UNIVERSITY OF CATANIA

Dept. of Physics and Astronomy

PhD in Physics

PhD Thesis

Marilena Bandieramonte

Muon Portal project: Tracks reconstruction, automated object recognition and visualization techniques for muon tomography data analysis

Supervisors

Prof. **Francesco Riggi**

University of Catania

Ing. **Ugo Becciani**

Catania Astrophysical Observatory - INAF

A.A. 2013-2014

Mostrai il mio capolavoro alle persone grandi, domandando se il disegno li spaventava. Ma mi risposero: " Spaventare? Perché mai, uno dovrebbe essere spaventato da un cappello?" Il mio disegno non era il disegno di un cappello. Era il disegno di un boa che digeriva un elefante. Affinché vedessero chiaramente che cos'era, disegnai l'interno del boa. Bisogna sempre spiegargliele le cose, ai grandi.

Antoine De Saint-Exupéry, "Il Piccolo Principe"

"O brothers, I said, who through a hundred thousand
dangers have reached the west,
to this so brief vigil
of our senses that remains to us,
choose not to deny experience,
following the sun, of the world that has no people.
Consider your origin:
you were not born to live like brutes,
but to pursue virtue and knowledge."

Dante Alighieri, "The Divine Comedy"

"O frati, dissi, che per cento milia
perigli siete giunti all'Occidente,
a questa tanto picciola vigilia
de' nostri sensi eh' del rimanente
non vogliate negar l'esperienza,
Diretro al sol, del mondo senza gente.
Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza."

Dante Alighieri, "La Divina Commedia"

Abstract

The present Ph.D. thesis is contextualized within the *Muon Portal* project, a project dedicated to the creation of a tomograph for the control and scanning of containers at the border in order to reveal smuggled fissile material by means of the cosmic muons scattering. This work aims to extend and consolidate the research in the field of muon tomography in the context of applied physics. The main purpose of the thesis is to investigate new techniques for reconstruction of muon tracks within the detector and new approaches to the analysis of data from muon tomography for the automatic objects recognition and the 3D visualization, thus making possible the realization of a tomography of the entire container. The research work was divided into different phases, described in this thesis document: from a preliminary speculative study of the state of the art on the tracking issue and on the tracks reconstruction algorithms, to the study on the *Muon Portal* detector performance in the case of particle tracking at low and high multiplicity. A substantial part of the work was devoted to the study of different image reconstruction techniques based on the POCA algorithm (*Point of Closest Approach*) and the iterative EM-LM algorithm (*Expectation-Maximization*). In addition, more advanced methods for the tracks reconstruction and visualization, such as data-mining techniques and clustering algorithms have been the subject of the research and development activity which has culminated in the development of an unsupervised multiphase clustering algorithm (modified-*Friends-of-Friends*) for the muon tomography data analysis.

Keywords: ***Muon Portal*** *project, muon tomography, particle tracking at low and high multiplicity, tracks reconstruction algorithms, 3D visualization algorithms, multivariate data-mining techniques, clustering algorithms, ROOT and GEANT4 simulations.*

Abstract

La presente tesi di dottorato si inquadra all'interno del progetto *Muon Portal*, un progetto dedicato alla realizzazione di un tomografo per il controllo e la scansione di container alle frontiere, alla ricerca di materiale fissile di contrabbando attraverso lo scattering dei muoni cosmici. Questo lavoro mira ad estendere e consolidare la ricerca nel campo della tomografia muonica nel contesto della fisica applicata. Lo scopo principale di questa tesi è di investigare nuove tecniche di ricostruzione di tracce dei muoni all'interno del rivelatore e nuovi approcci all'analisi di dati da tomografia muonica per il riconoscimento automatico di oggetti e la visualizzazione 3D, rendendo possibile la realizzazione di una tomografia dell'intero container. Il lavoro di ricerca è stato articolato in più fasi, descritte nel presente documento di tesi: da un preliminare approfondimento speculativo dello stato dell'arte relativo alle problematiche del tracking e agli algoritmi di ricostruzione delle tracce, allo studio sulle performance del *Muon Portal* detector nel caso di tracciamento di particelle in ambienti a bassa e alta molteplicità. Una parte sostanziale del lavoro è stata dedicata allo studio di diverse tecniche di ricostruzione delle immagini, basate sull'algoritmo POCA (*Point of Closest Approach*) e sull'algoritmo iterativo EM-LM (*Expectation-Maximization*). Inoltre, metodi più avanzati per la ricostruzione e la visualizzazione di tracce, quali tecniche di data-mining e algoritmi di clustering sono stati oggetto dell'attività di ricerca e sviluppo che è culminata nello sviluppo di un algoritmo di clustering non supervisionato multifase (modified-*Friends-of-Friends*) per l'analisi di dati da tomografia muonica.

Keyword: *Progetto **Muon Portal**, tomografia muonica, tracciamento di particelle a bassa e alta molteplicità, algoritmi di ricostruzione tracce, algoritmi di visualizzazione 3D, tecniche di analisi dati multivariata e data-mining, algoritmi di clustering, simulazioni con ROOT e GEANT4.*

Contents

Introduction

Motivation

Since the beginning of time man has questioned the meaning of his existence and tried to understand and decode the world around him, led by an infallible and innate weapon: the desire for knowledge. Aristotle was sure that all human beings by nature tend to the knowledge and that this desire is at the basis of the philosophy. The need to go beyond is rooted in the mankind, it is a factor intrinsic to its nature, and it has led to the numerous scientific discoveries that permeate our daily life. It is the thirst for knowledge, the desire to go further, that led man, for example, to exceed the limit set by the sea, learning to navigate, the limit imposed from the air, learning to fly, and going beyond up to the space, the limit imposed by the visible microscopic universe, going beyond the atom and the particles that compose it. First of all, scientific research is a dynamic activity that aims to expand the horizon of knowledge, to answer the questions that man poses, and build a legacy of increasing knowledge for the common good. It is an activity that aims to give new life to the system, and is the guarantor of the development of society. The research is the result of a collaborative and cooperative process. The researcher with their skills is the main actor of the research, but he lives within a system, a community that makes it fruitful. Research and innovation become unthinkable as isolated activities.

This PhD thesis aims to extend and consolidate the research in the field of muon tomography in the context of Applied Physics. It is developed within the context of the Muon Portal project, a research project which aims to build a large area muon detector ($18\,m^2$) for the inspection of TEU (Twenty-foot Equivalent Unit) containers by means of muon tomography, in order to counteract the illegal trafficking of fissile material. This project is an example of how collaboration and sharing of expertise and different know-how are the real challenges and the road to innovation. In fact, innovation can be reached by working at the border region between different skill areas in order to achieve a common goal. In the transfer from theoretical research to applied research, innovation is achieved.

The *Muon Portal* project addresses the challenges of homeland security technology innovation gaps. In the last few years concern has been expressed about the close links between terrorism and international organized crime, focusing in particular on the illegal traffic of nuclear, chemical, biological and other potentially lethal materials. The attention to and support by authorities in the fight against terrorism has become a matter of great importance. Terrorist use of radioactive material is a serious threat to health and safety of the population. This traffic requires a broader international control and improved detection services in order to prevent the movement of terrorists and the supply of weapons through effective border controls. The more worrying scenario is undoubtedly linked to the use of nuclear weapons. As an alternative to traditional detection methods, it has been long suggested to employ the scattering process of the secondary cosmic radiation at the sea level (mainly muons). This process strongly depends on the atomic number of the traversed material, hence is particularly sensitive to high-Z fissile elements (U, Pu) or their shielding. Muon tomography employs the scattering of secondary cosmic muons inside a material to reconstruct a 3D image as close as possible to the true localization of the objects inside the volume to be inspected. To reach a good precision in the reconstruction of the tomographic image, a good tracking muon detector is required. It should be able to reconstruct on a event-by-event basis the track of the muon before and after traversing the volume, even in presence of multiple hits generated by the structure itself or by the surrounding materials.

Among the project challenges, the study, design and development of proper algorithms to produce a tomographic image of the container volume, able to signal the presence of hidden high-Z materials, have to be taken into consideration. Imaging algorithms of good quality and able to produce results in a small CPU time are indeed an essential tool for the reconstruction of tomographic images. Relatively short computing times are required at the inspection site, i.e. few minutes at most, in order to follow the real container flux without significantly interfering with the usual harbour activities. Due to the natural flux of secondary cosmic muons at the sea level $\phi_\mu = 1m^{-2}s^{-1}$, enough muon tracks are expected to cross the large volume of a container (of the order of $50m^3$ for a 20-foot container) in a reasonable amount of time, thus producing enough collected events to reconstruct a tomographic image with good precision.

The original and substantial contribute of this thesis is related to the investigation of new techniques for the muons tracks reconstruction within the detector and of new approaches to the analysis of muon tomography data for the automatic objects recognition and the 3D visualization, thus making possible the realization of a tomography of the entire container. The main purposes of the thesis are to gain an understanding of the tracking issue and of the tracks reconstruction algorithms in the case of particle tracking at low and high multiplicity and to demonstrate how the application of more advanced methods for the tracks reconstruction and visualization, such as data-mining techniques and clustering algorithms can successfully elaborate muon tomography data, allowing to infer the implicit information in the data thus simplifying the complexity of big datasets.

Outline

This PhD thesis is structured as follows.

- In Chapter 1 the principles of Muon Tomography are outlined. A detailed description of Cosmic Rays, of their origin and composition and flux is carried out focusing in particular on the secondary component of cosmic radiation and on the most penetrating part: muons. The physics of muons and their interactions with matter are described together with the principles of muon tomography, starting from the multiple scattering effect. An overview on the state of the art, of the existing projects and future developments on muon tomography is also given.

- In Chapter 2 the attention is focused on the description of the *Muon Portal* project, of its basic idea and motivation to prevent the movement of terrorists and the supply of weapons through effective border controls. The project aims to develop a system able to detect radioactive materials inside containers, to counteract the illegal transfer, by means of muon tomography. A detection system based on the muon tomography is then described in all its packages: starting from the mechanical structure and the design of the apparatus, focusing on the modules design, to the design choices and the description of the construction and assembly phases. Also the electronic read out system is presented explaining how data acquisition is made from the electronic point of view. A description of all the software packages developed within the project, starting from the control and monitoring machine, going to the description of the reconstruction, elaboration and visualization package and to the software simulation activities is given.

- Chapter 3 shows a study with detailed simulations that have been carried out in order to understand the performance of the *Muon Portal* project detector in presence of low and high multiplicity events and to optimize the algorithms for the reconstruction of the tomographic images. An introduction to the Extensive Air Showers (EAS) phenomenon is given, together with a first approach to the track reconstruction problem with the Least Square method. In the final sections the procedure used to solve the high multiplicity tracks

reconstruction issue and simulation results on the performance of the Muon Portal project detector are described.

- In Chapter 4 the reconstruction and visualization algorithms for muon tomography imaging designed and implemented within the project are explained, starting from the POCA (*Point of Closest approach*), the EM-LM (*Expectation Maximization likelyhood*) and the 2pt-ACF (*Two-points Correlation Function*). The attention is then focused on the investigation of new approaches to the muon tomography data analysis. An introduction on the clustering analysis applied on muon tomography is presented with a literature review. Then follows the description of the FOF (*Friends-of-Friends*) algorithm and of the modified-*Friends-of-Friends* version. Finally some simulation results applied to the context of the muon tomography imaging and some conclusions are drawn.

- To conclude the thesis, in Chapter 5 the actual status of the *Muon Portal* project is presented together with the ongoing activities that include some predictions of completion and new perspectives about the usage of the detector for purposes more closely related to studies of the cosmic rays physics.

CHAPTER 1

Muon tomography: State of the Art

This chapter outlines the principles of Muon Tomography. In Section 1.1 a detailed description of Cosmic Rays, of their origin and composition and flux is carried out. Then in Section 1.2, the focus is on the secondary component of cosmic radiation and in particular on the most penetrating part: muons. The angular distribution, energy, flux, and their abundance at sea level are described in detail. Section 1.3 is dedicated to the physics of muons and their interactions with matter. The principles of muon tomography are described, starting from the multiple scattering effect. Finally, Section 1.4 is an overview on the state of the art, existing projects and future developments on muon tomography.

1.1 Cosmic Rays

Cosmic rays are extremely high-energy radiations of unknown extraterrestrial origin that continually strike the Earth. The discovery of cosmic rays was made by Victor Hess in 1912 [1]. Hess conducted a series of experiments to observe the production of ions in an electrometer. These measurements were carried out on a manned balloon at different altitudes. Hess noted that the ion production doubled as the balloon rose from 1.000 to 4.000 meters above the ground. This indicated the presence of some type of ionizing radiation that penetrated the atmosphere. Measuring the ionization rate at 5.300 meters, he discovered that it

had tripled compared to the sea level. Thus began research on cosmic rays, a very active field of research since the beginning and still widely studied. Many are, in fact, ongoing studies to measure the composition, flux and energy spectrum of cosmic rays. Thanks to the studies on cosmic rays, a world of particles beyond the boundaries of the atom has been discovered: the first particle of antimatter, the positron (anti-electron), for example, was discovered by C. Anderson in 1932 [2], the muon (Anderson and Neddermeyer [3], 1936), followed by the pion (Lattes, Occhialini, Powell and Muirhead, 1947 [4]), the kaon and many others strange particles (Rochester e Butler, 1947 [5]). Until the advent of high-energy particle accelerators in the early 1950s, this natural radiation provided the only way to study this growing family of particles. The physics of cosmic rays is still developing and still under investigation, and many are the open questions related to cosmic rays. The source of cosmic rays, for example, is not yet certain.

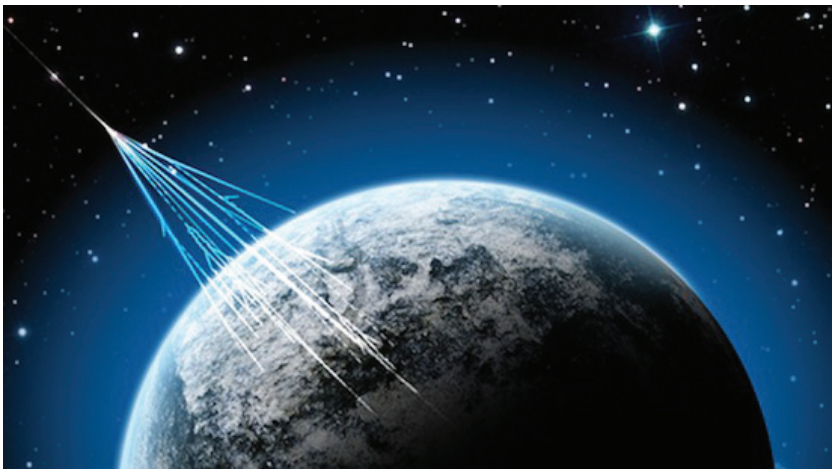


Figure 1.1: Visual reconstruction showing cosmic rays hitting Earth. Credit NSF / J. Hang.

In its active phases the Sun emits low energy cosmic rays, but these events are far too infrequent to explain the large amount of cosmic rays revealed. Even the other Sun-like stars are not sources of sufficient energy. The three basic hypotheses on the origin of cosmic rays are the “local”, the “galactic” and the “universal” hypothesis. These assumptions correspond to a first rudimentary location of the place of origin: the solar system, the galaxy and the entire universe. The second hypothesis is the one that collects the consent of the majority of researchers. Similar to the “shooting stars”, cosmic rays penetrate the atmosphere of our planet

colliding atoms in it. If our eyes had the sensitivity of a detector of Cerenkov light, we would continuously see the bluish glow in the night sky.

1.1.1 Historical Outline

When cosmic rays were discovered was the 1912. At that time, the only known particle was the electron. Niels Bohr had not yet presented his atomic theory and the quantum description of the microcosm was still far away. At that time the investigation on the origin of the radioactive phenomena, which were ubiquitous and unavoidable even with shields, was carried out. The Austrian Victor Hess began a systematic series of experiments reaching high altitudes with balloons, to identify the source of radioactivity. Detailed measurements allowed Hess to declare with certainty: "The results of these observations can be explained by assuming the presence of an extremely energetic radiation which penetrates the atmosphere and, interacts with it, causing ionization of the air as it is observed." For Hess they were gamma rays, the same observed in the natural radioactive decay; of the same idea was Robert Millikan, who in 1925 called them, in fact, cosmic radiation or cosmic rays. This analogy accompanied the scientists, who, in those years, were trying to understand the nature of the atom. We were still far from the construction of large accelerators, there were then only pioneering detectors consist of cloud chambers, where the particles left their trace in droplets passing through a supersaturated gas, and Geiger counters, in which the particles produce an electrical discharge. Stacking Geiger counters to form "telescopes", Walther Bothe and Werner Kolhorster in 1928 proved the particle nature of cosmic rays; their result was confirmed and deepened in Florence by Bruno Rossi, who developed the electronic circuits for systematic studies of cosmic rays. Until the 50s, cosmic rays were the only natural source of high-energy particles, that can produce new material species. They indirectly led to the first experimental observation of two fundamental discoveries in the field of particle physics: antimatter and the process of pion decay. In 1932 Carl Anderson observed positively charged particles, which left in the cloud chamber the same track of the electrons. His results were validated in 1933 by Patrick Blackett and Giuseppe Occhialini who recognized in them the antielectron or positron theoretically proposed by Paul

Dirac, observing the conversion of high energy photons into electron-positron pairs. Particle predicted in 1936 by Hideki Yukawa, the pion was experimentally observed in 1947 by Cecil Frank Powell, Occhialini and Cesar Lattes, using special photographic emulsions to record the pion production by cosmic rays and their subsequent decay into muons, which in turn decay into electrons (or positrons) and neutrinos (invisible). The observation of swarms of particles produced in cloud chambers suggested that the same cosmic rays, as they reach the ground, had to be the product of interactions and subsequent decays generated in the interaction with the upper atmosphere. The distinction between primary cosmic rays and secondary, tertiary and subsequent levels emerged quickly. It inevitably poses the question of the origin of the primary rays. The supported assumptions about the possible source and acceleration location of cosmic rays, were indicating initially the sun (Teller), then the galaxy (Fermi). Finally in 1956, Cocconi noticed that the most energetic part had extragalactic characteristics. It was necessary to wait for the beginning of the conquest of space for a deeper understanding of the origin of cosmic rays. Forty years of space missions are not many, but they suggested scientists to seek and verify basic properties of the universe in the study of celestial objects, such as blacks holes, neutron stars and extragalactic sources distant in space and time. Cosmic rays have the honor to be the harbingers of that information that the technology "hardly" will be able to achieve, a time machine, efficient and inexpensive, available to scientists and all those interested in reading the book of the universe, its laws and its history.

1.1.2 Composition, flux and energy spectrum of Cosmic Rays

Primary cosmic rays incident at the Earth's upper atmosphere include all of the stable charged particles and nuclei with lifetime of the order of 10^6 years that have been accelerated to enormous energies by astrophysical sources somewhere in our universe. Almost the total majority of cosmic rays is generated within our galaxy, and a low fraction of them (with energies less than a few GeV for protons) are from the atmosphere of the Sun (flares). They are charged particles as the acceleration mechanism is probably of electromagnetic origin and because their charge interacts with matter and produce the effects that we can easily observe here on Earth. They

cover range of energies from $10^9 eV (1GeV)$ up to $10^{20} eV (10^8 TeV)$.

We distinguish between "primary" and "secondary" cosmic rays. The primary are the set of charged particles accelerated at astrophysical sources which are distributed uniformly in space and which affect the most external layers of the atmosphere. They consist mainly of free protons (86-89)%, of He nuclei (9%), and for the rest they consist of more heavy nuclei, electrons, neutrinos and high-energy gamma.

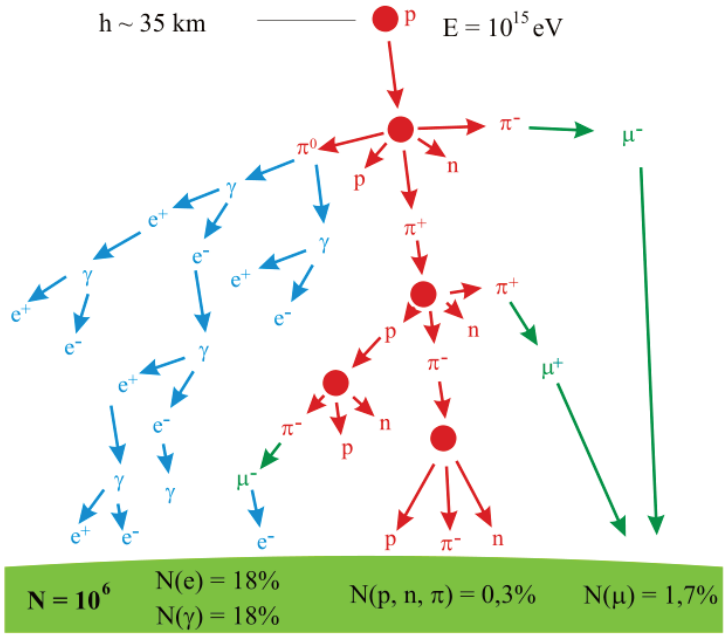


Figure 1.2: Schematization of the production of a swarm as a consequence of the interaction of a high-energy proton with an atomic nucleus in the upper atmosphere. In the picture are shown the various processes that give rise to swarms composed of different particles, such as the muonic, hadronic and electromagnetic components. (p) proton, (n) neutron, ($\pi\pm$) and (π_0) pions, ($\mu\pm$) muons, ($e\pm$) positron and electron, (ν) neutrino, (γ) gamma.

The composition of the cosmic rays varies in the path from the source to the Earth due to interactions with the interstellar medium, that give rise to the production of many other particles called secondary cosmic rays. Primary cosmic rays, in fact, will almost never reach the Earth surface, because they most likely will interact before with a nucleus of the air, usually tens of kms from the sea level.

Through these collisions, many new particles are created and many of the colliding nuclei evaporate.

In Fig. 1.2 is showed an example of swarm produced by the interaction of a high energy proton with a nucleon of Earth's atmosphere and are displayed separately the component of electrons and photons, the hadronic and the muon ones. Most of the new particles are π-mesons (pions). The neutral pions decay very quickly, usually in two γ-rays. Also charged pions decay, but after a longer time. Some of the pions may collide with another nucleus before they decay, originating a muon and a neutrino. Incoming nucleus can interact again, producing new particles. The γ-rays of neutral pions can also create new particles, an electron and a positron by pair-creation process. Electrons and positrons in turn can produce more γ-rays through bremsstrahlung process. The number of particles generated increases rapidly with the development of this, so called "shower" or "cascade" of particles downwards. Along the path and at each interaction, the particles lose energy, until, at a certain point they are no longer capable of generating other particles. After the point of maximum extension of the shower, the number of particles that stops becomes greater than those that are created and the size of the cascade decrease. Only a small fraction of particles, which depends on the energy and the type of the incident primary cosmic ray reaches the Earth surface. These amounts are subject to large variations. In fact, only the products of a low fraction of primary cosmic rays are able to reach up to the level of the ground. But since the flux of the primary is very high, the secondary flux is such that an area the size of a hand is hit by about one particle per second.

The secondary cosmic rays constitute about one-third of the natural radioactivity. When many thousands (sometimes millions or even billions) of particles arriving at ground level, perhaps on a mountain, this is called an extensive air shower (EAS). The majority of these particles will fall in the point that is located within a few hundred meters from the axis of motion of the original particle, which is defined as the axis of the shower. Sometimes the particles can also be found ams away. Along the axis, most of the particles can move nearly the speed of light inside a kind of disc that has a thickness of a few meters. This disc is slightly bent, and the particles away from the axis traveling at lower speeds. The thickness of

the disc also increases with the distance from the axis.

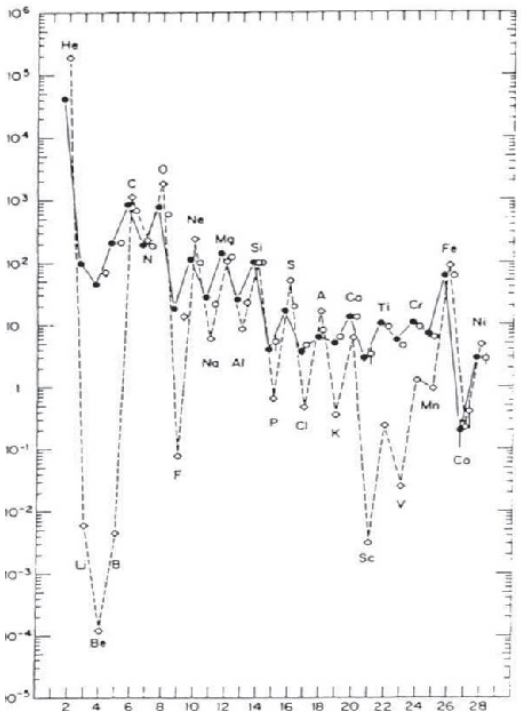


Figure 1.3: Cosmic Rays and Solar System relative abundances as function of atomic number Z.

The graph in Fig. 1.3 shows the relative abundances in cosmic rays and in the solar system, as a function of atomic number Z, measured at the Earth related to the abundance of silicon that is placed at 100%. It can be noticed an irregular pattern, common to both samples, which shows an odd-even effect: nuclei with even Z and/or A are more connected and therefore more abundant in nature, and in this way reduce the repulsion effect between charges of the same sign in the nucleus. Lighter nuclei, H and He are much less abundant in the cosmic rays. This can be due to the fact that fission mechanisms, which lead to the production of light atoms, are much less frequent in the Cosmic rays. Among the remarkable differences, the biggest one is an enormous enrichment in the cosmic ray abundances for some elements. In fact, elements such as Li, Be, B, Sc, V, Cr, Mn are much more abundant in cosmic rays compared to the matter of the solar system. Those elements would be virtually absent in the final stage of the process of stellar nucleosynthesis. In the cosmic rays are present as the results of the nuclear reactions of oxygen (Li,

Be, B) and iron content in cosmic rays (Sc, V, Cr, Mn) and the interstellar matter (ISM). This leads to a hypothesis about the residence time of cosmic rays in the galaxy: in order to explain the presence of these nuclei, cosmic rays have to pass through a quantity of matter equal to $X \simeq 5 g cm^{-2}$. Knowing the thickness X it can be determined the time, τ, spent by cosmic rays in the galaxy. Knowing that:

$$X = \rho_p c \tau \simeq 5 g cm^{-2} \tag{1.1.1}$$

it can be obtained:

$$\tau \simeq 3 \cdot 10^6 years \tag{1.1.2}$$

where $\rho_p \sim 1.67 \cdot 10^{-24} g cm^{-3}$ is the density if the interstellar gas, hypothesizing that it is essentially composed by protons. So the thickness of material traversed by cosmic rays in the Galaxy is:

$$l = \frac{X}{m_p \rho} = 1000\, kpc \tag{1.1.3}$$

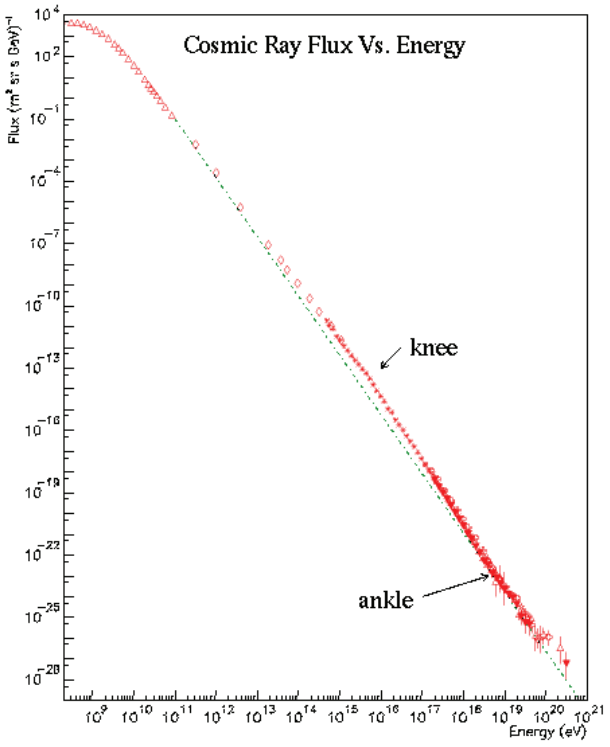


Figure 1.4: Differential flux of cosmic rays versus particles energy.

The thickness of the galactic disk $d = 0.2kpc$ that is a lot less than l, this implies that the cosmic rays in the Galaxy are diffused many times by the galactic magnetic field, $B_{gal} \simeq 3\mu G$ before getting out. Abundances of even Z elements with $30 \leq Z \leq 60$ are in reasonable agreement with solar system abundances.

Fig. 1.4 shows the differential flux of the cosmic rays, i.e. the number of incident particles per unit of energy, of time, of area and solid angle. The evolution of the flux in the energy function can be expressed by a power law of the type:

$$\frac{dN}{dE} \propto E^{-\alpha} \tag{1.1.4}$$

where the spectral index α takes different values at different energies:

$$\alpha = \begin{cases} 2.7, & \text{if } E < 3 \cdot 10^{15} eV \\ 3, & \text{if } 3 \cdot 10^{15} eV < E < 10^{18} eV \\ 2.7, & \text{if } E > 3 \cdot 10^{18} eV \end{cases} \tag{1.1.5}$$

Up to energies of order $10^{14} eV$ the cosmic flux is high enough to allow direct measurements of the primary cosmic in the upper atmosphere, using detectors located, for example, in balloons or satellites. But as the energy increases, the rate is lowered to levels too small to allow the use of these methodologies. Then indirect measurement techniques are used, conducted at the Earth surface, that by measuring the secondary products of cosmic radiation, try to extrapolate information about the primary who generated that particle shower. The first point of change of slope $E \simeq 3 \cdot 10^{15} eV$ is usually called "knee" and the second at $E \simeq 10^{18} eV$ is said "ankle". At the knee energy for instance, the flux is limited to 1 $particle/(m^2\, sr)$ per year, and only arrays with very large areas can detect an adequate number of events. In the region of low energy $E \leq 30 GeV$, the energy spectrum is not simply a power law but has a curvature in the log - log graph. The cosmic ray fluxes have a time dependence due to the modulations produced by the time variation of the solar wind intensity. The new measurements with magnetic spectrometers have significantly reduced the uncertainties of the flux below $100 GeV$ and measures taken at different times allow us to study the solar modulation extracting the interstellar flux. The trends observed can be explained assuming that the cosmic rays would be accelerated through the Fermi acceleration

mechanisms. This mechanism requires that the spectrum at the source of cosmic rays has a trend proportional to E^{-2}. The effect of confinement due to the galactic magnetic field, more efficient at low energies, distorts the spectrum of accelerated particles causing those we observed to follow a power law of the form $E^{-\gamma}$, with $\gamma \simeq 2.6 - 2.7$.

The charged particles in a magnetic field undergo the Lorentz force:

$$\overrightarrow{F_{Lorentz}} = q \cdot \overrightarrow{B} \hat{} \overrightarrow{v} \tag{1.1.6}$$

that is orthogonal to their velocity vector and then perform arcs or spiral shape. The radius of the trajectory is called Larmor radius or gyroradius and can be expressed by the relation:

$$R_{Larmor} = \frac{E}{Z\,|e|\,\overrightarrow{B}\,\beta c} \tag{1.1.7}$$

that for a relativistic proton $(z = 1, \beta \simeq 1)$ takes the form:

$$R\,[m] = \frac{E\,[GeV]}{0.3\,B\,[T]} \tag{1.1.8}$$

Knowing that the average magnetic field of the Galaxy B is equal to $B = 0.3\mu G$, with an energy of the order $E = 10^{16} eV$, we can obtain a Larmor radius of $\sim 4pc$ (where $1pc = 3.086 \cdot 10^{16} m$). Basing on the above considerations, we can assume that, at least up to the knee, so up to $(10^{15} - 10^{16}) eV$, the cosmic rays are predominantly of galactic origin: in fact, after the knee an increasing fraction of the cosmic rays escapes the galactic magnetic field, up to the ankle where the gyroradius is greater than the thickness of the galactic disk; over this energy the component of extragalactic origin is growing. Over $10^{20} eV$ almost all experiments (apart from AGASA) show a rapid decrease in the flux of cosmic rays, recently confirmed from AUGER experiment. This phenomenon has been predicted in the 60s by Greisen, Zatsepin and Kuzmin and is known as the GZK effect [6] [7]. This reduction of the flux of primary protons, known as the GZK cutoff is due to the fact that the cosmic rays of these energies can interact with photons of the microwave background at $2.73K$ to photoproduce the Δ^+resonance:

$$p + \gamma \rightarrow \Delta^+ \rightarrow \begin{cases} n + \pi^+ \\ p + \pi^- \end{cases} \tag{1.1.9}$$

The threshold energy for this reaction is very high; the preservation of the square four-momentum gives (considering $c = 1$):

$$E_p \geq \frac{2m_p m_\pi + m_\pi^2}{4E_{CMBR}} \approx 5.0 \cdot 10^{19} eV = 50 EeV \tag{1.1.10}$$

when the energy corresponding to the peak for E_{CMBR} is taken. The flux of cosmic rays with energies above this threshold is then attenuated along the way, and this limits the observable distance of the most powerful sources. At the same time, a few tens of Mpc can be observed, however, a greater flux of cosmic rays for energies immediately below the threshold, and this is due to the products of the reaction, see formula (1.1.9). It should be noted that among the secondary products, there are also high energy neutrinos.

1.1.3 Major experiments on cosmic rays

Measurements of the energy and arrival directions of the ultra-high energy primary cosmic rays by the techniques of ”density sampling” and ”fast timing” of extensive air showers were first carried out in 1954 by members of the Rossi Cosmic Ray Group at the Massachusetts Institute of Technology [8]. The experiment employed eleven scintillation detectors arranged within a circle 460 meters in diameter on the grounds of the Agassiz Station of the Harvard College Observatory. From that work, and from many other experiments carried out all over the world, the energy spectrum of the primary cosmic rays is now known to extend beyond $10^{20} eV$. High-energy gamma rays (>50 MeV photons) were finally discovered in the primary cosmic radiation by an MIT experiment carried on the OSO-3 satellite in 1967 [9]. Components of both galactic and extra-galactic origins were separately identified at intensities much less than 1% of the primary charged particles. Since then, numerous satellite gamma-ray observatories have mapped the gamma-ray sky. The most recent is the Fermi Observatory, which has produced a map showing a narrow band of gamma ray intensity produced in discrete and diffuse sources in our galaxy, and numerous point-like extra-galactic sources distributed over the celestial sphere. A huge air shower experiment called the Pierre Auger Observatory Project is currently operated at a site on the pampas of Argentina by an international consortium of physicists, led by James Cronin, 1980 Nobel

Prize in Physics of the University of Chicago and Alan Watson of the University of Leeds [10]. Their aim is to explore the properties and arrival directions of the very highest-energy primary cosmic rays. The results are expected to have important implications for particle physics and cosmology, due to a theoretical Greisen-Zatsepin-Kuzmin limit to the energies of cosmic rays from long distances (about 160 million light years) which occurs above $10^{20} eV$ because of interactions with the remnant photons from the big bang origin of the universe. In November 2007, the Auger Project team announced some preliminary results. These showed that the directions of origin of the 27 highest-energy events were strongly correlated with the locations of active galactic nuclei (AGNs). The results support the theory that at the centre of each AGN is a large black hole exerting a magnetic field strong enough to accelerate a bare proton to energies of $10^{20} eV$ and higher [11].

1.2 Secondary Cosmic Rays and Muons

The secondary radiation at sea level is constituted by two components (soft and hard) that have different behavior when crossing very dense materials (i.e. iron, lead). The soft component (approximately 30% of the secondary radiation), composed of electrons and photons and in small part by protons, kaons and nuclei, is capable of crossing only a few cm of absorber layer. The hard component (about 70%), composed of muons, do not interact strongly with matter, and can travel through the atmosphere and penetrate below ground. It is estimated that the average flux at sea level has an average energy of $3\,GeV$. The production of secondary muons from cosmic ray interactions is of particular interest to the topic of our research. Muons are created in a cosmic ray shower from the decay of hadronic secondary particles; they are mostly produced by charged pions and kaons. The interaction of nucleons with nitrogen and oxygen nuclei in the upper atmosphere produce pions (π) and kaons (K). Fig. 1.2 showed a schematic of the development of a typical cosmic-ray shower. The production of charged meson feeds the creation of muons that are the most prevalent decay product of charged pions and kaons:

$$\pi^+ / K^+ \longrightarrow \mu^+, \nu_\mu \tag{1.2.1}$$

$$\pi^- / K^- \longrightarrow \mu^-, \nu_\mu^- \tag{1.2.2}$$

Additional interactions in the upper atmosphere can take place between charged pions, with energies above 100 GeV, and nitrogen and oxygen. In this way they produce additional cascades of secondary particles before the pions and kaons decay. Instead charged pions of lower energies, are generally subjected to the weak force and decay before they can interact with the atmosphere. In fact, the average life of a charged pion is about 26 ns and this translates itself into a mean free path of 55 meters for energies around $1GeV$. These interactions are therefore unlikely since the pions cross only a small fraction of the atmosphere. The primary decay (99.9877% probability) for the charged pion is to produce a muon and a neutrino. In Fig. 1.5 is showed the Feynman diagram of the primary decay of charged π^+:

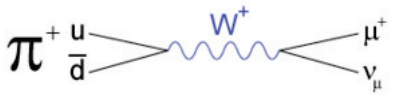


Figure 1.5: π^+ muon decay.

The lepton μ (muon) is an elementary particle with spin $1/2$, mass $105,66\ MeV/c^2$ about two hundred times the mass of the electron, and average lifetime τ equals to $2.197\mu s$. It exists in two charge states (μ^+/μ^-) and beyond the gravitational force it experiences two types of interaction: the electromagnetic and weak interactions. At the time of the production they are characterized by relativistic speeds so that their average life observed from Earth is greater than that observed in a system in which they are at rest, in agreement with the time dilation phenomenon theorized by the theory of relativity. Thanks to this phenomenon, a substantial fraction of muons produced in the upper atmosphere can reach the Earth surface before decaying, so it is thus possible to detect them at the ground. They reach the sea level, where it is noted that the μ^+ are about 20% more than the μ^- . Muons are, after the electrons, the lightest charged particles and therefore a muon can decay only into an electron or into other particles with null overall charge. Almost every time, they decay into an electron, intro an electron antineutrino and into a muon neutrino. The antimuon instead decay into a positron, into an electron neutrino and into an muon antineutrino:

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu, \quad \mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \tag{1.2.3}$$

In Fig. 1.6 is showed the flux of various components of cosmic ray showers as a function of atmospheric depth. Lower energy charged pions tend to decay to muons/neutrinos around 10 to 15 km above the earth. The figure shows vertical fluxes of cosmic rays in the atmosphere with energy greater than $1GeV$. The points show measurements of negative muons with energy greater than 1 GeV. The

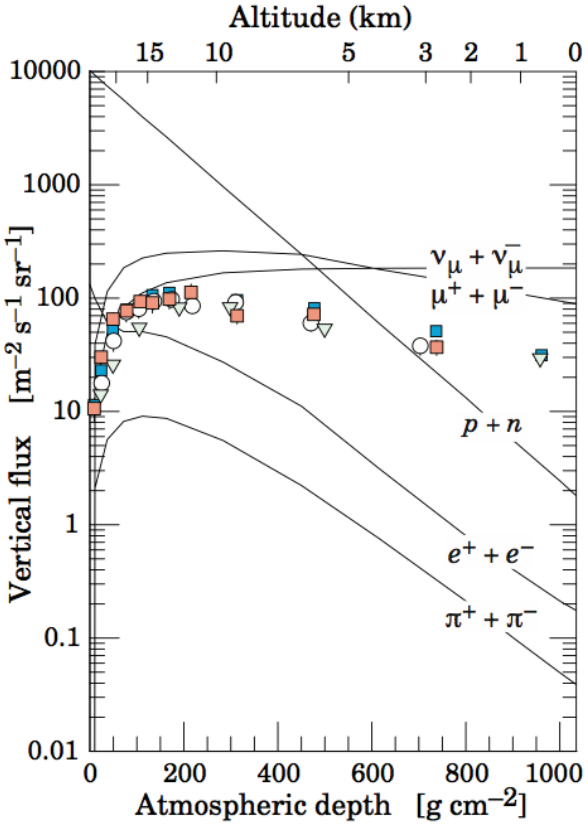


Figure 1.6: Flux of various components of cosmic ray showers as a function of atmospheric depth.

muons can penetrate deep into the atmosphere because they have an average life of 2200 ns. Those with energies above $5GeV$ can penetrate very deep levels in the underground. From the information shown in Fig. 1.6 it can be evinced that the muons are the product of the primary cosmic rays that can more easily be detected at the level of the Earth surface. The flux of muons at sea level is sufficiently high,

they hit the ground continuously with a rate of 1 muon per cm^2 per minute on average. The energy spectrum of muons at sea level at two different zenith angles can be observed in Fig. 1.7. In the figure we can notice a higher average muons energy at large zenith angles. This is explained by the fact that at high zenith angles low energy muons decay before they can reach the ground, and pions at high energies decay before they can interact with the atmosphere. The dark points

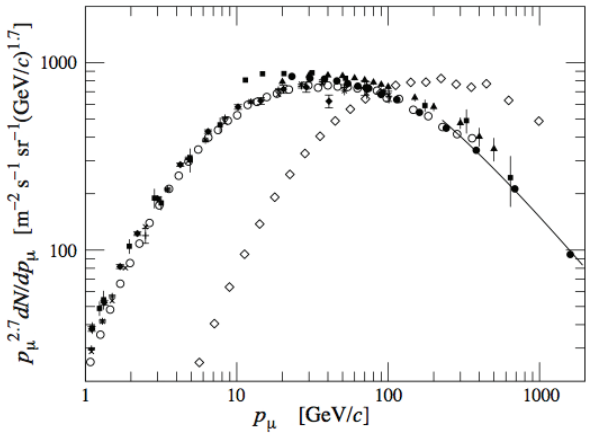


Figure 1.7: Muons Spectrum measured from sea level to $\theta = 0°$ and $\theta = 75°$, as measured by various experiments.

are for zenith angles of $0°$, while the hollow points are referred to $75°$ zenith angle. The overall angular distribution of muons at sea level is proportional to $cos^2\theta$, with θ being the zenith angle. In Fig. 1.7 it can be seen as a greater number of high energy muons per unit area is detected at large zenith angles. Instead, the total number of muons per unit surface is greater for low zenith angles, as expected from the distribution.

1.2.1 Angular distribution, energy intensity and muons abundance at sea level

Referring to [12], to describe the angular distribution and energy intensity of cosmic rays, both primary and secondary, we introduce the following quantities:

- the **directional intensity**, or just intensity, represents the number of particles, dN, per unit surface, dA, per unit of time, dt, per unit solid angle $d\Omega$, and is

written:

$$I(\theta,\phi) = \frac{dN}{dA\,dt\,d\Omega} \quad [cm^{-2}\,s^{-1}\,sr^{-1}] \tag{1.2.4}$$

The intensity depends on zenith angle θ, on the azimuthal angle ϕ and on the energy E. An important quantity in the study of cosmic rays is the vertical intensity $I_v \equiv I(\theta = 0)$.

- the **flux**, J_1, represents the number of particles that pass through, from the top downwards, an infinitesimal surface element, dA, per unit of time, dt , and is linked to the intensity by the relation:

$$J_1 = \int_{\theta \leq \frac{\pi}{2}} I(\theta,\phi)\,cos(\theta)\,d\Omega \quad [cm^{-2}\,s^{-1}] \tag{1.2.5}$$

- the **integrated intensity**, J_2, is obtained by integrating the intensity I in all the solid angle:

$$J_2 = \int I(\theta,\phi)\,d\Omega \quad [cm^{-2}\,s^{-1}] \tag{1.2.6}$$

According to the definition $J_2 \geq J_1$ always.

- the **differential energy spectrum**, also known as differential flux or differential intensity, is defined by the relation:

$$j(E) = \frac{dN(E)}{dA\,d\Omega\,dE\,dt} \quad [cm^{-2}\,s^{-1}\,sr^{-1}\,GeV^{-1}] \tag{1.2.7}$$

- Finally, we define the **integral spectrum in energy**, also called integral flux or integral intensity, $J(\geq E)$, as the integral of the differential spectrum, $j(E)$, extended to all particles with energy greater than E:

$$J(\geq E) = \int_E^{\infty} j(E)\,dE \quad [cm^{-2}\,s^{-1}\,sr^{-1}] \tag{1.2.8}$$

For energies above a few GeV the energy spectrum of cosmic rays can be expressed by a power law of the type:

$$J(\geq E) = C\,E^{-\gamma} \qquad j(E) = A\,E^{-(\gamma+1)} \tag{1.2.9}$$

where C and A are two constants. The exponent γ is said *integral spectral index*, while $\alpha = \gamma + 1$ is called the *differential spectral index*; in repeat with the spectrum of primary cosmic radiation, for energies between about $100GeV$ and the knee, the

spectral index is constant and has the value $\gamma = 2.7$. The dependence from the zenith angle, θ, for the secondary cosmic rays, can be expressed by the equation:

$$I(\theta) = I(0)\, cos^n(\theta) \tag{1.2.10}$$

where the exponent $n = n(X, E)$ is a function of the particle energy and the atmospheric thickness X. For what concern muons at sea level, in the literature there are many measures, but the values usually used for intensity, flux and integrated intensity for momenta greater than $350 MeV/c$ are the following:

$$I_v(> 0.35 GeV/c) = (0.94 \pm 0.05) \cdot 10^{-2} \;\; cm^{-2}\, s^{-1}\, sr^{-1} \tag{1.2.11}$$

$$J_1(> 0.35 GeV/c) = (1.44 \pm 0.09) \cdot 10^{-2} \;\; cm^{-2}\, s^{-1} \tag{1.2.12}$$

$$J_2(> 0.35 GeV/c) = (1.90 \pm 0.12) \cdot 10^{-2} \;\; cm^{-2}\, s^{-1} \tag{1.2.13}$$

In table 1.1 are reported values of the differential intensity and integral intensity for various momenta. The average muons energy at sea level is about $4 GeV$ and the spectrum is flat below $1 GeV$. For energies greater than $1 GeV$ the slope of the spectrum increases and the performance reflects that of the primary cosmic rays ($\gamma \simeq 2.7$); this increase is even more pronounced above $100 GeV$, because the pions reach the critical energy above which interact in the atmosphere before decaying.

The angular distribution of muons is proportional to $cos^n(\theta)$, with $n \simeq 2$, up to energies of the order of TeV, beyond which there is a trend proportional to sec (θ). At high-energies pions and kaons produced by the interaction of the primary in the atmosphere at angles $\theta > 0$ need longer trajectories and more time to go through a long column of air if compared to those produced with $\theta = 0$. Then the percentage of pions and kaons that decay rather than interact with the atmosphere increases with the angle θ and the energy; for example, around $10 GeV$ we have $I(65°)/I(0°) \simeq 0.5$ while at $100 GeV$ we have $I(65°)/I(0°) \simeq 2$. In Fig. 1.7 are shown two spectra of momentum for $\theta = 0°$ and $\theta = 75°$, which highlight the aspects discussed above. The azimuthal distribution is uniform except for low energies where it has to be considered the east-west asymmetry, an effect due to the presence of the geomagnetic field and the fact that particles that compose the

Pulse $[GeV/c]$	Differential Intensity $[\frac{cm^{-2}\,s^{-1}sr^{-1}}{(GeV/c)}]$	Integral Intensity $[cm^{-2}\,s^{-1}\,sr^{-1}]$
0.2	$3.73 \cdot 10^{-3}$	$9.94 \cdot 10^{-3}$
1.0	$2.79 \cdot 10^{-3}$	$7.22 \cdot 10^{-3}$
3.0	$1.06 \cdot 10^{-3}$	$3.72 \cdot 10^{-3}$
5.0	$4.97 \cdot 10^{-4}$	$2.26 \cdot 10^{-3}$
10	$1.33 \cdot 10^{-4}$	$9.42 \cdot 10^{-4}$
20	$2.70 \cdot 10^{-5}$	$3.21 \cdot 10^{-4}$
50	$2.36 \cdot 10^{-6}$	$5.93 \cdot 10^{-5}$
100	$3.04 \cdot 10^{-7}$	$1.38 \cdot 10^{-5}$

Table 1.1: Differential and integral intensity of vertical muons for various pulse [12].

primary radiation have predominantly positive charge. An analytical expression of the spectrum of muons, in good agreement with the experimental data for energies above a few GeV, is the following [13]:

$$\frac{dJ_\mu(E_\mu,\theta)}{dE_\mu} = AW_\mu(E_\mu + \Delta E_\mu)^{-\gamma}\left[\frac{r_\pi^{\gamma-1}\frac{B_\pi}{cos\theta}}{E_\mu + \Delta E_\mu + \frac{B_\pi}{cos\theta}} + 0.36b_r\frac{r_K^{\gamma-1}\frac{B_K}{cos\theta}}{E_\mu + \Delta E_\mu + \frac{B_K}{cos\theta}}\right] \tag{1.2.14}$$

where:

- W_μ is the muons *rate* of survival, which depends on the energy and the layer of atmosphere traversed;
- $r_\pi = 0.78$ is the ratio between the muon momentum and the momentum of the pion that generated it;
- $r_K = 0.52$ is the ratio between the muon momentum and the momentum of the kaon that generated it;
- $B_\pi = 90GeV$ is the energy of the pion corresponding to a decay length equal to the thickness of the layer of the atmosphere;
- $B_K = 442GeV$ is the energy of the kaon corresponding to a decay length equal to the thickness of the layer of the atmosphere;

- $b_r = 0.6355$ is the *branching ratio* of the decay $K^+ \rightarrow \mu^+ + \nu_\mu$;
- ΔE_μ is the energy lost by muons crossing the atmosphere.

The primary cosmic rays are composed almost exclusively of particles with a positive charge; during the interaction with the Earth atmosphere is produced a new generation of particles in which an excess of positive charges tends to decrease due to the charge symmetry of these particles production mechanisms. For this reason, the ratio between the number of positive and negative muons at sea level, N_μ^+ / N_μ^-, is on average in the range $1.25 - 1.30$ and depends on the energy and on the zenith and azimuth angles.

1.3 Principles of Muon Tomography

1.3.1 Muons interaction with matter

The energy loss of muons that propagate in the material takes place by means of the following mechanisms: ionization, atomic excitation, direct production of electron-positron pairs, bremsstrahlung and photo-nuclear interactions. In Fig. 1.8 is shown the energy loss per unit length (*stopping power*) of muons in copper as a function of $\beta_\gamma = p/Mc$, divided into various regions, in each of which a different approximation is used. The total energy loss can be written as:

$$-\frac{dE_\mu}{dX} = a(E_\mu) + b(E_\mu)E_\mu \tag{1.3.1}$$

where X is the thickness of material traversed, $a(E)$ is the energy loss by ionization and atomic excitation and the parameter $b(E)$ represents the fractional energy loss given by the sum of the contributions of the three radiative processes, i.e. bremsstrahlung, pair creation and photo-nuclear interaction, $b(E) = b_{br}(E) + b_{pp}(E) + b_{ni}(E)$.

The losses for ionization and excitation for a relativistic particle charge ze that moves in a material with atomic number Z and the mass number A, are described by the Bethe-Bloch formula, expressible in the following form:

$$-\frac{dE_\mu}{dX} = 4\,\pi\,N_A\,r_e^2\,m_e\,c^2\,z^2\frac{Z}{A}\frac{1}{\beta^2}\left[ln(\frac{2m_e c^2\gamma^2\beta^2}{I}) - \beta^2 - \frac{\delta}{2}\right] \tag{1.3.2}$$

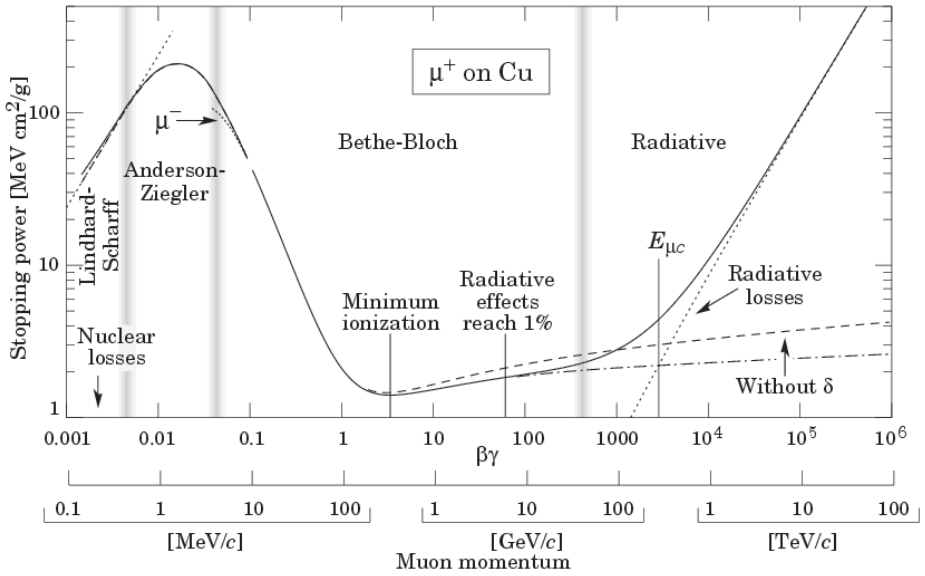


Figure 1.8: The graph represents the mass stopping power of muons in copper in a very wide range of momenta and shows which interaction processes are most important to the different energies.

where m_e is the electron mass, r_e the classical electron radius, N_A Avogadro's number, $\gamma = 1/\sqrt{1-\beta^2}$ the Lorentz factor, $\beta = v/c$, I the ionization potential and δ represents the correction due to the density Effect. The radiative losses become important at high energies but the expressions of the different terms are much more complicated; however for muons in a specific medium, the constants can be merged and we can obtain the practical expressions for the various types of interaction. Considering as medium the standard rock we obtain the following relations:

- **Losses due to ionization and excitation:** for energies $> 10^{11} eV$ has:

$$-\left(\frac{dE}{dx}\right)_{ion} = 1.84 + 0.076 ln \frac{E'_{max}}{m_\mu c^2} \ \ [MeV\, g^{-1}\, cm^2] \tag{1.3.3}$$

with:

$$E'_{max} \approx \frac{E_m^2 u}{E_m u + \left(\frac{m_\mu^2 c^2}{2 m_e}\right)} \tag{1.3.4}$$

where E_μ and m_μ are the energy and the mass of the muon and E'_{max} is the maximum transferable energy in an interaction.

At energies of the order of $10 GeV$ the previous relation is in agreement with experimental data within an uncertainty of about 5% [14]. For energies below a few GeV the relation (1.3.3) is not more valid because it is in the region of minimum ionization in which the stopping power has a minimum; at low energies still more losses increase rapidly, as shown in Fig. 1.8.

- **Bremsstrahlung losses:**

$$-(\frac{dE}{dx})_{br} \approx 1.77 * 10^{-6} E[MeV\, g^{-1}\, cm^2] \tag{1.3.5}$$

- **Losses for the direct production of pairs:**

$$-(\frac{dE}{dx})_{pp} \approx 2.40 * 10^{-6} E[MeV\, g^{-1}\, cm^2] \tag{1.3.6}$$

- **Losses due to photo-nuclear processes:**

$$-(\frac{dE}{dx})_{ni} \approx 0.48 * 10^{-6} E[MeV\, g^{-1}\, cm^2] \tag{1.3.7}$$

Summing these terms we obtain the following expression for the total energy loss:

$$-(\frac{dE}{dx})_{tot} = 1.84 + 4.65 * 10^{-6} E + 0.076\, ln\frac{E'_{max}}{m_\mu c^2}\ [MeV\, g^{-1}\, cm^2] \tag{1.3.8}$$

In Fig. 1.9 is shown the average stopping power of a muon as a function of the energy, in some different materials. As can be seen from the graph,with regard to the iron, the radiative contributions begin to be relevant for energies above a few tens of GeV and become more important for ionization losses around $300 GeV$; it can also be noticed how these energies depend on the absorber material. Furthermore from the graph it can be noticed how muons are particles that interact more with materials with high-Z, as can be the Uranium, especially at higher energies. For relativistic particles the energy loss in the material is equal to about $2.5\, MeV\, cm^2\, g^{-1}$ and is, in first approximation, independent of the type of particle and energy. This is true for particles with the same charge and if we are in the range of energies where radiative losses are negligible. In standard rock, in this approximation, there is an average energy loss of about $0.5\, TeV\, km^{-1}$.

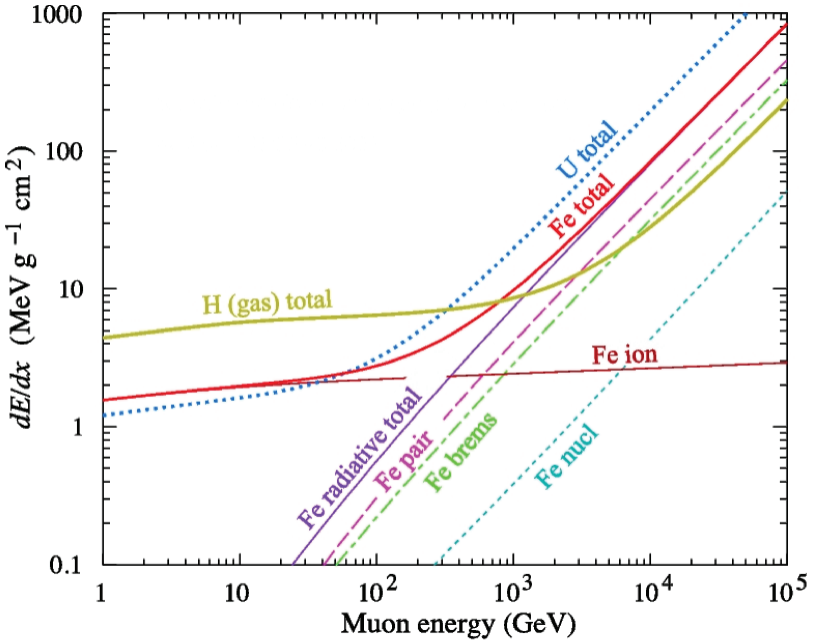


Figure 1.9: Trend of the average energy loss of muons in hydrogen, iron and uranium. In the case of iron, in addition to the total energy loss, are also shown the contributions of ionization and radiative processes separately.

1.3.2 The Multiple Coulomb Scattering Effect

A charged particle passing through a material undergoes a large number of deflections, most of which are due to Coulomb scattering with nuclei of material; this effect is then called multiple Coulomb scattering or simply multiple scattering. The angular distribution of the particle scattering, described by the theory of Moliere, is a Gaussian for small deflection angles, which tails get longer if we consider larger angles. In the approximation in which can be considered as a Gaussian, the distribution projected on a plane is the following:

$$\frac{1}{\sqrt{2\pi}\theta_0} e^{-\frac{\theta^2_{plane}}{2\theta^2_0}} d\theta_{plane} \tag{1.3.9}$$

where θ_0 is defined as:

$$\theta_0 = \theta^{rms}_{plane} = \frac{1}{\sqrt{2}} \theta^{rms}_{space} = \frac{13.6\, MeV}{\beta\, c\, p}\, z\, \sqrt{x/X_0}\, [1 + 0.038\, ln(x/X_0)] \tag{1.3.10}$$

where β, c, p and z are the speed, the momentum and the charge of the incident particle. The ratio x/X_0 defines the thickness of the material in unit of radiation length X_0. This approximation, obtained by fit with a Moliere distribution of particles with unitary charge and $\beta = 1$, describes the width θ_0 with an accuracy less than 11% within the limits $10^{-3} < x/X_0 < 100$ [15]. The radiation length is defined as the average distance after which an electron of high energy remains with $1/e$ times its initial energy. To give an estimate of the length of radiation in a material composed of a single element it might be used the following approximate formula:

$$X_0 = \frac{716.4\, g cm^{-2}\, A}{Z(Z+1)\, ln(287/\sqrt{Z})} \tag{1.3.11}$$

where A and Z are respectively the mass number and the atomic number of the element taken into consideration. To calculate X_0 for a compound, indicating with w_i the weight fraction of the i-th element and with X_i the corresponding radiation length, it can be used:

$$\frac{1}{X_0} = \sum_i \frac{w_i}{X_i} \tag{1.3.12}$$

The Earth has a chemical-physical composition variable that depends on many factors; as a first approximation it may be assumed that it is composed of the elements listed in table 1.2, with a density equal to $1.5\, g\, cm^{-3}$. Using the relations (1.3.10) and (1.3.11) and the data in table 1.2 we obtain: $X_0 \simeq 27.9\, g\, cm^{-2}$.

Element	**Weight fraction**
H	0.021
C	0.016
O	0.577
Al	0.05
Si	0.271
K	0.013
Ca	0.041
Fe	0.011

Table 1.2: The table shows the main elements that compose the soil with their abundances.

Then it can be calculated the average angle of scattering θ_0 for various energies

of the incident particle and for various distances from the point of incidence; the average displacement of the particle trajectory with respect to the direction of incidence projected in the plane orthogonal to that projected direction (to add image describing these quantities) is given by $y^{rms}_{plane} = \frac{1}{\sqrt{3}} x\,\theta_0$. This calculation is reported in Table 1.3 for $\beta \simeq 1$ and $z = 1$. Here to add some informations about Uranium and Lead.

Energy[GeV])	Distance [m]	$\theta_0(^\circ)$	y [cm]
	1	1	1
2	10	3.3	32.9
	20	4.7	95.8
	1	0.4	0.4
5	10	1.3	13.3
	20	1.9	38.1
	1	0.2	0.2
10	10	0.7	6.6
	20	0.9	19.2

Table 1.3: The table shows the values of θ_0 and y_{rms} for three different energies at a distance of 1, 10 and 20 m from the point of incidence in the material.

1.3.3 Muon tomography for imaging issues

The interest in the use of muons for imaging issues in the field of Earth Sciences, as said before, emerged soon after the discovery of these particles and their properties. Muons found in cosmic rays can go through hundreds of meters (the most energetics also through kilometers) of rock, undergoing to attenuations related to the quantity of matter traversed [16]. Initially, the muons imaging technique was born from the need to characterize from a geological point of view the structures in the underground labs that hosted particle detectors. Subsequently, when the devices used to carry out these measures became more portables, also measurements of geological structures such as mines, mountains and volcanoes were conducted, with the aim of monitoring natural phenomena such as earthquakes or eruptions.

The first idea to exploit the muon physical characteristics was, therefore, to use the information on the absorption of muons to measure the thickness of the material traversed by cosmic rays. The first application of cosmic rays to inspect large volumes was in 1955 when the thickness of a layer of rock overlying an underground tunnel was measured by E.P. George [17]. An application much more spectacular was conceived by the Nobel Prize for physics LW Alvarez [18] in 1970, that made a radiography of the pyramid of Khafre, in order to look for hidden chambers. Recently, the same technique was used to inspect the interior of volcanoes.

The operating principle is the same at the base of the X-ray radiography. The measured quantity is the flux attenuation of the muon that crosses the target that has to be studied; from this it can be derived the quantity of material encountered by muons along their trajectory. The interaction of the muons flux with the target depends both on the size of the target and on its density: with a small element and with low density only a few muons will interact, while a large and very dense body will absorb most of the striking muons. The imaging of the location of interest is realized as follows:

- Firstly, the muon flux is measured at different angles and locations over the target volume
- Then, differences in flux between particular angles and directions are used to produce a 3D image of the target volume using standard techniques of computer-assisted tomography, and reconstruction algorithms

In the muon tomography, one of the uncertainty is the assumption that the trajectory of the muon is constant, that is, that the direction given to the detector is a valid estimate of the direction from which the muon ultimately came. In reality, the muons ionize the crossed material, losing energy and also undergoing scattering. The probability of scattering decreases with increasing energy of muons. To minimize the localization errors due to the scattering deviation it might be thought of taking into consideration only the muons with an energy large enough in order to minimize the probability of leakage through the material considered. Muons lose approximately $2.2 MeV$ for every g/cm^2 of matter they travel through. Using this basic energy loss formula, it can be shown [19] that for a material with density

$\rho = 2.3 g/cm^3$, the cutoff muon energy should be around $2.5 GeV$. This insures that the muon spectrum is "hardened" at the detector. Alvarez accomplished this by surrounding his detector with blocks of iron. This way, lower energy muons were absorbed into the iron and not considered in the flux measurements.

From radiographies that use muons can be deduced the following quantity [20]:

$$\varrho(L) \equiv \int_L \rho(\xi)\, d\xi \tag{1.3.13}$$

where ϱ is said *opacity* (expressed in gcm^{-2}), ρ is the density and ξ represents the coordinate measured along the trajectory L of the muon. To determine the attenuation we compare the measured flux after crossing the geological structure of interest with the flux incident on it, which generally coincides with the flux of muons in the atmosphere at a given altitude. The measure of the flux into the atmosphere is an important parameter in order to obtain accurate images; although in the literature there are different parametrizations of muon flux, this is often measured in each new experiment. The main problem of these applications is the low flux of muons, which means high error bars on the measurement of opacity, due to poor statistics. For this reason, the measures have long duration, in order to acquire a sufficient statistic and be able to resolve an object with a given density. The flux of muons is modified by several environmental factors such as altitude, the geomagnetic cut-off, solar modulation and atmospheric conditions [21]. This research work is based on a new idea that has been developed by the *Muon Portal* project: the idea is to exploit the deflection of cosmic rays to make a radiography of a desired volume, in order to monitor its content. This novel approach comes from the need to analyze non-invasively, as the manual control requires too much time, sets of heterogeneous materials with the aim to identify specific types of materials. In particular, in our case, the final target is to create a container tomography for the detection of radioactive and high-Z materials such as *Uranium* and *Lead*. The first application that took advantage of muons as "materials inspectors" was based on the measurement of the absorption of cosmic rays to probe the interior volume inaccessible. Muon tomography is based on a specific property of muons: the *angular deviation* that their trajectory has undergone when crossing a material [22]. The trajectory of a charged particle within a material is the

result of the convolution of many small electrostatic deflections due to scattering of the particles by the atoms that constitute the material. The overall result of these diffusion processes depends strongly on the atomic number of the crossed material. Then by performing a measure of the deviation, is possible to obtain a reconstruction of the distribution of materials in three dimensions, achieving a result which can be assimilated from the point of view to a habitual medical TAC, that therefore, uses other scientific principles. This recent approach requires a lower number of muons and, therefore, a smaller waiting time in respect with the absorption method, but requires a more complex instrumentation because the direction, and not only the position, of each particle has to be measured. Two or more planes sensitive to muons are positioned around the volume to be analyzed, in order measure the position of muons in two orthogonal coordinates XY. The method will be described in details in the next section. The techniques of conventional radiography (for example, those with X-rays) make use of the phenomenon of the absorption of the radiation in the material, and therefore require to measure the flux of radiation transmitted (which is only a fraction of the incident one). They are limited by the capacity of penetration of the radiation X, which for energies of the order of MeV corresponds to about 2 cm of Pb. Greater thicknesses require higher energies, or very intense X-rays fluxes, dangerous in both cases. Tomography based on multiple scattering of cosmic muons represents an alternative to conventional radiographic techniques . One of the advantages of this technique is the ability to fully exploit the information provided by each particle (muon) that is diverted, with a reduction of measurement times and a significant increase of the statistical accuracy of the results. An added benefit, is not negligible, is related to the fact that the secondary cosmic radiation is a natural radiation that is constantly present in every place on Earth, and with which living beings coexist. A system based on muon tomography therefore does not introduce additional levels of radiation than those found in nature.

1.3.4 Description of the method

The concept of muon tomography is illustrated in Fig. 1.10. Two or more planes sensible to muons are positioned around the volume that has to be inspected.

Those detectors measure the position in which the muon crosses the planes in orthogonal coordinates. Muons that cross the volume are deviated from their original trajectory depending on the type of material of which the volume is composed. A discrete reconstruction of the volume is realized starting from data that are provided by different muons. Then, some different algorithms, that will be explained in the following chapters, are used in order to reconstruct the track of the particle and to calculate the scattering angle.

As explained above, after a series of interactions within the volume, even of small entity, the final effect is a scattering angle of the track from its original trajectory. The significant figure is the spread of scattering angles total considered for many muons, since the distribution depends on the length of radiation, X_0, which in turn depends on the material type, the atomic number (Z) and the density. An approximate expression for the width of this distribution θ_0, obtained by simplifying (1.3.10), is the following, where p is the muon momentum, x is the length of material traversed and X_0 is the radiation length.

$$\theta_0 \simeq \frac{13.6MeV}{p}\sqrt{\frac{x}{X_0}} \tag{1.3.14}$$

Typical radiation lengths and scattering angles are given in the table below. (aggiungere tabella, sono riportati in precedenza quelli dei materiali che compongono la terra, ma questa tabella va cambiata). It is clear that radiation length decreases for dense, high-Z material. It is this distinction which enables to determine the presence of dense materials such as uranium and plutonium. Thanks to position-sensitive detectors planes, the muon is tracked as it enters and leaves the inspection volume. The position of the muon is located as it passes through each of the detector planes, enabling a trigger, allowing to determine the incoming and outgoing muon trajectories.

Intersecting these two trajectories it is possible therefore to locate the interaction and to calculate the scattering angle. To avoid having “blind” zones within the volume and maximize the rate of muons measured, it is necessary that the detector is of a size substantially greater than the volume to be inspected. Therefore, the detectors for muon tomography would need to be not only position sensitive (with requirements of spatial resolution that can be sub mm), but also to be robust and

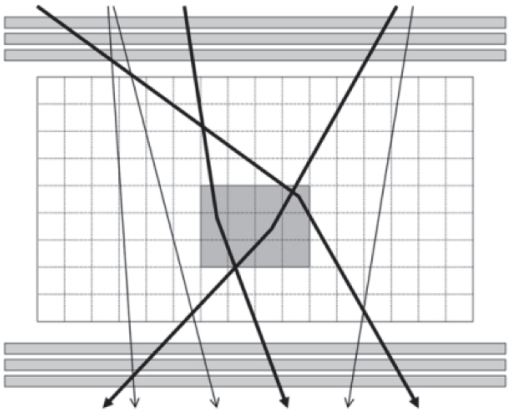


Figure 1.10: The figure shows the concept of muon tomography. The layers at the top are the planes that will detect the incoming muon, then there is the area to be imaged and, at the bottom, the ongoing muon detector.

scalable for large surfaces, all at reasonable costs. These requirements are very similar to those of the detectors used in high energy physics experiments, like the Large Hadron Collider at CERN, within which the skills in the construction of large trace detectors have been developed over the years. There are different technologies that can detect the passage of muons: these detectors include drift chambers, resistive plate chambers and segmented plastic scintillators. The intersection of the incoming and outgoing muon trajectories is defined as the point of maximum approach. This is because in three dimensions, it is unlikely that the vectors are coplanar and intersect in exactly one point. Then is it reasonable to choose the midpoint of the minimum segment joining the two lines. Different tracking and reconstruction techniques will be accurately described in the following.

1.4 Tracking detectors for muons: State of the Art and existing projects

1.4.1 Muon tomography for Earth Sciences purposes

Since its discovery, muon tomography has been used for different purposes in different scientific fields. One of the active fields, as mentioned before, is the geoscience and volcanology one. The interest on muon tomography for Earth

Sciences purposes, in fact, grew up soon after the discovery of cosmic rays and of the muon. There have been some attempts to exploit the immense penetrating power of cosmic ray muons to radiographically probe the internal structure of very large objects ranging from ancient pyramids to active volcanoes. The muons cross-section at those typical energies makes it a perfect probe since it is able to cross hundreds of meters of rock with an attenuation related to the amount of matter along its trajectory.

Volcanoes are more and more targeted by muon tomographic imaging since the alternative methods often reveal limited or difficult to work out in their environments. The *DIAPHANE* [23] is the first European project of tomography applied to volcanology. It started in 2008 with a collaboration between three French institutes: IPG Paris, IPN Lyon and Geosciences Rennes to promote muon tomography in the French Earth Science and Particle Physics communities.

In[24] the application of muon tomography to monitor and image the internal structure of volcanoes in the Lesser Antilles is discussed. Particular focus is directed towards the three volcanoes that fall under the responsibility of the Institut de Physique du Globe of Paris, namely La Montagne Pelee in Martinique, La Soufriere in Guadeloupe, and the Soufriere Hills in Montserrat. The detection matrices consist of scintillator bars arranged in two orthogonal series of X rows and Y columns. The number of bars depends on several parameters:

- the detection surface needed to measure a sufficient number of muon trajectories in a reasonable amount of time;
- the weight constraint discussed above;
- the 64 channels permitted by each photomultiplier (PM) and its associated electronic board;
- the desired angular resolution.

The matrices constructed to date are either 16×16 or 32×32 matrices respectively made with $1 \times 5 \times 80\,cm^3$ or $1 \times 5 \times 160\,cm^3$ bars of plastic scintillator (Fig. 1.11). The matrices then count either 1024 or 256 pixels of $5 \times 5\,cm^2$ in size.

An application related to the scanning of volcanoes can be found also in [21]. Here the authors present an alternative and complementary tomography

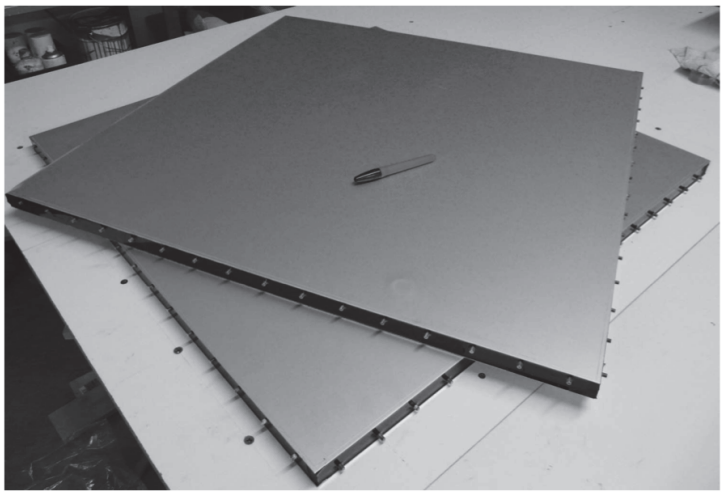


Figure 1.11: View of the two 16 × 16 matrices [24]. The optical plugs are visible on two sides of each matrix and are to be connected to the clear fibers guiding the light onto the channels of the photomultiplier. The thickness of the matrices is 27 mm and their individual total weight is 20 kg.

method based on the measurement of the cosmic muons flux attenuation through the geological structures. They use photo-active detectors in order to monitor, understand and better predict the behavior of volcanoes, with an obvious and crucial societal impact. Also in [25] the authors used cosmic rays muons for the measurement of the internal structure of the active volcano Mt. West Iwate. They use muon radiography as a radiographic probe that can provide useful information that complements the gravimetric data. They developed a detection system built with a set of two position-sensitive segmented detectors by the multiplicity analysis of a soft-component background after passing through an Fe plate between the detectors.

Similar projects are the MU-RAY project [26] and the [27]. The MU-RAY project aims at the construction of muon telescopes with angular resolution comparable with that obtainable with emulsions, but with real-time data acquisition and larger sensitive area. Volcano-eruption dynamics mostly depend on the gas content, the chemical composition of the magma and the conduit dimensions and shape. With respect to the last two items, traditional measurement methods (gravimetric, seismological and electromagnetic), can achieve resolutions of the order of several

100 m, in optimal acquisition conditions. Muon radiography can improve resolutions by one order of magnitude. Besides the scientific interest, most pronounced in the case of the so called *Strombolian* eruptions, understanding eruption-dynamics can have relevant social impact too, as in the case of Mt. Vesuvius, the highest volcanic risk in Europe with about 600,000 people living in the red zone around the volcano or on its slopes. The muon telescope will be composed by two X-Y

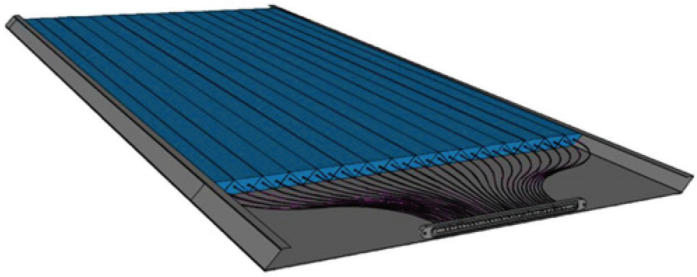


Figure 1.12: Schematic view of a 32 bars module of the MU-RAY project.

stations with a sensitive area of 4 m^2. A third station will be used to study possible backgrounds, as the one induced by cosmic-ray showers. The planes will be composed of four modules, each composed of 32 bars, 2 m long and $\simeq 0.5\,m$ wide (see Fig. 1.12). The telescopes are required to be able to work in harsh environment imposing a modular structure, each module being light enough to be easily transported by hand. Further requirements for the project are mechanical robustness and easy installation. Improvements to background suppression could come from good time resolution allowing to measure the muon time of flight.

Another very interesting work is [28]. It gives interesting future perspective to the application of muon tomography towards the prediction of volcanic eruptions. In their application authors use cosmic-ray muons arriving nearly horizontally along the earth as a probe of the inner-structure of a gigantic geophysical substance. They developed a simple detection system comprising a plastic scintillator hodoscope which is expandable to a larger scale. The first measurements of the inner-structure of Mt. Tsukuba are successfully described in the paper.

1.4.2 Non conventional muon techniques

Non conventional muon tomography techniques, used for imaging problems, are also described in [29]. Here cosmic-ray muons are assessed for their practical use in the tomographic imaging of the internal composition of large vessels over 2 m in diameter. The technique is based on the attenuation and scattering of cosmic-ray muons passing through a vessel and has advantages over photon-based methods of tomography that it is extendable to object containing high-density materials over many tens of meters. The main disadvantage is the length of time required to produce images of sufficient resolution and hence cosmic ray muon tomography will be most suited to the imaging of large structures whose internal composition is effectively static for the duration of the imaging period. The primary issue with using muons as a radiographic source is the determination of their incoming direction. It is proposed that two pixellated detectors be used to determine the presence of a muon with a vector than traverses the object.

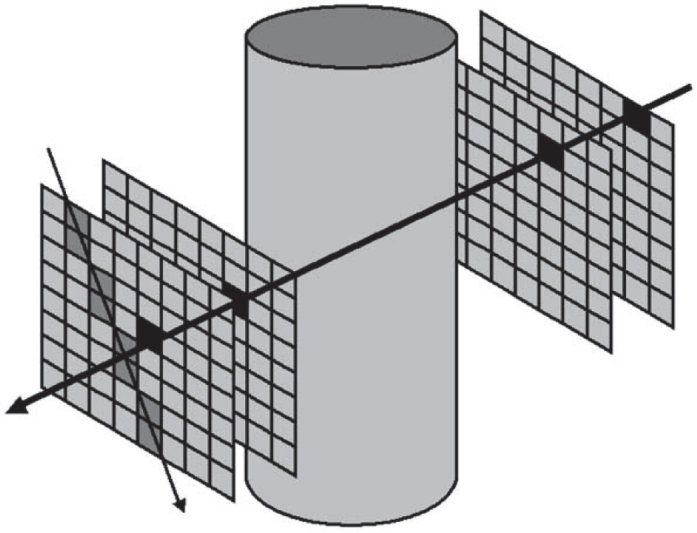


Figure 1.13: A muon attenuation CT system [29].

Two subsequent pixellated detectors are used to detect whether the muon has passed through the object or has been attenuated, see Fig. 1.13. The four detectors are run in a timing coincidence modality to ensure that muons which do not pass through the object are rejected. Each radiograph will have to be collected over a large period of time to ensure that the cosmic-ray muon.

1.4.3 Muon tomography for nuclear waste problems

Muon tomography is successfully applied also to problem related to the nuclear waste. The same kind of technology that is used in the *Muon Portal* project, has been used recently to study new kind of tracker prototype for the cosmic ray muon tomography of legacy nuclear waste containers, to assess the potential of this technology in the identification and characterization of high-Z materials [30]. The University of Glasgow is presenting the prototype of a muon tomography system, consisting of four scintillating-fibre tracker modules. Each module comprises two orthogonal planes of Saint Gobains plastic scintillating fibers with $2mm$ pitch and 97% active cross-sectional area. One detection plane contains a single layer of 128 fibers optically bonded onto low-Z, machine-grooved Rohacells support sheets. First results from the project have confirmed the high-Z material detection capabilities of this detector system which have verified the results of promising, initial simulation studies. Studies are ongoing using the small-scale prototype in Glasgow. The results of these will influence the design and construction of

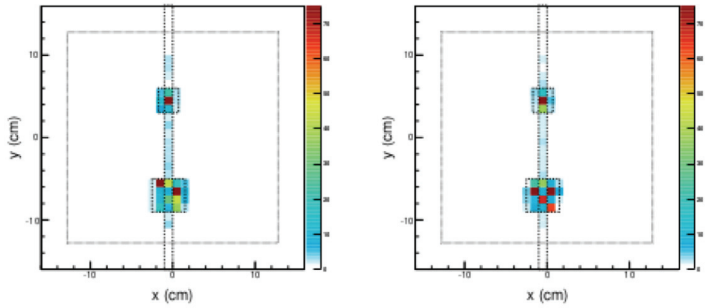


Figure 1.14: Reconstructed images comparison for experimental data taken from several weeks of exposure to cosmic-ray muons (left) and Geant4 simulation (right). A 1 cm slice in the xy-plane is shown, i.e. parallel to the detector modules. The dashed lines provide an estimation of the location and dimensions of the test objects within the volume.

a large-scale detector system which will be capable o accommodating an ILW container for assay. It is foreseen that with further development, this system will be employed in the future to assay the contents of legacy nuclear waste containers at Sellafield, and in doing so, significantly impact upon future storage policy by helping to mitigate the risks inherent with the long- term storage of nuclear

waste. First image reconstruction results from experimental data taken using this prototype detector system are presented in Fig. 1.14 in comparison to the corresponding simulation data for the same geometry and duration. In both the results from experimental and simulated data, sensitivity to atomic number Z and discrimination between the λ values of the stainless-steel bar, the two high-Z materials, and the surrounding air is observed. The non-uniformity of the λ values of the reconstructed high-Z objects is attributed to a combination of possible factors; a spread in λ values for high-density materials as a result of increased Coulomb scattering and non-uniform voxel coverage of the objects (e.g. for the uranium object, the central voxel with a reconstructed λ in excess of $70\, mrad^2 cm^{-1}$ is assumed to fully occupy the uranium, whereas the surrounding voxels occupy a combination of uranium and air which acts to dilute the reconstructed λ value). In these tomograms, the high image-resolution of this detector system in the xy-plane is observed. Smearing and stretching of the image in the z-direction is also noted, though not presented.

In [31] the attention is focused on the potential use of muons for examining very large industrial structures where long measurement times are acceptable, and such possibilities are investigated. They validate Monte Carlo predictions through a cosmic ray telescope and recording system.

1.4.4 Muon Tomography for security problems

In 2003, the scientists at Los Alamos National Laboratory (LANL) developed the muon scattering imaging tomography technique. In [32] the authors present a muon radiography technique which exploits the multiple Coulomb scattering of these particles for non-destructive inspection without the use of artificial radiation.

To demonstrate proof of principle, they constructed a small experimental apparatus built with a set of four position-sensitive delay line readout drift chambers. Two groups of detectors, each measuring particle position in two orthogonal coordinates, were placed above an object region and two groups were placed below, according to the configuration shown in Fig. 1.15, except that no momentum-measuring planes were used. Two plastic scintillators in coincidence with the outer- most drift chambers provided a timing trigger required by the delay line

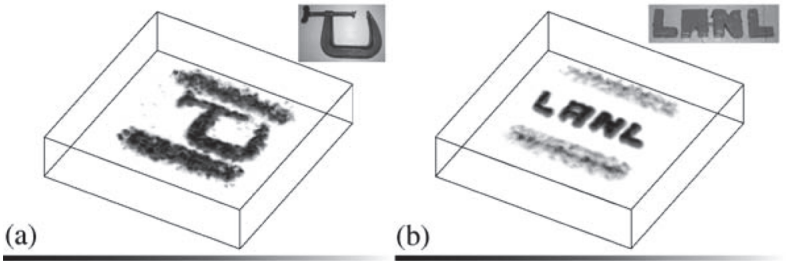


Figure 1.15: Experimentally produced cosmic ray muon radiographs of (a) a steel c-clamp, and (b) LANL constructed from 100 lead stock. The bar-like features result from steel beams used to support a plastic object platform [32].

detectors.

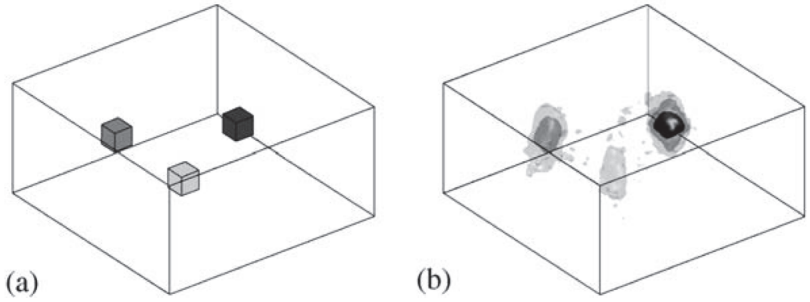


Figure 1.16: Discrimination of material Z (simulated). Objects are shown in (a) three $5 \times 5 \times 5\,cm^3$ cubes, one of tungsten (high-Z; shown darkest), one of iron (medium-Z; less dark), and one of carbon (low-Z; shown lightest). A simulated muon radiograph is shown in (b). The position and Z-level of all three objects is clearly represented.

Signals from the detectors were amplified and discriminated in standard NIM electronics, were digitized in FERA ADCs, and read into a computer using a PC-based data acquisition system. The detectors measured position to a precision of about $400\mu m$ full-width at half-maximum (FWHM), and angles to about $2mrad$ FWHM. The solid angle of this apparatus limited event rate to only a fraction of the available muon rate, and the lack of momentum measurement limited precision of object reconstruction, but the device was sufficient for proof of principle. Two test objects and corresponding cosmic ray muon radiographs of those objects are shown in Fig. 1.16.

An Italian collaboration, among which members of the Fundamental and Applied Physics group of the Department of Mechanical and Industrial Engineering of the University of Brescia, Department of Physics of the University of Padova, INFN Padova and others, recently built and operated the first large scale (about $10m^3$) prototype and demonstrated that indeed it is possible to reconstruct dimensions, position and possibly also the composition of the material contained in the volume under inspection [33].

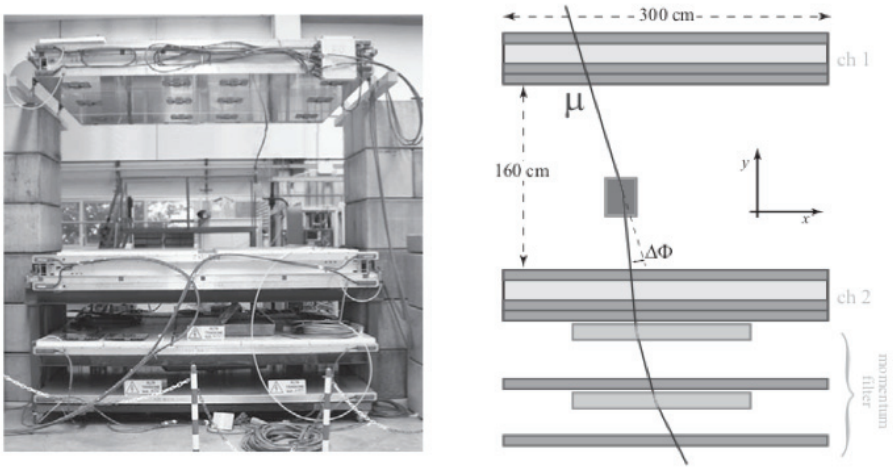


Figure 1.17: The prototype of the muon tomography system [33]. The sketch on the right shows, not in scale, the main elements of the setup. The z- axis is perpendicular to the drawing plane.

Experimental results obtained with a scanning system prototype, assembled using two large-area CMS Muon Barrel drift chambers are presented. The capability of the apparatus to produce 3D images of objects and to classify them according to their density is described. Among the drawbacks the results show that the absorption of low-momentum muons in the scanned objects produces an underestimate of their scattering density, making the discrimination of materials heavier than lead more difficult. Basing on the results of the prototype the Research Fund for Coal and Steel of the European Community decided to approve the Mu-Steel project [34], that started in July 2010 and ended in December 2012.

They used two CMS Muon Barrel drift chambers in the setup shown in Fig. 1.17. The chambers are supported by a concrete and iron structure, leaving a gap of 160 cm in between. Two additional drift chambers have been placed underneath

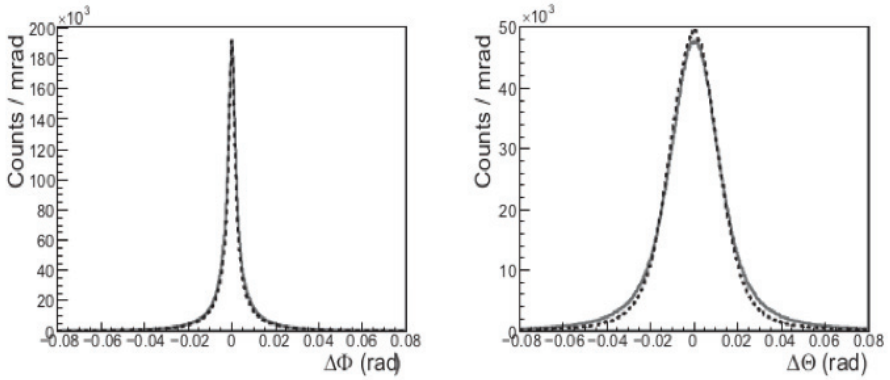


Figure 1.18: Difference between the angles measured by the upper and lower chambers. Left curve Φ angle, right curve Θ angle. The full line refers to experimental data, the dotted line to Monte Carlo predictions.

the lower detector and used, together with iron absorbers, as momentum filter.

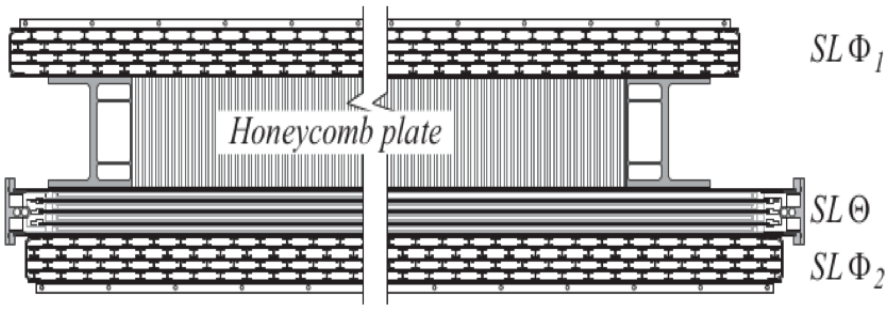


Figure 1.19: Cross section through the CMS drift chamber.

The CMS Muon Barrel drift chambers are described in detail elsewhere [35; 36]. Only their main features are summarized here. The chambers used in this study have dimensions of $300 \times 250\,cm$ and are $29cm$ thick. Each chamber consists of 3 independent units, called Super Layers (SL), structurally connected to an aluminum honeycomb panel. Each SL is composed by 4 planes, called layers, of parallel drift tubes with $43 \times 13mm$ cross section, filled with a Ar (85%) + CO2 (15%) gas mixture at atmospheric pressure. Each layer is staggered by half a cell with respect to the contiguous ones. The two external SLs (called Φ_1 and Φ_2 in Fig. 1.19) have wires in the same direction and measure the muon trajectory in the x-y plane (so called Φ-view), y being the vertical axis. The central SL, called Θ,

has wires perpendicular to the Φ SLs and measures the trajectory in the z-y plane (Θ-view). Each Φ SL contains 286 drift tubes, each Θ SL 227 tubes, for a total of 799 channels per chamber.

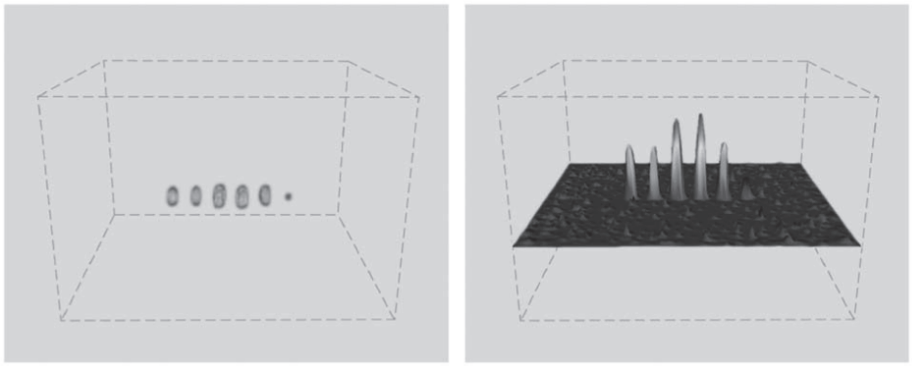


Figure 1.20: Image of the six cubes obtained with the LMIA algorithm. The cube order is (from left): brass, Cu, Pb, W, Fe, Al. Left: 3D view. Right: reconstructed scattering density across a horizontal plane passing through the cube centers.

In the Φ-view the muon track is measured by 8 points, while in the Θ-view only 4 points are available. Linear fits of the measured points provide the projected angles Φ and Θ of the reconstructed muon trajectory. The reconstruction of the material distribution inside the volume between the two chambers is based on the measurement. As an example, Fig. 1.20 shows the results of the reconstruction with the LMIA algorithm for a data set (run) collected with the W sample of 8 cm thickness in the vertical direction. The left image shows the 3D distribution of the reconstructed voxel densities. The right figure is a 2D histogram of the scattering density reconstructed in the voxels belonging to the horizontal plane crossing the block centers.

In [37] and in [38] Micro Pattern Gaseous Detector (MPGD) technologies, in particular Gas Electron Multiplier (GEM) detectors are used as excellent candidates for Muon tomography. Detection and imaging of high-Z materials with a Muon Tomography Station (MST) using GEM detectors was studied by an international collaboration, among which members of the U.S. Department of Homeland Security, ALICE DAQ group and members of the Florida Institute of Technology (FIT). Thanks to this technology position-sensitive detectors with high spatial resolution are builded for optimal tracking of incoming and outgoing cosmic ray

muons. The authors have built and operated a minimal MTS prototype based on $30cmx30cm$ GEM detectors for probing targets with various Z values inside the Muon Tomography System volume.The results are successful in the detection and imaging of medium-Z and high-Z targets of small volumes. The Gas Electron Multiplier (GEM) detector is a micro-pattern gas detector for charged particles [39]. It uses a thin foil of Kapton coated with copper layers on both sides and pierced by a regular array of chemically etched holes, typically $140\mu m$ apart. A voltage is applied across such a GEM foil and the resulting high electric field in the holes can make an avalanche of ions and electrons pour through the holes. The electrons are collected by a suitable device, here a readout plane with x-y strips. They used a $30 \times 30cm^2$ GEM foils (see Fig. 1.21) based on an upgraded version of the original foils used by the COMPASS experiment at CERN [40], but with the central foil area also sensitive to traversing particles. All components used for the GEM detector construction material and the readout electronics have been selected to minimize the mass and consequently multiple scattering in the GEM detectors themselves.

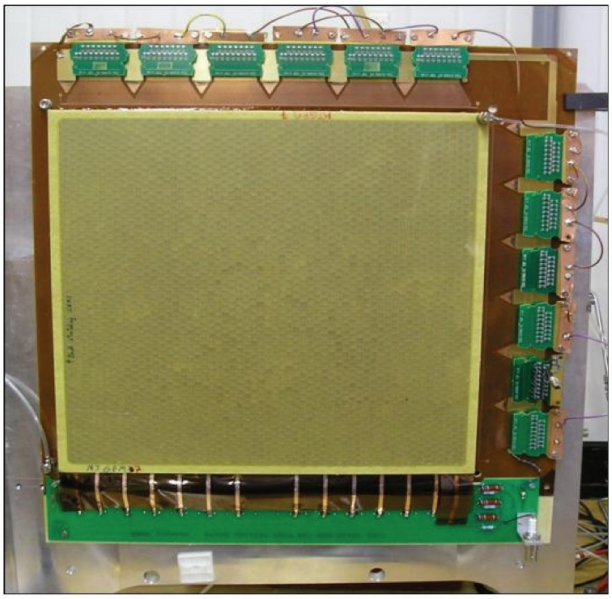


Figure 1.21: Triple-GEM Detector.

An interesting application is showed in [41]. In this paper the authors explore the use of cosmic ray muons to image the contents of shielded containers and detect high-Z special nuclear materials inside them. In particular, they investigate

how radiographic and tomographic techniques can be effective for non-invasive nuclear waste characterization and for nuclear material accountancy of spent fuel inside dry storage containers. The results show that the tracking of individual muons can potentially improve the accuracy and availability of data on nuclear waste and the contents of Dry Storage Containers (DSCs) used for spent fuel storage.

CHAPTER 2

A project for the construction of a muon tomography scanner

In recent years an increasing interest has grown within the international community towards the prevention of terroristic threatenings. Efforts have focused primarily on the illegal trafficking of nuclear, chemical, biological and other potentially deadly weapons. This traffic requires a broader international control and it becomes necessary to enhance systems for borders security in order to prevent the movement of terrorists and the supply of weapons through effective border controls. The more worrying scenario is undoubtedly linked to the use of nuclear weapons, typically composed of a core of fissionable substance (uranium or plutonium) surrounded by moderator material and properly shielded. Typically for border control are used the conventional X-ray techniques. Currently, muon tomography is under study to be used as an experimental technique to be implemented in devices able to perform an efficient scanning to detect radioactive materials inside different kind of systems, like shipping containers or cargos, for borders control. In this chapter the *Muon Portal* project, a project funded by the Italian Ministry of Research, that aims to develop a system able to detect radioactive materials inside containers, to counteract the illegal transfer, by means of muon tomography is described. A detection system based on the muon tomography exploits the phenomenon of cosmic muons deflection when crossing a material. The trajectory of a charged particle within a material is the result of the convolution of many small deflections due to the particles diffusion by the atoms that constitute the material and the overall result of these diffusion processes depends strongly on the atomic number of the crossed material. The chapter is structured as follows: in Section 2.1 the *Muon Portal* project motivation and goal are described,

in section 2.2 the mechanical structure and the design of the apparatus is depicted, showing the actual status of the overall detector. In Section 2.3 the attention is focused on the modules design, together with some design choices and the description of the construction and assembly phases. Then, in Section 2.4 the electronic read out system is presented explaining how data acquisition is made from the electronic point of view. Sections 2.5 and 2.6 present the software packages developed within the project, starting from the control and monitoring machine, going to the description of the reconstruction, elaboration and visualization package, and to the software simulation activities.

2.1 The *Muon Portal* project: basic ideas for a new tomographer

The huge flow of goods today, particularly through containers transported by land or sea, imposes the need of even more stringent control of the content, especially taking into account the security objectives. Over 95 % of the containers are now transported by sea, and only 1 % of them is checked to verify that the content matches the description of the goods. A recent United States law requires, since 2012, a check of all containers directed to American ports, in order to counter the possible illicit trafficking of fissile, radioactive and nuclear material. Such control requires monitoring systems and measures on a large scale, since the total number of containers transported are of the order of tens of millions per year. The American initiative has resulted in bilateral agreements with other countries, in order to check the containers traveling to an US destination in the place of departure firstly. If we also consider the huge amount of passenger and commercial vehicles in transit across borders per each year (more than 100 million at the USA borders), it can be understood that the control of these vehicles looking for nuclear/radioactive material is an issue very difficult to solve.

The *Muon Portal* project [42; 43; 44; 45] has as its target the construction of a prototype of a large area detector for cargo containers inspection in ports in order to counteract the illicit carrying of fissile material. The originality and strength of the project lie in the use of the muon tomography technique to identify the presence of suspicious radioactive material and to reconstruct the container contents, discriminating the type of materials without the need to open it for inspection.

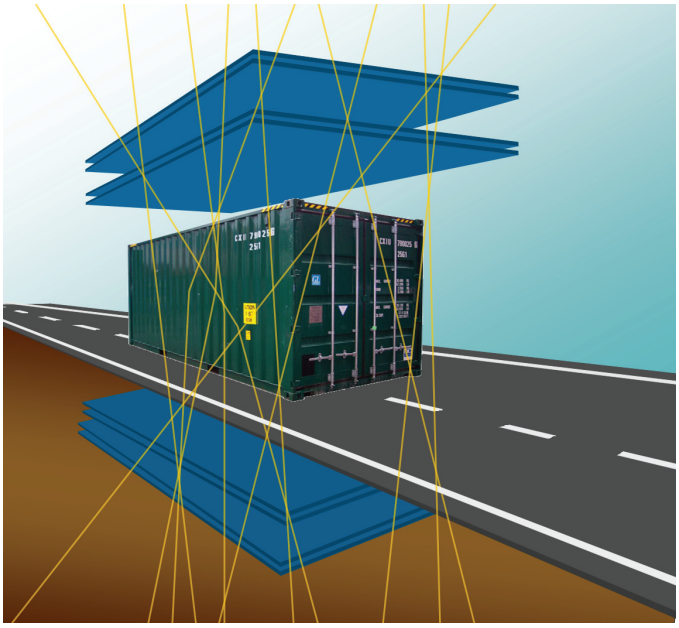


Figure 2.1: Basic idea under the *Muon Portal* project. Artist view of a Muon Tomograph: Two XY logical detection planes are placed above and below the container at a relative distance of the order of 140 cm, in order to reconstruct the muon tracks before and after traversing the container volume, searching for muon scattering from high-Z objects.

As explained in the previous chapter, in literature there are some works that, as an alternative to traditional techniques, propose the use of cosmic radiation for detection purposes. Some use the muon X-ray for border control [33; 46], others suggest to realize muon tomography but using different detection systems, such as gas detectors [37; 38]. Furthermore cosmic radiation is proposed for different aims, for example for volcanoes internal composition detection [13; 32]. The *Muon Portal* project is the only one that aims to realize a large-scale working prototype of a high-Z detection system based on cosmic muons physical interaction with matter.

The scanner prototype, built within the project, overcomes the limitations of traditional methods, obtaining tomographic images using the scattering of cosmic radiation as it crosses each cargo container. In Fig. 2.1 an artist view of a muon tomograph is showed: two XY logical detection planes are placed above and below the container in order to reconstruct the muon tracks before and after traversing

the container volume, searching for muon scattering from high-Z objects. As thoroughly described in the Chapter 1 in fact, the trajectory of a charged particle through any material is the result of the convolution of many small deflections due to Coulomb scattering from the charge of the atomic nuclei in the medium. Due to the dependence of the scattering angle on the atomic number Z of the material, muon tomography is particularly promising to search for the presence of high-Z materials inside large volumes - such as containers - even in presence of additional, low- and medium-Z, objects and shielding materials. This project demonstrates that, when coupled with passive radiation detection, muon interrogation could contribute to safe and robust border protection against nuclear devices or material in occupied vehicles and containers.

As a detection system, a large muon tracker consisting of 8 planes, 6 m long and 3 m wide, large enough for the inspection of standard containers, is under construction. A considerable amount of theoretical analysis with the GEANT code has been conducted. The detector technology used in the project consists of strips of extruded plastic scintillators coupled with a very sensitive photon detector, the Silicon Photomultiplier (SiPM). For the reconstruction of tomographic images, tracking algorithms and suitable imaging software were also developed. Preliminary results of simulations demonstrated the possibility of reaching detection times of few minutes.

Many are the challenges of the project: for example, spatial and angular resolutions are mandatory in this respect, in order to provide a good description of the incoming and outgoing muon tracks. Such performances might be achieved with several detection technologies, some of which have inherent good spatial resolution (for instance multi-wire gas detectors), whereas others with worse intrinsic resolution (such as segmented scintillator strips) may reach good results depending on the number of detection planes and on the relative distance between them. On the other side, imaging algorithms of good quality and able to produce results in a small CPU time are an essential tool for the reconstruction of tomographic images. In the following a detailed description of all the packages that constitute the project and of all the design choices made after an intense R&D activity, is provided.

2.2 Design of the apparatus and mechanical structure

The acquisition of a tomographic image of a container requires the use of a large tracking detector, able to reconstruct the incoming and outgoing muon directions. After more than one year of R&D activity, the *Muon Portal Collaboration* defined the structure of the detection setup which is based on eight position-sensitive physical planes (corresponding to four logical XY planes). Two logical planes are placed above and two planes below the container volume to be inspected (see Fig. 2.2). The detector covers an overall area of 18 m^2 ($3m \times 6m$) and is suitable for a full inspection of a real TEU (Twenty- foot Equivalent Units) container ($244 \times 259 \times 610\, cm^3$).

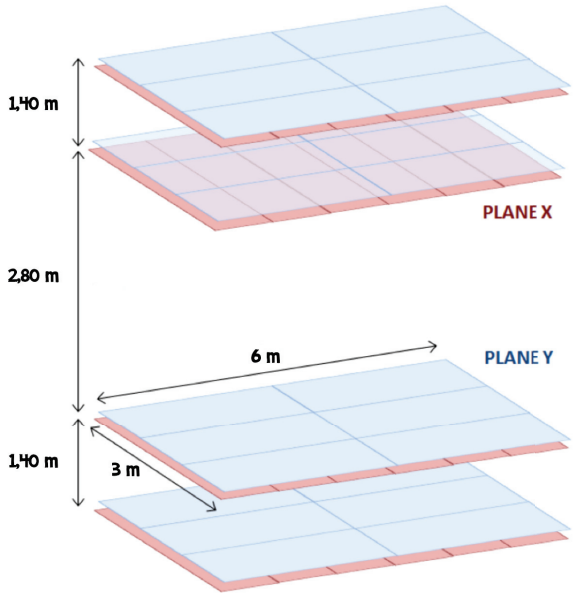


Figure 2.2: Planes schema of the detector structure. Each physical plane is segmented in six modules, placed with different orientations, depending on the coordinate X or Y.

The detector has a modular and scalable structure in order to facilitate the engineering of the entire apparatus and its maintenance. Each plane consists of six identical modules ($1 \times 3\, m^2$ each), suitably placed with different orientations for the X and the Y planes, in order to minimize the dead area, as shown in Fig. 2.2. A total number of 48 modules is necessary for the overall detector. Such geometry implies a slight drop of the detection efficiency at the container borders.

There are several reasons for choosing identical $1 \times 3\,m^2$ modules. Firstly, it costs less to assemble the modules if strips are of the same length and a single assembly tool is used. Secondly, shorter strips (if compared to the overall length required, 6 m) are more efficient. Moreover, shorter strips are easier to handle and so there is less chance of them being damaged while the modules are assembled.

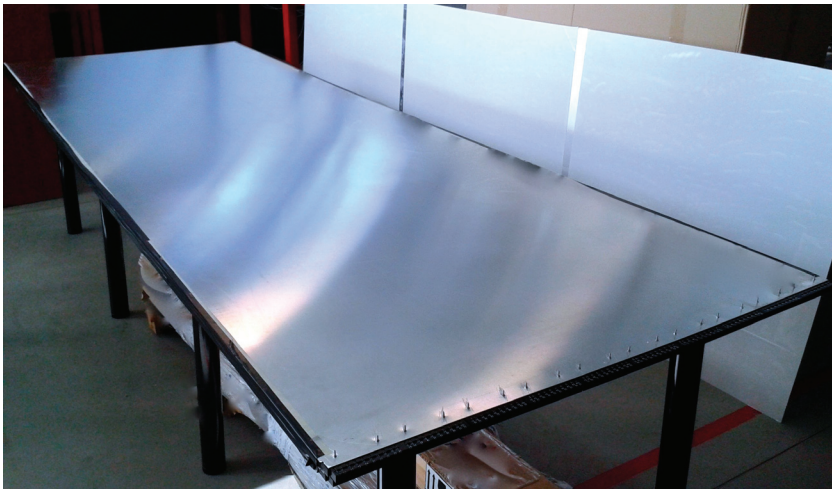


Figure 2.3: Picture of a single complete module. Each physical plane is constituted of six identical modules appropriately positioned: 6×1 for the X-plane, 2×3 for the Y-plane.

In Fig. 2.3 is showed the picture of a single complete module. Each module is made of 100 extruded plastic scintillator strips ($1 \times 1 \times 300\,cm^3$) which are coated with a reflective material ($Ti\,O_2$), with 2 embedded wavelength-shifting (WLS) fibers placed within a superficial groove in the strip and placed in a horizontal position relative to one another. Each fiber collects the light produced inside the scintillator bars and transports the photons emitted to the Silicon Photomultiplier (SiPM) which is optically coupled with the fiber and is located at one of its two ends.

The SiPMs are designed ad-hoc for the project to maximize the light yield with reasonable cost requirements. The spatial resolution, in the order of a few mm, is suitable to provide a good tracking ability of charged particles (electrons and muons), allowing the reconstruction of the incoming and outgoing tracks and, consequently, the scattering angle with a geometrical angular resolution of about $3\,mrad$. Details about each detector component are given in the next sections. The

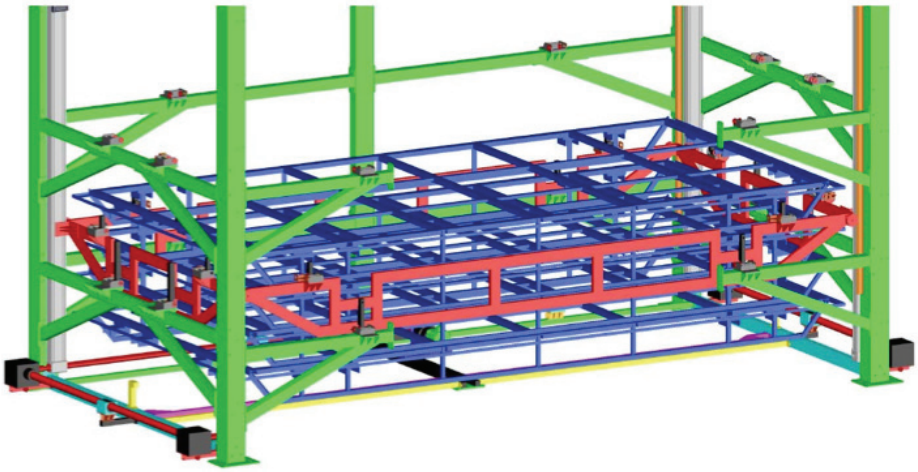


Figure 2.4: Sketch of the mechanical structure of the *Muon Portal* detector. The structure (size $6m \times 3m \times 7m$, is formed by frames (in blue) which will serve as trays for the detector modules and they will be dynamically placed with the aid of an elevator floor (in red).

two upper and lower detection planes are 140 *cm* spaced, while the internal part of the detector is 280 *cm* spaced, to allow the insertion of a standard container. A customized mechanical structure has been designed and already realized by MIWT (partner of the project), to provide adequate support for the detector planes, yet minimizing the quantity of material crossed by cosmic muons, thus avoiding undesired scattering. In Fig. 2.4 a design schema of the mechanical structure is depicted. It is $6m \times 3m \times 7m$ wide and it is formed by frames which will serve as trays for the detector modules as they will be dynamically placed with the aid of an elevator floor. Mechanical stresses due to the weight of the various components and to temperature variations have been studied, together with the alignment problems during assembly phases of the tracking modules. Sensors for alignment and alarms are installed in the structure to monitor temperature variations and mechanical stresses due to the weight of the various components. In Fig. 2.5, a picture of the real mechanical structure of the detector is shown. The structure, which can be controlled through an electrical panel, consists of movable planes (blue) which, once placed via the elevator (yellow) to the loading position, will be hooked up through the use of pistons automatically handled. Some functional

Figure 2.5: Picture of the actual mechanical structure of the *Muon Portal* detector. The structure, which can be controlled through an electrical panel, consists of movable planes (blue) which, once placed via the elevator (yellow) to the loading position, will be hooked up through the use of pistons automatically handled.

tests on the mechanical structure and on the control system have been run after the final installation of the system at the local dedicated within the Physics Dept. of University of Catania. The functional conditions both in terms of mechanical skills and of proper functionality of the automation system have been already checked and verified. In mid-November of this year the first logical plane of the detector (two physical X-Y planes) was assembled and mounted on the mechanical

structure. The Fig. 2.6 shows the entire plane and a particular of it.

(a) (b)

Figure 2.6: Picture of the first logical plane (X-Y) which has been recently assembled and positioned on the mechanical structure (a), particular (b).

2.3 Design and realization of the detection modules

2.3.1 Scintillator strips and fibers: collection and detection of the scintillation light

As mentioned before, each detection module is segmented into 100 strips of extruded plastic scintillator ($1 \times 1 \times 300\,cm^3$), with wavelength-shifting (WLS) fibers to transport the light produced in the scintillator to the photosensors placed at one of the fiber ends. The choice of extruded plastic scintillator strips for the detection of cosmic muons is due to a reasonable compromise in terms of efficiency, light yield, emission and absorption spectra, aging properties, maintenance, cost and availability on the market. Several prototypes of scintillator strips and WLS fibers from different suppliers have been tested by measuring their light response to the passage of cosmic muons, in order to choose the configuration for the final design of the detector, and optimize the light collection at one end of the strip. During the design phase the following parameters were considered: type of material, section and dimensions of the strips, type, number and placement of the fibers on the strip. In particular for what concerns the strips, the following ones have been taken into consideration and have been tested:

- Plastic scintillator strips provided by Uniplast (Vladimir, Russia) [47] , with dimensions $10\,mm \times 7\,mm \times 600\,mm$
- Plastic scintillator strips provided by Fermi National Accelerator Laboratory (Chicago, USA) [48], with dimensions $10\,mm \times 10\,mm \times 600\,mm$
- Plastic scintillator strips provided by AMCRYS-H (Ucraina) [49], with dimensions $26\,mm \times 10\,mm \times 600\,mm$.

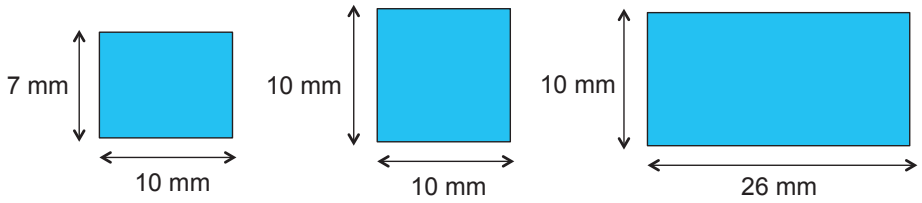


Figure 2.7: Different strips design configurations with different sections: respectively $7\,mm \times 10\,mm$, $10\,mm \times 10\,mm$ and $10\,mm \times 26\,mm$.

A schematic draw of the three different sections is showed in Fig 2.7. The differences between the strips that determine different physical properties are, referenced in [50] and in [51]. They are related to the chemical composition of the bulk, the composition and concentrations of the fluorescent dopant, the extrusion technique, the realization of space for the fibers and the technology used to coat the strips with diffusive reflective materials.

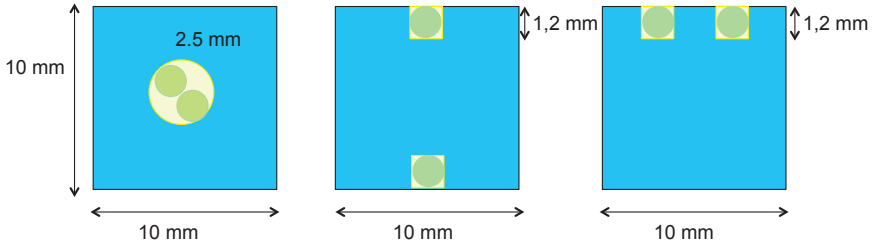


Figure 2.8: Strips design configuration with different fibers positioning on the strip, respectively: two fibers placed in a central hole, two fibers placed in two extruded groove positioned at the two opposite side of the strip, and two fibers positioned on two grooves on the same upper-side of the strip.

Since the collection of the light also depends on the emission and absorption spectra of the WLS fibers used, the strips were tested with different kind of multi-

cladding fibers, positioned with different configurations:

- WLS BCF-91A Saint-Gobain fibers [52] (USA) multicladding, $n = 1.42$, $\oslash = 1\,mm$, $L = 3.5\,m$
- WLS Y11 (200) M Kuraray fibers [53] (Japan) multicladding, $n = 1.42$, $\oslash = 1\,mm$, $L = 3\,m$.

They were tested using as SiPM the StMicroelectronics model 4SPM20-62N, 2mm x 2mm (PDE @ 490nm = 28%). Fig. 2.8 shows such possible designs for the fibers: two fibers placed in a central hole, two fibers placed in two extruded groove positioned at the two opposite side of the strip, and two fibers positioned on two grooves on the same upper-side of the strip.

A measurement campaign was conducted with different strips and fibers varying the over voltage Ov and the threshold Th_{pe} measured in photoelectrons (p.e.). No glue was used to fix the fibers in the groove or in the center hole because this is an expensive, time-consuming and not indispensable operation. The setup employed for such tests included a dark box with the strip, fibers and SiPM readout modules, together with a small-area scintillator to define a region along the strip, where to measure the light yield detected for each SiPM. The SiPM charge spectrum was acquired by the use of a digital oscilloscope. In order to reduce the noise due to the SiPM dark count rate, the coincidence between the two SiPMs was imposed during the acquisition.

The goal of the analysis was to choose the basic elements of the detector and optimize the operating conditions in order to have the highest efficiency possible that should also be consistent with the maximum dark rate that the system can support. Measures and tests were related to the study on the scintillation light yield and the attenuation length: the scintillation light produced, in fact, crossing the WLS fiber undergoes an attenuation which depends on the distance (x) between the SiPM and the point where the light is produced. It can be described by an approximate expression of the type:

$$I(x) \simeq I_0\, e^{\frac{-x}{\lambda_l}} \tag{2.3.1}$$

Where $I(x)$ is the intensity of the light that reaches the SiPM, I_0 is the emitted light intensity, x is the covered distance and λ_l is the long attenuation length.

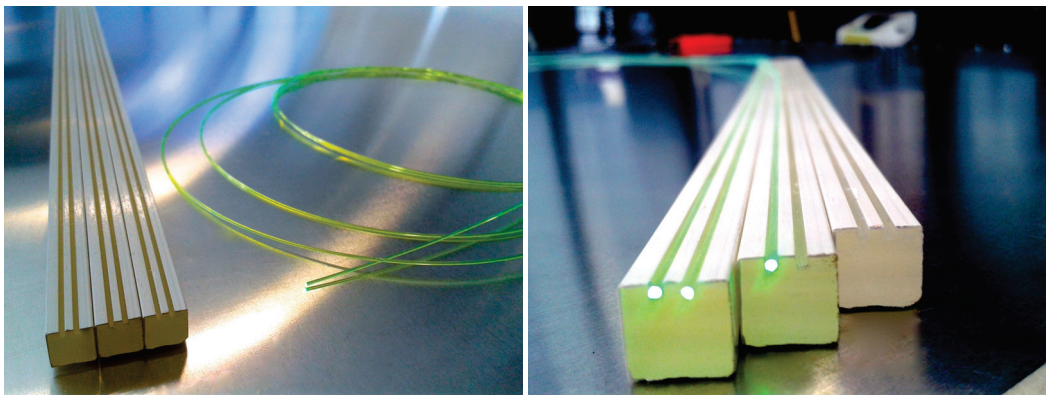


Figure 2.9: Strips-fibers final configuration chosed for the *Muon Portal* detector: Amcrys strip $1\,cm \times 1\,cm \times 300\,cm$ with two Kuraray WLS fibers, positioned on two grooves at the same upper-side surface.

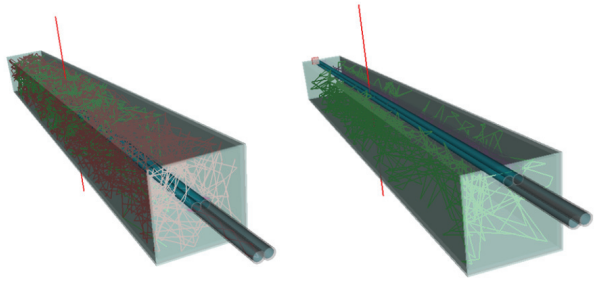


Figure 2.10: Two different extruded scintillator strip designs, one with the fibers placed at the center of the strip, the other one with fibers placed in a groove on the upper surface of the strip, together with typical GEANT4 simulated events generated by the passage of cosmic muons, with optical photons being transported inside the detector.

Indeed one of the simulation activities made with Geant4 was related to the choice of the geometry of the strip and the type of fiber that are the key elements for the performance of the tomograph. Before purchasing of all the materials it was necessary to create simulations and accurate tests. In Fig. 2.10 are showed simulation images related to the choice of the geometry of the grooves. The simulations with Geant4 took into account all the parameters of the scintillator and

fiber. In the example we have considered the strip type Amcrys and WLS fiber Kuraray. The figure shows a picture of two strip prototypes designs tested in our laboratories: on the left the design with a unique central hole is depicted, while, on the right the design with two groove on the upper side is showed. In both case we decided to test the behavior and the light gain that we have using two WLS fibers. Two fibers by $\phi = 1\,mm$ are inserted in various geometric configurations in a strip from $1 \times 1 \times 300\,cm^3$. It was determined the number of photons which protrudes from one end. The other end was covered with reflective material with $R = 80\%$. It was considered the effect of cosmic muons in various positions in the vertical strip. The physical and geometrical parameters of the strip were recruited on the manufacturer's data Amcrys: the reflectivity of the walls was assumed equal to 95%. The glue has been inserted into the groove.

The tests results together with detailed Geant4 simulation studies of the transport of the light inside the strip, suggested to choose for the final design, as a good compromise between the efficiency and the costs, $1 \times 1cm^2$ strips from Amcrys in order to preserve the detection efficiency along the strip above 90%, with two Kuraray WLS fibers, placed on the grooves at the same upper surface of the strip (see Fig. 2.9). For more details see [54].

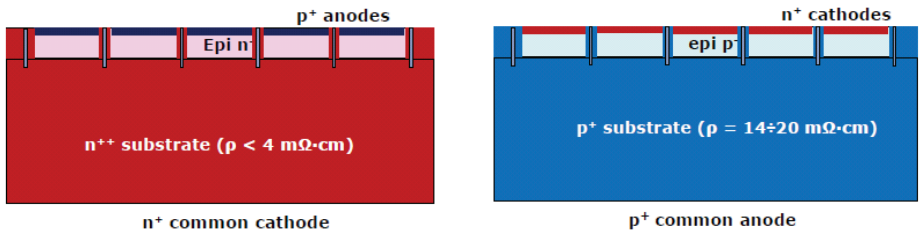
2.3.2 Silicon Photomultiplier photo-sensor design and characterization

For the readout of the scintillation light produced in the strips and transported by the WLS fibers a new type of sensor is being used: the Silicon Photomultiplier (also known in the literature with the acronym SiPM). A detailed description of the characteristics of SiPMs, which follows below, will make clear the many advantages that lead the *Muon Portal Collaboration* to prefer the use of the SiPM, rather than more traditional photosensors as photomultiplier tubes (PMT), photodiodes, phototransistors or APDs. A first important point is related to the economic aspect. In fact, a solution that uses the PMT (single or multi-anode type) implies much higher costs. Secondly, the SiPM is a solid semiconductor device and therefore is possible to create a very compact block scintillator-photodetector. This also translates into indisputable advantages deriving from the reduction of the weight of the structures to be implemented and by a concrete simplification of

the structures themselves. Compared to the solution with the PMT, the supply voltages are greatly reduced (few tens of volts, instead of about a thousand). This naturally translates into a reduction of the heat dissipation. Compared with other semiconductor light sensors (photodiodes, phototransistor, APD), SiPMs have a much higher gain. This simplifies the readout electronics and provides good sensitivity for the detection of the luminous flux of very low intensity, provided in our apparatus. Compared to the APD, SiPMs have a much better gain stability with respect to the supply voltage. Other advantageous aspects of the SiPM, such as reduced rise time and smaller signal recovery time, are of little interest in this application. In the following there is a brief description of the device.

The Silicon Photo-Multiplier: functioning principles

The Silicon Photomultiplier is a silicon sensor that consists of an array of photodiodes (microcells) operating in Geiger mode. The anodes and cathodes of microcells are simultaneously polarized in the same conditions of overvoltage via metallization anode and cathode common areas. Each of the photodiodes of the SiPM is connected to a resistor of quenching ($R_{Quenching}$) that shuts down the event of an avalanche triggered by a photo-generated or thermally generated e-h pair, on each device. The potential drop across the resistor (of value between some hundred $K\Omega$ to a few $M\Omega$) leads the voltage across the junction of the switched on device just above the breakdown threshold. In this way, the current flux in the photodiode at shutdown is lower than a threshold (called current latching) below which the number of charge carriers to the junction is so low that it is statistically unlikely that they could trigger a new avalanche event. Switch off occurs in a time ranging from a few hundred picoseconds to a few nanoseconds, according to the value of the capacitance of the depletion region of the pixels and stray capacitance. After switching off the voltage across the junction, an overvoltage value is imposed by the biasing circuit with external recovery time that depends on the resistance shutdown value, as well as by the capacity values of which has been said before. Typical values of recovery time are in the range between a few hundred nanoseconds up to a few microseconds. The operation mode of the single pixel is called Geiger. The presence of a quenching resistor for each individual



(a) (b)

Figure 2.11: Two different SiPM technologies developed at StMicroelectronics: n/p structure (a) and p/n structure (b)

pixel of the array makes each photodiode of the SiPM decoupled from all the other devices of the matrix. Indeed each resistor switches off only the photodiode connected to itself. The other pixels continue to be polarized to the overvoltage value imposed by the biasing external circuit until a couple of photo-generated or thermally generated carriers, will not trigger an avalanche event. We can therefore say that the SiPM is constituted by a number of photodiodes, each of which functions as an independent photon counter. The typical number of microcells in a SiPM ranges from a few hundred to a few thousand per mm^2, according to the size of the active area of the single microcell and the distance between adjacent microcells (pitch). The current output signal from the pixels of the SiPM is integrated in a certain time interval, called gate (typically of the order of some tens of nanoseconds) and is read as voltage fall across a signal resistor (Rout), of value between a few dozen and a few hundred Ohm. The potential drop at the terminals of this resistor is proportional to the product of the number of pixels turned on and the amount of charges produced in each event of an avalanche. It therefore provides information on the number of photons incident on service integration in the gate array provided. This number is roughly less than or at most equal to the number of pixels of the SiPM. When the photon flux incident on the array becomes intense, it may manifest effects of non-linearity (saturation) for which the number of photons detected (i.e. the number of pixels turned on) is less than the number of photon absorbed by the sensor. The SiPM operates in analog mode since the output provides a signal proportional to the flux of the incident photons, while the individual microcells operate in digital mode. The Silicon photomultipliers have

a high gain at low voltages polarization (a few tens of volts), therefore, usually is not necessary that the external electronic amplify the output signal from the photosensor. The sensitivity at low light fluxes (down to the individual photon), the high speed of response to the incident optical signal and low recovery times are other important features of a Silicon Photomultiplier, that lead our choice in the design phase.

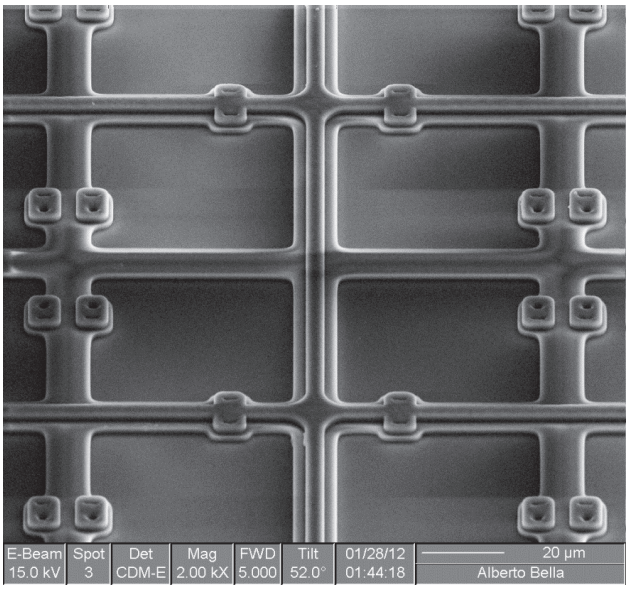
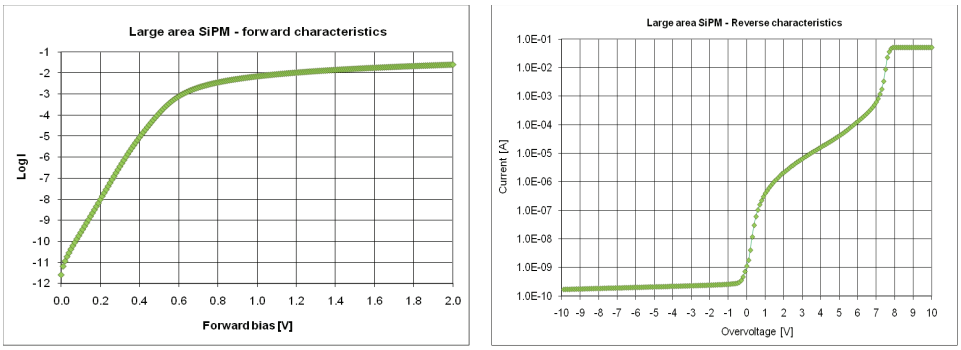


Figure 2.12: SiPM layout from StMicroelectronics, compatible with both p/n and n/p technologies.

Design choices and tests

Different SiPM technologies were developed and tested by STMicroelectronics [55; 56; 57], which is one of the industrial partner of the Project. The goal of such activity was to arrive at the production of devices with a high photon detection efficiency (PDE), a high fill factor, and a low cross-talk. Several prototypes, as shown in Fig. 2.11, both with the P on N (p/n) and N on P (n/p) technologies, were tested in a preliminary phase. The SiPM layout which is compatible with both technologies is visible in Fig. 2.12. In Fig. 2.13 are showed the forward and the reverse characteristics measured on a n/p SiPM device, Mod. 4SMP20-62. In (a) we can see the forward characteristic, in ordinate we have the logarithm

of the current, while in the abscissa we have the voltage. The curve show the ideality factor of the diode. In (b) is reported the current (A) in function of the overvoltage (V). The breakdown voltage (BV) is at 28 $Volt$ and the bias range is $\simeq 5\,V$, before the high bias discharge, being the working overvoltage (OV) up to $\simeq 10\,Volt$ (this is depending on SiPM size and on the cell pitch). The gain is greater than 10^6 (also depending on the cell size and the overvoltage), and the dark current density is equal to $500\,nA/mm^2$ with $OV = 5Volt$, while the dark noise rate density $< 0.65\,MHz/mm^2$ with $OV = 5Volt$. Further studies were carried out to determine exactly the bias point of the region in which make the SiPM to work, choosing a trade-off between detection efficiency and noise due to the dark current, as explained in what follows.



(a) (b)

Figure 2.13: Typical I-V forward and reverse characteristics for the SiPM Mod. 4SMP20-62, which explain the functioning of the silicon photo-multiplier.

Fig. 2.14 shows the photo-detection efficiency (PDE) measured on devices with cell fill factor equals to 45% and on the two different configurations at 4 V of overvoltage. The graph is showing the PDE, that take into consideration also the given cell fill factor, at the different wavelenght. As it can be seen, the p/n structure has its maximum around the frequency of the blue light (around the 430 nm), while the n/p has better performance for the detection of the green/red light, having its maximum around a wavelenght of 500 nm. This consideration lead the collaboration to choose the n/p configuration, given that the WLS fibers emit the light in the green wavelenght.

In the graph is present a secondary effect, due to a delayed cross-talk effect,

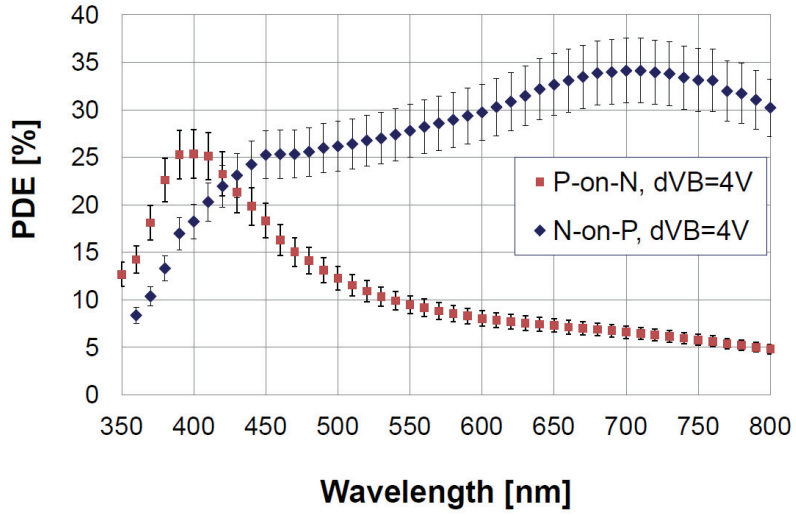


Figure 2.14: Photo detection Efficiency for the two technologies p/n (red) and n/p (blue) measured at Ov=4V.

which justify the unusual behavior in the wavelengths around 700 nm.

In Fig. 2.15 are showed some results related to the SiPM dependence on the temperature, related to the n/p technology. In the figure we have, in clock-wise orientation: the variation of the breakdown voltage (V), as a function of the temperature (C°). As it can be seen we have a linear trend, that shows a weak dependence of the breakdown voltage on the temperature, being the gradient $\Delta V_{BD}/\Delta T \simeq 30\, mV/K$. Then the current is showed at different temperatures. The green line is relative to the room-temperature, and it has to be noted that the graph is normalized on the breakdown voltage, which is growing with the temperature. Then the quenching resistance (Ω) as a function of the temperature (C°) is also showed.

Fig. 2.16 shows a summary of PDE measurements as a function of the light wavelength carried out at 410 nm on several devices with different SiPM technologies, as a function of the applied overvoltage. In the range $\Delta\lambda = 500 - 550\, nm$ (which corresponds to the main emission region of the Kuraray WLS fiber) the best device shows a PDE of the order of 35%.

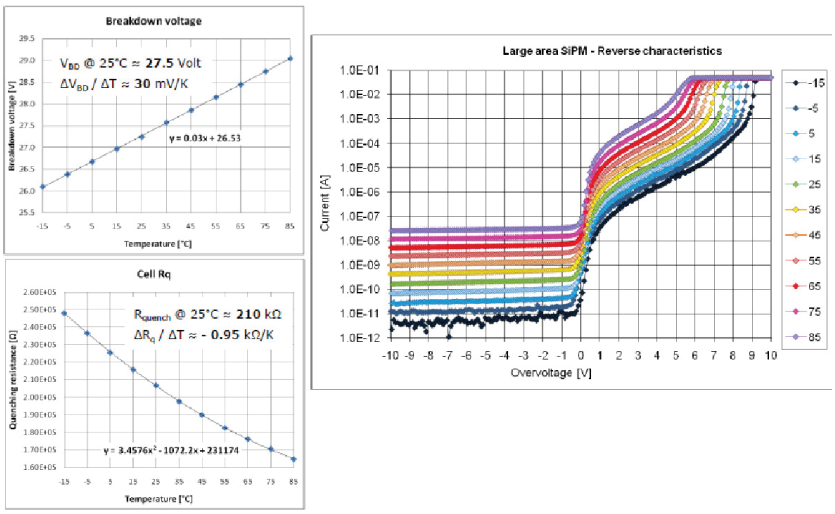


Figure 2.15: In clock-wise orientation: Variation of the breakdown voltage, as a function of the temperature. SiPM dependence dependence of the breakdown voltage on the temperature, the gradient is $\Delta V_{BD}/\Delta T \simeq 30\,mV/K$. And quenching resistance as a function of the temperature.

SiPMs characterization and classification

The layout of the final chip is based on a N^- medium doping technology and embeds 4 independent round shaped SiPMs. The evaluated chip contained two types of SiPMs, which differ for their cell pitch: the MUON-60 type has 548 cells with 60µm pitch, a cell fill factor of about 67% and a photon detection efficiency of 35% for the green light; the MUON-75 type has 320 cells with 75µm pitch, a cell fill factor of about 74% and a photon detection efficiency of 38%. In Fig. 2.17 the chip with the SiPMs matrix is visible while the table 2.1 reports the main features of the chosen SiPMs.

An intense work of characterization of such devices, both from the electrical and optical point of view has been made, at different temperatures and operating conditions, also making use of digital pulse processors. Simulations of the devices under development are also being considered and comparison of the data with Monte Carlo and SPICE calculations have been carried out to improve the SiPM design.

The plot in Fig. 2.18 reports the PDE of the chosen device with values corrected for the dead time and without cross-talk effect. The V_{BD} was measured from the

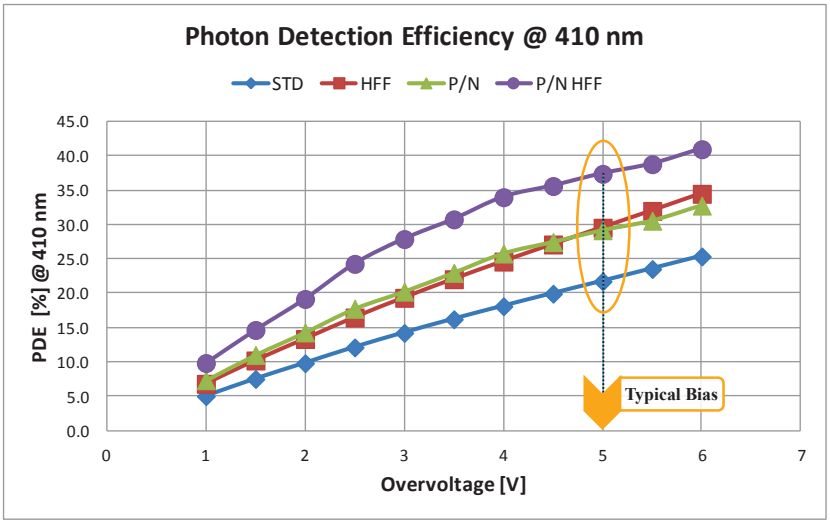


Figure 2.16: Photon Detection Efficiency for different Silicon photomultipliers technologies tested by STMicroelectronics.

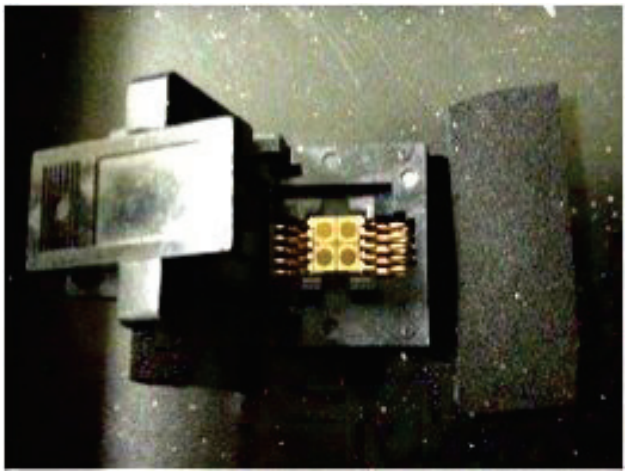


Figure 2.17: Black box used by the experimental apparatus. The socket within which is accommodated the SiPM matrix is visible.

voltage-current measurements and tracing the intercept between the line of best fit (range from 1 mA to 2 mA) and the x-axis.

The SiPM prototype designed for the Muon Portal project maximize the photon detection efficiency (PDE) and the cell fill factor, as well as ensuring a low cross-talk and dark count rate. In the final design the SiPMs will be optically coupled to the polished fibers and will be enclosed inside a hermetic box, in order to be operated in dark conditions at a controlled temperature. Moreover, to improve

Parameter	MUON60	MUON75
Sensitive area size (μm^2)	1.97	1.80
Number of cells	548	320
Cell fill factor %	67.4	73.8
Cell size μm^2	60×60	75×75
Quencing resistor squares number	28	28
Quencing capacitor area μm^2	26	26
Cell active area μm^2	2427	2427
Cell perimetral area μm^2	844	844
Diode bonding pad area μm^2	140×140	140×140
Diode bonding pad area μm^2	140×140	140×140

Table 2.1: Specifications of the SiPMs SPM10H5-60N (MUON-60) and SPM10H5-75N (MUON-75), designed ad-hoc for the *Muon Portal* project.

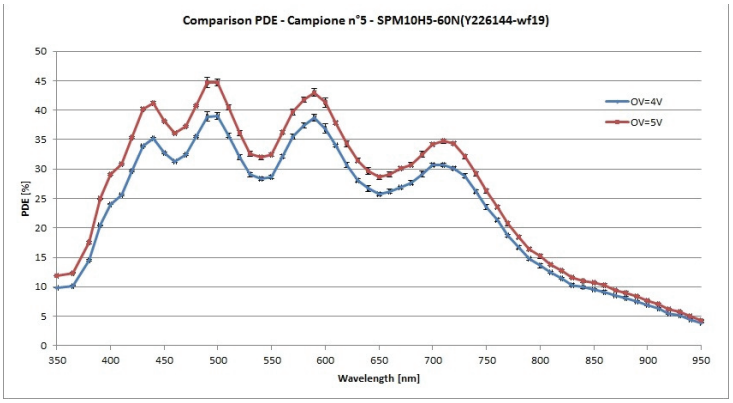


Figure 2.18: Photo detection efficiency comparison for the SPM10H5-60N (MUON-60) device at two different over voltages (4V and 5V).

the uniformity of the SiPMs response along the detection modules, the devices will be divided in groups, depending on their breakdown voltage. SiPMs with similar characteristics will be installed in the same region of the detector in order to set, for group of 10 SiPMs, the same bias voltage and threshold level. Finally, the thresholds will be remotely controlled and automatically varied to ensure the gain stability in case of temperature variation. For this reason, the characterization of all the devices is necessary for their classification before the final assembling of

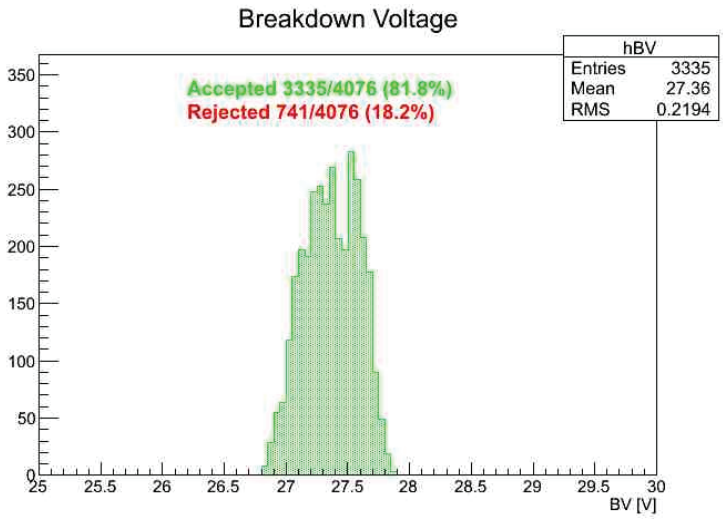
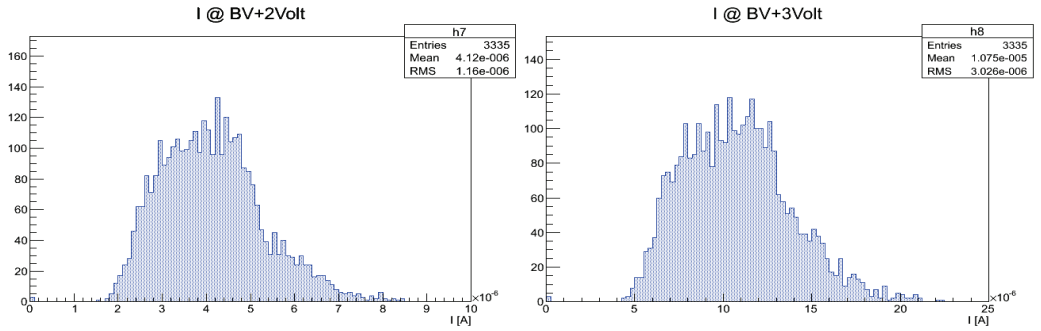


Figure 2.19: Distribution of breakdown voltages related to a sample of 4076 matrices tested. It is shown the percentage of SiPM suitable (green) and rejected (in red).

the detector.



(a) (b)

Figure 2.20: Current distributions at 2 (a) and 3 (b) Volts of Over Voltage. The graphs are based on a sample of 4076 matrices tested.

A batch of $\simeq$ 10000 SiPMs has been produced and encapsulated in a SMD optical package. A custom procedure has been implemented for the characterization of such devices and a LabVIEW program, running on a computer which communicates by RS-232 port with a Keithley picoammeter/voltage source, were used. Figures 2.19 and 2.20 show the histograms, respectively, of the distribution of the breakdown voltages, and of the currents in the two values of over voltage.

These distributions are related to the suitable SiPMs which are a subset of the total analyzed. A large fraction of the overall number of SiPM has been already tested and the procedure is being completed allowing to sort all the SiPMs which will be used in the detector, with an overall photo-sensors yield of about 82%.

2.3.3 Detection Modules assembly phases and tests

The detection modules have been designed according to a layered structure showed in Fig. 2.21: the base is an aluminum sheet of 1mm thickness, on which is

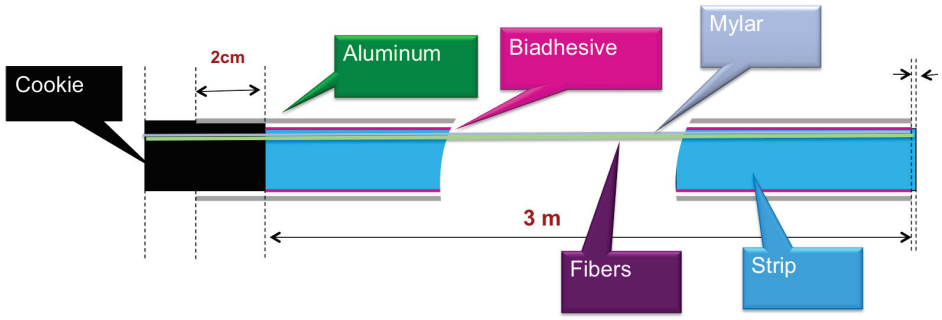


Figure 2.21: Schematized section of the structure of a single module. In order, from the bottom: aluminum sheet 1mm thick, layer of biadhesive, strips containing the fibers, layer of aluminized mylar, biadhesive and aluminum foil.

attached a layer of bi-adhesive. On this layer, the strips and the fibers are placed. Above the strips and the fibers is then attached a sheet of aluminized mylar, which has the aim to improve the light yield of the upper surface of the strip. For each group of 10 strips, each with 20 fibers, a black PVC adapter, called cookie (see Fig. 2.22), is used. It has 20 channels such that the fibers can be inserted on them and can emerge from the cookie having the same length and being placed in a precise location.

The fibers are glued at the bottom of the groove and protrude a few *mm* at most. In order to obtain a cookies+fibers surface perfect to have an optimal match with the SiPMs, cookies are lapped on the side from which the fibers emerge. It has been designed a card-carrying SiPM that brings on board over 10 SiPMs and also the amplifier circuitry and multiplier. SiPMs are welded at the same height and in the positions corresponding to the grooves of the cookie. Finally, the module is closed with another sheet of aluminum.

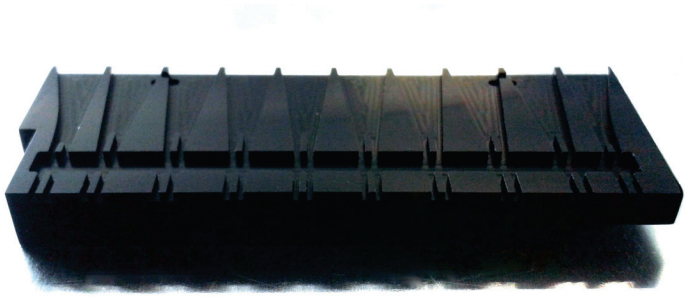


Figure 2.22: A black PVC adapter (cookie) specifically designed for the project in order to ensure the desired fibers positioning so that they can be optimally coupled with the SiPMs.

Assembly phases

For the modules assembly phase, the table shown in Fig. 2.23 of size of $3.1m \times 1.06m$, is used. On one side is mounted the motor that moves the diamond head for the lapping operations. On the other side is mounted a drill for the realization of the holes in the two aluminum plates to fix the cookies. Beside there are two 3 meters long sides. One of them is mobile and permits packing the 100 strips in order to obtain an exact overall dimensions of $1000\,mm$ as required.

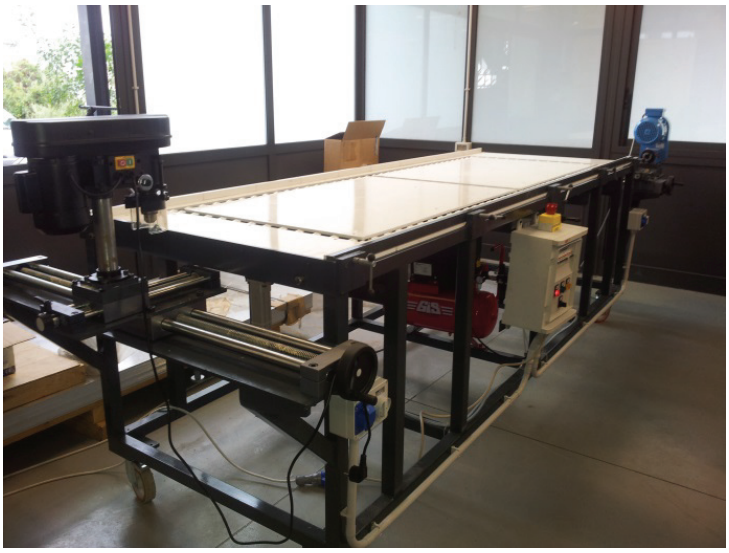
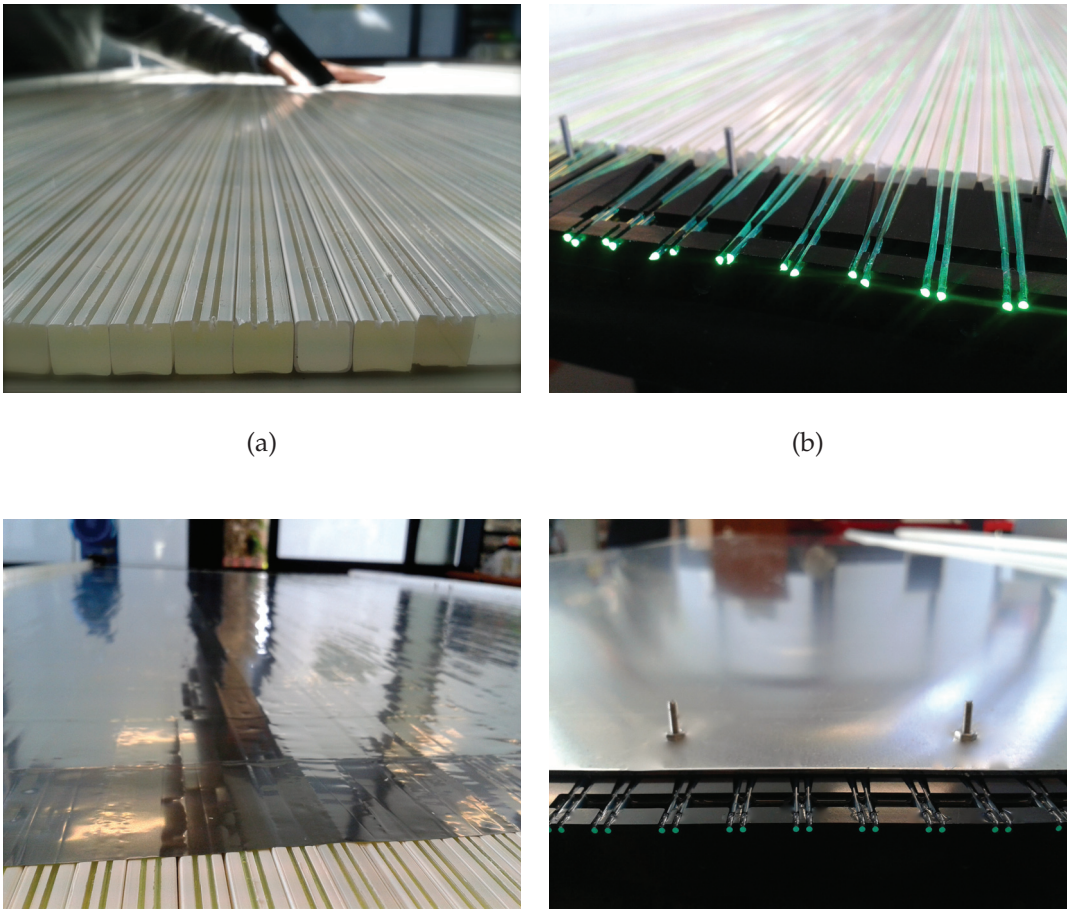


Figure 2.23: Assembly tool used for the realization of the modules. It is equipped with a pneumatic system to lift and rotate the module, a precision drill and a motor with a lapping system achieved by a diamond tip.



(a) (b)

(c) (d)

Figure 2.24: Some assembly steps of the module: After lapping, the strips are glued on the first sheet of aluminum through the double-sided adhesive (a), the fibers are channeled through cookies, stretched out on the strips and glued in some spots (b), a layer of reflective mylar is placed over the strips with fibers (c), the module, closed with another aluminum sheet is completed by applying a special glue to the fibers and by a final lapping operation to the fibers+cookies (d).

The main assembly stages of each module are described in the following:

- The aluminum sheets are perforated so that the cookies can be afterwards fastened.
- The strips are positioned on the plane and aligned to one of two ends. A first coarse leveling operation on the strips is executed using the lapping tool.

- The bi-adhesive is applied to the aluminum foil so then is possible to proceed to gluing the strips on the sheet, Fig. 2.24 (a).

- On the side where the aluminum protrudes, 10 cookies are glued so that they lean out from aluminum of about 14 mm.

- To each strip are stretched out the 2 WLS fibers of 1mm, making them protrude from both sides of the module, Fig. 2.24 (b). This is done to enable the final lapping procedure to completely equalize the two ends of the module. In the front part the fibers are positioned on the cookie channels, so that the fibers fit perfectly in the groove avoiding folds that would compromise the optical behavior. In the figure fibers inserted into the grooves and strips and cookies which have been glued, are clearly visible.

- To improve the light yield the upper surface of the strip is covered with aluminized mylar, Fig. 2.24 (c).

- The other aluminum sheet is then glued with double-sided adhesive on top of the module. In this was it is realized a sort of *sandwich* aluminum/bi-adhesive tape/mylar/(strip and cookies)/bi-adhesive tape/aluminum, tightening all the module with 20 bolts of 2 mm passing through the aluminum holes.

- At this point, the motor with the diamond head, positioned so that it paves only a few tenths of mm, is used, in order to polish the cookies with the fibers. Once this operation is finished, the module from the side of the cookies, is as shown in the Fig. 2.24 (d).

- The fibers inserted in the grooves are glued well at the bottom of the groove so that they are all in the same position.

- Then is possible to proceed to the flattening of the back of the module and to the application of the aluminized mylar at the end of the module so that the light can be reflected back in order to improve efficiency. The edges of the module, except from the side with cookies, are protected from the light with a black tape. In Fig. 2.25 the completed module is shown.

Figure 2.25: Picture of the competed module. The side with cookies and fibers is visible.

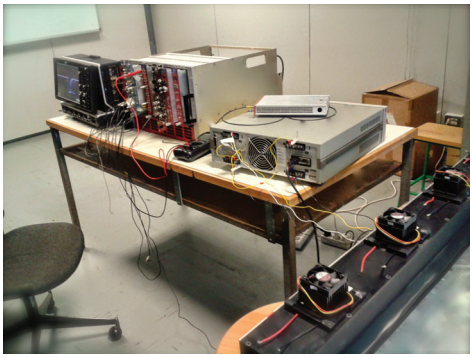
The module is now ready to be inserted into the electronics container shown in Fig. 2.26. In the picture it can be seen the long slot into which the module is inserted. The cooling fans of the Peltier cells, predisposed to maintain the temperature inside the box below 20°, are also evident. A special system, reliable and easy to assemble is designed in order to realize an optical connection.

Figure 2.26: Box used as container of the electronics. The single module is inserted in the box which contains a cooling system properly designed for the Peltier cells.

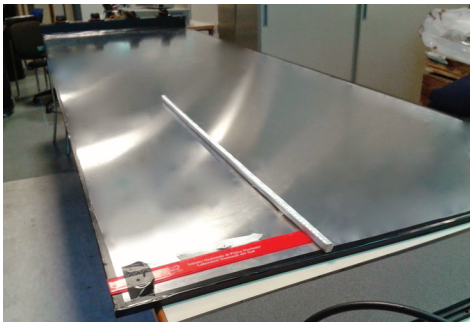
Detection modules response tests

The first modules produced were tested in the laboratory through the experimental apparatus shown in the Fig. 2.27 to measure the detection capability of the single strip and the response of the strip to cosmic muons. Hereafter a few

(a)



(b)

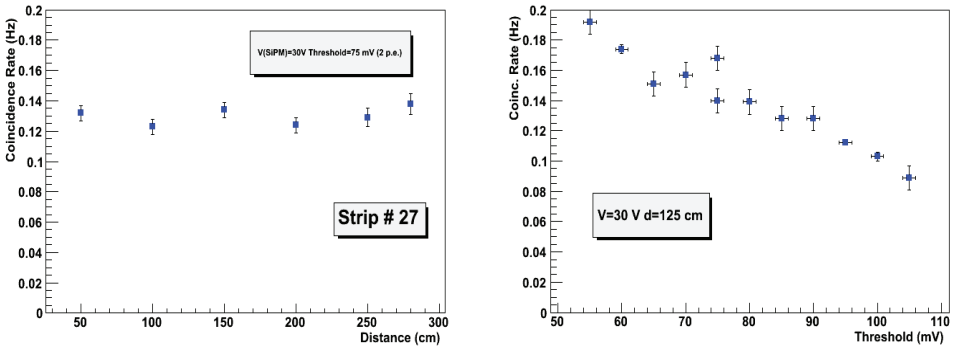


(c)

Figure 2.27: Laboratory measuring operations to test the response of the detection modules to cosmic muons in coincidence with a the external detector (a), the electronic measuring apparatus (b), the setup apparatus used for the module response testing phase (c).

illustrative results are reported.

In Fig. 2.28 the rate of coincidence between the strip #27 of the scintillator module and an external scintillator ($12cm \times 12cm$), positioned above the strip in question at a vertical distance of about $5\,cm$. This coincidence rate is shown as a function of the distance along the module and at $30V$ with a threshold= 2 p.e. in (a), while in (b) is plotted the same rate as a function of the threshold (at a distance $d = 125m$ set). The trend of the graphs is almost constant, considering also the error bars, so it can be concluded that the response of the module is almost independent of the distance of the source, and that the signal does not undergo large attenuations even within 3 meters. This result is certainly also due to the presence of the layer of reflective mylar placed at the end of the strip on



(a) (b)

Figure 2.28: Coincidence rate (Hz) measured vs the distance of the source (a), and varying the threshold once the distance is fixed ($d = 125m$).

the opposite side with respect to the SiPMs. The last figure, Fig. 2.29, shows the signal amplitude distribution of the SiPM for a fixed supply voltage ($30V$) and at a threshold equal to $80mV$. It can be seen the typical muon spectrum, with a mean value which is around $200mV$ that corresponds to $6 - 7pe$.

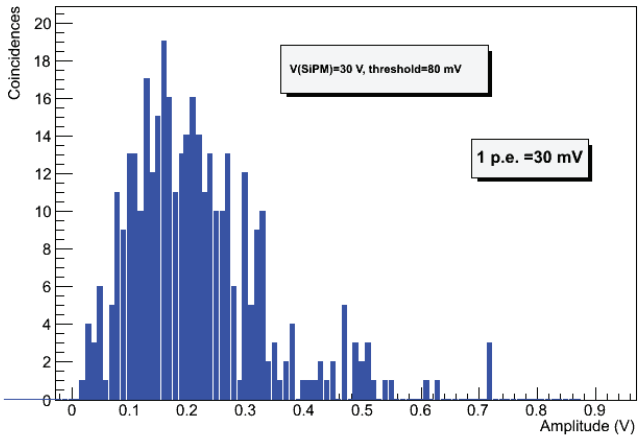


Figure 2.29: Distribution of signal amplitude of the SiPM with supply voltage $= 30\,V$ and Threshold $= 2p.e.$, when taking the signal in coincidence with an external detector.

2.4 Electronic readout and data acquisition

The light emitted when muons cross the strips is collected using two WLS fibers running along the same strip (for a total of 9600 WLS fibers) and going to an equal number of SiPMs. The SiPM signal is then amplified. The amplification, of a factor of 30, is obtained by a very fast operation amplifier, the THS3201 (1.8 GHz GBW). The resulting single photoelectron signal is about 27 mV in amplitude and 20 ns wide. Considering that each scintillating bar has two wavelength shifters, in consequence two SiPMs will give as output two almost synchronous signals when the corresponding bar is crossed by a muon. From the readout electronics point of view, the detector presents a very high number of channels to be read: 100 strips per module (each with two fibers WLS) and 6 modules per each physical plane, in total 9600 channels. In order to simplify the system, a smart readout strategy, [58], which reduces such number of channels to a suitable level for the electronics and data acquisition, is adopted. The output channels for each module are so reduced from 200 to 20, reducing the total number of channels to read by a factor of 10.

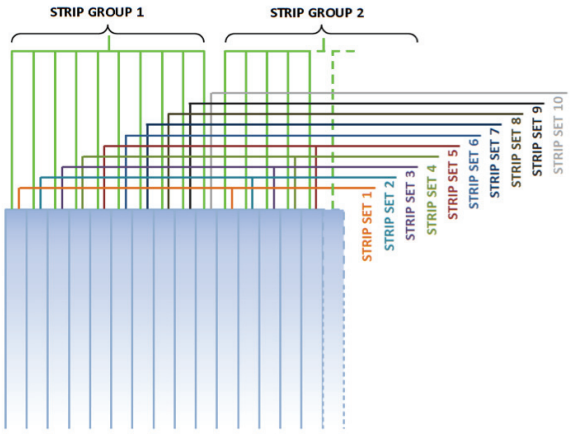


Figure 2.30: The read-out channel reduction system, for a module containing sixteen strips.

An example of the read-out channel reduction system, is shown in Fig. 2.30. In this example, each bar, or strip, has two wavelength shifter fibers read-out by two SiPMs. Following the schema, the analog signals coming from the amplifier output is properly combined in groups and sets. Specifically: for each module we

have 100 strips which are combined in 10 groups of 10 contiguous strips, using the first fiber of each strip and combining them all together with the very fast operation amplifier. The output signal is corresponding to the number of the group $n_{group} \in [1-10]$. The second fiber of each strip is then numbered, for each group, from 1 to 10 and combined with the amplifier to the other corresponding second fibers of the other groups (i.e the fiber number one of the first group with the corresponding fiber number one of the other groups), giving the number of the set $n_{set} \in [1-10]$. In this way the single strip hit will be uniquely identified by a pair of values (n_{group}, n_{set}). Their combination is able to identify the interested strips inside each module by the following formula:

$$Strip_{hit} = n_{set} + 10 * (n_{group} - 1) \tag{2.4.1}$$

In this way the number of channels to each module drops considerably resulting in 20 channels per module, reaching a compression factor equal to $2\sqrt{N}$ if N is the number of strips. Their combination is able to identify the interested strips inside each module.

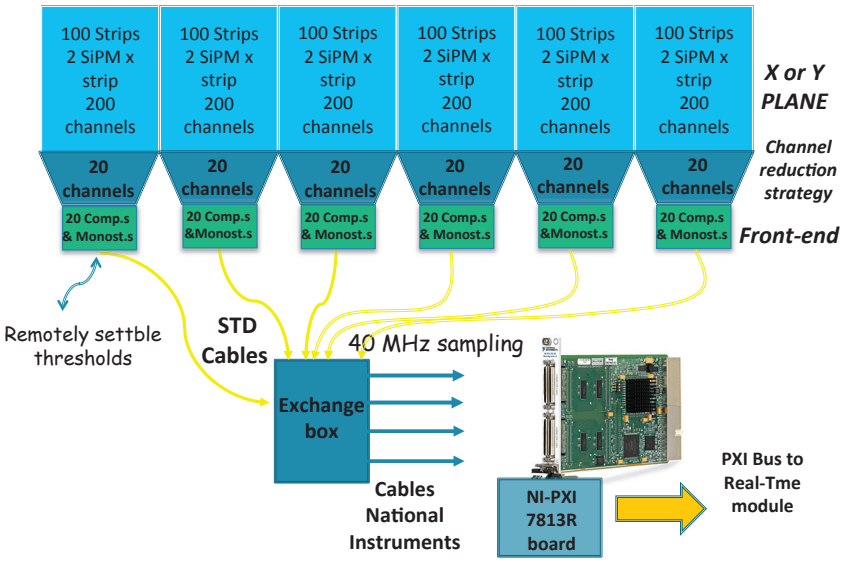
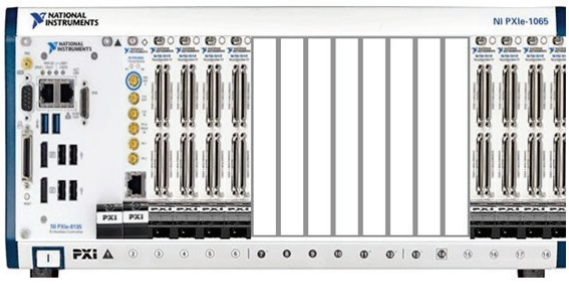


Figure 2.31: Block diagram of the front-end and read-out electronics.

In Fig. 2.31 the scheme of the front-end and read-out electronics is explained in details. In our system we have two signals related to the same strip. As explained before, in order to achieve the channels compression, the analog output of one of the two SiPMs is summed to the nine neighboring analogue signals of the same group, and then amplified by a factor 30, by the THS3201. Instead, the analog output of the other SiPM is summed to the first analog signals of the other groups of ten neighboring signals. The same scheme is repeated for the second module, and so on. The resulting analog signals are successively compared to a remotely settable threshold for binary conversion. After the reduction, there are in total 960 digital channels coming from a fully completed detector. Each module produces flows of 20 bit words. Shielded, ruggedized cables connect each module to the read-out electronics [59]. The PXI-7813 digital sampling board by National Instruments [60], serves as Data acquisition (ACQ). It has 160 digital input channels that are simultaneously sampled at 40 MHz and is connected to six detection modules, which means one complete X or Y detection layer. The eight PXI-7813 needed for the whole detector are, then connected by PXI bus in DMA-access to the real time module in a crate. The read-out is performed in real-time with a module that



n.1 NI Real-time PXIe-8135 module
n.8 NI PXI-7813R boards
n.1 Solid State Disk
n.1 GPS NI PXI-6682 timing board
Flexible and expandable chassis

Figure 2.32: The read-out crate including the real-time module, the GPS timing module and the eight PXI-7813 digital sampling board.

calculates the entrance and exit tracks for each event. A picture of the complete read-out and ACQ system is shown in Fig. 2.32. In order to correlate the arrival

of muons in the Muon Portal detector to additional detectors located around it, a GPS time-stamping unit is also planned in the acquisition architecture. This will allow offline correlation of the events measured in independent detectors, with a time precision of about 40ns.

2.5 Detector control software for the acquisition and monitoring

In view of the full operation of the *Muon Portal* detector, we foreseen the development of a control and monitoring system (hereafter *CCM*), which directly communicate with the data acquisition system software (*ACQ*), together with an imaging elaboration system (hereafter *SEI*). The former is designed to control the data acquisition procedures (i.e. run start/stop, calibration modes, etc.) and to collect, handle and store the monitoring observables generated from different devices in the portal. The acquisition and monitoring software is intended to perform the actual investigation of the scanned object and is also used for checking the correct and full functionality of the detection planes of the scanner. The imaging elaboration system is responsible for the tomographic analysis process, the image visualization and storage and the alarms identification. The measurement investigation is made with advanced graphics multidimensional visualization tools, in order to identify, inside the container, the number and the placement of elements that contain suspicious material with high atomic number (Uranium, Plutonium, Lead). So there were designed some software modules, with different levels of integration and addressed to specific targets, that manage the specifics necessary for the project, for the data acquisition, the control and monitoring of the data quality as well as the graphic visualization of the tomography images. Linked to this goal, it is also provided an efficient organization of the acquired data realizing a specific database, accessible for real-time analysis and also for post-processing phases.

A sketch of the prototyped system and its working schema is shown in Fig. 2.33 The imaging elaboration system (SEI) is made up of a set of computing machines (24 cores in our preliminary design), connected through a local LAN

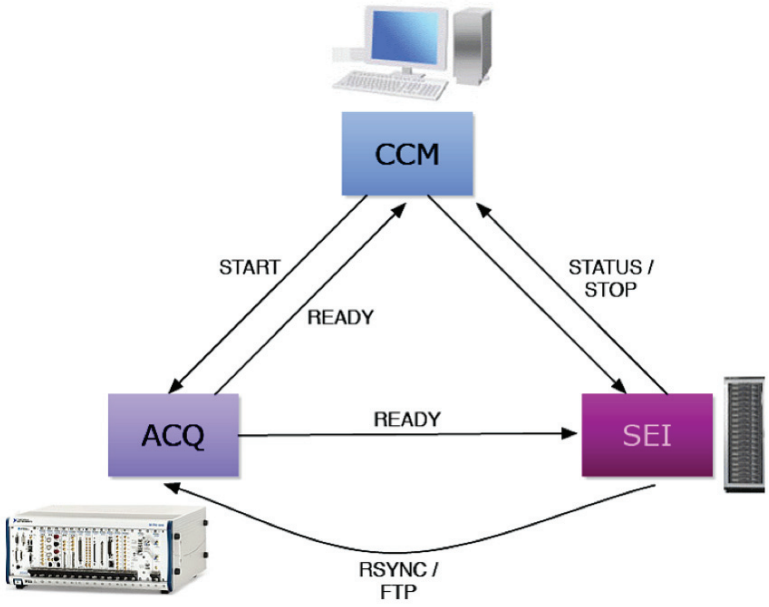


Figure 2.33: Schema of inter-communication between all available devices installed in the Muon Portal project for the online tomography imaging.

network to the data acquisition crate (ACQ) and the controller machine (CCM). The CCM sends a start message to the ACQ that initiates the data acquisition and at the same time also to SEI to inform it of the current status. After a pre-specified scanning time, the ACQ produces the first bunch of raw data and sends a ready message to SEI, which in turn starts transferring the raw data and the starts the image elaboration. At elaboration finished, with or without an alarm status, the SEI communicates the status of the image processing to the CCM. The above procedure is planned to be repeated in case of an alarm status. The communication from/to the CCM and ACQ machines is effectively realized by exchanging structured messages in JSON format [61]. With JSON is possible to manage the communication between machines in a modular, efficient and reliable way. The communication happens through web services and TCP/IP socket protocols respectively, in a simple server/client strategy. This ensures the reliability of the data sent and received as well as guaranteeing a proper time sequence. The code produced for the simulation purposes is written in C/C ++, in order to be independent and compatible with the POSIX standard and to ensure

the code execution on guest machines dedicated to the project. In Fig. 2.34 an example of the inter-process communication via HTTP, between CCM, ACQ and SEI is showed. The START message, for the acquisition procedure, is sent from the CCM to the ACQ, which then starts the scan of the container. When the scan is completed, the ACQ sends a READY message both to the CCM, and to the SEI, that can starts to download the tomography data and starts the elaboration and reconstruction phase.

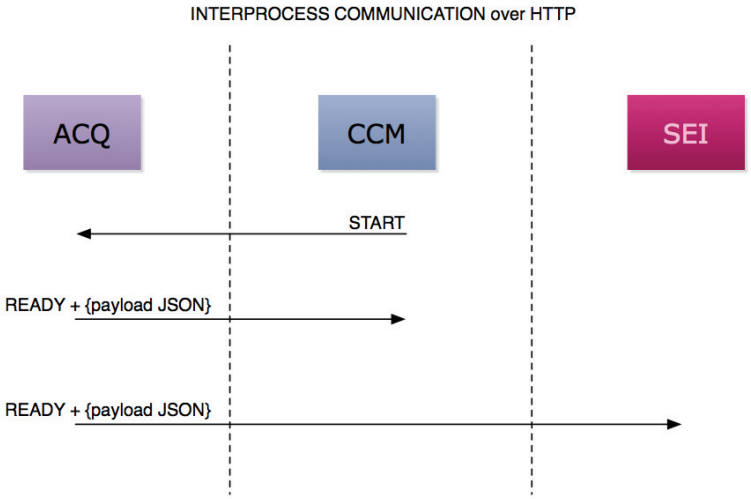


Figure 2.34: Model of the inter process communication between CCM, ACQ and SEI. The READY message is sent at the same time to CCM (for knowledge) and to SEI.

All the turning on and off operations related to the logical/physical equipment that constitutes the detector, are performed by the CCM software module, which communicates with the module that starts the actual data acquisition system. The CCM in addition to managing power supply of the detector and the associated electronics, performs the control and regulation of the temperature and of the thresholds, measuring the voltage levels, and anything else needed for the housekeeping of the detector.

2.5.1 Acquisition and control software module

The goal of this module is to start the data acquisition session from the detector. In particular:

- It checks the ready status by the electronic control module (Adam [1]).
- It starts the acquisition session.
- It collects and stores the data in a specific section of the database.
- It finishes the acquisition section manually and/or after a set time.

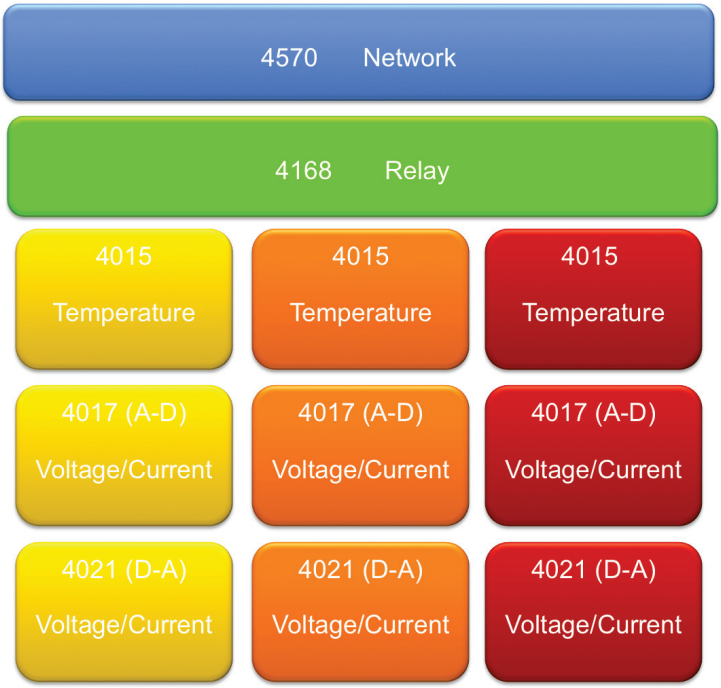


Figure 2.35: The Adam module architecture with communication protocols.

Fig. 2.35 shows the architecture of the Adam module and defines the communication protocols. The communication happens through the exchange of text messages via remote COM port. These messages are described in detail in the official documentation of *Advantech* [63]. Some tests using proprietary software *Advantech* have been made to verify the remote access to Adam systems: to this purpose some configuration commands and request data have been sent and all tests gave the expected outcome. The connection tests phase was followed by a

[1]The Adam [62] is an electronic component, chosen as electronic control module, that contains sensors and actuators and has the purpose of controlling and physically adjust the voltages, currents, and temperatures of the apparatus.

period of study in order to identify an optimal architecture for integrating middleware *WhereX* [64] within the *Muon Portal* system. *WhereX* is an event-based middleware, a software layer interposed between the control software and the physical devices and provides interfaces to and from the outside world, with the possibility of filtering and processing of the input data.
The main components are:

- **Drivers**: those modules able to query an external data source using a specific communication protocol; they perform periodic queries, and generate the Event.
- **Hubs**: as the drivers, they are used to generate events, and also they interface via communication protocols but, unlike the drivers, they wait for external operations.
- **Filters**: modules able to apply operations on the input Events in a timed way.
- **Rules**: are special modules that allow the end user to write (using a simple programming language) operations that have to be performed at each input event.
- **Dispatchers**: are the processing flow terminals and interface to the outside world (databases, file systems, actuators in general, erp systems, dashboards, etc ..) through communication protocols predefined.

As part of the *Muon Portal* project, the *WhereX* middleware is inserted as a data collector from different electronic devices. It was therefore necessary to define a structure that could handle heterogeneous messages/commands coming from different devices, with respect to reception, processing and distribution of them. In Fig. 2.36 is depicted the final architecture that illustrates the communication between the *WhereX* and the Adam module. The same architecture can be easily integrated with the other physical modules.

The access points are two: the Hub that receives requests from external components, such as SEI module, and/or user commands, and forwards them to the second access point, the driver, which cyclically requests to the ADAM module,

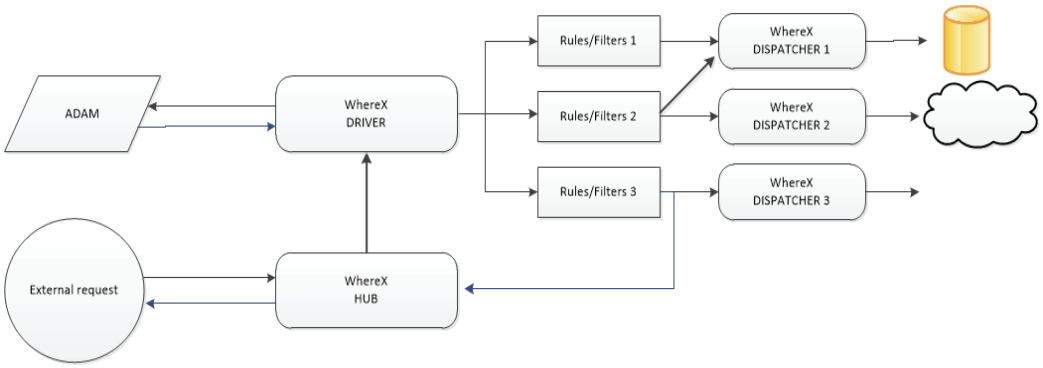


Figure 2.36: Final communication architecture between the *WhereX* and the Adam module.

information regarding operational parameters of the system (i.e. temperature, current and voltage). The Driver will have a structure like in the following figure, Fig. 2.37.

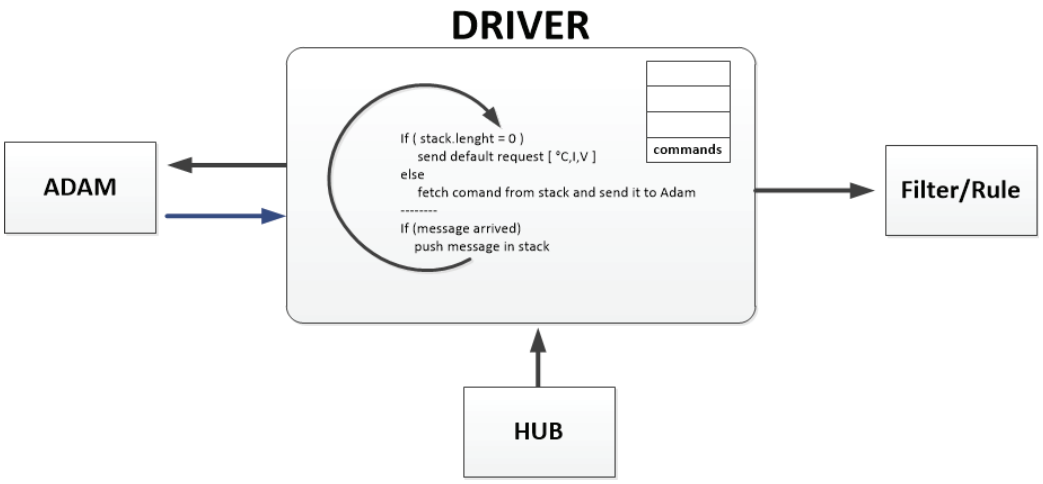


Figure 2.37: Schematized driver structure, the connection with the other components are described.

During each cycle, a command will be picked up by a special buffer and forwarded to the Adam module. The reply is sent in a broadcast to the various Filters/Rules. In this step, the event is analyzed and edited, and finally forwarded to the appropriate Dispatcher which sends the data to the DB, to the dashboard or to the external module that has requested it. In case that the Event submitted to the Rule is an external request, the package comes back to the Hub which will

send the corresponding reply.

All the components are under construction: The Driver, described above, that interfaces with the Adam module; the Hub that waits for external commands and forwards them to the Driver; the Dispatcher that connects to the database MongoDB, the Rules and the Filters. All of the components are written in Java language, according to the middleware *WhereX* specifications.

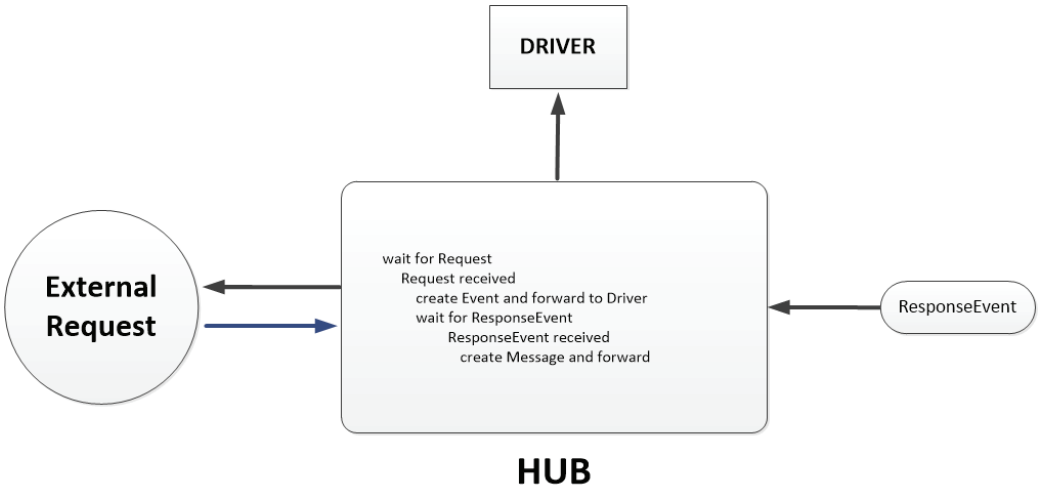


Figure 2.38: Hub functioning schema.

2.5.2 Monitoring software module

The monitoring module deals with the grouping of all the micro-events that were collected by the acquisition module, with the same temporal tag. In this way, the response of both individual strip, and physical planes can be checked in details. By setting the number of strips affected by an event, it is possible to get a distribution in which the total multiplicity should be centered to 8 under conditions of good functioning of the system. The second step has the purpose of extracting how many strips for each physical plane have given signal for each event. This kind of processing allows to obtain the information concerning the number of strips affected for each event on each of 8 planes. In full and efficient functionality operating conditions we expect to have a distribution with a peak in correspondence to the value 1: in fact, the majority of events will affect only one channel for each plan. Only of a small fraction of events is possible to have

multiplicity equal to 2 or greater.

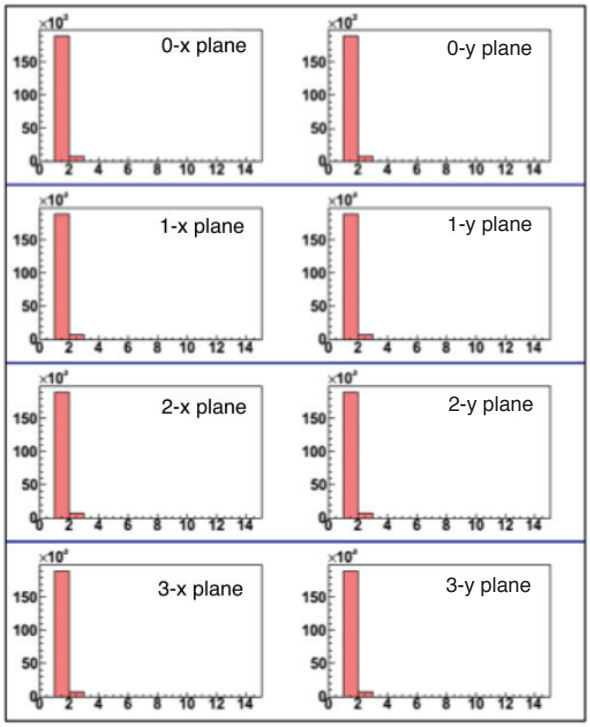


Figure 2.39: Distribution of the relevant channels for each detector plane.

Fig. 2.39 shows the multiplicity of individual planes (always extracted from the simulations), thus corresponding to what is supposed to have in proper operating condition of the detection planes. In the case where the experimental distributions would give substantial deviations compared to those provided by simulations, showing, for example, very high multiplicity or many zero values, this would indicate the presence of a malfunctioning of the planed and suggest the need for a check of the apparatus.

The module is also used to make an uniformity analysis. If the channels response is uniform, then the distribution of the relevant channels in each plane will be uniform, similar to the one obtained from the simulations (Fig. 2.40).

Possible malfunctioning problems, such as the presence of channels that give no response or particularly noisy channels are immediately identified from this type of analysis. The software is able to indicate which channels are showing abnormalities. Another analysis is the one made on the tracks: extracting the

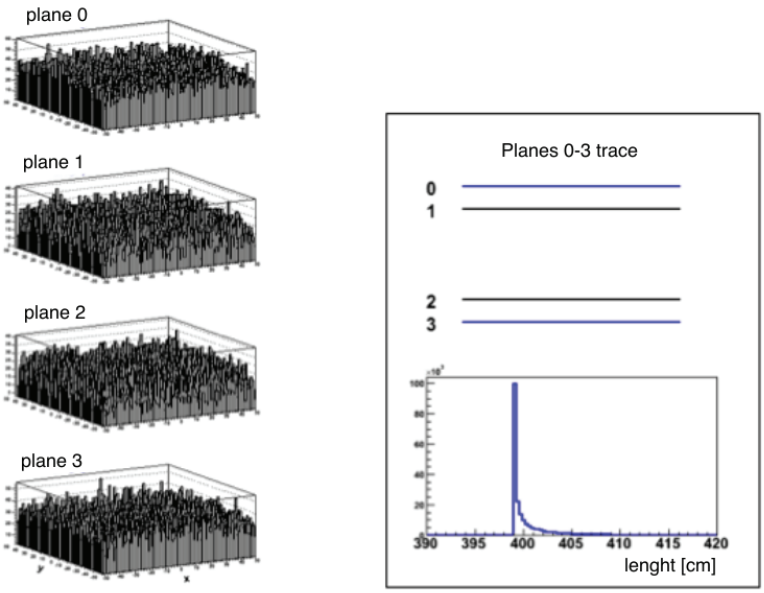


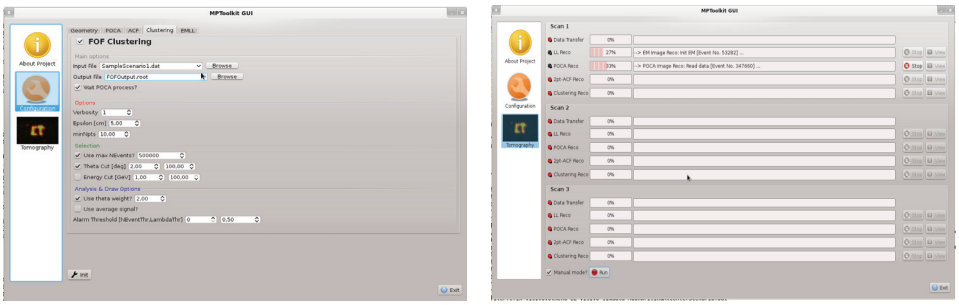
Figure 2.40: Distribution of channels affected in each detection plane and the distribution of the length of the tracks measured between the outer planes.

coordinates (X, Y) on each plane for each logical event with multiplicity 8, it is possible to derive the length of the muon track between the various pairs of planes, as the distance between the points (X, Y, Z) in the corresponding planes. Also the distribution of the lengths must assume a well defined characteristic shape, evaluated according to the geometry of the detector. For the prototype simulated with GEANT, the trace lengths distribution, measured between the outer planes, always appears as shown in Fig. 2.40. The same graph can be realized considering different pairs of planes: depending on the pair of planes taken into account, the minimum value of the length of the track is determined by the distance on which are placed the selected planes. The plot of these distributions will be assessed and displayed, and will be used, as well as in the process of tuning the apparatus, also for a periodic monitoring of the quality of the acquired data and for an event selection according to set quality criteria. The graphics can be optimized so as to make the interpretation as simple as possible to the user and thus facilitate the diagnosis of malfunctions. The monitoring software module is able to produce the plot shown above that are then stored together with the monitoring data.

2.6 Reconstruction, analysis, and imaging software

In order to make an analysis of data coming from the scanner and a 3D visualization of muon tomography, it has been developed a software for the analysis of complex multivariate data using techniques of data exploration and tools with advanced multidimensional graphic. In particular, tools have been developed for advanced 3D, multi-D and immersive graphics that took their origin from medical applications, with aspects of computed tomography. The softwares used as a starting point for the design and development of the imaging software module of the project (SEI), are VisIVO and VisIVOServer [65; 66; 67]. They use advanced graphic libraries for complex analysis (QT and VTK [68; 69]). In particular, the mentioned softwares, use sophisticated methods for the detection and analysis of the extension of clustered regions without specific limitations on the amount of data managed, which are typical of science simulations and application that are very similar to this development objective. A QT-based GUI (see a screenshot in Fig. 2.41) was developed in order to provide an easy and fully customization of the tomography algorithms for non-expert users, and it allows to perform a real-time tomography analysis either in a manual or driven mode.

The output of the GEANT4 simulations has been used to implement several



(a) (b)

Figure 2.41: A view of the Graphical User interface panel available for the online tomography imaging. In (a) a sketch of the panel used to tune the reconstruction algorithm and in (b) the elaboration panel with the status of the elaboration of each algorithm.

algorithms for the reconstruction of the tomographic images. Chapter 4 is focused

on the description of the algorithms implemented within the project, as the POCA (Point Of Closest Approach) and EM-ML (Expectation Maximization- Maximum Likelihood) methods, and it will mainly be focused on the description of the study of new clustering techniques applied to the muon tomography problem. In the following a summarized description of all the implemented methods.
The simplest method to reconstruct the tomographic image is the POCA algorithm. This is a purely geometric algorithm that ignores any underlying physics of scattering and assumes a muon scattered at a single point. By projecting the incoming and outgoing tracks, it is possible to find the points where they came closest and estimate the scattering point as the midpoint of the line between the points of closest approach. A more sophisticated algorithm is the Maximum Likelihood method. It is an iterative algorithm based on the subdivision of the entire volume to be inspected in k voxels (characterized by a density of scattering λ_k). Thanks to iterative procedures to maximize a log-likelihood function, it is possible to find the best set of the parameters λ_k. Even if the computation time is still prohibitive for a real application, a parallel implementation (applied both in the initialization and imaging step of the algorithm) allows a real time application of the method even with a modest number of computing machines. Cluster analysis is successfully used to elaborate data in the *Muon Portal* project. This technique, in fact, meets the need to make the tracks reconstruction and the visualization of the container content be independent from the grid and the 3D-voxels. The presence of a three-dimensional grid indeed limits the automatic object identification process. The problem is relevant in scenarios where the threat to be identified has a comparable size (or even smaller) to those of the single voxel and is located in a position not aligned with the grid. Clustering techniques, working directly on points, help to detect the presence of suspicious items inside the container, acting, as it will be shown, as a filter for a preliminary analysis of the data. The algorithms have been tested and optimized over different simulation scenarios with containers filled with different types of material. In Chapter 4 the results of such simulations and the 3D visualization will be showed.

The output of the reconstruction algorithms is then passed as input to the 3D viewer. The viewer, based on QT+VTK, has been implemented for real-time

visualization i.e. volume renderings, tomographic planes, isosurfaces extraction, cutting plane and so on, within the volume represented. The image represented is available through an interactive and user-friendly graphical interface, with the use of preset views, and allows the image manipulation with rotating and zooming motion interactively using the mouse to locate accurately the areas of interest (see Fig. 2.42).

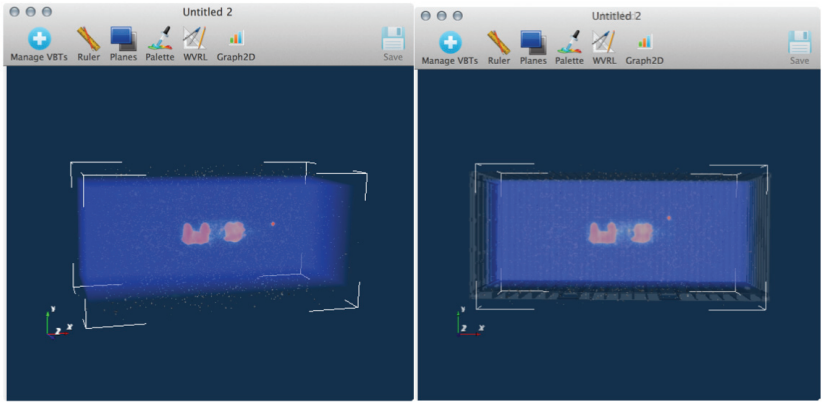


Figure 2.42: Example of the 3D viewer, performing a real time rendering.

It was also included a dynamic system for coloring rendered volumes by a tool of palette/color (see Fig. 2.43). This feature allows to dynamically recolor, on the basis of one or more values, acquired portions of volumes rendered. This makes possible to inspect containers by making transparent the less interesting areas and coloring the important ones.

The software for analysis and visualization is able to identify heavy elements with a precision of the order of dm^3 within the container volume. Through the analysis and visualization module is possible to exactly locate the material with a high atomic number inside the container. The time necessary for the scan was determined by a further study carried out by means of simulations, with different configurations of materials, atomic numbers and dislocations inside the container. The visualization made available from the data analysis and visualization module consists of 3D maps, which depend on the type of reconstruction algorithm used, that might represent the point of minimum approach between the muon track or the density of the scattering of the individual voxel, conditioned by a selection of events according to the relative angle θ between the two tracks. This type of

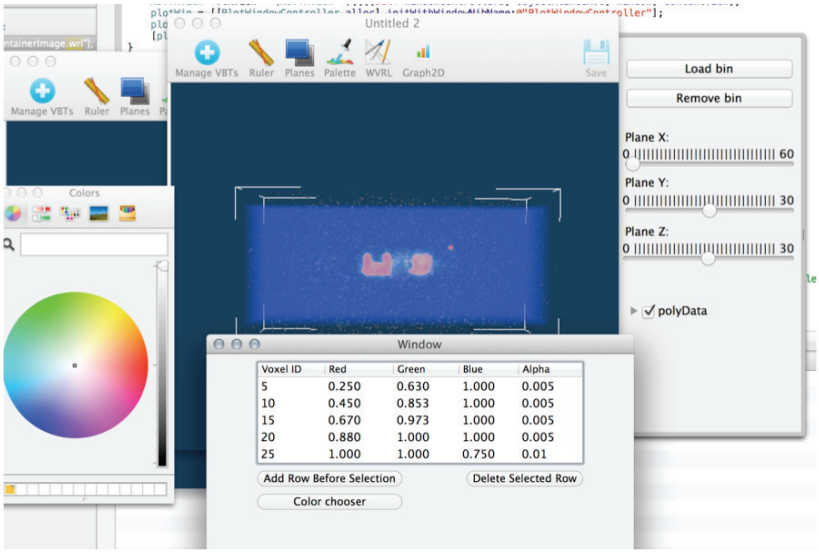


Figure 2.43: The color picker, a dynamic system for coloring rendered volumes by a tool of palette/color.

graphical visualization is the most advanced tool to proceed to the location of the threat with high atomic number. The detailed results will be shown in the following chapters.

2.7 Simulations of the prototype software

A full simulation code written with GEANT4, which takes into account the structure of the detector and the surrounding materials, in order to optimally define the geometry of the detector itself, the effect of the support structure and interposed materials, to create databases of simulated data and to develop the necessary tools for the reconstruction and visualization of tomographic images, has been implemented. Detailed simulations have been carried out in order to understand the performance of our detector and to optimize the algorithms for the reconstruction of the tomographic images. The main elements that were simulated are the following:

- Implementation of a full replica of the detector geometry and of the corresponding material and the structure of the container. The description includes the structure of the detector in its actual size, the segmentation of

the detection modules, for a total of 4800 strip of plastic scintillator, the mechanical structure supporting the detector, the material forming the container, the ground underneath the detector itself.

- Generation of a realistic distribution of cosmic rays (using the CORSIKA code) to simulate the development of extensive atmospheric showers in the atmosphere and generate secondary cosmic muons interacting with the detector of this project. This event generator has been used under different conditions, so as to generate a database of cosmic swarms at various energies and angles of inclination of the primary cosmic. This has required a massive use of computing resources and the preparation of specific tools to manage computer simulations provided.

- Analysis of the calculations made with the event generator CORSIKA, in order to extract the distributions in energy and angle of the secondary particles produced in each swarm and therefore assess the types of particles that interact with the portal.

- Evaluation and reconstruction of hits and clusters products in the interaction of these secondary particles with the detector, including those due to electromagnetic showers. Such multiplicity distributions are essential to understand the type of events that can be measured with the detector and to develop reconstruction algorithms.

- Evaluation of the energy and angular resolution relative to the muon, particularly in condition of containers "empty", which represents a kind of background for the actual measurement.

- Implementation of tracking algorithms to reconstruct the tracks in the detector and evaluate the deviation of each track.

- Production database of simulated events, taking into account the structure of the detector, according to various possible scenarios: empty container, container with light elements, materials with intermediate and heavy materials.

- Analysis of the data produced by macros written in ROOT environment.

Some of the simulation activities as, for example, the computation of the acceptance of the detector and the implementation of tracking algorithms will be thoroughly described in the next chapter.

CHAPTER 3

Studies on the detector performances and simulations of the apparatus

A full simulation code written in C++ and using ROOT libraries, which takes into account the structure and the design of the detector, in order to optimally define the geometry of the detector itself and to develop the necessary tools for the reconstruction and visualization of tomographic images, has been implemented. Detailed simulations have been carried out in order to understand the performance of our detector and to optimize the algorithms for the reconstruction of the tomographic images. In Section 3.1 the study on the geometrical acceptance of the detector is presented, in 3.2 an introduction to the *Extensive Air Showers* (EAS) phenomenon is given, together with a first approach to the track reconstruction problem with the Least Square method. Finally Section 3.3 presents the procedure used to solve the high multiplicity tracks reconstruction issue together with some simulation results on the performance of the *Muon Portal* project detector.

3.1 Study on the geometric acceptance and on the reconstruction ability of the detector

A preliminary study phase on the detector was necessary to investigate its intrinsic resolution, its geometrical acceptance and the event reconstruction ability on the single event/track and in presence of high multiplicity events. In order to

study the detector characteristics from the geometrical point of view, a simulation software, which reproduces the main geometrical characteristics of the detector, has been build using the ROOT environment.

3.1.1 Study on the Acceptance

A first study done on the detector has consisted of determining the geometrical acceptance $\mathcal{A}$, in order to define the reconstruction capacity of the detector and to choose the optimal distance at which to place the tracking planes. The geometrical acceptance, in fact, depends on the size of the detector, in particular on the extension of the planes and on the distance D between the first and the last plane. In order to calculate $\mathcal{A}$, the geometrical structure of the detector was simulated with planes of size $300\,cm \times 600\,cm$ and varying the distance between the external planes. The program, written in C++, and using ROOT libraries, simulates the generation of particles on an imaginary plane placed at the center of the detector, at $z = D/2$ and the origin of the reference system is placed at the center of this plane, as it can be seen in Fig. 3.1.

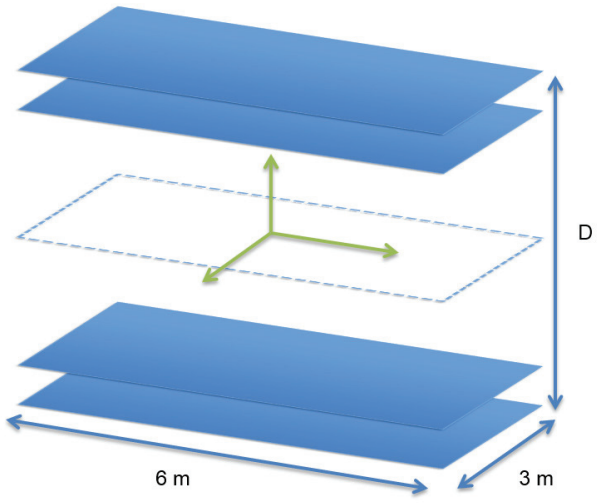


Figure 3.1: Imaginary plane at the center of the apparatus, used to generate a flux of particles in order to evaluate the acceptance $\mathcal{A}$.

It should be remembered that the solid angle element is defined by $d\Omega = sin(\theta)d\theta d\phi = d(cos(\theta))d\phi$, where θ is the zenith angle, measured from the vertical direction and ϕ is the azimuthal angle. Therefore the particle generation has been

made starting from uniform distributions in ϕ and $cos(\theta)$. Initially an isotropic particle distribution was generated with $cos\theta \in [0,1]$ and $\phi \in [0,2\pi]$ and uniformly distributed on the X-Y plane, with $x \in [-300,300]$ and $y \in [-150,150]$. The number of generated particles was $N_{gen} = 10^7$. The program generates randomly particles and the coordinates of the generation point on the imaginary plane (x_0, y_0, z_0) with $x_0 \in [-300,300]$, $y_0 \in [-150,1500]$, and $z_0 = D/2$, where D is distance between the first and the last plane. Once θ and ϕ are also generated, according to the different chosen distributions, the intersection points of the particles on the external planes of the detector are geometrically individuated. The particle is marked as accepted if both the intersection points are within the area of the external planes, otherwise is considered as rejected.

The geometrical acceptance $\mathcal{A}$ is defined, per each angular interval (bin), as the rate between the number of muons reconstructed (N_{rec}) and the number of muons generated (N_{gen}), divided by the orthogonal surface where the impact points are generated, multiplied by the solid angle correspondent to the considered angular interval ($d\Omega$). Therefore it is measured in $m^2\,sr$ and is calculated as:

$$\mathcal{A} = \frac{N_{rec}}{N_{gen}} \times S \times d\Omega \quad \left[m^2\,sr\right] \tag{3.1.1}$$

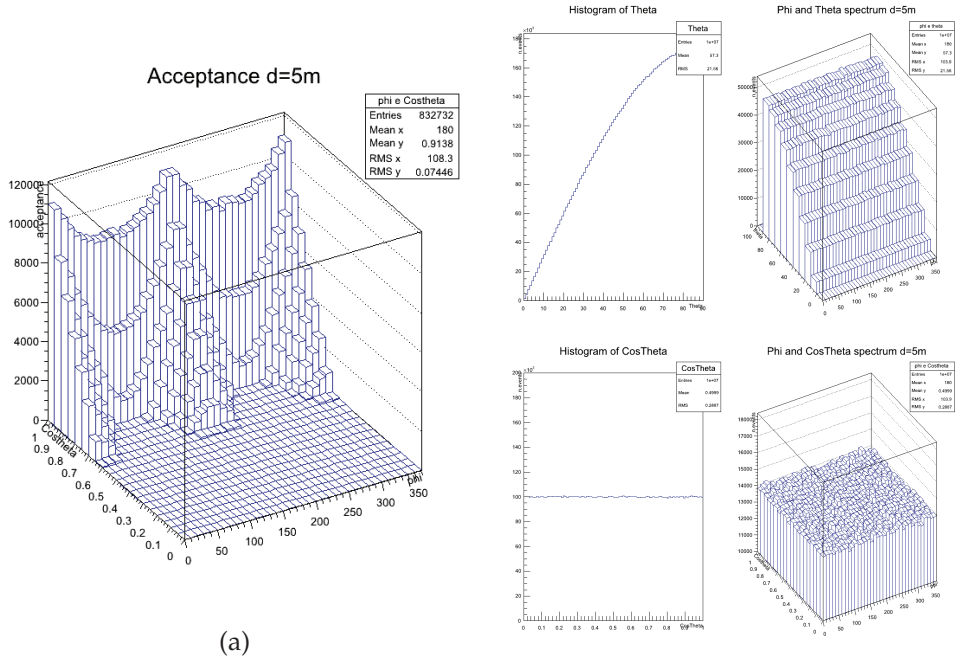
where N_{rec} is the number of reconstructed tracks, and N_{gen} is the total number of generated tracks. In table 3.1 the acceptance values calculated varying the distance between the planes, are shown. From the results shown, it is clear that a good compromise between the efficiency and the mechanical issues related to the detector design, would be to have a total distance between the planes that lies between 5 and 6 meters in total.

In Fig. 3.2 the acceptance of the apparatus, at the more conservative distance $D = 5m$ as a function of $cos\theta$ and ϕ, is plotted when the distribution of the particles is considered isotropic. In (b) the histogram of θ, and the ϕ-θ spectrum and the histogram of $cos\,\theta$ and the ϕ-$cos\,\theta$ spectrum, for $D = 5m$ are showed.

Actually the angular distribution of cosmic muons at sea level is uniform in the azimuthal angle ϕ, but is proportional to $cos^2(\theta)$. Therefore, in order to calculate the expected muon rate (number per unit time) on the detector, a number $N_{gen} = 10^7$ of muon particles were again generated with these angular distributions. Fig. 3.3 is related to the more realistic particle generation, which follows the distribution

Distance [m]	Acceptance [m^2 sr]
6	7.01
5	9.89
4	12.9
3	19
2	29.56
1	50.59

Table 3.1: Values of the geometric acceptance in function of the distance between external planes.

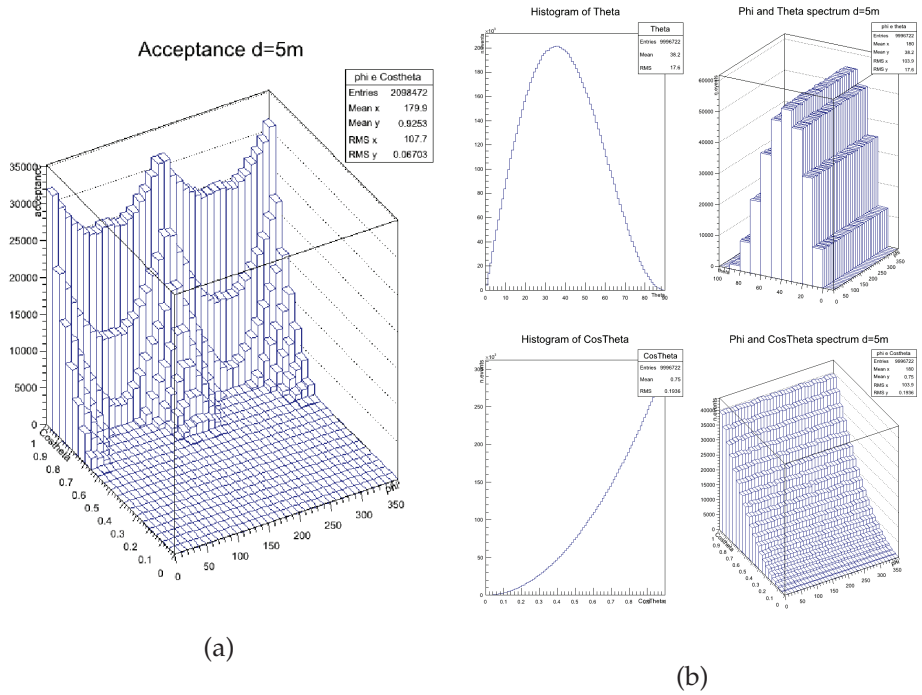


(a)

(b)

Figure 3.2: Acceptance of the apparatus, as a function of $cos\theta$ and ϕ, (a). Histogram of θ, and the ϕ-θ spectrum for a distance $D = 5m$. Histogram of $cos\,\theta$ and the ϕ-$cos\,\theta$ spectrum, for $D = 5m$ (b).

of the secondary cosmic rays at the sea level. In this case ϕ varies uniformly in the $[0, 2\pi]$ range, while θ is varying according to a $cos^2(\theta)$ distribution. The acceptance of the apparatus at the chosen distance $D = 5m$ as a function of $cos\theta$ and ϕ is then



(a) (b)

Figure 3.3: Acceptance with a distance $D = 5m$ and with a particles distribution uniform in ϕ and with θ varying like $cos^2(\theta)$ (b).

plotted. In (b) the histogram of θ, and the ϕ and θ spectrum and the histogram of $cos\,\theta$ and the ϕ and $cos\,\theta$ spectrum, for $D = 5m$ are showed.

The expected detector acceptance $\mathcal{A}$ to the flux of cosmic ray muons has been so evaluated from the simulations, yielding $\mathcal{A}$=10 m^{2}sr, corresponding to a number of expected events of $\sim 2\times10^5$ for a scanning time of Δt=5 minutes and a standard cosmic ray flux ϕ_μ= 1 m^{-2}s^{-1} integrated over the solid angle. The angular accuracy for muon track reconstruction has been found of the order of 0.25° from toy simulations in which only the impact of the position resolution is considered. It increases at $\sim$0.5° in detailed GEANT4 simulations including also multiple scattering effects.

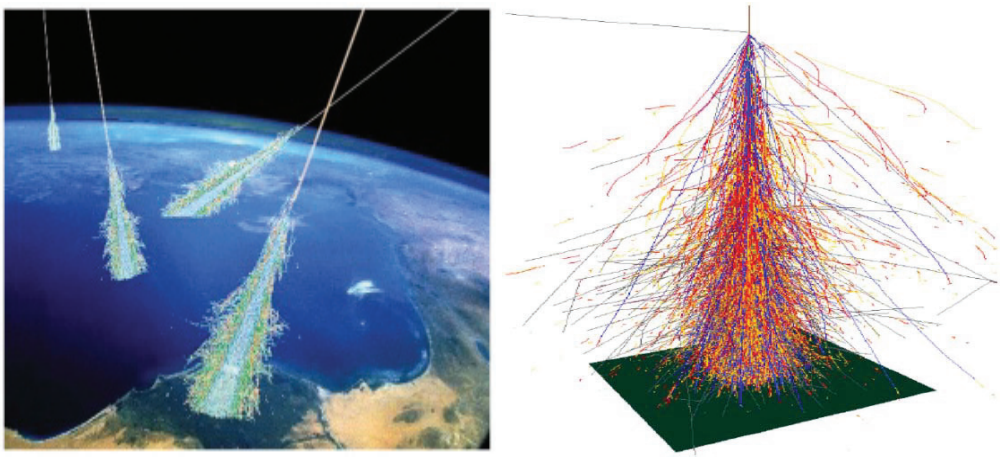
The precision on the determination of the scattering data, scattering angle and lateral displacement, which are relevant for tomography imaging studies, needs to be estimated too. The analysis yields a scattering angle uncertainty of $\sim$0.7° and a lateral displacement uncertainty of the order of 2 cm. This imposes a limit on the minimum size of the threat objects that can be identified with reasonable accuracy

inside the container volume and within reasonable scanning times, typically not smaller than 5 cm. A dummy mechanical structure to be inserted between the intermediate detector planes to emulate a real container volume, has also been designed.

3.2 The tracking issue at low and high multiplicity

A second phase of the study involved the reconstruction ability of the detector itself (without considering the presence of the container, or of radioactive material placed at the center of planes) through the application of reconstruction and fit algorithms. As well known, the muons that arrive to our detector, are part of the secondary cosmic radiation. The production process of the secondaries begins with the collision of the primary cosmic ray with a nucleus near the top of the atmosphere. Typically the first collision produces secondary particles (from a few tens at low primary energy to millions for the most energetic showers), the majority of which are pi-mesons (generally called pions), that through various collisions lead to *hadron particle showers* and to *electromagnetic showers*. The secondaries, in fact, are spread over significant areas at the observational level. This is due to the transverse momentum acquired by secondary particles at production and to the scattering which the shower electrons, in particular, undergo through interactions with the material of the atmosphere. This phenomenon, depicted in Fig. 3.4 , is called Extensive Air Shower (EAS), which indicates the nearly-simultaneous arrival of many particles over a large area on the Earth surface. At 10^{15} eV around 10^6 particles cover approximately $10^4\, m^2$ while at 10^{20} eV some 10^{11} particles are spread over about $10\, km^2$ [70]. The french physicist Pierre Victor Auger discovered this phenomenon in the 1930's [71]. As an EAS develops into the atmosphere, more and more particles are produced. Two important properties of the shower that have to be mentioned are the number of particles and the slant depth. At its maximum, an EAS typically contains $\sim 1 - 1.6$ particles for every 1GeV (10^9 eV) of energy carried by the primary cosmic ray, and secondly, the average slant depth at which the shower maximum occurs, varies with the energy of the primary cosmic ray. Given the standard cosmic rate flux $\phi_\mu = 1 m^{-2} s^{-1}$ at sea level, integrated over the solid angle and considering the overall acceptance $\mathcal{A}$=10 m^{2}sr, if for Δt=5

minutes we expect to have a number of events of $\sim 2 \times 10^5$, this means that we expect a rate of 5×10^2 Hz of accepted events on the detector. For this reasons we are interested in studying also the response of the detector in the presence of events at high rate and high multiplicity.



(a) (b)

Figure 3.4: Extensive air showers generated by the interaction of a primary cosmic ray with the atmosphere.

This study has been made through a software that simulates the detector and that has the following characteristics:

- 4 logical X-Y-Z planes, size $600\,cm \times 300\,cm \times 2\,cm$ each, numbered from top to bottom as: Plane 1, Plane 2, Plane 3 and Plane 4.
- Each plane is considered to be segmented in strips of $1\,cm \times 1\,cm$ section, which are positioned following opportune directions in order to have the x-y coordinates.
- The distance between the inner planes 2-3 equal to $d_{2,3} = 300\,cm$, and between the outer planes 1-2 and 3-4 equal to $d_{1,2} = d_{3,4} = 100\,cm$. So the total distance between the first plane and the last one is $d_{1,4} = 500\,cm$.
- Energy loss effects are not taken into account.
- The trigger of the detector happens if the track crosses all 4 planes.

In order to determine the type of tracks accepted by our detector, a first simulation step was to generate the particles, according to different types of distribution and check how many of them are accepted and are therefore potentially reconstructed from the tracker. Considering what stated in the Chapter 1, we expect that the particle follows a trajectory almost rectilinear, so that its track can be considered a straight line, and that the speed of the muon is nearly equal to that of the light in the vacuum. The track is so completely identified by the position $\vec{p} = [x_0, y_0, z_0]$ of the muon at the references time t_0 and by the direction cosines:

$$\vec{d} = [cos\phi\, sin\theta,\ sin\phi\, sin\theta,\ cos\theta] \tag{3.2.1}$$

where ϕ and θ in a Cartesian reference system (x, y, z) are respectively the azimuth and the angle to the zenith. For this reason we want to perform a fit of simulated experimental data with a 3D straight line. The procedure used is the following:

- Generates N_{gen} tracks $genTr_k$ with $k \in [0, N_{gen} - 1]$ uniform in (x, y) and isotropic in (θ, ϕ). The simulation was performed with a number of generated tracks $N_{gen} = 10^5$. The tracks have been generated by extracting (x, y) pairs always with a uniform distribution, i.e. $\{(x, y)\ | x \in [-300, 300],\ y \in [-150, 150]\}$ and the first time considering an isotropic distribution in (θ, ϕ) while the second time with θ varying according to a $sin\theta cos^2\theta$ distribution and with ϕ varying uniformly in $[0, 2\pi]$. In Fig. 3.5 plots of the generated points on the first plane and of the impact points of the generated tracks on the following planes are showed. It should be noticed that the density of the points is lower on the middle planes than on the external ones. This is probably due to geometrical reasons, and to the fact that only a nearly vertical track could trigger the detector crossing all the four planes near the borders.

- Iteratively and for all the tracks, by means of the direction cosines, verifies that the i-th track crosses the 4 planes, and therefore calculates the intersection points with every plane $(p_{i,1}, p_{i,2}, p_{i,3}, p_{i,4})$, with $i \in [0, N_{accepted} - 1]$.

- Introduces a Gaussian spread, in order to simulate measurements errors, on the coordinates of the impact points for each plane. In our case the uncertainty introduced by the detector resolution depends on the segmentation

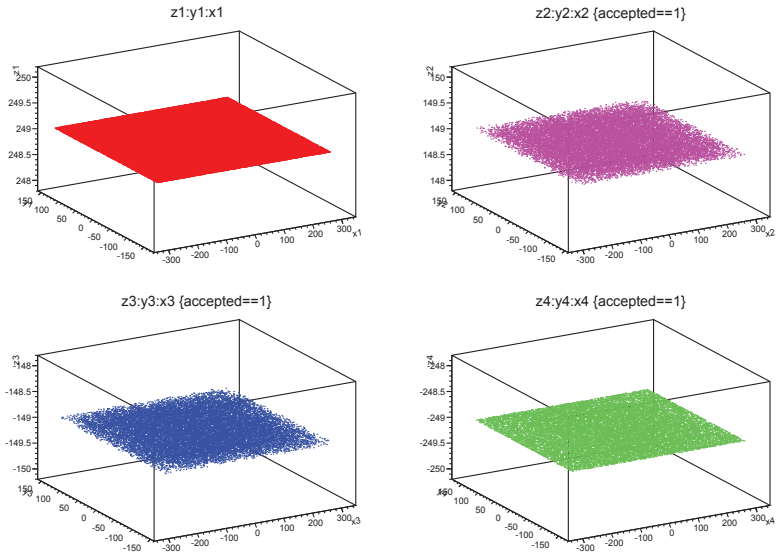


Figure 3.5: Plots of the generated points on the first plane and of the impact points of the generated tracks on the following planes.

of the same into $1cm \times 1cm \times 3m$ strips. The Gaussian spread introduced will be then of $\sigma_x = \sigma_y = \frac{1}{\sqrt{12}} \simeq 3mm$ and $\sigma_z = \frac{2}{\sqrt{12}} \simeq 6mm$. In this way we obtain the hits $(h_{i,1}, h_{i,2}, h_{i,3}, h_{i,4})$ which can be considered as if they were experimental data.

- Starting from the hits, reconstructs, through the least squares method, the fitted 3D straight line $fitTr_i$.
- Compares the fitted line $fitTr_i$ with the generated one $genTr_i$, calculating the θRel_i.

In the following paragraph the linear fitting of 3D Points procedure is explained.

3.2.1 Linear Fitting of 3D Points of Form $(x, y, f(x, y))$

The simplest and most commonly applied form of linear regression which provides a solution to the problem of finding the best fitting straight line through a set of points, is the linear least squares fitting technique [1]. Usually, the points are given on a 2D space and the analytical solution is well-known in this case. More complicated fitting problems are solved in practice by numerical methods

leading to excellent sub-optimal solutions [2-3]. On a more theoretical point of view, the analytical solution of the 3-D Linear Least Squares Fitting can be obtained, being the developments much more arduous than in the 2D case. We evaluate different procedures for the 3D fitting and the chosen one consists in splitting the fit into two 2D fits. Starting from the set of the available hits $(h_{i,1}, h_{i,2}, h_{i,3}, h_{i,4})$ with $i \in [0, N_{accepted} - 1]$ and where for example, $h_{i,1} = (x_{i1}, y_{i1}, z_{i1})$, we consider the orthogonal projections of these 3D points on the two planes x-z and y-z, so as to obtain the pairs: (x_{i1}, z_{i1}) and (y_{i1}, z_{i1}). In this way we will have to deal with two linear independent 2D fits. Let us assume that for the x-z plane is $z = Ax + B$ while in the case of the plane y-z is $z = A^{'}y + B^{'}$. The four, so calculated parameters $(A, B, A^{'}, B^{'})$, are then combined to find the fitted 3D line.
Let us remanding the least squares fit by a 2D line working principle.

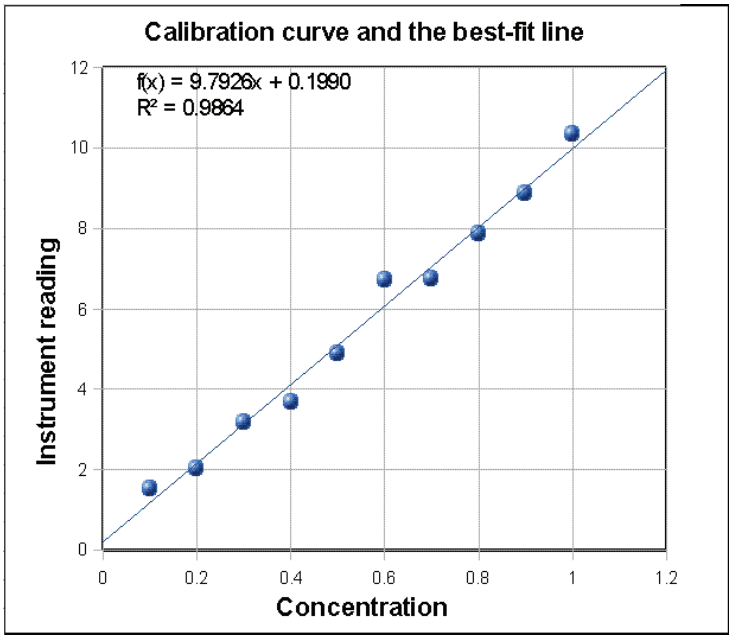


Figure 3.6: Example of a 2D best fit: a calibration curve and its best-fit line.

Usually, the data represents measurements where the y-component is assumed to be functionally dependent on the x-component. Given a set of samples $\{(x_i, y_i)\}_{i=1}^{m}$, the goal is to determine A and B so that the line $y = Ax + B$, best fits the samples. This means that the sum of the squared errors between the y_i and the line values $Ax_i + B$ has to be minimized. It must be underlined that the

error is measured only in the y-direction. In fact, amongst the method assumptions the are that the measures are independent between each others, and that the $\sigma_y/y >> \sigma_x/x$. The Fig. 3.6 is showing an example of 2D best-fit.

Let us define a nonnegative function $E(A,B) = \sum_{i=1}^{m} [(Ax_i + B) - y_i]^2$. Its graph is a paraboloid and its vertex occurs when the gradient satisfies $\nabla E = (0,0)$. This leads to a system of two linear equations in A and B which can be easily solved. In particular,

$$(0,0) = \nabla E = 2 \times \sum_{i=1}^{m} [(Ax_i + B) - y_i] (x_i, 1) \tag{3.2.2}$$

and so

$$\begin{bmatrix} \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} 1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} y_i \end{bmatrix} \tag{3.2.3}$$

The solution provides the least squares solution $y = Ax + B$. In Fig. 3.7 two examples of 2D best-fit results in the x-z and y-z planes, which are considering 4 hits on our detector (a,b). It can be noticed that the points are not always lying on a straight line (c,d). In what follows the results of the application of the described method is showed, together with the extensive results.

3.2.2 Detector tracking reconstruction ability with single tracks

We are interested in studying the reconstruction ability of the whole detector, as a tracker, given its intrinsic resolution and considering the totality of 4 planes, starting from events containing a single track.

Fig. 3.8 shows the distribution of the angles θ and ϕ of the generated tracks (a) and of those accepted by the detector (b).

It should be noted the distributions of θ which has a peak at 38° for the generated tracks and at $\sim 20°$ for the reconstructed ones, as expected. The detector accepts tracks with angles θ up to 50°, while greater angles are rejected. The trend of accepted ϕ refers to the shape of the planes, which is rectangular.

Fig. 3.9 shows the results of the 3D fit, realized with the two 2D Least Squares Method procedures (described before). The trend of the graphs shows that the reconstruction of the angles is faithful, giving an overall idea. In order to have an accurate assessment it is necessary to look at the θ_{rel} and $\Delta\theta$ and at the $\Delta\phi$, i.e. the

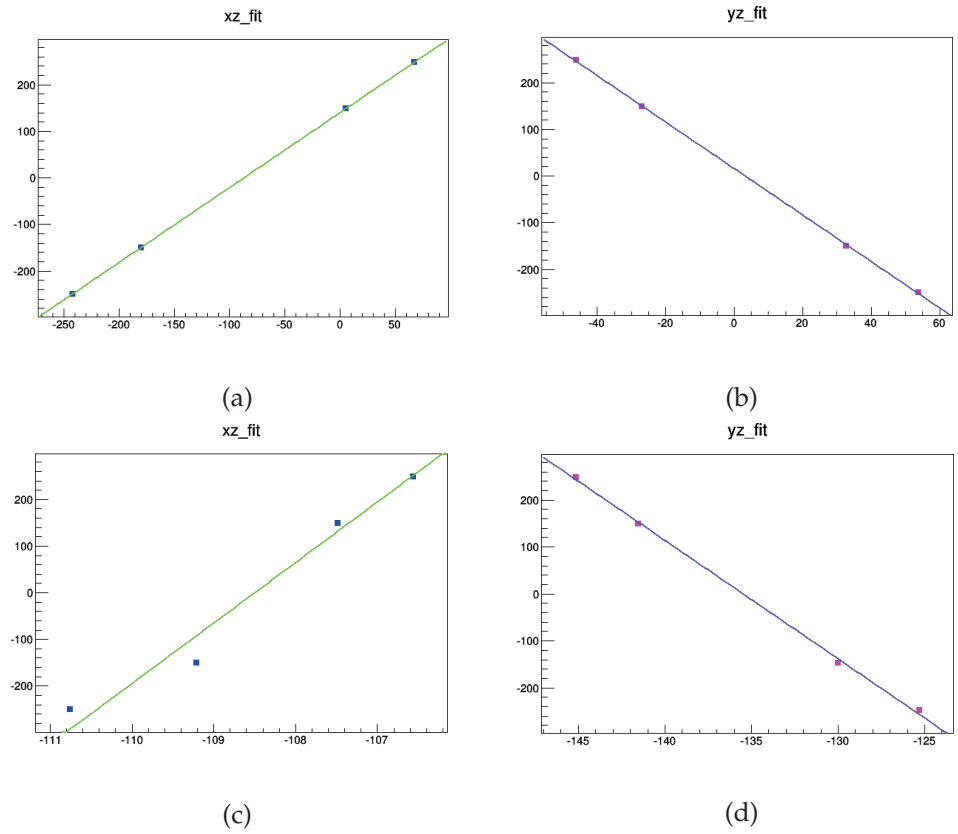


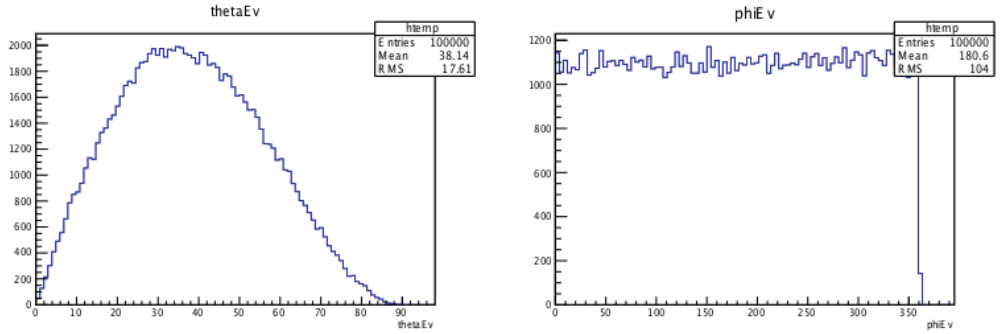
Figure 3.7: Two examples of 2D best-fit results in the x-z and y-z planes, applied to 4 simulated hits on our detector (a,b). The points are not always lying on a straight line (c,d).

difference between the angles generated and reconstructed on an event-by-event basis. The θ_{rel} is calculated with the formula:

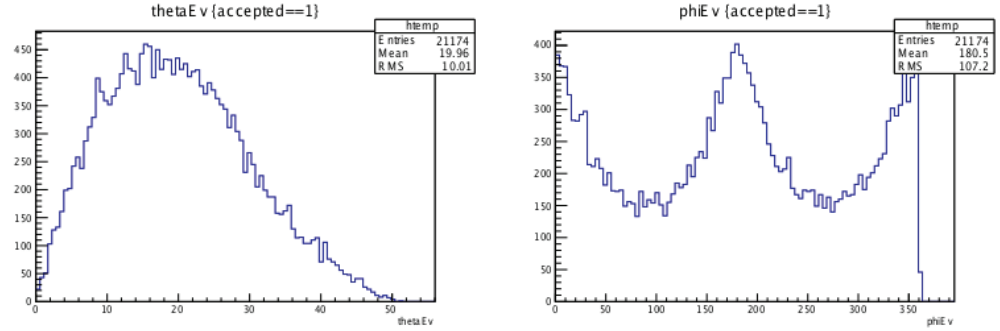
$$\theta_{rel} = \arccos(\cos(\theta_1)\ \cos(\theta_2)) + \sin(\theta_1)\ \sin(\theta_2)\ \cos(\phi_1 - \phi_2) \tag{3.2.4}$$

where (θ_1, ϕ_1) and (θ_2, ϕ_2) are respectively the zenith and azimuthal angles of the generated track and of the fitted one.
In Fig. 3.10 we can see the distribution of θ_{rel} considering that a Gaussian spread of $3mm$ in x and y, calculated as $1cm/\sqrt{12}$, and of $5mm$ in the z direction, calculated as $2cm/\sqrt{12}$, has been applied. A base 10 logarithmic scale is used for the y- axis and a linear scale for the x-axis with values in degree ($^\circ$).

In Fig. 3.11 the same distribution of θ_{rel}, always considering a Gaussian spread, but this time filtering the values higher than $\theta_{rel} > 5^\circ$. The y-values are always in



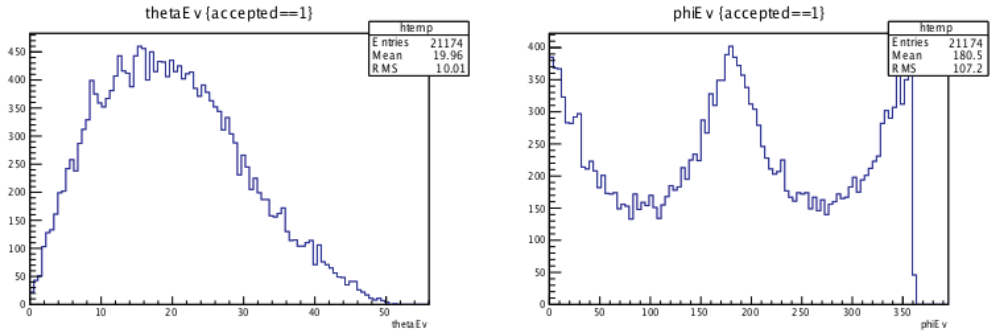
(a)



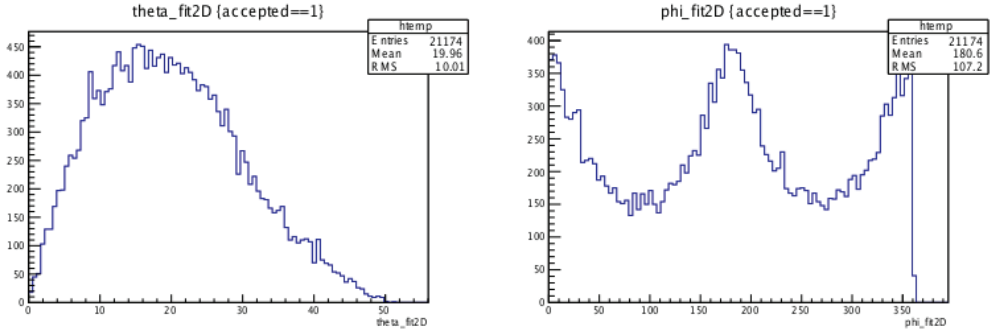
(b)

Figure 3.8: θ and ϕ distribution for 10^5 Events: Generated Tracks (a), and Accepted Tracks (b).

a log-scale. In Fig. 3.12 the punctual difference $\Delta\theta$ between θ angles of the original tracks and of the reconstructed ones is plotted (a) and in a log scale in (b). In (a) it is possible to notice a distribution centered nearly at zero (mean: 0,001 209), as expected, and RMS: 0,047 43. We note, however, the presence in the graph of some values that deviate from zero, this is clear if we analyze the histogram in (b). In Fig. 3.13 the punctual difference between ϕ angles of the original tracks and the reconstructed ones is plotted. Also in this case it is visible a distribution centered at zero (average: 0,032 54), as it is expected, but there is also a very high RMS=7,832 5. If we apply here also a filter $\theta > 10°$, we obtain in Fig. 3.13 a decrease in the RMS = 3,858, that still maintains a very high value.



(a)



(b)

Figure 3.9: Results of the 3D fit with the projection procedure. 10^5 Events simulated and reconstructed with the Least Square Method. In (a) θ and ϕ distributions of the generated tracks, in (b) the ones of the fitted tracks.

3.3 Tracking reconstruction ability with multiple tracks

In the case of events with multiple simultaneous tracks, which always correspond to a single event (a single swarm generated in the atmosphere by a single primary), we are interested to reconstruct ideally all tracks. We want to assess how many simultaneous tracks can be reconstructed and how well. Let us remember that the single event is recorded by the detector if it crosses all the 4 planes. Therefore in the presence of an event we will then have elaborate at least four hits, one on each plane. The trajectory of the muon is assumed to be linear, especially in absence of a magnetic field. In the case of a single track the reconstruction is trivial, but when the number of tracks increases, the use of an interpolation method becomes necessary.

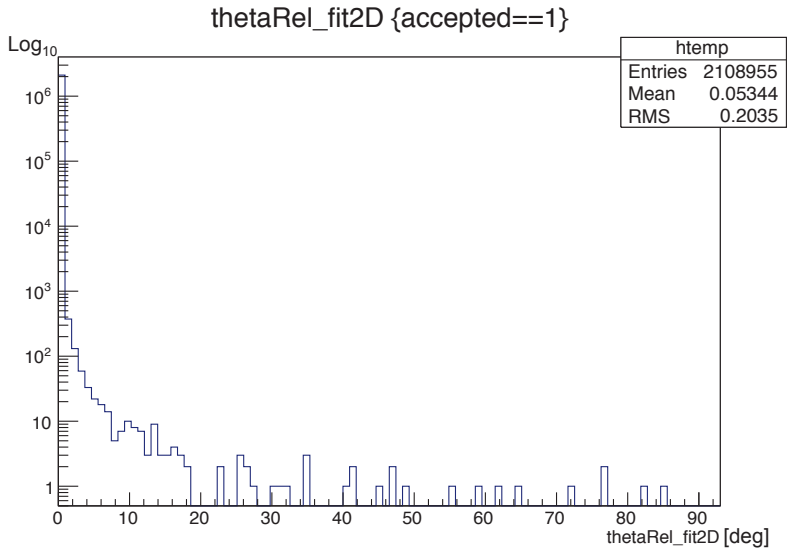


Figure 3.10: θ_{rel} between the generated track and the fitted one, applying a spread of 3mm to the x and y coordinates and of 5mm to the z one, simulation with $10^7 Events$.

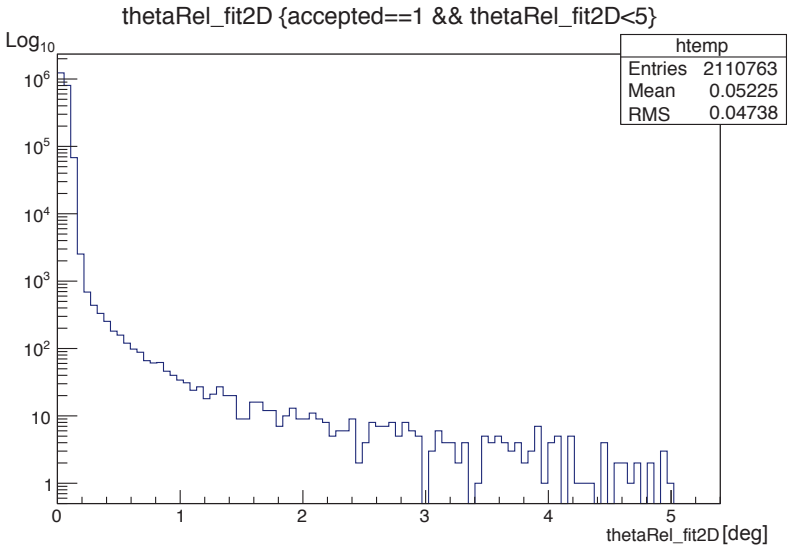
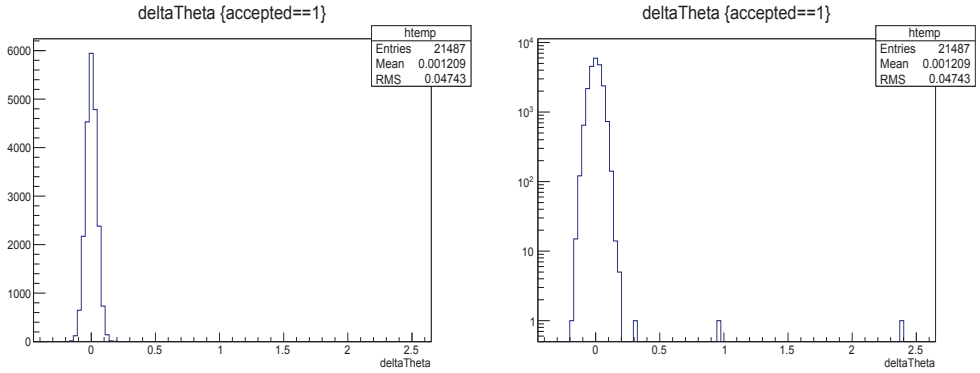


Figure 3.11: θ_{rel} between the generated track and the fitted one, applying a spread of 3mm to the x and y coordinates and of 5mm to the z one, filtered for values $< 5°$, simulation with $10^7 Events$.

The problem is always to fit the hits in order to determine the best line that connects the experimental points (the hits on the planes), but, this time, in presence



(a) (b)

Figure 3.12: Histograms showing the $\Delta\theta$, punctual difference between the generated θ and the reconstructed one (a), and with log scale on the y-axis (b).

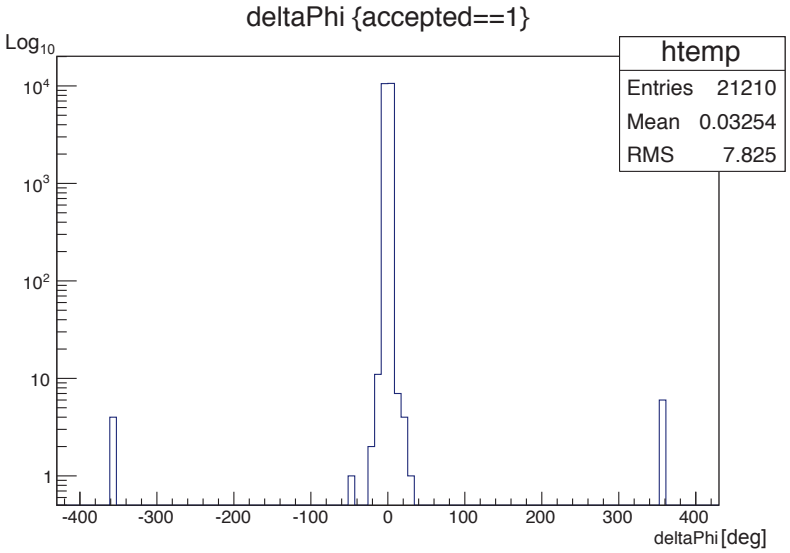
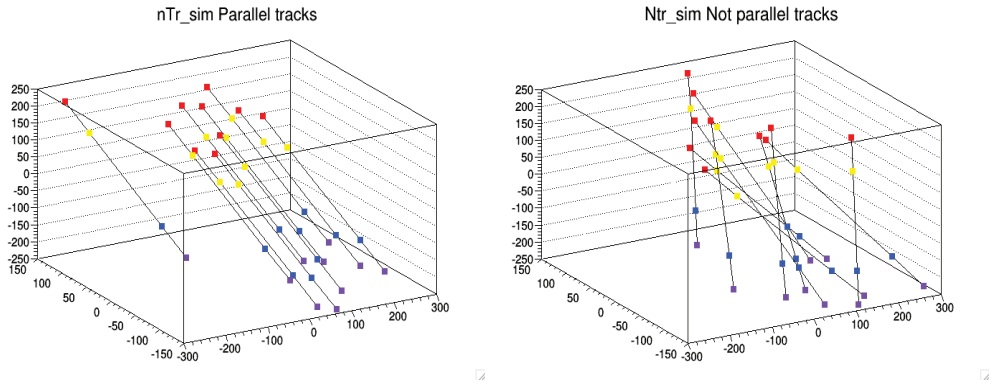


Figure 3.13: Histograms showing the $\Delta\phi$, punctual difference between the generated ϕ and the reconstructed one.

of multiple hits per plane that have to be combined in the correct order. In Fig. 3.14 a visualization showing 10 simultaneous generated tracks, within the volume of the scanner. parallel between each others (a) and with a gaussian spread of 20° in (b). In the graph the hits on different planes are indicated by different colors: red, yellow, blue and violet, respectively for the first, second, third and fourth plane. The simplified procedure that has been used to make the fit of all combinations,



(a) (b)

Figure 3.14: Visualization showing 10 simultaneous generated tracks, within the volume of the scanner, parallel between each others (a) and with a gaussian spread of 20° in (b).

varying the number of simultaneous tracks, is described below. Considering to have n accepted simultaneous tracks, for each plane (1-4) we will have a n-tuple of hits: $(h_{11}, h_{12}, ..., h_{1n})$, $(h_{21}, h_{22}, ..., h_{2n})$, $(h_{31}, h_{32}, ..., h_{3n})$, $(h_{41}, h_{42}, ..., h_{4n})$ (see Fig. 3.15 (a)).

The goal is to find the optimal combination between the hits in order to reconstruct the original tracks. To implement the matching between the different hits on the 4 planes, we proceed in a simplified way considering at the first stage only the hits of the first and the last planes. Therefore, the combination of each *i-th* hit of the first plane with every *j-th* hit of the last one, for n simultaneous tracks, leads to $n \times n$ possible combinations, or candidate tracks, instead of n^m combinations (being m the number of planes, 4 in the case of our detector). As shown in see Fig. 3.15 (b), for each *i-j* combination (h_{1i}, h_{4j}) with $(i, j) \in [1, n]$, we draw the line between the two points, which is the candidate track r_{ij}. Then the hits belonging to the planes in the middle have to be chosen. This is done by calculating the intersection points of the candidate track with the middle planes $(p2_{ij}, p3_{ij})$ (Fig. 3.15 (c,d)) and by evaluating if in a neighborhood of the intersection points there are some hits. In particular the neighborhood is chosen to be the circle centered on the intersection point, with radius that should be calculated as $r = 3 * \sigma$, where σ is the spatial resolution given from $\sigma = \sqrt{(\sigma_x^2 + \sigma_y^2)}$, being σ_x e σ_y the resolutions

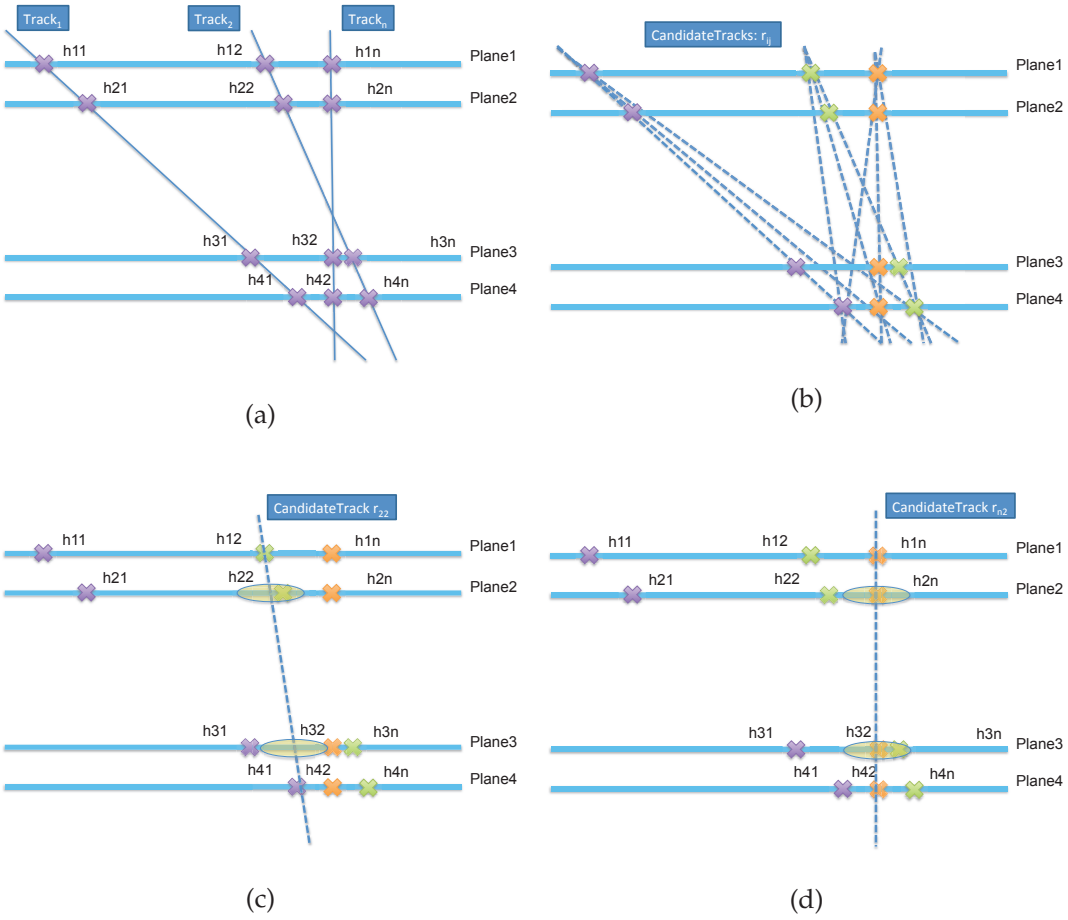


Figure 3.15: Explanatory diagram of the method used to couple hits and discard some in order to perform the best-fit procedure. In (a) the simultaneous tracks, in (b) the candidate tracks for the fit procedure are generated combining all the hits of the first and the last plane, leading to $n \cdot n$ combinations. Choice of the hits in the intermediate planes (c,d) evaluating how many hits are in the circle of radius $r = 3\sigma = 1.2cm$ centered on the intersection point, and consequently giving rise to candidate tracks to be fitted.

in x and in y. In our detector, which is a segmented one with a $1\,cm \times 1\,cm$ section in x-y: $\sigma_x = \sigma_y = 1/\sqrt{12}\,cm = 0.29cm$, so that $\sigma = 0.4\,cm$, which means $r = 1.2\,cm$. If there are no points in that neighborhood (c), then the algorithm proceed to examine the next hits combination between the two external planes. If there is only one matching hit for each of the two intermediate planes, then we have found the four hits associated with the combination, and they can be fitted. Instead,

when there are more than one matching hits, all the possible combinations are considered as candidate tracks (d). The choice between them is made using the chi-square test, which is described in the following. We expect that with the tracks initially assumed to be parallel, the reconstruction is made correctly unless with an unrealistically high number of tracks. With not parallel tracks, the reconstruction could be difficult even with a reduced number of tracks.

3.3.1 Metrics used and analysis of the results

The goodness of fit of a statistical model is used to define how well it fits a set of observations. Measures of goodness of fit are typically used to summarize the discrepancy between observed values and the values expected under the theory of the model in question. A statistical measure called χ^2, or chi-square [1], is a quantity commonly used to test if any given data are well described by some hypothesized function or theoretical model. Such a determination is called a chi-square test for goodness of fit. In the following, we discuss χ^2 and its statistical distribution, and show how it can be used as a test for goodness of fit. If ν independent variables x_i are each normally distributed with mean μ_i and variance σ_i^2, then the quantity known as chi-square is defined by:

$$\chi^2 = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} + \ldots + \frac{(x_\nu - \mu_\nu)^2}{\sigma_{nu}^2} = \sum_{i=1}^{\nu} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \qquad (3.3.1)$$

Ideally, given the random fluctuations of the values of x_i about their mean values μ_i, each term in the sum will be of order unity. Therefore, if we have chosen the μ_i and the σ_i correctly, we may expect that a calculated value of χ^2 will be approximately equal to ν. If it is, then we may conclude that the data are well described by the values we have chosen for the μ_i, that is, by the hypothesized function. If a calculated value of χ^2 turns out to be much larger than ν, and we have correctly estimated the values for the σ_i, we may possibly conclude that our data are not well-described by our hypothesized set of the μ_i. The analysis of the goodness of fit in a first stage is made evaluating the χ^2. Since there are two

[1] The notation of χ^2 is traditional. It is a single statistical variable, and not the square of some quantity χ. It is therefore not chi squared, but chi-square. The notation is merely suggestive of its construction as the sum of squares of terms.

independent fits, whether with χ^2 in the two independent fit we mean (naming a, b, c and d the coefficients of the two fits):

$$\chi^2_{zx} = \sqrt{\frac{\sum_{i=1}^{N}(z_{exp}(i) - a - bx_{exp}(i))^2}{\sigma_{exp}(i)^2}} \tag{3.3.2}$$

$$\chi^2_{zy} = \sqrt{\frac{\sum_{i=1}^{N}(z_{exp}(i) - c - dy_{exp}(i))^2}{\sigma_{exp}(i)^2}} \tag{3.3.3}$$

we can consider a single χ^2, given by the following formula, which is considering not the sum of the two values but the square root of the sum 2 values for each fit. In this way is like if the sum is extended to 8 points, instead of only four:

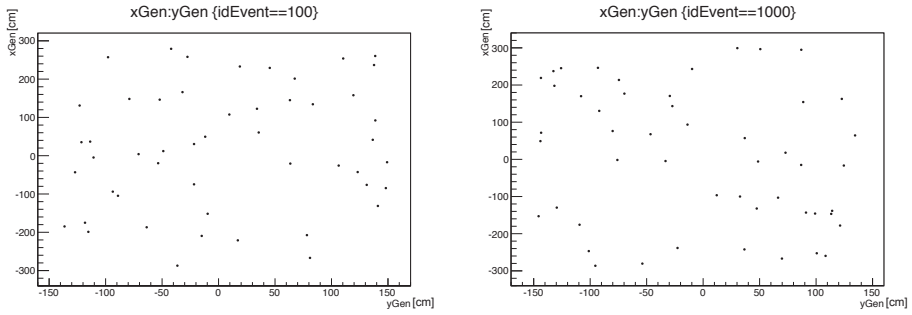
$$\chi^2 = \sqrt{\frac{\sum_{i=1}^{N}((z_{exp}(i) - a - bx_{exp}(i))^2 + (z_{exp}(i) - c - dy_{exp}(i))^2)}{\sigma_{exp}(i)^2}} \tag{3.3.4}$$

In these expressions of the χ^2, the $\sigma_{exp}(i)$ is an estimate of the error on $z_{exp}(i)$, that it can be assumed to be the same for all the points, and equal to $2/\sqrt{12}\,cm$. The quantity that is then used to assess the goodness of fit is the reduced-χ^2, given by the ratio between the of the χ^2, and the difference between the points used in the fit (8 in our case) and parameters used in the fit (4 in our case). So the reduced-χ^2=$\chi^2/4$. In our case, if we assume that the errors are evaluated correctly, we would expect a reduced-$\chi^2 = 1 \div 10$.

To evaluate the performance of the entire procedure we are interested in evaluating the efficiency ϵ, defined as the ratio between the well fitted tracks $N_{wellRec}$(the one with the reduced-χ^2 that is below a certain threshold, and/or those whose hits really belonged to the starting track) and total generated tracks N_{gen}.

3.3.2 Simulations results

This section shows the results of the simulation work done with events that contain nTr_{sim} simultaneous tracks. A total of $N_{ev} = 10^4$ events is generated, so that we have $10^4 * nTr_{sim}$ tracks in total. Fig. 3.16 is showing two plots of the points distribution on the generation plane. In the two plots two different events are

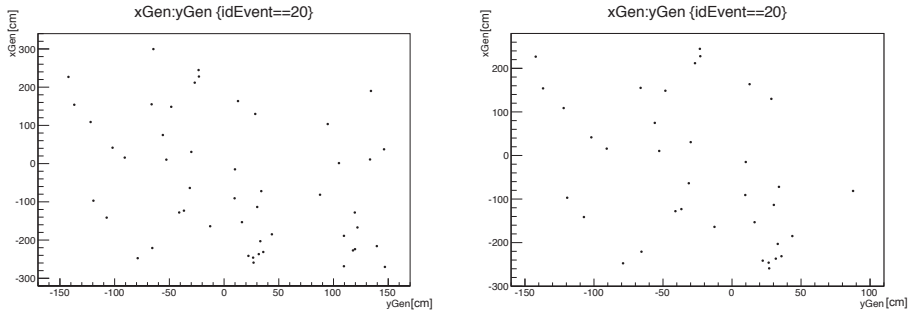


(a) (b)

Figure 3.16: Plot of the generated pairs (x,y) for a total of 50 points (simultaneous tracks). idEvent 100 (a) and 1000 (b).

showed (idEvent=100, and idEvent=1000) for a total of 50 simultaneous tracks per each event.

In Fig. 3.17 a 2D plot of the generated pairs (x,y) always with a total of 50 simultaneous tracks (a), and only of the accepted tracks (b), is visible.

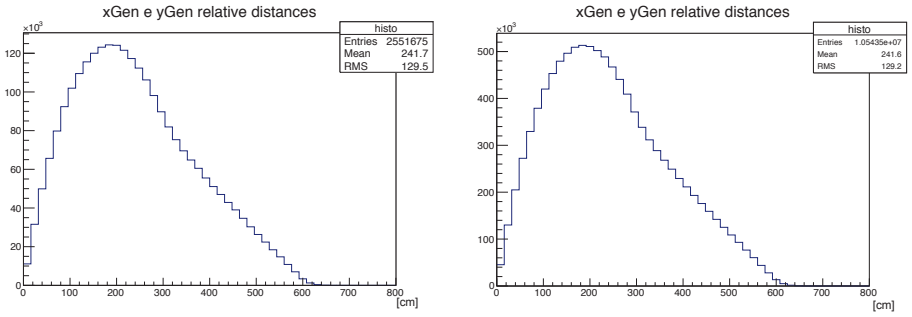


(a) (b)

Figure 3.17: 2D plot of the generated pairs (x,y) with a number of simultaneous track=50 for the idEvent=20 (a), and only the accepted tracks (b).

Histograms showing the relative distances between the generated particles on the generation plane for 50 simultaneous tracks (a) and 100 (b) are plotted in Fig. 3.18.

In this first phase, the tracks are considered to be all parallel to each other and to have different impact points on the detector. The code keeps track of the indices of the original tracks, and then, when a fit is made, it is possible to determine



(a) (b)

Figure 3.18: Histogram showing the relative distances between the generated particles on the generation plane for 50 simultaneous tracks (a) and 100 (b).

whether it was correct or not (a fit is considered to be correct if all the hits of the fitted tracks belong to the original track) and then to increment a counter. At the end, the percentage of tracks properly rebuilt $N_{wellRec}$ is calculated, compared to the tracks generated initially $N_{Tot} = nTr_{sim} \cdot N_{ev}$ and also the percentage of reconstructed tracks (those that have the χ^2 below the threshold) compared to the traces generated initially . There are three parameters that we can tune at this stage:

- *distMin*: it represents the radius of the circumference delimitating the minimum distance within which a hit must be taken into consideration as a possible combination
- *reduced-χ^2-Threshold*: maximum value for the reduced-χ^2. Above this threshold the reconstructed tracks are rejected
- nTr_{sim}: number of simultaneous tracks within a certain event.

Fig. 3.19 shows the reconstruction efficiency $\epsilon = N_{wellRec}/N_{Tot}$ when varying the distMin parameter, fixing the optimal conditions for the other two parameters i.e. a very high value for the reduced-χ^2-Threshold = 1000 and nTr_{sim} = 1, in order to see how the distMin parameter affects the reconstruction. We can see that with a 2.3 cm radius all the tracks are always reconstructed, but already with a value of 1.5 cm we reach an efficiency $\epsilon = 99.48\%$.

Fig. 3.20 shows the trend of the reconstruction varying the reduced-χ^2-Threshold

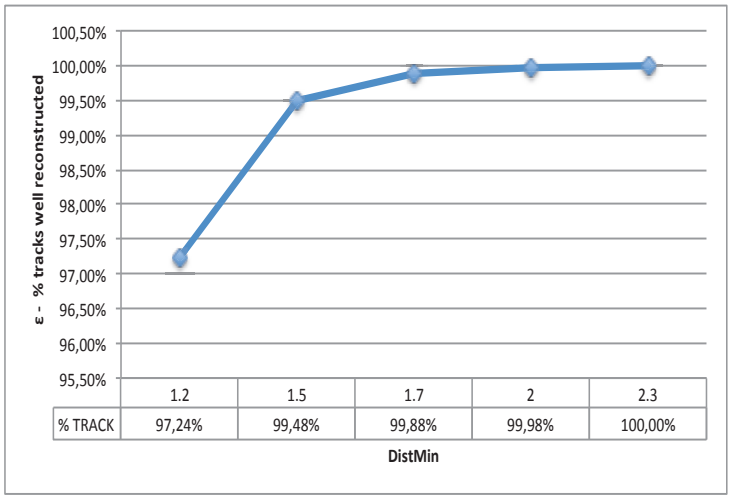


Figure 3.19: Efficiency ϵ of the reconstruction when varying parameter *distMin*, fixing the optimal conditions for the other two parameters (reduced-χ^2-Threshold = 1000 and $nTr_{sim} = 1$).

parameter and setting the other parameters to optimal values (i.e. $distMin = 2.3$cm and $nTr_{sim} = 1$). With a threshold value of 20 it is always possible to reconstruct all the tracks, but even putting the threshold at smaller values, like for example 4, we have that more than 95% of the tracks are correctly reconstructed.

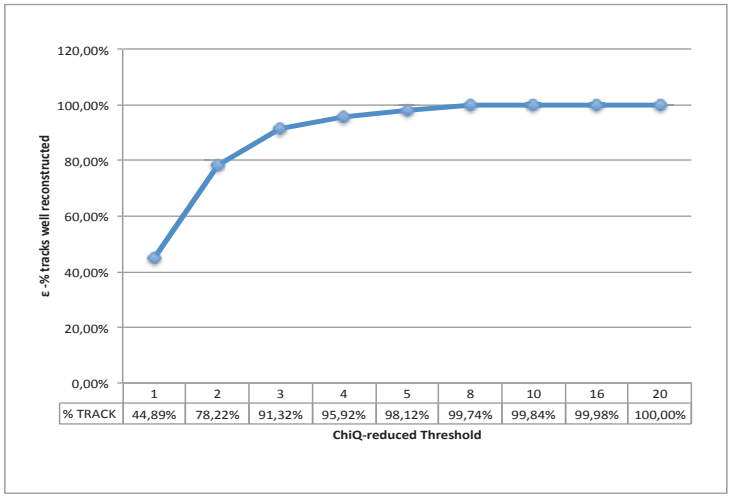


Figure 3.20: Reconstruction efficiency ϵ when varying the reduced-χ^2-Threshold parameter with $distMin = 2.3$ cm and $nTr_{sim} = 1$.

Fig. 3.21 shows the efficiency ϵ of reconstruction when the nTr_{sim} parameter

is varying, setting the other parameters to the optimal values (i.e. $distMin$ = 2.3 cm and reduced-χ^2-Threshold = 20). As it can be seen from the figure we reach

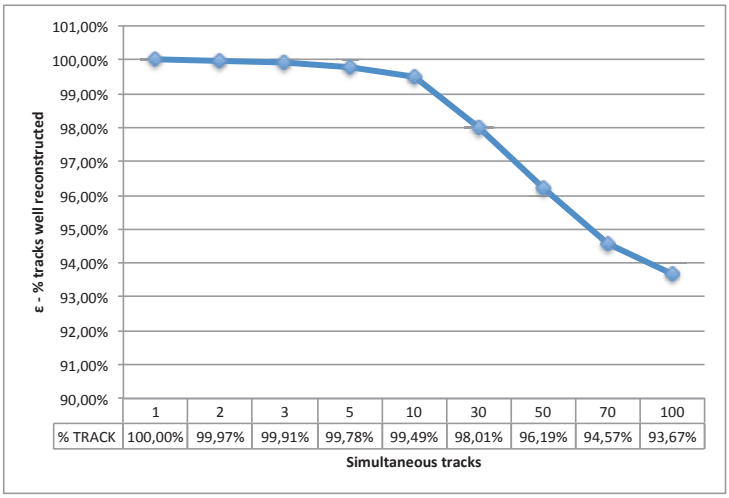


Figure 3.21: Efficiency ϵ of reconstruction procedure with a varying number nTr_{sim} of simultaneous parallel tracks. with $distMin$= 2.3 cm and reduced-χ^2-Threshold=20.

good performance that only degrade with a number of simultaneous tracks that goes above the expected number of tracks on our detector. The analysis was then repeated considering more realistic simultaneous tracks, i.e. not perfectly parallel. The $\theta_{primary}$ of the primary cosmic event is generated and then a Gaussian spread=20° is applied to obtain the $\theta_{muonTrack}$ of the single track. Fig. 3.22 shows the efficiency ϵ of the reconstruction procedure with a varying number nTr_{sim} of simultaneous not parallel tracks with $distMin$ = 2.3 cm and reduced-χ^2-Threshold=20. In this case we obtain better performance and even with nTr_{sim}=100 the efficiency ϵ is of 99, 38%.

This strange trend could be explained by the fact that the procedure used for the generation of nTr_{sim} simultaneous tracks, requires that all the nTr_{sim} tracks are accepted by the detector, thus imposing constraints not only on the $\theta_{muonTrack}$ but also on the spatial density and the relative distance of the generated and accepted tracks. In the case of parallel tracks it is more likely that, in order for them to be all accepted, the chosen one are those generated with a relative distance which is smaller with respect to the case of not parallel tracks. The analysis was

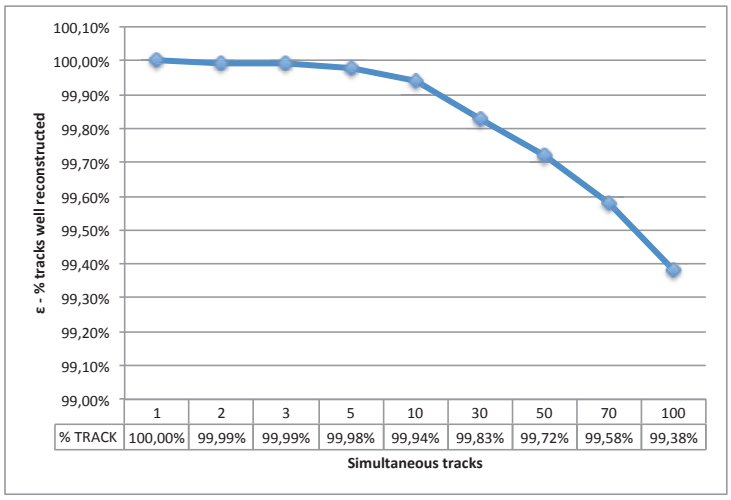


Figure 3.22: Efficiency ϵ of reconstruction procedure with a varying number nTr_{sim} of simultaneous not parallel tracks with $distMin$ = 2.3 cm and reduced-χ^2-Threshold = 20.

then repeated, this time not imposing necessarily that nTr_{sim} simultaneous tracks were accepted, but generating a total nTr_{sim} simultaneous tracks, which were not necessarily all accepted. Once tracks don't have the limits given by the geometric acceptance, we get the behavior shown in Fig. 3.23. The red line indicates the parallel tracks while the blue one the not parallel ones.

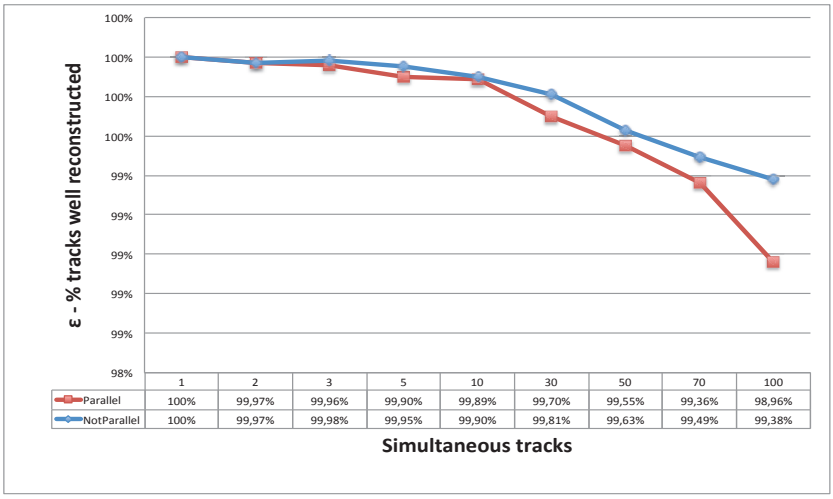


Figure 3.23: Comparison between the efficiency of the reconstruction with multiple tracks, parallel (red), and not parallel (blue).

This time, in the two cases we obtain similar performance, slightly worse in the case of perfectly parallel tracks. By the study carried out it is therefore possible to gather that the detector Muon Portal shows considerable capability to reconstruct simultaneous events, both at low and high multiplicity.

CHAPTER 4

Clustering analysis for the particle tracking and image reconstruction

Clustering analysis is a set of multivariate data analysis techniques through which is possible to gather statistical data units into groups, in order to minimize the *logical distance* within each group and to maximize the one between different groups. The *logical distance* is quantified by measures of similarity/dissimilarity between defined statistical units. Clustering techniques are traditionally applied to problems like pattern recognition, image classification and color quantization. These techniques allow to infer the implicit information into the data so they are used as a data mining technique to simplify the complexity of big datasets. In this chapter is presented a novel approach to the muon tomography data analysis based on clustering algorithms, together with the other reconstruction algorithms implemented within the project (POCA, EM-LM, Autocorrelation analysis). Cluster analysis is successfully used to elaborate data in the *Muon Portal* project. This technique, in fact, meets the need to make the tracks reconstruction and the visualization of the container content be independent from the grid and the 3D-voxels. The presence of a three-dimensional grid indeed limits the automatic object identification process. The problem is relevant in scenarios where the threat to be identified has a comparable size (or even smaller) to those of the single voxel and is located in a position not aligned with the grid. Clustering techniques, working directly on points, help to detect the presence of suspicious items inside the container, acting, as it will be shown, as a filter for a preliminary analysis of the data. The chapter is structured as follows: in Section 4.1 the reconstruction and visualization algorithms for muon tomography imaging

implemented within the project are explained. In Section 4.2 an introduction on the clustering analysis applied on muon tomography is presented, in Section 4.3 a literature review is presented. Section 4.4 is dedicated to the description of the *Friends-of-Friends* algorithm and of the modified version here presented. Section 4.5 shows some simulation results applied to the context of the muon tomography imaging and finally some conclusions are drawn in Section 4.6.

4.1 Reconstruction and visualization algorithms for muon tomography imaging

The track reconstruction is a challenging task and consists in the elaboration of data from the planes, the x and y coordinates of each logical level, in order to obtain information on the deflection occurred by the muons within the volume scanned. As described in the Chapter 1, when crossing a medium or high-Z material, the scattering angle of a muon from the original trajectory has a Gaussian distribution, with mean value 0 and RMS depending on:

- the inverse of the muon momentum $1/p$
- the material thickness x
- the radiation length X_0, which depends on $1/Z$.

Different materials mean different scattering angles, then it becomes necessary to measure these angles to determine the presence, position, shape and type of materials. In order to assess it, it is necessary to have a system that can detect the passage of the particle and reconstruct the trajectories. The container is treated as a black box: an analysis of its content is executed considering the behavior of particles crossing it, without any other information about what happens inside. Through the coordinates of the points of impact of the particle on the planes of the detector, it is then possible to derive the trajectories of the incident ray and of the outgoing one, so as to determine if scattering has occurred within the container and of which magnitude. This process is accomplished through the application of tracks reconstruction algorithms and 3D visualization techniques, which allow to realize the tomography of the container. Different statistical algorithms have been developed within the project for tomographic image reconstruction [72; 73].

In this section these algorithms are described, starting from the simplest and most common POCA - *Point of Closest Approach* algorithm and the more complex and accurate EM-LM *Expectation Maximization- likelihood* method, going to the *Autocorrelation Analysis*. An extra section is dedicated to the description of the Clustering approach with the *Friends of Friends* algorithm. All the algorithms have been developed in C++ using links to GEANT4, [74], and ROOT, [75], frameworks for detector geometry building and navigation and mathematical routines.

4.1.1 Point of Closest Approach algorithm

The simplest algorithm in the field is the *Point Of Closest Approach* (*POCA*). It is a purely geometric algorithm and it starts from a simplified assumption: each scattering is a single event and a single atomic nucleus is involved in the process, so that the muon scattering occurs in a single-point, ignoring the multiple scattering physical phenomenon. The algorithm take as input a list of two incoming sensor points (a_i, b_i) and two outgoing sensor points (c_i, d_i) for the i-th muon and return as output the corresponding list with the pairs $(P_{poca}, \theta_{scatt})_i$. Starting from the coordinates of the impact points at the planes, the incoming t_{in} and the outgoing t_{out} tracks are calculated. In a three-dimensional environment the two tracks, in most cases, will not be coplanar because of the scattering and of possible errors introduced by the measure, due for example to the angular resolution of the detector, so that they do not intersect at a single point. Consequently, basing on the projection of t_{in} and t_{out}, the algorithm searches for the geometrical point of closest approach P_{poca} between the incoming u_{in} and outcoming u_{out} reconstructed track directions with respect to the inspected volume using a linear algebraic formulation (see Fig. 4.1).

The scattering angle θ_{scatt} for each muon is also calculated. For each event the algorithm steps are:

1. Starting from the impact points on each plane, create the incoming track t_{in_i} and the outgoing one t_{out_i}

2. Using analytical formula find the closest points P_{in_i} and P_{out_i} for t_{in_i} and t_{out_i}

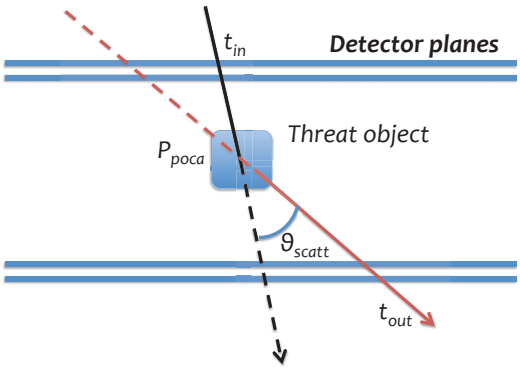


Figure 4.1: The working principle of the POCA algorithm.

respectively

$$\mathbf{P}_{in,out} = \mathbf{P}_{0\,in,out} + k_{in,out}\mathbf{u}_{in,out} \tag{4.1.1}$$

$$k_{in} = (be - cd)/\Delta \tag{4.1.2}$$

$$k_{out} = (ae - bd)/\Delta \tag{4.1.3}$$

where $\mathbf{P}_{0\,in,out}$ are two points on the incoming and outgoing tracks, $a = \mathbf{u}_{in} \cdot \mathbf{u}_{in}$, $b = \mathbf{u}_{in} \cdot \mathbf{u}_{out}$, $c = \mathbf{u}_{out} \cdot \mathbf{u}_{out}$, $d = \mathbf{u}_{in} \cdot \mathbf{w}$, $e = \mathbf{u}_{out} \cdot \mathbf{w}$, $\Delta = ac - b^2$, $w = \mathbf{P}_{0\,in} - \mathbf{P}_{0\,out}$.

3. Compute the mid-point P_{poca_i} between P_{in_i} and P_{out_i}
4. Compute the scattering angle θ_{scatt_i} between the tracks
5. Return the list of $\{(P_{poca}, \theta_{scatt})_i | \ 1 \leq i \leq N\}$ where the angle is not very close to zero (POCA point does not exist for parallel lines or where a muon has crossed without scattering).

The complexity of POCA is $O(N)$ if N is the number of tracks. Such method is of easy implementation and provides fast results, useful as first-order approximation to the problem or as a input for more detailed algorithms. However, it neglects the multiple scatterings through the volume material and therefore has the drawback of providing poor-resolution images, it is quite sensitive to the presence of shield materials located above or below the potential threat and cannot localize very well materials at the volume borders. This motivates the implementation of the log-likelihood algorithm discussed below, which is based on more realistic physical

and statistical assumptions and allows to face the problems encountered with the POCA algorithm. It is also desirable to have additional "grid-free" statistical methods for tomography analysis, which do not require to assume a predefined grid dividing the inspected volume into three-dimensional voxels. The clustering analysis will be helpful in this direction.

4.1.2 Autocorrelation analysis

The two-point autocorrelation function, hereafter denoted *2pt-ACF* for brevity, is one of the main statistics generally used to describe the distribution of galaxies and to search for localized excess of data observations at certain scales in a volume with respect to a homogeneous random distribution. It is therefore well suited also for the problem of tomography imaging where we need to search for a density excess of POCA observations inside the container with respect to a normal situation, for instance an empty container.
Following Peebles [76] the 2pt-ACF $\xi(r)$ defines the probability dP to find simultaneously two objects at a distance r from each other within two volume elements dV_1 and dV_2 in a data sample with event density n:

$$dP = n^2[1 + \xi(r)]dV_1dV_2 \tag{4.1.4}$$

A positive correlation ($\xi > 0$) at distance r indicate clustering at such scale, anti-correlation ($\xi < 0$) indicate that the objects tend to avoid each other, while $\xi \sim 0$ is relative to an homogeneous distribution without significative clusters.
For practical purposes the *2pt-ACF* can be computed from a sample of objects counting the pairs of observations at different separations r. Four estimators are generally used in literature: Peebles-Hauser $\hat{\xi}_{PH}$ [77], Davis-Peebles $\hat{\xi}_{DP}$ [78], Hamilton $\hat{\xi}_{H}$ [79] and Landy-Szalay $\hat{\xi}_{LS}$ [80]. They require to calculate the number of pairs $DD(r)$, $RR(r)$, $DR(r)$ at distance r respectively present in the data set (data-data), in a random data set (random-random) and in the data-random set

(data-random):

$$\hat{\xi}_{PH}(r) = \frac{N_{RR}}{N_{DD}}\frac{DD(r)}{RR(r)} - 1 \quad (4.1.5)$$

$$\hat{\xi}_{DP}(r) = \frac{N_{DR}}{N_{DD}}\frac{DD(r)}{DR(r)} - 1 \quad (4.1.6)$$

$$\hat{\xi}_{H}(r) = \frac{N_{DR}^2}{N_{DD}N_{RR}}\frac{DD(r)RR(r)}{[DR(r)]^2} - 1 \quad (4.1.7)$$

$$\hat{\xi}_{LS}(r) = 1 + \frac{N_{RR}}{N_{DD}}\frac{DD(r)}{RR(r)} - 2\frac{N_{RR}}{N_{DR}}\frac{DR(r)}{RR(r)} \quad (4.1.8)$$

with N_D, N_R total number of observations present in the data and random data sets and where $N_{DD} = N_D(N_D-1)/2$, $N_{RR} = N_R(N_R-1)/2$ and $N_{DR} = N_D N_R$ are the total number of corresponding pairs in the data-data, random-random, data-random catalogues. Such estimators take into account the edge effect due to the fact that it is not always possible to fit in complete spheres of radius r at every position within a survey volume, for example at the container borders.
The above estimators define spatial correlation only. To search for both spatial and angular correlations we introduced a weight $w_{ij} = \theta_i^\alpha + \theta_j^\alpha$ (α=2) for each observation pair ij with scattering angles (θ_i, θ_j) and computed a weighted correlation estimator $\xi_w(r)$.

4.1.3 Maximum likelihood algorithm

A better statistical treatment of the scattering processes can be done using a log-likelihood approach. It assumes the volume to be imaged divided into N_{voxels} three-dimensional voxels or pixels of size $N_x \times N_y \times N_z$. Following the well known Rossi's formula, describing the variance of the scattering angle of a particle of momentum p_0 traversing a material of radiation length X_0, a scattering density λ is defined for each voxel and given by:

$$\lambda(X_0) = \left(\frac{15\ \text{MeV}}{p_0}\right)^2 \frac{1}{X_0} \quad (4.1.9)$$

The determination of λ_j (j=1,... N_{voxels}) can be done by fitting the scattering data $\mathbf{x}_i$= ($\Delta\theta_{x,y}$, $\Delta_{x,y}$) for each i-th muon event for both x and y coordinates. Such joint distribution for a given scattering layer is with good approximation[1] modeled

[1] Actually long tails are present in the distribution and ∼2% of the data cannot be well described by the single gaussian assumption. A gaussian mixture is often used to reproduce the tails.

with a bivariate gaussian with covariance matrix $\mathbf{\Sigma}_i$ given by:

$$\Sigma_i = E_i + p_{r,i}^2 \sum_{j=1}^{N_{voxels}} W_{ij}\lambda_j \tag{4.1.10}$$

where E_i is the measurement error matrix, W_{ij} is the scattering covariance matrix through the j voxel and $p_{r,i} = p/p_0$ is the ratio between the muon momentum p and the reference momentum p_0.

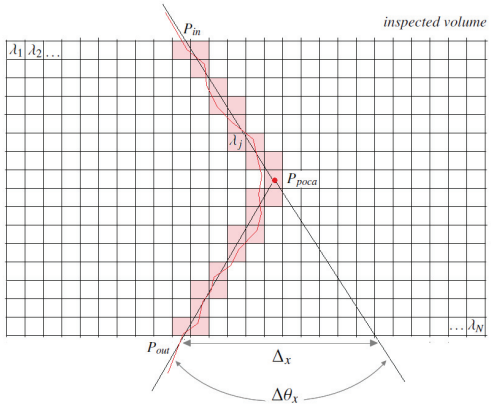


Figure 4.2: A sketch illustrating the working principle of the EM-ML method.

The log-likelihood $\mathcal{L}$ of a data sample of N muon events is therefore given by:

$$P(\mathbf{x}|\lambda) = \prod_{i=1}^{N} \frac{1}{2\pi|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}_i^T \mathbf{\Sigma}_i \mathbf{x}_i\right) \tag{4.1.11}$$

$$\mathcal{L}(\mathbf{x}|\lambda) = \frac{1}{2}\sum_{i}^{N}(\log|\Sigma_i^{-1}| - y_i^T\Sigma_i^{-1}y_i) \tag{4.1.12}$$

The scattering densities λ_j are estimated by maximizing the above log-likelihood. Traditional algorithms, such as those based on Newton-Raphson optimization, are limited by the large number of parameters to be determined, i.e. $\sim 5\times10^4$ for voxels of size 10 cm, and by considerable computation and storage required to compute the Hessian matrix. Schultz et al [81] provided a closed form solution to the problem in the EM formulation, leading to the following iterative estimation for λ_j:

$$\lambda_j^{(k+1)} = \frac{1}{M_j}\sum_i S_{ij}^{(k)} \tag{4.1.13}$$

$$S_{ij} = 2\lambda_j^{(k)} + p_{r,i}^2(\lambda_j^{(k)})^2(y_i^T\Sigma_i^{-1}W_{ij}\Sigma_i^{-1}y_i - \mathrm{Tr}(\Sigma_i^{-1}W_{ij})) \tag{4.1.14}$$

where M_j is the number of events traversing voxel j. A formula to compute W_{ij} is also available in [81]. In the following we will therefore denote this method as *EM-ML* for brevity. The algorithm is composed by the following steps:

Init

1. Reconstruct the scattering data $(\Delta\theta_{x,y}, \Delta_{x,y})_i$ for each event;

2. Compute the weight matrices W_{ij} for each event i on the basis of the muon path length through the j voxel. The latter can be estimated assuming a straight line connecting entrance and exit points from the inspected volume, eventually passing from the POCA point, if this is available or trustable. Fig. 4.2 shows a sketch of the algorithm raytracing principle. The path length calculation is achieved with a standalone GEANT4 navigator allowing a fast navigation through the container voxelized geometry;

Imaging

1. Assume an initial estimate λ_j^0 for λ_j;

2. Iterate formula (4.1.13) until convergence or early stopping;

It is well known that with iterative algorithms the image reconstruction can be deteriorated as the iteration proceeds. An early stopping criterion is therefore often used. Here we decided to stop the iterative procedure when the average relative λ_j variation drops below a prespecified threshold ε, namely:

$$\frac{1}{N_{voxels}} \sum_{j}^{N_{voxels}} \frac{\lambda_j^{(k+1)} - \lambda_j^{(k)}}{\lambda_j^{(k)}} < \varepsilon \tag{4.1.15}$$

where we assumed ε=1% and we required the criterion to be fulfilled during a given number of consecutive iterations (i.e. 5). Typically 20-30 iterations are needed to match the above criterion for the considered tomographic scenarios.

4.2 Clustering analysis for muon tomography data elaboration

Clustering or group analysis is a set of multivariate data analysis techniques that aims to select and to group homogeneous items in a data set. Since its appearance in the work of R. C. Tyron in 1939 [82], it was experienced with extended applications, starting from the '60s. Clustering techniques involve the division of the data into homogeneous subgroups. Through this multivariate data analysis technique, statistical units can be grouped together, in order to minimize the *logical distance* within each group and to maximize the one between different groups. The *logical distance* is quantified by measures of similarity/dissimilarity between the statistical units. Informally, the goal of this partitioning is twofold: data elements within a cluster must be similar to each other, while those within different clusters should be dissimilar, in such a way that the observations are the most possible homogeneous within classes and the most possible uneven between the different classes. The concept of homogeneity is specified in terms of distance and several criteria can be used, as it will be clarified later. Each of these classes is known as a *cluster*. It defines also a region where the objects density is locally higher than in other regions. Clustering techniques are successfully applied to problems as classification, pattern recognition and multivariate analysis. In this chapter a density-based clustering algorithm and its application to the analysis of muon tomography data is presented. Muon tomography, as stated in the previous chapters, is a technique that, using the secondary particles of cosmic radiation and their interaction properties with matter, has recently been successfully applied to the problems of scanning and detection of radioactive material, and the *Muon Portal* project is an example of application of this technique. The idea of using clustering algorithms to process data in the *Muon Portal* project arised from the need to make the tracks reconstruction and the visualization of the container content, be independent from the grid and the 3D-voxels as well. The clustering algorithms, working directly on points, are useful to detect the presence of suspicious items inside the container, and act as a filter for a preliminary analysis of the data. Several algorithms have been tested. In the following a modified

Friends-of-Friends (hereafter *FOF*) [83; 84; 85] algorithm is presented and simulations results with simulated scenarios are also showed. The Friends of Friends algorithm is a domain-specific clustering algorithm; it is a percolation algorithm normally used to identify dark matter halos from N-body simulations. We implemented a multiphase unsupervised clustering version of the algorithm which perform the clustering in two different steps, using at the first phase the scattering angle θ_{scatt} and the euclidean distance at the second step. The algorithm has been optimized by using space-partitioning data structures for organizing points in a k-dimensional space (kd-trees, [86])and further optimization strategies to achieve a $O(N\,log(N))$ complexity which makes it reliable for real time analysis.

4.3 Literature review on clustering

The clustering problem has been addressed in many contexts and by researchers in many disciplines; it has been effectively applied in a variety of engineering and scientific disciplines such as psychology, biology , medicine, computer vision , communications, and remote sensing (e.g. see [87; 88; 89]). Cluster analysis organizes data (a set of patterns where each pattern could be a vector measurements) by abstracting the underlying structure, playing therefore a vital role in the field of data mining. The clustering method implemented should be very fast, efficient, and robust. In literature there are several clustering algorithms which can be roughly grouped into 4 categories: partitioning methods, hierarchical methods, grid-based and density-based clustering (for further references see the [90] extended survey). They differ not only in their algorithm principles (which determine the behavior at runtime and the scalability of the algorithm itself), but also in many of their most basic properties such as the type of data processed, the assumptions on the cluster shape, the final form of the partitioning or the parameters that must be provided as input. Partitioning clustering algorithms generate a single partition, with a specified or estimated number of non-overlapping clusters, of the data in an attempt to recover natural groups present in the data. Depending on the kind of prototypes, we can distinguish k-means [91], k-modes [92] and k-medoid [93] algorithms. In the k-mean the similarity between clusters is measured with respect to the mean value of the objects belonging to a cluster. The

k-modes extends the k-means paradigm to categorical domains. The k-medoid algorithms use a prototype, called the medoid, that is one of the objects located near the center of a cluster, the so-called *centroid*. The algorithm *Clarans* introduced in [94] is an improved k-medoid type algorithm restricting the huge search space by using two additional user-supplied parameters. It is significantly more efficient than the well-known k-medoid algorithms *Pam* and *Clara* presented in [93], nonetheless producing a result of nearly the same quality. Hierarchical clustering algorithms construct a hierarchy of partitions, represented as a dendrogram in which each partition is nested within the partition at the next level in the hierarchy [95]. A common way to find regions of high-density in the data space is based on grid cell densities [96]. An histogram is constructed by partitioning the data space into a number of non-overlapping regions or cells. Cells containing a relatively large number of objects are potential cluster centers and the boundaries between clusters fall in the "valleys" of the histogram. The success of this method depends on the size of the cells which must be specified by the user. The problem is that cells of small volume give a very "noisy" estimate of the density, whereas large cells tend to overly smooth the density estimate. Density-based approaches apply a local cluster criterion and are very popular for the purpose of database mining. Clusters are regarded as regions in the data space in which the objects are dense, and which are separated by regions of low object density (noise). These regions may have an arbitrary shape and the points inside a region may be arbitrarily distributed. In [97], a density-based clustering method is presented which is not grid-based. The basic idea for the algorithm *Dbscan* is that for each point of a cluster the neighborhood of a given radius ε, has to contain at least a minimum number of points N_{min} where ε and N_{min} are input parameters. In [98] the density-based algorithm, named *DenClue*, is proposed. This algorithm uses a grid but is very efficient because it only keeps information about grid cells that do actually contain data points and manages these cells in a tree-based access structure. This algorithm generalizes some other clustering approaches which, however, result in a large number of input parameters. Another recent approach to clustering is the *Birch* method [99]. It cannot entirely be classified as a hierarchical or partitioning method: it constructs a CF-tree which is a hierarchical data structure designed for

a multiphase clustering method. First, the database is scanned to build an initial in-memory CF-tree which can be seen as a multi-level compression of the data that tries to preserve the inherent clustering structure of the data. Then, an arbitrary clustering algorithm can be used to cluster the leaf nodes of the CF-tree. Because *Birch* is reasonably fast, it can be used as a more intelligent alternative to data sampling in order to improve the scalability of clustering algorithms.

Another density-based approach is *WaveCluster* [100; 101], which applies wavelet transform to the feature space. It can detect arbitrary shape clusters at different scales and has a time complexity of O(n). The algorithm is grid-based and only applicable to low-dimensional data. Input parameters include the number of grid cells for each dimension, the wavelet to use and the number of applications of the wavelet transform.

Also the density- and grid-based clustering technique *Clique* [102] has been proposed for mining in high-dimensional data spaces. Input parameters are the size of the grid and a global density threshold for clusters. The major difference to all other clustering approaches is that this method also detects sub-spaces of the highest dimensionality such that high-density clusters exist in those subspaces.

In [103] an application of a clustering algorithm based on the relative distance between the scattering vertices, to a muon tomography project for the detection of high-Z materials inside containers, is presented. According to the results, obtained from simulations using RPCs with a realistic intrinsic resolution of $450\mu m$ and also using the momenta of the particles as information, the algorithm is capable of identifying the presence of a block of high-Z material, such as uranium, in various large-scale configurations. It must be said that the knowledge of the particle momentum introduces significant advantages, acting as a filter on the data, and allowing to discard many false positives that could be caused by the presence of large scattering angles generated by low energy muons. The algorithm was shown to be fast and efficient in a number of approximations of real-life situations, but it does not provide useful information for the reconstruction and visualization. It only provides a practical information on the absence of an alarm or the need to make further investigations, lengthening the scanning time of the container.

4.4 Multivariate Data Analysis Techniques: a multi-phase unsupervised FOF algorithm

Clustering analysis provides a very sensible tool to detect "objects" within a spatial domain. It is a customary tool adopted in Astrophysics, for instance, to quantify the hierarchical evolution of gravitationally bound systems like clusters and superclusters of galaxies, for instance. And it has a significant advantage over other tools because of its sensitivity: it can detect even objects (clusters) composed of a small number of units.
The reason of this sensitivity lies in its scaling with the number of composing units N_c, $O(N_c^2)$. This arises because clustering algorithms all are based on pair counting, and the number of pairs in a given set of N_c objects is given by: $(N_c((N_c - 1)/2)$. This scaling however is also computationally very expensive, in general. For this reason all the numerical implementations of the clustering algorithms have to be thoroughly tested to determine the minimum/optimal value(s) of N_c, as a result of a trade off between statistical significance and numerical complexity.
Common to all clustering algorithms is the comparison between the number of pairs in the sample with that in a random comparison sample, prepared under controlled conditions. Two questions arise naturally:

- How much "random" should the control sample be?
- How large should N_c be to guarantee a reasonable accuracy?

There is no universal answers to these two questions. What instead one often does is of testing many samples carrying a given signal against the null hypothesis, and use this statistics to determine the level of significance of a given problem. This is also what we have done in the present work.
Probably the most widely known clustering algorithm is the *Friends-of-Friends*, which is also the simplest algorithm, because it depends on a single parameter, i.e. the *linking length* r_{ll}. Its principle is very easy: starting from an arbitrary initial event it looks for all the events contained within a distance r_{ll}: if $\mid r_i - r_j \mid \leq r_{ll}$ the two events are members of a new object. It continues with all the remaining $(N - 2)$ events, and once completed the search of neighbours of the first object it

continues with the events which are not part of the first object.
The complexity of this algorithm grows *at most* as $N(N-1) \sim O(N^2)$: in practice, as events are added to the objects' list they are subtracted from the "active" list so that, at each step, the actual complexity grows as $N_{act}(N_{act}-1)$, where $N_{act} \leq N$.

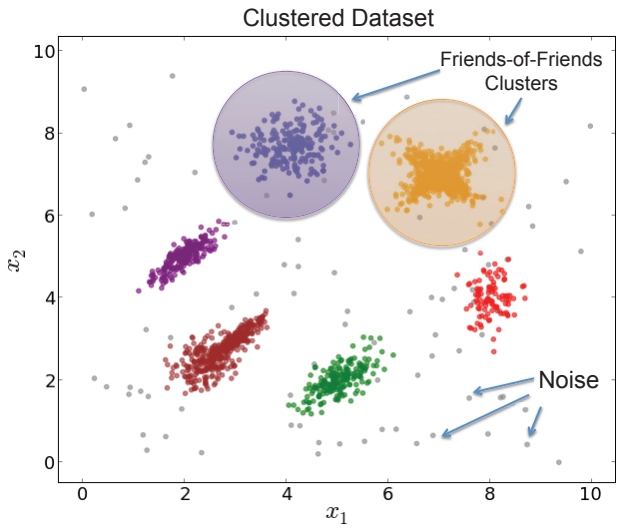


Figure 4.3: Image showing the result of the application of the FOF to a two variables dataset. The (x_1, x_2) pairs are considered as part of the same cluster if the relative distance is smaller than the defined r_{ll}.

The only freedom left lies in the choice of r_{ll}. A common choice is to take it as the average of the distance between events for a representative subset. This proves to be a very effective choice, as it has also been demonstrated that both the final list of objects and their average properties are largely independent from the choice of r_{ll}, provided that the input events catalogue is an unbiased, statistically representative sample. *FOF* is thus stable to its only parameter.
Another parameter in FOF algorithm is the minimum number of particles N_{min}, in the input list. This is particularly important to reject spurious clusters, i.e. transient objects which have a low statistical significance. Choosing N_{min} sufficiently large allows one to eliminate spurious clusters. In fact it is much more likely that a spurious cluster (noise) involves a small number of events and not viceversa. In the Fig. 4.3 a two variables clustered dataset obtained with the FOF algorithm is showed. Pairs (x_1, x_2) that do not match the algorithm clustering criteria are

rejected and considered as noise.

To summarize, the FOF requires r_{ll} and N_{min} as input parameters and it is based on the following steps:

1. Choose an arbitrary unvisited data point as starting point;
2. Find the neighborhood of this point, e.g. all points within the radius r_{ll}. If more than N_{min} neighborhoods are found around this point, a cluster is started and the point marked as visited, otherwise the point is labelled as noise;
3. If a point is found to be a part of the cluster then its r_{ll} neighborhood is considered as part of the cluster and the above procedure from step 2 is repeated for all r_{ll} neighborhood points. This is repeated until all points in the cluster are determined.
4. A new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.
5. This process continues until all points are marked as visited.

The algorithm defines uniquely groups that contain all the particles separated by a distance smaller than a given linking length r_{ll}. Once the length is defined the algorithm identifies all pairs of particles which have a mutual distance smaller than the linking one. These pairs are designated friends, and clusters are defined as sets of particles that are connected by one or more of the friendly relations, so that they are friends of friends.

4.4.1 A modified multivariate unsupervised FOF algorithm

An important property of many real datasets is that their intrinsic clusters structure can not be characterized by a single global density parameter. Very different local densities may be needed to reveal clusters in different regions of the data space. Furthermore one of the main weaknesses of clustering algorithms is that most of them are supervised and require the inclusion of a set of input parameters, which are dependent on the single dataset. In order to perform an optimized exploratory data analysis on the tomographic data from the *Muon*

Portal detector, we implemented a modified version of the FOF: an unsupervised and multi-phase FOF clustering version [73; 104; 105]. To make the algorithm independent from the setting of the input parameters and in order to identify the density-based clustering structure, we implemented an augmented sorting technique of the dataset, on the basis of Optics [106]. With this method, the points in the dataset are ordered linearly according to the criteria of a spatial distance. The points that are spatially close, become next to each other in the ordered database. During the data sorting operation, a special distance is stored for each point. It represents the density that has to be accepted for a cluster in order to be sure that two successive points are considered part of the same cluster. This process can be visualized through a dendrogram. It is a versatile base both for the automatic and interactive analysis of the cluster, useful to extract the intrinsic structure of the clusters themselves. The grouping operation is performed in two different steps: first the scattering angle θ_{scatt} is used as a criterion for grouping, performing an initial screening of the dataset and filtering out the noise from the date of interest. Then the Euclidean distance is used in the second phase, in order to distribute the clusters in the spatial domain. The *FOF* requires the computation of the nearest neighbour of each point in the volume, so it has, in its original implementation, a computational complexity of $(O(N^2))$. The modified version of the algorithm has been optimized by using kd-trees and further optimization strategies to achieve a $O(N\,log(N))$ complexity which makes the algorithm reliable for real time analysis.

A first stage of the analysis concerns the application of the algorithm Optics. In Fig. 4.4 is shown the Reachability plot relative to a dataset from Muon tomography taken as example and concerning a scenario with two high-Z threats. It shows on the x-axis data of our set reordered according to a distance criterion defined initially as spatial distance. A single line represents a single point, and its length stores the distance of the point from the previous one. The y-axis shows the distance values (according to the metric set), that give rise to the presence of clusters. The presence of dips in the plot indicates the existence of a cluster. The analysis made with Optics shows that the value of the ordinate, which is useful in the identification of two clusters, stood in the range of $5 - 10$.

The parameter r_{ll} thus identified, can be passed as input parameter to our

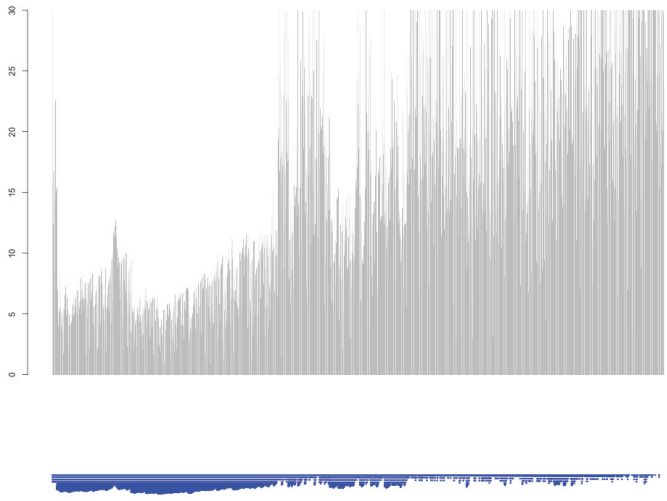


Figure 4.4: Reachability-plot for a simulation dataset with two blocks of high-Z material.

modified and multiphase FOF. In Fig. 4.5 are shown some plots which represent, by means of histograms, the results that we have from FOF by setting the parameter $r_{ll} = 5$, and assigning to N_{min} different values, respectively: 5, 20 and 30. On the ordinate we have the cardinality of elements for each cluster, while in the abscissa we find the cluster id. As it can be seen from the image, the parameter N_{min}, affects only the total number of clusters identified, but in all three cases, the presence of suspicious material is revealed. In the following sections some concrete results on simulated scenarios will be showed, together with a 3D visualization of the points filtered by the clustering and the volume rendering of the scenario. As it will be shown the method is successfully applied to the analysis, and the high-z material is then successfully identified and plotted.

4.5 Muon tomography simulation results

A detailed GEANT4 simulation of the Muon Portal detector, that considers all detector elements (scintillators, WLS fibers) and also the relevant mechanical structures, it has been developed in order to test the imaging methods presented above. Cosmic ray muons are injected in the detector with realistic energy and angular distributions, as derived from CORSIKA [107] simulations for proton-induced showers, generated for the Catania location (sea level, 37°30'4"68 N, 15°4'27"12

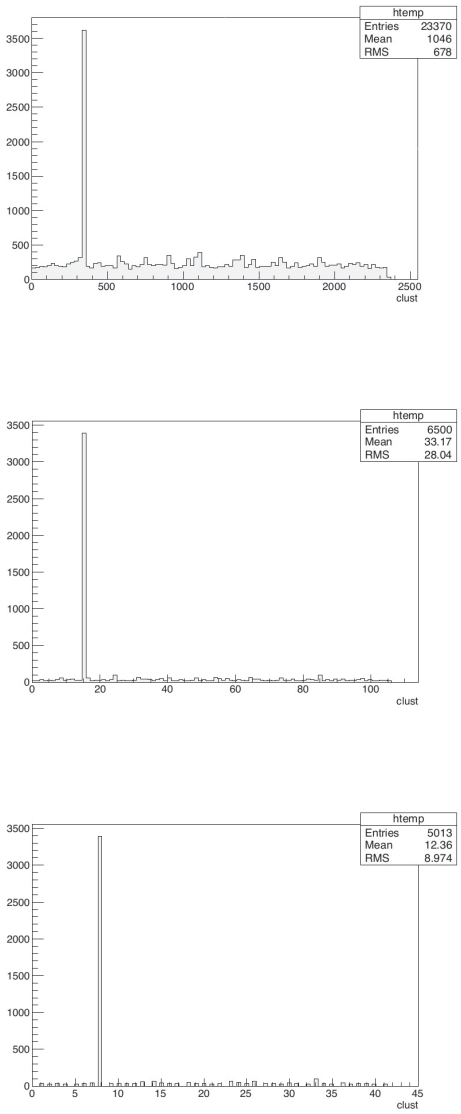


Figure 4.5: Results of the modified Friends-of-Friends applied on a dataset with high-Z material using a $r_{ll} = 5$ and N_{min} respectively equals to 5, 20, 30 points.

E, (B_x,B_z)= (27.16,-35.4)) with a $E^{-2.6}$ energy spectrum in the range 10^9-10^{15} eV and isotropic angular distribution. The energy distributions of the secondaries is approximately log-normal peaked at $\sim$3 GeV. The angular distributions are $\propto \sin\theta \cos^2\theta$, peaked at $\sim 30^\circ$.

Two kind of simulations are available. Full simulations include the explicit trans-

port of optical photons inside the scintillator bars and WLS fibers and are typically used only for detector design studies. Fast simulations, in which optical processes are switched off, are instead used for event reconstruction studies as well as to provide tomographic scenarios to test the imaging algorithms. We validated the implemented algorithms using GEANT4 simulations with different tomographic scenarios. The most relevant are shown in Fig. 4.6:

- *Scenario A*: Four threat boxes (*W*, *U*, *Pb*, *Sn*) of size 10 cm×10 cm×10 cm inserted at the center of a empty container. The container load relative to the scene is ∼100 kg.

- *Scenario B*: The simulated dataset refers to a scenario in which we have a container, at the center of which blocks of different materials have been placed. They create a MUON shape built with voxels of size 10cm $*$ 10cm $*$ 10 cm. Each letter is made of different materials: M= Uranium, U= Iron, O= Lead, N= Aluminum. The container load relative to the scene is $\simeq$ 480kg.

- *Scenario C*: Same of scenario B. A denser environment is assumed inside the container volume, filled with layers of washing machine-like elements. These are made by an aluminum casing with an iron engine inside with relative support bars and a concrete block. The container load relative to the scene is ∼3500 kg.

4.5.1 Event reconstruction and selection

The event reconstruction procedure is done according to the following stages:

  – *Hit selection*: Strips are considered to be triggered if the particle energy deposit (or the number of produced photoelectrons) is above a pre-specified threshold. A trigger threshold of 1 MeV is assumed for strip triggering. Furthermore we select strips triggering within a given time interval Δt. Events triggering at least four planes, hereafter denoted as 4-fold events, are selected and hits from each X-Y planes are collected to form a list of candidate track points. Each of these points is smeared

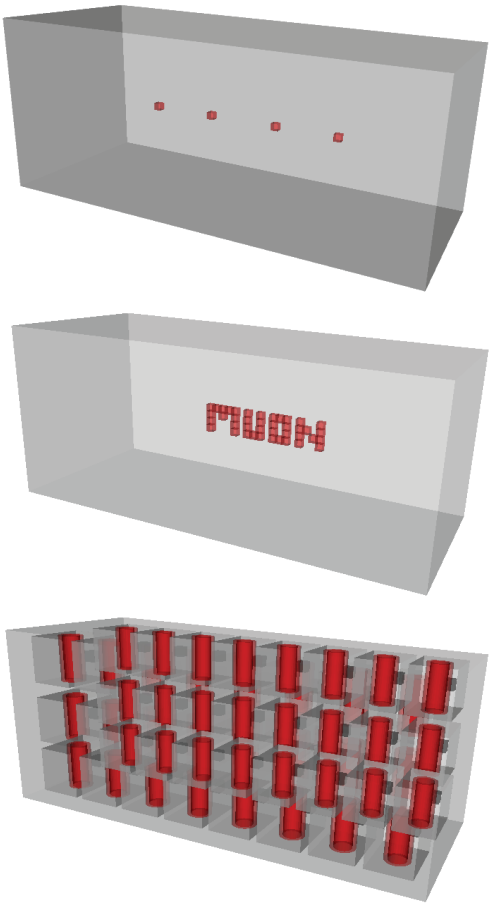


Figure 4.6: Four threat boxes (*W*, *U*, *Pb*, *Sn*) of size 10 cm×10 cm×10 cm inserted at the center of a empty container scenario (a). Simulated scenario with MUON writing at center made of different materials: M= Uranium, U= Iron, O= Lead, N= Aluminum (b). Simulated dense scenario with "washing-machine" like elements and MUON writing at center made of different materials (c).

with a Gaussian of width equal to the detector position resolution σ=1 cm/$\sqrt{12}$ ∼2.9 mm.

– *Cluster finding*: The list of candidate track points in each plane is scanned to find cluster candidates, defined by adjacent hit strips. The obtained clusters can be eventually re-splitted in single hits afterwards if the cluster multiplicity is above a given threshold. Finally single hits are replaced with the cluster barycenter and passed to the track finding stage.

– *Track finding*: Valid track candidates (at least one hit in each plane, $\theta_{rel} < 60^{\circ}$) are collected and then reconstructed using a Kalman-Filter approach [108]. In case of multiple track candidates, the track selection is done according to minimum χ^2 criterion.

For tomography studies a practical choice is to select events with only one cluster per plane. Multi-cluster events with larger multiplicity will be anyway recorded for cosmic ray physics studies.
To reject spurious events and therefore to reduce the chance of getting false positive in the imaging phase, we applied in the reconstruction algorithms a further quality selection to the available data:

– POCA, Clustering, 2pt-ACF: Events with scattering angles θ larger than 2 degrees are selected to reduce the noise due to other scatterers;

– EM-ML: Only events crossing the entire container from the top plane to the bottom plane are considered. To limit the chances of incorrect raytracing, leading to misidentifications and fakes, the POCA information is used to define the muon path inside the container only for events with a "trustable" POCA reconstruction, e.g. those preliminarily selected in the clustering analysis stage. After the selection chain, only a small percentage ($\sim$5%) of the total events was found to be rejected.

A number of muon events of 5×10^5, corresponding to $\sim$10 minutes scanning time, has been simulated for each scenario using a realistic energy spectrum with range 0.1-100 GeV.
In Fig. 4.7 we report the results relative to the POCA method for the three scenarios under test. On the left panels we report the tomographic XY section of the container for a fixed z level equal to z= 5cm (volume center). The right panels show a 3D volume rendering of the entire container. The volume has been divided into cells of volume 10 cm$\times$10 cm$\times$10 cm and the color scale represents the POCA signal for each bin i, namely $\sum_{j=1}^{N_i} \theta_j^2$ with N_i number of POCA events in bin i. As can be seen, all three scenarios are successfully

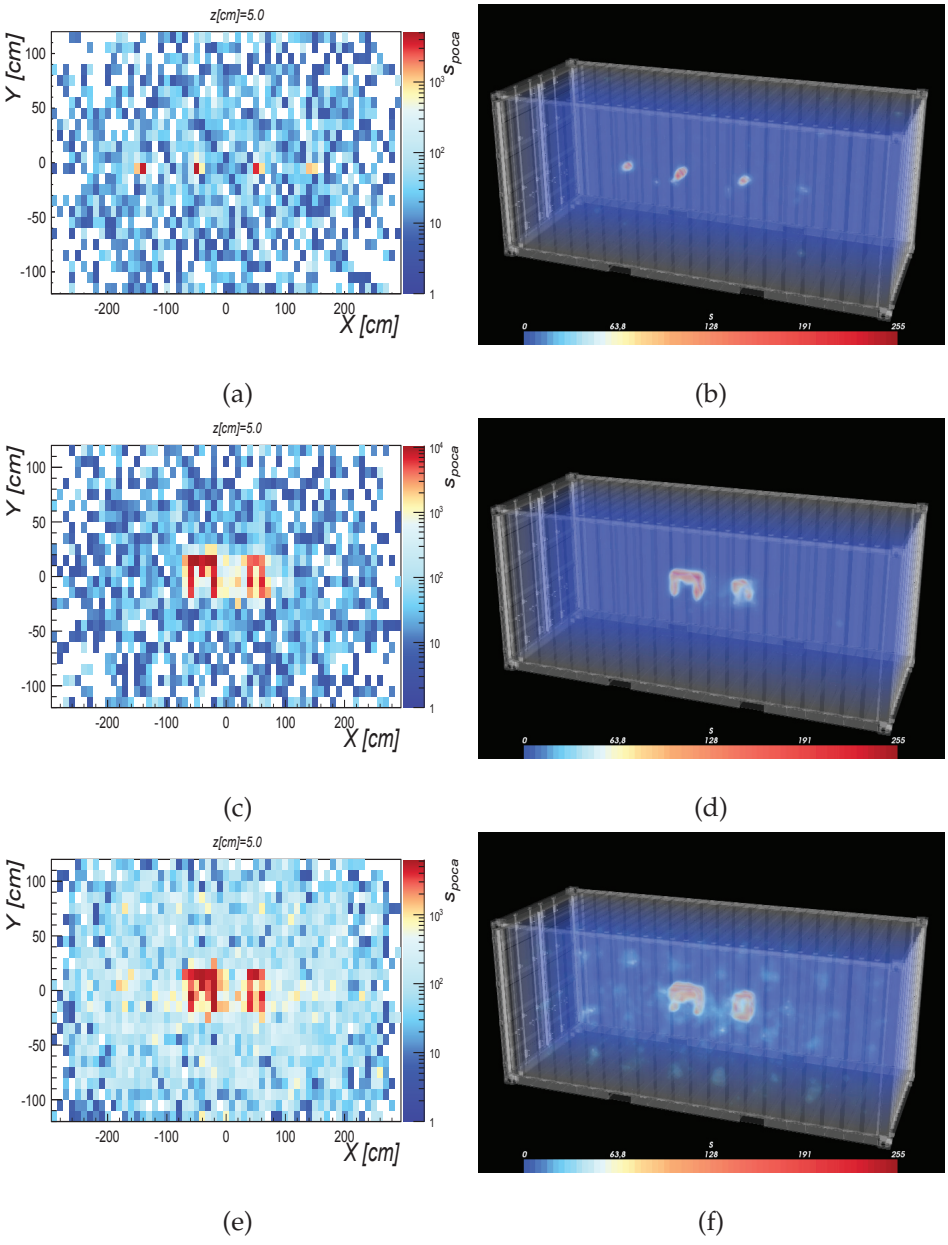
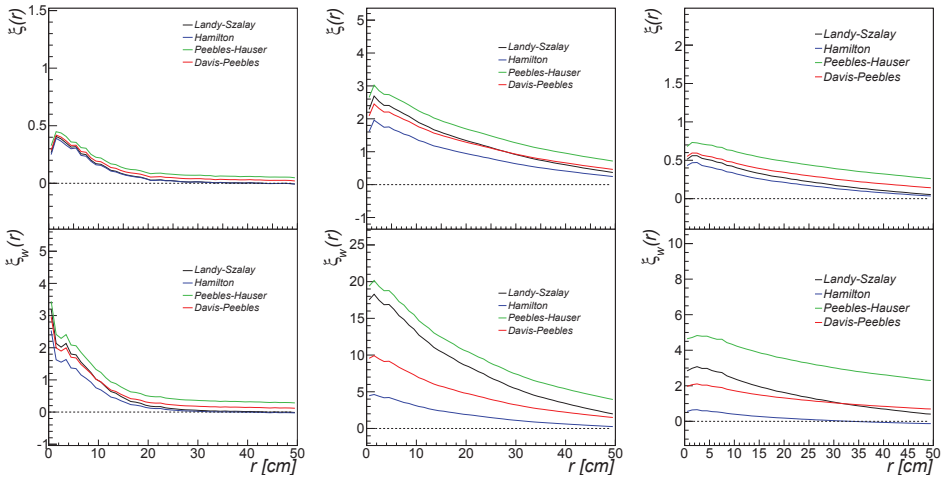


Figure 4.7: Tomographic imaging of the three simulated scenarios (from top to bottom panels) obtained with the POCA method. In the left panels the XY tomography view for a fixed z depth (z= 5 cm) is shown while in the right panels a 3D volume rendering of the entire container is reported.

identified. Due to the intrinsic resolution of the POCA method[2] a persistent

[2]We performed dedicated simulations of cosmic ray muons traversing a single uranium layer of thickness 10 cm and reconstructed the POCA information for each event. About 20% of the events

halo is present and consequently the imaged objects are slightly increased in size with respect to the real dimensions, particularly along the vertical z-axis. A considerable noise, related to the engine elements, is present in the dense environment scenario. In such case we needed to adopt a more stringent quality cut in the scattering angle ($\theta > 6^\circ$), to achieve the identification of the threat objects.



(a) (b) (c)

Figure 4.8: 2pt-ACF computed for the three tomographic scenarios from left to right panels. The colored lines refer to the four estimators described in the section 4.1.2.

We report in Fig. 4.8 the results relative to the autocorrelation analysis, computed using a random data sample of 5×10^6 simulated events in an empty container volume, e.g. ten times larger than the data sample under investigation. Upper plots refer to the standard 2pt-ACF estimators, reported with different color lines, while in the bottom panels we report the weighted correlation function. As can be seen, in all cases we obtain a significant excess with respect to the background at distance scales around 5 cm. The excess can be largely enhanced by using the scattering angle information (weighted

have the POCA information reconstructed outside the expected threat volume, falling in particular in the first surrounding voxels.

2pt-ACF) together with the spatial information.

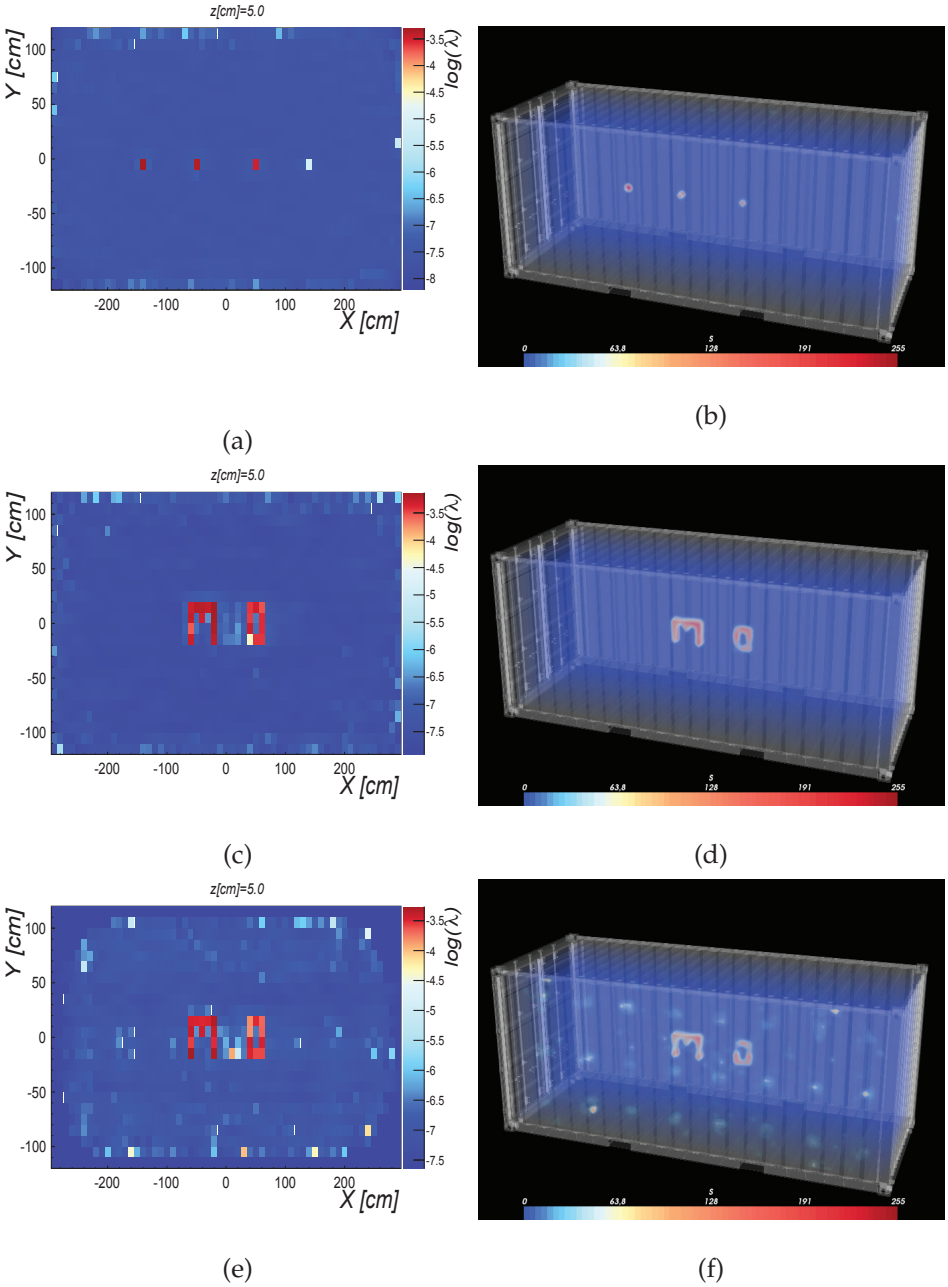


Figure 4.9: Tomographic imaging of the three simulated scenarios (from top to bottom panels) obtained with the EM-ML method. In the left panels the XY tomography view for a fixed z depth (z= 5 cm) is shown while in the right panels a 3D volume rendering of the entire container is reported.

In Fig. 4.9 we report the results obtained with the EM-ML method. As can be observed the target objects are reconstructed with a considerably better resolution. The halo responsible for the object deformation observed with the POCA reconstruction is almost absent. As already discussed for the other algorithms, the last scenario required a different strategy with respect to the scenarios A and B. In particular we observed that in presence of noise the convergence criterion adopted for the other reconstructions (ϵ=1%) is not sufficient to achieve a tomography image with quality comparable to scenario B. A smaller tolerance parameter was therefore assumed (ϵ =0.05%) to achieve better results (see Figs. 4.9 (e) and (f)), at the cost of significantly increasing the number of required iterations and the computing times. This scenario demonstrates that a finer tuning of the likelihood algorithm is currently desirable over different, more realistic, noisy scenarios with the aim of determining a unique configuration for real time analysis.

Finally in Fig. 4.10 we report the results obtained with the clustering method for the three scenarios. The analysis made with Optics provides a useful estimation of the linking length parameter ϵ to be used in the clustering reconstruction. We assumed ϵ=5 cm and a minimum point threshold of N_{pts}= 30. The color scale for the i-th cluster indicates the cluster weight w_i, defined as $w_i = \sum_{j=1}^{N_i} \theta_j^2 / N_j$ with N_i number of POCA points in cluster i. As it can be seen, a good accuracy is achieved in the identification of the high-Z material and the noise due to the washing machines engine in scenario (c) is significantly reduced. The clustering algorithm in fact acts as a filter, removing all spurious hits even if they have high scattering angles, because they are spatially isolated from the other POCA events. The FOF, however, basing on the output data from the POCA presents the same limitations.

To assess the reconstructed image quality we made use of the structural similarity index *SSIM* introduced in [109] for two-dimensional images, extended to our three-dimensional images. It allows luminosity, contrast and structure comparisons on a local basis between the reconstructed tomographic image and a reference image, built by considering the known scattering densities in each scenario as defined in formula 4.1.9.

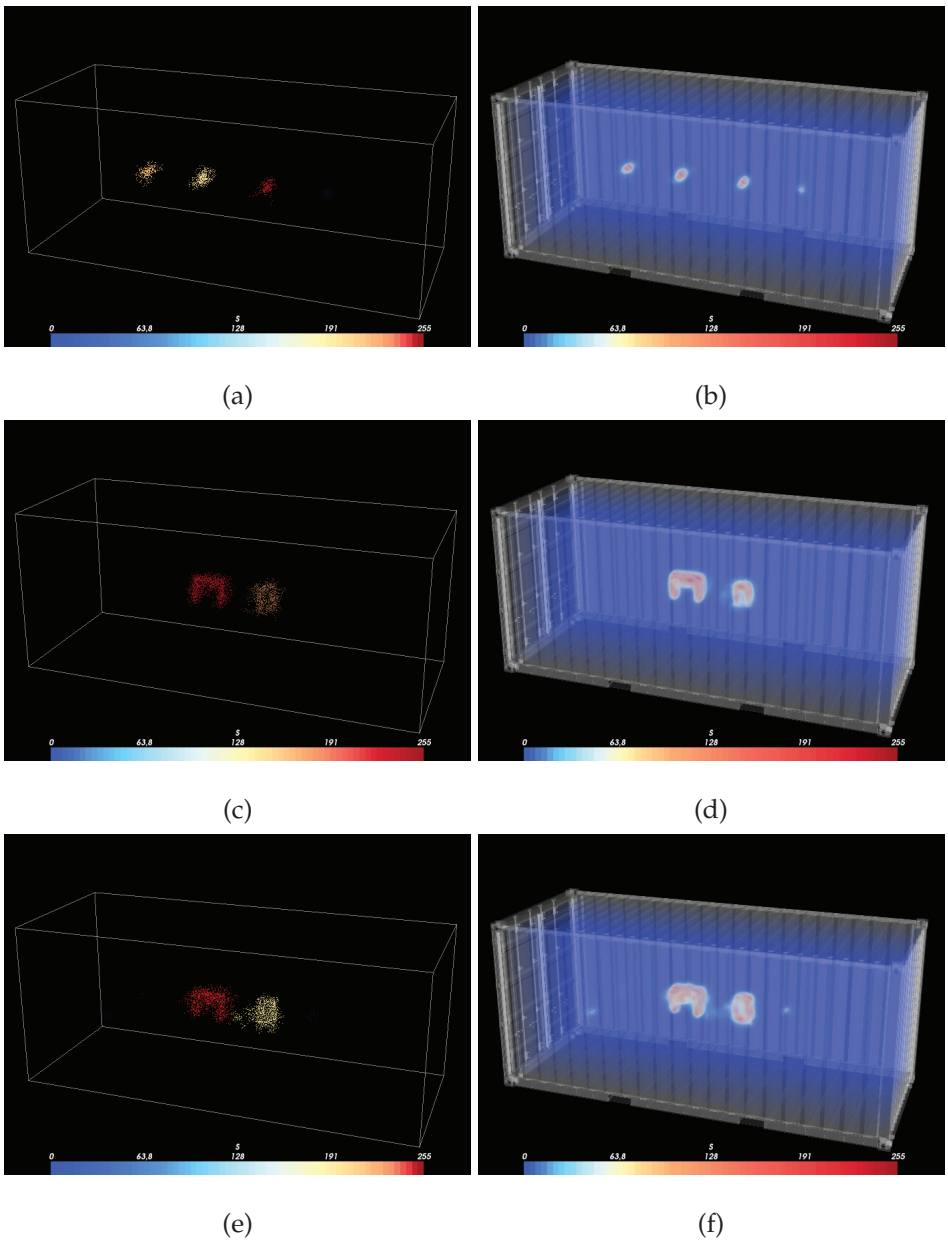


Figure 4.10: Tomographic imaging of the three simulated scenarios (from top to bottom panels) obtained with the Clustering method. In the left panels the visualization of the points belonging to the identified clusters while in the right panels a 3D volume rendering of the entire container is reported.

After proper normalization of both maps to the same range, we considered a comparison window around each image pixel j and calculated over it the pixel means ($\mu_{j,rec}$, $\mu_{j,ref}$), standard deviations ($\sigma_{j,rec}$, $\sigma_{j,ref}$) and covariance

$\sigma_{j,rec-ref}$ for the reconstructed and reference images respectively. The *SSIM* index for pixel j is then defined as:

$$SSIM_j = \frac{(2\mu_{j,rec}\mu_{j,ref} + C_1)(2\sigma_{j,rec-ref} + C_2)}{(\mu_{j,rec}^2 + \mu_{j,ref}^2 + C_1)(\sigma_{j,rec}^2 + \sigma_{j,ref}^2 + C_2)} \tag{4.5.1}$$

where C_1 and C_2 are small constants introduced to avoid numerical instabilities when $(\mu_{j,rec}^2 + \mu_{j,ref}^2)$ or $(\sigma_{j,rec}^2 + \sigma_{j,ref}^2)$ are very close to zero. An index close to one indicates strong agreement with the reference image, while, on the contrary, a nearly null index is symptomatic of a bad reconstruction.

Scenario	**MSSIM**		
	POCA	*Clustering*	*EM-ML*
A	0.94	0.94	0.99
B	0.89	0.89	0.98
C	0.83	0.83	0.89

Table 4.1: Similarity index SSIM obtained with the POCA, clustering and EM-ML methods for the three tomographic scenarios under analysis.

It is possible to define also a mean similarity index *MSSIM*, obtained by averaging the previous index over all pixels or over a given region of interest. In Table 4.1 we report the mean similarity index computed for the three tomographic scenarios under study and for the POCA, clustering and EM-ML methods. We considered a 3×3×3 pixel window around each pivot pixel and we calculated the mean similarity index over a spatial region surrounding the threats. As the visual analysis already suggested, the computed indexes are close to unity, indicating an overall accurate reconstruction. The computed index effectively reflects the better reconstruction performances of the EM-ML method with respect to the other implemented methods. A comparison between the algorithms behavior on the simulated scenario (c) can be seen in 4.11.

Additional considerations must be made on the execution time of imaging methods. We must consider that our application is time-critical in that, to

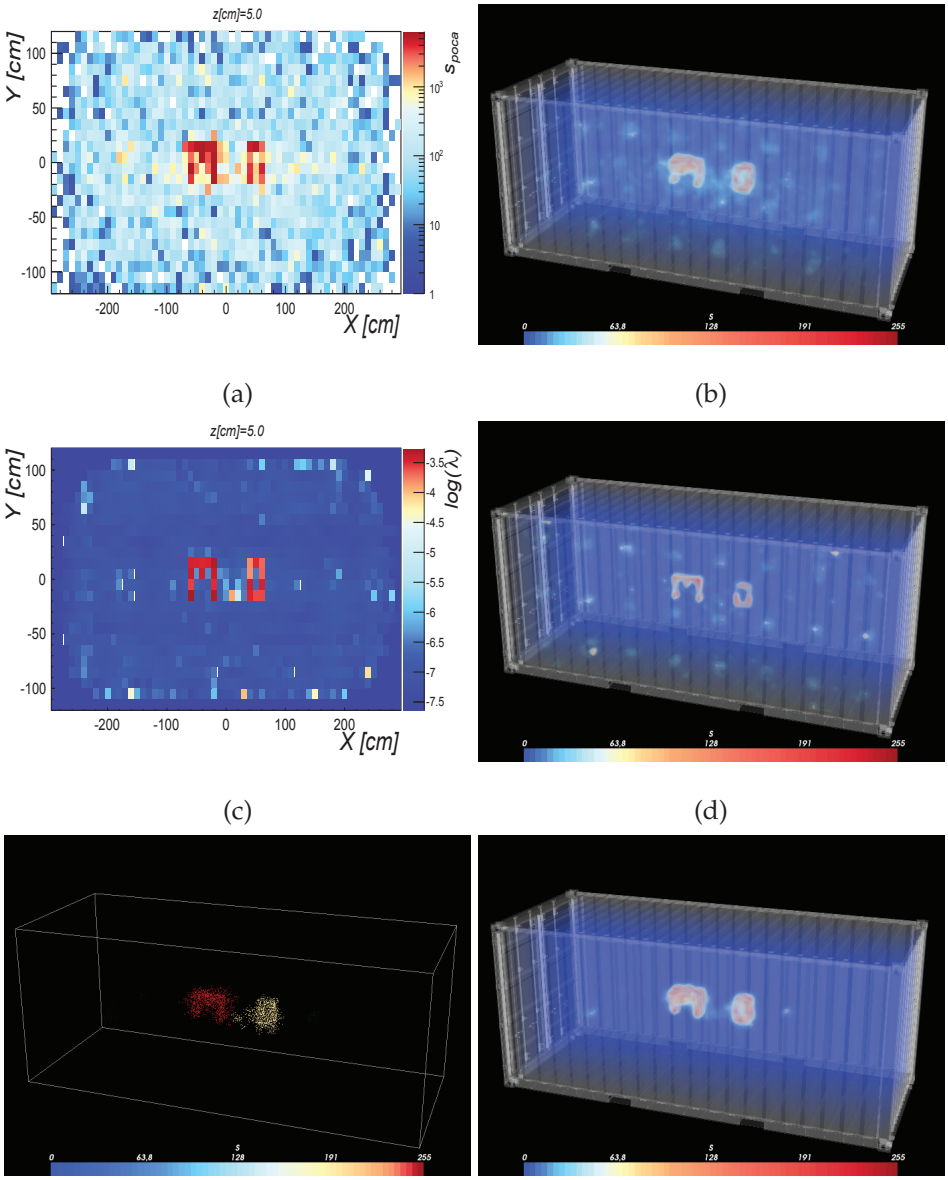


Figure 4.11: Comparison between the three different main approaches. Results of the reconstruction algorithms applied to the dense scenario with washing machines and the MUON writing hidden in the container. POCA algorithm results are showed in (a) and (b), EM-LM X-Y tomography and volume rendering are showed in (c) and (d) and modified-FOF clustering algorithm results are shown in (e) with points 3D-visualization and in (f), with volume rendering.

support the flow of containers in real harbour, it is necessary that the tomographic analysis is performed in a reasonably short time, a few minutes at most. The POCA algorithm, as expected, guarantees the smallest computation times, perfectly matching the port requirements. No optimizations are therefore needed in such case. The EM-ML algorithm, instead, typically requires ∼30 minutes on a Xeon QuadCore E5620 2.40Ghz processor for a typical scanning run of ∼ 10 minutes and 20-30 iterations, and therefore cannot match the requirements of a real time image processing, at least in its serial implementation. However, both the init and imaging step of the algorithm are embarrassingly parallelizable as being based on independent event loops. We therefore implemented also a parallel version of the algorithm using the MPI library [110]. In Fig. 4.12 a speed-up curve measurement for

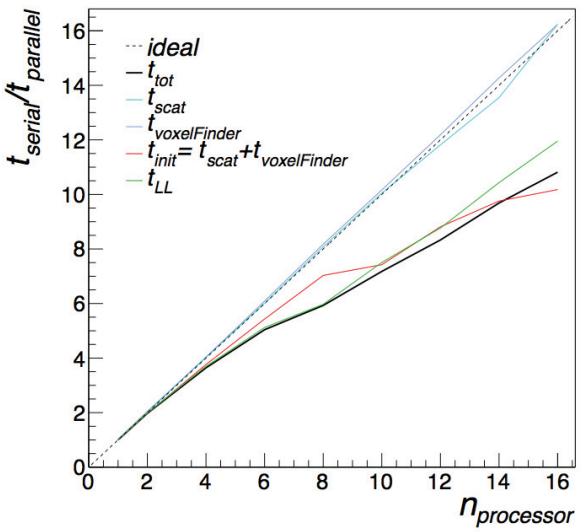


Figure 4.12: Speed-up curve for the EM-LM code with respect to the ideal curve. t_{tot}= total execution time, t_{scat}=time spent to calculate the scattering angle, $t_{voxelFinder}$= time to calculate the path within voxels, $t_{init} = t_{scat} + t_{voxelFinder}$=initialization time (takes into account also communication operations), t_{LL}= time for the likelihood iterative operation.

the EM-LM code, parallelized with MPI, with respect to the ideal curve is showed. The total time is divided into its components: t_{tot}= total execution time, t_{scat}=time spent to calculate the scattering angle, $t_{voxelFinder}$= time to

calculate the path within voxels, $t_{init} = t_{scat} + t_{voxelFinder}$=initialization time, which takes into account also communication operations, t_{LL}= time for the likelihood iterative operation. The achieved speed-up with respect to a serial implementation is remarkable. The parallel implementation allows a real time application of the method even with a modest number of computing machines.

The density-based clustering algorithm requires the computation of the nearest neighbor of each point in the volume having a computational complexity of(O(N^2)). The algorithm has been optimized by using kd trees [86; 111] and further optimization strategies to achieve a O($N \log(N)$) complexity which makes the algorithm reliable for real time analysis.

4.6 Conclusions

In this chapter it has been shown a study on the application of density based clustering algorithms based on a muon tomography dataset in the context of the *Muon Portal* project, together with the others algorithm implemented within the project for the reconstruction and imaging purposes. The standard algorithms have already been implemented and they are suitable to be used in a real time processing of a tomographic image. The use of alternative, more accurate, algorithms requires additional work, in the direction of parallelization at different levels, to provide good performance in terms of time consuming. A modified version of the FOF algorithm, multi-phase and unsupervised, has been implemented and the simulation results were shown. The clustering method proposed can be used both as a stand-alone tool to analyze and characterize a large dataset, for example for data mining applications and data processing, or it is suitable to be used as a filter, in pre-processing phases for other algorithms that operate, subsequently, on the identified clusters. The implemented method has many advantages: firstly, there is a linear scaling behavior, which makes it interesting for the application on big datasets. Secondly, the nature of the algorithm makes it quite robust to the effects of outliers within the data. Also, the proposed

clustering method has the ability to work with all types of datasets that can be described by a metric, and does not impose any assumptions about the shape of the clusters within which it works, being free from the limits of three-dimensional grids. Finally, an important aspect of the algorithm is its ability to automatically determine the number of clusters, being an unsupervised one. In practical situations, for example in a freight port, the need is to scan the maximum number of containers in the minimum time, and quickly exclude the presence of threats, so as not to overcrowd the unloading queue. The clustering method implemented is able to provide fast results about the absence or presence of the suspected threats. Thanks to the short runtime of the clustering algorithm, the time required to make a first screening does not go beyond the time necessary for the data acquisition, so the modified-*Friends-of-Friends* could be used for a first check of the container. For more densely loaded containers it would be necessary a longer scan which involve also the other reconstruction techniques.

CHAPTER 5

Status of the project and future perspectives

5.1 Status of the *Muon Portal* project and future perspectives

The *Muon Portal* project started in October 2011. The first year and a half of work has been devoted to research and development activity that has led to the definition of the basic parameters of the detector construction, many of which have been described in Chapter 2. The first phase was certainly dedicated to simulation activities. In fact the design of a complex particle detector usually requires a detailed simulation of the detector response, in order to evaluate the performance of the whole system. It is an usual procedure in the nuclear and particle physics field, in which the detectors are devices of a remarkable complexity, consisting of hundreds or thousands of elements and where the performances of the system depend on a large number of parameters. For this reason, simulation codes, specially oriented to the study of the detector, have been developed.

The performances of the detector that have been evaluated in the simulations phase included:

- **Sensitivity**: capacity of the apparatus and of the reconstruction software to detect the presence of a given quantity of material with high atomic number (eg. Uranium or Lead).
- **Scan time**: the time required to perform a full scan of the container, in relation to the accuracy and sensitivity required.
- **Angular and spatial resolution**: accuracy in tracking cosmic muons, necessary to obtain a good tracks reconstruction before and after crossing the container.
- **Location capabilities in space**: the ability of the apparatus and of the software to properly reconstruct the spatial position of the object hidden within the volume of the container.

For what concerns the technical solutions to be adopted, a number of possible options were evaluated before choosing the final design of the entire system. This process took into account the experience gained in areas where muon detectors of similar type have already been used. The evaluated options were:

- Possible solutions for the realization of the **detection planes** (i.e. dimensions, geometrical structure, etc.).
- **Materials** to be used in relation to the overall objectives.
- Solutions for the **optical coupling** between the detector and the photosensors.
- Type of **front-end and control** electronics to design and to implement.
- **Tracks reconstruction algorithms** and package for the visualization of 3D images.

To this day, the following research and development activities are all completed and some of these are referred below, in particular, for what concerns the realization of the single module.

Mechanics

This activity included among others, the definition of the type of media to create and the type of material to be used to glue the strip. It was chosen whether to proceed to the reading from both sides of the strip or from one side only, and if the reading were to occur from the front end directly coupled to each physical plane, or if the light signals were to be transported all in one place where to place the front-end.

Scintillators

Simulations and tests were carried out to define the type of material and the thickness of the strip and the groove, and verify with the suppliers, the ability to produce materials with the required specifications. Analysis of the strip have been conducted to understand, to vary the position of the track is along the strip that moving transversely, if there was a lowering of light output and loss of efficiency.

WLS fibers

Simulations and tests to define the type of fiber and its diameter were made. The choice of the fiber has been mainly due to the emission characteristics of the scintillator and the sensitivity spectrum of SiPM, in addition to its attenuation length and stability in conditions of exposure to high heat changes. Analysis related to the coupling between the fiber and the WLS SiPM have led to the choice of a direct optical coupling.

Photosensors

In view of the use of photo sensors for the detection of photons emitted by scintillators, a customization of the technology of manufacture of the SiPM was made in order to optimize the performance of the device in this particular application.

Design, implementation and test of the integrated revelation module

One of the testing stages of the final prototype concerns the testing of the individual detection module. This analysis was performed on the first modules already made and is in progress at the time in which this thesis is being written. The results of laboratory tests on the first modules were presented in Chapter 2.

5.1.1 Ongoing activities

Currently underway is the construction of the four integrated planes of the tomograph. 48 identical detection modules complete with electronic cards which house SiPMs with the front-end control and transmission electronics have to be constructed. At the moment about 1/3 of the total modules has already been built. Each of the 48 modules will be individually tested both with respect to its detection capability and to the transmission and control capabilities. The cooling system will also be tested. It should be remembered that the realization of the scanner prototype, consists of a set of 8 planes of the described size, and it will require a series of functional tests, designed to verify the operation of the apparatus, as well as to evaluate performances, comparing with the design specifications. For the evaluation of the full functionality of the apparatus it will be necessary to make a complex series of measures designed to improve the working conditions of the detector and to evaluate performances against the referenced specifications. The tests include, for example, studies related to the alignment of the planes. In fact, it will be necessary to check the flatness of each plane (horizontal), the relative distance between the physical XY planes of each logical plane, the distance between the different detection planes and their parallelism, with good precision. Furthermore the edges of the different planes have to be aligned, with precision of the order of mm. The influence of any misalignment (both along the vertical and horizontal direction) has to be examined by means of simulations, to evaluate and quantify the performance degradation as a function of the value of such misalignments. When all the planes will be

assembled and the sensors will be connected to the optical fibers, as well as the connection between the sensors and the front-end electronic will be realized, the first tests of functionality will concern the correct operation of all detection channels. For these tests the same secondary cosmic radiation might be used, in order to realize monitoring plots related to the channels distribution and the multiplicity distribution. It will be verified that all the sensors are properly connected, that there are no "dead" channels, and that the counting rate from individual channels is not abnormal. Further tests will be conducted to assess the detection efficiency of each plane, and therefore produce a map of efficiency as a function of the coordinates (X, Y). When all the planes will be arranged in operating conditions, the detector will be used to acquire data for a sufficient time (of the order of a day), to analyze the data and extract the angular distributions in the polar angle and azimuth angle, in order to assess the correctness of the system. Depending on the spatial resolution obtained, it will be possible to extract also the angular resolution. After completing the procedures to verify the full detection functionality, the following step will be the acquisition, with good statistical accuracy, of a series of tomographic images in the absence of materials with high atomic number, to build a database of images of the background signal noise. These measures should be used to evaluate the effect of the mechanical structure which supports the detector, the walls of the container, and the possible content of a container containing typical materials used in the transportation of goods. In this regard, a large series of measurements under various conditions, as realistic as possible, will be conducted. In fact a structure with the features of a real container is under construction. Finally, it will be possible to proceed to the tests for the identification and localization of hidden material, interposing blocks of known material (of different atomic number, from Aluminum, Iron, Lead) and of different sizes, among the intermediate levels, in various geometrical positions, and verifying the capabilities of the system and of the associated software of reconstructing the existence and locating the unknown material.

5.1.2 An extensive air shower trigger station for the *Muon Portal* detector

Apart from muon tomography, the *Muon Portal* has in itself the potential to detect single and multi-muon events over a large sensitive area and with a good angular resolution, due to its four X-Y detection planes [112]. This type of activity has already been foreseen for the *Muon Portal* detector, which might also be used to study the cosmic rays physics. As explained before, the distance between top and bottom planes is adjustable and can vary between 5 and 6 m. In Chapter 3 , some of the properties of the detector to reveal events to multiple tracks (multi-muon events) have been described and studied. A good angular resolution (of the order of 0.1°) may be achieved by the detector, if we consider to place the top and bottom planes at 6 m distance between each others, when the Muon Portal is used simply as a muon detector, without any heavy material interposed between the detection planes. As seen, single rates of the order of 500 Hz may be expected with the fully operational detector, allowing a large number of events $\simeq 4 \times 10^7 / day$ to be collected. This should give a reasonable statistics also on multi-muon events and any orientation- dependent cosmic ray study. In order to detect and reconstruct the extensive air showers (EAS) produced in the atmosphere by the interaction of high energy particles, usually detector arrays, placed at the sea level or moderate altitudes, are used. The performance of such arrays are defined by the number of detectors, by they sensitive area and the relative distance between them. These parameters condition the geometrical acceptance, the capability to reconstruct orientation and core of the primary particle which produced the shower, and the primary particle energy range. Large arrays for the detection of high energy cosmic rays may require thousands detectors covering a huge area, with a trigger given by the simultaneous detection of secondary particles above a given threshold. The most famous example of such an experiment is the Auger experiment [10]. The construction of a complex array for the detection of extensive air showers is not easily achievable but, a simple trigger station giving the signature of the arrival of a shower may be built in a relatively easy way using

the *Muon Portal*, providing the conditions for looking at other aspects of the shower itself, for instance at the multiplicity of muons in a muon tracker detector.

In this respect, a basic trigger station was designed with three large area scintillation counters which can be moved around, to select extensive air showers of low to medium energy and complement the detection capabilities of the *Muon Portal* detector. Due to its modularity, it is planned to extend the station with additional detectors, to provide a larger coverage of the area around the muon tracker and to better define the properties of the showers being detected. Some preliminary results have already been published and few of them are reported in what follows. At the first set of tests and data taking, the detectors were located at only $5\,m$ relative distance. This is a suitable solution to provide a trigger event by the use of a single acquisition card. Such solution may be employed even at distances of the order of a few tens meters. To test the overall working conditions, the three detectors were placed at the corners of an equilateral triangle of 5 m side. The PMTs were biased at $1250\,V$ and thresholds around $40 - 60\,mV$ were chosen, in order to roughly have the same counting rate in each module. In such conditions, the three-fold coincidence rate was found to be $0.011\,Hz$, i.e. about 900 events/day. In a time interval of 20 days, about $2 \cdot 10^4$ events were collected and analyzed, as an overall test of the working conditions of the setup. Within statistical fluctuations, this rate was nearly constant along the time, as it is shown in Fig. 5.1, which reports the number of detected coincidences in each time bin of 12 h. Such station could be used either as a stand-alone detector or in coincidence with the *Muon Portal* tracking detector, where single muon and multi-muon events may be observed, thus permitting a variety of studies in cosmic ray physics. Studies of muon bundles either in inclusive mode or triggered by low and medium energy air showers, as well as muon flux anisotropy studies, are in principle within the capabilities of the *Muon Portal* detector, once complemented by such station, and will be pursued in the near future.

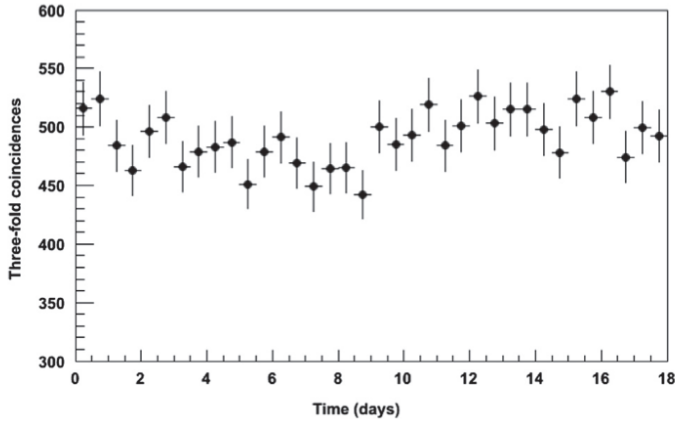


Figure 5.1: Number of observed three-fold coincidences (in a time bin of 12 h) as a function of time, for a period of approximately 20 days.

5.1.3 Towards a real-time tomographic analysis

Concerning the software tools, two working lines are currently almost completed within the project in view of the complete operation of the Muon Portal, one aiming to complete the graphical user interface for the tomography and visualization tasks, and the other focusing on the optimization of the designed algorithms for real-time application.

To be compatible with the real container traffic at the harbors, the tomography analysis must be performed in reasonable small times, few minutes at most. The POCA algorithm, with its simplicity, already guarantees the smallest computation times, perfectly matching the port requirements. No optimizations were therefore needed in such case. This was not the case for the other designed algorithms.

The EM-ML algorithm typically requires $\sim$30 minutes on a Xeon QuadCore E5620 2.40Ghz processor for a typical scanning run of $\sim$ 10 minutes and 20-30 iterations, and therefore cannot match the requirements of a real time image processing, at least in its serial implementation. However, both the init and imaging step of the algorithm are embarrassingly parallelizable as being based on independent event loops. Therefore we implemented also a parallel version of the algorithm using the MPI library [110], as showed in Chapter 4. The achieved speed-up with respect to a serial implementation is

remarkable. The parallel implementation allows a real time application of the method even with a modest number of computing machines.

The computation of the 2pt ACF is a very time-consuming task, proportional to N^2 (N size of the data sample). Optimized serial implementations, based on building a kd-tree [111] with the data, allow to drop the algorithm complexity at the level of $N\log(N)$. Significative speed-up can then be obtained afterwards with ad hoc optimization and parallelization techniques [113; 114] or by making use of GPUs [115]. At the present status a brute implementation is available. To maintain the computation time at reasonable level, the pair calculation is limited to adjacent three-dimensional voxels with size matching the maximum desired correlation scale and to observations with scattering angles larger than a predefined threshold. An optimized version of the algorithm is however currently being designed.

The density-based clustering algorithm suffers from the same problematic discussed for the ACF computation, as it requires the computation of the nearest neighbor of each point in the volume ($O(N^2)$). The algorithm has already been optimized by using kd-trees and further optimization strategies to achieve a $O(N\log(N))$ complexity which makes the algorithm reliable for real time analysis. In conclusion we pursued a large efforts in combining different reconstruction and visualization tools for a reliable and fast image processing of a muon tomography. Standard algorithms have already been implemented and their use in a real time processing of a tomographic image may be achieved even by a single standard processor thanks to the effort on parallelization. The use of alternative, more accurate, algorithms required additional work, but in the end is providing a competitive fast tool. The proposed modified-*Friends-of-Friends* can be used both as a stand-alone tool to analyze and characterize a large dataset, for example for data mining applications and data processing, or it is suitable to be used as a filter, in pre-processing phases for other algorithms that operate, subsequently, on the identified clusters. Furthermore, the clustering method proposed is able to provide fast results about the absence or presence of the suspicious threats. Thanks to the short runtime of the clustering algorithm the time required to

make a first screening would not be greater then the time necessary for the data acquisition.

5.1.4 Outreach and dissemination activities

Among the collateral activities of the project, some outreach and dissemination activities have also to be reported. A large-size box (dimensions $4\,m \times 4\,m \times 4\,m$) is being built for these purposes, to publicize the project and to be used for dissemination activities and educational purposes.

Figure 5.2: Imaginary view of the box designed for outreach and dissemination activities purposes.

The box is covered with solar panels, which are used as electricity supply of the box itself and that produce also part of electricity necessary for the detector functioning. The photovoltaic production system that powers the battery pack has an output of $9.3\,kWp$ [1]. Inside the box, there will be a room

[1] In the photovoltaics field, this value is used to indicate the power provided by one module or by a photovoltaic cell if subjected to the standard conditions of: irradiance of $1000 W/m^2$, cell

with video screens that will project information on the detector. In Fig. 5.2 an artistic image of the realization of the box.

5.2 Concluding remarks

The present Ph.D. thesis was framed within the *Muon Portal* project, a project dedicated to the creation of a tomographer for the control and scanning of containers at the border in order to reveal smuggled fissile material by means of the cosmic muons scattering. This work aimed to the extension and consolidation of the research in the field of muon tomography in the context of applied physics. The main purpose of the thesis was to investigate new techniques for reconstruction of muon tracks within the detector and new approaches to the analysis of data from muon tomography for the automatic objects recognition and the 3D visualization, thus making possible the realization of a tomography of the entire container. This research activity has addressed the tracking particles problem through an analysis of the detector performance in the case of particle tracking at low and high multiplicity and through the study of different image reconstruction techniques. The R&D activity was mainly based on the study, development and testing of different well-known reconstruction algorithms like the POCA algorithm (*Point of Closest Approach*) and the iterative EM-LM algorithm (*Expectation-Maximization*) and on a deeper investigation on new methods and technologies. In particular, more advanced methods for the tracks reconstruction and visualization, such as data-mining techniques and clustering algorithms have been the subject of the research and development activity which has culminated in the development of an unsupervised multiphase clustering algorithm (modified-*Friends-of-Friends*) for the muon tomography data analysis. These algorithms were tested using simulated data with the GEANT4 code, basing on a number of scenarios in which it was assumed the presence of high

temperature of 25 , position of the sun at $1.5AM$, where $AM \simeq 1/\cos(\theta)$ is short for Optical Air Mass, which is a parameter indicating the apparent position of the sun. This magnitude is related that the sun angle θ formed with the zenith.

Z material (Uranium, Plutonium, Lead, ..) along with lighter materials, to evaluate the ability of each algorithm to properly reconstruct the image of the hidden volume within the container. The tests demonstrate the validity of visualization methods and reconstruction techniques used, confirming the effectiveness of the muon tomography in automatic objects targeting and in discriminating the different types of material. These algorithms are now operative and are part of the official software of reconstruction and analysis within the project. At the time when the present thesis is being written the *Muon Portal* project is in full construction phase, and it is expected to be completed and operational around mid-2015.

Bibliography

[1] V. F. Hess. The electrical conductivity of the atmosphere and its causes. by v. f. hess. translated from the german by l. w. codd. london (constable co.), 1928. 8vo. pp. xviii. and 204. 12s. *Quarterly Journal of the Royal Meteorological Society*, 55(231):320–322, 1929. (Cited on page 7)

[2] Carl Anderson. The positive electron. *Phys. Rev.*, 43:491–494, Mar 1933. (Cited on page 8)

[3] Carl Anderson and Seth Neddermeyer. Cloud chamber observations of cosmic rays at 4300 meters elevation and near sea-level. *Phys. Rev.*, 50:263–271, Aug 1936. (Cited on page 8)

[4] C.M.G. Lattes, H. Muirhead, G.P.S. Occhialini, and C.F. Powell. Processes involving charged mesons. *Nature*, 159:694–697, 1947. (Cited on page 8)

[5] G.D. Rochester and C.C. Butler. Evidence for the Existence of New Unstable Elementary Particles. *Nature*, 160:855–857, 1947. (Cited on page 8)

[6] Kenneth Greisen. End to the cosmic-ray spectrum? *Phys. Rev. Lett.*, 16:748–750, Apr 1966. (Cited on page 16)

[7] Zatsepin G.T. and Kuzmin V.A. Upper limit of the spectrum of cosmic rays. *JETP Lett.*, 4:78–80, 1966. (Cited on page 16)

[8] B.B. Rossi. *Cosmic Rays: A Dramatic and Authoritative Account*. McGraw-Hill, 1964. (Cited on page 17)

[9] W. L. Kraushaar, G. W. Clark, G. P. Garmire, R. Borken, P. Higbie, V. Leong, and T. Thorsos. High-Energy Cosmic Gamma-Ray Observations from the OSO-3 Satellite. , 177:341, November 1972. (Cited on page 17)

[10] M.T. Dova. Survey of the Pierre Auger Observatory. pages 699–702, 2001. (Cited on pages 18 and 160)

[11] Pierre Auger Collaboration. Correlation of the highest-energy cosmic rays with nearby extragalactic objects. *Science*, 318:938–943, Nov 2007. (Cited on page 18)

[12] Peter K.F. Grieder. Preface. In Peter K.F. Grieder, editor, *Cosmic Rays at Earth*, pages v – vii. Elsevier, Amsterdam, 2001. (Cited on pages 21 and 24)

[13] H. Tanaka, K. Nagamine, N. Kawamura, S.N. Nakamura, K. Ishida, and K. Shimomura. Development of a two-fold segmented detection system for near horizontally cosmic-ray muons to probe the internal structure of a volcano. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 507(3):657 – 669, 2003. (Cited on pages 24 and 51)

[14] Thomas K. Gaisser. *Cosmic Rays and Particle Physics*. Cambridge University Press, 1990. (Cited on page 27)

[15] Particle Data Group. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012. (Cited on page 29)

[16] Kanetada Nagamine. *Introductory Muon Science*. Cambridge University Press, 2003. Cambridge Books Online. (Cited on page 30)

[17] E.P. George. Cosmic rays to inspect large volumes. *Commonwealth Engineer*, 455, 1955. (Cited on page 31)

[18] LW Alvarez, JA Anderson, FE Bedwei, J Burkhard, A Fakhry, A Girgis, A Goneid, F Hassan, D Iverson, G Lynch, et al. Search for hidden chambers in the pyramids. *Science (New York, NY)*, 167(3919):832, 1970. (Cited on page 31)

[19] Eric T Wright. A detector for muon tomography: Data acquisition and preliminary results. (Cited on page 31)

[20] N. Lesparre, D. Gibert, J. Marteau, Y. Declais, D. Carbone, et al. Geophysical muon imaging: feasibility and limits. *Geophys.J.Int.*, 183:1348–1361, 2010. (Cited on page 32)

[21] J. Marteau, D. Gibert, N. Lesparre, F. Nicollin, P. Noli, and F. Giacoppo. Muons tomography applied to geosciences and volcanology. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 695(0):23 – 28, 2012. New Developments in Photodetection - NDIP11. (Cited on pages 32 and 36)

[22] H.A. Bethe. Moliere's theory of multiple scattering. *Phys.Rev.*, 89:1256–1266, 1953. (Cited on page 32)

[23] Jourde Kevin, Gibert Dominique, Marteau Jacques, de Bremond d'Ars Jean, Gardien Serge, Girerd Claude, Ianigro Jean-Christophe, and Carbone Daniele. The muon tomography diaphane project : recent upgrades and measurements. *Earth Planets Space*, 62:153–165, 2010. (Cited on page 36)

[24] Dominique Gibert, François Beauducel, Y. Declais, Nolwenn Lesparre, J. Marteau, F. Nicollin, and A. Tarantola. Muon tomography: Plans for observations in the lesser antilles. *Earth Planets Space*, 62:153–165, 2010. (Cited on pages 36 and 37)

[25] H.K.M. Tanaka, K. Nagamine, S.N. Nakamura, and K. Ishida. Radiographic measurements of the internal structure of mt. west iwate with

near-horizontal cosmic-ray muons and future developments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1-2):164 – 172, 2005. (Cited on page 37)

[26] G. Ambrosi, F. Ambrosino, R. Battiston, A. Bross, S. Callier, F. Cassese, G. Castellini, R. Ciaranfi, F. Cozzolino, R. DAlessandro, C. de La Taille, G. Iacobucci, A. Marotta, V. Masone, M. Martini, R. Nishiyama, P. Noli, M. Orazi, L. Parascandolo, P. Parascandolo, G. Passeggio, R. Peluso, A. Pla-Dalmau, L. Raux, R. Rocco, P. Rubinov, G. Saracino, G. Scarpato, G. Sekhniaidze, P. Strolin, H.K.M. Tanaka, M. Tanaka, P. Trattino, T. Uchida, and I. Yokoyamao. The mu-ray project: Volcano radiography with cosmic-ray muons. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 628(1):120 – 123, 2011. VCI 2010 Proceedings of the 12th International Vienna Conference on Instrumentation. (Cited on page 37)

[27] H.K.M. Tanaka, T. Nakano, S. Takahashi, J. Yoshida, and K. Niwa. Development of an emulsion imaging system for cosmic-ray muon radiography to explore the internal structure of a volcano, mt. asama. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 575(3):489 – 497, 2007. (Cited on page 37)

[28] K Nagamine, M Iwasaki, K Shimomura, and K Ishida. Method of probing inner-structure of geophysical substance with the horizontal cosmic-ray muons and possible application to volcanic eruption prediction. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 356(2-3):585 – 595, 1995. (Cited on page 38)

[29] P.M Jenneson. Large vessel imaging using cosmic-ray muons. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 525(1-2):346 – 351, 2004.

Proceedings of the International Conference on Imaging Techniques in Subatomic Physics, Astrophysics, Medicine, Biology and Industry. (Cited on page 39)

[30] D.F. Mahon, A. Clarkson, D.J. Hamilton, M. Hoek, D.G. Ireland, J.R. Johnstone, R. Kaiser, T. Keri, S. Lumsden, B. McKinnon, M. Murray, S. Nutbeam-Tuffs, C. Shearer, C. Staines, G. Yang, and C. Zimmerman. A prototype scintillating-fibre tracker for the cosmic-ray muon tomography of legacy nuclear waste containers. *Nuclear Instruments and Methods in Physics Research. Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 732:408–411, December 2013. (Cited on page 40)

[31] W.B. Gilboy, P.M. Jenneson, S.J.R. Simons, S.J. Stanley, and D. Rhodes. Muon radiography of large industrial structures. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 263(1):317 – 319, 2007. Industrial Radiation and Radioisotope Measurement Applications Proceedings of the 6th Topical Meeting on Industrial Radiation and Radioisotope Measurement Applications 6th Topical Meeting on Industrial Radiation and Radioisotope Measurement Applications. (Cited on page 41)

[32] L.J. Schultz and et Al. Image reconstruction and material z discrimination via cosmic ray muon radiography. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 519(3):687 – 694, 2004. (Cited on pages 41, 42 and 51)

[33] S. Pesente, S. Vanini, M. Benettoni, G. Bonomi, P. Calvini, P. Checchia, E. Conti, F. Gonella, G. Nebbia, S. Squarcia, G. Viesti, A. Zenoni, and G. Zumerle. First results on material identification and imaging with a large-volume muon tomography prototype. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 604(3):738 – 746, 2009. (Cited on pages 43 and 51)

[34] M. Furlan, A. Rigoni, S. Vanini, G. Viesti, G. Zumerle, M. Benettoni, P. Checchia, F. Gonella, M. Pegoraro, G. Bonomi, D. Cambiaghi, L. Dassa, A. Donzella, M. Subieta, V. Villa, A. Zenoni, P. Zanuttigh, G. Calvagno, P. Calvini, and S. Squarcia. Muons scanner to detect radioactive source hidden in scrap metal containers: Mu-steel eu project. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, pages 601–606, Nov 2013. (Cited on page 43)

[35] S Chatrchyan, G Hmayakyan, V Khachatryan, AM Sirunyan, W Adam, T Bauer, T Bergauer, H Bergauer, M Dragicevic, J Erö, et al. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, 2008. (Cited on page 44)

[36] M. Aguilar-Benitez, J. Alberdi, M. Arneodo, K. Banicz, M. Benettoni, et al. Construction and test of the final CMS Barrel Drift Tube Muon Chamber prototype. *Nucl.Instrum.Meth.*, A480:658–669, 2002. (Cited on page 44)

[37] Kondo Gnanvo, Leonard V. Grasso III, Marcus Hohlmann, Judson B. Locke, Amilkar Quintero, and Debasis Mitra. Imaging of high-z material for nuclear contraband detection with a minimal prototype of a muon tomography station based on gem detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 652(1):16 – 20, 2011. Symposium on Radiation Measurements and Applications (SORMA) XII 2010. (Cited on pages 45 and 51)

[38] Kondo Gnanvo, Patrick Ford andJennifer Helsby andRichie Hoch and-Debasis Mitra, and Marcus Hohlmann. Performance expectations for a tomography system using cosmic ray muons and micro pattern gas detectors for the detection of nuclear contraband. *In IEEE Nucl. Sci. Symp.02*, 2008. (Cited on pages 45 and 51)

[39] Fabio Sauli. Gem: A new concept for electron amplification in gas detectors. *Nuclear Instruments and Methods in Physics Research Sec-*

tion A: Accelerators, Spectrometers, Detectors and Associated Equipment, 386(2):531–534, 1997. (Cited on page 46)

[40] P Abbon, E Albrecht, V Yu Alexakhin, Yu Alexandrov, GD Alexeev, MG Alekseev, A Amoroso, H Angerer, VA Anosov, B Badełek, et al. The compass experiment at cern. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 577(3):455–518, 2007. (Cited on page 46)

[41] G. Jonkmans, V.N.P. Anghel, C. Jewett, and M. Thompson. Nuclear waste imaging and spent fuel verification by muon tomography. *Annals of Nuclear Energy*, 53(0):267 – 273, 2013. (Cited on page 46)

[42] C Pugliatti, V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M Belluso, S Billotta, A A Blancato, D L Bonanno, G Bonanno, A Costa, G Fallica, S Garozzo, V Indelicato, P La Rocca, E Leonora, F Longhitano, S Longo, D Lo Presti, P Massimino, C Petta, C Pistagna, M Puglisi, N Randazzo, F Riggi, S Riggi, G Romeo, G V Russo, G Santagati, G Valvo, F Vitello, A Zaia, and G ZappalÃ . Design of a muonic tomographic detector to scan travelling containers. *Journal of Instrumentation*, 9(05):C05029, 2014. (Cited on page 50)

[43] P La Rocca, V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M Belluso, S Billotta, A A Blancato, D Bonanno, G Bonanno, A Costa, G Fallica, S Garozzo, V Indelicato, E Leonora, F Longhitano, S Longo, D Lo Presti, P Massimino, C Petta, C Pistagna, C Pugliatti, M Puglisi, N Randazzo, F Riggi, S Riggi, G Romeo, G V Russo, G Santagati, G Valvo, F Vitello, A Zaia, and G Zappala'. Search for hidden high-z materials inside containers with the muon portal project. *Journal of Instrumentation*, 9(01):C01056, 2014. (Cited on page 50)

[44] U Becciani, V Antonuccio, M Bandieramonte, F Belluomo, M Belluso, S Billotta, DL Bonanno, G Bonanno, A Costa, G Fallica, et al. The muon portal project: A dedicated muon detector for the inspection of shipping containers. In *Technologies for Homeland Security (HST), 2013*

IEEE International Conference on, pages 423–428. IEEE, 2013. (Cited on page 50)

[45] S Riggi, V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M Belluso, S Billotta, G Bonanno, B Carbone, A Costa, et al. A large area cosmic ray detector for the inspection of hidden high-z materials inside containers. In *Journal of Physics: Conference Series*, volume 409, page 012046. IOP Publishing, 2013. (Cited on page 50)

[46] Konstantin N. Borozdin, Gary E. Hogan, Christopher Morris and-William C. Priedhorsky, Alexander Saunders, Larry J. Schultz, and Margaret E. Teasdale. Surveillance: Radiographic imaging with cosmic-ray muons. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 422(277), 2003. (Cited on page 51)

[47] Vladimir Uniplast Factory. Ul, Grande Tverskaya, 77, Russia. (Cited on page 58)

[48] Fermilab National Laboratory. Batavia,Illinois, U.S.A. (Cited on page 58)

[49] AMCRYS-H. Lenin ave, Kharkov, 310001, Ukraine. (Cited on page 58)

[50] T Adam, E Baussan, K Borer, JE Campagne, N Chon-Sen, C de La Taille, N Dick, M Dracos, G Gaudiot, T Goeltzenlichter, et al. The opera experiment target tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 577(3):523–539, 2007. (Cited on page 58)

[51] V Senchyshyn, B Grynyov, S Melnychuk, V Lagutin, M Dracos, A Olchevski, Y Gornushkin, A Nozdrin, and A Sadovski. Influence of polystyrene scintillator strip methods of production on their main characteristics. *Radiation Measurements*, 42(4):911–914, 2007. (Cited on page 58)

[52] Saint Gobain Crystals. Scintillating Product, Plastics Scintillating Fibers Brochure - http://www.detectors.saint-gobain.com. (Cited on page 59)

[53] Methacrylic Resin Division Kuraray Co. Methacrylic Resin Division, 8F, Maruzen Building, 3Ð10, 2-Chrome, Hihonbashi, Chuo-ku, Tokyo, 103-0027, Japan - http://kuraraypsf.jp/psf/ws.htm. (Cited on page 59)

[54] G.V. Russo, D. Lo Presti, D. Bonanno, V. Indelicato, P. La Rocca, E. Leonora, F. Longhitano, C. Pugliatti, N. Randazzo, F. Riggi, G. Santagati, G. Zappalá, and A.A. Blancato. Strip detectors for a portal monitor application. *Journal of Instrumentation*, 9(11):P11008, 2014. (Cited on page 61)

[55] F Zappa, S Tisa, A Tosi, and S_ Cova. Principles and features of single-photon avalanche diode arrays. *Sensors and Actuators A: Physical*, 140(1):103–112, 2007. (Cited on page 64)

[56] Massimo Mazzillo, Giovanni Condorelli, Delfo Sanfilippo, Giuseppina Valvo, Beatrice Carbone, Giorgio Fallica, Sergio Billotta, Massimiliano Belluso, Giovanni Bonanno, Luigi Cosentino, et al. Silicon photomultiplier technology at stmicroelectronics. *Nuclear Science, IEEE Transactions on*, 56(4):2434–2442, 2009. (Cited on page 64)

[57] S Cova, M Ghioni, A Lacaita, C Samori, and F Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics*, 35(12):1956–1976, 1996. (Cited on page 64)

[58] D. Lo Presti, S. Aiello, D.L. Bonanno, G.A.P. Cirrone, E. Leonora, F. Longhitano, C. Pugliatti, N. Randazzo, F. Romano, G.V. Russo, M. Russo, C. Stancampiano, and V. Sipala. Offset: Optical fiber folded scintillating extended tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 737(0):195 – 202, 2014. (Cited on page 78)

[59] Bel Stewart Con. Belfuse - http://belfuse.com. (Cited on page 80)

[60] National Instruments. 1150 N Mopac Expy, Austin, TX 78759, U.S.A. - www.ni.com. (Cited on page 80)

[61] JSON protocol. JSON - http://www.json.org. (Cited on page 82)

[62] Adam module Advantech technologies. Adam - http://www.advantechdirect.com/eMarketingPrograms. (Cited on page 84)

[63] Advantech technologies. Adv - http://www.advantech.com/products/. (Cited on page 84)

[64] Antonio Puliafito, Angelo Cucinotta, Antonino Longo Minnolo, and Angelo Zaia. Making the internet of things a reality: The wherex solution. In *The Internet of Things*, pages 99–108. Springer, 2010. (Cited on page 85)

[65] U Becciani, A Costa, V Antonuccio-Delogu, G Caniglia, M Comparato, C Gheller, Z Jin, Mel Krokos, and P Massimino. Visivo–integrated tools and services for large-scale astrophysical visualization. *Publications of the Astronomical Society of the Pacific*, 122(887):119–130, 2010. (Cited on page 90)

[66] E. Sciacca, U. Becciani, A. Costa, F. Vitello, P. Massimino, M. Bandieramonte, M. Krokos, S. Riggi, C. Pistagna, and G. Taffoni. An integrated visualization environment for the virtual observatory: Current status and future trends. *Astronomy and Computing*, in press, 2014. (Cited on page 90)

[67] U Becciani, A Costa, M Comparato, G Caniglia, C Gheller, M Krokos, A Grillo, R Munzone, and F Vitello. Visivo: New integrated services. In *Astronomical Data Analysis Software and Systems XVIII*, volume 411, page 555, 2009. (Cited on page 90)

[68] QT. Qt cross-platform application - qt-project.org/. (Cited on page 90)

[69] VTK. The Visualization Toolkit - www.vtk.org/. (Cited on page 90)

[70] Karl-Heinz Kampert and Alan A Watson. Extensive air showers and ultra high-energy cosmic rays: a historical review. *The European Physical Journal H*, 37(3):359–412, 2012. (Cited on page 102)

[71] Pierre Auger, Paul Ehrenfest, Roland Maze, Jean Daudin, and Robley A Fréon. Extensive cosmic-ray showers. *Reviews of Modern Physics*, 11(3-4):288, 1939. (Cited on page 102)

[72] S Riggi, V Antonuccio-Delogu, M Bandieramonte, U Becciani, A Costa, P La Rocca, P Massimino, C Petta, C Pistagna, F Riggi, et al. Muon tomography imaging algorithms for nuclear threat detection inside large volume containers with the muon portal detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 728:59–68, 2013. (Cited on page 124)

[73] M. Bandieramonte. Muon tomography: Tracks reconstruction and visualization techniques. *Il nuovo cimento C*, 4:181–191, 2013. (Cited on pages 124 and 138)

[74] Geant4 collaboration. Geant4 - a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. (Cited on page 125)

[75] Rene Brun and Fons Rademakers. Root - an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1-2):81 – 86, 1997. New Computing Techniques in Physics Research V. (Cited on page 125)

[76] P.J.E. Peebles. *The Large-scale Structure of the Universe*. Princeton series in physics. Princeton University Press, 1980. (Cited on page 127)

[77] P.J.E. Peebles and M.J. Hauser. The large-scale structure of the universe. *ApJS*, 28(19), 1974. (Cited on page 127)

[78] Marc Davis and P.J.E. Peebles. A Survey of galaxy redshifts. 5. The Two point position and velocity correlations. *Astrophys.J.*, 267:465–482, 1982. (Cited on page 127)

[79] A.J.S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. *Astrophys.J.*, 417:19, 1993. (Cited on page 127)

[80] S.D. Landy and A.S. Szalay. The large-scale structure of the universe. *ApJ*, 412(L64), 1993. (Cited on page 127)

[81] Larry Joe Schultz, Gary Stephen Blanpied, Konstantin N Borozdin, Andrew M Fraser, Nicolas W Hengartner, Alexei V Klimenko, Christopher L Morris, C Oram, and Michael J Sossong. Statistical reconstruction for cosmic ray muon tomography. *Image Processing, IEEE Transactions on*, 16(8):1985–1993, 2007. (Cited on pages 129 and 130)

[82] R.C. Tryon. *Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality*. Edwards brother, Incorporated, lithoprinters and publishers, 1939. (Cited on page 131)

[83] Marc Davis, George Efstathiou, Carlos S. Frenk, and Simon D.M. White. The Evolution of Large Scale Structure in a Universe Dominated by Cold Dark Matter. *Astrophys.J.*, 292:371–394, 1985. (Cited on page 132)

[84] Cedric G. Lacey and Shaun Cole. Merger rates in hierarchical models of galaxy formation. 2. Comparison with N body simulations. *Mon.Not.Roy.Astron.Soc.*, 271:676, 1994. (Cited on page 132)

[85] Surhud More, Andrey V Kravtsov, Neal Dalal, and Stefan Gottlöber. The overdensity and masses of the friends-of-friends halos and universality of halo mass function. *The Astrophysical Journal Supplement Series*, 195(1):4, 2011. (Cited on page 132)

[86] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, September 1975. (Cited on pages 132 and 152)

[87] Eva Sciacca, Salvatore Spinella, Dino Ienco, and Paola Giannini. Annotated stochastic context free grammars for analysis and synthesis of proteins. In Clara Pizzuti, MarylynD. Ritchie, and Mario Giacobini, editors, *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, volume 6623 of *Lecture Notes in Computer Science*, pages 77–88. Springer Berlin Heidelberg, 2011. (Cited on page 132)

[88] H. Frigui and R. Krishnapuram. A robust competitive clustering algorithm with applications in computer vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(5):450–465, May 1999. (Cited on page 132)

[89] S. Bandyopadhyay, U. Maulik, and A. Mukhopadhyay. Multiobjective genetic clustering for pixel classification in remote sensing imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(5):1506–1511, May 2007. (Cited on page 132)

[90] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323, September 1999. (Cited on page 132)

[91] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967. (Cited on page 132)

[92] Zhexue Huang. A fast clustering algorithm to cluster very large categorical data sets in data mining. In *In Research Issues on Data Mining and Knowledge Discovery*, pages 1–8, 1997. (Cited on page 132)

[93] Leonard Kaufman and Peter J. Rousseeuw. *An Introduction to Cluster Analysis*, pages 320–331. John Wiley Sons, Inc., 2008. (Cited on pages 132 and 133)

[94] Raymond T. Ng and Jiawei Han. Efficient and effective clustering methods for spatial data mining. In *Proceedings of the 20th International*

Conference on Very Large Data Bases, VLDB '94, pages 144–155, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc. (Cited on page 133)

[95] StephenC. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967. (Cited on page 133)

[96] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988. (Cited on page 133)

[97] Martin Ester, Hans peter Kriegel, Jrg S, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996. (Cited on page 133)

[98] Alexander Hinneburg, Er Hinneburg, and Daniel A. Keim. An efficient approach to clustering in large multimedia databases with noise. pages 58–65. AAAI Press, 1998. (Cited on page 133)

[99] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. Birch: An efficient data clustering method for very large databases, 1996. (Cited on page 133)

[100] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. (Cited on page 134)

[101] Erica Kolatch. Clustering algorithms for spatial databases: A survey. (Cited on page 134)

[102] Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, and Prabhakar Raghavan. Automatic subspace clustering of high dimensional data. *Data Mining and Knowledge Discovery*, 11(1):5–33, 2005. (Cited on page 134)

[103] C Thomay, J J Velthuis, P Baesso, D Cussans, P A W Morris, C Steer, J Burns, S Quillin, and M Stapleton. A binned clustering algorithm to

detect high-z material using cosmic muons. *Journal of Instrumentation*, 8(10):P10013, 2013. (Cited on page 134)

[104] Marilena Bandieramonte, Vincenzo Antonuccio Delogu, Ugo Becciani, Alessandro Costa, Piero Massimino, Costantino Pistagna, Simone Riggi, Eva Sciacca, Fabio Vitello, Paola La Rocca, et al. Automated object recognition and visualization techniques for muon tomography data analysis. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, pages 517–522. IEEE, 2013. (Cited on page 138)

[105] Marilena Bandieramonte, Vincenzo Antonuccio Delogu, Ugo Becciani, Alessandro Costa, Piero Massimino, Costantino Pistagna, Francesco Riggi, Simone Riggi, Eva Sciacca, Fabio Vitello, Paola La Rocca, and Catia Petta. Clustering analysis for muon tomography data elaboration in the muon portal project. In *International workshop on Advanced Computing and Analysis Techniques in physics research (ACAT)*, 2014. (Cited on page 138)

[106] Mihael Ankerst, Markus M. Breunig, Hans peter Kriegel, and Jrg Sander. Optics: Ordering points to identify the clustering structure. pages 49–60. ACM Press, 1999. (Cited on page 138)

[107] D. Heck and J. Knapp. Upgrade of the Monte Carlo code CORSIKA to simulate extensive air showers with energies gt; 10**20-eV. 1998. (Cited on page 139)

[108] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl.Instrum.Meth.*, A262:444–450, 1987. (Cited on page 143)

[109] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004. (Cited on page 147)

[110] Jack Dongarra, Rolf Hempel, Tony Hey, and David W. Walker. MPI. (Cited on pages 151 and 162)

[111] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975. (Cited on pages 152 and 163)

[112] F. Riggi, A.A. Blancato, P. La Rocca, S. Riggi, and G. Santagati. An extensive air shower trigger station for the muon portal detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 764(0):142 – 149, 2014. (Cited on page 160)

[113] Joshua Dolence and Robert J Brunner. Fast two-point correlations of extremely large data sets. (Cited on page 163)

[114] *KDD '12: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2012. ACM. 618122. (Cited on page 163)

[115] Rafael Ponce, Miguel Cardenas-Montes, Juan Jose Rodriguez-Vazquez, Eusebio Sanchez, and Ignacio Sevilla. Application of GPUs for the Calculation of Two Point Correlation Functions in Cosmology. *ASP Conf.Ser.*, 461:73, 2012. (Cited on page 163)

[116] Schultz and et al. *Cosmic Ray Muon Radiography for Contraband Detection.* Jul 2003.

[117] Thomas Hebbeker and Charles Timmermans. A Compilation of high-energy atmospheric muon data at sea level. *Astropart.Phys.*, 18:107–127, 2002.

[118] Anna Pla-Dalmau, Alan D. Bross, and Victor V. Rykalin. Extruding plastic scintillator at Fermilab. 2003.

[119] BicronCorp. 12345 Kinsman Road, Newbury, Ohio44065,U.S.A.

[120] A. Dyshkant, V. Rykalin, A. Pla-Dalmau, and D. Beznosko. Extruded scintillator for the calorimetry applications. *AIP Conf.Proc.*, 867:513–520, 2006.

[121] Yu.G. Kudenko, L.S. Littenberg, V.A. Mayatsky, O.V. Mineev, and N.V. Ershov. Extruded plastic counters with WLS fiber readout. *Nucl.Instrum.Meth.*, A469:340–346, 2001.

[122] G. Bonanno, D. Marano, M. Belluso, S. Billotta, A. Grillo, S. Garozzo, G. Romeo, and M.C. Timpanaro. Characterization measurements methodology and instrumental set-up optimization for new sipm detectors x2014;part i: Electrical tests. *Sensors Journal, IEEE*, 14(10):3557–3566, Oct 2014.

[123] G. Bonanno, D. Marano, M. Belluso, S. Billotta, A. Grillo, S. Garozzo, G. Romeo, and M.C. Timpanaro. Characterization measurements methodology and instrumental set-up optimization for new sipm detectors x2014;part ii: Optical tests. *Sensors Journal, IEEE*, 14(10):3567–3578, Oct 2014.

[124] V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M Belluso, S Billotta, A Blancato, DL Bonanno, G Bonanno, B Carbone, et al. Design of a large area tomograph to search for high-z materials inside containers by cosmic muons. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, pages 5–8. IEEE, 2012.

List of websites consulted

- Muon Portal Project – `www.muoni.oact.inaf.it:8080`
- Root – `www.root.cern.ch`
- Geant4 – `www.geant4.cern.ch`
- Google scholar – `www.scholar.google.com`
- Research Gate – `www.researchgate.net`
- Science Direct – `www.sciencedirect.com`
- Wikipedia – `www.wikipedia.org`

– Journal of Research in Science – www.onlinelibrary.wiley.com

– Pierre Auger Observatory experiment – www.auger.org

Publications

[1] Marilena Bandieramonte, Vincenzo Antonuccio Delogu, Ugo Becciani, Alessandro Costa, Piero Massimino, Costantino Pistagna, Simone Riggi, Eva Sciacca, Fabio Vitello, Paola La Rocca, et al. Automated object recognition and visualization techniques for muon tomography data analysis. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, pages 517-522. IEEE, 2013.

[2] Ugo Becciani, Eva Sciacca, Marilena Bandieramonte, Alberto Vecchiato, Beatrice Bucciarelli, and Mario G Lattanzi. Solving a very large-scale sparse linear system with a parallel algorithm in the gaia mission. In *High Performance Computing Simulation (HPCS), 2014 International Conference on*, pages 104-111. IEEE, 2014.

[3] Marilena Bandieramonte, Ugo Becciani, Alberto Vecchiato, M Lattanzi, and Beatrice Bucciarelli. Avu-gsr gaia mission: A hybrid solution for hpc and grid-mpi infrastructures. In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012 IEEE 21st International Workshop on*, pages 167-172. IEEE, 2012.

[4] Alberto Vecchiato, Ummi Abbas, Marilena Bandieramonte, Ugo Becciani, Luca Bianchi, Beatrice Bucciarelli, Deborah Busonero, Mario G Lattanzi, and Rosario Messineo. The global sphere reconstruction for the gaia mission in the astrometric verification unit. In *SPIE Astronomical Telescopes+ Instrumentation*, pages 84513C-84513C. International Society for Optics and Photonics, 2012.

[5] Marilena Bandieramonte, Vincenzo Antonuccio Delogu, Ugo Becciani, AlessandroCosta, Piero Massimino, Costantino Pistagna, Francesco Riggi, Simone Riggi, Eva Sciacca, Fabio Vitello, Paola La Rocca, and Catia Petta. Clustering analysis for muon tomography data elaboration in the muon portal project. In *International workshop on Advanced Computing and Analysis Techniques in physics research (ACAT)*, 2014.

[6] Eva Sciacca, Marilena Bandieramonte, Ugo Becciani, Alessandro Costa, Mel Krokos, Piero Massimino, Catia Petta, Costantino Pistagna, Simone Riggi, and Fabio Vitello. Visivo workflow-oriented science gateway for astrophysical visualization. In Parallel, Distributed and Network-Based Processing (PDP), 2013 21st Euromicro International Conference on, pages 164-171. IEEE, 2013.

[7] U Becciani, V Antonuccio, M Bandieramonte, F Belluomo, M Belluso, S Billotta, DL Bo- nanno, G Bonanno, A Costa, G Fallica, et al. The muon portal project: A dedicated muon detector for the inspection of shipping containers. In Technologies for Homeland Security (HST), 2013 IEEE International Conference on, pages 423-428. IEEE, 2013.

[8] M. Bandieramonte. Muon tomography: Tracks reconstruction and visualization techniques. Il nuovo cimento C, 4:181-191, 2013.

[9] C Pugliatti, V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M

Belluso, S Billotta, A A Blancato, D L Bonanno, G Bonanno, A Costa, G Fallica, S Garozzo, V Indelicato, P La Rocca, E Leonora, F Longhitano, S Longo, D Lo Presti, P Massimino, C Petta, C Pistagna, M Puglisi, N Randazzo, F Riggi, S Riggi, G Romeo, G V Russo, G Santagati, G Valvo, F Vitello, A Zaia, and G Zappala' . Design of a muonic tomographic detector to scan travelling containers. *Journal of Instrumentation*, 9(05):C05029, 2014.

[10] P La Rocca, V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M Belluso, S Billotta, A A Blancato, D Bonanno, G Bonanno, A Costa, G Fallica, S Garozzo, V Indelicato, E Leonora, F Longhitano, S Longo, D Lo Presti, P Massimino, C Petta, C Pistagna, C Pu- gliatti, M Puglisi, N Randazzo, F Riggi, S Riggi, G Romeo, G V Russo, G Santagati, G Valvo, F Vitello, A Zaia, and G ZappalÌ . Search for hidden high-z materials inside containers with the muon portal project. *Journal of Instrumentation*, 9(01):C01056, 2014.

[11] Ugo Becciani, Eva Sciacca, Alessandro Costa, Piero Massimino, Costantino Pistagna, Simone Riggi, Fabio Vitello, Catia Petta, Marilena Bandieramonte, and Mel Krokos. Science gateway technologies for the astrophysics community. *Concurrency and Computation: Practice and Experience*, pages n/a-n/a, 2014.

[12] P Massimino, A Costa, U Becciani, M Krokos, M Bandieramonte, C Petta, C Pistagna, S Riggi, E Sciacca, and F Vitello. Learning astrophysics through mobile gaming. In *Astronomical Society of the Pacific Conference Series*, volume 475, page 113, 2013.

[13] A Costa, M Bandieramonte, U Becciani, M Krokos, P Massimino, C Petta, C Pistagna, S Riggi, E Sciacca, and F Vitello. Visivo: A web-based, workflow-enabled gateway for astrophysical visualization. In *Astronomical Society of the Pacific Conference Series*, volume 475, page 109, 2013.

[14] Marilena Bandieramonte, Antonella Di Stefano, and Giovanni Morana. Grid jobs sche- duling: the alienated ant algorithm solution. *Multiagent and Grid Systems*, 6(3):225-243, 2010.

[15] V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M Belluso, S Billotta, A Blancato, DL Bonanno, G Bonanno, B Carbone, et al. Design of a large area tomograph to search for high-z materials inside containers by cosmic muons. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2012 IEEE, pages 5-8. IEEE, 2012.

[16] Ugo Becciani, Eva Sciacca, Alessandro Costa, Piero Massimino, Costantino Pistagna, Simone Riggi, Fabio Vitello, Catia Petta, Marilena Bandieramonte, and Mel Krokos. Science gateway technologies for the astrophysics community. *Concurrency and Computation: Practice and Experience*, 2014.

[17] S Riggi, V Antonuccio, M Bandieramonte, U Becciani, F Belluomo, M Belluso, S Billotta, G Bonanno, B Carbone, A Costa, et al. A large area cosmic ray detector for the inspection of hidden high-z materials inside containers. In *Journal of Physics: Conference Series*, volume 409, page 012046. IOP Publishing, 2013.

[18] E. Sciacca, C. Pistagna, U. Becciani, A. Costa, P. Massimino, S. Riggi, F. Vitello, M. Ban- dieramonte, and M. Krokos. Towards a big data exploration framework for astronomical archives. In *High Performance Computing & Simulation* (HPCS), 2014 International Conference on, pages 351-357. IEEE, July 2014.

[19] Marilena Bandieramonte, Antonella Di Stefano, and Giovanni Morana. An aco inspired strategy to improve jobs scheduling in a grid environment. In *Algorithms and Architectures for Parallel Processing*, pages 30-41. Springer, 2008.

[20] S Riggi, V Antonuccio-Delogu, M Bandieramonte, U Becciani, A Costa, P La Rocca, P Massimino, C Petta, C Pistagna, F Riggi, et al. Muon tomography imaging algorithms for nuclear threat detection inside large volume containers with the muon portal detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 728:59-68, 2013.

[21] Antonella Di Stefano and Giovanni Morana. A bio-inspired scheduling algorithm for grid environments. In *Remote Instrumentation Services on the e-Infrastructure*, pages 113-128. Springer, 2011.

[22] Marilena Bandieramonte and Antonella Di Stefano. Evaluating the robustness of the alienated ant algorithm in grids. In *Enabling Technologies: Infrastructures for Collaborative Enterprises* (WETICE), 2010 19th IEEE International Workshop on, pages 272-274. IEEE, 2010.

[23] Eva Sciacca, Marilena Bandieramonte, Ugo Becciani, Alessandro Costa, Mel Krokos, Piero Massimino, Catia Petta, Costantino Pistagna, Simone Riggi, and Fabio Vitello. Visivo science gateway: a collaborative environment for the astrophysics community. In IWSG, 2013.

ACKNOWLEDGEMENTS

I am very attached to the character of Ulysses, who with his words: *"You were not born to live like brutes, but to pursue virtue and knowledge"*, has been exhorting me since when I was attending "Liceo Classico G. Verga of Adrano" secondary school. Every morning I used to see that epigraph at the entrance of the school and since then it resonates in my mind. I believe that the desire for knowledge is what pushed me so far, after many years.

I believe in knowledge as interdisciplinarity and as relation between different expertise. My training started from classical studies, developed with engineering studies before moving to the world of physics. I believe it was not a coincidence. I believe in knowledge as a common good and as a path to freedom and liberation. Through the words of Ulysses is felt that knowledge has neither age nor limits. Ulysses and his are now old. But it is worth leaving again. ALWAYS.

This doctoral experience was for me another stage of my journey, which led me to know and work with exceptional people from the professional point of view but, above all, from the human one.

First and foremost I want to thank my two advisors, starting with prof. Francesco Riggi. It has been an honor to be his Ph.D. student. He has taught me how good experimental physics is done and what does the words dedication and passion for work mean. I appreciate all his contributions of time, ideas, patience and passion.

I am very thankful to the other supervisor Ing. Ugo Becciani. To be part of his working group has been one of the most stimulating and productive experience of my Ph.D. The joy and enthusiasm he has for his research activity was contagious and motivational for me, even during tough times.

I would like to acknowledge also all the members of the *Muon Portal* collaboration that I have had the pleasure to work with or alongside of: Dept. of Physics & Astronomy of University of Catania, INAF- Astrophysical Ob-

servatory of Catania, StMicroelectronics s.r.l. Catania, Meridionale Impianti Welding technologies and Insirio SPA.

I would like to really thank all the members of the group at the Astrophysics Observatory of Catania: Fabio, Eva, Salvo, Simone, Alessandro, Piero e Costantino. They have contributed immensely to my personal and professional Ph.D. experience. The group has been a source of friendships as well as good advice and collaboration.

I am especially grateful to all the colleagues of the Observatory, friends now, and with whom I had the pleasure of sharing meals and breaks: Elisa, Katia, Mariachiara, Gaetano and Gianfranco. Thanks for enriching and stimulating discussions.

SPECIAL THANKS

My Ph.D. experience was made unique in large part due to the many friends that became a part of my life: Cristina, Fabio, Eva, Stefano, Daniele, Carlo. I am grateful for the time spent together, for all the lunches and dinners, for the sport activities organized together, for our memorable trip into the mountains (!), for all the laughs and for many other memories. Without this experience I would not have met all of you!

I would like to say a special thanks to all the friends that have always supported me, especially during the final stages of this Ph.D. experience.

Laura, you always find a way to push away the bad thoughts. Thank you for being always from my side. Giusi, always present and caring friend. Thank you for being here and thanks for that afternoon at the spa together! Cristina, can you imagine that when I met you I thought you were nasty? It was enough to attend a conference together and from there a friendship was born. We shared this journey end, in everything: joys and sorrows. Thanks for the laughs, the time spent together, our motto and our madness! Without you everything would have been much more difficult and less fun! Stefano, we started together this path of doctorate. I still remember the day of the exam, our tension, and the promise of a beer to celebrate in case of success .. here we are! We made it! Fabio, thanks for your sincere friendship. I know I can always count on you. Daniele, your care and your attention in taking care of others are astounding. Thanks for all your invaluable help and support. Giuseppe, although my phone decided to refuse you, you managed to make me feel your presence and your affection, thank you! Benjamin, though from very far away, you have been close as a few in this period. Thanks for being supportive and encouraging and for your comic emoticon-faces!

Lastly, I would like to thank my family for all their love and encouragement. For my parents Alfio and Tina, my grandmother Lina, my brother Giuseppe and my sister Laura.

Thank you for the Love.