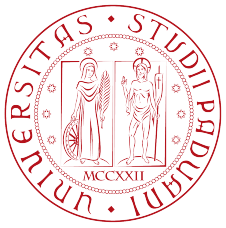


UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Ph.D. School in PHYSICS

Doctoral Thesis

Development of a Proton Tomography scanner

Ph.D. School Coordinator:
Prof. Franco Simonetto

Supervisor:
Prof. Piero Giubilato

Candidate:
Filippo Baruffaldi
Matricola: 1203329

XXXIV Cycle

Abstract

iMPACT, innovative Medical Proton Achromatic Calorimeter and Tracker, is a University of Padova and INFN research project, funded by the European Research Council. The project aims to design, develop and prototype a fast and accurate proton Computed Tomography (pCT) Scanner, with the ultimate goal of demonstrating the technology necessary to realize a clinically viable pCT system. The overall development, current state, and projected performances of the scanner will be illustrated and discussed. Monte Carlo simulation, a selection of data collected with cosmic rays, and tests with a proton beam will be reviewed as well to quantitatively assess the performance of the apparatus. Preliminary studies on proton track reconstruction, based on a Maximum Likelihood path formalism, will be also presented, together with a supporting object shape identification algorithm.

The iMPACT scanner is essentially made by a multi-layer silicon pixels sensors tracker stage using the ALPIDE sensors, and a scintillators-based range calorimeter. There will be an in-depth review of the innovative, highly segmented structure of the calorimeter, based on multiple, orthogonal scintillating elements, and of its read-out architecture, which exploits massive FPGAs parallelism and distributed memory to achieve the triggering and data collection performance necessary to cope with the extremely high event-rate requested by pCT applications. On the tracker side, an overview of the ALPIDE sensor, developed within the ALICE Collaboration for its Inner Tracking System (ITS), and currently adopted for the prototyping phase of the iMPACT tracker, will be illustrated as well, together with the general tracker layout and operations. In parallel, in order improve upon the techniques and methods used in particle physics for tracking purposes, specific studies have been performed to optimize the ALICE ITS alignment, which results will be also presented.

Finally, a brief mention will be given to the INFN project ARCADIA, focused on the development of innovative Monolithic Active Pixel Sensors characterized by fully depleted substrate to improve the charge collection efficiency and timing characteristics over a wide range of operational and environmental conditions. The iMPACT project in fact plans to employ the ARCADIA technology to build a pixel detector more suited for the pCT application compared to the ALPIDE sensor.

Contents

1 Introduction

Hadron-therapy is a leading edge procedure for the treatment of tumors, one of the main causes of death in every Country of the world; it exploits the ballistic precision of protons and ions to focus the dose deposition in a confined volume, thus sparing healthy tissues surrounding the tumor. This technique requires precise 3D maps of the biological tissues along the entire particle path leading up to the tumor, with exact Stopping Power measurements, which are essential to tune the beam energy for an effective and accurate energy deposition. Nowadays, these Stopping Power maps are obtained with X-ray Computed Tomography (X-ray CT), which is however not capable of delivering measurements accurate enough to fully exploit the intrinsic precision of the hadron treatment, mainly due to the different physical interactions that photons and charged particles undergo.
Proton Computed Tomography (pCT) is a technique which is expected to increase the precision of the treatments plans, by using protons also for the imaging procedure; pCT has also been proven to be advantageous, with respect to X-ray CT, in dosimetric terms. Several prototypes are being developed around the world, however, pCT is still yet not clinically available, for technological, physical, and economical limitations.
iMPACT, *innovative Medical Proton Achromatic Calorimeter and Tracker*, is a University of Padova and INFN project, funded by the European Research Council, which aims at developing and prototyping an extremely fast and accurate pCT scanner, with ultimate goal of representing a concrete step towards the viability of the method in a medical environment. The scanner includes a silicon pixels sensor tracker stage and a highly-segmented scintillator-based calorimeter.
This thesis begins by displaying both the benefits and the limitations of the hadron-therapy procedure, especially compared with the more widespread X-ray radiotherapy; the concepts behind pCT are then presented, with an overview of the current state-of-the-art pCT prototypes, with a highlight on the still present limitations inhibiting the technique from being extensively applied.
The thesis will then focus on the design characteristics of the iMPACT scanner concept, with a particular attention on the technical details, and the solutions adopted in order to achieve the ambitious goal that the project has set. The development of a Monte Carlo simulation, based on the GATE-Geant4 framework, is then illustrated. The simulation has been used to validate the calorimeter architecture, guide its development, and provide experimental-like datasets for the development of the analysis tools; it will be also used in the future for the complete calibration of the apparatus. Results from test beams are presented, showing the potentiality of the iMPACT scanner as an innovative pCT apparatus.
A table-top versatile acquisition system, developed in the iMPACT framework for hardware test employing cosmic rays, is presented, together with a summary of the measurements, prototyping assessments, and hardware characterizations performed with the system. The thesis then focuses on the most recent advancements in the iMPACT calorimeter read-out architecture implementation, based on a heavily-parallel structure of FPGAs, performing triggering and data-formatting tasks.
During the first half of the 2020, given the global CoViD situation and the difficulties in accessing physical laboratories, the focus of the project shifted towards studying imaging techniques, which will be eventually employed once the system reaches its full functionality. Applications of proton track reconstruction techniques have been investigated, using Monte Carlo simulations modeling the iMPACT scanner layout and performances, and the results of these studies are presented, as well as a proposed algorithm for a preliminary imaging reconstruction.
The iMPACT tracking system is based, for its first prototype, on the ALPIDE sensor, designed by the ALICE Collaboration at CERN; members of the iMPACT project have been actively participating in the development and commissioning efforts of the sensor. As member of ALICE, the author of this thesis has performed studies on the assembly alignment of the Inner Tracking

System of the ALICE detector, which is composed of the pixel sensor in use for iMPACT; this service work also represented an excellent opportunity to get familiar with the techniques and methods used in particle physics for tracking purposes, allowing for the eventual application of the gained competences in the iMPACT tracking tasks.
Finally, the development of a novel silicon Monolithic Active Pixel Sensor is presented, with a brief overview on an innovative read-out architecture, in the framework of the ARCADIA INFN Collaboration; the new sensor is designed for a wide range of application, in the medical, space, and particle physics fields, and will be employed in future versions of the iMPACT scanner.

2 Proton-Therapy

2.1 Protons and Hadrons in cancer treatment

Cancer is one of the main causes of death, in every country of the world. In fact, from a 2019 estimate from the World Health Organization (WHO), tumors are the first or the second cause of premature death in the majority of the counties (112 out of 183), and between the third and the fourth cause in 23 additional countries [1] [2]. As highlighted in Fig. 1, cancer ranks higher especially in higher income countries, while in developing countries other diseases prevail. Moreover, the incidence of cancer-related deaths is rapidly growing worldwide, as the mortality rates of stroke and coronary heart disease are declining [1].

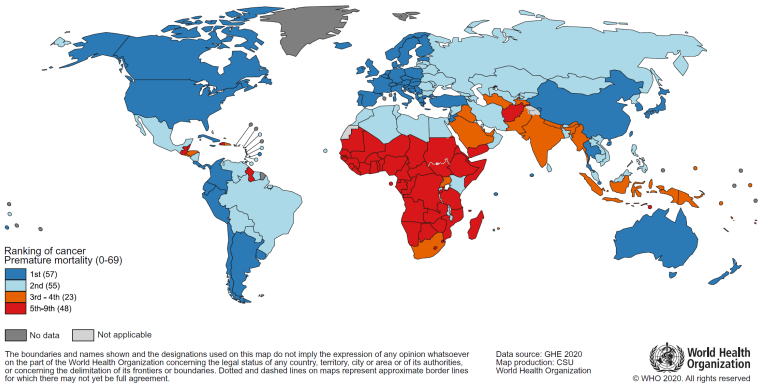


Fig. 1: National ranking of cancer as a cause of death before 70 years of ages in 2019. The legend include the numbers of countries represented in each ranking group. From: World Health Organization (WHO) [1] [2].

The most common tumor treatment procedures are surgery, chemotherapy, hormone therapy [3], and therapy with ionizing radiation (also *radiotherapy*), often combined to increase the chance of success. The overall success rate of tumor treatments is around 45%. In more than 50% of the cases a treatment with X-ray radiation is performed [4] [5].
The goal of radiation therapy is to deliver a sufficient amount of ionizing radiation dose (deposited energy per unit mass) to the cancerous tissues to kill the harmful cells; in particular, the objective is to cause damage to the DNA strains within the cell, rendering it unable to reproduce. At the same time, dose deposition in the surrounding healthy tissues has to be limited as much as possible. In the vast majority of the cases photons are chosen to deposit the required dose, primarily X or gamma rays up to a few MeV, however electron beams with energies of the MeV order are also used, especially to treat superficial skin carcinomas [6]. X-rays and gamma rays are relatively easy to produce, and can be generated with compact radiogenic systems: electrons are accelerated with a voltage difference and produce photons from bremsstrahlung while stopping in a target made of metals, such as tungsten or molybdenum or alloys [5].
The intrinsic limitation of X-ray and gamma radiation therapy comes from the energy deposition profile of the photons, shown in Fig. 2. The energy deposition for photons reaches its maximum a few cm under the surface and then decreases exponentially with the depth, as photons are absorbed by the medium. This behaviour compromises the possibility of depositing energy in a well defined volume, without considerably affecting the surrounding tissues; this characteristic represents a significant limitation especially for the treatment of deep-seated tumors. To alleviate the effect and to spread the collateral energy deposition over a larger volume, the tumor area is often targeted from different directions, see Fig. 4.
Fig. 2 also shows the energy deposition profile of a 107 MeV proton beam (blue curve) and a 200 MeV u^{-1} ^{12}C beam (red curve). In the energy range of interest of ion therapy, the ionizing

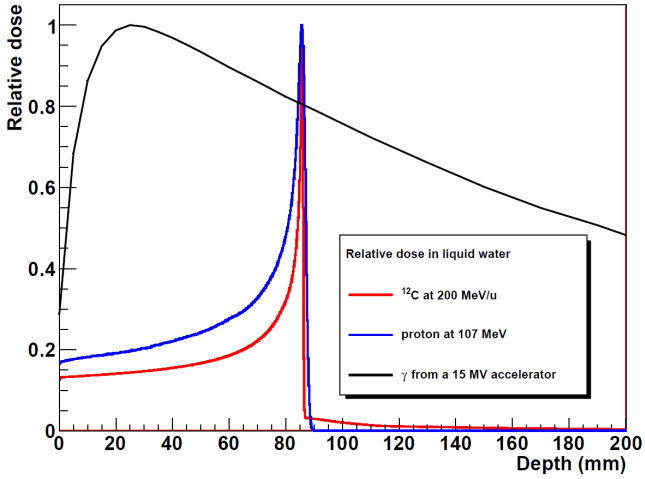
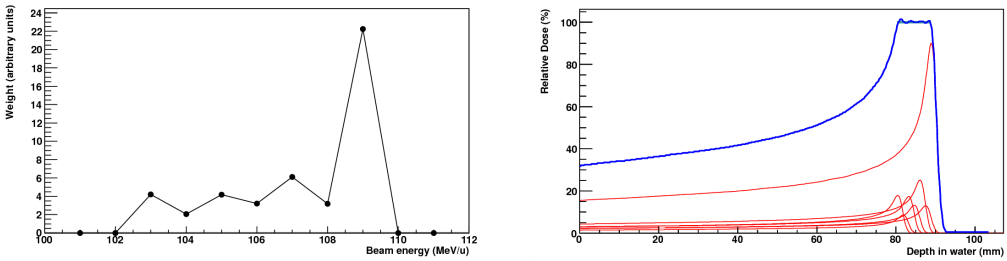


Fig. 2: Comparison of the dose depth profile in water for carbon ions (red curve), protons (blue) and X-rays (black) [5].

energy deposition of charged particles is inversely proportional to the square of their velocity:

$$\left\langle \frac{dE}{dx} \right\rangle_{el} \propto \frac{z^2}{v^2} \quad , \tag{1}$$

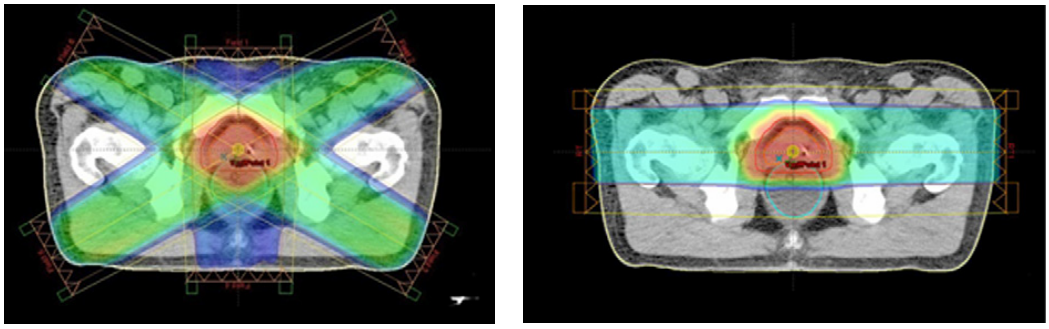
where z and v are the particle charge and velocity, therefore the dose profile presents a sharp maximum in proximity of the end of the particle path. This final spike is know as the Bragg peak, after William Henry Bragg who studied the phenomenon in 1904 [7]. The complete expression of the ionizing energy loss of charged particles is described by the Bethe-Bloch formula in Eq. (6). The energy deposition at the start of the particle path is relatively low, about 20% of the maximum, and it is almost absent after the peak for proton; for carbon and other heavy ions a long tail is present after the peak, caused by the deposited energy of the fragments generated in nuclear collisions. On the other hand, the peak is visibly sharper for the carbon beam with respect to the protons one; in fact, the fluctuations in the stopping position for protons is more than 3 times larger than carbon ions (see Sec. 2.2.2).
The depth at which the Bragg peak occurs can be tuned by adjusting the initial energy of the particle beam, while its height in a given material is approximately fixed, as it is regulated by the energy loss rate (as described in Sec. 2.2.1). Protons and ions beams can be therefore employed to target tumor tissues, with a better control on the affected volume with respect to X-rays, by precisely adjusting the peak deposition to match the tumor depth inside the body; the tissues preceding the tumor receive only a fraction of the dose, and the deposition in the volumes behind the tumor is limited. The procedure of treating cancer with ions is named *hadron-therapy*, or *proton-therapy* if limited to protons. Hadron-therapy with ions mainly employs ^{12}C, and very rarely ^{20}Ne [8]. Ion therapy is used, in particular, for the treatments of types of pediatric cancers, ocular melanomas, chordomas and chordrosarcomas (slow growing bone cancers) [9].
To cover the entire volume of a tumor with a uniform dose distribution, ion beams with different energy are used, in the process called *tumor painting*: as shown in Fig. 3 (a), the energies and the relative intensities of the beams can be precisely calibrated, so that the resulting energy deposition profile, which is the convolution of the single curves, presents a plateau region matching the size of the tumor (Fig. 3 (b)). The uniform dose region is also known as Spread-Out-Bragg-Peak, or SOBP; in the aggregated dose profile the fraction of the energy deposited in the initial portion of the curve is higher with respect to single energy Bragg peak, depending on the size of the tumor and the number of beams used to paint the volume, up to 50% of



(a) (b)

Fig. 3: (a) Energy weight distribution of various proton beams, in order to create a Spread-Out-Bragg-Peak (SOBP), (b) corresponding energy deposition profile of the SOBP; from [5].

the maximum [5]. The slight variations on the plateau are derived by the discretization of the energies. The beams of different energies are generated either actively, adjusting the energy of the accelerator energy, or passively using absorber layers of calibrated thicknesses.
The tweaking of the energies of the single beams and their weight, as well as their direction is planned by computed software, the Treatment Planning System (TPS), which includes information of the 3-dimensional distribution of the biological tissues in the patient, the target volume, the parameters of the specific apparatus used to administer the treatment, and models taking into account the physical behaviours of the primary beam in addition to the effects of secondary particles generated inside the patient [4]. The TPS software computes the optimal treatment plan that delivers the sufficient dose to the tumor, while minimizing the collateral exposure of healthy tissues; an example of a clinically employed TPS is the GSI TRiP software [10]. Fig. 4 shows a comparison between a plan for a treatment of the same pathology with X-rays, in (a), and with protons, in (b); the color palette indicates the predicted deposited dose, with the color blue and red indicating areas affected by a qualitatively lower and higher dose respectively.



(a) (b)

Fig. 4: Example of a treatment plan for prostate cancer, with dose distribution. (a) Axial view of X-ray radiotherapy plan. (b) Axial view of proton plan. The color palette indicates the predicted dose, qualitatively blue and red corresponds to lower and higher doses respectively. A red line outlines the target area. Images from [11].

For the X-ray planning, the tumor, placed at the centre of the body, is targeted from a number of different directions, and the dose deposition is relatively high along the entire beam path, affecting not only the designed area, outlined by a red line, but also a large portion of the surrounding tissues. In this case, the area where the highest energy deposit is supposed to take place is roughly defined by the overlap of the various beams, which does not match perfectly with the target area.
In the treatment planning with protons, Fig. 4 (b), the tumor is targeted from two directions

and the unwanted energy deposition in the surrounding tissues is both less broad and of lower intensity than the previous case with X-rays. With proton beams, the tumor spot can be covered more precisely, and the dose distribution can be more sharply defined.

2.1.1 Biological effects of radiation

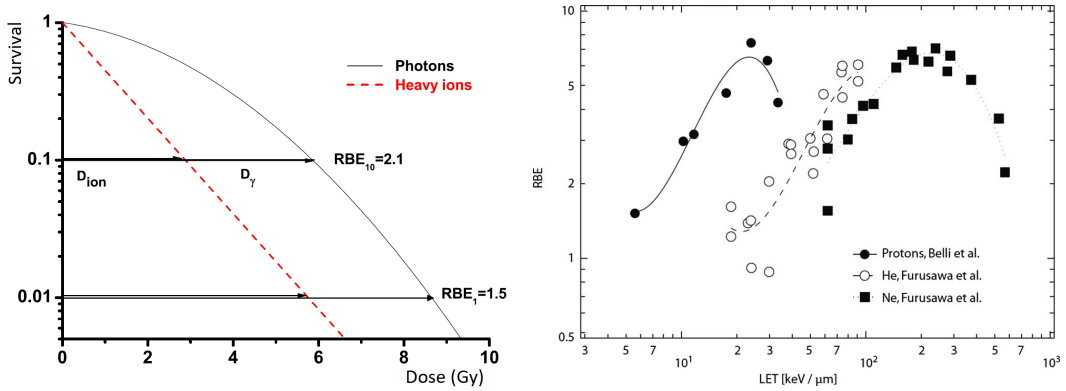
The damage to biological tissues occurs in different ways; moreover, damage to various cell structures leads to different consequences. Photons traversing matter interact mainly with the electrons of the medium, via the: photoelectric effect, Compton scattering, or pair production. In all these scenarios, the photon energy is totally or partially transferred to electrons, which then travel and further ionize the atoms of the material. Protons and ions inside matter mainly lose energy to electrons, causing ionization; they also interact with the nuclei of the material, via Coulomb interactions, which cause deviations in the path of particle. Additionally, protons and ions can undergo nuclear interactions; both the projectile and the target nucleus can fragment and produce secondary particles, such as neutrons, protons, light ions and gamma rays. The secondary particles produced via nuclear interactions are responsible of about 10% of the total dose deposited by a 200 MeV proton beam [12].
The ionization process can occur directly in a structure of the cell, hindering the chemical stability of the bonds, or indirectly by creating chemically reactive free radicals [13]. Not all structural damages are lethal for the cell, as cells are able to repair most of the damage; the most vulnerable component of a cell is its DNA, located in the nucleus. In particular, damage occurring simultaneously to both the strands in a DNA double-helix has been proven to be the most effective, while strand damage can still be repaired by the cell if the opposing strand remains intact [4].
Besides the greater dose deposition precision, an additional advantage of proton and hadrontherapy, over conventional radiotherapy, is given by their higher biological effect on cells. In fact, in order to kill the same fraction of living cells, ions require less deposited dose than X-rays [5]. This effect is quantified with the Relative Biological Efficiency factor (or RBE), which is defined as the ratio between the dose deposited by photons D_γ and the dose deposited by a given hadron D_{hadr} in order to cause the same biological damage to the tissue:

$$\text{RBE} = \frac{D_\gamma}{D_{\text{hadr}}} \tag{2}$$

where the biological damage is evaluated by the fraction of the number of cells S that survive after the deposition of a dose D; dose is often measured in Gray, 1 Gray = 1 J/kg. RBE factors are usually reported for survival rates of 10% and 1%. The survival fraction S can be parameterized by a linear-quadratic model:

$$S(D) = \exp(-\alpha D - \beta D^2) \tag{3}$$

where α and β are experimental determined parameters [14]. The linear-quadratic model has been proven to be reasonably accurate in the range of clinical applications for tumor treatments [15].
Figure 5 (a) shows typical trends of survival fractions for photons (black solid line) and for a generic heavy ion (red dashed line). The RBE depends on a variety of factors, such as the particle type, the deposited dose, the tumor type, and the characteristics of the targeted biological cells. On average, the RBE for protons is estimated to be around 1.1, while for heavier ions it is larger; in particular carbon has an average RBE of 3, making it more effective than both protons and photons [5].
Another key quantity related to biological damage caused by ionizing radiation is the Linear Energy Transfer (LET). While the Bethe-Bloch ionization curve (or Stopping Power, Eq. (6)) quantifies the energy loss by the particle in traversing a material, the LET (commonly expressed



(a) (b)

Fig. 5: (a) Survival fraction (number of surviving cells over the total) as a function of deposited dose, for photons and a generic ion. The RBE factor (Relative Biological Efficiency) is often indicated for two reference levels: 10% and 1% of survival fraction. (b) $\mathrm{RBE}_\alpha = \alpha_{ion}/\alpha_\gamma$ as a function of LET for different particles; α_{ion} and α_γ are the linear part of the survival curve for photons and ions, respectively, defined in Eq. (3), retrieved from experimental data with V79 hamster cells [16] [4].

in keV µm^{-1}) indicates the energy absorbed by the medium, per unit distance, considering only energy transfers below a certain threshold Δ_E:

$$\mathrm{LET}_{\Delta_\varepsilon} = \frac{dE_{\Delta_\varepsilon}}{dl} \quad , \tag{4}$$

where dE_{Δ_ε} is the average local energy absorbed, through collisions involving transfer energies lower than a specific value Δ_ε [17]. If Δ_ε is set to infinite, LET results numerically equal to Stopping Power. The value of the energy threshold limits on the range of the secondary particles generated by the interactions; LET is therefore an estimation of the local ionization taking place in the close proximity of the particle trajectory. Usually, the energy threshold is set to exclude delta-rays, i.e. electrons freed by ionization but with enough energy to cause further ionization.

Protons and ions offer a higher LET with respect to photons; this is because the nature of the interaction of charged particle with matter implies that their energy deposition is continuously distributed along the particle trajectory, and concentrated around it. In fact, even for protons in the medical domain (200 MeV), the maximum energy transferred to secondary electrons is about 500 keV, which translates in a range of around 2 mm in water [12]. Photons, instead, tend to produce more scattered and point-like interactions. For these reasons, photons are often referred to as low-LET radiation, while protons and heavier ions as high-LET radiation; moreover, the heavier the ion, the higher LET it can produce [4].

Fig. 5 (b) shows the RBE of different ions, as a function of their LET. Generally higher LET values lead to a stronger biological effect, as higher energy deposition causes more damage to the biological tissues. However, as LET increases RBE reaches a maximum, followed by a decrease of the biological effect of the radiation. The position of the maximum RBE is related to an average distance between ionizing interactions that is around 2 nm, corresponding to the average separation between the two strands forming a DNA helix, therefore increasing the chance of a double DNA break. At higher LET, a single ionization is sufficient to fatally damage the cell, so the energy deposited by successive ionization is wasted; additionally, it requires a lower number of higher LET particles to reach the total planned dose, thus the number of spared cells increases, and so does RBE [4]. The position of the maximum RBE peak is located at higher LETs for heavier ions.

Another crucial factor in cancer treatment is the oxygenation level of the cells under irradiation.

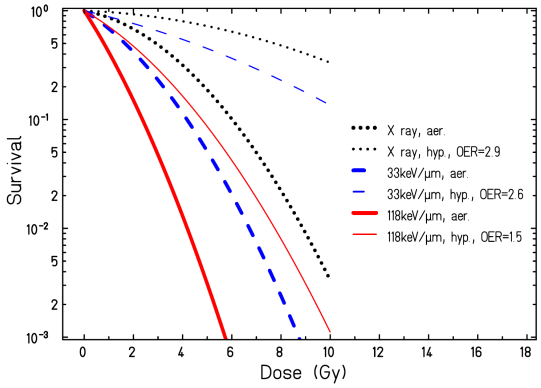


Fig. 6: Survival fraction as a function of the dose of human kidney cells after irradiation with X-rays (dotted black lines) and carbon ions with different LET (dashed blue lines and solid red lines); the survival fraction trends are shown for hypoxic (bold lines) and normally oxygenated cells; the lack of oxygen hinders the cell elimination. Experimental data [4] [18]

As tumors grow in size, to sustain their increasing mass, they require increasingly more oxygen, which the body struggles to provide. For this reason, areas with a lower oxygen level tend to form, especially at the centre of larger tumors. Hypoxic conditions have been seen to lead to a higher biological resistance to radiation damage, with respect to normally oxygenated (aerobic) tissues [4], i.e. healthy tissues are more sensitive to radiation than tumor tissues. This effect can be quantified by the Oxygen Enhancement Ratio (OER):

$$\mathrm{OER} = \frac{D_{\mathrm{hypoxic}}}{D_{\mathrm{aerobic}}} \tag{5}$$

where D_{hypoxic} and D_{aerobic} are the dose required to have the same survival fraction on hypoxic and aerobic cells respectively. A widely accepted explanation for the radiation resistance caused by low oxygenation is still missing [4].
Fig. 6 shows survival ratio curves for X-rays and carbon ions with a lower, and a higher LET, relative to cells with normal or low oxygenation; the results are retrieved from experimental data on human kidney cells [18]. As shown in the graph, the effectiveness of radiation is reduced when dealing with hypoxic cells, however the difference is less pronounced for the higher-LET particle. For X-rays the average OER is on average 3: in hypoxic conditions it takes three times the dose, with respect to aerobic conditions, to achieve the same survival fraction. For particles with low-LET values, as ^{12}C at high energies, the OER value is around 2.6, while for high-LET values (i.e. ^{12}C close to the Bragg peak), the OER value decreases down to 2 [5]. The OER decreases with the ion mass.

2.2 The physics of proton and hadron-therapy

Protons and heavy ions in matter travel along a roughly straight trajectory, continuously losing energy by Coulomb inelastic interactions with atomic electrons; in fact, the electron mass is too low to substantially deviate a proton ($m_p/m_e = 938.272\mathrm{MeV}/0.511\mathrm{MeV} \approx 1836$), let alone heavier ions. The energy-loss rate of a charged particle via ionization is described by the Bethe-Bloch formula, Eq. (6).
At the same time, proton interact with the positive-charged atomic nuclei via Coulomb elastic scattering, which causes the deflection of the particle trajectory. The probability distribution of the deflection angle is parameterized by the Rutherford's cross-section, Eq. (18)).
Hadrons can also undergo nuclear inelastic interactions with the nuclei; nuclear interactions are less frequent; around 20% of 200 MeV protons undergo nuclear inelastic reactions along their

path [5]. In this case, the projectile overcomes the Coulomb barrier of a positively charged nucleus and gets in contact with the nucleons, producing secondary particles (see Sec. 2.2.3).

2.2.1 Energy loss and mean Range

The mean energy loss rate per unit length, or Stopping Power S_p ($S_p = -\mathrm{LET}_{\Delta_\varepsilon \to \infty}$), of a projectile particle in a material via inelastic Coulomb interactions with electrons can be described by the Bethe and Bloch formula [19], [20] (a quantistic extension of the Bohr calculation [21]):

$$S_p = -\left\langle \frac{dE}{dx} \right\rangle_{el} = \frac{4\pi}{m_e c^2} \cdot \frac{nz^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\varepsilon_0} \right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I \cdot (1-\beta^2)} \right) - \beta^2 - \frac{\delta}{2} - \frac{C}{Z} \right] \tag{6}$$

where m_e is the electron mass, β is the projectile speed in units of c, z is the the projectile electric charge, I is the mean excitation potential of the material, δ is the outer electrons shielding correction, and C is a shell-correction parameter; the last two terms were added as relativistic corrections [22]. The electron density n can be calculated with:

$$n = \frac{N_A \cdot Z \cdot \rho}{A \cdot M_u} \tag{7}$$

where N_A is the Avogadro number, Z and A are the material atomic and mass number respectively, ρ is the mass density of the absorbing material, and $M_u = 1$ g/mol. The maximum energy loss is achieved for particles velocity of:

$$v \approx z^{\frac{2}{3}} v_0 \quad , \tag{8}$$

where $v_0 = e^2/\hbar$ is the Bohr velocity, for which the corresponding β is $e^2/\hbar c = 1/137$ [4]. At particle speeds lower than $z^{\frac{2}{3}} v_0$, the particle charge has to be substituted [23]:

$$z \to z_{\mathrm{eff}} = z \left[1 - \exp\left(-125 \beta z^{-\frac{2}{3}} \right) \right] \tag{9}$$

to consider the decrease of mean projectile particle charge due to ionization and recombination processes. The energy-loss expression in Eq. (6), is an average quantity for a particle beam. Due to statistical fluctuations in the single ionization processes, a single particle experience an actual energy loss which is distributed as a Landau distribution, around the mean value defined by the Bethe-Bloch formula (see Sec. 2.2.2).
The precise estimation of the parameter I, the mean excitation potential, is fundamental for the range calculation of a particle in a given material; in a compound, it can be estimated as:

$$\ln(I) = \left(\sum_i \frac{\omega_i Z_i}{A_i} \ln(I_i) \right) \left(\sum_i \frac{\omega_i Z_i}{A_i} \right)^{-1} \quad , \tag{10}$$

where Z_i and A_i are the atomic charge and mass of the i-element, ω_i is its mass ratio and I_i is its mean excitation potential. Values of I for water, obtained with energy loss measurements with different particles, range from 75 eV to 79.7 eV [5].
The Bethe-Bloch equation (Eq. (6)) shows how the Stopping Power is strongly related to the properties of the absorbing material, especially depending on the material density. In fact, the factor Z/A varies less than 20% for biologic materials, and, although the excitation potential I can vary from 19 eV for ^{1}H to 820 eV for ^{208}Pb, the dependence on this parameter is logarithmic. However, energy loss is linearly proportional to the mass density, whose values can cover a range of three orders of magnitude in a human body: from air in the lung (0.0012 g/cm^3) to cortical bone (up to 2 g/cm^3). Clearly, that precise knowledge of the tissue density distribution is essential to plan a proton therapy treatment with sufficient accuracy [12].

Often the Stopping Power of a given material $S_p{}^{\mathrm{mat}}$ is expressed in terms of the relative stopping power RSP:

$$\mathrm{RSP} = \frac{S_p^{\mathrm{mat}}(E)}{S_p^{\mathrm{H_2O}}(E)} \quad , \tag{11}$$

where $S_p^{\mathrm{H_2O}}$ is the Stopping Power in water; both the Stopping Powers are considered at the same particle energy. The Relative Stopping Power is a convenient quantity, as water is taken the reference in many aspects of radiobiology; RSP is practically independent from the energy of the particle, in medical applications: Fig. 7 shows the dependence of RSP on the proton energy, for three different materials. The materials considered are: compact bone (ICRU-119), with a density of $1.8\,\mathrm{g\,cm^{-3}}$, and including a number of different elements such as H, C, N, O, Mg, P, S, and Ca; skeletal muscle (W&W type 1), with a density of $1.058\,\mathrm{g\,cm^{-3}}$ and including, among others elements, K, S, P, and Cl; polyvinyltoluene (PVT, C_9H_{10}, ICRU-570) with a density of $1.032\,\mathrm{g\,cm^{-3}}$; dry air (ICRU-104) with a density of $1.63 \times 10^{-3}\,\mathrm{g\,cm^{-3}}$. The plots in Fig. 7 were produced from Stopping Power table calculated with the Monte Carlo software SRIM [24].

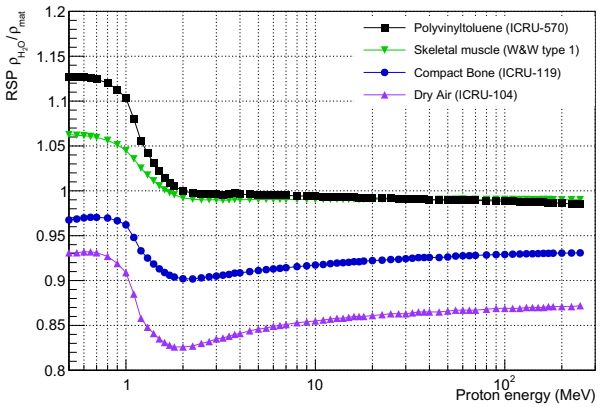


Fig. 7: Relative Stopping Power for different materials, compact bone (ICRU-119), skeletal muscle (W&W type 1), polyvinyltoluene (ICRU-570), dry air (ICRU-104), with respect to water (ICRU-276). Stopping Power tables calculated with the Monte Carlo software SRIM [24]. RSP is re-scaled a factor $\rho_{\mathrm{H_2O}}/\rho_{\mathrm{mat}}$, to count for the large difference in density.

The Relative Stopping Power shows a small variation with energy, about 1%, over two orders of magnitude from 2 to over 250 MeV, for muscle and PVT; for materials as compact bone, RSP varies up to 3% over the same energy range. More noticeable differences are visible below 2 MeV, at which however the proton range in water is about 78 µm. The data in Fig. 7 are obtained from Stopping Power tables calculated with the SRIM Monte Carlo software [24]; RSP values are re-scaled by a factor $\rho_{\mathrm{H_2O}}/\rho_{\mathrm{mat}}$, to take into account the differences in density between the considered materials.

Range is defined as the depth inside a material where half of the particles have come to rest, as presented in Fig. 8, which shows the normalized fluence of a proton beam in water. The mean range is defined as the depth at which the beam fluence reaches the half-height. Due to statistical fluctuations affecting the energy-loss of single particles, the rest points of the particles are distributed around the mean range value. The concept of range does not apply to a single particle, but only for an ensemble beam. Moreover, for particle as photons, which get absorbed with an exponential law, the range can not be defined in these terms.

From the beam entrance point up to the fall-off point, the fluence undergoes a constant reduction, which is caused by the removal of particles due to nuclear interactions with the nuclei of

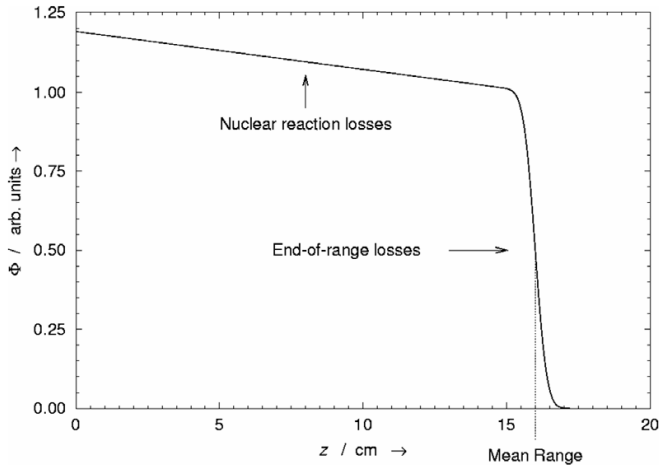


Fig. 8: Normalized fluence Φ of a proton beam as a function of the depth in water. The progressive loss of particles along the path is given by nuclear interactions, while the final steep fall-off is given by particles coming to rest after losing all their kinetic energy to electrons. The normalization reference is set near the end of the path after nuclear losses and before the final fall-off [12].

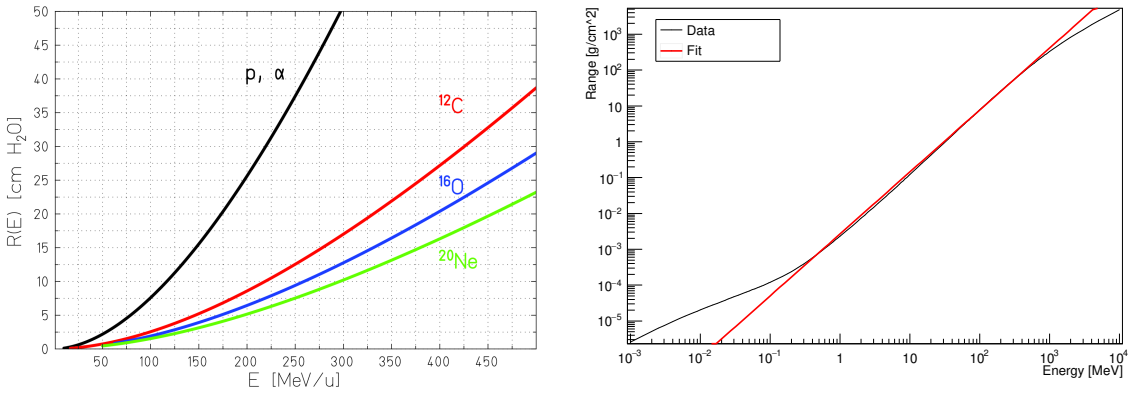
the medium. For the beam range estimates, only particles not affected by nuclear interactions are considered [12].
The range of a mono-energetic beam can be defined as:

$$R(E) = \int_0^E \left\langle \frac{dE'}{dx} \right\rangle_{el}^{-1} dE' \approx \sum_0^E \left\langle \frac{dE'}{dx} \right\rangle_{el}^{-1} \Delta E' \quad , \tag{12}$$

where E is the initial energy of the beam, and the sum is a discrete step approximation. To be precise, the previous equation defines the particles average path-length, which is equal to the average range when the particles path is considered strictly linear, thus neglecting lateral MCS scattering (which will be discussed in Sec. 2.2.2). In reality, due to multiple scattering, a particle beam progressively spreads when travelling in a medium; therefore the range, which is the lateral projection of the path length, is slightly shorter than the latter (usually of less than 0.5%).
Particles ranges can be estimated from an analytical relation, which was empirically identified by Bragg and Kleeman, in 1905. Their formula associates, through a power law, the particle range with energy in a given target material [7]:

$$R(E) = \alpha E^p \tag{13}$$

where α and p are particle and material-dependent parameters, to be derived experimentally [25] [26]. Figure 9 (a) shows range-energy curves for typical ions used in hadron-therapy, from Monte Carlo simulations. The range for ions with the same energy per nucleon scales with a factor of A/Z^2. [4].
Range estimations using the analytical approximation given by the Bragg-Kleeman rule, Eq. (13), are more practical than calculating the Bethe-Bloch numerically. From experimental measurements, the power law is correct in the energy range between 1 and 300 MeV, as visible is Fig. 9 (b), which is more than adequate for biological applications; in fact in this field, particle ranges of interest typically span between 1 mm (the typical voxel size in medical imaging) to about 30 cm (the deepest site in the human body), corresponding to about 11 and 220 MeV respectively for protons in water [12].



(a) (b)

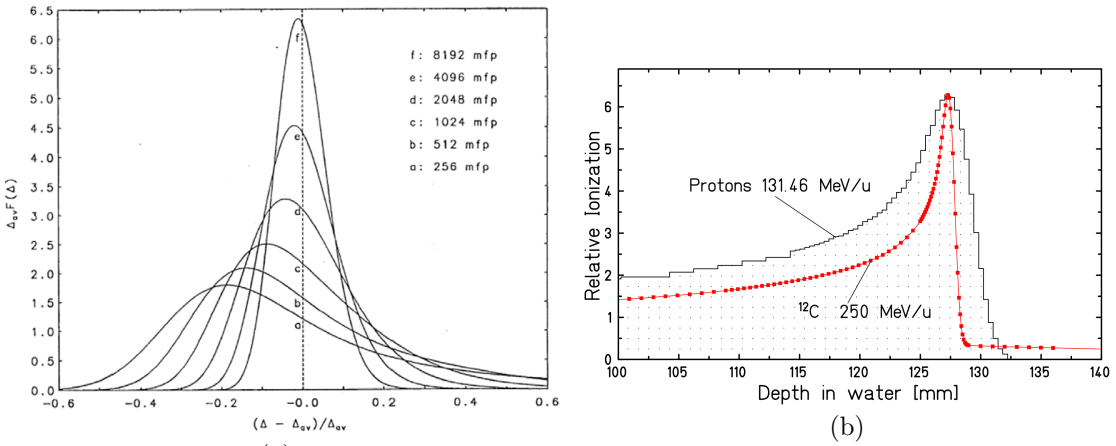
Fig. 9: (a) Simulated (SRIM) range in water as a function of energy, of different particles of interest for hadron-therapy [4]. (b) Range vs proton energy in polystirene-based scintillator material (from NIST data tables [27]) and fit with the Bragg-Kleeman rule (Eq. (13), $R = \alpha E^p$); fit performed in the [1; 300] MeV range. ($\alpha = 0.00279 \pm 0.00006$) g/cm^2, $p = 1.726 \pm 0.004$.

2.2.2 Range and lateral straggling

As introduced previously, the energy loss rate calculation of Bethe-Bloch formula, Eq. (6), describes an average behaviour over a number of particles.The stochastic nature of the single inelastic Coulomb scattering, integrated over a wide number of collisions, leads to statistical fluctuations of the energy loss rate, called energy straggling. The energy loss fluctuations cause variations in the total length of the particles trajectories, thus a dispersion of their stopping position and a broadening of the Bragg peak; this effect is named range straggling.
A particle passing through a layer of material, suffers an energy loss ΔE, which is distributed according to a probability density function (PDF), that was studied by Vavilov [28] and Landau [29], for thin and intermediate thicknesses respectively. Fig. 10 (a) shows the normalized energy loss distribution for protons in absorbers of different thicknesses; the thickness is expressed in terms of mean-free-path. In the graph, the energy loss distributions relative to thin and intermediate thicknesses (a to d) show an asymmetric shape, with the most probable value placed at lower energies than the mean value and a long tail at higher energies. The shape of the energy loss distribution is described by a Landau distribution. The mean value is related to the Bethe-Block formula (6). The distribution shape is given by the fact that the energy loss is dominated by by small energy transfer interactions, lower-bound by the minimum energy necessary to cause a ionization, in addition to less-likely higher-energy interactions which generate δ-rays.
In thicker absorbing layers, the number of single interactions increases, and for the central limit theorem, the Vavilov-Landau distribution describing the energy loss can be approximated by a Gaussian distributions (as studied by Bohr [31, 32]). This effect is visible in Fig. 10 (a), where the distributions relative to the thicker absorbers (f and e) show a distribution progressively more symmetrical. The selection parameter for an appropriate theory, for a given absorber thickness, is often based on a reduced energy loss factor:

$$k = \frac{\xi}{\Delta E_{max}} \quad , \tag{14}$$

where ξ is the mean energy loss and ΔE_{max} is the maximum energy loss in a single event; the transition between the Landau (lower values of k) and the Bohr/Gaussian (higher k) regimes is commonly placed at $k = 10$ [12].
According to Bohr approach, the energy loss distribution can be approximated by a Gaussian



(a) (b)

Fig. 10: (a) Energy loss Probability Density Distributions (PDF) of 10 MeV protons for water absorbers of various thickness, expressed in units of mean-free-path (mfp $\approx 1.34 \cdot 10^{-6}$ g/cm^2) [30]. The single event energy losses are scaled for the mean energy loss in the entire absorber Δ_{av}; PDFs are normalized with unitary integral. (b) Measured energy loss curves for protons and ^{12}C ions with the same mean range in water. The plot shows the smaller energy loss straggling for the heavier ion, which leads to a more defined Bragg Peak [4]

with a mean value $\langle dE/dx \rangle \Delta x$ and a variance $(\sigma_E^{\text{stragg}})^2$, given by:

$$(\sigma_E^{\text{stragg}})^2 = 4\pi z_{\text{eff}} Z e^4 n \Delta x \left[\frac{1 - \beta^2/2}{1 - \beta^2} \right] \quad , \tag{15}$$

where n is the electron density, defined in Eq. (7), z_{eff} the effective projectile charge (Eq. (9)), and Δx the layer thickness [5].
The particles stopping position distribution can be similarly described by a Gaussian, centred at the mean range, and dispersion σ_R^{stragg}; the range straggling is related to the energy straggling σ_E^{stragg} by the expression:

$$(\sigma_R^{\text{stragg}})^2 = \int_0^E \left(\frac{d\sigma_{E'}^{\text{stragg}}}{dx} \right) \left(\frac{dE'}{dx} \right)^{-3} dE' \quad . \tag{16}$$

The ratio between range straggling and range has been shown to be kinetic energy dependent by Bruno Rossi [33], and the relation can be described by:

$$\frac{\sigma_R^{\text{stragg}}}{R} = \frac{f}{\sqrt{m}} \left(\frac{E}{mc^2} \right) \quad , \tag{17}$$

where m is the particle mass and f is a slowly varying function depending on the absorber [5]. For light ions stopped in water the relative straggling $\sigma_R^{\text{stragg}}/R$ is of the order of 10^{-3}. The relative range straggling is smaller for heavier ions, due to the $1/\sqrt{m}$ factor; e.g. for carbon the relative range straggling is a factor 3.5 smaller with respect to protons [4]. This effect is visible in Fig. 10 (b) where the energy loss profile for a proton and a carbon beam with the same mean range in water are presented; the carbon beam show a noticeably more sharp Bragg peak than the proton beam.
Experimental values of range and range straggling have been measured and tabulated by J. Janni [34], for protons with different energies in various materials. For protons in the energy

range of interest for therapy ($\sim$ 200 MeV), the relative range straggling σ_R^{stragg} is distributed between 0.9 and 1.2%. Range straggling determines an intrinsic limit to the precision with which particle ranges can be measured and predicted.

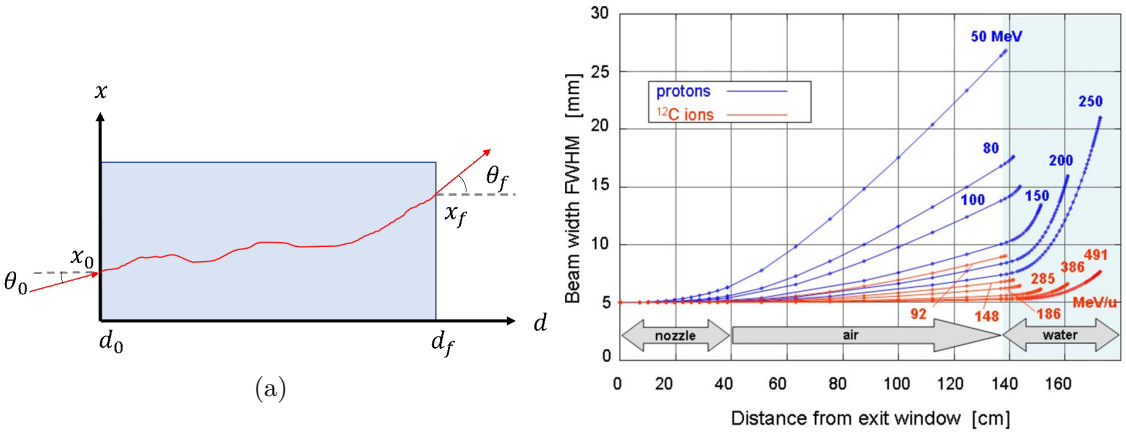


Fig. 11: (a) Schematic diagram of a particle trajectory deviated by Multiple Coulomb Scattering while traversing a material. (b) Calculated beam spread for proton and ^{12}C ion beams in a typical treatment beam line. The initial beams lateral dimension is 5 mm. The beams exit the vacuum beam and travel 1 m in air before entering a water absorber (modelling a patient) [4].

Protons and ions, while travelling in matter, interact multiple times with the positively-charged nuclei of the medium, via elastic Coulomb scattering, which causes deviations in the particle trajectory; in this Coulomb scatterings, the ion loses a negligible quantity of energy [12]. The deviation angle distribution of a single Coulomb scattering is described by the Rutherford differential cross-section [35]:

$$\frac{d\sigma}{d\Omega} = z^2 Z^2 r_e^2 \left(\frac{m_e c}{4\beta p}\right)^2 \sin^{-4}\left(\frac{\theta}{2}\right) \quad , \tag{18}$$

where θ is the final scattering angle, Ω is the solid angle, $r_e = 2.8179 \cdot 10^{-15}$ m is the classical electron radius and p is the particle momentum; given the inverse proportionality on θ, small angular deflections are favored.
When dealing with thick targets, as in most clinical situations, the particle interacts with a large number of nuclei along its trajectory, and a single scattering approximation is not adequate; a schematic diagram of a particle trajectory undergoing Multiple Coulomb interactions in a medium is shown in Fig. 11 (a). In 1948, G. Molière developed a complete and complex Multiple Coulomb Scattering (MCS) formulation, including an analytical solution for the distribution $F(\theta_f, d)$ of the final scattering angle θ_f at depth d; however, in a small-angles approximation, Molière's solution can be expressed as a Gaussian distribution [36], with a standard deviation given by:

$$\sigma_\theta = \frac{1.41\text{MeV}}{\beta p c} z \sqrt{\frac{d}{L_{\text{rad}}}} \left[1 + \frac{1}{9}\log_{10}\left(\frac{d}{L_{\text{rad}}}\right)\right] \quad , \tag{19}$$

where L_{rad} is the material radiation length. Values for radiation lengths in different materials can be found tabulated in [37]; for water L_{rad} is 36.8 g/cm^2. Fig. 11 (b) shows the transverse dimension of proton and ^{12}C beams traversing various materials in a common therapy line layout; it can be noted as denser materials tend to cause a larger lateral spread, such as water with respect to air, while heavier ions, as carbon, suffer a lower deviation than a lighter particle, such as proton. Moreover, particles with lower energies are, on average, deviated by

larger angles. For instance, a 200 MeV proton pencil beam in water at the end of its range, (256 mm) has a lateral spread of 5.5 mm.

2.2.3 Hadronic interactions

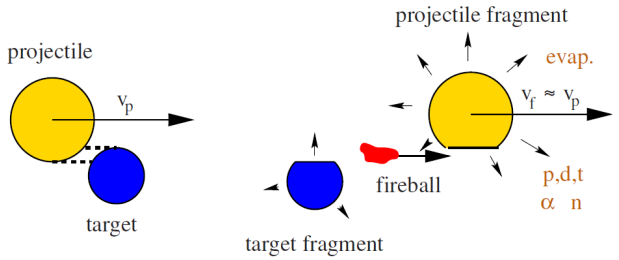
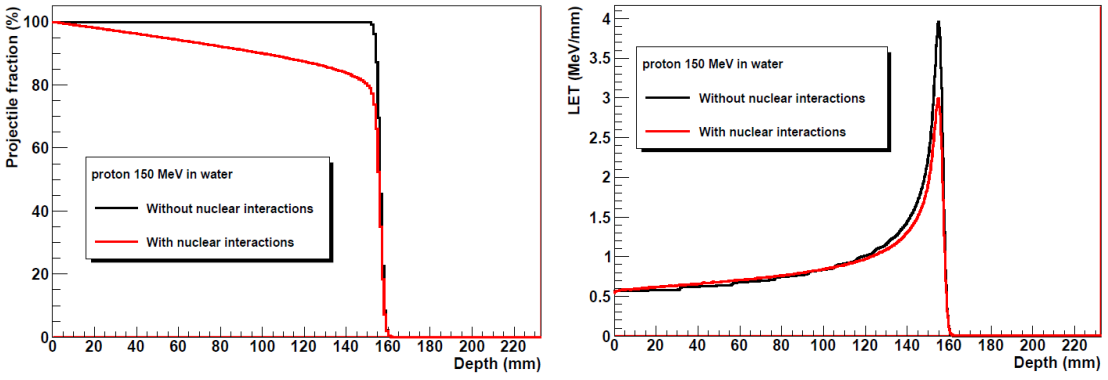


Fig. 12: Schematic representation of an inelastic nuclear collision, with fragment generation and de-excitation via evaporation and gamma decay [4]

Protons and ions in matter can interact via hadronic interactions with the nuclei; these interactions can be either elastic or inelastic. In hadronic collisions, the projectile has enough energy to overcome the Coulomb barrier and interact with the nucleus via the short range strong interaction; if the collision is inelastic, light fragments from the hit nucleus, such as protons, neutrons, deuterons, tritons, and alpha, are emitted promptly. If the primary particle is also a heavy ion, the projectile is likely to undergo fragmentation as well. Both the quasi-projectile and quasi-target nuclei produced in the collision, can be generated in an excited state, which de-excite by light particles evaporation or gamma emission. The process is schematized in Fig. 12.



(a) (b)

Fig. 13: (a) Beam intensity as a function of depth in water of a proton beam, with or without nuclear interactions (black and red lines respectively); (b) Dose profile. Data from Monte Carlo simulations [5].

Inelastic hadronic interactions remove particles from the beam, as can be seen in Fig. 13 (a), in which the intensity of a proton beam of the 200 MeV order decreases of about a 20% factor before reaching the end of its range. Heavier ions are more subject to nuclear interaction, as a ^{12}C with the same mean range loses almost half of its particles before reaching the Bragg peak [5]. The loss of particles along the path causes a decrease in energy deposition in the Bragg peak area, as visible in Fig. 13; heavy localized ionization takes place in the nuclear interaction points, due to the recoiling target nucleus and charged fragments. Fragments are lighter than

the original primary particle, but can be produced with high velocities; being lighter, the fragments can have ranges longer the original particle, and can ionize beyond the Bragg peak, as visible in the dose-depth profile of ^{12}C in Fig. 2. The effect on the total dose deposition from secondary particles is small but not negligible, as secondary protons are responsible for about 10% of the total deposited energy in a proton beam treatment; the dose generated by heavier fragments adds up to about 1% of the total [12].
Secondary neutrons can be dangerous, as they are charge-neutral and can travel tens of centimetres before interacting again with a nucleus in an unpredictable position; in that location, localized ionization spots are created. At the same time, depending on the position of their generation, neutrons can exit the patient's body without interacting; it has been found that the dose from neutrons is a factor 10 lower than the dose produced by charged fragments [12].
Hadrons can also interact with the material nuclei via elastic nuclear collisions, however the contribute of this process to the energy loss begins to dominate at very low energies $E \lesssim 10$ keV/u, namely in the last microns at end of the particle path; this process is therefore neglected in most of particle therapy applications [5].

2.3 Brief historical overview of the hadrotherapy

The X-rays were discovered by Wilhelm Conrad Röntgen in November 1895 at Würzburg, Germany [38]; by the first months of the following year, the newly identified form of radiation was already being used to treat skin conditions, without any understanding of the physics, or the possible side effects. The radiation therapy with X-rays continued to improve during the 20$^{\text{th}}$ century, from the Röntgen discovery, up until the 80s, where the technique reached its maturity with the development of the Intensity-Modulated X-ray Therapy (IMXT), the technique still in use today [39].
Ion therapy was first proposed by Robert R. Wilson in 1946, suggesting the use of machine-accelerated protons for medical purposes [40]. Wilson received his PhD under the supervision of Ernest O. Lawrence, who founded the Lawrence Berkeley National Laboratory in Berkeley, California, and developed the first cyclotron. The clinical use of protons started in 1954, with a limited number of experimental treatments [41], while a couple of years later C. O. Tobias began studies on the medical applications of heavier ions, such as ^{20}Ne [42]; the nuclear physicist was interested in the biological effects of ionizing radiation on living cells, and grasped the advantages of ions over other types of radiation [39] [43].
Over the following decades, numerous experimental physics facilities developed the possibility of administering proton therapy and the technique began to spread: the second laboratory after Berkeley was in Uppsala, Sweden in 1957; in 1961 proton neurological radiosurgery was performed at the Massachusetts General Hospital, while other intracranial adenomas were treated at Harvard in 1963 [39]. Early proton therapy facilities were instituted in Russia: Dubna in 1967, Moscow in 1969 and at St. Petersburg in 1975; slightly later in Japan at Chiba in 1979 and Tsukuba in 1983. In 1985 treatments began in at the Paul Scherrer Institute, Switzerland [39].
A dedicated hospital facility was designed in 1970 at Loma Linda University Medical Center (LLUMC), Loma Linda, USA, while the previous treatment facilities were born in physics laboratory; the LLUMC centre also pioneered computed-assisted treatment planning, in conjunction with computed tomography imaging methods (see Fig. 14), which significantly improved the accuracy and effectiveness of the treatment [39].
In addition to protons and heavier ions, other charged hadrons were investigated for medical use: in 1961 negative pions were proposed [44], and tests were conducted in a few centres around the world (as PSI [45]), however no clinical centers were realized [39].

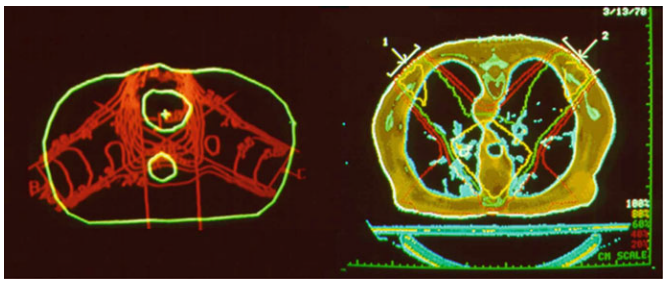
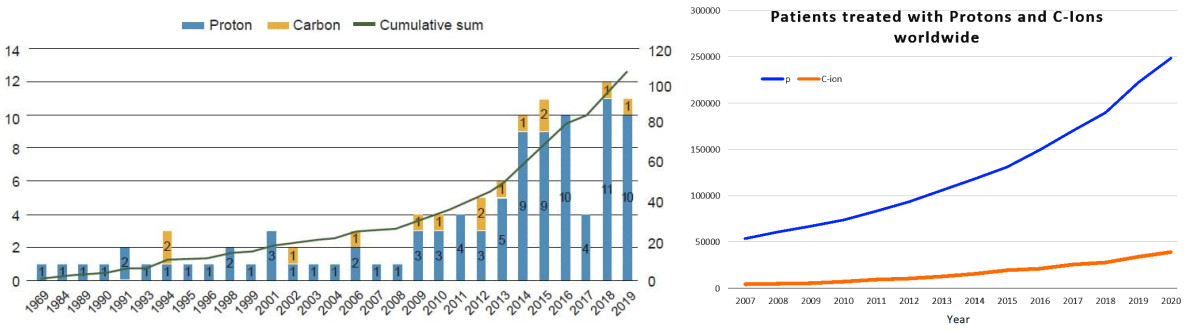


Fig. 14: Example of the first computer-assisted treatment planning graphic visual output developed at LLUMC in the 1970s (left), the patient's body image is generated by ultrasound imaging. Treatment planning from the second LLUMC-developed planning software (right), in 1978, employing a Computed Tomography scan as image source [39].

2.3.1 The present of hadron-therapy

By the end of 2020 more than 290,000 overall patients have been treated with Particle Therapy, worldwide, of which about 250,000 with protons, close to 40,000 with C-ions and about 3,500 with He, pions, and other ions [8].
Hadron-therapy has steadily grown over the last decades, in terms of both number of treated patients and facilities in operation, as shown by the trends in Fig. 15. At the present day, more than 100 hadron-therapy facilities are operational in the world, with the vast majority exclusively with protons; a total of 12 facilities are equipped to deliver ^{12}C ions too [8]. At the top of the ranking for number of centres, are the USA, with 41 and Japan with 24 clinical locations; about 40 centres are operative in Europe, with three in Italy: INFN-LNS in Catania [46], APSS in Trento [47], and CNAO in Pavia, which is the only one capable of producing carbon ions [48]. Moreover, worldwide, around 40 new facilities are planned, or in the construction phase.



(a) (b)

Fig. 15: (a) Number of new hadron-therapy facilities each year and cumulative sum, from 1969 to 2019 [49]. (b) Number of total patients treated with protons and carbon separately, from 2007 to 2020 [8].

Despite the number of available facilities and the increasing number of patients with hadron-therapy, the common X-ray radiotherapy remains vastly more employed; in fact, currently more than 7600 centres administering X-ray radiotherapy are active worldwide, and the number of patients annually treated is of the order of 5 million [50] [51].
The main factor limiting the diffusion of hadron-therapy is the high construction and operation costs of these facilities, including accelerators, beam lines, and patient delivery systems (gantries), together with their maintenance. The cost of a proton treatment, in fact, was calculated to be roughly twice than the cost of a photon treatment [52]. In recent years, however,

the development of more compact and cost-competitive solutions for proton accelerators has shown promising results [53].
Moreover, as will be discussed in more details in Sec. 3, hadron-therapy requires a delicate planning phase and preliminary imaging procedures, as the effectiveness of the process is affected by the tuning precision of the beam energy and aim. X-ray radiotherapy, producing a more broad and diffused energy deposition, is less susceptible to planning imprecisions.
Finally, the advantages of hadron-therapy over radiotherapy are fully exploited only in a fraction of the tumors; in the majority of the cases, in fact, when tumors develop in less delicate areas or are more superficial, the intrinsic advantages in terms of precision offered by the ion therapy is still overshadowed by the cost effectiveness of the X-ray therapy [9].

3 Proton Computed Tomography

3.1 The medical reasons

The effectiveness of proton therapy for tumor treatment is intrinsically related to the accuracy of the treatment planning, which is based on the information on the Stopping Power (SP) of the biological tissues. In fact, to precisely match the depth of the Bragg peak with the position of the targeted area, the exact tissue distributions and their Stopping Powers, along the entire particle path within the body, have to be known with precision, in order to tune the beam energy accordingly [54]. Currently, the SP measurements are obtained with conventional X-ray Computed Tomography (CT). An X-ray tomographer consists of a rotating structure including an X-ray beam generator and a position-sensitive photon detector, usually scintillator-based. The photon beam invests the patient, placed at the centre of the apparatus, while the detectors record the intensity of the transmitted beam; a 2-dimensional absorption map is therefore obtained for each imaging angle. The scanning proceeds with the generator-detector pair spiraling along the rotation axis, to cover the patient's volume of interest [55]. The series of absorption maps are then processed with a reconstruction software, which from line integrals produces a 3-dimensional map of the local absorption coefficient of the tissue; usually the scanned area is divided into sub-millimeter cubic boxes (*voxels*) in which the material is considered homogeneous. The local absorption coefficients are expressed as Hounsfield Units (HU, also known as *CT numbers*), defined as:

$$HU = 1000 \cdot \left(\frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \right) \quad , \tag{20}$$

where μ is the linear attenuation coefficient of the given material, μ_{water} and μ_{air} are the attenuation coefficients of distilled water and air at standard pressure and temperature; since the attenuation coefficient of air is nearly zero, a variation of one Hounsfield Unit corresponds to a change of 0.1% of the water attenuation coefficient [56].
For proton therapy, the HU values have to be converted into Stopping Power values, using an empirical calibration curve, as the one presented in Fig. 16; therefore, the quality of the curve directly affects the accuracy of the treatment [54]. The accuracy of the calibration curve is affected by various drawbacks. The curve is outlined from a series of empirical measurements of the HU-SP correlation in different samples, and the intervals between the points are interpolated; the result of the HU measurements, however, depend on the X-ray spectrum at the point of interest, which is affected by both the source type and the materials traversed by the photons up to the measurement depth. This latter effect is also referred to as *beam hardening*, which leads to the fact that the HU-SP calibration curve has to be traced for each specific CT-scanner and should also contain a dependence on, at least, the dimensions of the scanned object [54].
Moreover, the linear attenuation coefficient for photons depends on the atomic number Z, while the proton Stopping Power depends on Z/A (see the Bethe-Bloch Eq. (6)), causing a non unique-correspondence between HU and SP. In fact, the linear attenuation coefficient for a polyenergetic X-ray beam with a peak energy less than 1.02 MeV in a composite material can be estimated as:

$$\langle \mu_x \rangle = \rho_e \left(K^{ph} \tilde{Z}^{3.62} + K^{coh} \hat{Z}^{1.86} + K^{KN} \right) \quad , \tag{21}$$

where ρ_e is the material electron density, K^{ph}, K^{coh}, and K^{KN} are the constants for the photoelectric interaction, coherent scattering and Compton scattering respectively; $\tilde{Z}$ and $\hat{Z}$ are two slightly differently defined effective atomic numbers [57]. The K^{ph}, K^{coh}, and K^{KN} coefficients are specific for each CT scanner. As visible in Fig. 16, the HU curve covers a large variety of tissues, and different interpolation lines are present, especially where tissues of different nature (e.g. muscular, bone, or adipose tissues) meet, as can be noted in the enlargement box of the figure. All these complications lead to an accuracy of the proton range prediction of about

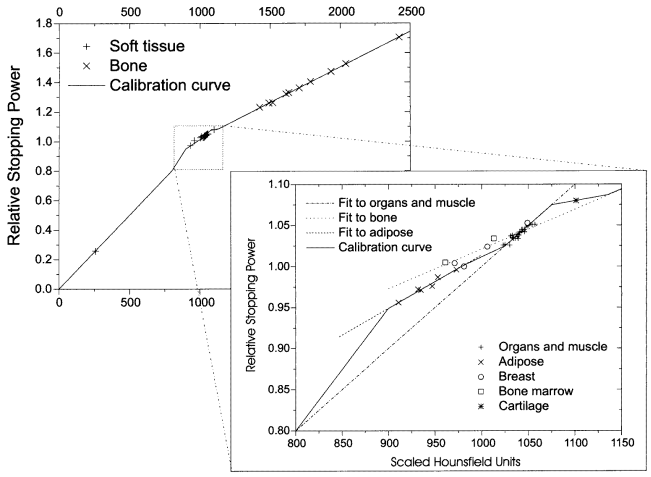


Fig. 16: Calculated (HU scaled, SP) pairs for biological tissues grouped into soft tissue and bone (behind). In the enlarged graph the soft tissue group is again split into five categories. The different lines show the three linear fits to the calculated soft tissue, adipose and bone (HU, SP) pairs and the chosen calibration curve [54].

1.1% for soft tissues and 1.8% for bone, which converts to a precision of 1-3 mm in most applications [54]; however range uncertainties in certain regions of the head can be as large as 15 mm [56] [58].
Different approaches have been explored in order to overcome these limitations. One of the most promising ones is the Dual-Energy CT (DECT), which operates similarly to regular X-ray CT, but it uses two photon sources with two different spectra [59]. It has been shown that DECT has the potential of lowering the range prediction accuracy by up to a half percentage point [60] [61].
Despite all the efforts to improve the proton Stopping Power values estimation from the X-ray CT, the ideal way to obtain those values is to bypass the HU-RSP conversion and directly measure them from proton energy loss measurements [54]. The procedure consists in using protons for imaging purposes is known as *proton Computed Tomography*, or pCT. Proton CT has the potential of obtaining range predictions with a lower and more consistent uncertainty than the currently used X-ray CT, closer to the physical limit of $\sim$1% given by range straggling [58]. Studies have found that the accuracy of the RSP values can be potentially a factor 2.5 better for pCT respect to X-ray CT [62].
Moreover, numerous studies proved the advantage of pCT over X-rays CT, in terms of dose delivered to the patient, up to a factor 100 lower for a scan of the same area [58] [62] [63] [64].
Nowadays several proton Computed Tomography scanner prototypes are being studied around the world, mainly in the USA, Europe and Japan. The vast majority of the current setups have converged to a configuration similar to the scheme presented in Fig. 17: a proton tracking system, to measure the positions and directions of the individual particles entering and exiting from the object of the imaging, coupled with a residual energy detector [65]. Every tomographer foresees the possibility of rotating either the entire scanner around the object or the object around its vertical axis.
The sensors employed for the tracking system are, in most cases, based on Silicon Strip Detectors (SSD) technologies, as in [66]–[69], or also $x-y$ Scintillating Fiber (Sci-Fi) planes [70]. For the residual energy detector, scintillator-based systems are employed almost exclusively, due to their lower cost with respect to other detector designs, intrinsic fast timing response, and possibility to cover large areas; an exception is represented by the PRaVDA Collaboration, which has opted for a solid-state residual energy detector [68] and will be presented more in

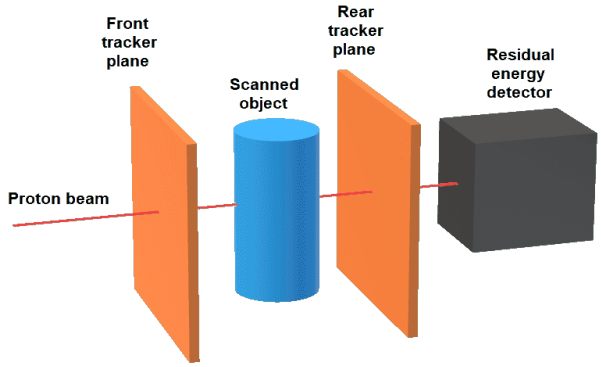
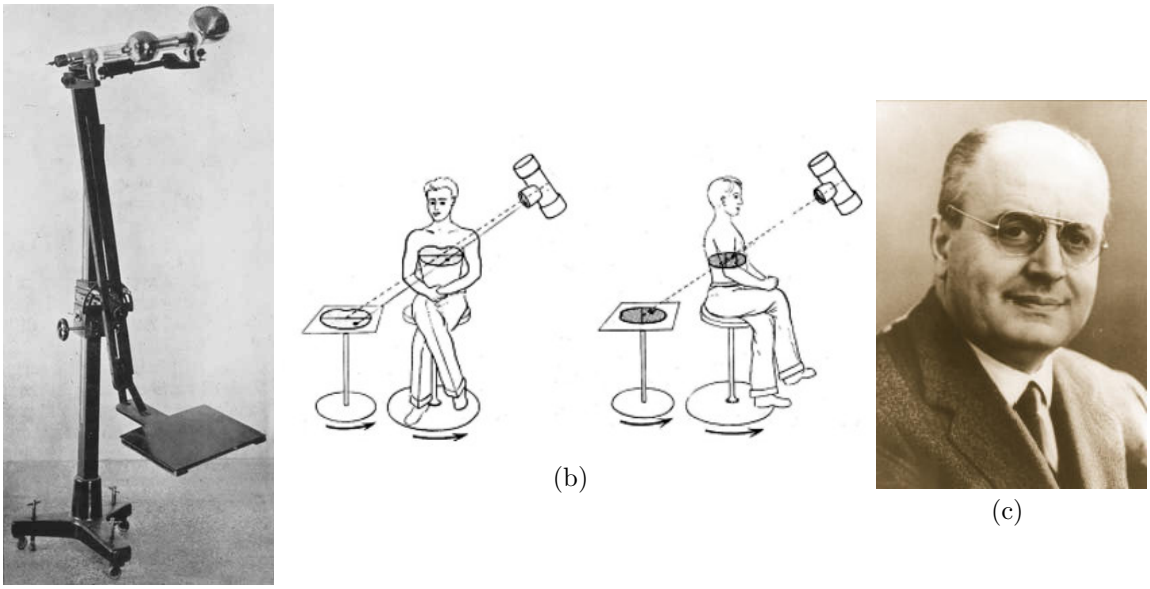


Fig. 17: Schematic representation of typical pCT scanner layout.

detail in Sec. 3.3.3. In many cases the energy detector is operated as a Range Telescope.
The 3D image reconstruction is performed in two main steps: first, the most probable path of each proton inside the scanned object has to be calculated, starting from the recorded hits on the tracker [71] (an implementation of this procedure is presented in Sec. 7); then an iterative Algebraic Reconstruction Technique (ART) is usually employed to combine the path and energy-loss informations from the total number of protons into a progressively more refined image [64].
In the following section, a brief historical digression on the Computed Tomography technique is given, with a focus on pCT, mentioning some of the crucial efforts and the influential figures that characterized the evolution of the method. In the paragraphs 3.3.1-3.3.3 the state-of-the-art of the pCT will be presented, with the description of some of the currently most promising pCT prototypes, with an overview of their concept and performances.
Proton tomography can also be effective as an imaging technique proceeding carbon-based therapy, as the RSP values can be easily scaled to account for the carbon nuclear charge from the Bethe-Bloch formula (Eq. (6)). On the contrary, carbon tomography is not viable: with respect to protons, carbon causes more damage to healthy tissue and bring more secondary dose due to nuclear interactions (fragmentation and neutrons).

3.2 History of proton Tomography

The first ancestor of the tomographic technique can be traced back to the so-called *Focal Plane Tomography*, or *Stratigraphy* [72], presented at the Genova Hospitals Congress (*Congresso Sanitario degli Ospedali di Genova*) in 1930 by Alessandro Vallebona, an Italian radiologist and professor [73]. Vallebona developed the Stratigraphy method in order to dissociate the shadows of surrounding tissues from the areas of interest of a radiography. In Fig. 18 (a) the Vallebona Stratigraphy apparatus is shown: an X-ray tube and a radiography plate are situated at the opposing ends of a rod, which is able to rotate around a fulcrum, while the area to scan is placed at the centre of the system. After a small number of different projections, on the plate a sharp image of the area around the rotation axis would be generated, while further tissues would appear increasingly blurred and spread. The operating principle of the Vallebona apparatus is schematized in Fig. 18 (b).
Closer to the modern notion of X-ray tomography, William H. Oldendorf obtained a USA patent for a "radiant energy apparatus for investigating selected areas of interior objects obscured by dense material" in 1963 [75] [76], while the first commercially viable Computed Tomography scanner, also known as the EMI CT 1000 scanner, was developed by Godfrey Hounsfield in 1967 [75]. Electric and Musical Industries (EMI) was a large business group and record label



(a) (b) (c)

Fig. 18: (a) Vallebona Stratigraphy apparatus. The X-ray tube and the radiography plate are placed at the opposing ends of a rod; a rotating movement could be impressed to either the object or the tube-plate structure. (b) Operating principle of the Stratigraphy. (c) Dr. Alessandro Vallebona (1899-1987) [74] [73].

conglomerate in the music industry, known for having signed the Beatles[1] and also conducting research in electrical instrumentations, and computers. The Hounsfield scanner was installed at the Atkinson Morley Hospital in Wimbledon, England, and the first patient brain-scan was done on 1 October 1971 [75]. The images were computer-processed, using algebraic reconstruction techniques developed by Allan M. Cormack in 1961 [78]. These images had, relative to modern standards, an extremely low resolution, with only 80×80 pixels, as visible in Fig. 19 [75]. Both Cormack and Hounsfield received the Nobel Prize in Medicine in 1979, "for the development of computer assisted tomography" [79].

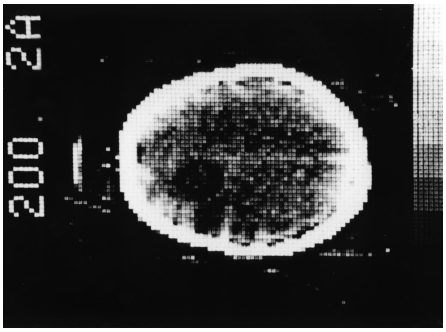


Fig. 19: First Computed Tomography on a human patient, performed at the Atkinson Morley Hospital in 1971, by G. Hounsfield with an EMI CT 1000 scanner. A cyst is visible in the right frontal lobe of the brain. The diagnosis was confirmed by the subsequent surgical findings [80].

The proposed usage of protons for imaging purposes can be traced back to the 60s, when Allan Cormack, in 1963, suggested exploiting charged particles to obtain density map measurements

[1] It is reported that Paul McCartney personally funded the Hounsfield CT project [77]

in biological tissues. In fact, the relatively high variability of density, even in tissues of the same nature, made it difficult for the radiotherapists to precisely calibrate the Bragg peak depth for the recently born proton therapy [78] [81]. In 1968 A. M. Koehler produced a radiography of an object, using 137 MeV protons at the Harvard University Cyclotron Laboratory, noting a higher image contrast but worse spatial resolution with respect to photons [82]. M. Goiten of the Lawrence Berkeley Laboratory, USA, in 1972 presented the first 3-dimensional tomographic reconstruction obtained from charged particles with an iterative algorithm; in particular a 840 MeV alpha beam was used on a plastic phantom containing inserts of different materials [83]. In the same article he recognized the potentiality of the technique for measuring 3-dimensional stopping power distributions as a direct support for ion therapy. Four years later Cormack and Koehler investigated using protons to measure fine density variations in tissues, and were able, with 158 MeV protons, to sharply distinguish density differences as low as 0,5%, even with an admittedly rudimentary equipment; the Cormack and Koehler apparatus, operating at the Harvard cyclotron, consisted of a plastic phantom containing sugar solutions and two sodium iodine scintillator readout by photomultiplier tubes [84].
Extensive studies on different aspects of proton tomography were conducted between 1977 and 1982, by K. N. Hanson at the Los Alamos Meson Physics Facility (LAMPF) in New Mexico, USA. Phantoms with different inserts were scanned using 192 and 204 MeV protons, a multi-wire chamber as tracker and a HPGe detector for calorimetry; the images were reconstructed employing a filtered back-projection algorithm that considered proton paths as straight lines [63] [85]. The same phantoms were also scanned using an EMI 5005 X-ray CT scanner; the resulting X-ray and proton images are presented in Fig. 20 (a) and (b) respectively.

(a) (b)

Fig. 20: (a) 320 × 320 pixels display from the EMI 5005 X-ray CT scanner. (b) Proton CT reconstructions of a 30 cm phantom performed by Hanson in 1979 at Los Alamos Meson Physics Facility (LAMPF). Images from [63]. Note the higher contrast for protons but the better X-ray resolution.

Hanson reported the substantial advantage in terms of deposited dose for the proton CT with respect to X-ray CT in order to obtain the presented scans, with the latter offering however a better spatial resolution and image sharpness. Moreover, he recognized that for the realization of a competitive and commercially viable pCT scanner the technology of the time was not yet sufficient, considering also the costs and impracticability for the production of a proton beam [63]. The same group a few years later performed, with an apparatus including a range telescope, the first proton tomography on human organs, in particular a human heart and brain from a recently deceased person; comparing the results with images obtained with commercial CT scanners, the proton tomography showed less visible reconstruction artifacts [86] [87].
In 1994, U. Schneider and E. Pedroni, at Paul Sherrer Institute (PSI, Switzerland), showed that proton radiography could represent a useful tool to use in proton therapy, in conjunction with regular CT, to overcome the limitation of the latter in recognizing tissues inhomogeneities; the proton range prediction improvement was estimated to be better than a factor 2 when including proton measurements [88]. Exactly ten years later, a larger group from the same institute, performed the first proton radiography on a live patient: a dog that was being treated with proton therapy for a nasal tumor. The dose that the dog received during exposure was reported

to be approximately a factor 50–100 smaller than for a comparable X-ray image [62]. Finally, the first cone beam proton CT was presented in 2000, operating with 160 MeV at the Harvard cyclotron; the scanner included a Gd_2O_2S:Tb screen, which produced a light spark with intensity proportional to the proton energy, and recorded by a CCD camera. The Stopping Power measured with protons were found in better agreement with the expected values than the ones obtained with conventional X-ray CT, once the effects of MCS blur was taken into account [89].

3.3 Proton Tomography: the state-of-the-art

3.3.1 UCSC-LLU-NIU Phase II

In 2003 a collaboration, aimed at developing a proton tomography scanner prototype, was created, with the joint efforts of the University of California Santa Cruz (UCSC), the Loma Linda University (LLU) and the Northern Illinois University (NIU) [65]; the first prototype was built and tested in 2007 [90]. In 2012 an upgrade of the original apparatus, labeled Phase II, was assembled and tested. As presented in Fig. 21, Phase II includes a tracking section and a scintillator-based energy detector.

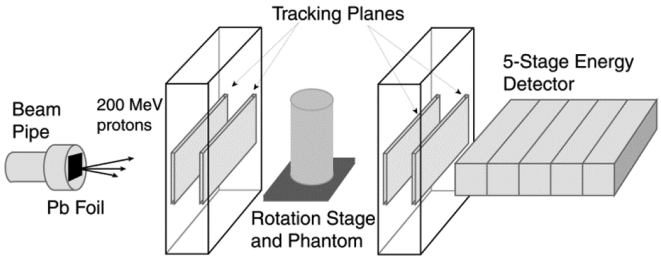
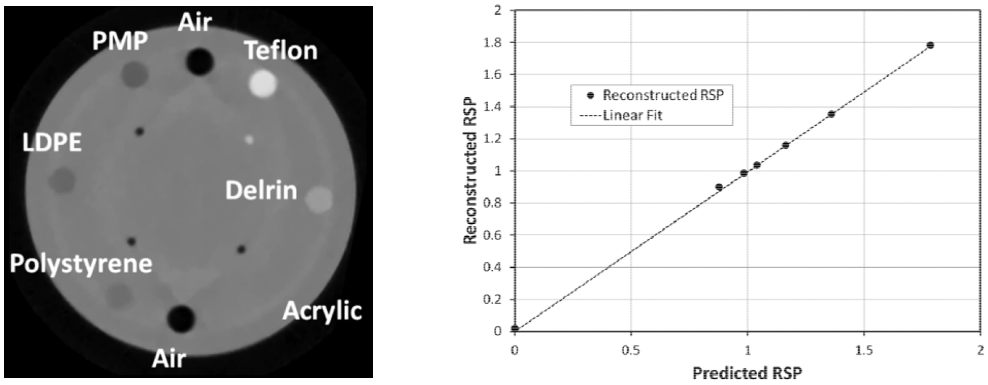


Fig. 21: A schematic representation of the Phase II pCT scanner, showing the tracking system and the five blocks calorimeter. Image from [66].

The tracking detector of Phase II employs silicon strip detectors; in particular, the scanner uses single-sided sensors produced for the Fermi-LAT telescope [91]. The strip sensor is 400 µm thick, with a 228 µm pitch; additionally, the silicon sensors are required to be enclosed in a light-tight case, equipped with 50 µm thick aluminum foils covering the particles entry and exit windows [66].
The residual energy calorimeter consists in a UPS-923A polystyrene-based scintillator volume, divided along the depth into five blocks, each 5.1 cm thick and read out by individual photomultiplier tubes. The calorimeter is operated as a hybrid energy-range detector, where the total range of the particle is obtained by summing the depth of the completely traversed blocks plus the residual range in the last traversed block, calculated from an energy measurement. The calorimeter can distinguish one proton at the time, with an upper limit of the acquisition rate was measured at 1.2 MHz [66].
Tests with 200 MeV protons carried out at Loma Linda University Medical Center (LLUMC). Fig. 22 (a) shows and example of image reconstruction performed with the Phase II scanner, the object is a CatPhan 404 phantom; the image is obtained with $1.5 \cdot 10^6$ protons per each of the 90 rotation angles. The phantom contains a number of biological equivalent samples and the obtained RSP values are compared with expected values retrieved from Monte Carlo simulations, as shown in Fig. 22 (b). The linear fit of the points in the graph indicates that RSP measurements have discrepancies from the predicted values on average smaller than 1%.

3.3.2 The PRIMA Collaboration

The PRIMA (PRoton IMAging) collaboration is an all-italian project carried out in the INFN Commission CSN5 (Commissione Scientifica Nazionale 5) and additionally funded by MIUR



(a) (b)

Fig. 22: (a) An example of image reconstructed with data from the Phase II scanner; the scanned object is a CatPhan 404 phantom: a plastic cylinder with several cylindrical inserts of differing biological equivalent materials. (b) The measured RSP of the phantom inserts, correlated with the predicted values from Monte Carlo. The dashed line is a linear fit passing through the origin, with a slope of 0.994 ± 0.009 [66].

(Italian Ministry of Instruction, University, and Research) PRIN 2006. The PRIMA Collaboration includes personnel from the Universities of Florence, Catania and Sassari, INFN sections of Firenze, Catania, Cagliari, and Catania National Laboratories (INFN-LNS) [92].

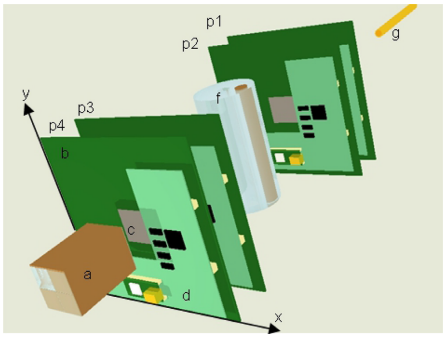
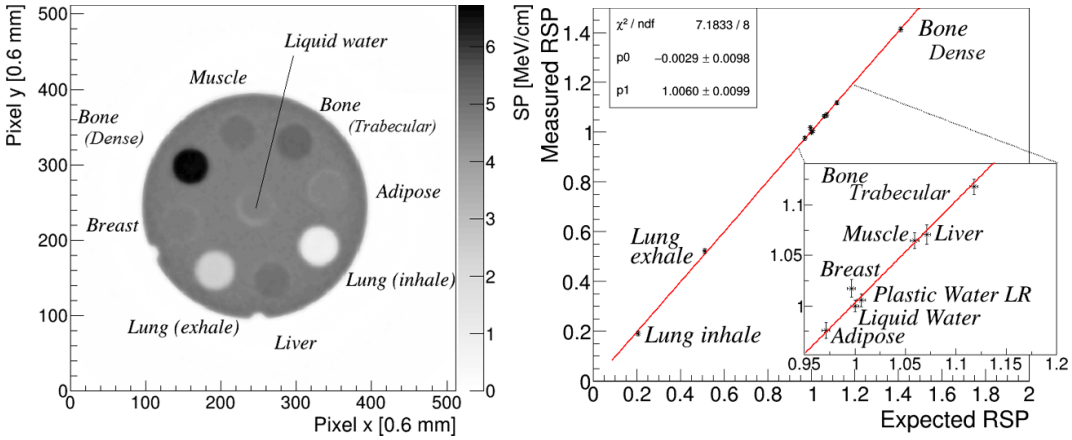


Fig. 23: Schematic representation of the PRIMA pCT apparatus, including the scintillator-based calorimeter (a), tracker front-end boards (b), silicon microstrip sensor (c), tracker digital boards (d), the complete 4 tracking planes (p1-p4), a phantom (f), the beam pipe (g). Image from PRIMA Collaboration [67].

The PRIMA apparatus includes a silicon microstrip-based tracking section and a segmented YAG:Ce calorimeter, for residual energy measurements [93]. In Fig. 23 a rendering of the apparatus is shown. The tracker is composed of four $x-y$ planes, that exploit the single-sided silicon microstrip technology by Hamamatsu, each plane including 256 strips with 200 µm pitch [67]. The thickness of a single direction strip is 200 µm [93]. The surface covered by the tracker planes has an area of 5.1×5.1 cm^2.
The calorimeter, in the first iteration of the PRIMA scanner, included four scintillating volumes, each with a 3×3 cm^2 cross section and 10 cm depth, arranged as a 2×2 matrix. The scintillators are made of YAG:Ce (Yttrium Aluminum Garnet doped by Cerium), characterized by a decay constant of 70 ns, and read-out by Hamamatsu silicon photodiodes [94]; given the decay time of the scintillators, the calorimeter is able to sustain a 1 MHz particle rate [93]. The depth of the calorimeter is tuned to completely absorb 230 MeV protons [67].
An upgrade of the initial PRIMA apparatus includes a larger field-of-view for both the tracker,

up to 5.1×20 cm^2, and consequently the calorimeter, 6×21 cm^2 [93].
The scanner has been tested with low energy protons (62MeV) at LNS-INFN in Catania, and at Svedberg Lab in Uppsala, Sweden, with higher energies (180MeV) [94] [67]. Using radiographic images of custom PMMA phantoms, the spatial resolution of the apparatus has been estimated to be below 300 µm [95]. The energy resolution of the PRIMA calorimeter has been measured to be around 2.6 MeV in the range 70–211 MeV. In recent test beams the system has been proven to reconstruct the correct tissues RSP with a maximum deviation of 1.6% between the measured and expected values [96]. The peak sustainable data rate is 80 kHz [96].



(a) (b)

Fig. 24: (a) The central 1.65 mm thick slice of the reconstructed image of a phantom tomography using 180 MeV protons. The grey levels are proportional to the absolute SP. (b) Comparison between measured RSP values expected values from a Geant4 simulation of the inserts materials. The red line represent a linear fit of the data. In the box the region close to RSP = 1 is shown. Images from PRIMA Collaboration [96]

3.3.3 PRaVDA

The PRaVDA consortium, constituted in 2013, takes a peculiar approach to the pCT challenge, developing a system completely based on solid-state detectors, instead of the common configurations that include a scintillating calorimeter. The PRaVDA detector employs silicon microstrip technologies developed for the ATLAS experiment at the High Luminosity Large Hadron Collider (HL-LHC, CERN) [68] by University of Liverpool. The silicon sensors are composed of a total of 2048 microstrips, with 90.8 µm pitch and a 150 µm thickness, for a covered area of 93×96 mm^2. The sensors can be read-out with a 26 MHz frequency, for up to 8 strips in parallel.

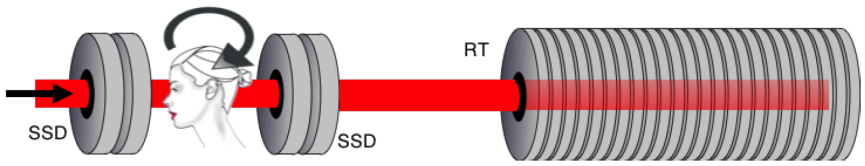


Fig. 25: A representation of the PRaVDA pCT system design. Two sets of three microstrip detectors are placed before and after the patient, in addition to a Range Telescope (RT), formed by interspaced microstrip planes and layers of passive PMMA absorber. Image from [97].

The PRaVDA tracker includes two stations formed each by three planes of strip detectors, ar-

ranged in a $x-u-v$ configuration with a 60° rotation between each plane, and a sensible area of 8.5×8.5 cm^2. This configuration was adopted to allow higher particle rates and to reduce the number of ambiguous hits in correlating coordinates between planes [68]. In addition to the tracking system, the apparatus includes a Range Telescope, based on the same sensor design. The PRaVDA telescope is formed by 21 microstrip planes interleaved with 2 mm passive PMMA absorber sheets, equivalent to 2.6 mm of water; the proton range is estimated as the depth of the last hit silicon plane, with an uncertainty evaluated equal to 1.3% of the beam range. The apparatus has been tested with 125 MeV protons at iThemba LABS, South Africa; tests reported a RSP accuracy between 0.7 and 1.6% for different biological equivalent tissues, including water, adipose and bone tissues [68]. The PRaVDA range telescope is designed to sustain acquisition rates of the order of the MHz [98].

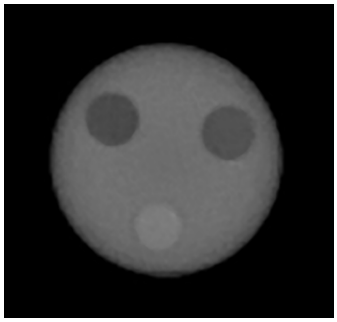


Fig. 26: A slice of the tomographic image of a spherical phantom containing inserts of three different materials (top left: water equivalent, top right: adipose equivalent, bottom: average bone equivalent). Image from PRaVDA Collaboration [68].

3.4 Present limitations of proton Tomography

Despite the long history, from its proposal in 1963 to the present days, and the numerous efforts and promising prototypes, the proton tomography procedure is still not clinically viable, as opposed to the widely established X-ray CT technique, due to physical, technological and accessibility limitations.

The main physical drawback of pCT with respect to X-ray CT is given by the Multiple Coulomb Scattering that protons undergo when traveling inside matter, while photons travel basically in a straight line [99]. In a pCT image reconstruction, considering the proton path as linear leads to a clinically insufficient spatial resolution; therefore, a most likely path calculation has to be performed for each proton [99]. Employing proton path reconstruction algorithms, a spatial resolution as low as 100-150 µm can be achieved, which is sufficient for clinical uses. Moreover, the proton spatial resolution is not uniform, as the reconstruction is expected to be less accurate near the centre of the scanned object [71]; the resolution is also dependent on the beam energy: higher energy protons have a lower uncertainty; however, the maximum energy of the beam in proton therapy facilities is generally limited to 250-300 MeV, as higher energies are not necessary for the treatment [47]. Given the non-linear path, proton hits in the upstream section of the tracker have to be associated to hits in the downstream section, a procedure that leads to a non-perfect reconstruction efficiency, depending on the number of simultaneous tracks recorded. Therefore, it is necessary to record more particles, to take into account the efficiency loss introduced by the reconstruction. Moreover, to achieve a lower spatial uncertainty, it is suggested to reject events with large scattering angles, thus further diminishing the fraction of acceptable particles [71]. Finally, although path reconstruction techniques can be expected to improve in the future, the spatial resolution achievable using protons will be limited by their physical behaviour, intrinsically worse than photons [99] [100]. An application of a proton path

reconstruction method, with an evaluation on the calculation error, is presented in Sec. 7.
Another physical limitation is represented by nuclear interactions; as introduced earlier, protons in the 200-250 MeV can interact with the nuclei of the medium, producing in most cases neutrons, protons, or gamma rays. These types of events are of no interest for proton tomography, as only ionizing energy loss to electrons has to be considered, and have to be discarded from the analysis. On the other hand, they are a nuisance as they can generate additional signals in the detectors, either the tracker or the calorimeter, making the track recognition procedure more difficult. Finally, secondary particles from nuclear interactions constitute an additional source of dose for the patient; nevertheless, as discussed in Sec. 3.1, the expected dose for a pCT is still at least an order of magnitude lower than a X-ray CT.
At the present state another considerable limitation of the proton CT is the slow acquisition rate of the current prototypes, as even the fastest scanners nowadays can reach a sustainable particle rate of the MHz order [98]. In proton CT, in fact, each proton has to be recorded individually, which is a particularly challenging task to perform at high speeds. On the contrary, in X-ray CT, this requirement does not exist, as photons are not detected individually, and the transmitted beam intensity is measured instead. To perform an satisfactory pCT scan of an object with the dimensions of a typical human head, at least 10^9 acceptable and completely reconstructed protons events are needed, net of the inefficiencies mentioned above [101]. Even with the fastest available prototype, a complete pCT scan would take several minutes, which is not clinically acceptable. In fact, with such long scan durations, the patient's motion, especially due to breathing, will negatively affect the quality of the image. Shorter scanning times, in the order of 10 s or less, would allow the patient to hold their breath for the duration of the procedure, a measure already applied in certain instances for X-ray CT [102]. Therefore, the goal acquisition rate for a viable pCT scanner would be of the order of hundreds of MHz.
Current state-of-the-art proton tomography scanner, to cope with the rate and precision requirements, are inevitably forced to employ frontier technologies in the field of particle detectors and engineering; such necessity usually comes with a high price tag. It is also worth noting that in the future, once a scanner design would have successfully completed the prototyping phase, the subsequent steps toward a commercialisation of the apparatus and an industrialization of the manufacturing process, would likely lead to a cut of the costs.
Finally, as discussed in Sec. 2.3.1, the proton CT is intended to be used in conjunction to proton therapy, which, despite its fast growth over the last decades, is still not as widespread as the more common radiation therapy with X and gamma rays: at the moment only about 100 hadron-therapy facilities are in function worldwide, compared with more than 7600 centres administering photon radiotherapy [8] [50]. In the future, a greater accessibility of the proton and hadron therapy solutions would most likely induce a growth in demand for proton tomography.

4 The iMPACT project

4.1 The concept

The *innovative Medical Proton Achromatic Calorimeter and Tracker* (iMPACT) is a project funded by the European Research Council (ERC), through a Consolidator Grant [103], hosted by University of Padova and supported by the INFN Section of Padova. The goal of the project is to develop a fast and high resolution proton Computed Tomography scanner, capable of obtaining a range prediction resolution near the $\sim 1\%$ theoretical limit; it represents a competitive solution in terms of construction, instrumentation and running costs. One of the main targets of the project is to reach an acquisition rate as close as possible to 10^9 recorded proton events in about 10 s, the minimum requirement for clinical viability [101]; the acquisition rate constitutes, in fact, one of the principal reasons why the pCT method remains nowadays still non suitable for actual medical applications (as discussed in Sec. 3.4). The ultimate objective of the iMPACT prototype is to demonstrate that the pCT technique can be actually used in real clinical applications, even from an economical point of view. In order to achieve these challenging goals, iMPACT is designed to exploit technologies developed and currently in use for particle physics detectors.

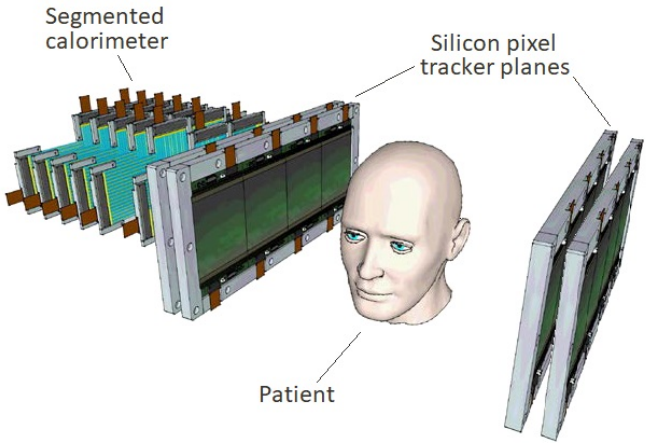
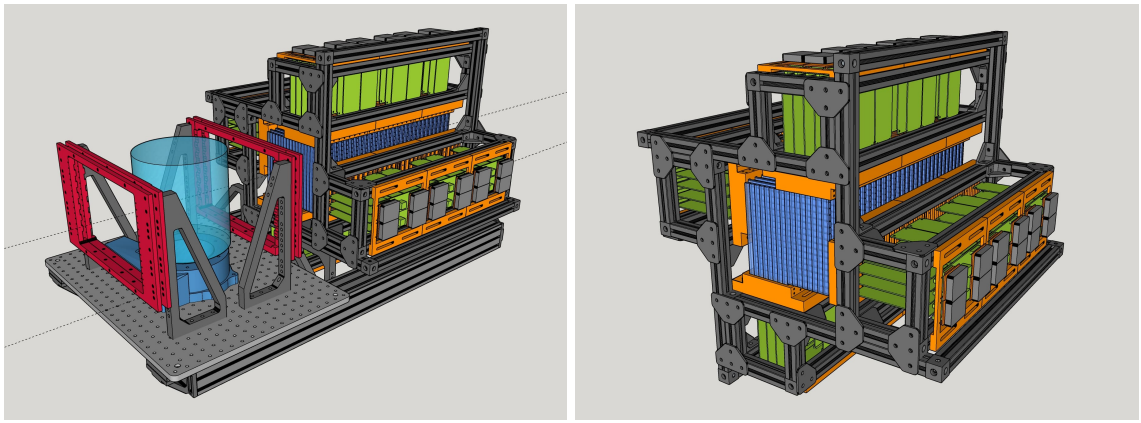


Fig. 27: Rendering of the iMPACT project scanner layout.

A schematic layout of the iMPACT scanner is presented in Fig. 27; the proton Tomography apparatus shares the basic structure of the majority of the current pCT prototypes, presented in Sec. 3.3, including 4 silicon tracking planes grouped in two stations, and a plastic scintillator calorimeter. An element that differentiates iMPACT from most of the other prototypes is the use of Monolithic Active Pixel Sensors (MAPS) for its tracking section, when the majority of the other projects employs silicon strip sensors; the tracking section is presented in detail in Sec. 4.3.1. Moreover, as for most of the other prototypes, the iMPACT calorimeter is scintillator-based; however, it is designed with a high level of segmentation, in both depth and transversely, and is operated as an advanced version of a range telescope, using deposited energy information to improve the proton stopping point measurement. The calorimeter system design will be treated in more details in Sec. 4.2.
The iMPACT project was proposed in 2015 [104] [105], was approved by ERC in 2016, within the call Horizon 2020, and received a Consolidator Grant (agreement n. 649031) of about 1.8 million euro. A team formed by an associate professor, two Ph.D. students and a technician, and led by Prof. P. Giubilato is currently involved in the project, while INFN personnel is active in hardware development tasks. The INFN Section of Padova is also responsible for the purchase of the silicon Sensor employed for the tracking system, from the ALICE Collaboration

at CERN, which accounts for a sensible portion of the budget. Members of the iMPACT project also participated in the efforts of the ALICE Collaboration, for the development and commissioning of the sensors. Additionally, iMPACT members participate in a national collaboration between different INFN Sections (ARCADIA), with the goal of developing and producing an innovative silicon Sensor, aimed at improving the performances of future versions of the iMPACT scanner, among with a large range of other scientific and industrial applications; the future Sensor design is discussed in Sec. 9.
The iMPACT duration was initially planned to last until the end of 2020, however the project was extended by an additional year, due to the international pandemic situation, which unavoidably affected the work progress. In 2017 the Italian Ministry for Instruction and Research (MIUR), through the FARE program, granted additional funding to the project. Currently the project is its final assembly phase, and during the following years intensive tests are planned. Fig. 28 (a) shows a 3-dimensional view of the final design of the complete iMPACT scanner; it includes the frames for the silicon pixel sensors (in red), a Catphan 600 biological phantom [106] (blue cylinder), and the calorimeter assembly. Fig. 28 (b) shows in detail the modular design of the calorimeter: it includes the scintillating elements (blue), electronics boards (green), and custom 3D printed structural elements (orange). The mechanical stability of the entire structure is ensured by heavy-duty Thorlabs mechanical elements (black components in the visualization) [107].



(a) (b)

Fig. 28: (a) Visual 3-dimensional representation of the final iMPACT scanner layout; the frames for the silicon pixel sensors are represented in red, the biological phantom (Catphan 600 [106]) as a blue cylinder. (b) particular of the modular design of the calorimeter: the scintillating elements are represented in blue, electronics boards are in green, and custom 3D printed structural elements in orange. Heavy-duty mechanical elements from Thorlabs [107] are visualized in black.

4.2 The iMPACT Calorimeter

One of the two main components of the iMPACT proton tomography scanner is a highly-segmented scintillator-based calorimeter. The calorimeter is designed to operate as an advanced version of a range telescope, measuring the stopping position of each particle, instead of measuring the residual energy. A range telescope design was adopted also in other pCT prototypes, such as [68] [69]; many projects instead employ energy calorimeters [90] [94]. A range telescope has the advantage of being more straightforward to calibrate, with respect to an energy calorimeter, however the high number of channels represents a challenge in the read-out process.
The iMPACT calorimeter layout is shown in Fig. 29; the sensible volume is segmented in depth

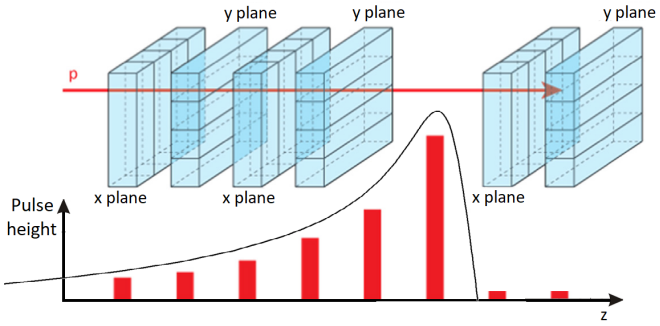
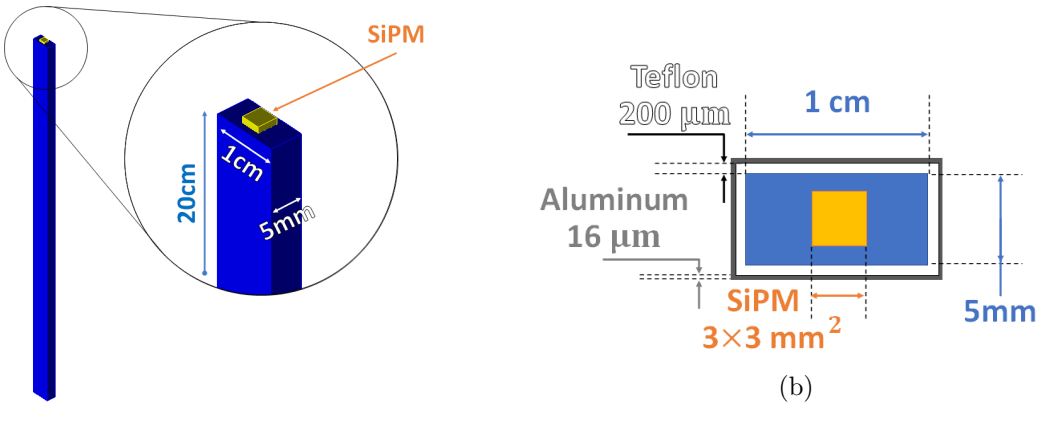


Fig. 29: Conceptual layout of the iMPACT calorimeter, based on PVT scintillating elements (*fingers*); the calorimeter volume is segmented along the depth (into 5 mm thick z-planes) and alternatively into the x and y directions. Red bars represent the expected signal amplitudes along the planes.

(z direction) into 5 mm planes; a total of 60 planes is foreseen, enough to stop a 230 MeV proton beam. Each plane is formed by a total of 20 high aspect-ratio scintillating elements (named *fingers*), each one 1 cm wide and 20 cm long. The scintillating fingers are arranged in planes aligned along the x and y directions alternatively. In the current design, the calorimeter can cover a 16×16 cm^2 surface area; however, the feasibility of the concept is currently being demonstrated with a 8×8 cm^2 surface prototype. In the future, the sensible area can be increased up to 32×32 cm^2 in a straightforward way, employing the same basic construction elements.
Each finger is singularly read-out by a Silicon Photo Multiplier (SiPM), placed on one of the finger end surfaces, as shown in Fig. 30; the SiPM technology is presented more in detail in Sec. 4.2.1.



(a) (b)

Fig. 30: (a) Visualization of a single scintillator finger, with the relative SiPM; particular of the SiPM. (b) Section of a 5×10mm^2 scintillator, with the 3×3mm^2 SiPM, the teflon (with 200 µm thickness) and aluminum (with 16 µm thickness) wrapping layers. The scintillator and SiPM are in scale, while the wrapping layers thicknesses are exaggerated for visualization purposes.

The scintillating fingers are individually wrapped in a thin, highly-reflective teflon layer, to increase the light collection efficiency, and an external layer of aluminium, to obscure the scintillator from external light and cross-talk; the wrapping materials and procedures have been optimized to reach the required performances, while minimizing the material budget. Each plane, in the direction of the particle beam, includes 5 mm PVT, two 200 µm Teflon layers and two 16 µm aluminum foils, for a total thickness of 5.432 mm; the equivalent water thickness has

been estimated to be 5.877 mm, using a Monte Carlo software[2]. The equivalent water thickness will be compared with an experimental measurement in a future test beam. The optimization procedure of the wrapping method is presented in Sec. 5.2.
The calorimeter segmentation pitch along the depth has been tuned in order to provide an accurate estimate on the proton path length. As introduced in Sec. 2.2.2, the proton range is affected by a intrinsic statistical fluctuation given by the range straggling, which for the energy of interest (230 MeV) is $\sigma_E^{\text{stragg}} \approx 3$ mm; with the chosen z-pitch, the accuracy of a depth estimation, of the order of $5\,\text{mm}/\sqrt{12} \approx 1.4$ mm, will not be the dominant contribution to the total path length uncertainty. More details on the path length estimation accuracy are presented in Sec. 4.2.3.
The segmentation in the transverse plane has the double purpose of allowing for a higher sustainable particle rate, as multiple simultaneous events recorded in different points of the calorimeter can be distinguished (up to certain multiplicities), while at the same time provide coarse information on the particle trajectories.
The calorimeter can provide proton stopping positions measurements with three different, and progressively more accurate, approaches:

- In the most straightforward use, the $x - y$ segmentation of the planes is ignored, and the proton path length is estimated as the mid-point of the last plane recording a signal. The position uncertainty in this case is given by the z-pitch (5 mm) divided by $\sqrt{12}$. In this modality, the energy information from each finger is reduced to a binary hit-miss logic, with just a fixed threshold indicating the passage of a proton in a given plane.

- From this basic mode of operation, the procedure can be refined by exploiting the transverse segmentation of each plane. In fact, the proton path can be coarsely followed inside the calorimeter volume, with a sensibility of 1 cm in both x and y coordinates. The proton path length can therefore be more accurately estimated as the sum of the single segments connecting the mid-points of the traversed fingers. This, however, is only a marginal improvement, as the difference between the the path length and its z-projection is less than 0.5% (see Sec. 11). The main improvement that the tracking capabilities of the calorimeter brings, is that protons undergoing large deviations inside the calorimeter volume, which can be an indication of inelastic nuclear scatterings, can be identified and rejected from the analysis.

- Finally, as suggested in Fig. 29, the energy signals in the individual fingers can be exploited, to outline the Bragg Peak curve shape and refine the proton stopping point depth estimation with an accuracy, which can be up to 30% better with respect to the one given by the z-pitch. The uncertainties achievable with the last methodology are discussed in Sec. 4.2.3.

4.2.1 The calorimeter hardware

For the scintillating elements, Polyvinyltoluene-based (PVT, C_9H_{10}) plastic scintillators have been chosen, due to their fast timing characteristics, with less than 1 ns raising time and good light output, more than 11000 photons/MeV. Additionally, the low density of PVT (1.032 g/cm^3) makes for less stringent constraints in terms of structural load. Two different solutions for PVT scintillators have been tested during the initial phase of the project, BC-420 and BC-408, both manufactured by Saint-Gobain Crystals [108, 109], with the second one being the preferred choice, based on its higher light yield and measured faster response (despite the tabulated values from the manufacturer, see Sec. 4.2.1 and 5.2). For the full-scale prototype construction, Eljen EJ-200 scintillators have been selected, as the EJ-200 is equivalent to the

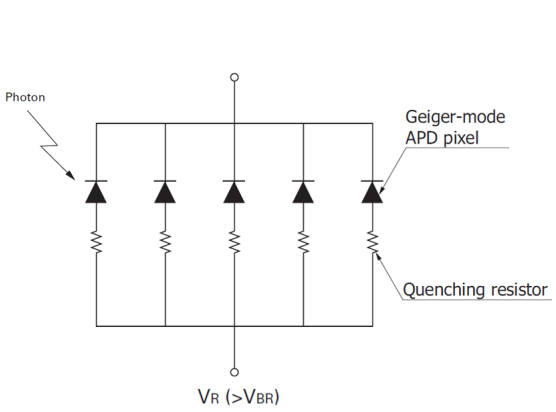
[2]The water equivalent thickness of a wrapped finger has been estimated with the SRIM Monte Carlo Software, comparing the range of a proton beam in water and in a volume formed by stacked wrapped scintillators. The water equivalent thickness of a wrapped scintillator varies less than 0.2% in the energy range 70-230 MeV.

BC-408 in terms of performance, while representing a more cost-efficient solution [110]. Tab. 1 summarizes the characteristics of the PVT scintillators considered, as declared by the respective manufacturers.

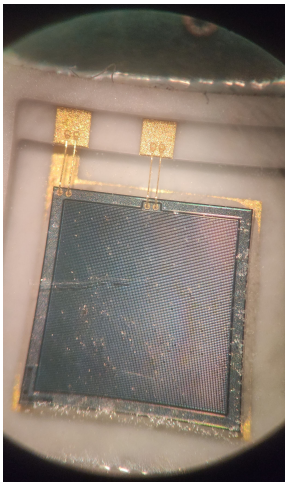
Parameter	BC-420	BC-408/EJ-200
Density (g/cm^3)	1.032	1.032
Refractive index	1.58	1.58
Light yield (% Anthracene)	64%	64%
Light yield (photons/MeV)	11136	11136
Max emission wavelength (nm)	391	425
Raise time (ns)	0.5	0.9
Decay time (ns)	1.5	2.1
Light attenuation length (cm)	110	380

Tab. 1: BC-420 and BC-408 polyvinyltoluene scintillators declared parameters. Data retrieved from Saint-Gobain data sheets [108] [109]. The Eljen Technologies EJ-200 is equivalent to the BC-408 [110].

Each scintillating finger is equipped with a Silicon Photo Multiplier (SiPM), which converts the optical photons, produced by the scintillator, into an analog electric signal. SiPM, also named Multi Pixel Photon Counter (MPPC), are solid-state devices designed for low intensity photon detection in the optical range, developed as an alternative to the Photo-multiplier Tube (PMT) technology. SiPM offer detection efficiencies up to 40%, extremely compact design with respect to PMT, a lower operating voltage (less than 100 V), time resolution up to 250 ps, and are not affected by magnetic fields [111].
The SiPM surface is composed of a matrix of independent Avalanche PhotoDiodes (APD), as represented in Fig. 31 (a), biased with a reverse voltage V_R higher than the break-down voltage of the junction (V_{BR}); in this regime, even the charge deposition of a single photon can initiate a Geiger avalanche discharge, with charge gain of the order of $10^5 - 10^6$. The current generated by a photon detection is constant and independent of the deposited charge; moreover, during the duration of the discharge, multiple photons detected in the same APD cell do not result in additional signal. Each ADP cell includes a quenching resistor in series, so the current flow generated by a discharge generates a voltage drop across the resistor and suppresses the avalanche. In Fig. 31 (b) the surface of one of the SiPMs employed in the iMPACT calorimeter is shown, through a microscope; the sensible area surface is 3×3 mm^2, the 25 µm pitch cells are visible
The current generated in each parallel ADP cell is summed; the current is generally loaded onto a resistor, across which the current signal is converted into a voltage signal. The voltage amplitude of the output signal is proportional to the number of simultaneous ADP cells discharging. The output signal amplitude is proportional to the number of detected photons as long as the incident light intensity is low enough for the probability of multiple photons hitting the same cell can be neglected.
The detection efficiency of a SiPM in given by the product of three factors: quantum efficiency, avalanche probability, and fill factor. Quantum efficiency quantifies the probability for an incident photon to release charge carriers, avalanche probability is the probability for the freed charges to generate an avalanche (it is dependent on the bias voltage), and the fill factor is the ratio between active and total area.
For the iMPACT calorimeter Hamamatsu S12572-025c SiPM have been chosen; the model presents a 3×3 mm^2 total area and 25 µm pitch APD cells. The efficiency of the SiPM model, as a function of the wavelength, is shown in Fig. 32 (a), compared with the light output spectrum of the considered scintillator models. The emission of BC-408 and EJ-200 (the final preferred choice for the calorimeter) peaks at 425 nm, which matches better with the SiPM efficiency than BC-420 (peaking at 391 nm), resulting in a higher number of detected photons.
A digitized signal measured directly from the SiPM model, without the read-out shaping electronics, is presented in Fig. 32; after the signal peak, the amplitude decreases exponentially,



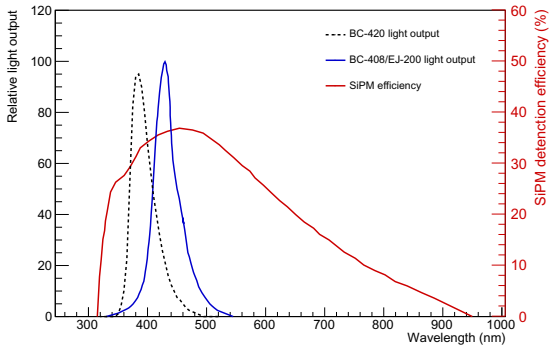
(a)



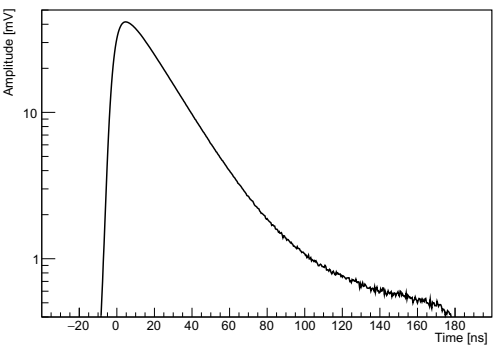
(b)

Fig. 31: (a) Configuration with multiple Avalanche Photo Diode (APD) in parallel forming a SiPM matrix; each cell include a quenching resistor. The SiPM is biased with a reverse voltage V_{R} higher than the break-down voltage of the junction (V_{BR}); adapted from [111] (b) A single 3×3 mm^2 Hamamatsu S12572-025c SiPM with 25 µm cell pitch, under a microscope.

returning below 1 mV in about 100 ns. Based on this measured quenching time, the effect of the pile-up of multiple proton events on the same scintillator has been studied.
Once an APD cell in a SiPM detects a photon, the cell is blind to additional photons for the duration of the avalanche; therefore if multiple protons are detected in a 100 ns window, the SiPM signal output progressively saturates, as an increasing number of cells is blinded. Using a Monte Carlo simulation taking into account the photons detection position on the SiPM surface, it has been estimated that at 100 MHz (i.e. 10 protons in a 100 ns window on the same scintillator), at the Bragg peak (where the photon intensity is at its highest) the photons are detected by the SiPM with an efficiency of about a 70%; at the initial part of the Bragg curve the efficiency is 95%. More details on the simulation setup are presented in Sec. 4.4, while the efficiency estimation procedure can be found in [112] and [113].



(a)



(b)

Fig. 32: (a) BC-420 [108], BC-408 [109] and EJ-200 [110] scintillators light output spectrum (primary axis), compared with Hamamatsu S12572-025c SiPM detection efficiency (secondary axis) [111]. (b) Measured average Hamamatsu S12572-025c SiPM output signal, probed before read-out electronics.

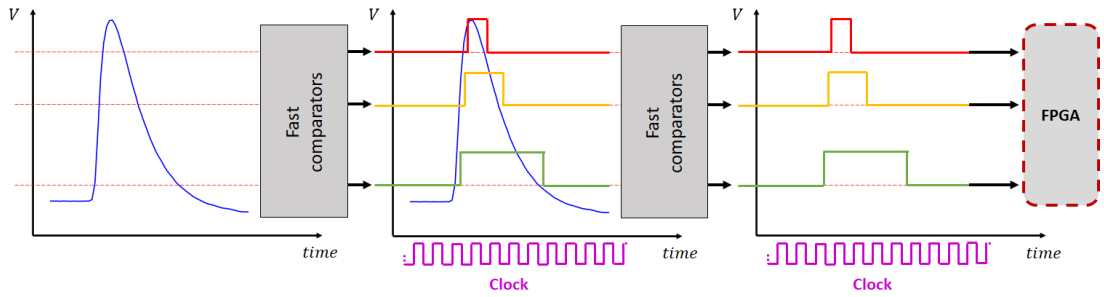


Fig. 33: Concept on a read-out based on three threshold. Each channel is equipped with fast comparators with programmable levels, which provide a digitization of the waveform (blue curve). The digital information are processed by FPGAs and stored.

The complete iMPACT calorimeter includes a large number of channels to be read-out: 480 for the 8×8 cm^2 prototype, 960 for the bigger scale 16×16 cm^2 version, and 3840 for the full-scale scanner. An analog read-out of such a large number of channels at GHz sampling frequencies would be unsustainable in terms of instrumentation costs and power consumption, as well as being highly impractical and unnecessary. For these reasons, a threshold-based digital read-out architecture has been developed, as outlined in Fig. 33. Each calorimeter channel includes three fast comparators (timing precision ≈ 1 ns), which samples the analog signal generated by the SiPM coupled to the scintillating finger; the comparators provide a digital output for the time duration in which the analog input signal exceeds a programmable threshold. The procedure is equivalent to digitizing the waveform amplitude with a 2-bit ADC. The digitized signals are sent to FPGAs for data formatting and storage. The hardware and technical details of the digital read-out architecture will be dealt with in Sec. 6.

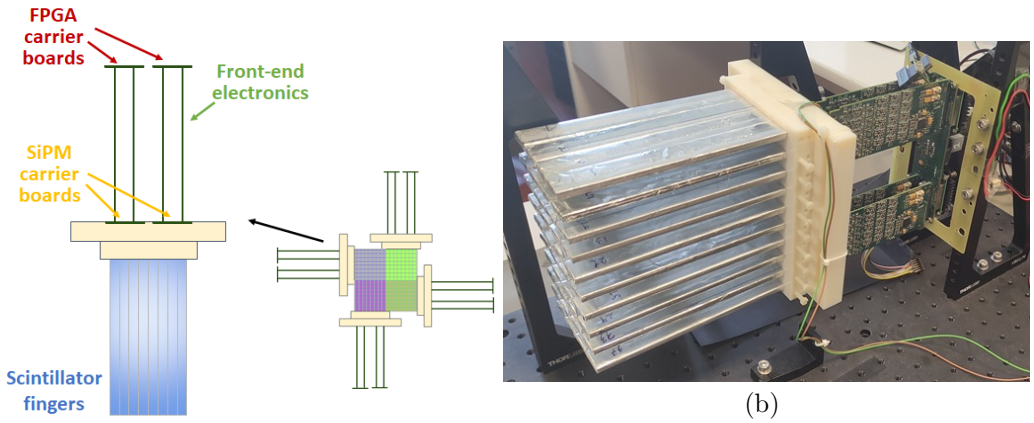


Fig. 34: (a) Single iMPACT calorimeter module, including 80 scintillator fingers arranged in 10 planes, two SiPM carrier boards, 8 front-end electronics boards, and 4 FGPAs; arrangement of four modules to form the $x - y$ structure. (b) Photo of an assembled module.

The iMPACT calorimeter has a completely modular layout, which contributes to containing the costs and ensuring scalability. The basic module is shown in Fig. 34: it includes a total of 80 scintillating fingers, arranged in 10 parallel planes. The module structure is mechanically self-sustained, and holds the SiPMs, the read-out electronics for each channel, and the FPGA for the event data formatting. The module includes: two electronic boards housing 4×10 SiPM each (visible in Fig. 35), 8 front-end boards, capable of reading out 10 SiPMs each, and two

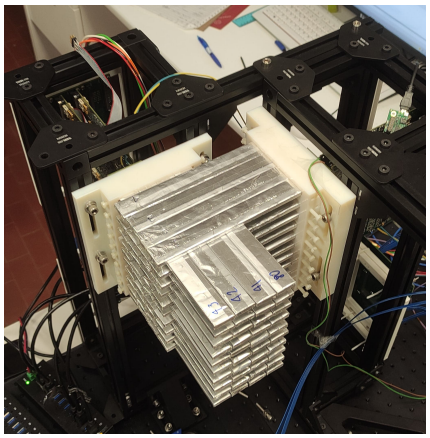
FPGA carrier boards (see Fig. 70). Each channel of the front-end boards includes an analog stage, composed by a high-pass filter and an amplifier, followed by the three digital comparators for the signal digitization; the electronic components dedicated to the digital read-out of the calorimeter are presented in detail in Sec. 6.1.
The entire sensible calorimeter volume is formed by a set of modules, arranged in orthogonal directions, as outlined in Fig. 34 (a). The required calorimeter depth of 30 cm is obtained with a stack of 6 modules. The small-scale and full-scale prototypes (with 8×8 cm^2 and 16×16 cm^2 surfaces respectively) can both be assembled with a repetition of the basic module. The future 32×32 cm^2 full scanner would share the very same mechanical and electronics components of the smaller prototypes, however arranged in a different basic module. Tab. 2 summarizes the components included in a module, as well as in the small-scale, and full-scale prototypes.

Element	Single Module	Small-scale prototype (6 Module)	Full-scale prototype (12 Modules)
Scintillators	80	480	960
SiPM carrier boards	2	12	24
SiPM	80	480	960
Front-end boars	8	48	96
FPGA boards	2	12	24
FPGA	4	24	48

Tab. 2: Number of scintillators, mechanical, and electronics elements in a calorimeter module, and in the small-scale (8×8 cm^2 surface) and full-scale (16×16 cm^2) prototypes.

Fig. 35 (a) shows the 4×10 SiPM array, housed by the SiPM carrier board, next to a full calorimeter module; Fig. 35 (b) shows two complete modules arranged perpendicularly to form the $x - y$ structure of the calorimeter.

(a)



(b)

Fig. 35: (a) Picture of the SiPM array housed by the SiPM carrier board. (b) Picture of two full calorimeter modules arranged to form an $x - y$ plane.

4.2.2 Calorimeter electronics

Fig. 36 shows a picture of a front-end board, including the analog shaping-amplifying stage and electronics for the signal digitization. A single front-end board is designed to read-out 10 scintillators channels, and it is connected on the left side to the SiPM board, and on the right side to the FPGA carrier board (the structure of a calorimeter module was shown in Fig. 34).

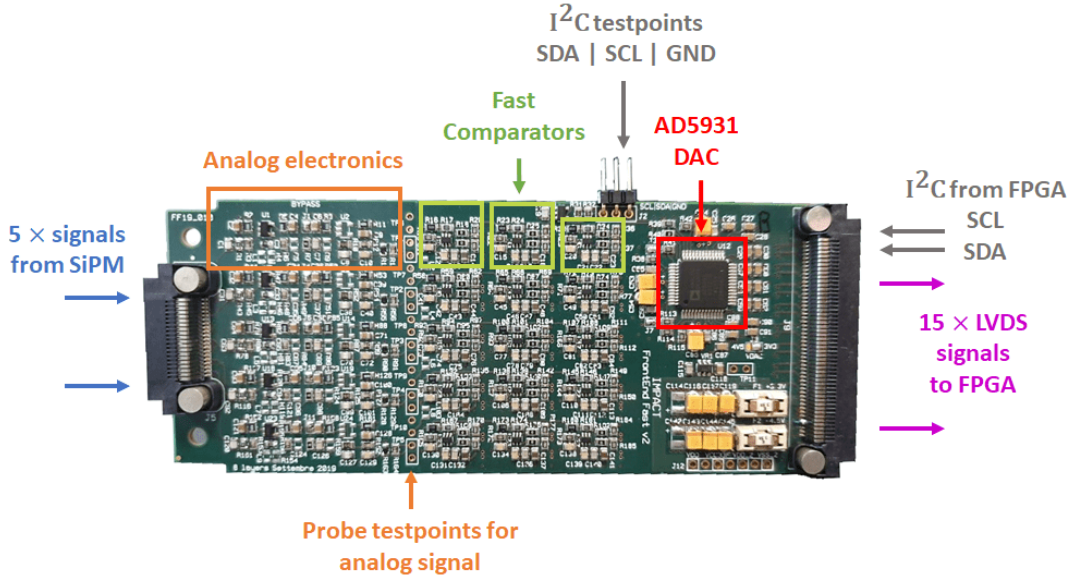
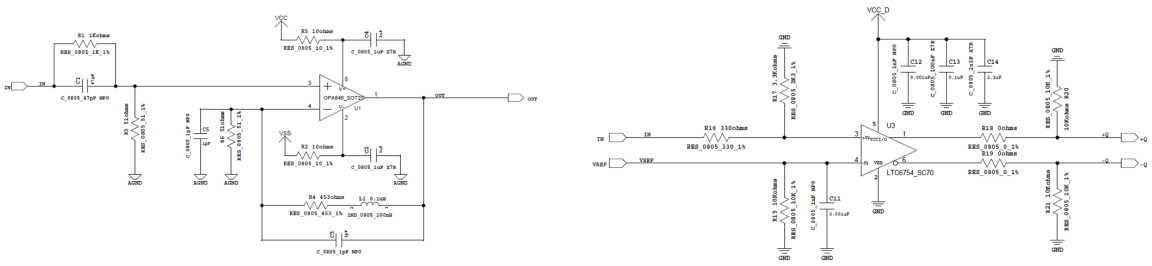


Fig. 36: Layout of the iMPACT calorimeter front-end board. The same layout is replicated on both faces of the front-end board. Each face of the board features 5 identical channels (10 in total), formed by an analog amplifying and shaping stage, and 3 Linear Technology LTC6754 fast comparators [114]. One Analog Devices AD5391 DAC [115] on each face is used to set the comparators reference voltages.

Fig. 37 shows the schematics of the electronics present in each channel. The first stage is a high-pass filter with a cut-off frequency of ≈400 MHz, which lets through the first transient of the SiPM signal, cutting the long tail. Following the high-pass filter, a non-inverting amplifying stage is present; the use of a second shaping-amplifying stage was investigated and discarded, see Sec. 5.2. Following the first tests on the calorimeter read-out, illustrated in Sec. 4.5 and



(a) (b)

Fig. 37: Schematics of the main components of iMPACT Calorimeter front-end electronics: a) analog amplifying and shaping stage; b) Single voltage comparator stage featuring a Linear Technology LTC6754 [114].

5.2, a 1 kΩ resistor R_1 in parallel with the high-pass filter capacitor was added to the front-end design, to limit the signal undershoot. A comparison between analog waveforms produced by the original and the improved front-end designs is shown in Fig. 38. The first front-end version produces a signal that presents a sensible negative undershot, corresponding to about 25% of the positive peak value, and requiring around 100 ns for the baseline to restore back to zero. On the contrary, the improved version shows no appreciable undershoot, allowing for a higher sustainable particle rate.

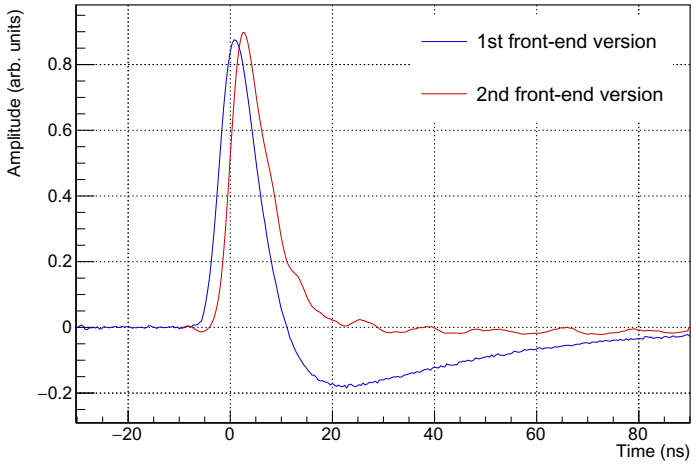


Fig. 38: .

A front-end board employs two digital-to-analog converters (DAC), one per side, to set the voltage references for the fast comparators; the schematic of a voltage comparator stage featuring a Linear Technology LTC6754 [114] is shown in Fig. 37 (b). The DAC model used is the Analog Devices AD5391BSTZ, which controls up to 16 independent voltages, with a precision of 12-bits in the range 0-2.5 V. Each DAC sets 15 voltage levels, 3 per channel, serving therefore 5 channels [115]. The output voltage levels can be set through a serial interface compatible with the $\mathtt{I^2C}$ protocol by the FGPAs present in the calorimeter modules. $\mathtt{I^2C}$ (also `IIC` or `I2C`) is a low-speed, synchronous, single-ended, serial communication bus invented in 1982 by Philips Semiconductors [116]. More details on the implementation of the $\mathtt{I^2C}$ communication for the digital threshold setup are presented in Sec. 6.5.

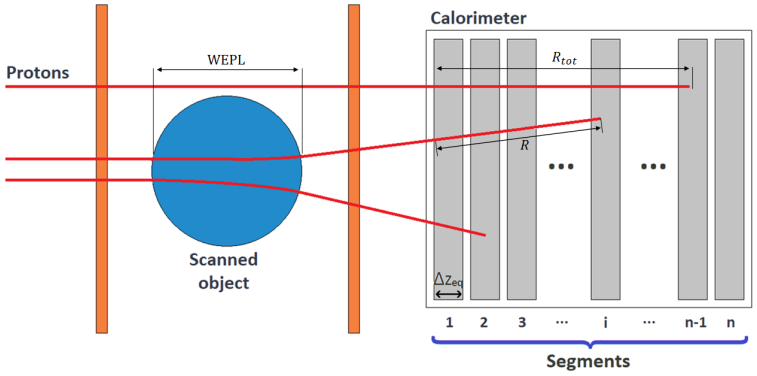
4.2.3 Estimated accuracy of the calorimeter

The 3-dimensional maps of Relative Stopping Power (RSP), obtained from a modern pCT scanner, are based on the evaluation of the Water Equivalent Path Length quantity (also referred to as WEPL) of each recorded proton [117]. The Water Equivalent Path Length can be expressed as the line integral of of RSP along the total proton path L up to its complete stop:

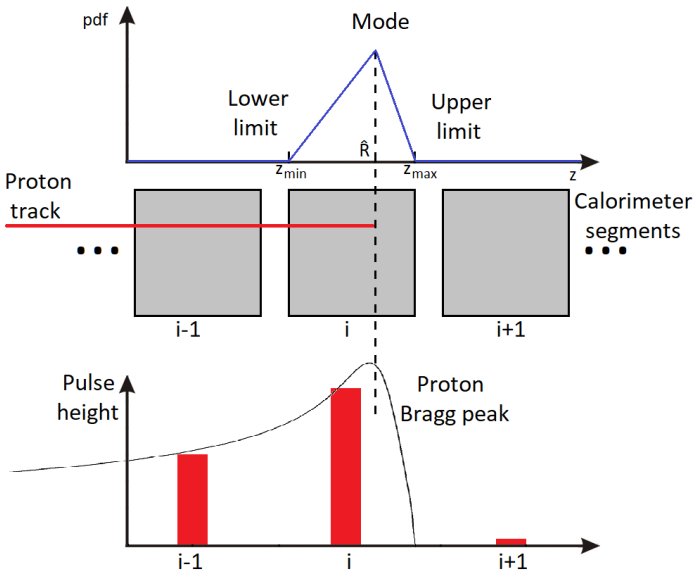
$$\mathrm{WEPL} = \int_L \mathrm{RSP}(l) dl \quad , \tag{22}$$

where $\mathrm{RSP}(l)$ is the relative stopping power at l distance along the path; RSP is defined in Eq. (11).
The iMPACT range-calorimeter, as shown in Fig. 39 (a), can be schematized as a volume divided into n consecutive segments along the beam direction z, each one with a water-equivalent thickness Δz_{eq}. The total calorimeter depth $n\Delta z_{eq}$ is sufficient to completely stop the proton beam.
The path length of a proton inside the scanned object can be estimated as the difference between the water-equivalent range of the full-energy beam inside the calorimeter R_{tot} (as defined in Eq. (12)) and the measured distance travelled by the given proton in the calorimeter:

$$\mathrm{WEPL} = R_{\mathrm{tot}} - R \quad . \tag{23}$$



(a)



(b)

Fig. 39: (a) Schematic view of iMPACT pCT scanner with segmented calorimeter; (b) Concept of an improved range calorimeter, exploiting the energy information to improve the estimation on protons stopping position given by the last hit plane. The pdf describing the knowledge of the stopping position, based on energy deposition in the calorimeter layers, is the triangular distribution.

The WEPL estimation is affected mainly by two sources of uncertainty: the experimental uncertainty on the individuation of the proton stopping point, and the intrinsic range straggling; in the energy range of interest the range straggling is estimated to be 1.1% of the range value [117]. Assuming the two contributes uncorrelated, the total WEPL uncertainty can be estimated as:

$$\sigma_{\mathrm{WEPL}} \approx \sqrt{(0.011 \times R_{\mathrm{tot}})^2 + (\sigma_{\mathrm{exp}})^2} \quad . \tag{24}$$

When the iMPACT calorimeter is operated as a simple range counter, the proton stopping point is estimated as the mid-point of the last i-th plane registering a signal (as shown in Fig. 71 (a)); the path length inside the scanned object can be then obtained as:

$$\mathrm{WEPL} \approx R_{\mathrm{tot}} - \left[(i-1)\Delta z_{eq} + \frac{\Delta z_{eq}}{2}\right] \quad ; \tag{25}$$

in this case, the lateral deviation of the track is not considered, as it is a minor contribution. The experimental uncertainty in the determination of the stopping point is simply $\Delta z_{eq}/\sqrt{12}$, as no finer estimations can be given on the stopping point inside the last segment, therefore its *pdf* (probability density function) is to be considered uniform . The total WEPL uncertainty of a Range Telescope (RT) can therefore be expressed as:

$$\sigma_{\mathrm{WEPL,RT}} \approx \sqrt{(0.011 \times R_{\mathrm{tot}})^2 + \left(\frac{\Delta z_{eq}}{\sqrt{12}}\right)^2} \quad , \tag{26}$$

which is independent on the WEPL value. The range straggling contribution is intrinsic to the physics process and is unavoidable; at 230 MeV, the target energy for iMPACT, the term $\sigma_E^{\mathrm{stragg}} \approx 3.58$ mm $(0.011 \times R_{tot})$, while $\Delta z_{eq}/\sqrt{12} = 5.877$ mm$/\sqrt{12} \approx 1.70$ mm. The sum in quadrature of the two terms yields $\sigma_{\mathrm{WEPL}} \approx 3.96$ mm, which is only 9.6% higher than the straggling term alone; therefore, even when operated in a simple range counter mode, the experimental uncertainty from the iMPACT calorimeter is a minor contribution compared to the intrinsic range straggling.
The iMPACT calorimeter, however, is able to provide also measurements of the energy deposited in each plane. As shown in Fig. 39 (b), the last $i-$th hit plane set the limits for the proton stopping position in the interval $[(i-1)\Delta z_{eq}\ ;\ i\Delta z_{eq}]$. By evaluating the signals on few planes before the last one, the proton stopping point estimation inside the i-th layer can be further refined during the data analysis stage[3]. In this case, the pdf of a stochastic variable with known lower and upper bounds (z_{min} and z_{max} respectively), and the mode as an educated guess ($\hat{R}$) is the triangular distribution, also known as the "lack of knowledge" distribution [118, 119].
The standard deviation of this distribution can be calculated as:

$$\left(\sigma_R^{\mathrm{triang}}\right)^2 = \frac{{z_{\mathrm{min}}}^2 + {z_{\mathrm{max}}}^2 + \hat{R}^2 - z_{\mathrm{min}} z_{\mathrm{max}} - \hat{R}(z_{\mathrm{max}} + z_{\mathrm{min}})}{18} \quad . \tag{27}$$

The WEPL can then be calculated as:

$$\mathrm{WEPL} \approx R_{\mathrm{tot}} - (i-1)\Delta z_{eq} - \hat{R} \quad , \tag{28}$$

and its uncertainty can be estimated with:

$$\sigma_{\mathrm{WEPL,iMPACT}} \approx \sqrt{(0.011 \times R_{\mathrm{tot}})^2 + \left(\sigma_R^{\mathrm{triang}}\right)^2} \quad . \tag{29}$$

Fig. 40 shows the total WEPL uncertainty achievable by the iMPACT calorimeter $\sigma_{\mathrm{WEPL,iMPACT}}$, as well as the various contributes $\sigma_E^{\mathrm{stragg}}$ and $\sigma_R^{\mathrm{triang}}$. In the same graph the experimental uncertainty when using the calorimeter as a simple Range Telescope σ_R^{RT} is also plotted for comparison. The experimental uncertainty $\sigma_R^{\mathrm{triang}}$ ranges between 1.20 mm and 1.39 mm, lower than the Range Telescope uncertainty, which is constant at 1.70 mm; the energy information can therefore improve the proton stopping point estimation up to 30%. The total WEPL uncertainty achievable by iMPACT fluctuates around 3.8 mm, which is 6% higher than the straggling term alone. Further improvements on the experimental precision, such as a finer segmentation and a more complex read-out system, would not lead to a substantial increase on the total accuracy, as the uncertainty is already dominated by the intrinsic variance of the physical process.
The WEPL uncertainty, as visible in Fig. 40, is essentially constant over the entire WEPL range, from 0 to R_{tot}; therefore, the response of the calorimeter does not depend on the energy of the particle, hence the *achromatic* attribute in the iMPACT acronym. The periodic substructure of the uncertainty is due to the calorimeter segmentation along the depth.
It is worth mentioning that the analysis procedure of the calorimeter signals is not yet completely defined at the present state of the project; therefore, the expression describing the

[3]The method currently under investigation, is based on a Monte Carlo to correlate the signal sequence in the last 6 hit planes, with the most probable proton stopping point.

experimental uncertainty, Eq. (27) for this estimation was chosen to be as generic as possible, and could be subject to modifications in the future.
Detailed comparison of the iMPACT calorimeter WEPL accuracy with other calorimeter architectures can be found in [112].

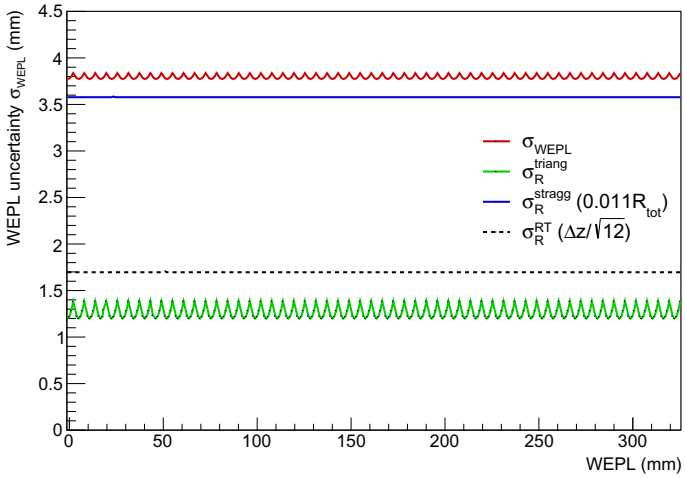


Fig. 40: Total WEPL uncertainty achievable by the iMPACT calorimeter $\sigma_{\mathrm{WEPL,iMPACT}}$, and the various contributions: $\sigma_E^{\mathrm{stragg}}$ and $\sigma_R^{\mathrm{triang}}$. σ_R^{RT} is the experimental uncertainty when using the calorimeter as a simple Range Telescope.

4.3 The iMPACT tracker

The majority of the present pCT prototypes employs silicon strip detectors (SSD) for particle tracking purposes, as illustrated in Sec. 3.3. Strip-based detectors measure particle hit positions along one dimension at the time, therefore two planes, or more (as in the PRAvDA prototype, Sec. 3.3.3), are required to retrieve a 2-dimensional coordinate. However, SSD-based trackers are limited in the sustainable particle rate, as multiple particles passing through the tracking stage, in the same time window, inevitably lead to ambiguities in the assignment of the correct coordinates to each proton track, due to the projective read-out. The sustainable acquisition rate can be increased employing pixel detectors, which feature 2-dimensional coordinate measurements in the same plane, avoiding therefore possible ambiguities. At the same time, pixel detectors allow for a substantial reduction in the material budget, as an effective measurement is performed with a smaller number of planes (each time the sensor thickness increases by a factor 2, the angular scattering increases by a factor $\approx \sqrt{2}$, as parameterized in Eq. 19).
While several pixel sensor design have been explored for the iMPACT tracker, the ALPIDE sensor, developed by the ALICE Collaboration, is adopted for the first prototype; a more suited solution is currently under development, and will be briefly presented in Sec. 9.

4.3.1 The ALPIDE sensor

The ALICE (A Large Ion Collider Experiment) Collaboration at CERN has recently undergone a major upgrade of its detector systems, in order to sustain the higher rate of interactions provided by the upgrade of the LHC particle accelerator (High-Luminosity LHC). These upgrades have been installed and, at the time of the writing, are being commissioned. In particular, the Inner Tracking System (or ITS), which is the innermost detector in the ALICE, is required to

improve its performances in terms of read-out rate as well as position and momentum resolutions [120]. At the same time less stringent requirements are imposed on the power consumption and radiation tolerance. The ALICE experiment and the ITS upgrade are presented in more detail in Sec. 8.2 and 8.3. In Tab. 3 the requirements for a pixel chip to be used in the upgraded ALICE detector are summarized; the specifications are differentiated for the two portions of the detector: the innermost Inner Barrel and the surrounding Outer Barrel.

Parameter	Inner Barrel	Outer Barrel
Sensor thickness (µm)	50	50
Spatial resolution (µm)	5	10
Dimensions (mm^2)	15×30	15×30
Power density (mW cm^{-2})	300	100
Time resolution (µs)	30	30
Detection efficiency (%)	99	99
Fake hit rate	10^{-5}	10^{-5}
TID radiation hardness (krad)	2700	100
NIEL radiation hardness (1 MeV$n_{eq}$$cm^{-2}$)	$1.7 \cdot 10^{13}$	10^{12}

Tab. 3: Requirements for a pixel sensor to be used in the ALICE Inner Tracking System Upgrade for Inner Barrel and Outer Barrel. Values for radiation hardness include a factor 10 safety margin [120].

The ALPIDE sensor (*ALICE Pixel Detector*) was one of the pixel chips developed as candidates for the ALICE ITS upgrade [121], and is the final choice for the task [122]. ALPIDE, designed using a MAPS (*Monolithic Active Pixel Sensor*) architecture, is a 50 µm thick pixel sensor with a sensible surface of 15×30 mm^2 and 28 µm pixel pitch, arranged in a 512×1024 matrix. MAPS detectors are charged particle detectors, almost exclusively used for tracking purposes, where the energy-deposition volume and the front-end read-out electronics are integrated on the same silicon die [123]. The integrated electronics is usually designed with commercial sub-micron CMOS processes. The MAPS architecture removes the need for post processing operations like bump bonding, and offers generally a lower material budget than equivalent hybrid-designed sensors.

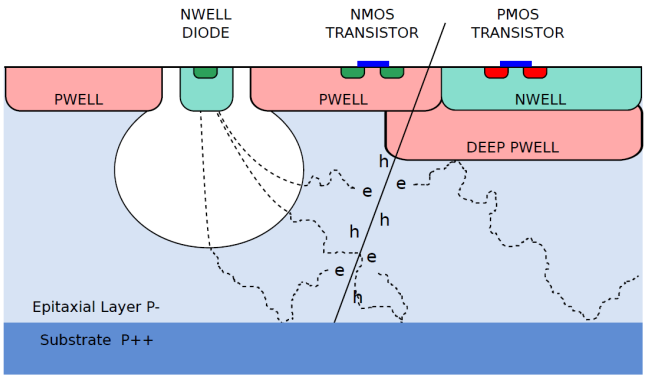


Fig. 41: Schematic view of the ALPIDE Monolithic Active Pixel Sensor (MAPS), designed in the TowerJazz 180 nm imaging CMOS process, together with a schematic representation of charge collection process [121].

ALPIDE is manufactured by TowerJazz Semiconductors [124] using a 180 nm CMOS Imaging Process, whose features are schematized in Fig. 41. The process provides an oxide thickness below 4 nm, which is expected to give a substantially better tolerance to Total Ionizing Dose (TID), with respect to solid-state technologies currently in use in particle physics experiments

sensors [121]. The six metal layers are adequate for high density and low power digital circuits. A high-resistivity epitaxial layer, with resistivity ranging between 1 kΩ cm and 6 kΩ cm depending on the wafer substrate choice, allows to generate sizable depleted volumes at low voltages (0V - 20 V), and consequently improve the signal-to-noise ratio. A reverse bias of 6 V has been shown to improve the charge collection efficiency [120]. The access to deep p-wells in the design opens up possibilities for enhanced functionalities within a pixel. In many commonly available MAPS technologies, the n-well is used as sensing diode; as a consequence, only NMOS transistors can be placed in the pixel area. In such technologies, any PMOS would require an additional n-well, which would compete with the charge collection in the sensing volume [121]. Therefore the in-pixel front-end electronics must rely on NMOS devices only, limiting the design to simple low-gain amplifiers or source followers [121]. More complex blocks, as hit discrimination or more advanced circuitry, need to be confined at the periphery of the pixel matrix. On the other hand, the TowerJazz process allows for the implementation of a deep p-well, on top of which an n-well can be created to accommodate PMOS transistors, as shown in Fig. 41. The electric potential between the epitaxial layer and the deep p-well ensures the isolation of the n-well. The presence of full CMOS circuitry inside the pixel allows for more efficient and less power-consuming read-out strategies [120]. Finally, stitching technologies allow the production of large area sensors [121].

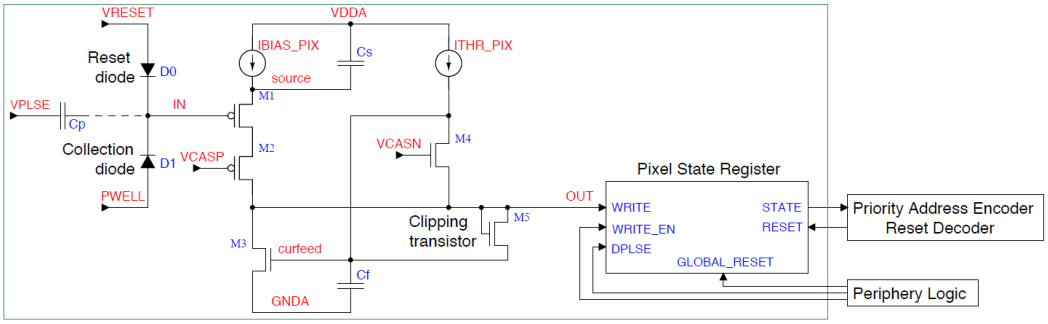


Fig. 42: Reset, amplification and discrimination circuit of ALPIDE [121].

The in-pixel front-end circuitry and signal processing chain is shown in Fig. 42. The charge generated by the particles is collected by the collection electrode, and the signal input stage can be reset via a diode connected to a static voltage V_{reset}, set by an on-chip DAC. The front-end electronics includes a high-gain amplifier and current comparator circuit, with a power consumption of about 40 nW. The output of each pixel is digital, being either 1 or 0 according to whether the deposited charge is higher than a programmable threshold. The rise-time of the amplifier is between 1 and 2 µs, with a total pulse duration of around 4 µs; the sensor time resolution is given by the rise time [120] [125]. The event is saved into a Pixel State Register, when the logic AND of the output of the comparator and an external STROBE signal is true, and stored until a read-out takes place; the register can buffer up to 3 hits on the pixel. The output of the comparator is latched. The assertion of the STROBE signal has a latency that matches the front-end peaking time. The STROBE signal is kept in a high state for 100 ns, in order to minimise the time in which the circuit output is integrated and to reduce the possibility of spurious events [121]. A configuration signal is used to enable the external simulated event injection on the pixel, either before the analog front-end (APulse) or after (DPulse); a second configuration can be set to mask the pixel, to disable noisy pixels.
Pixels are logically organized in 512 identical double-columns, each containing 512 + 512 pixels. The pixel hit statuses within a double column are encoded by an asynchronous priority encoder system, the Address Encoder and Reset Decoder (AERD). The pixel matrix read-out architecture is sketched in Fig. 43. The task of the AERD is to transfer to the matrix periphery only

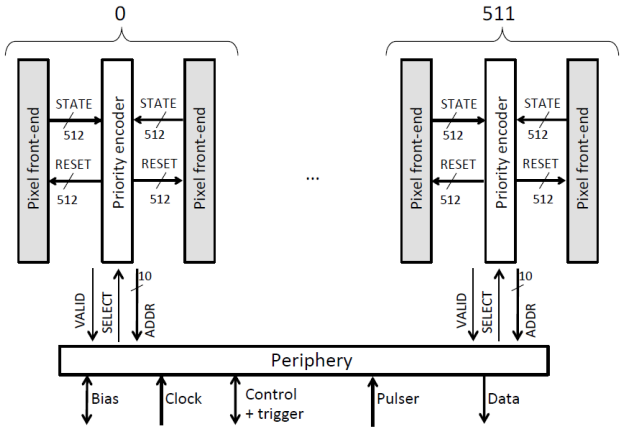
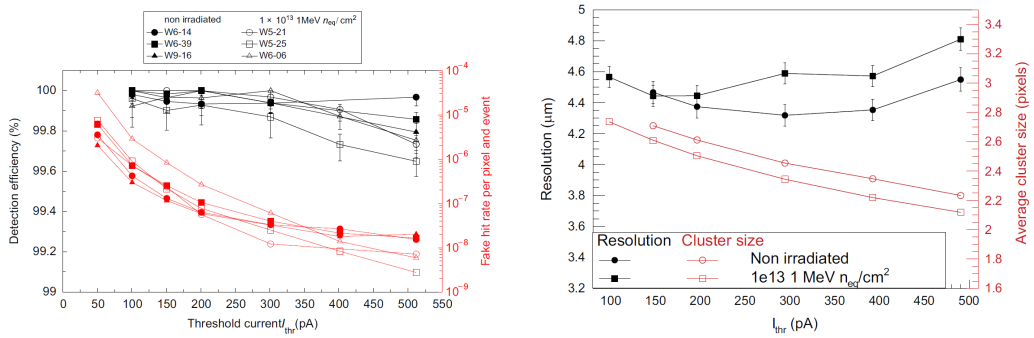


Fig. 43: ALPIDE matrix readout architecture. [125]

the addresses of relevant (hit) pixels. The AERD asserts a `VALID` signal whenever a pixel has been hit and has not been read-out yet [121], in addition to the address of the highest priority pixel on the address bus (`ADDR`). The periphery reads-out one pixel at each `CLOCK` cycle, and then drives the `SELECT` line to signal the AERD to reset the read-out pixel; the periphery `CLOCK` frequency is 40 MHz. This read-out procedure is designed to be efficient, as only the hit pixels are read-out [121]. Moreover, the entire matrix is divided into 32 regions that can be read in parallel [125]. Finally, the periphery assembles and compresses the data, which is possible as particles usually generate hits in a cluster of adjacent pixels [121].
The power consumption is around 25 mW for the entire pixel matrix, and only 3 mW for the AERD read-out, even assuming the maximum expected sensor occupancy [125]. The consumption of the entire digital periphery adds up to 102 mW, and the DACs further absorb only 0.72 mW. The overall power consumption per unit area can therefore be estimated to be around 28 mW cm^{-2} [125].



(a) (b)

Fig. 44: (a) Detection efficiency (primary axis) and fake-hit rate (secondary axis), (b) spatial resolution (primary axis) and average cluster size (secondary axis) measured for different ALPIDE prototypes, before and after neutron irradiation; reverse bias at −3 V [120].

The performances of the ALPIDE sensor have been tested experimentally with a number of prototypes, using pions and electrons of wide energy ranges; the results show a detection efficiency well above 99.5%, and a fake-hit rate significantly below 10^{-5} over a large interval of threshold currents [120]. Low fake-hit rates can be further improved using the in-pixel masking

configuration, designed to block out noisy pixels. With cluster sizes between 2 and 4 pixels, the sensor can achieve a sub-pixel spatial resolution of approximately 4.7 µm, compatible with the required specifications. The tests reported also negligible radiation effects on the sensor performances up to 10^{13} $1-\mathrm{MeV}n_{eq}$ cm^{-2}, mainly evident as a slightly smaller average cluster size. The results of the tests on ALPIDE prototypes are summarized in Fig. 44.

4.3.2 Mechanics of the tracking section

The ALPIDE sensors are arranged in structural and logic elements, called *Staves*. Each Stave is a an independent unit, containing all the mechanical, electrical and cooling elements that the sensors need in order to operate [121]. In the ALICE ITS two flavours of Staves are present, which design varies depending on which part of the detector they have to cover, the Inner and Outer Barrels of the ITS respectively. The tracking section of the iMPACT prototype employs only Inner Barrel (also referred to as IB) Staves, which will be therefore presented in this section; the design of the Outer Barrel (or OB) Staves will be discussed in Sec. 8.3.

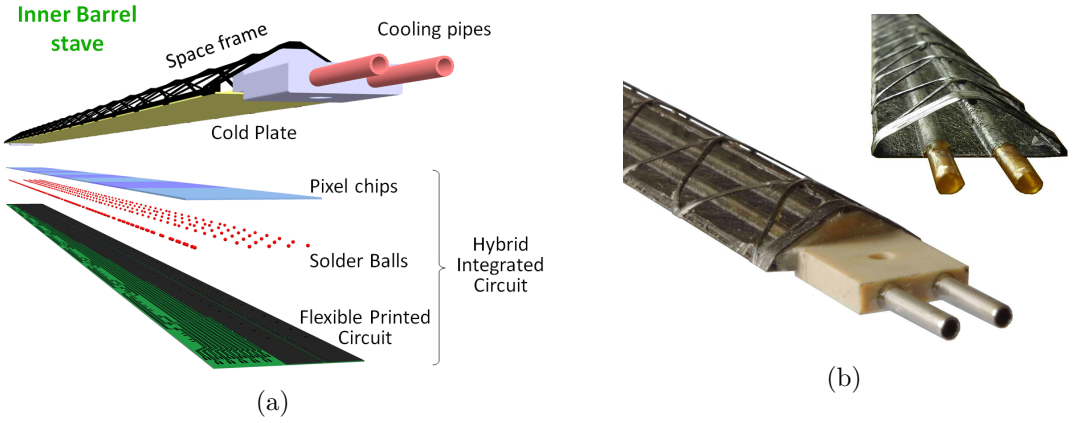
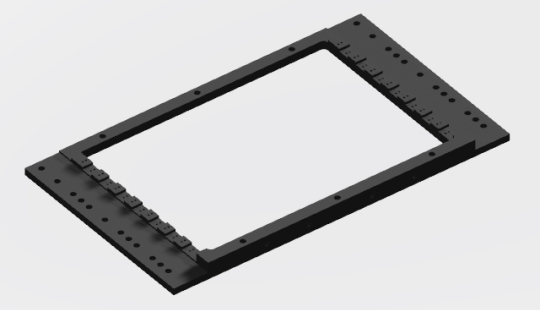
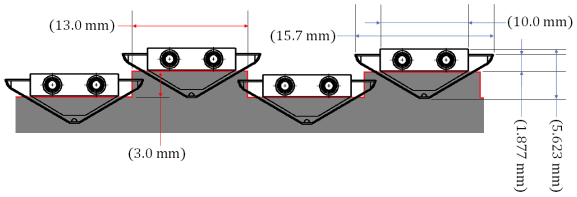


Fig. 45: (a) Schematic layout of an Inner Barrel Stave and its components, from the ITS upgrade at ALICE. (b) Pictures of a Inner Barrel Stave prototype [121].

The layout of the ITS Inner Barrel Staves is presented in Fig. 45, as well as a schematic of its components. The main mechanical element of Stave, either Inner or Outer, is the *Space Frame*, a light-weight carbon-fiber structure based on thin threads arranged in triangular patterns. Underneath the frame a sheet of high-thermal conductivity carbon fibre, named *Cold Plate*, is installed, with two polyimide cooling pipes running within. For the Inner Barrel Staves the Cold Plate is integrated within the Space Frame. The Cold Plate is in direct thermal contact with the entire surface of the sensor chips to ensure heat dissipation [121]. The *Hybrid Integrated Circuit* (HID) is an assembly of a polyimide *Flexible Printed Circuit* (FPC) on which a number of ALPIDE chips are bonded, via multiple (3 per pad) wire bonding, together with some passive components. Each Inner Barrel Stave includes a total of 9 chips arranged in line, for a total area of 15×270.8 mm^2. Between each adjacent chip a 100 µm gap is present. The data connection, power supply contacts and cooling are accessible from the same side of the Stave. Ledges extend at both ends to secure the Stave to a mounting structure, as visible in Fig. 45.
For the installations in the ITS, the Staves are arranged in barrel structures (es. in Fig. 107), while for iMPACT a custom support with a planar geometry has been designed. Figure 46 (a) shows the 3D rendering of the Stave mechanical support for the iMPACT tracker.
The tracker structural component has been designed in-house and was manufactured by the Mechanical Workshop of the Physics and Astronomy Department of the University of Padova



(a)



(b)

Fig. 46: (a) 3D representation on the Staves arranging mechanical support for the iMPACT tracking section. (b) Side view of the arrangement of the Staves; the Staves are staggered in height to allow for an overlap on the adjacent sensible surface.

and INFN-Padova [126], using a low thermal-expansion aluminum alloy. The structure is designed to host a single plane of tracking sensors; four identical planes are necessary to form the iMPACT tracking section. The structure can arrange a total of 14 Inner Barrel Staves per plane, with adjacent Staves staggered in height by 3 mm, to allow for a slight overlap of the sensing surfaces of neighboring chips, in order to avoid blind areas in the plane. The lateral view of the iMPACT Stave support is shown in Fig. 45 (b), where the height staggering is visible. With ALPIDE being 15 mm wide, the arrangement pitch is 13 mm, so the covered area of a plane is 270.8 mm wide and 182 mm high. A consequence of this design, is that about 7% of the total area of a plane is covered by overlapping sensors, so around the same fraction of particles are expected to produce one additional hit point on a plane. These additional points would facilitate the process of grouping aligned hits to identify proton tracks.

4.4 Monte Carlo simulation development

A detailed Monte Carlo simulation of the iMPACT calorimeter has been developed, in order to optimize the apparatus design in the first stages of the project, and, in more advanced stages, to provide experimental-like datasets for tests on the read-out chain, the data analysis software, and, finally, the image reconstruction techniques. The simulation was developed using the GATE framework [127], an advanced application optimized for Monte Carlo simulations in medical imaging and radiation therapy, based on the Geant4 toolkit [128].
The simulation in centred around the modeling of the single scintillating elements, parameterizing the entire signal generation mechanism, from the scintillation process, to the behaviour of the analog read-out electronics. Complex multiple-scintillator setups are then simulated including multiple instances of the single basic element.
The scintillator finger is modeled as a $0.5 \times 1 \times 20$ cm^2 box of polyvinyl toluene material, with defined density ($1.032\,\mathrm{g\,cm^{-3}}$) and stoichiometric composition (C_9H_{10}). The optical and physical properties properties of the PVT material, such as light yield, scintillation decay time constants, refractive index, light attenuation length, and light emission spectrum (Fig. 32 (a)), are also defined, based on the manufacturer tabulated parameters, as illustrated in Sec. 4.2.1. Other parameters, such as the energy resolution, had to be calibrated from experimental measurements (see [112]). The proprieties of the internal surfaces, as the teflon reflectivity, are also modeled.
Fig. 47 illustrates a simulation of a particle (a 200 MeV proton), interacting with the scintillator material: optical photons are generated isotropically around the proton trajectory, according to the defined spectrum, in number proportional to the deposited energy. Each individual optical photon is then transported inside the material, reflected by the surfaces, and possibly absorbed.

The photons reaching the SiPM surfaces have a probability to be detected; the SiPM efficiency as a function of the wavelength is parameterized from the manufacturer data sheet (see Fig. 32).

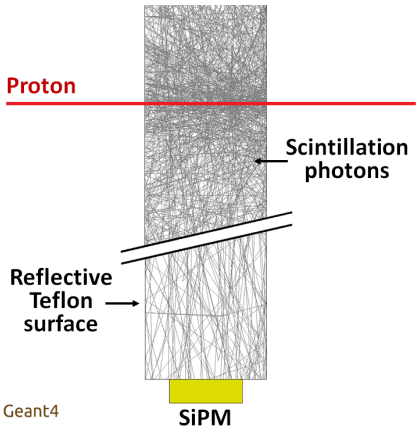
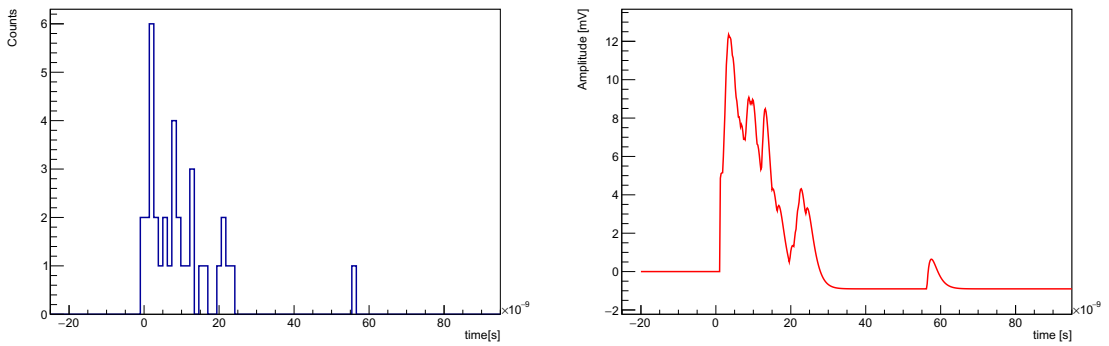


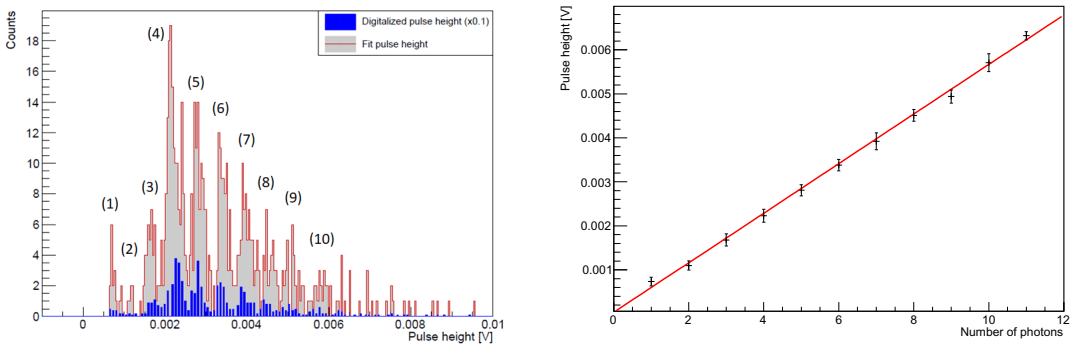
Fig. 47: Simulation of the scintillation process for a proton energy loss (red) in a PVT scintillator, transportation and reflection of optical photons (grey), and photon detection in a SiPM; obtained with a GATE-Geant4 Monte Carlo simulation. In the image the number of generated photons has been reduced by a factor ∼1000 for visualization purposes.

The photons detected by the SiPM are then used to simulate an experimental-like analog signal. The simulation provides the arrival time of each photon on the SiPM surface, Fig. 48 (a) shows an example of photon arrival time distribution (with a limited number of photons, for sake of clarity). The photon detection time distribution is a convolution of the material scintillation time constant and the photons different flight times. As illustrated in Sec. 4.2.1, the SiPM surface is formed by a matrix of Avalanche Photo Diodes (APD), each producing an identical impulse every time a photon is detected; a general multiple-photon analog signal can therefore be simulated as the pile-up of a single-photon impulse per each detected photon. Fig. 48 (b) shows an example of a simulated analog signal, built summing single-photon pulses, with the starting time of each pulse given by the photon detection time shown in Fig. 48 (a). This assumption is valid up to the point where the light intensity is high enough to cause multiple photon hits on the same APD cell; deviations from proportionality between the incident light intensity and the amplitude of output signal, in the iMPACT conditions, has been estimated to be between 1% to 5%, at the Bragg curve plateau (lowest energy deposition) and peak (maximum energy deposition) respectively [112].
The single-photon impulse, generated by the APD cell in the SiPM and then shaped by the read-out electronics, has been measured experimentally. The layout for the single-photons measurement, included a pulsed low intensity light source (at 600 nm), directed towards a teflon reflective surface; a SiPM, together with its complete read-out chain, was positioned to detect the diffused light. The setup was optimized in order to minimize the light intensity on the SiPM, at the point of being able to distinguish a countable number of photons. Fig. 49 shows the measured signal-height spectrum for this extremely low-intensity light setup. In the spectrum equally spaced peaks are visible, each one associated with an increasing integer number N of simultaneous detected photons. The spectrum trend resembles a Poisson distribution with an average number of events ≈ 4. The maximum voltage amplitude of the signals was obtained with a Gaussian fit around each waveform peak, to overcome the coarse digitization of such low-intensity voltage signals; the spectrum obtained from the digitized peak values is superimposed in blue.
Fig. 49 (b) shows the spectrum peaks position as a function of the associated number of pho-



(a) (b)

Fig. 48: (a) Example of distribution of optical photons detection time on the SiPM surface, GATE-Geant4 simulation; (b) corresponding simulated signal, as pile-up of multiple single-photon response. The number of photons generated has been reduced for visualization purposes.



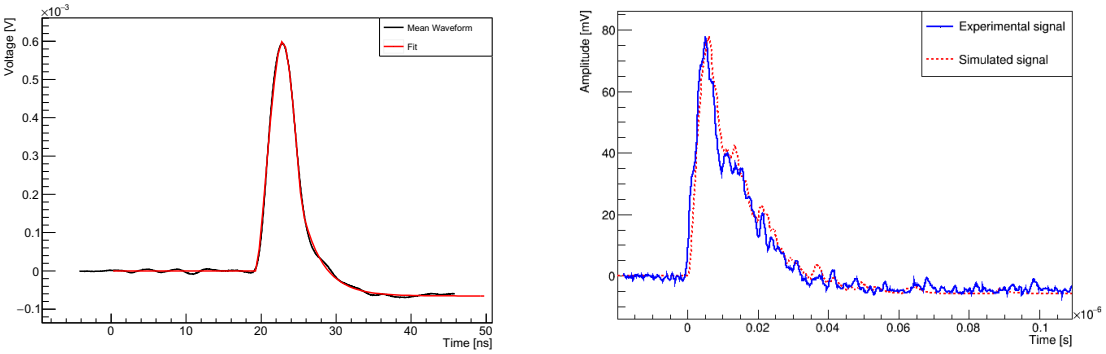
(a) (b)

Fig. 49: (a) Experimental signal amplitude spectrum of countable numbers of photons (in parenthesis) detected by a SiPM; red spectrum is obtained from a fit on the waveform peak, blue spectrum includes the maximum digitized waveform value. (b) Countable photons signal height as a function on the number of photons; a linear fit in shown in red, with (0.561 ± 0.008) mV/photon proportionality coefficient and (0.02 ± 0.06) mV intercept.

tons. The dependence is linear, and the point are interpolated with a line, the proportionality coefficient is (0.561 ± 0.008) mV/photon and the (0.02 ± 0.06) mV intercept is compatible with zero.

The average single-photon waveform is obtained by averaging the few-photons signals, each normalized by their respective number of photons N; the resulting average single-photon waveform has been parameterized with an empirical piecewise-defined custom function. The experimental average single-photon impulse is shown in Fig. 50 (a), together with the empirical parameterization (red). More details on the experimental procedure and analysis can be found in [112].

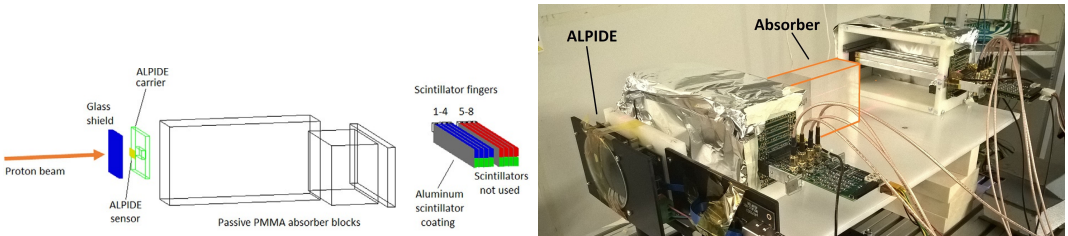
The parameterized single-photon impulse is used to build the simulated analog waveforms. The simulation has been calibrated and validated using experimental data from two separate test beams: with 3 to 5 MeV protons at INFN-LNL (Legnaro, Italy), and with 70-230 MeV protons at the TIFPA Proton Therapy Centre experimental beam line (Trento, Italy); the results of the latter test are briefly reported in Sec. 4.5. Fig. 50 (b) shows a comparison between an experimental analog waveform, produced by a 5 MeV proton, and a simulated signal in the same conditions. The simulated signal reproduces effectively both the height and the duration of the experimental waveform, as well as the ripple structure.



(a) (b)

Fig. 50: (a) Average experimental single-photon impulse (black) and empirical parameterization with a custom function. (b) Comparison between an experimental (blue) and a simulated (red dashed) waveforms, generated by a 5 MeV proton.

4.5 Tests with proton beams



(a) (b)

Fig. 51: (a) Schematic representation of the test setup, at INFN-TIFPA APSS proton beam line, including 8 active scintillators and passive PMMA absorber blocks; the relative position of the elements were varied between runs, in order to sample the Bragg curve at different depths. (b) Picture of the test setup.

Preliminary commissioning tests were conducted at the Trento Institute for Fundamental and Applied Physics (TIFPA) experimental line, located in the Proton Therapy facility [47], Trento, managed by the Azienda Provinciale per i Servizi Sanitari (APSS). The facility is based around an IBA cyclotron, that delivers proton beams of variable energy between 230 MeV and 70 MeV, which can be lowered down to 15 MeV using special passive degraders.
During the test beam, a setup including a small number of scintillating fingers were studied. A total of 8 scintillators, were arranged in two stages of four scintillators each, including SiPM, read-out electronics and input-output boards. A number of PMMA (Poly-methylmethacrylate) blocks were used as passive absorbers; the relative position between the active scintillators and the passive absorber could be changed in order to sample the Bragg curve at different equivalent depth. The test setup is presented in Fig. 51, while the arrangement used for different test runs are summarized in Tab. 4. The proton beam energy was set to 230 MeV. The analog signals of the 8 channels were recorded with two PSI DRS4 boards [129], equipped with 5 GS/s, 8-bit ADCs. A single ALPIDE silicon MAPS sensor, presented in Sec. 4.3.1, was positioned in front of the setup, as shown in Fig. 51, for an independent characterization. For the precise experimental procedure and data analysis see [130] and [131].

Run	Arrangement		
Run 4	4 scintillators	272 mm PMMA	4 scintillators
Run 5		272 mm PMMA	4+4 scintillators
Run 7		282 mm PMMA	4+4 scintillators
Run 8	4+4 scintillators		

Tab. 4: Summary of the test beam runs, including the arrangement of scintillators and passive PMMA absorber. $E_{beam} = 230$ MeV.

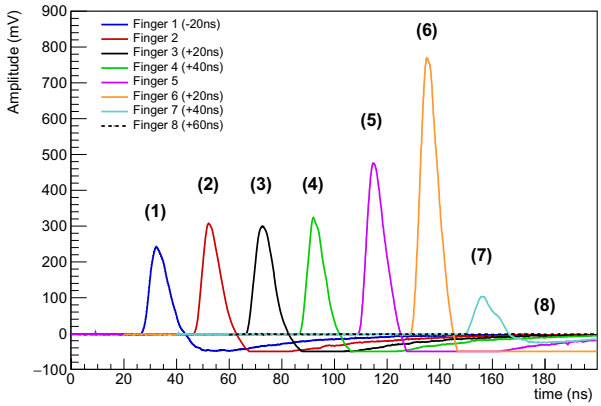
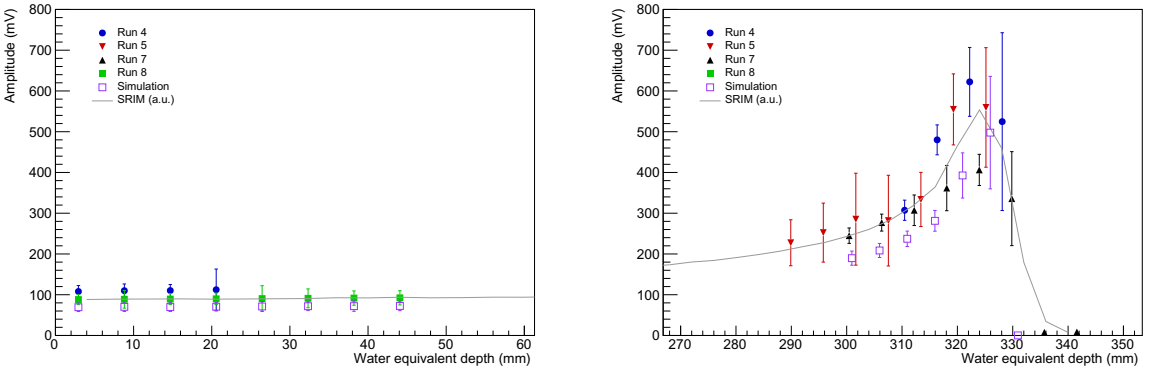


Fig. 52: Measured analog signals from the 8 consecutive scintillator fingers, relative to a single proton event, outlining a Bragg Peak trend. For the purpose of visualization, the signals are time-shifted by tens of ns (as reported in the legend). The negative portion of the waveforms is clipped due to the set dynamic range on the digitizer.

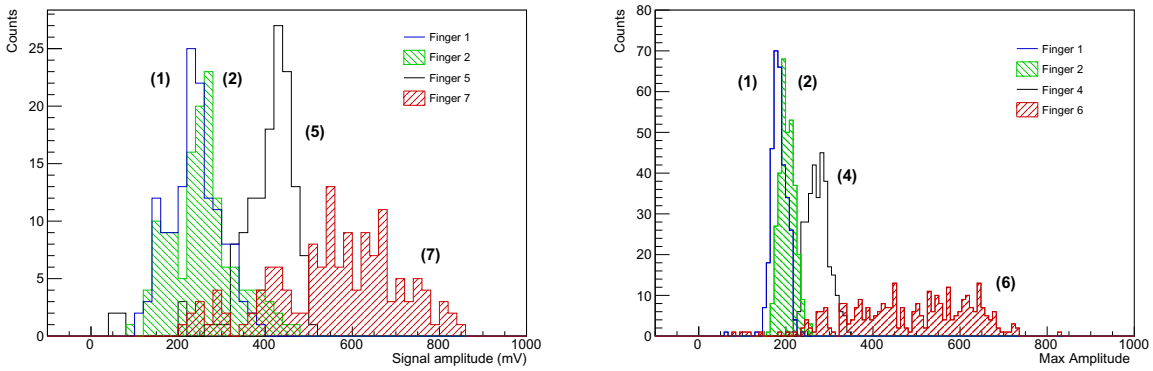
Fig. 52 shows the digitized analog signals produced by a single proton event in the 8 consecutive scintillators; the signals are shifted in time for visualization purposes. The signal height sequence outlines the trend of a proton Bragg Peak. The signal height in the first four scintillators is almost constant, while it reaches its maximum in the 6^{th} scintillator. The proton then travels a small distance inside the 7^{th} scintillator and therein comes to rest; in fact, in the last scintillator (the 8^{th} one) no significant signal is registered. The scintillator in which the proton stops is therefore clearly recognizable; moreover, using the energy deposition sequence in the scintillators, a finer estimation of that position is possible.
Fig. 54 summarizes the data collected during the test beam at TIFPA. The z-coordinate of each experimental point has been expressed as water equivalent depth, based on the absorber and scintillator elements configurations in each run, as presented in Tab. 4; more details on the data analysis procedure can be found in [113] and [132]. The energy loss profile retrieved from the SRIM software is also superimposed to provide a visual guide, and empirically scaled to overlap the data points. Fig. 54 (a) highlights the initial portion of the energy deposition curve, where the trend is approximately constant in the first 5 cm, at around 100 mV. Fig. 54 (b) shows the 230 MeV proton Bragg peak, located at a 325 mm of water-equivalent depth. The shape and position of the peak is visible from the data points, with a peak amplitude roughly 5 times higher than the signals registered in the initial plateau. Both experimental and simulated data points present large error bars, especially at higher depths, due to range straggling that affect the protons near the end of their paths; this effect is also visible in Fig. 54. Nevertheless Fig. 53 and 54 shows the capability of the iMPACT calorimeter to recognize the protonic Bragg peak in the energy range of interest and, most importantly, to identify the stopping depth of single protons with the required precision. Some discrepancies are visible in the depth of the



(a) (b)

Fig. 53: Summary of the measurements performed at the TIFPA beam line, with a combination of passive PMMA absorber blocks and 8 scintillators, outlining the Bragg curve for a 230 MeV proton beam. In (a) and (b) the initial and final portions of the energy deposition curve are highlighted, respectively. The runs configurations are summarized in Tab. 4. Points obtained with a Monte Carlo simulation of the setup are also included (details on the simulation procedure can be found in Sec. 4.4); an energy deposition curve obtained from SRIM are superimposed to the points, arbitrarily scaled to overlap the data points.

peak between different runs; this can be due to an inaccurate absorber density or composition description, which can lead to imprecise equivalent depth calculations.

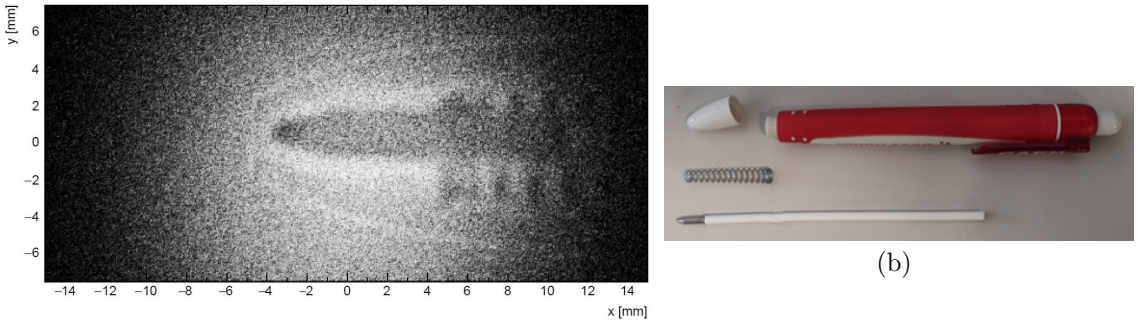


(a) (b)

Fig. 54: (a) Signal amplitude spectrum of scintillator fingers 1, 2, 5, and 7 in the Run 5 configuration (the runs configurations are summarized in Tab. 4); the fingers 1 and 2 sample the initial part of the Bragg peak build-up, while the 7^{th} one is located at the depth at which the majority of the 230 MeV protons stop. (b) Examples of spectra obtained from the Monte Carlo simulation, roughly corresponding to the depth of the experimental ones. The error bars correspond to the distributions rms.

Finally, Fig. 54 (a) shows examples of experimental energy spectra recorded in four different scintillators, in the Run 5 configuration, while Fig. 54 (b) shows simulated energy spectra. In the two graphs, scintillators located ad the initial part of the Bragg peak build-up (labelled 1 and 2) shows relatively defined peak, around 200 mV. In deeper scintillators (as number 4 and 5) the peaks move towards higher values, as the average energy deposition increases; the

scintillators in which the majority of the protons come to rest (the 7^{th} one in (a) and the 6^{th} in (b)) present a particularly broad energy distribution, as the protons are affected by range straggling, and the path length travelled by the particles in the last scintillator can vary considerably (the range straggling of ≈3.6 mm is comparable with the scintillator thickness of 5 mm). It can be noted that the simulated distributions reproduce the experimental behaviours.
During the test at the TIFPA beam line, a single ALPIDE sensor was tested, which was read-out independently from the calorimeter section; the goal of the ALPIDE test was to demonstrate the usability of a MAPS sensor in an imaging application. ALPIDE operates with an external trigger (Sec. 4.3.1); in this case the sensor was triggered continuously with a an impulse generator set at a regular 2 kHz frequency. The bias voltage was set at 0 V. Fig. 55 shows a demonstrative radiography performed on a ball-point pen. The tip of the pen was fixed in front of the silicon sensor, and irradiated with a 70 MeV proton beam. The plot shows a simple map of the number of hits detected by each pixel in the 1.5×3 cm^2 active area of the sensor. The grey scale is proportional to the rate, with a lighter colour indicating a higher number of hits. The plot includes a total of $6 \cdot 10^5$ events. In the image, the internal components of the pen, such as the plastic cap, the metal spring and core are visible; in particular the point sphere, which is the denser element, appears the darkest. In fact, materials with a higher density or higher atomic number, generally deviate protons at larger angles, projecting therefore a darker shadow on the sensor surface. The image definition is, however, affected by considerable Poisson noise, as the pixel with the highest multiplicity contains only about 80 hits. Fig. 55, although preliminary, shows the high image contrast achievable with protons, as well as displaying the potential use of MAPS sensors in medical imaging.



(a) (b)

Fig. 55: (a) Proton radiography of the tip of a ball-point pen, obtained with the ALPIDE sensor, with a 70 MeV proton beam. (b) Picture of the pen.

5 Scintillators electronics development and characterization

5.1 The iLDA setup

During the development and prototyping phase of the iMPACT pCT scanner, multiple choices in terms of instrumentation, materials, and electronics design were required, in order to optimize the performance of the apparatus. Moreover, the characterization of the behaviour of the different elements included in the system was needed, in order to parameterize their response. These prototyping tests were carried out exploiting cosmic ray muons, which readily provide a repeatable and well-known particle source to generate reference datasets. The iLDA setup (iMPACT LabVIEW Data Acquisition), a ready-to-use table-top flexible data acquisition system, was developed within the iMPACT framework to conduct such studies. Besides preliminary measurements presented in this Section, the iLDA setup is also used during the final assembly of the iMPACT calorimeter for quality assessment, in particular to calibrate and verify the uniformity of signal read-out in each of the 480 channels included in the present prototype, as illustrated in Sec. 4.2.1.

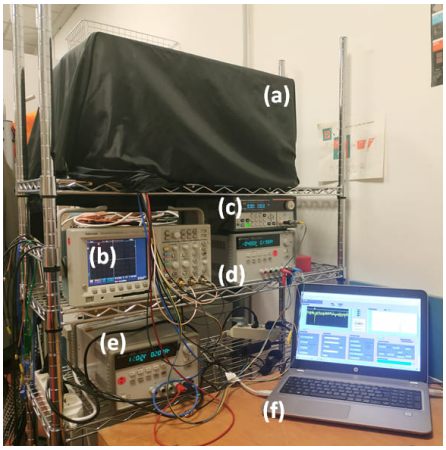


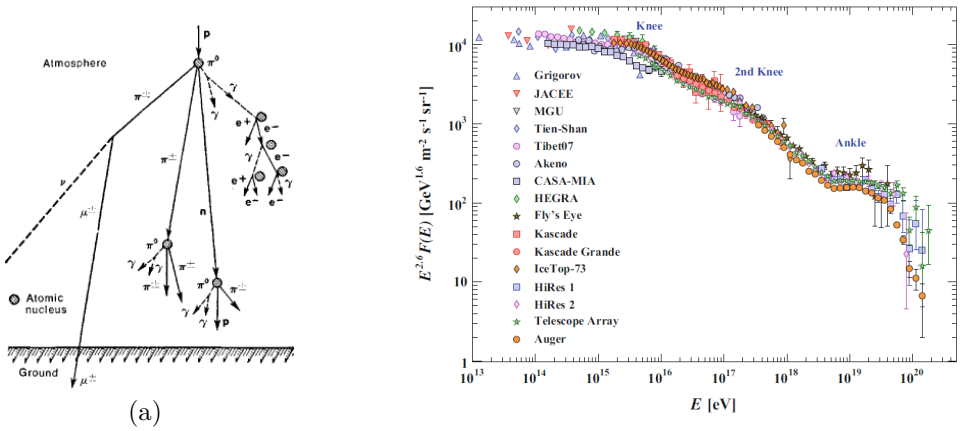
Fig. 56: Photo of the iLDA (iMPACT Labview Data Acquisition) system: (a) dark box where the scintillators and front-end electronics are placed; (b) Tektronix TDS3034B Oscilloscope (4 channels, 2.5 GS/s, 300 MHz); (c) power Supply for SiPM bias; (d) power supply for read out electronics; (e) power supply for PMT bias; (f) acquisition computer with LabView software.

The iLDA setup is shown in Fig. 56; it can simultaneously handle up to 4 scintillator channels, together with the SiPMs, and a complete analog read-out chain, arranged in a modular setup that allows for a quick replacement of single elements for comparative studies. The scintillator, SiPMs and electronics are housed in a dark box (indicated with (a) in Fig. 56), in order to shield the elements from external light; the data waveforms digitization is performed with a Tektronix TDS3034B Oscilloscope (b), featuring 2.5 GS/s over 4 channels. A custom LabVIEW-based software was developed to control the data acquisition procedure; via the software the oscilloscope setup can be managed, even remotely, with real-time acquisition diagnostics and data storage. The system was developed for long-duration continuous data acquisitions, up to several days, which is essential to collect sufficient cosmic rays statistics. Finally, different trigger structures can be implemented via the iLDA software, using different logical combinations of the available channels. The setup includes all the necessary power supply to operate different type of instrumentation (from SiPM to PMT tubes) and electronics.
In the iMPACT framework, this cosmic rays setup was used to compare different scintillator models and dimensions, read-out electronics design, as well as to optimize the scintillator wrap-

ping materials and procedure; the results are presented in Sec. 5.2. Moreover, the same setup was employed to verify the uniformity of response of the scintillators along their length and to tune the Monte Carlo simulation parameter to match the experimental behaviours; these test are presented in Sec. 5.3.

5.1.1 Cosmic rays

At sea level, radiation can be categorized into two groups: natural radioactivity of elements of the Earth surface, and radiation in cosmic ray showers produced by high-energy primaries interacting in the atmosphere, as shown schematically in Fig. 57 (a). The energy distribution of the primary cosmic rays, estimated from shower measurements at the surface of the Earth, can be described by a power law E^{-p}, with p typically between 2.7 and 3.3 [133]. Fig. 57 (b) show the energy spectrum of the cosmic rays that reach the planet surface, as measured by various experiments.
At lower energies (less than 1 GeV) the particle flux is particularly intense, with thousands of particles per square meter per second, while the spectrum present a cut-off at energies between $10^{19}-10^{20}$ GeV; particles with energies above 10^{11} GeV are extremely rare, with one particle per square kilometer per century [133]. As visible in Fig. 57 (a), the energy spectrum presents three changes of slope, called *knees* and *ankle*. Below a few GeV the flux is dominated by particles originated from the sun, with the intensity of these lower energy particles being variable in time; the first knee, located around $5 \cdot 10^{15}$ eV, indicates a steepening of the trend from $p = 2.7$ to 3.1, and separates the region relative to a galactic (at lower energies) and a extra galactic origin of the particles. The spectrum trend further steepens to $p = 3.3$ in proximity of the second knee located at $4 \cdot 10^{17}$ eV, and finally returns flatter after $5 \cdot 10^{18}$ eV, i.e. the ankle. The origin of the second knee and ankle is not yet clear [133].



(a) (b)

Fig. 57: (a) Schematic representation of an atmospheric cosmic ray shower: the primary particle (usually a proton) collides with a nucleus of the atmosphere generating a cascade of secondary particles, such as secondary protons, neutrons, neutral and charged π mesons, antineutrons, antiprotons, hyperons, and heavy mesons. Gamma rays are generated as product of neutral mesons decay, initiating electromagnetic showers [134]. (b) Energy spectrum of cosmic rays on Earth surface, measured by various experiments; the spectrum has been multiplied by $E^{2.6}$ to highlight the different regions [135].

The acceleration of particles in the universe is believed to take place in the proximity of gigantic gravitational collapses, such as in supernovae and supermassive black holes, which, due to the spatial scales available in space, are able to impress to particles kinetic energies of magnitude unattainable in human facilities [133].
Protons account for the majority (around 88%) of the high-energy primary cosmic rays at

the atmosphere entrance; an additional 10% are alpha particles and 1% are heavier ions or neutrons. The remaining 1% includes electrons, photons and neutrinos. The composition of cosmic rays changes drastically when the primary particles encounter the Earth atmosphere. The primary particles, in fact, interact with the nuclei of the atmospheric atoms, mainly via hadronic inelastic collisions; the most abundant particles generated from these collisions are pions ($\pi^{\pm}$ or π^0), but kaons, hyperons, charmed particles and nucleon-antinucleon pairs are also produced [136]. Most of the secondary particles decay or further interact before reaching the surface. In particular, charged pions almost exclusively decay into muons with a 0.999877 branching fraction and mean lifetime τ=2.6 $\cdot$ 10^{-8} s:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$$

with the second most likely decay being into electrons (0.000123 fraction). Neutral pions instead decay into gamma rays with a particularly short lifetime (τ=8.4 $\cdot$ 10^{-17} s):

$$\pi^0 \rightarrow \gamma + \gamma$$

and can start electromagnetic showers. Muons further decay into electrons:

$$\mu^+ \rightarrow e^+ + \bar{\nu}_e + \nu_\mu$$

$$\mu^- \rightarrow e^- + \nu_e + \bar{\nu}_\mu$$

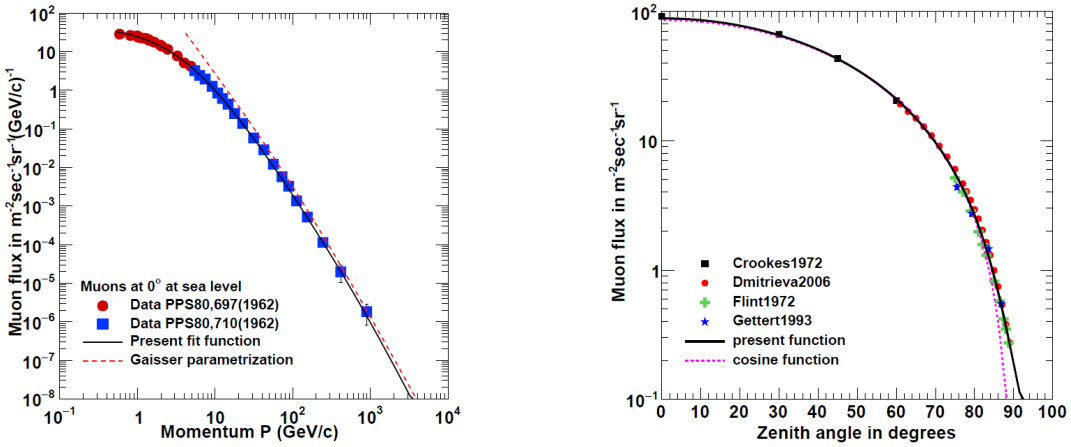
with a mean lifetime of τ=2.196 $\cdot$ 10^{-6} s (the equality between the positive and negative muon lifetime has been establish to better than one part in 10^4 [137]). Secondary muons are produced at about 10 to 15 km of height, and Lorentz dilation is large enough to allow them to reach the Earth's surface, taking into account special relativity; during their flight, muons lose about 2 GeV of energy before reaching the ground [136]. Charged particles at sea level are mostly muons, with a 60 $\mathrm{m^{-2}\,s^{-1}\,sr^{-1}}$ flux, and a mean energy of about 4 GeV [133]. The muon energy spectrum (shown in Fig. 58 (a)) below 1 GeV is approximately flat, while it follows the falling primary particle spectrum trend in the 10-100 GeV range; at higher energies the spectrum steepens, as pions of higher energies are progressively more likely to interact before decaying [136]. The zenith angular distribution follows roughly a $\cos^2\theta$ trend, although the distribution is energy-dependent: at lower energies the distribution decreases faster with the angle, as low-speed muons at larger angles are more likely to decay in flight. The average muons angular distribution is shown in Fig. 58 (b).
The general energy loss for muons can be described as the sum of the electronic energy loss (defined by the Bethe-Bloch formula, Eq. (6)) and a radiative term:

$$-\left\langle \frac{dE}{dx} \right\rangle_\mu = -\left\langle \frac{dE}{dx} \right\rangle_{el} + b(E)E \quad , \tag{30}$$

where $b(E)$ includes bremsstrahlung, pair production, and nuclear contributions. The radiative term, for muon energies lower than 100 GeV, accounts for less than 1% of the total energy loss [138]. At high energies, the electronic energy loss and the radiative term $b(E)$ are approximately constant, and the total muon energy loss can be expressed as:

$$-\left\langle \frac{dE}{dx} \right\rangle_\mu \approx \frac{1}{b} \ln\left(1 + \frac{E}{E_{\mu c}}\right) \quad , \tag{31}$$

where the critical energy $E_{\mu c}$ is defined as the energy at which the radiative term equals the electronic energy loss. Fig. 59 shows the energy loss as a function of the particle momentum for positive muons in copper, vertical bands define the regions in which the different terms dominate: in the central region the energy loss substantially follows the Bethe-Block formula,



(a) (b)

Fig. 58: (a) Measured muon momentum distribution at 0 zenith angle, at sea level. (b) Average measured muon angular distribution at sea level [136].

while above the muon critical energy (around hundreds of GeV) the radiative term prevails.
In the very low-energy regime, muons are slower than the valence electrons, and the stopping power in this region has been modelled by Lindhard, describing the material electrons as a Fermi gas [139]: the energy loss turns out to be proportional to the muon velocity. The second region has been described by phenomenological fits to experimental data by Andersen and Ziegler [140].
Cosmic muon, which have an average energy between 1 and 10 GeV, can be considered *Minimum Ionizing Particles* (MIP), with an average energy loss of 1.66 MeV g^{-1} cm^2.

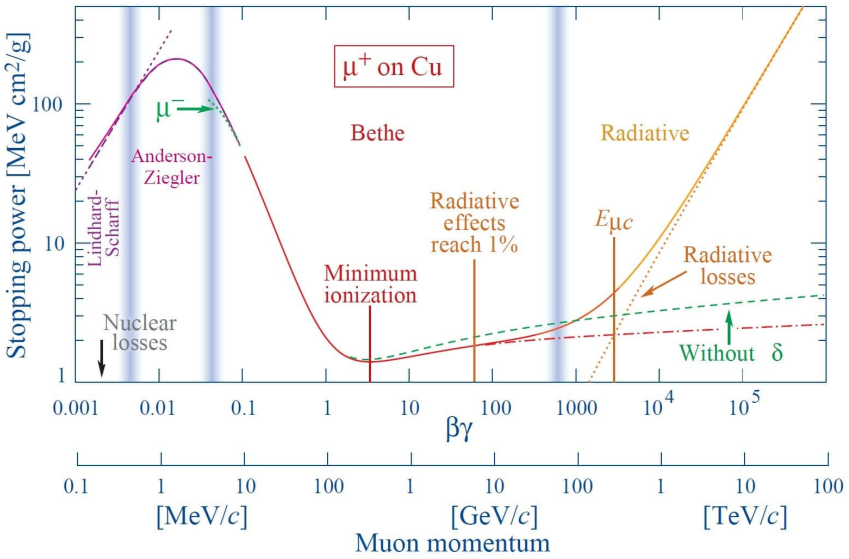
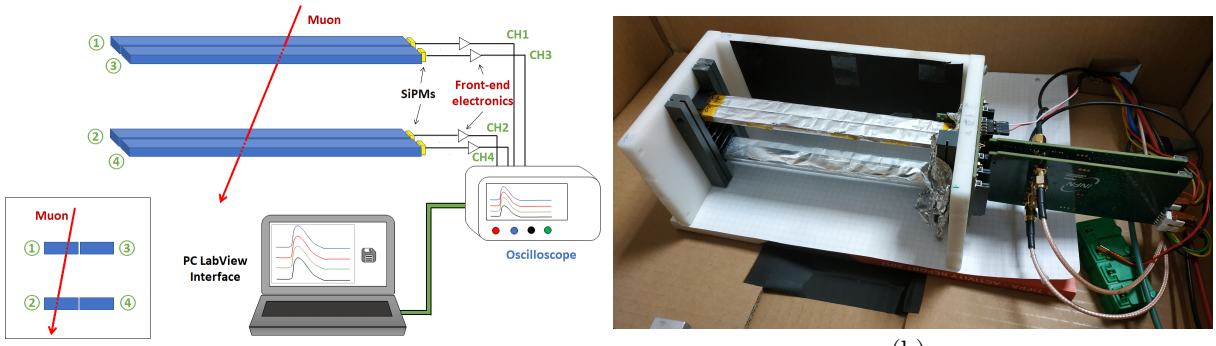


Fig. 59: Stopping power for positive muons in copper as a function of momentum and $\beta\gamma$ parameter; from [141].

5.2 Prototyping tests with cosmic rays



(a) (b)

Fig. 60: (a) Layout of the iLDA setup in the configuration for the prototyping test. A total of four scintillators are arranged in pairs on two levels (labelled 1-3 on the top and 2-4 on the bottom), as highlighted in the box. (b) Photo of the scintillator arrangement in this configuration.

The basic configuration for prototyping tests is presented in Fig. 60 (a). It includes four scintillator fingers, arranged in two pairs, placed on two levels; the two scintillators on the top level are labelled with 1 and 3, while the bottom ones with 2 and 4. The four scintillators are placed horizontally inside the dark box, as visible in Fig. 60 (b). A plastic structure ensures the mechanical stability of the system and houses the four SiPMs; two read-out boards, containing shaping and amplifying analog electronics, retrieve the signal from the SiPMs and send them to the oscilloscope for the data acquisition. The acquisition of an event is triggered by a coincident signal in two scintillators vertically aligned (e.g. 1 `AND` 2 or 3 `AND` 4), in order to select vertical muon tracks. With the coincident trigger other cosmic rays components, as gamma and neutrons, are suppressed, as these neutral particles have a negligible probability of interacting in both the triggered scintillators.
In Fig. 61 some results of the prototyping tests of the calorimeter are summarized: the graph shows the average waveforms retrieved with four different hardware variations. The waveform generated by the candidate scintillator model BC-420 is drawn in blue, compared with the BC-408 model in red. The BC-420 produces a signal that is 5 to 10 ns longer than the counterpart. Therefore, the BC-408 was preferred for the calorimeter assembly, also because, the amplitude of the signal is abut 10% higher, as its light output spectrum better matches the SiPM efficiency (visible in Fig. 32 (a)).
The behaviour of a shorter scintillator was also studied; the signal generated by a BC-408 scintillator cut in half is shown in Fig. 61 with a green line. The waveform relative to the half-length scintillator appears marginally shorter in duration, with respect to the full-length element, and with a slightly higher peak; this is because the generated optical photons, in the shorter scintillator, travel a shorter distance before being detected, and the probability of being absorbed along their path diminishes. These minor advantages, however, do not justify a modification in the calorimeter layout to host shorter-length scintillating elements.
Finally, a read-out electronics including a second derivative stage (see Sec. 6.1.1) was also tested; the generated waveform is shown in Fig. 61 with a black line. In this configuration, the waveform, after the peak, reaches significant negative voltages, more than half the peak value. Despite the overall shorter pulse duration, the second derivative stage was discarded, as large negative voltages could damage other electronic components, especially the fast comparators.

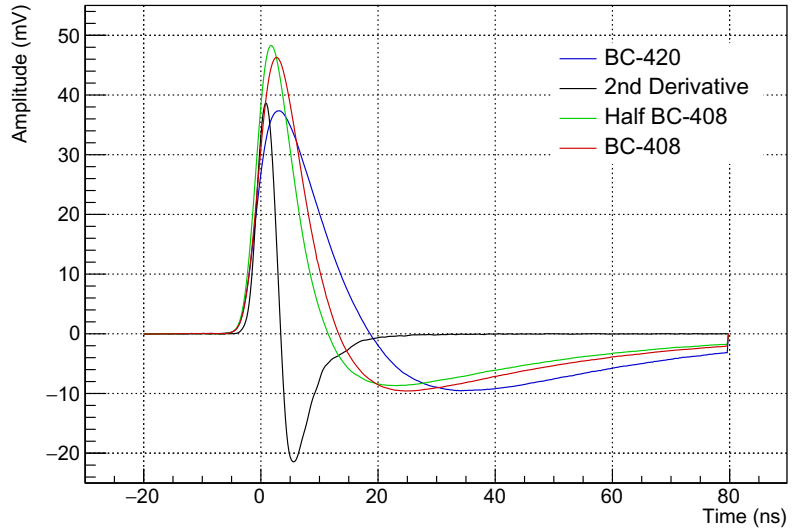


Fig. 61: Average waveforms from cosmic ray acquisitions, with four different hardware variations: data retrieved with a BC-420 scintillator model is shown in blue, a proposed version of read-out electronics employing a second derivative stage is shown in black, data from a half-length BC-408 scintillator is in green, reference acquisition with a BC-408 is shown in red.

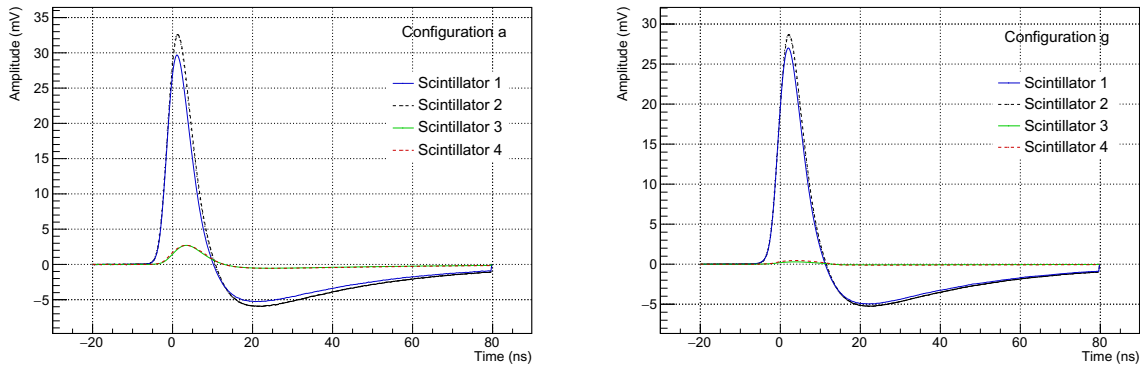
5.2.1 Optimization of the wrapping procedure

Cosmic rays were also employed to study the most efficient materials and procedure for the wrapping of the scintillating elements. The scintillator wrapping represents a key component for the effective operation of the iMPACT calorimeter, as it has to ensure an adequate internal reflectivity with a low light absorption on its internal surface; these characteristics are essential for an efficient collection of the photons produced by the passage of particles. At the same time the wrapping material needs to provide optical isolation between adjacent scintillators to avoid optical cross-talk, and isolation from external light sources, which can cause noise or even damage the SiPMs. Finally, the wrapping material thickness should be minimized, as it constitutes a passive material in the calorimeter, causing undetected energy loss and resulting in blind spots between adjacent scintillators forming a plane.
The materials considered for the wrapping are commercial aluminum and teflon. Teflon (Polytetrafluoroethylene or PTFE) has an extremely high reflectivity ($\approx 99\%$) in the optical wavelength range [142], therefore it is best suited as the innermost layer in direct contact with the scintillator surface. Aluminum has a worse reflectivity than teflon (between 85 and 87% [143]), however for its opacity to optical light, ductility, and availability in thin thicknesses ($\approx$16 µm), it represents a convenient solution for the most external wrapping layer. The iLDA system has been used to evaluate the combination of teflon and aluminum layers, to ensure optimal light collection efficiency and minimize the material budget. At the same time, studies on the magnitude and origin of possible cross-talk (either electronic or optical) between adjacent channels have been conducted.
Different configurations have been tested, based on the essential layout presented in Fig. 60, formed by four scintillators arranged in two pairs. The idea is that, triggering on an actual particle event in one scintillator, any cross-talk effect in the adjacent one can be observed in the form of an induced signal. For each configuration two datasets are collected, one triggering with the coincidence of channels 1 and 2, observing the cross-talk in channels 3 and 4, and vice-versa. The specular acquisitions allow to even out possible non-uniformities between channels,

in terms of SiPM gain, electronic components tolerances, or manual assembly. The following acquisition configurations have been studied:

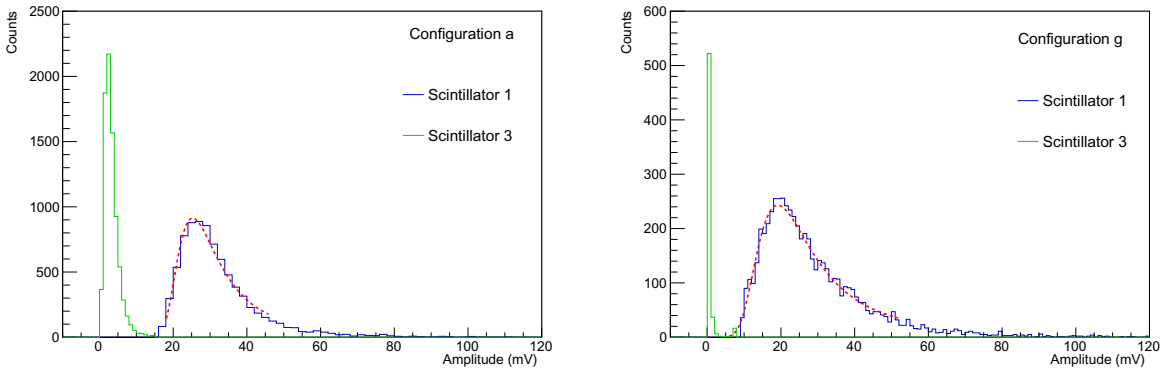
- configuration a (c_a): the four scintillators were all uncovered; in this configuration the optical cross-talk is maximised, as only a fraction of the optical photons remains inside the scintillating volume in which they were generated (for PVT the minimum angle for total internal reflection is 39° from the surface normal);
- configuration b (c_b): the triggered scintillators were uncovered, while the other two scintillators were wrapped in aluminum and the sides facing the SiPMs were covered with black tape in order to prevent them from detecting scintillation photons; this configuration provides a background reference, where the optical cross-talk is eliminated, and only the cross-talk due to adjacent SiPMs and analog electronics chains is present. In particular, the avalanches generated in a SiPM could induce a signal in neighboring electronic components;
- configuration c (c_c): all four channels were triggered with an external impulse generator at a constant frequency, to measure the electronics noise floor and SiPM dark counts, without any particle events;
- configuration d (c_d): triggered scintillators are uncovered. The non triggered scintillators are wrapped in aluminum and their corresponding SiPMs are removed; this configuration allows for an estimation of the cross-talk caused purely the analog electronics of adjacent channels, excluding the contribution from the SiPM;
- configuration e (c_e): one teflon layer separates the adjacent scintillators;
- configuration f (c_t): two teflon layers separate the adjacent scintillators;
- configuration i (c_g): each scintillator is wrapped individually in a single teflon layer and an aluminum layer separates adjacent scintillators.

Fig. 62 (a) and (b) shows the average waveforms for the four channels in two particular configurations: c_a and c_g respectively. In the first configuration, the scintillators are completely uncovered, therefore photons can leave the scintillating volume in which they were generated and get detected in the adjacent SiPM; in Fig. 62 (a), in fact, the non triggered channels (3 and 4) present a visible induced signal in correspondence of the primary signal. The configuration presented in Fig. 62 (b), c_g, includes an aluminum foil between adjacent scintillators and the induced signal is almost absent; the small induced signal can be due to small openings in the aluminum wrapping or to electronic cross-talk.
Fig. 63 shows signal amplitude spectra, in the same considered configurations, where the main signal spectrum is presented in blue, while the induced spectrum is in green. The spectra can be described by a Landau distribution, typical of the energy deposition of minimum ionizing particles in thin absorbers; the red dashed lines in the plot represent, in fact, fits to Landau functions. In the c_a configuration (Fig. 63 (a)) the cross-talk signal spectrum peaks around 5 mV, compared with the main Landau signal spectrum that peaks around 30 mV. On the other hand, the induced cross-talk (Fig. 63 (b)) distribution in the configuration c_g appears closer to zero and narrower.
For each configuration, the signal amplitude spectrum was interpolated with a Landau distribution, while for the spectra including only electronic noise a Gaussian interpolation was performed. The results are summarized in Tab. 5. Data distributions are reported with the median of their Landau fit, with the upper and lower errors marking the 34% of the distribution integral on either side of the median. Data distributions were chosen to be described with the median, instead of the mean, as the latter is more shifted towards the high-energy tail and further away from the maximum value. In Tab. 5, the average ratio between the induced signal



(a) (b)

Fig. 62: (a) Average waveform registered in each channel, in a configuration with a sensible optical cross-talk (configuration c_a); the trigger is set in scintillators 1 and 2, induced signal is visible in scintillators 3 and 4. (b) Average waveforms in a configuration with negligible optical cross-talk (configuration c_g)



(a) (b)

Fig. 63: (a) Signal amplitude spectrum in scintillators 1 (triggered, in blue) and 3 (not triggered, in green), in configuration c_a; the induced signal in scintillator 3 is visible. The spectrum of scintillator 3 is rescaled in height by a factor 1/2 for visualization purposes. (b) Signal amplitude spectrum in a configuration with negligible optical cross-talk (c_g); the spectrum of scintillator 3 is rescaled in height by a factor 1/10 for visualization purposes. The red dashed lines show a interpolation with a Landau distribution.

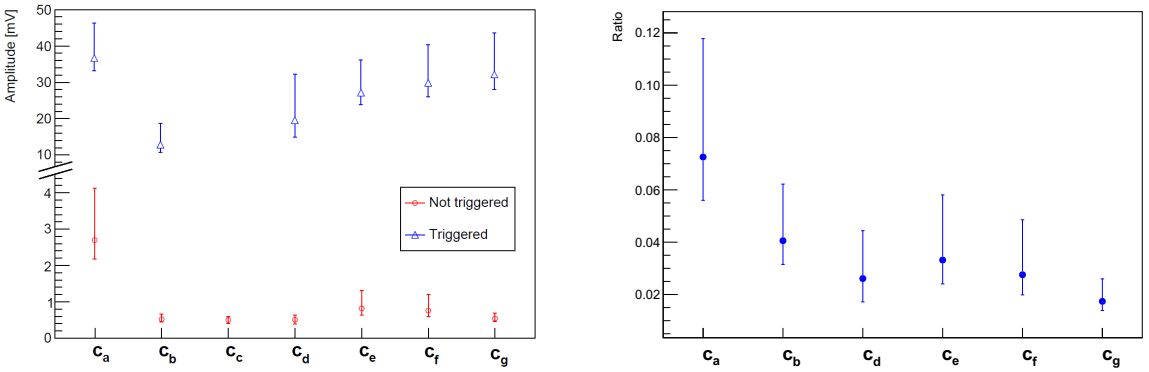
over the main signal is also presented. Additional details on the experimental and analysis procedure can be found in [144].

The results are also summarized in Fig. 64 (a), which shows for each configuration the average main triggered signal (blue points) and the average induced non-triggered signals; Fig. 64 (b) shows the induced/main signal ratio.

The configuration c_g, which includes one teflon layer around each scintillator and a single aluminum between them, appears the best wrapping solution, as it minimize the cross-talk while employing a moderate amount of material. In this configuration the main signal is on average higher than 30 mV, inducing in adjacent scintillators less than 2% of the primary signal. Moreover, the absolute value of the induced signal in this configuration is 0.55 mV, which is compatible with the noise floor due to electronic noise and SiPM dark counts, measured in c_c.

Configuration	Triggered channels	Main signal amplitude (mV)	Cross-talk signal amplitude (mV)	Ratio
a	1-2	34^{+13}_{-5}	$2.8^{+2.2}_{-0.8}$	$0.07^{+0.05}_{-0.02}$
	3-4	39^{+14}_{-5}	$2.3^{+1.8}_{-0.7}$	
b	1-2	13^{+8}_{-3}	$0.54^{+0.25}_{-0.09}$	$0.041^{+0.022}_{-0.009}$
	3-4	13^{+9}_{-3}	$0.5^{+0.1}_{-0.1}$	
c	ext	–	$0.50^{+0.09}_{-0.09}$	–
d	1-2	19^{+13}_{-5}	$0.5^{+0.1}_{-0.1}$	$0.03^{+0.02}_{-0.01}$
e	1-2	23^{+11}_{-4}	$0.9^{+0.8}_{-0.3}$	$0.033^{+0.025}_{-0.009}$
	3-4	31^{+14}_{-5}	$0.8^{+0.6}_{-0.2}$	
f	1-2	30^{+14}_{-5}	$0.8^{+0.7}_{-0.3}$	$0.028^{+0.021}_{-0.008}$
	3-4	29^{+16}_{-6}	$0.7^{+0.5}_{-0.2}$	
g	1-2	28^{+16}_{-6}	$0.6^{+0.3}_{-0.1}$	$0.017^{+0.009}_{-0.004}$
	3-4	37^{+16}_{-6}	$0.5^{+0.1}_{-0.1}$	

Tab. 5: Summary of the cross-talk studies results, including main and induced (cross-talk) signal amplitude, triggered channels, and ratio between induced and main signal, for each setup configuration. Reported values represent the median of the distribution, with the interval containing ±34% of the distribution [144].



(a) (b)

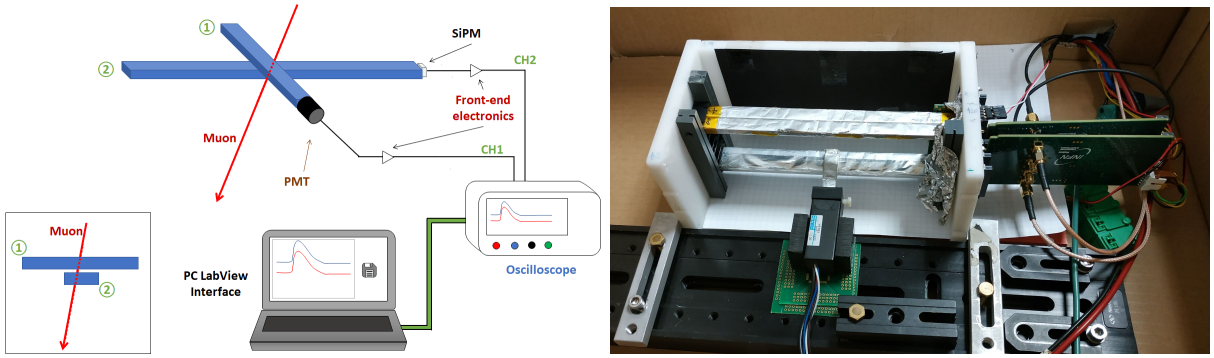
Fig. 64: (a) Signal amplitude of the main signal (blue points) and induced signal (red points), in the different acquisition configurations; the points relative to the induced signals are shown in a enlarged scale, to facilitate the visualization. (b) Ratio between induced over primary signals, in the various configurations. Data are summarized in Tab. 5 [144].

No additional teflon layers appear necessary. In any configuration not including an aluminum layer (c_a, c_e, and c_f) the induced signal is higher than the noise baseline, therefore these configurations were not considered acceptable. In particular the first configuration, c_a, produced the highest level of cross-talk, with $\approx 7\%$ of the main signal.
In the dataset c_b, the contribution to the cross-talk due to neighboring electronic components and adjacent SiPMs is considered, while the optical contribute was removed; in dataset c_d the SiPM were removed, therefore only the electronic contribute is present. The noise levels in this two configurations is compatible with the background measured in c_c, therefore the electronics

and SiPM-generated cross-talk can be considered negligible.
Once this wrapping configuration, including one teflon and one aluminum layer, was verified and accepted, the wrapping operations were performed manually on all the 480 scintillating fingers necessary to complete the calorimeter; to complete the task about 160 man-hours full-time were spent.

5.3 Verification on the response uniformity

Given the high aspect ratio of the scintillating element forming the calorimeter ($1\times0.5\times20$ cm^3), the uniformity of their response along the entire length is a critical aspect, as it ensure the response uniformity inside the calorimeter volume. The signal amplitude, in fact, should be independent from the particle detection position, to avoid possible deformations in the track reconstruction. Moreover, preliminary studies, based on the Monte Carlo simulation described in Sec. 4.4, showed a strong dependence of the signal amplitude on the position, with signals generated by particles passing closer to the SiPM being more than 40% higher than signals generated by particles passing at the opposite end of the scintillator. More details on this topic can be found in [144].
In order to characterize the response uniformity of the scintillating elements, verify the simulated results, and possibly tune the simulation parameter to match the experimental behaviour, the iLDA cosmic rays setup was adapted for position-sensitive measurements.



(a) (b)

Fig. 65: (a) Representation of the iLDA setup for position scans. The configuration includes a shorter scintillator (labelled 1) which can be moved along the length of a full-length scintillator finger (labelled 2), with the two scintillators placed perpendicularly to one another; the coincident signal between the two channels define the track position along the scintillator under study. (b) Picture of the setup in the described configuration.

In Fig. 65 the modified layout is shown. The setup includes two scintillators: one short-scaled scintillator (labelled 1 in the figure), read-out by a PMT, can be precisely moved along an aluminum rail to cover the entire length of a full-scale scintillator, labelled 2; the two scintillators are perpendicular to one another. The shorter scintillator is read-out by a PMT. The event recording was setup with a coincident trigger between the two scintillators, in order to limit the acquisition to particles intercepting the scintillators in the 1 cm^2 overlapping area, thereby selecting a defined position.
Fig. 66 summarizes the experimental results obtained from a position scan along the scintillator under study; the points represent the signal amplitude as a function of the position, expressed as the distance from the SiPM surface, measured as the central point of the perpendicular scintillator. The data points are represented as the fitted Landau distribution median, with the error bars marking the 68% interval. The experimental data show no noticeable dependence on

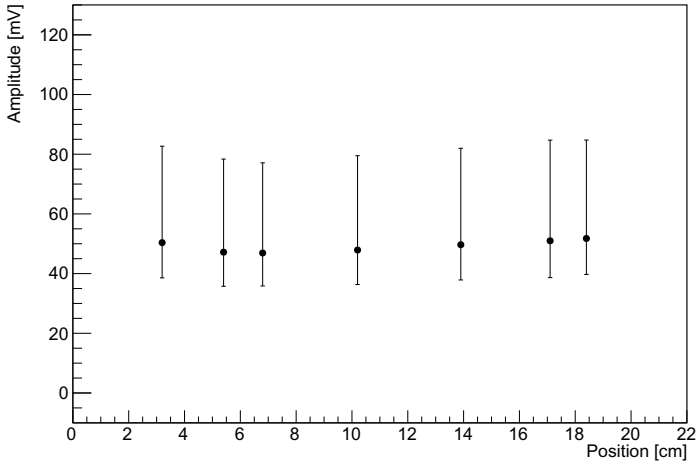
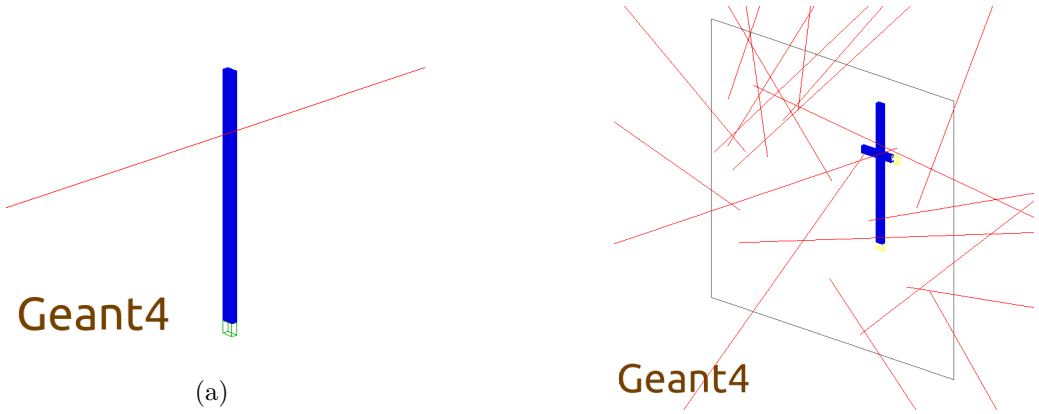


Fig. 66: Experimental measurements of signal amplitude as a function of the position along a scintillator.

the position, as all the points are compatible with each other within the error bars. Therefore, the experimental data dispels the concerns brought on by the preliminary simulation.
The parameters of the simulation were then adjusted, in order to obtain a better agreement with the observed behaviour; in particular the reflectivity of the internal teflon surface, and the scintillator material absorption length were both increased from the default values. The modified simulation layout is presented in Fig. 67 (a): a single scintillator is included, and the particle source was defined as a muon pencil beam; the muon energy spectrum was parameterized based on the muons spectrum, shown in Fig. 58 (a). The position of the muon source can be changed to span along the scintillator length and perform the positional scan.
A second simulation version was setup, to implement a more realistic modeling of the setup, as shown in Fig. 58 (b). The improved simulation included both the scintillators, and the particle source was defined as a plane emitting muons isotropically, in the direction of the scintillators; the same muon energy spectrum of the previous simulation was adopted. The realistic angular distribution of the muon (proportional to $\cos^2\theta$) was not taken into account, as the solid angle covered by the overlapping scintillators is limited.
Fig. 68 shows the results of the two simulation setups, compared with the experimental data. Both simulated data series present no strong dependence between the signal amplitude and the position along the scintillator, as observed in the experimental data. The data retrieved from the first simulation version (labelled *v1* in the figure) are systematically sightly lower than the experimental values, while the second version of the simulation reproduces the experimental data more consistently, not only in the trend but also quantitatively. In the first simulation the particles pass through the scintillator with a trajectory perpendicular to its surface; in the second simulation particles travel on average a larger distance inside the scintillating volume, as their are produced isotropically, and generate a larger energy deposition. The Monte Carlo data are overall in good agreement with the experimental results; this is a significant achievement, given that the simulated particles (muons with $E > 1$ GeV) are totally different from the particles used for the simulation calibration, i.e. protons in a totally different energy range (70 to 230 MeV).
The results presented in Fig. 68 confirm that the signal amplitude shows no dependence where the charge is deposited inside the scintillator, therefore the response in the entire iMPACT calorimeter volume can be considered uniform.
The Monte Carlo simulation, together with the analog signal reconstruction method (illustrated



(a) (b)

Fig. 67: (a) Rendering of the cosmic rays simulation in the first version, including a single scintillator and a pencil beam. (b) Layout of the second version of the simulation, including two perpendicular scintillators and a planar cosmic rays source [144].

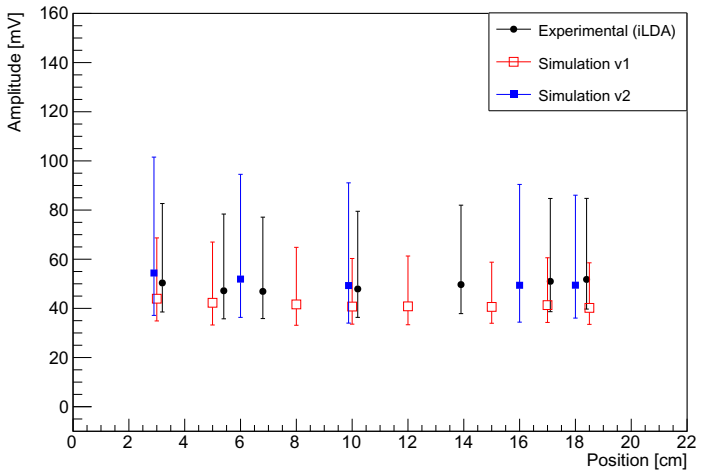


Fig. 68: Experimental measurements of signal amplitude as a function of the position along a scintillator, compared with simulated data; the two simulation setups are presented in Fig. 67 [144].

in Sec. 4.4), has proven to be particularly effective in reproducing both proton and cosmic rays datasets, in particular, accurately modeling the spatial response of the single scintillating elements forming the entire calorimeter.
In the first commissioning tests of the complete iMPACT scanner, cosmic rays will be employed for calibration and performances evaluations, see Sec. 6.6. The simulation will be used to scale the results and the calibration parameters obtained from cosmic acquisitions, to values suitable for proton acquisitions, in view of future test beams.

6 iMPACT digital read-out architecture

6.1 Calorimeter read-out structure

The task of interfacing the calorimeter analog and digital front-end electronics, which produce the data, with the computer in which the data is stored, is assigned to a number of FPGAs, in a single-master multiple-slaves structure.
The schematic of the read-out and control structure of the calorimeter is shown in Fig. 69. In this structure, only the slave FPGAs are connected to the front-end boards and can read-out data. Each FPGA can receive commands via a custom USB protocol from the computer (see Sec. 6.4 for the commands communication protocol); the master receives commands concerning the acquisition control, such as start, stop acquisition and reset. The master can drive 2 signals (`start` and `reset`), according to the particular command issued by the PC, in addition to a `clock` which is always generated, independently from the status of the acquisition. These 3 signals are then sent to a synchro-board, which clones and distributes them to each slave FPGA; this structure ensures the simultaneous reception of acquisition-control signals between the slaves.
Slave FPGAs can also receive commands via the same USB protocol that controls the master. The USB communication is, by design asynchronous and several orders (at least 100 times, depending on the task) slower than `start`, `reset` and `clock` signals that travel through the physical connections; thus only slow control commands without timing requirements are sent through the USB protocol. Via USB rather complex 32-bits commands can be transferred: commands to slaves are used to set the comparators thresholds to either pre-calibrated values, including different acquisition regimes (cosmic rays or proton beam), or to user-defined values; local FIFO memory resets can also be asserted. The computer controlling the acquisition can read-out the data stored inside each slave FPGA FIFO memory via the USB protocol using a custom-defined data packet formatting, which is described in Sec. 6.3.

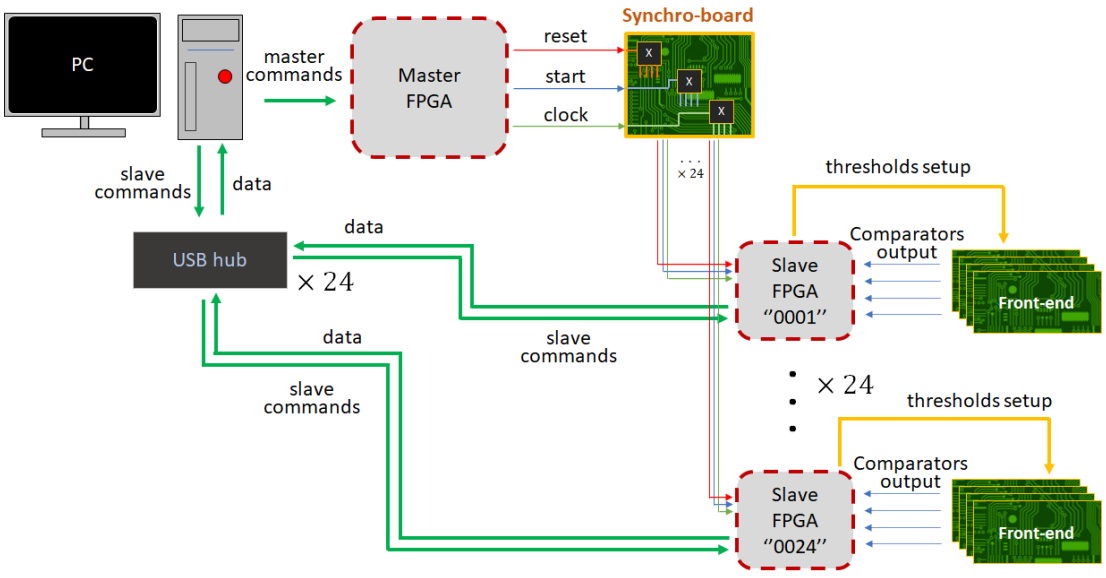
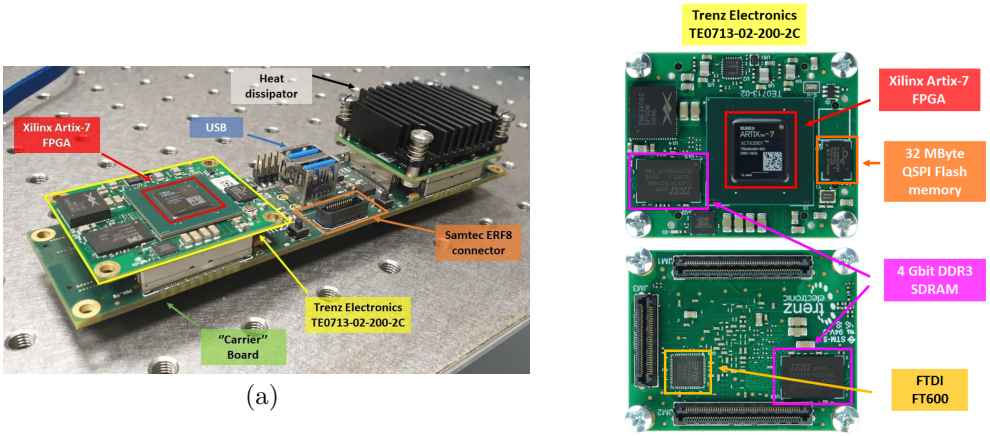


Fig. 69: Configuration scheme of the read-out and control structure of the iMPACT calorimeter. The read-out infrastructure is formed by a master FPGA which receives commands from a computer and generates two control signals (`start` and `reset` and `clock`) accordingly. The control signals are distributed to a number of slave FPGAs which perform the data read-out from the calorimeter front-end. All the FPGAs are synchronized by means of a `clock` signal. Only the main connections are presented in this scheme; for a complete diagram of each FPGA input and output signal see Fig. 72.

6.1.1 Calorimeter digital electronics

As presented in Sec. 4.2.2 a single front-end board is designed to read-out 10 scintillators channels, it is connected to the SiPM board and to the FPGA carrier board (the structure of a calorimeter module was shown in Fig. 34).



(a) (b)

Fig. 70: FPGA Carrier board. The board connects 4 front-end boards (Fig. 36) with 2 Xilinx Artix-7 FPGAs; it features a Samtec ERF8 connector [145], for the synchronization signals, 2 USB 3.0 connectors [146].

A front-end board employs a digital-to-analog converters on each side (Analog Devices AD5391 DAC [115]), to set the voltage references for the 15 fast comparators. The output voltage levels can be set through a serial interface compatible with the I^2C protocol; more details on the implementation of the I^2C communication for the digital threshold setup are presented in Sec. 6.5.
An FPGA carrier board, shown in Fig. 70 (a), collects the output of the comparators for the final data read-out and transmission; a carrier board hosts two FPGAs, each one responsible for the read-out of 20 channels. The Xilinx Artix-7 XC7A200T FPGAs are embedded in a commercial Trenz TE0713 module (Fig. 70 (b)), which includes auxiliary components, such as a 4 Gbit DDR3 SDRAM, a QSPI flash memory, and an USB interface FTDI FT600 chip [146]. The task of distributing the synchronous control signals (i.e. `start`, `reset`, and `clock`) from the master FPGA to the slaves is performed by the synchro-board, shown in Fig. 71. The synchro-board design is based on three CDCLVD1216 Low Jitter differential LVDS Buffers, one for each signal, from Texas Instruments [147]. The buffers replicate an input clock, up to 800 MHz, to 8 differential outputs with a lower than 300 fs RMS jitter; the input connector (from the master) and the output connectors (to the slaves) are Samtec ERF8 [145].

6.2 FPGA firmware overview

The structure and operations of the firmware, developed for the FPGAs controlling the calorimeter data acquisition, will be presented in this Section. After an overview of the input, output, control, communications and peripherals management signals, the general description of the firmware structure and functions will follow, before analyzing each firmware section and branch in-depth (see Sec. 6.3-6.5).
In the following sections, a group of parallel signals (or *bus*), will be indicated with the notation: `data[MSB:LSB]`, where the the most significant bit of the array is specified first and the least significant bit second. This notation will also be used to identify a sub-group of bits in a larger bus. Moreover, numbers in different bases will be expresses with the following notations: `0b` or

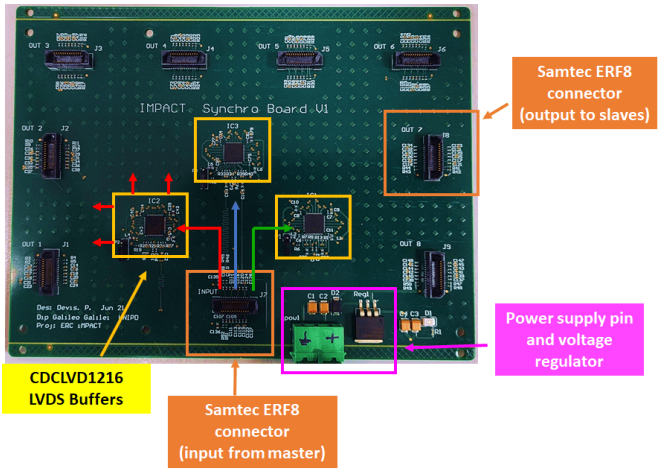


Fig. 71: Synchronization board (*synchro-board*), including 3 CDCLVD1216 Low Additive Jitter LVDS Buffer [147], used to multiply and propagate the `start`, `reset`, and `clock` signals up to 8 times. The signal input from the master FPGA and the output towards the slave FGPAs is performed through Samtec ERF8 connectors.

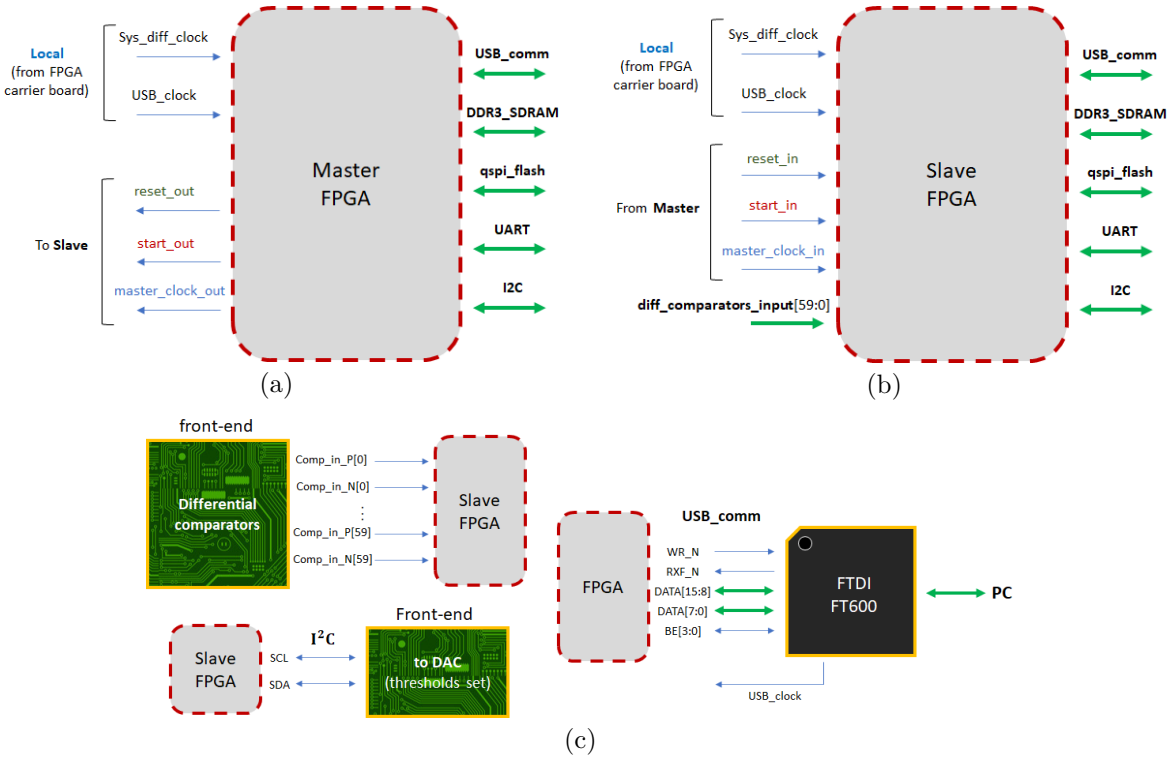


Fig. 72: Overview of the I/O ports of the calorimeter a) master and b) slave FPGAs; single signals are indicated by thin arrows, multiple signals (buses) by thick arrows. b) Detailed diagram of the main signal buses. Details on the USB communication protocols are available in [148]

$\ldots_2$ for binary, `0x/0h` or $\ldots_{16}$ for hexadecimal.
Fig. 72 shows an overview of the input and output signals, connected to the FPGAs, presented for both master (a) and slave (b) configurations; in Fig. 72 (c)) the main buses and groups of

signals are presented in more detail.
Each FPGA (regardless if slave or master) receives a differential clock (indicated in Fig. 72 as `Sys_diff_clock`) from their own Trenz TE0713 board (Fig. 70 b)); this clock is therefore not synchronized between FPGAs. The `Sys_diff_clock` is a 100 MHz clock that is used by each FPGA mainly for soft-processor (MicroBlaze) operations (see Sec. 6.5).
The `USB_clock` is a 100 MHz clock generated by the FTDI FT600 chip [148], installed in each FPGA carrier board; the syncrhonization with this clock is mandatory for the FPGA in order to communicate via USB through the FT600 chip with a computer. For this reason this `USB_clock` is also used by the firmware section responsible of data taking and formatting, as well as the section responsible of receiving commands from PC. This clock is also independent for each FPGA and independent from the `Sys_diff_clock`.
As introduced in Sec. 6.1, each slave FPGA receives a third clock, `master_clock` in the design Fig. 72, from the master. This clock is generated by the FTDI FT600 chip of the master, distributed via the synchro-board, and used by the slaves only for the internal timestamp counter, in order to ensure the timing synchronization between modules, and the correct reconstruction of the events recorded through different calorimeter modules. The signal `reset_out` is generated by the master FPGA, and set high when a reset of the timestamp counter is required by the PC. The signal `start_out` is again propagated from the master to the slaves (received as `start_in`) in order to issue the initiation of the data acquisition.
Each slave FPGA receives data from the front-end boards (see Sec. 4.2.2) as a couple of differential signals for each of the 60 comparators (20 scintillator fingers with 3 thresholds each) that a FPGA has the task of reading out, for a total of 120 signals (see Sec. 4.2.1). The duty of each slave FPGA is formatting, storing and eventually sending the data to the computer. The communication between the FPGAs and the computer is made possible by the FTDI FT600 chip, present in the FPGA Module Board TE0713-02-200-2C (see Fig. 70), which acts as interface bridge for both directions of communication. The USB bus (indicated in Fig. 72) as `USB_comm`) includes a 16-bit data bus and various signals (`WR_N`, `RXF_N` and `BE`) used for read/write requests and acknowledgments (more details on the USB communication will be presented in Sec. 6.4). The I^2C bus, introduced in Sec. 4.2.2, is used for setting the thresholds of the comparators present in the front-end board; it includes two signals: `SCL` (clock) and `SDA` (data). The `DDR3_SDRAM` bus establishes the communication with the RAM memory present on the FPGA carrier board; this memory is currently employed only by the MicroBlaze soft-processor, and its working operations; in future versions of the firmware the RAM is planned to be used also for data storage. The `qspi_flash` bus communicates with the non-volatile Quad SPI memory present in the FPGA carrier board; versions of firmware and software are stored in this memory and automatically booted at the device switch-on. Finally, the `UART` bus is a low-speed USB bus that can be used by the MicroBlaze processor to print strings onto a terminal in a computer; this link is usually used for debugging and status checks.
In Fig. 73 the block design of the firmware programmed in the FPGAs is presented, as visualized in the Xilinx Vivado Design Suite [149]. The block design of a firmware project provides a visual representation of the various logic modules, either proprietary or user-defined, and the signals that link them. The colours given to the blocks in Fig. 73 give an indication of the functions that a particular section or group of modules serve, although in some cases the categorization is not unique. The modules dedicated to the triggering and data processing are indicated by the blue colour; this section is formed by the front-end LVDS comparators input, input buffers, a 40-bit counter for the timestamp to associate to each recorded event, a State Machine controlling the data recording, and a data FIFO (First In First Out) where events are stored waiting to be transferred to the computer via USB (see Sec. 6.3 for more details). The State Machine that manages the transmission of commands and data via the USB FTDI FT600 chip to and from the computer is highlighted in green, together with its auxiliary modules (Sec. 6.4 for complete description). The orange blocks indicate the section responsible for receiving and actuating the acquisition-control and setup commands from the

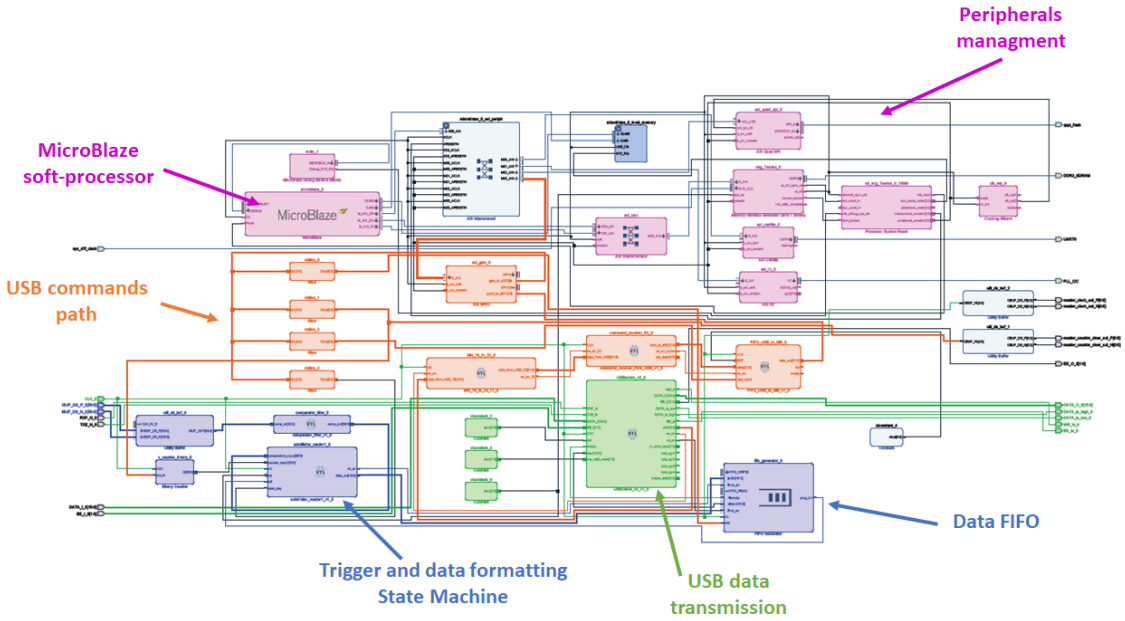


Fig. 73: General view of the calorimeter slave-FPGA Block Design, in Xilinx Vivado Design Suite [149]. Colours highlight different sections and functions: blocks related to the data flow, triggering and formatting are highlighted in blue; blocks related to USB transmission of data and the USB State Machine are highlighted in green; the orange section manages the commands reception from PC to FPGA; in violet, the sections related to the operations of the soft-processor MicroBlaze [150] [151] and peripherals controls (I^2C, `UART`, `qSPI` and `DDR3_SDRAM`).

computer; this sector is formed essentially by a 16 to 32-bit data parallelizer, a State Machine handling the command receiving protocol, and a commands FIFO (details in Sec. 6.4). Finally, the magenta modules and blocks are related to the soft-processor implemented in the FPGA fabric (MicroBlaze) [150] [151] and its operations; in particular, the soft-processor is used to handle the peripherals control, such as the I^2C to program the comparators threshold, and to actuate the actions stated in the USB commands. Using the MicroBlaze soft-processor in the calorimeter FPGAs firmware allows to split tasks on two levels: the tasks that need to be performed online, with strict time constraints, such as triggering and data formatting, are performed inside the FPGA fabric, where digital instructions can be executed in real-time at hardware speed; more complex tasks, easier to implement in a software environment, are instead handled by the soft-processor, which is intrinsically slower than the FPGA fabric. Additionally, the Xilinx soft-processor is equipped with a stable and reliable set of methods and functions which makes peripherals management more straightforward to program. As stated previously, the data modules and the MicroBlaze operate with two independent clocks in a parallel fashion. The functions and the operations of the soft-processor are illustrated in details in Sec. 6.5.

6.3 Trigger and data flow in the FPGAs

In Fig. 74 the block design of the firmware modules that are related to the acquisition of data are shown; the graphic representation of the design is produced by the Xilinx Vivado Design Suite [149], the software framework which allows to design and program Xilinx FPGAs. Each FPGA in the calorimeter is responsible for the read-out of 20 SiPMs, for a total of 60 voltage comparators (as seen in Sec. 4.2.2). The output of the comparators are fed into an FPGA as 60 LVDS differential signal couples; in Fig. 74 these input ports are appointed

as `Comp_IN_P/N` and are indicated as two 60-lines wide buses. An Input Buffer commutes the double-ended signals into single-ended signals, which then pass through a custom debug module (`comparator_filter`). In the current implementation, the debug module does not affect the signals and it simply propagates the input to the output; however, during the development of the firmware masking or swapping certain lines was performed to assess the correct behaviour of the downstream modules.
The Module indicated as `Binary Counter` is a Xilinx proprietary simple 40-bit counter triggered by the positive edge of the master Clock signal; as introduced in Sec. 6.1, the master Clock is the clock generated by the FTDI FT600 USB chip [148] of the master FPGA and propagated via hardware to every calorimeter slave FPGA, Fig. 69. With a 100MHz clock, the 40-bit counter overflows in about 3 hours, which is compatible with the duration of an acquisition with cosmic rays. The timestamp counter can be reset to 0 by a timestamp reset signal, also generated by the master. Therefore, the counter module provides the event timing information, which is synchronized among every slave read-out FPGA. The module Data FIFO is a memory queue implemented inside the FPGA fabric [152]; it provides ordered storage and retrieval of the events data. This module supports 16-bit words streams. This FIFO is the memory location in which data is stored locally on each slave FPGA, waiting for the acquisition computer to issue a read request and perform the transfer of data. The FIFO works with a single local clock, generated by each FPGA FTDI FT600 chip, therefore read and write are synchronous. The write (from FPGA into the FIFO) interface consists of a 16-bit bus (`din[15:0]`, data in) and a write command signal, `wr_en`; both the `din[15:0]` and `wr_en` signals are driven by the State Machine inside the `scintillator_reader` module (Fig. 77). The read interface allows for the retrieval of data from the local FIFO to the computer, and consists, similarly to the write interface, of a 16-bit bus (`dout[15:0]`, data out) and a read request signal, `rd_en`, which is issued by the State Machine managing the USB communication with the computer (see Sec. 6.4). Additionally, the FIFO can be reset synchronously, to delete its content, via the `rst` port, which is issued by a computer command and actuated by the Soft-processor MicroBlaze (see Sec. 6.5 for more details). Finally the FIFO module possesses two status signals: the `empty` pin, which flags when the memory does not contain any event, and it is interrogated during the data retrieval procedure; the `prog_full` pin, which signals when the FIFO contains a predetermined number of words, is checked before filling the memory with a new event. The need for a programmable threshold, instead of a basic `full` flag, is given by the fact that, in the present implementation, a single event is saved as triplets of sequential 16-bit words; this will be clarified throughout this section. Therefore the effective value of the FIFO depth has

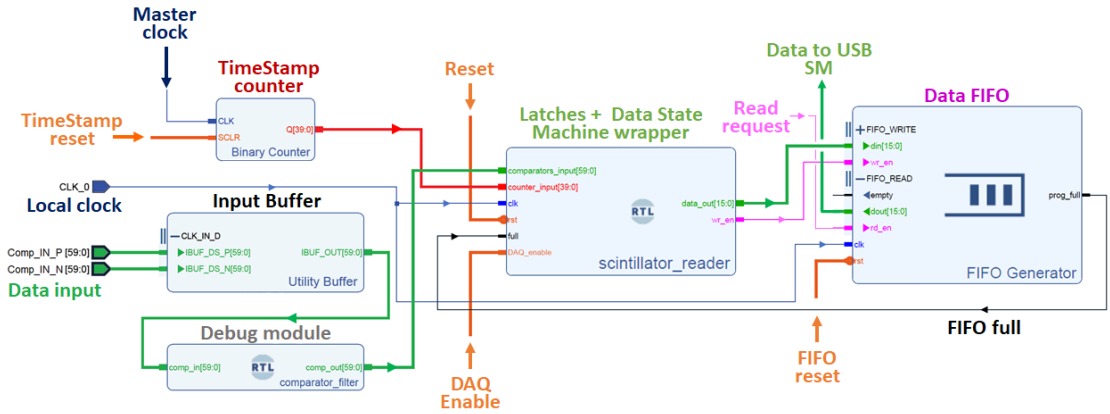


Fig. 74: Block Design Trigger and Data flow. The sub-structure of the module `scintillator_reader`, containing the latches for the SiPM signals and the Data State Machine, is shown in Fig. 75.

to be divisible by 3, as opposed to the maximum FIFO capacity which can be allocated only as powers of 2. Implementing an effective 3-multiple deep FIFO prevents cases in which the full FIFO capacity is reached, while having written a partial event; this would causes erroneous read-out operations, as the grouping on words into an event would be mis-aligned. In the present conditions, the FIFO memory is set to store up to 2001 16-bit words. Further events, after the set quantity, are discarded until data are read-out and a sufficient number of memory slots is freed. The order of magnitude of the FIFO depth was tuned to comply with the amount of data generated during an acquisition based on cosmic rays, allowing the acquisition control computer to retrieve the data sequentially from each FPGA avoiding any of the local memories to saturate.

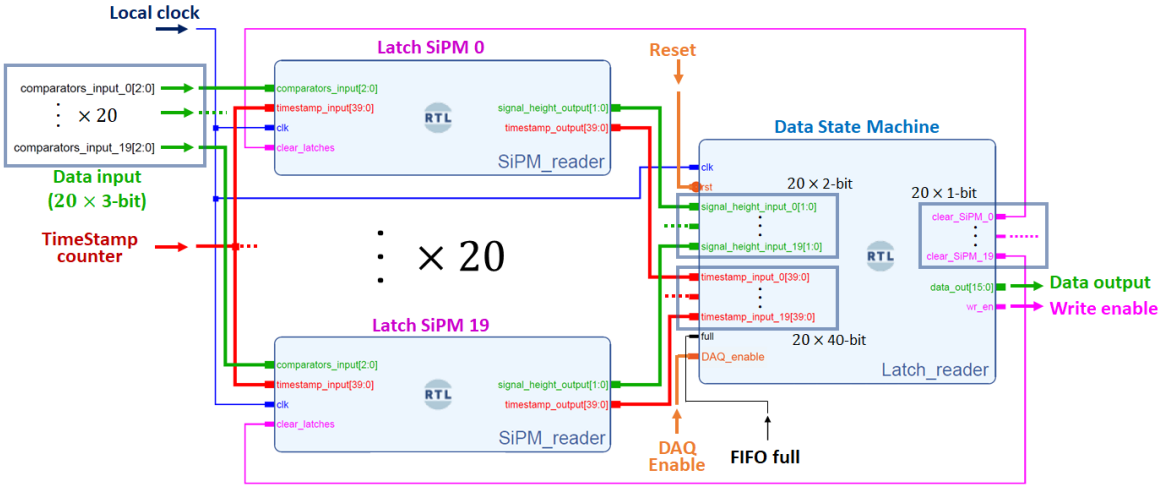


Fig. 75: Diagram of the `scintillator_reader` module, containing Latches (`SiPM_reader`) and data State Machine `Latch_reader`.

The central control module in charge of the data managing section of the calorimeter read-out firmware is the `scintillator_reader`. This module acts as a wrapper enclosing a number of sub-modules and its structure is schematized in Fig. 75. The `scintillator_reader` contains a total of 20 latch modules (named `SiPM_reader`), one for each SiPM, and a State Machine (named `Latch_reader`) that performs the actual read-out and reset of these latches, as well as managing the data formatting and the local data storage. Each `SiPM_reader` has as inputs (compare Fig. 74 and Fig. 75): the triplet of single-ended signals corresponding to the output of the 3 comparators reading a single SiPM; the 40-bit timestamp counter timed by the master Clock; the Local Clock and a `clear_latches` signal, generated by the `Latch_reader` State Machine, which resets a particular SiPM latch after a read-out.
In Fig. 76 the operations of a single `SiPM_reader` is presented, in the form of a behavioural logic simulation in Xilinx Vivado [153]. The `comparators_input` waveform (green) shows the typical behaviour of the input analog signal, sampled by 3 comparators, visualized with three thresholds. In the simulation, for each input signal the maximum height is randomized between 1 and 3, the distance between the signals is extracted uniformly between 0 and 50 ns, and the duration of the signal is generated in the range 3 to 10 ns. The rise and fall times are set respectively to 1 ns and 4 ns per exceeded threshold, to mimic the shape of an experimental signal.
At the point of the first marker, at 240 ns of the simulation, a `clear_latches` signal is received and the latch is released from the previous event; when no data is stored in the latch, the output timestamp is put at a defined value (all bits at 1 =`0xFFFF`), and the signal height

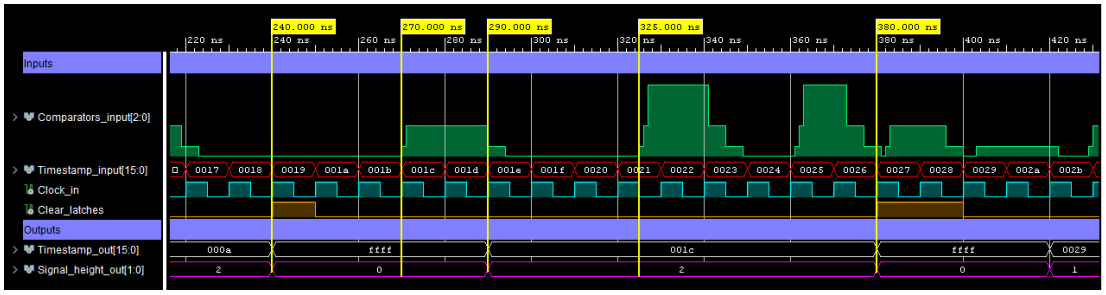


Fig. 76: Behavioural simulation of a single SiPM latch (`SiPM_reader`). The values in `timestamp_input` and `timestamp_out` are expressed in hexadecimal base for a more concise representation.

(`signal_height_output`) is set to 0. An event arrives at 270 ns, as soon as a comparator fires it is asynchronously latched, and the current timestamp saved; in this case the value `0x01C` ($= 28_{10}$) is saved. Then, starting from the first following positive edge of the clock, two clock periods are waited, in order to let the analog signal reach its maximum value; the highest exceeded threshold during this period is latched. In fact, experimental signals present rise times between 5 and 10 ns, shorter than the 2 cycles waiting window, equal to 20 ns, at 100 MHz. At the third cycle, at 290 ns, the stored values are presented as outputs: in this way the `timestamp_out` (grey) corresponds to the signal arrival time, while the `signal_height_output` (purple) reports the overall maximum signal height, which would occur most likely at a later timestamp. The height of the signal is output as 2 bits (up to `0b11` $= 3_{10}$), as opposed to the 3-bit comprators in input. The output values are maintained until a `clear_latches` signal is issued by the `Latch_reader` State Machine after a read-out of that particular `SiPM_reader`: if a second event takes place before a latch clear, as at the 325 ns marker, the latter signal is ignored even if higher than the first one. Having a waiting window of 2 cycles is completely sustainable when dealing with cosmic muons, the rate of which is expected to be around $2/s$ on a single scintillator, and feasible for proton beam acquisitions up to 20 MHz for each scintillator. The `Latch_reader` sub-module, contained in the `scintillator_reader` module, (see Fig. 74 and Fig. 75) receives the information on the signal height and timestamp from each of the 20 `SiPM_reader` sub-modules and it produces, as output, a 16-bit `data_out` bus, which is sent to the input the FIFO memory. Other inputs of the State Machines are: the local clock; the master Reset; the `DAQ_enable` signal generated by the master FPGA, which indicates the active acquisition status; and the FIFO `full` level, at which write operations are skipped. As additional output ports the `Latch_reader` transmits 20 `clear_latches` signals, each connected to a different `SiPM_reader`. The logic of the operations of the `scintillator_reader` State Machine is schematized in Fig. 77.

The `Latch_reader` Finite State Machine (FSM) currently operates over 5 states; at every positive edge of the local clock the conditions defining the state are verified, and possibly the state is changed. At the moment of power up the initial state is `STOP`: every output port and internal register is zero. If the acquisition enable signal, `DAQ_enable`, is low the `STOP` state is maintained. On the contrary, if a Enable DAQ command is received, the state transitions to `RESETALL`: a clear signal is sent to every `SiPM_reader` latch module, so any event possibly stored before the DAQ initialization is discarded; every other output and internal register is set to zero. The State Machine then transitions into the `STEADY` state, in which the `clear_latches` signal are brought down to zero, so that events can be stored and the acquisition can be initiated. The State Machine then enters into the `CYCLE` state, in which the actual read-out of the data is performed. While in the `CYCLE` state, the State Machine monitors the output of a single `SiPM_reader` sub-module, during each clock cycle; if no event is stored in that particular latch, namely the `signal_height_output` is zero and the timestamp is all 1 bits, the State Machine takes no action. During the following clock cycle, the next subsequent input channels

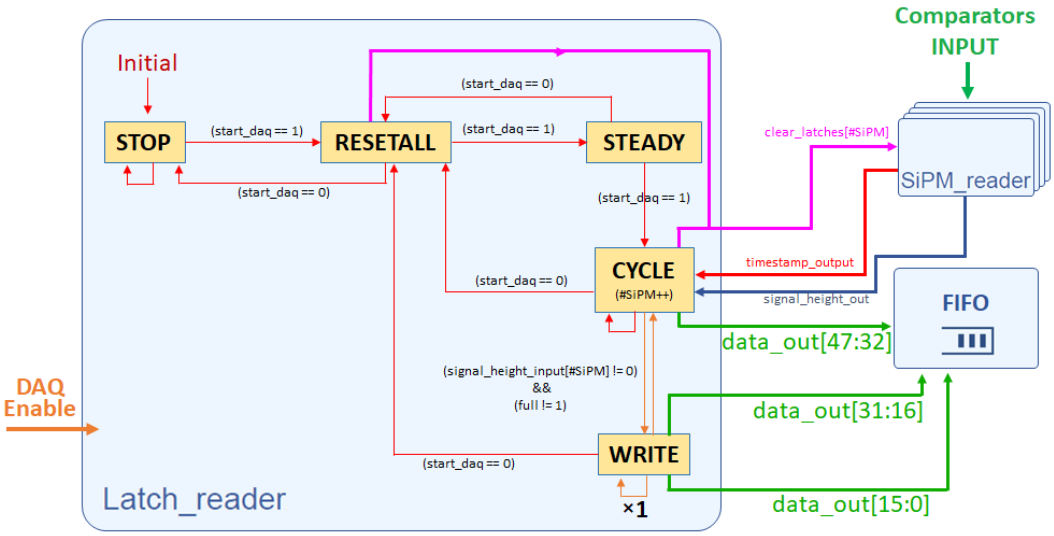


Fig. 77: Diagram of the states in the Data management State Machine. Yellow boxes indicate the states of the Finite State Machine; red thin arrows illustrates transitions between the states; thick arrows represent signals and data busses between the State Machine and other logic sub-modules, as `SiPM_readers` and the data `FIFO`.

is checked, and so on. Therefore, a complete monitoring of every SiPM requires, in the scenario of no recorded events, 20 clock cycles, for a total of 200 ns, which gives a 5 MHz frequency.
If the `SiPM_reader` that is being surveyed, in a particular iteration of the `CYCLE` state, presents a stored event, the State Machine initiate a write routine. The `signal_height_output` and `timestamp_output` values are stored in an internal register of the State Machine and arranged in the structure illustrated in Fig. 78.

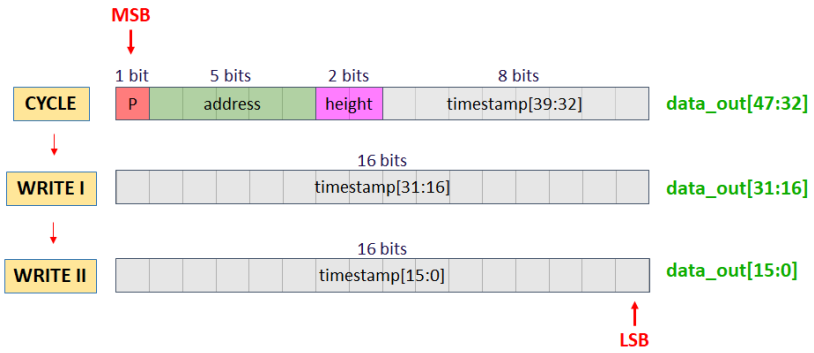


Fig. 78: Representation of the data transfer protocol via USB, based on 16-bit packets.

The event information is formatted into a total of 48-bit: the 40-bit timestamp is placed at the end of the pattern from the 9^{th} to the least significant bit (i.e. from `data_out[39]` to `data_out[0]`); the 2-bit `signal_height_output` takes the place of `data_out[41:40]` and it is preceded by the 5-bit address of the SiPM that has generated the event (values can go from 0_{10} to 19_{10} = `0b10011`). The Most Significant Bit, `data_out[47]`, is reserved for a parity bit, serving as a transmission integrity check, throughout the entire subsequent read-out chain.
The local memory FIFO can support only 16-bit words storage, so the event is sliced into 3 words that are sent during consecutive clock cycles. During the `CYCLE` state, in which an event is

found stored in a `SiPM_reader` sub-module, the upper snippet of that event (`data_out[47:32]`) is sent to the data FIFO; then state switches to the state named `WRITE`. The `Latch_reader` Machine stays in the `WRITE` state for two cycles, during which the following two remaining segments of the event data (`data_out[31:16]` and `data_out[15:0]`) are sent out. During the 3 clock cycle long data transmission, the write enable signal, `wr_en`, is set to high, to notify the FIFO that a valid word is meant to be stored; at the end of the second `WRITE` period this signal is lowered again. Additionally, during the `CYCLE` state, the latch corresponding to the SiPM that generated the event is cleared, via the specific `Clear_latches` signal; during the following `WRITE` state the reset signal is lowered, so that the `SiPM_reader` is ready to store a new event while the previous one is still being written.
As illustrated in this paragraph, an event write operation requires two additional clock cycles, in addition to the single initial cycle. In the worst case scenario, when every channel has stored events, a second consecutive read-out operation on the same SiPM would occur 60 clock cycles after the first one, for a total of 600 ns, thus a 1.7 MHz. This serial read-out mode is intended to be used during a cosmic rays acquisition.
Finally, at the end of a data acquisition, i.e. when the signal `DAQ_enable` is lowered, the state machine transits through the `RESETALL` state and ends back to the `STOP` state; the only exception occurs when the acquisition is ended while a write routine is in progress: in this case the complete event is written before proceeding to the `RESETALL` state, in order to avoid transmitting an incomplete data sequence.
For future tests and for the final application of the scanner with a proton beam, some adjustments on the data formatting will be made, in order to sustain the higher expected particle flux, with respect to cosmic rays. In the first place, the 40-bit timestamp counter could be reduced in length, to comply with acquisition periods of minutes, reducing the amount of data to be stored and transferred. Furthermore, in the current implementation the *Latch_reader* State Machine passes through each of the 20 SiPMs in a serial sequence, while in a proton beam environment a parallel read-out, with zero-suppression, would be preferable; in this second case the status of the fired channels could be transmitted at a given time, with the benefit of having to save only one timestamp. Considering a parallel read-out of the 20 SiPM, the read-out and reset of a hit SiPM can be expected to be performed in 5 clock cycles, allowing for a 20 MHz rate.

6.4 USB data transmission

The section of the firmware responsible for the transmission of data from and to the FPGA will be illustrated in this Section. The acquired data, which are stored temporarily in the FPGA fabric FIFO, are periodically transferred to the acquisition control Computer for longer term storage and analysis. At the same time, the computer can send streams of words to each FPGA, to assert a variety of instructions. For a schematic of the modules, connections and ports that make up the USB data transmission firmware section see Fig. 81.
The data transfer, in either direction, is performed via a USB bus, and it is managed by a FTDI (Future Technology Devices International) FT600 USB chip [148], present on the Trenz Electronics TE0713 board [146] (see Fig. 70). The FT600 is a USB to FIFO interface bridge chip, supporting USB 3.0 Super Speed (5 Gbps), USB 2.0 High Speed (480 Mbps) and a built-in 16 kB FIFO data buffer RAM. The bus protocol supports up to 8 channels (4 in and 4 out), however in this application only one channel is used. The FT600 operates with a 100 MHz clock, generated internally, which is also made available on an output pin, to provide synchronization with the connected peripheral. As introduced in Sec. 6.2, the data reading section of the firmware on each FPGA is also fed on the FT600 clock.
The FT600 USB bridge operates as an interface between the FPGA and the computer. The chip acts as a slave bus, while the FPGA performs master operations. Likewise the FT600 chip, the computer assumes the part of the bus slave: read or write requests by the computer are received by the FT600 and passed on, via appropriate signals, to the FPGA, which then can initiate the transmission, as the bus master.

As introduced previously in Sec. 6.2 (in particular Fig. 72), the USB bus includes the signals:

- `WR_N` is the bus master (FPGA) to bus slave (the computer via the FT600 chip) data transaction request signal, for either read or write, and it is active low;
- `RXF_N` is the bus slave to bus master data receive acknowledge signal, also active low;
- two 8-bit data buses `DATA[15:8]` and `DATA[7:0]`, that are used as the 16-bit data bus during the data transfer phase; in the idle state `DATA[15:8]` is driven by the slave to provide the FIFO status to the bus master, the upper nibble (`DATA[15:12]`) provides the 4 OUT channels FIFO status while the lower nibble (`DATA[11:8]`) provides the 4 IN channels FIFO status. They are all active low. `DATA[15:8]` are used by the FT600 chip to communicate to the FPGA write or read requests from the computer;
- `BE[1:0]`, the Read/Write command: it is driven by the FPGA at the transmission initialization as either `0b00` or `0b11` indicating a master read or write respectively; additionally the `BE[1:0]` bus is driven by the slave (FT600) to indicate a valid word, during a master read sequence;

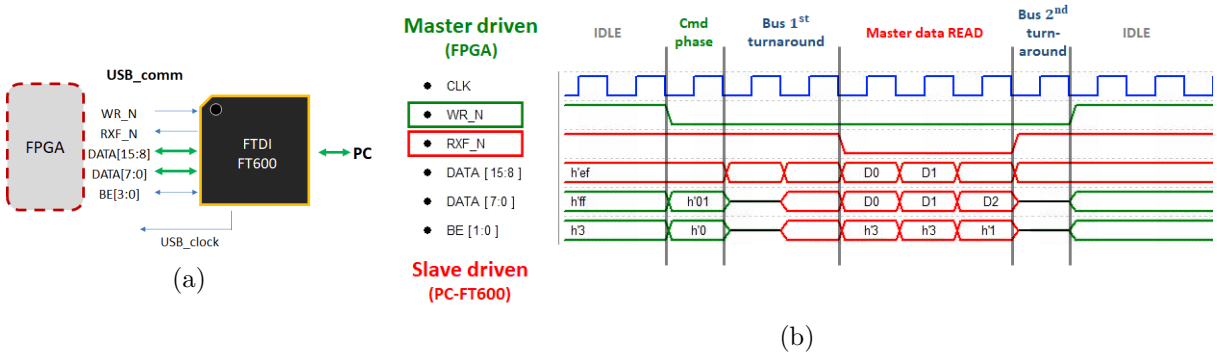


Fig. 79: (a) Remind of the FTDI FT600 USB chip communication interface (b) Sequence of a master read transaction: the FPGA reads data sent by the computer.

In Fig. 79 (a), the structure of the FTDI FT600 communication interface is recalled, while in (b) an example of read operation from the PC (the bus slave) to the FPGA (bus master) is presented. This communication protocol defines the signal sequence that the USB chip expects to observe, in order to perform a successful data transfer; it is presented in details, as it defines the behaviour of the FPGA firmware and it has a direct impact on how the data formatting and transmission need to be structured, and ultimately it has an immediate repercussion on the achievable data transfer rate.
As illustrated in Fig. 79(b), in the Idle state, the bus master controls the `WR_N`, `DATA[7:0]` and `BE[1:0]` all at a high state (active low); while `RXF_N` and `DATA[15:8]` are driven by the bus slave; the upper 4 bits of `DATA[15:8]` are set to `0hE` = `0b1110`, indicating that some data can be read from the channel 1 of the slave. The master can drive `WR_N` low to command a transmitting sequence, setting at the same time in `DATA[7:0]` the addressed channel number, (`0h01` for channel 1 in the example), and `BE[1:0]` = `0b00` to indicate that the requested transmission is a reading, i.e. data from slave to master. The command phase is followed by an interval where the control of data buses and `BE[1:0]` is released by the master and acquired by the slave, to allow the latter to write its data; the end of this turnaround phase is confirmed by the slave, via the lowering of the `RXF_N` line. For each following clock cycle, a total of 16 bits are written in the data buses; the 2-bit `BE[1:0]` is used by the slave to ensure that each 8-bit half-word is valid, asserting a 1 value in the corresponding bit. The transmission can be interrupted by either the master, bringing the `WR_N` to a high state, or by the slave, when for

instance there are no more data to transmit, via the `RXF_N` signal. The control of the buses and signals are then handed back to the original owner, and the interface status returns to the initial idle state.

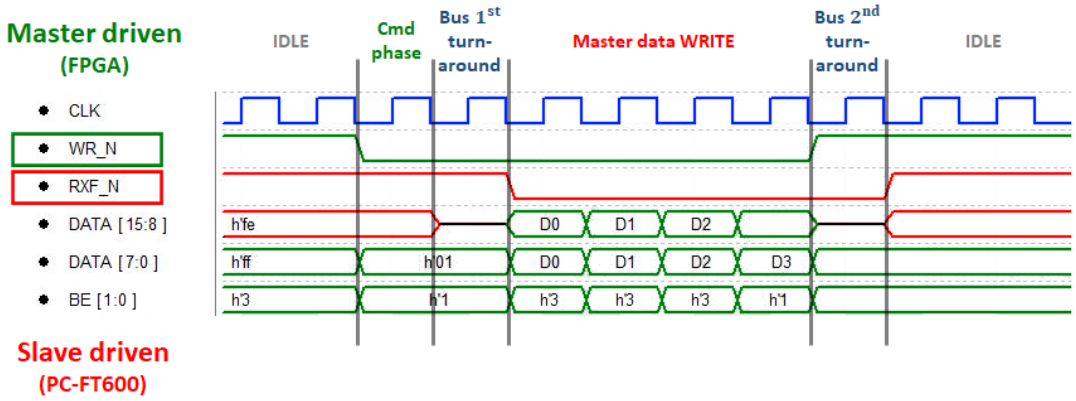


Fig. 80: FTDI FT600 master write sequence

The USB master write sequence is illustrated in Fig. 80: it is executed when data are transferred from the FPGA to the Computer. The master write protocol is similar to the master read, presented previously; in the case in Fig. 80, during the idle state the `DATA[15:8]` bus is driven by the slave to `0hFE` = `0b1111 1110`, which indicates that the computer is requesting to receive data on the channel 1. The FPGA can issues a transmission by setting `WR_N` low and `BE[1:0]` to `0x01`, to initiate a writing sequence. The appropriate turnaround then occurs: the bus master maintains the control over `DATA[7:0]` and `BE[1:0]`, while acquiring the control on `DATA[15:8]`. The data buses are then filled with the data to transmit while the `BE[1:0]` bus indicates the validity of the delivered words. At the end of the transmission, the USB bus returns in the idle state.

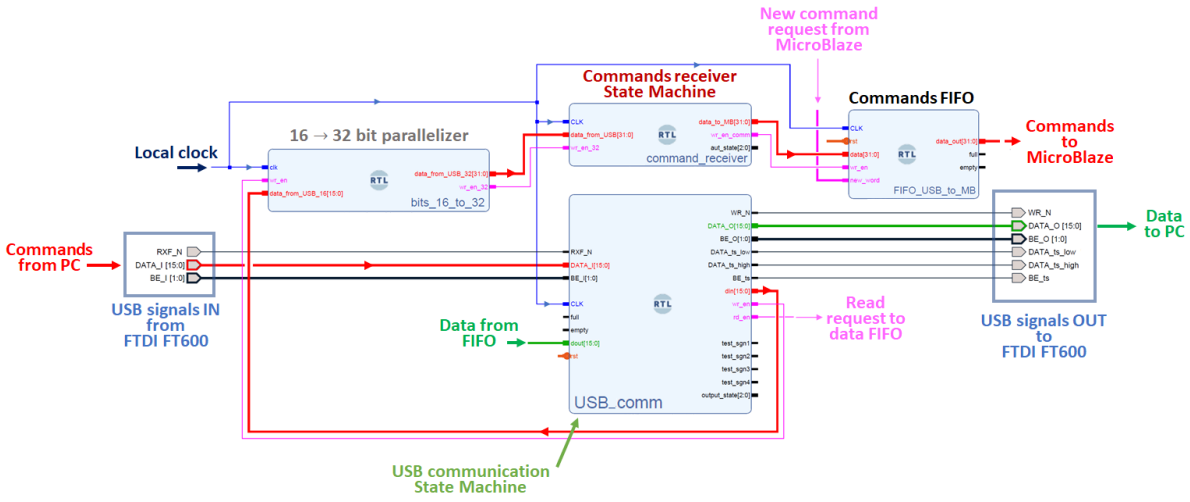


Fig. 81: Block Design USB State Machine. USB bus signals are enclosed in grey boxes.

In the FPGA firmware, the module which acts as the master of the USB bus and initiates the data transmission, to and from the computer, is named `USB_comm`, as illustrated in Fig. 81. The scheme the `USB_comm` State Machine is connected to a series of input and output ports

that correspond to the signals needed for the USB communication. Bidirectional signals are present separately as an input and output ports, generally indicated with the suffix `_I` and `_O` respectively, with the addition of a corresponding number of output ports marked with `_ts` (tri-state). The two mono-directional ports are then managed in a logic wrapper, that encloses all the logic HDL (Hardware Description Language) modules, which acts as a tri-state switch (for an example see Fig. 82): the `DATA_OUT[7:0]` and `DATA_IN[7:0]` are recognized by the FPGA as two separate mono-directional signal buses, which are alternatively connected to the bi-directional line by a switch node, controlled by a binary signal (`DATA_ts_low`) generated by the State Machine, based on the direction required in a particular transmission sequence (see Fig. 79 and Fig. 80). All the bi-directional lines are managed in an equivalent way.

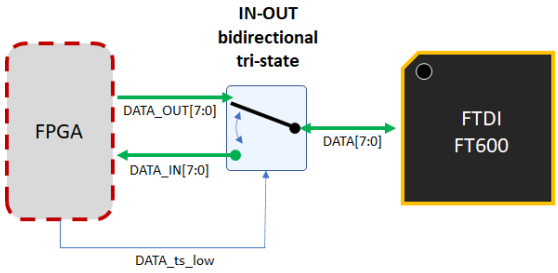


Fig. 82: Implementation of a in-out bidirectional port using a tri-state switch; in the FPGA firmware an inout port is described

Connections between the `USB_comm` State Machine module and the data FIFO are defined (see also Fig. 74) as: 16-bit `Data_out` from the FIFO, containing the events to send out; read request signal to the FIFO (`rd_en`); `empty` flag, indicating that no data is present in the memory.
At the same time, the words that the `USB_comm` State Machine receives from the computer are moved from the FPGA fabric, in the form of a 16-bit bus (red arrow in Fig. 81) and a write enable signal (violet arrow). Defined sequences of words received by an FPGA, through the USB connection, describe commands, for acquisition setup, resets and acquisition status. These words sequences are managed by three HDL custom modules: a bit parallelizer named `bits_16_to_32`, `command_receiver` Finite State Machine and a Commands FIFO.
The operating principle of the parallelizer module is illustrated in Fig. 83 (a). The parallelizer module receives from the `USB_comm` state machine the 16-bit packets during succeeding clock cycles; a valid packet is indicated by the write enable `wr_en` signal as input. The parallelizer module then stacks a pair of consecutive 16-bit words into a single 32-bit word, while generating as output another write enable signal, to indicate to the subsequent module that a stitching has been completed and a word is ready to be taken over. The following `command_receiver` State Machine, receives the 32-bit words, and checks if their sequence complies with the defined commands custom protocol, illustrated in Fig. 83 (b): the transmission sequence starts with two specific handshake words, the second one containing the number of commands to be expected in that particular communication round. Subsequent words define the actual commands, which can be either simple statements or more articulated parametric commands encoded in the 32-bit structures. Different commands are intended for the calorimeter master or slave FPGAs. A list and description of the currently supported commands is presented in Section 6.5.
After the expected number of commands are received, the communication is closed by two 32-bit footers; if the entire sequence is compliant with the protocol, the `command_receiver` State Machine moves the actual commands into a FIFO. The Commands FIFO is polled by the embedded microprocessor, via a new-command-request signal, and then parsed (see Sec. 6.5 for this procedure).
The command transmission protocol requires a relevant number of protocol words, in order to dispatch a single command: in the worst case, for a 32-bit meaningful packet, 4 times ancillary

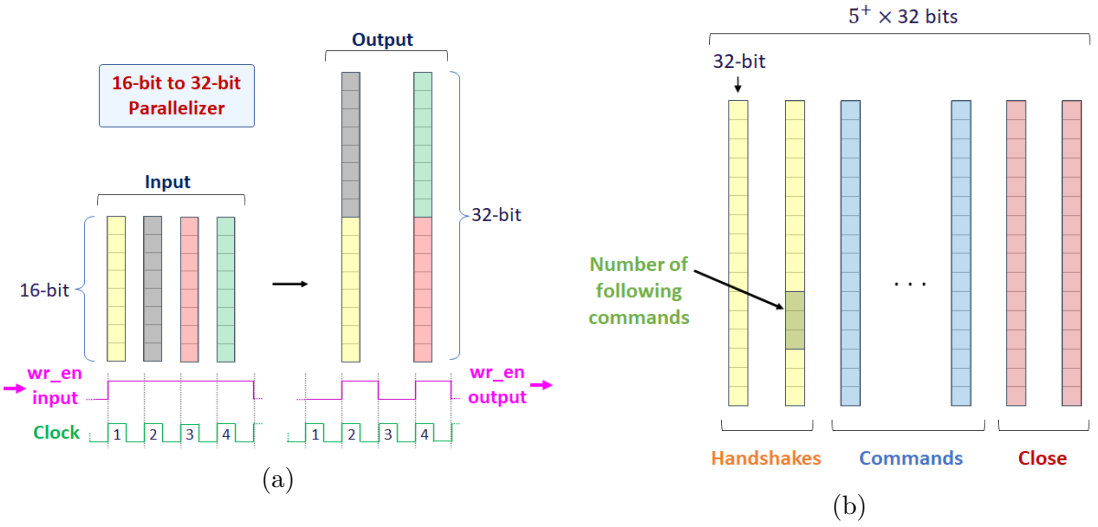


Fig. 83: (a) Schematization of the 16-bit to 32-bit parallelizer operation (b) Sequence of 32-bit packets forming a USB commands protocol transmission.

data are required for handshakes and transmission closing. The Embedded Microprocessor, however, requires about 0.7 µs to read a command and perform the required task, thus the transmission over-time due to ancillary words results essentially negligible. On the other hand, the redundancy of the protocol makes it robust against errors and/or transmission glitches.

6.5 MicroBlaze soft-processor operations

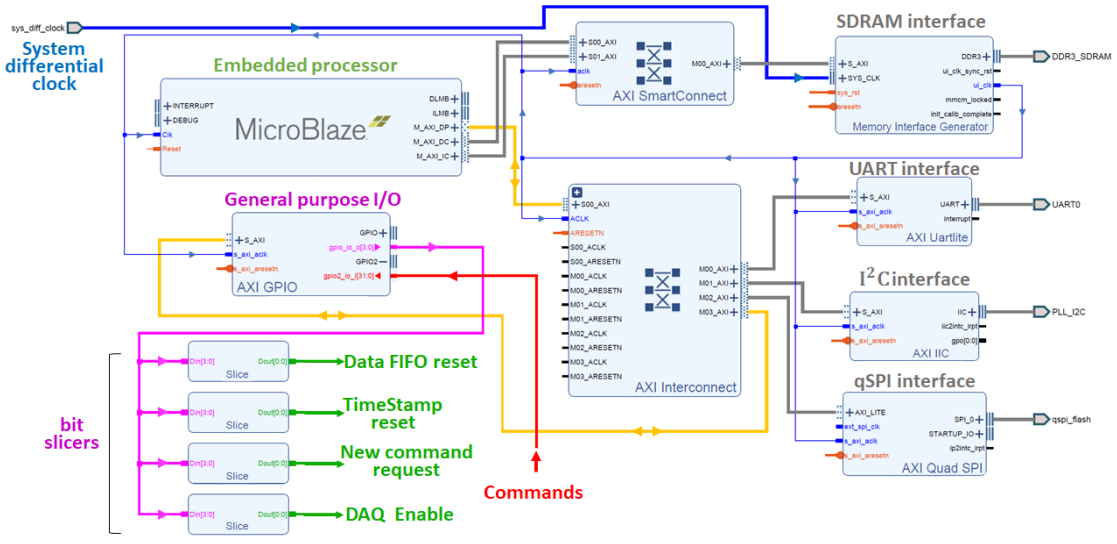


Fig. 84: Block Design relative to the Microblaze soft-core processor, and peripherals management blocks.

The whole communication firmware is implemented as logic modules in the matrix of the FPGA. The modules operate in a highly parallel and synchronized structure, driven by the common

100 MHz clock. Most of the logic operations are well-suited for an implementation through a Hardware Description Language (HDL), as they can be reduced to a linear combinations of logic or boolean operations, bits manipulations and hardware digital input/output management. The more complex tasks necessary to perform a complete acquisition management and system setup are more easily implemented in a software environment. Software programs, executed in a processor, present significantly poorer performances in terms of execution speed with respect to digital circuitry, as well as limited parallel and time-critical synchronization. For these reasons the firmware operates on two levels: the data flow, from the front-end electronics to the storage computer, is completely implemented in the FPGA fabric, ensuring read-out speed, transmission bandwidth, rigorous parallelism and precise timing. More complex management and setup tasks are instead performed and handled in the software environment, running on the soft-processor. The processor and logic levels operate asynchronously, communicating through a set of interconnecting signals between them, such as the 32-bit commands bus or the `DAQ_enable` and `reset` signals, introduced earlier.
The soft-processor is not a separate physical chip, but it is implemented into the FPGA fabric itself. The embedded soft-core processor MicroBlaze is designed by Xilinx to exploit the logic fabric and the general-purpose memory available in the FPGA [150] [151]. The Microblaze embedded processor is a Reduced Instruction Set Computer (RISC) core. A RISC architecture is optimized to operate with a large number of registers, and supports a small set of highly optimized instructions, ensuring a low number of clock cycles per instruction. The MicroBlaze can, in many applications, maintain a single-cycle per instruction regime. An instruction is a basic action, such as a mathematical operation, logic operation, or writing/reading an output/input register. Elementary operations usually require few cycles: for instance, an addition, a logic AND or a comparison between numbers, require a single cycle; reading or writing a streaming or cache register require between 1 and 3 cycles, depending on optimization parameters. On the contrary, more complex instructions need more cycles to complete: a Floating Point Arithmetic Square Root or Division require, for example, up to 30 cycles. A software function or API normally consists in a number of basic instruction, hence, depending on the complexity of the function itself, the Microblaze can operate, usually, only at a fraction of the FPGA fabric speed.
Being implemented in a programmable logic matrix, many parameters of the MicroBlaze can be configured by the user: cache size, pipeline depth, embedded peripherals, memory management unit, and bus-interfaces can be customized. Compilers for `C` and `C++` producing the assembly code for the MicroBlaze are available, allowing for a variety of embedded applications. MicroBlaze's primary input/output bus is a master–slave capable transaction bus with memory mapping, named `AXI` interconnect (Advanced eXtensible Interface) [154].
In the FPGA market, nowadays, there are many available products, commonly referred to as Systems on a Chip (SoC), that include in the same board an FPGA and a physical processor, usually ARM architecture based. Having a separate physical processor improves the performances over a softcore processor implemented in the FPGA matrix. Examples of SoC are the Intel Altera Cyclone family [155] and the Xilinx Zynq range [156]. For the iMPACT calorimeter, it was chosen not to employ such devices, naturally more powerful but at the same time more expensive, as the need for particularly complex on-line calculations was not so stringent.
In Fig. 84 the logic structure of the firmware section concerning the Embedded Microprocessor and the peripheral management is schematized. Different peripheral modules, for `SDRAM`, `UART`, `I`^{2}`C` and `qSPI`, are present; these modules are connected to the Microblaze Embedded Processor module via `AXI`-type links, passing through interfaces multi-plexing modules (`AXI SmartConnnect/Interconnect`). Peripheral modules allow to control and manage various kinds of devices and external integrated circuit chips, present on the FPGA board, which become accessible by software programs running in the Microblaze processor, via device-specific APIs and functions.
The `DDR3-SDRAM` interface module, in Fig. 84, is a Memory Interface Generator (MIG) [157],

which serves as a controller for interfacing FPGA designs and AXI slave interfaces to DDR3 SDRAM devices (Synchronous Dynamic Random-Access Memory). This module allows the Microblaze processor to access the 4 Gbit volatile DDR3 SDRAM component present in the Trenz TE0713 FPGA board (see Fig. 70 in Sec. 4.2.2).
The qSPI interface is a module that translates AXI protocol into quad SPI-protocol. The SPI protocol (Serial Peripheral Interface) is a synchronous serial communication interface, used for short-distance communication, developed by Motorola in the 1980s. The Quad version is an enhancement of the original SPI protocol, which allows for wider data bus. The qSPI interface communicates with the 32 MByte QSPI Flash memory present in the TE0713 board (Fig. 70). On-board QSPI flash memory is used to store the initial FPGA configuration, which loads automatically at the device power-up; it contains both the firmware configuration (bitstream) and the software program that runs in the soft-processor. This automatic setup avoid the need of re-programming each FPGA after a reboot, and it is essential for the practical use of the apparatus.
The UART-AXI (Universal Asynchronous Receiver-Transmitter) interface is used by programs in the Microblaze to print information that can be read on a computer terminal. The UART protocol is serial and asynchronous, the data format and transmission speeds (baud rate) are configurable; the data stream is let out through a USB cable, which can be connected to a computer serial port and decoded by a serial receiver software (such as Tera Term [158]). The UART connection is used in the developing phase for single modules debugging purposes, while during acquisitions it is employed to monitor the acquisition status and to receive feedback and reports from the master FPGA.
An additional interface module is represented by the General-Purpose Input Output (AXI GPIO) module [159]. Through this module the MicroBlaze can communicate to and from the FPGA logic fabric. The GPIO acts as multi-channel 32-bit interface between the FPGA fabric lines and the AXI bus, which is then accessible to the Microblaze. The GPIO lines are accessible from a Microblaze software via polling of the ports: in order for a program running inside the processor to have access to the value of the GPIO input, a read function must be called. In fact, changes in the values of the input GPIO lines can not be detected by the software. Furthermore, as introduced earlier, the speed at which the Embedded processor executes routines is orders of magnitude (microseconds vs nanoseconds) slower than the FPGA clock, so any value presented from the fabric to a GPIO input has to be latched, to allow the software to perform a poll of the input port and retrieve the value. In the same fashion, output values from the embedded software to the FPGA fabric are set by calling a read API from inside the Microblaze. The GPIO input port accepts the 32-bit wide commands that the FPGA receives from the computer, via the protocol illustrated in Sec. 6.4. Currently supported commands are:

- EnableAcquisition (master only), it commands the master FPGA to start the acquisition, setting the start line to a high state for the duration of the acquisition; this signal is then time-deterministically propagated in parallel to every slave FGPA (as schematized in Fig. 69); the start signal corresponds to the DAQ_enable on the data acquisition State Machine Fig. 74;
- ResetCounter (master only), it commands the master FPGA to drive the reset signal, which resets to zero the timestamp counter in each slave FPGA; this command should be called before an acquisition start to synchronize every module of the calorimeter;
- StartDataTaking (master only) is a composite command containing the previous two operations, to setup and initiate an acquisition;
- StopAcquisition (master only), it commands the start line to a low state, to signal the end of the data taking period;
- ResetFIFO (slave only), it resets the content of the data storage FIFO in a calorimeter slave FPGA;

- `SetSingleThreshold` (slave only), this command is sent to a slave FPGA, which communicate with the front-end boards, to set a particular DAC threshold to the transmitted value. In the 32-bit command the 12-bit DAC value and the 6-bit comparator address (from 0 to 59) are encoded;
- `SetAllThresholds` (slave only), it request the setup of all the thresholds of a given slave FPGA to the same value, which is transmitted in the command;
- `SetAllThresholdsDefault` (slave only), with this command the comparator thresholds of a slave FPGA are set to the values which have been pre-calibrated singularly for each channel (different default values can be set for either proton or cosmic rays acquisitions); it consists, in practice, of a series of single `SetSingleThreshold` commands, where the values are stored in calibration tables in the Computer memory;
- `StopAndErase` is another composite command stating an acquisition stop and a complete data memory reset; it would be called in case of errors.

The software program inside the Microblaze cyclically interrogates the input for a command, and, once a command is received, the corresponding action is started. The execution of the commands is actuated by setting to a particular value the 4-bit output port of the GPIO: the most significant output bit indicates the reset of the Data FIFO; the next bit assert the reset of the timestamp counter, and it is propagated simultaneously to every slave FPGA in the calorimeter; the third bit requests a new command from the Command FIFO once the previous command has been carried out; the last bit corresponds to the DAQ enable signal. The 4-bit output value is sliced into the individual bits, which are then connected to the corresponding modules or external outputs.
The commands initiating a DAC threshold setting do not use the 4-bit `GPIO` output; instead, the `AXI-I2C` interface is employed, to control the Analog Devices AD5391 DAC. In order to program a certain voltage threshold the communication sequences presented in Fig. 85 occurs [115]. The I^2C transmission, as introduced in Sec. 4.2.2, runs on two bidirectional lines: `SCL` and `SDA`. `SCL` acts as a clock, driven by the master device (in this case the FPGA) and defines the transmission speed: one data bit is transferred per `SCL` cycle. The `SCL`-clock is driven only when an operation is in execution; when the bus is not busy both signals remain pulled in a high-state. `SDA` is the data line and can be driven by either master or slave, depending on the data direction. The data on `SDA` must remain stable during the high period of the `SCL` clock pulse. Changes in `SDA` while `SCL` is high determine control signals, as START and STOP. In this application the I^2C communication operates with a data transfer rate of 100 kbit/s and voltage reference of 2.5 V. The DAC implements two hardware external pins, `AD0` and `AD1`, reserved for the device addressing on the I^2C bus; setting on the hardware either pin to a low or high state allows a single FPGA to address up to 4 DACs. Therefore, in the iMPACT front-end configuration, the I^2C bus comprehends an FPGA, which acts as the I^2C-master and 4 AD5391 DACs as slaves.
There are three specific modes in which data can be written to the AD5391 DAC: 4-byte, 3-byte and 2-byte modes; in the present configuration of the first mode is employed. A 4-byte mode transmission to set a DAC analog output voltage is outlined in Fig. 85; in this mode, for each threshold, the DAC I^2C bus address, the threshold channel pointer and the output value have to be transmitted.
To open a transmission to the AD5391, the START condition is issued, identified by a high-to-low transition on `SDA` line, driven by the master, while `SCL` is high. The Start condition is followed by the 7-bit slave address on the I^2C bus, formed by a fixed sequence `0b10101` plus the two address bits `AD1` and `AD0` of the AD5391; practically the 4 ADCs on the same bus are identified by addresses from `0x54` to `0x57`. The 8^{th} bit of the first byte is a read/write bit (`R/`$\bar{\text{W}}$) set to 0 to indicate a write request from the master. If the addressed DAC on the bus receives the correct sequence, it generates an acknowledge bit (`ACK`) pulling the `SDA` low during the ninth clock period, and the master then regains control of the line.

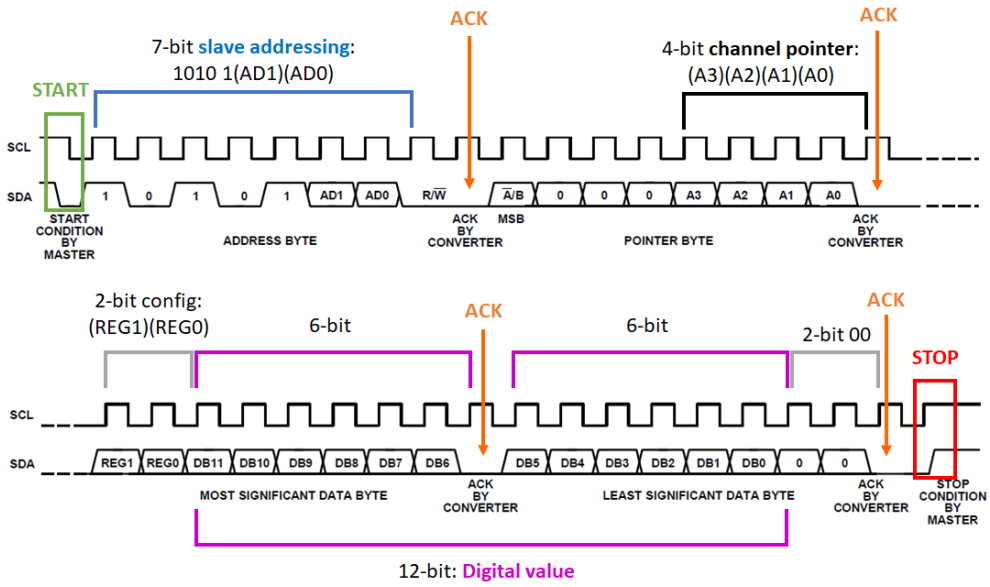


Fig. 85: AD5391 4-Byte Mode I2C Write Operation [115]

The first bit of the second transmitted byte (`A/B`) in standard operations is default at 0, followed by 3 other 0 bits; the byte is completed by the 4-bit pointer to the specific DAC channel to set. The correct sequence is acknowledged back by the DAC. The address and the pointer byte are followed by two bytes in which the digital value of the threshold is specified. In the AD5391 the 12-bit digital value is split in half, it is preceded by two bits, `REG1` and `REG0`, which when set to `0b11` allow to access the output data register. The analog output voltage is related to the digital value written in the register as:

$$\text{Analog Output Voltage} = 2 \times V_{ref} \times \frac{\text{(12-bit Digital Input Value)}}{2^{12}} \tag{32}$$

with $V_{ref} = 2.5$ V. After the reception of each byte the DAC responds with an acknowledge bit. A STOP condition, corresponding to a low-to-high transition on the `SDA` while `SCL` is high, releases the bus, allowing to initiate further transmissions. In Fig. 86 a single threshold set transmission between an FPGA and a AD5391 DAC is showed, recorded by an oscilloscope. The data transfer divided into 4×8-bit bursts is visible.

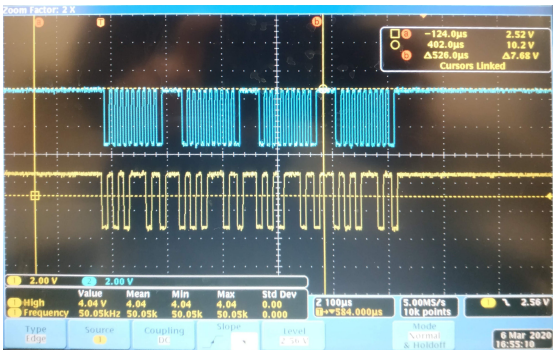
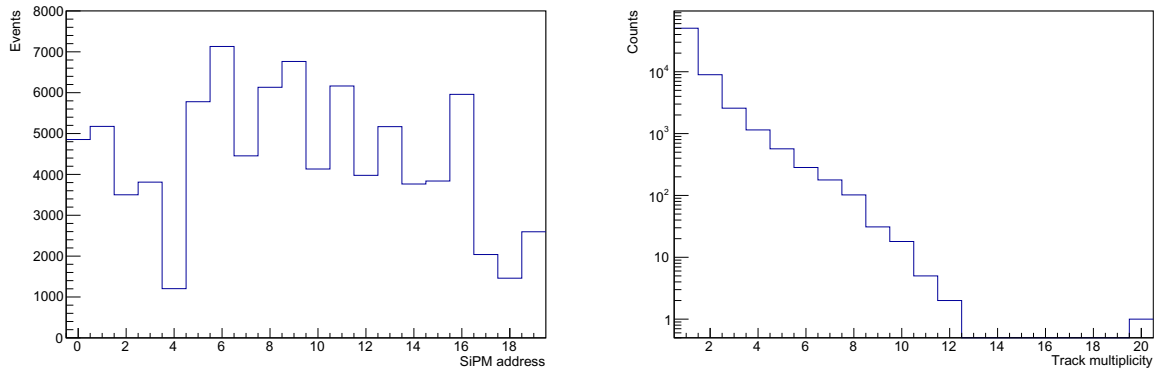


Fig. 86: Example of an AD5391 4-Byte Mode I^2C Write Operation recorded with an Oscilloscope. `SCL` is in blue, while `SDA` in yellow.

6.6 Digital read-out chain tests

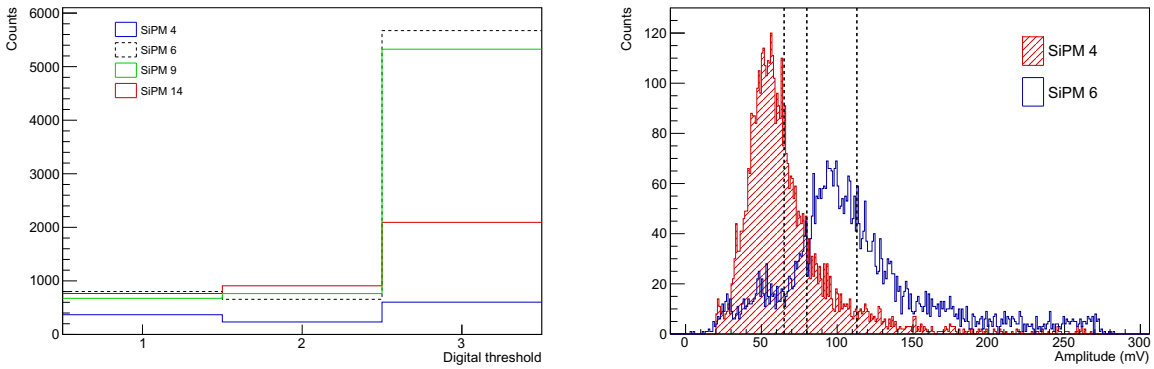
In order to test the whole calorimeter acquisition setup and the complete digital FPGA-based read-out chain, preliminary commissioning tests with cosmic rays were performed. The three digital thresholds were set at exploratory values of 65 mV, 80 mV, and 115 mV respectively. The final threshold values will have to be tuned for each channel, to account for intrinsic anisotropies, such as differences in SiPM gain, electronics components tolerances, noise and dark counts levels, irregularity in the wrapping, and manual assembly of the elements. During the initial commissioning test the threshold will be tuned with cosmic rays, then scaled up for protons, and finally calibrated in a proton test beam. These preliminary tests were initially performed on a 2×10 scintillators array, arranged in a calorimeter Module vertically, and read-out by a single FPGA.



(a) (b)

Fig. 87: (a) Event counts for an approximately 3 hours long cosmic ray digital acquisition with the iMPACT calorimeter, for each of the 20 channels. (b) Recorded track multiplicity, within a 5 timestamps window (50 ns).

Fig. 87 (a) shows the events count for each of the 20 active scintillators during a 3 hours cosmic rays acquisition, using the lowest of the three thresholds. Sensible disparities are observable between different channels, with event rates spanning from 0.1 1/s to 0.65 1/s, over the 1×20 cm^2 area of the individual scintillators. In channels with a higher event rate, such as #6 and #9, the set thresholds are likely to be low enough to partially pick up noise-generated events; on the contrary, in lower rate channels, such as #4 and #18, the lowest threshold is likely to be too high, with respect to the output of the channels, to trigger lower-energies events.
Events recorded in different scintillators within a 5 timestamp window, corresponding to 50 ns, are grouped together into track candidates. Fig. 87 (b) shows the track multiplicities of the recorded acquisition: most of the time windows include only a recorded hit, most likely due to noise-generated events. Events featuring higher-multiplicity, up to 12, are present, which is compatible with the 2×10 scintillators layout. In a single event all the 20 available channels were triggered, due to an external electromagnetic interference, which have been observed in coincidence with the power-on of other instruments, plugged on the same power outlet of the calorimeter power supplies. A power-supply interference filter has been installed afterwards, and has been observed to eliminate such kind of noise.
Fig. 88 (a) shows the digital thresholds spectrum for a selection of channels, the numbers 4, 6, 9 and 14. In channels with higher event rates, such as #6 and #9, the third threshold is fired the most, showing that the largest fraction of the energy spectrum is located at higher voltages with respect to the threshold value. Fig. 88 (b) shows the analog spectrum of channels #4 and #6, which feature the lowest and highest rate respectively, acquired with the iLDA



(a) (b)

Fig. 88: (a) Examples of digital threshold spectrum for channels 4, 6, 9, and 14. (b) Analog energy spectrum for channels 4 and 6; the levels of the three digital thresholds (65 mV, 80 mV, and 115 mV) are marked with vertical dashed lines.

setup illustrated in Sec. 5.1; three vertical dashed lines marks the digital threshold values. The analog acquisitions are performed with four vertically aligned scintillators, triggering on the coincidence between the top and bottom one, to select vertical tracks. The two analog spectra are qualitatively comparable, showing a Landau distribution trend, however the two peaks are shifted from one another by about 50 mV; for channel #4 the bulk of the distribution lays lower than the first threshold, therefore the acquisition triggers only on events on the higher-energy tail. On the contrary, the chosen levels appear well suited for channels #6, where the lowest threshold discriminates between proper events and lower-amplitude noise.
Preliminary tests were also performed on a larger-scale section of the calorimeter, to ensure the correct functionality of the synchronization network, with a 4×10 array, read-out by two FPGAs. Fig. 89 shows three examples of cosmic ray tracks, outlined by the scintillators array, grouped within 5 timestamp windows; the 40 scintillators are represented, with realistic proportions, as grey boxes labelled from 0 to 39. The channels recording a signal are highlighted with a color corresponding to the triggered level: green (65 mV), yellow (80 mV), and red (115 mV); in the central example a compatible muon track is superimposed. These preliminary measurements (representing de-facto the first example of tracking performed by the iMPACT calorimeter) demonstrate the correct functionality of the entire read-out chain, from the SiPM signal generation and the analog electronics, to the FPGA-based parallel acquisition architecture, as well as showing the correct functionality of the synchronization structure.

6.7 Outlook

The calorimeter read-out architecture has been developed with a highly parallel structure, based on identical Modules. The calorimeter is currently able to cope with an acquisition rate per scintillator between 1.7 and 5 MHz, designed for cosmic rays acquisitions; with minor modifications on just a few logical blocks, the same infrastructure is planned to be adapted for proton beams acquisitions, with a sustainable rate up to 20 MHz per scintillator, employing a parallel read-out of the SiPMs.
The iMPACT calorimeter has been proven to be able to perform particle tracking with cosmic rays, even before a complete calibration of the digital thresholds. The acquisition and analysis parameters, such as threshold values are currently being calibrated with cosmic particles on a channel-by-channel basis; the obtained parameters will be re-scaled to proton-suited values using the Monte Carlo simulation described in Sec. 4.4, and finally tuned during a proton test

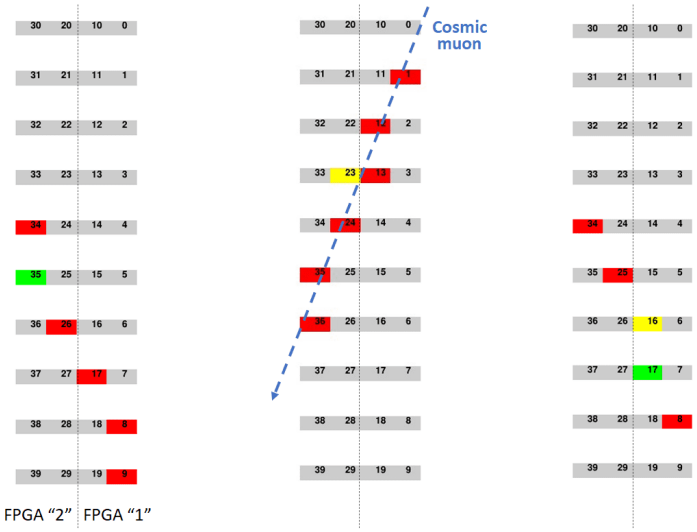


Fig. 89: Examples of three different cosmic tracks recorded with the calorimeter, employing the digital read-out architecture, with a total of 40 scintillators. The two halves of the scintillator array are recorded by different FPGAs, showing the correct functionality of the synchronization structure. Hits are grouped within a 5 timestamp window (50 ns). A muon track compatible with the recorded hits is shown (blue dashed). The scintillator positions are outlined with grey numbered boxes; channels in which a hit has been recorded are highlighted with the color corresponding to the digital threshold fired (green 65 mV, yellow 80 mV, red 115 mV).

beam. In proton beams, the signal is foreseen to be between 2 to 10 times higher (in the Bragg plateau and peak respectively), with respect to cosmic muon MIPs, therefore the signal discrimination from noise is expected to be greatly simplified.
The mechanical assembly of both the calorimeter and tracker is currently approaching its completion, and commissioning test including the entire scanner are planned to be initiated in the near future. In particular, the mechanical alignment of the various elements forming the scanner, as well as efficiency studies, are planned to be evaluated, employing cosmic tracks.

7 Proton track reconstruction

7.1 The Most Likely Proton Path Formalism

One of the main disadvantages of the proton Computed Tomography technique, with respect to the conventional x-ray CT, is the worse spatial resolution [71]. Protons along their trajectory in a medium are, in fact, subject to Multiple Coulomb Scattering (MCS), which primarily causes a large number of small angle deflections. On the other hand, x-rays, if not absorbed, travel in a straight lines.
The majority of current pCT prototypes (see Sec. 3.3) employ silicon-based tracking apparatuses, which can achieve a spatial accuracy of tens of microns or lower. However, the spatial resolution of the resulting image is largely dominated by the accuracy of the reconstruction of the proton track inside the scanned object, which has to rely on mathematical formalisms to model the stochastic nature of MCS process [71].
Studies on the mathematical problem of modelling the MCS effects on a proton trajectory have been conducted by Schneider and Pedroni [160], who proposed a formalism based on the generalized Fermi-Eyges theory of MCS [161]. This formalism is based on the knowledge of the particle entry and exit point, without relying on the information on the entering and exiting directions, although the method was expanded later to also include these parameters [71]. Later, Williams derived a method based on χ^2 to calculate the most probable trajectory of protons in matter, knowing the position of both entrance and exit points, plus the direction in the latter, as the incident beam was assumed to be parallel to the system axis [162] . The Williams formalism, although not applicable in cases of incomplete entrance and exit information, was proven to offer a sensible advantage in terms of the spatial resolution of the final image, with respect to Schneider's formalism [163]. In 2008 R. W. Schulte *et al.* proposed a matrix formalism, obtained from a scattering model, similar to the one used by Williams, employing a Bayesian statistics to calculate the most likely proton lateral displacement and direction in a material at a given depth. In addition, the derived formalism allows to take into account for general proton incident beams, resulting in a more compact expression compared to the previous mentioned ones [71].
In this Section the Schulte formalism will be briefly presented, along with an example of its application in an setup similar to the iMPACT scanner, using a Monte Carlo simulation to assess the effectiveness of the method, by comparing the calculated tracks with the simulated ones.
Indicating with z the depth inside the object, identified by the direction of the beam, the Coulomb scattering in the x and y directions can be considered statistically independent from each another. The scattering process can therefore be studied in a single longitudinal plane, for instance the $x-z$ plane as represented in Fig. 90. A homogeneous object composed by water will be considered for the following calculations, although the algorithm could be in theory applied iteratively with updated information on the object composition at each more accurate iteration; for different materials the mathematical expressions would be formally identical, necessitating only the adjustment of a few numerical parameters.
At a given depth z_1 inside the object, a proton track can be described by a vector containing the track transverse coordinate x_1 and its direction projected on the $x-z$ plane θ_1:

$$\mathbf{t}_1 = \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix} \tag{33}$$

The problem can be articulated in a Bayesian framework: the probability of observing the proton track at a depth z_1 with position x_1 and direction θ_1, given the information on the entering parameters $\mathbf{t}_0$, is related to what can be seen as a *prior* likelihood:

$$\mathcal{L}(\mathbf{t}_1|\mathbf{t}_0) \quad , \tag{34}$$

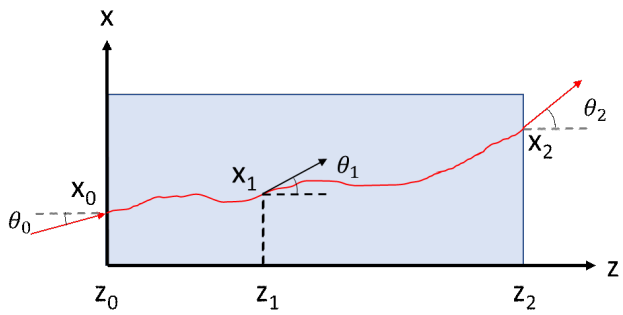


Fig. 90: Scattering geometry of a particle track (red) in the $x - z$ plane. With the indexes 0 and 2 the initial and final coordinates respectively are indicated. With 1 the coordinates at an intermediate depth z_1.

which is a function of x_1 and θ_1, The likelihood of the proton exiting with parameters $\mathbf{t}_2$, given the intermediate position and direction $\mathbf{t}_1$, can be expressed as:

$$\mathcal{L}(\mathbf{t}_2|\mathbf{t}_1) \quad . \tag{35}$$

Indicating the *posterior* likelihood of finding the proton in the intermediate position given the exiting measurements as:

$$\mathcal{L}(\mathbf{t}_1|\mathbf{t}_2) \quad , \tag{36}$$

the relation between the terms (34), (35) and (36) can be describes by the Bayes theorem:

$$\mathcal{L}(\mathbf{t}_1|\mathbf{t}_2) = \mathcal{L}(\mathbf{t}_2|\mathbf{t}_1)\mathcal{L}(\mathbf{t}_1|\mathbf{t}_0) \quad , \tag{37}$$

where the normalization factor $\mathcal{L}(\mathbf{t}_2|\mathbf{t}_0)$ is considered equal to 1. The most probable coordinates $\mathbf{t}_1^{\mathrm{MLP}}$ at z_1 depth are the ones that maximize the posterior likelihood:

$$\mathcal{L}(\mathbf{t}_1^{\mathrm{MLP}} \mid \mathbf{t}_2) = \max[\mathcal{L}(\mathbf{t}_1|\mathbf{t}_2)] \quad , \tag{38}$$

or equivalently:

$$\nabla\mathcal{L}(\mathbf{t}_1|\mathbf{t}_2) = \begin{pmatrix} \partial x_1 \\ \partial\theta_1 \end{pmatrix} \mathcal{L}(\mathbf{t}_1|\mathbf{t}_2)|_{\mathbf{t}_1=\mathbf{t}^{\mathrm{MLP}}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad . \tag{39}$$

The MLP coordinates calculated at different depth points along the z direction outline the most probable proton path.
This MLP formalism is based on the Fermi-Eyges theory [161], which is a Gaussian approximation of the more complete Moliere theory [164]. The Gaussian approximation is expected to be accurate enough for proton Tomography, because the large angle scattering events that populate the tails, caused by Coulomb, nuclear elastic or nuclear inelastic scattering, will be rejected, as they are expected to worsen the image reconstruction accuracy [71].
At first, the particle can be assumed to enter the object at $x_0 = 0$ with a trajectory parallel to z. The prior likelihood function can be expressed, in the Fermi-Eyges model, as a 2-dimensional normal distribution, which in matrix notation is:

$$\mathcal{L}\left(\mathbf{t}_1|\mathbf{t}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \exp\left(-\frac{1}{2}\mathbf{t}_1^T\boldsymbol{\Sigma}_1^{-1}\mathbf{t}_1\right) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_1) \quad , \tag{40}$$

where Σ_1 is the scattering matrix, containing the variance of x_1 and θ_1, and their covariance, calculated between the points at z_1 and z_0:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1\theta_1}^2 \\ \sigma_{x_1\theta_1}^2 & \sigma_{\theta_1}^2 \end{pmatrix} \quad , \tag{41}$$

which is symmetric and positive-definite.
The elements of the covariance matrix are calculated with the expressions:

$$\sigma^2_{x_1}(z_1, z_0) = E_0^2 \left(1 + 0.038 \ln \frac{z_1 - z_0}{X_0}\right)^2 \times \int_{z_0}^{z_1} \frac{(z_1 - z)^2}{\beta^2(z) p^2(z)} \frac{dz}{X_0} \quad , \tag{42}$$

$$\sigma^2_{\theta_1}(z_1, z_0) = E_0^2 \left(1 + 0.038 \ln \frac{z_1 - z_0}{X_0}\right)^2 \times \int_{z_0}^{z_1} \frac{1}{\beta^2(z) p^2(z)} \frac{dz}{X_0} \quad , \tag{43}$$

$$\sigma^2_{x_1\theta_1}(z_1, z_0) = E_0^2 \left(1 + 0.038 \ln \frac{z_1 - z_0}{X_0}\right)^2 \times \int_{z_0}^{z_1} \frac{z_1 - z}{\beta^2(z) p^2(z)} \frac{dz}{X_0} \quad , \tag{44}$$

in which $\beta = v/c$ is the relative velocity with respect to light and p is the relativistic momentum, both as functions of the depth z in water; $E_0 = 13.6$ MeV/c and 0.038 are empirical parameters [165], and $X_0 = 36.1$ cm is the radiation length in water. The expressions in (42)-(43) have been derived, without the logarithmic term, by Eyges [161] as analytical solutions of the equations in the Fermi MCS model [166]. The Eyges expressions also introduced the modelling of the energy loss of the particles, while in Fermi's treatment the energy loss was neglected. Additionally, the expression include a logarithmic thickness-dependent correction [36]. The expressions in (42)-(43) have been proven to describe accurately the lateral profile of a proton beam in water [167].
For a pCT application, the incident particles angles can be assumed to be relatively small, so that the small-angle approximation can be employed:

$$\sin\theta_0 \approx \theta_0 \quad ; \quad \cos\theta_0 \approx 1 \quad . \tag{45}$$

In order to use the expression of a bi-variate Gaussian centered in 0 as in Eq. (40), the coordinate vector $\mathbf{t}_1$ needs to be transported in the reference system in which $x_0 = 0$ and $\theta_0 = 0$:

$$\mathbf{t}'_1 = \mathbf{t}_1 - \mathbf{R}_0 \mathbf{t}_0 \quad , \tag{46}$$

in which the rotation matrix transformation $\mathbf{R}_0$ is:

$$\mathbf{R}_0 = \begin{pmatrix} 1 & z_1 - z_0 \\ 0 & 1 \end{pmatrix} \quad . \tag{47}$$

With the defined transformation the prior likelihood in Eq. (40) can be expressed as:

$$\mathcal{L}(\mathbf{t}_1|\mathbf{t}_0) = \exp\left(-\frac{1}{2}\left(\mathbf{t}_1^T - \mathbf{t}_0^T \mathbf{R}_0^T\right) \mathbf{\Sigma}_1^{-1} \left(\mathbf{t}_1 - \mathbf{R}_0 \mathbf{t}_0\right)\right) \quad . \tag{48}$$

A similar expression for $\mathcal{L}(\mathbf{t}_2|\mathbf{t}_1)$, i.e. the likelihood of the exit coordinates $\mathbf{t}_2$ given $\mathbf{t}_1$ at the intermediate depth z_1, can be obtained (replacing the 1 and 0 indexes with 2 and 1, respectively). The complete expression for $\mathcal{L}(\mathbf{t}_2|\mathbf{t}_1)$ can be found in Eq: (80)-(85) in Appendix B.
The formulation of the Bayes theorem (37) can therefore be obtained combining the expressions in Eq.(48) and (80):

$$\begin{aligned} \mathcal{L}(\mathbf{t}_1|\mathbf{t}_2) &= \exp\left(-\frac{1}{2}\left(\mathbf{t}_1^T - \mathbf{t}_0^T \mathbf{R}_0^T\right) \Sigma_1^{-1} \left(\mathbf{t}_1 - \mathbf{R}_0 \mathbf{t}_0\right) - \frac{1}{2}\left(\mathbf{t}_2^T - \mathbf{t}_1^T \mathbf{R}_1^T\right) \mathbf{\Sigma}_2^{-1} \left(\mathbf{t}_2 - \mathbf{R}_1 \mathbf{t}_1\right)\right) \\ &\equiv \exp\left(-\chi^2\right) \end{aligned} \quad , \tag{49}$$

where the argument of the exponential is indicated as a χ^2. Maximizing the likelihood is then equivalent to maximizing the χ^2. Setting the derivative of χ^2 to zero:

$$\nabla\chi^2 = (\mathbf{\Sigma}_1^{-1} + \mathbf{R}_1^T \mathbf{\Sigma}_2^{-1} \mathbf{R}_1)\mathbf{t}_1 - \mathbf{\Sigma}_1^{-1} \mathbf{R_0} \mathbf{t}_0 - \mathbf{R}_1^T \mathbf{\Sigma}_2^{-1} \mathbf{t}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad , \tag{50}$$

and solving for $\mathbf{t}_1$, the matrix expression of the most likely position and direction $\mathbf{t}_1^{\text{MLP}}$ at a given depth z_1 can be found [71]:

$$\mathbf{t}_1^{\text{MLP}}(z_1|\mathbf{t}_0, z_0, \mathbf{t}_2, z_2) = \begin{pmatrix} x_1^{\text{MLP}} \\ \theta_1^{\text{MLP}} \end{pmatrix} = (\mathbf{\Sigma}_1^{-1} + \mathbf{R}_1^T \mathbf{\Sigma}_2^{-1} \mathbf{R}_1)^{-1} (\mathbf{\Sigma}_1^{-1} \mathbf{R}_0 \mathbf{t}_0 + \mathbf{R}_1^T \mathbf{\Sigma}_2^{-1} \mathbf{t}_2) \quad . \tag{51}$$

The explicit expression for x_1^{MLP} can be obtained expanding this matrix expression. In particular it results:

$$x_1^{\text{MLP}} = \frac{a_{22} b_1 - a_{12} b_2}{a_{11} a_{22} - a_{12}^2} \quad , \tag{52}$$

where:

$$a_{11} = \frac{\sigma_{\theta_1}^2}{\det \mathbf{\Sigma}_1} + \frac{\sigma_{\theta_2}^2}{\det \mathbf{\Sigma}_2} \tag{53}$$

$$a_{12} = \frac{\sigma_{\theta_2}^2 (z_2 - z_1) - \sigma_{x_2\theta_2}^2}{\det \mathbf{\Sigma}_2} - \frac{\sigma_{x_1\theta_1}^2}{\det \mathbf{\Sigma}_1} \tag{54}$$

$$a_{22} = \frac{\sigma_{x_1}^2}{\det \mathbf{\Sigma}_1} + \frac{\sigma_{\theta_2}^2 (z_2 - z_1)^2 - 2\sigma_{x_2\theta_2}^2 (z_2 - z_1) + \sigma_{x_2}^2}{\det \mathbf{\Sigma}_2} \tag{55}$$

$$b_1 = \frac{\sigma_{\theta_1}^2 (x_0 + \theta_0 (z_1 - z_0)) - \theta_0 \sigma_{x_1\theta_1}^2}{\det \mathbf{\Sigma}_1} + \frac{x_2 \sigma_{\theta_2}^2 - \theta_2 \sigma_{x_2\theta_2}^2}{\det \mathbf{\Sigma}_2} \tag{56}$$

$$b_2 = \frac{\theta_0 \sigma_{x_1}^2 - \sigma_{x_1\theta_1}^2 (x_0 + \theta_0 (z_1 - z_0))}{\det \mathbf{\Sigma}_1} + \frac{(x_1 \sigma_{\theta_2}^2 - \theta_2 \sigma_{x_2\theta_2}^2)(z_2 - z_1) + \theta_2 \sigma_{t_2}^2 - x_2 \sigma_{x_2\theta_2}^2}{\det \mathbf{\Sigma}_2} \tag{57}$$

$$\det \mathbf{\Sigma_1} = \sigma_{x_1}^2 \sigma_{\theta_1}^2 - \sigma_{x_1\theta_1}^4 \quad ; \quad \det \mathbf{\Sigma_2} = \sigma_{x_2}^2 \sigma_{\theta_2}^2 - \sigma_{x_2\theta_2}^4 \quad . \tag{58}$$

The explicit calculations can be found in Appendix B, in particular Eq. (87), (88) and (89). With the expression in Eq. (89), x_1^{MLP} can be calculated at a series of depths, to outline the most likely path of a proton inside an object. For instance, when considering a voxelization of the object, the trajectory could be evaluated at the boundaries of every voxel, in order to calculate the particle path length inside each subdivision.

The variance of the trajectory estimation of Eq. (51) can be derived as well. In fact, knowing the distribution of the possible trajectories around the most likely one could be useful during the image reconstruction phase, to have an assessment on the confidence belt of the track estimation. Moreover, it has been suggested that image reconstruction algorithms could be based on integrating the contribution of a single proton over its trajectory likelihood distribution, rather than being based solely on the one-dimensional trajectory [162].

The bi-variate Gaussian likelihood distribution can be described by the error matrix $\boldsymbol{\varepsilon}_{x,\theta}$, which corresponds to the inverse of the curvature matrix $\alpha_{x,\theta}$ [71]:

$$\boldsymbol{\alpha}_{x,\theta} \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial x_1 \partial \theta_1} \quad , \tag{59}$$

where

$$\frac{\partial^2 \chi^2}{\partial x_1 \partial \theta_1} = \mathbf{\Sigma}_1^{-1} + \mathbf{R}_1^T \mathbf{\Sigma}_2^{-1} \mathbf{R}_1 \quad , \tag{60}$$

so the 2×2 error matrix can be found as:

$$\boldsymbol{\varepsilon}_{x_1,\theta_1}(z_1) = 2(\mathbf{\Sigma}_1^{-1} + \mathbf{R}_1^T \mathbf{\Sigma}_2^{-1} \mathbf{R}_1)^{-1} \quad , \tag{61}$$

in which the element in position $(1;1)$ corresponds to the variance of x_1^{MLP}. The expression in Eq. (61) is almost equal to the first term in Eq. (51), so the variance of x_1^{MLP} can be calculated as:

$$\sigma_{x_1^{\text{MLP}}}^2 = 2 \frac{a_{22}}{a_{11} a_{22} - a_{12}^2} \quad . \tag{62}$$

7.2 Simulations and application

The calculation of the MLP trajectory for a proton requires the computation of the integral terms σ^2 in Eq. (42)-(43) and (83)-(85): for each point along the track 12 integrals have to be calculated, 6 for the x displacement and 6 for the y. However, numerical integrations are particularly heavy operations in terms of computational resources, and the computational time increases linearly with the track reconstruction granularity and with the total number of protons. This could have a substantial impact on the image reconstruction computational time, especially if more iterative reconstruction steps are required, and could hinder the overall feasibility of the MLP method.
In particular in the σ^2 expressions, the term:

$$\frac{1}{\beta^2(z)p^2(z)} = \frac{(E(z)+m_p)^2c^2}{(E(z)+2m_p)^2E^2(z)} \quad , \quad m_p = 938.272\ \frac{\text{MeV}}{c^2} \tag{63}$$

can not be integrated analytically, in this form, and has to be calculated numerically. However, it has been shown that this term can be fitted from simulated data with a 5-th degree polynomial with satisfying results [71] [162]:

$$\frac{1}{\beta^2(z)p^2(z)} \sim a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 \quad , \tag{64}$$

allowing for an analytical integration of all the σ^2 terms; the computation of this resulting 6-th degree polynomial is significantly faster (around 200 times) than the numerical integration of the same term in the original form (63). The coefficients of the fit are specific for a particular beam energy and material. It has to be noted that the polynomial fit of the $\beta^{-2}(z)p^{-2}(z)$ term is increasingly less adequate at larger depths (see Fig. 91), so a different type of approximation should be considered in case of larger objects.
To evaluate the coefficients of the polynomial approximation (64) a Monte Carlo simulation has been set up, using the GATE framework, based on the Geant4 toolkit [127] [128]. The Monte Carlo includes a monochromatic proton pencil beam incident on an homogeneous water volume, deep enough to stop protons up to 250 MeV (≥ 40 cm). The values of $\beta^{-2}(z)p^{-2}(z)$ as a function of the depth z are plotted in Fig. 91 (a) for a 200 MeV and (b) 250 MeV proton beam; the data points are interpolated with the polynomial Eq. (64). In the graphs the residual proton energy as a function of the depth is also shown. The horizontal axis in the plots corresponds to the actual distance traveled by the proton inside the object, estimated as the progressive sum of the distance between the particles positions along the sampling of their trajectory. From the graphs it can be observed that 200 MeV protons have a range in water of about 26 cm, which extends to 37 cm for the 250 MeV beam.
In the graph corresponding to the 200 MeV proton beam, in addition to the fit obtained from Monte Carlo data, the function with coefficients derived from [71] is also plotted (dashed line). For the 200 MeV fit (Fig. 91 (a)) the range considered is up to 15 cm, while for 250 MeV up to 25 cm (Fig. 91 (b)). For both beam energies only events subject to MCS and energy loss to electrons are considered: any proton undergoing non-ionizing interactions, either nuclear elastic or inelastic, at any point during its motion is discarded. The discarded events represent the 54% and 48% of the total for 200 MeV and 250 MeV respectively. From the graphs it can be noted that the 5-th grade fit fails to describe the data for higher depths; for 200 MeV the interpolations becomes inconsistent after about 20 cm, while for 250 MeV the fit fails from $25-30$ cm and above. Moreover, given the stochastic nature of the energy loss, at increasing depths the dispersion of the points grows larger; consequently, track predictions are expected to be less precise when affected by bigger the energy losses, as suggested in [160].
In Tab. 6 the coefficients of the polynomial interpolation Eq. (64) performed on the Monte Carlo data, shown in Fig. 91, are presented. The coefficients of the interpolation retrieved from [71] are also tabulated, for reference.

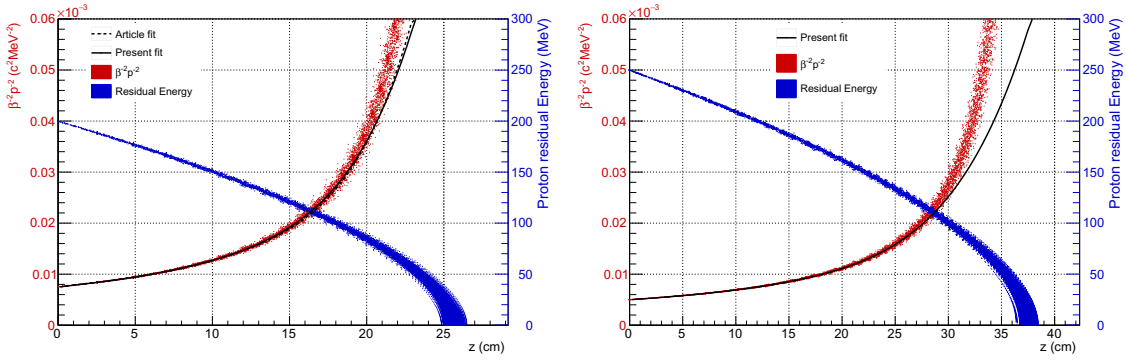
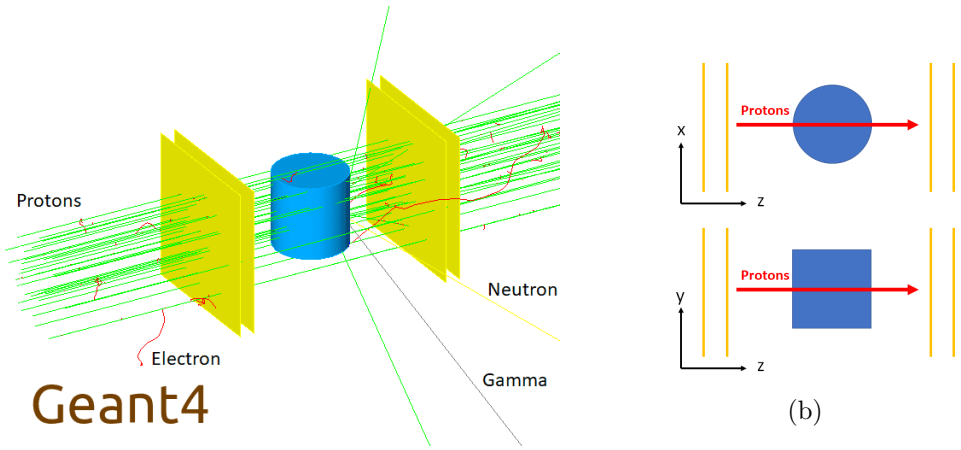


Fig. 91: Trend of the quantities $\beta^{-2}(z)p^{-2}(z)$ (red) and residual energy (blue) as functions of the depth z inside a homogeneous water object for a monochromatic proton beam. Data generated from Monte Carlo simulations based on the GATE-Geant4 toolkit; the proton beam energies are: (a) 200 MeV and (b) 250 MeV, the polynomial is fitted up to 15 cm in (a) and up to 25 cm in (b). Values for the interpolation coefficient for both energies are tabulated in Tab. 6. In both the 200 and 250 MeV cases only completely non-hadronic events are considered.

Coefficient	Literature (200 MeV)	Fit 200 MeV	Fit 250 MeV
a_0 $(c^2\text{MeV}^{-2})$	$7.457 \cdot 10^{-6}$	$7.48 \cdot 10^{-6} \pm 8 \cdot 10^{-8}$	$4.98 \cdot 10^{-6} \pm 3 \cdot 10^{-8}$
a_1 $(c^2\text{MeV}^{-2}\text{cm}^{-1})$	$4.548 \cdot 10^{-7}$	$3.7 \cdot 10^{-7} \pm 7 \cdot 10^{-8}$	$1.6 \cdot 10^{-7} \pm 2 \cdot 10^{-8}$
a_2 $(c^2\text{MeV}^{-2}\text{cm}^{-2})$	$-5.777 \cdot 10^{-8}$	$-2 \cdot 10^{-8} \pm 2 \cdot 10^{-8}$	$-3 \cdot 10^{-9} \pm 4 \cdot 10^{-9}$
a_3 $(c^2\text{MeV}^{-2}\text{cm}^{-3})$	$1.301 \cdot 10^{-8}$	$7 \cdot 10^{-9} \pm 3 \cdot 10^{-9}$	$1.1 \cdot 10^{-9} \pm 4 \cdot 10^{-10}$
a_4 $(c^2\text{MeV}^{-2}\text{cm}^{-4})$	$-9.228 \cdot 10^{-10}$	$-5 \cdot 10^{-10} \pm 2 \cdot 10^{-10}$	$-6 \cdot 10^{-11} \pm 2 \cdot 10^{-11}$
a_5 $(c^2\text{MeV}^{-2}\text{cm}^{-5})$	$2.687 \cdot 10^{-11}$	$1.7 \cdot 10^{-11} \pm 3 \cdot 10^{-12}$	$1.4 \cdot 10^{-12} \pm 3 \cdot 10^{-13}$

Tab. 6: Coefficients of the 5-th degree polynomial function (64) for the fit of $\beta^{-2}(z)p^{-2}(z)$. Literature values are referring to [71] for a 200 MeV proton beam. Fit values are from interpolations performed with data from Monte Carlo simulations with 200 and 250 MeV proton beam, shown in Fig. 91.

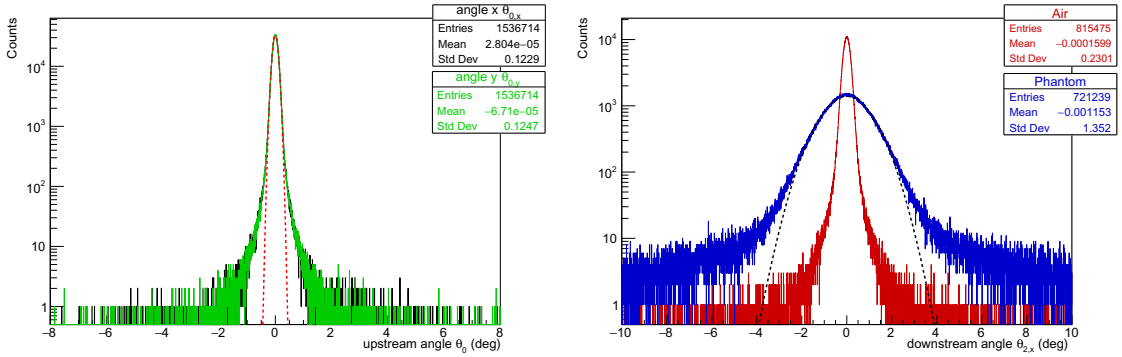
In order to study the performance of the Maximum Likelihood Path formalism and its applicability in a setup similar to the iMPACT scanner, a GATE-Geant4 Monte Carlo simulation including the iMPACT tracker section has been designed. The layout of the simulation is shown in Fig. 92. The entire system is placed in air. The tracker planes are modelled as 20×20 cm^2 silicon planes, with 100 µm thickness; the 4 tracker planes are placed at $z = \pm 14$ and ± 16 cm. The simulated sensor thickness represent a worst-case scenario, as the actual thickness of ALPIDE is 50 µm. A homogeneous water cylinder, with 5 cm radius and 10 cm height, placed in the centre of the reference system, constitutes the target object, or the *phantom*. The simulation includes a uniform circular proton source, with 8 cm radius, and monochromatic 200 MeV energy; the particle direction at the source is parallel to the z-axis. The source is placed 24 cm before the first silicon plane, in order to simulate a realistic test beam in which the beam travels a certain distance in air before encountering the apparatus. The phantom is completely contained in the field of view of the proton source.
In Fig. 92 (a) proton tracks are visible in green, while red tracks represent electrons generated by the ionizing interaction of protons with atoms. Some neutrons and gamma particles are also generated as products of nuclear interactions. The output of the simulation includes the points on the silicon planes hit by the particles; to each point a Gaussian blurring of $\sigma_{x,y} = 10$ µm in both directions is added, in order to simulate the effect of the pixel sensor spatial resolution. The value of 10 µm is an underestimation of the resolution of a realistic sensor, as the ALPIDE



(a) (b)

Fig. 92: Layout of the GATE-Geant4 simulation for the application of the MLP formalism with iMPACT tracker. (a) A 3D rendering of the layout generated by Geant4, with a number of particle tracks. Green tracks represent protons, red ones are electrons, the yellow one is a neutron and the grey one is a gamma particle; (b) top and lateral view of the simulation layout. The simulation includes four 20×20 cm^2, 150 µm thick, silicon planes and a cylindrical water phantom, with 5 cm radius and 10 cm height. The proton beam is, uniform, parallel and circular, with 8 cm of radius.

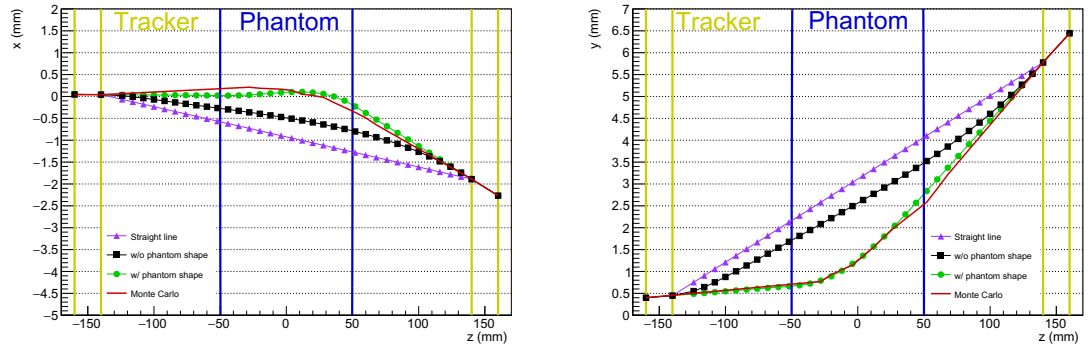
chip used for the iMPACT prototype ensures a spatial resolution of about 5 µm (see Sec. 4.3.1). The actual trajectory of the proton inside the water phantom is also recorded, in order to compare it with the calculated MLP track. The simulation is further divided into 36 steps; at the end of each step the phantom is rotated by 10° around its vertical axis. The rotation of the phantom simulates a tomography procedure. In Fig. 93 the distributions of the mea-



(a) (b)

Fig. 93: (a) Distribution of the direction of the proton trajectories, as measured by the upstream tracker section (before the phantom), the black histogram refers to the angle defined by the projection of the track on the $x - z$ plane ($\theta_{0,x}$), while the green histogram is related to the projection angle on the $y - z$ plane ($\theta_{0,y}$); the proton source is located at a 24 cm distance from the first tracker plane; a Gaussian fit (dashed red line) is superimposed. (b) Distribution of the proton trajectories, as measured by the downstream tracker section (after the phantom); only the angle projection on the $x - z$ plane is shown ($\theta_{2,x}$), the blue histogram is due to the protons that passed through the phantom, the red histogram is due to protons that did not intercept the phantom; 28 cm separate the inner tracking planes; a Gaussian fit (dashed black line) is also shown.

sured particle angles θ in the Monte Carlo Simulation are shown, as detected by the upstream tracking planes in (a) and by the downstream tracking planes in (b). The angle measurements are affected by an uncertainty deriving by the tracker spatial resolution, in this case equal to $\sigma_\theta = \frac{2}{|\Delta z|}\sigma_x = 1$ mrad $= 0.06°$, where $\Delta z = z_B - z_A = z_D - z_C = 20$ mm is the distance between two tracking planes. In the first graph the projection of the angles on the $x - z$ (horizontal) and $y - z$ (vertical) planes are presented separately. As it can be seen, the distributions of $\theta_{0,x}$ and $\theta_{0,y}$ are equivalent, as the deviation due to MCS in the horizontal and vertical planes can be seen as independent processes. The angular spread of the beam, measured by the first two tracking planes, is caused mainly by the deviation the particles undergo in the 150 µm of silicon of the first plane, in addition to the deviation suffered traversing the 24 cm of air between the MC source and first tracking plane. The beam spread at the first tracking stage is relatively limited, as 90% of the particles have a measured angle smaller than $0.155° = 2.27$ mrad. The central part of the upstream angle distribution has been interpolated with a Gaussian function, shown in Fig. 93 (a) with a dashed red line; as expected, at higher angles the distribution of MCS angles deviates from a Gaussian behaviour. The Gassian fit gives a standard deviation of $\sigma_{\theta_0} \approx 0.093° = 1.62$ mrad; the deviation caused by the silicon plane has been estimated to be around $\sigma_{\mathrm{Si,200MeV}} \approx 0.065° = 1.13$ mrad, which indicates that it constitutes a larger contribution to the beam spread with respect to the air.
In Fig. 93 (b) the measured downstream angles θ_2 are separated between the particles that encountered the phantom along their path (blue distribution) and the ones that missed the phantom and only travelled in air (red distribution); for simplicity's sake only the angle projection on the $x - z$ plane is shown, as it is equivalent to the $y - z$ projection. As expected, the protons passing through the phantom are more likely to be deviated to a higher angle. In fact, while the 90% of the downstream particles that did not encountered the phantom have an angle $\leq 0.26° = 4.54$ mrad and 90% of the distribution of the particle encountering the object is contained in the interval $\pm 1.9° = 33.2$ mrad. The standard deviations of Gaussian fits of the central areas of the distributions are $\sigma^{air}_{\theta_2,x} \approx 0.150° = 2.62$ mrad and $\sigma^{phantom}_{\theta_2,x} \approx 0.980° = 17.1$ mrad. In the simulation setup, for geometric reasons, about one half of the particles passes through the phantom.
In Fig. 94 an example of a simulated proton track is shown, in his $x-z$ (a) and $y-z$ projections (b). The volume of the phantom is included between the blue vertical lines; the tracking planes are indicated by four yellow lines. The considered track undergoes a lateral displacement of about 2.25 mm and 6.5 mm in x and y respectively. In the graphs, three reconstruction methods are compared, in order to determine the most efficient one in estimating the exact particle trajectory, taking into particular consideration the accuracy of the reconstruction inside the phantom volume.
The most basic tracking algorithm considered consists in tracing a straight line connecting the hits on the two innermost silicon planes; this method is particularly fast in terms of computational resources, but at the same time it is expected to give worse results with respect to other methods. The straight line reconstruction is shown in Fig. 94 with violet triangles. This type of reconstruction for the considered track is able to estimate the proton path with an error up to 1.25 mm in x and 1.8 mm in y.
The second method considered employs the MLP formalism in an basic way: the reconstructed track $\mathbf{t}_1^{\mathrm{MLP}}(z_1, \mathbf{t}_0, \mathbf{t}_2)$ is calculated using Eq. (52) at regular intervals along z. In this example, for display purposes, the z-pitch of the tracking was set particularly coarse, to 8 mm, although for an adequate imaging application the track is required to be evaluated with a spacing of the order of the millimeter or even smaller. The parameters describing the initial coordinates of the particle x_0 and θ_0 are considered as the position measured in the second tracking plane and the upstream direction measured by the first two planes. Similarly, the final coordinates are set as the position of the hit on the third plane and the downstream measured angle. No information on either the shape or positioning of the phantom is used. As can be noted in both Fig. 94 (a) and (b), this basic MLP method performs better than the straight line: the reconstruction



(a) (b)

Fig. 94: Example of the reconstruction of a simulated proton track using different methods: (a) projection of the track in the $x-z$ (horizontal) and (b) $y-z$ (vertical) plane. The four tracker planes are indicated by yellow vertical lines, while the internal volume of the water phantom is delimited by the two blue lines. The straight line reconstruction method is shown with violet triangles. The MLP method without considering the informations on the phantom shape is shown as black squares. In in this example, the z-pitch of the reconstruction is 8 mm. The track reconstruction using the MLP formalism with information on the phantom geometry is shown with green circles. The vertical range is equal between the two graphs.

error inside the phantom is at worst 0.6 mm in x and 1.4 mm in y, for the shown case. However, while this algorithm is able to follow more closely the curvature of the real trajectory, the MLP calculation allows the reconstructed track to match the initial and final directions of the particle by distributing the deviation smoothly along the entire track. The MLP calculation, in fact, assumes an homogeneous medium, while in reality most of the deviation takes place inside the object, and the trajectory in air is close to a straight line.

A natural improvement of this simple use of MLP, consists in taking advantage of a-priori knowledge of the geometrical characteristics of the phantom. As shown in Fig. 94 (a) and (b), for the reconstruction with informations on the phantom (green circles), the track is propagated forward, from the upstream tracker, and backwards, from the downstream tracker, as straight lines according to the respective measured directions. The track inside the phantom is calculated with the MLP method Eq. (52), using the intersections of the propagated forward and backward linear tracks with the phantom outer surface as entrance and exit points. As can be seen in both graphs, this last algorithm provides a more accurate result than the two previous ones: in both direction the maximum error is around the 200 to 300 µm order. For the MLP calculation with phantom information, exact knowledge of the geometrical shape of the object is assumed, which, although possible in a simulation, it would not be readily feasible in a real-life application with organic bodies; for this reason, in the following section 7.3, a method for a preliminary target shape recognition will be proposed.

Fig. 95 shows the distributions of the reconstruction errors, for the three aforementioned methods, obtained from a complete simulation with the setup presented in Fig. 92. Protons undergoing nuclear interactions in the phantom volume (about 20% of the total) are discarded, as this type of event is recognizable by a anomalous number of hits in the tracker: more than one interaction points per plane indicates the production of secondary particles, while a missing hit in the downstream planes signals that the proton has been deviated enough to miss the planes. In particular, the histograms in Fig. 95 (a) and (c) show the distributions of the maximum error per each track, in linear and logarithmic scales on the vertical axis respectively; errors in x and y are considered independent and included as different entries. Only the portions of the tracks internal to the phantom volume have been considered. The maximum error histograms

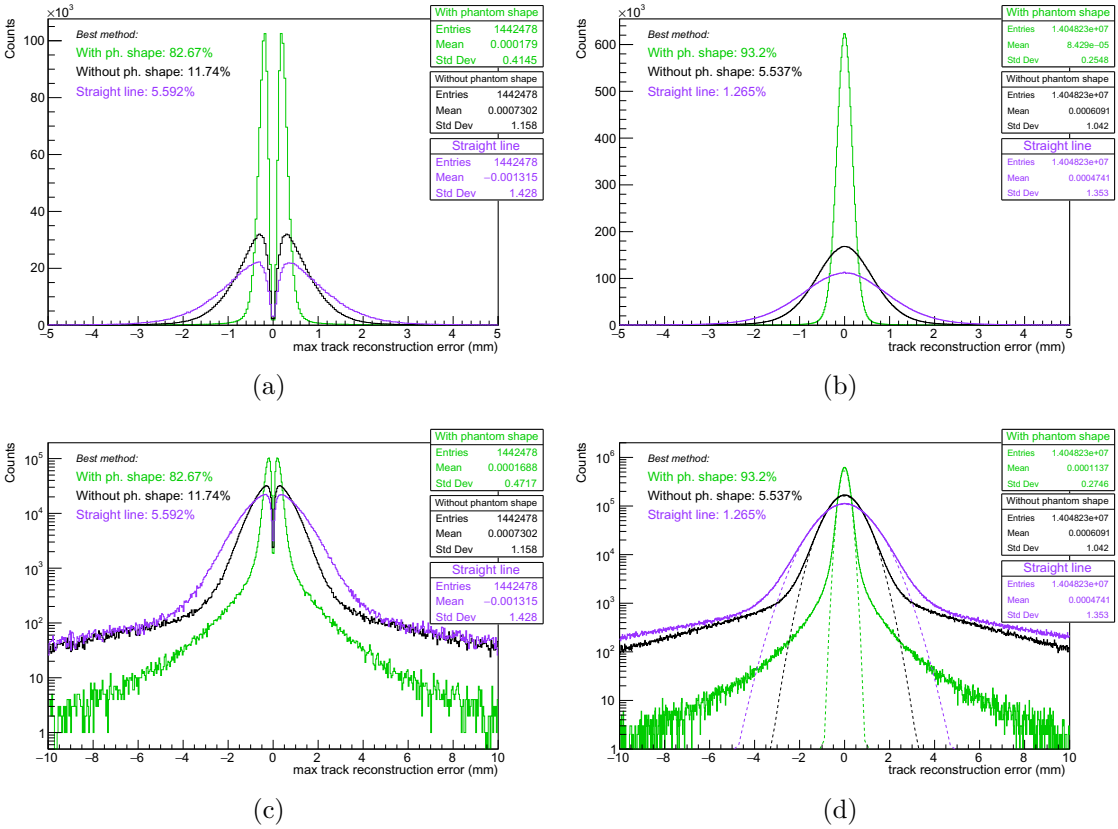
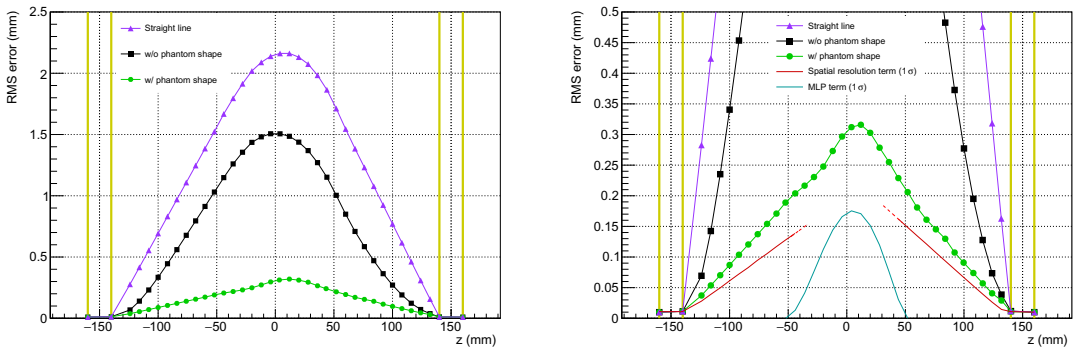


Fig. 95: (a) and (c): distribution of the maximum track reconstruction error per each particle, in linear and logarithmic scale respectively. (b) and (d): distribution of the reconstruction error in each step along the track, in linear and logarithmic scale respectively. The green histogram is related to the MLP reconstruction method in which the shape of the phantom is taken into account, the black one is related to the basic MLP track reconstruction that does not take into account the phantom shape, while the violet one refers to tracks consisting in a simple straight line connecting the hits on the inner tracking planes. The errors are considered only when the track is inside the phantom volume. The fraction of the tracks in which a particular method resulted the most accurate is also reported on the graphs. In (d) Gaussian fits are also shown with dashed lines.

appear as bi-modal symmetric distributions with a local minimum at 0. It is apparent that the algorithm based on the phantom shape knowledge presents generally smaller errors with respect to the other methods, with a narrower bulk and lower tails; the modes of the distribution (i.e. the most probable error) are placed at about $\pm 180\,\mu m$. Additionally, as expected, the straight line calculation provides the worst reconstruction accuracy out of the tree methods. In fact, for almost 83% of the particles, the track calculated with phantom shape knowledge has the smallest maximum distance to the Monte Carlo track; at the same time the calculation without phantom knowledge has the smallest error of about 13% of the cases. Finally, the straight line tracking has the smallest error only 5.6% of the time.

Fig. 95 (b) and (c) show the distribution of the reconstruction error inside the phantom, calculated in the various sampling steps along the z direction. Once more the calculation that takes into account the phantom shape appears to be the most accurate of the three: Gaussian fits performed on the bulk of the histograms show a standard deviation of $178\,\mu m$ for the mentioned method, $664\,\mu m$ for the basic MLP calculation and $984\,\mu m$ for the straight line tracking. On the logarithmic plot, the Gaussian trends are also superimposed to the histograms, which show

large deviations of the distributions tails from the Gaussian ones, as expected from the physics of MCS. To identify the most accurate tracking method for each particle, the squared-sum of the errors along the track steps is compared. Using this metric, the reconstruction that takes into account the shape produces the most accurate estimation in the vast majority of the cases, 93%. The simple MLP calculation is best in the 5.5% of cases, while the straight line wins only 1.3% of the time.



(a) (b)

Fig. 96: (a) Root Mean Square (RMS) of the reconstruction error as a function of z, for the tree methods: straight line (violet triangles), MLP without phantom shape knowledge (black squares) and MLP with shape knowledge (green circles). Only tracks traversing the phantom volume are included. (b) Highlight on the RMS error related to the MLP reconstruction with phantom shape. The 1 σ term deriving from the entry and exit points uncertainty is shown in red; in cyan the estimated uncertainty $\sigma^2_{x_1^{\mathrm{MLP}}}$, from Eq. (62).

Fig. 96 shows the Root Mean Square (RMS) error of the three track reconstruction methods as a function of the z coordinate. Only particles passing through the phantom at any point during their path are considered. Given the dimensions of the MC proton source (8 cm of radius), the plot includes particles that traversed the phantom at its maximum thickness (10 cm at the centre) down to just few millimeters at the lateral edges of the cylinder. The errors trend grows larger with the distance from the tracking planes, reaching their maximum near the centre of the phantom, although slightly downstream, as the energy loss negatively affects the accuracy of the track reconstruction, as suggested in [160]. The straight line calculation presents a RMS at the phantom centre (on axis) of about 2.2 mm, which is not adequate for imaging applications, considering typical voxel sizes of the order of 500 µm to 1 mm. The basic MLP without shape information performs better, with an average error at the centre of about 1.5 mm, which is again too large for a definitive imaging, but can be acceptable as a preliminary iteration of the reconstruction procedure, that would need to be further refinements with more precise methods. Finally, the MLP calculation with shape knowledge shows a substantially lower error, slightly above 300 µm at its maximum; this means that, considering 500 µm voxels, in most cases a track would be associated to the correct voxel, or at least to an adjacent one. The precision the entrance and exit points are located, considering the depths $z = \pm 50$ mm, is calculated with the Monte Carlo data to be around 200 µm RMS, slightly lower for the entrance point (194 µm) and higher for the exit point (222 µm). The main contributions to these uncertainties are due to: the tracker spatial resolution, which also causes uncertainty on the angle measurements; the deviation in the second and third 100 µmthick silicon planes, that can not be measured by either the upstream or the downstream tracking sections; and, with a smaller weight, the beam spread caused by the air. The uncertainty caused by the tracker

spatial resolution can be estimated, by error propagation, from the formula of a straight line:

$$\left(\sigma_{x_{in}}^{spatial}\right)^2 \approx \left[2\left(\frac{z_{in}-z_B}{|\Delta z|}\right)^2 + 2\frac{z_{in}-z_B}{|\Delta z|} + 1\right]\sigma_x^2 = 71\,\mu\text{m} \tag{65}$$

where the value has been calculated for $z_{in} = -50$ mm (the tracker spatial resolution is $\sigma_{x,y}$ =10 µm). The same expression stands for the exit point $z_{out} = 50$ mm, substituting z_C to z_B, which provides the same numerical result. The term related to the deviation of the protons inside the second tracker plane can be estimated with a simple projection of the angle for the travelled distance (as previously $\sigma_{\text{Si,200MeV}} \approx 0.065° = 1.13$ mrad):

$$\sigma_{x_{in}}^{\text{Si, up}} \approx (z_{in}-z_B)\sigma_{\text{Si,200MeV}} = 102\,\mu\text{m} \quad ; \tag{66}$$

for the exit point the expression remains the same, although the deviation inside the silicon is on average larger, as the particles underwent energy loss inside the phantom. The average residual energy of the proton beam after traversing 10 cm of water is about 150 MeV, while $\sigma_{\text{Si,150MeV}} \approx 0.086° = 1.5$ mrad; the resulting term is:

$$\sigma_{x_{out}}^{\text{Si, down}} \approx (z_C - z_{out})\sigma_{\text{Si,150MeV}} = 135\,\mu\text{m} \quad . \tag{67}$$

The term related to the beam spread caused by the air is similar in form, and it gives a smaller contribution:

$$\sigma_{x_{in}}^{\text{Air}} \approx (z_{in}-z_B)\sigma_{\text{Air,90mm}} = 39\,\mu\text{m} \quad . \tag{68}$$

The sum in quadrature of the terms for the entrance and exit points respectively is:

$$\sigma_{x_{in}} = \left[\left(\sigma_{x_{in}}^{spatial}\right)^2 + \left(\sigma_{x_{in}}^{\text{Si, up}}\right)^2 + \left(\sigma_{x_{in}}^{\text{Air}}\right)^2\right]^{\frac{1}{2}} = 130\,\mu\text{m} \quad , \tag{69}$$

$$\sigma_{x_{out}} = \left[\left(\sigma_{x_{out}}^{spatial}\right)^2 + \left(\sigma_{x_{out}}^{\text{Si, up}}\right)^2 + \left(\sigma_{x_{in}}^{\text{Air}}\right)^2\right]^{\frac{1}{2}} = 157\,\mu\text{m} \quad . \tag{70}$$

The estimated values for the entrance and exit points uncertainties are of the same order of the Monte Carlo calculated RMS of Fig. 96, although smaller, which is acceptable as the error distributions are not Gaussian. To estimate the accuracy of the track reconstruction method, the calculation based on a Monte Carlo simulation seems therefore the more reasonable choice, although comparing the different contributions to the overall uncertainty could give an indication of the more critical sources of error. For instance, it can be noted that improving the tracker resolution (smaller pixels) would not lead to a significant improvement of the overall uncertainty, as the MCS inside the silicon planes contributes for a larger factor. For that reason, a thinner sensor is preferable, as it would improve the overall track reconstruction uncertainty (with half the thickness, as 50 µm in ALPIDE, the scattering would decrease by a factor $\approx \sqrt{2}$).

In Fig. 96 (b), the 1 σ uncertainty of the MLP calculation, from Eq. (62), is also shown; the MLP uncertainty is calculated considering particles entering in the phantom at the centre, travelling the longest possible path inside the object and therefore presenting the highest possible MLP uncertainty. As expected, the MLP uncertainty presents its maximum near the centre of the phantom, slightly downstream, similarly to the Monte Carlo RMS, with the highest value being 175 µm. This uncertainty is however unavoidable, as it derives from the intrinsic nature of MCS. Nevertheless, as shown in [71], the overall accuracy of the track reconstruction can be improved by discarding particles that underwent higher deviation angles, therefore considering events that lie inside the range where the MCS can be considered Gaussian distributed, as the entire MLP formalism is based upon that assumption.

Fig. 97 shows the error distributions at different depths along z, for the most precise reconstruction algorithm. The broader distribution is the one related to $z = 0$ (black line), corresponding

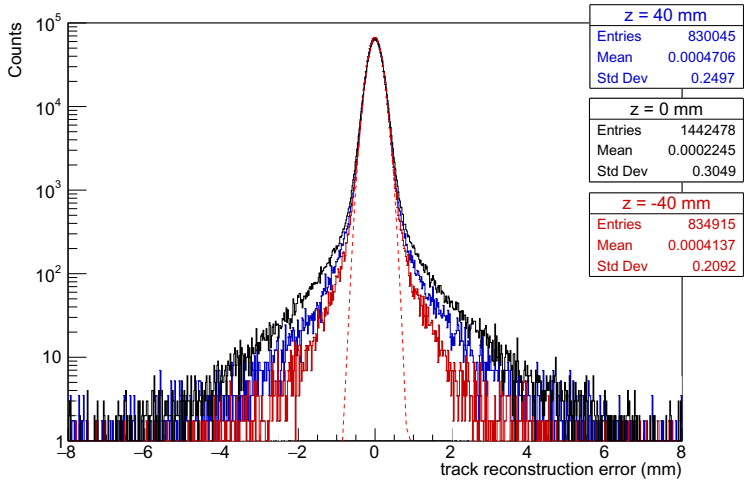


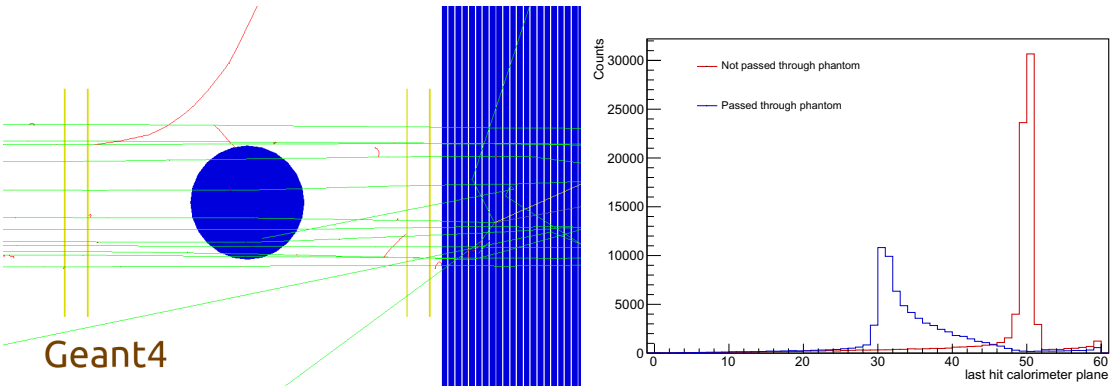
Fig. 97: Reconstruction error distribution at given z, obtained with the MLP calculation with shape informations. Black histogram is related to $z = 0$, corresponding to the centre of the phantom; blue histogram is related to $z = 40$ mm, downstream near the outer surface of the cylinder (radius $r = 50$ mm); red histogram is related to $z = -40$ mm, upstream. A Gaussian fit of the latter distribution is shown with a dashed red line.

to the centre of the phantom. The other distributions are related to symmetric positions inside the phantom volume, but close to the outer surface, respectively upstream, $z = -40$ cm (red line) and downstream, $z = 40$ cm (blue line). As already highlighted in Fig. 96, the downstream errors are generally larger with respect to their symmetric position upstream, as the energy loss degrades the accuracy of the reconstruction. Finally, the error distributions shows the non-Gaussian tails typical of MCS; furthermore, a Gaussian fit on the bulk of the distribution for $z = -40$ cm (red histogram) present a σ parameter of 176 µm, while the histogram calculated RMS (Std Dev in the plot) is 209 µm.

7.3 Shape recognition

As shown in the previous section, performing the particle trajectory estimation exploiting the knowledge of the object shape critically improves the accuracy of the reconstruction. In the case of a Monte Carlo simulation with a simple phantom, the exact shape of the object is known a-priori. However, when the simulated object is more complex or during an actual application, it is not straightforward to determine where each track enters and exits the object volume. In this section a method to recognize the external shape of the scanned target is proposed; such reconstruction, although coarse, could also represent the first step of a complete and refined imaging process.

The shape recognition method is based on the information retrieved by the calorimeter, which gives an indication on the residual energy of each particle. The initial GATE Geant4-based simulation, presented in Fig. 92, has been updated with the addition of a calorimeter replicating the iMPACT one. In the simulation the calorimeter is modeled as a total of 60 PVT planes, each one with a thickness of 5 mm, placed downstream after the second tracking section, as shown in Fig. 98 (a). For each particle, the last plane detecting any amount of deposited energy is registered, regardless whether the particle is the primary proton or a secondary particle produced in an interaction. The simulated calorimeter presents only the segmentation along the depth z, while it does not include the $x - y$ segmentation that the actual iMPACT calorimeter has.

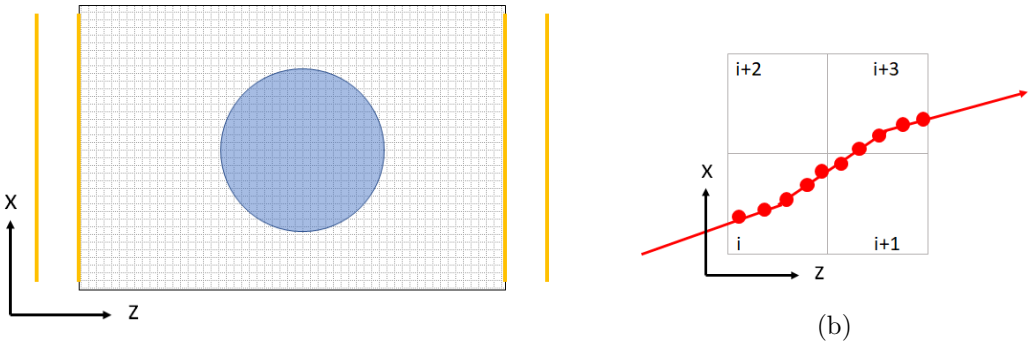


(a) (b)

Fig. 98: (a) Layout of the Monte Carlo simulation with the addition of the calorimeter (of which the initial planes are visible), made up of 60 PVT planes with 5 mm thickness; every other parameter is identical to the previous simulation in Fig. 92. (b) Distributions of the last calorimeter plane hit by a 200 MeV proton beam. Events are separated whether the proton passed through the phantom (blue histogram) or not (red histogram).

Fig. 98 (b) shows the distribution of the last plane registering deposited energy for each primary proton. The 200 MeV protons that do not pass through the phantom, included in the red histogram, present a sharp peak around the plane 50, corresponding to 25 cm in PVT, compatible with 26 cm of range in water observed in Fig. 91 (a).
Protons that do encounter the phantom present a generally lower range than the unaffected 200 MeV protons; the distribution is also broader, as the particles pass through the phantom at different thicknesses. The peak of the distribution in this case is around the plane 30, which is compatible with particles passing through the centre of the phantom, as it is placed about 10 cm (20 planes) before the full-energy peak.
In some cases, however, protons can interact via elastic or inelastic nuclear interaction with the nuclei in the calorimeter, which causes them to come to rest at a lower depth than usual or either generate secondaries that can deposit energy up to the last calorimeter plane. Moreover, in nuclear interaction events, secondary particles, especially gammas and neutrons, are likely to escape the calorimeter volume.
The proposed phantom shape recognition algorithm is based on the fraction of the protons that do not pass through the phantom. The idea is that these protons can be used to exclude from the total volume the areas where the particles have passed, while not losing a sensible quantity of energy, obtaining, in practice, a negative image of the object. The method has been based on full-energy protons for basically three reasons. These protons are easier to identify, in fact, considering the events with last hit plane between 48 to 51 (Fig. 98 (b)), only a fraction equal to 1.4% of the particles does not represent a full-energy proton. Additionally, exploiting the $x - y$ segmentation of the iMPACT calorimeter, not yet included in tis simulation, some secondary events generated by nuclear interactions could be recognized and excluded, cleaning up the spectrum from spurious events and improving the identification of full-energy protons. Secondly, the last hit calorimeter plane distribution of full-energy protons is independent from the phantom, as it is related only to the beam energy, making the method viable for a large variety of applications without the need of adjustments. Finally, to perform the object shape recognition, it would be necessary to employ the simple MLP track calculation, as the phantom would not be known yet; in this case the track reconstruction for particles travelling in air would be more accurate with respect to the ones encountering the phantom, as their trajectory is generally more linear.
To identify the phantom, the entire volume between the inner tracking planes is divided into

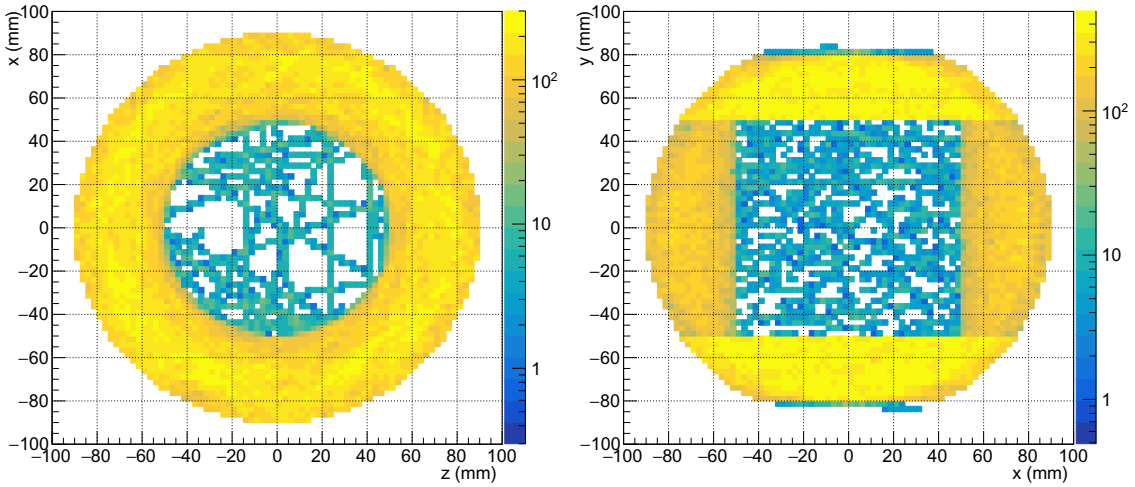
cubic voxels, as visualized in Fig. 99 (a). The dimensions of the voxels for this preliminary imaging step can be larger than the dimensions of the voxels in the definitive imaging. For this application the voxel pitch has been chosen to be 2.5 mm. The goal is to identify whether a given voxel contains, on average, a material denser than air.



(a) (b)

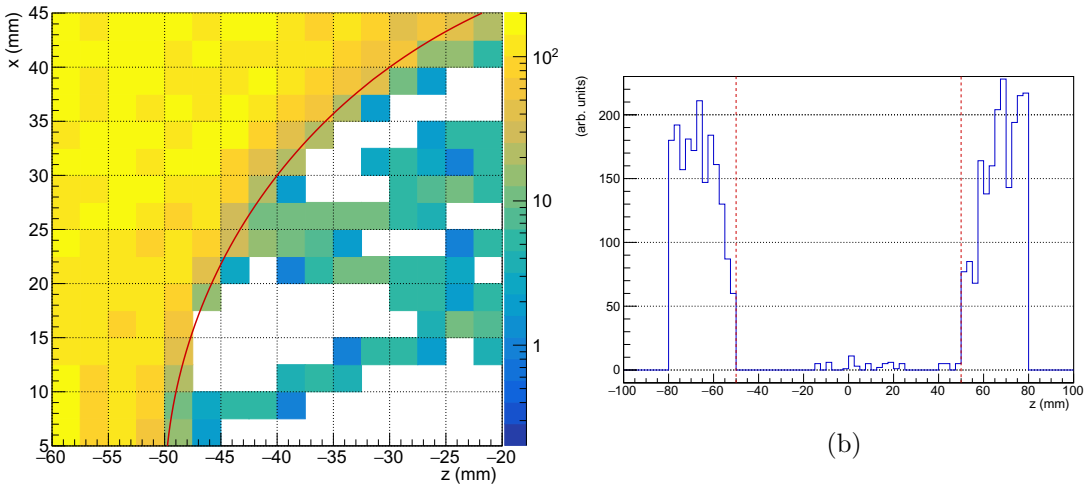
Fig. 99: (a) $x - z$ visualization of the voxelization of the volume between the inner tracking planes. (b) Representation of a trajectory calculation along z-steps and association of a hit point to a definite voxel.

Only the particles stopping between the calorimeter planes 48 and 51 are selected, and their trajectory is estimated using the simple MLP method without shape knowledge. The z-pitch of the trajectory calculation is chosen to be 5 times smaller than the voxel dimensions, equal to 500 µm in this example. A counter, initially equal to zero, is associated to each voxel. At each step of the track calculation, the voxel in which the point lies is identified, and its counter is increased by 1. A track can be found inside the same voxel for multiple consecutive steps, so the relative counter gets increased accordingly; in this way a higher weight is associated to the voxels in which the particle travels for a longer path, and vice versa. This procedure in schematized in Fig. 99 (b).
Fig. 100 (a) represents a 2D map of the voxel track counter values; the plot includes a single voxel thick slice in the $x - z$ plane at $y = 0$. In the graph, lower values correspond to higher probability of presence of material. Outer areas, with $x^2 + y^2 + z^2 > (80 \text{ cm})^2$, are excluded, as they are outside of the proton beam field of view. At the centre of the 2D map the circular shape, with 50 mm radius, of the cylinder top view is recognizable. A few of the tracks selected with the condition on the calorimeter are incorrectly identified, so some trajectories can be seen passing through the object. In Fig. 100 (b) a $x - y$ slice of the 3D map is shown; the slice is one voxel thick and it corresponds to the $z = 0$ coordinate. Similarly, in this second graph the square shadow of the lateral view of the cylinder is recognizable, with a small number of track passing through the object.
Fig. 101 (a) shows a detail of the $x - z$ slice of the object recognition 3D map, related to the outer circular surface of the cylindrical phantom. In the graph a circle with 50 mm of radius is superimposed, which highlights the real position of the cylindrical surface. The plot shows qualitatively how the color distribution follows the outline of the phantom shape. In Fig. 101 (b), the plot represents a 1-dimensional slice along z of the 3D object map, at $x = 0$ and $y = 0$, which shows the sharp transitions that takes place at the object boundaries, highlighted by the dashed lines at $z = \pm 50$ mm.
The graphs in Fig. 100 and 101 have been generated with a total of $3.6 \cdot 10^5$ simulated particles, of which about a half passing through the phantom; moreover, about 20% of them underwent nuclear interaction inside the phantom and were discarded. Considering instead a real application, the imaging of a similar target would require a number of acceptable particles in the order of 10^9. Even factoring, for the real application, a voxelization with 1 mm spacing, as opposed



(a) (b)

Fig. 100: (a) 2D $x - z$ representation of a slice of the 3D map of the voxel track counter values; the chosen slice is 1 voxel thick in the y direction and corresponds to $y = 0$. (b) 1 voxel thick 2D $x - y$ slice, at depth $z = 0$. The parameter value represented by the colour palette is shown in logarithmic scale.



(a) (b)

Fig. 101: (a) Particular of the 2D slice in the $x - z$ plane from Fig. 100; the red line outlines the phantom cylindrical surface. (b) 1D slice along z ($x = y = 0$), the vertical axis represents the shape recognition value (track counter) per each voxel; red dashed lines indicates the cylindrical phantom boundaries.

to 2.5 mm used in the simulation, the available statistic would be around 300 times higher than the one used to produce the plots presented here. Therefore, in a concrete application, the shape recognition is expected to be even sharper, with respect to the one shown in Fig. 100 and 101.

Based on the 1D slice in Fig. 101 (b), a voxel is considered internal to the phantom when its track counter value is lower than 30. In Fig. 102 a 3D representation of the voxels that satisfy the condition is shown. The shape outlined by the voxels looks very close to the original cylinder.

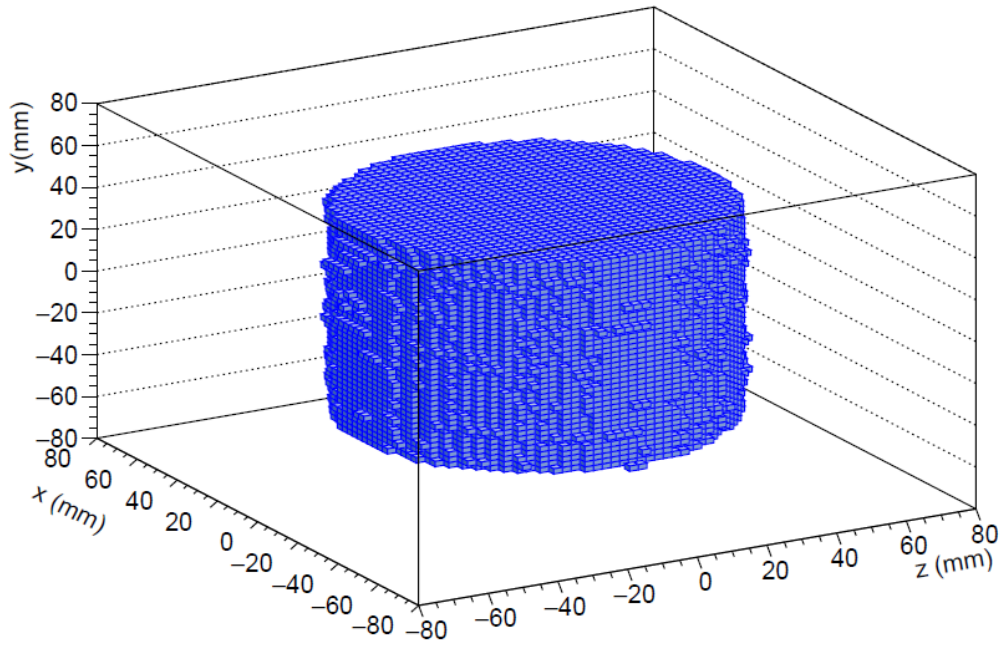


Fig. 102: 3D representation of the reconstructed cylindrical phantom, visualized as the set of voxels that have been identified as containing a denser material than air.

The shape recognition method has also been tested with a more complex and less symmetric shape. An homogeneous water box was defined in the simulation, as the phantom. The dimensions of the box are $10 \times 10 \times 4$ cm^3. Additionally, the phantom includes a cavity that passes through its entire depth, with $5 \times 5 \times 4$ cm^3 dimensions. The cavity is translated by 1 cm towards the positive x, to increase the asymmetry for rotations around the vertical axis. In Fig. 103 the plots relative to the track counting 3D map are shown. In particular Fig. 103 (a) shows the 2D $x - z$ slice at $y = 0$, with superimposed the outline (red dashed line) of the object section; (b) shows the $x - y$ slice at $z = 0$, (c) shows the 1D slice along x at $z = y = 0$, with the object section (red dashed lines); finally (d) shows the 3D visualization of the voxels containing material.
The shape recognition is not able to perfectly identify the sharp edges of the object. This can be explained by the fact that the calorimeter, with its 5 mm of z-segmentation, is not able to distinguish too small energy losses. On the other hand, the performances of method are not affected by the asymmetry of the object.
After the object recognition procedure, the particle trajectory can be estimated using the more refined MLP algorithm presented in Sec. 7.2, which has been shown to provide sensibly more accurate results than a calculation not based on phantom shape knowledge.

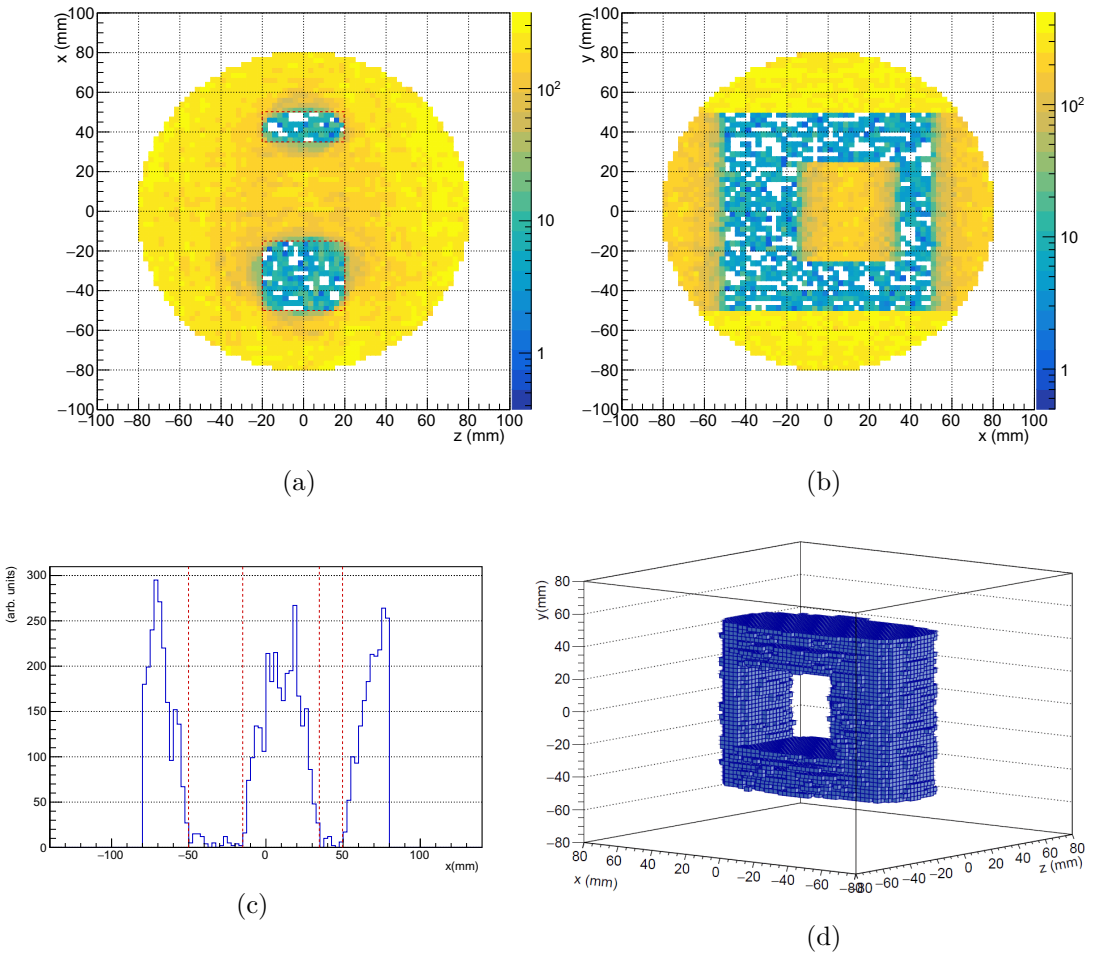


Fig. 103: (a) 2D $x - z$ representation of a slice of the 3D map of the voxel track counter values; the slice is 1 voxel thick in the y direction and corresponds to $y = 0$. (b) 1 voxel thick 2D $x - y$ slice, at depth $z = 0$. The parameter value represented by the colour palette is shown in logarithmic scale. (c) 1D slice along x ($y = z = 0$), the vertical axis track counter per each voxel; red dashed lines indicates the phantom boundaries. (d) 3D representation of the reconstructed phantom, visualized as the set of voxels that have been identified as containing a denser material than air.

7.4 Outlooks

The Most Likely Proton path formalism has been implemented in a setup modeling the iMPACT scanner; a 5-degree polynomial parameterization of the $\beta^2(z)p^2(z)$ is appropriate in the energy range of interest (between 200 MeV and 250 MeV), which substantially reduce the computational time. The proton trajectory reconstruction is expected to be accurate enough to be effectively employed, with an uncertainty lower than 300 µm at the centre of the phantom. Moreover, in the simulation the sensor thickness has been overestimated (100 µm vs 50 µm of ALPIDE), therefore the actual accuracy is expected to be better. The spatial resolution of ALPIDE is more than enough to be a minor contribution to the overall reconstruction accuracy.
To achieve these high reconstruction accuracies, a level-zero imaging procedure, to evaluate the shape of the phantom, is necessary. The proposed shape-recognition algorithm is able to precisely identify objects even with asymmetrical features; the method is based on a negative image, recognizing volumes where no material is present, therefore it is expected to be suitable

for a large variety of objects. The algorithm efficiency can be further improved by employing the $x - y$ segmentation of the iMPACT calorimeter, allowing for the identification of nuclear interactions and a more defined energy spectrum. In a real-world application the available particle statistics is also expected to be orders of magnitude higher than the shown Monte Carlo simulation, allowing for smaller voxels and a more defined image. Application on more complex objects are planned.
The presented implementation of Most Likely Proton path formalism will represent the starting point of the data analysis of the complete iMPACT scanner, with the object-shape algorithm being the first step in the image reconstruction procedure. These calculation are planned to be implemented in a parallel GPU-based computational environment.

8 Alignment of the ALICE ITS with cosmic-ray tracks

As introduced in Sec. 4.3.1, the iMPACT scanner employs the ALPIDE sensor in the tracking layers of the first prototype. The ALPIDE sensor has been developed by the ALICE collaboration for the upgrade of the Inner Tracking System (ITS), and it is the result of an international effort by many research institutions. The iMPACT group actively participated to the ALPIDE development and commissioning, as well as to the ITS design and building. As the iMPACT tracker basically acts as a small-scale high-energy physics tracker, exploiting the techniques and methods used in particle physics for the own iMPACT tracking purposes was a natural choice. To better familiarize with and understand such techniques, specific studies on the completed ALICE ITS were performed, focusing in particular on whole assembly sensors alignment in 3D space.

8.1 The LHC and its experiments

LHC (Large Hadron Collider) is the largest and highest energy particle accelerator in history. The beam pipe ring has a diameter of about 27 km and it is installed underground, at around 100 m of depth. LHC has been built at CERN (Conseil Européen pour la Recherche Nucléaire) between the French-Swiss border, near the city of Geneva [168]. The underground gallery hosting the LHC was originally excavated to host the Large Electron Positron Collider (LEP), decommissioned in 2001 to allow the construction of the present accelerator. The LHC main purpose is to study proton-proton, proton-ion and ion-ion collisions, up to Pb nuclei. At different point along the ring four large experiments have been installed. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are general-purpose detectors optimized for proton-proton collisions, searching for new-physics phenomenon. The two collaborations announced the experimental observation of the Higgs boson in 2012 [169] [170]. The LHCb experiment (Large Hadron Collider beauty) is instead specialized in the study of CP violation in interactions of b-hadrons, i.e. hadrons containing the *bottom* (or *beauty*) quark. Finally, the ALICE experiment (A Large Ion Collider Experiment) is devoted to high-multiplicity events detection, and the detailed study of Quark-Gluon Plasma (QGP). In particular, ALICE is designed to look at Pb-Pb collision events: inside the LHC ring, two particle beams travel in opposite directions at the same time, in separate beam pipes. The high energy particle beams are bent to follow the LHC circumference by magnetic fields produced within super-conducting magnets: the two beams are deviated at for points along the LHC ring, to have them colliding head to head at the interaction points of the four experiments. LHC is able to produce a luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} at a centre-of-mass energy of $\sqrt{s} = 13$ TeV [171] for proton-proton collisions, and a luminosity of $\mathcal{L} = 10^{27}$ cm^{-2}s^{-1} for Pb-Pb at $\sqrt{s_{NN}} = 5.02$ TeV/nucleon [172].
In Fig. 104 the LHC injection system is shown, as well as the positioning of the four main experiments along the ring. The main accelerator operates in conjunction with many auxiliary accelerators, such as the LINear ACcelerator 2 (LINAC2), the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS); these accelerators are responsible for the pre-acceleration of protons before their injection into the LHC. Pb ions instead pass through the LINAC3 and Low Energy Ion Ring (LEIR) before being injected into PS and SPS.

8.2 The ALICE experiment

ALICE is a general-purpose, particles detector which optimized for the study of strongly interacting matter and Quark-Gluon Plasma (QGP) at extremely high energy densities and temperatures in nucleus-nucleus collisions. ALICE focuses in harvesting the experimental necessary to improve our knowledge of the Quantum Chromo-Dynamics (QCD), the branch of the Standard Model which describes the strong interactions [174]. It is designed to detect particles,

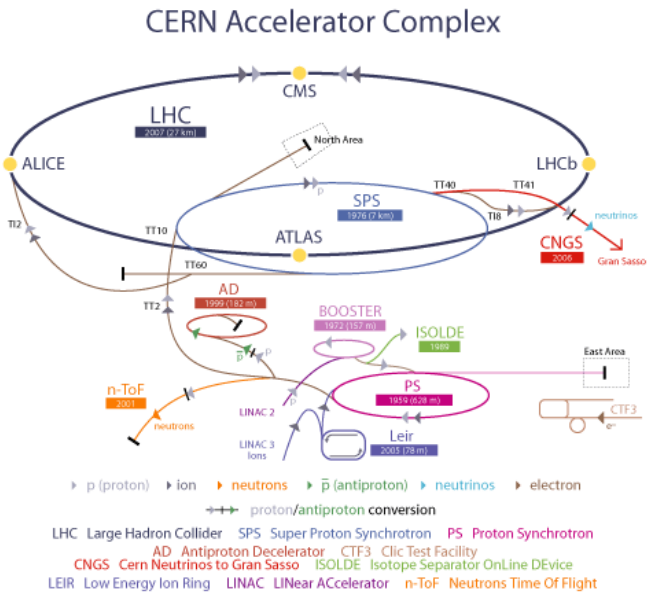


Fig. 104: LHC injection system and position of the main experiments [173].

such as hadrons, electrons, muons, and photons, produced in the collision of heavy nuclei, in particular Pb-Pb, with the challenge of dealing with the highest multiplicities at LHC. The physics goals also foresees the study of lighter ions collisions, as well as lower energy events, and dedicated proton-nucleus and proton-proton runs. Due to its focus on ion-ion collisions, strong-interaction phenomenology, and on the intermediate rapidity range, ALICE can be seen as a complementary experiment respect to the other LHC detectors [174].
The first idea of a heavy-ion detector at LHC was expressed in 1990 [175], and the experiment was approved in 1997 [174]. The expected performances and the physics plan are presented in the Physics Performance Reports [176] [177].

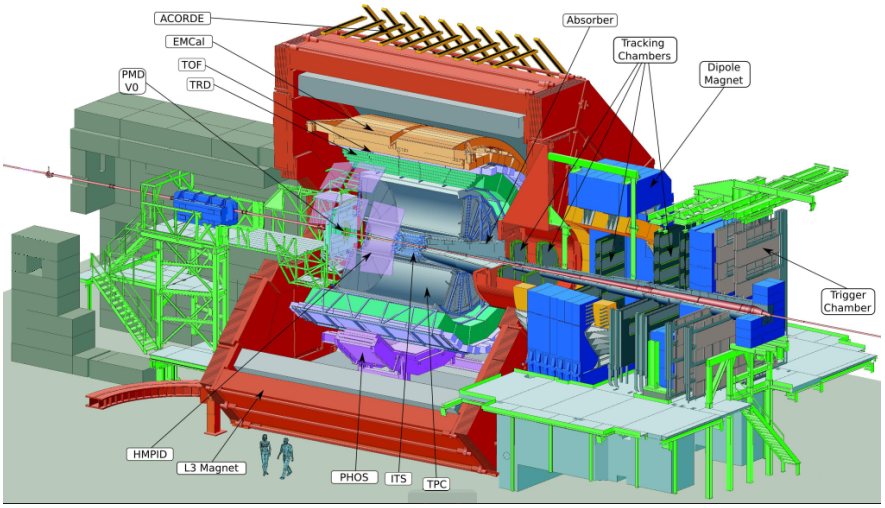


Fig. 105: Schematic diagram of the ALICE experiment.

In Fig. 105 the schematic diagram of the ALICE experiment is shown. The detector dimensions are $16 \times 16 \times 26$ m^3, and it weights about 10^7 kg. The ALICE layout includes a central barrel,

for detecting hadrons, electrons, and photons, covering polar angles between 45° and 135°, and a forward muon spectrometer [174]. The central barrel is enclosed in a solenoid magnet, inherited from the L3 experiment at LEP, which can generate a magnetic field up to 0.5 T. The barrel contains several detector systems, from the closest to the interaction point: the Inner Tracking System (ITS), presented more in details in Sec. 8.3, a cylindrical Time Projection Chamber (TPC), three particle identification arrays of Time-of-Flight (TOF) detectors, a Ring Imaging Cherenkov (HMPID), Transition Radiation (TRD) detectors, and two electromagnetic calorimeters (PHOS and EMCal). The forward muon spectrometer consists of a dipole magnet, absorbers, and fourteen planes of tracking and triggering chambers. On top of the magnet, an array of plastic scintillators (ACORDE) measures cosmic rays for commissioning, calibration and dedicated cosmic-ray studies.
Concurrently to the LHC upgrade to High-Luminosity LHC (HL-LHC), planned to become operational during 2022 [178], the ALICE experiment went through a major upgrade of the original detector, by enhancing its low-momentum vertexing and tracking capability, and improving the whole data taking operations to cope with the substantially higher rates that the upgraded LHC will deliver, up to 50 kHz of Pb-Pb [121].

8.3 The Inner Tracking System ITS

The ITS is the innermost detector system present in ALICE, placed in close proximity to the pipe and surrounded by the Time Projection Chamber (TPC). Its main tasks include localizing the primary vertex with a resolution better than 100 µm, reconstructing the secondary vertices, tracking and identifying particles with momentum below 200 MeV/c. Furthermore, it operates in conjunction with the TPC to resolve and reconstruct particles momentum and trajectory. The ITS contributes to practically all physics channels investigated by the ALICE experiment measurements [174].
The first version of the ITS had six cylindrical layers of silicon detectors of different technologies. The first two layers were equipped with Silicon Pixel Detectors, the following two did use Silicon Drift Detectors, and the outermost two layers were instrumented with double-sided Silicon micro-Strip Detectors. Additionally, the four outermost layers of original ITS were equipped with analog read-out, so that energy deposition measurements were possible.
The ITS upgrade, part of the general ALICE upgrade, had to meet the new challenges posed by the increased track density and collision rate. The different silicon detector technologies of the old ITS have been replaced by a single sensor solution, where the same Monolithic Active Pixel Sensors (MAPS) chip equips the whole system. ALICE has been the first of the four large experiments at the LHC to devise a large silicon tracker (10 m) entirely based on this technology [179]. A number of different MAPS implementations were considered and tested, and ultimately the ALPIDE chip was chosen for the task [122]. The ALPIDE sensor has been presented in Sec. 4.3.1.
Using monolithic sensors allows reducing the overall material budget per layer by a factor of seven in comparison to the previous ITS (50 µm instead of 350 µm), accounting for the silicon only, allowing for a significant improvement on the tracking performance and momentum resolution, while at the same time increasing the pixel density by a factor 50 [121]. On the other hand, the ALPIDE digital readout does not allow for an energy loss measurement.
In Fig. 106 and 107 the layout of the new ITS is shown. The system includes seven concentric cylindrical layers, of which the innermost is placed 22 mm from the centre of the beam line, and the outermost at 430 mm. The three innermost Layers (L0 to L2) are included in the Inner Barrel (or IB), while the next two (L3 and L4) and the remaining two (L5 and L6) are coupled in the so-called Middle and Outer Layers respectively (ML and OL); ML and OL are often grouped together as Outer Barrel (or OB). Single ALPIDE sensors are arranged along the beam directions in mechanically independent structures called staves, which cover the entire length of the respective layer; staves forming the Inner and Outer Barrel present some construction differences; Inner Barrel staves have been presented in Sec. 4.3.2.

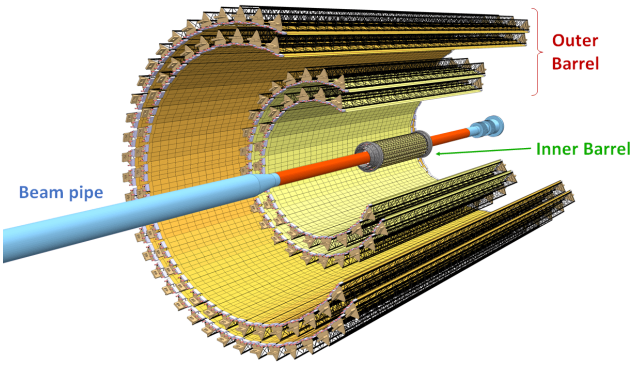


Fig. 106: 3D visualization of the upgraded ALICE ITS (Inner Tracking System) [121]

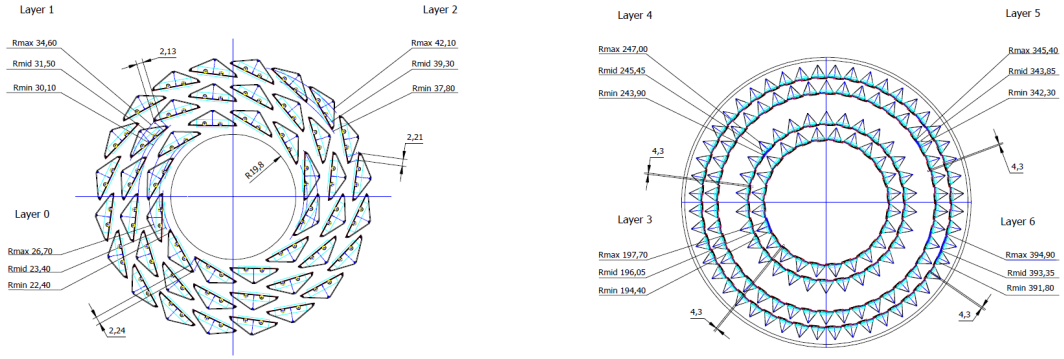


Fig. 107: Schematic view of the cross section of the ITS (a) Inner Barrel (IB) and (b) Outer Barrel (OB) [121]. The Inner Barrel includes Layer 0, 1 and 2; the Outer Barrel includes Layers 3, 4, 5 and 6.

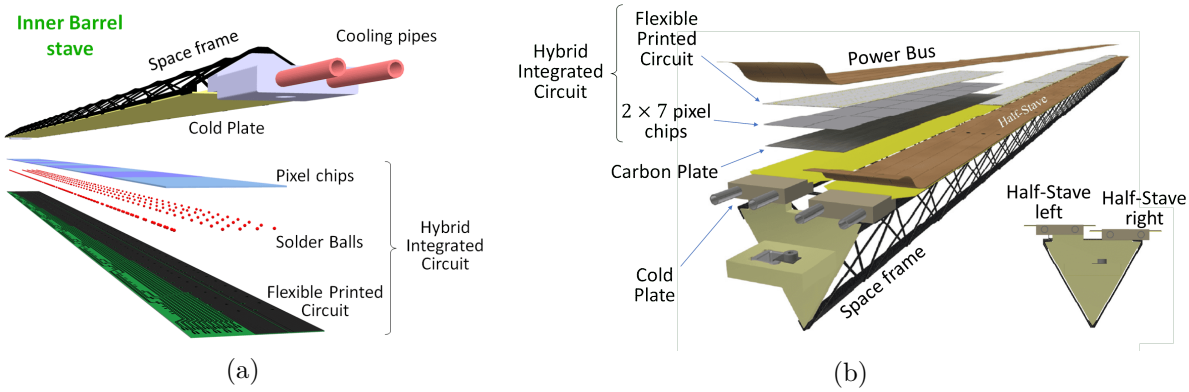


Fig. 108: Schematic drawing of the Outer Barrel staves (right), compared with the Inner Barrel [121].

Fig. 107 shows the schematic layout of an Outer Barrel stave and its components, compared with an Inner Barrel stave. As for the Inner Barrel, the mechanical stability of the Outer Barrel staves is provided by a truss-like lightweight structure based on carbon fiber material. staves from the Outer Barrel are further segmented azimuthally into two half-staves, which are places side by side along the stave length, as illustrated in Fig. 107 (b). In the Outer Barrel, each stave includes two cold plates, one per half-stave, which contain cooling pipes. In Outer Barrel

staves the two Cold Plates are attached underneath the Space Frame, while in Inner Barrel staves it is embedded in the frame [121].
Half-staves are additionally divided longitudinally into *modules*, which contain an array of 2×7 single sensors. Each module is an independent entity, and includes its own Hybrid Integrated Circuit (HIC), a Flexible Printed Circuit (FPC) and a carbon plate, which ensures mechanical rigidity to the module itself. On the contrary, in the Inner Barrel staves, a single FPC running for the entire stave length connects all the 9 sensors directly mounted on the stave space frame (Fig. 107 (a)). The Outer Barrel staves (L6 and L5) share the same basic structure with the staves in Layer 4 and 3 (ML), although the former are almost twice as long as the latter. staves of the Outer Layers and the Middle Layers include, respectively, seven and four sensor modules. Given the longer length of Outer Barrel staves, with respect to Inner Barrel, the former require an additional *Power Bus* to distribute digital and analog power, and overcome the voltage drop along the whole length of the half-stave. The geometrical parameter, and sensor counts for each layer, are summarized in Tab. 7.

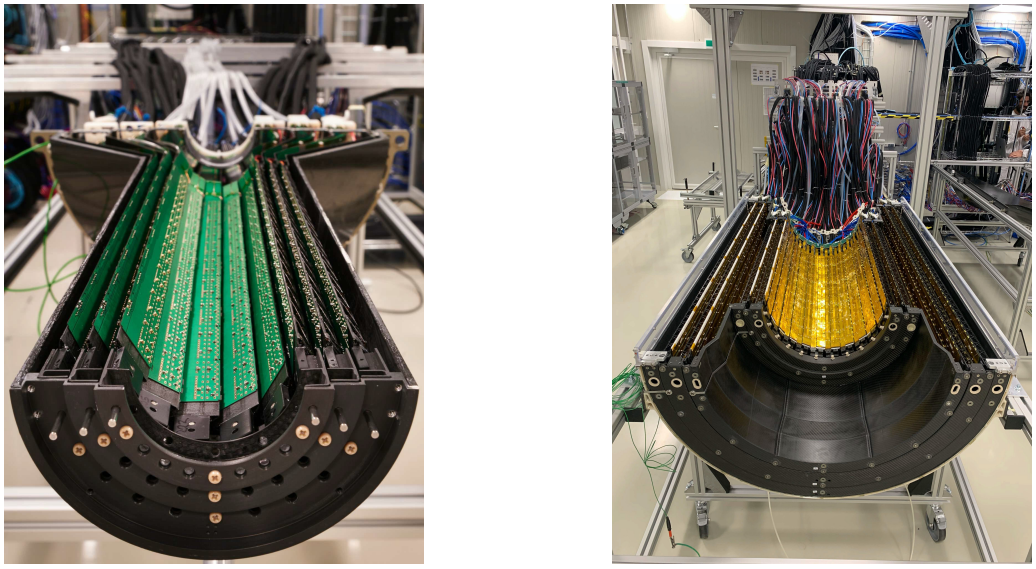
	Layer	Length (mm)	Radius (mm)			Number of			
			min	mid	max	Staves	Half-Staves	Modules	sensors
IL	0	271	22.4	23.4	26.7	12	/	/	108 ⟨9⟩
	1	271	30.1	31.5	34.6	16	/	/	144 ⟨9⟩
	2	271	37.8	39.3	42.1	20	/	/	180 ⟨9⟩
ML	3	843	194.4	196.05	197.7	24	48	192 ⟨8⟩	2688 ⟨112⟩
	4	843	243.9	245.45	247	30	60	240 ⟨8⟩	3360 ⟨112⟩
OL	5	1475	342.3	343.85	345.4	42	84	588 ⟨14⟩	8232 ⟨196⟩
	6	1475	391.8	393.35	394.9	48	96	672 ⟨14⟩	9408 ⟨196⟩
	tot.					192	288	1692	24120

Tab. 7: Summary of the geometrical parameters of the ITS, and elements count. The length measures the stave dimension along the beam direction z; the radii are listed separately as the distance from the beam pipe centre of the innermost and outermost points of each stave's sensible surface, as well as the average distance. The number of modules and sensors are tabulates as the total number in the layer and in parenthesis the number per stave [180].

The construction of the Inner Barrel, Middle Layers and Outer Layers, including spares, finished in July 2019, October 2019 and December 2019 respectively [180]. Initially the Barrels have been assembled in the form of two Inner Half-Barrels and two Outer Half-Barrels, and installed in a clean room at the main CERN site, for the individual on-surface commissioning. In Fig. 109 two assembled Inner and Outer Half-Barrels are shown, respectively in (a) and (b). The on-surface commissioning included verification of the power consumption, the control communication, and the high speed read-out links. Additionally, threshold tuning and fake-hit rate measurements were performed. After the readout chain for both the Inner and Outer Barrels has been demonstrated to work, a campaign to observe the first particle tracks from cosmic muons have been initiated [180]. The final assembly of the ITS and its installation in the centre of the ALICE experiment took place between spring and summer 2021.

8.4 ITS Alignment survey with cosmic-ray tracks

The alignment procedures of a tracking detector are necessary to ensure that the reconstructed track parameters accuracy is not significantly affected by the inevitable mismatches between the ideal, as-per-design, and the real, post-assembly overall 3D geometry of the whole assembly. For the ALICE ITS, this maximum acceptable degradation has been conventionally set to 20% [181].
To better gauge the challenge, that is equivalent to measure the position of every single sensor within a volume of about 1.5 m in length and 1 m in diameter with a precision better than 5 µm



(a) (b)

Fig. 109: Completely assembled Inner (a) and Outer (b) Half-Barrels, store on-surface in a cleanroom at the main CERN site [180].

for the innermost sensors. During the staves assembly phase, the correct positioning of every single sensors was ensured by using lithography reference marks present on the visible side of each chip. The positioning of the reference marks is strictly related to the pixel matrix, and the points were scanned by 3-dimensional Control Measuring Machines (CMM) to determine the relative position of the sensors in a stave [121]. After the assembly of the staves in Half-Barrels structures, a similar procedure was performed using reference marks present on the barrels end-wheels, on the FPC, and the accessible marks on the sensors [121]. Additionally, barrel end-wheels include spots for contact-measurements. After the installation of the ITS in the cavern, the position of the elements has been surveyed by a laser system, through the line of sight along the beam pipe, although the small angle of view makes this measurement more challenging. These metrology and survey data serve as starting points for the final alignment, which is based on tracks reconstruction. The goal of the alignment procedures is to evaluate the position of the silicon sensors with an accuracy of the order of few microns for the Inner Barrel and few tens of microns for the Outer Barrel [121].
During the 2020-2021 on-surface commissioning of the ALICE ITS, once the correct operation of the single elements was assessed, the open half-barrels were put through their paces with the detection of cosmic tracks. In addition to sensors detection efficiency, fake-hit rate and calibration tests, the cosmic-ray tracks were employed to initiate the evaluation of the alignment precision of the mechanical assembly of the Half-Barrel structures. These studies will be reported in the present Section. This on-surface alignment survey constitutes a preliminary assessment on the assembly precision, as the final and systematic study can only be performed once the entire system is installed in its definitive location, inside the experiment. Moreover, given the partial assembly in Half-Barrels, an on-surface study is not able to provide information on the final relative position of these portions of the ITS, let alone the positioning of the ITS with respect to the other ALICE detector systems. Additional displacements are also to be expected as the result of the relocation of the assembled Half-Barrels from the CERN main site to the ALICE location, a few kilometers away. Nonetheless, the preliminary alignment survey helped developing alignment methods and procedures, and highlighted critical aspects that will be useful in the final alignment phase. The final alignment will further exploit particle tracks from collision

events, also with the magnetic field active. The relative position of the elements is anyway foreseen to shift with time due to factors as vibration, thermo-elastic movement, weight load, and settling of different parts. Alignment runs will therefore be periodically repeated, and given the complications of accessing the various elements for optical measurements once the entire ALICE detector is in place, particle tracks will be the main tools for such periodic alignment re-assessment [121].
The data used to test the preliminary alignment methods presented in this section are relative to the on-surface commissioning phase; in particular, the data used for the Inner Barrel (either IB-bottom or IB-top) have been acquired in the period between September and October 2020, while the data employed for the Outer Barrel have been recorded between November and December of the same year. The data analysis has been performed using the software framework O^2 (Online - Offline computing system) designed for the ALICE upgrade; the O^2 framework includes the general computing functionalities needed in a particle physics experiment, relative to read-out, event building, data recording, detector calibration, data reconstruction, physics simulation and analysis [182].
Raw data for analysis are aggregated as clusters of adjacent hit pixels; the spatial coordinates of the hit are calculated as the approximated geometrical centre of the cluster (due to the binary output of the ALPIDE sensor, a true center-of-mass derivation is not possible). Clusters coordinates are stored in the *Local Frame* of reference of the sensor that registered the hit; this reference system is schematized in Fig. 110: the Local Frame has its origin in the centre of the sensor volume, at half silicon depth (25 µm), the $x_{loc} - z_{loc}$ plane is parallel to the sensor surface, with z_{loc} parallel to the global z (beam direction) along the 30 mm side of the sensor. For each sensor a different Local Frame is defined.

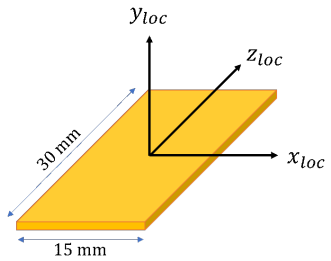
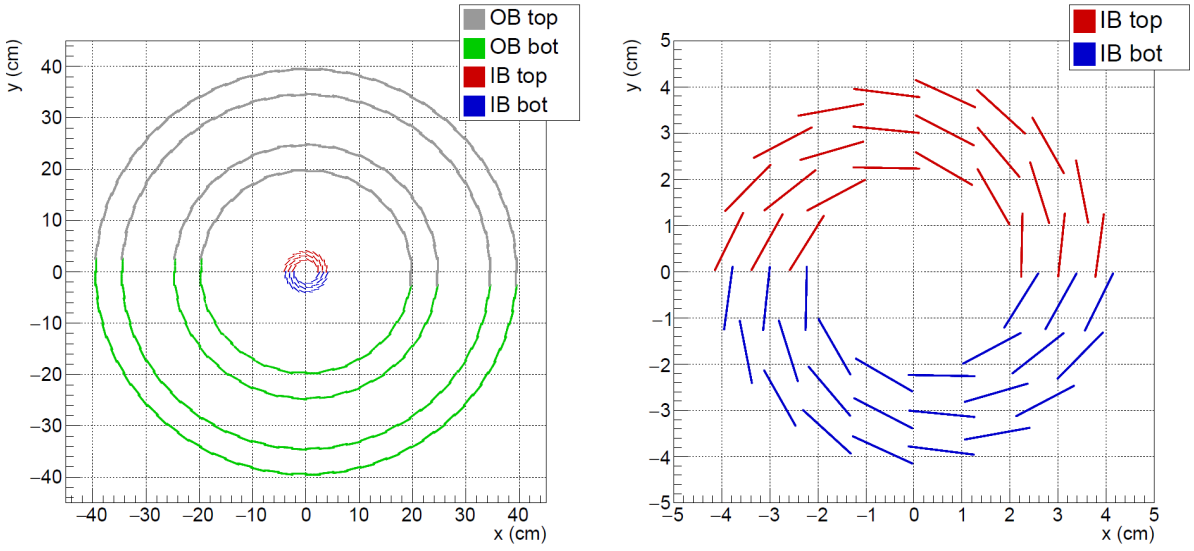


Fig. 110: Local Frame of reference of a given sensor. The frame origin is placed at the geometric centre of the sensor volume, the $x_{loc} - z_{loc}$ plane lays parallel to the sensor surface at half of the silicon depth (25 µm), z_{loc} is parallel to the global z and y_{loc} is normal to the surface.

The position of a cluster in the *Global Frame* is retrieved by applying roto-translation transformations described by matrices, which are calculated from a geometry `.root` file. In the geometry file the position of each sensor is tabulated. For the alignment analysis, the ideal geometry is considered, while after the definitive alignment survey of the ITS, a final geometry file describing the real positions of the various elements will be produced and used for the following analysis. In the ALICE software framework O^2, a third system of reference, the *Tracking Frame*, is defined: the Tracking Frame is obtained from the Global Frame by rotating around the z axis until the x_{track} direction is normal to the sensor plane.
Fig. 111 (a) shows a $x - y$ overview of the positions of the detected clusters, which outlines the stave placement in the ITS, as well as the divisions of the layers into Half-Barrels. The three Inner Layers are highlighted in Fig. 111 (b), which shows the spiral arrangement of the staves, also known as *Turbo* design, which avoids blind tracking spots.
In the on-surface commissioning datasets, the threshold scans and the noisy pixels rejection were not definitive yet, so some spurious clusters were present. To eliminate these fake hits, 1-cluster events were not considered for the calculations. This would, consequently, also reject some actual events; however, in the alignment analysis, rejecting spurious events is more critical

than achieving a high reconstruction efficiency, i.e. it is preferable to reject an actual track than accepting a fake one.



(a) (b)

Fig. 111: (a) Global $x - y$ view of the clusters position; the distribution outlines the shape of the 7 layers forming the ITS. (b) Highlight of the 3 layers of the Inner Barrel and the spiral arrangement of the IB staves. During the on-surface commissioning the four sections (IB-bottom, IB-top, OB-bottom and OB-top) are assembled and read-out independently.

As introduced earlier, during the commissioning the ITS was divided into four Half-Barrels, which are read-out independently; therefore, to produce datasets including the entire system, multiple independent datasets need to be merged.
As mentioned in Sec. 4.3.1, in the ALPIDE sensor the hit storage is determined by an external trigger signal; during the acquisitions the trigger is sent at a regular frequency, with a period longer than the front-end electronic shaping time. For the cosmics acquisition, the trigger period was set to 22 µs. Clusters are labeled accordingly to these time windows, or Time Frames, with no additional timing information. In the datasets, some events were recorded twice in adjacent Time Frames, which is a foreseen occurrence when the trigger falls during the event signal formation. Hence, to avoid considering twice the same event, if a Time Frame contains a track candidate, the following Time Frame is rejected. This causes virtually no loss of actual events, as the cosmic tracks detection rate was calculated around the order of 2/s on the entire ITS.
The basic quantity used for the alignment study is the residual, i.e. the point-to-point distance between the measured position of a cluster and the coordinates of the expected hit position on the same sensor, assuming an ideal as-per-design geometry. The concept of the residual calculation is sketched in Fig. 112: given a particle track, producing in this particular case three clusters on different Layers, and assuming a displacement of the sensor on Layer 1, the measured cluster would be generated at the interception between the track and the real position of the sensor. The expected cluster position is then derived considering the interception between the track and the ideal position of the sensor plane. Finally, the residual is calculated as the difference between the coordinates of the expected and measured points. Residuals are preferably expressed in the Local Frame of the sensor, as it indicates more intuitively the direction of the displacement.
The magnitude of a residual calculated for a given sensor is not a direct measure of the dis-

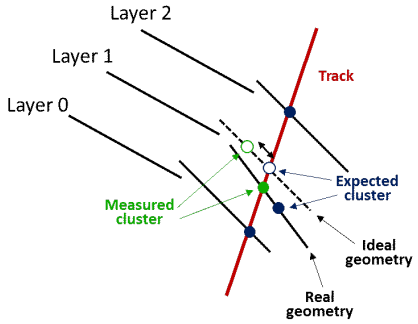


Fig. 112: Concept of residual used to estimate sensors alignment. The expected cluster position is calculated as the interception between the track and the ideal plane of the sensor. The residual is calculated, in local coordinates, as the spatial difference between the expected and measured positions.

placement of the sensor itself but, more precisely, the convolution of the displacements of all the sensors traversed by the track. For instance, considering the case illustrated in Fig. 112, to calculate the residual on the sensor on L2, the track would be traced using the measured cluster on L1; the resulting large residual on the L2 sensor, would be in this case caused by the sensor on L1, projecting its own displacement onto the facing sensors on L2, and not by an actual displacement on the latter. After the residual calculation, an iterative procedure is therefore necessary to disentangle the various contributions, and estimate the displacements of each different sensors.
The first step for the residual calculation is the track-finding procedure. At the moment of the alignment analysis, the datasets were not post-processed, and no track-finding was performed; moreover, no definitive tracking reconstruction algorithm was yet defined in the updated ALICE O^2 software framework. A basic custom tracking algorithm was therefore built for the task of identifying cosmic tracks.

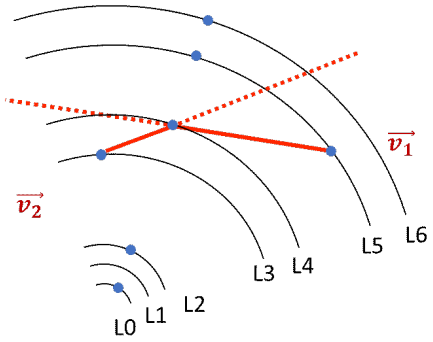


Fig. 113: Sketch of the track-finding procedure. The alignment of a clusters triplet is measured using the angle between the two identified vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, as expressed in Eq. (71).

Tracks candidates are built from the clusters registered inside the same Time Frame. Fig. 113 shows the situation in a typical Time Frame, where a number of clusters are distributed between the 7 Layers; the goal is to identify the clusters originated by cosmic muons tracks, discriminating the hits generated by noise or other particles. Cosmic muons are expected to generate an approximately straight track, aside from scattering inside the various materials present in ITS, spatial resolution of the sensor, and the aforementioned assembly misalignment.
The procedure starts by identifying a track *seed*, defined by three aligned clusters in three dif-

ferent Layers. The algorithm cycles over all the possible cluster triplets in a single TF, and calculates the angle between the two vectors $\mathbf{v}_1$ and for each triplet $\mathbf{v}_2$ identified by the three points, as illustrated in Fig. 113:

$$\mathbf{v}_1 \times \mathbf{v}_2 = \begin{pmatrix} x_b - x_a \\ y_b - y_a \\ z_b - z_a \end{pmatrix} \times \begin{pmatrix} x_c - x_b \\ y_c - y_b \\ z_c - z_b \end{pmatrix} = \hat{n}|\mathbf{v}_1||\mathbf{v}_2| \sin\alpha \quad , \tag{71}$$

where $x_{a,b,c}$, $y_{a,b,c}$ and $z_{a,b,c}$ are the coordinates of the three clusters ordered in a innermost to uttermost fashion, $\hat{n}$ is the versor perpendicular to the plane identified by $\mathbf{v}_1$ and $\mathbf{v}_2$. The angle α between the two vectors is found as:

$$\frac{|\mathbf{v}_1 \times \mathbf{v}_2|^2}{|\mathbf{v}_1|^2|\mathbf{v}_2|^2} = \sin^2\alpha \quad . \tag{72}$$

The clusters triplet with the lowest $\sin^2\alpha$ in a Time Frame is taken as a track seed, which is in turn used to build a complete track by attaching additional aligned clusters to it. Only triplets with a $\sin^2\alpha$ lower than a given threshold, empirically set at 0.1, are considered.

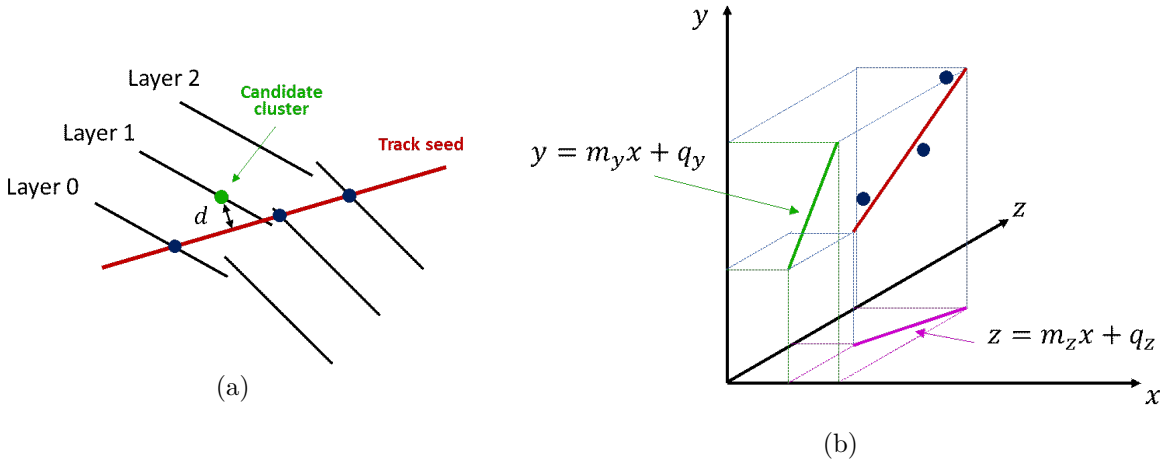
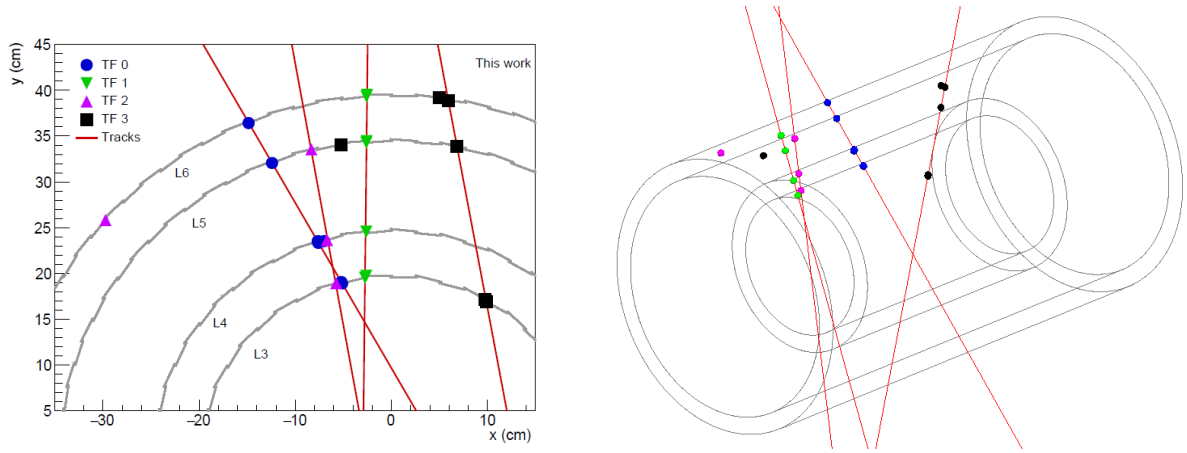


Fig. 114: (a) Candidate clusters are attached to a track seed if the 3D point-line distance d is < 1 mm. (b) The track seed parameters are obtained fitting the triplet separately in the $x - y$ (green line) and $x - z$ (violet line) projections.

For every remaining clusters present in the Time Frame the 3-dimensional line-point distance from the track seed is calculated, as represented in Fig. 114 (a). The track seed parameters are obtained by fitting the cluster triplet separately in the $x - y$ and $x - z$ projections, as shown in Fig. 114 (b). Two equations with a total of four parameters are derived:

$$\begin{cases} y = m_y x + q_y \\ z = m_z x + q_z \end{cases} \tag{73}$$

and the 3D distance is calculated from the point-line distances in the two planes. A candidate cluster is accepted and attached to the track seed if the distance $d < 1$ mm. Once all the clusters in the Time Frame have been tested, the procedure continues searching for another track seed in the same Frame, while the clusters forming the previous complete track are excluded from the subsequent searches. When no more tracks are found, the following Time Frame is considered, and so on.
Fig. 115 shows an example of clusters recorded in four different Time Frames, and the performance of the track recognition algorithm. In the first Time Frame (TF 0, blue dots in the



(a) (b)

Fig. 115: Example of clusters recorded in four different Time Frames (points) and the identified tracks, (a) $x - y$ view (b) axonometric view. The Outer Barrel structure is outlined in grey.

graph) a total of 5 clusters are visible, and the tracking procedure individuates a 4-cluster track discarding the remaining cluster on Layer 4. In both TF 2 and TF 3 some clusters are correctly discarded, as they are visibly not aligned with the identified track. In TF 1 multiple clusters on overlapping adjacent sensors are present. In TF 2 and TF 3 the expected clusters are not detected, respectively, in Layer 6 and Layer 4, which indicates that a sensor or a stave was not active in that particular acquisition or, less likely, that due to the detection inefficiency those hits were lost.

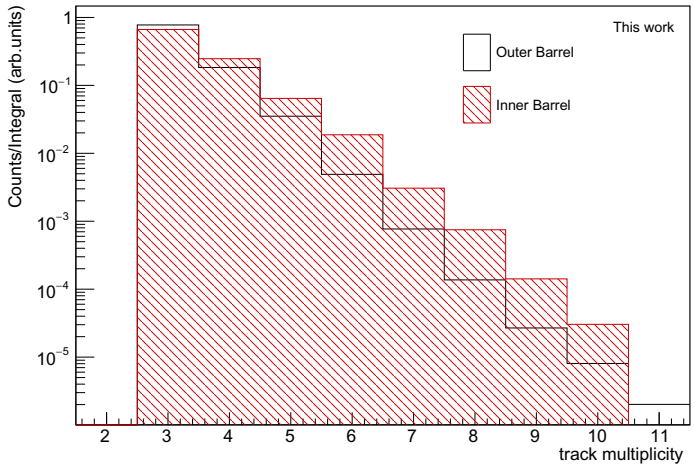
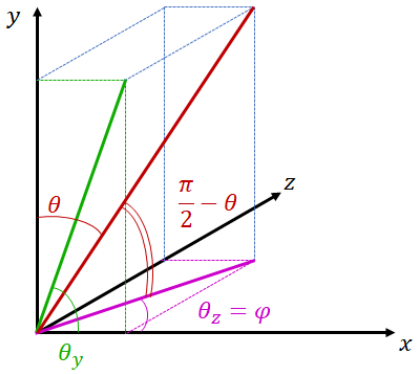


Fig. 116: Multiplicity distribution of the recognized tracks, detected by the ITS. Data from Inner Barrel (red filled histogram) and Outer Barrel (black histogram) separated.

Fig. 116 shows the distribution of the track multiplicity, i.e. the number of clusters in a track. Tracks with three clusters are the most probable, while the counts of events with higher multiplicities diminish exponentially with the number of clusters. Higher multiplicities are hindered by the fact that in this acquisition campaign the Half-Barrels are separated; with the full ITS higher multiplicities are to be expected. Moreover, the Inner Barrel presents slightly higher

multiplicities with respect to the Outer Barrel, although the latter includes more Layers; this effect is caused by the proportionally more overlapped area between adjacent sensors in the IB, and the IB much tighter curvature.



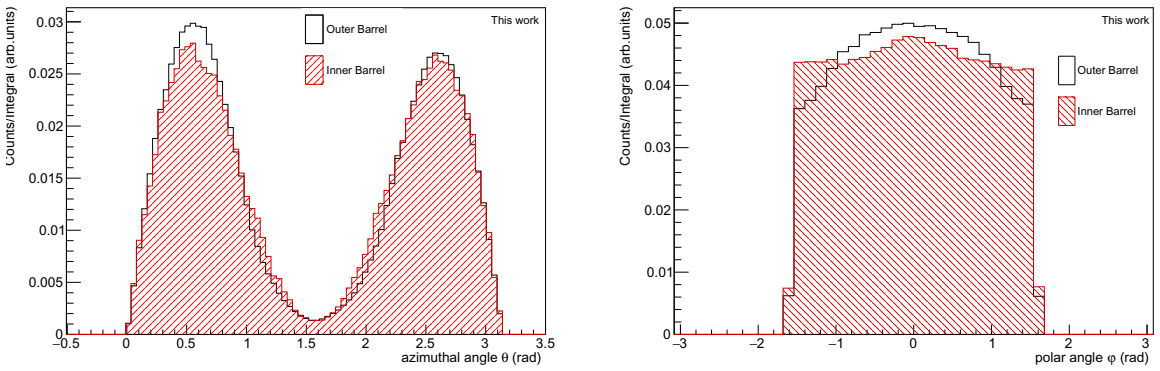
$$\begin{aligned} \theta_y &= \arctan(m_y) \\ \theta_z &= \arctan(m_z) = \varphi \\ \theta &= \tfrac{\pi}{2} - \arctan(m_y \cos(\theta_z)) \end{aligned} \tag{74}$$

Fig. 117: Scheme of the azimuthal angle θ and polar angle φ derivation from the track parameters. Relations between the track parameters in Eq. (73) and the azimuthal and polar angles.

Apart from the visual case-by-case test displayed in Fig. 115, another check of the accuracy of the track recognition, is to observe whether the angular distribution of the reconstructed tracks is compatible with a cosmic origin of the events. The relations between the azimuthal and polar angles in the Global Frame of reference and the track fit parameters are shown in Fig. 117 and Eq. (74). In Eq. (74) θ_x in the angle formed by the track projection onto the $x-y$ plane with respect to the x axis, θ_z is the angle on the horizontal plane $x-z$ and it corresponds to the polar angle ϕ; the azimuthal angle θ is the angle that the track forms with the vertical y axis. The distributions of the azimuthal angle θ and the polar angle φ are presented in Fig. 118 (a) and (b) respectively. The azimuthal angle presents two maximal points near 0 and π and a minimum in proximity of $\pi/2$, which is compatible with the mostly vertical cosmic tracks; Inner and Outer Barrel present comparable behaviours. The polar angle φ is the angle in the $x-z$ (horizontal) plane with respect to the x-axis, expressed in the range $(-\pi/2;\ \pi/2)$. The distribution of φ is basically flat, as cosmic rays have, in an acceptable approximation, no privileged polar direction. Nonetheless, in the measured distribution a slight preference for polar angles near to 0 is visible in both IB and OB, due to the higher angular acceptance of the Half-Barrels in the x direction, with respect to the z direction.

After the tracks recognition procedure, the residuals between the expected and measured cluster positions are calculated. The residuals are obtained using two independent methods, in order to cross-check the results. The first method uses a basic analytic geometry calculation to find the intersection point of the track with the sensor plane, while the second one defines a physical particle track and propagates it through the detector, employing functions developed in the O^2 software framework for particle analysis.

The analytical residual calculation procedure is outlined in Fig. 119. To calculate the expected position of a given cluster belonging to an identified track, the positions of the remaining clusters in the track are transformed into the Local Frame of reference of the considered cluster. The track parameters in the Local Frame are obtained by interpolating the clusters new transformed positions, using the fit method illustrated in Fig. 114 (b) and Eq. (73); the cluster under study is excluded from the fit, to avoid biasing the interpolation results. Then, the expected cluster position $(x_{loc}^{expected}, z_{loc}^{expected})$ is obtained as the intersection of the track with the sensor



(a) (b)

Fig. 118: (a) Distribution of the azimuthal angle θ of the recognized cosmics tracks, measured by the ITS Half-Barrels. (b) Distribution of the polar angle φ. Data from Inner Barrel (red filled histogram) and Outer Barrel (black histogram) separated.

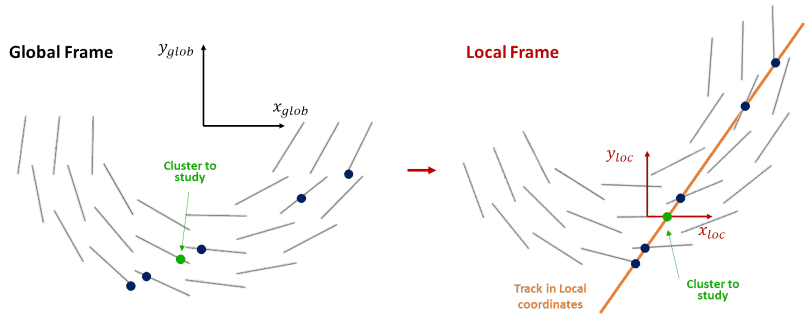


Fig. 119: Residual calculation using the analytic geometry method: the clusters belonging to a track are transformed into the Local Frame of the considered cluster (green), and the track parameters are calculated in this coordinate system. The expected cluster position is calculated as the intersection point between the track and the plane $y_{loc} = 0$.

plane, i.e. by setting the vertical coordinate $y_{loc} \equiv 0$:

$$\begin{cases} x_{loc}(t) = t \\ y_{loc}(t) = q_{y,loc} + t\ m_{y,loc} \equiv 0 \\ z_{loc}(t) = q_{z,loc} + t\ m_{z,loc} \end{cases} \Rightarrow \begin{cases} x_{loc}^{expected} = \frac{q_{y,loc}}{m_{y,loc}} \\ z_{loc}^{expected} = q_{z,loc} + q_{y,loc}\frac{m_{z,loc}}{m_{y,loc}} \end{cases} . \tag{75}$$

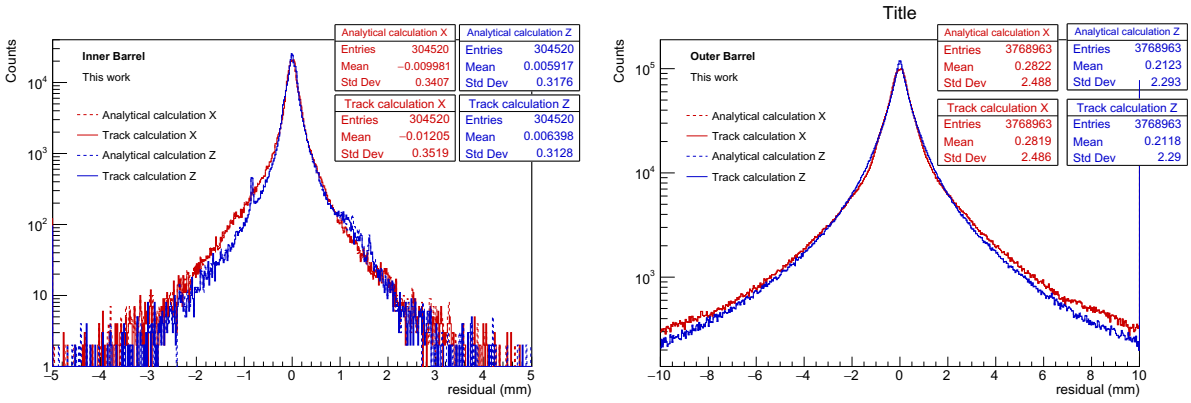
Finally the residual is calculated simply as the difference:

$$(x_{loc}, z_{loc})_{expected} - (x_{loc}, z_{loc})_{measured} \quad , \tag{76}$$

and the procedure is repeated for each cluster included in a track and for every identified track. The second residual calculation method consists in defining a physical particle track by means of an origin point and a momentum vector. In this case the clusters positions are interpolated in the Global Frame, excluding successively the cluster under study. The track is then uniquely defined by a point and a vector:

$$\mathbf{x} = \mathbf{x_0} + t\mathbf{p} = \begin{pmatrix} 0 \\ q_y \\ q_z \end{pmatrix} + t \begin{pmatrix} 1 \\ m_y \\ m_z \end{pmatrix} \quad , \tag{77}$$

which contains the parameters of the track interpolation. The origin point of the track and its momentum are not meaningful measurements of the real physical particle, as the energy of a cosmic ray is not measurable without a magnetic field or auxiliary detectors. However, they are sufficient to define its position and direction in the considered case, as only linear tracks are considered. The track position is then propagated from its origin point onto the surface of the sensor where the cluster under study has been detected. The propagation performed with the O^2 software framework official methods takes into account the energy loss of the particle traversing every material volume along its path, being it either associated with a sensing or structural element, and bends the trajectory according to the magnetic field. During the on-surface commissioning no magnetic field was present, therefore the propagation of the particle track results in a linear transportation. The residual is finally obtained transforming both the measured and expected cluster positions into the Local Frame of the sensor and calculating their distance, similarly to Eq. (76). The second method for calculating the residuals has the advantage of being adaptable for future applications, when the magnetic field will be engaged, while the analytical method is suitable exclusively for cosmic tracks propagating strictly in a straight line.

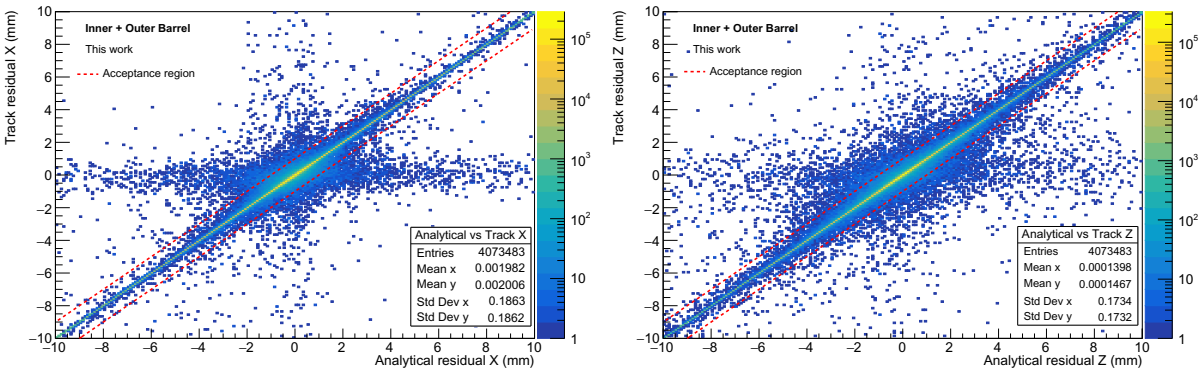


(a) (b)

Fig. 120: Histograms containing residuals relative to (a) Inner Barrel and (b) Outer Barrel. Residuals in the x direction are plotted in red, z residuals in blue. The histograms relative to residuals calculated with the O^2 framework defining physical tracks are plotted with solid lines, the histograms relative to the analytical residuals calculations are indicated with dashed lines.

In Fig. 120 the residuals distributions are presented, separately for the Inner and Outer Barrel and for both the x and z directions. The residuals in the y directions, i.e. affecting the sensors positioning radially, are not considered in the present stage. This is because radial displacements of the sensors tend to have a smaller effect on the track parameters accuracy, with respect to misalignments in the x or z directions, especially for collision-originated particles, which usually form a small angle with the sensor plane normal [181]. Moreover, an incorrect radial positioning of the sensors does not affect the reconstructed tracks fit resolution, but instead biases the measured track curvature, and therefore the particle momentum estimate [181]. A displacement in the y direction, as well as other *weak modes* such as z-dependent deformations and rotations, can be more effectively estimated using physical observables as a reference [181]. Finally, for either residual calculation method adopted in this work, the y coordinate represents the independent parameter which is set in order to retrieve the plane-track intersection, so a different approach would be necessary to study residuals in this direction.
The residuals distributions in Fig. 120 present central values within less than 2 microns from 0, and the residuals in both x and z directions show overall similar behaviours. The distributions relative to the Inner Barrel appear narrower with respect to the Outer Barrel, which is com-

patible with the larger distances between the layers of the latter. The z distribution relative to the Inner Barrel, Fig. 120 (a), shows a defined peak around -1 mm in addition to other substructures between 1 and 2 mm. These events could be an evidence of a small number of displaced staves. The differences between the analytical and track-based residual calculations (solid vs dashed lines) are noticeable only in a few bins of the histograms in both Inner and Outer Barrel cases.



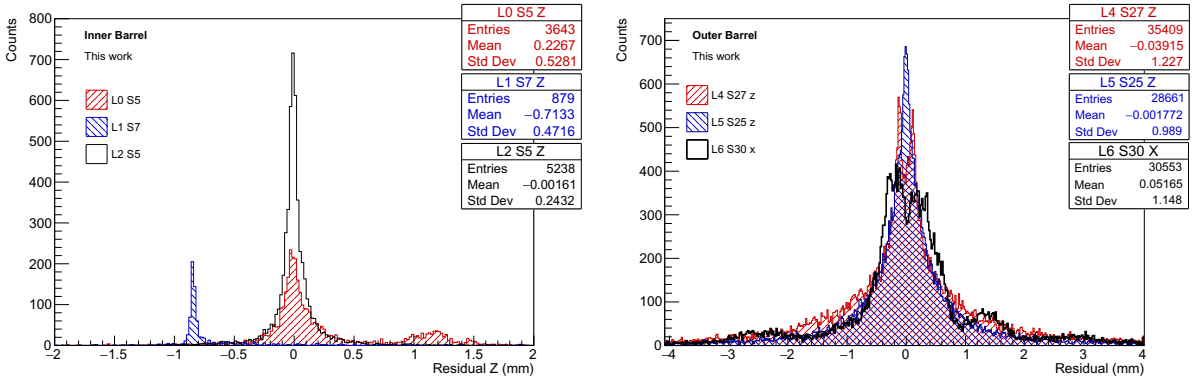
(a) (b)

Fig. 121: Scatter plots showing the correlation between the analytical (horizontal axis) and O^2 framework track-based (vertical axis) residual calculations, for the (a) x and (b) z directions. Inner and Outer Barrels are here considered collectively. The dashed lines delimit the area in which the residuals are accepted.

Fig. 121 shows scatter plots correlating the residual values obtained with the two calculation methods: by the analytic fit in the horizontal axis and by the O^2 physical track propagation in the vertical axis, separately for the x and z directions, in (a) and (b) respectively. The vast majority of events lays in close proximity to the bisector, while for a relatively small fraction of events the two methods yield different results. Divergent results can be generated from tracks intercepting the sensors in close proximity to the edge of the sensing area, for which even a small variation of the track parameters can lead to the track missing the sensor surface, and therefore an incorrect interception point calculation. Moreover, errors in the residual calculation can also be the result of an incorrect track interpolation procedure; additionally, the tracking methods in the O^2 framework are optimized for collision-generated tracks, with particles originated in the proximity of the centre of the beam pipe, and then propagating outwards, while in this case they have been adapted to describe cosmic particles with an external origin, and propagating throughout the entire detector, so in a small number of instances the track propagation could fail. Finally, as the data were collected during the commissioning period of the Inner Tracking System, the read-out procedure from the sensors was not completely immune to glitches or incorrect data formatting. In order to avoid considering inaccurate events, the residuals falling outside the belt of ± 1 mm around the bisector are discarded, as indicated in Fig. 121. From this point on, the results are calculated taking into consideration the residuals calculated with the O^2 method.

In this preliminary alignment status survey, the smallest basic movable element is considered to be the stave, as the positioning of each sensor relatively to the stave has been evaluated in the assembly phase. Once the positions at the stave-level would have been assessed and corrected, the definitive alignment survey is likely to be performed at the sensor-level.

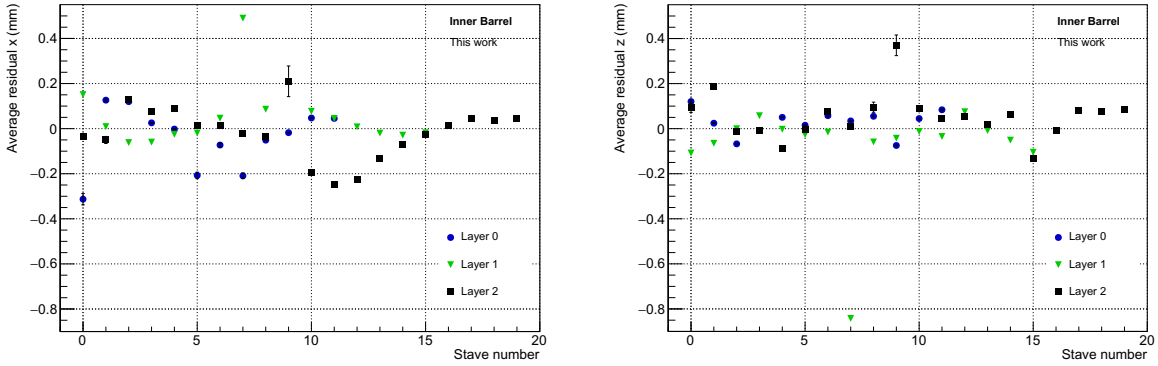
For each individual stave the distributions of the residuals are obtained, in order to observe possible biases. In Fig. 122 some examples of residuals distributions for selected staves are presented. The distributions qualitatively show three identifiable behaviours. In particular, for



(a) (b)

Fig. 122: Examples of residuals distributions for selected staves: (a) for the Inner Barrel the histograms relative to the z-residual in stave 5 of Layer 0 (L0 S5), stave 7 of Layer 1 (L1 S7) and stave 5 of Layer 2 (L2 S5) are shown; (b) for the Outer Barrel L4 S27 z, L5 S25 z and L6 S30 x have been selected.

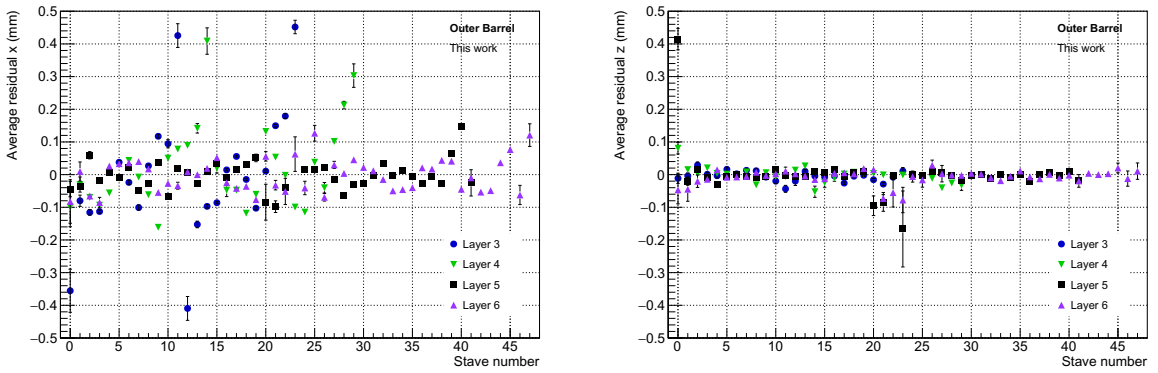
a fraction of the staves the distribution appears centered near 0, similarly to the histogram relative to the stave 5 of Layer 2 (or L2 S5) shown in Fig. 122 (a), which sports a clear peak around $-1\,\mu$m. Distributions relative to other staves show similarly defined peaks, however not compatible with 0, as the stave L1 S7 included in the same plot, which is centered around $-700\,\mu$m. This result is a candidate signature of a displaced stave. Finally, the stave L0 S5 has been selected to illustrate a different behaviour: in addition to the central peak, a second population of events concentrating far from 0 is recognizable. This suggests a correct positioning of the selected stave, but the presence of a displacement on a surrounding stave, with the latter projecting its own incorrect positioning onto the former. It can be noticed that some of the spikes visible in the general residuals histogram relative to the z direction in Fig. 120 (a) are compatible with the isolated histograms per stave in Fig. 122 (a).
In Fig. 122 (b) some examples of the single-stave residuals distributions for the Outer Barrel are illustrated. In the Outer Barrel case, the distributions present similar behaviours with respect to the Inner Barrel: for instance, the presented L5 S25 is characterized by a well-defined single peak with an average value of about $0.2\,\mu$m, while some histograms, as the one relative to the stave L4 S27, show a bi-modal symmetrical distribution. The second case may originate from a minor relative misplacement of the two half-staves forming a single Outer Barrel stave. In the Outer Barrel case, the residuals relative to the x directions present a generally wider distribution and, where present, a number of substructures and spikes, as visible in the L6 S30 in Fig. 122 (b). The more complex distributions from the Outer Barrel staves, especially in the x direction, could be caused by its larger dimensions with respect to the Inner Barrel: the higher weight of the staves could provoke a sagging effect of increasing magnitude the further the position is from the Barrel end-wheels. This effect would have an impact on the x but not on the z residuals.
The residuals histogram for each stave is interpolated with a Gaussian, to estimate the central value of the distribution. Fig. 123 show the central value parameter from the Gaussian fit as a function of the stave number in the Inner Barrel, separately for the x, z direction. The average x residual for the Inner Barrel (Fig. 123 (a)) for the majority of the staves is distributed in the $\pm 200\,\mu$m range, with a small number of staves presenting higher residuals, such as L0 S0, which show a mean value of $-313\,\mu$m, L1 S7 with $490\,\mu$m and L2 S9 with $210\,\mu$m. It can be noted that the staves included in the second Half-Barrel of L2 (from S10 to S19) share a common trend in the average residuals, which could be due to an assembly effect. The average z residuals for the Inner Barrel (Fig. 123 (b)) are overall smaller with respect to the x, and for the most



(a) (b)

Fig. 123: Central value parameter obtained from a Gaussian fit of the residuals distribution for each stave in the Inner Barrel; residuals relative to (a) the x direction and (b) the z direction. The exact values are tabulated in Tab. 9 in Appendix C. The error bars correspond to the uncertainty on the central value parameter obtained from the fit.

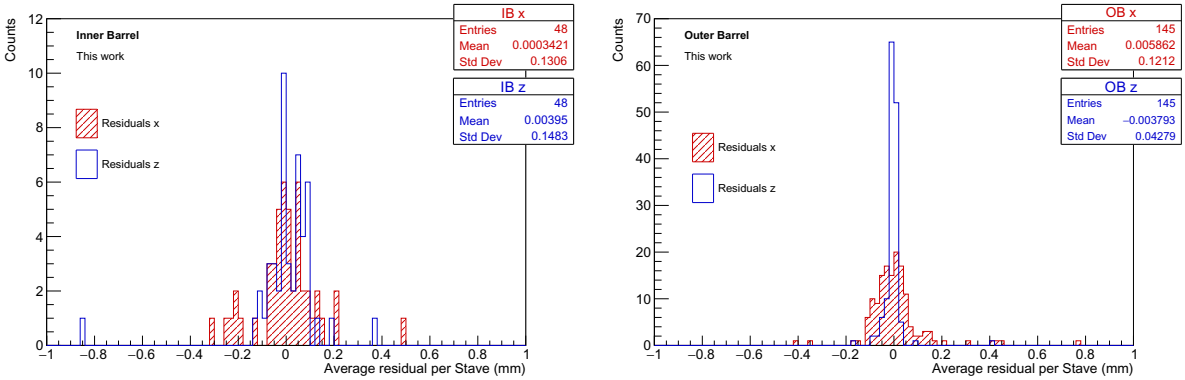
part limited in the $\pm 150\,\mu m$ range. Exceptions are represented by the L1 S7 ($-313\,\mu m$) and L2 S9 ($-313\,\mu m$), which have shown particularly high average residuals also in the perpendicular direction (Fig. 123 (a)).



(a) (b)

Fig. 124: Central value parameter obtained from a Gaussian fit of the residuals distribution for each stave in the Outer Barrel; residuals relative to (a) the x direction and (b) the z direction. The exact values are tabulated in Tab. 10 in Appendix C. The error bars correspond to the uncertainty on the central value parameter obtained from the fit.

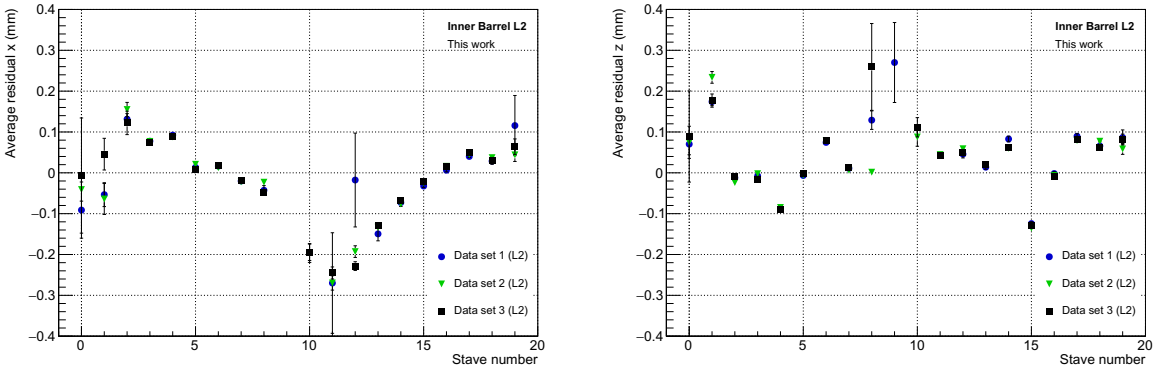
Fig. 124 illustrate the average residual per stave in the Outer Barrel, obtained from a Gaussian fit of the residuals distributions. Average residuals appear more dispersed in the x direction (Fig. 124 (a)), while in the z direction (Fig. 124 (b)) the values are for the most part smaller than $50\,\mu m$ in absolute value. It can be observed that generally, for both IB and OB (Fig. 123 and Fig. 124 respectively) error bars, corresponding to the uncertainty on the central value parameter obtained from the fit, are larger for the first, last and middle staves in a Layer. These, in fact, correspond to the staves whose sensors surfaces are more vertical, hence collecting less statistics from cosmic particles.
Fig. 125 includes the histograms containing the central values of the residuals distribution,



(a) (b)

Fig. 125: Histograms including the average residuals per staves in x and z directions, respectively for (a) Inner and (b) Outer Barrel. Values represented in Fig. 123 and Fig. 124; tabulated in Tab. 9 and 10 in Appendix C.

represented in Fig. 123 and 124.



(a) (b)

Fig. 126: Central value parameter obtained from a Gaussian fit of the residuals distribution for each stave in the Layer 2 of Inner Barrel; residuals relative to (a) the x direction and (b) the z direction. Three distinct datasets are presented, in order to observe the consistency of the residuals behaviour.

In order to verify the consistency of the residuals calculation, and to confirm that the misalignment effects identifiable in certain staves are not statistical fluctuations, the procedure has been performed on three disjoint datasets. The results are shown in Fig. 126 for the Layer 2 of the Inner Barrel. The thee datasets show compatible trends with each other, with a few exceptions, exclusively in staves affected by low statistics. This is an indication that the residual calculation highlights, in fact, actual and reproducible features of the ITS geometry.

8.5 Re-alignment algorithm

In order to identify the displaced staves and to estimate the necessary corrections to the staves positioning, an iterative, automated algorithm is here proposed. The results of this method could provide a starting point for the official ALICE re-alignment tool Millepede [48].
The goal of the proposed re-alignment procedure is to build a table of corrections for the position

of each stave, in both the x and z directions, with an iterative algorithm. The procedure attempts to identify the displaced staves, operating on a single direction (x or z) at a time, minimizing a global χ^2 parameter. The proposed re-alignment procedure develops in a series of steps:

- The first step is to assess the initial alignment condition, performing the complete procedure of track-finding and residual calculations. A quality parameter, comparable to a global normalized χ^2, is defined:

$$\chi^2 \equiv \sum_{i=0}^{N} \frac{r_i^2}{N} \quad , \tag{78}$$

  where r_i are the calculated residuals and N the total number of residuals. This first step can either follow a preliminary re-alignment procedure, or start from a totally blank correction table.

- The staves are ordered with respect to their average residual, from the largest to the smallest; the stave with the largest average residual is then taken as the displaced candidate and a tentative correction is applied. The magnitude of the correction can be a fraction of the average residual (e.g. 1/2 of the value) or a fixed quantity.

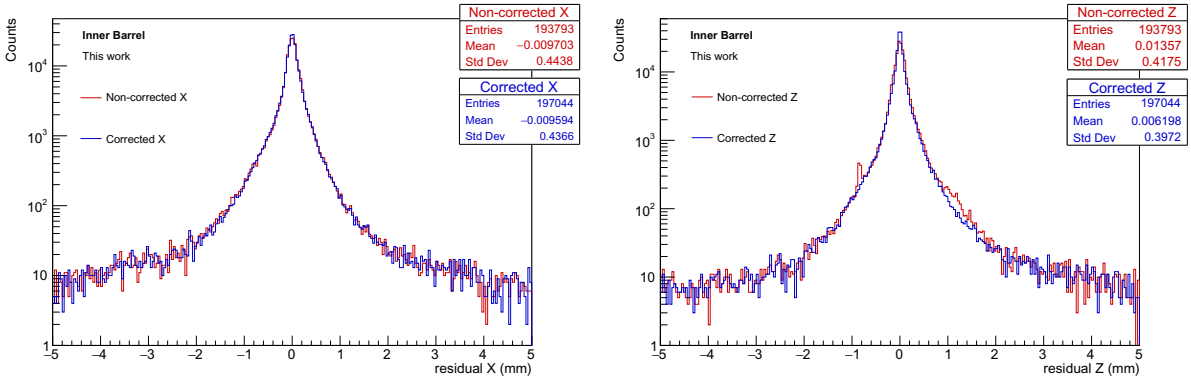
- The track-finding and residuals calculation procedure is repeated, applying the tentative correction: the coordinates of the clusters belonging to the candidate stave are shifted according to the magnitude of the correction. The new χ^2 parameter is calculated, with the updated set of r_i and N.

- If the new χ^2 is smaller than the one calculated in the previous iteration, the correction is accepted, the correction table is updated to be used in all the next iterations, and the subsequent iteration is tested against the new χ^2.

- If the new χ^2 is instead larger than the previous one, the correction is rejected. The stave with the second highest average residual is then considered, and the relative correction is tried. Subsequent staves are in order tested, until a correction is accepted.

- The algorithm stops when all the staves have been tested and no correction was accepted. For a finer re-alignment the procedure can be performed with increasingly smaller correction steps.

The algorithm was tested with the Inner Barrel data, given the more manageable number of degrees of freedom. The correction step per iteration was first set as 1/2 of the staves average residual, and then as a fixed step of 5 µm, for a finer tuning; the algorithm was applied for the z and x direction separately. The algorithm procedure was not carried out to the full completion of the χ^2 minimization, mainly for computational time reasons[4]; if deemed effective, the code can be further optimized for better run times and performed on more powerful machines.
Fig 127 shows the complete distributions of the residuals, for the x and z directions, in (a) and (b) respectively, calculated by applying the corrections table obtained from the iterative algorithm, and compared with the initial residuals without the corrections.
In the x direction, the corrected residuals have an average value slightly closer to zero, a smaller standard deviation, and the central peak of the distribution appears marginally higher; however, the re-alignment procedure could be iteratively applied to further improve the results. In the z direction the effect of corrections is more noticeable: the distribution mean is more than halved, and the standard deviation is appreciably lower; additionally, the central peak of the distribution appears higher. Moreover, the smaller peaks originally present around -1 mm and between 1 and 2 mm, due to displaced staves (as shown in 122 (a)), have disappeared. Finally,



(a) (b)

Fig. 127: Histograms with all the residuals relative to (a) x and (b) z direction. The red histograms include the original residuals calculated without the corrections; blue histograms include the residuals calculated applying the corrections obtained from the iterative algorithm.

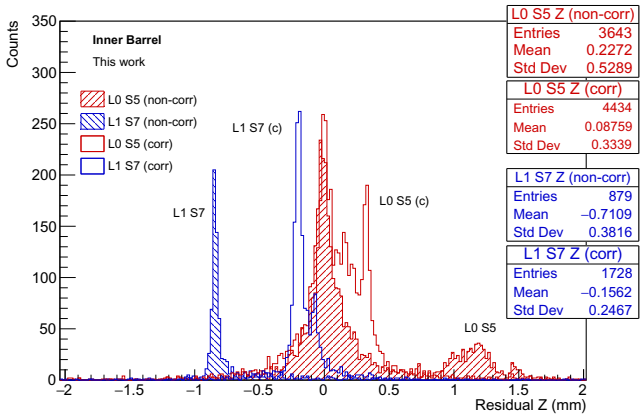
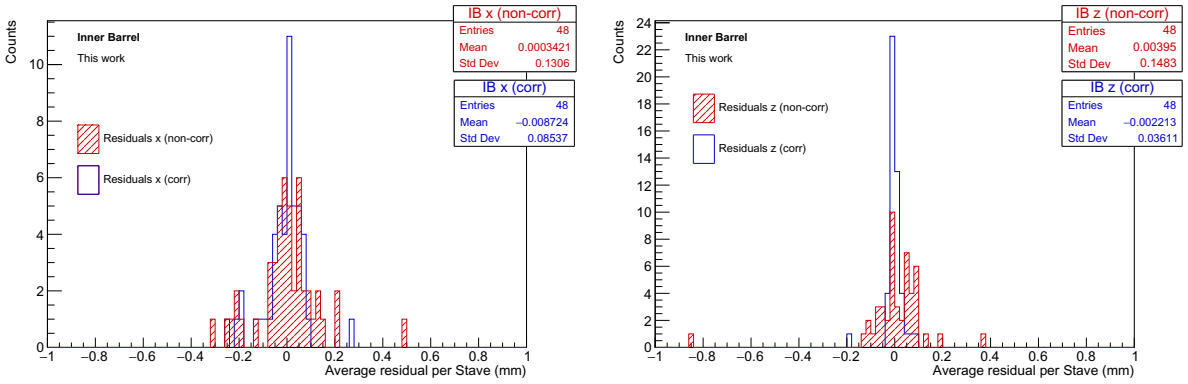


Fig. 128: Examples of residuals distributions for selected staves: stave 5 of Layer 0 (L0 S5, blue) and stave 7 of Layer 1 (L1 S7, red); the histograms before (filled) and after (blank) the corrections are compared.

after the corrections, the total number of clusters associated to tracks is 1.6% higher.
Fig. 128 shows the residuals distributions for the staves L0 S5 and L1S7, previously identified as staves with evident displacement effects (see Fig. 122), before and after the corrections, in order to observe the effects of the re-alignment procedure. The stave L1 S7, before the corrections, presented a distribution centered at 711 µm, while after the re-alignment the peak has shifted significantly towards zero, and is now at −156 µm; additionally, the number of clusters belonging to a track has increased by almost 22%. The stave L0 S5 initially featured a bi-modal distribution, with a distinct peak around 1.2 mm; after the correction, the central part of the distribution is still present, while the additional peak has shifted heavily towards zero, and now can be found around 350 µm. Also in this case, the number of recognized clusters have increased significantly. The outcomes on this two particular staves show how the automatic re-alignment procedure has, in fact, proceeded in the appropriate direction, reducing the magnitude of the residuals, while also visibly correcting the effects of the displacements.

[4]The computation has been performed with the author's personal laptop.



(a) (b)

Fig. 129: Histograms including the average residuals per staves in (a) x and (b) z directions. Results obtained without the corrections (red filled histograms) and with the corrections (blue blank histograms) are compared.

Fig. 129 show the distributions of the average residual per stave, before (red histograms) and after corrections (blue histograms). In both the x and z directions, after the re-alignment procedure, the distributions are narrower and more concentrated around zero; in both cases the dispersion have been reduced, and the more extreme points have been brought closer to the centre. The proposed algorithm has successfully corrected the more problematic aspects caused by the assembly displacement of the staves in the ITS; moreover, it can be further applied to refine the the corrections.

8.6 Outlook

The preliminary alignment status of the ITS has been surveyed, showing a more than acceptable condition: in most cases the average residual per stave is within 150 µm for the Inner Barrel and 100 µm for the Outer Barrel. Staves presenting higher than average residuals have been identified; these cases represent only a minor fraction of the total. The alignment method, based on the calculation of residuals, appears to be effective for this task, robust, and it has the potential to be applied also for a survey of the ITS after its installation in the cavern. The method is intrinsically extendable also for non-cosmic particles, such as collision-generated secondaries, to periodically assess the alignment status during the entire experiment life-time, when physical access to the apparatus would be problematic.
At the same time, an iterative re-alignment procedure has been proposed and tested on the Inner Barrel staves. The method appears an effective tool to speed-up the alignment procedure, and it can be applied to the entire Tracking System. The algorithm has, in fact, efficiently identified and corrected the position of displaced elements, visibly improving the quality of the tracking. The same concept can also be extended to a large set of applications. In particular, the alignment of the iMPACT tracking section, which is equally composed by IB staves, is planned to be evaluated employing an adapted version of the presented algorithm.
Overall, the opportunity to work on the ALICE ITS has been an invaluable chance to acquire key competences on the particle tracking field, and to grow familiar with the very same sensor that, currently, is one of the main component of the iMPACT medical scanner.

9 ARCADIA

ARCADIA (*Advanced Readout CMOS Architectures with Depleted Integrated sensor Arrays*) is an INFN national project, started in January 2019, as an evolution of the SEED sensor [183], involving various INFN Sections: Bologna, Milano, Padova, Pavia, Perugia, Torino and Trento (TIFPA). The project goal is to design and develop a novel Full Depleted Monolithic Active Pixel Sensor (FD-MAPS), with active sensor thickness in the range of 50 µm to 300 µm, characterized by a scalable readout architecture with ultra-low power capability (of the order of 10 mW cm^{-2}). The ARCADIA sensor is implemented in the 110 nm CMOS technology from LFoundry [184], with backside lithography; the monolithic process allows for in-pixel implementation of complex digital functions [185]. The new sensor is designed for a wide range of application, in the medical, space, and particle physics fields.

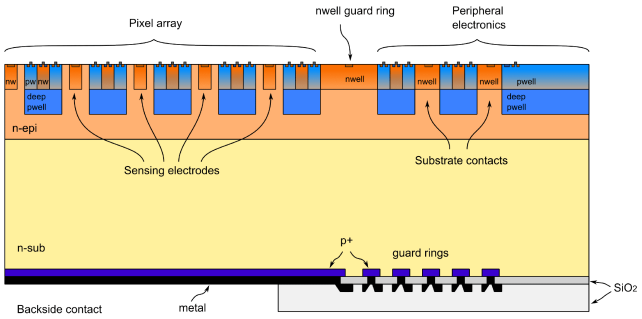


Fig. 130: Cross section of the ARCADIA monolithic sensor [183].

Fig. 130 shows a schematic cross section of the sensor design. The sensor features a high resistivity (2 kΩ cm^{-1}) n-type floating zone substrate with deep p-well implants, to isolate the n-wells for the PMOS transistors from the epitaxial layer. An n-type epitaxial layer, with a lower resistivity with respect to the substrate, separates the deep p-wells from the substrate, and allows for the implementation of 1.2 V transistor circuitry on the sensor front side [183]. The substrate high-resistivity allows to reach full depletion with relatively high voltages 100 V while keeping a low bias current.

The sensor exploits wafer-thinning and back-processing methods to implant a boron-doped p-region and an electrode node on the backside of the substrate, allowing to back-bias the the sensor with an uniform electric field, so creating an homogeneous depletion region through the whole substrate. The sensor full depletion, achievable at voltages of the order of 100 V, leads to a faster charge collection, as the carriers motion is dominated by drift rather than diffusion, as well as a higher radiation tolerance, and a higher efficiency [185]. The depletion from the bottom makes the sensor suitable, via back-illumination, for the detection of extremely low range particles, which would otherwise be absorbed in the front-side epitaxial layer, including near-UV photons. The sensor full depletion also has a focusing effect on the charge cloud generated by passing and/or impinging particles, thus limiting the pixels clusters multiplicity, a feature especially advantageous when detecting non-MIP particles (such as protons in the medical energy range). The dopant concentration of the n-type epitaxial layer has been tuned to hinder punch-through current from the back to the front-side at full depletion conditions. To prevent parasitic currents, a system of guard rings has been added around the back-side electrode.

A 576 pixels sensor (24×24), named MATISSE, was designed to assess the performances of the described technology [186]; the pixel matrix is divided into 4 vertical sectors with 6 columns each, featuring different sensing diode geometry. Both analog front-end and digital electronics are included inside the 50 µm-pitch pixels, with a 20 µm×20 µm area at the centre reserved for

the charge-collecting diode. The four sectors are managed by an End-of-Column (EoC) block, controlling pixel configuration and performing the read-out; the entire pixel array can be read in less than 30 µs. MATISSE operates with an analog read-out, performing sampling of the baseline and the signal, on a pixel-level.

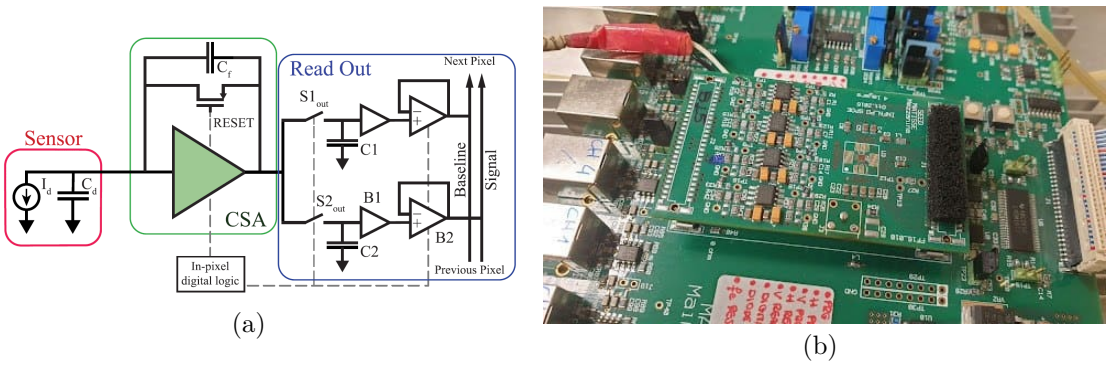


Fig. 131: (a) Schematic of the MATISSE in pixel read-out front-end [183]. (b) Picture of the setup for MATISSE characterization in Padova.

Fig. 131 (a) shows the in-pixel front-end, including a Charge-Sensitive Amplifier (CSA), a sample-and-hold circuit performing Correlated Double Sampling (CDS) of both signal and baseline. Metal–Insulator–Metal (MIM) capacitors are used to locally store the signal; two analog buffers allows for the transmission of the analog signals along the column [183]. The power consumption of the array has been estimated between 3.8 mW and 6.4 mW [185]. Tests with a ^{55}Fe source, producing 5.9 keV and 6.5 keV photons. with range in Si of the order of 10 µm [187], were performed in Padova. With that source and a voltage bias of 200 V, the observed cluster size was limited to 2×2 pixels, with clusters with more than 4 pixel being observed in less than 1% of the cases. No significant differences were observed between front and back-illumination, confirming the full depletion of the sensor and the fact that charge motion is dominated by drift rather than diffusion. An average 40 electrons noise, at room temperature, was also estimated [185].

9.1 Innovative read-out architecture

Particle positions on a pixel detectors are obtained from the charge cluster signals released by the passing particle. In most practical applications, the matrix occupancy is usually less than a few percent; in traditional pixel detectors the hit pixel information is retrievable only from a complete read-out of the whole matrix, ultimately limiting the data acquisition rate and weighing on the power consumption [188]. As a solutions to these drawbacks, hybrid-pixel sensors developed within the High Energy Particles community have been developed, where the sensitive pixels layer is separated from the advanced per-pixel readout circuits, which is optimized to perform data sparsification within the sensor itself [189], drastically reducing the data throughput; intrinsic limits to the hybrid design are, however, the higher manufacturing complexity, cost, material budget and power consumption. More recently, such sparsifying architectures have been successfully embedded within Monolithic Active Pixel sensors (MAPS) [120]. Yet, the implementation of complex architectures within a monolithic sensors imposes several limits in therms of complexity and, even more critically, being the power consumption minimization one of the major advantages of monolithic sensors over the hybrid hones, a complex readout architecture would waive this key advantage of the monolithic technology.
Trying a different approach to the readout architecture problem, the OrthoPix project proposed an innovative read-out architecture scheme, relying on lossy data compression, a concept which is planned to be eventually implemented in the ARCADIA sensor. The OrthoPix architecture

is based on passively-addressed pixels, instead of the active pixel-driven read-out, adopted by traditional pixel detectors [188]. Additionally, after the compression logic, the data output produced by a particle hit is constant, thus simplifying data transmission. The lossy compression is highly efficient at lower particle rates, while it starts to cause data loss for higher matrix occupancies.
The OrthoPix architecture in based on an evolution of a projective read-out; as illustrated in Fig. 132, a matrix of N^2 pixels can be read-out relying on two projections the horizontal and vertical coordinate $x - y$. The position of a hit can be obtained by observing only $2N$ elements, with a data compression of $N/2$ with respect to a pixel-per-pixel read-out. The natural limitation of a 2-dimensional projection is that multiple simultaneous hits generate spurious combinations that can not be discriminated from the real coordinate positions, if no other information is available.

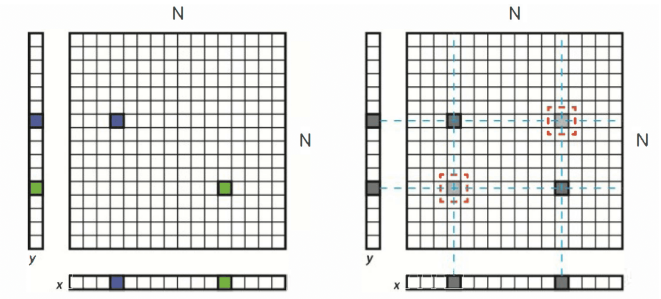


Fig. 132: Schematic representation of a projection-based read-out on two coordinates; multiple simultaneous hits produce ambiguous reconstructions of the original positions [188].

This limit puts a serious constraint on the sustainable particle rate R of the detector. Considering a Poisson statistic, if the detector allows for only a single hit in a time frame to be correctly reconstructed, to reach a 99% efficiency, the particle rate has to be $\approx R/10$ [188].
As shown in Fig. 133, if the same N^2 pixel matrix is equipped with a additional projection, in this case a diagonal one, two or more simultaneous hits can be correctly distinguished. This simple solution can be expanded to consider an arbitrary number n of maps (non necessarily geometrical projections), which link any pixel to an output element. The principle is illustrated in Fig. 134, where the space of the pixels P is mapped by a set of maps π_i into sub-spaces $O_i = \pi_i(P)$.
To maximize the output occupancy, each one of these maps π_i needs to be defined as that every output element is linked to the same number of input elements. Using the simple projection analogy, an equal number of pixels are to be projected into each x or y coordinate, as, in fact, happens is square matrices; on the contrary, the basic diagonal projection does not meet this

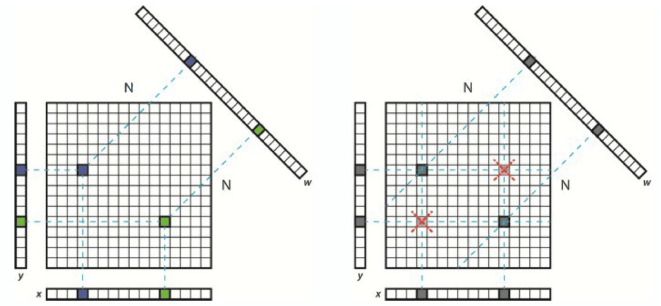


Fig. 133: Schematic representation of a projection-based read-out with an additional (diagonal) projection, which allow to correctly reconstruct the positions of more than one event [188].

requirement. Additionally, to ensure the optimization of the decomposition procedure, given any couple of maps, their intersection must be represented by a single element.

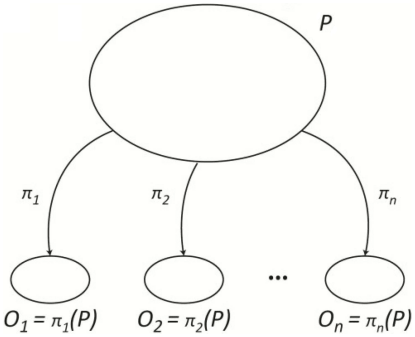


Fig. 134: Representation of the set of maps π_i linking elements of the pixel array P to output elements; O_i represents the image of the map π_i. [188].

It can be demonstrated that the reconstruction efficiency, i.e. the fraction of hits correctly recognized over the total, in a N^2 pixel matrix can be described by the expression:

$$f = 1 - \left[1 - \left(1 - \frac{1}{H}\right)^{\frac{(H^2-H)^2}{N^2}}\right]^n \quad , \tag{79}$$

where H is the number of hit pixels; the expression is valid assuming uniformly distributed hits over the matrix, and small hits H compared to N, thus the probability of hits sharing coordinates is negligible [188].
For a practical application in a pixel sensor, the number of possible projection is limited by the space available in the pixel, as the implementation of the projection logic, to achieve the highest efficiency, should be realized physically inside the matrix. In a realistic implementation, a total of $n = 4$ projections is foreseeable; Fig. 135 shows the results of a Monte Carlo simulating the hit reconstruction efficiency for different matrix sizes, considering four projections, compared with the estimation in Eq. (79). The plot shows how extremely high reconstructions efficiency ($\approx 99\%$) can be achieved, given a low enough hit rate.

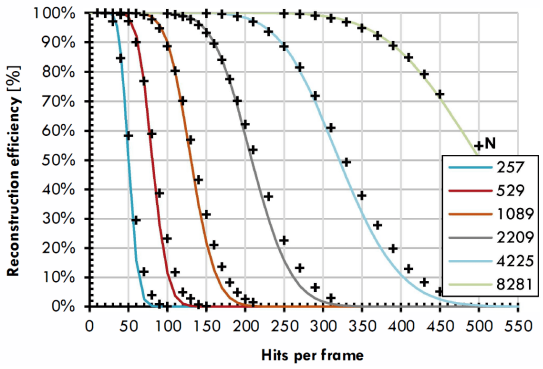


Fig. 135: Reconstruction efficiency as a function of simultaneous hits in a single time frame for different N^2 matrix sizes, with a $n = 4$ projections architecture; points represent Monte Carlo calculations, while lines show estimations from Eq. (79) [188].

The physical implementation of the architecture on a sensor exploits a synchronous read-out with a pixel static addressing system. The projections are physically embodied by metal lines

connecting different pixels patterns via logical OR gates: whenever any of the connected pixels is hit, the entire projection is set to a high logical state. This implementation relies on a trade-off between the limited amount of total data transferred to the periphery, hence a saving in terms of power, and the area cost of having more physical lines inside the sensor. At the End-of-Column a State Machine is responsible of de-constructing the projective coordinates, to retrieve the actual position of the hit pixel. Given the lossy nature of the data compression, the reconstruction of the original position could return a wrong address; however, as shown in Fig. 135, for low enough hit rates, the reconstruction inefficiency is negligible.

10 Conclusions

Hadron-therapy is a consolidated technique for tumor treatment, which has experienced a steady growth in both number of facilities and treated patients during the last decades. This technique has clear advantages over the traditional, and far more widespread, X-ray radiotherapy, due to the particular dose profile of protons and ions; in fact, these particles reach the maximum energy deposition right at the end of their path, allowing for a more precise targeting of the tumor, a feature especially advantageous in the treatment of deep-seated cancers in close proximity to vital organs; X-rays, instead, produce an energy deposition that peaks a few millimeters inside the body and than decays exponentially with depth, making it difficult to target a defined volume without affecting also surrounding healthy tissues.
To fully exploit the precision hadrons offer, accurate 3-dimensional maps of Relative Stopping Power (RSP), along the entire path of the particles leading up to the tumor, are necessary. Nowadays, the only clinical way to obtain such maps is via X-ray Computed Tomography; this technique, however, suffers the intrinsic drawback of estimating quantities for hadrons, deriving them from measurements with photons, which undergo completely different physical interactions. Proton Computed Tomography (pCT) is a developing technique that is expected to produce 3D RSP maps with much higher accuracy, as it uses for the imaging process the very same particles employed for the treatment. Such high resolution RSP maps will make it possible to fully exploit the intrinsic precision of hadron-therapy, coming close to the physical limit given by range straggling. Despite the numerous efforts and working prototypes all around the world, the pCT technique is not yet ready for clinical application, mainly due to the low acquisition rates of the current experimental systems, which lead to too long imaging times that are humanly unsustainable.
The iMPACT project, hosted by the University and INFN Section of Padova, and funded by the European Research Council, aims at developing a fast and accurate pCT scanner, with the goal of producing a concrete technological advancement towards the realization of a clinically viable pCT technique. The iMPACT scanner relies on state-of-the-art sensors, currently in use for particle physics, and it includes a tracking and a calorimeter sections: the tracker currently uses the ALPIDE MAPS sensor, developed by ALICE collaboration at CERN, while a highly segmented hybrid energy-range telescope based on plastic scintillators and SiPMs readout provides the energy information. The calorimeter has an estimated accuracy that is only 6% higher than the physical limit, with a target acquisition rate of 100 MHz. The sensible area of the current prototype is 8×8 cm^2, easily expandable by four times due to its modular design.
To perform preliminary studies and components characterization, a versatile desktop setup optimized for cosmic rays data acquisition has been specifically developed. Furthermore, a small portion of the calorimeter and the tracking sensor have been tested with 70 to 230 MeV protons at the TIFPA beamline facility. Test beam results confirmed the calorimeter prototype met the design goals in term of energy resolution and acquisition rate, and that large area MAPS could indeed meet the requirement for medical imaging applications. In more recent development, a highly-parallel digital read-out architecture of the entire calorimeter has been developed and implemented, using custom-design electronics boards and COTS FPGAs. In its present version, the read-out is able to cope with a 5 MHz acquisition rate, optimized for cosmic rays acquisition, with the possibility to easily extend to at least 20 MHz per single scintillator with minimal firmware modifications.
At the time of writing, the final assembly of the iMPACT scanner is being completed; soon to follow, an intense campaign of commissioning tests, initially with cosmic rays, and eventually with proton beams will be performed. Preliminary tests with cosmic rays did already show the capability of the calorimeter to perform coarse particle tracking, even before a complete channel-by-channel calibration of the digital thresholds. The cosmic rays runs will generate the calibration data necessary to tune the entire setup: the data will be re-scaled to proton-suited values using the presented simulation, which employs a combination of Monte Carlo generated

events with the experimentally-measured single-photon response. In fact, the simulation has been proven to successfully reproduce both the cosmic muons and the proton data.
In-depth studies were performed on the implementation of a Most Likely Proton path formalism in a setup modeled on iMPACT. The proposed method proved able to approximate the protons trajectory with an high enough accuracy (300 µm or better) to generate suitable medical images; it is therefore planned to use it as the standard reconstruction algorithm for the analysis of the iMPACT scanner datasets. The ALPIDE tracking sensor spatial resolution represents a negligible contribution to the total reconstruction uncertainty, as well as the very low material budget. A sensible increase in proton path reconstruction precision is achievable by adding a-priori knowledge of the phantom shape: to exploit this opportunity, an object boundaries identification algorithm, which uses the very same calorimeter data, has therefore been devised and successfully tested on differently-shaped objects. The method is used as the starting point of a multiple-stage reconstruction technique.
Members of the iMPACT project actively participated to ALICE Collaboration, in both the development and commissioning of the ALPIDE sensor and the Inner Tracking System (ITS); in particular, the preliminary alignment status of the ITS has been assessed. The studies did show an acceptable condition, with average residuals per stave within 150 µm in most cases. A limited number of particularly displaced staves were identified: to recover these mis-alignments, an iterative algorithm has been developed, showing promising results in correcting the more problematic cases. Such automatic correction procedure finds applications in a wide range of cases, such as the alignment of the sensor elements of the iMPACT scanner, for which ad adapted version of the algorithm is planned to be employed. Furthermore, the expertise acquired in the particle tracking field, during the service work performed for the ALICE Collaboration, is expect to be extremely valuable for the continuation of the iMPACT project.
Finally, an innovative Fully-Depleted Monolithic Active Pixel Sensor, in development within the INFN ARCADIA Collaboration, is foreseen to instrument the future versions of the iMPACT pCT scanner. An innovative read-out architecture based on passive pixel addressing and projective data compression will allow for an efficient and low-power sensor operation.
Overall, the performed activities on the iMPACT proton Computed Tomography scanner have brought very important results, showing the real potential for a decisive breakthrough towards the clinical application of the pCT method, in terms of significant improvement of the accuracy of the hadron-therapy planning and, as a direct consequence, an advancement in the efficiency of the tumor treatment. As a consequence, the development of proton tomography could further accelerate and spread hadron-therapy techniques.

Appendices

A Proton and hadron-therapy

	Particle facility				Photon facility
	Combined carbon-ion and proton	*Ratio to photon*	Proton-only	*Ratio to photon*	
Capital costs					
Base case scenario	€ 138,600,000		€ 94,930,000		€ 23,430,000
Optimistic scenario (−20%)	€ 110,880,000		€ 75,944,000		€ 18,744,000
Pessimistic scenario (+20%)	€ 166,320,000		€ 113,916,000		€ 28,116,000
	Cost per fraction		Cost per fraction		Cost per fraction
Base case scenario	€ 1128	4.8	€ 743	3.2	€ 233
Optimistic scenario (−20%)	€ 1100	4.8	€ 724	3.2	€ 229
Pessimistic scenario (+20%)	€ 1156	4.9	€ 762	3.2	€ 237
Operational costs					
Base case scenario	€ 32,138,027		€ 21,800,383		€ 8,800,850
Optimistic scenario (−20%)	€ 25,710,421		€ 17,440,306		€ 7,040,680
Pessimistic scenario (+20%)	€ 38,565,632		€ 26,160,459		€ 10,561,020
	Cost per fraction		Cost per fraction		Cost per fraction
Base case scenario	€ 1128	4.8	€ 743	3.2	€ 233
Optimistic scenario (−20%)	€ 931	4.9	€ 613	3.2	€ 190
Pessimistic scenario (+20%)	€ 1325	4.8	€ 872	3.2	€ 276
Lifecycle of a facility	Cost per fraction		Cost per fraction		Cost per fraction
Base case: 30 years	€ 1128	4.8	€ 743	3.2	€ 233
5 years	€ 1837	5.6	€ 1213	3.7	€ 328
40 years	€ 1093	4.8	€ 719	3.2	€ 228
No. of patients per year[a]					
Base case scenario	1810		1681		2287
Optimistic scenario (+20%)	2172		2017		2744
Pessimistic scenario (−20%)	1448		1345		1829
Pessimistic scenario (−30%)	1267		1176		1601
Pessimistic scenario (−50%)	905		840		1143
	Cost per fraction		Cost per fraction		Cost per fraction
Base case scenario	€ 1128	4.8	€ 743	3.2	€ 233
Optimistic scenario (+20%)	€ 940	4.8	€ 619	3.2	€ 194
Pessimistic scenario (−20%)	€ 1410	4.8	€ 928	3.2	€ 291
Pessimistic scenario (−30%)	€ 1612	4.8	€ 1061	3.2	€ 333
Pessimistic scenario (−50%)	€ 2256	4.8	€ 1485	3.2	€ 466
Average time per fraction (particle)	Cost per fraction		Cost per fraction		Cost per fraction
Base case: 18 min	€ 1128	4.8	€ 743	3.2	–
Optimistic scenario (−4 min)	€ 877	3.8	€ 578	2.5	–
Pessimistic scenario (+4 min)	€ 1379	5.9	€ 908	3.9	–
Hours per day available	Cost per fraction		Cost per fraction		Cost per fraction
Base case: 14 h/day	€ 1128	4.8	€ 743	3.2	€ 233
8 h/day	€ 1974	4.8	€ 1300	3.2	€ 407
16 h/day	€ 987	4.8	€ 650	3.2	€ 204
Treatment room availability	Cost per fraction		Cost per fraction		Cost per fraction
Base case: 95%, 98%, 98%	€ 1128	4.8	€ 743	3.2	€ 233
90% all facilities	€ 1191	4.7	€ 809	3.2	€ 253
Treatment room utilization (particle)	Cost per fraction		Cost per fraction		Cost per fraction
Base case: 98%	€ 1128	4.8	€ 743	3.2	–
90%	€ 1228	5.3	€ 809	3.5	–
Energy costs	Cost per fraction		Cost per fraction		Cost per fraction
Base case: 100 € /MWh/a	€ 1128	4.8	€ 743	3.2	€ 233
150 € /MWh/a	€ 1147	4.9	€ 749	3.2	€ 236
200 € /MWh/a	€ 1165	4.9	€ 755	3.2	€ 238
Interest payment	Cost per fraction		Cost per fraction		Cost per fraction
Base case: 5% over 20 years	€ 1128	4.8	€ 743	3.2	€ 233
5% over 25 years	€ 1089	4.8	€ 716	3.1	€ 228
5% over 30 years	€ 1063	4.7	€ 700	3.1	€ 224
No interest payment	€ 787	4.2	€ 516	2.8	€ 187
T4 category photon treatments only [20]	Cost per fraction		Cost per fraction		Cost per fraction
Base case: standard treatment	€ 1128	4.8	€ 743	3.2	€ 233
Weight factor 1.5	–	3.2	–	2.1	€ 350
Weight factor 2.9	–	1.7	–	1.1	€ 676
Weight factor 5.0	–	1.0	–	0.6	€ 1165

[a] Assuming an average number of fractions per patient (specific for each facility).

Tab. 8: Estimation of the cost per treatment fraction of X-ray radiotherapy, proton therapy, and combined proton-carbon therapy. Table from [190].

B Image reconstruction

From Sec. 7 the term $\mathcal{L}(\mathbf{t}_2|\mathbf{t}_1)$ from Eq. (37), representing the likelihood of finding a proton with exiting coordinates $\mathbf{t}_2 = (x_2 \;\; \theta_2)^T$ after being at coordinates $\mathbf{t}_1$ inside the object, can be expressed as:

$$\mathcal{L}(\mathbf{t}_2|\mathbf{t}_1) = \exp\left(-\frac{1}{2}\left(\mathbf{t}_2^T - \mathbf{t}_1^T\mathbf{R}_1^T\right)\boldsymbol{\Sigma}_2^{-1}\left(\mathbf{t}_2 - \mathbf{R}_1\mathbf{t}_1\right)\right) \quad . \tag{80}$$

The matrix $\mathbf{R}_1$ describes the reference transformation defined by x_1 and θ_1, assuming valid the small angle approximation:

$$\mathbf{R}_1 = \begin{pmatrix} 1 & z_2 - z_1 \\ 0 & 1 \end{pmatrix} \quad , \tag{81}$$

the covariance matrix $\boldsymbol{\Sigma}_2$ has is defined similarly to $\boldsymbol{\Sigma}1$ in Eq. (41):

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} \sigma^2_{x_2} & \sigma^2_{x_2\theta_2} \\ \sigma^2_{x_2\theta_2} & \sigma^2_{\theta_2} \end{pmatrix} \quad , \tag{82}$$

where its elements are calculated as in Eq. (42)-(44) with z_2 and z_1 as extremes of the integration:

$$\sigma^2_{x_2}(z_2, z_1) = E_0^2\left(1 + 0.038\ln\frac{z_2 - z_1}{X_0}\right)^2 \times \int_{z_1}^{z_2} \frac{(z_2 - z)^2}{\beta^2(z)p^2(z)}\frac{dz}{X_0} \quad , \tag{83}$$

$$\sigma^2_{\theta_2}(z_2, z_1) = E_0^2\left(1 + 0.038\ln\frac{z_2 - z_1}{X_0}\right)^2 \times \int_{z_1}^{z_2} \frac{1}{\beta^2(z)p^2(z)}\frac{dz}{X_0} \quad , \tag{84}$$

$$\sigma^2_{x_2\theta_2}(z_2, z_1) = E_0^2\left(1 + 0.038\ln\frac{z_2 - z_1}{X_0}\right)^2 \times \int_{z_1}^{z_2} \frac{z_2 - z}{\beta^2(z)p^2(z)}\frac{dz}{X_0} \quad . \tag{85}$$

The explicit expression of the Maximum Likelihood Path calculation will be obtained. The MLP coordinates $\mathbf{t}_1^{\text{MLP}}$ at z_1 depth, given initial coordinates $\mathbf{t}_0$ at z_0 and final coordinates $\mathbf{t}_2$ at z_2, is given by the matricial expression:

$$\mathbf{t}_1^{\text{MLP}}(z_1|\mathbf{t}_0, z_0, \mathbf{t}_2, z_2) = \begin{pmatrix} x_1^{\text{MLP}} \\ \theta_1^{\text{MLP}} \end{pmatrix} = (\boldsymbol{\Sigma}_1^{-1} + \mathbf{R}_1^T\boldsymbol{\Sigma}_2^{-1}\mathbf{R}_1)^{-1}(\boldsymbol{\Sigma}_1^{-1}\mathbf{R}_0\mathbf{t}_0 + \mathbf{R}_1^T\boldsymbol{\Sigma}_2^{-1}\mathbf{t}_2) \quad . \tag{86}$$

From the definitions of $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\mathbf{R}_1$, the first term can be calculated:

$$\boldsymbol{\Sigma}_1^{-1} + \mathbf{R}_1^T\boldsymbol{\Sigma}_2^{-1}\mathbf{R}_1 =$$

$$\frac{1}{\det\boldsymbol{\Sigma}_1}\begin{pmatrix} \sigma^2_{\theta_1} & -\sigma^2_{x_1\theta_1} \\ -\sigma^2_{x_1\theta_1} & \sigma^2_{x_1} \end{pmatrix} + \frac{1}{\det\boldsymbol{\Sigma}_2}\begin{pmatrix} 1 & 0 \\ z_2 - z_1 & 1 \end{pmatrix}\begin{pmatrix} \sigma^2_{\theta_2} & -\sigma^2_{x_2\theta_2} \\ -\sigma^2_{x_2\theta_2} & \sigma^2_{x_2} \end{pmatrix}\begin{pmatrix} 1 & z_2 - z_1 \\ 0 & 1 \end{pmatrix} =$$

$$\frac{1}{\det\boldsymbol{\Sigma}_1}\begin{pmatrix} \sigma^2_{\theta_1} & -\sigma^2_{x_1\theta_1} \\ -\sigma^2_{x_1\theta_1} & \sigma^2_{x_1} \end{pmatrix} +$$

$$\frac{1}{\det\boldsymbol{\Sigma}_2}\begin{pmatrix} \sigma^2_{\theta_2} & \sigma^2_{\theta_2}(z_2 - z_1) - \sigma^2_{x_2\theta_2} \\ \sigma^2_{\theta_2}(z_2 - z_1) - \sigma^2_{x_2\theta_2} & \sigma^2_{\theta_2}(z_2 - z_1)^2 - 2\sigma^2_{x_2\theta_2}(z_2 - z_1) + \sigma^2_{x_2} \end{pmatrix} \equiv$$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad , \tag{87}$$

where $\det\boldsymbol{\Sigma_1} = \sigma^2_{x_1}\sigma^2_{\theta_1} - \sigma^4_{x_1\theta_1}$ and similarly for $\det\boldsymbol{\Sigma_2}$.
The second term can be expressed as:

$$\boldsymbol{\Sigma}_1^{-1}\mathbf{R}_0\mathbf{t}_0 + \mathbf{R}_1^T\boldsymbol{\Sigma}_2^{-1}\mathbf{t}_2 =$$

$$\frac{1}{\det \mathbf{\Sigma}_1}\begin{pmatrix}\sigma^2_{\theta_1} & -\sigma^2_{x_1\theta_1}\\ -\sigma^2_{x_1\theta_1} & \sigma^2_{x_1}\end{pmatrix}\begin{pmatrix}1 & z_1-z_0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_0\\ \theta_0\end{pmatrix}+$$
$$\frac{1}{\det \mathbf{\Sigma}_2}\begin{pmatrix}1 & 0\\ z_2-z_1 & 1\end{pmatrix}\begin{pmatrix}\sigma^2_{\theta_2} & -\sigma^2_{x_2\theta_2}\\ -\sigma^2_{x_2\theta_2} & \sigma^2_{x_2}\end{pmatrix}\begin{pmatrix}x_2\\ \theta_2\end{pmatrix}=$$
$$\frac{1}{\det \mathbf{\Sigma}_1}\begin{pmatrix}\sigma^2_{\theta_1}(x_0+\theta_0(z_1-z_0))-\theta_0\sigma^2_{x_1\theta_1}\\ \theta_0\sigma^2_{x_1}-\sigma^2_{x_1\theta_1}(x_0+\theta_0(z_1-z_0))\end{pmatrix}+$$
$$\frac{1}{\det \mathbf{\Sigma}_2}\begin{pmatrix}x_2\sigma^2_{\theta_2}-\theta_2\sigma^2_{x_2\theta_2}\\ (x_1\sigma^2_{\theta_2}-\theta_2\sigma^2_{x_2\theta_2})(z_2-z_1)+\theta_2\sigma^2_{t_2}-x_2\sigma^2_{x_2\theta_2}\end{pmatrix}\equiv\begin{pmatrix}b_1\\ b_2\end{pmatrix}\quad . \tag{88}$$

Then from Eq. (86), (87) and (88):

$$\mathbf{t}_1^{\mathrm{MLP}}(z_1|\mathbf{t}_0,z_0,\mathbf{t}_2,z_2)=$$
$$(\mathbf{\Sigma}_1^{-1}+\mathbf{R}_1^T\mathbf{\Sigma}_2^{-1}\mathbf{R}_1)^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{R}_0\mathbf{t}_0+\mathbf{R}_1^T\mathbf{\Sigma}_2^{-1}\mathbf{t}_2)=$$
$$\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}^{-1}\begin{pmatrix}b_1\\ b_2\end{pmatrix}=$$
$$\frac{1}{a_{11}a_{22}-a_{12}^2}\begin{pmatrix}a_{22}b_1-a_{12}b_2\\ a_{11}b_2-a_{12}b_1\end{pmatrix}=\begin{pmatrix}x_1^{\mathrm{MLP}}\\ \theta_1^{\mathrm{MLP}}\end{pmatrix}\quad , \tag{89}$$

where the first element is therefore the most likely lateral position at depth z_1.

C ALICE

L	S	r_x	σ_{r_x}	r_z	σ_{r_z}
0	0	-3.13E-1	2.5E-2	1.20E-1	7.0E-3
0	1	1.26E-1	4.8E-3	2.42E-2	1.9E-3
0	2	1.20E-1	2.0E-3	-6.75E-2	1.5E-3
0	3	2.57E-2	1.8E-3	-7.49E-3	1.4E-3
0	4	-1.93E-3	1.4E-3	5.02E-2	1.2E-3
0	5	-2.07E-1	1.4E-2	1.52E-2	3.2E-3
0	6	-7.28E-2	8.5E-3	5.77E-2	5.7E-3
0	7	-2.09E-1	1.2E-2	3.44E-2	2.3E-3
0	8	-5.11E-2	2.5E-3	5.49E-2	1.8E-3
0	9	-1.79E-2	1.4E-3	-7.46E-2	1.8E-3
0	10	4.73E-2	1.6E-3	4.48E-2	1.7E-3
0	11	4.54E-2	2.2E-3	8.37E-2	2.8E-3
1	0	1.51E-1	1.0E-2	-1.08E-1	6.6E-3
1	1	9.07E-3	3.5E-3	-6.50E-2	2.3E-3
1	2	-6.18E-2	1.8E-3	-3.04E-4	9.0E-4
1	3	-5.97E-2	8.8E-4	5.73E-2	5.4E-4
1	4	-2.53E-2	1.0E-3	-2.65E-3	7.4E-4
1	5	-1.95E-2	7.3E-4	-2.68E-2	9.1E-4
1	6	4.64E-2	5.4E-4	-1.53E-2	9.3E-4
1	7	4.90E-1	5.9E-3	-8.42E-1	1.1E-3
1	8	8.60E-2	9.9E-3	-5.89E-2	5.2E-3
1	9	2.04E-1	4.4E-3	-4.24E-2	1.1E-3
1	10	7.72E-2	2.4E-3	-1.23E-2	6.6E-4
1	11	4.53E-2	9.8E-4	-3.48E-2	9.5E-4
1	12	7.94E-3	8.8E-4	7.52E-2	7.2E-4
1	13	-1.99E-2	8.4E-4	-8.40E-3	9.0E-4
1	14	-2.90E-2	7.0E-4	-5.10E-2	1.0E-3
1	15	-1.82E-2	2.3E-3	-1.05E-1	2.8E-3
2	0	-3.56E-2	1.2E-2	9.33E-2	2.1E-2
2	1	-5.04E-2	1.6E-2	1.87E-1	6.8E-3
2	2	1.31E-1	8.4E-3	-1.30E-2	2.9E-3
2	3	7.59E-2	2.8E-3	-8.08E-3	1.7E-3
2	4	9.03E-2	1.9E-3	-8.65E-2	1.4E-3
2	5	1.37E-2	1.3E-3	-5.44E-3	1.3E-3
2	6	1.57E-2	1.0E-3	7.68E-2	1.4E-3
2	7	-1.97E-2	1.9E-3	1.15E-2	2.0E-3
2	8	-3.61E-2	5.0E-3	9.52E-2	2.2E-2
2	9	2.10E-1	6.8E-2	3.70E-1	4.6E-2
2	10	-1.92E-1	1.3E-2	9.15E-2	1.3E-2
2	11	-2.48E-1	9.2E-3	4.40E-2	3.0E-3
2	12	-2.23E-1	7.1E-3	5.39E-2	2.1E-3
2	13	-1.33E-1	2.9E-3	1.94E-2	1.4E-3
2	14	-7.20E-2	2.2E-3	6.08E-2	2.1E-3
2	15	-2.42E-2	1.5E-3	-1.31E-1	1.6E-3
2	16	1.42E-2	1.6E-3	-7.11E-3	1.8E-3
2	17	4.61E-2	1.2E-3	8.13E-2	2.3E-3
2	18	3.40E-2	3.4E-3	7.58E-2	4.2E-3
2	19	4.50E-2	9.1E-3	8.44E-2	1.0E-2

Tab. 9: Values of average residual $r_{x,z}$ per each Stave for the Inner Barrel, obtained as the central value of a Gaussian fit of the residuals distributions. $\sigma_{r_{x,z}}$ is the error on the central value parameter obtained from the fit. Values are all expressed in mm. Data plotted in Fig. 123.

L	S	r_x	σ_{r_x}	r_z	σ_{r_z}
3	0	-3.55E-1	6.6E-2	-1.14E-2	1.6E-2
3	1	-8.00E-2	1.7E-2	-4.19E-3	9.4E-3
3	2	-1.16E-1	8.5E-3	2.92E-2	9.4E-3
3	3	-1.13E-1	4.6E-3	1.04E-3	6.1E-3
3	4	1.01E-2	5.0E-3	-3.31E-3	7.0E-3
3	5	3.78E-2	4.9E-3	1.62E-2	6.6E-3
3	6	-2.39E-2	6.4E-3	2.49E-3	6.3E-3
3	7	-1.01E-1	7.8E-3	1.32E-2	6.0E-3
3	8	2.68E-2	5.0E-3	1.18E-2	6.0E-3
3	9	1.17E-1	5.5E-3	-7.69E-3	5.9E-3
3	10	9.45E-2	1.2E-2	-1.98E-2	8.0E-3
3	11	4.26E-1	3.6E-2	-4.35E-2	1.1E-2
3	12	-4.10E-1	3.6E-2	-1.97E-2	1.3E-2
3	13	-1.53E-1	1.0E-2	9.76E-3	6.3E-3
3	14	-9.71E-2	5.3E-3	-4.47E-3	4.6E-3
3	15	-8.61E-2	3.3E-3	-7.50E-3	4.1E-3
3	16	1.44E-2	3.7E-3	-4.16E-3	4.1E-3
3	17	5.56E-2	3.0E-3	-2.61E-2	4.0E-3
3	18	-1.48E-2	3.4E-3	-4.90E-3	4.5E-3
3	19	-1.03E-1	4.6E-3	-2.01E-3	3.9E-3
3	20	1.06E-2	3.8E-3	-1.57E-2	3.9E-3
3	21	1.50E-1	4.1E-3	-2.90E-2	4.5E-3
3	22	1.79E-1	7.0E-3	-7.99E-3	7.0E-3
3	23	4.52E-1	2.1E-2	1.11E-2	1.1E-2
4	0	-9.93E-2	6.0E-2	8.08E-2	1.7E-2
4	1	-3.04E-2	2.3E-2	1.53E-2	1.0E-2
4	2	-6.94E-2	8.3E-3	3.32E-3	7.3E-3
4	3	-9.34E-2	5.8E-3	2.06E-2	6.7E-3
4	4	-5.60E-2	3.4E-3	3.93E-3	4.5E-3
4	5	-1.17E-2	5.3E-3	1.39E-3	5.2E-3
4	6	4.34E-2	3.6E-3	-5.54E-3	5.8E-3
4	7	-7.47E-3	4.4E-3	8.31E-3	5.6E-3
4	8	-6.15E-2	5.4E-3	-3.13E-2	5.4E-3
4	9	-1.61E-1	7.1E-3	5.87E-3	4.9E-3
4	10	5.05E-2	3.6E-3	-5.57E-3	4.5E-3
4	11	7.82E-2	3.6E-3	-6.02E-3	4.1E-3
4	12	8.96E-2	7.0E-3	1.38E-2	6.5E-3
4	13	1.42E-1	1.5E-2	2.68E-2	1.0E-2
4	14	4.09E-1	4.0E-2	-5.24E-2	1.7E-2
4	15	1.96E-2	1.5E-2	-2.43E-2	1.4E-2
4	16	-4.75E-2	2.0E-2	-9.66E-3	1.1E-2
4	17	-4.74E-2	6.4E-3	2.02E-3	5.1E-3
4	18	-1.18E-1	3.4E-3	7.40E-3	3.6E-3
4	19	-6.03E-2	2.7E-3	2.88E-3	3.0E-3
4	20	1.32E-1	7.0E-3	3.67E-3	3.2E-3
4	21	5.35E-2	2.2E-3	-8.30E-4	3.2E-3
4	22	-2.04E-3	2.7E-3	-3.58E-3	3.6E-3
4	23	-9.93E-2	3.8E-3	-1.31E-3	4.2E-3
4	24	-1.15E-1	4.4E-3	-5.73E-3	3.1E-3
4	25	3.80E-2	2.6E-3	-7.95E-3	2.9E-3
4	26	-4.01E-2	1.6E-2	-1.15E-2	4.5E-3
4	27	1.02E-1	4.6E-3	-4.03E-2	6.4E-3
4	28	2.13E-1	1.2E-2	-2.57E-2	7.6E-3
4	29	3.03E-1	3.6E-2	-3.25E-2	1.6E-2
5	0	-4.42E-2	2.4E-2	4.15E-1	3.3E-2
5	1	-3.59E-2	8.6E-3	-2.06E-2	1.2E-2
5	2	5.93E-2	1.1E-2	1.53E-2	6.9E-3
5	3	-1.74E-2	7.9E-3	-8.40E-3	6.6E-3
5	4	7.48E-3	4.0E-3	-3.11E-2	4.6E-3
5	5	-7.69E-3	4.0E-3	-6.81E-3	5.0E-3
5	6	2.29E-2	4.1E-3	1.30E-3	4.8E-3
5	7	-4.75E-2	4.2E-3	-5.87E-3	4.9E-3
5	8	-2.81E-2	3.0E-3	-9.08E-3	5.0E-3
5	9	3.76E-2	3.5E-3	-5.72E-3	6.1E-3
5	10	-6.58E-2	5.8E-3	1.67E-2	6.5E-3
5	11	1.81E-2	2.8E-3	-1.65E-3	4.6E-3
5	12	6.84E-3	3.4E-3	-6.69E-3	4.0E-3
5	13	-2.65E-2	4.0E-3	-4.95E-3	4.1E-3
5	14	1.01E-2	4.6E-3	9.23E-3	4.8E-3
5	15	3.26E-2	3.3E-3	6.16E-3	4.5E-3
5	16	-9.71E-3	4.2E-3	1.44E-2	5.5E-3
5	17	1.54E-2	6.7E-3	-4.89E-3	6.7E-3
5	18	3.20E-2	9.1E-3	6.11E-3	7.8E-3
5	19	5.11E-2	1.2E-2	8.11E-3	1.2E-2
5	20	-8.44E-2	5.5E-2	-9.53E-2	3.0E-2
5	21	-9.81E-2	1.8E-2	-8.68E-2	2.5E-2
5	22	-3.90E-2	6.5E-3	-4.16E-3	9.5E-3
5	23	7.67E-1	6.5E-1	-1.66E-1	1.2E-1
5	24	1.70E-2	5.7E-3	-6.91E-4	4.4E-3
5	25	1.52E-2	3.5E-3	-3.78E-3	3.7E-3
5	26	2.19E-2	5.5E-3	9.96E-3	3.3E-3
5	27	-1.43E-2	2.8E-3	7.96E-3	3.2E-3
5	28	-6.26E-2	2.8E-3	-4.10E-3	3.4E-3
5	29	-3.13E-2	1.9E-3	-1.75E-2	3.2E-3
5	30	-2.78E-2	2.2E-3	-2.63E-3	3.1E-3
5	31	-3.67E-3	2.5E-3	-2.64E-5	3.7E-3
5	32	3.43E-2	2.3E-3	-1.20E-2	3.0E-3
5	33	-3.86E-3	2.2E-3	5.40E-4	2.6E-3
5	34	1.20E-2	2.6E-3	-9.09E-3	2.9E-3
5	35	-7.07E-3	3.0E-3	-3.28E-4	3.0E-3
5	36	-2.63E-2	4.5E-3	-2.14E-2	3.8E-3
5	37	-6.13E-3	3.1E-3	1.85E-3	4.0E-3
5	38	-2.67E-2	4.5E-3	7.19E-3	4.5E-3
5	39	6.51E-2	6.5E-3	-3.05E-3	4.7E-3
5	40	1.47E-1	6.8E-3	8.55E-3	8.9E-3
5	41	-2.52E-2	4.0E-2	-1.93E-2	1.9E-2
6	0	-8.24E-2	6.7E-2	-4.68E-2	4.3E-2
6	1	9.45E-3	2.9E-2	-4.47E-2	3.7E-2
6	2	-6.56E-2	9.6E-3	-2.03E-2	1.9E-2
6	3	-8.57E-2	1.6E-2	-1.43E-2	8.7E-3
6	4	2.62E-2	8.9E-3	1.48E-2	8.7E-3
6	5	3.34E-2	5.0E-3	-8.11E-3	6.4E-3
6	6	3.62E-2	5.8E-3	-7.33E-3	6.2E-3
6	7	4.00E-2	5.9E-3	-5.64E-3	6.5E-3
6	8	1.73E-2	5.0E-3	7.09E-3	6.4E-3
6	9	-5.46E-2	4.0E-3	-2.79E-3	6.8E-3
6	10	-2.69E-2	3.7E-3	2.28E-3	6.4E-3
6	11	-3.39E-2	1.1E-2	1.07E-2	1.2E-2
6	12	8.56E-3	4.5E-3	2.43E-3	6.3E-3
6	13	-1.56E-4	3.6E-3	-4.25E-3	5.8E-3
6	14	1.90E-2	4.5E-3	-1.51E-2	5.6E-3
6	15	5.18E-2	6.9E-3	-1.26E-2	6.1E-3
6	16	-2.47E-2	6.1E-3	3.70E-3	5.9E-3
6	17	-4.21E-2	4.4E-3	5.59E-3	6.5E-3
6	18	-3.59E-2	5.0E-3	2.02E-3	7.4E-3
6	19	-7.66E-2	7.6E-3	1.84E-2	8.5E-3
6	20	5.59E-2	1.4E-2	1.50E-2	9.3E-3
6	21	-3.22E-2	1.5E-2	-7.31E-2	1.8E-2
6	22	-5.15E-2	4.0E-2	-5.52E-2	4.4E-2
6	23	6.27E-2	5.3E-2	-7.78E-2	3.9E-2
6	24	-4.09E-2	2.1E-2	-5.95E-3	1.7E-2
6	25	1.27E-1	2.4E-2	-1.61E-2	2.2E-2
6	26	-6.95E-2	1.1E-2	2.86E-2	1.6E-2
6	27	2.97E-2	1.1E-2	2.82E-3	2.5E-2
6	28	3.56E-3	7.1E-3	-1.05E-3	5.7E-3
6	29	4.54E-2	4.4E-3	-2.27E-3	5.0E-3
6	30	2.20E-2	3.5E-3	3.79E-3	4.1E-3
6	31	1.17E-2	3.9E-3	1.79E-3	4.5E-3
6	32	-1.49E-2	3.3E-3	-9.95E-3	4.1E-3
6	33	-4.97E-2	2.4E-3	-1.86E-2	4.4E-3
6	34	-4.58E-2	2.6E-3	-4.97E-3	4.4E-3
6	35	-4.00E-2	3.3E-3	1.14E-2	4.5E-3
6	36	2.08E-2	3.0E-3	-6.87E-3	4.3E-3
6	37	1.71E-2	2.5E-3	-1.28E-2	4.1E-3
6	38	4.36E-2	3.8E-3	7.57E-4	3.5E-3
6	39	4.11E-2	4.6E-3	-1.04E-2	4.0E-3
6	40	-4.48E-2	4.2E-3	-1.79E-3	3.9E-3
6	41	-9.83E-3	4.1E-3	-1.93E-2	5.9E-3
6	42	-5.36E-2	3.9E-3	3.78E-3	5.3E-3
6	43	-4.90E-2	5.8E-3	1.23E-4	5.8E-3
6	44	3.64E-2	9.6E-3	2.72E-3	6.2E-3
6	45	7.65E-2	9.3E-3	2.04E-2	1.5E-2
6	46	-6.23E-2	2.9E-2	-1.22E-2	2.3E-2
6	47	1.21E-1	3.5E-2	1.07E-2	2.5E-2

Tab. 10: Values of average residual $r_{x,z}$ per each Stave for the Outer Barrel, obtained as the central value of a Gaussian fit of the residuals distributions. $\sigma_{r_{x,z}}$ is the error on the central value parameter obtained from the fit. Values are all expressed in mm. Data are plotted in Fig. 124.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] W. H. O. (WHO) *et al.*, "Global health estimates 2020: Deaths by cause, age, sex, by country and by region, 2000-2019," *World Health Organization*, 2020, [accessed 18-December-2021]. [Online]. Available: who.int/data/gho/data/theme%20s/mortality-and-global-health-estimates/ghe-leading-causes-of-death.

[3] L. S. Mørch, E. Løkkegaard, A. H. Andreasen, S. Krüger-Kjær, and Ø. Lidegaard, "Hormone therapy and ovarian cancer," *Jama*, vol. 302, no. 3, pp. 298–305, 2009.

[4] D. Schardt, T. Elsässer, and D. Schulz-Ertner, "Heavy-ion tumor therapy: Physical and radiobiological benefits," *Reviews of modern physics*, vol. 82, no. 1, p. 383, 2010.

[5] D. Cussol, "Hadron therapy," in *Ecole Joliot-Curie (30 years)" Physics at the femtometer scale"*, 2011, 46–p.

[6] C. Griep, J. Davelaar, A. N. Scholten, A. Chin, and J.-W. H. Leer, "Electron beam therapy is not inferior to superficial x-ray therapy in the treatment of skin carcinoma," *International Journal of Radiation Oncology* Biology* Physics*, vol. 32, no. 5, pp. 1347–1350, 1995.

[7] W. Bragg and R. Kleeman, "On the otparticles of radium, and their loss of range in passing through various atoms and molecules," *Philos. Mag. & J. Sci*, pp. 318–340, 1905.

[8] *Particle Therapy Co-Operative Group*, [accessed 18-December-2021]. [Online]. Available: https://www.ptcog.ch/index.php/facilities-in-operation.

[9] R. Mohan and D. Grosshans, "Proton therapy–present and future," *Advanced drug delivery reviews*, vol. 109, pp. 26–44, 2017.

[10] M. Krämer, O. Jäkel, T. Haberer, G. Kraft, D. Schardt, and U. Weber, "Treatment planning for heavy-ion radiotherapy: Physical beam model and dose optimization," *Physics in Medicine & Biology*, vol. 45, no. 11, p. 3299, 2000.

[11] M. Yoon, S. H. Ahn, J. Kim, D. H. Shin, S. Y. Park, S. B. Lee, K. H. Shin, and K. H. Cho, "Radiation-induced cancers from modern radiotherapy techniques: Intensity-modulated radiotherapy versus proton therapy," *International Journal of Radiation Oncology* Biology* Physics*, vol. 77, no. 5, pp. 1477–1485, 2010.

[12] W. D. Newhauser and R. Zhang, "The physics of proton therapy," *Physics in Medicine & Biology*, vol. 60, no. 8, R155, 2015.

[13] D. Chiappara, "Preliminary modelling for the proton boron capture therapy (PBCT)," 2019.

[14] E. J. Hall and A. J. Giaccia, *Radiobiology for the Radiologist*. Lippincott Williams & Wilkins, 2006.

[15] D. J. Brenner, "The linear-quadratic model is an appropriate methodology for determining isoeffective doses at large doses per fraction," in *Seminars in radiation oncology*, Elsevier, vol. 18, 2008, pp. 234–239.

[16] M. Scholz, "Effects of ion radiation on cells and tissues," *Radiation effects on polymers for biological use*, pp. 95–155, 2003.

[17] R. D. Rivarola, M. E. Galassi, P. D. Fainstein, and C. Champion, "Computation of distorted wave cross sections for high-energy inelastic collisions of heavy ions with water molecules," *Advances in Quantum Chemistry*, vol. 65, pp. 231–267, 2013.

[18] E. A. Blakely, C. A. Tobias, T. C. Yang, K. C. Smith, and J. T. Lyman, "Inactivation of human kidney cells by high-energy monoenergetic heavy-ion beams," *Radiation research*, vol. 80, no. 1, pp. 122–160, 1979.

[19] H. Bethe, "Zur theorie des durchgangs schneller korpuskularstrahlen durch materie," *Annalen der Physik*, vol. 397, no. 3, pp. 325–400, 1930.

[20] F. Bloch, "Zur bremsung rasch bewegter teilchen beim durchgang durch materie," *Annalen der Physik*, vol. 408, no. 3, pp. 285–320, 1933.

[21] N. Bohr, "Lx. on the decrease of velocity of swiftly moving electrified particles in passing through matter," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 30, no. 178, pp. 581–612, 1915.

[22] U. Fano, "Penetration of protons, alpha particles, and mesons," *Annual Review of Nuclear Science*, vol. 13, no. 1, pp. 1–66, 1963.

[23] W. H. Barkas, "Nuclear research emulsions," 1963.

[24] J. F. Ziegler, M. D. Ziegler, and J. P. Biersack, "Srim–the stopping and range of ions in matter (2010)," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 268, no. 11, pp. 1818–1823, 2010.

[25] H. Bethe and J. Ashkin, *Passage of radiations through matter," experimental nuclear physics, vol. 1, edited: E. segre*, Wiley, New York, 1953.

[26] W. Ulmer and E. Matsinos, "Theoretical methods for the calculation of bragg curves and 3d distributions of proton beams," *The European Physical Journal-Special Topics*, vol. 190, no. 1, pp. 1–81, 2010.

[27] NIST (National Institute of Standards and Technology) database, http://physics.nist.gov/PhysRefData/Star/Text/PSTAR.html.

[28] P. Vavilov, "Ionization losses of high-energy heavy particles," *Soviet Phys. JETP*, vol. 5, 1957.

[29] L. D. Landau, "On the energy loss of fast particles by ionization," *J. Phys.*, vol. 8, pp. 201–205, 1944.

[30] M. J. Berger, M. Inokuti, H. H. Andersen, H. Bichsel, D. Powers, S. M. Seltzer, D. Thwaites, and D. E. Watt, "6. energy-loss straggling," *Reports of the International Commission on Radiation Units and Measurements*, vol. os-25, no. 2, pp. 61–68, 1993. DOI: 10.1093/jicru_os25.2.61. eprint: https://doi.org/10.1093/jicru_os25.2.61. [Online]. Available: https://doi.org/10.1093/jicru_os25.2.61.

[31] N. Bohr, "Scattering and stopping of fission fragments," *Physical Review*, vol. 58, no. 7, p. 654, 1940, APS.

[32] S. P. Ahlen, "Theoretical and experimental aspects of the energy loss of relativistic heavily ionizing particles," *Reviews of Modern Physics*, vol. 52, no. 1, p. 121, 1980.

[33] B. B. Rossi, "High-energy particles," 1952.

[34] J. F. Janni, "Proton range-energy tables, 1 kev-10 gev, energy loss, range, path length, time-of-flight, straggling, multiple scattering, and nuclear interaction probability. part i. for 63 compounds," *Atomic data and nuclear data tables*, vol. 27, p. 147, 1982.

[35] E. Rutherford, "Lxxix. the scattering of α and β particles by matter and the structure of the atom," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 21, no. 125, pp. 669–688, 1911.

[36] V. L. Highland, "Some practical remarks on multiple scattering," *Nuclear Instruments and Methods*, vol. 129, no. 2, pp. 497–499, 1975.

[37] Y.-S. Tsai, "Pair production and bremsstrahlung of charged leptons," *Reviews of Modern Physics*, vol. 46, no. 4, p. 815, 1974.

[38] W. C. Röntgen, "On a new kind of rays," *Science*, vol. 3, no. 59, pp. 227–231, 1896.

[39] J. M. Slater, "From x-rays to ion beams: A short history of radiation therapy," in *Ion Beam Therapy*, Springer, 2012, pp. 3–16.

[40] R. R. Wilson, "Radiological use of fast protons," *Radiology*, vol. 47, no. 5, pp. 487–491, 1946.

[41] M. L. Boone, J. H. Lawrence, W. G. Connor, R. Morgado, J. A. Hicks, and R. C. Brown, "Introduction to the use of protons and heavy ions in radiation therapy: Historical perspective," *International Journal of Radiation Oncology* Biology* Physics*, vol. 3, pp. 65–69, 1977.

[42] C. A. Tobias, J. Roberts, J. Lawrence, B. Low-Beer, H. Anger, J. Born, R. McCombs, and C. Huggins, "Irradiation hypophysectomy and related studies using 340-mev protons and 190-mev deuterons," 1955.

[43] C. A. Tobias and P. W. Todd, "Heavy charged particles in cancer therapy.," Univ. of California, Berkeley, Tech. Rep., 1967.

[44] P. H. Fowler and D. H. Perkins, "The possibility of therapeutic applications of beams of negative *pi*-mesons," *Nature*, vol. 189, 1961.

[45] R. Greiner, C. Von Essen, H. Blattmann, U. Studer, A. Zimmermann, G. Bodendörfer, and G. Schmitt, "Results of curative pion therapy at sin.," *Strahlentherapie*, vol. 161, no. 12, pp. 797–800, 1985.

[46] G. Cirrone, G. Cuttone, P. Lojacono, S. L. Nigro, V. Mongelli, I. Patti, G. Privitera, L. Raffaele, D. Rifuggiato, M. Sabini, *et al.*, "A 62-mev proton beam for the treatment of ocular melanoma at laboratori nazionali del sud-infn," *IEEE Transactions on Nuclear Science*, vol. 51, no. 3, pp. 860–865, 2004.

[47] F. Tommasino, M. Rovituso, S. Fabiano, S. Piffer, C. Manea, S. Lorentini, S. Lanzone, Z. Wang, M. Pasini, W. Burger, *et al.*, "Proton beam characterization in the experimental room of the trento proton therapy facility," *Nuclear Instruments and Methods in Physics Research A*, vol. 869, pp. 15–20, 2017.

[48] S. Rossi, "The status of cnao," *The European Physical Journal Plus*, vol. 126, no. 8, pp. 1–39, 2011.

[49] Y. Han, "Current status of proton therapy techniques for lung cancer," *Radiation oncology journal*, vol. 37, no. 4, p. 232, 2019.

[50] *IAEA-DIRAC: International Atomic Energy Agency - Directory of Radiotherapy Centres*, [accessed 18-December-2021]. [Online]. Available: https://dirac.iaea.org/.

[51] M. L. Yap, E. Zubizarreta, F. Bray, J. Ferlay, and M. Barton, "Global access to radiotherapy services: Have we made progress during the past decade?" *Journal of global oncology*, vol. 2, no. 4, pp. 207–215, 2016.

[52] M. Goitein and M. Jermann, "The relative costs of proton and x-ray radiation therapy," *Clinical Oncology*, vol. 15, no. 1, S37–S50, 2003.

[53] G. Vilches-Freixas, M. Unipan, I. Rinaldi, J. Martens, E. Roijen, I. P. Almeida, E. Decabooter, and G. Bosmans, "Beam commissioning of the first compact proton therapy system with spot scanning and dynamic field collimation," *The British journal of radiology*, vol. 93, no. 1107, p. 20 190 598, 2020.

[54] B. Schaffner and E. Pedroni, "The precision of proton range calculations in proton radiotherapy treatment planning: Experimental verification of the relation between ct-hu and proton stopping power," *Physics in Medicine & Biology*, vol. 43, no. 6, p. 1579, 1998.

[55] P. Mah, T. Reeves, and W. McDavid, "Deriving hounsfield units using grey levels in cone beam computed tomography," *Dentomaxillofacial Radiology*, vol. 39, no. 6, pp. 323–335, 2010.

[56] U. Schneider, E. Pedroni, and A. Lomax, "The calibration of ct hounsfield units for radiotherapy treatment planning," *Physics in Medicine & Biology*, vol. 41, no. 1, p. 111, 1996.

[57] M. Yang, X. R. Zhu, P. C. Park, U. Titt, R. Mohan, G. Virshup, J. E. Clayton, and L. Dong, "Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration," *Physics in Medicine & Biology*, vol. 57, no. 13, p. 4095, 2012.

[58] H.-W. Sadrozinski, V. Bashkirov, B. Keeney, L. R. Johnson, S. G. Peggs, G. Ross, T. Satogata, R. W. Schulte, A. Seiden, K. Shanazi, *et al.*, "Toward proton computed tomography," *IEEE Transactions on Nuclear Science*, vol. 51, no. 1, pp. 3–9, 2004.

[59] T. R. Johnson, "Dual-energy ct: General principles," *American Journal of Roentgenology*, vol. 199, no. 5_supplement, S3–S8, 2012.

[60] N. Hudobivnik, F. Schwarz, T. Johnson, L. Agolli, G. Dedes, T. Tessonnier, F. Verhaegen, C. Thieke, C. Belka, W. H. Sommer, *et al.*, "Comparison of proton therapy treatment planning for head tumors with a pencil beam algorithm on dual and single energy ct images," *Medical physics*, vol. 43, no. 1, pp. 495–504, 2016.

[61] E. Bär, A. Lalonde, G. Royle, H.-M. Lu, and H. Bouchard, "The potential of dual-energy ct to reduce proton beam range uncertainties," *Medical physics*, vol. 44, no. 6, pp. 2332–2344, 2017.

[62] U. Schneider, J. Besserer, P. Pemler, M. Dellert, M. Moosburger, E. Pedroni, and B. Kaser-Hotz, "First proton radiography of an animal patient," *Medical physics*, vol. 31, no. 5, pp. 1046–1051, 2004.

[63] K. M. Hanson, "Proton computed tomography," *IEEE Transactions on Nuclear Science*, vol. 26, no. 1, pp. 1635–1640, 1979.

[64] C. Civinini, D. Bonanno, M. Brianzi, M. Carpinelli, G. Cirrone, G. Cuttone, D. L. Presti, G. Maccioni, S. Pallotta, N. Randazzo, *et al.*, "Proton computed tomography: Iterative image reconstruction and dose evaluation," *Journal of Instrumentation*, vol. 12, no. 01, p. C01034, 2017.

[65] V. Giacometti, "Modelling and improvement of proton computed tomography," 2016.

[66] R. P. Johnson, V. Bashkirov, V. Giacometti, R. F. Hurley, P. Piersimoni, T. E. Plautz, H. F.-W. Sadrozinski, K. Schubert, R. Schulte, B. Schultze, *et al.*, "A fast experimental scanner for proton ct: Technical performance and first experience with phantom scans," *IEEE transactions on nuclear science*, vol. 63, no. 1, pp. 52–60, 2015.

[67] M. Scaringella, M. Brianzi, M. Bruzzi, M. Bucciolini, M. Carpinelli, G. A. P. Cirrone, C. Civinini, G. Cuttone, D. L. Presti, S. Pallotta, *et al.*, "The prima (proton imaging) collaboration: Development of a proton computed tomography apparatus," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 730, pp. 178–183, 2013.

[68] M. Esposito, C. Waltham, J. T. Taylor, S. Manger, B. Phoenix, T. Price, G. Poludniowski, S. Green, P. M. Evans, P. P. Allport, *et al.*, "Pravda: The first solid-state system for proton computed tomography," *Physica Medica*, vol. 55, pp. 149–154, 2018.

[69] F. Ulrich-Pur, T. Bergauer, A. Burker, A. Hirtl, C. Irmler, S. Kaser, and F. Pitters, "A proton computed tomography demonstrator for stopping power measurements," *arXiv preprint arXiv:2106.12890*, 2021.

[70] G. Poludniowski, N. Allinson, and P. Evans, "Proton radiography and tomography with application to proton therapy," *The British journal of radiology*, vol. 88, no. 1053, p. 20 150 134, 2015.

[71] R. Schulte, S. Penfold, J. Tafas, and K. Schubert, "A maximum likelihood proton path formalism for application in proton computed tomography," *Medical physics*, vol. 35, no. 11, pp. 4849–4856, 2008.

[72] B. Kevles, *Naked to the bone: Medical imaging in the twentieth century*. Rutgers University Press, 1997.

[73] F. Bistolfi, "Alessandro vallebona 1899-1987. ricordo di un grande radiologo e del suo contributo allo sviluppo delle scienze radiologiche," *Fisica in Medicina*, vol. 2, pp. 115–23, 2005.

[74] *Alessandro vallebona — Wikipedia, the free encyclopedia*, [accessed 12-December-2021]. [Online]. Available: https://it.wikipedia.org/wiki/Alessandro_Vallebona.

[75] E. C. Beckmann, "Ct scanning the early days," *The British journal of radiology*, vol. 79, no. 937, pp. 5–8, 2006.

[76] W. H. Oldendorf, "The quest for an image of brain: A brief historical and technical review of brain imaging techniques," *Neurology*, vol. 28, no. 6, pp. 517–517, 1978.

[77] S. Zannos, *Godfrey Hounsfield and the Invention of CAT Scans*. Mitchell Lane Pub., 2003.

[78] A. M. Cormack, "Representation of a function by its line integrals, with some radiological applications. ii," *Journal of Applied Physics*, vol. 35, no. 10, pp. 2908–2913, 1964.

[79] G. Di Chiro and R. A. Brooks, "The 1979 nobel prize in physiology or medicine," *Science*, vol. 206, no. 4422, pp. 1060–1062, 1979.

[80] P. N. T. Wells, "Sir godfrey newbold hounsfield kt cbe. 28 august 1919–12 august 2004: Elected frs 1975," *Biographical Memoirs of Fellows of the Royal Society*, no. 51, pp. 221–235, 2005.

[81] Cormack, Allan Macleod, "Representation of a function by its line integrals, with some radiological applications," *Journal of applied physics*, vol. 34, no. 9, pp. 2722–2727, 1963.

[82] A. Koehler, "Proton radiography," *Science*, vol. 160, no. 3825, pp. 303–304, 1968.

[83] M. Goitein, "Three-dimensional density reconstruction from a series of two-dimensional projections," *Nuclear Instruments and Methods*, vol. 101, no. 3, pp. 509–518, 1972.

[84] A. Cormack and A. Koehler, "Quantitative proton tomography: Preliminary experiments," *Physics in Medicine & Biology*, vol. 21, no. 4, p. 560, 1976.

[85] K. Hanson, J. Bradbury, T. Cannon, R. Hutson, D. Laubacher, R. Macek, M. Paciotti, and C. Taylor, "Application of protons to computer tomography," Los Alamos Scientific Lab., N. Mex.(USA), Tech. Rep., 1977.

[86] K. Hanson, J. Bradbury, T. Cannon, R. Hutson, D. Laubacher, R. Macek, M. Paciotti, and C. Taylor, "Computed tomography using proton energy loss," *Physics in Medicine & Biology*, vol. 26, no. 6, p. 965, 1981.

[87] K. Hanson, J. Bradbury, R. Koeppe, R. Macek, D. Machen, R. Morgado, M. Paciotti, S. Sandford, and V. Steward, "Proton computed tomography of human specimens," *Physics in Medicine & Biology*, vol. 27, no. 1, p. 25, 1982.

[88] U. Schneider and E. Pedroni, "Proton radiography as a tool for quality control in proton therapy," *Medical physics*, vol. 22, no. 4, pp. 353–363, 1995.

[89] P. Zygmanski, K. P. Gall, M. S. Rabin, and S. J. Rosenthal, "The measurement of proton stopping power using proton-cone-beam computed tomography," *Physics in Medicine & Biology*, vol. 45, no. 2, p. 511, 2000.

[90] V. Bashkirov, R. Schulte, G. Coutrakon, B. Erdelyi, K. Wong, H. Sadrozinski, S. Penfold, A. Rosenfeld, S. McAllister, and K. Schubert, "Development of proton computed tomography for applications in proton therapy," in *AIP Conference Proceedings*, American Institute of Physics, vol. 1099, 2009, pp. 460–463.

[91] J. Bregeon, "Design and performance of the silicon strip tracker of the fermi large area telescope," *Journal of Instrumentation*, vol. 6, no. 12, p. C12043, 2011.

[92] G. A. P. Cirrone, G. Candiano, G. Cuttone, S. L. Nigro, D. L. Presti, N. Randazzo, V. Sipala, M. Russo, S. Aiello, M. Bruzzi, *et al.*, "The italian project for a proton imaging device," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 576, no. 1, pp. 194–197, 2007.

[93] V. Sipala, N. Randazzo, D. L. Presti, C. Stancampiano, M. Bruzzi, M. Bucciolini, S. Pallotta, M. Scaringella, C. Talamonti, M. Tesi, *et al.*, "Upgrade of the proton computed tomography system of the prima project," *PoS (RD11)*, vol. 13, 2011.

[94] M. Bruzzi, M. Brianzi, M. Bucciolini, M. Carpinelli, G. Cirrone, C. Civinini, G. Cuttone, D. L. Presti, S. Pallotta, C. Pugliatti, *et al.*, "The prima (proton imaging) collaboration: Status of the development of a proton computed tomography scanner," in *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, IEEE, pp. 4314–4317.

[95] C. Talamonti, M. Bruzzi, M. Bucciolini, M. Carpinelli, G. Cirrone, C. Civinini, G. Cuttone, D. L. Presti, S. Pallotta, N. Randazzo, *et al.*, "Prima proton imaging for clinical application," in *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, IEEE, 2012, pp. 2218–2221.

[96] C. Civinini, M. Scaringella, M. Brianzi, M. Intravaia, N. Randazzo, V. Sipala, M. Rovituso, F. Tommasino, M. Schwarz, and M. Bruzzi, "Relative stopping power measurements and prosthesis artifacts reduction in proton ct," *Physics in Medicine & Biology*, vol. 65, no. 22, p. 225 012, 2020.

[97] M. Esposito, T. Anaxagoras, P. Evans, S. Green, S. Manolopoulos, J. Nieto-Camero, D. Parker, G. Poludniowski, T. Price, C. Waltham, *et al.*, "Cmos active pixel sensors as energy-range detectors for proton computed tomography," *Journal of Instrumentation*, vol. 10, no. 06, p. C06001, 2015.

[98] T. Price, M. Esposito, G. Poludniowski, J. Taylor, C. Waltham, D. Parker, S. Green, S. Manolopoulos, N. Allinson, T. Anaxagoras, *et al.*, "Expected proton signal sizes in the pravda range telescope for proton computed tomography," *Journal of Instrumentation*, vol. 10, no. 05, P05013, 2015.

[99] U. Schneider, E. Pedroni, M. Hartmann, J. Besserer, and T. Lomax, "Spatial resolution of proton tomography: Methods, initial phase space and object thickness," *Zeitschrift für Medizinische Physik*, vol. 22, no. 2, pp. 100–108, 2012.

[100] T. E. Plautz, V. Bashkirov, V. Giacometti, R. Hurley, R. Johnson, P. Piersimoni, H.-W. Sadrozinski, R. W. Schulte, and A. Zatserklyaniy, "An evaluation of spatial resolution of a prototype proton ct scanner," *Medical physics*, vol. 43, no. 12, pp. 6291–6300, 2016.

[101] J. Liang, T. Li, R. Schulte, T. Satogata, D. Williams, and H. Sadrozinski, *Cancer imaging: Instrumentation and application vol. 2*, 2008.

[102] M. Hayat, *Cancer imaging: instrumentation and applications*. Academic Press, 2007, vol. 2.

[103] European Research Council, ERC Consolidator Grants 2014 results, https://erc.europa.eu/sites/default/files/document/file/erc_2014_cog_full_results_by_domain.pdf.

[104] S. Mattiazzo, D. Bisello, P. Giubilato, D. Pantano, N. Pozzobon, W. Snoeys, and J. Wyss, "Advanced proton imaging in computed tomography," *Radiation protection dosimetry*, vol. 166, no. 1-4, pp. 388–392, 2015.

[105] P. Giubilato, W. Snoeys, S. Mattiazzo, N. Pozzobon, D. Pantano, D. Bisello, T. Kugathasan, J. Rousset, H. Mugnier, and C. M. Tobon, "Impact: Innovative pct scanner," in *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2015, pp. 1–2.

[106] The Phantom Laboratory, NY, USA, https://www.phantomlab.com/catphan-600.

[107] Thorlabs, Inc., https://www.thorlabs.com/.

[108] Saint-Gobain Crystals BC420 Data Sheet, https://www.crystals.saint-gobain.com/radiation-detection-scintillators/plastic-scintillators/fast-timing-bc-418-bc-420-bc-422-bc-422q.

[109] Saint-Gobain Crystals BC408 Data Sheet, https://www.crystals.saint-gobain.com/radiation-detection-scintillators/plastic-scintillators/bc-400-bc-404-bc-408-bc-412-bc-416.

[110] Eljen Technology, EJ-200 Datasheet, https://eljentechnology.com/products/plastic-scintillators/ej-200-ej-204-ej-208-ej-212.

[111] *Hamamatsu sipm data sheet*, hamamatsu.com.cn/UserFiles/DownFile/Product/s12572-025_etc_kapd1043e03.pdf.

[112] F. Baruffaldi, "Studies for a proton tomography scanner," 2017.

[113] C. Bonini, "Sviluppo di un calorimetro acromatico per protoni in ambito medico," 2018.

[114] Linear Technology (Analog Devices), "LTC6754: High Speed Rail-to-Rail Input Comparator with LVDS Compatible Outputs," https://www.analog.com/media/en/technical-documentation/data-sheets/6754f.pdf.

[115] Analog Devices, "AD5390/AD5391/AD5392: 8-/16-Channel, 3 V/5 V Serial Input, Single-Supply," https://www.analog.com/media/en/technical-documentation/data-sheets/AD5390_5391_5392.pdf.

[116] J. Irazabel and S. Blozis, "Philips semiconductors,"i2c-manual," application note, ref," *AN10216-0, March*, vol. 24, 2003.

[117] V. Bashkirov, R. W. Schulte, R. Hurley, R. Johnson, H.-W. Sadrozinski, A. Zatserklyaniy, T. Plautz, and V. Giacometti, "Novel scintillation detector design and performance for proton radiography and computed tomography," *Medical physics*, vol. 43, no. 2, pp. 664–674, 2016.

[118] S. Kotz and J. R. Van Dorp, *Beyond beta: other continuous families of distributions with bounded support and applications*. World Scientific, 2004.

[119] R. Hesse, "Triangle distribution: Mathematica link for excel," *Decision Line*, vol. 31, no. 3, pp. 12–14, 2000.

[120] M. Mager, A. collaboration, *et al.*, "Alpide, the monolithic active pixel sensor for the alice its upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 824, pp. 434–438, 2016.

[121] B. Abelev, A. Collaboration, *et al.*, "Upgrade of the alice experiment: Letter of intent," *Journal of Physics G: Nuclear and Particle Physics*, vol. 41, no. 8, p. 087 001, 2014.

[122] A. Di Mauro, A. Collaboration, *et al.*, "The new inner tracking system for the alice upgrade at the lhc," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 625–629, 2019.

[123] W. Dulinski, J.-D. Berst, A. Besson, G. Claus, C. Colledani, G. Deptuch, M. Deveaux, A. Gay, D. Grandjean, Y. Gornushkin, *et al.*, "Cmos monolithic active pixel sensors for minimum ionizing particle tracking using non-epitaxial silicon substrate," *IEEE Transactions on Nuclear Science*, vol. 51, no. 4, pp. 1613–1617, 2004.

[124] , "Tower Semiconductor,", https://towersemi.com/.

[125] P. Yang, G. Aglieri, C. Cavicchioli, P. Chalmet, N. Chanlek, A. Collu, C. Gao, H. Hillemanns, G. Huang, A. Junique, *et al.*, "Maps development for the alice its upgrade," *Journal of Instrumentation*, vol. 10, no. 03, p. C03030, 2015.

[126] "Servizio di Officina Meccanica," *Dipartimento di Fisica e Astronomia "G.Galilei", Università degli Studi di Padova; INFN Sezione di Padova*, http://www.dfa.unipd.it/servizi/officine/officina-meccanica/, https://www.pd.infn.it/it/servizioofficinameccanica/.

[127] D. Strulab, G. Santin, D. Lazaro, V. Breton, and C. Morel, "Gate (geant4 application for tomographic emission): A pet/spect general-purpose simulation platform," *Nuclear Physics B-Proceedings Supplements*, vol. 125, pp. 75–79, 2003.

[128] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, *et al.*, "Geant4—a simulation toolkit," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[129] PSI DRS4 Evaluation Board, https://www.psi.ch/drs/evaluation-board.

[130] F. Baruffaldi, "Simulations and test beam studies of the impact calorimeter," *Il nuovo cimento C*, vol. 41, no. 3, pp. 1–10, 2018.

[131] S. Mattiazzo, F. Baruffaldi, D. Bisello, B. Di Ruzza, P. Giubilato, R. Iuppa, C. La Tessa, D. Pantano, N. Pozzobon, E. Ricci, *et al.*, "Impact: An innovative tracker and calorimeter for proton computed tomography," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 2, no. 4, pp. 345–352, 2018.

[132] N. Pozzobon, F. Baruffaldi, D. Bisello, C. Bonini, B. Di Ruzza, P. Giubilato, S. Mattiazzo, D. Pantano, L. Silvestrin, W. Snoeys, *et al.*, "Calorimeter prototyping for the impact project pct scanner," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 1–4, 2019.

[133] A. De Angelis and M. Pimenta, *Introduction to particle and astroparticle physics: multimessenger astronomy and its particle physics foundations*. Springer, 2018.

[134] A. M. Prokhorov *et al.*, "Great soviet encyclopedia," 1970.

[135] M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka, K. Agashe, G. Aielli, C. Amsler, *et al.*, "Review of particle physics," *Physical Review D*, vol. 98, no. 3, p. 030 001, 2018.

[136] P. Shukla and S. Sankrith, "Energy and angular distributions of atmospheric muons at the earth," *International Journal of Modern Physics A*, vol. 33, no. 30, p. 1 850 175, 2018.

[137] G. Bardin, J. Duclos, A. Vitale, E. Zavattini, A. Bertin, M. Piccinini, M. Capponi, J. Martino, and A. Magnon, "A new measurement of the positive muon lifetime," *Phys. Lett. B*, vol. 137, no. CERN-EP-84-09, pp. 135–140, 1984.

[138] D. E. Groom, N. V. Mokhov, and S. I. Striganov, "Muon stopping power and range tables 10 mev–100 tev," *Atomic Data and Nuclear Data Tables*, vol. 78, no. 2, pp. 183–356, 2001.

[139] J. Lindhard, M. Scharff, and H. Schiøtt, "Kgl. danske videnskab. selskab, mat," *Fys. Medd*, vol. 28, no. 8, p. l954, 1954.

[140] H. H. Andersen and J. F. Ziegler, "Hydrogen. stopping powers and ranges in all elements," 1977.

[141] K. A. Olive, K. Agashe, C. Amsler, M. Antonelli, J.-F. Arguin, D. M. Asner, H. Baer, H. R. Band, R. Barnett, T. Basaglia, *et al.*, "Review of particle physics," *Chinese physics C*, vol. 38, no. 9, p. 090 001, 2014.

[142] C. F. P. d. Silva, "Reflection distribution of the fluoropolymers for the xenon scintillation light," *Ph. D. Thesis*, 2010.

[143] V. Pozzobon, W. Levasseur, K.-V. Do, B. Palpant, and P. Perré, "Household aluminum foil matte and bright side reflectivity measurements: Application to a photobioreactor light concentrator design," *Biotechnology Reports*, vol. 25, e00399, 2020.

[144] S. Sartore, "Characterization with cosmic muons of plastic scintillators for the impact project calorimeter," 2019.

[145] Samtec, ERF 8, 0.80 mm Edge Rate® Rugged High Speed Socket, https://www.samtec.com/products/erf8.

[146] Trenz Electronic, "Trenz Electronic TE0713 Technical Reference Manual, October 10, 2019," https://shop.trenz-electronic.de/trenzdownloads/Trenz_Electronic/Modules_and_Module_Carriers/4x5/TE0713/REV01/Documents/TRM-TE0713-01.pdf.

[147] Texas Instruments, "CDCLVD1216: 2 to 16 Low Additive Jitter LVDS Buffer," https://www.ti.com/product/CDCLVD1216?keyMatch=CDCLVD1216RGZT&tisearch=search-everything&usecase=OPN.

[148] Future Technology Devices International Ltd. FT600Q-FT601Q IC Datasheet (USB 3.0 to FIFO Bridge), http://www.ftdichip.com/Support/Documents/DataSheets/ICs/DS_FT600Q-FT601Q%20IC%20Datasheet.pdf.

[149] Xilinx, Inc., "Vivado Design Suite User Guide. Release Notes, Installation, and Licensing. UG973 (v2020.2) February 3, 2021," https://www.xilinx.com/support/documentation/sw_manuals/xilinx2020_2/ug973-vivado-release-notes-install-license.pdf.

[150] Xilinx, Inc., "Embedded System Tools Reference Manual. UG1043 (v2019.2) October 30, 2019," https://www.xilinx.com/content/dam/xilinx/support/documentation/sw_manuals/xilinx2020_1/ug1043-embedded-system-tools.pdf.

[151] Xilinx, Inc., "MicroBlaze Processor Reference Guide Embedded Development Kit EDK UG081 (v14.7)," https://www.xilinx.com/support/documentation/sw_manuals/xilinx14_7/mb_ref_guide.pdf.

[152] Xilinx, Inc., "FIFO Generator v13.1. LogiCORE IP Product Guide. PG057 April 5, 2017," https://www.xilinx.com/support/documentation/ip_documentation/fifo_generator/v13_1/pg057-fifo-generator.pdf.

[153] Xilinx, Inc., "Vivado Design Suite Tutorial. Logic Simulation. UG937 (v2019.1) June 4, 2019," https://www.xilinx.com/support/documentation/sw_manuals/xilinx2019_1/ug937-vivado-design-suite-simulation-tutorial.pdf.

[154] Xilinx, Inc., "AXI Reference Guide. UG761 (v13.1) March 7, 2011," https://www.xilinx.com/support/documentation/ip_documentation/ug761_axi_reference_guide.pdf.

[155] Intel Corporation, "SoC System on Module," https://www.intel.com/content/www/us/en/programmable/products/soc/ecosystem/system-on-modules.html.

[156] Xilinx, Inc., "Xilinx Adaptive SoCs," https://www.xilinx.com/products/silicon-devices/soc.html.

[157] Xilinx, Inc., "Zynq-7000 SoC and 7 Series Devices Memory Interface Solutions. User Guide, UG586 December 5, 2018 (v4.2)," https://www.xilinx.com/support/documentation/ip_documentation/mig_7series/v4_2/ug586_7Series_MIS.pdf.

[158] "Tera Term," https://ttssh2.osdn.jp/index.html.en.

[159] Xilinx, Inc., "AXI GPIO v2.0. LogiCORE IP Product Guide. PG144 October 5, 2016," https://www.xilinx.com/support/documentation/ip_documentation/axi_gpio/v2_0/pg144-axi-gpio.pdf.

[160] U. Schneider and E. Pedroni, "Multiple coulomb scattering and spatial resolution in proton radiography," *Medical physics*, vol. 21, no. 11, pp. 1657–1663, 1994.

[161] L. Eyges, "Multiple scattering with energy loss," *Physical Review*, vol. 74, no. 10, p. 1534, 1948.

[162] D. Williams, "The most likely path of an energetic charged particle through a uniform medium," *Physics in Medicine & Biology*, vol. 49, no. 13, p. 2899, 2004.

[163] T. Li, Z. Liang, J. V. Singanallur, T. J. Satogata, D. C. Williams, and R. W. Schulte, "Reconstruction for proton computed tomography by tracing proton trajectories: A monte carlo study," *Medical physics*, vol. 33, no. 3, pp. 699–706, 2006.

[164] H. A. Bethe, "Moliere's theory of multiple scattering," *Physical review*, vol. 89, no. 6, p. 1256, 1953.

[165] G. R. Lynch and O. I. Dahl, "Approximations to multiple coulomb scattering," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 58, no. 1, pp. 6–10, 1991.

[166] B. Rossi and K. Greisen, "Cosmic-ray theory," *Reviews of Modern Physics*, vol. 13, no. 4, p. 240, 1941.

[167] S. Safai, T. Bortfeld, and M. Engelsman, "Comparison between the lateral penumbra of a collimated double-scattered beam and uncollimated scanning beam in proton radiotherapy," *Physics in Medicine & Biology*, vol. 53, no. 6, p. 1729, 2008.

[168] "LHC," http://www.hep.upenn.edu/johnda/Papers/vC/LHC.pdf.

[169] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. A. Khalek, A. A. Abdelalim, R. Aben, B. Abi, M. Abolins, O. AbouZeid, *et al.*, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012.

[170] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, *et al.*, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012.

[171] "LHC performance reaches new highs," https://home.cern/news/news/accelerators/lhc-performance-reaches-new-highs.

[172] L. Evans and P. Bryant, "Lhc machine," *Journal of instrumentation*, vol. 3, no. 08, S08001, 2008.

[173] "LHC," https://public-archive.web.cern.ch/en/research/AccelComplex-en.html.

[174] A. Collaboration *et al.*, "The alice experiment at the cern lhc," *Jinst*, vol. 3, no. 420, S08002, 2008.

[175] H. Specht, "Experimental aspects of heavy ion physics at lhc energies," 1990.

[176] F. Carminati, P. Foka, P. Giubellino, A. Morsch, G. Paic, J. Revol, K. Safarik, Y. Schutz, U. A. Wiedemann, A. Collaboration, *et al.*, "Alice: Physics performance report, volume i," *Journal of Physics G: Nuclear and Particle Physics*, vol. 30, no. 11, p. 1517, 2004.

[177] B. Alessandro, F. Antinori, J. Belikov, C. Blume, A. Dainese, P. Foka, P. Giubellino, B. Hippolyte, C. Kuhn, G. Martinez, *et al.*, "Alice: Physics performance report, volume ii," *Journal of Physics G: Nuclear and Particle Physics*, vol. 32, no. 10, p. 1295, 2006.

[178] I. B. Alonso, O. Brüning, P. Fessia, L. Rossi, L. Tavian, and M. Zerlauth, "High-luminosity large hadron collider (hl-lhc): Technical design report," *CERN Yellow Reports: Monographs, CERN, Geneva*, 2020.

[179] G. A. Rinella, A. Collaboration, *et al.*, "The alpide pixel sensor chip for the upgrade of the alice inner tracking system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 845, pp. 583–587, 2017.

[180] J. P. Iddon, "Commissioning of the new alice inner tracking system," *Journal of Instrumentation*, vol. 15, no. 08, p. C08009, 2020.

[181] A. collaboration *et al.*, "Alignment of the alice inner tracking system with cosmic-ray tracks," *Journal of Instrumentation*, vol. 5, no. 03, P03003, 2010.

[182] P. Buncic, M. Krzewicki, and P. Vande Vyvre, "Technical design report for the upgrade of the online-offline computing system," Tech. Rep., 2015.

[183] L. Pancheri, J. Olave, S. Panati, A. Rivetti, F. Cossio, M. Rolo, N. DeMaria, P. Giubilato, D. Pantano, and S. Mattiazzo, "A 110 nm cmos process for fully-depleted pixel sensors," *Journal of Instrumentation*, vol. 14, no. 06, p. C06016, 2019.

[184] LFoundry S.r.l, `http://www.lfoundry.com/en`.

[185] L. Pancheri, R. A. Giampaolo, A. Di Salvo, S. Mattiazzo, T. Corradino, P. Giubilato, R. Santoro, M. Caccia, G. Margutti, J. E. Olave, *et al.*, "Fully depleted maps in 110-nm cmos process with 100–300-μm active substrate," *IEEE Transactions on Electron Devices*, vol. 67, no. 6, pp. 2393–2399, 2020.

[186] S. Panati, J. Olave, A. Rivetti, L. Pancheri, F. Cossio, P. Giubilato, D. Pantano, S. Mattiazzo, M. D. R. Rolo, and N. Demaria, "Matisse: A versatile readout electronics for monolithic active pixel sensors characterization," in *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2017, pp. 1–4.

[187] L. Fernandes, F. Amaro, A. Antognini, J. Cardoso, C. Conde, O. Huot, P. E. Knowles, F. Kottmann, J. Lopes, L. Ludhova, *et al.*, "Characterization of large area avalanche photodiodes in x-ray and vuv-light detection," *Journal of Instrumentation*, vol. 2, no. 08, P08005, 2007.

[188] P. Giubilato and W. Snoeys, "Orthopix: A novel compressing architecture for pixel detectors," in *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, IEEE, 2012, pp. 1735–1741.

[189] N. Wermes, "Pixel detectors for tracking and their spin-off in imaging applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 541, no. 1-2, pp. 150–165, 2005.

[190] A. Peeters, J. P. Grutters, M. Pijls-Johannesma, S. Reimoser, D. De Ruysscher, J. L. Severens, M. A. Joore, and P. Lambin, "How costly is particle therapy? cost analysis of external beam radiotherapy with carbon-ions, protons and photons," *Radiotherapy and oncology*, vol. 95, no. 1, pp. 45–53, 2010.