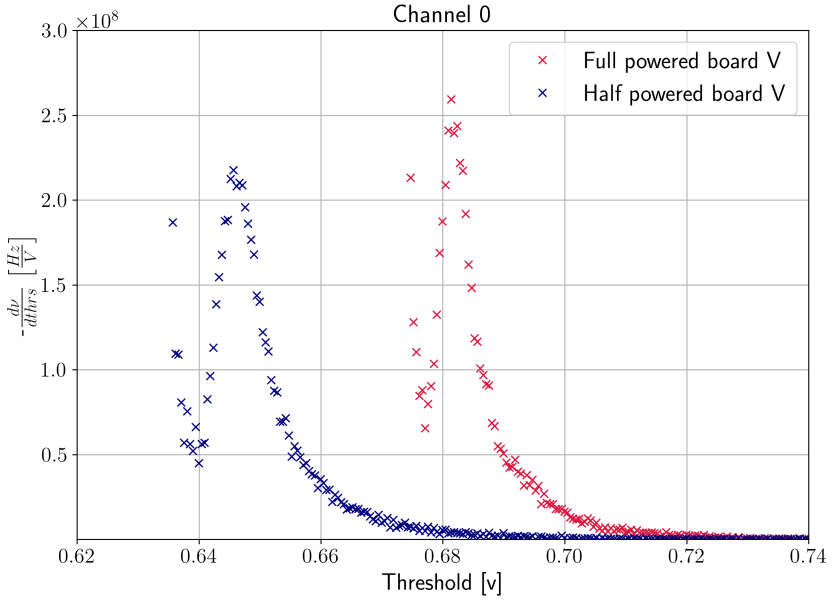


Figure 4.37: Offset observed in the threshold scan of channel 0 using a proton energy of 228 MeV. In blue, the threshold scan was performed with half of the ESA_ABACUS turned on, while in red, the threshold scan was performed with the board fully powered. The bias was set to 220 V in both cases.

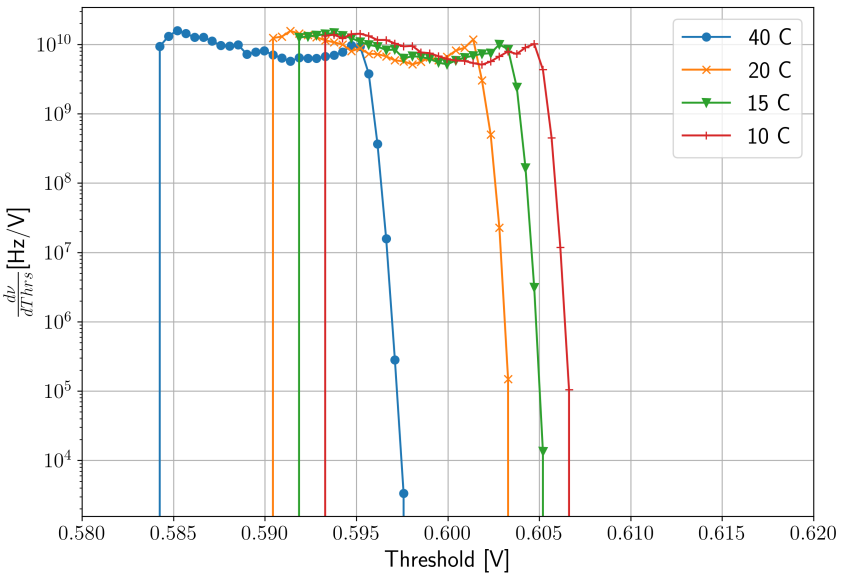


Figure 4.38: Noise region of the threshold scan performed in climatic chamber at different temperature of: 40, 20, 15, 10 ° C.

4.5 Appendix

4.5.1 Landau distributions noise processing and fitting

As shown in Figure 4.36, the Landau distribution of energy deposition in LGADs is noisy. This can make it difficult to automate the fitting process: the Landau distribution to fit depends heavily on the proton energy and on the voltage bias. As a result, it is difficult to identify fixed initial points to help the minimization algorithm converge. In addition, the Landau distributions are two orders of magnitude below the noise, as shown in Figure 4.29a, making the fit more prone to divergence. Therefore, it is important to provide good initial fit parameters, such as a guess of the mean, variance, and fit interval, to make the fit converge. Simply finding the first local maximum by evaluating each value of the data from right to left is not robust, as noise can create fake peaks.

The main goal is to automate this process, since the total number of fits to be performed is considerable (hundreds). To smear the noise from the distributions, I performed a convolution with an analytical approximation of the Landau function[Behrens and Melissinos, 1981]:

$$c(x, a, \mu, \sigma) = a \exp\left(-\frac{\frac{x-\mu}{\sigma} + \exp\left(-\frac{x-\mu}{\sigma}\right)}{2}\right). \tag{4.10}$$

This convolution process, shown in Figure 4.39, with a good guess of μ, σ and a from Equation 4.10, can highlight the signal of interest of the charge probability distribution. By scan-

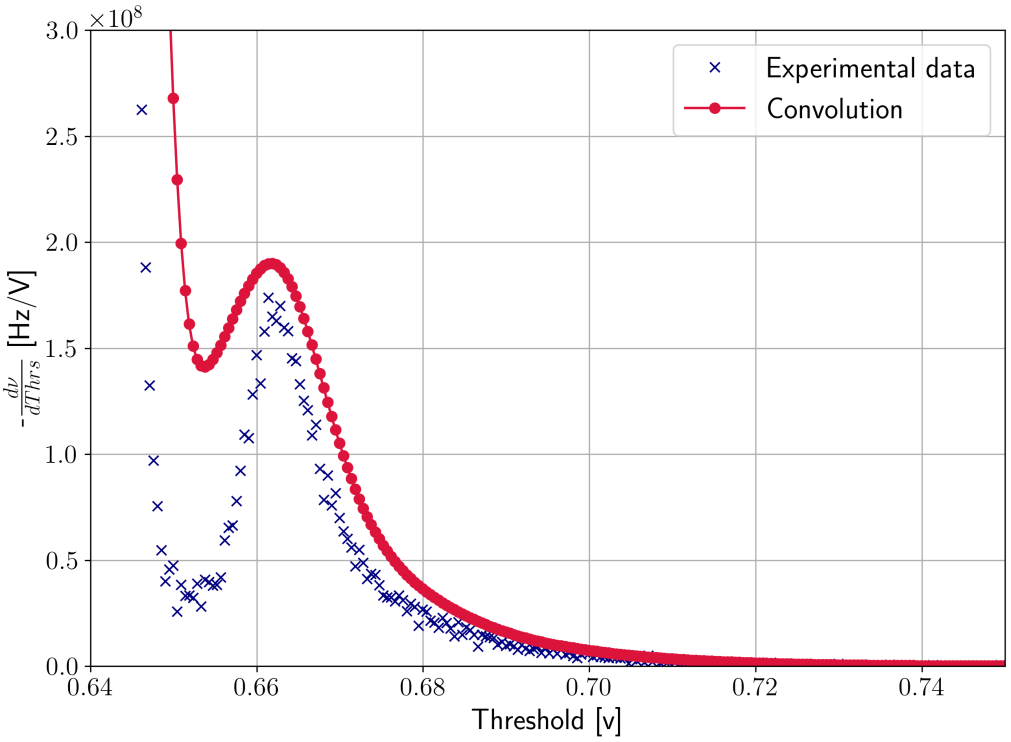


Figure 4.39: Example of convolution of the raw signal with (blue cross) with Equation 4.10, resulting in the red, connected, circle line. Threshold scan obtained from 179 MeV with 220 V as bias for the LGADs.

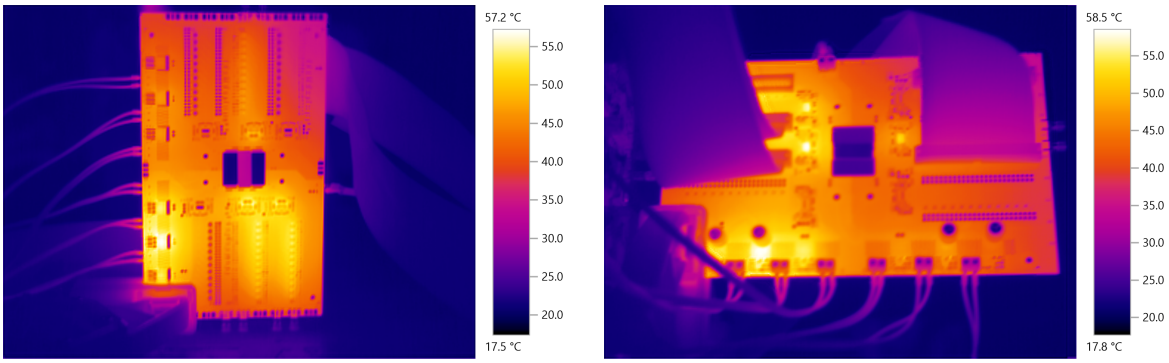
ning the points from right to left, it is now possible to identify fixed markers on the convoluted

shape, such as the aforementioned first local maxima, used as the starting point for the μ parameter, and the first local minima, used as the lower limit of the fit boundary. Identification of these two parameters alone significantly improves the fitting process. This makes it almost unsupervised. However, choosing the main parameters used in the Landau used as convolution filter is tricky. The fit is performed using Equation 4.10, and, to automate the process and ensure convergence, χ^2 is evaluated for each fit. If the χ^2 value is above a defined threshold, meaning that the fit has failed, the convolution and fitting process is repeated, changing the parameters of the Landau used as convolution function. The choice of such parameters is predefined. If the fit does not converge after all these parameters have been evaluated, the specific signal is flagged and sent to the user for further investigation.

In summary, this algorithm has reduced and improved the charge probability distribution fitting process. By calculating the convolution of the raw signal with a Landau distribution, the resulting curve is less susceptible to noise artifacts. The latter can be used to identify important initial parameters such as μ and the fit boundary of Equation 4.10, which helps the fit convergence.

4.5.2 Thermal measurements

Thermal measurements were performed with a thermal camera model "testo 883". The following plots were acquired immediately after irradiation at the proton therapy center in Trento.



(a) Front view (b) Back view

Figure 4.40: Thermal camera image of the ESA_ABACUS with fully alimented.

Chapter 5

Tissue Equivalent Proportional Counter Readout for HDM

In this chapter, I will illustrate the novel data acquisition system that employs Field Programmable Gate Array (FPGA) technology for the Tissue Equivalent Proportional Counter (TEPC) detector. I will structure this discussion using a bottom-up approach, beginning with the detector signal and ending with the data analysis process, with the FPGA serving as the final component in the acquisition chain. First, I present an introductory overview of the acquisition chain, less detailed than what I covered in section 2.2, since it is common to both systems. This is followed by an in-depth examination of the hardware used for the FPGA implementation, focusing on the Eclypse Z7 development board. Both the Programmable Logic (PL) and Programmable System (PS) implementations will be thoroughly examined, as this development board hosts a dual-core Cortex-A9 processor integrated into the FPGA. Finally, I move on to the data analysis and calculation of the microdosimetric spectra.

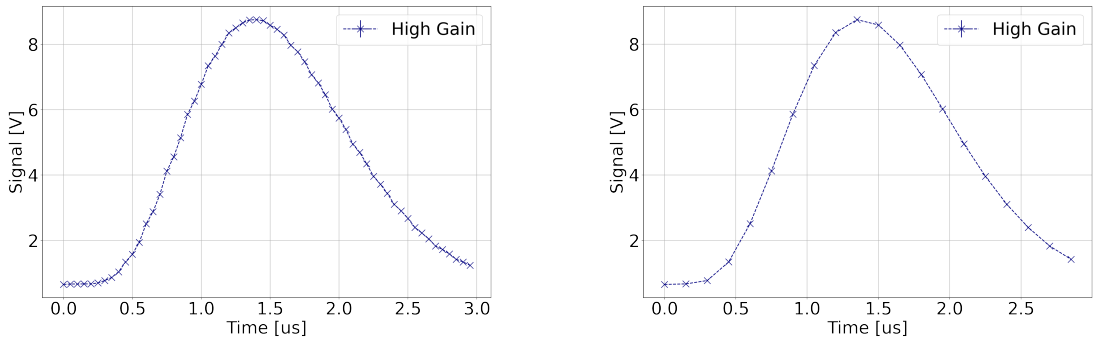
5.1 Introduction

This section offers a compact overview of the acquisition chain; a more comprehensive discussion is available in section 2.2.

5.1.1 Acquisition chain

The main goal of the acquisition chain is to convert the signal from the Tissue Equivalent Proportional Counter (TEPC) into a usable signal from which useful information can be extracted.
The TEPC signal comes from the secondary charged particles produced when radiation interacts with the gas inside the detector. However, the current intensities produced by radiation and collected are usually extremely small (few tens of µA [Rossi et al., 1996]), and for this reason cannot be used directly in the data analysis. Therefore, the first stage of the acquisition chain is a charge-sensitive preamplifier, model A422A from CAEN [CAE, a]. The signal is then amplified in three different gain branches: high gain with a gain of ~ 1000, medium gain with a gain of ~ 100, and low gain with a gain of ~ 10. This is because the dynamic range of microdosimetry can span several orders of magnitude while requiring high resolution. The selection of gains was made to ensure some overlap between the high-medium and medium-low signal ranges. Consequently, certain signals are shared between high and medium gain, while others are shared between medium and low gain. Each gain stage is performed by shaping amplifiers that shape the signal to a time constant of 2 µs. Two CAEN model N968 shaping

amplifiers [CAE, b] were used for high and medium gain, while an Intertechnique 7243E amplifier was used for low gain. As a reference, a graph of the signal shaped after the shaping amplifier, amplified by the high gain, is shown in Figure 5.1a.



(a) (b)

Figure 5.1: TEPC signal acquired after the high gain (×1000 gain) shaping amplifier. The shaping constant was set to 2 µs. In 5.1a a total of 60 samples were acquired with a sampling frequency of 20 MHz. While in 5.1b, a total of 20 samples were acquired with a sampling frequency of 6.667 MHz. Error bars are not visible. Lines are drawn as a guide for the reader's eye.

In the previous acquisition chain, after the amplification stage, the signal was sent to Multichannel Analyzers (MCAs) for data acquisition and subsequent processing. In the new acquisition setup, however, the last stage is replaced by the Eclypse Z7 Development Board. Instead of being processed by MCAs, the three signals from the amplification stages are acquired by three analog-to-digital converters (ADCs) integrated on the Eclypse Z7 Development Board. This design offers a huge potential for customization in data processing. Although the MCAs process the signal in a predefined manner that remains inaccessible to the user, storing the complete time series using ADCs allows the user to process the data according to their specific needs. By incorporating an FPGA to oversee both the ADCs and the data flow, the users gain the capability to implement real-time data processing directly on the board. This means that they can execute customized data processing tasks and algorithms in real time, harnessing the computational power of the FPGA to tailor data analysis to their exact requirements. This level of flexibility and control enhances the versatility of the system, allowing for on-the-fly data manipulation and analysis, which can be particularly advantageous in dynamic and rapidly evolving research scenarios. This level of flexibility is imperative for the successful implementation of the Hybrid Detector for Microdosimetry (HDM). The ability to timestamp the data, tailor trigger mechanisms to specific conditions, and transmit signals generated through the use of FPGA for acquisition is absolutely essential. In order to meet the precise synchronization requirements of HDM between the TEPC and LGAD layers, as well as other components in general, signal pulses must be generated at specific times and therefore the system must be adapted to send such signals. All this flexibility comes at the cost of actually implementing this functionality on the FPGA. As usual, this is not a trivial task, since the FPGA programming language directly targets the hardware that has to be physically implemented; moreover, debugging is difficult and requires a thorough understanding of both hardware and software.

5.1.2 Eclypse Z7 Development board

Eclypse Z7 Development board, developed by Diligent, hosts a Zynq-7000 as the central logical core, the same unit available in the ZC702 development boards used for the LGADs readout described in subsection 4.1.4. Using an FPGA based on the Zynq architecture, both a Programmable Logic (PL) and a Programmable System (PS) are present.

Within the PL real-time logic is implemented from the user, executing time-critical operations that leverage the precise timing of custom-designed circuitry. It is worth highlighting that when programming an FPGA, the result is effective creation of an actual circuit within the hardware, embodying the specific functionality and performance characteristics needed for the application. The PS component, on the other, integrates a Central Processing Unit or CPU, such as a dual core Cortex-A9 in the case of the Eclypse Z7 Development board, which provides the flexibility to run software applications, interface with external devices, and manage system-level tasks. In this domain, the user can easily implement more complex logic typically executed on CPU. The Eclypse Z7 is also equipped with 1G Ethernet port, micro USB ports, micro SD card reader, two PMOD ports [PMO,] and two ZMOD ports, as well as other interfaces listed here [Dig,]. The ZMOD ports are following with the SYZYGY

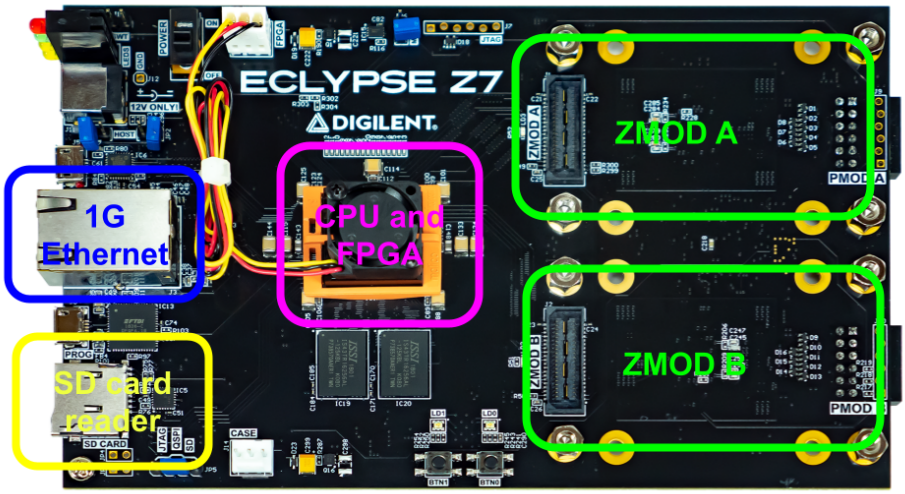


Figure 5.2: Top view of the Eclypse Z7 Development board. In the picture. the important connectors are highlighted: 1G Ethernet port in blue, SD card reader in yellow, FPGA and CPU in magenta and the ZMOD connectors in light green.

Standard interface specification [SYZ,], enabling the connection of external hardware. A picture of the Eclypse Z7 with the main ports highlighted is shown in Figure 5.2. Multiple devices are compatible with the ZMOD port standard. One component that stands out is the Zmod Scope 1410, which hosts a Dual-channel 14-bit Analog-to-Digital Converter Module [ZMO, d]. In this device, an AD9648 [AD9, a] integrate circuit is hosted. The AD9648 serves the actual purpose of ADC but, being an integrated circuit, it requires proper power supply and configuration. The Zmod 1410 provides the necessary service power and connections to the AD9648 to function properly, as well as the main connections of the device to the FPGA. The AD9648 is a complex circuit capable of sampling analog signals in different ranges and with different sampling frequencies. It can digitalize two channels, named channel A and channel B. To communicate with the user and configure the device options, the serial communication protocol is used. Such options relevant to this application are: clock divider, calibration, coupling, and range.

- The clock divider defines a multiplication factor that determines the sampling frequency of the ADC. The AD9648 can be clocked up to 100 MHz, but the clock is not used directly for sampling. Instead, the device uses a fast clock to maintain the timing precision of operations. It is then automatically down-converted to the desired sampling frequency by multiplying the fast clock by the clock divider factor. In this way, a lower sampling frequency can be obtained while maintaining the timing accuracy of a 100 MHz clock.

- Calibration is necessary to correct for drift or inaccuracy of the ADC output. A linear correction to the output values is applied:

$$y_{\text{calibrated}} = y_{\text{acquired}} \cdot m + q \,, \tag{5.1}$$

  with m and q calibration constant. The vendor already provides m and q specific to each of the two outputs of the AD9648. This operation corresponds to correct the gain (m) and the offset (q).

- Channel coupling can occur in AC (alternating current) or DC (direct current). If the channel is coupled in AC, any offset of the channel input is removed. This is the equivalent of filtering the continuous frequency of the signal. On the contrary, in DC coupling, the continuous component of the signal bandwidth is preserved.

- Range defines the maximum input of the analog signal. Two ranges are available: ± 25 V or ± 1 V. Both ranges have the same resolution of 14 bits. As a result, if a range of ± 25 V is selected, the resolution is 50 V (the spanned signal range) divided by the ADC bits $50\,\text{V}/2^{14} = 0.00305\,\text{V}$. In the other case of ± 1 V, the resolution is 0.1220 mV.

The user can configure these options via serial communication interface, with a SPI, CLK, and ENABLE lines. The communication protocol is described in [AD9, b]. Finally, the AD9648 output data can be formatted in different multiplex options. The AD9648 output configuration is versatile and can be adjusted to operate in "interleaved" output mode. In this interleaved mode, data from both channels is combined and sent over a single output bus. This configuration helps minimize the overall number of traces needed for data transmission. Figure 5.3 shown the how the interleaved output mode operates. This mode involves the transmission of two sets of data over a single line. To maintain synchronization, the data transmission is aligned with a clock signal generated by the AD9648 itself, named DCOA/DCOB. During the rising edge of the DCOA/DCOB clock signal, channel B data is transmitted, and conversely, during the falling edge of DCOA/DCOB, channel A data is sent. This alternating pattern continues, with each subsequent iteration transmitting the next set of data from both channel B and channel A.

Another important aspect of the AD9648 is a SYNC input. This input serves as a key component in achieving synchronized operation when using multiple ADCs in a system. It offers users the flexibility to align the timing of sample clocks, a critical factor in applications requiring precise data acquisition. The SYNC input can be configured in two modes: single occurrence synchronization and continuous synchronization. In the single occurrence mode, the ADC synchronizes with the SYNC signal once, allowing for coordinated data acquisition at specific time point. On the other hand, in continuous synchronization mode, the ADC aligns itself with every rising edge of the SYNC signal, ensuring ongoing and continuous synchronization. It is worth noting that achieving tight synchronization between multiple AD9648 units in a system requires external synchronization of the SYNC input signal. This extra level of precision ensures that all ADCs in the setup operate in perfect harmony, minimizing timing uncertainties, and ensuring reliable data acquisition across the system.

The vendor of the Zmod Scope 1410, Digilent, offers a valuable IP core, named *Zmod ADC 1410 Low Level Controller* [ZMO, c] designed to simplify the control and integration

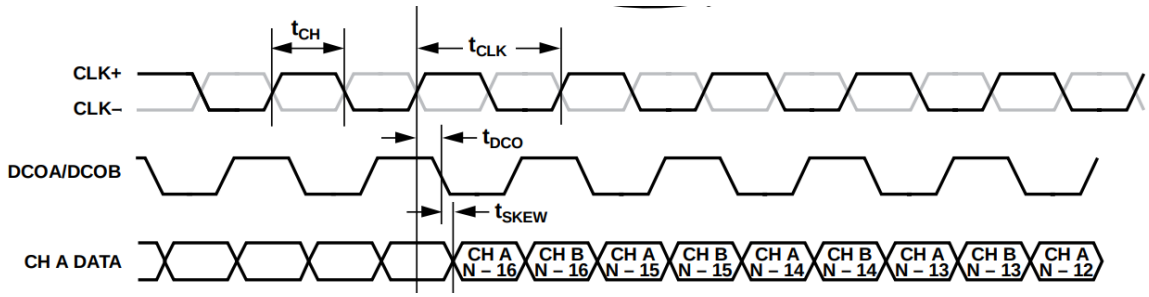


Figure 5.3: Interleaved output mode for the AD9648. Output data is sent in a single transmission line in sync with the DCOA/DCOB clock signal, generated directly by the AD9648. At the rising edge of DCOA/DCOB channel B data is sent, while at the falling edge of DCOA/DCOB channel A data is transmitted. This sequential alternation persists, where each successive cycle transmits the subsequent data sets from both channel B and channel A. Picture from [AD9, a].

of the AD9648 into applications. IP cores are pre-designed intellectual property blocks that serve as building blocks for FPGA-based designs. The Zmod ADC 1410 Low Level Controller is responsible for several features. It begins by initializing the hardware on the Zmod ADC 1410 by providing users with the option to overwrite the initial ADC configuration. As a consequence, it is possible to easily set initial parameters, such as calibration, coupling, and range. The IP core also provided demultiplexing of the data stream. Thanks to the Input Double Data-Rate Register (IDDR) primitive [xil, b], the IP core converts the data outputted as in Figure 5.3 into a two separate data stream for channel A and channel B in sync with a clock.

Another important feature of this IP core is that it can be controlled by another optional IP core, again provided by Digilent. This IP core, named *Zmod ADC 1410 AXI Adapter* [ZMO, b], includes the capability to control the Zmod ADC 1410 through a set of control and status registers accessible via the Advanced eXtensible Interface Lite 4 (AXI4-L) interface. The AXI4 Lite is a simplified variant of the full AXI (Advanced eXtensible Interface) protocol used for communication between components within digital systems, particularly in FPGAs. A description of the AXI4 lite protocol, used in a different implementation to configure the thresholds of the LGADs readout for HDM, can be found in section 4.2.1. The Zmod ADC 1410 AXI Adapter also incorporates a fundamental trigger mechanism for data acquisition, the implementation of a $16384 = 16 \cdot 2^{10}$ bits sample buffer for each of the two channels within the Zmod ADC 1410, and a streamlined AXI Stream interface that simplifies data transfer through a Direct Memory Access (DMA) engine. A similar implementation that relies only on the AXI Strem and DMA, without the use of the buffer and trigger mechanism, was used and described for the LGAD readout in section 4.1.4. The buffer is based on Block Random Access Memory (BRAM). Each bit of coming from channel A and channel B of the AD9648, further converted in the two data stream by the Zmod ADC 1410 Low Level Controller, is mapped in a BRAM of dimension 16384 bits. As a result, a total of 14 BRAM, each with a width of 16384 bits, are allocated for each bit of channel A, and an additional 14 BRAM of dimension 16384 bits, are allocated for each bit of channel B. The BRAM is controlled through the specific BRAM IP core [xil, a]. When selecting the BRAM dimension of 16384 bits, an address of 14 bits ($2^{14} = 16384$) is required to span the entire memory range. The address selects which of the 16384 bits of the BRAM will be read or written. A diagram of the design is shown in Figure 5.4. Each bit of the 14 bits of one channel is sent to its dedicated BRAM at a specific address. The data buffer is then transmitted according to the AXI stream protocol to an external DMA IP core, once a specific trigger condition is met. The

number of ADC samples sent to the DMA can be configured by the user through a dedicated register: called *AXIS_S2MM_LENGTH*. It can vary from a minimum of 1 bits to a maximum of 16384 bits, the size of the entire buffer. To define when the buffer is sent to the DMA, the Zmod ADC 1410 AXI Adapter offers the possibility of setting a trigger condition. In this way, the acquisition can be run in either triggered or triggerless mode. In triggerless mode, data is acquired continuously and sent to a DMA without waiting for specific trigger conditions. On the other hand, the trigger mode allows the user to define various trigger parameters, including the trigger channel (A or B), the trigger edge (rising or falling), the trigger amplitude level, and the position of the triggering sample within the acquisition window. This makes it possible to trigger the acquisition to be at a specific point within the time series. The trigger sample position is configured by setting the *Trigger Position Register*. By adjusting the trigger position register in conjunction with the AXIS_S2MM_LENGTH parameter, users can determine the number of samples to be acquired both before and after the trigger event.

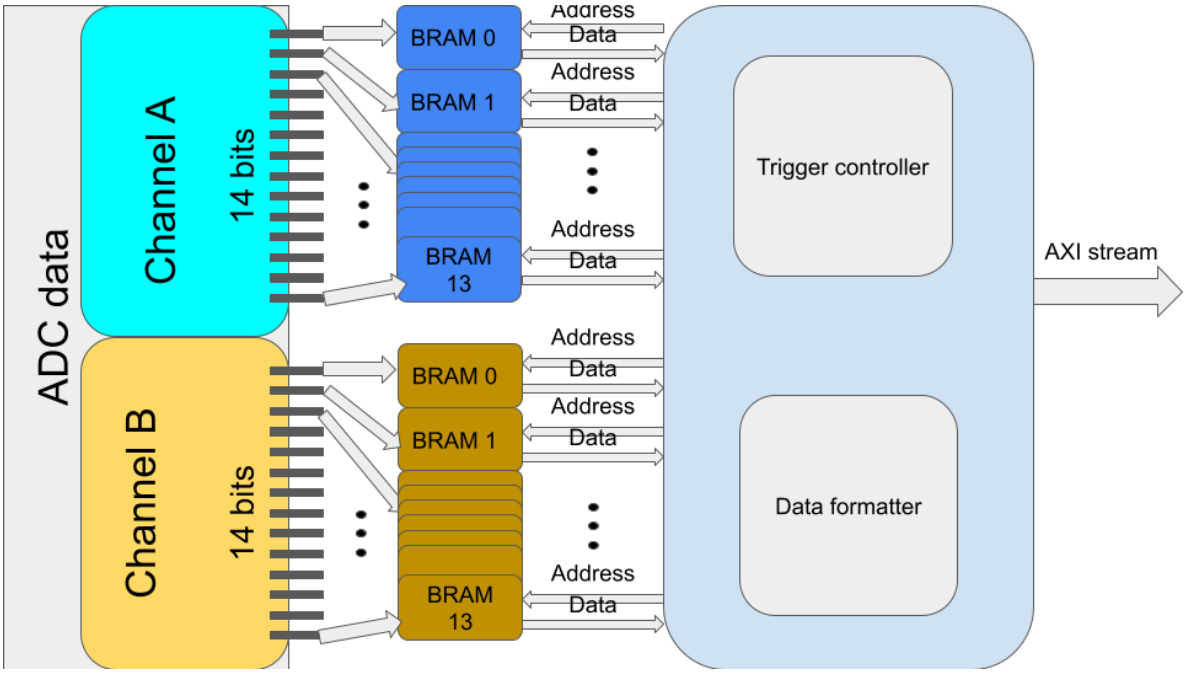


Figure 5.4: This diagram illustrates the operation of the Zmod ADC 1410 AXI Adapter. It shows how the 14 bits from each channel are stored in BRAM. In this configuration, the BRAM has a dimension of 16384 bits. It is possible to write individual specific addresses depending on the selected address, which ranges from 0 to 16383. Once the trigger condition has been met and the buffer is filled with *AXIS_S2MM_LENGTH* elements, the data is formatted and transmitted according to the AXI stream protocol to another device, typically a DMA, for further processing.

A simplified representation of the Finite State Machine (FSM) governing the AXI Adapter IP core is depicted in Figure 5.5. The `Idle` state is the initial state after a reset condition, and it remains in this state until the RunStop bit in the control register is set. Depending on the selected trigger mode, if the acquisition is in triggered mode, the state machine enters the `Arm` state (green arrow). If the acquisition is in triggerless mode, the state machine advances to the `Fill buffer none` state (red arrow). While in the `Arm` state, the state machine loads a number of samples into the BRAM buffer equal to the value set in the Trigger Position register. After this operation, it enters the `Wait trigger` state. This state ensures that a minimum number of samples, as specified in the Trigger Position Register, are captured before

the trigger condition is detected. In the `Wait trigger` state, the machine records incoming samples while monitoring the trigger source for a trigger condition. If the trigger condition is satisfied and the required buffer is already full, i.e. the trigger position coincides with the last buffered value, the state machine goes directly to the `Send buffer` state (transition not shown in the figure). On the contrary, if the value of the Trigger Position Register is less than the requested sample length, the state machine enters the `Fill buffer normal` state. In this state, the machine loads the missing buffer points to reach the requested AXIS_S2MM_LENGTH. Once the missing data is collected, a flag is set that indicates that the core has acquired a valid data set according to the requested sample length and trigger conditions. The machine then enters the `Send buffer` state. In this state, the requested data is sent from the BRAM to the DMA engine using the AXI Stream protocol. When the transfer is complete, the machine enters the initial `Idle` state. In the case of triggerless acquisition, the `Fill buffer none` state is designed to load the required number of samples according to the AXIS_S2MM_LENGTH register. When data acquisition is complete, the device simply enters the `Send buffer state`, where the data can be sent to the DMA.

Finally, Digilent provides libraries to control the Zmod ADC 1410 AXI Adapter. These libraries are available for both a Linux implementation, running a Linux Operating System on the PS, and for a bare metal implementation [ZMO, a]. They are capable of controlling and setting all the main parameters of the acquisition system, including: trigger condition, channel coupling, calibration, and DMA. In this design, the DMA is controlled by a program that runs in the userspace, unlike what was implemented in section 4.1.4, where the DMA was mainly controlled by a kernel driver. This approach allows the user to perform operations directly on the DMA, making debugging easier. On the contrary, direct access to the resource requires the user to take special care when requesting operations. The reason for this is that if the user has unrestricted direct access to the hardware, the user can cause damage to the device in the event of incorrect operations. In fact, Digilent preferred to use an existing implementation of DMA libraries [DMA,] that have been tested and validated. It is worth noting that this driver does not support scatter-gather; therefore, with this design, memory can only be written to contiguous addresses. A common approach is to define a memory buffer of the appropriate size where the data is expected to be written by DMA. In this way, the user can then read this memory region and extract the data sent by the PL.

To assess the seamless integration of all the components mentioned above, namely the Zmod ADC 1410 Low-Level Controller, Zmod ADC 1410 AXI Adapter, and a Linux OS, Digilent offers a fully functional example [ZMO, a]. This working example encompasses a hardware implementation within Vivado 2019.1, capable of operating a single ZMOD ADC 1410. The management of IP cores is facilitated through the provided libraries. Users have the flexibility to choose between executing the program in either a bare metal implementation or within a Linux OS environment. In the latter scenario, Digilent provides an SD card image containing a bootable Linux OS. Like the LGAD readout, Linux-based designs are preferable due to their flexibility and ability to implement complex operations. The core functionality of this example is to drive the ADC at a sampling frequency of 100 MHz, capturing 1024 samples, and setting the range to ± 25 V. This acquisition operates in a triggerless mode and, leveraging the Linux system, writes the sampled data on the SD card in a Comma Separated Value (CSV) format. Figure 5.6 shows a diagram of the data and configuration flow developed in the provided example. The data path is represented by the red arrows: the data is sampled by the AD9648 (not shown in the figure) and immediately processed by the Zmod ADC 1410 Low-Level Controller, which decouples the data in the two channels, A and B. The Zmod ADC 1410 Low-Level Controller is also responsible for the data decoupling and serial communication with the AD9648. This device, as explained before, can operate in different ways, and for this reason, must be configured properly. The communication and configuration

data flow are shown by the light green arrows. Moving to the Zmod ADC 1410 AXI Adapter, it is used to configure the Zmod ADC 1410 Low-Level Controller via the AXI4-L protocol. In this way, it is possible to perform the required operations from Linux user space by simply writing memory addresses corresponding to specific configuration registers. The 1410 AXI Adapter also processes the received data by sending it to Direct Memory Access (DMA) in compliance with the AXI4 stream protocol. This allows data to be quickly moved to memory in a buffer area. A 32 bit buffer, corresponding to C++ type `int32_t`, was used to store the ADC data from both channels as each channel samples 14 bits for every point, for a total of $14\,\text{bits} \cdot 2 = 28$ bits. The remaining 4 bits were unused. The buffer can then be read by the user, providing the acquired ADC data.

I used this example as a starting point for implementing the TEPC readout.

5.2 Implementation

In this section, I will go into the practical implementation of the TEPC readout system. Beginning with a concise overview of the requisite adaptations to the provided example by Digilent, I illustrate the implementation process, adhering to a bottom-up methodology. I start by describing the programming logic (PL) implementation and end with the processing system (PS) implementation.

The example provided by Digilent and illustrated in the previous section, subsection 5.1.2, was used as a starting point to develop the final design of the TEPC acquisition system. Hoverer, the example does not meet all of the requirements necessary to replace the MCAs. Some adaptations are required:

1. Expanding to two ZMOD ADC 1410s: To align with the demands of TEPC acquisition, which necessitates at least three channels (one for each amplifier), it is worth noting that a single ZMOD ADC 1410 can support up to two channels. Consequently, the example design requires expansion to incorporate two ZMOD ADC 1410s, thereby accommodating a total of four potential input channels. As a result, special attention must be paid to the synchronization of the two ZMOD ADC 1410s. As mentioned above, the ZMOD ADC 1410 is already predisposed with a sync input to synchronize with multiple devices. Another challenge arises from the triggering condition. Triggering is handled by the 1410 AXI Adapter for just one ZMOD ADC 1410, implying the utilization of only one of the two available channels in the AD9648 as the trigger input. To ensure triggering for all signals, regardless of their amplification level, it is advisable to designate the high gain line as the trigger channel. This choice is favored because the smallest signals generated by the TEPC can be effectively observed through the high gain channel, thanks to its amplification capabilities. Consequently, both the medium and low gain amplification signals must be acquired synchronously with the high gain trigger for comprehensive signal capture. Therefore, the 1410 AXI Adapter requires a modification of the trigger condition. The modification of the AXI Adapter will consequentially requires the modification of the provided libraries to govern the configuration via AXI4-L from the Linux OS.

2. Sampling frequency optimization for TEPC Signals: Considering the well-defined signal duration of approximately 3-4 microseconds achieved through shaping amplifiers, the sampling frequency of 100 MHz (equivalent to 1/(10 nanoseconds)), adopted in the example, is excessive. Oversampling in this context is not only resource-intensive but also leads to unnecessary data points. A lower sampling frequency should be implemented, which reduces the number of data points to store. For instance, Figure 5.1a is acquired

at a sampling rate of 20 MHz, necessitating a total of 60 data points to cover a time span of 3 microseconds (60/ (20 MHz)). In contrast, Figure 5.1b displays the same signal but with an altered sampling configuration, specifically 20 data points sampled at a rate of 6.667 MHz. The signal remains recognizable despite the under-sampling. Consequently, it becomes evident that lower sampling frequencies can be effectively employed, leading to reduced data points without a significant loss of signal fidelity. This realization prompts the exploration of optimized sampling frequencies to enhance the efficiency of data acquisition and storage.

3. Data storage optimization: Although the provided example utilizes the Comma Separated Value (CSV) format for data storage, which is user-friendly and easily readable, it poses a significant bottleneck in terms of acquisition speed. To improve both storage efficiency and write speed, it is advisable to switch to binary data format.

5.2.1 Programmable Logic - PL

The Programmable Logic (PL), is where the actual FPGA design is realized. In this domain, the easiest way to meet the channel extension requirement is simply to duplicate the existing sample design. The first step is to add a second Zmod Scope 1410 to the unused ZMOD port available on the Eclypse Z7 Development board. Following the same architecture as shown in the example, this Zmod Scope 1410 is connected to a Zmod ADC 1410 Low-Level Controller, which in turn is connected to the Zmod ADC 1410 AXI Adapter. With this overall implementation, the design shown in Figure 5.6 is replicated in its exact copy; the only difference comes from connecting to the other ZMOD port. To keep the architecture as modular as possible, a second DMA engine is used in the design. Consequently, the use of a second DMA, requires the extension of the libraries provided in the example. To achieve synchronized triggering across multiple Zmod ADC 1410 AXI Adapters, direct modifications to the IP core are necessary. This new enhanced version of the IP core incorporates a configurable flag that allows the selection between the conventional internal triggering mode (named master triggering mode) and an alternative mode based an external logic signal (named slave triggering mode). In master triggering mode, the Zmod ADC 1410 AXI Adapter operates as originally designed, initializing data transfer when the specified trigger condition is met, with the only difference that a logical pulse is generated when the trigger condition is met. Conversely, in slave triggering mode, a Zmod ADC 1410 AXI Adapter depends on another Zmod ADC 1410 AXI Adapter configured as the master. In this alternate mode, instead of waiting for the trigger condition as in the `Wait for trigger` state depicted in Figure 5.5, the state transitions to the `Fill buffer` state upon receiving a logical pulse signal from the master. This adjustment ensures synchronized trigger operation across multiple units, aligning their actions effectively. To ensure consistency of this system, the state machines shown in Figure 5.5 implemented in both Zmod ADC 1410 AXI adapters and the AD9648 device require precise synchronization. In the absence of such synchronization, a critical issue arises where the samples loaded into the BRAM (Block RAM) buffer of the two Zmod ADC 1410 AXI adapters may become temporally misaligned. This timing mismatch makes the acquired data unusable. Therefore, it is paramount to ensure perfect synchronization between these components. Two strategies are employed to ensure proper synchronization. The first involves utilizing the same clock line, while the second relies on the dedicated sync port of the two AD9648. By sending appropriate sync signals to the Zmod ADC 1410 Low-Level Controllers and, consequently, to the AD9648s, the acquisition of samples is ensured to be synchronized. Furthermore, by using the "continuous synchronization" mode, it is possible to continuously ensure the correct timing throughout the acquisition process. A SERDES (Serializer/Deserializer) primitive, previously described in section 4.1.4, is used in serialized mode to generate a pulse signal that is precisely

aligned to the clock line. The SERDES is configured in a 4:1 ratio, where all input bits except the fourth are at zero logic level. In this way, it is possible to create a pulse signal that is four times faster than the input clock rate, with exact timing synchronization. The pulse is then sent to the AD9648s sync input. Using the SERDES to generate a pulse signal faster than the clock is a strategy to improve the timing alignment of the AD9648. In fact, by providing a faster signal, the circuit actually implemented in the FPGA takes extra care to propagate the signal with minimum delay and skew, stressing the design demands. The result is improved timing accuracy due to the faster signal requirements and, as a consequence, precise sampling time synchronization between the AD9648s. Although precise time alignment can be achieved by increasing the clock frequency of the signals, another limitation must be taken into account: lowering the AD9648's sampling frequency. As observed in Figure 5.1a, a lower sampling frequency is compatible with the acquisition. This offers advantages in terms of resource optimization. Both the amount of data transferred to the DMA and the overall memory and storage requirements are optimized by tuning the sampling frequency. To facilitate this adjustment, I leverage the AD9648 clock multiplication factor. This approach is essential because the actual FPGA implementation necessitates adherence to precise timing requirements concerning skew and delay. However, accommodating a wide range of clock frequencies, from the original 100 MHz in the example to the approximately 6 MHz used to sample Figure 5.1b, which is sufficient while maintaining signal reconstruction, is a formidable challenge. Vivado was unable to find an implementation of this design that maintained signal propagation integrity due to the very tight time constraint. This is primarily due to the necessity for all clocks to remain synchronized within the implemented FPGA design, a task that becomes increasingly challenging as timing accuracy must be rigorously maintained for all the clock frequencies. To ensure a realizable FPGA design without timing synchronization failure, I adjusted the main clock to operate at 50 MHz. Simultaneously, by configuring the AD9648's internal clock multiplication factor to 8, I achieved a sampling frequency of 50 MHz divided by 8, resulting in 6.25 MHz or 1/(0.16 µs). Remarkably, this adjustment allows to maintain precise timing accuracy. In fact, the sync signal can be generated from the 50 MHz clock using the SERDES in a 4:1 configuration, resulting in a pulse of 5 ns. In this way, it is possible to synchronize several AD9648 with sufficient accuracy.

Overall, with this modifications the design is now capable of satisfying two of the three requirements highlighted in the previous section:

1. 4 input channels capable of operating in tight time synchronization with a shared trigger condition.

2. Optimized sampling frequency to 6.25 MHz or 1/(0.16 µs), while maintaining a time accuracy of 200 MHz or 1/(5 ns). As a consequence, to span at least 3 µs, a minimum of 19 sample points are required.

The last requirement has been implemented in the Processing System (PS) and will be addressed in the next section.

5.2.2 Processing System - PS

The Processing System (PS) represents a critical component in this setup. It serves as the control center, responsible for managing various aspects of the TEPC readout system. This includes interfacing with the FPGA to initiate and oversee data acquisition, processing, and communication with external devices. The PS also plays a pivotal role in configuring system parameters, such as trigger settings and acquisition modes, ensuring seamless operation. In addition, the PS is used to manage DMA, a critical component responsible for sending the data stream from the PL to memory. By modifying the PL implementation, the libraries

provided in the example [ZMO, a] need to be adapted. With the incorporation of two Zmod ADC 1410 AXI adapters into the design, the example libraries need to be extended to control both. To achieve this, each function has been rewritten and renamed by adding a subscript "_2" to perform the required task not only on one Zmod ADC 1410 AXI Adapter, but on both. Special attention was dedicated on the function configuring the triggering. This because the Zmod ADC 1410 AXI adapters has been modified to work either in master configuration or slave configuration. Consequentially, only the master Zmod ADC 1410 AXI adapter must be configured with proper triggering condition since the second Zmod ADC 1410 AXI Adapter, is configured salve mode at the hardware level. When dealing with DMA transfers, two Zmod ADC 1410 AXI adapters would be able to perform DMA transfers by using two DMAs individually and independently. This is not the case. The libraries used by Diligent [DMA,] incorporate a DMA transfer mechanism that employs blocking code. In practical terms, this implies that the program must wait for each instruction to finish execution. Consequently, when the second Zmod ADC 1410 AXI adapter makes requests, it is forced to wait for the completion of the requests made by the first Zmod ADC 1410 AXI adapter. This synchronization constraint results in a performance limitation for DMA transfer requests and data reading. The sequential nature of DMA operations can cause bottlenecks, affecting the overall data acquisition efficiency and speed. However, by modifying the example, a significant improvement in performance has been achieved. Improvement includes:

- multithreading execution,
- a buffer for incoming data,
- further optimizing the example code, including compiling with the -O2 optimization flag.

Leveraging multithreading capabilities, I employed two distinct threads, each assigned a specific task, thereby fully harnessing the potential of the dual-core Cortex-A9 processor. This approach involved binding one thread to each core for efficient execution. The first thread was exclusively dedicated to handling DMA functions and preliminary data storage operations. By doing so, I minimized downtime between successive DMA operations, optimizing the overall system performance. The acquired data was initially stored in a RAM buffer, which was shared with the second thread. At the same time, the second thread is in charge of performing critical pre-run activities, including tasks such as triggering setup and calibration, followed by the execution of subsequent data processing operations. This thread was responsible for the smooth and efficient transfer of data from the shared buffer to permanent disk storage in the required format. I took special care to avoid typical problems that can occur when multiple threads share resources. Because I used a shared buffer and variables, it was critical to manage how threads accessed these resources. To make sure everything worked smoothly, I took steps to coordinate how threads interacted with shared resources. This prevented conflicts that could arise from multiple threads accessing the same resources at the same time. Finally, the data is stored on the SD card in binary format, which has proven to be the best choice for storage because it excels in both space efficiency and write speed. I formatted the binary file to consist of `int16_t` chunks, each representing a 16-bit integer, taking advantage of the DMA data transfer that occurs within a `int32_t` buffer containing information from both channels. Within the binary file, the initial element encodes the count of samples within each captured window. Subsequently, all other elements correspond to the data points acquired by the two ZMOD 1410 modules. The channel associated with each sample can be identified because the order in which the data is written to the file remains consistent. For visual reference, Equation 5.2 presents the first ten elements of the binary file, where "SP" denotes the sample

count per acquisition window, and C_M^N represents the 16 bits integer carrying the M^{th} sample from channel N. As an example, C_1^3 corresponds to the 3^{rd} ADC sample from the 1^{st} channel.

$$\mathrm{SP}, C_0^0, C_0^1, C_0^2, C_0^3, C_1^0, C_1^1, C_1^2, C_1^3, C_2^0 \ldots \tag{5.2}$$

For increased reliability in data corruption scenarios, users have the option to specify the maximum number of items to be stored within a file. When the maximum number of elements in a file is reached, the program automatically creates a new file with the same structure and a different name. This precaution ensures that if a file is lost due to unforeseen circumstances, the integrity of the remaining acquired data is maintained without critical disruption.

In general, the measures implemented have led to significant improvement in data acquisition speed. An entire system benchmark has been conducted with all 4 channels with a DMA transfer of 16 bits per channel. When acquiring 19 samples according to the time windows required, an average trigger frequency of $\simeq$ 18kHz was obtained. With this information, it is possible to estimate the average data throughput via Equation 5.3:

$$\overbrace{4}^{\text{channels}} \underbrace{19}_{\text{samples}} \overbrace{16}^{\text{bits}} \underbrace{18\,\mathrm{kHz}}_{\text{average trigger frequency}} \simeq 20\,\mathrm{Mbit/s} \tag{5.3}$$

This average transfer rate is typical for DMA transfer, as a similar value was obtained with a low gain avalanche detector readout implementation in subsection 4.3.1. However, the trigger rate of 18 kHz is already in the pileup range for a TEPC detector LET-1/2 as shown in section 2.4. Therefore, the acquisition rate is more than sufficient to operate the detector properly. In order to monitor the trigger rate and thus avoid potential pileup conditions, the new acquisition offers the possibility to monitor the trigger rate every second. This is achieved by a spawning a dedicated thread in the PS that simply checks every second how many triggering events have been recorded. However, the acquisition saturates above the $\sim$ 18 kHz limit. Therefore, it is not possible to distinguish between trigger frequencies above this limit.

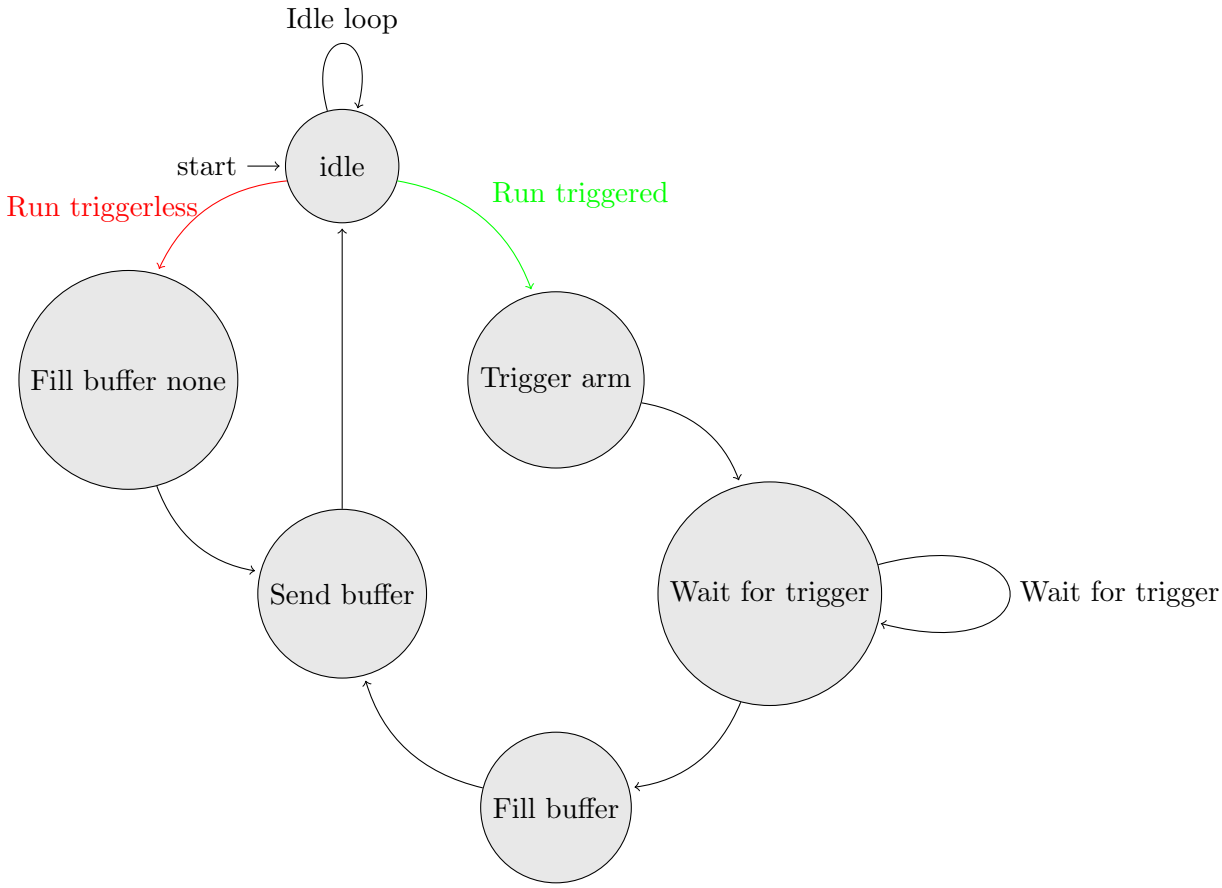


Figure 5.5: Simplified scheme of the Finite State Machine (FSM) implemented in the AXI Adapter IP. From the `Idle` state, depending on whether the acquisition is triggerless or triggered, the machine can enter the `Fill buffer none` state (red arrow - triggerless) or the `Arm` state (green arrow - triggered). From the `Fill buffer none` state, AXIS_S2MM_LENGTH data samples are loaded into the BRAM. Once all samples have been loaded, the device changes to the `Send buffer` state where the acquired amount of data is sent to the DMA according to the AXI stream protocol. In the case of triggered acquisition, the FSM enters the `Arm` state where a number of points equal to the trigger position are loaded into the BRAM. Upon compilation, the FSM changes to the `Wait Trigger` state where the trigger condition is continuously evaluated on new samples. Once the trigger condition is satisfied, the FSM is updated to the `Fill buffer` state where the remaining points are collected for a total of AXIS_S2MM_LENGTH points. After the remaining points are captured in the BRAM, the FSM changes to the `Send buffer` state where the captured data is sent to the DMA according to the AXI Stream protocol. When the transfer is complete, the FSM returns to `Idle` state.

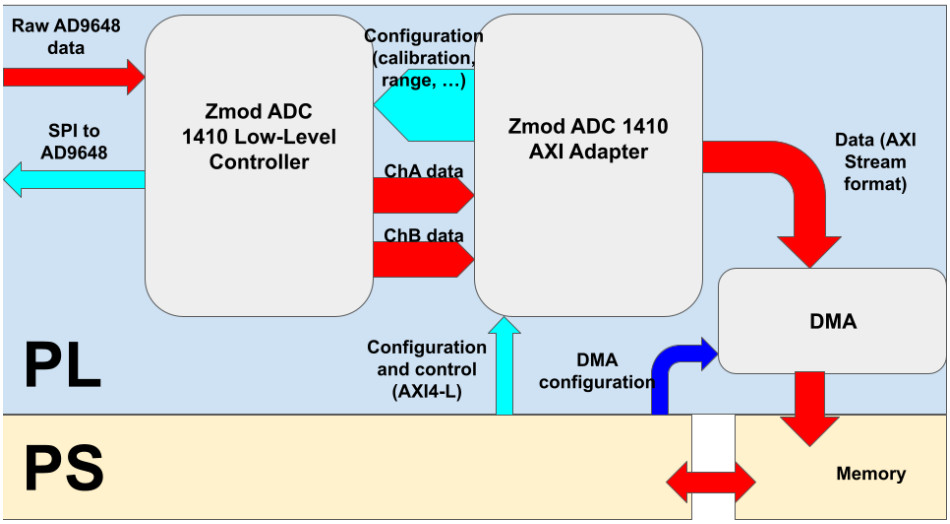


Figure 5.6: The figure illustrates the data and configuration flow implemented in the example provided by Digilent [ZMO, a]. The data path is depicted by the red arrows, starting with data sampling by the AD9648 (not shown). The sampled data is immediately processed by the Zmod ADC 1410 Low-Level Controller, which separates it into two channels, A and B. This controller also manages the serial communication and configuration with the AD9648 for configuration. The communication and configuration data flow is represented by the light green arrows. To pilot the Zmod ADC 1410 Low-Level Controller, the Zmod ADC 1410 AXI Adapter employs the AXI4-L protocol. The 1410 AXI Adapter further processes the received data by transmitting it to the Direct Memory Access (DMA) system using the AXI4 stream protocol. This facilitates rapid data transfer to memory, specifically to a designated buffer area. This buffer is readily accessible to the user, providing easy access to the acquired ADC data.

5.3 Data analysis

In this section, I will explore the data analysis techniques used to analyze the data provided by this new acquisition. Continuing with the bottom-up approach, I now assume that the data is correctly acquired in the form of 3 time series, one for each gain level: high, medium, and low. However, the experimental data was captured with a sampling frequency of 20 MHz, as in Figure 5.1a, without the optimized sampling frequency of 6.667 MHz mentioned in the previous section. This is because this optimization was not implemented at the time of the acquisition of the experimental data. In addition, with more sampling points in the time series, the data analysis can be fine-tuned and presented to the reader in a more comprehensive way.

I define a time series with n elements as $x_0, x_1, \dots x_{n-1}$, where x_i is the i-th element of the time series. The value of x_i is represented by a 16-bit integer, depending on the range configured in the ZMOD 1410s, it is possible to convert this value into a voltage signal.

5.3.1 Time series processing

Time series processing is essential to extract important information or features of information encapsulated in the time series. In particular, the amplitude of the acquired formed signal is relevant. The reason for this is that in the acquisition chain described in subsection 5.1.1, sharper amplifiers encode the information of the released energy in the detector as the amplitude of the output signal. Ideally, the difference between the minimum and maximum elements of the time series can be used to estimate an amplitude: $A := \max x_i - \min x_i$. Unfortunately, the time series can be altered by multiples such as noise, signal saturation, etc., making this simplistic approach ineffective. A more robust data analysis must be implemented in the selection of the minimum and maximum points to use as the difference. In addition, I want the data analysis to be able to distinguish and classify events based on some features such as the aforementioned saturation or excessive noise contribution.

Saturation detection

The first stage of data analysis is the identification of the saturating signals. As reported in the shaping amplifier manual [CAE, b], the maximum output value is 10 V. I found that this limit is not precise as higher absolute amplitudes of the signal were detected. This might depend on the offset of the baseline that has not yet been taken into account. To correctly classify saturating events, I count the number of points of the time series element above a defined threshold. In fact, signals that are saturating have a flat maximum, as in the case of Figure 5.13a where the high gain channel is saturating. If more than 2 points are found above the defined threshold, the time series is classified as a saturating signal and is labeled as "-saturation_by_max". This procedure was repeated for the time series of each channel. Figure 5.7 shows 5 sample time series that were found to be saturating using the previously described algorithm. Unfortunately, there is no direct way to evaluate the efficiency of this method, since when dealing with experimental data, the only option is to manually check each time series for correct classification. This is not possible since a typical acquisition stores millions of time series.

A threshold value of 13.15 V was chosen through experimental investigation. This value reflects a conservative choice that allows for the misclassification of non-saturating events as saturating events. This is an acceptable choice since, by exploiting the data redundancy available when a channel is close to saturation provided by the same signal at a lower gain, there is no concern about data loss. If a channel is marked as saturating, it is likely that the lower amplification channel of the same time series can be used instead.

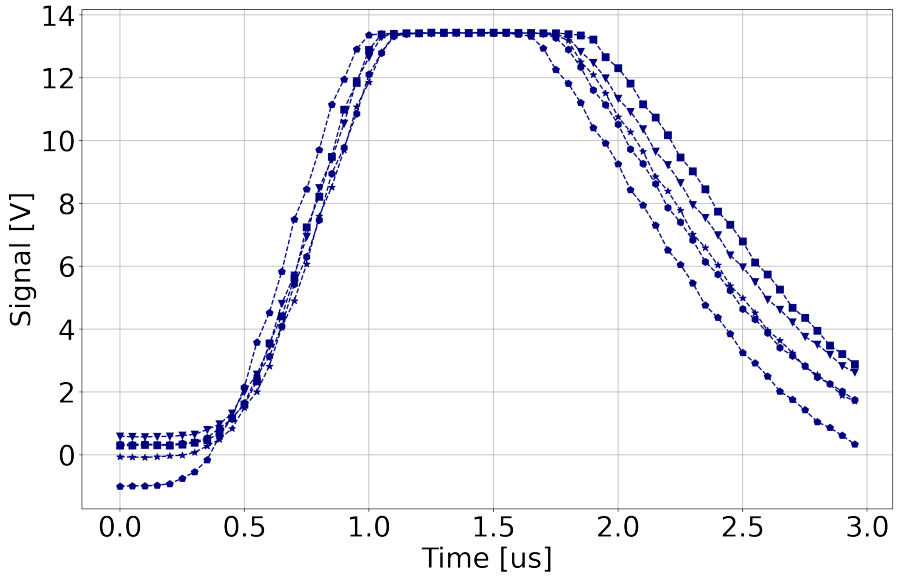


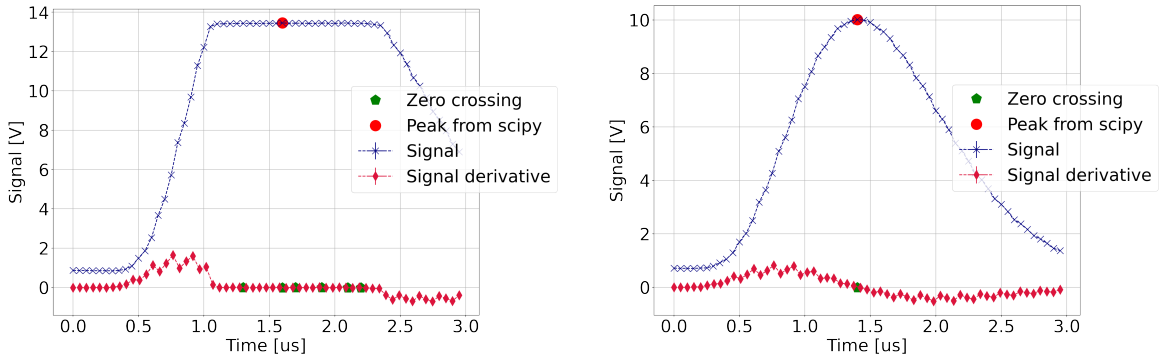
Figure 5.7: 5 time series acquired from the high gain channel. All time series displayed were classified as "saturation_by_max" and therefore expected to be saturating. Signals were generated by a 70 MeV protons beam impinging in 37 mm of RW3 water equivalent material to lower the protons energy. Error bars are not visible. Lines are drawn as a guide for the reader's eye.

Peak detection

Once the saturating time series have been classified and removed, the next step in data analysis is peak identification. Peak identification is performed in two ways: i) using the Python library Scipy and its peak finder function [Virtanen et al., 2020], or ii) by calculating the first derivative of the signal with respect to time and checking for zero crossing. Peak finding is the most critical step because there can always be multiple peaks in a time series. In addition, the trigger condition implemented in the Zmod ADC 1410 AXI Adapter simply relies on the positive crossing of a specified trigger threshold value. Once this value is exceeded by the time series, the acquisition is triggered. However, multiple hits may occur within the time series, or the acquisition may simply be triggered by the noise. For these reasons, the correct detection of the peaks is of paramount importance to avoid losing or distorting the acquired events. When coming to the Scipy peak finder function, a set of parameters are required to correctly identify time series peaks. If the parameters are not set correctly, real peaks can either be missed or misclassified from the time series noise. Again, by exploring different parameter values, the ideal choice was found to have a `height` = 1.4 and a minimum peak distance of at least `distance` = 40 points according to the definition of the `find_peaks` function [Sci,]. The function returns the index of the time series in which the peak is found. In this way it was possible to classify events having one or more than one peaks. The probability of having more than one peaks in the time series is directly related to the particles rate at the detector: higher particles rate are likely to increase the probability of having multiple peaks in a single time series. On the contrary, if a lower particle rate is selected, it is likely that single peaks are present in a time series. To benchmark this method, the first derivative with respect to the time is computed on the time series. This can be calculated numerically as the difference of adjacent elements:

$$x'_i = x_{i+1} - x_i \,. \tag{5.4}$$

The denominator can be omitted as it functions as a constant multiplier. In an ideal time series with at least one peak, once the signal reaches its maximum, it begins to decrease. This change after the maximum is detectable in the first derivative of the signal as it crosses the $y = 0$ axis from positive to negative values. Hence, the goal lies in identifying the zero crossing of the first derivative, or mathematically, finding j such that $x'_j > 0$ and $x'_{j+1} \leq 0$. The value of 'j' corresponds to the index in the time series where the signal peak occurs. Figure 5.8 shows two different time series with their respective first derivatives calculated with Equation 5.4. The Figure 5.8a shows the importance of finding time series with saturation in advance. In this particular case, the signal is saturating, making the first derivative method unsuitable. In contrast, in Figure 5.8b the time series exhibits a clean artifact-free trend. Both methods, either based on the Scipy peak detection function or based on the first derivative, correctly identify the time series peak.



(a) (b)

Figure 5.8: Time series acquired from the high gain channel using both peak detection algorithms. In Figure 5.8a the time series was previously classified as "saturation_by_max", while Figure 5.8b shows a normal peak, without any artifact. Peaks found using the Scipy library are represented by a red circle. To help the reader, the first derivative of each time series is also shown. The points where the first derivative crosses zero and changes sign from positive to negative are marked with a green pentagon. Error bars are not visible. Lines are drawn as a guide for the reader's eye.

For each time series, both methods are used to analyze it. The number of peaks and their locations are recorded with both approaches. In this way, it is possible to have a further benchmark on the effectiveness and reliability of the adopted peak finding algorithms. If more then one peak is found, the event is labeled as "-pileup_by_peak" as further processing is required to extract the amplitude information.

Baseline estimation

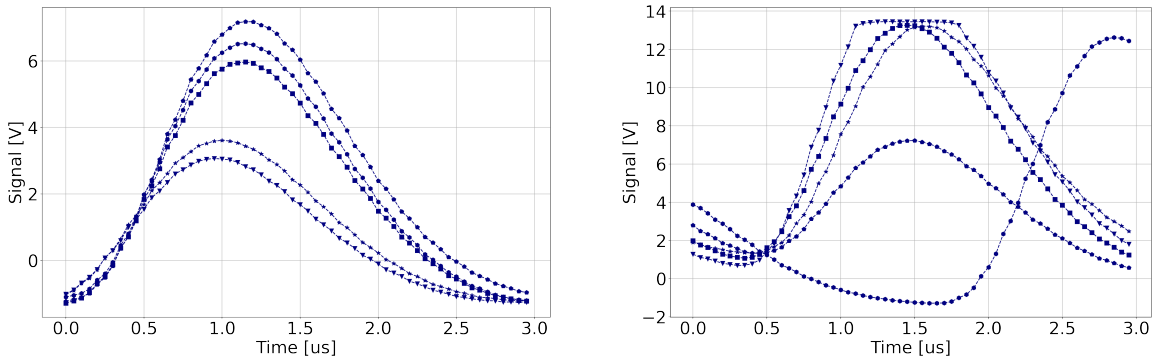
Once the peaks are found, the problem shifts to defining the correct baseline. To find a reliable baseline value, the acquisition uses the *Trigger Position Register* parameter. This register implements the ability to acquire samples of the time series before the trigger condition is met. The purpose of this is to evaluate those points of the time series before the trigger point where no signal is expected. Experimentally, I found that 8 points before the trigger point is enough to get a significant sample of points. To quantitatively characterize the baseline estimate, I compute both the average and the standard deviation of the aforementioned points. A reliable baseline typically exhibits a relatively low standard deviation. Elevated standard deviation values may indicate that the time series is affected by noise or that the tail of a

previous signal persists. In addition to that, to have a more comprehensive description, a linear fit with equation $y = mx + q$ is performed on the baseline points. The resulting offset value q can be used as an estimate of the baseline and compared with the previously calculated averages. The angular coefficient estimated from the fit, m, instead gives an estimate of the trend of the points: whether they are increasing or decreasing and how intensely. The number of time series points to be included in the calculation must be chosen carefully. Ideally, a large number of points is required to make the linear fit reliable and to reduce the error in the mean and standard deviation. However, the limited number of sampling points must be considered. Figure 5.9 shows how these methods can be used to classify different signals. In Figure 5.9a time series featuring a high standard deviation (above 0.5) in the points used for the baseline estimation are shown. A high standard deviation in the first points is a good indicator of a time series with scattered values. This is exactly the case as the first elements of the depicted time series are already showing a positive tread typical of the initial part of the signal. As a consequence, the estimation of the baseline is inaccurate when considering these points. These time series typically exhibit an initial segment with values below zero, signifying a shifted baseline toward lower values. Consequently, when these signals eventually reach the trigger amplitude level, they are already partially formed due to their earlier initiation. This results in a delayed trigger event, since the entire time series is already partially formed. Therefore, the baseline value cannot be adequately estimated by analyzing the initial points. One possible option is to check the end points of the time series. If the acquisition is triggered in advance by a partially formed time series, it is likely that the last points of the series where the signal returns to the baseline will also be captured. In fact, Figure 5.9a also captured the tail of the signal. This tail is more stable compared to the initial points of the time series, where the signal is sharply rising. As a consequence, the tail points of the time series represent a better, more stable, data source in estimating the baseline. Finally, by experimental trial, standard deviations above 0.5 were classified as "-bad_by_std", requiring additional dedicated analysis to estimate the baseline.

Figure 5.9b instead shows the results obtained by analyzing the fit parameters of the first time series points. In particular, Figure 5.9b shows time series where the linear fit has negative slope values. As before, this may indicate the tail of a previous signal. In such cases, the baseline cannot be reliably estimated because the time series does not contain points where there is no signal. Unlike what was observed in Figure 5.9a, the tail of Figure 5.9b signals is not usable because the signal contribution is still too high. These events were labeled as "-bad_by_m". Now that two estimations of the baseline are available, either via the intercept coefficient estimated using the linear fit or the average performed on the same points it is finally possible to define a signal amplitude. For each time series, the amplitude is calculated simply as the difference between the point identified as the peak and the baseline. I found no significant discrepancies in using q or the average value as a baseline. Therefore, the two methods are considered equivalent.

Advanced signal classification

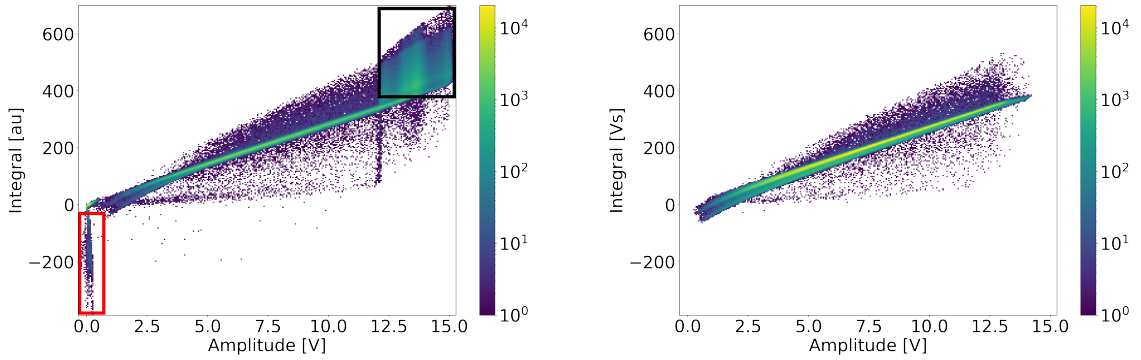
Calculating the integral of the time series is another way to further verify that the time series contains artifacts. By comparing the previously estimated signal amplitude with the integral of the time series, another discrimination can be performed to further distinguish unexpected behavior. There are several ways in which this integral can be calculated numerically. One of the most common methods, also used in this work, is the use of the `trapezoidal rule` [Atkinson, 1991]. Prior to computing the integral, it is essential to determine the baseline value, as the entire time series can potentially be shifted. By subtracting the estimated baseline values using the aforementioned methods, the correct integral value of the time series



(a) (b)

Figure 5.9: Time series obtained from the high gain channel and classified based on baseline parameters. This evaluation involved considering the standard deviation and applying a linear fit to the first eight data points. The set of five time series shown in Figure 5.9a had a significantly high standard deviation exceeding 0.5. In contrast, Figure 5.9b shows the five time series where the linear fit indicated a negative intercept. In both scenarios, relying solely on the initial data points of the time series proved insufficient for an accurate baseline estimation. Error bars are not visible. Lines are drawn as a guide for the reader's eye.

can be calculated. This integral value plays a crucial role in the identification of complex scenarios that involve multiple peaks. For instance, consider a scenario where two signals occur very closely in time, resulting in a single peak being visible in the time series. In such cases, the inversion in the first derivative may fail to detect this situation, and conventional peak search algorithms may prove ineffective. However, by evaluating the integral value, it becomes possible to discern this condition. The close superposition of several events is expected to yield a longer signal, consequently resulting in a larger integral value compared to a single, normal signal. On the contrary, a typical signal of exactly the same amplitude is expected to have a smaller integral. Generally, a well-defined proportional relationship between amplitude and integral is expected. If this proportionality is not satisfied, either the time series is not being processed correctly, or some artifact is present. In Figure 5.10, I show the previously mentioned relationship between signal amplitude and integral. For every high gain time series, both the amplitude and the integral were computed. The figure visually represents these values for each time series, forming a 2D histogram to enhance readability. The color map reflects the density of points at specific integral and amplitude values. To illustrate how this method can detect misclassification or signal affected by artifacts, Figure 5.10a shows the plot obtained considering all events, regardless of how they were labeled. Therefore, this plot includes saturating events, double hits, and other time series affected by artifacts. Figure 5.10b, on the other hand, is formed only by events classified as good events. By comparing the plots in Figure 5.10, some particle populations can already be identified. In particular, the absence of the particle population enclosed in a red triangle of Figure 5.10a suggests that they were removed due to classification. Judging by the negative value of the integral characterized by this population, it is suggested that these time series might have a negative initial slope. The low amplitude value suggests that these time series may represent the residual tail of a particularly long signal. The acquisition was then triggered by noise.
An interesting subset, also removed due to labeling, is enclosed by a black square in Figure 5.10a. This subset has a limited amplitude range, but its integral continues to rise, distinguishing it from the broader signal populations. The limited amplitude range suggests that this subset consists primarily of saturating signals. The observed increase in integral



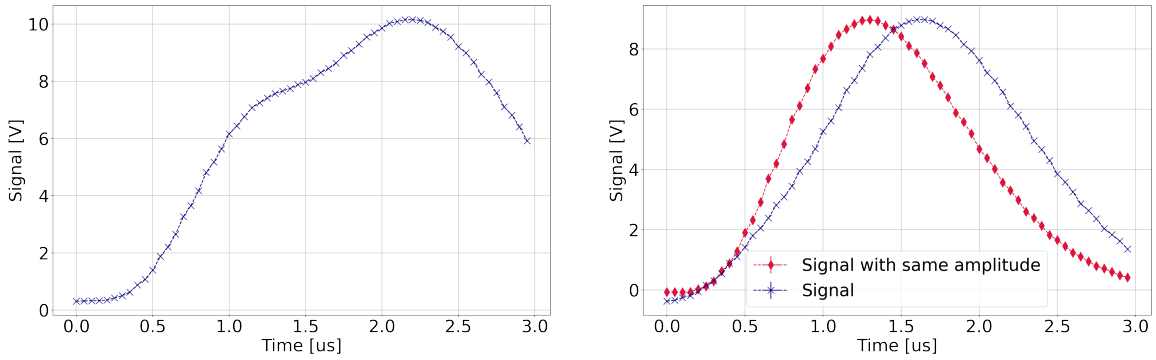
(a) (b)

Figure 5.10: 2D histogram calculated on the high gain time series. For each time series, the numeric integral and the amplitude were calculated and associated with a two-dimensional bin of the histogram. The color map represents the density of the points for each bin. Figure 5.10a contains 1382215 time series, without any distinction in classification. On the contrary, Figure 5.10b is composed of events classified as having no artifacts, totaling 950808 time series.

can be attributed to the larger area encompassed by these highly saturating signals. It can be noted that, even after the removal of events, there is a considerable population of events that deviate from the main trend. As an example, Figure 5.11 shows two signals deviating from the main tread reported in Figure 5.10b. The time series shown in Figure 5.11a clearly show the contribution of two events. Combining these temporally closed events produces a hard-to-detect double peaked time series. Because the second event prevents the signal from decaying, the first events do not produce a distinct peak. Such signals invalidate the efficiency of both previously described peak detection methods.

Moving our attention to the second plot in Figure 5.11b, the time series represented by the blue cross seems to show typical behavior. However, a comparison with a time series of the same amplitude, represented by the red diamonds in Figure 5.11b, shows that the signal width is significantly different. Several factors could contribute to this discrepancy. I speculate that this situation is similar to the extreme condition shown in Figure 5.11a, where two events occur very close to each other. This close proximity may interfere with the proper processing of the signals by the shaping amplifier, resulting in the observed differences in signal width.

It is safe to assume that the time series of Figure 5.10b that deviates from the ideal amplitude-integral relationship requires further investigation. To automatically identify and classify these points, a linear fit of the form $y = mx + q$ is performed on all points of Figure 5.10b, providing a linear estimate of the amplitude-integral behavior. Figure 5.12 shows the linear fit performed as a red line. Once the line equation is available, it is possible to add an offset of any value. This new shifted line can be used as a boundary to classify events that deviate from the ideal behavior. Of course, the offset cannot be chosen arbitrarily, but must be determined in some analytical and ideally automated way. To achieve this objective, I have determine the boundary intervals that encompass all the data points. To do so, I identify the farthest point above the linear fit, denoted as (x_u, y_u), and the farthest point below the fit, denoted as (x_b, y_b). Subsequently, I calculate two linear equations, both parallel to the fitted line equation. One line passes through the point (x_u, y_u), while the other passes through (x_b, y_b). This process yields two lines that are parallel to the fitted line equation and collectively enclose all the data points. Ultimately, I can determine the displacement between the fitted line equation and the two boundaries, denoted o_u for the upper boundary and o_b for the lower boundary. Subsequently, the user has the option of establishing a new parallel line,



(a) (b)

Figure 5.11: Two time series that have been selected from the subset of events in Figure 5.10b that exhibit deviations from the expected amplitude-integral relationship. In Figure 5.11a, two distinct peaks are evident, but these are not correctly identified by the implemented peak detection methods; only the second peak is accurately recognized. On the other hand, Figure 5.11b presents two time series with identical amplitudes. Notably, the time series represented by the blue cross exhibits a longer signal duration compared to the time series depicted by the red diamonds. This extended signal duration results in a divergent integral value from the expected behavior, which is consistently followed by the red diamond time series. Lines are drawn as a guide for the reader's eye.

modulating its offset from the fitted line to a value within the range of o_u. The same procedure must be iterated to select another parallel line located below the fitted line, with an offset not exceeding o_b. This approach allows the user to establish a new interval, offering the flexibility to include all points or select only a specific subset from the plot in Figure 5.12. Furthermore, the boundaries of the interval are intentionally aligned with the primary trend observed in Figure 5.12, allowing the user to focus on isolating anomalous points that deviate from this overarching pattern with a degree of precision. To illustrate this technique, Figure 5.12 shows an example in which the two black lines parallel to the red ideal behavior line delineate the boundaries that encompass the selected points. Events that are outside the defined interval and therefore, anomalous, are labeled as "-bad_by_integral". It is worth mentioning that the events that were discarded with this method, using the boundaries of Figure 5.12 represent only the 1.5% of all the events in the figure. However, detection of these points required more complex tools, with specific techniques.

Thus, after applying all the above classification methods, each time series is assigned a label based on the various classification outcomes. For example, a label such as "-bad_by_integral-bad_by_m" means that the event failed both the integral-based classification and the baseline estimation. Time series with an empty label and a single peak are used as starting data to calculate the microdosimetric spectra.

From a coding optimization perspective, to enhance efficiency, the code has been optimized to leverage multithreading. This approach significantly reduces the analysis time for a dataset, allowing for swift processing. For example, when using a 32-core cluster to analyze a dataset containing $\sim$ 446000 time series, the computational time is minimized to less than one minute. This is an important aspect of the analysis, since a preliminary data evaluation can help to identify potential problems of acquisition or setup in advance.

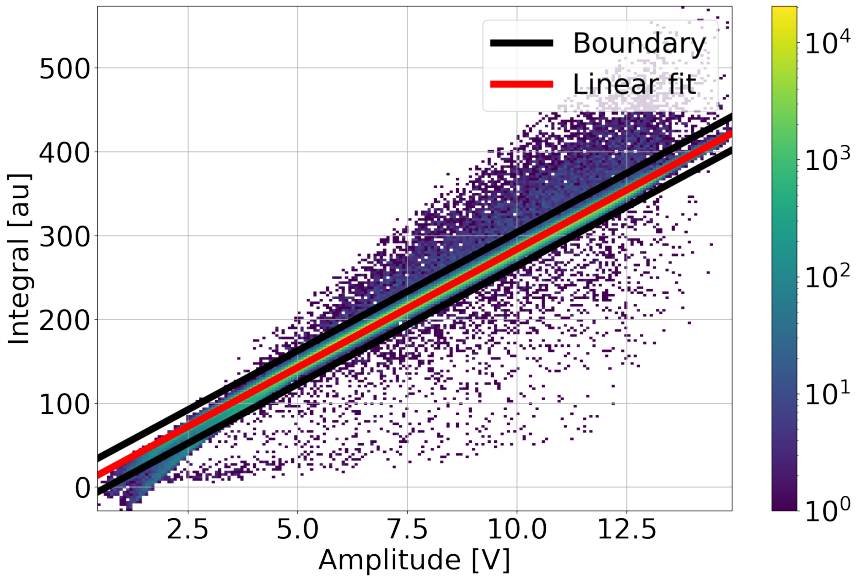
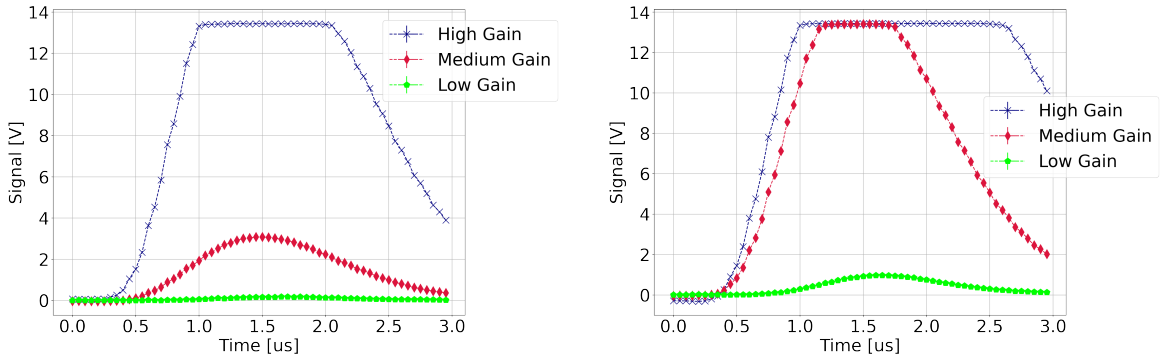


Figure 5.12: 2D histogram calculated on the high gain time series. For each time series, the numeric integral and the amplitude were calculated and associated with a two-dimensional bin of the histogram. Events where artifacts were identified were removed. In addition, a linear fit is performed and sown by the red line. The calculated fit coefficients are $m = 28.207 \pm 0.003$ and $q = 1.86 \pm 0.03$. Black lines parallel to the aforementioned fit were selected as boundaries between the accepted events and the anomalous events.

5.3.2 Intercalibration and joining

Thanks to the three gain stages and proper channel synchronization, no information is lost in the signal amplitude when a signal reaches saturation due to the limitations of the amplifier. This is achieved by carefully selecting the three different gains so that the same saturated signals can be handled by analyzing a different channel with a different (lower) gain. Figure 5.13 shows the three time series acquired with the new acquisition system, associated with the high- medium- and low amplification. In Figure 5.13a the high gain signal is saturating as a constant flat region is visible as the maximum of the time series. As mentioned above, if the high gain is saturating, the signal information can be recovered by observing a channel with lower amplification, the medium gain channel in this case. Finally, in Figure 5.13b, both the high and medium gain channels are saturating, leaving the low gain as the only usable time series for data analysis. When no saturation is present, but the same signal visible in at least two channels with different gain, it is possible to correlate the signal amplitude of visible channels. This process is called *intercalibration*, and was a critical procedure that allows the joining of the information extracted from the three different gains. Having now the possibility of analyzing the amplitude of event by event between the different gains, the intercalibration can be calculated directly during data acquisition. However, the signals coming from the detector must cover a wide range of amplitudes to collect sufficient statistics of common events between gains: high-medium and medium-low. In the unfortunate case that there are no common events, the correlation between the gain channels cannot be estimated. This correlation is uniquely dependent on the shaping amplifier configuration, ensuring that the same intercalibration can be reused as long as the shaping amplifier parameters are not changed. It is recommended practice to calculate the intercalibration each time data is analyzed to ensure that the acquisition process is operating as expected. Any deviation from the



(a) (b)

Figure 5.13: Time series acquired with the new acquisition of the three gain: high, medium, and low. The signal was generated using a pulser to emulate the detector signals. Figure 5.13a shows the case of high gain time series saturation and therefore the medium gain can be used to recover the information. Figure 5.13b shows the case of saturation for high and medium gain signals. In this case, the low gain time series must be used to process the data. Error bars are not visible. Lines are drawn as a guide for the reader's eye.

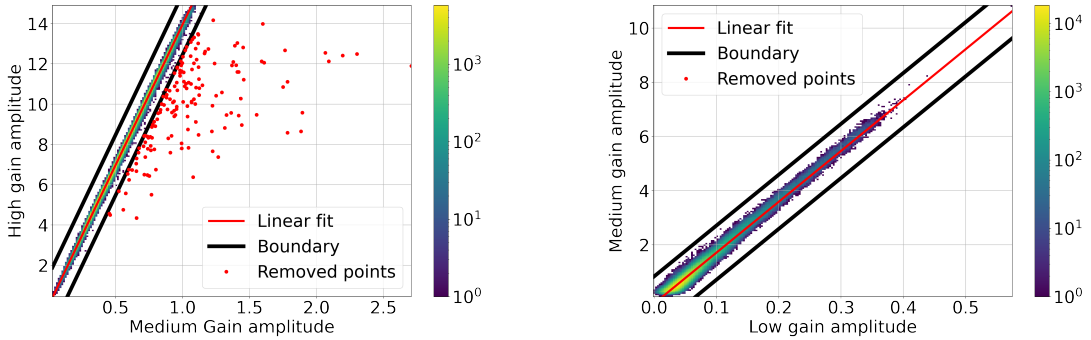
expected intercalibration may be a symptom of incorrect signal processing. It should be noted that this procedure was not feasible with the previous data acquisition method discussed in section 2.2 using MCAs as the final component of the chain. In fact, it was not possible to uniquely associate the same event on multiple MCAs. As a consequence, intercalibration must be performed before data acquisition with a pulser, as described in section 2.2. Figure 5.14 shows the above mentioned amplitude relationship for the high-medium (Figure 5.14a) and medium-low (Figure 5.14b). If the same signal is present in two time series, the amplitude estimate can be used as a point in Figure 5.14. In addition, a linear fit can be used to extract a relationship between channels with different gain. The performed linear fit is shown in red in Figure 5.14. Again, there are points that deviate significantly from the ideal behavior. To mark and exclude these points, a procedure similar to the one described in the section 5.3.1 was used. Two parallel lines to the ideal fit are drawn. The user can offset them at will to define a boundary of acceptable points, forming the black lines of Figure 5.14. This delimits the region of ideal behavior, allowing for the exclusion of points, which are marked with red circles, from being used. In addition, with the fit parameters obtained, it is possible to shift and correct the amplitudes estimated by the medium gain as they were captured by the high gain amplification. This procedure can also be iterated for the low to medium gain signals, and then again for the medium to high gain signals. Mathematically, using the calculated coefficients from the linear fits of Figure 5.14:

$$A_{\text{medium to high}} = A\_\text{medium} \cdot m_{\text{h-m}} + q_{\text{h-m}} \,, \tag{5.5}$$

A_medium is the amplitude evaluated from the medium gain time series, while $m_{\text{h-m}}$ and $q_{\text{h-m}}$ are the coefficients of the linear fit obtained from Figure 5.14a. In this way, a new amplitude $A_{\text{medium to high}}$ can be calculated. The same procedure can be iterated for the low to medium conversion:

$$A_{\text{low to medium}} = A\_\text{low} \cdot m_{\text{m-l}} + q_{\text{m-l}} \,. \tag{5.6}$$

This time A_low is the amplitude estimated from the low gain time series. $m_{\text{m-l}}$ and $q_{\text{m-l}}$ are instead the coefficient obtained from Figure 5.14b. $A_{\text{low to medium}}$ can be further converted from a medium gain signal to a high gain signal using Equation 5.5. At this point, when



(a) (b)

Figure 5.14: Amplitude relationship of time series belonging to the same event but processed by a differed gains. Figure 5.14a shows the amplitude relationship between the high and medium gain, while Figure 5.14b shows the amplitude relationship between the medium and low gain. A linear fit, shown as a red line, is calculated to estimate the main trend. Two lines parallel to the ideal fit, shown in black, are used to define the boundaries of an interval. Points outside this interval, marked by red circles, are discarded from further analysis because of their unexpected behavior. Parameters obtained from the linear fit of Figure 5.14a $m_{\text{h-m}} = 13.9628 \pm 0.0007$, $q_{\text{h-m}} = -0.0222 \pm 0.0004$. Parameters obtained instead from the linear fit of Figure 5.14b $m_{\text{m-l}} = 18.769 \pm 0.003$, $q_{\text{m-l}} = -0.1653 \pm 0.0002$.

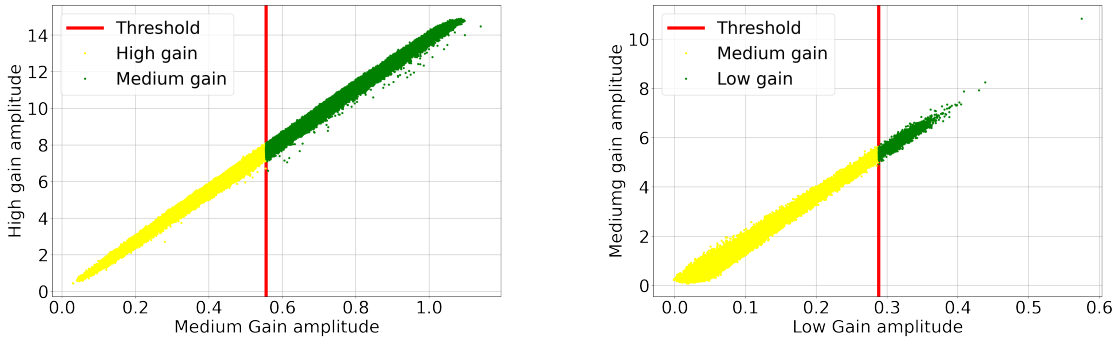
all the amplitudes are related to the high gain amplification, it is possible to build a unique histogram with the information from all the amplifications.

Since intercalibration events are correctly processed by two channels at different amplifications, it is important to define a rule to determine which event is to be considered. This prevents double counting of these common events. A simple vertical line in the intercalibration plot, defined by the user, is used as a separation threshold. Figure 5.15 shows the discrimination obtained for the high-medium gain in Figure 5.15a, and for the medium-low gain in Figure 5.15b. The threshold level was selected as half of the amplitude range on the x-axis. Having now all the time series with an amplitude related to the high gain amplification, without the issue of double counting, it is possible to create the first histogram that covers the entire acquisition amplitude range. Figure 5.16 shows the histograms of the amplitudes, obtained by joining the three different gain. Each gain component is highlighted in different colors. This is the last step of the intercalibration and joining, as now it is possible to build a spectrum. However, a calibration is missing to finally obtain a microdosimetric spectrum with physical meaning.

5.3.3 Calibration

The calibration procedure was conducted based on the procedure of [Moro et al., 2015]. This procedure essentially identify fixed points in the spectra that can be used as calibration points. One of these points is the proton edge: the maximum amount of energy that a proton can release inside the microdosimeters. Therefore, to satisfy this condition, the end of the proton range must coincide with the far end of the detector boundary [Moro et al., 2015] .

To calibrate the detector, an experimental campaign was carried out at the Proton Therapy Center in Trento. The proton beam was degraded from an initial energy of 70 MeV by 37 mm of RW3 water equivalent material. The TEPC was placed at the isocenter (defined 125 mm away from the beam exit) and data was acquired and analyzed with the aforementioned procedures.



(a) (b)

Figure 5.15: Amplitude relation of time series belonging to the same event but processed by differed shaping amplifier with different gain. Points marked as anomalous of Figure 5.14 were omitted. A red vertical line is drawn as a threshold. This threshold defines the origin of the signal, distinguishing whether it is selected from one channel with a certain gain or from the other channel with a different gain. This step is essential to avoid double counting. Figure 5.15a shows the threshold separating high- and medium- gain events. Figure 5.15b shows instead the threshold separating medium- and low- gain events.

Simulations were performed implementing the same experimental setup in Geant4. Then the microdosimetric spectra was calculated. It was found that 37 mm of RW3 are the best choice to obtain protons whose energy is likely to perform proton-edge depositions.
Once the experimental spectra are acquired and analyzed, it is possible to identify the region of spectra typical of the proton edge. This region is expected to follow a Fermi-like function [Moro et al., 2015]:

$$f(x) = \frac{A}{1 + \exp\left[B\left(x - C\right)\right]}\,. \tag{5.7}$$

Subsequently, I can perform a fitting operation on this specific spectral region using Equation 5.7. In Figure 5.17, I present the processed experimental data, highlighting the region of spectra where the proton edge contributes. The outcome of the fitting process for the Fermi-like function is represented by the red line in Figure 5.17. It has been proven that the best reference point, in terms of stability and precision, can be derived from the Fermi-like function. Such a point represents the intercept of the tangent through the inflexion point with the x-axis [Conte et al., 2013, Moro et al., 2015] and can be written as:

$$x_{\mathrm{TC}} = \frac{2}{B} + C \tag{5.8}$$

This reference value can then be used as calibration point. I used as reference value the one obtained via Monte Carlo simulations performed with Geant4 in the same experimental configuration. The same procedure was adopted to find the $x_{\mathrm{TC}}^{\mathrm{sim}}$ of the simulation. Having acquired this information, the calibration can be performed simply by multiplying the estimated experimental amplitudes by the ratio of the x_{TC} coefficients [Conte et al., 2013, Moro et al., 2015]:

$$A_{\mathrm{calibrated}} = A \cdot \frac{x_{\mathrm{TC}}^{\mathrm{sim}}}{x_{\mathrm{TC}}^{\mathrm{exp}}} \tag{5.9}$$

Subsequently, it becomes possible to establish a correlation between the signal amplitude A and the line energy y. This correlation is visually presented in Figure 5.18a, where the ratio

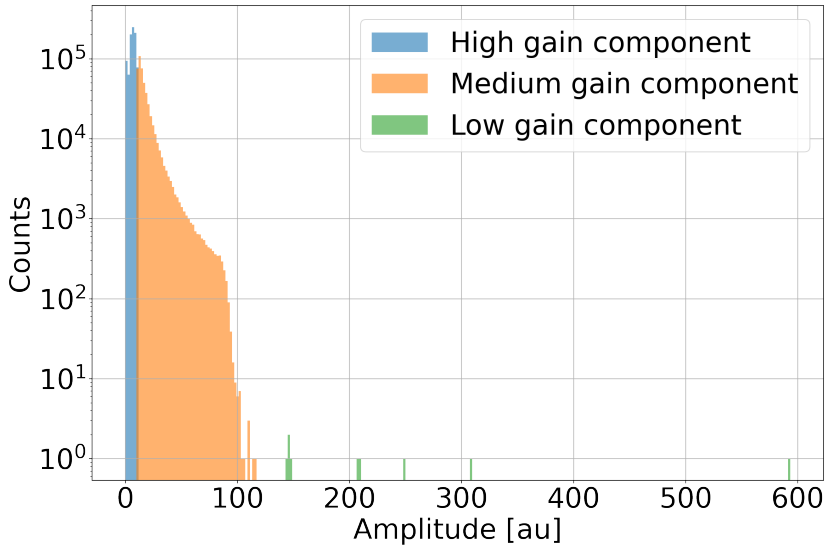


Figure 5.16: Histogram of the intercalibrated and joined amplitudes from different gains.

$\frac{x_{\mathrm{TC}}^{\mathrm{sim}}}{x_{\mathrm{TC}}^{\mathrm{exp}}} = 1.3067\,\mathrm{keV/(umV\,V)}$ is used to compute the line. Ultimately, the impact of calibration on the uncalibrated spectra can be observed, and this effect can be compared to the simulated spectra. After calibration, the experimental spectra are expected to match the simulated spectra. This alignment is evident in Figure 5.18b, where the uncalibrated experimental spectra are colored orange. When calibration was applied, the calibrated spectra (in green) coincide with the simulated spectra (in blue).

Once the amplitudes are calibrated, obtaining the values of y, it is possible to calculate the typical microdosimetric spectra and its derived quantities and distributions described in chapter 1. To verify that the entire procedure is consistent with the previously adopted acquisition chain and data analysis, I compared a microdosimetric spectrum acquired and analyzed from an AbBe radioactive source with both acquisitions. The two $yd(y)$ spectra are shown in Figure 5.20, where the green "Mastro" line refers to the acquisitions previously adopted, while the red "ADC" line was obtained using the methods described in this chapter. The experimental setup is shown in Figure 5.19. The TEPC was located at 20 cm from the AmBe source. Additionally, the TEPC was surrounded by plastic material to henance neutron interactions.

5.4 Discussion and Conclusion

This chapter described the new TEPC acquisition system based on FPGA for HDM. This new acquisition system replaces the MCAs of the original acquisition chain and represents a necessary step towards the integration of the HDM tracking component. In fact, with a dedicated custom acquisition, it is now possible to integrate other detectors. In addition, this new system represents a step forward with respect to the original acquisition system, since, as shown in Figure 5.8b for example, the entire time series containing all the necessary information is captured. On the contrary, with the previous acquisition system, the MCAs directly processed the time series, preventing the user from accessing it. However, by removing the MCAs, the users is faced with the challenge of properly analyzing and extracting the information from the time series to build a microdosimetric spectra. This process required

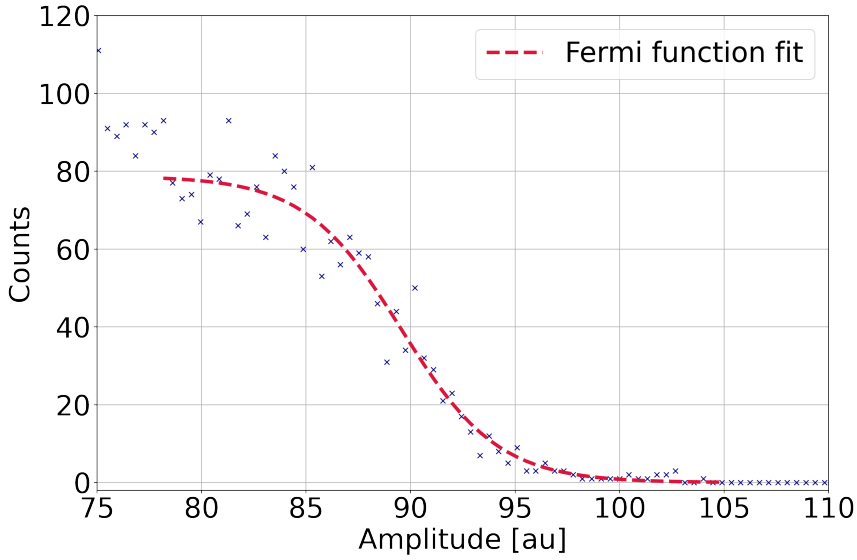
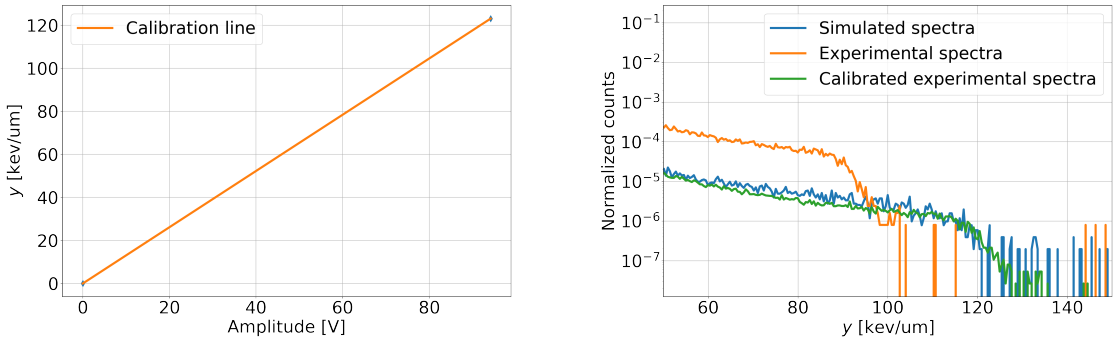


Figure 5.17: Zoomed-in view of the experimental spectra resulting from 70 MeV proton interactions with a 37 mm thick RW3 water equivalent material. The distinctive Fermi-like shape of the proton edge is clearly visible and has been fitted using Equation 5.7, yielding the red line depicted in the figure. The obtained fit coefficients according with Equation 5.7 are: $A = 78 \pm 2$, $B = 0.43 \pm 0.04$ and $C = 89.3 \pm 0.3$

the classification of events that are suitable for analysis and do not contain artifacts such as saturation (Figure 5.13). In addition, calibration and intercalibration are required as a fundamental step in building a microdosimetric spectrum. The ability to derive intercalibration directly from the acquired data is one of the major advantages of this acquisition. On the contrary, the calibration process has to be performed specifically. This resulted in a calibration coefficient $A = 1.3067\,\mathrm{keV/(umV\,V)}$, which corresponds to an ideal high gain vertical resolution of $50/2^{14} \cdot A = 0.004\,\mathrm{keV}\ /\mu\mathrm{m}$. The resolution can be enhanced even further by fitting the peak of the time series, as opposed to simply selecting the time series value corresponding to the peak. By adopting this approach, the chosen point is not limited by the ADC resolution and can take on any continuous value. It is important to note, however, that this method demands additional computational resources, which in turn increases the necessary processing time.

This new acquisition and data analysis has been shown to be capable of subtitling the MCA as shown in Figure 5.20, where the two acquired spectra, one with the old acquisition system labeled “Maesto” in green and the one described in this chapter labeled “ADC” in red, overlap. Consequently, this acquisition method can potentially serve various microdosimeters that utilize the same acquisition chain, thereby enhancing data quality by preserving all relevant information. This is achieved through a customized system that can be further fine-tuned as required.
In fact, some improvements can be further implemented by taking advantage of the high customizability of the system. The first improvement is to send a timestamp along with the ADC data in the DMA. This additional reference improves the estimation of the trigger rate and helps to define the temporal structure of the trigger events. In this way, it is possible to determine when each event was acquired and whether the acquisition suffered from particular events close in time. Another potential improvement involves the software used to acquire



(a) (b)

Figure 5.18: Figure 5.18a shows the calibration relation built from the ratio of the $\frac{x_{\mathrm{TC}}^{\mathrm{sim}}}{x_{\mathrm{TC}}^{\mathrm{exp}}}$ coefficients. Figure 5.18b shows the effect of the calibration in the region of the spectra where the proton edge can be observed. The uncalibrated experimental spectra are displayed in orange. Following the application of calibration, the calibrated spectra (illustrated in green) align remarkably well with the simulated spectra (depicted in blue).

data. Using the available Ethernet port on the Eclypse-Z7 board, it is possible to implement data transfer. This allows data to be stored on a PC away from a potentially hazardous radiation environment, reducing the risk of data corruption due to radiation. When it comes to data analysis, in the case of signal saturation, as in Figure 5.7, by knowing the signal shape, it is possible to use the available non-saturating points to estimate the signal shape of the saturating points. In this way, it is possible to reconstruct the shape of the time series with saturation. This can be used as a double check to confront the signal amplitude at a lower gain level. Additionally, to avoid signals of Figure 5.9 without a good baseline estimation, the acquisition windows can be extended to acquire a few more points. This can already be selected as a parameter in the acquisition program, especially during acquisitions where the ambient noise is significant and could lead to false triggering or difficult baseline estimation. In the case of double hits, another step can be taken in the data analysis. As in the case of Figure 5.11a, if the two signals are still distinguishable, it is possible to fit these signals with a function that is the sum of two signals. In this way, the original amplitudes of the two signals that can be decoupled are estimated.

Finally, when I come to Figure 5.20, it is possible to observe how the new acquisition system can correctly detect lower values of y when compared with the acquisition based on Maestro. Lower y values are associated to small signals and therefore, assuming uniform noise contribution, a small signal to noise ratio. As a consequence, the minimum detectable threshold defines how the acquisition system is subjected to noise.
The noise level of the new acquisitions is essentially determined by the trigger threshold. Once the trigger threshold is set to a certain value, it is expected that signals with an amplitude greater than this value will be acquired. However, this is not always the case, as shown in Figure 5.9a a baseline shift can invalidate this assumption on the trigger. However, the trigger condition can be fully customized in the PL. To make the trigger more complex and selective, it is possible to include multiple points of the time series. Using the High Level Synthesis (HLS) language, it is possible to write a custom trigger condition directly in C++ syntax and implement it as an IP core on the FPGA [Daoud et al., 2019]. However, this must be selected carefully as two aspects have to be considered: 1. The implementation of a strict trigger condition can lead to what is known as event biasing. Event biasing is the

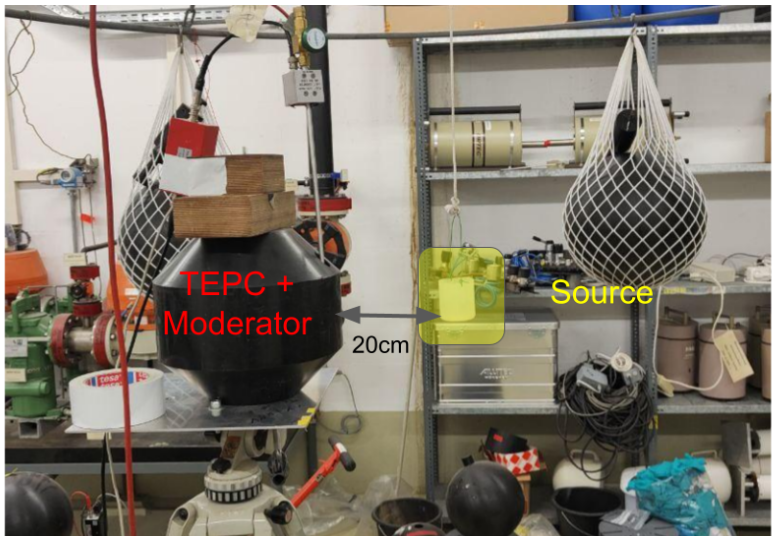


Figure 5.19: Experimental setup for the acquisition with the AmBe source. The TEPC was placed 20 cm away from the radioactive source. Additionally, the TEPC was surrounded by plastic material to enhance neutron interactions.

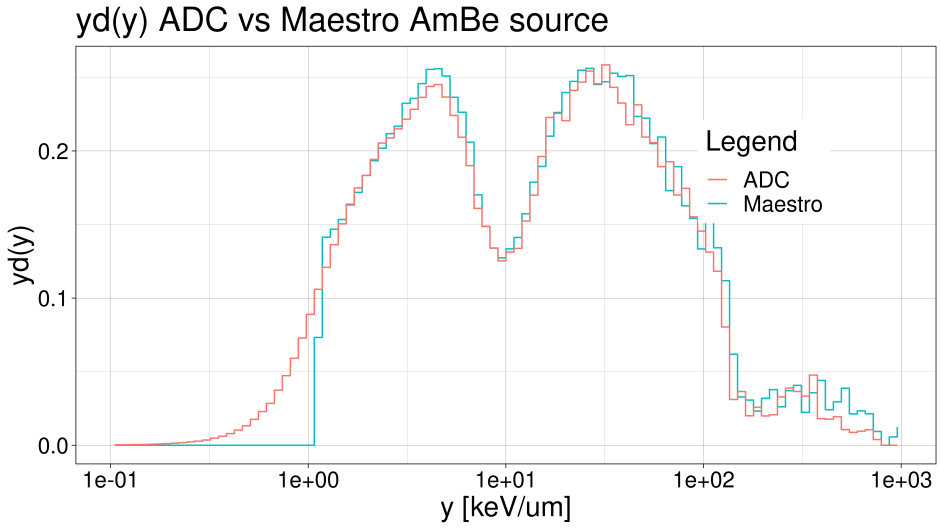


Figure 5.20: $yd(y)$ microdosimetric spectrum acquired from an AmBe source. The two acquisition systems were tested in the same experimental setup. The green line labeled "Maestro" refers to the original acquisition system, while the red line labeled "ADC" refers to the acquisition and data analysis shown in this chapter.

consequence of erroneously sampling only a category of events that is not representative of the entire population. This can result in a biased microdosimetric spectrum. 2. With a custom trigger condition, the level of the trigger amplitude can be lowered, because the trigger should be able to better distinguish between noise and signal. It is important to note that once the time series meets the trigger conditions, it is captured. As a consequence, events lost due to an incorrect trigger condition implementation are lost forever. On the other hand, noise acquired by a trigger level that is too low can always be processed and discarded in the data analysis. However, this approach requires acquisition resources to store time series without relevant information. This delicate balance requires extensive experimentation with a new trigger condition implemented to evaluate its benefits.

Regarding HDM integration, it is important to emphasize that this new implementation allows for complete customization of the acquisition triggering. It can accommodate external sources or transmit signals to synchronize other equipment when the triggering condition is satisfied.

In addition, a comprehensive acquisition latency study must be performed to quantify the exact time it takes for a sample from the ADC to become available on the PL. Latency information is critical, especially when coordinating multiple detectors, where knowing the event processing time is essential for synchronization. This is a fundamental requirement for associating the recorded energy position with the other HDM stage responsible for particle tracking.

Conclusions and future perspectives

This work has focused on experimental microdosimetry, with the goal of improving the available tools of microdosimetry. As more complete and comprehensive tools become available, I hope that microdosimetry will gain both popularity and users in the scientific community. In fact, microdosimetry, over the past several years, has already gained popularity as a tool to measure radiation quality. When compared to the LET, microdosimetry offers a more accurate description of radiation quality, accounting for the stochastic nature of energy deposition.
The use of microdosimetry in ion therapy is becoming increasingly important. By measuring the lineal energy distribution, it provides a critical resource for quality assurance in treatment planning. In addition, it provides a strong link between physical characterization and biological effect. Ultimately, achieving the most accurate representation of radiation quality is the basis for predicting any biological outcome.

To improve the description of the radiation quality provided by microdosimetry, a new hybrid detector is proposed. This detector, named Hybrid Detector for Microdosimetry is composed of a commercial TEPC microdosimeter with the task of recording the energy deposition in tissue equivalent material, followed by 4 layers of LGADs strips that serve as a tracker. In this way, it is possible to define and measure a new quantity, y_T which replaces the mean chord length $\bar{l}$ with the estimated track length inside the microdosimeter l. The concept of this new detector, the Hybrid Detector for Microdosimetry (HDM), has been validated using Monte Carlo simulations in chapter 3. Discrepancies have been observed in the $yd(y)$ and the $y_T d(y_T)$ microdosimetric spectra, especially in the high y_T regions, demonstrating the limitations of the mean chord length approximation, particularly for radiobiologically relevant events at high y. In addition, a more complete description of the radiation field quality can be achieved by incorporating the real track length. This new approach eliminates the dependence on detector geometry and assumptions about the radiation field that may not always hold. Despite all these advantages, the main weakness of HDM is its tracking efficiency. To mitigate this problem and improve the tracking accuracy, a dedicated study has been conducted using machine learning methods [Missiaggia et al., 2022].

The development of HDM necessitates the implementation of a dedicated readout system chapter 4 is dedicated to a readout system for the LGADs based on FPGA. A preliminary version of the readout capable of handling a single layer with 24 channels has been developed and tested with an experimental campaign at the Protonterapy Center in Trento. The results indicated the ability of the readout system to effectively detect protons within the clinical energy range.
To handle 71 channels, as required by HDM, the readout can be extended without much effort, since the core part of the code has already been implemented and validated. The biggest change is expected in the threshold control handling, since a single FPGA model ZC702 can handle up to 72 channels, making it impossible to handle the 71 LGADs strips and the 15 threshold control lines at the same time. To solve this problem, it is necessary to integrate a fifth FPGA dedicated to threshold control.
One of the biggest challenges of this implementation was the transfer of data from the PL to the end user. This problem was solved by implementing DMA transactions that allow for fast and efficient data transfer from the FPGA's internal logic to the end user.
In general, the core of such an implementation can be reused for other readout implementations that require consistent and fast data flows from the PL to the end user.

A second chapter chapter 5 is dedicated to make the existing TEPC acquisition compatible with HDM. The new acquisition digitizes and stores the entire time series of the signal in memory, so all the data and associated information is always available. Although this approach requires the implementation of a dedicated data analysis process, it gives users

the freedom to act directly on the signal time series and perform sophisticated data analysis operations. One of the major benefits of this new acquisition is that the intercalibration is automatically obtained from the data analysis and can be used as a consistency check of the acquisition.
When it comes to performance evaluations, the new readouts have proven to withstand detection rates in the detector's operating range and even at higher particle rates where pileup can occur. In general, this acquisition can be used for different detectors that share a similar acquisition chain, and if specific requirements arise, it can be adapted within the limits of the hardware capabilities. One of the major limitations of data analysis is the peak detection algorithm, which has been proven as a critical point in data analysis. New algorithms based on machine learning techniques for peak identification are under development. Another related improvement comes from the trigger condition available in the FPGA, as a simple threshold is not the best option. Today, with the help of the High-Level Synthesis (HLS) tools, it is possible to create a complex IP core with advanced and custom trigger functionality simply by writing C++ code. The development of a dedicated IP core in charge of the triggering condition will consequentially improve the quality of the acquired data.

Finally, in order to synchronize the two stages of the HDM, the latency of the two systems must be carefully characterized. In particular, in the case of the TEPC, the time between the transmission of a pulse and the actual digitization of this pulse must be measured accurately. This latency must be taken into account when implementing a common trigger condition. Regarding the LGAD readout, the implemented design is expected to have negligible latency (few nanoseconds).
The mean chord length approximation will be used for particles that cannot be fully tracked by the LGAD layers. It is therefore fundamental that the new acquisition of TEPC triggers the whole tracking system, as most of the particles will traverse the TEPC before the LGADs layers. Upon energy deposition detection in the TEPC, the acquisition must send a triggering signal to the LGADs tracking FPGA. To account for the estimated latency, the LGADs acquisition data is stored in a circular buffer with the appropriate size and delay. When triggered by the TEPC, the buffer contents are subsequently transferred to the PS memory. In this way, it is possible to acquire energy deposition in the TEPC and correlated the activation of LGADs strips.

In parallel to the implementation of the readout system for HDM, I have developed a dedicated software called Mandarina to perform the analysis of microdosimetric spectra. This software, described in section 2.3, has a graphical user interface that allows the user to monitor in real time how the parameters required by the data analysis modify the final spectrum. In this way, the user can adjust these parameters in an efficient and easy way with direct feedback from the program.
Mandarina uses as input data spectra acquired with a TEPC microdosimeter connected to the acquisition chain described in section 2.2. However, since the acquisition chain itself can be used on different microdosimeters, as long as the last stage of the acquisition chain consisting of multichannel analyzers are used, Mandarina can also be used to analyze the data. Furthermore, if it is necessary to modify the current data analysis or to add a new feature, the simplicity of C# allows it to be integrated without any particular effort.

Finally, section 2.4 is dedicated to the long-standing pileup problem affecting microdosimetric spectra. With the increasing use of microdosimetry in clinical applications, particularly for quality assurance procedures, the challenges posed by high particle rates can limit the application of microdosimetric measurements. To address this issue, an algorithm has been developed to estimate the probability of pileup that affects a microdosimetric spectrum. Through the development of a new methodology, we have successfully estimated the pileup contribution in microdosimetric spectra. Although this methodology has been applied for

proton energies of 70 MeV and 11 MeV, the pileup-rate response follows the same trend, independently of the energy. Consequently, for proton energies within this interval and for TEPC LET-1/2 detectors, the estimated pileup rate curves can be used as reference values. Nevertheless, the presented methodology can be applied in the case of different detectors and different energies.

In conclusion, it is my aspiration that my thesis and the research I have conducted will facilitate the broader adoption of microdosimetry among the scientific community. Through the development of new tools and enhancements to existing ones, I anticipate that microdosimetry will achieve increased accessibility and effectiveness, fully unlocking its potential in both clinical and non-clinical applications.

Bibliography

[AD9, a] AD9648 datasheet. https://www.analog.com/media/en/technical-documentation/data-sheets/AD9648.pdf. Accessed: 2023-08-27.

[AD9, b] AD9648 registers configuration. https://www.analog.com/media/en/technical-documentation/application-notes/AN-877.pdf?doc=AD9648.pdf. Accessed: 2023-08-27.

[DMA,] AXI DMA user space driver. https://github.com/bperez77/xilinx_axidma/tree/master. Accessed: 2023-08-27.

[xil, a] Block RAM ip core user guide. https://docs.xilinx.com/v/u/8.3-English/pg058-blk-mem-gen. Accessed: 2023-08-27.

[CAE, a] Caen A422A charge sensitive preamplifier. https://www.caen.it/products/a422a/. Accessed: 2023-08-27.

[CAE, b] Caen N968 shaping amplifier. https://www.caen.it/products/n968/. Accessed: 2023-08-27.

[Dig,] Digitilent Eclypse Z 7 developement board. https://digilent.com/reference/programmable-logic/eclypse-z7/reference-manual. Accessed: 2023-08-27.

[IMB,] IBM-CNM centro nacional de microelectronica. https://github.com/bperez77/xilinx_axidma/tree/master. https://www.imb-cnm.csic.es/en.

[xil, b] IDDR xilinx primitive. https://docs.xilinx.com/r/en-US/ug953-vivado-7series-libraries/IDDR. Accessed: 2023-08-27.

[ORT, a] Maestro software ortec. https://www.ortec-online.com/products/application-software/maestro-mca. Accessed: 2023-09-30.

[ORT, b] Multichannel Analyzer 927 ortec. https://www.ortec-online.com/products/electronics/multichannel-analyzers-mca/basic-analog/aspec-927. Accessed: 2023-09-30.

[PMO,] PMOD Port digilent standard. https://digilent.com/reference/_media/reference/pmod/pmod-interface-specification-1_3_1.pdf. Accessed: 2023-08-27.

[Sci,] Scipy peak find function. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html. Accessed: 2023-09-12.

[STD,] Standard Library chrono. https://en.cppreference.com/w/cpp/header/chrono. Accessed: 2023-08-16.

[SYZ,] SYZYGY port standard specifications. https://syzygyfpga.io/specification/. Accessed: 2023-08-27.

[VIT,] VITA fmc specification. https://www.samtec.com/standards/vita/fmc/. Accessed: 2023-08-16.

[Xil, a] XILINX DMA ip core guide. https://docs.xilinx.com/r/en-US/pg021_axi_dma. Accessed: 2023-08-16.

[Xil, b] Xilinx ZC702 developement board. https://www.xilinx.com/products/boards-and-kits/ek-z7-zc702-g.html. Accessed: 2023-08-16.

[ZMO, a] ZMOD 1410 Library user guide - digilent. https://digilent.com/reference/zmod/scope/zmodadc1410libraryuserguide. Accessed: 2023-08-27.

[ZMO, b] Zmod ADC 1410 AXI Adapter user guide. https://github.com/Digilent/vivado-library/blob/zmod/v1/2019.1-2/ip/Zmods/AXI_Zmod_ADC1410/docs/ZmodADC1410AxiAdapter.pdf. Accessed: 2023-08-27.

[ZMO, c] Zmod ADC 1410 Low Level Controller ip core user guide. https://github.com/Digilent/vivado-library/blob/zmod/v1/2019.1-2/ip/Zmods/ZmodADC1410Controller/docs/ZmodADC1410LLC.pdf. Accessed: 2023-08-27.

[ZMO, d] ZMOD ADC digilent. https://digilent.com/reference/zmod/scope/reference-manual. Accessed: 2023-08-27.

[ICR, 1983] (1983). Microdosimetry. icru report 36.

[doi, 2007] (2007). Prescribing, recording, and reporting proton-beam therapy: Contents. *Journal of the ICRU*, 7(2). PMID: 24171083.

[Abolfath et al., 2020] Abolfath, R., Helo, Y., Carlson, D. J., Stewart, R., Grosshans, D., and Mohan, R. (2020). A new approach to modeling the microdosimetry of proton therapy beams. *Medical physics*, 47(7):3184–3190.

[Agosteo et al., 2006] Agosteo, S., Colautti, P., Fazzi, A., Moro, D., and Pola, A. (2006). A solid state microdosimeter based on a monolithic silicon telescope. *Radiation protection dosimetry*, 122(1-4):382–386.

[Agosteo and Pola, 2011] Agosteo, S. and Pola, A. (2011). Silicon microdosimetry. *Radiation protection dosimetry*, 143(2-4):409–415.

[Agostinelli et al., 2003] Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Axen, D., Banerjee, S., Barrand, G., Behner, F., Bellagamba, L., Boudreau, J., Broglia, L., Brunengo, A., Burkhardt, H., Chauvie, S., Chuma, J., Chytracek, R., Cooperman, G., Cosmo, G., Degtyarenko, P., Dell'Acqua, A., Depaola, G., Dietrich, D., Enami, R., Feliciello, A., Ferguson, C., Fesefeldt, H., Folger, G., Foppiano, F., Forti, A., Garelli, S., Giani, S., Giannitrapani, R., Gibin, D., Gómez Cadenas, J., González, I., Gracia Abril, G., Greeniaus, G., Greiner, W., Grichine, V., Grossheim, A., Guatelli, S., Gumplinger, P., Hamatsu, R., Hashimoto, K., Hasui, H., Heikkinen, A., Howard, A., Ivanchenko, V., Johnson, A., Jones, F., Kallenbach, J., Kanaya, N., Kawabata, M., Kawabata, Y., Kawaguti, M., Kelner, S., Kent, P., Kimura, A., Kodama, T., Kokoulin, R., Kossov, M., Kurashige, H., Lamanna, E., Lampén, T., Lara, V., Lefebure, V., Lei, F., Liendl, M., Lockman, W., Longo, F., Magni, S., Maire, M., Medernach, E., Minamimoto, K., Mora de Freitas, P., Morita, Y., Murakami, K., Nagamatu, M., Nartallo, R., Nieminen, P., Nishimura, T., Ohtsubo, K., Okamura, M., O'Neale, S., Oohata, Y., Paech, K., Perl, J., Pfeiffer, A., Pia, M., Ranjard, F., Rybin, A., Sadilov, S., Di Salvo, E., Santin, G.,

Sasaki, T., Savvas, N., Sawada, Y., Scherer, S., Sei, S., Sirotenko, V., Smith, D., Starkov, N., Stoecker, H., Sulkimo, J., Takahata, M., Tanaka, S., Tcherniaev, E., Safai Tehrani, E., Tropeano, M., Truscott, P., Uno, H., Urban, L., Urban, P., Verderi, M., Walkden, A., Wander, W., Weber, H., Wellisch, J., Wenaus, T., Williams, D., Wright, D., Yamada, T., Yoshida, H., and Zschiesche, D. (2003). Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303.

[Ahlen, 1980] Ahlen, S. P. (1980). Theoretical and experimental aspects of the energy loss of relativistic heavily ionizing particles. *Reviews of Modern Physics*, 52(1):121.

[Allison et al., 2006] Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce Dubois, P., Asai, M., Barrand, G., Capra, R., Chauvie, S., Chytracek, R., Cirrone, G., Cooperman, G., Cosmo, G., Cuttone, G., Daquino, G., Donszelmann, M., Dressel, M., Folger, G., Foppiano, F., Generowicz, J., Grichine, V., Guatelli, S., Gumplinger, P., Heikkinen, A., Hrivnacova, I., Howard, A., Incerti, S., Ivanchenko, V., Johnson, T., Jones, F., Koi, T., Kokoulin, R., Kossov, M., Kurashige, H., Lara, V., Larsson, S., Lei, F., Link, O., Longo, F., Maire, M., Mantero, A., Mascialino, B., McLaren, I., Mendez Lorenzo, P., Minamimoto, K., Murakami, K., Nieminen, P., Pandola, L., Parlati, S., Peralta, L., Perl, J., Pfeiffer, A., Pia, M., Ribon, A., Rodrigues, P., Russo, G., Sadilov, S., Santin, G., Sasaki, T., Smith, D., Starkov, N., Tanaka, S., Tcherniaev, E., Tome, B., Trindade, A., Truscott, P., Urban, L., Verderi, M., Walkden, A., Wellisch, J., Williams, D., Wright, D., and Yoshida, H. (2006). Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1):270–278.

[Allison et al., 2016] Allison, J., Amako, K., Apostolakis, J., Arce, P., Asai, M., Aso, T., Bagli, E., Bagulya, A., Banerjee, S., Barrand, G., Beck, B., Bogdanov, A., Brandt, D., Brown, J., Burkhardt, H., Canal, P., Cano-Ott, D., Chauvie, S., Cho, K., Cirrone, G., Cooperman, G., Cortés-Giraldo, M., Cosmo, G., Cuttone, G., Depaola, G., Desorgher, L., Dong, X., Dotti, A., Elvira, V., Folger, G., Francis, Z., Galoyan, A., Garnier, L., Gayer, M., Genser, K., Grichine, V., Guatelli, S., Guèye, P., Gumplinger, P., Howard, A., Hřivnáčová, I., Hwang, S., Incerti, S., Ivanchenko, A., Ivanchenko, V., Jones, F., Jun, S., Kaitaniemi, P., Karakatsanis, N., Karamitros, M., Kelsey, M., Kimura, A., Koi, T., Kurashige, H., Lechner, A., Lee, S., Longo, F., Maire, M., Mancusi, D., Mantero, A., Mendoza, E., Morgan, B., Murakami, K., Nikitina, T., Pandola, L., Paprocki, P., Perl, J., Petrović, I., Pia, M., Pokorski, W., Quesada, J., Raine, M., Reis, M., Ribon, A., Ristić Fira, A., Romano, F., Russo, G., Santin, G., Sasaki, T., Sawkey, D., Shin, J., Strakovsky, I., Taborda, A., Tanaka, S., Tomé, B., Toshito, T., Tran, H., Truscott, P., Urban, L., Uzhinsky, V., Verbeke, J., Verderi, M., Wendt, B., Wenzel, H., Wright, D., Wright, D., Yamashita, T., Yarba, J., and Yoshida, H. (2016). Recent developments in geant4. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835:186–225.

[Atkinson, 1991] Atkinson, K. (1991). *An introduction to numerical analysis*. John wiley & sons.

[Barkas and Evans, 1963] Barkas, W. and Evans, D. (1963). *Nuclear Research Emulsions: Techniques and theory*. Nuclear Research Emulsions. Academic Press.

[Behrens and Melissinos, 1981] Behrens, S. and Melissinos, A. (1981). Univ. of rochester preprint.

[Bellinzona et al., 2021] Bellinzona, V., Cordoni, F., Missiaggia, M., Tommasino, F., Scifoni, E., La Tessa, C., and Attili, A. (2021). Linking microdosimetric measurements to biological effectiveness in ion beam therapy: a review of theoretical aspects of mkm and other models. *Frontiers in Physics*, 8:578492.

[Bernstein, 2021] Bernstein, H. (2021). *Measuring electronics and sensors: basics of measurement technology, sensors, analog and digital signal processing*. Springer Nature.

[Bethe, 1953] Bethe, H. A. (1953). Molière's Theory of Multiple Scattering. *Phys. Rev.*, 89.

[Bianchi et al., 2021] Bianchi, A., Selva, A., Colautti, P., Petringa, G., Cirrone, P., Reniers, B., Parisi, A., Vanhavere, F., and Conte, V. (2021). Repeatability and reproducibility of microdosimetry with a mini-tepc. *Frontiers in Physics*, 9:727816.

[Bichsel, 1968] Bichsel, H. (1968). Charged-particle interactions. *Radiation dosimetry*, 1:157–224.

[Bisht et al., 2022] Bisht, A., Borghi, G., Boscardin, M., Centis Vignali, M., Ficorella, F., Hammad Ali, O., and Paternoster, G. (2022). Development of lgad sensors at fbk. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1039:167018.

[Bolst et al., 2017] Bolst, D., Guatelli, S., Tran, L. T., Chartier, L., Lerch, M. L., Matsufuji, N., and Rosenfeld, A. B. (2017). Correction factors to convert microdosimetry measurements in silicon to tissue in 12c ion therapy. *Physics in Medicine & Biology*, 62(6):2055.

[Booz et al., 1983] Booz, J., Braby, L., Coyne, J., Kliauga, P., Lindborg, L., Menzel, H., and Parmentier, N. (1983). Report 36. *Journal of the International Commission on Radiation Units and Measurements*, (1):NP–NP.

[Braby, 2015] Braby, L. (2015). Experimental microdosimetry: history, applications and recent technical advances. *Radiation protection dosimetry*, 166(1-4):3–9.

[Casado et al., 2022] Casado, M., Group, A. H., et al. (2022). A high-granularity timing detector for the atlas phase-ii upgrade. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1032:166628.

[Chiriotti et al., 2015] Chiriotti, S., Moro, D., Colautti, P., Conte, V., and Grosswendt, B. (2015). Equivalence of pure propane and propane TE gases for microdosimetric measurements. *Radiation protection dosimetry*, 166(1-4):242–246.

[Cms, 2019] Cms, C. (2019). A mip timing detector for the cms phase-2 upgrade. Technical report.

[Colautti et al., 2020] Colautti, P., Magrin, G., Palmans, H., Cortés-Giraldo, M. A., and Conte, V. (2020). Characterizing radiation effectiveness in ion-beam therapy part ii: Microdosimetric detectors. *Frontiers in Physics*, 8:451.

[Comfort et al., 1966] Comfort, J. R., Decker, J. F., Lynk, E. T., Scully, M. O., and Quinton, A. R. (1966). Energy Loss and Straggling of Alpha Particles in Metal Foils. *Phys. Rev.*, 150.

[Conte et al., 2020] Conte, V., Agosteo, S., Bianchi, A., Bolst, D., Bortot, D., Catalano, R., Cirrone, G., Colautti, P., Cuttone, G., Guatelli, S., et al. (2020). Microdosimetry of a therapeutic proton beam with a mini-tepc and a microplus-bridge detector for rbe assessment. *Physics in Medicine & Biology*, 65(24):245018.

[Conte et al., 2013] Conte, V., Moro, D., Grosswendt, B., and Colautti, P. (2013). Lineal energy calibration of mini tissue-equivalent gas-proportional counters (TEPC). *AIP Conference Proceedings*, 1530(1):171–178.

[Cruz et al., 2001] Cruz, G. S., Palmer, M., Matatagui, E., and Zamenhof, R. (2001). A theoretical model for event statistics in microdosimetry. i: Uniform distribution of heavy ion tracks. *Medical physics*, 28(6):988–996.

[Daoud et al., 2019] Daoud, L., Hussein, F., and Rafla, N. (2019). Optimization of advanced encryption standard (aes) using vivado high level synthesis (hls).

[De Nardo et al., 2004] De Nardo, L., Cesari, V., Donà, G., Magrin, G., Colautti, P., Conte, V., and Tornielli, G. (2004). Mini-tepcs for radiation therapy. *Radiation protection dosimetry*, 108(4):345–352.

[Delaney et al., 2005] Delaney, G., Jacob, S., Featherstone, C., and Barton, M. (2005). The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 104(6):1129–1137.

[Dicello et al., 1980] Dicello, J., Amols, H., Zaider, M., and Tripard, G. (1980). A comparison of microdosimetric measurements with spherical proportional counters and solid-state detectors. *Radiation Research*, 82(3):441–453.

[Dieter Schardt, 2010] Dieter Schardt, D. S.-E. (2010). In *Heavy-ion tumor therapy: physical and radiobiological benefits*. Rev.Mod.Physics 82.

[Eadie et al., 1971] Eadie, W. T., Drijard, D., and James, F. E. (1971). Statistical methods in experimental physics. *Amsterdam: North-Holland*.

[Farah et al., 2015] Farah, J., Mares, V., Romero-Expósito, M., Trinkl, S., Domingo, C., Dufek, V., Klodowska, M., Kubancak, J., Knežević, Z., Liszka, M., Majer, M., Miljanić, S., Ploc, O., Schinner, K., Stolarczyk, L., Trompier, F., Wielunski, M., Olko, P., and Harrison, R. M. (2015). Measurement of stray radiation within a scanning proton therapy facility: Eurados wg9 intercomparison exercise of active dosimetry systems. *Medical Physics*, 42(5):2572–2584.

[Farahmand et al., 2004] Farahmand, M., Bos, A., De Nardo, L., and Van Eijk, C. (2004). First microdosimetric measurements with a tepc based on a gem. *Radiation protection dosimetry*, 110(1-4):839–843.

[Fausti et al., 2018] Fausti, F., Mazza, G., Cartiglia, N., Ferrero, M., Giordanengo, S., Ali, O., Mandurrino, M., Monaco, V., Olave, J., Sacchi, R., et al. (2018). Characterization of a silicon detector and front-end electronics prototype for single ion discrimination in hadrontherapy. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, pages 1–3. IEEE.

[Fausti et al., 2021] Fausti, F., Olave, J., Giordanengo, S., Ali, O. H., Mazza, G., Rotondo, F., Wheadon, R., Vignati, A., Cirio, R., Monaco, V., and Sacchi, R. (2021). A single ion discriminator asic prototype for particle therapy applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 985:164666.

[Ferrero et al., 2017] Ferrero, M., Ali, O. H., Arcidiacono, R., Boscardin, M., Cartiglia, N., Cenna, F., Cirio, R., Costa, M., Dalla Betta, G. F., Ficorella, F., Giordanengo, S., Mandurrino, M., Monaco, V., Obertino, M. M., Pancheri, L., Paternoster, G., Sacchi, R., Siviero, F., Sola, V., Staian, A., and Vignati, A. (2017). Developments in the FBK production of ultra-fast silicon detectors. In *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–5.

[Ferrero et al., 2019] Ferrero, M., Arcidiacono, R., Barozzi, M., Boscardin, M., Cartiglia, N., Betta, G. D., Galloway, Z., Mandurrino, M., Mazza, S., Paternoster, G., and et al. (2019). Radiation resistant LGAD design. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 919:16–26.

[Fiorino et al., 2020] Fiorino, C., Guckenberger, M., Schwarz, M., van der Heide, U. A., and Heijmen, B. (2020). Technology-driven research for radiotherapy innovation. *Molecular Oncology*, 14(7):1500–1513.

[Forouzan, 2002] Forouzan, B. A. (2002). *TCP/IP protocol suite*. McGraw-Hill Higher Education.

[Fowler, 1989] Fowler, J. F. (1989). The linear-quadratic formula and progress in fractionated radiotherapy. *The British journal of radiology*, 62(740):679–694.

[Giacomini, 2023] Giacomini, G. (2023). Lgad-based silicon sensors for 4d detectors. *Sensors*, 23(4):2132.

[Goldhaber and Heckman, 1978] Goldhaber, A. S. and Heckman, H. H. (1978). High energy interactions of nuclei. *Annual Review of Nuclear and Particle Science*, 28(1):161–205.

[Goodhead, 1988] Goodhead, D. (1988). Spatial and temporal distribution of energy. *Health Physics*, 55(2):231–240.

[Goodhead, 2013] Goodhead, D. (2013). https://melodi-online.eu/wp-content/uploads/2021/04/02-What-is-radiation-quality-Goodhead.pdf.

[Gunzert-Marx et al., 2008] Gunzert-Marx, K., Iwase, H., Schardt, D., and Simon, R. (2008). Secondary beam fragments produced by 200 mev u- 1 12c ions in water and their dose contributions in carbon ion radiotherapy. *New journal of physics*, 10(7):075003.

[Hawkins, 1994] Hawkins, R. B. (1994). A statistical theory of cell killing by radiation of varying linear energy transfer. *Radiation research*, 140(3):366–374.

[Hazdra et al., 2018] Hazdra, P., Popelka, S., and Schöner, A. (2018). Optimization of sic power p-i-n diode parameters by proton irradiation. *IEEE Transactions on Electron Devices*, 65(10):4483–4489.

[Hejlsberg et al., 2003] Hejlsberg, A., Wiltamuth, S., and Golde, P. (2003). *C# language specification*. Addison-Wesley Longman Publishing Co., Inc.

[Horst, 2020] Horst, F. E. (2020). Measurement of nuclear reaction cross sections for applications in radiotherapy with protons, helium and carbon ions.

[Hüfner et al., 1975] Hüfner, J., Schäfer, K., and Schürmann, B. (1975). Abrasion-ablation in reactions between relativistic heavy ions. *Phys. Rev. C*, 12:1888.

[ICRU, 1964] ICRU (1964). Physical Aspects of Irradiation Report 10b. *International Commission on Radiological Units and Measurements*.

[Kellerer et al., 1985] Kellerer, A. M. et al. (1985). Fundamentals of microdosimetry. *The dosimetry of ionizing radiation*, 1:77–162.

[Knoll, 2010] Knoll, G. F. (2010). *Radiation detection and measurement*. John Wiley & Sons.

[Kramberger et al., 2015] Kramberger, G., Baselga, M., Cindro, V., Fernández-Martínez, P., Flores, D., Galloway, Z., Gorisek, A., Greco, V., Hidalgo, S., Fadeyev, V., Mandić, I., Mikuz, M., Quirion, D., Pellegrini, G., Sadrozinski, H.-W., Studen, A., and Zavrtanik, M. (2015). Radiation effects in low gain avalanche detectors after hadron irradiations. *Journal of Instrumentation*, 10:P07006–P07006.

[Langen et al., 2002] Langen, K., Binns, P., Lennox, A., Kroc, T., and DeLuca Jr, P. (2002). Pileup correction of microdosimetric spectra. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 484(1-3):595–612.

[Letzter and Webster, 1970] Letzter, S. and Webster, N. (1970). Noise in amplifiers. *IEEE spectrum*, 7(8):67–75.

[Lim et al., 2009] Lim, W. H., Ziebell, A. L., Cornelius, I., Reinhard, M. I., Prokopovich, D. A., Dzurak, A. S., and Rosenfeld, A. B. (2009). Cylindrical silicon-on-insulator microdosimeter: Design, fabrication and tcad modeling. *IEEE Transactions on Nuclear Science*, 56(2):424–428.

[Luoni et al., 2021] Luoni, F., Horst, F., Reidel, C., Quarz, A., Bagnale, L., Sihver, L., Weber, U., Norman, R., De Wet, W., Giraudo, M., et al. (2021). Total nuclear reaction cross-section database for radiation protection in space and heavy-ion therapy applications. *New Journal of Physics*, 23(10):101201.

[Magrin et al., 2023] Magrin, G., Palmans, H., Stock, M., and Georg, D. (2023). State-of-the-art and potential of experimental microdosimetry in ion-beam therapy. *Radiotherapy and Oncology*, page 109586.

[Marti Villarreal et al., 2023] Marti Villarreal, O., Vignati, A., Giordanengo, S., Abujami, M., Borghi, G., Centis Vignali, M., Data, E., Ferrero, M., Ficorella, F., Galeone, C., Hammad Ali, O., Mas Milian, F., Medina, E., Menzio, L., Montalvan Olivares, D., Peroglio Carus, G., Cirio, R., Monaco, V., Sacchi, R., Donetti, M., and Pullia, M. (2023). Characterization of thin lgad sensors designed for beam monitoring in proton therapy. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1046:167622.

[Mayles et al., 2007] Mayles, P., Nahum, A., and Rosenwald, J.-C. (2007). *Handbook of radiotherapy physics: theory and practice*. CRC Press.

[Mazzucconi et al., 2019] Mazzucconi, D., Bortot, D., Agosteo, S., Pola, A., Pasquato, S., Fazzi, A., Colautti, P., Conte, V., Petringa, G., Amico, A., et al. (2019). Microdosimetry at nanometric scale with an avalanche-confinement tepc: response against a helium ion beam. *Radiation protection dosimetry*, 183(1-2):177–181.

[McMahon, 2018] McMahon, S. J. (2018). The linear quadratic model: usage, interpretation and challenges. *Physics in Medicine & Biology*, 64(1):01TR01.

[Missiaggia et al., 2020] Missiaggia, M., Cartechini, G., Scifoni, E., Rovituso, M., Tommasino, F., Verroi, E., Durante, M., and La Tessa, C. (2020). Microdosimetric measurements as

a tool to assess potential in-and out-of-field toxicity regions in proton therapy. *Physics in Medicine & Biology*.

[Missiaggia et al., 2023] Missiaggia, M., Cartechini, G., Tommasino, F., Scifoni, E., and La Tessa, C. (2023). Investigation of in-field and out-of-field radiation quality with microdosimetry and its impact on relative biological effectiveness in proton therapy. *International Journal of Radiation Oncology* Biology* Physics*, 115(5):1269–1282.

[Missiaggia et al., 2021] Missiaggia, M., Pierobon, E., Castelluzzo, M., Perinelli, A., Cordoni, F., Centis Vignali, M., Borghi, G., Bellinzona, E., Scifoni, E., Tommasino, F., et al. (2021). A novel hybrid microdosimeter for radiation field characterization based on the tissue equivalent proportional counter detector and low gain avalanche detectors tracker: a feasibility study. *Frontiers in Physics*, 8:578444.

[Missiaggia et al., 2022] Missiaggia, M., Pierobon, E., La Tessa, C., and Cordoni, F. G. (2022). An exploratory study of machine learning techniques applied to therapeutic energies particle tracking in microdosimetry using the novel hybrid detector for microdosimetry (hdm). *Physics in Medicine & Biology*, 67(18):185002.

[Mohamad et al., 2018] Mohamad, O., Yamada, S., and Durante, M. (2018). Clinical indications for carbon ion radiotherapy. *Clinical Oncology*, 30(5):317–329.

[Moliere, 1948] Moliere, G. (1948). Theory of scattering of fast charged particles ii multiple and multiple scattering. *Journal for Natural Research A*, 3(2):78–97.

[Moro et al., 2015] Moro, D., Chiriotti, S., Conte, V., Colautti, P., and Grosswendt, B. (2015). Lineal energy calibration of a spherical TEPC. *Radiation Protection Dosimetry*, 166(1-4):233–237.

[Norbury et al., 2012] Norbury, J. W., Miller, J., Adamczyk, A. M., Heilbronn, L. H., Townsend, L. W., Blattnig, S. R., Norman, R. B., Guetersloh, S. B., and Zeitlin, C. J. (2012). Nuclear data for space radiation. *Radiation measurements*, 47(5):315–363.

[Otten, 1989] Otten, E. (1989). Treatise on heavy-ion science, vol. 8. *DA Bromley Plenum, New York*.

[Paganetti, 2018] Paganetti, H. (2018). *Proton therapy physics*. CRC press.

[Pellegrini et al., 2014] Pellegrini, G., Fernández-Martínez, P., Baselga, M., Fleta, C., Flores, D., Greco, V., Hidalgo, S., Mandić, I., Kramberger, G., Quirion, D., et al. (2014). Technology developments and first measurements of low gain avalanche detectors (lgad) for high energy physics applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 765:12–16.

[Powers et al., 1968] Powers, E. L., Lyman, J. T., and Tobias, C. A. (1968). Some effects of accelerated charged particles on bacterial spores. *International Journal of Radiation Biology*, 14(4):313–330.

[PTCOG, 2020] PTCOG (2020). Ptcog website. https://www.ptcog.ch/index.php/ptcog-patient-statistics.

[Rosenfeld, 2016] Rosenfeld, A. B. (2016). Novel detectors for silicon based microdosimetry, their concepts and applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 809:156–170. Advances in detectors and applications for medicine.

[Rosenfeld et al., 2000] Rosenfeld, A. B., Bradley, P., Cornelius, I., Kaplan, G., Allen, B., Flanz, J., Goitein, M., Van Meerbeeck, A., Schubert, J., Bailey, J., et al. (2000). A new silicon detector for microdosimetry applications in proton therapy. *IEEE Transactions on nuclear science*, 47(4):1386–1394.

[Rossi et al., 1996] Rossi, H., Zaider, M., and Zaider, M. (1996). *Microdosimetry and Its Applications*. John Libbey.

[Rossi, 1959] Rossi, H. H. (1959). Specification of radiation quality. *Radiation Research*, 10(5):522–531.

[Rossi and Rosenzweig, 1955] Rossi, H. H. and Rosenzweig, W. (1955). Measurements of neutron dose as a function of linear energy transfer. *Radiation research*, 2(5):417–425.

[Rovituso and La Tessa, 2017] Rovituso, M. and La Tessa, C. (2017). Nuclear interactions of new ions in cancer therapy: impact on dosimetry. *Transl Cancer Res*, 6:S914–33.

[Sacchi et al., 2020] Sacchi, R., Ganjeh, Z. A., Arcidiacono, R., Attili, A., Cartiglia, N., Donetti, M., Fausti, F., Ferrero, M., Giordanengo, S., Ali, O. H., et al. (2020). Test of innovative silicon detectors for the monitoring of a therapeutic proton beam. In *Journal of Physics: Conference Series*, volume 1662, page 012002. IOP Publishing.

[Sacchi et al., 2021] Sacchi, R., Monaco, V., Vignati, A., Cirio, R., Sola, V., et al. (2021). Energy measurement of clinical proton beams with a telescope of ultra-fast silicon detectors. *Il nuovo cimento C*, 44(143):1–4.

[Sadrozinski et al., 2016] Sadrozinski, H.-W., Anker, A., Chen, J., Fadeyev, V., Freeman, P., Galloway, Z., Gruey, B., Grabas, H., John, C., Liang, Z., Losakul, R., Mak, S., Ng, C., Seiden, A., Woods, N., Zatserklyaniy, A., Baldassarri, B., Cartiglia, N., Cenna, F., Ferrero, M., Pellegrini, G., Hidalgo, S., Baselga, M., Carulla, M., Fernandez-Martinez, P., Flores, D., Merlos, A., Quirion, D., Mikuž, M., Kramberger, G., Cindro, V., Mandić, I., and Zavrtanik, M. (2016). Ultra-fast silicon detectors (ufsd). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 831:18–23. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors.

[Santa Cruz et al., 2001] Santa Cruz, G., Palmer, M., Matatagui, E., and Zamenhof, R. (2001). A theoretical model for event statistics in microdosimetry. ii: Nonuniform distribution of heavy ion tracks. *Medical physics*, 28(6):997–1005.

[Schardt, 2016] Schardt, D. (2016). Hadrontherapy. In *Basic Concepts in Nuclear Physics: Theory, Experiments and Applications*, pages 55–86. Springer.

[Schardt et al., 2010] Schardt, D., Elsässer, T., and Schulz-Ertner, D. (2010). Heavy-ion tumor therapy: Physical and radiobiological benefits. *Reviews of modern physics*, 82(1):383.

[Schardt et al., 2008] Schardt, D., Steidl, P., Krämer, M., Weber, U., Parodi, K., and Brons, S. (2008). Gsi scientific report 2007. *GSI Report*, 1.

[Scholz, 2003] Scholz, M. (2003). Effects of ion radiation on cells and tissues. *Radiation effects on polymers for biological use*, pages 95–155.

[Scholz, 2006] Scholz, M. (2006). Dose response of biological systems to low-and high-let radiation. In *Microdosimetric response of physical and biological systems to low-and high-LET radiations*, pages 1–73. Elsevier.

[Serber, 1947] Serber, R. (1947). Nuclear reactions at high energies. *Physical Review*, 72(11):1114.

[Shannon, 1949] Shannon, C. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.

[S.M. Sze, 2006] S.M. Sze, K. K. N. (2006). *Physics and Properties of Semiconductors—A Review*. John Wiley & Sons, Ltd.

[Smathers et al., 1977] Smathers, J. B., Otte, V. A., Smith, A. R., Almond, P. R., Attix, F. H., Spokas, J. J., Quam, W. M., and Goodman, L. J. (1977). Composition of a-150 tissue-equivalent plastic. *Medical physics*, 4(1):74–77.

[Sola et al., 2019a] Sola, V., Arcidiacono, R., Boscardin, M., Cartiglia, N., Dalla Betta, G.-F., Ficorella, F., Ferrero, M., Mandurrino, M., Pancheri, L., Paternoster, G., et al. (2019a). First FBK production of 50 μm ultra-fast silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 924:360–368.

[Sola et al., 2019b] Sola, V., Arcidiacono, R., Boscardin, M., Cartiglia, N., Dalla Betta, G.-F., Ficorella, F., Ferrero, M., Mandurrino, M., Pancheri, L., Paternoster, G., and Staiano, A. (2019b). First fbk production of 50 μm ultra-fast silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 924:360–368. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[Sola et al., 2019c] Sola, V., Arcidiacono, R., Boscardin, M., Cartiglia, N., Dalla Betta, G.-F., Ficorella, F., Ferrero, M., Mandurrino, M., Pancheri, L., Paternoster, G., and Staiano, A. (2019c). First FBK production of 50 μm ultra-fast silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 924:360 – 368. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[Srdoč, 1970] Srdoč, D. (1970). Experimental technique of measurement of microscopic energy distribution in irradiated matter using rossi counters. *Radiation research*, 43(2):302–319.

[Steidl et al., 2008] Steidl, P., Schardt, D., Iancu, G., Kraemer, M., and Weber, U. (2008). Precision measurements of bragg curves of light-ion beams in water. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 43.

[Tarasov and Bazin, 2008] Tarasov, O. and Bazin, D. (2008). LISE++: Radioactive beam production with in-flight separators. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 266(19-20):4657–4664.

[Tommasino and Durante, 2015] Tommasino, F. and Durante, M. (2015). Proton radiobiology. *Cancers*, 7(1):353–381.

[Tommasino et al., 2017] Tommasino, F., Rovituso, M., Fabiano, S., Piffer, S., Manea, C., Lorentini, S., Lanzone, S., Wang, Z., Pasini, M., Burger, W., et al. (2017). Proton beam characterization in the experimental room of the trento proton therapy facility. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 869:15–20.

[Tommasino et al., 2015] Tommasino, F., Scifoni, E., and Durante, M. (2015). New ions for therapy. *International journal of particle therapy*, 2(3):428–438.

[Vignati et al., 2022] Vignati, A., Abujami, M., Bersani, D., Borghi, G., Vignali, M. C., Data, E., Ficorella, F., Galeone, C., Garbolino, S., Giordanengo, S., Ali, O. H., Villarreal, O. M., Milian, F. M., Mazza, G., Staiano, A., Wheadon, R., Cirio, R., Sacchi, R., and Monaco, V. (2022). Monitoring therapeutic proton beams with lgad silicon detectors. *Journal of Instrumentation*, 17(11):C11001.

[Vignati et al., 2020] Vignati, A., Donetti, M., Fausti, F., Ferrero, M., Giordanengo, S., Ali, O. H., Villarreal, O. M., Milian, F. M., Mazza, G., Monaco, V., et al. (2020). Thin low-gain avalanche detectors for particle therapy applications. In *Journal of Physics: Conference Series*, volume 1662, page 012035. IOP Publishing.

[Vignati et al., 2017] Vignati, A., Monaco, V., Attili, A., Cartiglia, N., Donetti, M., Mazinani, M. F., Fausti, F., Ferrero, M., Giordanengo, S., Ali, O. H., et al. (2017). Innovative thin silicon detectors for monitoring of therapeutic proton beams: preliminary beam tests. *Journal of Instrumentation*, 12(12):C12056.

[Virtanen et al., 2020] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

[Wigley et al., 2009] Wigley, A., Sutton, M., Wheelwright, S., Burbidge, R., and MacLoed, R. (2009). *Microsoft® NET Compact Framework (Core Reference)*. O'Reilly Media, Inc.

[Wilkinson, 1996] Wilkinson, D. (1996). Ionization energy loss by charged particles part i. the landau distribution. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 383(2):513–515.

[Wilson, 1946] Wilson, R. R. (1946). Radiological use of fast protons. *Radiology*, 47(5):487–491.

[Zaider et al., 1996] Zaider, M., Rossi, B. H. H., and Zaider, M. (1996). *Microdosimetry and its Applications*. Springer.