UNIVERSITÀ
degli STUDI
di CATANIA

Dipartimento
di Fisica
e Astronomia
"Ettore Majorana"

PhD programme in physics

Giuseppe Piparo

Analysis of very forward neutral particle spectra with the LHCf experiment at the LHC

PhD thesis

Supervisor:
Chiar.ma Prof.ssa A. Tricomi

---

academic year 2023/2024

Contents

Abstract

Cosmic rays, especially ultra-high energy cosmic rays (UHECRs), have been the subject of extensive study due to their relevance to astrophysics and particle physics. Although substantial progress has been made through ground-based experiments, like the Pierre Auger and Telescope Array Observatories, several uncertainties remain due to the dependence on hadronic interaction models for data interpretation. These models are critical for understanding cosmic ray composition and the mechanisms behind their production and acceleration. The LHC-forward (LHCf) experiment at the Large Hadron Collider (LHC) plays a key role in addressing these uncertainties by providing valuable measurements of forward particle production in proton-proton and proton-ion collisions. This thesis presents a detailed overview of the LHCf experiment, its role in forward physics, and its recent contributions to the so-called "Muon Puzzle" in hadronic interaction models. Chapter 1 presents an overview of cosmic ray physics and the challenges in detecting UHECRs, including the simulation and experimental techniques used in extended air shower models. Special attention is given to the hadronic interaction models and their limitations, which are crucial for interpreting UHECR data. Chapter 2 discusses the Large Hadron Collider (LHC) and its experiments, with a focus on their results concerning the Muon Puzzle and the LHCf experiment, detailing its experimental apparatus, data acquisition systems, and published results. A major part of this work is the analysis of the production of the η meson in proton-proton collisions at $\sqrt{s} = 13$ TeV with the LHCf-Arm2 detector, detailed in Chapter 3. The production rate as a

function of the Feynman-x variable is measured, and the results are compared with predictions from widely used hadronic interaction models. The future prospects of this analysis, leveraging data collected by LHCf during the LHC RUN III at $\sqrt{s} = 13.6$ TeV, are discussed, highlighting the necessity for the calibration of the LHCf-Arm2 detector, described in Chapter 4, with a focus on the SPS H2 beam test, which assesses energy resolution, linearity, and systematic uncertainties affecting the detector absolute energy scale. Finally, Chapter 5 presents the conclusions of this thesis and outlines future perspectives, including two ongoing works, the measurement of the η/π^0 production ratio and the application of machine learning techniques to enhance the reconstruction of multiple calorimetric clusters in the LHCf experiment.

Chapter 1

Introduction

1.1 Cosmic ray physics

Cosmic rays are charged particles that arrive at Earth with relativistic energies from outer space. Victor Hess detected these particles for the first time in 1912 by using an air balloon which ascended to an altitude of 5 kilometres, where he recorded an increase in ionization levels with higher altitudes. From these observations, he proposed that a form of highly penetrating radiation entered the atmosphere from outer space. Today, it is understood that cosmic rays comprise a wide range of atomic nuclei and it is thought that they encompass all stable elements in their composition. The most abundant particles are protons (86%) and helium nuclei (11%), with heavy nuclei making up 1%, electrons 2%, and a small percentage of antimatter [1]. The relative chemical abundance of cosmic rays arriving at Earth is similar, though not identical, to the solar metallicity at low energies. The abundance of rare elements such as Lithium (Li), Beryllium (Be), and Boron (B) is significantly higher in cosmic rays than in the solar system, as shown in Figure 1.1 [2]. Additionally, the isotopic composition shows an excess in neutron-rich isotopes. These excesses can be attributed to spallation reactions, where cosmic ray nuclei interact with hydrogen in the interstellar medium. Specifically, Li, Be, and B are produced from the spallation of Carbon (C) or Oxygen (O), while elements like Scandium (Sc), Titanium (Ti), Vanadium (V), and Magnesium

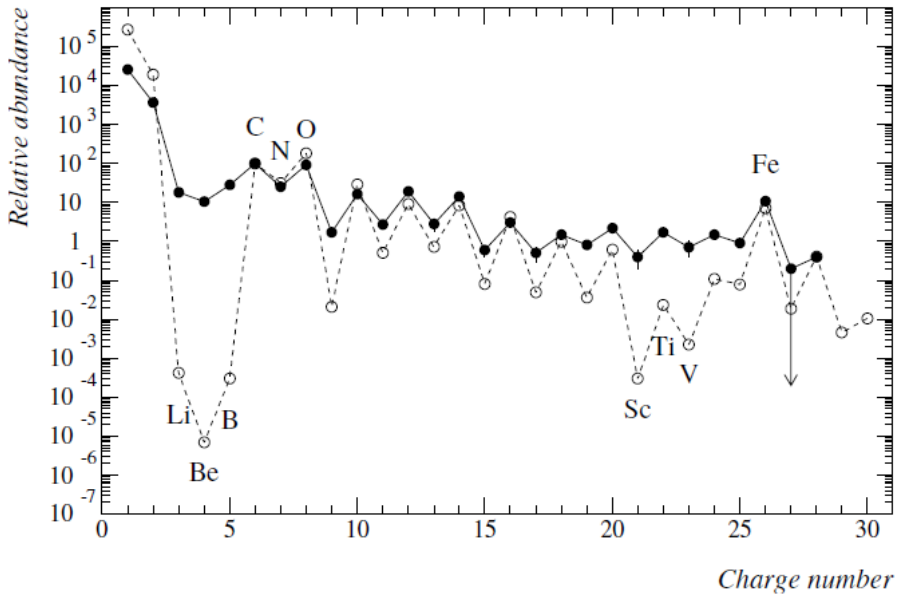

Figure 1.1: Comparison of solar abundances (open circles) and cosmic ray abundances (filled circles) across various elements, plotted as a function of atomic charge number. The y-axis represents the relative abundance on a logarithmic scale. The plot highlights the discrepancies between the abundances of certain elements in cosmic rays and those in the solar system. This enhancement is likely due to spallation processes where cosmic rays interact with the interstellar medium [2].

(Mg) are produced from the spallation of Iron (Fe). The arrival directions of cosmic rays on Earth are almost isotropic, particularly at lower energies, due to deflections caused by the galactic magnetic field, which ranges from 1 to 4 μG [3]. This deflection destroys the original paths of cosmic rays, making them appear uniformly distributed. For instance, a proton with a momentum of around 1 TeV/c has a radius of curvature far smaller than the typical galactic travel distance ($\approx 10^6$ parsecs [4]), causing significant scattering. However, at energies above the EeV scale, this deflection diminishes [5, 6, 7, 8], enabling the detection of anisotropies. This phenomenon is crucial for ultra-high energy cosmic ray (UHECR) astronomy, where the reduced scattering allows for the potential identifi-

cation of point-like sources. Cosmic rays present a spectrum of energies spanning below 10^9 eV to beyond 10^{20} eV. This is shown in Figure 1.2, where the differential flux [1] as a function of the energy, measured by different experiments, is shown. The particle with the highest energy ever reported was a cosmic ray observed in the 1995 by the Fly's Eye experiment [9], with an energy of (320 ± 90) EeV = $(3.2 \pm 0.9) \cdot 10^{20}$ eV [10]. Recently, the Telescope Array experiment [11] observed the second most energetic cosmic ray with an energy of $(244 \pm 29\ (\text{stat.})\ ^{+51}_{-76}\ (\text{syst.}))$ EeV = $(2.44 \pm 0.29\ (\text{stat.})\ ^{+0.51}_{-0.76}\ (\text{syst.})) \cdot 10^{20}$ eV [12]. The arrival direction of this particle points back to a void in the large-scale structure of the Universe. Possible explanations for this observation include a significant deflection by the foreground magnetic field, the presence of an unidentified source in the local extragalactic neighbourhood, or gaps in our current knowledge of particle physics. Understanding this observation is of fundamental importance for pointing and characterising astrophysical sources by exploiting flux anisotropies. The differential flux distribution of cosmic rays follows a power-law relationship given by:

$$F \propto E^{-\gamma}, \tag{1.1}$$

The spectral index, γ, varies depending on the energy range. It is possible to identify four discontinuities in the cosmic ray spectrum, including a steepening known as the "knee" at $E_{\text{knee}} \approx 10^{15.6}$ eV, a flattening called the "ankle" at $E_{\text{ankle}} \approx 10^{18.7}$ eV, another steepening referred to as the "second knee" at $E_{\text{2nd knee}} \approx 10^{17.2}$ eV (reported for the first time by the Kascade-Grande experiment [13]), and a rapid suppression known as the Greisen-Zatsepin-Kuzmin (GZK) cutoff around $E_{\text{GZK}} \approx 10^{19}$ eV [8]. The spectral index γ assumes different values in various energy regions, as described by the following:

$$\gamma \approx \begin{cases} -2.87 & \text{if } E > E_{\text{knee}} \\ -3.0 & \text{if } E_{\text{knee}} < E < E_{\text{2nd knee}} \\ -3.29 & \text{if } E_{\text{2nd knee}} < E < E_{\text{ankle}} \\ -2.53 & \text{if } E > E_{\text{ankle}} \end{cases}. \tag{1.2}$$

[1] The differential flux is defined as the number of particles crossing a surface normal to the direction of propagation per unit of time, area, solid angle and energy.

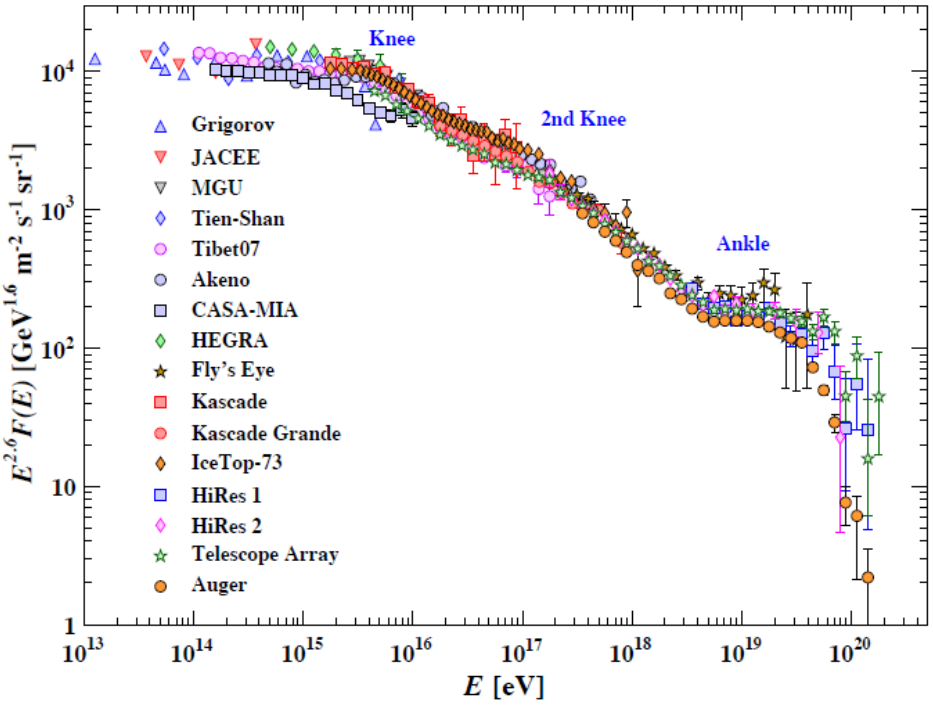

Figure 1.2: The differential flux of charged cosmic rays as a function of energy, measured by various experiments. The flux is multiplied by $E^{2.6}$ to compress the large scale and highlight slope changes. Key features labelled in the spectrum are the "Knee," "2*nd* Knee," and "Ankle," which mark significant changes in the flux slope. The high-energy end of the spectrum shows a suppression, attributed to the GZK cutoff.

These variations in γ reflect changes in the acceleration mechanisms and propagation effects at different energies. Cosmic rays in the same energy range exhibit a consistent acceleration and/or propagation mechanism, contrasting with those found in other energy regions. This observation allows for the development of models that can replicate the observed differential flux by assuming specific mechanisms for acceleration and propagation within each energy range. Four main astrophysical sources have been identified as candidates for the acceleration of high-energy cosmic rays ($E > 10^{12}$ eV): plasma clouds in the interstellar medium, shock waves

in supernova remnants, active galactic nuclei (AGN) and starburst galaxies (SGB) [14]. In the cases of plasma clouds and supernova shock waves, acceleration occurs via the Fermi mechanism [2], which involves statistical acceleration through multiple interactions. This mechanism is of the 2^{nd} order (Type I) for plasma clouds and the 1^{st} order (Type II) for shock waves. It successfully explains the power-law distribution observed in the energy flux of cosmic rays up to approximately 10^{15} eV. However, it fails to account for the existence of high-energy cosmic rays. Additionally, the galactic magnetic field is unable to confine particles with such energies, suggesting an extra-galactic origin for these particles. The confinement within the galaxy depends on the rigidity of the particles R[2], leading to an expected change in the composition of cosmic rays around the region of the knee. Above the ankle region, the contribution from extra-galactic sources becomes predominant. Candidates for accelerating such high-energy particles are thought to be AGNs and SBGs. As depicted in Figure 1.2, the differential flux of cosmic rays decreases dramatically with increasing energy. This suggests a progressive shutdown of low-energy acceleration mechanisms. Two primary models describe this transition: the mixed composition model [15] and the proton composition model [16]. In the mixed composition model, the ankle marks the transition between galactic and extra-galactic sources, while the knee indicates the energy limit of most galactic cosmic accelerators. In the proton composition model, the knee distinguishes galactic from extra-galactic cosmic rays, and the ankle represents the energy threshold for proton interactions with the cosmic microwave background radiation (CMBR), leading to electron-positron pair production:

$$p + \gamma \rightarrow e^- + e^+ + p. \tag{1.3}$$

The ankle region dip structure is a strong indicator that high-energy cosmic rays primarily consist of extra-galactic protons, with a negligible contribution from other nuclei [17]. This finding is crucial because it suggests that measuring high-energy cosmic ray composition can help determine

[2]The rigidity, denoted as R, is defined as $R = p/q$, where p represents the momentum and q is the charge of the particle. The gyroradius r_L of a charged particle with rigidity R in a magnetic field B is given by $r_L = R/B$.

which model is more accurate. Another significant region where composition measurements are critical is the one of the expected GZK cutoff around $5 \cdot 10^{19}$ eV. This cutoff energy corresponds to the threshold for the interaction of protons with CMBR photons, leading to pion production:

$$p + \gamma \rightarrow \begin{cases} n + \pi^+ \\ p + \pi^0 \end{cases} . \tag{1.4}$$

Given that the mean free path for this interaction is around 3 Mpc, which is significantly shorter than typical intergalactic distances, most ultra-high energy protons are expected to interact before reaching Earth. If cosmic rays in this energy region are primarily composed of extra-galactic protons, a cutoff at this energy is expected. However, this cutoff can also arise from photo-dissociation in the mixed composition model. Confirming measurements in this energy range is challenging due to the extremely low expected flux, approximately 1 particle per square kilometer per century. The first observation of the GZK cutoff is reported in [8]. Understanding the composition of high-energy ($E > 10^{15}$ eV) and ultra-high energy ($E > 10^{18}$ eV) cosmic rays is crucial not only for validating theoretical models like the proton composition model and the mixed composition model but also as an alternative method to studying the sources and acceleration mechanisms. By investigating the composition, it is possible to gain insights into the types of particles involved and their origins, complementing studies of anisotropies in the cosmic ray flux, which also aim to uncover the sources and mechanisms behind cosmic ray acceleration. The following sections will further explore these topics in detail, by discussing the detection techniques of cosmic rays in the different energy ranges, with a focus on the ultra-high energy region, and by providing insights on the obtained results and the observed discrepancies.

1.2 Cosmic rays detection techniques

At low energies, up to about 10^{15} eV, cosmic rays can be directly measured by detectors on satellites or high-altitude balloons. These direct measurements capture the particles before they interact with the Earth

atmosphere, allowing for precise energy and composition analysis. However, as energy increases, the flux of cosmic rays decreases significantly, necessitating different detection strategies for higher energies. Indeed, achieving the necessary large acceptance and extended data collection periods is impractical due to significant limitations in payload mass, volume, and lifetime. Stratospheric balloons (e.g., ATIC [18], CREAM [19]) and space satellite experiments (e.g., PAMELA [20], AMS-02 [21], CALET [22], DAMPE [23]) have been designed to measure cosmic rays below approximately 10^{12} eV for electrons and positrons, and 10^{14} eV for protons and nuclei. In earlier satellite experiments like PAMELA and AMS-02, the main detector was a magnetic spectrometer, whereas more recent experiments like CALET and DAMPE utilize an electromagnetic calorimeter. This shift is due to the improved resolution at high energies provided by calorimeters, where the energy resolution scales inversely with the square root of the energy compared to the linear dependence in spectrometers. Furthermore, calorimeters, unlike spectrometers, are not subject to stringent geometric constraints, allowing for optimization to maximize acceptance. Several recent projects, including HERD [24], are exploring new designs to further enhance performance. For a recent review of direct cosmic ray detection techniques and results see [25]. Despite potential advancements in energy resolution and geometric acceptance, extending direct detection of cosmic rays beyond 10^{15} eV remains currently unfeasible. Due to the aforementioned reasons, indirect methods are used to detect UHECRs by observing cascades of particles, the so-called extended air showers (EASs), initiated by cosmic rays hitting atmospheric nuclei, primarily Nitrogen (N) and Oxygen (O). By reconstructing the characteristics of EASs, it is possible to obtain information on the energy and composition of primary UHECRs. This approach allows the study of cosmic rays beyond the direct measurement limit, using the atmosphere itself as a detector. In the following sections, we will discuss how these measurements can be performed, the main results obtained and the observed discrepancies. In Section 1.2.1, the modelling behind EAS will be explained. In Section 1.2.2 we will outline the detection techniques that exploit the information from EASs to infer the characteristics of primary UHECRs. In Section 1.2.3, analysis methods and experimental results of hybrid ground-based experiments will be presented and discussed.

1.2.1 Extended air showers modelling

UHECRs generate EASs in the Earth atmosphere, with lateral spreads reaching tens of kilometers. Initially, the particles produced have a very narrow angular spread due to relativistic effects, but this spread increases as their energy decreases. These particles form a shower axis, characterized by two angles: the zenith angle (θ), which measures the axis inclination (0° being a vertical shower), and the azimuthal angle (ϕ). At any moment, the shower particles are concentrated in a thin, rapidly moving disc perpendicular to the shower axis, with the fastest particles (mainly muons) forming the shower front. The intersection of the shower axis with the ground is called the shower core. Air showers with energies above PeV exhibit a high degree of regularity, a concept known as shower universality [26]. While early-stage interactions introduce statistical fluctuations, the later stages of the shower become uniform due to the large number of particle interactions averaging out these fluctuations. An EAS can be fully characterized using measurements from only three spatially separated points. The angles θ and ϕ can be accurately determined through triangulation of arrival times, and the shower core and density profile can be inferred from local particle densities. Following the initial interaction, the EAS splits into two partially coupled components: the hadronic and electromagnetic showers. The hadronic shower involves interactions of long-lived hadrons, while the electromagnetic shower is driven by photons mainly produced from neutral pion and η decays. The electromagnetic shower progresses through electron pair production and bremsstrahlung. EASs produce various particles, including long-lived hadrons, electrons, photons, muons, and neutrinos. Most muons are generated at the end of the hadronic cascade when decays become more likely than further interactions, with a small fraction produced directly via pair production. Neutrinos, once produced, are decoupled from the shower. Some muons decay, contributing to the electromagnetic and neutrino components, while muons overall trace the hadronic shower development. A schematic representation of a typical EAS development is depicted in Figure 1.3. When charged particles from the EASs ionize air molecules, they lose energy. These ionized molecules eventually fall to their ground state, emitting energy in the form of fluorescence light.

This light, which radiates in all directions, can be detected by sensitive telescopes from several kilometers away during dark nights. The benefit of using fluorescence light is that it allows the observation of the shower from the side, enabling a single telescope to monitor a large air volume. Additionally, air showers produce Cherenkov light flashes that form a cone around the shower axis. Due to its smaller detection area, the Cherenkov light technique is primarily used for identifying sub-EeV air showers.

1.2.2 UHECR Detection techniques

A UHECR event is completely characterized by its energy (E), mass (A), and arrival direction, which is determined by the zenith (θ) and azimuthal (ϕ) angles, as well as its impact point (x, y) on the ground, known as the shower core. Unlike measurements obtained via balloons and satellites, the mass inferred from EAS characteristics is influenced by stochastic fluctuations due to inherent variations in the shower. These fluctuations can obscure the minor average differences between adjacent elements. Therefore, at energies above the PeV scale, the composition is often represented by a single value, the mean-logarithmic mass, denoted as $\langle \ln A \rangle$. The gold standard to observe a UHECR-induced EAS at the ground level is to combine the information from a large array of surface detectors and a few stations for measuring fluorescence light:

- **Surface detectors (SD)**: This technique was first pioneered by Kolhörster et al. in 1938 [27] and further developed by Auger in 1939 [28]. SDs permit the sample of the particles that reach the Earth's surface, offering a near-continuous observation cycle, allowing for precise determination of the shower arrival direction and the measurement of particle composition at the surface, particularly the number of muons, providing data on the EAS lateral size and, by extension, the primary particle energy. They also deduce the direction of the shower through the precise timing of the particles arrival. Since only a single slice of the shower longitudinal development is captured, energy measurements can be highly model-dependent. This issue can be mitigated by placing the ground array at high altitudes, bringing the observed slice closer to the shower maximum,

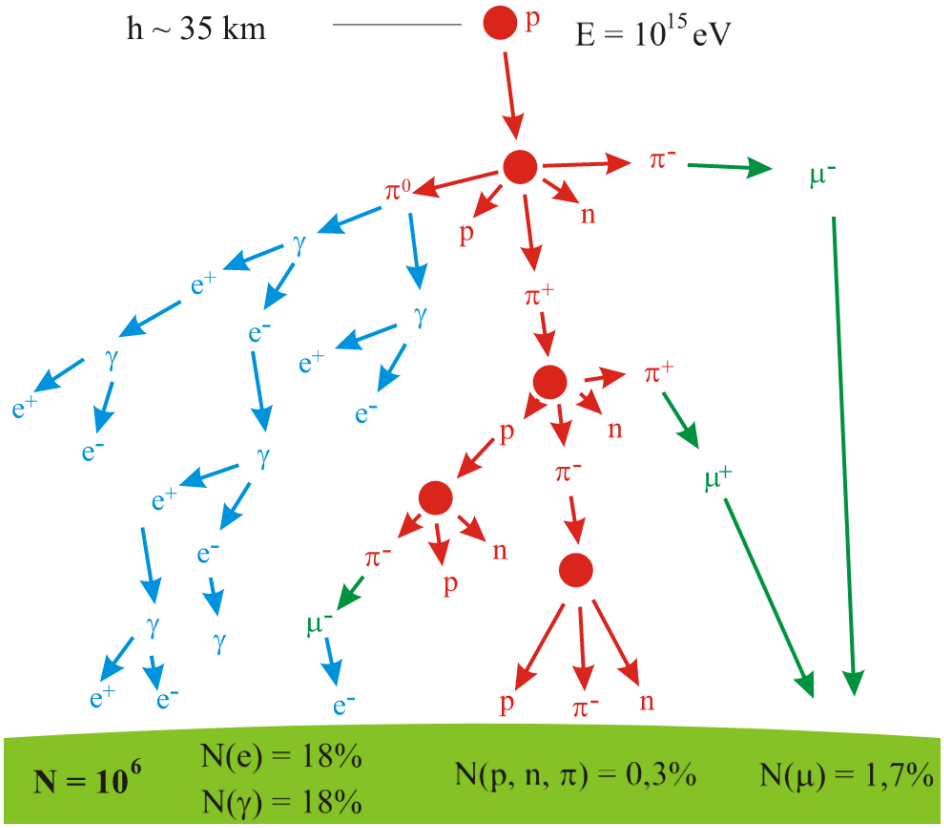

Figure 1.3: Schematic of an Extensive Air Shower (EAS) initiated by a high-energy cosmic ray interacting with the Earth atmosphere. The diagram shows the development of hadronic (red) and electromagnetic (blue) showers, resulting in the production of various particles such as protons (p), neutrons (n), pions (π), muons (μ), electrons (e^-), positrons (e^+), and photons (γ).

increasing the detectable particle density and reducing systematic uncertainties in energy measurement.

- **Fluorescence detector (FD)**: The use of fluorescence telescopes to observe air showers was first implemented by the Fly's Eye experiment [9]. FDs provide a near-calorimetric measurement of the shower energy because a significant portion of the kinetic energy of

UHECRs is transferred to the atmosphere, with a predictable fraction being emitted as detectable faint ultraviolet light. The EAS development over time can also give information on the mass composition, as explained later. However, the duty cycle of these detectors is limited to around 10% since it requires observations during dark nights.

Both methods can generally be powerful enough to determine the primary energy and composition of UHECRs, but it is possible to obtain a three-dimensional reconstruction of the EAS by combining the information given by the two kinds of detectors. There currently are two big state-of-art experiments dedicated to observing UHECRs by using both techniques, the Pierre Auger Observatory (PAO) [29] and the Telescope Array (TA) [11] experiment. PAO is the largest ground-based experiment currently operating. It is located on the Pampa, near Malargüe in the province of Mendoza, Argentina, at an altitude of about 1,400 meters. It is composed of 1660 water-Cherenkov detector stations arranged in a triangular grid with 1.5 km spacing, covering an area of $\approx$ 3000 km^2, and 27 fluorescence telescopes positioned at four peripheral locations. The water-Cherenkov stations are highly sensitive to muons and, due to their 1.2 m height, can also detect horizontal air showers. The AMIGA array features a denser infill array with 750 m spacing and buried scintillator detectors to measure the isolated muon component in vertical showers [30]. This infill array is monitored by high-elevation telescopes optimized for low-energy showers. The ongoing AugerPrime upgrade includes equipping the water-Cherenkov stations with scintillators. This upgraded surface detector array will enable model-independent measurements of the muon content in vertical showers with a 100% duty cycle [31]. A schematic view of the PAO experimental apparatus is shown in Figure 1.4. TA is located in the western desert of Utah, USA, comprising a $\approx$ 762 km^2 array of 507 scintillator detectors spaced 1.2 km apart, overseen by fluorescence three telescope stations. Due to the lack of acceptance for horizontal air showers by scintillators, stations positioned far from the shower axis are used to measure muon density, achieving a muon purity of 70% [33]. TA features a low-energy extension known as TALE, which measures cosmic ray flux and composition from 2 PeV to 2

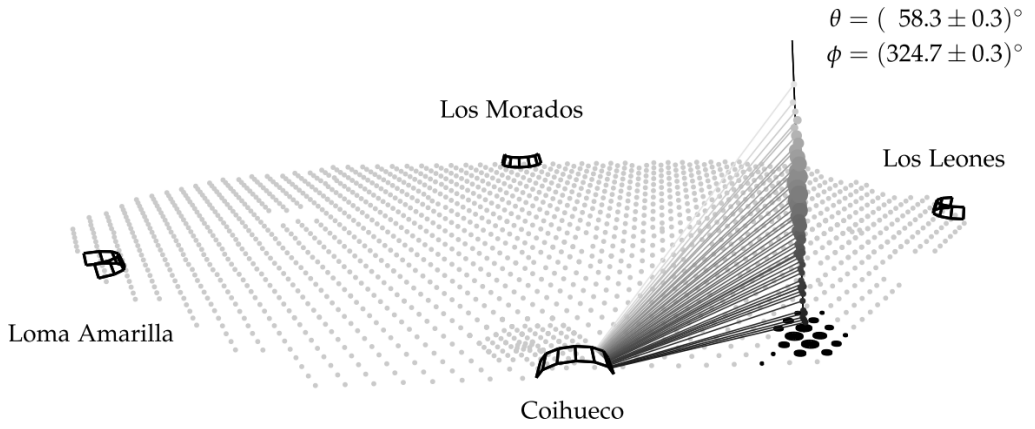

Figure 1.4: Schematic view of the Pierre Auger Observatory (PAO) experimental apparatus. The array consists of 1660 water-Cherenkov detector stations (grey dots) arranged in a triangular grid with 1.5 km spacing, covering an area of approximately 3000 km^2. The array is overseen by 27 fluorescence telescopes located at four peripheral sites: Los Leones, Coihueco, Los Morados, and Loma Amarilla. The diagram illustrates the detection of a cosmic-ray air shower. Figure from [32].

EeV [34, 35]. Additionally, a non-imaging optical array named NICHE is being constructed [36], and the main array is undergoing an upgrade to TA*4, expanding its coverage area by four times to enhance acceptance for energies above 50 EeV [37]. A schematic view of the TA experiment is shown in Figure 1.5.

1.2.3 Experimental results of ground-based experiments

Ground-based hybrid observatories can measure the energy of the primary UHECR by combining data from different detectors and using different types of parameterizations. As it will be discussed in Section 1.3, Hadronic interaction models are essential to relate the observable parameters of EASs to the primary particle energy. In the PAO and TA experiments, both SDs and FDs are employed for energy reconstruction. For FDs, the energy of the primary cosmic ray is derived by observing the

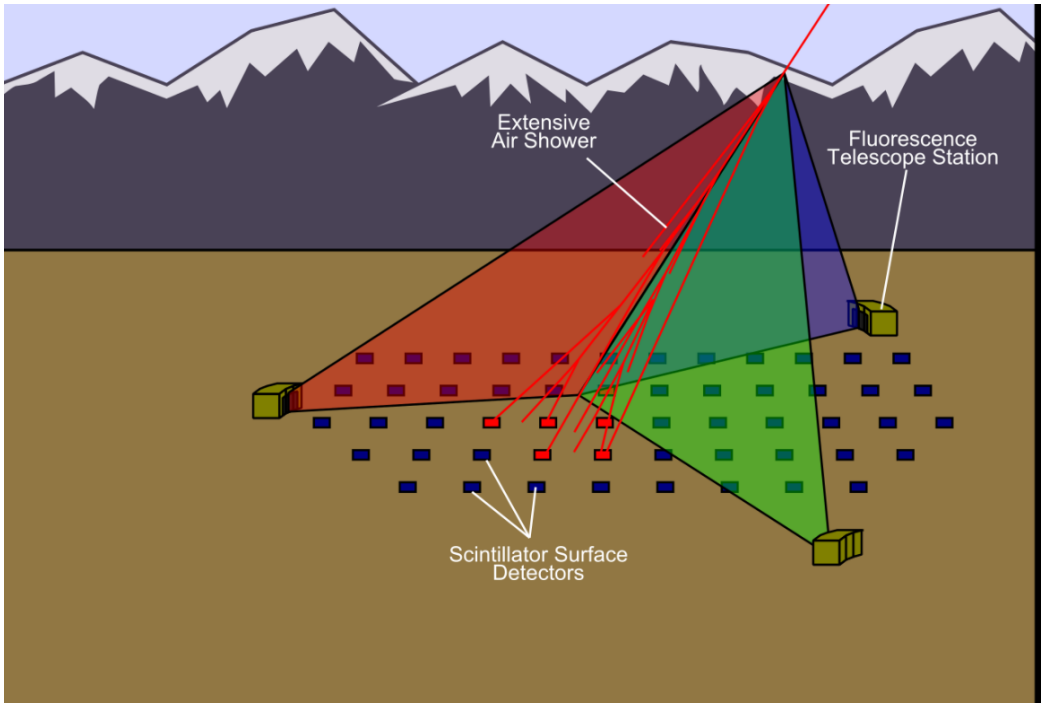

Figure 1.5: Schematic view of the Telescope Array (TA) experimental setup. The array consists of 507 scintillator surface detectors (blue and red squares) arranged in a grid with 1.2 km spacing, covering an area of ≈ 762 km^2. The detectors are overseen by three fluorescence telescope stations (yellow boxes) that monitor EASs. The figure illustrates the detection process, where the scintillator detectors measure the particle density on the ground, and the fluorescence telescopes capture the development of the shower in the atmosphere.

fluorescence light emitted in the sky, correcting for the fluorescence efficiency at different depths, reconstructing the longitudinal profile, fitting it with the Gaisser-Hillas function, and estimating the total atmospheric energy deposit. Due to the FD limited duty cycle ($\approx 10\%$) and the nearly continuous operation of the SD ($\approx 100\%$), most events use SD for energy estimation. This is performed in two distinct ways: for vertical events ($\theta < 60^\circ$) by measuring the lateral muon density at the ground, ρ_μ, which is correlated to the energy of the primary particle. For the inclined events

($62^\circ < \theta < 80^\circ$) the energy is determined by comparing the observed muon content at ground level to the predictions made by simulations for a proton primary with an energy of 10^{19} eV. The energy scale of SD is calibrated using events that are detected simultaneously by both the FD and the SD. The energy reconstruction from the FD is used as a reference to adjust and validate the SD energy measurements. A recent combined measurement of the energy spectrum of UHECR by PAO and TA is shown in Figure 1.6 and discussed in [38]. As with the results shown in Figure 1.2, the result is compatible with a GZK scenario, in which a pure-proton composition is assumed. Even if energy reconstruction by ground-based experiments requires hadronic interaction models, their effect on the final uncertainty is not as large as in the case of composition measurements. The determination of the UHECR composition is performed by using two different methods:

- The first method is performed using an estimator called X_{max} $[g/cm^2]$, which is the maximum depth of longitudinal shower development [39]. Indeed, in this context, we consider the interaction between a primary nucleus and the atmosphere as the superposition of single nucleon interactions, each one carrying a part of the primary energy, so it is naturally deduced that a nucleus-induced shower has a smaller X_{max} with respect to a proton-induced one. The estimator can be measured using FD detector, observing the light emission produced by the excitation and de-excitation of the Nitrogen molecules, which is correlated to the electromagnetic components in air-shower, that carry about 90% of the primary energy. Comparing the experimental mean and RMS of X_{max} with the hadronic interaction model results it's possible to establish how heavy the composition is.

- The second method exploits the study of the evolution of the hadronic component of the EAS by inferring the number of muons at ground N_μ, a quantity that provides critical insights into the primary cosmic ray composition [40]. This is because different types of primary particles (protons, heavy nuclei) interact with the atmosphere in distinct ways, leading to variations in the development of the EAS. Primaries like protons tend to produce showers with fewer

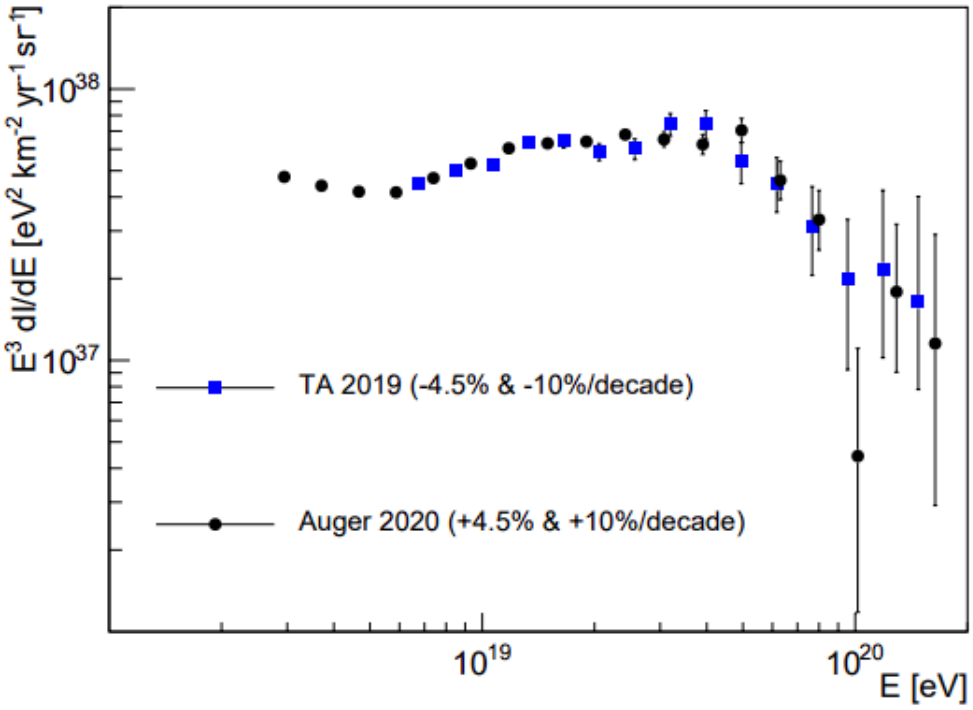

Figure 1.6: Comparison of the energy spectra of UHECR measured by PAO and TA. The plot shows data with adjustments of -4.5% and +10% per decade for TA, and +4.5% and +10% per decade for Auger, respectively. The measured spectra are compatible with a GZK scenario. For more details see [38].

muons compared to heavier nuclei. This is due to the fact that heavier nuclei have more nucleons, leading to more interactions and a higher probability of producing secondary particles that decay into muons. Consequently, showers initiated by heavier nuclei will generally have a higher muon content than those initiated by lighter particles. By measuring the muon density ρ_μ and integrating it to determine the number of muons N_μ it is possible to estimate the mean composition.

The two bands on the right panel of Figure 1.7 represent a range of mea-

surements conducted by various air shower experiments [39]. Composition estimates derived from N_μ measurements have particularly large systematic uncertainties and barely constrain the models of production and acceleration of UHECRs, reported for comparison in the left panel of Figure 1.7. Also, a tension between the results of the two techniques is observed, with the results obtained from the measurement of N_μ that at very high energies ($E > 10^{19}$) seem to favour a composition of UHECRs made up of heavy nuclei, in contrast to the GZK scenario favoured by the energy spectrum measurements. Moreover, The large error bands on the composition estimated by measuring N_μ are also problematic because this method better distinguishes between light and heavy primaries on a shower-by-shower basis at the EeV scale [41] and it is beneficial to gather large statistics, especially above $10^{19.5}$ eV, where observations of X_{max} with fluorescence telescopes become statistically poor. These telescopes can only be operated during dark nights, resulting in a duty cycle of about 10%, whereas muons can be observed with a duty cycle of 100%. These results do not allow an understanding of the origin of UHECRs, which is one of the main open problems in astroparticle physics. Most of the observed uncertainty is not due to experimental factors. Ground-based experiments are capable of making highly precise air shower measurements, with an experimental uncertainty in the measured value of N_μ that is about 10%, which is 2.5 to 4 times smaller (depending on the energy) than the width of the band depicted in Figure 1.7. Instead, the bulk of the uncertainty is theoretical, stemming from the air shower simulations used to derive the mean logarithmic composition from X_{max} and N_μ [42]. These simulations are crucial for interpreting EAS measurements, as there is no astrophysical source with a known mass composition in the sky that can be used for calibration. The uncertainty does not stem from the transport of particles in the atmosphere, as this process is relatively well understood. Indeed several air shower simulation programs are available and have been cross-checked with one another and they exhibit only minor differences of about 5% concerning N_μ [42]. Instead, The main source of uncertainty lies in the modelling of the hadronic interaction that governs the EAS development and the subsequent muon production, since showers are primarily driven by hadronic interactions involving small momentum transfer, which cannot be accurately calcu-

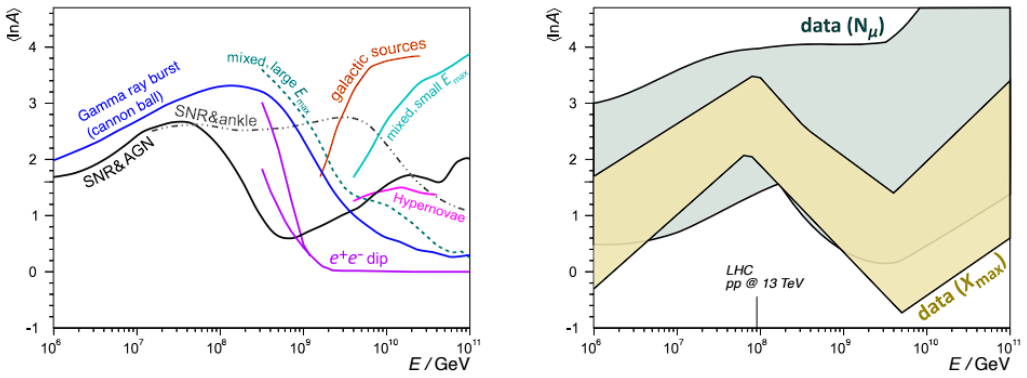

Figure 1.7: Left: Predictions of the mean-logarithmic mass <lnA> of cosmic rays as a function of their energy from several theories. Right: Two bands that cover the ranges of measurements, grouped by the mass-sensitive variable used (X_{max} or N_μ). The vertical line indicates the equivalent energy of p-p interaction at 13 TeV at the LHC. The width of the data bands is dominantly caused by theoretical uncertainties of forward hadron production. These uncertainties prevent the exclusion of theories on the origins and acceleration mechanism of UHECRs. Data and model lines were taken from [39].

lated using perturbative quantum chromodynamics (pQCD), rather effective theories and phenomenological approaches are employed to estimate the rates of these interactions and the spectra of secondary particles they generate. These approaches are implemented as software tools known as hadronic interaction models (HIM). Several HIM are actually in use, each based on different concepts and yielding varying predictions. It was observed that HIMs predict fewer muons than what is measured in EASs by ground-based hybrid experiments, and this discrepancy, which is called the "Muon Puzzle" [42], has a significant contribution to the uncertainties linked to the measurements of the composition of UHECRs. In the next Section 1.3 we describe the HIM, the effect of their discrepancies on the muon production in EAS on the mass composition measurements uncertainties and the possible solutions for the improvement of models.

1.3 Hadronic Interaction Models

1.3.1 Overview of Hadronic Interaction Models

HIMs are a fundamental component in astroparticle physics, providing a theoretical framework to predict the outcomes of hadronic interactions. These interactions are particularly crucial in the study of cosmic rays, where a primary particle from space collides with a nucleus in the Earth atmosphere producing an EAS. The accurate modelling of these air showers is essential for understanding the composition, energy, and therefore the production and acceleration mechanism of UHECRs. Several HIMs are widely used in astroparticle physics, including QGSJET [43, 44, 45, 46, 47], EPOS [48, 49, 50, 51], DPMJET [52, 53, 54, 55, 56] and SIBYLL [57, 58, 59, 60]. These models are constructed from a blend of theoretical calculations and experimental data, primarily sourced from particle accelerators like the Large Hadron Collider (LHC) [61]. The LHC, with its capability to reach unprecedented collision energies in proton-proton and proton-ion collision, has played an important role in refining the newest versions of HIMs, the so-called post-LHC models, improving their accuracy and extending their predictive power to cover a broader range of high-energy interactions. However, strong discrepancies are still present [62, 63, 64]. HIMs are involved in the prediction of various processes within EAS, such as the production of secondary particles, the fragmentation of nuclei, and the scattering of particles. These processes are governed by QCD. However, due to the non-perturbative nature of QCD at the energy scales relevant to these interactions, where the strong coupling constant α_S is large, the direct calculation of these processes from first principles is not feasible. Instead, these models rely on effective theories and phenomenological approaches. The parameters within these models are fine-tuned using experimental data, particularly those from high-energy proton and heavy-ion collisions. Each model, while based on the Gribov-Regge theories [65], adopts different approaches to model the "soft" QCD processes, which dominate hadronic interactions at low momentum transfer. These differences lead to variations in the predictions of secondary particle production, which are crucial for accurately modelling EASs. QGSJET and SIBYLL, for example, are primarily focused on simu-

lating air showers, with a limited set of parameters specifically tuned to replicate the physics that directly impacts shower development. In contrast, models like DPMJET and EPOS have a broader scope, capable of simulating a wide range of minimum-bias proton-proton and heavy ion collisions, making them applicable to both high-energy physics and astroparticle physics. These models involve larger parameter sets and are constrained by a more extensive set of experimental data. The predictive power of these models is critical, as they are used to extrapolate to center-of-mass energies that far exceed those achievable by current colliders, and to regions of phase space, particularly forward rapidities, that are not well covered by existing experiments. The challenge of simulating interactions at these extreme energies, coupled with the need to handle a variety of projectiles (including protons, nuclei, charged pions, and kaons) and targets (such as nitrogen, oxygen, and argon), underscores the complexity of these models. Moreover, there has been significant progress in bridging the gap between HIMs used in high-energy physics and those used in extensive air shower simulations. For instance, PYTHIA, a model widely used in electron and proton colliders, has recently expanded its focus to include heavy-ion collisions through the addition of the Argantyr model [66]. Although PYTHIA was not initially designed to frequently switch between different collision energies and particle types, a feature necessary for EAS simulations, there are ongoing efforts to address this limitation. PYTHIA has already been used in specialized studies comparing it with other EAS generators, as demonstrated in [67]. The continuous development of these models, particularly in light of new data from the LHC and other high-energy facilities, has significantly improved our ability to simulate and understand cosmic ray interactions. However, despite these advancements, HIMs remain the largest source of uncertainty in air shower simulations. Figure 1.8 provides a summary of the main features of these HIMs, highlighting the differences and commonalities among them ([42].

1.3.2 The Muon Puzzle in HIM

As introduced in Section 1.2.3, the Muon Puzzle arises from a discrepancy between the number of muons observed in EASs and the number

	DPMJET-III.19-1	EPOS-LHC	QGSJETII-04	SIBYLL2.3d	PYTHIA 8
Domain	EAS, HEP	EAS, HIC	EAS	EAS	HEP
Theoretical basis	GRFT + minijet	GRFT + energy sharing	GRFT + resummation	GRFT + minijet	parton model
Nuclear collisions	Glauber	extended GRFT	extended GRFT	extended superposition	Glauber via Argantyr
Pomeron	soft+hard	semi-hard	semi-hard	soft+hard	soft+hard
Energy evolution of parton densities	via $Q_0(s)$ cut	parameterised	Higher-order Pomeron graphs	via $Q_0(s)$ cut	via $Q_0(s)$ cut
Energy evolution of elasticity	constant	falling	falling	constant	—
Parton distributions	CT14	custom	custom	GRV	various
Non-diffractive remnant	—	multi-quark exchange (low to high mass)	one-quark exchange (low mass)	one-quark exchange (low mass)	low mass
Diffractive dissociation (low mass)	2-channel eikonal	diffractive Pomeron	3-channel eikonal	2-channel eikonal	Pomeron emission
Diffractive dissocation (high mass)	cut enhanced graphs	Pomeron exchange	cut enhanced graphs	Pomeron exchange	Pomeron exchange
String fragmentation (fitted data)	Lund (e^+e^-)	area law (e^+e^-)	custom (p-p)	Lund (p-p)	Lund (e^+e^-)
Forward-central correlation	weak	strong	strong	weak	strong
Charm production	pQCD (incomplete)	—	—	parameterised + intrinsic	pQCD
Collective effects	string fusion	core-corona (parameterised)	—	—	colour reconnection, rope hadronization, string shoving

Figure 1.8: Comparison of the theoretical approaches commonly used in HIMs [42].

predicted by HIMs. This suggests that the models are underestimating the production of muons in hadronic interactions, particularly at high energies. Solving the Muon Puzzle will likely involve enhancements to the HIMs, either by refining the existing models or by developing new models that can more accurately predict the production of muons. This will necessitate a combination of theoretical work, to deepen our understanding of the physical processes involved in hadronic interactions, and experimental work, to provide more precise data for model tuning and validation. In the context of the Muon Puzzle, the meta-analysis of muon measurements in EASs, from the Working group on Hadronic Interaction and Shower Physics (WHISP) which was formed in 2018 by members of

nine air shower experiments [42, 68, 69, 70], is particularly relevant. This analysis, which combines results from multiple experiments, provides a comprehensive view of the discrepancy between observed and predicted muon numbers in HIM, introducing a variable to make results from different experiments directly comparable, the logarithmic z-scale:

$$z = \frac{\ln N_\mu - \ln N_{\mu,p}}{\ln N_{\mu,Fe} - \ln N_{\mu,p}}, \tag{1.5}$$

where N_μ is the measured muon number while $N_{\mu,p}$ and $N_{\mu,Fe}$ are the corresponding simulated numbers (on detector-level) from EAS simulation with a particular HIM for pure proton and iron composition, respectively. The issue with this approach is that measurements can only be compared if air shower simulations with the same HIM exist. The muon number increases nearly linearly with shower energy, which implies that any systematic discrepancies in energy scales across different experiments affect the variable z. Differences in energy scales across experiments were corrected by using the uniformity of cosmic ray flux on Earth. This correction reduced the variation in z values. The results of this analysis are presented in Figure 1.9, where the z-scale muon productions of various experiments using different HIMs are shown. For comparison, the figure also includes the results from measurements of X_{max} and from the Global Spline Fit (GSF) model [71], a theoretical framework designed to provide a smooth, empirical fit to the muon production data, incorporating results from multiple experiments and energy ranges. For energies above 10 PeV the data points exceed the expected values with a significance level of 8σ. This significance is greater than that found in any single measurement, where the observed muon deficit significance does not exceed 3σ. This suggests that the discrepancy is not simply due to statistical fluctuations or systematic errors in the measurements, but rather to a fundamental issue with the HIMs. The meta-analysis shows that the HIMs tend to underestimate the production of muons, particularly at high energies. This indicates that the models may be missing some important aspects of hadronic interactions, such as the production of secondary particles or the fragmentation of nuclei. The number of muons produced in an EAS is sensitive to various parameters of hadronic interactions [72], including the inelastic cross-section (σ_{inel}), the hadron multiplicity (N_{mult}), the elas-

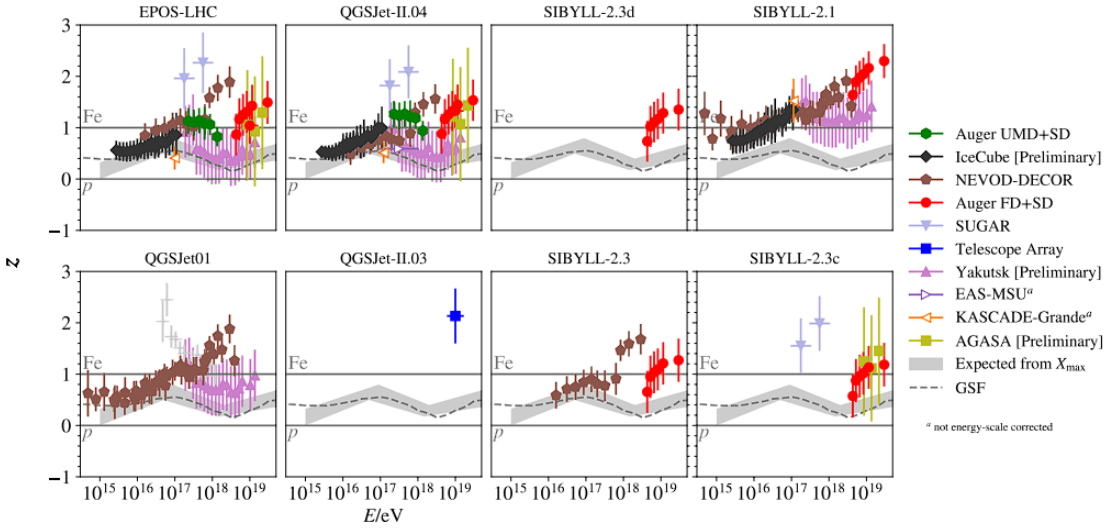

Figure 1.9: Measurements of muons produced in EASs from nine experiments after adjusting for energy-scale offsets (colored points) converted to the z-scale as a function of shower energy for different HIMs. Shown for reference are predicted z_{mass} values for proton and Fe primaries (solid lines) and from measurements of X_{max} (grey bands). GSF model predictions are reported as well for comparison (dashed lines). Figure from [70].

ticity (energy fraction carried by the most energetic particle ($E_{leading}/E_0$), and the ratio between the energy carried by photons from short-lived hadrons like the π^0 compared to the energy in long-lived hadrons (R). These parameters are particularly crucial for the most common π^0-air interactions in EASs. An ad-hoc model was proposed for EAS simulations where these parameters are adjusted during the simulation depending on the energy of the colliding hadron in the rest frame of the target [73]. This model uses the original predictions of a specific event generator as a baseline, which are scaled by an energy-dependent factor $f(E)$. This factor is set to 1 below a chosen energy threshold of 1 PeV and increases logarithmically above it. The reason is that while generators are well-constrained by accelerator data at low energies, they tend to diverge logarithmically when extrapolated to higher energies, where accelerator data have poor statistics. The modification's magnitude is determined by the parameter

f_{19}:

$$f(E) = 1 + (f19 - 1) \cdot \begin{cases} 0 & \text{if } E < 1 \text{ PeV} \\ \frac{\log_{10}(E/1 \text{ PeV})}{\log_{10}(10 \text{ EeV}/1 \text{ PeV})} & \text{if } E \geq 1 \text{ PeV} \end{cases}, \tag{1.6}$$

where $f(E)$ represents the modification factor for a hadron with an energy of 10 EeV = 10^{19} eV, which is an arbitrary scale. This equation is applied to each hadron collision within an air shower simulation to modify the respective parameters. Large modifications of $f(E)$ could suggest exotic modifications of QCD and might conflict with more recent LHC measurements, whereas small deviations may remain within the bounds of conventional scenarios and be consistent with LHC data. The results are presented in Figure 1.10 for the mean and standard deviation of the muon number N_μ and the depth of the shower maximum X_{max} for a $10^{19.5}$ eV primary proton shower as a function of the modification factor at the LHC energy scale, using the SIBYLL-2.1 model. In the original work, the π^0 fraction is used instead of the energy ratio R, however, the numerical effects of modifying this fraction and R are similar. The most effective way to increase the muon number in EASs is by reducing the π^0 fraction. A 10% reduction leads to a 13% increase in N_μ. Changes to the inelastic cross-section and elasticity have negligible effects on N_μ. The impact on the standard deviation of N_μ is also significant, and this has recently been measured for the first time by the Pierre Auger Observatory [64]. The data showed reasonable agreement with the post-LHC models EPOS-LHC, QGSJET II-04, and SIBYLL 2.3d, placing strong constraints on changes to elasticity, which is the only parameter among the four considered with a substantial impact on N_μ fluctuations. EAS simulations based on post-LHC HIM provide a reasonable description of X_{max}, indicating that parameters influencing X_{max} cannot deviate much from current models without losing consistency. X_{max} is highly sensitive to the inelastic cross-section, well-measured at the LHC, though some uncertainty remains when extrapolated to p-air and π-air interactions. The standard deviation of X_{max} is even more sensitive to the inelastic cross-section, with small to no impact from changes in other parameters. In Figure 1.11 these individual results are combined [74, 75] . The impact of changes in hadron multiplicity N_{mult} and the energy ratio R on the means of the

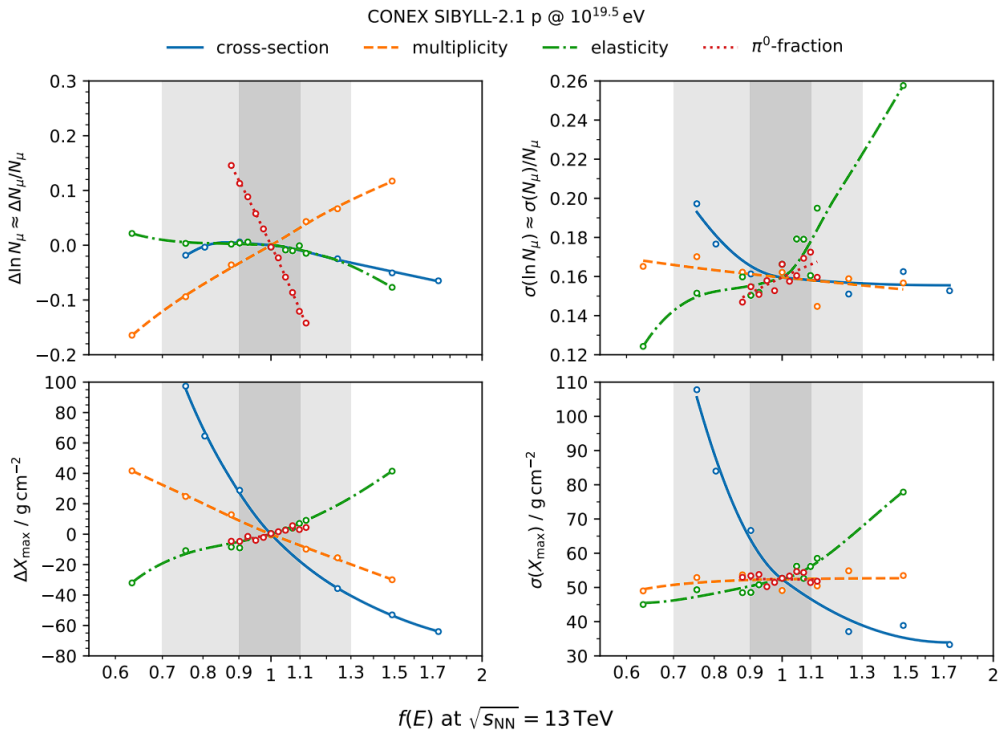

Figure 1.10: The alterations to the inelastic cross-section, the hadron multiplicity, the elasticity (energy proportion carried by the most energetic particle), and the proportion of neutral pions produced, influence the muon number N_μ and its variations, as well as the peak depth of the shower and its fluctuations, for a proton shower simulated at $10^{19.5}$ eV with SIBYLL-2.1 as the standard. These changes are presented in relation to the energy-dependent scale factor at the LHC energy level of 13 TeV. The points depict the results of the simulations, and the lines are heuristic fits to provide a visual guide. Figure from [42].

logarithm of the muon number and X_{max} is shown for 10^{19} eV primary and compared to measurements by the Pierre Auger Observatory. On this double-logarithmic scale, any potential mass composition of cosmic rays between the extremes of pure proton showers (bottom right) and pure iron showers (top left) produces a point on a straight line. The standard prediction by EPOS-LHC is indicated by a grey line (visible under the

other colored lines). The muon deficit in simulations explains why this line does not coincide with the data point. In addition to the standard prediction by EPOS-LHC, ad-hoc modified predictions are also shown. Changing the multiplicity (blue and red lines) shifts the lines along themselves and has no potential to solve the Muon Puzzle, while altering the ratio of energy going into neutral pions, or equivalent, a reduction in the ratio between the electromagnetic energy flow and the hadronic energy flow (yellow and green lines) has an effect perpendicular to the lines, significantly impacting data interpretation. This strongly suggests that the solution to the Muon Puzzle lies in modifying the energy ratio R. The results in Figure 1.9 show that the discrepancy has an early onset, at about $4 \cdot 10^{16}$ eV, which corresponds to a center-of-mass energy $\sqrt{s} = 8$ TeV in the first interaction. This indicates that the origin of the discrepancy should be observable at the LHC. There are several key measurements from the LHC experiments, including the inelastic cross-section, the hadron multiplicity, and the fraction of neutral pions produced. The Large Hadron Collider, the experiments it hosts, the impact of their measurements in the calibration of HIMs and in the resolution of the Muon Puzzle will be discussed in the next Chapter 2, with a focus on the LHCf experiment, its experimental apparatus and its experimental results.

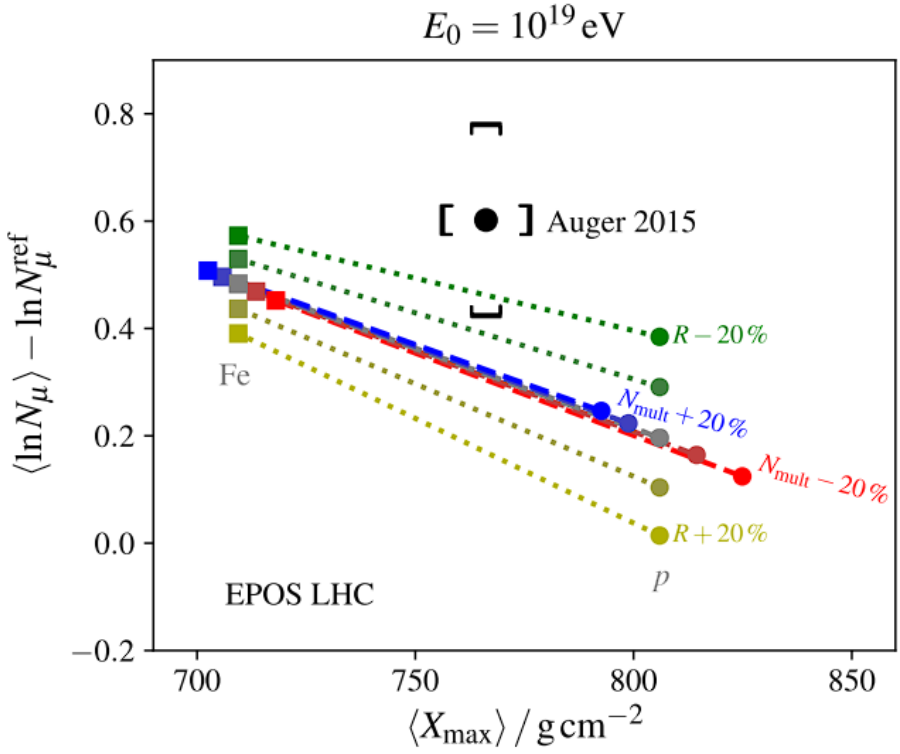

Figure 1.11: The effect of alterations to the hadron multiplicity and the energy ratio on N_μ and X_{max} for $10^{19.5}$ eV EASs are considered. The lines represent prediction from air shower simulations with EPOS-LHC. The grey line symbolizes the baseline model, while the coloured lines represent predictions from a modified model, with changes in increments of 10%. The data point is derived from the Pierre Auger Observatory [62].

Chapter 2

The LHCf experiment at LHC

2.1 The Large Hadron Collider

The LHC is currently the most powerful particle accelerator in the world. It is strategically located at CERN (European Organization for Nuclear Research), straddling the border between France and Switzerland. This machine, forming a 27-kilometer ring that hosts a hadron collider composed of two parallel circular beam pipes, boasts an expected peak collision energy of 14 TeV in the center-of-mass, setting the frontier for high-energy physics experiments. This energy is achieved thanks to a sophisticated system involving both powerful magnetic and electric fields. The magnet system, consisting of 1232 superconductive dipoles, guides each beam along the ring, while 392 quadrupoles focus them. Meanwhile, high-frequency electric fields in the radio-frequency (RF) cavities accelerate the protons to nearly the speed of light as they circulate within the collider. The LHC is the last of a series of accelerators, each designed to progressively increase the energy of particle beams before they are injected into the collider. This acceleration chain is crucial for achieving the high energies necessary for cutting-edge experiments. Protons, the primary particles used in the collider, begin their journey as ionized hydrogen. They are first accelerated to 50 MeV by LINAC 2, then boosted to 1.4 GeV in the Proton Synchrotron Booster (PSB). Subsequent stages involve the Proton Synchrotron (PS), which elevates the beam to 26 GeV,

and the Super Proton Synchrotron (SPS) which ramps it up to 450 GeV before finally reaching the LHC. For heavy ions, which currently include Lead and Xenon, with ongoing studies to incorporate additional types, the acceleration process differs significantly from that of protons. These ions are produced through the thermal ionization of isotopically pure sources and are initially accelerated to 4.5 MeV per nucleon in the LINAC 3. Subsequently, they are further accelerated in the Low Energy Ion Ring (LEIR) to approximately 72 MeV per nucleon. Following this, they traverse the same acceleration path as protons, passing through the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). The LHC is set to expand its ion capabilities with the inclusion of Oxygen ions, a significant addition scheduled for acceleration during the LHC Run III in 2025. This collision type will provide new opportunities to explore nuclear matter under different conditions, also allowing carrying out studies on HIMs under conditions similar to the Earth atmosphere. A schematic view of CERN accelerator system is shown in Figure 2.1. In addition to the energy of acceleration, a collider like the LHC is also characterized by its luminosity (L). Luminosity is defined as the ratio of the number of particle collision events (N) per unit time (t) to the interaction cross section (σ), expressed as:

$$L = \frac{1}{\sigma}\frac{dN}{dT}, \tag{2.1}$$

this parameter is directly proportional to the event rate (R), calculated by

$$R = L\sigma. \tag{2.2}$$

Luminosity itself can be derived from several machine features, including the number of colliding bunches n, particles per bunch in each beam, beam revolution frequency f, the Lorentz factor γ, beam emittance ϵ, the betatron function at the interaction point β^*, and a geometric luminosity reduction factor at the interaction point F:

$$L = \frac{nN_1N_2f\gamma}{4\pi\epsilon\beta^*}F. \tag{2.3}$$

The LHC is not just a singular experimental setup but a complex of experiments each located at different interaction points (IP) along the collider's ring. Major experiments include:

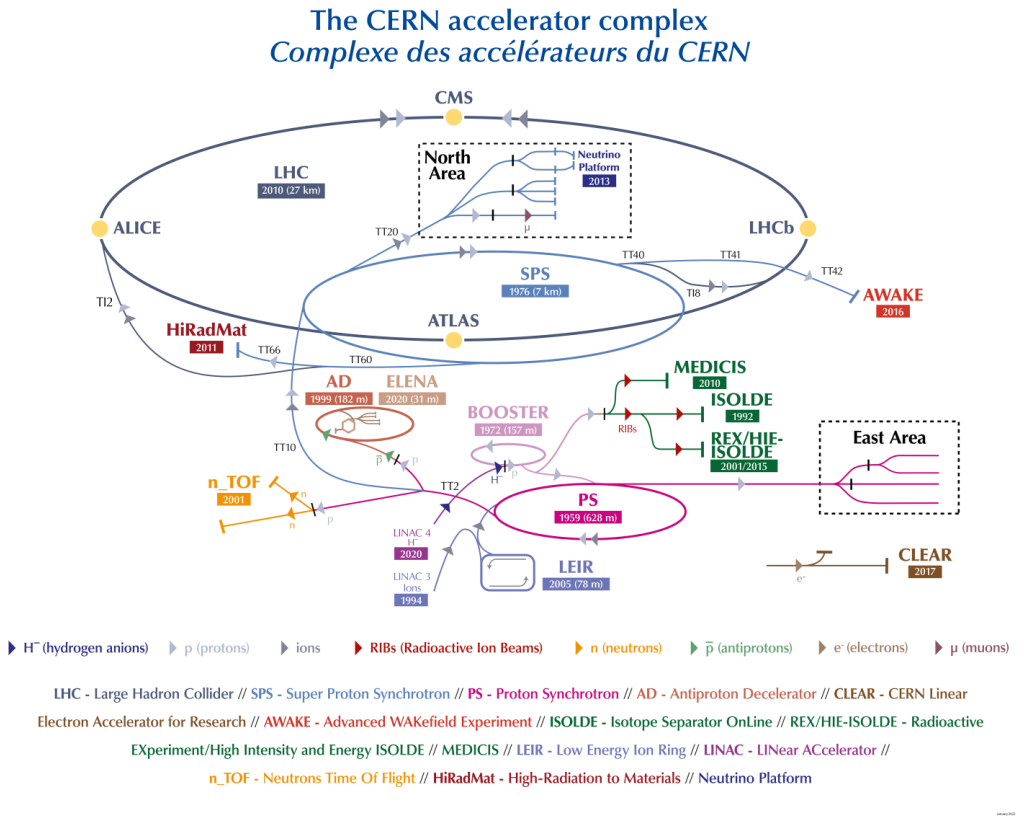

Figure 2.1: Schematic view of the CERN accelerators system. Only the four big LHC experiments are shown here [61].

- **ATLAS** (A Toroidal LHC ApparatuS [76]) and **CMS** (Compact Muon Solenoid [77]): They are installed at IP1 and IP5, respectively. Both are general-purpose detectors aimed at probing critical questions in Standard Model physics, such as electroweak interactions and the search for new physics phenomena.

- **ALICE** (A Large Ion Collider Experiment [78]): Located at the IP2, it is dedicated to the study of heavy ion physics, particularly the physics of quark-gluon plasma and hadronization processes.

- **LHCb** (Large Hadron Collider beauty [79]): Positioned at the IP8, it is designed specifically to study particles containing charm (c)

or bottom (b) quarks. The experiment focuses extensively on CP (Charge Parity) violation and hadronic spectroscopy.

Complementing these are smaller experiments such as:

- **LHCf** (Large Hadron Collider forward [80]): Positioned at IP1, near the ATLAS experiment, LHCf investigates particle production spectra in the very forward region to calibrate HIMs.
- **TOTEM** (Total Elastic and diffractive cross section Measurement [81]): Located at IP5, alongside CMS, TOTEM is mainly dedicated to measure the elastic and inelastic cross sections in high-energy proton-proton collisions.
- **MoEDAL** (Monopole and Exotics Detector at the LHC [82]): Installed at IP8 with LHCb, MoEDAL searches for magnetic monopoles and other highly ionizing massive particles.
- **FASER** (ForwArd Search ExpeRiment [83]): Situated at IP1, FASER seeks light, weakly interacting particles that could be potential candidates for dark matter. Additionally, FASER includes a subdetector, FASERν (FASER Neutrino Experiment), which is dedicated to studying high-energy neutrinos produced in LHC collisions, enhancing our understanding of neutrino physics.
- **SND@LHC** (Scattering and Neutrino Detector at the LHC [84]): Also located at IP1, SND@LHC aims to detect neutrinos and study neutrino interactions at TeV energies. This detector adds a unique perspective to the search for physics beyond the Standard Model by focusing on neutrino properties and interactions at high energies.

Each of these experiments is strategically positioned at specific interaction points around the LHC ring to optimize data collection and enhance the effectiveness of the research. This configuration allows the LHC to provide a thorough investigation of the fundamental aspects of particle physics. Since the beginning of operations, the LHC has completed several experimental runs, each of which was characterised by significant results. The RUN I began in 2010, reaching 3.5 TeV per beam, later increasing to 6.5 TeV per beam in the RUN II that began in 2015. These efforts led

to groundbreaking discoveries, including the Higgs boson, confirming the last unverified area of the Standard Model particle predictions. The RUN III of the LHC, started in 2022 and is still ongoing, aims to push the limits even further by increasing both collision energy and luminosity. This run will prepare the ground for the High-Luminosity LHC (HL-LHC), an upgrade planned for 2029, which promises to increase the luminosity by a factor of ten compared to the original LHC design. This upgrade will significantly increase the potential for physics discovery. The HL-LHC will allow to observe rare processes that occur beyond the current sensitivity of the LHC, offering insights into the mysteries of dark matter, supersymmetry and other theories beyond the Standard Model.

2.2 Forward physics at LHC

Experimental measurements at a hadron collider can be expressed using different kinematic quantities that describe the properties of particles produced in collisions. Typically, the z-axis is defined as parallel to the beam direction, with the x-y plane perpendicular to this axis. The momentum of a particle can be split into three components: p_x, p_y, and p_z. However, it is often more convenient to express these components in terms of transverse momentum p_T, azimuthal angle ϕ, and scattering angle θ. Transverse momentum, defined as $p_T = \sqrt{p_x^2 + p_y^2}$, is particularly important because it remains invariant under Lorentz transformations along the z-axis. The azimuthal angle ϕ is given by $\phi = \arctan\left(\frac{p_y}{p_x}\right)$, which describes the angle around the beam axis. The scattering angle θ, which describes the angle relative to the beam direction, is typically replaced by pseudorapidity η in high-energy physics. Pseudorapidity is defined as $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$, and it becomes a useful generalization of rapidity y when the particle mass is negligible compared to its momentum. Rapidity itself is defined as $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$, where E is the particle's energy and p_z is the longitudinal momentum. Pseudorapidity η converges to rapidity y in the high-energy limit where the particle mass is much smaller than its momentum. The preference for using p_T, y (or η and ϕ) over the direct momentum components stems from their simpler trans-

formation properties under Lorentz boosts, particularly along the beam axis. In the context of collider experiments, the central region (or mid-rapidity region) refers to the area perpendicular to the beam direction, where particles are produced at large scattering angles relative to the beam axis. This region corresponds to small values of pseudorapidity, typically around $|\eta| \approx 0$. It is well covered by most detectors and is crucial for studying a wide range of phenomena, including those related to the Standard Model and beyond. Moving away from the central region, the forward region refers to the area around the beam direction where particles are produced at smaller scattering angles relative to the beam axis, corresponding to large values of pseudorapidity $|\eta| > 2$. The very forward region extends even closer to the beam axis, typically associated with pseudorapidities $|\eta| > 6$. The mid-rapidity region is where measurements at the LHC are most precise, but it has only indirect relevance to air showers. This is shown in the left panel of Figure 2.2, which shows that the majority of muons in an air shower are generated by long-lived hadrons that are emitted in the forward region, specifically at $\eta > 2$. Also, measurements in the forward and very forward regions are crucial for studying hadronic interactions in EAS, since it is where a significant amount of the collision energy is carried away by particles moving close to the beamline (dashed black lines in the left panel of Figure 2.2), but at the same time, they are particularly challenging due to the high energy flow and consequent extreme radiation damage to which the instruments are subjected, thus requiring highly specialized detectors and data taking strategies. The acceptances of the LHC experiments are depicted in the right panel of Figure 2.2.

2.3 Relevant measurements for the Muon Puzzle at LHC

The LHC is an ideal facility for acquiring crucial data related to key parameters in the development of Extensive Air Showers (EAS). Proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 14$ TeV correspond to an energy of approximately (10^{17} eV in the reference frame where one proton is at rest. As discussed in Section 1.3.2, the discrepancy in the pre-

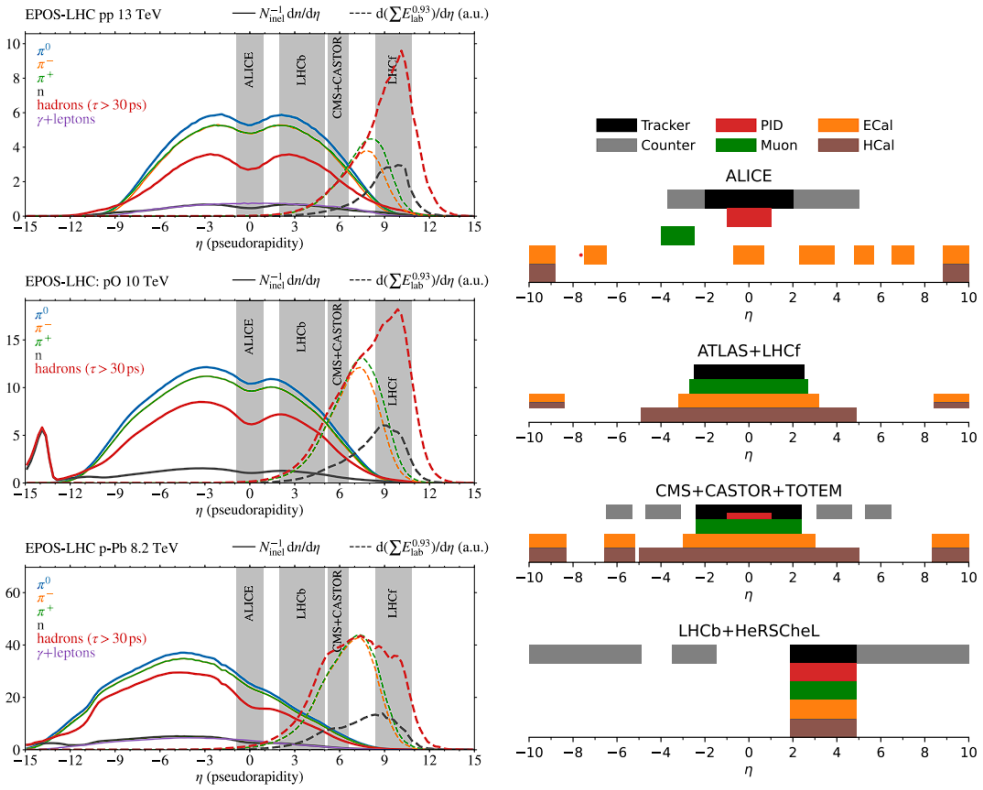

Figure 2.2: Left: Schematic view of the pseudorapidity distributions for various particles in proton-proton collisions at 13 TeV, proton-oxygen collisions at 10 TeV, and proton-lead collisions at 8.2 TeV, as modelled by EPOS-LHC. The distributions are shown for π^0, π^+, neutrons, hadrons with with life time $\tau > 30$ ps, and γ-leptons. The solid black lines represent the inclusive charged particle multiplicity $N_{\text{incl}}dn/d\eta$, while the dashed black lines represent the derivative of the sum of the transverse energy with respect to η. The grey-filled regions correspond to the pseudorapidity coverage of the ALICE, LHCb, CMS + CASTOR, and LHCf experiments. Right: The pseudorapidity coverages of the experiments in the four interaction points of LHC. Figure from [42].

dicted number of muon by HIM has an early onset, at about $4 \cdot 10^{16}$ eV, which corresponds to a center-of-mass energy $\sqrt{s} = 8$ TeV, so in principle

it is possible to observe its effect with the LHC experiments. While no single experiment at the LHC can comprehensively cover all parameters relevant to EAS physics, a combination of results from various collaborations allows for a robust understanding of air shower evolution. A comprehensive review on the measurement of experiments at the LHC for solving the muon puzzle can be found in [42] and related references. Below we provide a brief description of how individual experiments can contribute and some key results. ATLAS is a general-purpose detector with broad acceptance, featuring systems that cover increasing pseudorapidity ranges up to $|\eta| < 4.9$. ATLAS, thanks to the ALFA Roman Pot, has provided the most accurate measurements of the inelastic proton-proton cross-section and has played a key role in constraining hadron multiplicity through its measurements of charged particle spectra and forward energy flow. Moreover, it is capable of combining their measurements with the LHCf detectors to provide insights into diffractive dissociation and exotic forward particle production. CMS is another general-purpose detector, similar in design to ATLAS but with a slightly different emphasis. It features the CASTOR system, which has been pivotal in studying diffractive events, low-x parton physics, and forward energy flow. Despite the decommissioning of CASTOR for Run III, CMS continues to provide data for EAS physics through its broad acceptance and high-accuracy luminosity measurements. ALICE, primarily focused on heavy-ion physics, excels in high-resolution tracking and hadron particle identification at mid-rapidity. ALICE capabilities are crucial for tuning and validatingHIMs, particularly through its precise measurements of identified hadron spectra. The discovery of quark-gluon plasma (QGP) effects, such as strangeness enhancement, has significant implications for EAS physics. LHCb, designed for flavour physics, is particularly relevant for EAS physics due to its coverage of the forward region $1.9 < \eta < 4.9$. LHCb has been instrumental in measuring forward production cross-sections of D and B mesons, providing critical data for predicting the prompt atmospheric lepton flux, a key background for high-energy neutrino observatories. In addition to the big experiments, LHCf and TOTEM are specialized experiments that play a significant role in EAS physics. LHCf focuses on measuring the production cross-sections of photons, π^0 and η in the very forward region $|\eta| > 8.4$, crucial for understand-

ing forward particle production. Moreover, the very forward neutron spectra permit the infer the elasticity, a key parameter for the development of EAS. More details on LHCf will be provided in the next sections. TOTEM, which surrounds the CMS detector, is designed to measure the total proton-proton cross-section and study elastic scattering and diffractive dissociation. TOTEM's precise measurements of the inelastic cross-section and elastic scattering provide critical insights into hadronic interactions at high energies. Measurements at the LHC that are fundamental for improving EAS simulations primarily involve the average properties of light-flavor hadron production at low momentum transfer within the context of semi-hard and soft QCD processes. Events involving heavy or high p_T particles generally have small impact on air shower development, except forward heavy flavour production, which contributes to the prompt atmospheric neutrino flux. As discussed in Section 1.3.2 The most important measurements for EAS physics include the inelastic cross-section (σ_{inel}), hadron multiplicity (N_{multi}) across a wide rapidity range, the elasticity and the π^0 fraction. Other important parameters are the diffractive cross-sections, and the composition and spectra of light hadrons such as pions, kaons, and protons. Forward production cross-sections of specific particles, such as π^0, η, ρ^0, the lightest D and B mesons, as well as forward energy flow $(\mathrm{dE}/\mathrm{d}\eta)$, are also very important. The total energy flow, particularly the ratio of hadronic to electromagnetic energy flow, is a direct measure of the energy fraction (α) remaining in the hadronic cascade. This ratio is highly sensitive to the number of low-energy muons produced in EASs. Forward production of ρ^0 mesons, which increases α, cannot be directly measured with LHC beams, but the ratio of ρ^0 to π^0 production in proton-proton and proton-ion collisions provides an important benchmark for HIMs. While D and B meson production is not directly related to the Muon Puzzle, these measurements are essential for accurately predicting the inclusive atmospheric lepton fluxes above 1 PeV, which are crucial for high-energy neutrino observatories. The prompt component of the neutrino flux, dominant at these energies, originates from the production and decay of charm and beauty quarks in cosmic ray interactions with air. Production cross-sections for D and B mesons in proton-proton, proton-lead, and proton-oxygen collisions are key for constraining nuclear parton density functions, which

are necessary for flux calculations. Existing data from the LHC have already provided significant constraints on inelastic cross-sections, hadron multiplicity, diffractive cross-sections, and the production cross-sections for D and B mesons, including in the forward region. However, there is a notable lack of data on identified hadron spectra and strangeness production in the forward region, particularly in proton-ion collisions, which are expected to have a high impact on the Muon Puzzle, togheter with measurements of elasticity e and π^0 fraction. The inelastic cross-section is measured using two complementary techniques. One method involves counting empty events, which requires wide acceptance and precise beam luminosity measurements, relying on theoretical extrapolation to obtain the full inelastic cross-section. The other method, employed by TOTEM and ATLAS/ALFA, involves observing elastic scattering at very low momentum transfer, allowing the total cross-section to be calculated from the forward amplitude of elastic scattering using the optical theorem. Diffractive cross-sections and rapidity gaps are crucial for understanding air showers, as diffractive collisions produce large rapidity gaps devoid of particles. These gaps are significant in single and double diffractive events in air showers. Measurements of diffractive cross-sections have been conducted in proton-proton collisions by ALICE, CMS, and TOTEM, with ongoing studies in proton-lead collisions. Also, diffractive measurements by the joint analyses of LHCf and ATLAS are ongoing. Charged particle spectra are measured as a function of pseudorapidity and transverse momentum. These measurements are particularly valuable for EAS physics, especially forward measurements covering the relevant pseudorapidity regions. Forward energy flow, captured by calorimeters, measures energy deposits from both charged and neutral particles, with CMS-CASTOR calorimeter playing a key role in these measurements, especially in studying the electromagnetic to hadronic energy flow ratio, crucial for understanding EAS dynamics. Identified hadron spectra, measured by experiments with particle identification systems like ALICE and LHCb, provide critical data, especially in forward measurements. However, there is a gap in forward data for proton-lead collisions, important for understanding hadron composition in air showers. Inclusive photon, neutral pion, and η spectra are essential for understanding the electromagnetic component of air showers. LHCf has contributed significantly by mea-

suring these particles in the very forward region in proton-proton and proton-lead collisions, crucial for investigating the Muon Puzzle as they constrain the hadronic to electromagnetic energy flow ratio. Also, LHCf has measured very forward neutron spectra, providing insights into the energy distribution of forward-produced particles and the elasticity. The ratio of neutral pion to neutron spectra and η to neutral pions are particularly relevant for muon production in EASs. The experimental apparatus and results obtained by LHCf will be discussed in more detail in the following Section 2.4.

2.4 The LHCf experiment

The Large Hadron Collider forward (LHCf) is a unique experiment mainly dedicated to measuring the neutral particle production cross-section in the forward pseudorapidity region. This section will describe the experimental setup, the data acquisition system and the experimental published results obtained in the past works.

2.4.1 Experimental apparatus

The LHCf (Large Hadron Collider forward) experimental apparatus comprises two distinct detectors, called Arm1 and Arm2. Each detector is composed of two imaging shower calorimeters, named Large Tower (LT) and Small Tower (ST). The sizes of the calorimeters transverse to the beam direction are 20 mm x 20 mm and 40 mm x 40 mm in Arm1, 25 mm x 25 mm and 32 mm x 32 mm in Arm2, as shown in Figure 2.3. The different geometry between the Arms facilitates independent measurements across overlapping pseudorapidity regions while maximizing the span of geometrical coverage given their differing acceptances. Each Arm comprises a sequence of 16 scintillator layers, each 1 mm thick, alternating with 17 tungsten layers, culminating in a total length of about 210 mm, equivalent to 44 radiation lengths X_0 and 1.6 interaction lengths λ_I. EJ-260 plastic scintillators were used before the RUN II operations at the LHC as active materials. The initial 11 layers and the final one are 7 mm thick, with layers 12 through 16 being 14 mm thick, tailored for hadronic shower mea-

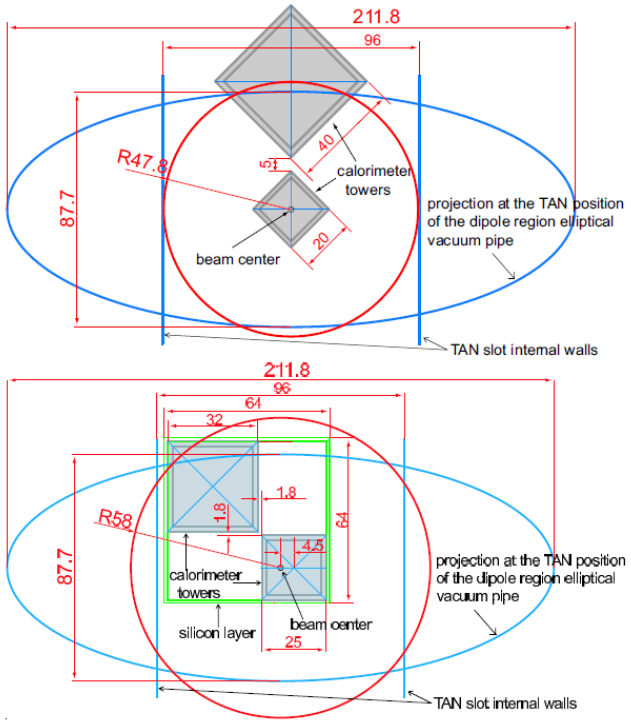

Figure 2.3: Cross section of Arm1 (top figure) and Arm2 (bottom figure) detectors inside the TAN region.

surements. Additionally, each calorimeter integrates 4 pairs of X-Y imaging layers at varied depths, constructed from plastic scintillating fibres (SciFi) with a 1 mm pitch in Arm1 (before LHC RUN II operations) and 0.16 mm pitch silicon microstrip detectors in Arm2. The optical signals from the scintillators in Arm1 are conveyed through quartz light guides and detected using HAMAMATSU R7400 photomultiplier tubes (PMTs), which feature a 64-anode section (MAPMT). The longitudinal structure of the two detectors before LHC RUN II operations is depicted in Figure 2.4.
The two detectors are positioned at a zero-degree collision angle approximately 141.5 meters away from Interaction Point 1 (IP1, see Figure 2.5), nestled within the Target Neutral Absorbers (TANs). The TANs serve the critical function of shielding the outer superconducting beam separation

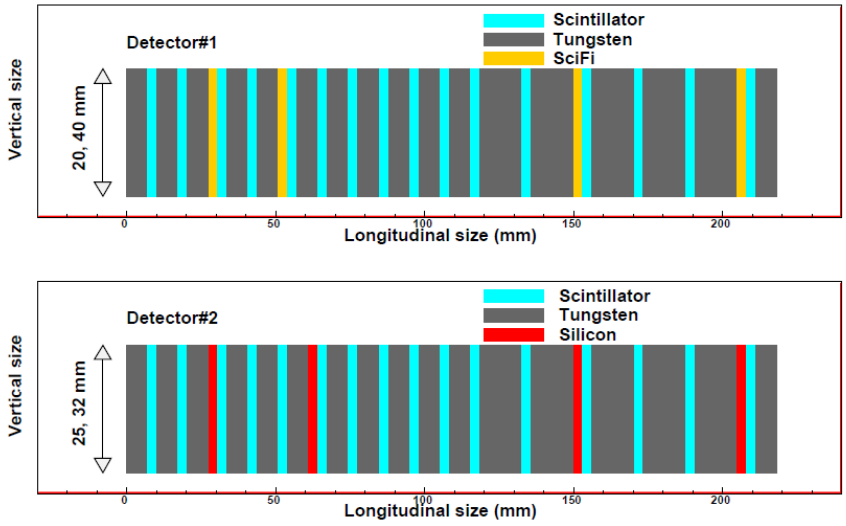

Figure 2.4: Longitudinal layouts of Arm1 (top figure) and Arm2 (bottom figure) used for the LHC-RUN I operations. The color coding for the layers in the images is as follows: grey for tungsten, cyan for scintillator, yellow for SciFi, and red for silicon layers. The upstream side of each detector is oriented to the left in the images.

dipoles (D2) from neutral particle debris emanating from the IP. The detectors are located during operations in a specific area of the TANs where the beam vacuum chamber transitions into a Y shape, diverging from a single beam tube near the IP to two separate tubes that link to the LHC arcs, as displayed in Figure 2.6. This optimal placement allows the detectors to cover an extensive pseudorapidity range from 8.4 to infinity. Positioned 55 meters beyond the D1 dipole magnets, which segregate the proton beams, the detectors exclusively encounter neutral particles. The detectors underwent significant upgrades in preparation for the LHC RUN II operations, to accommodate the elevated radiation levels encountered at higher energies of proton-proton and proton-lead collisions. Previous plastic scintillators were replaced with Gd_2Si0_5 (GSO, Hitachi Chemical) scintillators, capable of maintaining their light output up to 10^6 Gy [85], in contrast to the earlier plastic scintillators that degraded around a level

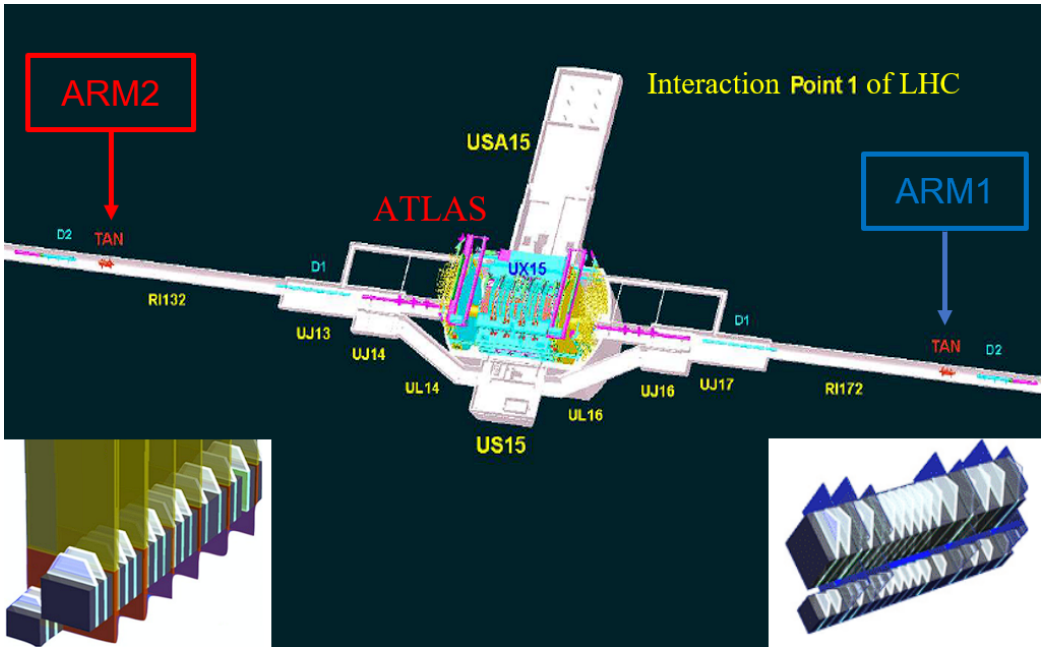

Figure 2.5: Geometry of the IP1 area at the LHC. The structure at the center highlights the ATLAS detector, which surrounds the interaction point. A line extending from the top-left to the bottom-right represents the LHC tunnel long straight section, and the "TAN" labelled areas on both side of IP1 indicate the location of the LHCf detectors. A schematic representation of the Arm1 (on the right) and Arm2 (on the left) detectors is also shown

.

of 10^2 Gy. Additionally, the silicon microstrip trackers in Arm2 were entirely replaced, incorporating a new bonding scheme that reduces the signal charge by approximately 50% compared to the prior setup. This alteration allows the readout electronics to handle higher energy saturation levels. These new silicon detectors were also repositioned longitudinally to function effectively as energy detectors [86], by detaching at different depths the last two X-Y pairs of silicon planes (see Figure 2.7). Moreover, the large tower in Arm2 was relocated to the opposite side to minimize background interference from high-energy protons in diffractive interactions, which the D1 dipole magnet does not adequately deflect. Each Arm

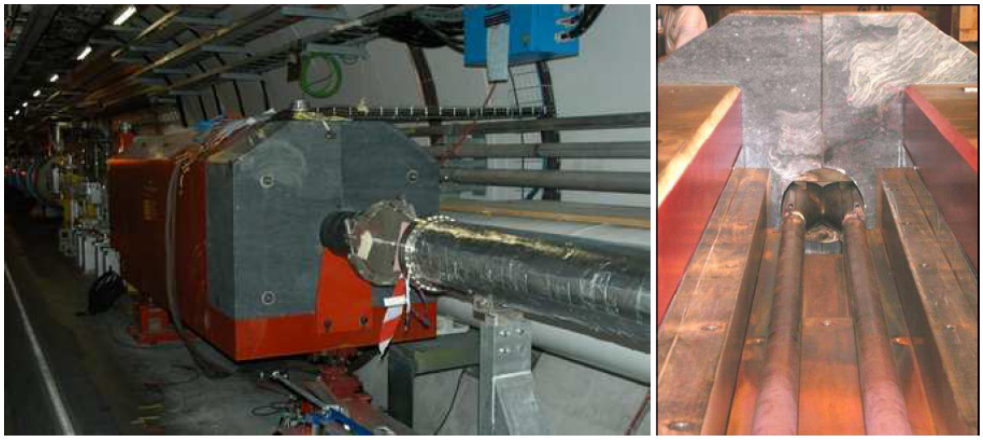

Figure 2.6: Pictures of the TAN absorber, positioned at about 141.5 meters from the interaction point IP1. On the left, it shows the TAN fully assembled in the LHC tunnel as viewed from the direction of the IP1. On the right, there is a depiction of the TAN during its assembly at CERN, viewed from above facing the IP. A 96mm gap between the two beam pipes is included to accommodate the installation of the LHCf detectors.

also features a sub-detector known as the Front Counter, consisting of two 2 mm plastic scintillators separated by a 0.5 mm copper plate, providing a total depth of 0.06 X_0 and a sensitive area of 8 cm x 8 cm. These Front Counters are crucial for determining the event rate and, with precise calibration using the Van Der Meer scan [87], can measure the instantaneous luminosity. Upgraded configurations of Arm1 and Arm2, depicted in Figure 2.8, achieve an energy resolution of 5% for photons and 40% for neutrons, and a position resolution for electromagnetic showers of 200 μm in Arm1 and 40 μm in Arm2 [88].

2.4.2 Data acquisition system and trigger logic

Overview of the DAQ system

Data acquisition for the LHCf experiment is performed in the ATLAS counting room (USA15) and is optimized for operation with 43 cross-

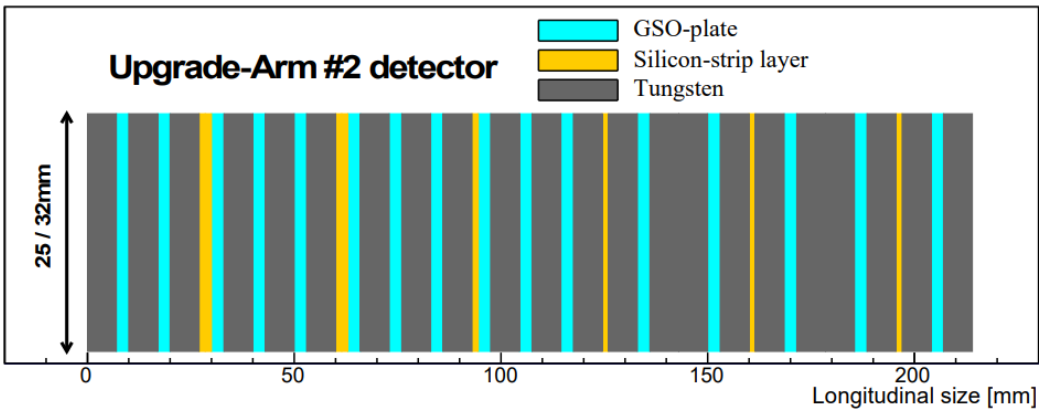

Figure 2.7: Longitudinal structure of upgraded Arm2 used from the LHC-RUN II operations. The color coding for the layers in the images is as follows: grey for tungsten, cyan for scintillator, yellow for silicon layers. The upstream side of the detector is oriented to the left in the image.

ing bunches at a luminosity of approximately 0.1 $\mu b^{-}1 s^{-}1$. The data acquisition (DAQ) flow adopted for operations during the LHC RUN II is depicted in Figure 2.9 and consists of two primary components: the PMT chain and the silicon microstrip chain. The PMT chain begins at the TAN site, where PMT signals are initially amplified by Technoland N-SE810 pre-amplifiers. These signals are then transmitted over 200 meters of coaxial cable to the USA15 counting room. There, the signals enter Technoland N-SE810 fan-out modules, which split each signal. The first output from each module is directed to a 12-bit ADC (CAEN V965), and the second output passes through a low pass filter (LHCF-LPF) and a discriminator (V814B), which together are responsible for generating the experimental trigger. The silicon microstrip chain involves signals that are amplified and digitized by dedicated electronics at the TAN site. Upon generation of a trigger by the PMT chain, this trigger is managed by the Trigger Sequencer Card (TSC [89]) and forwarded to the Front End Control Unit (FEC [90]). This unit then signals the control ring, which interfaces with four motherboards (MDAQ). These motherboards capture analog signals from the PACE3 chips [91] connected directly to the silicon microstrip layers within the detector. They then convert these analog

(a) Arm1 (b) Arm2

Figure 2.8: Photos of the upgraded Arm1 (left side) and Arm2 (right side). The two detectors are positioned at the bottom of the pictures.

signals into digital format and relay them to the VME Receiver. Additionally, the electronic chain includes a Power Supply SY257 to provide the necessary power and a Laser System for calibrating the PMTs. The entire data acquisition process utilizes the MIDAS package (Maximum Integration Data Acquisition System, which handles the integration and processing of the collected data.

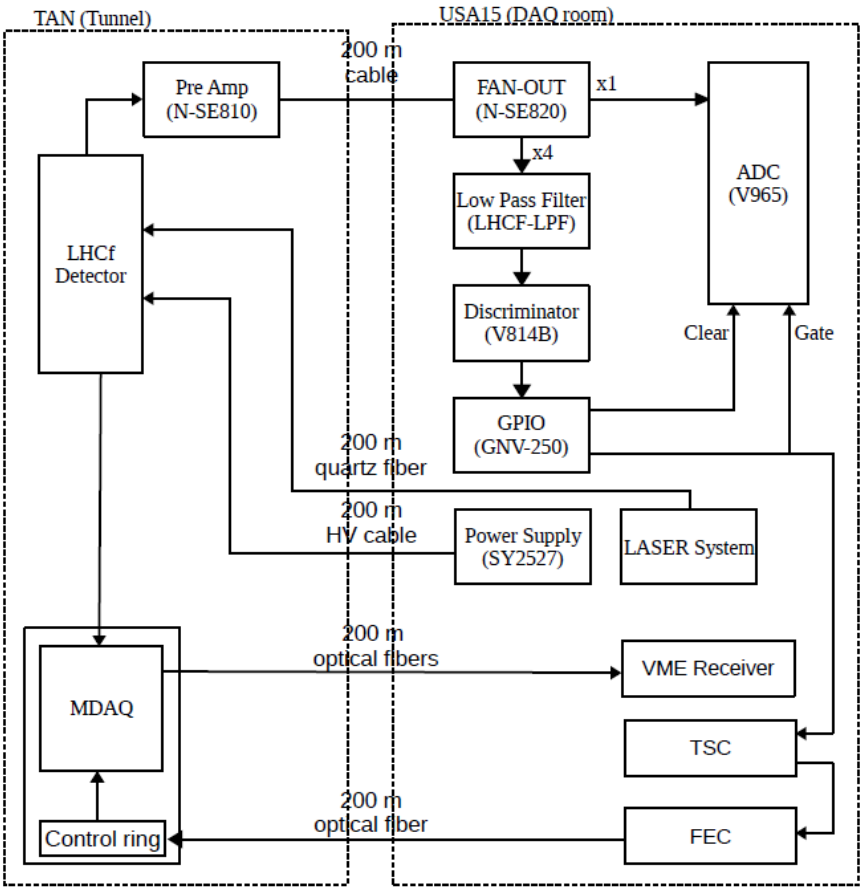

Figure 2.9: Diagrammatic representation of the LHCf Data Acquisition System: the upper and lower sections correspond to the scintillator and silicon components, respectively.

Overview of the trigger logic

The LHCf trigger system utilizes the 40 MHz LHC clock as its primary timing source. This system is engineered using FPGA technology, with the logic loaded onto an FPGA chip housed on a VME board (GN0324), which interfaces with the VME bus. The triggering mechanism comprises three distinct levels. The initial level, known as the first level trigger (L1T), is activated by a bunch crossing (BX) identified via two Beam Pick-

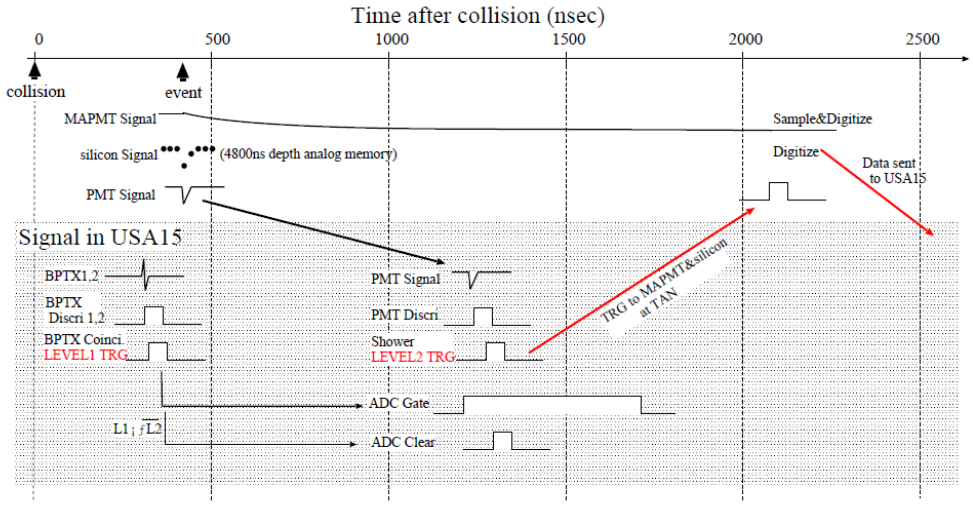

Figure 2.10: The timing diagram for the LHCf trigger and data acquisition system.

Up Timing for Experiments (BPTX) signals, which are produced by two Beam Position Monitors (BPMs) positioned 175 meters from the interaction point. Upon activation of the L1T, a gate signal of 500 nanoseconds (L1T-ENABLE) is dispatched to the PMT ADC. Subsequently, a second level trigger (L2T) is generated if more than three successive layers detect over 300 particles each, surpassing an energy threshold of approximately 200 GeV, thereby ensuring a 99% efficiency for photon detection. Should the L2T fail to activate within 1 microsecond following the L1T, the ADC is reset. The final stage, the third level trigger (L3T), synthesizes various types of triggers, including pedestal and LASER calibrations, among others. The assertion of the L3T results in the recording of the event. A timing diagram illustrating the trigger and data acquisition process is featured in Figure 2.10. The interval between two successive bunch crossings is about 2.5 microseconds, with an anticipated trigger rate of roughly 1 kHz. Under these operational conditions, the system live time is estimated to be about 50%.

Year	Energy (TeV)	Collisions	Detector	Papers			
				γ	n	π^0	η
2009	0.9	p-p	Arm1, Arm2	[92]			
2010	7	p-p	Arm1, Arm2	[93]	[94]	[95], [96]	
2013	2.76	p-p	Arm1, Arm2			[95], [97]	
2013	5.02	p-Pb	Arm2			[95], [97]	
2015	13	p-p	Arm1, Arm2	[98], [99]	[100]		[101]
2016	5.02	p-Pb	Arm2				
2016	8.16	p-Pb	Arm2				

Table 2.1: Summary of the LHCf runs at LHC and analysis matrix.

2.4.3 LHCf acquired data and published results

Since 2009 the LHCf experiment has been collecting data across various collision energies and using different types of colliding particles. Most operational phases have been conducted under specialized conditions to maintain a luminosity below 0.1 $\mu b^{-}1s^{-}1$, ensuring that pile-up remains under 1%. Additionally, a β^* value of around 20 m was maintained to ensure that the colliding protons remain mostly parallel, a crucial factor for accurately reconstructing the scattering angle. A summary of the LHCf operations and publications is presented in Table 2.1. The primary objective of the LHCf data analysis is to measure the energy and momentum spectra of neutral particles produced in the forward region. These measurements are essential for calibrating and validating the predictions of HIMs. They also provide insights into parameters like elasticity from neutron energy spectra and lateral development distributions from p_T spectra. Given that EASs primarily originate from proton-ion collisions, it is vital to study the variations in yields due to nuclear effects. Currently, lead is the only ion that can be produced and accelerated at the LHC. However, the potential use of the most abundant atmospheric light ion beams, such as oxygen, is being explored to allow for a more realistic examination of EAS development. To bridge the knowledge from the energy range of 10^{17} eV to 10^{20} eV, the application of a scaling law is necessary. Thus, another significant analytical goal involves testing key scaling laws, including Feynman scaling, p_T scaling, and the hypothesis of limiting fragmentation. To date, analyses have primarily focused on

the most abundantly produced particles in collisions: photons, neutrons, neutral pions and eta mesons. The properties of pions and etas are analyzed through the decay into two photons.

$$\pi^0/\eta \rightarrow \gamma\gamma. \tag{2.4}$$

Given the energy resolution for electromagnetic showers exceeds 2% above 100 GeV, neutral pions are excellent candidates for precise measurements of nuclear effects and for testing scaling laws. Moreover, the measurement of production rates is crucial, as neutral pions significantly influence EAS development by transferring energy from the hadronic to the electromagnetic channel. Neutrons also play an integral role in studying the muon component of the EAS, and are important for calibrating HIMs. The recent measurement of η mesons by LHCf will be discussed in detail in the next Chapter 3, providing further insights into the particle dynamics within the very forward region.

Chapter 3

Measurement of the forward η meson production rate

This chapter details the first measurement of forward η meson production in proton-proton collisions at $\sqrt{s} = 13$ TeV by the LHCf experiment. The production of η mesons in the forward region is particularly relevant for understanding the electromagnetic component in EASs, probing strangeness production in hadronization processes, and contributing to the study of muon production in cosmic rays, all of which are discussed in detail in the motivation section (Section 3.1). The analysis is based on data collected exclusively by the LHCf-Arm2 detector during a special low-luminosity run at $\sqrt{s} = 13$ TeV. The datasets used in this analysis are described in Section 3.2, which includes both the experimental data and the Monte Carlo simulations. Section 3.3 outlines the analysis framework, including the methods used for η meson reconstruction from photon pairs, the criteria for selecting candidate events, and the strategies employed for background subtraction. Given the challenges associated with measuring particles in the forward region, several corrections are applied to account for the experimental limitations, such as acceptance and selection inefficiencies. These corrections are discussed in Section 3.4. Additionally, the chapter examines the systematic uncertainties associated with these corrections and other aspects of the analysis in Section 3.5. The results of the η meson production rate measurement are presented in Section 3.6. These results are compared with predictions from

various HIMs, providing insight into the accuracy of these models in the forward region. To ensure the reliability of the analysis, validation tests were conducted, as discussed in Section 3.7. These tests confirm the effectiveness of the applied corrections and their alignment with the true particle-level distributions. Finally, Section 3.8 presents the conclusions of this analysis and discusses future prospects. This section reflects on the significance of the results, the comparison with theoretical models, and the expected improvements in statistical precision with future data, particularly with the new LHCf data from LHC RUN III.

3.1 Motivation of the measurement

The measurement of forward η meson production by LHCf-Arm2 in proton-proton collisions at $\sqrt{s} = 13$ TeV, presented in this Chapter, is important for several reasons:

- **Electromagnetic component contribution in EAS**: Photons produced in η mesons decay modes are the second most significant source of electromagnetic showers in EASs, following the ones produced in two-photon π^0 decays. The relative contribution of η mesons to electromagnetic showers with respect to π^0 one depends on the energy, the pseudorapidity and the specific HIM used for simulation. In the forward region, it is highly unconstrained due to the lack of data, inducing strong discrepancies between HIM predictions. For this reason, measuring the η/π^0 production ratio in the forward region is crucial to calibrate and test HIMs.

- **Strangeness Production**: The role of the strange (s) quark in hadronization processes can be explored by analyzing the production cross-sections of η mesons. The forward strange quark contribution is a key parameter in HIMs. Differences in this parameter result in large discrepancies in the predicted forward η production cross-sections across these models, with these variations being more significant than those observed for forward π^0 production cross-section, as shown in Figure 3.1, where the predicted differential cross-section

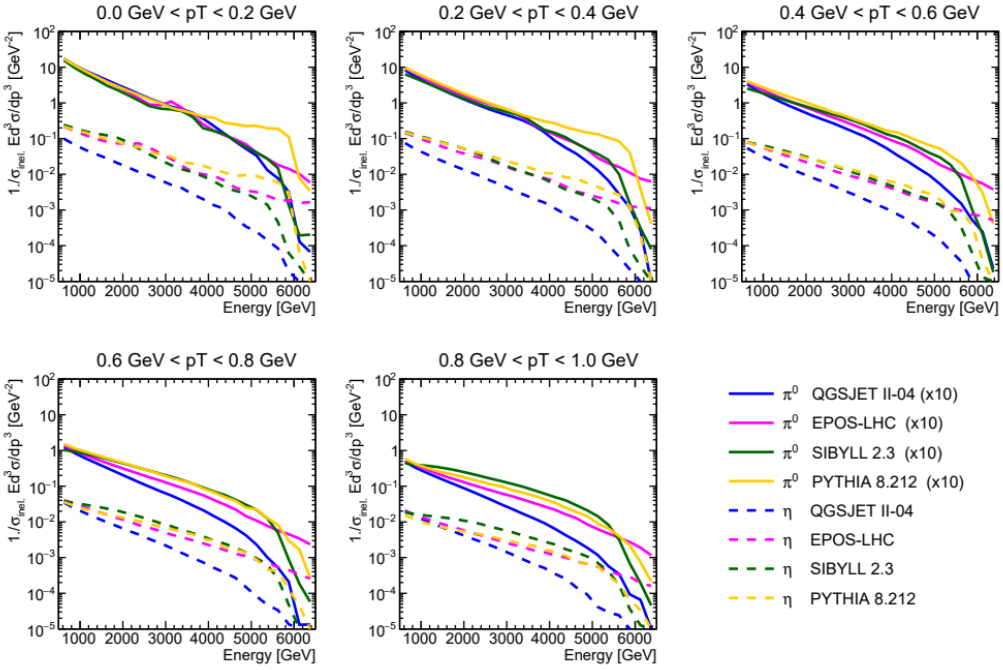

Figure 3.1: Differential cross-section of π^0 and η mesons predicted by several HIMs (QGSJET II-04, EPOS-LHC, SIBYLL 2.3, PYTHIA 8.212) at different transverse momentum ranges. Solid lines represent π^0 predictions, while dashed lines indicate η predictions. The π^0 predictions are scaled by a factor 10 [102].

of η and π^0 mesons at different transverse momentum ranges, according to various HIMs, is shown [102].

- **Muon Production in Extensive Air Showers**: η mesons are also an important source of muons in EAS, through charged decays, such as $\eta \rightarrow \mu^+\mu^-\gamma$ [103]. This is depicted in the left panel of Figure 3.2, where it is shown the contributions from the decays of various particles to the atmospheric $\mu^+ + \mu^-$ flux. One can note that the high energy prompt flux ($E_{\mu^+ + \mu^-} > 10^7 GeV$) is dominated by the unflavored component of the EAS (orange solid line), whose composition is given in Figure 3.3, where it is shown that η mesons are one of the main components (black solid line). The study of these

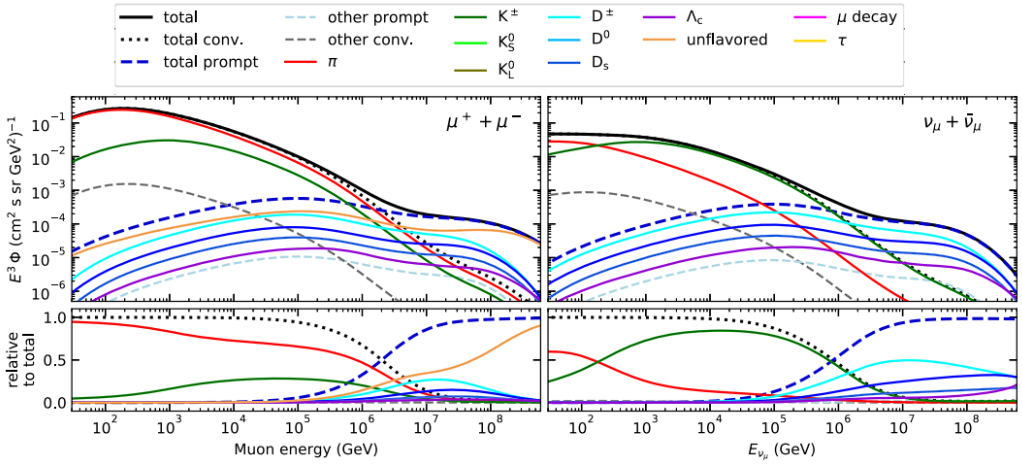

Figure 3.2: Left: Energy spectra of atmospheric $\mu^+ + \mu^-$ flux as a function of muon energy, showing contributions from the decays of various particles. Right: Contributions to the $\nu_\mu + \bar{\nu}_\mu$ flux. The solid lines represent the total fluxes, while the dashed lines indicate the contributions from conventional and prompt sources. Figure from [103].

particles, particularly in the forward region, is of great importance in the direct study of muon production in EAS, providing an important probe for testing and improving HIM predictions. This has a strong implication in solving the Muon Puzzle.

- **Absolute Energy Calibration of LHCf detectors**: By measuring both π^0 and η mesons, the LHCf experiment can calibrate the absolute energy scale of the detector using the invariant mass peaks of these mesons (depicted in Figure 3.4), as discussed in Section 3.5.1. The π^0 and η invariant mass peaks serve as known reference points, allowing for precise calibration of the detector energy measurements. Despite this motivation is not directly connected to the physics potential of this measurement, the absolute energy scale calibration is fundamental for ensuring the accuracy of the experimental results and for making reliable comparisons with theoretical predictions of HIMs.

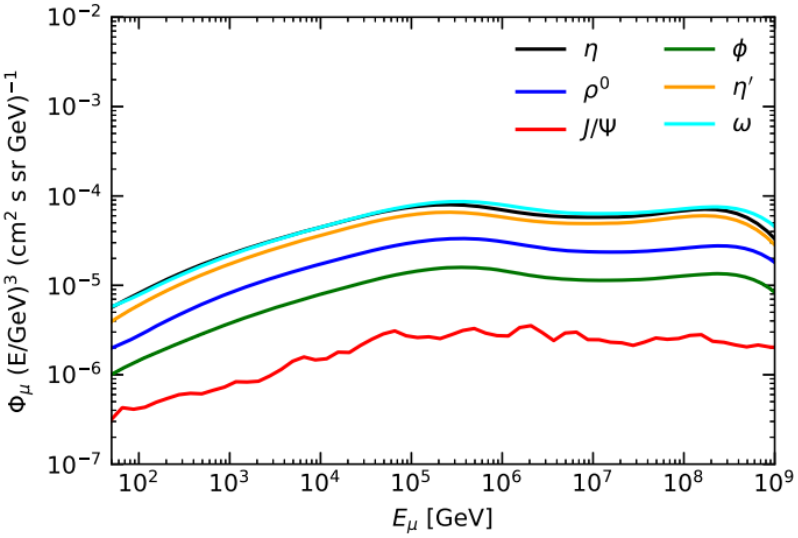

Figure 3.3: Contribution of various unflavored mesons to the high-energy prompt muon flux in EAS. The figure shows the differential muon flux as a function of energy, with different mesons contributing to the unflavored component, including η (black), ϕ (green), ρ^0 (blue), η' (orange), ω (cyan), and J/Ψ (red). Figure from [103].

3.2 Data analysis datasets

3.2.1 Experimental datasets

LHCf performed a data acquisition in proton-proton collisions at $\sqrt{s} = 13$ TeV that took place from June 10th to 13th, 2015. This special run, characterized by low luminosity to protect the experimental apparatus from radiation damage, involved 29 bunches colliding at IP1 with a half crossing angle of 145 μrad and a β^* of 19 meters. Additionally, 6 and 2 non-colliding bunches were circulating in the clockwise and counter-clockwise directions, respectively. These non-colliding bunches were instrumental in estimating the background noise from interactions between the particles produced and the residual gases within the beam pipe. Throughout the operation LHCf performed a shared data acquisition with the ATLAS

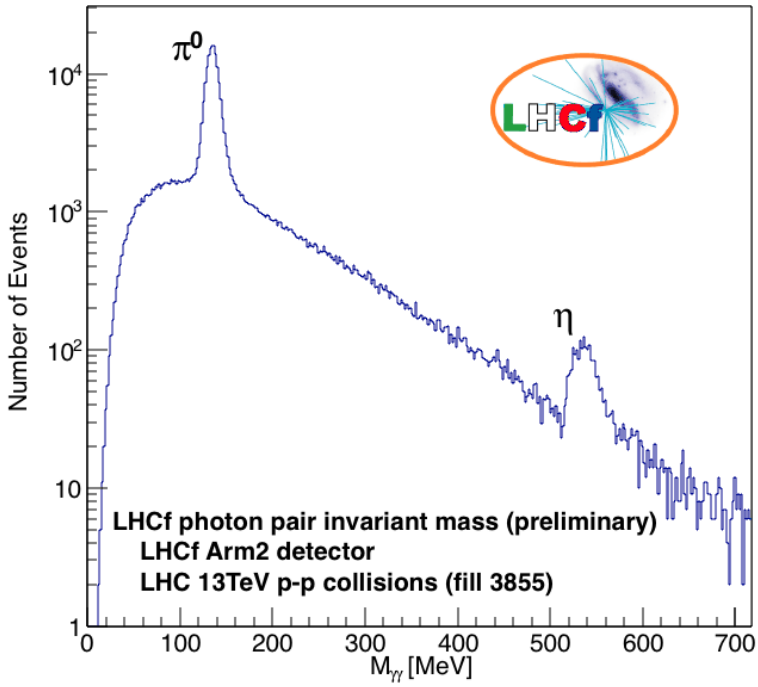

Figure 3.4: Invariant mass distribution of photon pairs measured by the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV (Fill 3855). Two distinct peaks corresponding to the π^0 and η mesons can be distinguished. These peaks are used to calibrate the absolute energy scale of the LHCf detectors, providing reference points for precise energy measurements.

experiment, exchanging the trigger in order to perform a combined data analysis. The integrated number of acquired triggers as a function of time for shower events and π^0 candidates is shown in Figure 3.5. During the 26 live-time hours of the acquisition, approximately $4 \cdot 10^7$ events were acquired. For the analysis, two datasets were utilized, corresponding to LHC Fill 3855 and LHCf Run 44299-45106. This Fill was chosen because it was the longest of the acquisition campaign (about 14 hours) and the one where the luminosity remained most stable. The two acquired datasets differ in the value of the average number of collisions per bunch cross-

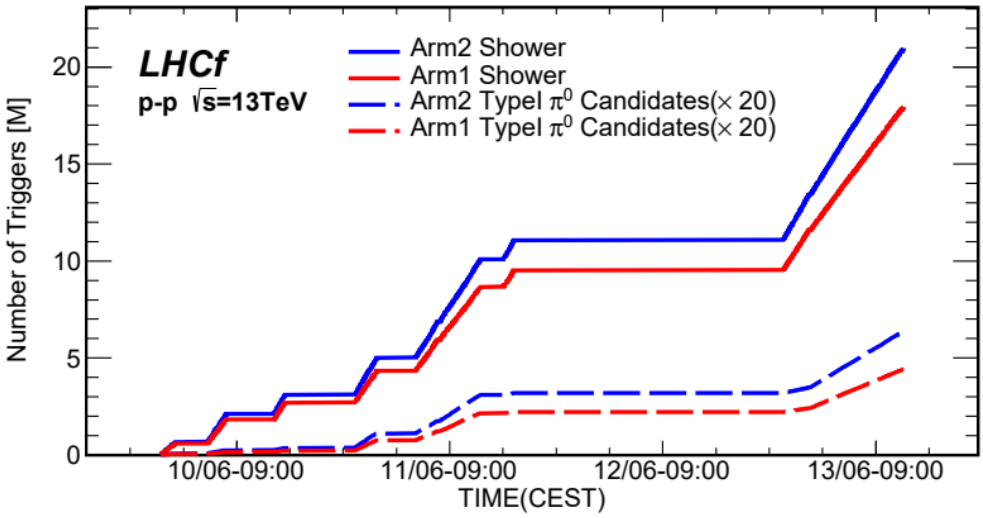

Figure 3.5: Number of integrated triggers recorded by the LHCf experiment during proton-proton collisions at $\sqrt{s} = 13$ TeV as a function of time. The solid blue and red lines represent the number of showers detected by LHCf-Arm2 and LHCf-Arm1, respectively. The dashed blue and red lines show the number of π^0 candidates detected by LHCf-Arm2 and LHCf-Arm1, respectively, with a scaling factor of 20 applied for clarity.

ing μ. The first dataset (LHCf Run 44299-44472) was collected from 22:32 to 1:30 (CEST) of June 12th to 13th, with an instantaneous luminosity of $L = (3-5) \cdot 10^{28} cm^{-2} s^{-1}$ (as measured by the ATLAS experiment [104]), with μ ranging from 0.007 to 0.012. The second dataset (LHCf Run 44482-45106) was recorded from 1:40 to 12:10 (CEST) on June 13th, with an instantaneous luminosity of $L = (13-17) \cdot 10^{28} cm^{-2} s^{-1}$ [104] and μ between 0.03 and 0.04. The instantaneous luminosities of the two datasets were multiplied for the DAQ efficiency of LHCf (40-50%) and integrated over time. The integrated luminosities for the first and second datasets were calculated to be approximately 0.194 nb^{-1} and 1.9378 nb^{-1}, respectively, with an uncertainty of 1.9%, taken into account in the analysis in Section 3.5.3. Together, these datasets contributed to a comprehensive analysis encompassing a total of 8.4 million triggered events.

3.2.2 Monte Carlo datasets

Two categories of Monte Carlo (MC) simulations were employed to support various aspects of the analysis, all set within the same experimental configuration present during the LHCf operations at LHC in proton-proton collisions at $\sqrt{s} = 13$ TeV:

- a) **Collision Generation Only**: This category focused on the generation of proton-proton collisions at $\sqrt{s} = 13$ and their collision products.
- b) **Full Detector Simulation**: These simulations involved the collision generation (as in point a), the propagation of produced particles from the IP1 to the LHCf detectors through the influence of magnetic fields and interactions with the beam pipe, and finally injecting them into the detector.

Group a) simulations were performed using the Cosmic Ray Monte Carlo package (CRMC [105]), which acts as a frontend for the models under consideration to generate collisions. CRMC is particularly favoured in cosmic ray physics for such purposes. Thus, we created an extensive dataset of simulations using the models typically employed for the simulation of hadronic interactions in the context of EAS physics: QGSJET II-04, EPOS-LHC, SIBYLL 2.3, and DPMJET 3.06. For each model, we considered only particles directly resulting from the collisions or from the decay of unstable particles with a decay length ($c \cdot \tau$) less than 1 cm. Details regarding the number of events and the inelastic cross-section for each model are provided in Table 3.1. Simulations of this type were used to generate the input events for the calculation of the acceptance correction as described in Section 3.4.1, to compare the final experimental results with HIM predictions as shown in Section 3.3 and to validate the analysis procedure by comparing them with the results of the detector-level simulations as discussed in Section 3.7. For category b), the full simulation consists of three steps:

- Hadronic interactions from proton-proton collisions at $\sqrt{s} = 13$ TeV at IP1, performed with the same methodology as simulations of the category a).

Model	QGSJET	EPOS	SIBYLL	DPMJET
σ_{inel} [mb]	80.17	79.98	79.86	80.14
N_{ev}	$9.96 \cdot 10^7$	$9.90 \cdot 10^7$	10^8	10^8

Table 3.1: The total inelastic cross-section (σ_{inel}) for proton-proton collisions at $\sqrt{s} = 13$ TeV and the number of events (N_{ev}) for each HIM employed in the comparison with experimental data. Specific version numbers were omitted. Table from [101].

- Transportation of the resulting particles from IP1 to the LHCf detector locations in the TAN regions. The decay probability of secondary particles was also taken into account, with the resultant daughter particles being further propagated.
- Interaction of the collision products inside the detectors.

The three components of the simulation process were performed using the simulation packages Cosmos 7.633 and EPICS 9.15 [106, 107]. The full simulations began with events produced by two HIMs, QGSJET II-04 and EPOS-LHC. Interactions between the collision products and the detector were modelled using DPMJET 3.04. As detailed in Sections 3.4 and Section 3.5, these complete simulation datasets were used for the calculation of certain correction factors and systematic uncertainties. Additionally, the complete simulation using QGSJET II-04 was employed to determine the energy-dependent cut function for particle identification, discussed in Section 3.3.3.

3.3 η meson reconstruction and selection

The LHCf-Arm2 detector can identify η mesons by reconstructing the two photons generated from the decay $\eta \to \gamma\gamma$, which has a branching ratio of $(39.41 \pm 0.18)\%$ [108]. These two photons can either enter different detector towers (Type I events, as shown in the left panel of Figure 3.6) or both enter the same tower (Type II events, as shown in the right panel of Figure 3.6). The methodologies for reconstructing and selecting η mesons are similar to those used for Type I π^0 analysis [95, 96, 97]. However, Type

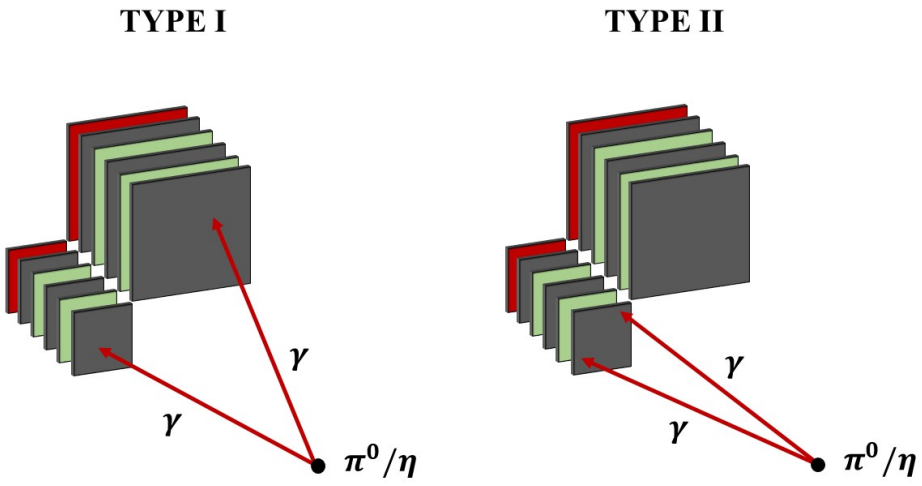

Figure 3.6: Diagrammatic illustration of a π^0 or η meson decay as detected by the LHCf-Arm2. In this setup, photons may either hit two separate towers (Type I event, shown in the left panel) or both hit the same tower (Type II event, shown in the right panel). Note that the depiction of LHCf-Arm2 is not to scale. Figure from [101]

II η studies are not feasible in this analysis due to the low acceptance of such events in the LHCf-Arm2 detector. indeed, The energy threshold for Type II η events is determined by the formula $E_{min} = 2M_\eta L/d_{max}$, where M_η represents the mass of the η meson (547.862 ± 0.018 MeV$/c^2$), L is the distance from Interaction Point 1 (IP1) to the LHCf detector, approximately 141.05 meters, and d_{max} is the maximum separation between the impact points of the two photons. Given that d_{max}, which corresponds to the diagonal of the large tower of LHCf-Arm2, is 39.6 mm, only Type II η mesons with an energy threshold of approximately 3900 GeV can be detected. Due to the lack of such events in our datasets, the analysis is limited to Type I events. The η mesons originating from the collisions decay proximate to the IP1, and the opening angle θ between the photon impact points at the LHCf-Arm2 detector is calculated based on the assumption that the decay occurs right at IP1. Consequently, the open-

ing angle is very small, constrained to $\theta \leq 0.6$ mrad. The kinematic variables of η mesons (energy, p_T, and p_z) are reconstructed using the energies and positions of the impacting photons. The minimum distance between photon impact positions is defined as $d_{min} = 2M_\eta L/E_{max}$, with $E_{max} = 6500$ GeV, resulting in a d_{min} of 23.85 mm. Given that this distance significantly exceeds the position resolution of the LHCf-Arm2 for electromagnetic showers above 200 GeV (approximately 40 μm, as detailed in Section 2.4.1), there is no significant loss of η mesons due to indistinguishable photon pairs. The inclusive production rate of η mesons is expressed using the Feynman-x variable, calculated as $x_F = 2p_z/\sqrt{s}$. While insightful data concerning scaling laws can be obtained from the x_F distribution of η mesons across various p_T bins, the limited statistics only permitted this distribution to be analyzed for a single bin, where $p_T < 1.1$ GeV/c. The data analysis algorithm is composed of five steps: hit position reconstruction, energy reconstruction, single photon identification, η meson reconstruction, and background subtraction.

3.3.1 Hit position reconstruction

The transverse position of particles hitting the LHCf-Arm2 detector is determined through the analysis of lateral profile distributions recorded by the detector position-sensitive silicon microstrip layers. A single silicon detector spans both the Small and Large Towers, hence when determining the position of a particle hitting a calorimeter, only the range of strips corresponding to the respective tower is taken into account. Particles that hit within 2 mm of the edges of the detector towers are rejected to minimize inaccuracies caused by lateral shower leakage. Utilizing an algorithm from the TSpectrum class within the ROOT analysis framework [109], it is possible to distinguish between events with a single particle hitting the tower (single-hit events) from that with more than one particle in the same tower (multi-hit events). Single-hit events exhibit only one peak in the lateral profile distribution in the region corresponding to one of the two towers, whereas multi-hit events display multiple peaks in the same tower region. As an example, Figure 3.7 depicts an event recorded by LHCf with a single peak, corresponding to a single particle reaching the detector, while Figure 3.8 represents an event with two peaks, one for

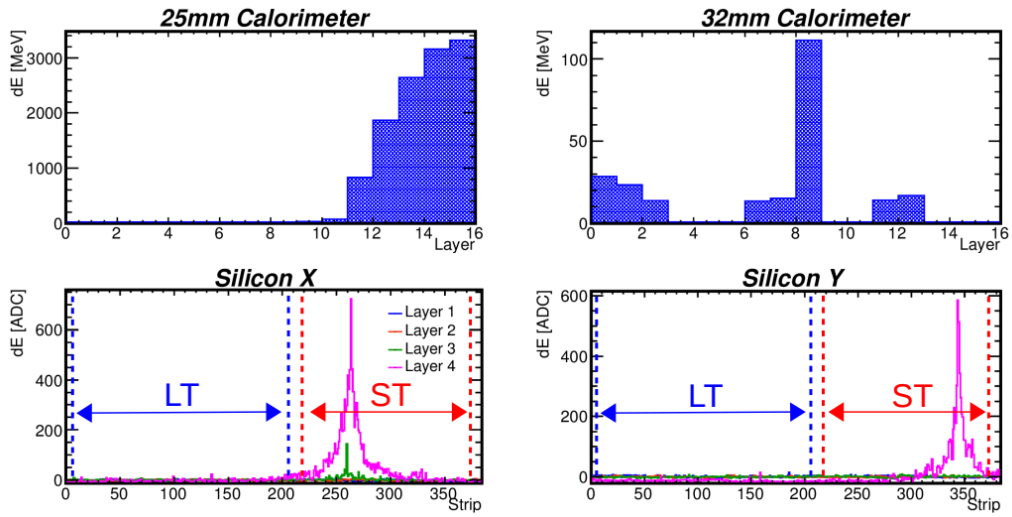

Figure 3.7: An event recorded by the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV. The upper panels illustrate the energy deposited across the layers of the 25 mm (Small Tower) and 32 mm (Large Tower) calorimeters, respectively. The lower panels display the lateral energy distribution recorded by the position-sensitive silicon microstrip layers along the X and Y axes. The clear, singular peak in the lateral profile indicates that this is a single-hit event, with the particle impacting the region corresponding to the Large Tower.

each tower, corresponding to a candidate π^0 or η of Type I. Both events are considered single-hit since both towers were hit by at most one particle. Events classified as multi-hit are excluded from the analysis. Corrections for the loss of η events due to this multi-hit exclusion are detailed in Section 3.4.2. The lateral distributions are fitted using a 3-component Lorentzian function to accurately determine the peak position, height, and width of each shower. The fit function is expressed in Equation 3.1:

$$f(x) = p_0 \left[\frac{p_2}{\frac{(x-p_1)^2}{p_3} + p_3} + \frac{p_4}{\frac{(x-p_1)^2}{p_5} + p_5} + \frac{1-p_2-p_4}{\frac{(x-p_1)^2}{p_6} + p_6} \right]. \tag{3.1}$$

In this context, x represents the transverse coordinate, either X or Y, depending on the orientation of the strips. The parameters p_0 through p_6

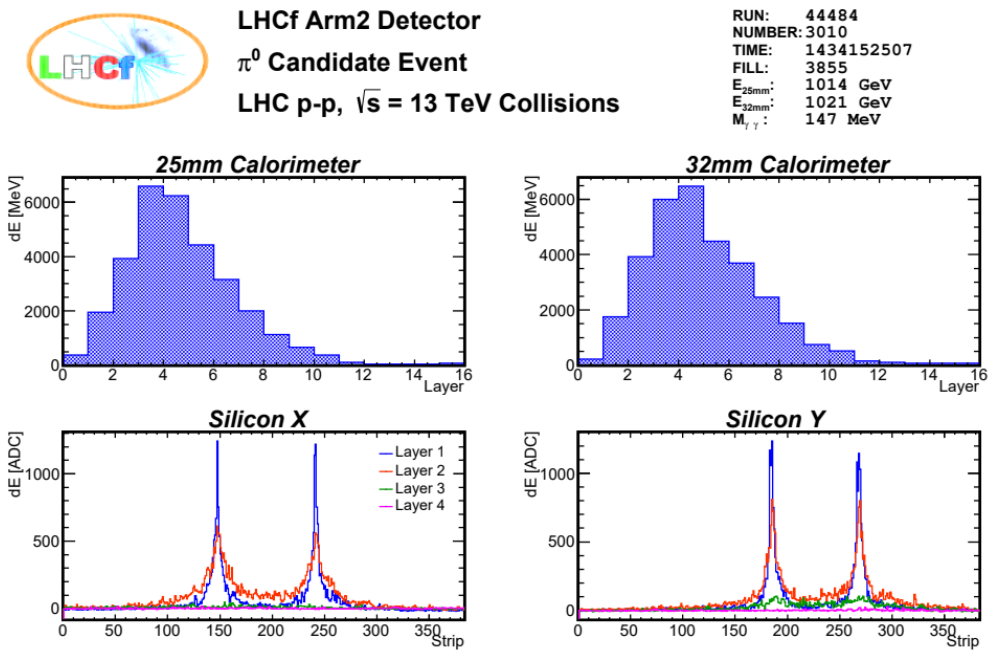

Figure 3.8: An event recorded by the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV. The upper panels illustrate the energy deposition across the layers of the 25 mm (Small Tower) and 32 mm (Large Tower) calorimeters, respectively. The lower panels display the lateral energy distribution recorded by the silicon microstrip layers along the X and Y axes. The two distinct peaks in the lateral profiles correspond to energy deposits in both the Small Tower (ST) and the Large Tower (LT), indicating a candidate Type I π^0 or η meson decay, where each photon from the decay enters a different tower.

are free parameters within the model. These parameters are initialized based on the results derived from the TSpectrum algorithm. The reconstructed position was defined as the parameter p_1 after the fit in the pair of X-Y detectors with the maximum energy deposit. For events with more than one particle hitting the same tower, the fit function is defined as the sum of N superimposed Lorentzian functions (Equation 3.1), where N is the number of peaks found by TSpectrum in the same tower.

3.3.2 Energy reconstruction

The reconstruction of particle energies from the energy deposited in scintillators involved several steps:

- **Energy Deposit Conversion:** The energy deposited in the scintillator layers, measured in ADC units, was converted to reconstructed energy using the formula:

$$dE_j^i = S_j^i \cdot \frac{G_j(HV_j)}{C_j} \cdot \frac{1}{P_j(x_i, y_i)} \cdot \frac{1}{A_j}, \tag{3.2}$$

  where:

  - S_j^i are the energies deposited in the j-th scintillator for the i-th event, adjusted by subtracting the pedestal.
  - C_j are the conversion factors determined during beam tests at SPS, as detailed in [88].
  - $G_j(HV_j)$ correct for the difference in PMT gain between 600 V (used at SPS) and the operational voltage HV_j at LHC, which varies from 375 V to 450 V across different layers.
  - $P_j(x_i, y_i)$ correct for non-uniformity in detector response at the reconstructed position (x, y), discussed in [88].
  - A_j are the attenuation factors for the analog signal along the approximately 200 meters of cable connecting the detector to the readout electronics, measured using a cable replica at SPS.
  - dE_j^i are the converted energy deposits for the j-th scintillator in the i-th event, expressed in GeV and corrected for position-dependent signal variations.

  This step is omitted when analyzing MC simulations.

- **Total Energy Release:** The total energy deposited in the calorimeter was calculated by summing the energy deposits across layers:

$$SumdE = \sum_{j=2}^{13} dE_j \cdot w_j, \tag{3.3}$$

where dE_j are the energies deposited in the j-th layer and w_j are weighting factors that equal 1 for layers up to 11th and 2 for layers 12th and beyond, to take into account differences in scintillator sampling steps. The 1st and the 14th to 16th layers were excluded to avoid noise and background contributions due to the low energy deposited in these layers by the showers.

- **Energy Estimation:** The sum of the deposits, $SumdE$, was then used to estimate the energy of the primary particle E as

$$E = J_k(SumdE), \tag{3.4}$$

where $J_k(SumdE)$ are the conversion functions (one for each of the two calorimetric towers) estimated by a simulation where electrons with six mono-energies of 100, 200, 500, 1000, 3000, and 6000 GeV injected in a 1 mm x 1 mm square at the center of the Small Tower ($k = 0$) and Large Tower ($k = 1$), respectively.

Events with energies below 200 GeV are excluded from the analysis to filter out particles produced through interactions in the beam pipe and to minimize uncertainties caused by trigger inefficiencies.

3.3.3 Single photon identification

The particle identification (PID) algorithm used in this analysis is designed to distinguish photons from the neutral hadron background, predominantly neutrons. This selection is achieved by utilizing a key feature of shower profiles, the longitudinal distance from the first calorimeter layer to the point where 90% of the shower total energy has been deposited. This measurement, referred to as $L_{90\%}$, is expressed in radiation lengths [X_0]. Previous studies have shown that $L_{90\%}$ is highly effective in discriminating between pure electromagnetic showers and those resulting from hadrons. Indeed, since Tungsten has a shorter radiation length compared to its interaction length, electromagnetic showers are expected to exhibit lower $L_{90\%}$ values compared to hadronic showers. This distinction is observable in Figure 3.9, which displays the $L_{90\%}$ distribution in units of radiation length X_0 for all showers developing in the LHCf-Arm2 Small Tower. The PID criteria are formulated as a function of the

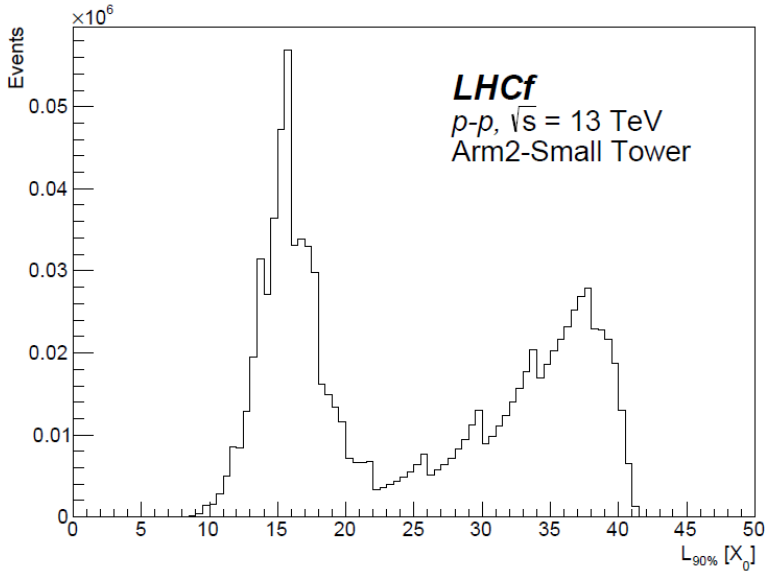

Figure 3.9: The $L_{90\%}$ distribution for events with energy above 200 GeV in the LHCf-Arm2 Small Tower, in proton-proton collisions at $\sqrt{s} = 13$ TeV. The peak at approximately 16 X_0 is associated with electromagnetic showers while the peak near 37 X_0 corresponds to hadronic showers. Small structures at 26, 30, and 34 X_0 in the distribution are attributed to the discrete sampling of the shower.

particle energy, $f_{L_{90\%}}(E)$, in order to ensure a constant selection efficiency of 90% across all photon energy ranges, that corresponds to about 95% efficiency on single photon selection. Two separate functions are derived for each tower using the full simulation based on the QGSJET II-04 model described in Section 3.2.2. To define these cut functions, we implemented all the single-photon selection criteria described in earlier sections and detailed in Table 3.2. Only photon pairs originating from η decays were considered in the simulation, as per the MC truth, to maintain controlled selection efficiency within the energy range significant for η detection. This control is crucial since the majority of photons meeting the selection criteria are typically products of π^0 decays, which due to mass differences, exhibit distinct kinematics from the ones produced by decay of η mesons, as shown in Figure 3.10, where the distributions in the left and

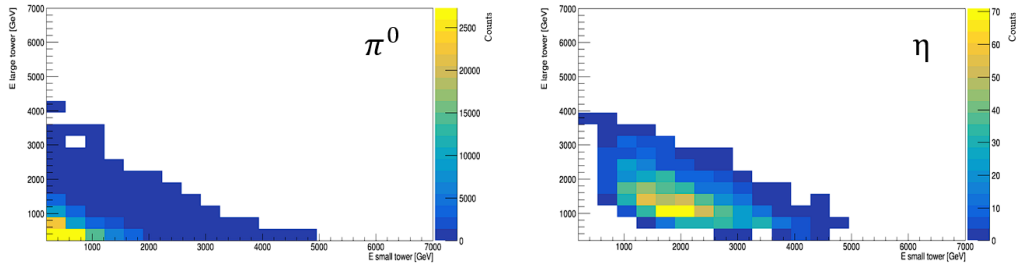

Figure 3.10: Scatter plots comparing the energy distributions of photon pairs detected in the Small Tower (x axis) and the Large Tower (y axis). The left panel shows the energy distribution for photon pairs originating from π^0 decays, while the right panel displays the corresponding distribution for photons from η decays.

right panels compare the Type I photon energy in the two towers produced in the decays of π^0 and η mesons, respectively. The cut functions are calculated by plotting the threshold values of $L_{90\%}$ for each energy bin to ensure a 90% η selection efficiency and then by fitting the obtained distributions with a function:

$$f_{L_{90\%}}(E) = t_0 \log(1 + t_1 E) + t_2 + t_3 e^{t_4 E}, \tag{3.5}$$

where E is the photon energy and $t_0 ... t_4$ are free parameters. In order to calculate the systematic uncertainty due to the particle identification method described in Section 3.5.2, this procedure was repeated by modifying the value of the efficiency required to calculate the threshold value of $L_{90\%}$ for each energy bin, to obtain two more pairs of functions to select

Event type	Type I
Number of hits	Single-hit for each tower
Incident position	Within 2 mm from the edge of the calorimeter
Energy threshold	E > 200 GeV
PID criteria	Photonlike [$L_{90\%} < f_{L_{90\%}}(E)$]

Table 3.2: List of single-photon selection criteria for the reconstruction of η mesons. Table from [101].

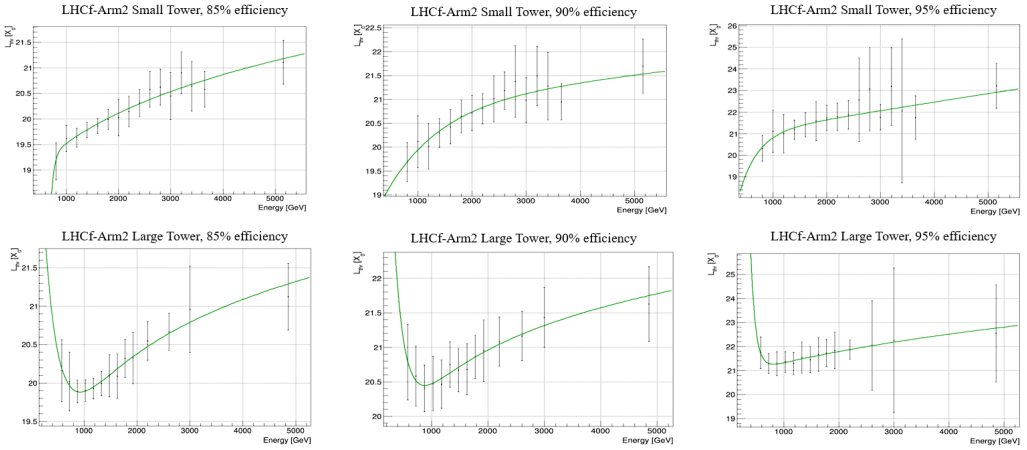

Figure 3.11: $f_{L_{90\%}}(E)$ distributions displayed for the LHCf-Arm2 Small Tower (top) and Large Tower (bottom) and for the three values of η selection efficiency used in the analysis, namely 85% (left), 90% (center) and 95% (right). The fit functions are illustrated in green.

with an efficiency of 85% and 95%, respectively. The values of the fit parameters obtained for the two towers of the LHCf-Arm2 detector and the three imposed detection efficiencies are given Table3.3, while the graphs of $f_{L_{90\%}}(E)$ together with the fit curves are shown in Figure 3.11. The residual hadron contamination, generally around 10%, can be estimated

Tower	Small Tower			Large Tower		
Efficiency	85%	90%	95%	85%	90%	95%
t_0	1.55	10.60	45.78	1.07	$9.50 \cdot 10^{-1}$	2.32
t_1	$8.52 \cdot 10^{-4}$	$1.32 \cdot 10^{-5}$	$8.97 \cdot 10^{-6}$	$1.04 \cdot 10^{-2}$	$4.87 \cdot 10^{-3}$	$2.88 \cdot 10^{-4}$
t_2	18.57	20.85	20.84	17.09	16.70	20.72
t_3	$-17.23 \cdot 10^{4}$	-2.82	-8.33	9.47	18.75	489.64
t_4	$-1.71 \cdot 10^{-2}$	$-9.88 \cdot 10^{-4}$	$-2.91 \cdot 10^{-3}$	$-3.83 \cdot 10^{-3}$	$-5.40 \cdot 10^{-3}$	$-1.15 \cdot 10^{-2}$

Table 3.3: Values of the fit parameters obtained for the two towers of the LHCf-Arm2 detector (Small Tower and Large Tower) and the three imposed detection efficiencies. The parameters corresponding to 90% efficiency were used to select individual photons, while those corresponding to 85% and 95% were used to calculate the systematic uncertainty of particle identification.

as a function of energy via MC simulation and is corrected alongside the 90% efficiency. This contamination is then accounted for during the background subtraction process as outlined in Section 3.3.4. Selection inefficiencies are corrected in Section 3.4.3.

3.3.4 η reconstruction and background subtraction

Candidates for η mesons are identified by the characteristic peak in the di-photon invariant mass spectrum around the η meson rest mass. The invariant mass $M_{\gamma\gamma}$ is calculated using the energies of the selected photon pairs and their positions according to the formula:

$$M_{\gamma\gamma} = \sqrt{2E_1E_2(1-\cos\theta)}, \tag{3.6}$$

where E_1 and E_2 represent the energies of the two photons, and θ is their opening angle in the laboratory frame. The invariant mass distribution in the region of the η rest mass for the experimental dataset is shown in Figure 3.12. The peak of the distribution was observed at $M_{\gamma\gamma} = (533.3 \pm 1.1)$ MeV/c^2. The world averaged rest mass of the η meson is $M_\eta = 547.86$ MeV/c^2 [108], indicating a deviation of (-2.65 ± 0.20)% in our measurements. This shift was confirmed to be consistent with that observed in the invariant mass peak from π^0 decays into two photons, each hitting different towers of the LHCf-Arm2 (Type I events), where a shift of (-2.57 ± 0.04)% with respect of the world averaged rest mass of $M_{\pi^0} = 134.98$ MeV/c^2 was observed, as depicted in Figure 3.13, where are presented the invariant mass distribution, in the respective region, of π^0 and η mesons, together with the results of the fit described later in this Section. As explained in section 3.5.1, given that the two shift values are within error margins and are compatible with the uncertainty on the absolute energy scale, we adjusted the energies of individual photons upwards by 2.65% to align the peak position with the reference value. Following this adjustment, the peak for the η meson was accurately positioned at $M_{\gamma\gamma} = (548.1 \pm 1.1)$ MeV/c^2. The errors reported for invariant mass values and shifts are statistical only. Due to the low statistics of η in the dataset, it was not possible to extract the x_F distribution and eliminate the residual background using a template fit for each x_F bin. Consequently, a sideband method [96] was adopted. Initially, we performed

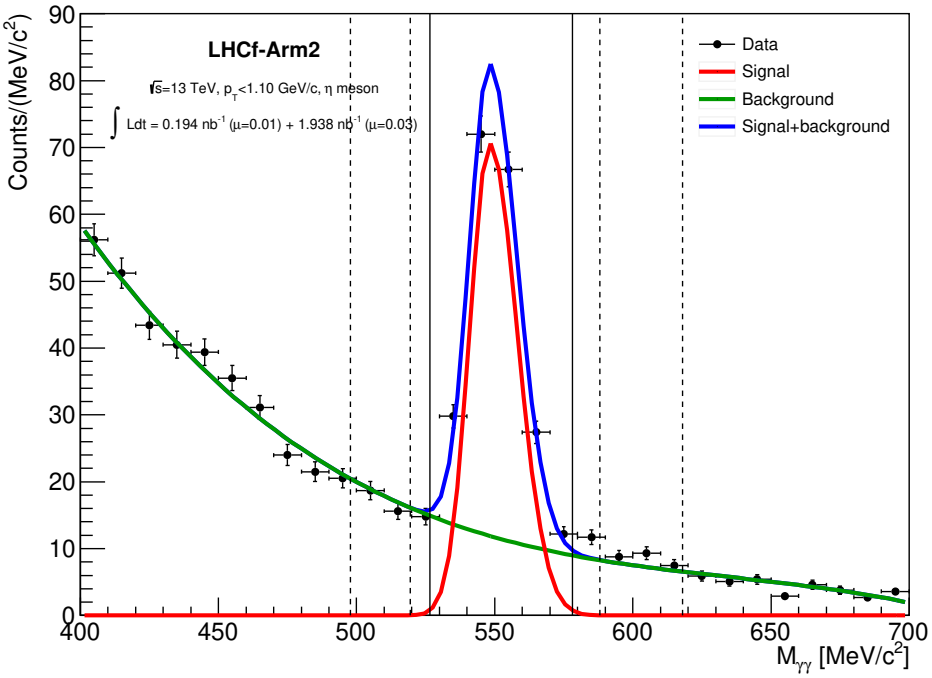

Figure 3.12: Invariant mass distribution of di-photon events ($M_{\gamma\gamma}$) detected by the LHCf-Arm2 detector inproton-proton collisions at $\sqrt{s} = 13$ TeV. The data points (black) represent the experimental measurements. The red curve indicates the signal component fit (asymmetric Gaussian) while the green curve indicate the background contribution fit (third-order Chebyshev polynomial). The blue curve represents the sum of the signal and background components. The vertical dashed lines indicate the signal region and the two background regions. Figure from [101].

a binned fit of the distribution using a composite model comprising an asymmetric Gaussian function for the signal:

$$S(x) = \begin{cases} Ae^{-\frac{(x-\mu)^2}{2\sigma_l^2}} & \text{if } x < \mu, \\ Ae^{-\frac{(x-\mu)^2}{2\sigma_r^2}} & \text{if } x > \mu \end{cases}, \tag{3.7}$$

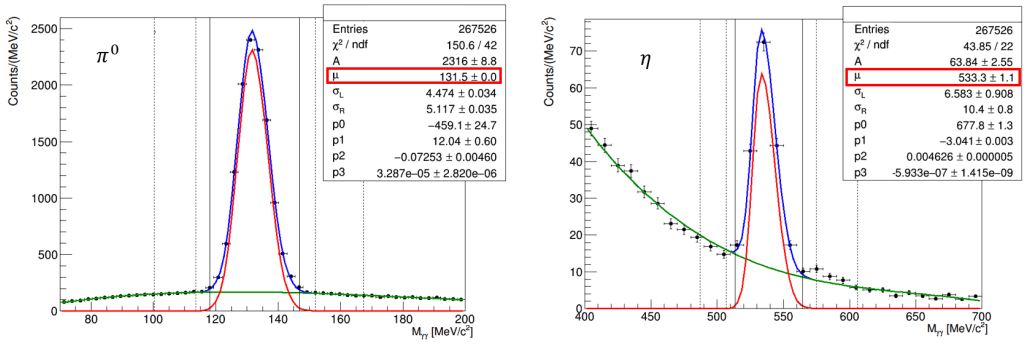

Figure 3.13: Invariant mass distributions of di-photon events ($M_{\gamma\gamma}$) for π^0 (left) and η (right) mesons detected by the LHCf-Arm2 detector during proton-proton collisions at $\sqrt{s} = 13$ TeV. The data points (black) represent the experimental measurements, with the red curve indicating the signal component fit (asymmetric Gaussian) and the green curve showing the background contribution fit (third-order Chebyshev polynomial). The blue curve represents the sum of the signal and background components. The peak positions (μ) of the fitted distributions are highlighted in red, showing a deviation from the world average rest masses of the π^0 and η mesons by $-2.57 \pm 0.04\%$ and $-2.65 \pm 0.20\%$, respectively.

and a third-order Chebyshev polynomial for the background:

$$B(x) = p_0 + p_1 x + p_2(x^2 - 1) + p_3(4x^3 - 3x), \tag{3.8}$$

where μ is expected the mean, σ_l and σ_r are the left and right 1σ deviations, and A, p_0, p_1, p_2, p_3 are free parameters. μ, σ_l and σ_r defined the signal region within $[\mu - 3\sigma_l, \mu + 3\sigma_r]$ and two background regions within $[\mu - 7\sigma_l, \mu - 4\sigma_l]$ and $[\mu + 4\sigma_r, \mu + 7\sigma_r]$. The x_F distributions for the signal, $\left[f^{sig}(x_F)\right]$, are calculated by subtracting the background distribution, $\left[f^{bkg}(x_F)\right]$, determined in the background regions, from the combined signal and background distribution, $\left[f^{sig+bkg}(x_F)\right]$, taken from the signal region. The results of the fits and the signal and background regions are depicted in Figure 3.12 for η only and in Figure 3.13 for both π^0 and η. The fraction of the background component within the signal region is estimated by normalizing $\left[f^{sig}(x_F)\right]$. The normalization factor $N(x_f, \mu, \sigma_l, \sigma_r)$ is derived from the likelihood function $\mathcal{L}_{bkg}(x_F, M_{\gamma\gamma})$,

characterized by the best-fit third-order Chebyshev polynomial function (Equation 3.8). Thus, the signal distribution with background subtracted is given by:

$$\left[f^{sig}(x_F)\right] = \left[f^{sig+bkg}(x_F)\right] - N(x_f, \mu, \sigma_l, \sigma_r)\left[f^{bkg}(x_F)\right], \tag{3.9}$$

where $N(x_f, \mu, \sigma_l, \sigma_r)$ is calculated as:

$$N(x_f, \mu, \sigma_l, \sigma_r) = \frac{\int_{\mu-3\sigma_l}^{\mu+3\sigma_r} \mathcal{L}_{bkg}\, dM_{\gamma\gamma}}{\int_{\mu-7\sigma_l}^{\mu-4\sigma_l} \mathcal{L}_{bkg}\, dM_{\gamma\gamma} + \int_{\mu+4\sigma_r}^{\mu+7\sigma_r} \mathcal{L}_{bkg}\, dM_{\gamma\gamma}}. \tag{3.10}$$

This method resulted in the identification of approximately 1500 η mesons in the experimental dataset. The uncertainty of the background subtraction method was assessed using the fully reconstructed QGSJET II-04 simulation, as detailed in Section 3.5.4. The same procedure was performed for the two full detector simulations described in Section 3.2.2, based on QGSJET II-04 and EPOS-LHC. The invariant mass distributions for the two models are depicted in Figure 3.14. The η detector-level x_F distributions for the data and the two MC datasets are presented in Figure 3.15. Note that the uncertainty in the Figure is obtained by summing in quadrature the statistical error and the contributions to the systematic errors due to the operative conditions, discussed in Section 3.5.

3.4 Corrections for experimental effects

This section discusses the corrections applied to the detector-level x_F distribution of η mesons depicted in Figure 3.15. Detailed descriptions of each correction are provided in Sections 3.4.1 to 3.4.3. For the selection and multi-hit corrections, we utilized the two full detector MC simulations, which are detailed in section 3.2.2 and based on the QGSJET II-04 and EPOS-LHC models. In this case, the final correction was obtained by averaging the results of the two simulations. This choice is supported by the fact that as shown in Figure 3.15, the distribution of the experimental data lies approximately in the middle between the results of the two

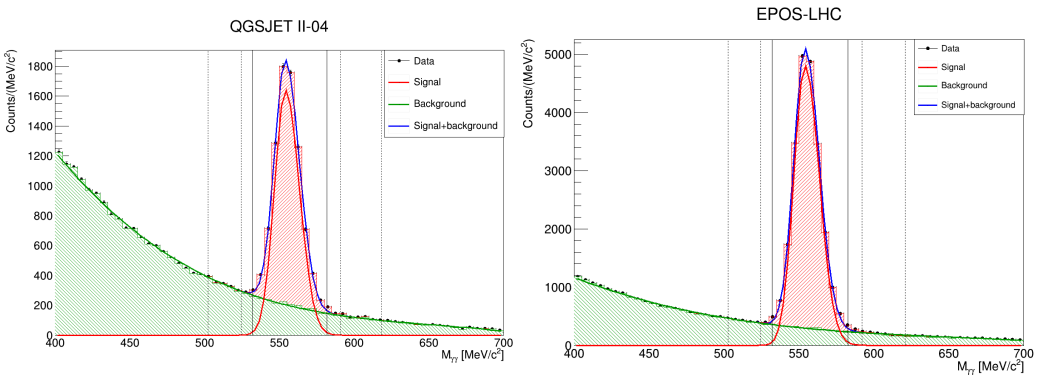

Figure 3.14: Invariant mass distributions of di-photon events ($M_{\gamma\gamma}$) from η meson decays in the simulations based on QGSJET II-04 (left) and EPOS-LHC (right) models. The data points are depicted in black, while the red curve indicates the signal component corresponding to the η meson peak. The green shaded area represents the background contribution, and the blue curve shows the sum of the signal and background components. The vertical dashed lines indicate the signal and background regions. The filled red and green areas represent the true distribution of signal and background, respectively.

models. On the other hand, for the acceptance correction, we used four toy MC simulations based on the collision generation results of all the models used in the analysis, namely QGSJET II-04, EPOS-LHC, SIBYLL 2.3, and DPMJET 3.06, described in Section 3.2.2. Figure 3.16 illustrates the distribution of correction factors as a function of x_F. To validate the accuracy of the correction process, a cross-check was performed where the particle-level distributions from the MC simulations were compared with the corrected detector-level MC distributions. This validation, discussed in Section 3.7, confirmed that the applied corrections accurately reproduced the true distribution of η mesons.

3.4.1 Acceptance and branching ratio correction

First, the detector-level distribution was corrected for the limited aperture of the LHCf-Arm2 detector, which does not encompass the full 2π azimuthal angle. This limitation introduce a bias in the measurement

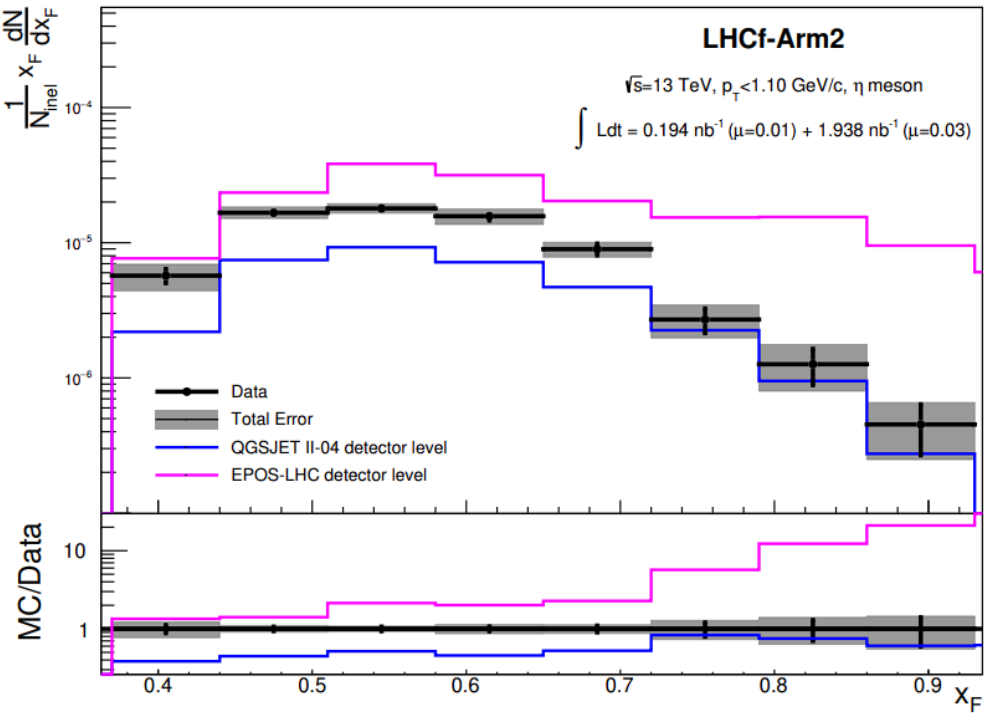

Figure 3.15: The detector-level x_F distributions for η mesons detected by the LHCf-Arm2 detector during proton-proton collisions at $\sqrt{s} = 13$ TeV. The black points represent the experimental data with the total error shown in grey. The solid blue and magenta histograms correspond to the detector-level x_F distributions from the QGSJET II-04 and EPOS-LHC full simulation, respectively. The lower panel shows the ratio of the MC predictions to the experimental data for both models, illustrating the level of agreement and deviations across the x_F range. As discussed in Section 3.5, only systematic uncertainties based on the operating conditions were considered in this stage. Figure from [101].

since only a fraction of the produced η mesons is detected. To accurately account for this effect, the acceptance correction factor was determined through toy MC simulations, using predictions from four HIMs: QGSJET II-04, EPOS-LHC, DPMJET 3.06, and SIBYLL 2.3 (collision generation only, as described in Section 3.2.2). In these simulations, the p_t-x_F

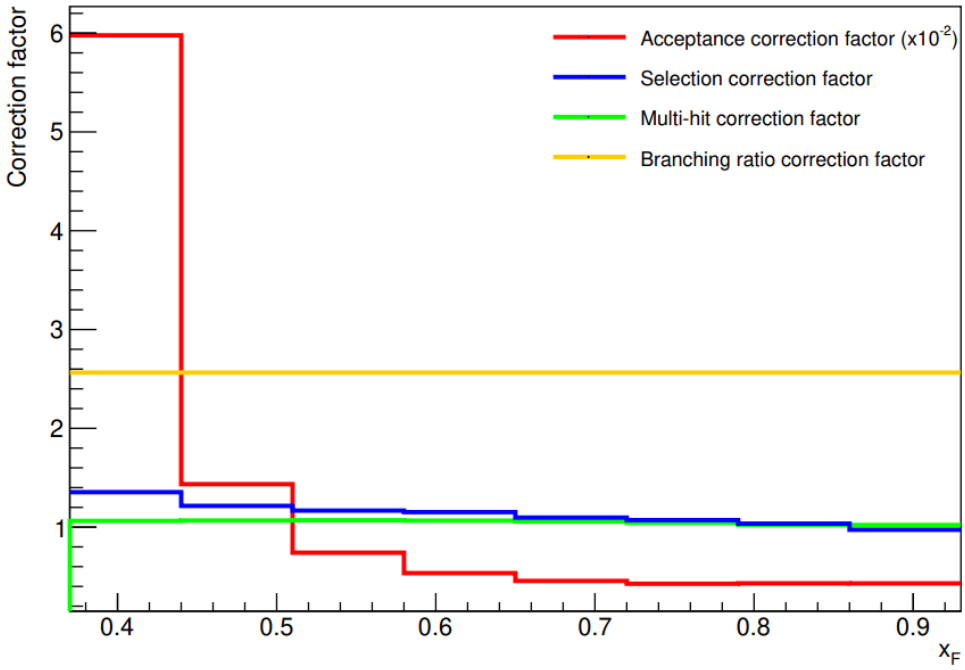

Figure 3.16: Summary of the correction factors applied to the experimental η meson detector-level distribution as a function of x_F. The figure includes the acceptance correction factor (red line), selection correction factor (blue line), multi-hit correction factor (green line), and branching ratio correction factor (yellow line). The acceptance correction factor is scaled by a factor of 10^{-2} for better visualization. Figure from [101].

phase space of η mesons was first generated according to the respective predictions of each HIM. The generated η mesons were then subjected to a simulation of their decay process $\eta \rightarrow \gamma\gamma$. After selecting only events with the two photons included in the geometric acceptance of LHCf-Arm2 and meeting the selection criteria listed in Table 3.2, we extracted their p_t-x_F phase space. Geometrical acceptance efficiency was calculated by dividing the phase space of accepted events by the phase space of the generated events. The four acceptance maps are depicted at the top panels of Figure 3.17, together with a red box indicating the analysis region in the p_t-x_F plane, defined as $p_T < 1.1$ GeV/c and $0.37 \leq x_F < 0.93$.

The acceptance correction distributions for the four simulations, depicted in the bottom panel of Figure 3.17, were calculated as the inverse of the geometrical acceptance efficiency. To align the simulation results with the experimental data and ensure consistency in the analysis, the acceptance maps shown in the top panels of Figure 3.17 were rebinned according to the binning scheme used in the experimental distribution. Specifically, the analysis was conducted within the $p_T < 1.1$ GeV/c range, aggregated into a single bin, and within the $0.37 \leq x_F < 0.93$ range, which was divided into 8 bins. This rebinning ensures that the acceptance corrections can be directly applied to the experimental data. The acceptance correction factors varied among the models due to differences in the p_t-x_F spectrum shapes within the x_F bins analyzed. The final acceptance correction was calculated by averaging the results of the four simulations, and it is presented in Figure 3.16 (red line). A systematic uncertainty was evaluated using the method outlined in Section 3.5.5 to accommodate differences between the simulations. Additionally, a correction was made for the branching ratio of η decay into two photons, which is $(39.41 \pm 0.18)\%$. This inefficiency was compensated by applying a constant factor across the entire x_F range. Given that the uncertainty of the branching ratio for the two-photon decay of η meson is under 0.5%, its impact on the total uncertainty was considered negligible. The x_F distribution of the branching ratio correction is shown in Figure 3.16 (yellow line).

3.4.2 Multi-hit correction

Second, the inefficiency arising from the rejection of multi-hit events was calculated. This inefficiency results from rejecting η meson events that involve more than one particle per tower in the LHCf-Arm2 detector, as detailed in Section 3.3.1. The multi-hit correction factor for each bin i of x_F is defined as:

$$f^i_{multi-hit} = \frac{N^i_{multi-hit} + N^i_{single-hit}}{N^i_{single-hit}}, \tag{3.11}$$

where $N^i_{multi-hit}$ is the number of multi-hit events, and $N^i_{single-hit}$ is the number of single-hit events, each within the respective bin i of x_F. Two

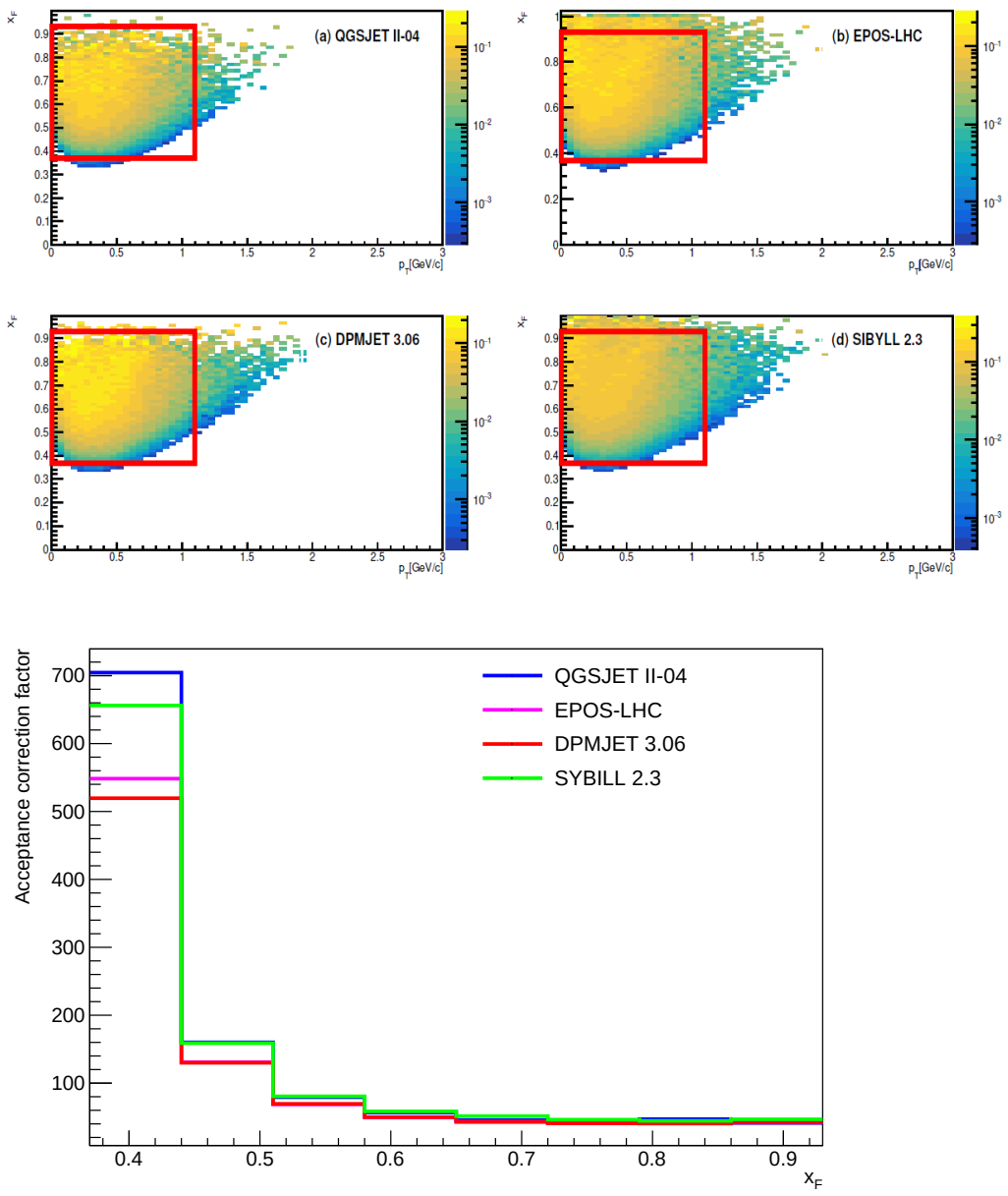

Figure 3.17: *Top:* Acceptance maps for detecting η mesons with the LHCf-Arm2 detector across the p_t-x_F phase space, generated using four different HIMs: (a) QGSJET II-04, (b) EPOS-LHC, (c) DPMJET 3.06, and (d) SIBYLL 2.3. The red box in each panel indicates the analysis region, defined by $p_T < 1.1$ GeV/c and $0.63 \leq x_F < 0.93$. The maps were rebinned to match the binning of the experimental data, with p_T integrated into a single bin and x_F divided into 8 bins across the range $0.37 \leq x_F < 0.93$. *Bottom:* Acceptance correction factor as a function of x_F for the four simulations. The correction factors are computed as the inverse of the geometrical acceptance efficiency depicted in the top panels. Figure from [101].

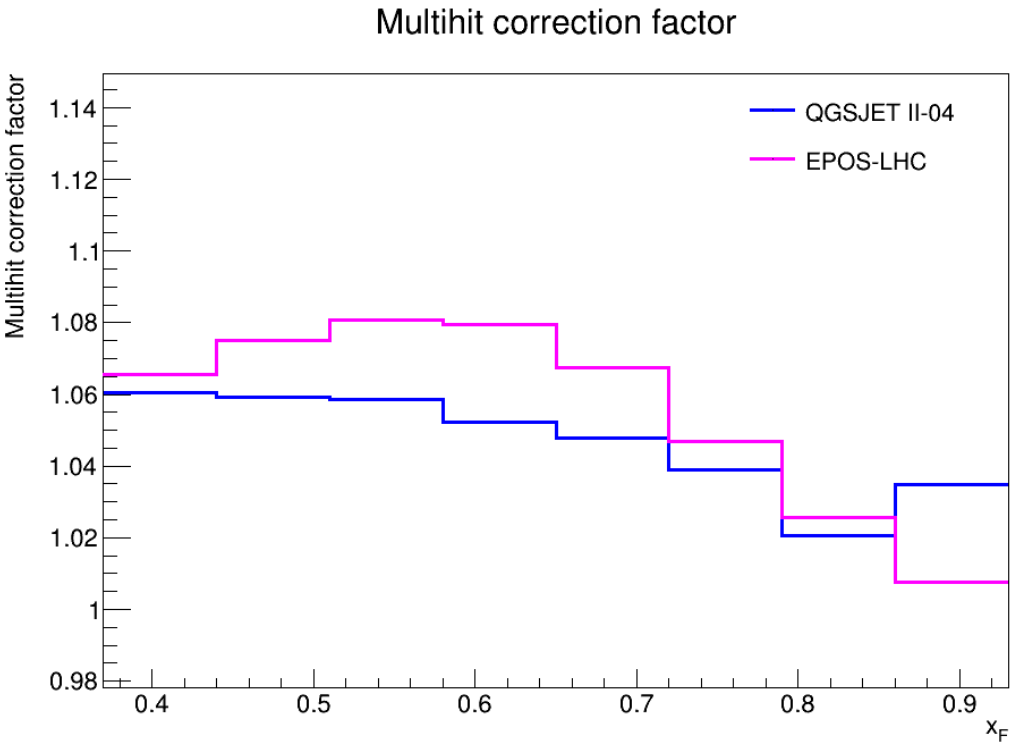

Figure 3.18: Multi-hit correction factors as a function of x_F, derived using the simulations based on QGSJET II-04 (blue) and EPOS-LHC (magenta) models. The correction factors were averaged to obtain the final multi-hit correction applied to the detector-level distribution. The differences between the models were considered as part of the systematic uncertainty.

corrections were derived from the results of the full detector simulations using the QGSJET II-04 and EPOS-LHC models described in Section 3.2.2. The correction applied was calculated as the average of the two corrections, depicted in Figure 3.18. To accommodate differences between these two models, the signal distribution was adjusted using the average of the correction factors from both models, and an additional systematic error was computed as outlined in Section 3.5.5. The x_F distribution of the multi-hit correction is shown in Figure 3.16 (green line).

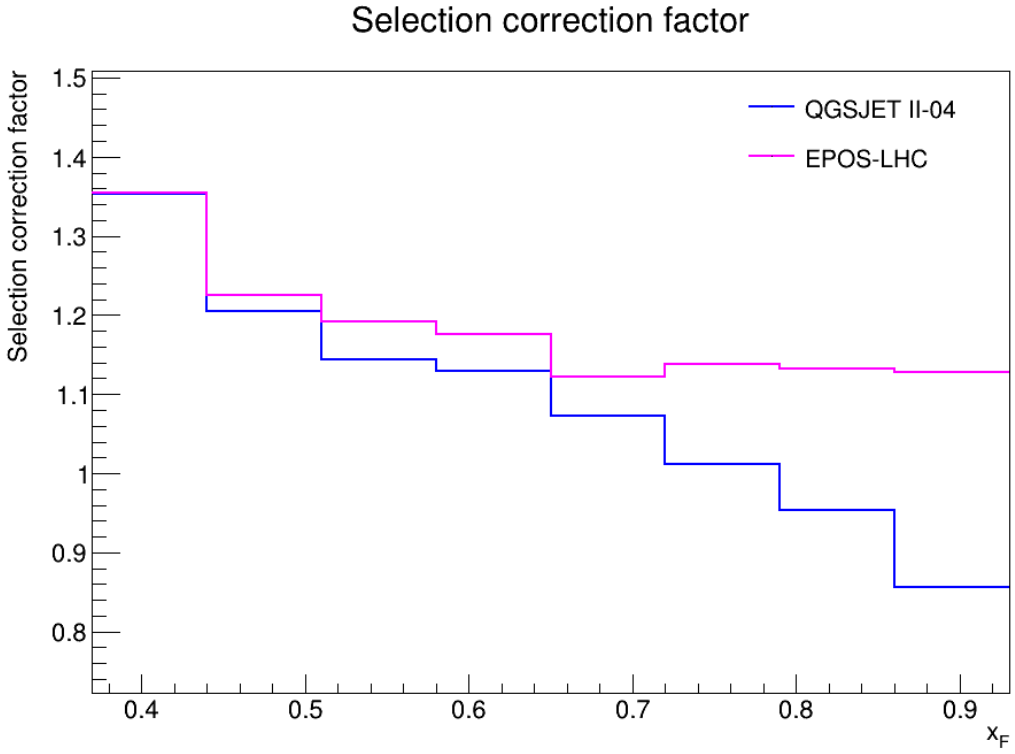

Figure 3.19: Selection correction factors as a function of x_F, derived using the simulations based on the QGSJET II-04 (blue) and EPOS-LHC (magenta) models. The correction factors were averaged to obtain the final selection correction applied to the signal distribution. Differences between the models were considered as part of the systematic uncertainty.

3.4.3 Selection correction

Finally, the signal distribution was corrected for both selection inefficiency and smearing effects. These corrections were applied simultaneously by utilizing the full detector simulations based on the QGSJET II-04 and EPOS-LHC models, as described in Section 3.2.2. For each model, we calculated the ratio between the detector-level x_F distribution of η candidates, derived using the same reconstruction algorithm as the experimental data, and the true x_F distribution of η mesons. The results of the two simulations are depicted in Figure 3.19. The final correction

factor was determined by averaging the ratios obtained from both models. The x_F distribution of the selection correction is shown in Figure 3.16 (blue line). To account for potential discrepancies between the QGSJET II-04 and EPOS-LHC models, a systematic uncertainty associated with the correction was estimated as detailed in Section 3.5.5.

3.5 Systematic uncertainties

This section discusses the estimation of the total uncertainty on the x_F distribution of η mesons. The systematic uncertainties can be categorized into two types: those related to the operating conditions of the experiment and those arising from the use of MC simulations in the data analysis. Each source of error is described in detail in Sections 3.5.1–3.5.5. The total systematic error is calculated by summing the contributions from each source quadratically, togheter with the statistical error. The uncertainties related to operating conditions include the energy scale, particle identification, beam-center, and luminosity errors, while those associated with MC simulations encompass acceptance correction, multi-hit correction, background subtraction, and selection correction errors. Note that only the errors related to operating conditions were used to calculate the total uncertainties of experimental data at the detector level (Figure 3.15), while all error sources were used to calculate the total uncertainty on the final corrected distribution. Figure 3.20 presents the estimated uncertainties as a function of x_F.

3.5.1 Energy scale

As discussed in Section 3.3.4, systematic shifts in the invariant mass peaks of π^0 and η mesons were observed relative to the world-averaged rest mass values of these particles. The discrepancies were found to be $(-2.57 \pm 0.04)\%$ for π^0 and $(-2.65 \pm 0.2)\%$ for η. These values were consistent within their errors and compatible with the uncertainty on the absolute energy scale, which was calculated using beam test data at SPS and comes from various contributions [98]:

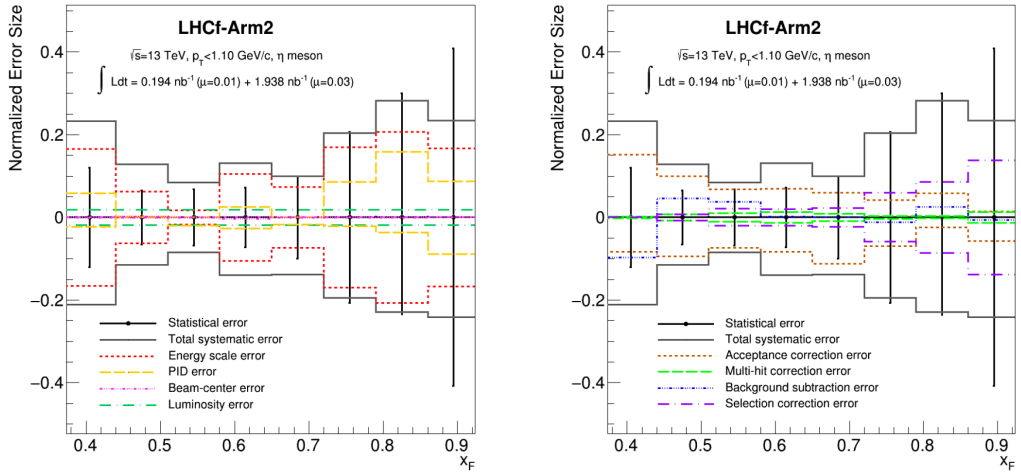

Figure 3.20: Relative uncertainties as a function of x_F for η mesons detected by the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV with $p_T < 1.1$ GeV/c. The left plot shows the contributions of the systematics related to the operating conditions: energy scale, PID, beam-center, and luminosity errors, while the right plot presents the uncertainties due to the use of MC simulations in the data analysis: acceptance correction, multi-hit correction, background subtraction, and selection correction errors. The total systematic error is calculated as the quadratic sum of these contributions, represented by the dashed lines. The statistical error is also shown for comparison (black markers). Figure from [101].

- **PMT gain calibration:** Uncertainty in the relationship between the high voltage applied and the gain of the PMTs.
- **Signal non-uniformity:** Uncertainty arising from the spatial dependence of the signal within the detector.
- **Conversion factor uncertainty:** Error associated with the calibration of conversion factors used to convert ADC signals into physical measurements [88].
- **LPM effect impact:** Uncertainty due to the Landau-Pomeranchuk-Migdal (LPM) effect, which reduces the pair production cross-section

Contribution	Error (%)
PMT gain table	1.8
Non-uniformity	1.1
Conversion factors	1.0
LPM effect (calibration)	0.8
LPM effect (energy reconstruction)	0.7
H.V. table	0.7
Cable attenuation	0.5
ADC linearity	0.1
Temperature correction	0.1

Table 3.4: List of all the sources contributing to the total uncertainty on the absolute energy scale.

for high-energy photons, affecting both the calibration and energy reconstruction processes.

- **High voltage measurement error:** Uncertainty in the PMT gain due to inaccuracies in the applied high voltage measurement.
- **Cable attenuation uncertainty:** Error in determining the signal attenuation introduced by the cables used in the detector setup.
- **ADC non-linearity:** Uncertainty due to deviations from linearity in the analog-to-digital converters (ADCs) used to digitize the detector signals.
- **Temperature correction error:** Uncertainty related to the correction made to account for temperature variations affecting the PMT gain.

The error contributions of all the effects described above are listed in Table 3.4. The total uncertainty on the absolute energy scale was calculated by summing in quadrature the single contributions and was found to be $\pm 2.7\%$. Since the invariant mass shift of the two mesons was compatible with the total uncertainty on the absolute energy scale, the peaks were restored to their proper positions by increasing the energies of individual photons by $+2.65\%$. To assess the uncertainty of the energy scale correction, we evaluated its stability as a function of energy. This was done

separately for each tower by analyzing Type II events, where all the energy is deposited in a single tower. Since Type II η mesons are mainly outside the detector acceptance, we used Type II π^0 mesons from the same dataset of η for this purpose (described in Section 3.2.1). First, we computed the invariant mass distributions for Type II π^0 events in several x_F bins for both the Small Tower (Figure 3.21) and Large Tower (Figure 3.22), using the same selection criteria listed in Table 3.2, except for the type of event requested (Type II instead of Type I). Following, we fitted all the distributions with the compound function described in Section 3.3.4, composed of an asymmetric Gaussian (Equation 3.7) and a third-order Chebyshev polynomial (Equation 3.8). Fit functions are depicted in Figure 3.21 and Figure 3.22 in red line. The central position of the peaks as a function of x_F is shown in Figure 3.23 (right panel for Small Tower and left panel for Large Tower). The mean values of $M_{\gamma\gamma}$ for the Type II π^0 mesons were found to be stable, with maximum variations within 1% across the entire x_F range for both towers. Finally, systematic errors on the energy scale were determined by generating two η mesons x_F distributions, artificially scaling the single-photon energies by +1% and -1%, respectively. The variation from the unscaled spectrum was used to estimate the systematic uncertainty. The upper and lower error bands were conservatively symmetrized by assigning the maximum of the two values to both sides for each bin of x_F. The distribution of the energy scale errors is depicted in the left panel of Figure 3.20 (red line). This is the first time this uncertainty has been calculated using this methodology. In the past, to quantify the systematic error on the energy scale, spectra were scaled by an amount equal to the uncertainty on the absolute energy scale ($\pm$2.7. %). This is no longer necessary, since for the first time, as discussed in Section 3.1, the simultaneous measurement of π^0 mesons and η mesons allowed for accurate calibration of the absolute energy scale, enabling a reduction in the associated statistical uncertainty.

3.5.2 Particle identification

The systematic uncertainty associated with the particle identification method used in this analysis (described in Section 3.3.3) was calculated for each bin of the x_F distribution by comparing spectra obtained with different

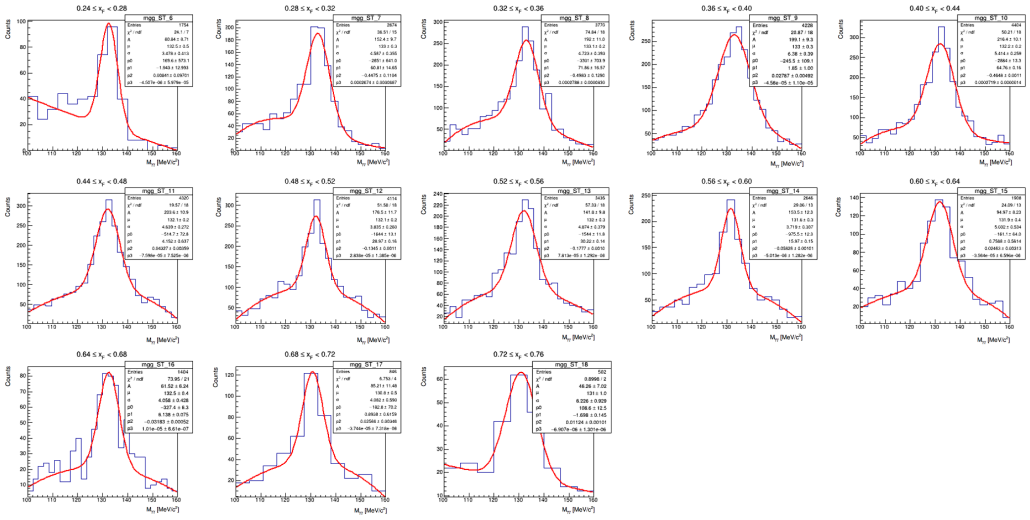

Figure 3.21: Invariant mass distributions for Type II π^0 events detected in the Small Tower across different x_F bins, as indicated in each subplot, in proton-proton collisions at $\sqrt{s} = 13$ TeV. The red curves represent the results of the fit using an asymmetric Gaussian function combined with a third-order Chebyshev polynomial. The fit parameters and corresponding statistical uncertainties are displayed within each subplot.

PID criteria. Two additional $L_{90\%}$ cut functions were computed with efficiency values of 85% and 95% instead of 90%, as described in Section 3.3.3. These limits were selected to ensure that the product of efficiency and purity remained above 75% across the entire energy range. The full analysis was then repeated using these different functions, and the PID error was estimated based on the relative deviation from the distribution obtained with $L_{90\%}$ the cut functions computed with efficiency values of 90%. The distribution of the particle identification errors can be visualized in the left panel of Figure 3.20 (yellow line).

3.5.3 Beam-center stability and luminosity

The upstream LHC reference system is defined with the z-axis aligned along the beam direction, with positive values extending from IP1 toward LHCf-Arm2. The y-axis is perpendicular to the ground, while the

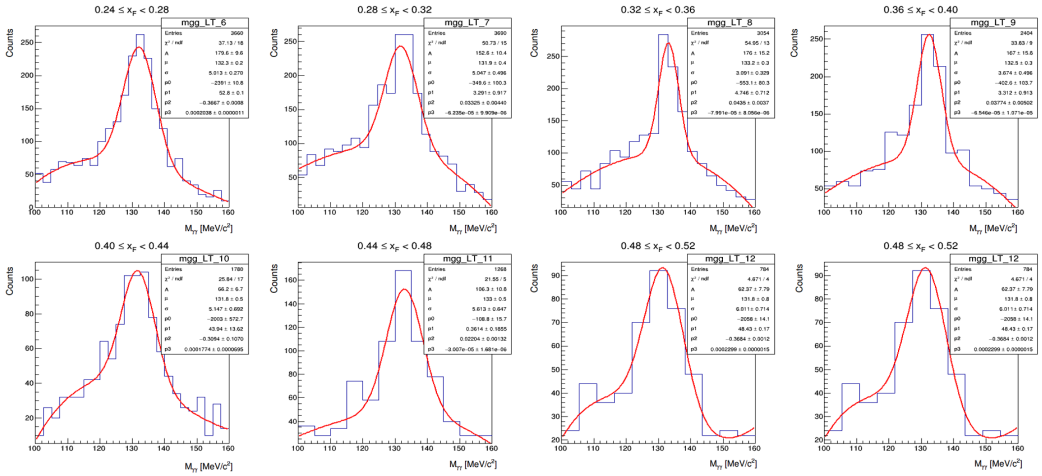

Figure 3.22: Invariant mass distributions for Type II π^0 events detected in the Large Tower across different x_F bins, as indicated in each subplot, in proton-proton collisions at $\sqrt{s} = 13$ TeV. The red curves represent the results of the fit using an asymmetric Gaussian function combined with a third-order Chebyshev polynomial. The fit parameters and corresponding statistical uncertainties are displayed within each subplot.

x-axis is perpendicular to the y-z plane, with positive values extending from LHCf-Arm2 toward the center of the LHC ring. It is fundamental to determine the coordinates representing the projection of the beam-center onto the LHCf detector to correctly define the analysis region. This can be done in two ways: the first method uses Beam Position Monitors (BPMSW) located at ± 21 m from IP1 while the second method takes advantage of the fact that the neutron position distribution is highly concentrated along the beam axis. The second method proved to be more accurate than the first. Therefore, in this work, BPMSW data were not considered. Instead, the beam-center projection on the detector was estimated by fitting the two-dimensional neutron position distribution. The bi-dimensional fit function used is:

$$f(x,y) = Ae^{-B\sqrt{(x-x_c)^2+(y-y_c)^2}}, \tag{3.12}$$

where A, B, x_c, and y_c are fit parameters. The (x_c, y_c) pair represents the coordinates of the beam-center projection. The optimal values for the

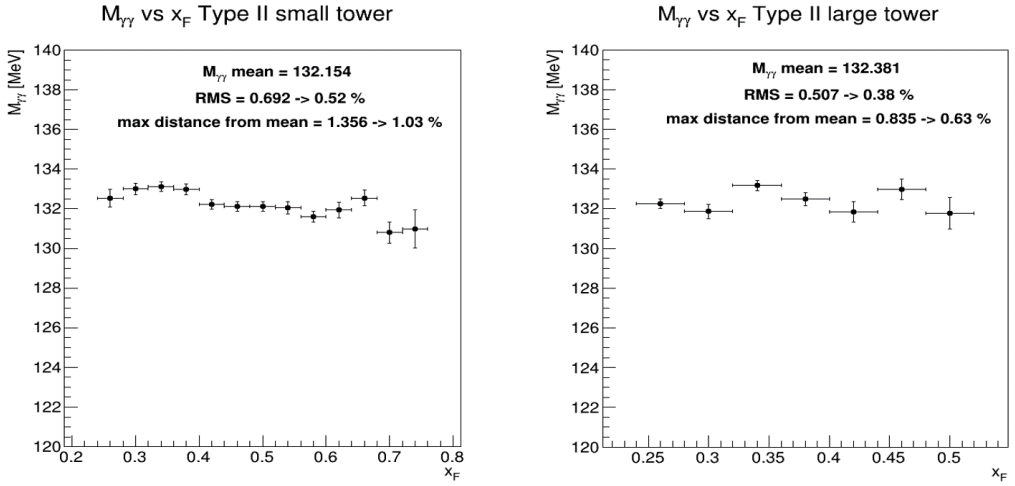

Figure 3.23: Invariant mass peak positions as a function of x_F for Type II π^0 events detected in the Small Tower (left) and Large Tower (right). The mean value, root mean square, and maximum deviation from the mean are indicated for each tower. The results show that the energy scale is stable within 1.03% for the Small Tower and 0.63% for the Large Tower across the measured x_F range, confirming the reliability of the energy scale correction applied to the data.

point (x_0, y_0) were obtained using hadrons with reconstructed energies above 1 TeV. The two-dimensional neutron position distribution, together with the fit function, is depicted in Figure 3.24. The variations of the beam center position during Fill 3855 were evaluated and are shown in Figure 3.25. It was observed that fluctuations of the order of 0.3 mm are present and are larger with respect to the uncertainty of the fit (red dashed lines in Figure 3.25). To account for this effect, a systematic error was estimated by shifting the beam center position by ± 0.3 mm in both the X and Y directions and repeating the analysis for the four obtained configurations. Then, the four resulting spectra were compared to the original spectrum and assigned the systematic uncertainties due to beam center stability based on the maximum relative deviations between them. The distribution of the beam center errors is shown in the left panel of Figure 3.20 (magenta line). As discussed in Section 3.2.1 the uncertainty

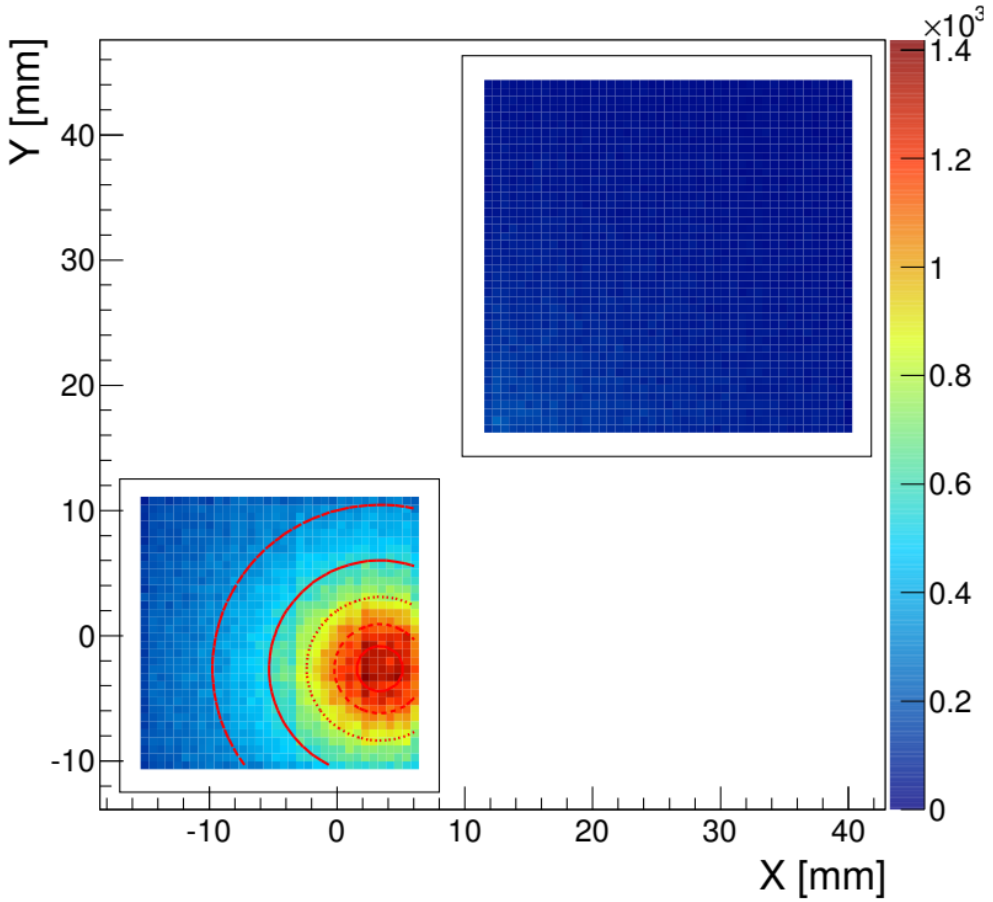

Figure 3.24: The two-dimensional neutron position distribution as observed in the LHCf-Arm2 detector in proton-proton collision at $\sqrt{s} = 13$ TeV, shown together with the fitted function (red contours). This projection was obtained using hadrons with reconstructed energies above 1 TeV.

on the integrated luminosity measured by ATLAS was estimated to be $\pm 1.9\%$. This value was derived using the methodology described in [104] for LHC Fill 3855. Notably, this is the only energy-independent systematic uncertainty. The distribution of the luminosity error can be visualized in the left panel of Figure 3.20 (teal line).

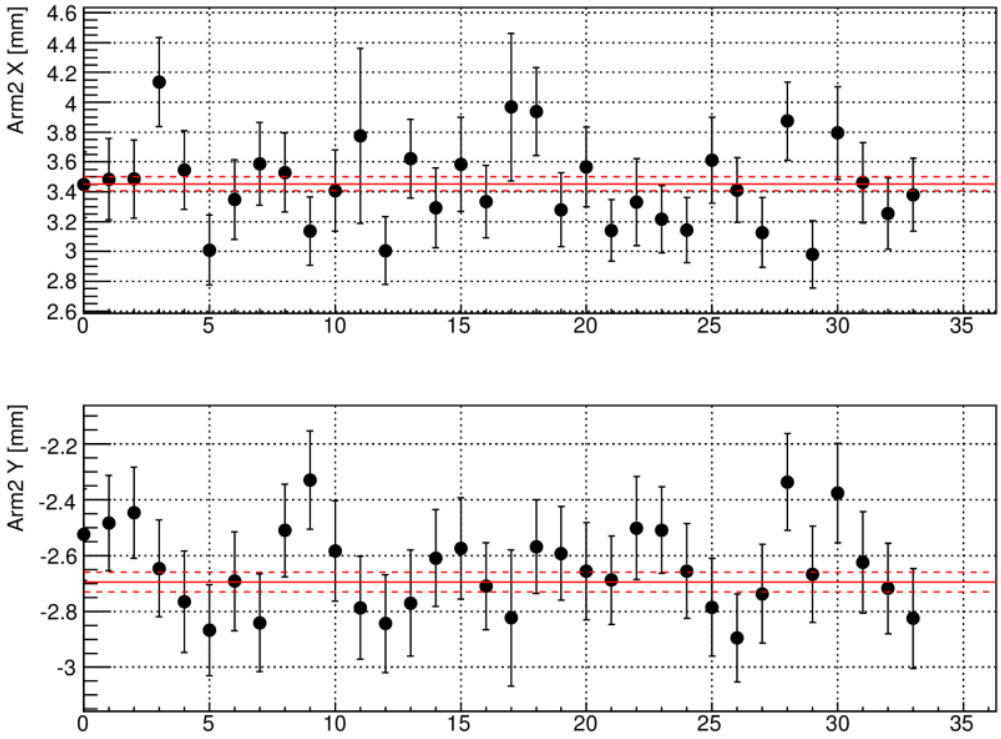

Figure 3.25: The fluctuations in the beam center position during Fill 3855, as measured by the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV. The acquired data were segmented into samples of approximately 125,000 triggered events, and the fit to determine the beam center was performed separately for each sample. The top panel shows the variation in the X-coordinate of the beam center, while the bottom panel shows the variation in the Y-coordinate. The red dashed lines represent the uncertainty from the bi-dimensional fit.

3.5.4 Background subtraction

The uncertainty associated with the background subtraction method explained in Section 3.3.4, is evaluated using the full MC detector simulation based on QGSJET II-04 described in section 3.2.2. The entire analysis procedure is applied to this MC dataset, up to the step where the x_F spec-

trum is extracted using the sideband method, as detailed in Section 3.3.4. Another x_F spectrum is generated by rejecting background events based on the MC truth information from the simulation. The relative variation between these two spectra is used to estimate the error associated with the background subtraction method, whose distribution is shown in the right panel of Figure 3.20 (blue line).

3.5.5 MC related correction

As described in Section 3.4, MC simulations were used to calculate several corrections, including selection, acceptance, and multi-hit corrections. A systematic error was calculated for each correction to mitigate model dependence on the final experimental distribution. The multi-hit and selection corrections were derived from full detector simulations based on QGSJET II-04 and EPOS-LHC, described in Sections 3.2.2. In these cases, errors were determined by the relative deviation between the correction values predicted by the two models, depicted in Figure 3.18 and Figure 3.19, for the multi-hit and selection corrections, respectively, and for the two models. The obtained error distributions are reported in the right panel of Figure 3.20 in green (multi-hit error) and purple (selection error) lines. Predictions from four different models (QGSJET II-04, EPOS-LHC, DPMJET 3.06, and SIBYLL 2.3) were used for the acceptance correction. The corresponding errors were calculated conservatively by considering the relative maximum and minimum differences between the model predictions (depicted in the bottom panel of Figure 3.17) and the mean correction value. The distribution of the acceptance correction error can be visualized in the right panel of Figure 3.20 (orange line).

3.6 Results

The x_F spectrum of η mesons measured by the LHCf-Arm2 detector in $p_T < 1.1$ GeV/c is presented in Figure 3.26. The inclusive production rate is given by the expression:

$$\frac{1}{\sigma_{inel}} x_F \frac{d\sigma}{dx_F}, \tag{3.13}$$

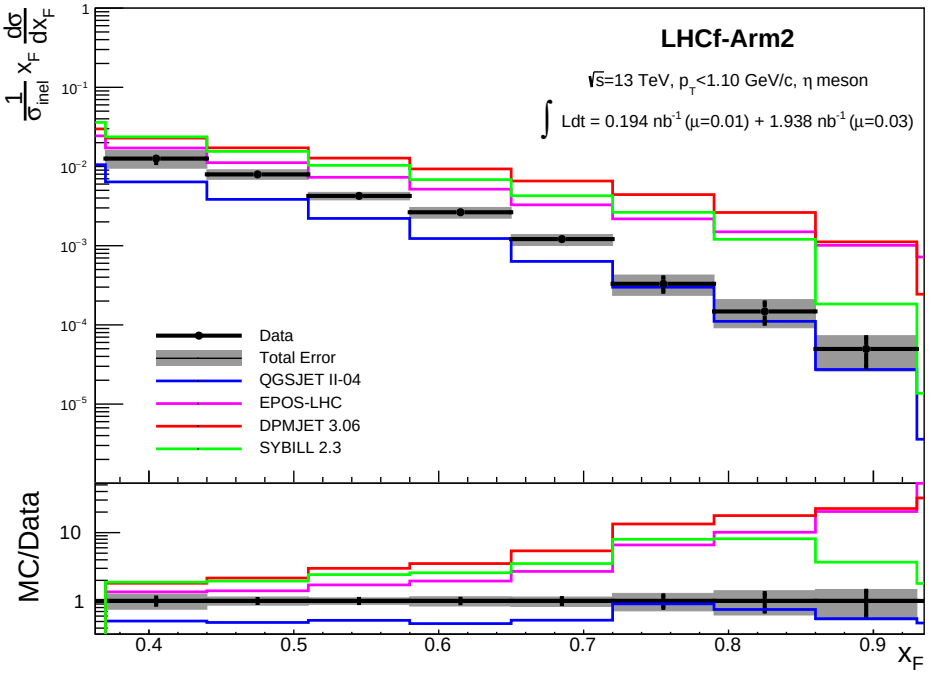

Figure 3.26: Inclusive η production rate as a function of x_F with $p_T < 1.1$ GeV/c for p-p collisions at $\sqrt{s} = 13$ TeV, measured using the LHCf-Arm2 detector. Black markers represent experimental data with statistical errors, while grey bands indicate the total uncertainties, calculated by summing statistical and systematic errors in quadrature. The data points are compared with the predictions at the generator level from the HIMs considered in this analysis: QGSJET II-04 (blue line), EPOS-LHC (magenta line), SIBYLL 2.3 (green line), and DPMJET 3.06 (red line). Figure from [101].

where σ_{inel} is the inelastic cross section for proton-proton collisions at $\sqrt{s} = 13$ TeV, measured by the TOTEM experiment as $\sigma_{inel} = (79.5 \pm 1.8)$ mb [110]. Even when considering measurements of the inelastic cross section by other LHC experiments [111, 112], the differences are minimal, rendering the resulting uncertainty on η production rate negligible. The quantity $x_F d\sigma/dx_F$ represents the differential cross section of η produc-

x_F range	$(x_F/\sigma_{inel})(d\sigma/dx_F)$
[0.37-0.44]	$(1.26^{+0.33}_{-0.31}) \times 10^{-2}$
[0.44-0.51]	$(0.79^{+0.11}_{-0.11}) \times 10^{-2}$
[0.51-0.58]	$(4.25^{+0.46}_{-0.46}) \times 10^{-3}$
[0.58-0.65]	$(2.64^{+0.40}_{-0.42}) \times 10^{-3}$
[0.65-0.72]	$(1.21^{+0.17}_{-0.21}) \times 10^{-3}$
[0.72-0.79]	$(3.30^{+0.96}_{-0.94}) \times 10^{-4}$
[0.79-0.86]	$(1.48^{+0.61}_{-0.56}) \times 10^{-4}$
[0.86-0.93]	$(0.50^{+0.23}_{-0.24}) \times 10^{-4}$

Table 3.5: The inclusive η production rate for each bin of x_F with $p_T <$ 1.1 GeV/c, measured using the LHCf-Arm2 detector in p-p collisions at $\sqrt{s} = 13$ TeV. Total uncertainties are also provided. Table from [101]

tion, with $d\sigma = dN_\eta / \int Ldt$, where dN_η is the number of η mesons and $\int Ldt$ is the integrated luminosity of the dataset. The black error bars in Figure 3.26 indicate the statistical uncertainties, while the grey shaded areas represent the total uncertainties, obtained by combining statistical and systematic errors in quadrature. The inclusive η production rate values for each x_F bin, along with the total uncertainties, are shown in Table 3.5. Figure 3.26 also displays the predictions of several HIMs at the generator level, QGSJET II-04, EPOS-LHC, DPMJET 3.06, SIBYLL 2.3, and their ratios to the experimental data. Among the models tested, QGSJET II-04 shows the best agreement with the LHCf data, especially for $x_F > 0.7$, although there is a discrepancy of about a factor of 2 at lower x_F. The other three models, EPOS-LHC, DPMJET 3.06, and SIBYLL 2.3, predict higher production rates and a harder spectrum compared to the experimental data across the entire x_F range. The ratios of the values for each model to the experimental data, for each x_F bin, are listed in Table 3.6.

3.7 Validation tests

To ensure the reliability of the corrections applied to the experimental data, discussed in Section 3.4, validation tests were conducted using the two simulations based on the QGSJET II-04 and EPOS-LHC models. In

x_F range	QGSJET II-04	EPOS LHC	SIBYLL 2.3	DPMJET 3.06
[0.37-0.44]	0.51	1.36	1.88	1.81
[0.44-0.51]	0.48	1.41	1.95	2.16
[0.51-0.58]	0.52	1.72	2.43	3.00
[0.58-0.65]	0.47	1.95	2.58	3.52
[0.65-0.72]	0.52	2.70	3.53	5.41
[0.72-0.79]	0.91	6.61	7.99	13.38
[0.79-0.86]	0.75	10.12	8.14	17.76
[0.86-0.93]	0.55	20.36	3.69	22.54

Table 3.6: Ratio of inclusive η production rates from HIMs to experimental data for p-p collisions at $\sqrt{s} = 13$ TeV for each x_F bin with $p_T < 1.1$ GeV/c. Table from [101].

these tests, the equivalence between the particle-level distributions and the corrected detector-level distributions was verified. The objective was to check whether the applied corrections accurately accounted for the effects of the detector response, ensuring that the reconstructed data correctly represents the underlying physical processes. The validation tests involved a direct comparison between the particle-level distributions, representing the true kinematic properties of the η mesons generated in the MC simulations and the corrected detector-level distributions, which were obtained by applying the same reconstruction algorithms and correction factors as used in the analysis of the experimental data. The comparison was carried out for the η mesons production rate distribution with transverse momentum $p_T < 1.1$ GeV/c. The results of these validation tests are depicted in Figures 3.27 and 3.28, corresponding to the QGSJET II-04 and EPOS-LHC models, respectively. In each figure, the black data points represent the corrected detector-level distributions, while the blue (for QGSJET) and magenta (for EPOS) lines denote the particle-level distributions. The error bars shown are statistical only. As shown in the two Figures, a good agreement is observed within the uncertainties, demonstrating that the corrections applied to the data effectively mitigate the detector effects, leading to a reliable reconstruction of the true particle-level distribution. While small discrepancies are visi-

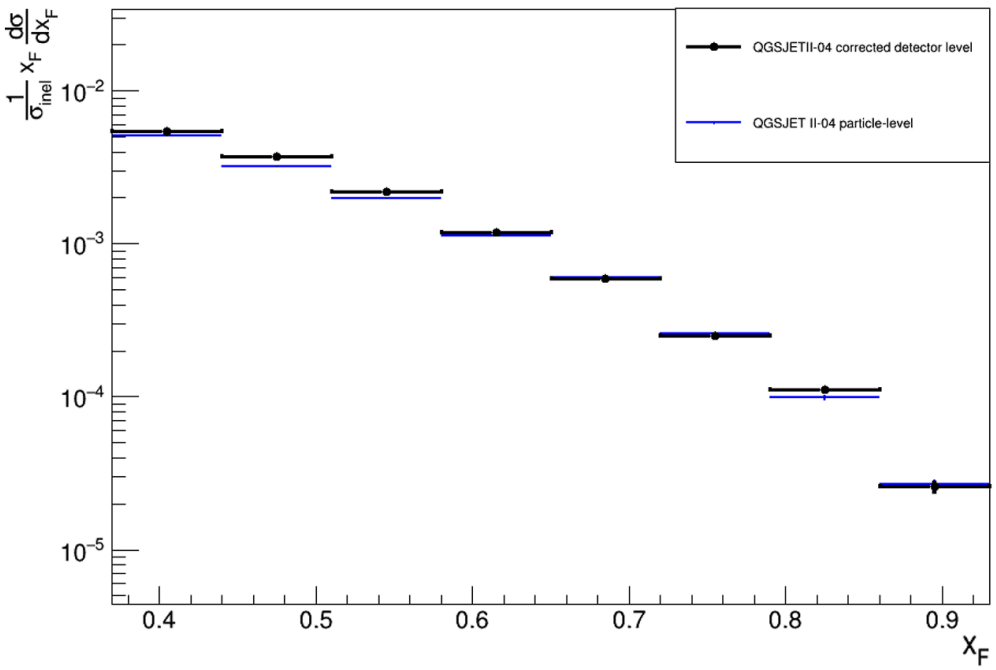

Figure 3.27: Comparison of the η meson production rate as a function of x_F between the particle-level (blue line) and corrected detector-level (black points) distributions for events with $p_T < 1.1$ GeV/c, as simulated by the QGSJET II-04 model. The particle-level distribution represents the true physical output of the simulation, unaffected by detector effects, while the corrected detector-level distribution has been reconstructed from simulated data, including corrections for detector response and smearing effects. The agreement between these two distributions within statistical uncertainties demonstrates the effectiveness of the applied corrections. The vertical error bars represent statistical uncertainties only.

ble in both figures, these are attributed to residual smearing effects that were not entirely corrected. However, these effects are minor and do not significantly impact the overall agreement between the particle-level and corrected detector-level distributions.

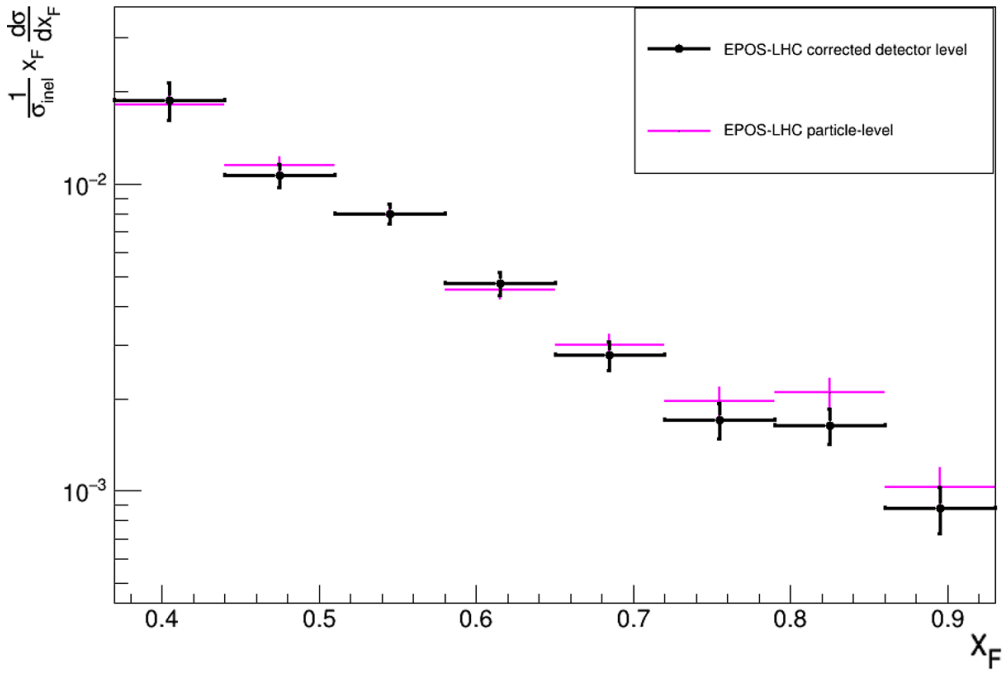

Figure 3.28: Comparison of the η meson production rate as a function of x_F between the particle-level (blue line) and corrected detector-level (black points) distributions for events with $p_T < 1.1$ GeV/c, as simulated by the EPOS-LHC model. The particle-level distribution represents the true physical output of the simulation, unaffected by detector effects, while the corrected detector-level distribution has been reconstructed from simulated data, including corrections for detector response and smearing effects. The agreement between these two distributions within statistical uncertainties demonstrates the effectiveness of the applied corrections. The vertical error bars represent statistical uncertainties only.

3.8 Conclusions

The LHCf experiment has successfully measured the inclusive production rate of η mesons in proton-proton collisions at $\sqrt{s} = 13$ TeV, focusing on events with transverse momentum $p_T < 1.1$ GeV/c. Approximately

1500 η meson candidates were identified in the dataset analyzed. This is both the first measurement of η mesons by the LHCf experiment and the first observation of a particle containing strange quarks in the very forward region of high-energy collisions. This measurement, motivated by the important role of η mesons in the electromagnetic component, strangeness production, and muon production in EAS, was introduced in Section 3.1. The analysis utilized data only from the LHCf-Arm2 detector, as described in Section 3.2. The data acquisition during LHC Fill 3855 was carefully selected to ensure stable operating conditions, which allowed for a precise analysis of the η meson production. In Section 3.3, the methodology for reconstructing and selecting η mesons was outlined. The reconstruction process involved hit position reconstruction, energy reconstruction, and particle identification, with a specific focus on events where the two photons from η meson decay enter different towers of the detector (Type I events). The background was subtracted using a sideband method, and the results were corrected for detector acceptance, selection inefficiency, and other experimental effects, as detailed in Section 3.4. Contribution to the total uncertainty of the measurements were discussed in Section 3.5. The results, presented in Section 3.6, were compared with predictions from several HIMs, including QGSJET II-04, EPOS-LHC, DPMJET 3.06, and SIBYLL 2.3. None of the models provided a perfect agreement to the experimental data across the entire x_F range, although QGSJET II-04 showed the best agreement, particularly at higher x_F values. The other models tended to overestimate the production rate, especially at higher x_F. These discrepancies highlight the need for further refinement in the modelling of forward η meson production. To ensure the reliability of the corrections applied to the data, validation tests were conducted, as discussed in Section 3.7. These tests compared the corrected detector-level distributions with the true particle-level distributions from the simulations, confirming that the corrections effectively accounted for detector effects and provided a reliable reconstruction of the true physical distribution. In conclusion, this analysis has provided the first measurement of forward η meson production by the LHCf experiment, offering valuable insights into the behaviour of hadronic interactions in the forward region. The results, while revealing some limitations of current models, also indicate the way for future stud-

ies with larger datasets. With the realisation of data-taking during LHC Run III, in proton-proton collisions at $\sqrt{s} = 13$ TeV, the LHCf experiment has significantly improved the statistics for η meson production. The upgraded LHCf-Arm2 detector, which benefited from enhancements in its silicon readout electronics, successfully recorded data in proton-proton collisions at $\sqrt{s} = 13.6$ TeV. This upgrade, which involved replacing the outdated 100 Mbit/s Fiber Optical Transmitter/Receiver Interface (FOXI) transmitters with a new configuration based on Gbit/s fast optical links, allowed the DAQ speed to increase from 0.5 to 1.5 kHz. This improvement led to an increase in statistics compared to the previous proton-proton operation during Run II of about eight times, as highlighted in [113]. The increased data statistics have already been analyzed, yielding preliminary distributions of the di-photon invariant mass measured by the LHCf-Arm2 detector during Run III, as shown in Figure 3.29. The π^0 and η peaks are clearly visible, indicating the improved statistics achieved with the upgraded detector and enhanced DAQ capabilities. The data collected in Run III will substantially reduce the experimental uncertainties that were previously encountered due to low statistics, allowing for more stringent tests of HIMs. These future measurements are expected to contribute significantly to our understanding of cosmic ray physics and high-energy particle interactions in the forward region. While the preliminary analysis of the data collected during LHC Run III shows a significant improvement in statistics, leading to clearer peaks for both π^0 and η mesons, it is crucial to perform a precise calibration of the LHCf-Arm2 detector before finalizing the analysis. This calibration is necessary to ensure that the energy measurements and other detector responses are accurately interpreted, which is essential for evaluating systematic uncertainties and validating the results. To achieve this, a dedicated beam test was conducted at the CERN Super Proton Synchrotron (SPS) shortly after the Run III data-taking period, in October 2022. This beam test, which employed electron and proton beams at various energies, was designed to calibrate the detector response by establishing accurate conversion factors between the collected charge signals and the energy deposits in the components of the detector (both calorimetric layers and position-sensitive layers). The calibration also involved evaluating the energy resolution of the calorimeters and ensuring the linearity of the detector response.

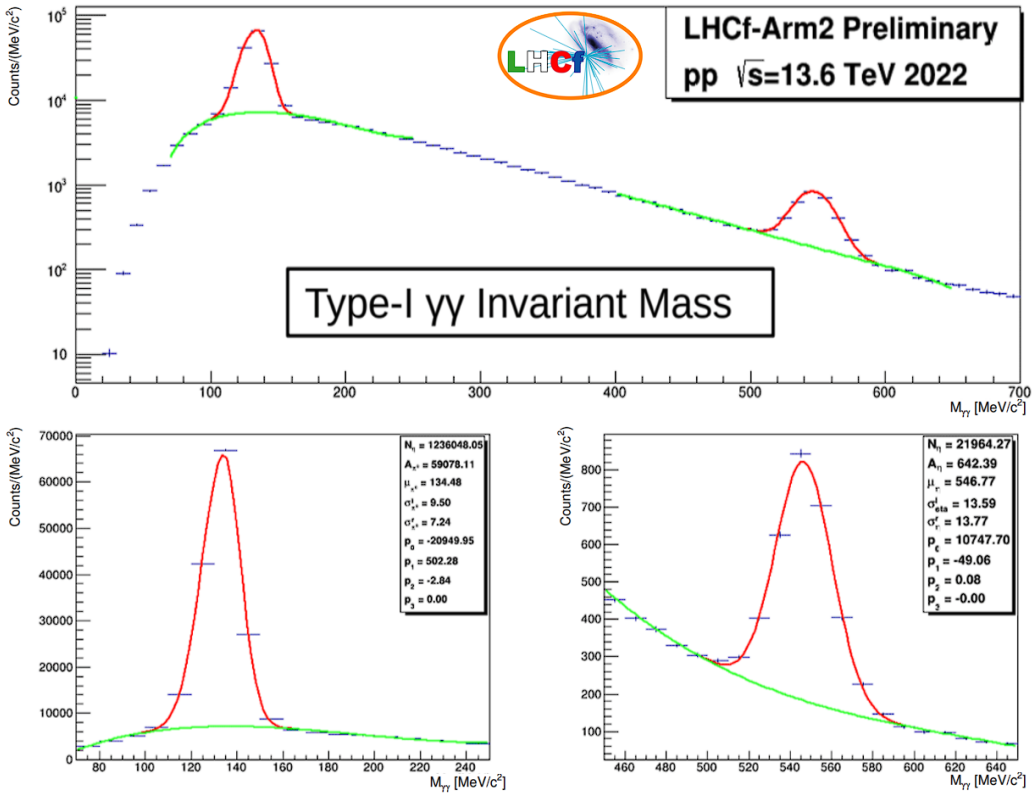

Figure 3.29: Invariant mass distributions of di-photon events measured by the LHCf-Arm2 detector during proton-proton collisions at $\sqrt{s} = 13.6$ TeV in LHC Run III. The top panel shows the overall distribution, with clear peaks corresponding to the π^0 and η mesons. The bottom left panel provides a zoomed view of the π^0 mass peak, while the bottom right panel focuses on the η mass peak. The red curves represent the signal component fits using an asymmetric Gaussian, and the green curves indicate the background fits using a third-order Chebyshev polynomial. The improved statistics, a result of the upgraded LHCf-Arm2 detector and enhanced DAQ system, allow for better resolution and reduced experimental uncertainties compared to previous runs. Figure from [113]

The detailed procedure and results of this calibration are presented in the following Chapter 4, where the experimental setup, data samples, and

calibration methodology are discussed.

Chapter 4

Calibration of the LHCf-Arm2 detector

During September 2022, the LHCf experiment carried out a data-taking operation in concomitance with the LHC RUN III, in proton-proton collisions at $\sqrt{s} = 13.6$ TeV. To calibrate the detector responses and evaluate the performances, in view of the analysis of collected data, a beam test was performed at the CERN Super Proton Synchrotron (SPS) right after, on October 12-19, 2022, using electrons and proton beams. The energies provided were 149.14, 197.32 and 243.61 GeV for the electron beam, 150 and 350 GeV for the proton beam. For this work, we present only the calibration of the LHCf-Arm2 detector GSO layers using electron data. The experimental setup is described in Section 4.1. Section 4.2 details the event sample for both data and MC simulations used in this analysis. Section 4.3 explains the procedure to determine the conversion factors for converting the collected charge signal into the energy deposit. Section 4.4 discusses the obtained energy resolution of the calorimeters and the linearity of the detector response. Finally, Section 4.5 provides an estimation of absolute energy scale systematic uncertainties contributions related to the energy dependence of calibration factors and the position dependence of the signal.

4.1 Experimental setup of the SPS H2 beam test

The beam test was carried out at the SPS-H2 beam line. The detectors were installed in an Aluminium box with temperature maintained by a chiller, ensuring a controlled temperature of 19 $^{\circ}C$, stable within variations of ± 1 $^{\circ}C$. The detectors and electronics were mounted on a movable table in order to scan the entire calorimeter surface along the plane perpendicular to the beam axis. A micro-strip silicon detector (ADAMO [114]) was positioned in front of the detector to measure the position of incident particles. Two plastic scintillators, 20 mm and 30 mm square, were placed between the detector and the beam pipe. The coincidence of these scintillator signals served as the data acquisition trigger. The beam test was also conducted for the first time with the detector rotated from its nominal position. This rotation allowed the deeper layers of the detector to be brought forward, enabling their calibration using electron beams rather than only proton beams. Typically, electromagnetic showers do not propagate through the entire detector, limiting calibration with electron beams to only the first 12 layers out of 16. The remaining layers were previously calibrated using protons, as hadronic showers can penetrate the entire detector. It is expected that utilizing electrons with the rotated detector to calibrate the deeper layers will produce more accurate results compared to using protons, which are more difficult to reconstruct inside the LHCf detectors and are affected by a non-negligible background contribution of pions. Figure 4.1 provides a schematic view of the beam test setup. The red arrow indicates the rotation of the LHCf detectors (360 degrees parallel to the movable table).

4.2 Data and MC samples

A sample of approximately 200000 triggered events at the center of each tower was recorded for each electron energy, for both front and back configurations. All events in each sample were acquired in consecutive runs. Additionally, about 2 million triggered events for each tower were collected at 197.32 GeV by scanning the entire area of the calorimeter acceptance to assess the position dependence of the signal, as described

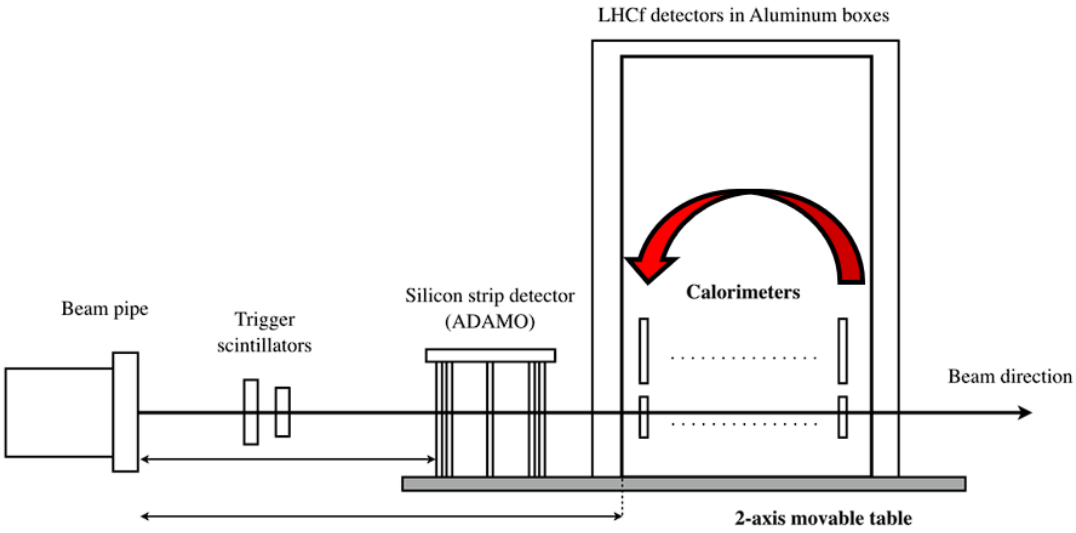

Figure 4.1: Schematic representation of the beam test setup at the SPS-H2 beam line. The detectors were installed inside Aluminium boxes. The entire setup was mounted on a 2-axis movable table to allow scanning of the calorimeter surface perpendicular to the beam axis. The setup includes a silicon strip detector (ADAMO) positioned before the calorimeters to measure the position of incident particles. Trigger scintillators were placed between the beam pipe and the detector to trigger data acquisition. The red arrow indicates the rotation of the LHCf detectors during the test, which was performed to allow the calibration of deeper detector layers using electron beams.

in Section 4.5.2. Since a comparison between data and simulations was required for the calibration (as explained in Section 4.3), a set of MC simulations was also generated with a methodology similar to that used for full LHC detector simulations, described in Section 3.2.2. Again, the Cosmos and EPICS packages were used for the simulation of mono-energy electron events, including transport and interaction with the detector. The configuration of the experimental setup, including the last 12 meters of the beam pipe, was accurately implemented in the simulation geometry. For each energy and configuration, 2.5 and 4 million events were simulated for Small and Large Towers, respectively, covering the whole surface of the two towers.

4.3 GSO scintillator layers calibration

To estimate the energy deposited in each layer from the measured charge (in ADC units), we compared the charge distributions for each GSO sampling layer with predictions from MC simulations using mono-energetic electron beams. This process involved two distinct configurations: one with the detector in the LHC nominal position (front configuration), to calibrate the first scintillators, and the other by rotating the detector (back configuration), as described in Section 4.1, to calibrate the deeper layers. Specifically, the first 12 scintillator layers were calibrated using the front configuration, while the deeper layers (13 to 16) were calibrated using the back configuration. This choice was made based on which of the two datasets (197.32 GeV electrons, front and back) exhibited a higher energy release in the corresponding layers. For data and simulation comparisons, each MC event energy deposit was scaled using a layer-specific conversion factor. The energy deposit E_{ij} (in GeV) of the i-th event in the j-th layer was multiplied by a conversion factor C_j (in ADC/GeV units). A term $\P_{ij}$ was added to account for pedestal fluctuations. This is described by the equation:

$$S_{ij} = C_j \cdot E_{ij} + \P_{ij}, \tag{4.1}$$

$\P_{ij}$ fluctuations were sampled from real pedestal events, which are randomly triggered without incident particles. An alternative strategy could be to generate independent Gaussian-distributed fluctuations for each

layer, but this method wouldn't account for layer-by-layer noise correlations. Energy deposit histograms were constructed with the scaled energy deposit S_{ij} for each layer. For the data, energy deposits remained unscaled. The agreement between data and simulation distributions was evaluated using a χ^2 test:

$$\chi^2_j(C_j) = \sum_k \frac{(Data^k_j - MC^k_j)^2}{(\sigma^k_{Data_j})^2 + (\sigma^k_{MC_j})^2}, \tag{4.2}$$

where $Data^k_j$ and MC^k_j are the k-th bin contents of the j-th layer for data and MC, respectively, and $\sigma^k_{Data_j}$ and $\sigma^k_{MC_j}$ are the statistical errors associated with the bin contents. To account for the difference between the position distribution of the MC events (generated with a uniform distribution) and the data, a correction was applied to the MC^k_j values obtained as the ratio of the 2D position distributions of the experimental and MC data, calculated in the region corresponding to the impact position of the MC event. Only bins with content above 30% of the distribution maximum were included in the calculation of the χ^2 to avoid contributions from tails influenced by hadron contamination. The entire process was repeated for various C_j values to find the configuration minimizing the χ^2. A second-order polynomial:

$$f(x) = p_0 \cdot x^2 + p_1 \cdot x + p_2, \tag{4.3}$$

was fitted around the minimum. The minimum $\chi^2_{j,\min}$ value determined the best conversion factor C_j. The error on C_j (ΔC_j) was the interval around the minimum where χ^2_j increased by less than 1 ($|\chi^2_j - \chi^2_{min,j}| < 1$), giving a 1-$\sigma$ confidence interval. Calibration events had to meet specific criteria:

1. The particle incident position was within a square of 5 mm × 5 mm for the Small Tower and 10 mm × 10 mm for the Large Tower around the calorimeter centers.

2. Over 90% of the total energy deposit was released within 20 radiation lengths to avoid hadron contamination.

3. Sum-dE within 3σ of the Sum-dE distribution mean. Initially, this distribution was obtained for events meeting the previous conditions, and then its mean and σ were measured.

Since the last two conditions required an estimation of C_j for all layers, the procedure for calculating the conversion factors was iterated multiple times. The results of the χ^2 minimization for each GSO layer in the front (layers 1 to 12) and back (layer 9 to 16) configurations of the 197.32 GeV datasets, together with the second-order polynomial fits and their results, are depicted in Figures 4.2 (Small Tower front), 4.3 (Large Tower front), 4.4 (Small Tower back), 4.5 (Large Tower back). The energy deposit distributions for each layer, comparing the experimental data and the scaled MC simulations, are shown in Figures 4.6 (Small Tower front), 4.7 (Large Tower front), 4.8 (Small Tower back), 4.9 (Large Tower back). The figures demonstrate that the MC simulations, once scaled by the appropriate conversion factors, closely match the experimental data across all layers. This consistency validates the calibration procedure, confirming that the derived conversion factors accurately translate the measured ADC values into deposited energy (in GeV). The histograms further reveal that the shape of the energy deposit distributions is well-reproduced by the simulations, with the small discrepancies mainly occurring in the tails of the distributions, likely due to residual hadron contamination or statistical fluctuations in the data. The final calibration results for each GSO layer are summarized in Table 4.1. The table presents the optimal conversion factors C_j (in ADC/GeV units) determined from the χ^2 minimization procedure, along with their associated uncertainties ΔC_j and the reduced χ^2 values. Layers 1 to 12 were calibrated using the front configuration, while layers 13 to 16 were calibrated using the back configuration. The reduced χ^2 values indicate a good agreement between the data and MC simulations, confirming the accuracy of the calibration procedure.

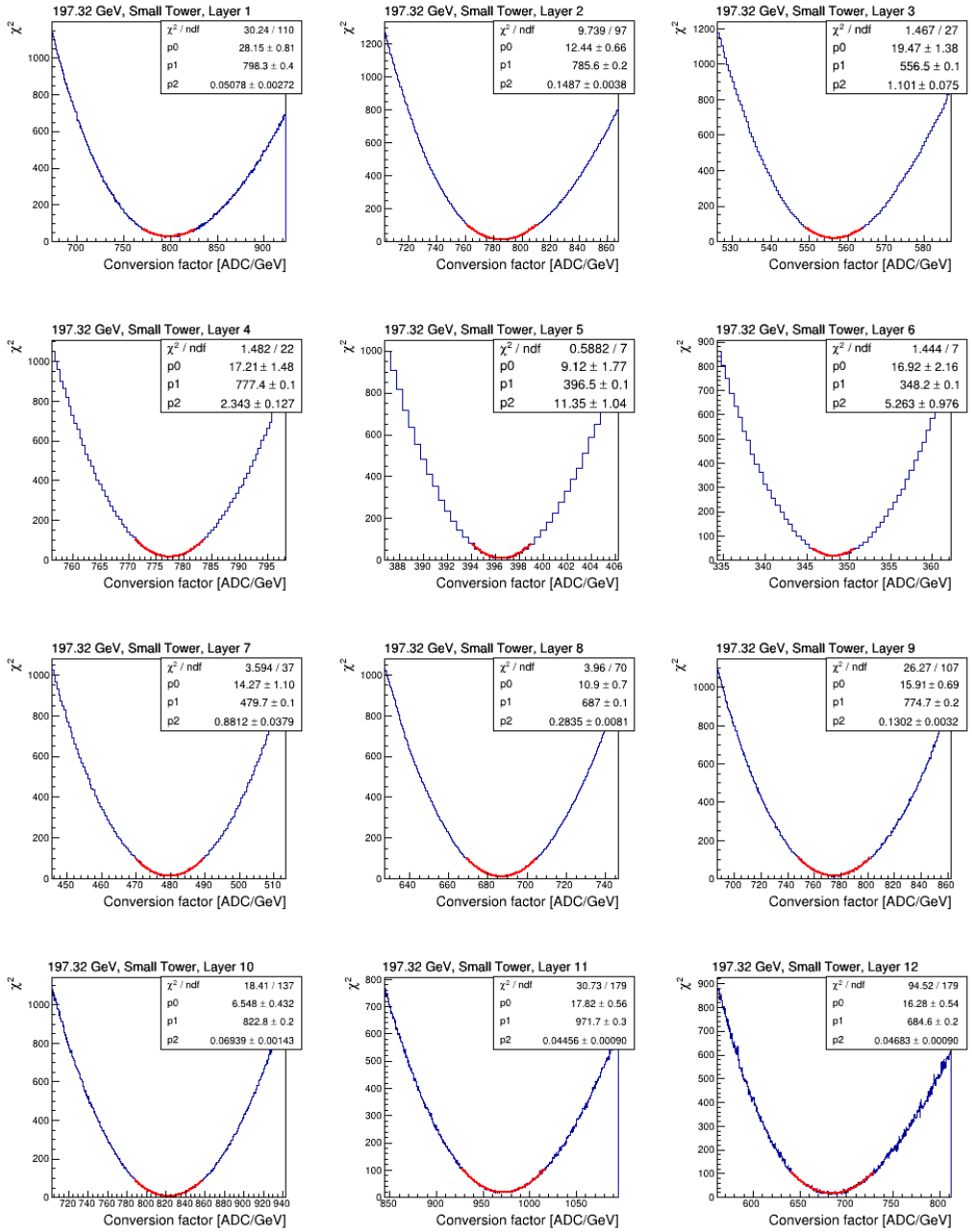

Figure 4.2: Results of the χ^2 minimization for each layer in the front configuration of the 197.32 GeV dataset for the Small Tower. The χ^2 values are shown as a function of the conversion factor (ADC/GeV) for each layer, with a second-order polynomial fit, represented by the red line, used to determine the optimal conversion factor. The fit parameters are provided in each subplot.

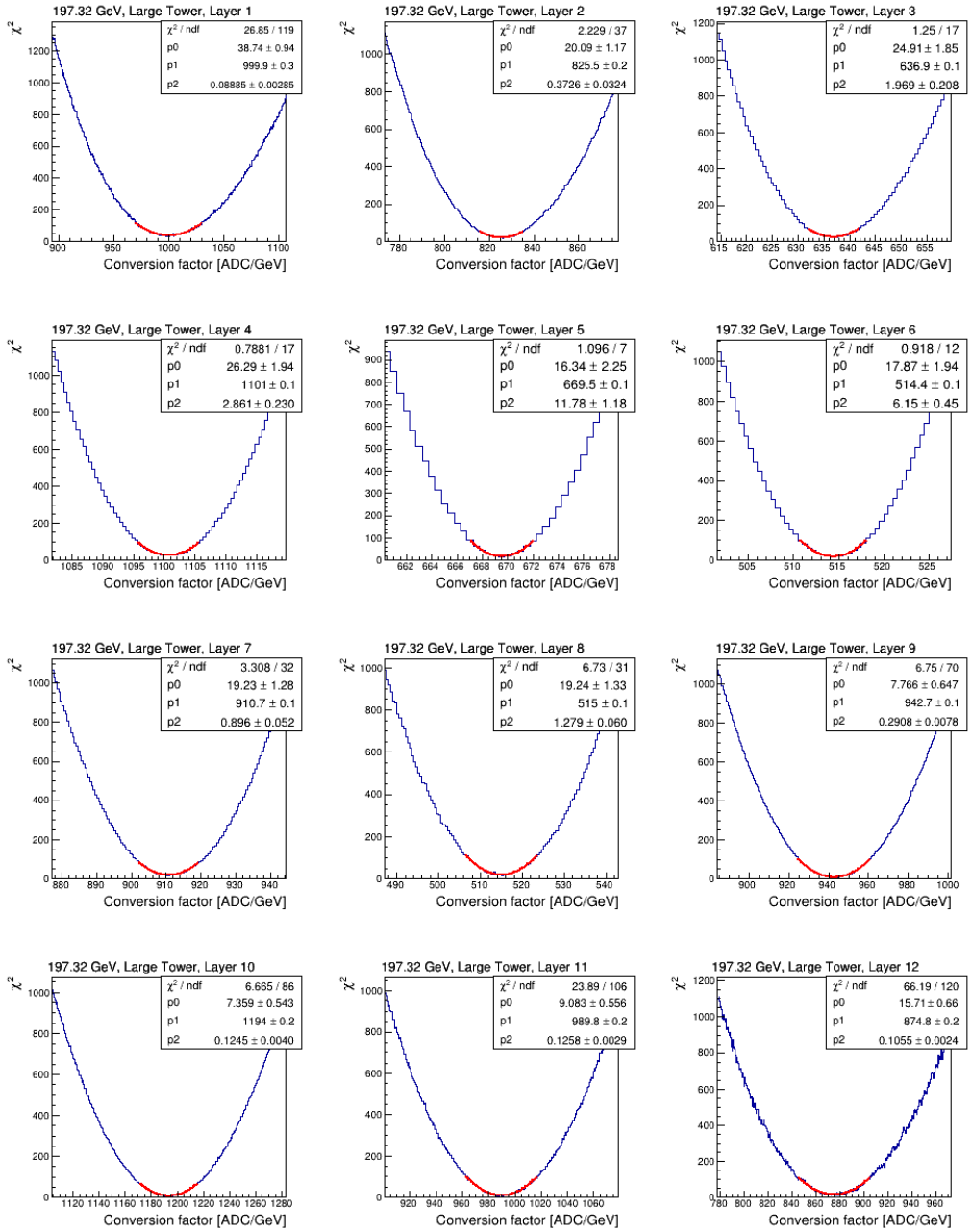

Figure 4.3: Results of the χ^2 minimization for each layer in the front configuration of the 197.32 GeV dataset for the Large Tower. The χ^2 values are plotted against the conversion factor (ADC/GeV) for each layer, with the optimal conversion factor determined by a second-order polynomial fit, shown as the red line. The fit parameters are displayed in each subplot.

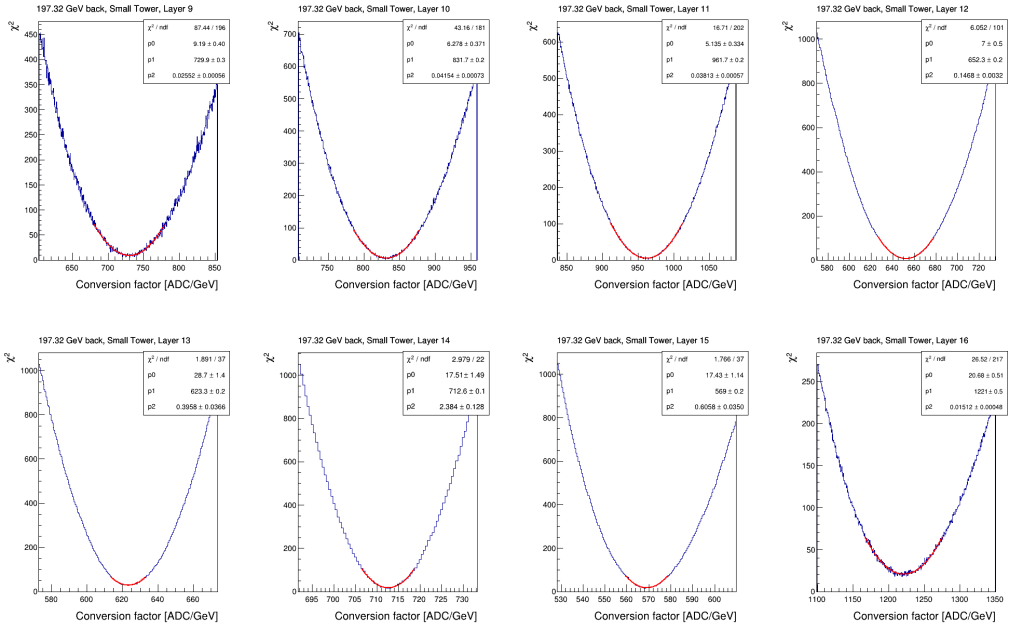

Figure 4.4: Results of the χ^2 minimization for each layer in the back configuration of the 197.32 GeV dataset for the Small Tower. The figure displays the χ^2 values as a function of the conversion factor (ADC/GeV) for each layer. The red line represents the second-order polynomial fit used to determine the optimal conversion factor, with the fit parameters shown in each subplot.

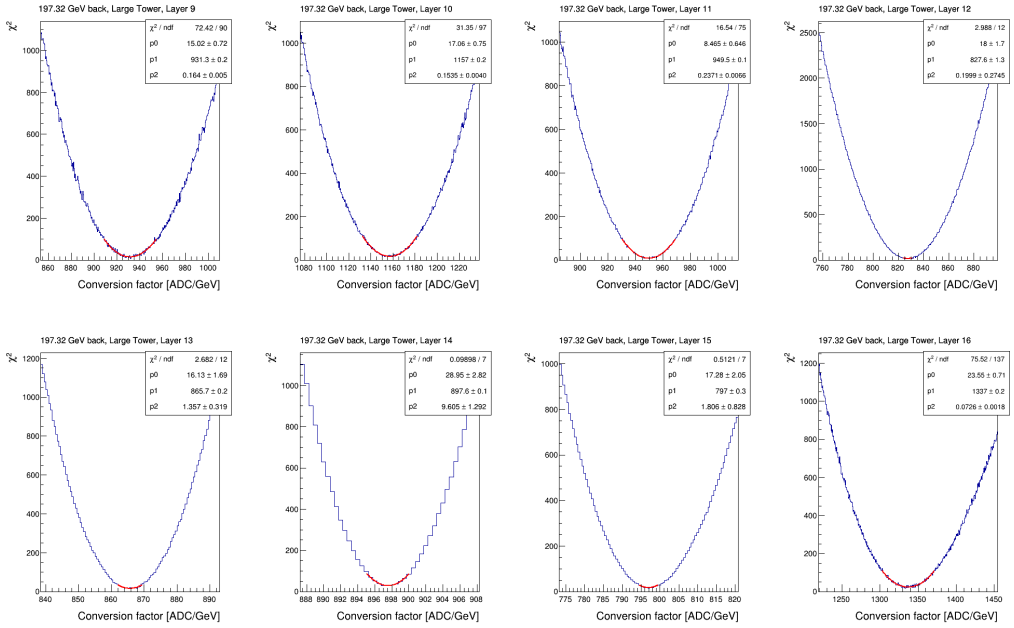

Figure 4.5: Results of the χ^2 minimization for each layer in the back configuration of the 197.32 GeV dataset for the Large Tower. The χ^2 values are plotted against the conversion factor (ADC/GeV) for each layer, with the second-order polynomial fit, represented by the red line, used to determine the optimal conversion factor. The fit parameters are provided in each subplot.

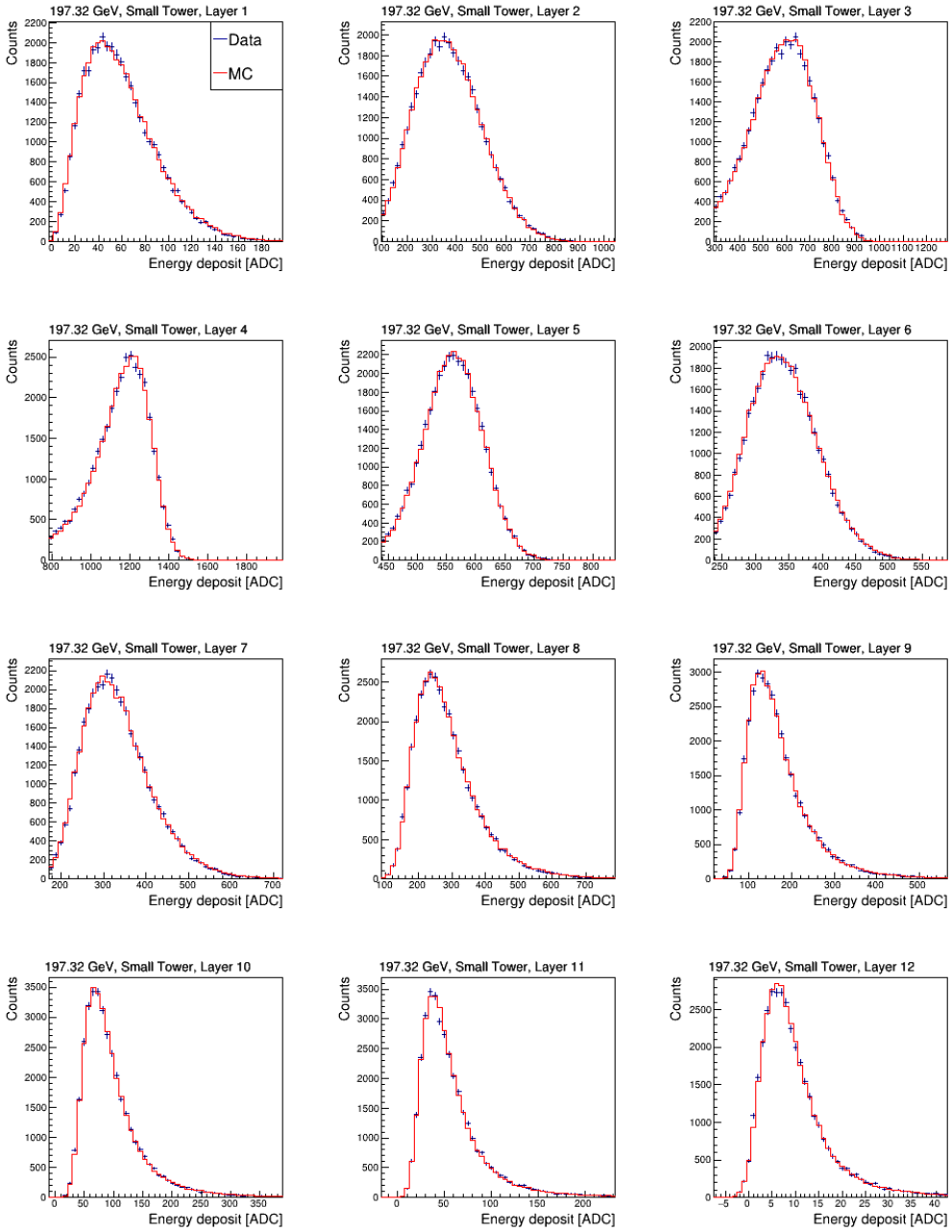

Figure 4.6: Energy deposit distributions for the Large Tower layers 1 to 12 in the front configuration for 197.32 GeV electrons. The experimental data (blue) are compared to the scaled MC simulations (red). The distributions show good agreement, validating the calibration procedure.

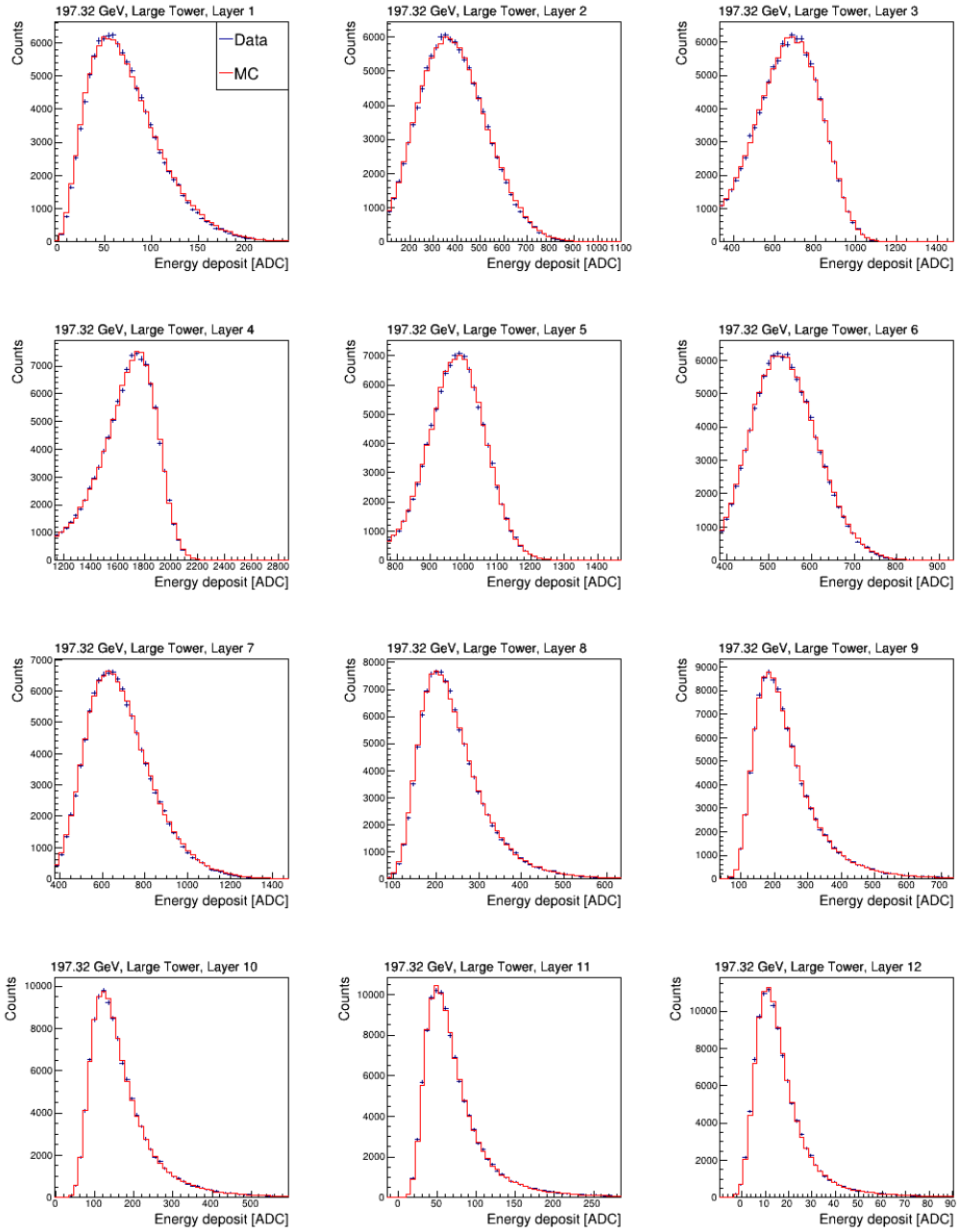

Figure 4.7: Energy deposit distributions for the Large Tower layers 1 to 12 in the front configuration for 197.32 GeV electrons. The experimental data (blue) are compared to the scaled MC simulations (red). The distributions show goof agreement, validating the calibration procedure.

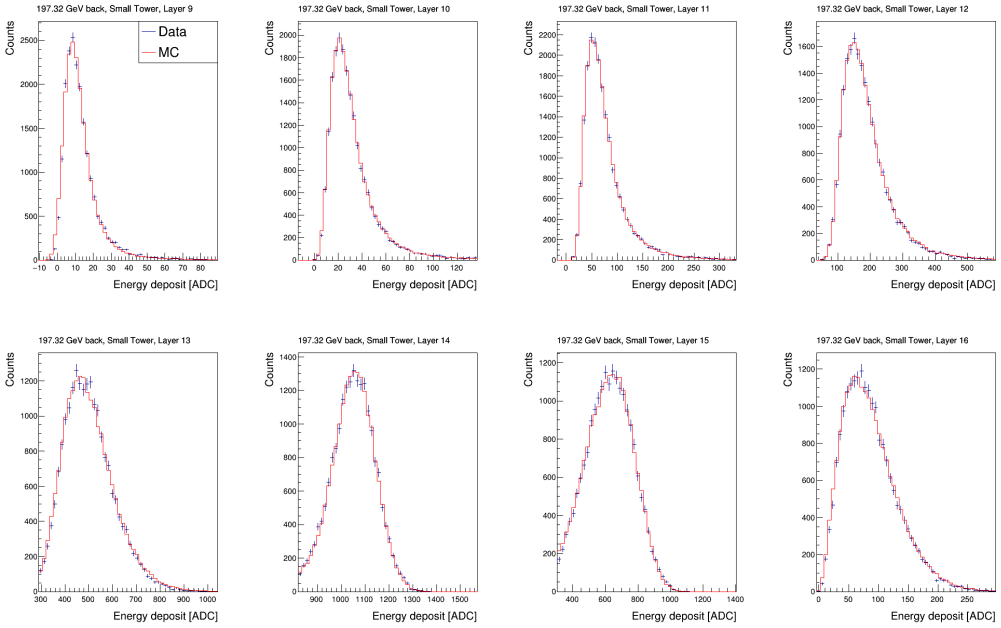

Figure 4.8: Energy deposit distributions for the Small Tower layers 9 to 16 in the back configuration for 197.32 GeV electrons. The experimental data (blue) and the scaled MC simulations (red) are compared, demonstrating that the simulation accurately models the energy deposit across the layers.

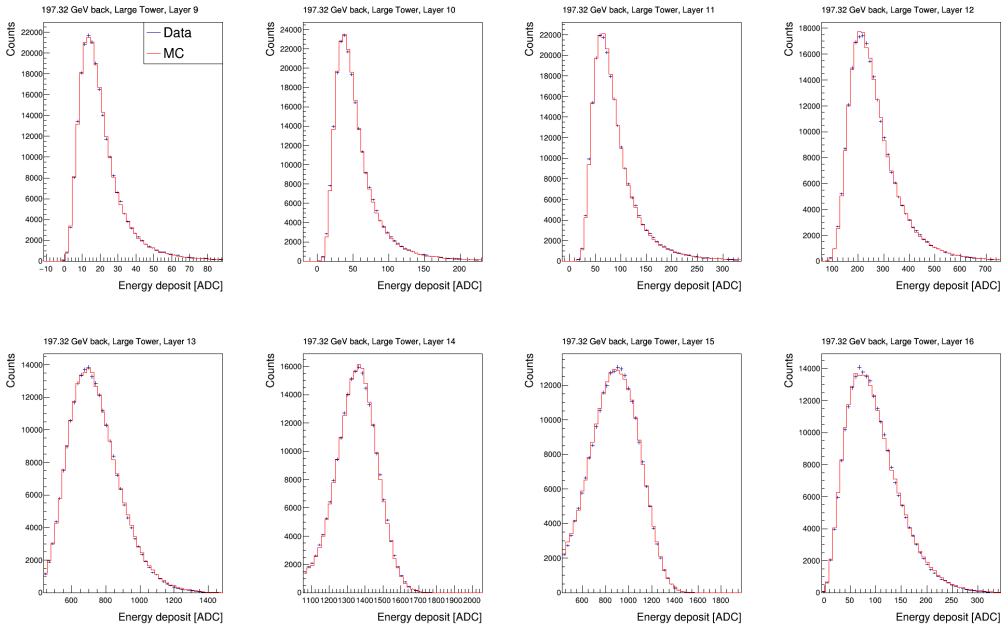

Figure 4.9: Energy deposit distributions for the Large Tower layers 9 to 16 in the back configuration for 197.32 GeV electrons. The experimental data (blue) are compared to the scaled MC simulations (red). The agreement between the distributions supports the reliability of the conversion factors used for the deeper layers.

Layer	Small Tower			Large Tower		
	C_j	ΔC_j	χ^2/ndf	C_j	ΔC_j	χ^2/ndf
1	799.28	±4.44	1.37	1000.88	±3.35	1.87
2	786.62	±2.59	0.57	826.48	±1.64	0.92
3	557.54	±0.95	0.90	637.87	±0.71	1.26
4	778.37	±0.65	1.31	1101.72	±0.59	1.44
5	397.50	±0.30	0.56	670.53	±0.29	1.19
6	349.19	±0.44	0.68	515.40	±0.40	0.71
7	480.68	±1.07	0.83	911.71	±1.06	0.95
8	688.02	±1.88	0.67	516.01	±0.88	1.06
9	775.74	±2.77	0.99	943.71	±1.85	0.60
10	823.80	±3.80	0.43	1194.57	±2.83	0.67
11	972.67	±4.74	2.11	990.81	±2.82	0.86
12	685.59	±4.62	0.97	875.81	±3.08	1.37
13	624.30	±1.59	1.77	866.70	±0.86	0.76
14	713.59	±0.65	0.84	898.55	±0.32	1.60
15	570.00	±1.28	0.86	798.00	±0.74	0.90
16	1221.90	±8.13	0.99	1337.91	±3.71	1.40

Table 4.1: Conversion factors C_j (in ADC/GeV units) for each GSO layer, derived from the χ^2 minimization process. The table also includes the corresponding uncertainties ΔC_j and the reduced χ^2 values. Layers 1 to 12 were calibrated using the front configuration, while layers 13 to 16 were calibrated using the back configuration.

4.4 Energy resolution and linearity for electromagnetic showers

The energy resolution and linearity of the calorimeter response for electromagnetic showers were evaluated for both the Small and Large Towers using the results of the three different electron beam energies recorded during the beam test, specifically 149.14, 197.32 and 243.61 GeV. The total energy deposit, or Sum-dE, was measured by integrating the energy deposited by electrons in the scintillator layers, specifically from the 2nd to the 13th layer, by using the conversion factor obtained with the 197.32

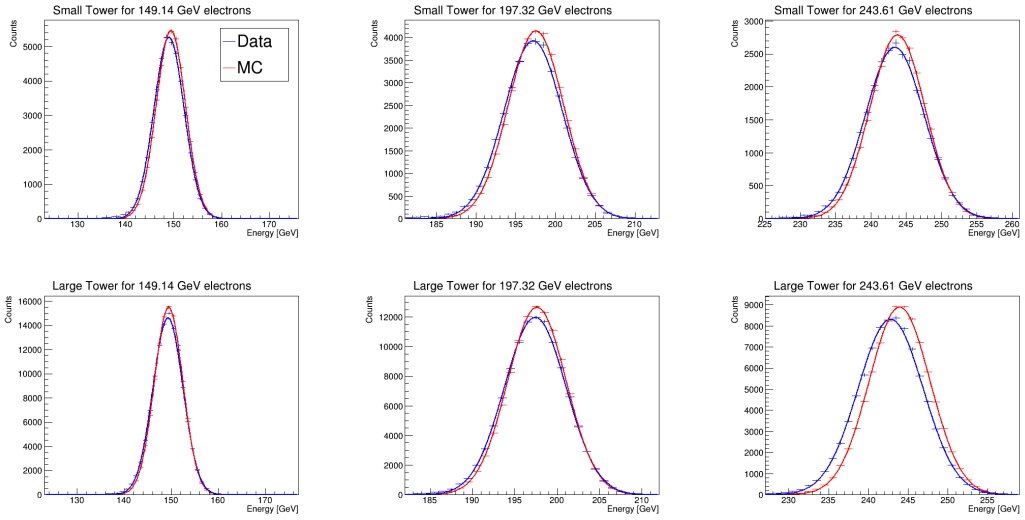

Figure 4.10: Distribution of the reconstructed energy for three different electron beam energies (197.32, 149.14 and 243.61 GeV) in both the Small Tower (top row) and Large Tower (bottom row). The comparison between experimental data (blue) and MC simulations (red) shows the reconstructed energy for each tested energy, illustrating the agreement between the data and the simulations across all energies and tower configurations.

GeV electron dataset, described in Section 4.3 and reported in Table 4.1. The same position and PID selection criteria described in Section 4.3 were used to calculate the total energy deposit. Then, we converted the sum-dE into reconstructed energy using the method described in Section 3.3.2. Figure 4.10 shows the distribution of the reconstructed energy for all tested energies in both the Small and Large Towers, comparing the results from data and MC simulations. The energy resolution was determined by fitting a Gaussian function to the reconstructed energy distributions, as shown in Figure 4.10. The resolution, defined as the ratio of the standard deviation σ to the mean of the energy distribution, was evaluated across the various incident energies. The energy dependence of the resolution

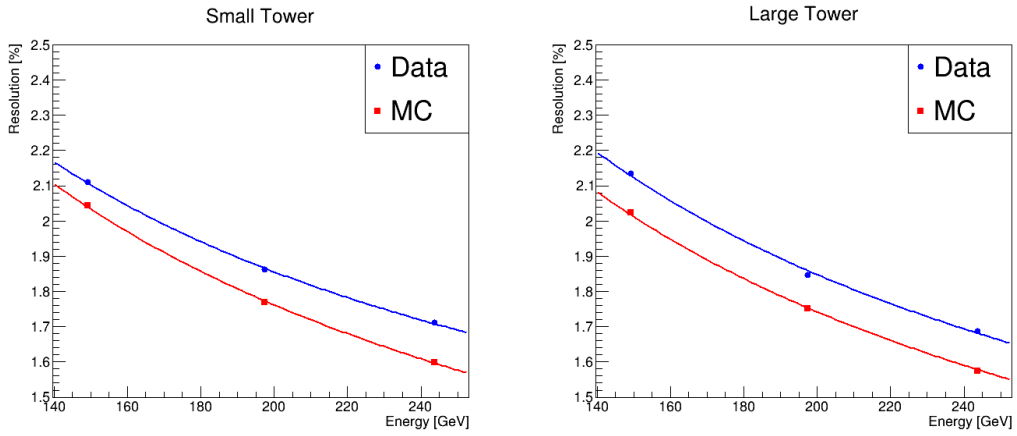

Figure 4.11: Energy resolution as a function of electron beam energy for the Small Tower (left) and Large Tower (right). The resolution, defined as the ratio of the standard deviation to the mean of the reconstructed energy distribution, is compared between experimental data (blue) and MC simulations (red). The data points are fitted using the parametrization given in Equation 4.4, which describes the energy dependence of the resolution.

was parametrized using the equation:

$$\frac{\sigma}{E} = \frac{b_0}{\sqrt{E/(100\ \text{GeV})}} \oplus b_1, \tag{4.4}$$

where σ/E is the resolution, E is the incident energy, and b_0 and b_1 are free parameters, with "$\oplus$" indicating a quadratic sum. The resolution as a function of the electron beam energy for both Small and Large Towers, as measured in the data and predicted by MC simulations, along with the fit results using Equation 4.4, are presented in Figure 4.11, while the results of fits are presented in Table 4.2. The data and MC simulations were compared, and while the fit function accurately describes the simulation results, discrepancies were observed in the data, particularly in the constant term b_1, which suggests additional contributions to the resolution in the experimental data not accounted for in the simulations. For both the Small and Large Towers, the parameter b_0 was consistent across data and simulations, indicating that the stochastic term of the resolution was

Tower	b_0 (%)	b_1 (%)
Small Tower Data	2.42 ± 0.02	0.72 ± 0.04
Small Tower MC	2.49 ± 0.01	0.09 ± 0.20
Large Tower Data	2.56 ± 0.06	0.38 ± 0.20
Large Tower MC	2.46 ± 0.006	0.005 ± 0.20

Table 4.2: Fit parameters for the energy resolution as a function of electron beam energy using the parametrization from Equation 4.4.

well-modelled. The linearity of the calorimeter response was assessed by plotting the relation between the reconstructed energy and the incident electron beam energy as shown in the top panels of Figure 4.12 (left panel for Small Tower and right panel for Large Tower). The residuals with respect to the incident electron beam energies were computed (bottom panel of Figure 4.12), demonstrating the good linearity of the two calorimeters, with deviations from a linear response being less than 0.4% for both the Small and Large Towers throughout the examined energy range.

4.5 Contributions to absolute energy scale error

Using SPS data, it is possible to estimate two contributions to the systematic uncertainty of the absolute energy scale. The uncertainty related to the energy dependence of conversion factors will be discussed in Section 4.5.1, while the error associated with the position dependence of the signal will be examined in Section 4.5.2.

4.5.1 Energy dependence systematic error

The conversion factors determined using different electron beam energies (197.32, 149.14, 243.61 GeV) and configurations (front and back for 197.32 GeV electrons) were compared to estimate the systematic uncertainties associated with the measurement of the factors C_j. Figure 4.13 shows the conversion factors for layers from the 1st to the 13th in the Small Tower (left panel) and Large Tower (right panel). The factors are

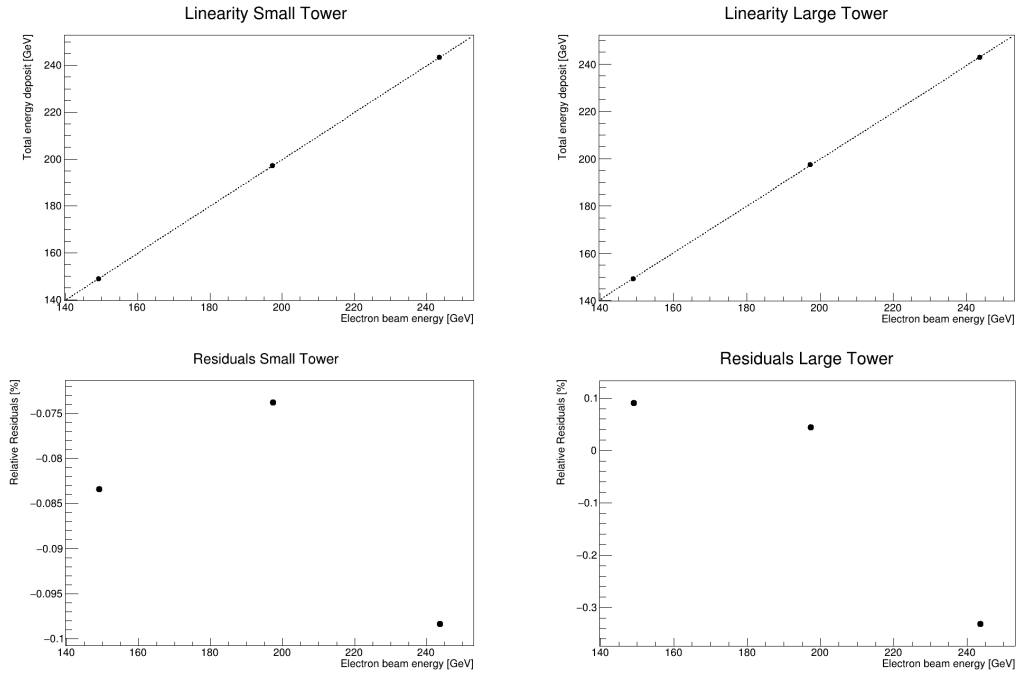

Figure 4.12: Linearity of the calorimeter response for the Small Tower (left) and Large Tower (right). The top panels show the relationship between the reconstructed energy and the incident electron beam energy. The bottom panels display the residuals, representing the deviations from the incident electron beam energy. The results indicate good linearity for both towers, with deviations from a linear response being less than 0.4% across the entire energy range.

normalized to those measured with the 197.32 GeV electron beam. The systematic errors associated with the conversion factors ware estimated both for electromagnetic and hadronic showers:

- For electromagnetic showers, the uncertainty was determined by the RMS of the distribution of the differences between the results at 149.14 GeV and 243.61 GeV, compared to those at 197.32 GeV, for layers 1-11. This uncertainty was estimated to be at the level of 0.7%.

- For hadronic showers, the uncertainty was determined by the RMS of the distribution of the maximum differences between the results

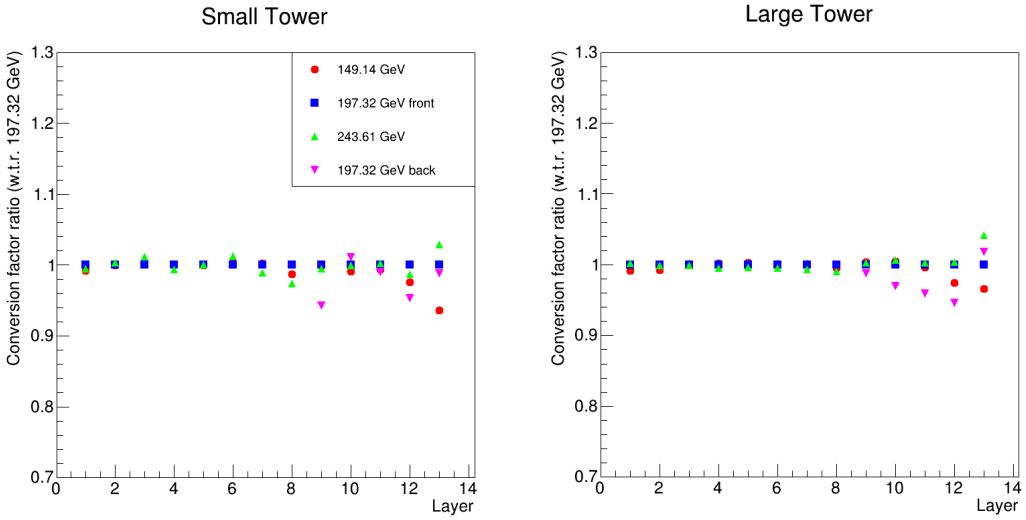

Figure 4.13: Comparison of conversion factors determined using different electron beam energies (149.14, 197.32, 243.61 GeV) and detector configurations (front and back for 197.32 GeV electrons) in the Small Tower (left panel) and Large Tower (right panel). The conversion factors are shown for layers 1 through 13 and are normalized to those measured with the 197.32 GeV electron beam.

at 149.14 GeV and 243.61 GeV, compared to those at 197.32 GeV, for layers 1-11. Additionally, the difference between the 197.32 GeV front and back configurations for layers 9-13 was considered. This uncertainty was estimated to be at the level of 1.9%.

4.5.2 Position dependence systematic error

Using the data sample from the scan of the entire surface of the towers with 197.32 GeV electrons, the systematic error due to the position dependence of the detector was estimated. The transverse surface of each tower was divided into a grid of 1 mm bins, and the mean reconstructed energy was computed for each position bin, excluding 2 mm from the border to avoid inaccuracies caused by lateral shower leakage. The maps of the position dependence of the reconstructed energy are shown in the top panels of Figure 4.14 for the Small Tower (left panel) and Large Towers

(right panel). In the case of Large Tower, an additional cut was applied to the position of events within 10 mm on the X-axis from the left border of the calorimeter. This is because the ADAMO detector covered this region during the beam test, which caused the electrons to interact before reaching LHCf-ARM2, distorting the uncertainty calculation. By the RMS of the residuals with respect to the mean reconstructed energy in the 2x2 mm^2 area in the center of the towers, we evaluated the systematic uncertainty connected to position dependence, which was found to be 0.64% for the Small Tower and 0.83% for the Large Tower, as shown in the bottom panels of Figure 4.14.

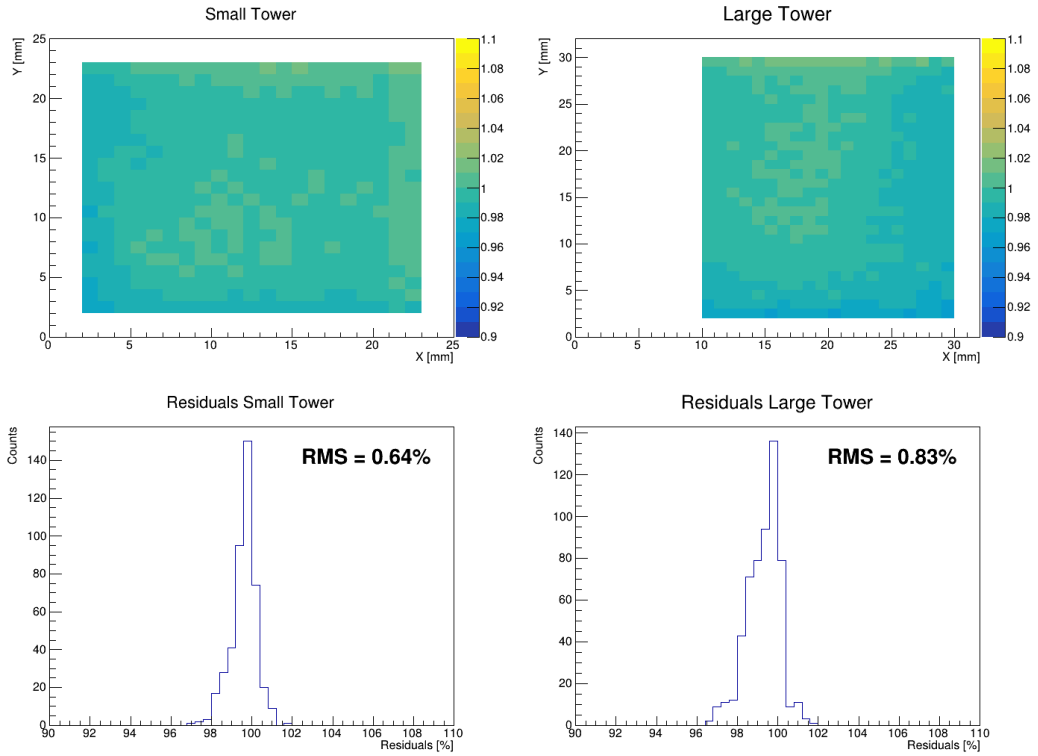

Figure 4.14: The top panels show the position dependence of the reconstructed energy across the Small Tower (left) and Large Tower (right) surfaces using a 197.32 GeV electron beam. The towers were divided into 1 mm bins, and the mean energy was calculated for each bin, excluding a 2 mm border to avoid edge effects. For the Large Tower, an additional cut was applied to events within 10 mm of the left border due to interference from the ADAMO detector during the beam test. The bottom panels display the histograms of residuals relative to the mean energy in a 2x2 mm^2 central area of each tower. The RMS of these residuals, representing the systematic uncertainty due to position dependence, was 0.64% for the Small Tower and 0.83% for the Large Tower.

Chapter 5

Conclusions and future prospects

This thesis provided a detailed overview of the LHCf experiment contributions to the study of ultra-high energy cosmic rays and the refinement of HIMs through very forward neutral particle measurements in proton-proton collisions at the LHC. As discussed in Chapter 1, UHECRs are central to our understanding of high-energy astrophysical phenomena, yet their interpretation relies on the accurate modelling of extensive air showers. HIMs are essential for simulating the complex particle cascades in extensive air showers, which in turn allow to determine crucial ultra-high energy cosmic ray properties, such as their energy spectrum and mass composition. However, these models require calibration and tuning against experimental data. In this context, the LHCf experiment, described in detail in Chapter 2, plays a fundamental role. Positioned to detect forward neutral particles produced in proton-proton and proton-ion collisions at the LHC, LHCf provides unique data to test and improve the performance of HIMs in the forward pseudorapidity region. This is particularly important for accurately modelling the hadronic interactions that dominate UHECR air showers, allowing for more precise simulations of cosmic ray events and helping to reduce uncertainties in their astrophysical interpretation. In this work, the production rate of the η meson was measured for the first time in the very forward region of proton-proton collisions at $\sqrt{s} = 13$ TeV, using the LHCf-Arm2 detector. This measurement, discussed in Chapter 3, provides valuable data for HIMs, particularly concerning the electromagnetic and strange quark

components, which are significant for understanding muon production in EAS. The results showed that while the QGSJET II-04 model closely matched experimental data at high x_F values, neither it nor the other HIMs under analysis (EPOS-LHC, DPMJET 3.06, and SIBYLL 2.3) could fully reproduce the η meson production across the entire x_F range. These discrepancies highlight the need for further refinement in the modelling of forward particle production, particularly in terms of strange quark contributions. The calibration of the LHCf-Arm2 detector, as discussed in Chapter 4 plays a key role in ensuring the accuracy and precision of the results, especially in preparation for the analysis of new data collected during LHC Run III in proton-proton collisions at $\sqrt{s} = 13$ TeV. The detailed calibration process, including energy resolution tests and corrections for systematic uncertainties, guarantees reliable detector performance. Looking towards the future, two ongoing preliminary studies, presented in the following sections, extend the work carried out in this thesis. The first, detailed in Section 5.1, involves the measurement of the η/π^0 production ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV. as discussed in Section 3.1, this observable is particularly important for calibrating HIMs, as it provides deeper insights into the relative production rates of neutral mesons in the forward region, which directly impacts the understanding of electromagnetic shower development in EAS. The preliminary results indicate discrepancies between the experimental data and model predictions, particularly at higher x_F values, pointing to the need for further analysis and larger datasets, which are expected to become available from LHC Run III. The second preliminary study, outlined in Section 5.2, explores the application of machine learning techniques to the reconstruction of multiple calorimetric clusters in the LHCf detector. This approach is motivated by the increasing complexity of the events observed in forward particle production, where traditional methods struggle to accurately reconstruct events involving multiple particles hitting the same detector. By leveraging machine learning models, the LHCf experiment aims to improve the precision of energy and position reconstruction for these events, which is particularly relevant for analyzing more complex particles like K_s^0 and Λ^0 in the upcoming data from LHC Run III. The initial results from ML applications demonstrate a significant improvement over standard methods, marking a promising di-

rection for future data analysis. In conclusion, the work presented in this thesis provides important contributions to the calibration and testing of HIMs through the measurement of forward η and π^0 meson production and also lays the groundwork for future studies that will exploit advanced techniques like machine learning to enhance the accuracy of particle reconstruction. The ongoing research, particularly with the improved detector performance in LHC Run III, will lead to more precise cosmic ray models and a deeper understanding of ultra-high energy cosmic rays phenomena, leading to significant advances in both astrophysics and high-energy particle physics.

5.1 Preliminary measurement of the forward η/π^0 ratio with the LHCf-Arm2 detector in p-p collisions at $\sqrt{s} = 13$ TeV

This section outlines the methodology used to extract the preliminary x_F distribution of the η/π^0 production ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV with the LHCf-Arm2 detector. The η/π^0 ratio is an important observable for calibrating and testing HIMs and understanding electromagnetic shower development in EASs, as discussed in Section 3.1. This analysis involves the reconstruction of π^0 mesons and the comparison of their production rates with respect to η meson (detailed in Chapter 3), using the same x_F and p_T binning, within the $p_T < 1.1$ GeV/c range, aggregated into a single bin, and within the $0.37 \leq x_F < 0.93$ range, which was divided into 8 bins. This ensures that the two datasets are directly comparable for calculating the η/π^0 ratio. Moreover, the same experimental and MC datasets, described in Sections 3.2.1 and 3.2.2, respectively, were used.

5.1.1 Preliminary measurement of the forward π^0 production rate in p-p collisions at $\sqrt{s} = 13$ TeV

The methodology used to extract the π^0 production spectrum is similar to the approach detailed in the η meson analysis described in Chapter 3.

However, the π^0 analysis includes the reconstruction of both Type I and Type II events, increasing the acceptance of the detector for π^0 candidates:

- **Type I Events:** In this case, the two photons produced from the $\pi^0 \to \gamma\gamma$ decay enter two separate towers of the LHCf-Arm2 detector. The same reconstruction and selection procedures used for the η meson apply here.
- **Type II Events:** Unlike the η meson, the π^0 analysis also considers Type II events, where both photons enter the same detector tower, thanks to the higher statistics. Type II events are further categorized based on whether the photons hit the Small Tower or the Large Tower of the LHCf-Arm2 detector. The energy deposited by each photon is inferred from the ratio of the peak heights in the silicon detector layers, particularly the layer with the highest energy deposition. This technique allows the sharing of the total energy between the two photon candidates.

Figure 5.1 displays the invariant mass spectra ($M_{\gamma\gamma}$) for the π^0 candidates for Type I (left panel), Type II Small Tower (central panel), and Type II Large Tower (right panel) events, respectively. The characteristic π^0 peaks are visible in all cases, with background events modelled by third-order Chebyshev polynomials (green curves), and the signal is fitted using an asymmetric Gaussian function (red and blue curves). In the case of Type II events, due to the low performances of the energy sharing method (discussed in Section 5.2) the peaks turned out to be enlarged, so an additional asymmetric Gaussian had to be added to the signal fit function to account for the spreading. From the fits, the signal and background regions were derived and the x_F distributions of the π^0 of each type were extracted, subtracting the background via the sideband method, in the same way as that described in Section 3.3.4. The x_F distribution of π^0 production, measured for both Type I and Type II events, is shown in Figure 5.2. The corrections and uncertainties were calculated in the same way as described in Sections 3.4 and 3.5, respectively. The π^0 production spectrum is compared with predictions from various HIMs, including QGSJET II-04, EPOS-LHC, SIBYLL 2.3, and DPMJET 3.06. Preliminary results indicate that none of the models is able to reproduce the shape of the experimental distribution in the whole x_F range.

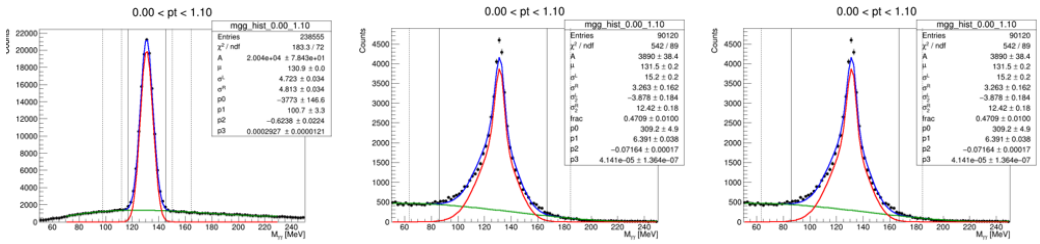

Figure 5.1: Invariant mass distributions of π^0 candidates for Type I (left), Type II events in the Small Tower (middle), and Type II events in the Large Tower (right) detected by the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV for $p_T < 1.1$ GeV/c. In all cases, the π^0 signal is fitted with an asymmetric Gaussian function (red and blue curves), while the background is modelled by third-order Chebyshev polynomials (green curves). The broadening of the signal for Type II events, particularly in the Small and Large Towers, is attributed to the lower accuracy of the energy-sharing method used to reconstruct the two photon candidates in the same tower and is taken into account by adding an additional asymmetric Gaussian function to the signal component.

5.1.2 Calculation of the η/π^0 production ratio

By the distributions of the η and π^0 production rate (Figures 3.26 and 5.2, respectively), the η/π^0 ratio was calculated for each x_F bin, separately for the three types of π^0 events (Type I and Type II for both Small and Large Towers), to cover a larger x_F range. For the first time this measurement is carried out in the forward region of high-energy collisions. Figure 5.3 shows the measured η/π^0 ratio for the different event types and compares these results with the predictions of the HIMs QGSJET II-04, EPOS-LHC, SIBYLL 2.3, and DPMJET 3.06. This preliminary result indicates that only QGSJETII-04 and DPMJET 3.06 are able to reproduce the shape of the experimental distribution, while EPOS-LHC and SIBYLL 2.3 predict a much larger value of the ratio, especially at high x_F. One possible reason for this discrepancy is the fact that EPOS-LHC and SIBYLL 2.3 have a greater emphasis on producing low-mass resonances that can decay into η mesons than QGSJETII-04 and DPMJET 3.06, although discussions on this discrepancy with model developers are currently ongoing.

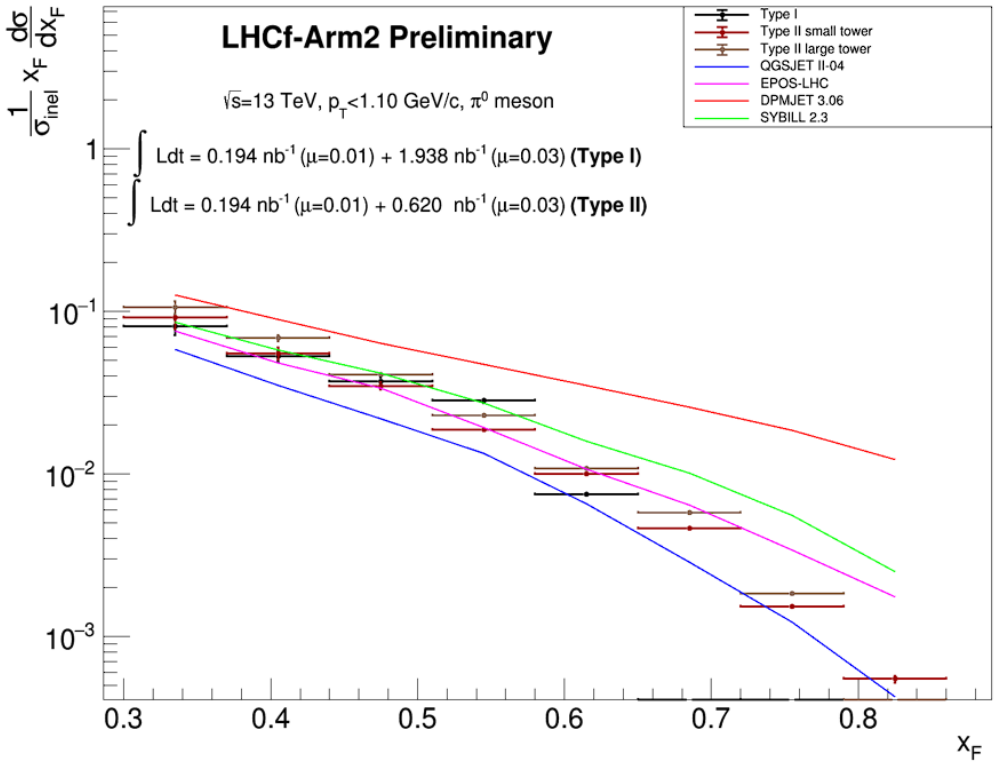

Figure 5.2: Inclusive π^0 production rate as a function of x_F for $p_T < 1.1$ GeV/c, measured using the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV. The experimental results are presented for Type I (black points) and Type II events (Type II in the Small Tower in red, and Type II in the Large Tower in brown). The predictions from various HIMs (QGSJET II-04, EPOS-LHC, DPMJET 3.06, and SIBYLL 2.3) are shown for comparison. None of the models is able to fully capture the experimental distribution across the entire x_F range.

The results presented in this section are currently preliminary, requiring refinements in the π^0 reconstruction methods (especially Type II) and in the calculation of systematic corrections and uncertainties. In addition, the new data acquired in the LHCf data holdings in Run III (discussed in Section 3.8), thanks to the increase in the statistics of the π^0 and η

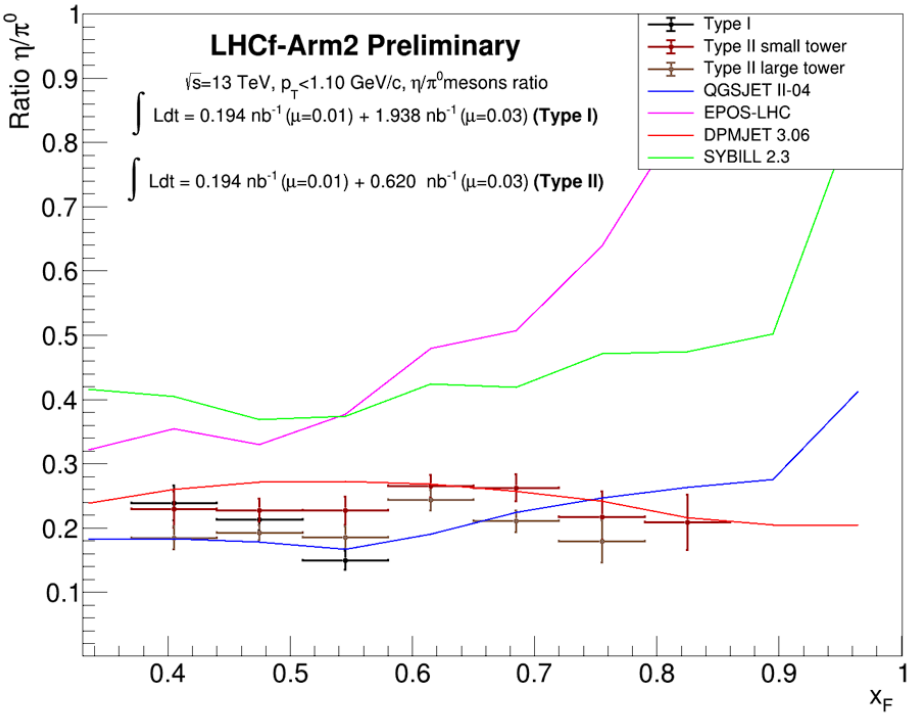

Figure 5.3: The η/π^0 production ratio as a function of x_F for $p_T < 1.1$ GeV/c, measured using the LHCf-Arm2 detector in proton-proton collisions at $\sqrt{s} = 13$ TeV. The experimental results are shown for π^0 Type I (black points) and Type II events (Type II Small Tower in red, and Type II Large Tower in brown). The predictions from various hadronic interaction models (QGSJET II-04, EPOS-LHC, DPMJET 3.06, and SIBYLL 2.3) are presented for comparison.

mesons, will allow an improvement of this analysis, making it possible to derive the value of the production ratio in various p_T bins, providing useful insights for the calibration and testing of HIMs.

5.2 Reconstruction of multiple calorimetric clusters in the LHCf experiment with machine learning techniques

One of the significant challenges in the LHCf experiment is the accurate reconstruction of multiple calorimetric clusters, especially when more than two particles hit the same detector tower or when the same cluster is produced by photons and neutrons at the same time. In recent years, machine learning (ML) techniques have been increasingly applied to the reconstruction of multiple calorimetric clusters [115, 116, 117]. This section details the ML approaches currently under development in the LHCf experiment, focusing on the reconstruction of events involving multiple hits, which is a significant challenge due to calorimetric signal overlap. Section 5.2.1 discusses the motivations for integrating ML techniques into the LHCf reconstruction pipeline, highlighting the limitations of traditional methods in handling complex multi-hit events. Section 5.2.2 outlines the ML pipeline being developed, the dataset preparation process and the architecture of the ML models. The results of the model evaluation are discussed in Section 5.2.3, where the performance of the ML models is compared with the standard LHCf energy-sharing method. Finally, Section 5.2.4 explores the prospects of ML in the LHCf experiment, identifying areas for improvement and potential new applications of ML techniques in the reconstruction pipeline.

5.2.1 Motivations for Machine Learning in LHCf

The primary motivation for integrating ML techniques into the LHCf reconstruction pipeline arises from the need to enhance the accuracy of cluster reconstruction in scenarios involving multiple particles. The traditional energy and position reconstruction methods, described in Section 3.3.2, respectively, are very efficient in the case of single-hit events, characterised by a single particle hitting the calorimetric towers, but they begin to lose in performance when there are two or more particles in a single tower. In the latter case, there is only one energy release corresponding to the total energy, so a method for dividing the energy between the inci-

dent particles is required. Regarding position reconstruction, it is possible to exploit the lateral distribution in position-sensitive detectors to identify the impact position of multiple particles by identifying peaks with TSpectrum and performing a fit on each one with a three-component Lorentian function (Equation 3.1). For one or two peaks the method works well, but shows large inaccuracies when there are three or more peaks. Moreover, as the number of peaks increases, the computational complexity increases. The implementation of ML methods for performing these tasks could improve the reconstruction efficiency of multiple calorimetric clusters, bringing benefits to various analyses:

- **Type II π^0 and η analysis:** These events occur when the two photons produced in the π^0 and η decays hit the same calorimetric tower. As discussed in Section 5.1.1, in this case the total energy in the tower is divided to the two photons using the ratio of the height of the corresponding peaks in the transverse profile measured by the position-sensitive detectors, in the layer with the maximum energy release. As shown in Figures 5.1, this method results in a reduction in the reconstruction efficiency of the individual energies, producing a broadening in the invariant mass peak of the two photons corresponding to the decay of π^0, leading to a deterioration in the reconstruction performance of the events.

- **K^0_s analysis:** The measurement of forward K^0_s production is one of the main analysis targets of the new LHCf data-taking in LHC-RUN III (described in Section 3.8). This would be the first observation of very forward K^0 in high-energy proton collisions and would have a strong impact on the calibration and testing of HIMs. Indeed, this measurement would improve the modelling of forward strange quark production and allow for the inference of total kaon production (both neutral and charged) in the EASs, which are among the primary sources of leptons [103]. Consequently, this observation would have a significant impact on both the Muon Puzzle and neutrino physics. The challenge in identifying K^0_s particles lies in the complexity of their decay topology, which can be reconstructed in LHCf. With a branching ratio of 30.69% [108], K^0_s decay into two π^0, which in turn decay into four photons in the final state. As shown in

Figure 5.4, which depicts possible K_s^0 event types in the Arm1 (top panels) and Arm2 (bottom panels) detectors and their respective acceptance maps, events with two photons per tower are very rare, while those requiring the reconstruction of three or four photons in a single tower are more common. This necessitates the enhancement of traditional position and energy reconstruction methods using ML.

- Λ^0 **analysis:** The study of forward Λ^0 production is crucial for understanding forward strange quark production and hadronization processes in high-energy collisions. Also, this particle was never measured in the very forward region of high-energy collisions and would have a critical impact on the calibration and testing of HIMs. Λ^0 decays are characterized by complex event topologies and can be identified by LHCf detector by the decay in a neutron and a π^0, having a two photons and one neutron in the final state. As shown in Figure 5.5, the different event types in the Arm1 (top panels) and Arm2 (bottom panels) detectors and their corresponding acceptance maps highlight that the accurate reconstruction of these events is challenging because the most common events are the one with one neutron and one or two photons in a single tower. Therefore, in addition to developing accurate methodologies for position reconstruction and energy sharing, it is also necessary to devise a method for identifying the neutron. Given the complexity of this task, developing ML techniques could be essential.

5.2.2 Machine Learning Approach

The ML pipeline for multiple calorimetric cluster reconstruction in LHCf, actually under development, will involve several key steps:

- Peak identification and inference on the number of hits.
- Position reconstruction of multi-hit events.
- Energy sharing between particles.

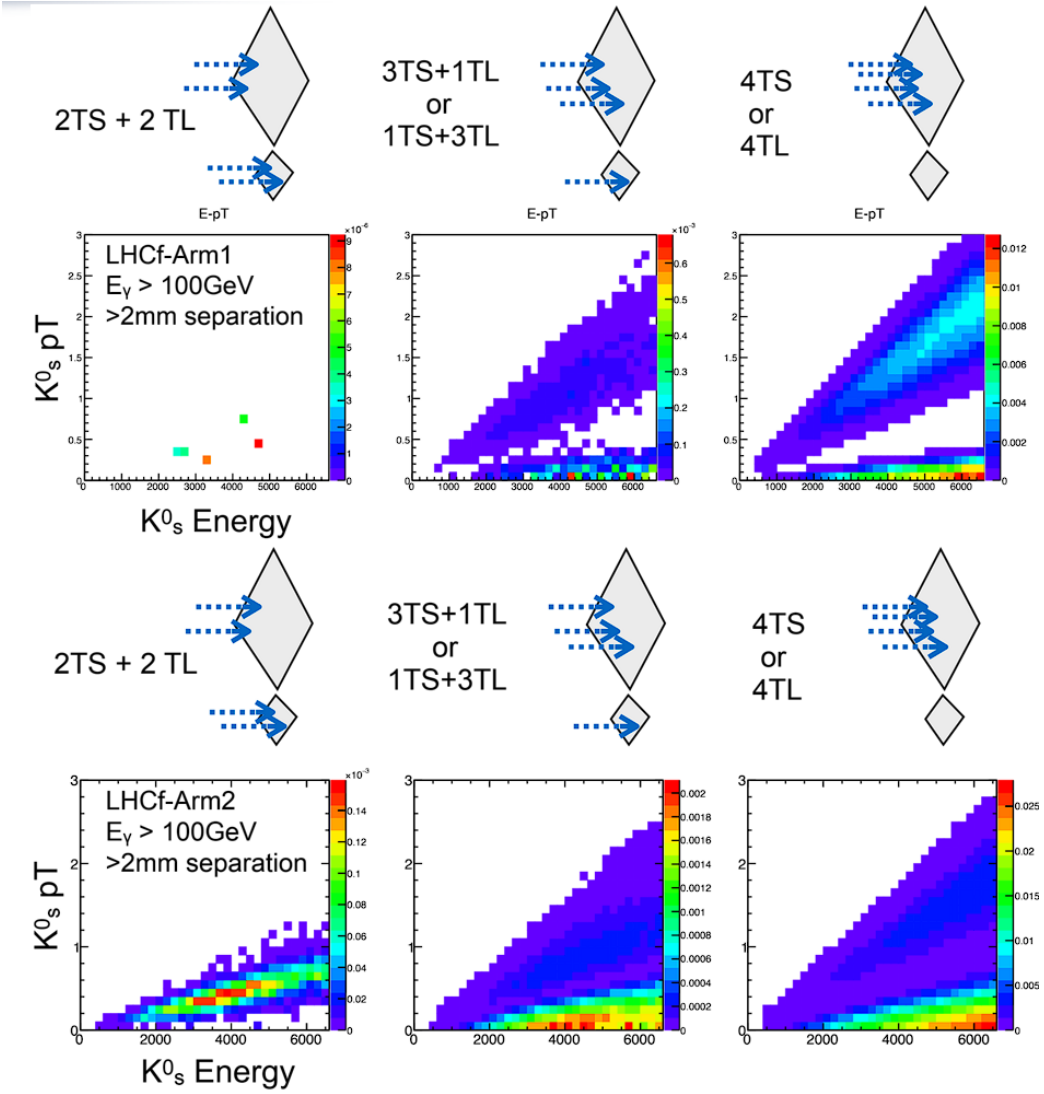

Figure 5.4: Acceptance maps and event topologies for K^0_s decays as observed in the LHCf detectors. The top panels represent the event types in the Arm1 detector, while the bottom panels correspond to Arm2. The diagrams illustrate different photon multiplicities per tower, with the left, middle, and right columns showing cases of 2 photons per tower, 3 photons in one tower and 1 in another, and 4 photons in a single tower, respectively. Notably, events with higher photon multiplicity in a single tower are more common.

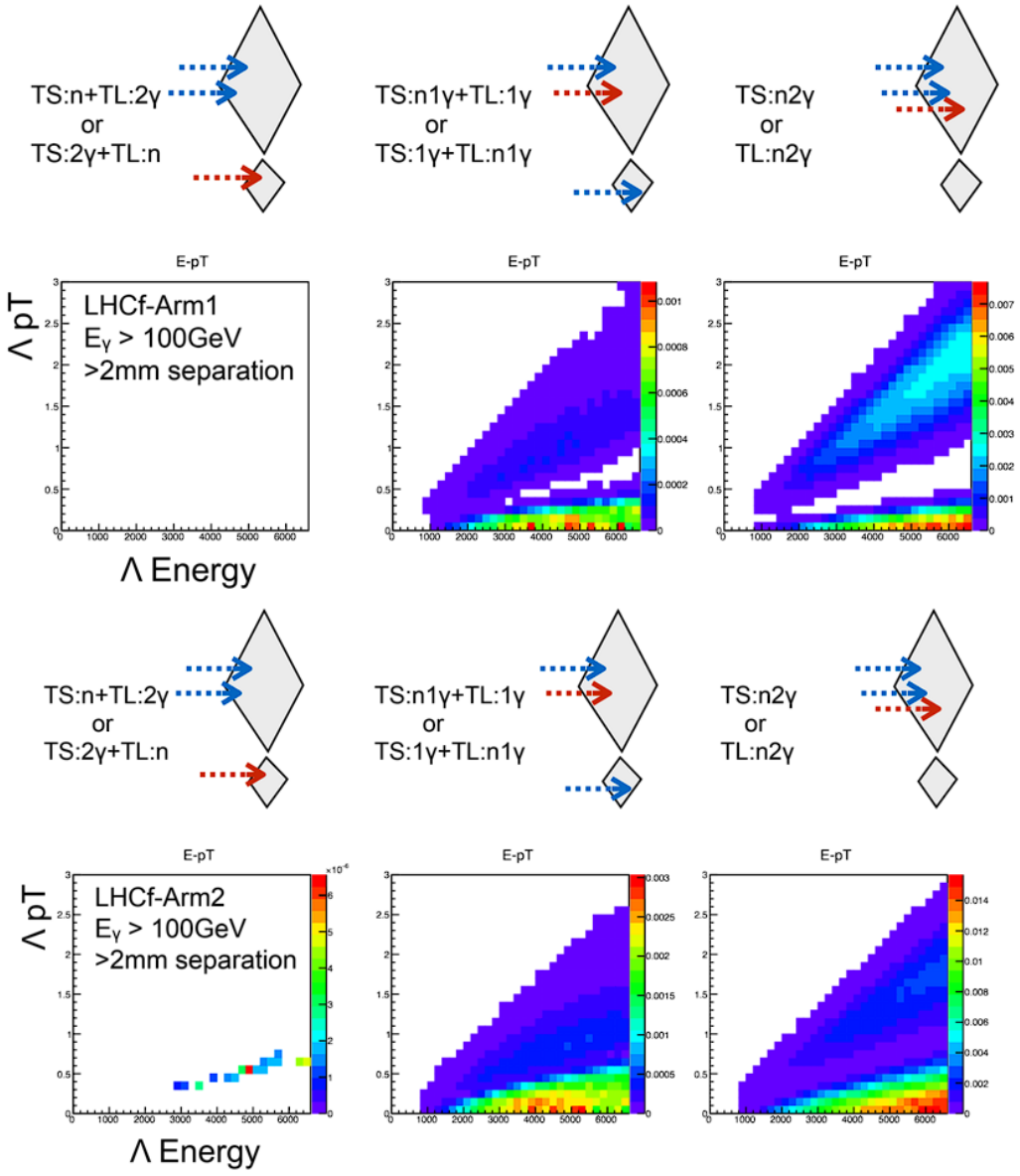

Figure 5.5: Acceptance maps and event topologies for Λ^0 decays observed in the LHCf detectors. The top panels represent the event types in the Arm1 detector, while the bottom panels correspond to Arm2. The diagrams illustrate various photon and neutron multiplicities per tower, with the left, middle, and right columns depicting different configurations. Accurate event reconstruction using ML could be crucial due to the complex topology of these decays.

This work is focused on the last point, in the case of two photons hitting the same tower of the LHCf-Arm2 detector. This will permit the improvement of the analysis of Type II π^0 and η and to find the best models and methods to analyze events with 3 or 4 photons or with the presence of a neutron for K_s^0 and Λ^0 analyses.

Dataset Preparation

To infer the fraction of the total energy shared between two-hit events, we used the results from the 3-component Lorentzian fit (Equation 3.1) of the silicon transverse profile in the first four layers as input for our models. The dataset was generated using a full detector MC simulation of proton-proton collisions at $\sqrt{s} = 13$ TeV, employing the QGSJETII-04 model as the event generator (the methodology of the simulation is detailed in Section 3.2.2). Separate models were constructed and trained for each tower in the detector. Specifically, the seven fit parameters for each particle, for both the X and Y views, were extracted from the first two silicon plane pairs and used as input variables, resulting in a total of 56 input variables. An example of an event used in the ML models is shown in Figure 5.6. This figure displays the fit results for each of the two particles under investigation and the fraction of the total energy carried by each. A total of approximately 130000 and 60000 events were used for the Small Tower and Large Tower, respectively. 70% of the events were used to train the models, while the remaining 30% were used to test performance.

Model Architectures

The ML pipeline utilized ensemble methods based on gradient-boosting decision trees (BDTs [118]). The architecture tested includes:

- **First Level BDTs**: Two separate BDT models were trained to predict the energy of each particle individually.
- **On-Top BDT**: A higher-level BDT model that combines the predictions of the first-level BDTs with the original input dataset to infer the total energy of all particles in the event.

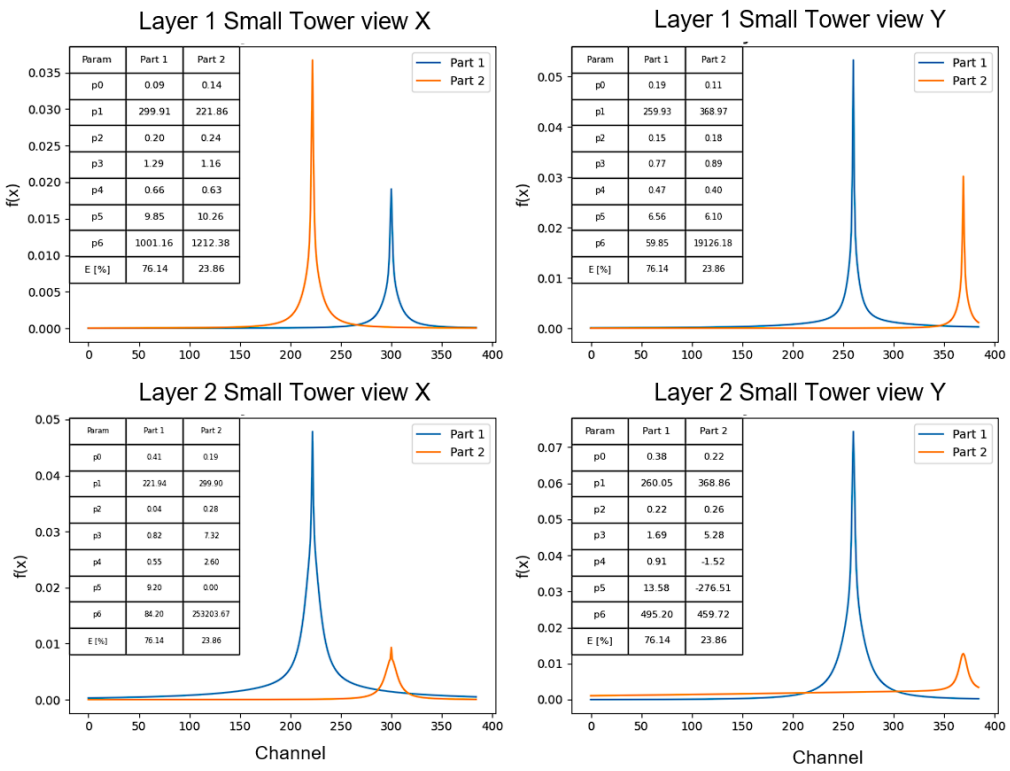

Figure 5.6: Lorentzian fits for the transverse profiles of two particles in the first two silicon layers (Layer 1 and Layer 2) of the Small Tower detector, shown for both X and Y views. The plots illustrate the fitting results of the 3-component Lorentzian function for each particle (denoted as Part 1 and Part 2). The table in each plot shows the corresponding fit parameters (p0-p6) and the fraction of total energy carried by each particle (E%). These results serve as input variables for machine learning models designed to infer the energy distribution between two-hit events.

Hyperparameter	Value
L2 Leaf Regularization	20
Random Strength	$1.44 \cdot 10^{-8}$
Learning Rate	0.0001
Iterations	300000
Depth	3
Minimum Data in Leaf	1

Table 5.1: Hyperparameters for the BDT Models.

A schematic view of the architecture can be visualized in Figure 5.7 Two BDT ensemble methods were trained using libraries XGBoost [119] and CatBoost [120]. The hyperparameters used for all the BDT models are detailed in Table 5.1. The models were evaluated using the Root Mean Square Error (RMSE) as the performance metric, to compare their results with the standard LHCf energy-sharing method described in Section 5.1.1.

5.2.3 Results

To evaluate the performance of the developed ML models for energy sharing in two-particle hit events, both the standard LHCf energy-sharing method (described in Section 5.1.1) and the BDT ensemble models based on XGBoost and CatBoost were compared. The performance was measured using the Root Mean Square Error (RMSE) as the evaluation metric, comparing the predicted values with the true values. Figures 5.8 and 5.9 show the results for the Small Tower and the Large Tower, respectively. Each figure presents scatter plots that compare the predicted values with the true values for the Baseline model (standard LHCf method), XGBoost, and CatBoost. As shown in Figure 5.8, the results for the Small Tower indicate a significant improvement in the performance of the machine learning models compared to the standard LHCf energy-sharing method. The XGBoost model achieved an RMSE of 169.78 for the first particle (E1) and 116.39 for the second particle (E2), significantly lower than the RMSEs of the Baseline model, which were 506.73 and 548.61, respectively. The CatBoost model achieved comparable results, with an

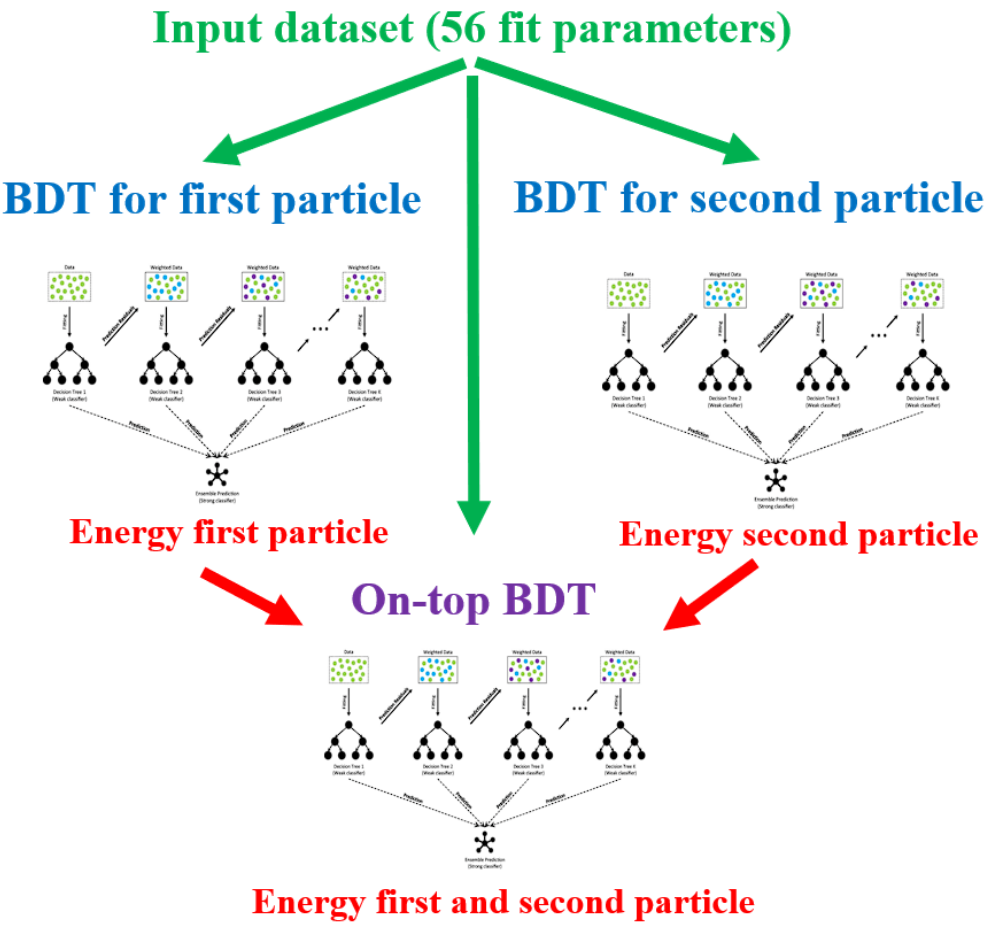

Figure 5.7: Schematic representation of the machine learning pipeline used for energy prediction in two-hit events. The input dataset, consisting of 56 fit parameters from the silicon transverse profiles, is fed into two separate BDT (Boosted Decision Tree) models. Each BDT is trained to predict the energy of a single particle in the event. The predictions from these first-level BDTs, along with the original input dataset, are then combined and passed into an "On-Top BDT." This final model infers the total energy shared between the two particles.

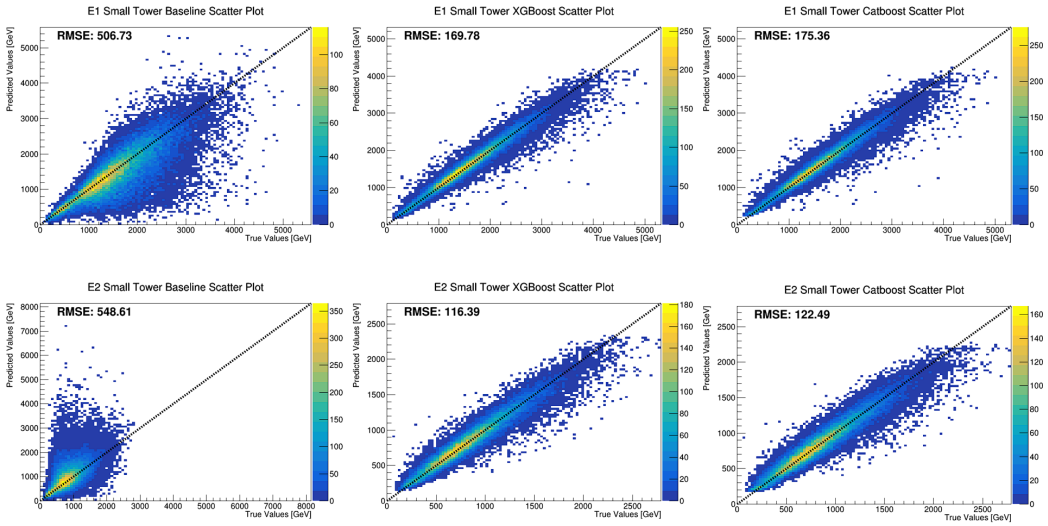

Figure 5.8: Comparison between predicted and true values for the Baseline, XGBoost, and CatBoost models applied to particles E1 (top panels) and E2 (bottom panels) in the Small Tower. The RMSE is reported for each model.

RMSE of 175.36 for E1 and 122.49 for E2, showing slightly lower performance than XGBoost but still significantly better than the standard method. The results for the Large Tower, shown in Figure 5.9, confirm the better performance of the ML models over the standard method. For E1, the XGBoost model achieved an RMSE of 135.94, compared to 369.60 for the Baseline model. Similarly, for E2, XGBoost obtained an RMSE of 92.08, versus 398.93 for the Baseline model. Similarly, CatBoost achieved an RMSE of 141.09 GeV for E1 and 97.22 GeV for E2, showing results very close to those of XGBoost and significantly better than the Baseline. The obtained results clearly show that the ML models offer a significant improvement in the accuracy of energy reconstruction for two-hit events in the LHCf towers. The significantly lower RMSE values compared to the Baseline method indicate that these models can handle the complexity of energy sharing among multiple particles, thereby reducing reconstruction errors and improving the quality of the physical analyses.

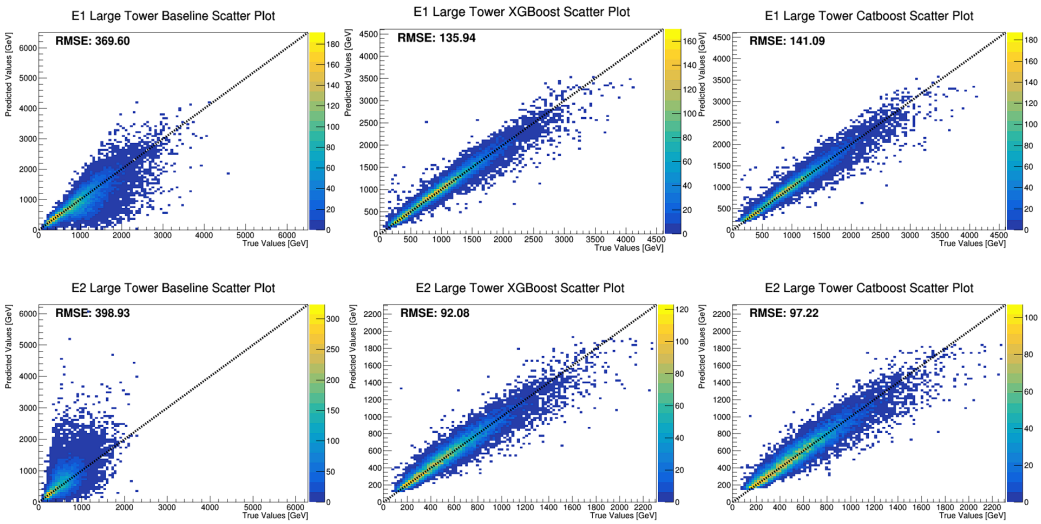

Figure 5.9: Comparison between predicted and true values for the Baseline, XGBoost, and CatBoost models applied to particles E1 (top panels) and E2 (bottom panels) in the Large Tower. The RMSE is reported for each model.

5.2.4 Future prospects

Current methods are primarily designed for events with up to two particles hitting the same detector tower. Extending these techniques to handle more complex scenarios with three or four hits is crucial for comprehensive multi-hit event reconstruction, required for the analysis of K_S^0 and Λ^0. Increased number of hits leads to significant signal overlap, making accurate deconvolution more challenging. By using the same models of the two-hit analysis, preliminary tests using a three-hit dataset were performed using a similar architecture, composed of three first-level BDTs to infer the single particle energy and one on-top BDT that combines the predictions of the first-level BDTs with the original input dataset to infer the total energy of all particles in the event. The CatBoost library was used in this case. Figure 5.10 shows an example of three-hit events together with the 3-component Lorentian function fit parameters. The results obtained with the three-hit dataset are depicted in Figure 5.11. These scatter plots show the predicted versus true energy values for three dif-

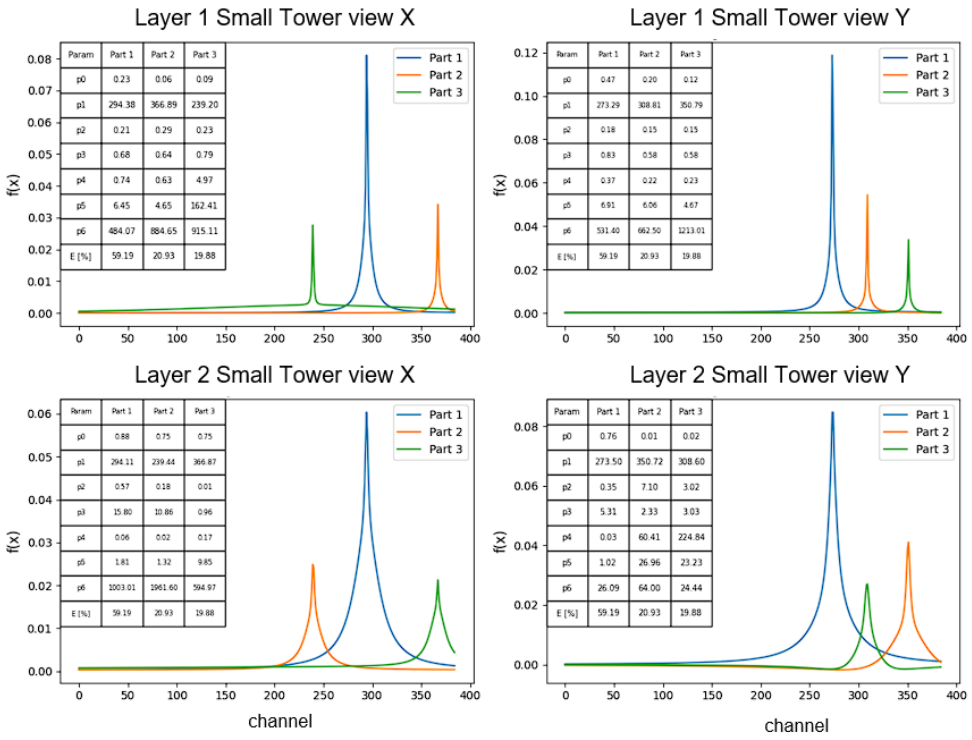

Figure 5.10: Lorentzian fits for the transverse profiles of three particles in the first two silicon layers (Layer 1 and Layer 2) of the Small Tower detector, shown for both X and Y views. The plots illustrate the fitting results of the 3-component Lorentzian function for each particle (denoted as Part 1, Part 2, and Part 3). The table in each plot shows the corresponding fit parameters (p0-p6) and the fraction of total energy carried by each particle (E%).

ferent events (E1, E2, and E3) in both the Small Tower and Large Tower. The scatter plots indicate that the CatBoost model can handle the case of three-hit events, although with varying degrees of accuracy. However, it is important to note that this dataset consists of only about 1000 events for both towers, with a 70/30 split between the training and testing datasets. Given the relatively low statistics, the current results are preliminary. We expect that increasing the dataset size could lead to significant improvements in model performance, as a larger dataset would provide more robust training and more reliable validation. Moreover, the traditional (baseline) method for energy sharing in three-hit events is still under development. Consequently, it is currently not possible to directly compare the ML results with those obtained using the conventional approach. The preliminary results presented demonstrate the potential of machine learning (ML) to significantly enhance the accuracy and efficiency of energy reconstruction in complex multi-hit events within the LHCf experiment. However, these initial studies also highlight several areas where further improvements and developments are necessary to fully leverage ML techniques. One of the primary future tasks involves extending the ML pipeline to efficiently handle events with three or more particles hitting the same detector tower. The initial results using a three-hit dataset have shown promise, but there is a clear need for further refinement. As the complexity of multi-hit events increases, traditional methods of energy sharing and position reconstruction become less effective due to significant signal overlap. ML models have the potential to deconvolute these overlapping signals more accurately, especially when trained on larger and more diverse datasets. Another promising strategy for improving the performance of ML models is the direct use of raw energy deposition distributions in the silicon detectors, rather than relying solely on fitted parameters. The raw signals provide a more detailed representation of the energy release patterns, which ML models can exploit to achieve more precise reconstructions. Incorporating these raw distributions as additional input features could allow the models to learn more intricate patterns associated with multi-hit events, potentially leading to even greater accuracy in energy sharing and position reconstruction tasks. Looking forward, several specific tasks within the LHCf reconstruction pipeline could benefit from the integration of advanced

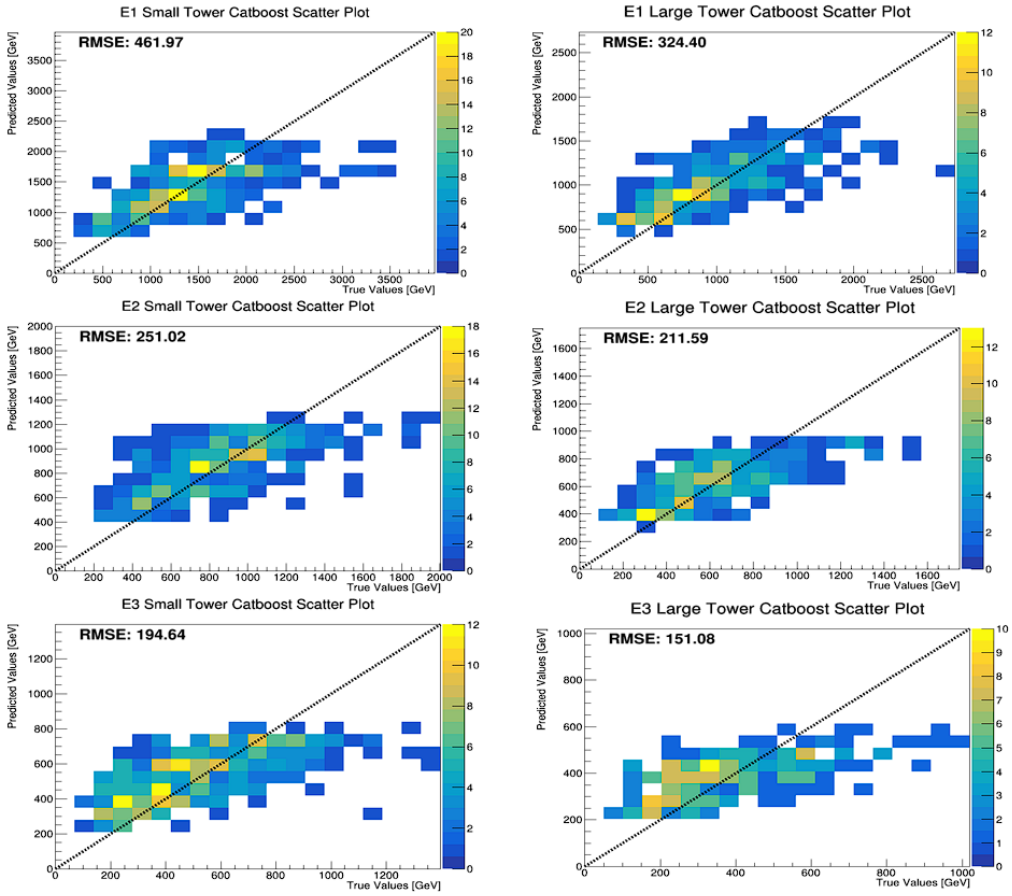

Figure 5.11: Scatter plots of predicted versus true energy values for three-hit events (E1, E2, E3) in the Small Tower (left) and Large Tower (right), using the CatBoost model. The RMSE for each particle is shown.

ML techniques:

- **Peak identification and classification**: ML models can be developed to automatically identify and classify peaks in the energy distributions, reducing reliance on semi-automated methods such as TSpectrum. This would be particularly useful in scenarios with multiple overlapping peaks, where traditional methods struggle.
- **Position reconstruction:** Accurately determining the position of multiple particles hitting the same tower is crucial for precise event

reconstruction. Traditional methods rely on fitting techniques like the Lorentzian function, which can become computationally expensive and less accurate as the number of particles increases.

- **Neutron identification:** In events where neutrons are involved alongside photons, accurately distinguishing between these different types of particles is crucial. ML models can be trained specifically to recognize the distinct signatures of neutrons and photons in the calorimetric data, improving the accuracy of particle identification and subsequent analysis.

To achieve these goals, it is essential to scale the current models by increasing the size and diversity of the training datasets. This scaling will help ensure that the models generalize well across different types of events and detector conditions. Additionally, ongoing developments in ML, such as deep learning architectures and reinforcement learning, could be explored to further enhance the capabilities of the reconstruction pipeline.

List of Figures

List of Tables

Bibliography

[1] T. Stanev, *High Energy Cosmic Rays*. Astronomy and Astrophysics Library. Springer, Berlin, 2004.

[2] J. A. Simpson, *Elemental and Isotopic Composition of the Galactic Cosmic Rays*, *Annual Review of Nuclear and Particle Science* **33** (1983), no. 1 323–382.

[3] P. Sokolsky, *Introduction to Ultrahigh Energy Cosmic Ray Physics*. Frontiers in Physics. Westview Press, Boulder and Oxford, 2004.

[4] T. K. Gaisser, *Cosmic Rays and Particle Physics*. Cambridge University Press, Cambridge, England; New York, 1990.

[5] A. Aab et al., *Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above* 8×10^{18} *eV*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017), no. 6 026.

[6] A. Aab et al., *Observation of a Dipole in the Arrival Directions of Ultra-high-energy Cosmic Rays*, *Science* **357** (2017), no. 6537 1266.

[7] A. Aab et al., *Multi-Messenger Observations of a Binary Neutron Star Merger*, *Astrophysical Journal Letters* **853** (2018), no. 2 L29.

[8] R. U. Abbasi et al., *First Observation of the Greisen-Zatsepin-Kuzmin Suppression*, *Astrophysical Journal Letters* **790** (2014), no. 2 L21.

[9] R. M. Baltrusaitis et al., *The Utah Fly's Eye detector*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **240** (1985), no. 2 410–428.

[10] D. J. Bird et al., *Detection of a Cosmic Ray with Measured Energy Well Beyond the Expected Spectral Cutoff due to Cosmic Microwave Radiation*, *The Astrophysical Journal* **441** (1995), no. 1 144–150.

[11] M. Fukushima et al., *Telescope Array Project for Extremely High Energy Cosmic Rays*, *Progress of Theoretical Physics Supplement* **151** (2003) 206–210.

[12] Telescope Array Collaboration et al., *An Extremely Energetic Cosmic Ray Observed by a Surface Detector Array*, *Science* **382** (2023), no. 6673 903–907.

[13] T. Antoni et al., *KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems*, *Astroparticle Physics* **24** (2005), no. 1 1–25.

[14] L. A. Anchordoqui, *Ultra-High-Energy Cosmic Rays*, *Physics Reports* **801** (2019) 1–93, [arXiv:1807.09645].

[15] M. Unger, G. R. Farrar, and L. A. Anchordoqui, *Origin of the Ankle in the Ultrahigh Energy Cosmic Ray Spectrum, and of the Extragalactic Protons Below It*, *Physical Review D* **92** (2015), no. 12 123001.

[16] R. Aloisio, D. Boncioli, A. di Matteo, A. F. Grillo, S. Petrera, and F. Salamida, *Cosmogenic Neutrinos Challenge the Cosmic Ray Proton Dip Model*, *Journal of Cosmology and Astroparticle Physics* **2015** (2015), no. 10 006.

[17] G. Giacinti, M. Kachelrieß, and D. V. Semikoz, *Escape Model for Galactic Cosmic Rays and an Early Extragalactic Transition*, *Physical Review D* **91** (2015), no. 8 083009.

[18] T. Guzik et al., *The ATIC long duration balloon project*, *Advances in Space Research* **33** (2004), no. 10 1763–1770.

[19] E. Seo et al., *Cosmic-ray Energetics and Mass (CREAM) balloon project*, *Advances in Space Research* **33** (2004), no. 10 1777–1785.

[20] P. Picozza et al., *PAMELA–A payload for antimatter matter exploration and light-nuclei astrophysics*, *Astroparticle Physics* **27** (2007), no. 4 296–315.

[21] S. Ting, *The Alpha Magnetic Spectrometer on the International Space Station*, *Nuclear Physics B - Proceedings Supplements* **243-244** (2013) 12–24.

[22] S. Torii et al., *The CALET, CALorimetric Electron Telescope, mission for the International Space Station*, *Nuclear Physics B - Proceedings Supplements* **113** (2002), no. 1 103–110.

[23] A. De Benedittis et al., *The DAMPE experiment: performances and results*, *Proceedings of Science* **364** (2020) 030.

[24] A. D. Angelis et al., *The HERD space mission*, *Nuclear Physics B - Proceedings Supplements* **243-244** (2017) 176–181.

[25] N. Tomassetti, *Direct Measurements of Galactic Cosmic Rays*, in *27th European Cosmic Ray Symposium (ECRS)*, 2023. Published on December 14, 2023.

[26] J. R. Patterson and A. M. Hillas, *The Extensive Air Showers Produced by Ultra-high-Energy Cosmic Rays*, *Journal of Physics G: Nuclear Physics* **9** (1983) 1433–1448.

[27] W. Kolhörster, I. Matthes, and E. Weber, *On the Nature of Extensive Air Showers*, *Naturwissenschaften* **26** (1938) 576.

[28] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and A. Robley, *Extensive Cosmic-Ray Showers*, *Reviews of Modern Physics* **11** (1939) 288.

[29] A. Aab et al., *The Pierre Auger Observatory Upgrade*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **798** (2015) 172–213.

[30] A. Aab et al., *The AMIGA Extension of the Pierre Auger Observatory: Performance and First Data*, *Journal of Instrumentation* **11** (2016), no. 02 P02012.

[31] A. Castellina, *The AugerPrime Upgrade of the Pierre Auger Observatory: Status and Perspectives*, *EPJ Web of Conferences* **210** (2019) 06002.

[32] A. Aab et al., *Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above* $10^{17.8}$ *eV*, *Physical Review D* **90** (2014), no. 12 122005.

[33] R. U. Abbasi et al., *Measurement of the Flux of Ultra High Energy Cosmic Rays by the Telescope Array Experiment*, *Physical Review D* **98** (2018), no. 2 022002.

[34] R. U. Abbasi et al., *Search for Large-scale Anisotropy in the Arrival Directions of Cosmic Rays with Energies above* 10^{19} *eV Detected at the Telescope Array Experiment*, *The Astrophysical Journal* **865** (2018), no. 1 74.

[35] R. U. Abbasi et al., *The Energy Spectrum of Ultra-High-Energy Cosmic Rays Measured by the Telescope Array and the Pierre Auger Observatory*, *The Astrophysical Journal* **909** (2021), no. 2 178.

[36] D. Bergman, J. F. Krizmanic, K. Nakai, Y. Omura, and Y. Tsunesada, *Recent Results from the Telescope Array Experiment*, *Proceedings of Science* **ICRC2019** (2020) 189.

[37] E. Kido, *Recent Results from the Telescope Array Experiment: Highlights of the Telescope Array*, *JPS Conference Proceedings* **19** (2018) 011025.

[38] Y. Tsunesada, *Measurement of UHECR Energy Spectrum with the Pierre Auger Observatory and the Telescope Array*, *Proceedings of Science* **ICRC2023** (2023) 406.

[39] K. H. Kampert and M. Unger, *Measurements of the Cosmic Ray Composition with Air Shower Experiments*, *Astroparticle Physics* **35** (2012), no. 10 660–678.

[40] A. Aab et al., *Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events*, *Physical Review D* **91** (2015), no. 3 032003.

[41] S. Müller, R. Engel, T. Pierog, and M. Roth, *Impact of Shower Fluctuations on the Reconstruction of the Mean Logarithmic Mass Composition of Cosmic Rays*, *Astroparticle Physics* **97** (2018) 174–183.

[42] J. Albrecht et al., *The Muon Puzzle in Cosmic-Ray Induced Air Showers and Its Connection to the Large Hadron Collider*, *Astrophysics and Space Science* **367** (2022), no. 3 1–50.

[43] S. Ostapchenko, *Nonlinear Screening Effects in High Energy Hadronic Interactions*, *Physical Review D* **74** (2006), no. 1 014026.

[44] S. Ostapchenko, *QGSJET-II: Towards Reliable Description of Very High Energy Hadronic Interactions*, *Physics Letters B* **636** (2006) 40–45.

[45] S. Ostapchenko, *Monte Carlo Treatment of Hadronic Interactions in Enhanced Pomeron Scheme: QGSJET-II Model*, *Physical Review D* **81** (2010) 114028.

[46] S. Ostapchenko, *Elastic and Diffractive Scattering in the QGSJET-II Model*, *Physical Review D* **83** (2011) 014018.

[47] S. Ostapchenko, *QGSJET-III: A New Model Framework for High Energy Cosmic Ray Studies*, *EPJ Web of Conferences* **208** (2019) 11004.

[48] K. Werner, F.-M. Liu, and T. Pierog, *Parton Ladder Splitting and the Rapidity Dependence of Transverse Momentum Spectra in Deuteron-Gold Collisions at RHIC*, *Physical Review C* **74** (2006) 044902.

[49] T. Pierog and K. Werner, *EPOS Model and Ultra High Energy Cosmic Rays*, *Nuclear Physics B - Proceedings Supplements* **196** (2009) 102–105.

[50] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, *EPOS LHC: Test of Collective Hadronization with LHC Data*, *Physical Review C* **92** (2015) 034906.

[51] T. Pierog, K. Werner, et al., *EPOS-LHC R: A Refined Model for the Description of Hadronic Interactions*, *Physical Review C* **110** (2024) 012345.

[52] J. Ranft, *Dual Parton Model at Cosmic Ray Energies*, *Physical Review D* **51** (1995) 64–79.

[53] J. Ranft, *DPMJET versions II.3 and II.4: Sampling of Hadron Hadron, Hadron Nucleus and Nucleus Nucleus Interactions at Cosmic Ray Energies According to the Dual Parton Model Description of the Model and Code Manual*, Technical Report INFN-AE-97-45, Gran Sasso Laboratory, 1997.

[54] J. Ranft, R. Engel, and S. Roesler, *DPMJET-III: A Hadronic Interaction Model for Simulating Cosmic Ray Air Showers*, *Nuclear Physics B, Proceedings Supplements* **122** (2003) 392–395.

[55] S. Roesler, R. Engel, and J. Ranft, *The Monte Carlo Event Generator DPMJET-III*, in *Proceedings of the International Conference on Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications (MC 2000)*, pp. 1033–1038, 2000.

[56] A. Fedynitch, R. Engel, T. K. Gaisser, F. Riehn, and T. Stanev, *Calculation of Conventional and Prompt Lepton Fluxes at Very High Energy*, *EPJ Web of Conferences* **99** (2015) 08001.

[57] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, *Cosmic Ray Interaction Event Generator SIBYLL 2.1*, *Physical Review D* **80** (2009) 094003.

[58] F. Riehn, H. P. Dembinski, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, *The Hadronic Interaction Model SIBYLL 2.3c and Feynman Scaling*, in *Proceedings of Science, ICRC2017*, vol. 301, 2017.

[59] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, *Hadronic Interaction Model SIBYLL 2.3d: An Update on the Charm Production and the Soft Hadronic Interactions*, *Physical Review D* **102** (2020), no. 6 063002.

[60] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, *Improved Description of Hadronic Interactions in SIBYLL 2.3d: Impact on Air Shower Simulations*, *Physical Review D* **102** (2020), no. 6 063002.

[61] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** (2008), no. 08 S08001.

[62] A. Aab et al., *Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above* $10^{17.8}$ *eV*, *Physical Review D* **91** (2015), no. 3 032003.

[63] A. Aab et al., *Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory*, *Physical Review Letters* **117** (2016), no. 19 192001.

[64] A. Aab et al., *Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory*, *Physical Review Letters* **126** (2021) 152002.

[65] V. N. Gribov, *A Reggeon Diagram Technique*, *Soviet Physics JETP* **26** (1968) 414–422.

[66] C. Bierlich, G. Gustafson, L. Lönnblad, and H. Shah, *The Angantyr Model for Heavy-Ion Collisions in PYTHIA8*, *Journal of High Energy Physics* **10** (2018) 134.

[67] D. d'Enterria, T. Pierog, and G. Sun, *Constraints from Very High Energy Hadronic Interactions on Nuclear Modifications of the Gluon Density in Lead*, *The Astrophysical Journal* **874** (2019) 152.

[68] L. Cazon et al., *EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, and Yakutsk EASArray Collaborations*, in *Proceedings of Science, ICRC2019*, vol. 214, 2019.

[69] L. Cazon, *Working Group Report on the Combined Analysis of Muon Density Measurements from Eight Air Shower Experiments*, *ArXiv preprint* (2020).

[70] D. Soldin, *Update on the Combined Analysis of Muon Measurements from Nine Air Shower Experiments*, *ArXiv preprint* (2021).

[71] H. P. Dembinski, *Data-driven Model of the Cosmic-ray Flux and Mass Composition from 10 GeV to 100 PeV*, *Astroparticle Physics* **102** (2018) 89–97.

[72] A. Fedynitch, J. B. Tjus, and P. Desiati, *Influence of Hadronic Interaction Models and the Cosmic Ray Spectrum on the High Energy Atmospheric Muon and Neutrino Flux*, *Physical Review D* **86** (2012) 114024.

[73] R. Ulrich, R. Engel, and M. Unger, *Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers*, *Physical Review D* **83** (2011) 054026.

[74] Z. Citron et al., *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC*, *CERN Yellow Reports: Monographs* **7** (2019) 1159.

[75] S. Baur, H. Dembinski, M. Perlin, T. Pierog, R. Ulrich, and K. Werner, *Core-corona Effect in Hadron Collisions and Muon Production in Air Showers*, *ArXiv preprint* (2019).

[76] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** (2008) S08003.

[77] S. Chatrchyan et al., *The CMS Experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08004.

[78] K. Aamodt et al., *The ALICE Experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08002.

[79] A. A. Alves Jr et al., *The LHCb Detector at the LHC*, *Journal of Instrumentation* **3** (2008) S08005.

[80] O. Adriani et al., *The LHCf Detector at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** (2008), no. 08 S08006.

[81] G. Anelli et al., *The TOTEM Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** (2008) S08007.

[82] J. Pinfold et al., *The MoEDAL Experiment at the LHC*, *Nuclear Instruments and Methods in Physics Research Section A* **A478** (2002) 140.

[83] A. Ariga et al., *The FASER Experiment*, *Journal of High Energy Physics* **1901** (2019) 075.

[84] C. Ahdida et al., *SND@LHC - Scattering and Neutrino Detector at the LHC*, *Journal of Instrumentation* **15** (2020) P09002.

[85] M. Tanaka et al., *Applications of Cerium-doped Gadolinium Silicate Gd2SiO5:Ce Scintillator to Calorimeters in High-Radiation Environment*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **404** (1998), no. 2 283–294.

[86] K. Noda et al., *Data Analysis of the LHCf Si Microstrip Sensors*, in *Proceedings of the Institute of High Energy Physics*, vol. 5, pp. 35–38, 2011.

[87] V. Balagura, *Notes on Van der Meer Scan for Absolute Luminosity Measurement*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654** (2011), no. 1 634–638.

[88] Y. Makino et al., *Performance Study for the Photon Measurements of the Upgraded LHCf Calorimeters with Gd2SiO5 (GSO) Scintillators*, *Journal of Instrumentation* **12** (2017) P03023.

[89] M. Ageron, *Trigger Sequencer Card User Manual*. IPNL Lyon, 2001.

[90] C. Ljuslin and C. Paillard, *Front End Control Unit for Embedded Slow Control*, 2003.

[91] P. Aspell et al., *PACE3: A Large Dynamic Range Analog Memory ASIC Assembly Designed for the Readout of Silicon Sensors in the LHC CMS Preshower*, in *Proceedings of the Tenth Workshop on Electronics for LHC and Future Experiments*, pp. 422–425, CERN, 2004.

[92] O. Adriani et al., *Measurement of Zero Degree Inclusive Photon Energy Spectra for $\sqrt{s} = 900$ GeV Proton-Proton Collisions at LHC*, *Physics Letters B* **715** (2012) 298–303.

[93] O. Adriani et al., *Measurement of Zero Degree Single Photon Energy Spectra for $\sqrt{s} = 7$ TeV Proton-Proton Collisions at LHC*, *Physics Letters B* **703** (2011) 128–134.

[94] O. Adriani et al., *Measurement of Very Forward Neutron Energy Spectra for 7 TeV Proton-Proton Collisions at the LHC*, *Physics Letters B* **750** (2015) 360–366.

[95] O. Adriani et al., *Measurements of Longitudinal and Transverse Momentum Distributions for Neutral Pions in the Forward-Rapidity Region with the LHCf Detector*, *Physical Review D* **94** (2016), no. 3 032007.

[96] O. Adriani et al., *Measurement of Forward Neutral Pion Transverse Momentum Spectra for $\sqrt{s} = 7$ TeV Proton-Proton Collisions at LHC*, *Physical Review D* **86** (2012) 092001.

[97] O. Adriani et al., *Transverse-Momentum Distribution and Nuclear Modification Factor for Neutral Pions in the Forward-Rapidity Region in Proton-Lead Collisions at $\sqrt{s_{NN}} = 5.02$ TeV*, *Physical Review C* **89** (2014), no. 6 065209.

[98] O. Adriani et al., *Measurement of Forward Photon Production Cross-Section in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV with the LHCf Detector*, *Physics Letters B* **780** (2018) 233–239.

[99] LHCf and ATLAS, *Measurement of Contributions of Diffractive Processes to Forward Photon Spectra in p-p Collisions at $\sqrt{s} = 13$ TeV*, Technical Report ATLAS-CONF-2017-075, CERN, Geneva, Nov, 2017.

[100] O. Adriani et al., *Measurement of Inclusive Forward Neutron Production Cross Section in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV with the LHCf Arm2 Detector*, *Journal of High Energy Physics* **11** (2018) 073.

[101] O. Adriani et al., *Measurement of the Forward η Meson Production Rate in p-p Collisions at $\sqrt{s} = 13$ TeV with the LHCf-Arm2 Detector*, *Journal of High Energy Physics* **2023** (2023) 169.

[102] O. Adriani et al., *LHCf - Technical Proposal for the LHC Run3*, Technical Report CERN-LHCC-2019-008 ; LHCC-P-014, CERN, 2019.

[103] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, *Hadronic Interaction Model SIBYLL 2.3c and Inclusive Lepton Fluxes*, *Physical Review D* **102** (2020) 063002.

[104] A. Collaboration, *Luminosity Determination in pp Collisions at $\sqrt{s} = 13$ TeV Using the ATLAS Detector at the LHC*, *European Physical Journal C* **83** (2023), no. 10 982.

[105] T. Pierog, C. Baus, and R. Ulrich, "CRMC Web Page." `https://web.ikp.kit.edu/rulrich/crmc.html`.

[106] K. Kasahara, *Introduction to Cosmos and Some Relevance to Ultra High Energy Cosmic Ray Air Showers*, in *Proceedings of the 24th International Cosmic Ray Conference*, vol. 1, p. 399, 1995.

[107] K. Kasahara, "Cosmos and Epics Web Page." `http://cosmos.n.kanagawa-u.ac.jp`. Accessed: 2024-08-20.

[108] R. Workman and O. (Particle Data Group), *Review of Particle Physics*, *Progress of Theoretical and Experimental Physics* **2022** (2022), no. 8 083C01.

[109] R. Brun and F. Rademakers, *ROOT - An Object-Oriented Data Analysis Framework*, *Nuclear Instruments and Methods in Physics Research Section A* **389** (1997) 81–86.

[110] G. Antchev et al., *First Measurement of Elastic, Inelastic and Total Cross-Section at $\sqrt{s}$ = 13 TeV by TOTEM and Overview of Cross-Section Data at LHC Energies*, *Eur. Phys. J. C* **79** (2019), no. 1 1.

[111] A. Collaboration, *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s}$ = 13 TeV with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** (2016) 182002.

[112] L. Collaboration, *Measurement of the Inelastic PP Cross-Section at a Centre-of-Mass Energy of 13 TeV*, *JHEP* **06** (2018) 100.

[113] A. Tiberio et al., *The LHCf Experiment at the Large Hadron Collider: Status and Prospects*, in *Proceedings of the 38th International Cosmic Ray Conference (ICRC2023)*, vol. 444, pp. 1–7, 2023. Presented at the 38th International Cosmic Ray Conference, Nagoya, Japan.

[114] L. Bonechi et al., *Development of the ADAMO Detector: Test with Cosmic Rays at Different Zenith Angles*, in *Proceedings of the 29th International Cosmic Ray Conference (ICRC2005)*, vol. 9, pp. 283–286, 2005. Presented at the 29th International Cosmic Ray Conference, Pune, India.

[115] A. Albertsson et al., *Deep Learning for Calorimeter Energy Reconstruction*, *JINST* **13** (2018) P12003.

[116] K. Pedro et al., *Machine Learning for High Energy Physics Calorimeter Reconstruction*, *EPJ Web Conf.* **150** (2017) 00016.

[117] L. Collaboration, *Deep Learning Approaches for LHCb ECAL Reconstruction*, *JHEP* **06** (2018) 100.

[118] J. H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, *Annals of Statistics* **29** (2001) 1189–1232.

[119] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016) 785–794.

[120] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, *CatBoost: UnbIased Boosting with Categorical Features*, *Advances in Neural Information Processing Systems* **31** (2018) 6638–6648.

Acknowledgements

First of all, I would like to thank, on behalf of the LHCf collaboration, the institutions and people who made the work discussed in this thesis possible: we thank the CERN staff and the ATLAS collaboration for their essential contributions to the successful operation of LHCf. We are grateful to S. Ostapchenko for useful comments about QGSJET II-04 generator and to the developers of CRMC interface tool for its implementation. This work was supported by several institutions in Japan and in Italy: in Japan, by the Japanese Society for the Promotion of Science (JSPS) KAKENHI (Grant Numbers JP26247037, JP23340076) and the joint research program of the Institute for Cosmic Ray Research (ICRR), University of Tokyo; in Italy, by Istituto Nazionale di Fisica Nucleare (INFN), by the University of Catania (Grant Numbers UNICT 2020-22 Linea 2) and by the Centro Siciliano di Fisica Nucleare e Struttura della Materia (CSFNSM). This work was partially supported by the ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union—NextGenerationEU—CUP F77G22000120006. This work took advantage of computer resources supplied by ICRR (University of Tokyo), CERN and CNAF (INFN).

To begin my personal thanks, I would like to express my deepest gratitude to my supervisor, Professor Alessia Tricomi, for her unwavering support even in difficult times, her guidance through which I have had a rare opportunity for professional and personal growth, and above all for all that I have been able to learn from her over the years, both directly

and by observing and being inspired by the way she carries out her work.

I am immensely grateful to all the members of the LHCf experiment for their collaboration and support, both professionally and humanely. Their collective expertise and dedication have been invaluable for this thesis work. Being part of this experiment is a source of great pride for me, and I am deeply grateful for the experiences it has allowed me to have. My time in Florence was made both productive and enjoyable thanks to the warm hospitality of my colleagues there, both from the LHCf but in general from all cosmic ray group components. In particular, I extend my heartfelt thanks to Professors Oscar Adriani, Lorenzo Bonechi, Raffaello D'Alessandro and Sergio Ricciarini, for their precious teachings and their closeness during my periods abroad. Alessio Tiberio and Eugenio Berti for making the day-to-day work pleasant, yet productive, and for their closeness in both happy and difficult moments. Pietro Betti for sharing the PhD experience together, and for broadening my horizons, professionally and otherwise. Sebastiano Detti and Monica Scaringella for their explanations on the wonders of electronics, and more. I am also deeply thankful to my Japanese colleagues for their warm welcome and hospitality during my stay in Nagoya. Their support has been crucial in advancing my research. I would like to express my gratitude to Professors Yasushi Muraki, Takashi Sako, Hiroaki Menjo and Katsuaki Kasahara for all the helpful advice and for strongly contributing to achieving my PhD goals. Moe Kondo, Ken Ohashi, Yuga Kitagami, Kosuke Kinoshita and Haruka Kobayashi for their cooperation, especially during the data-taking periods at CERN.

I would also like to thank the institutions that have most supported my research, namely the "Ettore Majorana" Department of Physics and Astronomy at the University of Catania, the Catania Section of the INFN and the Centro Siciliano di Fisica Nucleare e Struttura della Materia. Thanks go to all the teachers I have had over the years, each of them has played a role in my growth. A special thanks go to those who ignited in me the passion for particle physics, showing me the way and providing me with the tools to embark on this career: Professors A. Tricomi, S. Albergo, F. Riggi, P. La Rocca, C. Petta, V. Bellini, C. Tuvè, M. De Napoli, P.

Sapienza and R. Caruso. I also thank all the administrative collaborators, both from the University and INFN, who in recent years have strongly supported me in the bureaucratic management of research: R. Barbato, S. De Francisci, G. Vicari, A. Magrì, C. Lombardo, L. Romano. Thanks also to the Physics Department of the University of Florence, for hosting me and allowing me to participate in the courses during the years of my PhD.