UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Evaluation of particle identification performance using the upgraded Electromagnetic Calorimeter for the High Luminosity phase of LHC

Thesis supervisor

Prof. Donatella Lucchesi

Thesis co-supervisor

Dr. Lorenzo Sestini

Dr. Davide Zuliani

Candidate

Lisa Arnone

Academic Year 2023/2024

Contents

Introduction

The Large Hadron Collider (LHC) at CERN is the world's largest particle accelerator, whose goal is the exploration of fundamental physics, significantly contributing to our understanding of the Standard Model. However, to achieve more precise measurements and explore New Physics, the LHC must collect considerably more data, which has led to the High Luminosity project (HL-LHC) aimed at increasing its luminosity. This upgrade, while essential, presents two significant challenges:

- The elevated radiation levels expected during the High-Luminosity LHC (HL-LHC) phase will surpass the tolerance of the current detectors, necessitating their replacement to avoid irreversible damage.
- The rise in pile-up, with multiple proton-proton collisions occurring simultaneously, will complicate particle identification, making it more challenging to distinguish between different types of particles accurately.

These challenges are particularly relevant to the LHCb experiment, one of the four major experiments at the LHC. LHCb, with its unique forward acceptance range, plays a critical role in probing the electroweak sector of the Standard Model, investigating phenomena in a previously unexplored region. LHCb has achieved significant results, contributing to the measurements of inclusive vector boson production, forward-backward asymmetry, W boson mass measurements, and the determination of the Weinberg angle [22]. While these results are already competitive with those from other experiments, there remains considerable potential for further improvements, which are expected to be realized with the so-called Upgrade II, indicating the last step of the transition to the HL-LHC phase.

This thesis focuses on the Upgrade II of the LHCb, specifically on the development and implementation of advanced particle identification (PID) algorithms designed for the Electromagnetic Calorimeter (ECAL) in its new configuration. The challenges posed by increased pile-up and radiation dose demand greater efforts to preserve PID accuracy. To address these challenges, the thesis explores the integration of new technologies, including the Spaghetti Calorimeter (SpaCal) and the addition of timing information, which will enhance the performance of the ECAL modules through their longitudinal segmentation, which allows for a double-sided readout.

As part of this work, a detailed study of a SpaCal prototype was conducted using data from test-beam campaigns to evaluate its performance, particularly in terms of energy and time resolution, as well as its effectiveness in electron-pion discrimination. Additionally, the impact of the high-luminosity environment on Z boson production in electron channels is investigated through simulations, focusing on the challenges posed by the increased pile-up.

The thesis is structured in the following way:

1. Chapter 1: The theoretical framework, i.e. the Electroweak theory, is presented together with an overview of the state of the art of experimental results by LHCb.
2. Chapter 2: Here the LHC accelerator and the LHCb experiment are described, with the characterization of its several sub-systems, with particular attention to the electromagnetic calorimeter and its performance in energy reconstruction and charged particle identification.
3. Chapter 3: In this section, the LHCb detector in its upgraded configuration is presented, with

a particular focus on the Electromagnetic Calorimeter (ECAL) whose changes are described thoroughly together with their motivations. A detailed description of the new modules for the electromagnetic calorimeter is also presented.

4. Chapter 4 focuses on the development of the particle identification algorithm. The simulation framework is exploited to study both electron and pion reconstruction of the upgraded ECAL. Its performance in electron and pion discrimination is evaluated using a Boosted Decision Tree [31, 32] classifier.

5. Chapter 5 presents the test-beam analysis of a SpaCal prototype. Here the geometry of the prototype is described, together with the simulation procedure and test-beam results. Energy and time resolution are studied.

6. Chapter 6 is devoted to the electroweak analysis of Z vector boson production in the di-electron final state, using simulated samples in the upgraded configuration. The study explores the impact of the new high-luminosity environment on Z boson reconstruction, efficiency, and invariant mass resolution.

Chapter 1

Electroweak physics at LHCb

This Chapter aims to present an overview of electroweak physics, and the current state of LHCb measurements. The possibility of the LHCb experiment to contribute to these measurements will be highlighted as well as the current reconstruction performance of the detector, with particular attention to electron reconstruction with the Electromagnetic calorimeter (ECAL). Finally, the limitations of the current detector will be presented, with an opening to the possibilities that a future upgrade can bring.

In particular, this Chapter will be organized as follows:

- Section 1.1 will give an overview of the Standard Model of Particle Physics
- Section 1.2 goes into detail about the electroweak sector and describes the global electroweak fit
- Section 1.3 will present the capabilities and results of LHCb in terms of electroweak measurements
- Section 1.4 will describe the current LHCb limitations and the measurements that would benefit from a future upgrade.

1.1 The Standard Model of Particle Physics

The Standard Model (SM) [28,35] is the current theory that describes particles and their interactions. Like any model of particle physics, it should be formed by a gauge group, a particle content, and the possibility of Spontaneous Symmetry Breaking (SSB) into a smaller group, and it has to be renormalizable, meaning that the quantities that the theory predicts should be finite.

The Standard Model is a Quantum Field Theory represented by the gauge symmetry group

$$G = SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \tag{1.1}$$

where each group describes a different interaction between particles which is mediated by $N^2 - 1$ generators, where N is the dimension of the group:

- $SU(3)_c$ describes the strong force, and its generators are the eight gluons. Gluons mediate the interactions between quarks, which carry the color charge in Quantum Chromodynamics (QCD)
- $SU(2)_L$ describes the isospin invariance of the weak interactions
- $U(1)_Y$ is the gauge group of Quantum Electrodynamics (QED) and defines the local phase invariance of electromagnetic interactions

The last two groups combine to form the Electroweak theory [34], with $W^{\pm}$, Z, and γ bosons as mediators.

The particle content of the Standard Model includes many classes of elementary particles, grouped by different characteristics as shown in Fig. 1.1. The first subdivision is of matter and forces, respectively of fermions and bosons. Fermions are spin 1/2 particles and can be divided into leptons and quarks based on their carried charge and how they interact. Vector bosons are spin 1 particles that behave as the mediator of forces described by the Standard Model: electromagnetic, weak, and strong forces, respectively mediated by γ, Z and W, and gluons. There is also a scalar boson of spin 0, the Higgs boson, which gives masses to all the elementary particles after symmetry breaking [25].

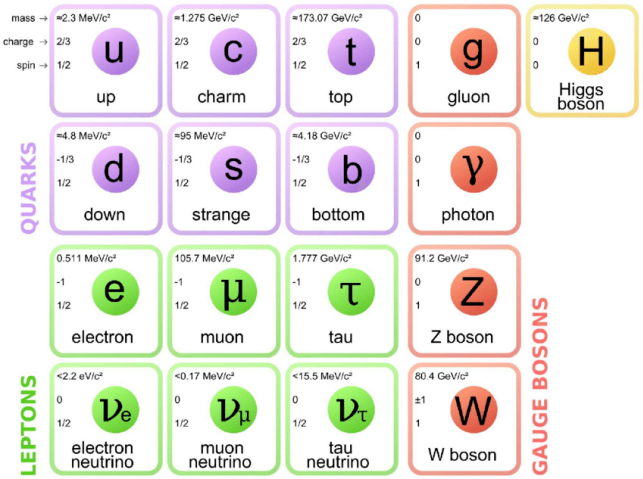

Figure 1.1: Standard Model particle content, where elementary particles are grouped into subcategories based on their characteristics.

1.2 The electroweak sector

Electroweak theory unifies two of the four fundamental forces, the weak and the electromagnetic force. The gauge group is

$$G_{EW} = SU(2)_L \otimes U(1)_Y \tag{1.2}$$

The electroweak theory is described by a Lagrangian $\mathcal{L}$, which can be decomposed into three main components:

$$\mathcal{L} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_f + \mathcal{L}_{\text{Higgs}} \tag{1.3}$$

Here, $\mathcal{L}_{\text{gauge}}$ describes the interactions through vector bosons, $\mathcal{L}_f$ describes the fermion content, and $\mathcal{L}_{\text{Higgs}}$ incorporates the Higgs mechanism, through which Standard Model particles acquire mass.

- The gauge Lagrangian, $\mathcal{L}_{\text{gauge}}$, is given by:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} W^i_{\mu\nu} W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \qquad i = (1,2,3) \tag{1.4}$$

  where $W^i_{\mu\nu}$ and $B_{\mu\nu}$ are the field strength tensors for the $SU(2)_L$ and $U(1)_Y$ gauge fields, respectively. They are defined as:

$$W^i_{\mu\nu} = \partial_\nu W^i_\mu - \partial_\mu W^i_\nu - g_W \epsilon^{ijk} W^j_\mu W^k_\nu$$

$$B_{\mu\nu} = \partial_\nu B_\mu - \partial_\mu B_\nu$$

  where g_W is the coupling constant for $SU(2)_L$, and ϵ^{ijk} is the Levi-Civita symbol.

- The fermion Lagrangian, $\mathcal{L}_f$, describes the coupling between massless bosons and fermions. Fermions are classified into left-handed (Ψ_L) and right-handed (ψ_R) components. Left-handed fermions form doublets under $SU(2)_L$, while right-handed fermions are singlets. The fermion Lagrangian is expressed as:

$$\mathcal{L}_f = \sum_{\text{fermions}} \bar{\Psi}_L \gamma^\mu D_\mu \Psi_L + \bar{\psi}_R \gamma^\mu D_\mu \psi_R \tag{1.5}$$

  with the covariant derivatives D_μ defined as:

$$D_\mu \Psi_L = \left(\partial_\mu + i g_W \frac{\tau^i}{2} W_\mu^i + i g_Y Y B_\mu \right) \Psi_L$$

$$D_\mu \psi_R = \left(\partial_\mu + i g_Y Y B_\mu \right) \psi_R$$

  where τ^i are the Pauli matrices, g_Y is the coupling constant for $U(1)_Y$, and Y is the weak hypercharge.

Upon symmetry breaking, the components of the Higgs field yield mass terms for the gauge bosons and fermions. The masses of the gauge bosons are derived as:

$$m_W = \frac{1}{2} \nu g_W$$

$$m_Z = \frac{1}{2} \nu \sqrt{g_W^2 + g_Y^2}$$

where θ_W is the Weinberg angle defined by $\cos\theta_W = \frac{g_W}{\sqrt{g_W^2+g_Y^2}}$ and ν is the Vacuum Expectation Value (vev) of the Higgs field.

The massless photon A_μ, and the massive $W^\pm$ and Z bosons emerge from the mixing of the original gauge fields:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{1.6}$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \pm i W_\mu^2) \tag{1.7}$$

This result in a constant relation between W and Z boson masses, which is presented by introducing the parameter ρ:

$$\rho = \frac{m_W^2}{m_Z^2 \cos^2\theta_W} = 1 \tag{1.8}$$

at tree level. An approximate global symmetry, known as custodial symmetry, protects this relation. It arises by the Higgs mechanism which, after electroweak symmetry breaking, preserves a residual global $SU(2)$ symmetry. This symmetry is indeed broken by hypercharge $U(1)_Y$ interactions and Yukawa couplings, but their effects only appear at the loop level and correspond to sub-percentage deviation from unity.

The ρ parameter is particularly important in the Standard Model, as it is highly sensitive to the symmetry structure of the Higgs sector, and if it were to deviate from 1, it would indicate the breaking of the custodial symmetry and point to interactions beyond the Standard Model [37].

1.2.1 The electroweak fit

As explained in the previous section, in the realm of particle physics, the electroweak theory unifies two of the four fundamental forces, namely the weak and the electromagnetic force. In order to test Standard Model parameters within this sector, precision measurements are needed. In fact, there are 26 fundamental parameters that cannot be derived by theoretical predictions and must be set by experimental values.

The Global Electroweak Fit is a powerful method used to test the consistency of the Standard Model by comparing theoretical predictions with a wide range of precision experimental measurements.

Since all these parameters are related to each other by theory, such as M_W, M_Z, and $\sin\theta_W$ which are related by eq. 1.8, the prediction of one observable is obtained by giving the measurement of all the others as input to the electroweak fit. This value is then compared to the measured value of that observable, and their discrepancy is used to evaluate the goodness of the model. Therefore, changing one observable, such as M_W, can introduce new tensions in the fit, potentially revealing inconsistencies or pointing to New Physics.

Figure 1.2 shows the current deviation of SM observables from their average fit values [19]. Some observables stand out, such as the forward-backward asymmetry measured by LEP or SLD, or the W boson mass. Recently, M_W was measured by CDF [14] with a very small uncertainty of 9.4 MeV and deviates by 7σ from the electroweak fit. This measurement was not taken into account in the plot shown.

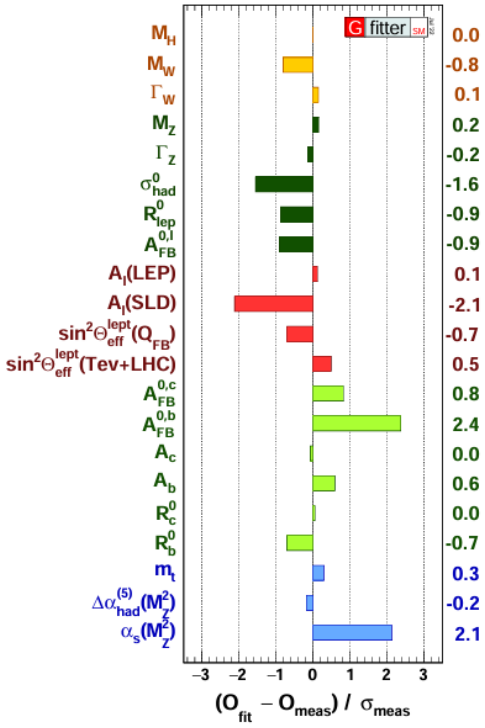

Figure 1.2: Deviation of measured values from average fit value divided by the measured uncertainty [19].

1.3 LHCb contribution to Electroweak physics

While there are already many experiments that are dedicated to this task, also LHCb can partake in electroweak measurements.

The detector apparatus [9] is a single-arm spectrometer with an angular coverage from 10 to 300 mrad and it extends longitudinally along the beam direction. Its pseudorapidity coverage is $2 < \eta < 5$

(where $\eta = -\ln(\tan(\theta/2))$ and θ is the polar angle with respect to the beam direction), complementary to the one of ATLAS and CMS.

Therefore, LHCb can study the regions of high x and low x and high Q^2, where x is the momentum fraction of the parton with respect to the proton momentum and Q^2 is the squared energy transfer during the scattering. Figure 1.3 shows those regions covered by LHCb compared to other experiments, it is noticeable that the region at low x is completely unexplored and that constitutes a great asset for LHCb. In fact, measurements in this region can be anti-correlated to the ones obtained by other experiments, contributing to lower Parton Distribution Function (PDF) uncertainties, as for the M_W described in the next section.

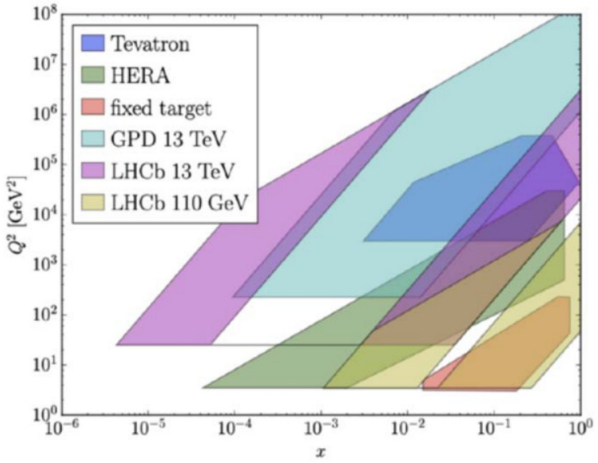

Figure 1.3: x - Q^2 plane showing the region covered by several experiments. LHCb acceptance is shown in violet, ATLAS, and CMS acceptance is included in the light blue region 'GPD'. [11]

1.3.1 LHCb EW Measurements

LHCb contributes to electroweak measurements by studying the production of W and Z bosons in hadronic and leptonic decay channels. This plays an important role in advancing our understanding of fundamental particles and their interactions due to their relation with some other important quantities such as lepton charge asymmetry or $\sin\theta_W$.

The weak mixing angle $\sin^2\theta_W^{\text{eff}}$ is measured through the forward-backward asymmetry with respect to the muon invariant mass exploiting $Z \to \mu^+\mu^-$ decays [4]. The differential cross section at leading order is expressed as:

$$\frac{d\sigma}{d\cos\theta^*} = A(1+\cos^2\theta^*) + B\cos\theta^*$$

Here, θ^* represents the polar angle of the negatively charged lepton in the Collins-Soper frame [16], where the z-axis is aligned with the difference between the incoming proton momentum vectors in the dimuon rest frame. A and B are coefficients that vary with the dimuon invariant mass, in particular, B depends on $\sin^2\theta_W$ and is linked to the forward-backward asymmetry, A_{FB}, which is defined as: $A_{FB} = \frac{N_F - N_B}{N_F + N_B}$ where N_F and N_B are respectively the numbers of forward and backward decays. Figure 1.4 shows the measurements of A_{FB} as a function of the dimuon invariant mass obtained at $\sqrt{s} = 8$ TeV.

A_{FB} is therefore used to measure $\sin\theta_W$. Lastly, $\sin^2\theta_W^{\text{eff}}$ is defined in terms of the ratio between the vector and axial-vector couplings of the Z boson to the involved fermions, and is proportional to $\sin^2\theta_W$.

Figure 1.5 shows the current measurements of $\sin^2\theta_W^{\text{eff}}$ compared to the value from electroweak fit. The biggest uncertainty of this measurement is given by the theoretical (PDF) uncertainty. The forward

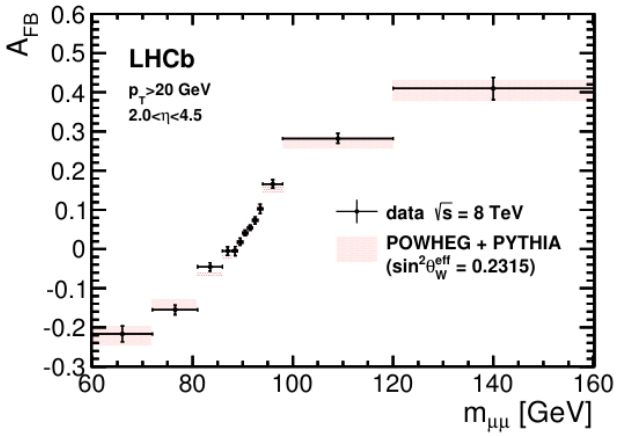

Figure 1.4: A_{FB} measurement and Standard Model predictions for data obtained at $\sqrt{s} = 8$ TeV [4].

direction gives the smallest PDF uncertainty, and, from this perspective, LHCb contribution to the measurement is valuable.

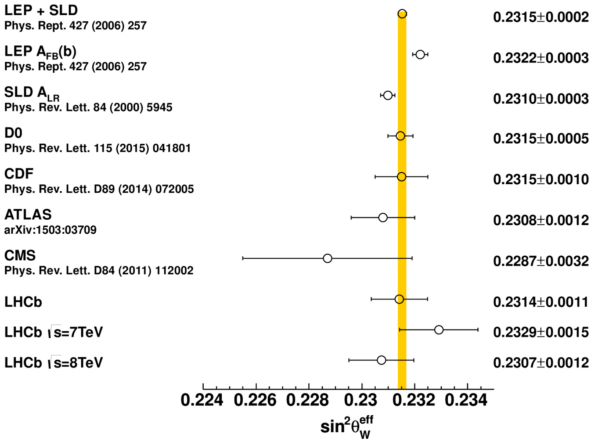

Figure 1.5: The measured value of $\sin^2\theta_W^{\text{eff}}$ by LHCb compared to measurements from other experiments (points with error bars) and from electroweak fit (yellow band) [4].

In addition, LHCb involvement in the measurement of the W mass is of great importance in the physics landscape, since the recent measurement of CDF seems to deviate from the expected value of the electroweak fit. Also, LHCb measurement has an anticorrelation of 63% from the ATLAS and CMS results, resulting in a partial cancellation of the PDF uncertainty. The current measurements, together with the LHCb measurement, are shown in Fig. 1.6.

The most recent measurement of M_W at LHCb exploits data from $W \to \mu\nu$ obtained in 2016 at $\sqrt{s} = 13$ TeV, corresponding to 1.7/fb of integrated luminosity [26]. LHCb result is compatible with the electroweak fit, and the statistical uncertainty of 32 MeV is expected to reach $\sim$ 20 MeV using the full Run 2 dataset.

1.4 LHCb limitations and future prospects

While LHCb is able to contribute to the measurement of electroweak observables, as described in the previous section, the detector has to face some limitations. In the analysis of W or Z boson production in lepton channels, only the channels with muon final states were used to contribute to the electroweak fit precision measurements. In fact, LHCb has a very good reconstruction performance for muon final states, while not the same could be said for electron final states.

Figure 1.7 shows the reconstructed Z invariant mass in dimuon or dielectron final states. This result

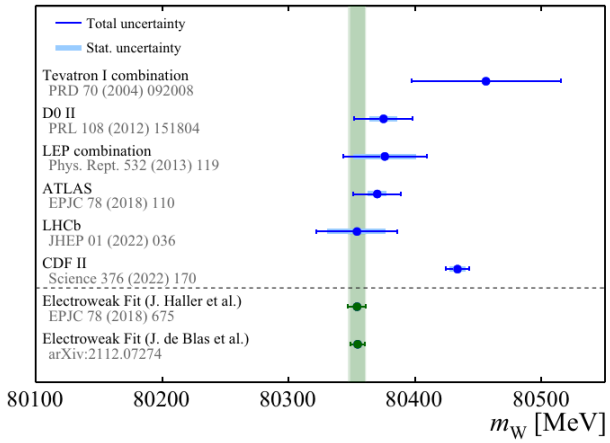

Figure 1.6: The measured value of M_W by LHCb compared to measurements from other experiments (blue) and the EW fit (green) [26].

is produced by a muon sample with 99.2% purity, while for electrons, only 92.2% purity is obtained [5]. While the invariant mass obtained with muon pairs can reconstruct a well-defined peak, the distribution obtained with electrons is very spread out, with FWHM/2 $\sim$ 16 GeV.

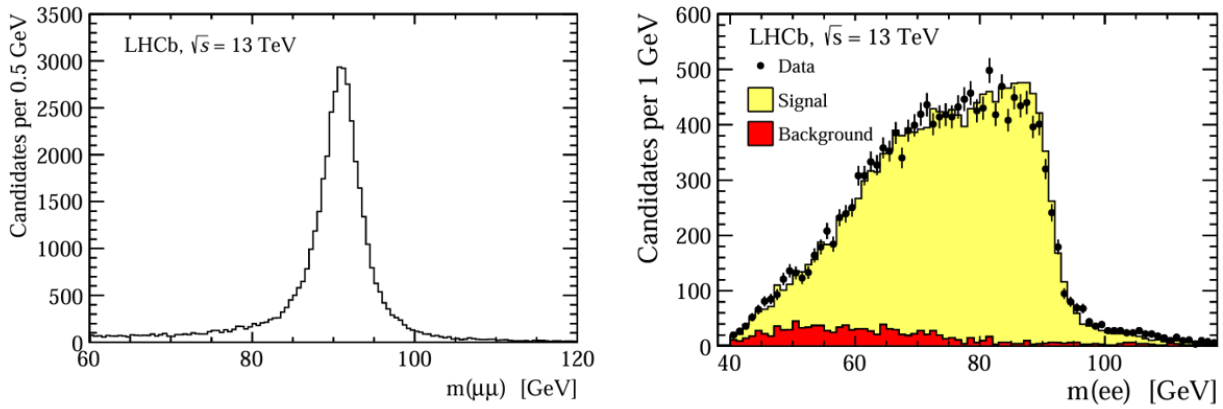

Figure 1.7: Invariant mass distribution of Z boson in dimuon $\mu^+\mu^-$ final state (left) or dielectron e^+e^- final state (right) obtained from $\sqrt{s} = 13$ TeV LHCb data corresponding to integrated luminosity of 294 ± 11 pb^{-1}. [7]

The principal cause of that is the electron reconstruction limitation of the current Electromagnetic Calorimeter. At the moment, its performance is limited to the ADC dynamic range, which causes calorimeter cells to saturate at a transverse energy of 20 GeV. Figure 1.8 shows the transverse momentum distribution for Z boson production in an electron-positron pair. The fiducial cut to remove background contributions given by same-sign candidates is $p_T > 20$ GeV, while a cut on the energy over momentum distribution, E/p, is performed to remove the hadronic background. However, this requirement strongly limits the statistics, resulting in poorer reconstruction efficiency with respect to muon candidates.

Given the possibilities of the LHCb measurements on electroweak physics, LHCb aims to improve its electron reconstruction precision in order to get closer to the one that it has for muons. If it were to use both electron and muon channels in its measurement, with the same reconstruction precision, would increase statistical sensitivity by a factor of $\sqrt{2}$ and lower systematic and statistical uncertainties.

In the next years, the LHC will start a high luminosity program (HL-LHC) where the event rate will increase significantly [6]. LHCb aims at increasing its total dataset from the current 9/fb to 300/fb of integrated luminosity, leading to enhanced measurement precision. This advancement comes with challenges, and LHCb intends to prepare to upgrade its detectors as well as its trigger system. The issue is addressed in detail in Chapter 2 and 3.

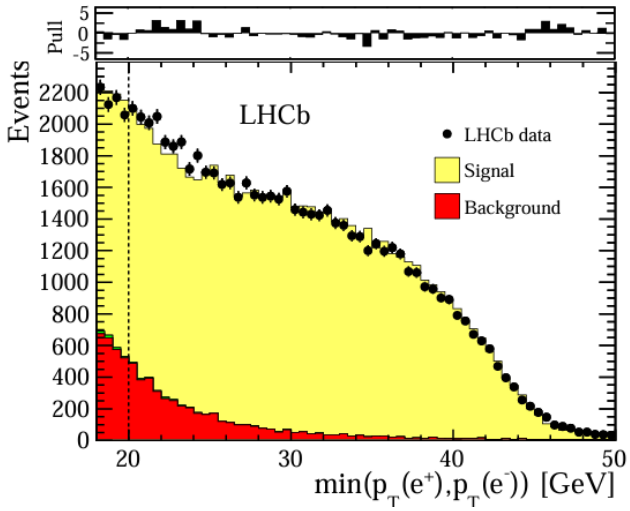

Figure 1.8: Comparison between data and simulation for the distribution of the minimum electron p_T obtained by $Z \to e^+e^-$ production at $\sqrt{s} = 8$ TeV [7].

The impact of an upgraded ECAL, specifically the one that will be described in Chapter 3, on electroweak measurements is expected to be great.

Statistical uncertainty of the W mass measurement mentioned above is expected to reach 10 MeV , as opposed to the current 32 MeV, with more statistic of proton collisions and including $W \longrightarrow e\nu$ channel.

$\sin^2\theta_W^{eff}$ result is statistically limited, therefore it would benefit both from the increasing integrated luminosity and from the addition of decay channels with electron final states, reaching a total uncertainty of $\sim$10% of the current value.

Chapter 2

The LHCb experiment

This Chapter presents an overview of the LHC machine, as well as a more detailed description of the LHCb experiment detector, with particular care to its Electromagnetic Calorimeter.

2.1 LHC overview

The Large Hadron Collider (LHC) [21] is the world's largest particle accelerator, with a circumference of 26.7 km. It operates at a depth of approximately 100 meters underground near Geneva, on the border between France and Switzerland. Constructed in 10 years on the site of the previous Large Electron-Positron (LEP) collider, the LHC has been in operation since 2008.

The LHC structure consists of two vacuum pipes where proton beams are accelerated in opposite directions. These pipes cross at four interaction points, hosting major experiments. The proton beams in the LHC consist of 2808 bunches, each containing about $1.5 \cdot 10^{11}$ protons. The bunches are spaced approximately 25 nanoseconds apart, with each bunch being around 30 cm long and having a transverse dimension of about 1 mm. At the interaction points, the bunches are squeezed to a transverse dimension of 16 μm, allowing high-frequency collisions at 40 MHz.

To reach the required center-of-mass energy, the acceleration process occurs in several stages. Protons extracted from hydrogen ionization are initially accelerated in a linear accelerator (Linac2) to 50 MeV, then further in the Proton Synchrotron Booster (PSB) to 1.4 GeV. Next, they are boosted to 28 GeV in the Proton Synchrotron (PS) and to 450 GeV in the Super Proton Synchrotron (SPS), the final stage before entering the LHC. Once inside the LHC, protons are further accelerated by 16 superconducting radio-frequency (RF) cavities housed in cryomodules, which maintain a stable electromagnetic field at extremely low temperatures of 1.9 K using liquid Helium. This setup allows the protons to reach energies of up to 7 TeV each, resulting in a center-of-mass energy of $\sim$ 14 TeV.

The primary experiments are four: ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid), ALICE (A Lead Ion Collision Experiment), and LHCb. ATLAS and CMS are general-purpose detectors designed for a wide range of experiments, including studying Higgs physics and exploring physics beyond the Standard Model. ALICE focuses on heavy-ion collisions and quark-gluon plasma studies, while LHCb specializes in forward-region collisions, focusing on the physics of b and c quarks.

The LHC's performance is measured by its luminosity L, defined as $L = N/\sigma$, where σ is the collision cross-section and N is the number of events per second generated by the collision. The LHC has a current peak luminosity of $2 \cdot 10^{34}$ cm^{2}s^{-1}.

2.2 LHCb detector

The LHCb experiment [9] is designed to study the behavior of b and c hadrons in the forward region of proton-proton collisions at the Large Hadron Collider (LHC). This unique single-arm spectrometer is positioned at LHC Point 8 and has an angular coverage ranging from 10 to 300 (250) mrad in the bending (non-bending) plane corresponding to a pseudo-rapidity (η) range between 2 and 5.

The LHCb detector is 21 meters long, 10 meters high, and 13 meters wide, with a total weight of about 5600 tons. It extends longitudinally along the z-axis. The detector comprises several key sub-detectors and systems for tracking and particle identification (PID) whose schematic representation is shown in Fig. 2.1 [15].

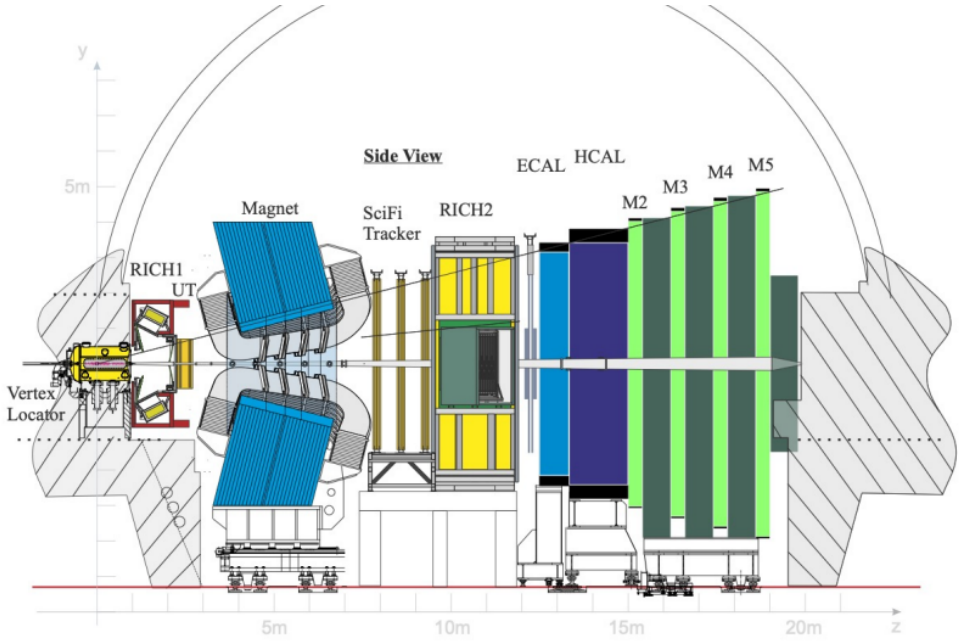

Figure 2.1: Schematic representation of the LHCb experiment in its Upgrade I configuration, showing each sub-detector. The z-axis runs parallel to the beam direction while the y-axis is parallel and opposite to the gravitational acceleration.

The Vertex Locator (VELO) is the innermost sub-detector, dedicated to vertex reconstruction. Beyond the VELO, there are four tracking stations: the upstream tracker (UT), located upstream of the magnet in the direction of increasing z, and three downstream stations (SciFi Tracker). Particle identification is achieved with two Ring Imaging Cherenkov (RICH) detectors, one upstream and one downstream of the magnet. The calorimeter system includes an Electromagnetic Calorimeter (ECAL), and a Hadronic Calorimeter (HCAL). Additionally, a muon system with four stations (M2, M3, M4, M5) alternating with iron shielding is employed for muon identification.

Unlike other LHC experiments, LHCb operates with reduced instantaneous luminosity through a technique known as luminosity leveling [27]. This technique is used to maintain a constant luminosity of approximately $2 \cdot 10^{33}$ cm^{2}s^{-1} by tilting the proton beams to increase the interaction area. The lower luminosity helps to reduce pile-up (the average number of proton-proton collisions per bunch crossing), providing cleaner collision events. The nominal pile-up for LHCb is kept around 5-6, compared to higher values in other LHC experiments such as ATLAS and CMS.

The schedule of activity of LHC is marked by periods of data acquisition and periods of stop, namely Long Shutdowns (LS), which are exploited to make changes to various components of the experiments. LHCb timeline follows the one of LHC and is shown in Fig. 2.2. The data-taking periods are called "Runs". In total, there are 6 Runs and 5 Long Shutdowns. Structural adjustments or repairs to detector components are made during those periods of stop. In particular, Run 3 and Run 4 belong to the so-called Upgrade I, while Run 5 and Run 6 belong to Upgrade II. The motivation and the

description of these changes is explained in the next chapter.

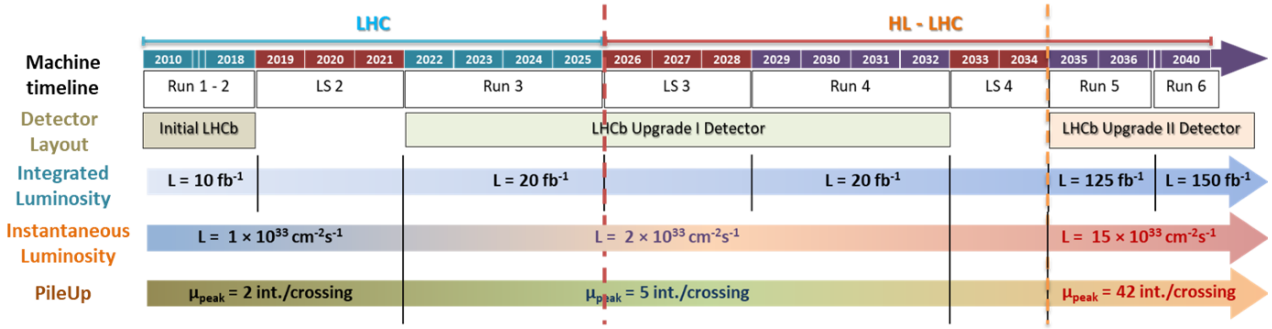

Figure 2.2: LHCb timeline with information on detector upgrades, luminosity, and pile-up. Vertical lines mark the evolution of the experiment in accordance with the LHC luminosity plan [23].

In the next section, there will be a detailed description of all the sub-detectors.

2.2.1 VErtex LOcator

The VErtex LOcator (VELO) is a critical sub-detector in the LHCb experiment. Its function is crucial for reconstructing the location of pp collision and distinguishing primary from secondary vertices, which is fundamental for flavour tagging of heavy-flavour hadrons.

It was completely replaced during Long Shutdown 2 (LS2) in 2022 with a new structure and technology (VeloPix).

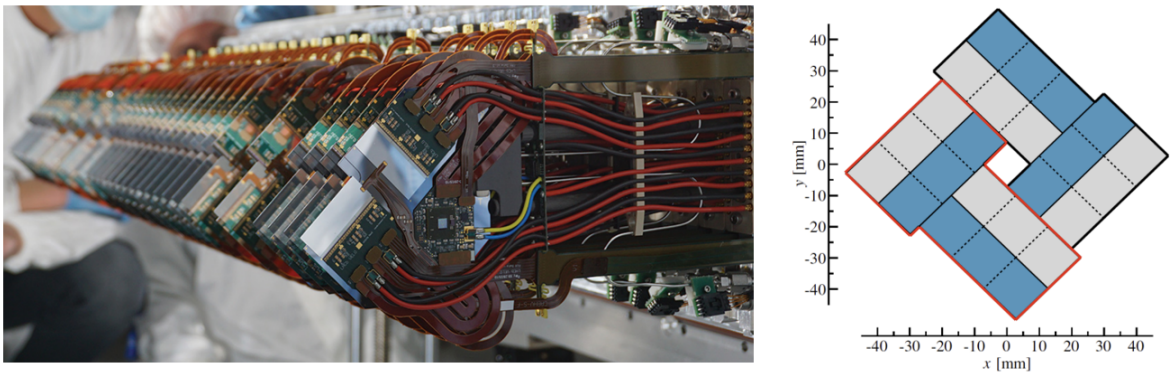

Figure 2.3: Perspective view of one-half of installed VELO modules (left). Schematic front view of the chips of VELO detector in the x-y plane in the closed configuration. The contour in red shows the components of the same module. (right)

The new VELO is located close to the interaction point and consists of 52 silicon modules arranged in two retractable halves, forming a parallelepiped-shaped structure that extends along the beam direction which is shown in Fig. 2.3. These modules consist of pixelated hybrid silicon detectors, different from the old $r - \phi$ strips.

The two halves can be remotely moved along the x-axis to allow beam injection and prevent damage during unstable beam conditions, while, for stable beam conditions, the modules are positioned close to the beamline, providing high spatial resolution. The silicon sensors are kept in a vacuum environment at -20 °C to minimize radiation damage. This system contributes to high tracking efficiency, with charge collection efficiency greater than 99% after 4 MGy.

2.2.2 Magnet

In the LHCb experiment, the momentum of charged particles is determined by measuring their curvature radius in a magnetic field. The relationship between the curvature radius and momentum is

governed by the Lorentz force, described by the equation:

$$p = qrB$$

where p is the momentum, q is the charge of the particle, r is the curvature radius, and B is the magnetic field strength.

This measurement is facilitated by a room-temperature dipole magnet placed between the Upstream Tracker (UT) and the first tracking station. The magnet, weighing approximately 1600 tons, consists of aluminum conical coils symmetrically positioned in the magnet yoke, with a unique shape designed to create slopes that cover the full detector acceptance. It has a 4 Tm bending power and generates a dipolar magnetic field in the y direction, allowing for an angular acceptance of ±250 mrad in the vertical plane and ±300 mrad in the horizontal plane.

The magnetic field is inverted regularly by reversing the direction of the current in the coils, allowing for the correction of potential misalignments in the detector.

2.2.3 Tracking chambers

The LHCb experiment employs a comprehensive tracking system alongside the Vertex Locator (VELO) to identify particle tracks and reconstruct their trajectories. This system consists of four planar tracking modules which include the Upstream Tracker (UT) and three additional tracking stations, T1, T2, and T3.

The UT module is positioned between RICH1 and the magnet and covers the entire geometrical acceptance of the LHCb detector. To achieve high purity, high hit efficiency, and resistance to radiation damage, a silicon microstrip technology was chosen. It consists of two substations, UTa and UTb with a gap of 205 mm in the z-axis. Each sub-station has a pair of layers, placed 55 mm from one another, with specific tilt orientations: UTa has X- and U-layers, while UTb has V- and X-layers, where V and U have strips inclined by 5° in the opposite directions. This design provides a single-hit spatial resolution of around 50 µm.

The other trackers are placed downstream of the magnet. They cover the entire LHCb acceptance range and consist of scintillating plastic fibers arranged in multilayered mats; hence the name SciFi Tracker. Each station has four layers displaced in an X, U, V, X configuration.

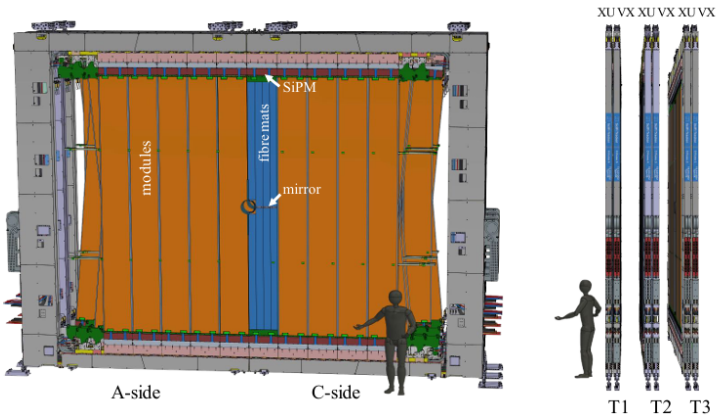

Figure 2.4: Schematization of the front and side view of the Scintillating Fiber Tracker detectors.

2.2.4 Ring Imaging Cherenkov Detectors

To identify particles like pions and kaons produced by the proton-proton collisions, two Ring Imaging Cherenkov (RICH) detectors play a critical role. RICH1 and RICH2 are positioned at different locations along the z-axis to cover varying momentum ranges and angular acceptances.

RICH1 is located upstream of the magnet, positioned after the VELO and before the UT station. It primarily detects low-momentum particles with a momentum range of about 2.6–60 GeV/c. RICH1 has a broader angular acceptance, covering from 25 to 300(250) mrad in the horizontal (vertical) plane. RICH2 is located downstream of the magnet and it has a reduced acceptance from 10 to 120(100) mrad. This limited angular range allows RICH2 to focus on particles with higher momenta up to 100 GeV/c.

Both RICH detectors employ a combination of spherical mirrors to focus the Cherenkov light and planar mirrors to redirect the light to the multi-anode photomultiplier tubes (MaPMTs) which replace the original Hybrid Photon Detectors (HPDs). They convert the collected light into electrons, forming an electric signal that can be analyzed to determine the properties of the particles.

The cone angle of Cherenkov radiation is related to the particle's velocity through the equation:

$$\cos(\theta_c) = \frac{c/n}{v}$$

where θ_c is the cone angle, c is the speed of light, v is the particle's velocity, and n is the refraction index of the radiator. By measuring the momentum p of the particle through the curvature of its trajectory and the cone angle θ_c, the mass m of the particle can be calculated to distinguish between different types of charged hadrons.

2.2.5 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) plays a crucial role in particle identification by measuring particle (electrons and photons) energies.

ECAL is positioned approximately at 12.5 m downstream from the interaction point. With respect to the previous configuration, two elements are removed, the Scintillating Pad Detector (SPD) and the Preshower (PS) which worked mainly as the trigger, while ECAL was left unchanged.

The current ECAL uses Shashlik [10] technology, a type of sampling calorimeter consisting of 66 alternating layers of 2 mm lead tiles and 4 mm scintillator tiles, separated by 120 µm of reflective paper. This creates a total depth of 42 cm, which is equivalent to 25 radiation lengths, allowing for efficient absorption and conversion of electromagnetic energy. The signal readout exploits wavelength-shifting fibers (WSL) to collect scintillation light and direct it to photomultipliers (PMTs), where it is converted into electrical signals. Figure 2.5 shows the front view of ECAL while fig. 2.6 shows a schematization of a single Shashlik module and a picture of three modules with different granularity.

The ECAL has variable granularity, divided into three regions: inner, middle, and outer. The inner region consists of cells measuring 4 x 4 cm^2, the middle region has cells of 6 x 6 cm^2, and the outer region has cells of 12 x 12 cm^2 which is shown in Fig. 2.7. This arrangement provides higher resolution and granularity where particle occupancy and radiation dose are higher i.e. near the beamline.

The energy resolution for a calorimeter is defined as the function:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.1}$$

where a is the sampling term, b is the noise term, and c is the constant term.

For ECAL, this is:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\% \tag{2.2}$$

where E is the energy in GeV, $\oplus$ denotes the quadratic sum of the terms, and the constant term on the right is given by structural imperfections in the calorimeter.

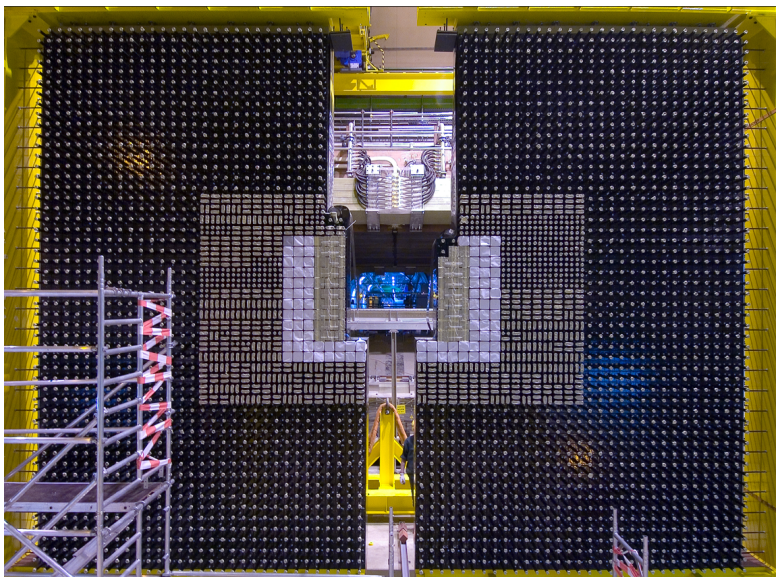

Figure 2.5: Front view of the ECAL with the two halves open where the 3 regions can be seen [12].

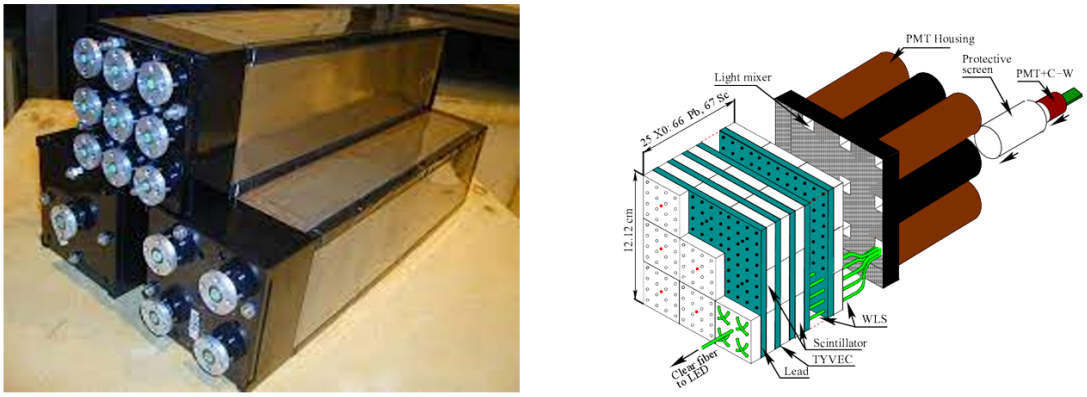

Figure 2.6: The ECAL modules with Shashlik technology with 9, 4, 1 cell per module (left). The schematization of a 9-cell Shashlik module structure (right).

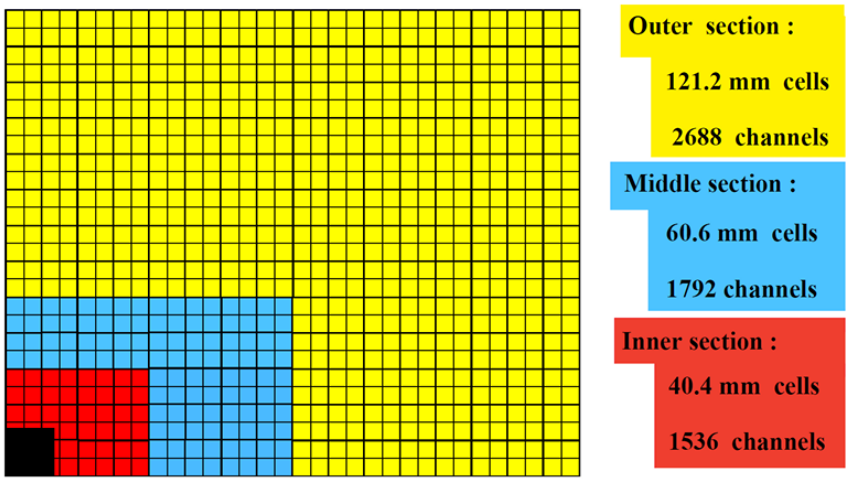

Figure 2.7: Segmentation of ECAL regions.

This energy resolution allows for accurate measurement, though there are limitations for high-energy electrons and photons due to the saturation of the ADC for transverse energies over 20 GeV. This was linked to the original objective of LHCb, so for b-physics, but now this limitation is addressed in the

context of an ECAL upgrade discussed in later chapters of the thesis.

Calorimeter performance Here the current ECAL performance is described in terms of energy resolution and particle identification.

The energy resolution curve of the outer Shashlik modules is shown in Fig. 2.8 where a sampling, constant, and noise term of respectively 9.4 %, 0.8%, and 0.15 GeV were obtained using electron beam [10].

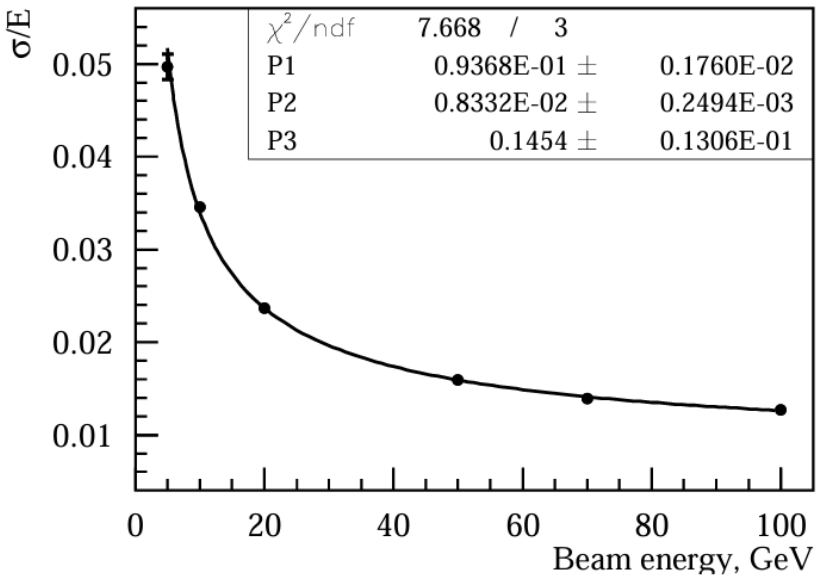

Figure 2.8: Energy resolution plot of outer ECAL modules obtained with an electron beam [8].

Figure 2.9 shows the difference in the distribution between electrons and pions on the quantity of energy over the particle momentum obtained by the 2011 (Run 1) data [8].

The reconstruction efficiency of charged particles with respect to particle momentum at the end of Run 1 is shown in Fig. 2.10 [1] on the left, while on the right side there is the preliminary plot of global particle identification efficiency of Run 3 where the old ECAL performance is compared to the new one with pile-up of 3 [3].

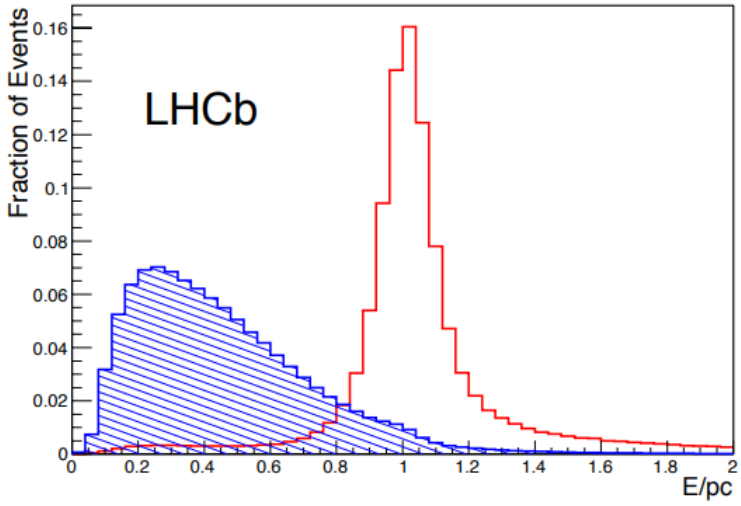

Figure 2.9: Distribution of E/p in the ECAL for electrons (red) and hadrons (blue) using 2011 data. [8]

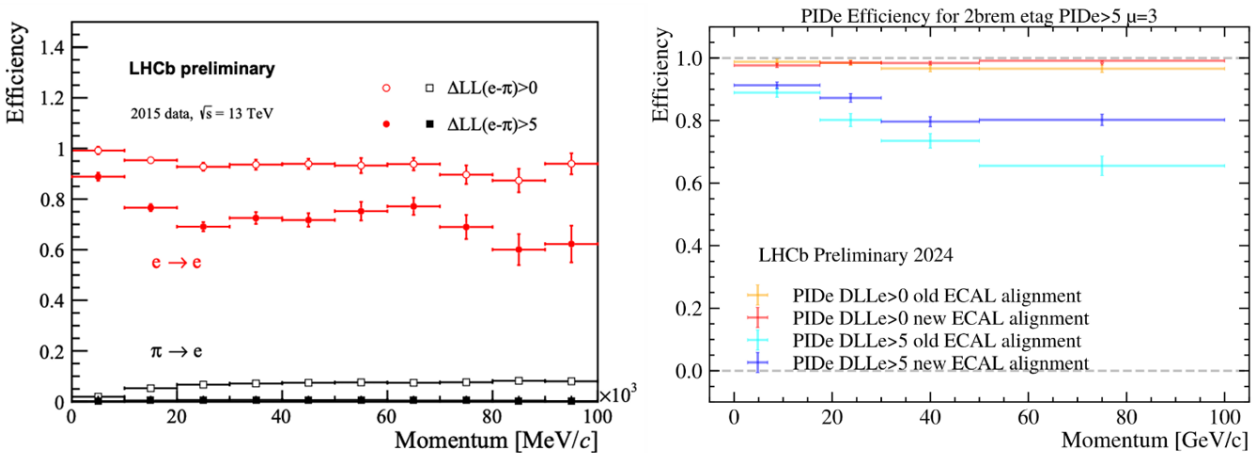

Figure 2.10: Global PID efficiency in electron momentum bins for 2015 data (left) [1]. PID efficiency in probe electron momentum bins for μ average of 3 comparing old and new ECAL alignment deployed in run 291593. Both electrons have a bremsstrahlung photon associated. The selection was done using linear cuts and a BDT trained with an MC simulation (right) [3].

2.2.6 Hadronic Calorimeter

The HCAL is placed downstream of the ECAL, approximately 13.33 m from the interaction point. It is designed to measure the energy of hadrons and to provide stopping power to contain hadronic showers before they reach the muon stations. The HCAL uses a similar structure to the ECAL but with a larger depth to contain hadronic showers, which are more extended than electromagnetic showers.

The HCAL consists of alternating layers of 1 cm thick iron absorber tiles and 4 mm thick polystyrene scintillator tiles, providing a total depth of 128.3 cm (5.6 interaction lengths). Like the ECAL, the HCAL uses WLS fibers to collect scintillation light and transmit it to PMTs for signal conversion. The HCAL has two regions with different granularity: the inner region has cells of $13 \cdot 13$ cm^2, while the outer region has cells of $26 \cdot 26$ cm^2. This arrangement reflects the larger size of hadronic showers compared to electromagnetic showers. This division is shown in Fig. 2.11.

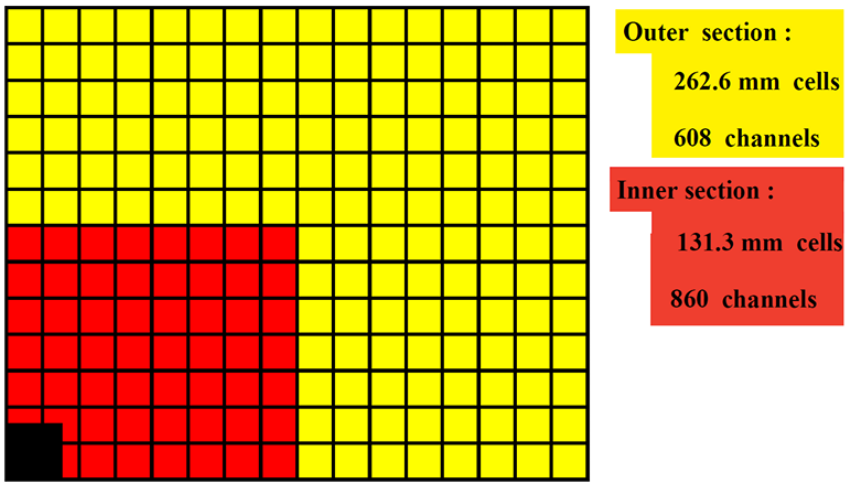

Figure 2.11: Segmentation of the HCAL.

The energy resolution for the HCAL is given by:

$$\frac{\sigma_E}{E} = \frac{(69 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\%$$

where again E is the energy deposited in GeV.

Currently, the possibility of removing HCAL and replacing it with iron shielding during LS3 has been investigated. This choice would impact PID performance in electron-hadron separation, especially for electrons at high transverse momenta, significantly increasing the background. For this reason, several ways to better reconstruct electrons, involving Machine Learning, are being studied. In fig. 2.12 the reconstruction of the D^0 invariant mass in an electron-positron pair is shown with its hadronic background. With 50% electron reconstruction efficiency, almost two times the background events are retained without HCAL, with the more difficult challenge due to their distribution peaking just under the signal one, proving its impact on electron-hadron discrimination.

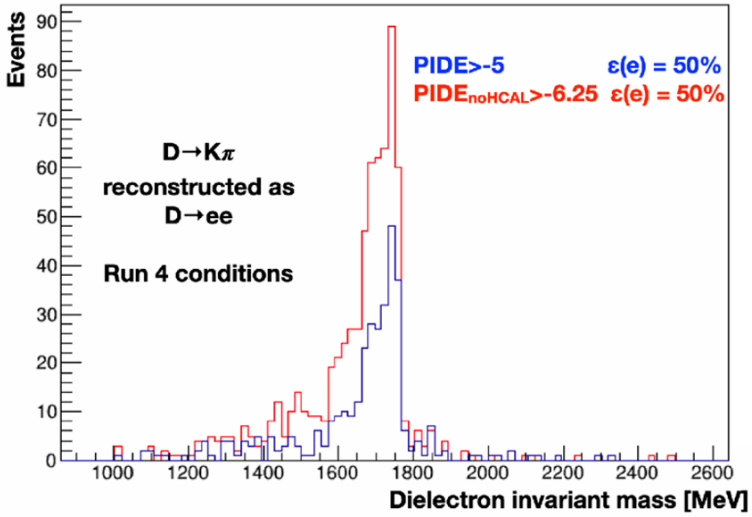

Figure 2.12: Reconstruction of the D^0 invariant mass in electron-positron pairs with LHCb Run 4 conditions. The blue and the red distributions correspond respectively to electrons and hadrons (pions and kaons). From ECAL Upgrade II Workshop.

2.2.7 Muon chambers

The muon system is placed downstream of the calorimeter system, at the end of the LHCb detector. It consists of four stations (M2-M5) with a rectangular structure in the plane perpendicular to the beam axis, with a surface area that increases with the distance from the interaction point.

These stations allow the detection of muons through gas ionization, where the electrons created by ionization are collected by multi-wire proportional chambers (MWPCs). They are interleaved with 80 cm thick iron absorbers with interaction length ~ 20 in order to filter low-energy particles and detect muons with momentum greater than 6 GeV/c.

Each muon station is then divided into four regions of varying sizes and granularity, depending on their distance from the interaction point and their role in muon identification.

The muon identification efficiency in the LHCb detector is high, averaging about 98%, with a misidentification rate for pions and kaons of less than 1%.

2.2.8 Trigger

The LHCb experiment operates in a lower luminosity environment with respect to the other LHC experiments ATLAS and CMS so that the bunch crossing rate is 40 MHz. This leads to a data volume of 4 TB/s which should be reduced to around 10 GB/s due to offline storage capacity. This reduction is achieved through a real-time selection process, using a fully software-based trigger that was developed for Run 3. Therefore, for this upgrade, the original hardware trigger was removed due to saturation.

The trigger system consists of two parts: HLT1 (High Level Trigger) and HLT2. A schematic representation of the trigger data flow is shown in fig 2.13. The HLT1 trigger reduces data volume by a factor of 20 by performing preliminary track reconstruction and allowing for real-time alignment and

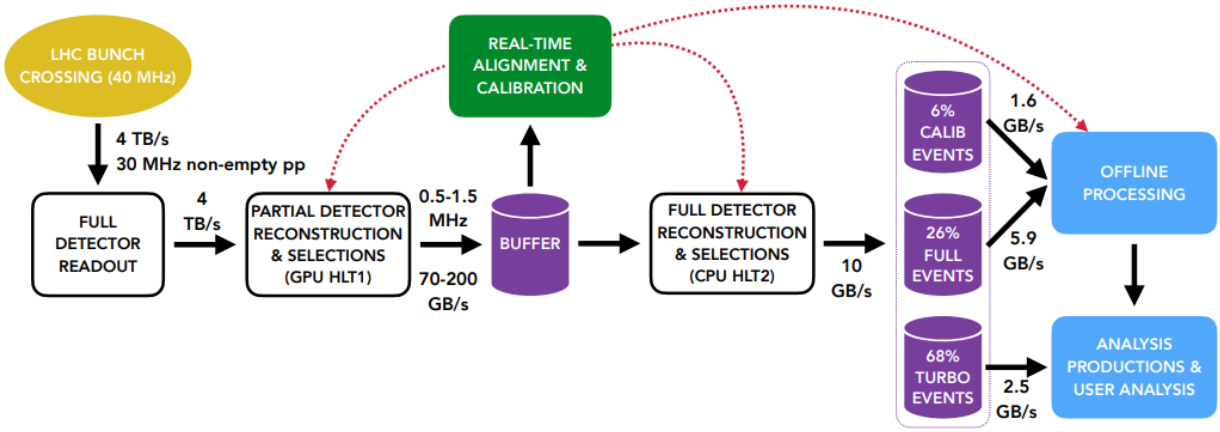

Figure 2.13: Scheme of the upgraded LHCb trigger.

calibration. It must be able to reconstruct tracks and track momenta at the percent level and also discriminate muons from non-muons.

Then HLT2 further reduces the data volume to reach the final amount of data which will be stored permanently by a selection of physics signatures with a full-offline event reconstruction. It implements various algorithms to select specific physics processes.

Chapter 3

Upgrade II: description and motivations

This Chapter aims to give an overview of the High Luminosity program of LHC, and how this program will affect the LHCb detector, with particular attention to its Electromagnetic calorimeter. The challenges that this detector has to face are thoroughly presented as well as the proposed modifications and performance of new modules:

- Section 1 introduces the High Luminosity program of LHC (HL-LHC);
- Section 2 presents the Upgrade II schedule, in accordance with the HL-LHC timeline, and dwells on the upgraded Electromagnetic Calorimeter with motivations and description;
- Section 3 describes the upgraded ECAL performance evaluation studies, both with simulation of physics processes and test-beam data.

3.1 The High Luminosity program

To investigate rare processes and potentially uncover New Physics beyond the Standard Model, the Large Hadron Collider (LHC) must collect significantly more data, requiring a substantial increase in luminosity. The High Luminosity LHC (HL-LHC) program [33] addresses this by significantly increasing the LHC's number of collisions, aiming to detect approximately 15 million Higgs bosons per year, which is about 5 times more than what they had in 2017.

To achieve nearly 10 times the current integrated luminosity, significant changes are needed in both the LHC machine and its detectors. The upgrades aim to maintain precision and resolution while handling the higher collision rates. Key components undergoing improvement include the superconducting quadrupole magnets that focus the particle beam to increase collision intensity. The new quadrupole magnets, made from a novel compound of Niobium (Nb) and Tin (Sn), generate a magnetic field of 12 T, as opposed to the current 8 T. Another implementation is the introduction of 'crab cavities' which tilt the particle beams to increase the collision area. Other enhancements to ensure a focused and stable beam while providing machine protection include improved dipole magnets, superconducting links, collimators, and a new injector chain with a linear accelerator (Linac4).

3.2 LHCb Upgrade II

The HL-LHC is affecting all the experiments, particularly LHCb since by design is operating at a lower luminosity with respect to ATLAS and CMS. Figure 3.1 shows the schedule of LHCb upgrade activities and the corresponding luminosity. The experiment is currently in its operational period, namely Run 3, while Run 4 corresponds to the so-called Upgrade I. It will be followed by the Upgrade II which includes Run 5 and Run 6 data taking periods. The new luminosity shall be obtained step

by step. It is expected to increase slightly during Upgrade I and then rise greatly. The dots indicate the instantaneous luminosity reached in that specific Run while the line is the integrated luminosity. At the end of the HL-LHC phase, LHCb is expected to collect almost 300 fb^{-1} of data.

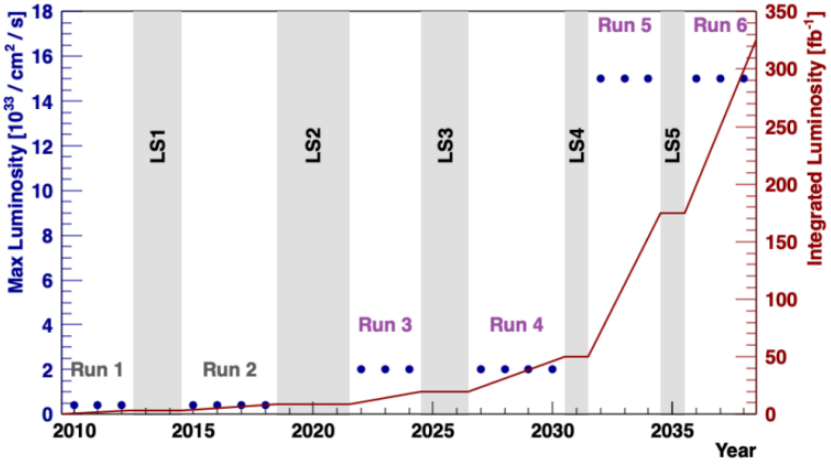

Figure 3.1: Schedule of LHCb operational and long shutdown periods and corresponding luminosity. Upgrade I corresponds to Run 4, and Upgrade II involves Run 5 and Run 6.

With the luminosity, also the pile-up μ, namely the number of interactions per bunch crossing, will increase. For the LHCb experiment, the average value in Run 1 and Run 2 was μ = 1-2, during Run 3 it is expected to reach $\mu \sim 5$ with 50 fb^{-1} of integrated luminosity, while for Upgrade II it is expected to reach values μ >40 with an instantaneous peak luminosity of $1.5 \cdot 10^{34}$ cm^2s^{-1} 2.2. The pile-up increase will crucially affect the experiment performance, particularly considering the process of primary vertices identification, therefore a totally new requirement will be introduced, that is the possibility to do precise time measurements with different detectors to be able to correctly identify particles interaction vertices and therefore suppress combinatorial background. To give an example, the distance between two primary vertices will become 1.5 mm instead of 4.2 mm of Upgrade I [2].

New technologies will be implemented, especially in those sub-detectors involved in particle identification (PID), namely RICH, ECAL but also VELO, which shall also be resistant to radiation damage due to the increased particle number. For this thesis, LHCb upgrades will be described in more detail, especially the ones that affect the calorimeter system.

Here is a short list of the main changes to the principal components of the LHCb detector [2]:

- The Vertex Locator (VELO) has already undergone significant changes for Upgrade II, as described in the previous chapter. The original radial (r,ϕ) strips were replaced by pixel sensors for improved precision. However, they should be redesigned to be radiation hard and to reach 20 ps of track time stamp resolution.

- The Ring-Imaging Cherenkov (RICH) detectors shall be changed primarily for photodetection. In fact, multi-anode photomultipliers (MAPMTs) will be replaced with new high-granularity photodetectors. A new requirement shall be fast timing capability.

- A new detector will be added, namely the Time Of internally Reflected CHerenkov (TORCH) detector, designed to measure time-of-flight from reflected Cherenkov light, and therefore improve particle identification for kaons and protons at low momenta. The time resolution requirement is of 70 ps per photon.

- The inner parts of the SciFi tracker will be replaced with a high-granularity silicon pixels to enhance tracking precision while being resistant to radiation damage. Instead, the UT station will not endure the conditions of Upgrade II, and various new technologies MAPS-based are under investigation.

- The inside walls of the magnet will be instrumented with scintillating bar detectors to extend the low momenta coverage of the tracking system.
- The calorimeters will undergo significant changes in Upgrade II. The Hadron Calorimeter (HCAL), previously used for hardware triggering, will be removed during Run 5 and replaced with 1.7 meters of iron to provide additional shielding for the muon chambers. The Electromagnetic Calorimeter (ECAL) will be completely redesigned.

3.2.1 Motivations for ECAL Upgrade II

The primary goals for the electromagnetic calorimeter (ECAL) improvements in the HL-LHC Upgrade II are achieving high performance in energy and time resolution while ensuring resistance to radiation damage [2]. As the LHC transitions to higher luminosity, rapid degradation of the innermost calorimeter modules is expected due to increased radiation exposure and pile-up. This requires substantial changes to ensure the ECAL can maintain accuracy and reliability under these demanding conditions.

The existing ECAL will not withstand the increased radiation levels expected in HL-LHC Upgrade II, especially in the innermost regions near the beam pipe, with the light collection efficiency dropping drastically as shown in Fig. 3.2. While in the outer modules, the efficiency remains close to 1, in the innermost region it drops well below 50%. The continuous and dashed lines show where the new technology will replace the currently used Shashlik one.

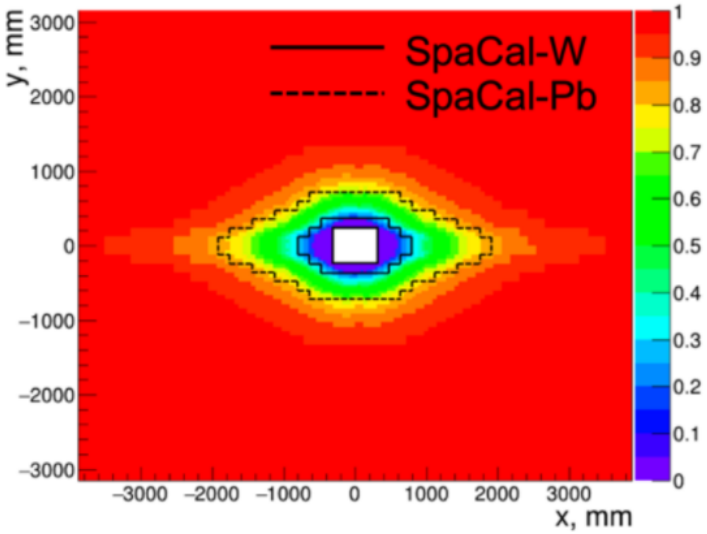

Figure 3.2: Expected efficiency of the ECAL modules after Run 3 (2025) corresponding to 48/fb of integrated luminosity. The lines represent the regions that will be replaced by Spaghetti Calorimeter technology.

Tests conducted on the calorimeter modules indicated that a performance loss of 40-50% occurs after a radiation dose of 10 kGy, with a 60-70% loss after 19 kGy [2]. Given these results, the current ECAL is expected to endure until Long Shutdown 3 (LS3), when the accumulated dose will reach 30-40 kGy. After LS3, the radiation dose will continue to rise, reaching 1-2 MGy during Upgrade II, requiring new materials and technologies to withstand this level of exposure.

The challenge with higher luminosity is the expected overlap of electromagnetic showers due to increased pile-up. This leads to degraded energy resolution and makes identifying primary vertices more challenging. To address this, the ECAL will require finer granularity by scaling down cell dimensions and introducing fast timing to help isolate particle clusters and reduce background from combinatorial effects.

Figure 3.3 shows the estimated radiation dose in Gray that will hit ECAL at the maximum luminosity of Upgrade II. Following its distribution, the calorimeter granularity is chosen to be as shown in the right image, where the dotted line represents the radiation limits for degradation of current Shashlik modules.

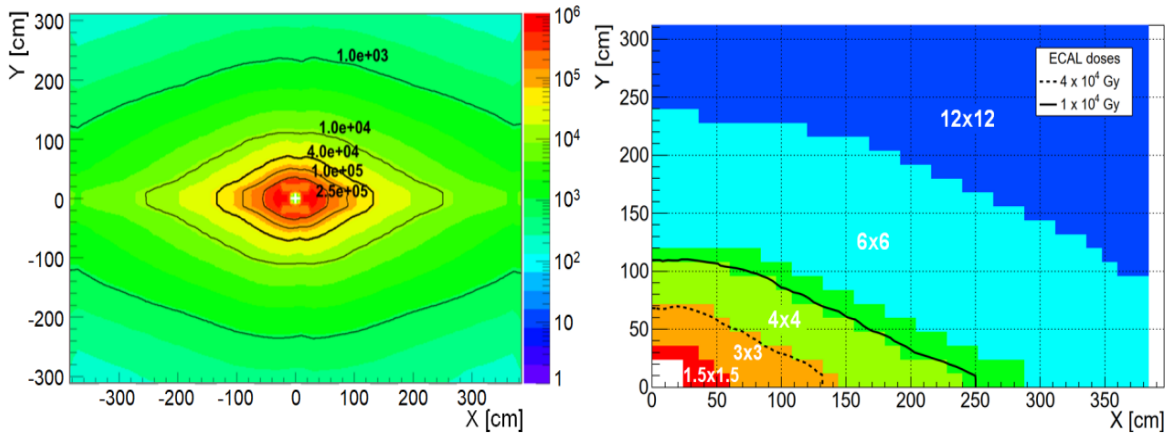

Figure 3.3: Expected accumulated radiation dose on ECAL after an integrated luminosity of 300 fb^{-1} (left). ECAL regions and cell sizes for Upgrade II with the SpaCal-Shashlik option. The dotted line delimits the current Shashlik radiation limit. (right)

To improve granularity and resist radiation, the Upgrade II design will use smaller cells and more resistant materials, such as tungsten or tungsten alloys. These changes will help reduce the Molière radius and the lateral spread of showers, allowing for better separation of distinct particle events. Accurate timing information with a precision of a few tens of picoseconds will be essential for associating clusters with their tracks, considering also the VELO time resolution feature, aiding in energy reconstruction.

The proposed modifications will involve creating more regions within the ECAL, each with varying granularity to handle different radiation profiles. The placement of the new modules will follow the radiation curves, thus resulting in a rhomboidal geometry. Figure 3.4 shows the difference between the layout of the modules in Upgrade I and Upgrade II. In this way, high performance even in high pile-up conditions is maintained, ensuring that the calorimeter can continue to deliver precise energy measurements. Additional adjustments to the calorimeter, such as attaching silicon layers to the cell structure for improved time resolution, are also under consideration [2].

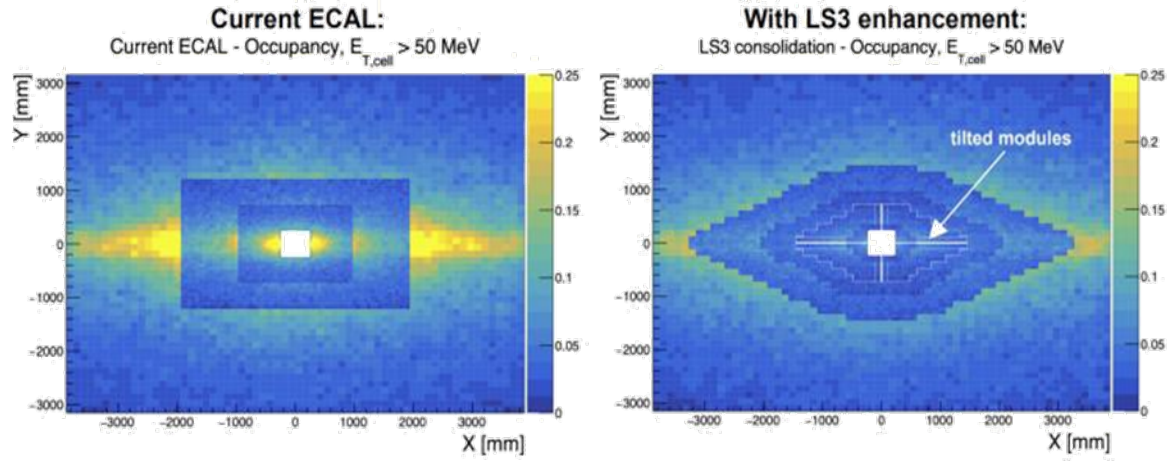

Figure 3.4: Module layout comparison versus beam intensity profile in the two ECAL configurations. Current ECAL granularity where the three regions are visible (left), upgraded ECAL module layout and granularity, following beam intensity profile (right).

3.2.2 Description of ECAL prototypes proposals

The central region of ECAL will be equipped with a Spaghetti Calorimeter (SpaCal) design. In SpaCal modules, the absorber is crossed by longitudinal crystal fibers (hence the name 'spaghetti') with respect to the beam direction that acts as both scintillators and light-transporting media, eliminating the need

for wavelength-shifting fibers which would not have sustained high radiation.

The design for the SpaCal modules includes longitudinal segmentation, with each module split into two sections, i.e. a front and a back section, around the shower maximum as shown in Fig. 3.6. This segmentation is motivated by the fact that the front section's readout is then less affected by pile-up because showers will be less developed upstream, while the back section timing information will improve clustering and reconstruction. This new feature of longitudinal segmentation is being implemented also for the Shashlik modules.

Because the fibers are placed longitudinally to the beam direction if the prototype were to be positioned orthogonal to the beam, only a few fibers would be hit by the shower, but with a great energy deposit. The compromise should be achieved by tilting the prototype so that more fibers would be crossed with not too small an energy deposit.

Baseline option The baseline configuration for Run 5 can be summarized with three modules which are arranged in 6 different sections instead of the current three as it will be described below. This configuration includes longitudinal segmentation for all modules and, consequently, a double-sided readout.

- **Region 1**: This region is the most external one, and therefore less susceptible to radiation damage. It will be equipped with refurbished Shashlik modules with WSL and double-sided readout. Here high granularity is not required, so each module will consist of a single cell, resulting in 1344 Shashlik modules/cells 12 × 12 cm^2 and 300 mm long.
- **Region 2**: As we progress towards the center, the granularity shall increase. Here the ECAL will be equipped with Shashlik 4-cells 12 × 12 cm^2 modules such as each cell area will measure 6 × 6 cm^2. The total number of modules will be 1344.
- **Region 3**: This region will be very thin and will be equipped with 176 9-cell Shashlik modules, with a cell measuring 4 × 4 cm^2.
- **Region 4**: This is the last Shashlik region, with 272 9-cell modules. After this region, Shashlik technology would not be able to endure the radiation dose which will go up 200 kGy. Figure 3.5 shows the layout of a Shashlik module, both longitudinally and orthogonally.

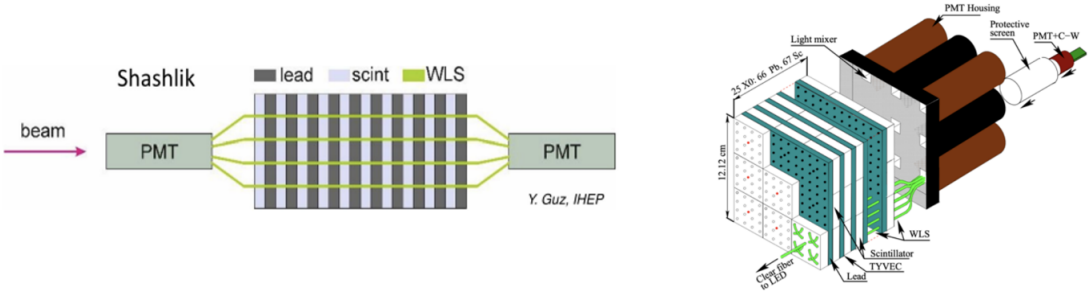

Figure 3.5: Profile and structure of Shashlik prototype for Upgrade II.

- **Region 5**: For intermediate radiation regions, with radiation doses between 40 and 200 kGy, ECAL will employ the new SpaCal design with a lead absorber combined with polystyrene fibers. These fibers have been studied thoroughly and Kuraray fibers (SCSF-78) gave great results with a light output of 10000 photons/MeV and a scintillation decay time of 2.8 ns. This setup consists of 144 modules, each with cell sizes of 3 × 3 cm^2. The module total length will be 150 mm split in two parts of 50 and 100 mm respectively.
- **Region 6**: In high-radiation central regions where the dose exceeds 200 kGy, the proposed SpaCal uses tungsten as the absorber material and garnet crystal fibers as the scintillator. This configuration includes 32 modules, each with cell sizes of 1.5 × 1.5 cm^2. The chosen material for the fibers was $Gd_3Al_2Ga_3O_12$ (GAGG). The light output was measured with a ^{137}Cs radioactive

source to range from 27900 and 49500 photons/MeV and the light rise and decay time were measured by gamma excitation to be around 70 and 50 ps respectively.

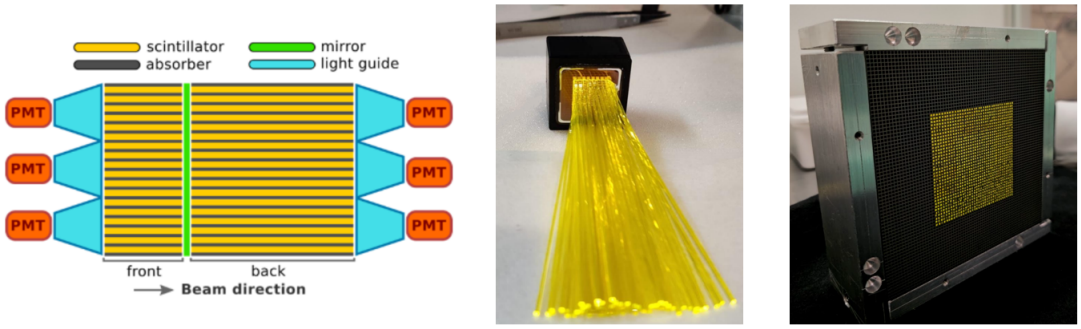

Figure 3.6: Schematization and picture of W/GAGG prototype used for region 6 of ECAL.

Research and development on precise timing indicate that both Shashlik and SpaCal technologies, whether using crystal or polystyrene fibers, offer properties that could achieve very good timing performance. Test beam measurements have demonstrated that these configurations can attain timing precision of around 15 picoseconds at high energy [2].

The proposed modifications for the ECAL in Upgrade II also have implications for the readout system. The strategy involves advanced waveform processing and using timing information to improve clusterization in the detector's backend electronics, thus optimizing readout bandwidth and supporting efficient online reconstruction and software triggers. These changes are designed to maintain high performance in energy and time resolution while ensuring resistance to the increased radiation levels expected during the HL-LHC's high-luminosity operations.

Downscoped option In parallel to the baseline configuration just described, a downscoped scenario has been investigated. It consists in reusing a large fraction of the existing Shashlik modules, and replacing only the innermost modules with SpaCal ones. This implies that the longitudinal segmentation will be present only in SpaCal modules, while Shashlik modules will remain with a single-sided readout.

This scenario would benefit in terms of material cost, however, it would result in a drop in performance, especially for time resolution [29]. The lack of upstream timing information would reduce the possibility of reconstructing overlapping showers in the back section of the calorimeter, an important factor in high pile-up conditions.

3.3 Review of studies for ECAL Upgrade II

In preparation to the Upgrade II, three different channels are explored in parallel:

1. The simulation of physics processes on the upgraded ECAL to assess particle reconstruction and identification performance
2. Simulation of single ECAL modules to extrapolate their energy and time resolution
3. Test-beam campaigns on single ECAL modules (prototypes), to compare with simulation results

The first channel studies and preliminary results are presented in the section below. Chapter 4 of this thesis is dedicated to the study of Run 5 PID performance using single particle simulation, while an original study on the $Z \to e^+e^-$ process is presented in Chapter 6 of this thesis.

The second and third channels are assessed in Chapter 5 of this work, where test-beam data are analyzed and compared with simulated data.

ECAL performance with physics processes To assess the performance of ECAL in the Upgrade II configurations, several studies have been carried on, exploiting simulation results to get information about particle reconstruction and identification.

Concerning physics performance, the electroweak decay $B^0 \to K^{*0}e^+e^-$ was chosen to be the benchmark channel for electron identification in b-physics, having a peaking background coming mostly from misidentified pions in $B^0 \to K^{*0}\pi^+\pi^-$. The usual figure of merit for PID is the total reconstructed energy over track momentum distribution (fig. 2.9), namely E/p, and the front side reconstructed energy over the total energy, E_F/E. Figure 3.7 shows PID performance comparison between different configurations of ECAL. The electron identification efficiency $\epsilon(ee)$ is shown with respect to the pion identification efficiency $\epsilon(\pi\pi)$. In Run 2, to $\epsilon(ee) = 80\%$ corresponded $\epsilon(\pi\pi) \sim 3 \cdot 10^{-3}$. Run 3 performance is worse, due to the higher pile-up environment and initial degradation of the modules more exposed to radiation. Green and blue lines are obtained by simulated data, corresponding to two different options of the Run 5 configuration. In order to achieve the same PID performance as for Run 2, some cuts on E/p and E_F/E distributions were made. The figure shows that this can be achieved by the Run 5 configuration by applying requirements on both distributions.

Another relevant physics channel in evaluating the electron reconstruction performance in the high luminosity case is $Z \to e^+e^-$. The resolution of its invariant mass is an important quantity for this purpose since the current measurement suffers from poor electron reconstruction due to ADC saturation (Ch. 1). M_Z resolution has been estimated to be 14 GeV, with a 23% signal efficiency for Upgrade II conditions [29]. In Chapter 6, Z boson reconstruction performance in the dielectron final state in Run 5 configuration is studied.

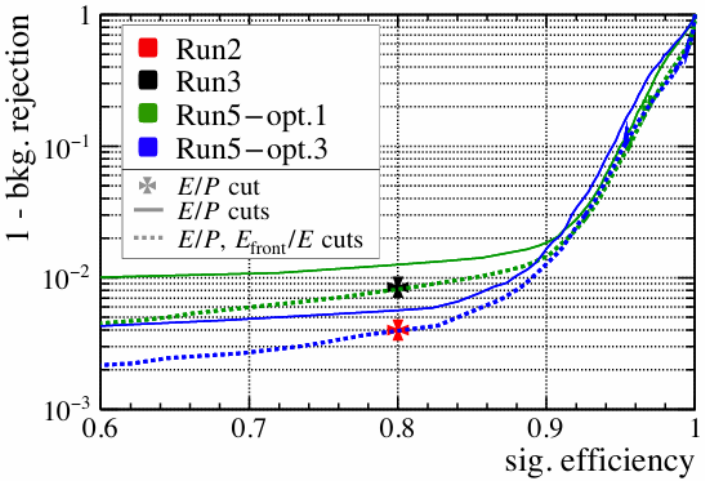

Figure 3.7: Performance comparison between Upgrade I and Upgrade II runs for the physics decay $B^0 \to K^{*0}e^+e^-$. Dashed and continuous lines represent different input variables to the identification algorithm [29].

Test-beam and simulation studies on ECAL modules Together with simulations, test-beam campaigns have been conducted at SPS and at DESY facilities to investigate the different modules performance.

Figure 3.8 shows energy and time resolution for SpaCal with lead absorber and Polystyrene fibers, with prototype tilted of 3°+3°on both azimuthal and polar planes. The energy resolution is compatible with simulations and it meets the requirement on sampling and constant term, with respectively 10% and 1.1%, while time resolution reaches below 25 ps for both front and back sections of the prototype for beam (electron) energies over 20 GeV, reaching below 10 ps for 100 GeV. Chapter 5 of this thesis is dedicated to the study of SPS test-beam campaign with SpaCal lead/polystyrene prototype.

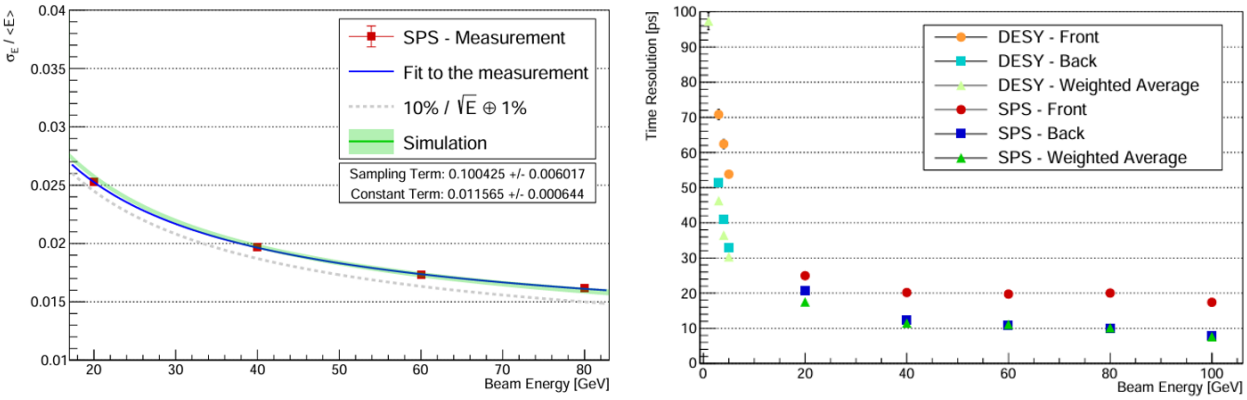

Figure 3.8: Energy resolution data and simulation for SpaCal lead/polystyrene prototype obtained at SPS (left). Time resolution for SpaCal lead/polystyrene prototype obtained at SPS and DESY (right) [29].

Chapter 4

Charged particle identification

In this Chapter, charged particle identification performance will be presented for ECAL in Run 5 conditions with simulations in both downscoped and baseline options.

The Chapter is structured as follows:

- In Section 1 there will be an introduction to the Hybrid MonteCarlo simulation framework
- Section 2 presents the event reconstruction framework with the cluster reconstruction algorithm
- Section 3 is dedicated to the evaluation of the reconstruction performance both in low pile-up and high pile-up conditions
- Section 4 describes the BDT algorithm used for electron-hadron discrimination
- In Section 5, the PID performance of various BDT models is compared in both baseline and downscoped option
- Section 6 shows the application of the BDT model to the $B^0 \to K^{0*}e^+e^-$ physics process.

4.1 The Hybrid MonteCarlo simulation framework

At the time of writing this thesis, the LHCb simulation framework has not been updated yet with the upgraded configuration of the experiment. Therefore, a standalone simulation framework has been developed to study the response of the whole LHCb ECAL as well as the single modules of Shashlik and SpaCal, for different configurations of the upgrades. It is a C++ Monte Carlo software based on the GEANT4 toolkit [20] which simulates the energy deposition in the calorimeter, the production and transport of Cherenkov and Scintillation light, the photodetector response, and the time pick-off [29].

The key features and steps of this simulation are the following and shown in figure 4.1:

- Geometry Configuration: The framework can reproduce calorimeter modules (SpaCal and Shashlik) with various geometries, using elementary geometries of GEANT4. Defined modules can be replicated in space to fill a volume, allowing the simulation of various configurations of the LHCb ECAL.
- Energy Deposition Simulation: The software simulates how particles deposit energy in the calorimeter, accurately modeling the interaction of particles with the detector using the FTFP_BERT [18] physics list of GEANT4 to simulate radiation-matter interactions. This process produces detailed 3D maps of energy deposits.
- Optical Calibration: Particles interacting with the detector can produce Cherenkov and Scintillation photons. The framework handles the simulation of the former with GEANT4's ray-tracing library, which models the travel of light within the medium and accounts for surface non-idealities.

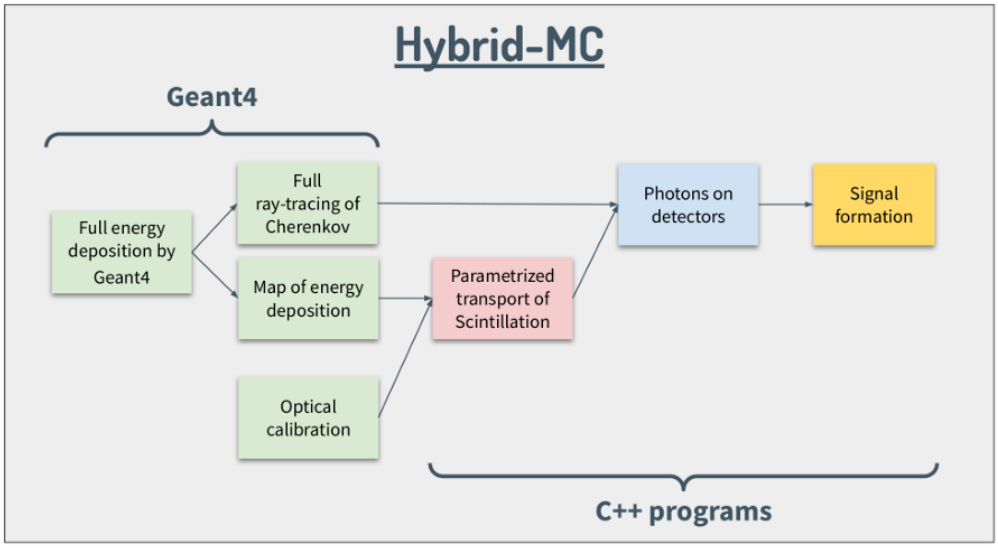

Figure 4.1: Schematic representation of the Hybrid-MC framework.

Instead, simulating scintillation photons is computationally demanding, as it requires full ray-tracing of millions of photons, making the simulation time-intensive (approximately 1-2 hours per GeV of deposited energy). To address this, a hybrid approach is used: scintillation photons are generated at various positions of each module tile, and then, the probability of light extraction is computed to create a wavelength-dependent spatial map and time-spread distribution. The energy deposits obtained in the previous step are converted into scintillation photons according to the light yield and emission spectrum of the active material.

This significantly reduces computation time while maintaining realism, allowing efficient and accurate simulation of scintillation photon transport. Figure 4.2 shows the performance comparison between full ray-tracing and Hybrid simulation. The hybrid approach proves to be reliable.

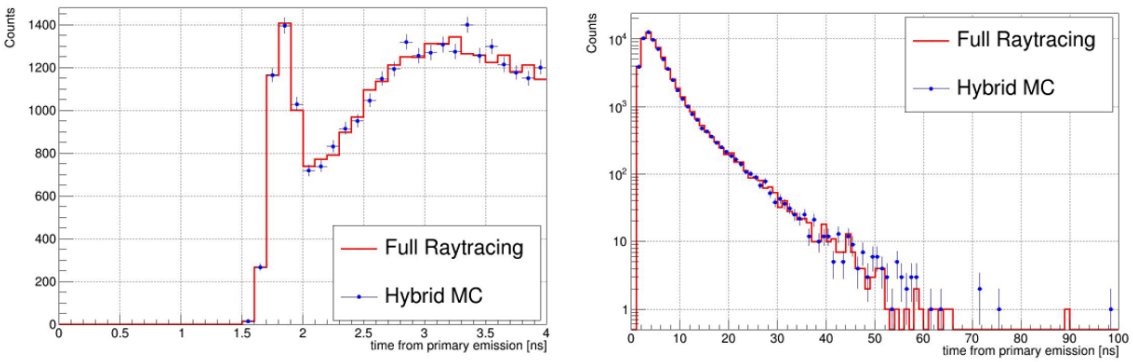

Figure 4.2: Detector time response from primary emission evaluated with full ray-tracing and with Hybrid MC simulation.

- Pulse Production and Time Pick-off: The final simulation step involves building the photodetector's waveform from the photon data. The quantum efficiency of the photodetector is applied to filter photons by wavelength, and additional losses due to a light guide are simulated. A digitized pulse is initialized, with electronic noise added as a Gaussian-distributed random voltage. Time smearing accounts for the photodetector's transit time spread, and the single photoelectron pulses are computed and summed to the digitized pulse, assuming linear detector behavior and infinite dynamic range of the electronics.

4.2 Event reconstruction framework

To reconstruct particle energy, a framework has been developed which includes the Hybrid simulation described in the section above. Its workflow is shown in Fig. 4.3, from the generation of events to the analysis, and it is described below:

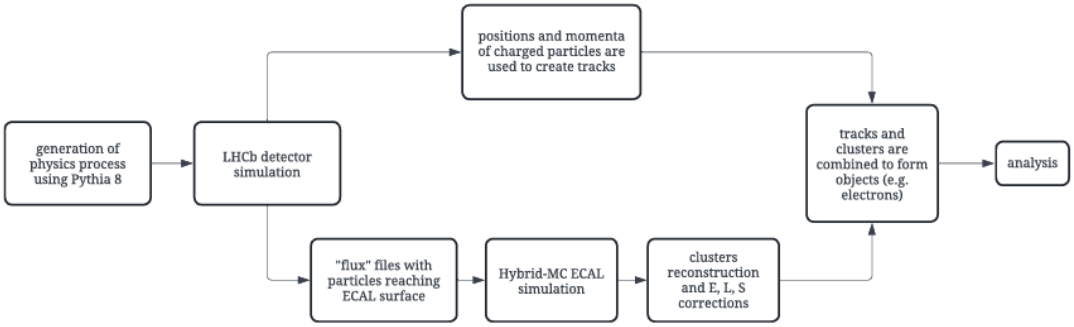

Figure 4.3: Schematization of the workflow for the simulation and event reconstruction.

1. The required physics process (either single particle guns or a specific physics process such as $Z \to ee$) is generated using PYTHIA8 [36].

2. The particle passes through the LHCb detector simulation, in its Run 5 configuration, and their interaction with matter is simulated using GEANT4. The interaction with all detectors before ECAL is simulated and the information on the particles that reach its surface is saved in a 'flux' file which is given as input to the Hybrid-MC simulation tool. This information includes the primary momentum of the particle, its position, and momentum, which are used to reconstruct the particle trajectory and so to create tracks.

3. The Hybrid simulation gives as output the energy collected by each ECAL cell, following the steps described in the previous section.

4. The cluster reconstruction algorithm is based on the cellular automaton approach [17], closely following the reconstruction approach used so far in the LHCb experiment. It searches for local maxima with transverse energy over 50 MeV, which are then considered "seed cells". Then it checks that other seed cells are not adjacent to each other and creates a cluster of 3 x 3 cells with the seed cell at the center as shown in Fig. 4.4.

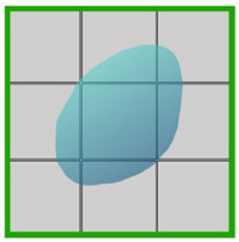

Figure 4.4: Representation of 3x3 cluster formed around a local maximum of deposited energy.

5. The energy and position of the cluster are corrected following the official LHCb framework procedure. L-, S-, and E-corrections are applied [29]. L- and S- corrections are applied respectively to the longitudinal and transverse position of the cluster. E-correction takes into account leakage effects and corrects the reconstructed cluster energy. Figure 4.5 shows the effect of E-correction on reconstructed energy.

6. Cluster position and energy are combined with flux information obtained to perform analysis. The flux file gives information on the position of particles just before arriving at the face of ECAL. For each event, the distance d between the particle hit (x_{track}, y_{track}) and the cluster

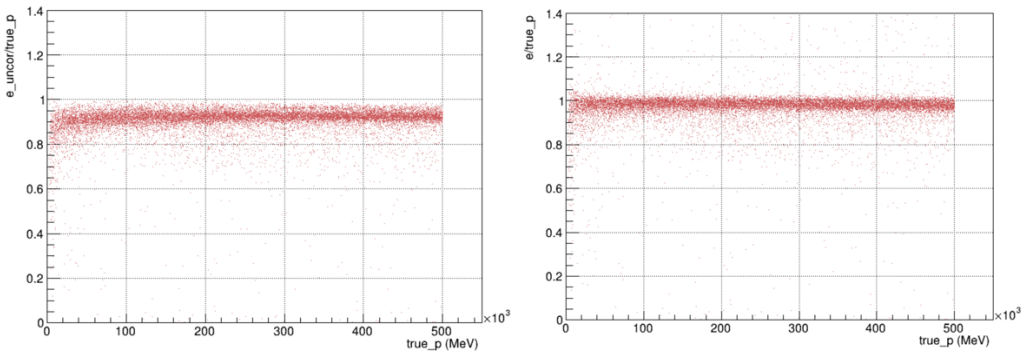

Figure 4.5: Reconstructed cluster energy over the momentum of a primary particle in the baseline configuration with $\mu = 1$ before correction (left) and after correction (right).

position $(x_{cluster}, y_{cluster})$ in the (x, y) plane is computed and minimized.

$$d = \sqrt{(x_{track} - x_{cluster})^2 + (y_{track} - y_{cluster})^2} \tag{4.1}$$

Only clusters with the closest position to the particle hits are matched with the corresponding particle, retrieving the primary particle momentum.

7. Once clusters are matched with tracks, their deposited energy cannot always reproduce the initial particle momentum, and there is some missing energy. This happens because the energy loss caused by Bremsstrahlung emission was not taken into account with this procedure.

   In the LHCb experiment, the main cause of Bremsstrahlung radiation is the interaction of the electron with the detector material. Two cases can be distinguished:

   - If the Bremsstrahlung emission happens after the magnet, the Bremsstrahlung photons are typically contained in the same ECAL cluster as the electron, and their trajectory would be almost identical to that of the original electron, allowing the event to reconstruct the correct energy as these photons would enter the electron cluster.
   - If the Bremsstrahlung emission happens before the magnet, the Bremsstrahlung photons might create another ECAL cluster separated from the electron cluster, given that electrons will bend inside the magnetic field, as shown in figure 4.6.

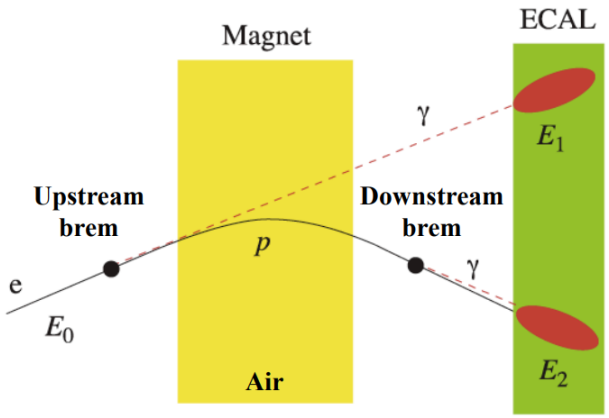

Figure 4.6: Schematic representation of the emission of Bremsstrahlung radiation before and after the magnet.

Therefore, clusters could be created either by the original electron or by a Bremsstrahlung photon emitted before the LHCb magnet, resulting in two clusters in the same event: one from the electron and one from the photon.

Figure 4.7 shows the photon production vertex positions for 100 $Z \to e^+e^-$ events. It is possible to recognize the position of the three tracking chambers downstream of the magnet (the white

area). This shows the effect of the neutron shielding placed before the front face of ECAL, which contributes to the creation of many photons with low energy that cannot be reconstructed separately because of their close production vertices. In a very good approximation, it is right to assume that those photons, i.e. the photons produced downstream of the magnet, will deposit energy in the same cluster of the electron, effectively contributing to the energy computation. On the other hand, upstream photons could be reconstructed separately as their production vertices are lower in number and more distant from each other.

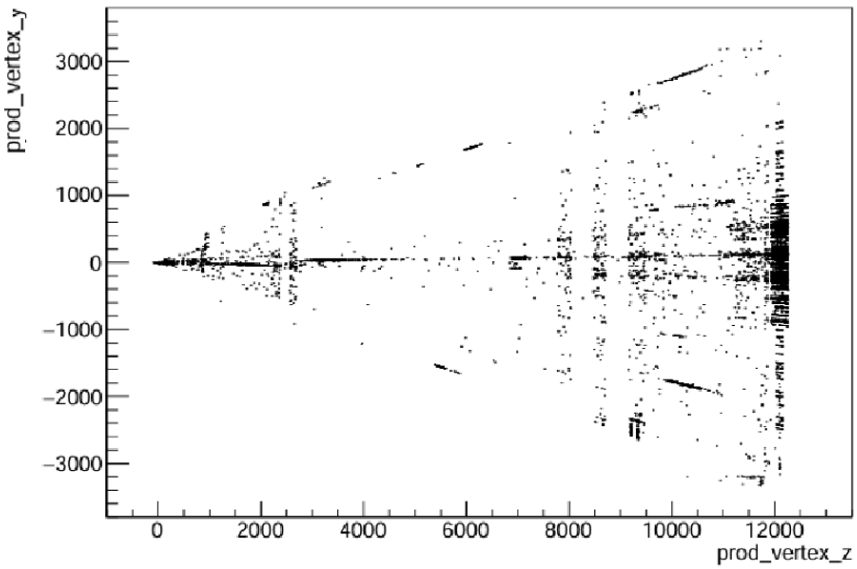

Figure 4.7: Photons production vertex position for 100 $Z \to e^+e^-$ events.

In the analysis presented in this Chapter, and later in Chapter 6, Bremsstrahlung recovery is handled in different ways. While downstream Bremsstrahlung photons are considered contained in the cluster, upstream Bremsstrahlung recovery is currently possible only in the case of low pile-up. In fact, the missing energy due to Bremsstrahlung can be reconstructed by looking at other clusters from the same event and matching their position to the track of the primary electron as shown in Fig. 4.6. In the high pile-up case, this matching procedure is very difficult, due to the substantial increase in the cluster number per event, therefore algorithms for Bremsstrahlung recovery for Run 5 are currently under investigation and will not be present in this thesis.

4.3 Evaluation of reconstruction performance

The aim of this study is to test the reconstruction algorithm performance in the Upgrade II configuration of the Electromagnetic Calorimeter and in the new higher luminosity environment exploiting the previously described simulation framework. Moreover, the quantities obtained from the reconstruction are investigated in order to understand the difference between pions and electrons expected from their different shower development and to exploit this difference to develop a discrimination algorithm based on Machine Learning. Finally, the efficiency of this algorithm is studied and the obtained model is applied to test-beam data and to the simulation of Z boson production in electron channel in the following Chapters.

The analysis was performed on 10^4 simulated single particle events with energy uniformly distributed in the (0, 500) GeV energy range both for pions and electrons in two different configurations, baseline and downscoped, described in Chapter 3. Also, this procedure was firstly done in the case without pile-up and then pile-up was added by simulating it with Minimum Bias, i.e. events for inelastic pp collisions selected with a very loose requirement, at $1.5 \cdot 10^{34}$ cm^{-2}s^{-1} luminosity. The Minimum Bias events were merged with the single particle at the pulse level. In this Chapter, the word "particle" refers to both pions and electrons if not specified otherwise.

4.3.1 Preliminary analysis with pile-up = 1 condition

At first instance, it was decided to test the simpler case of pile-up $\mu = 1$. This is a necessary step to develop and test the reconstruction and discrimination algorithms with the advantage of being in a controlled environment, and so where it is possible to track the original particles and correctly associate the cluster to the primary particle.

The reconstruction algorithm provides the energy of the clusters, deposited in the back and in the front (when present) sides of the calorimeter as well as the position of the clusters. To test the algorithm's effectiveness, these quantities were investigated with respect to the energy of the generated particle. Figure 4.8 shows the number of clusters reconstructed for each event. The two distributions are quite similar, with the majority of events producing only one cluster.

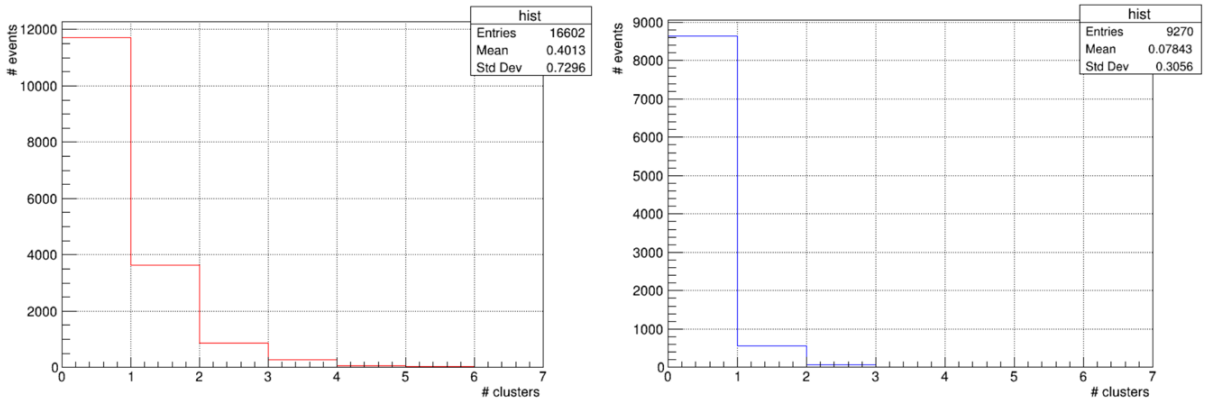

Figure 4.8: Number of cluster distributions for electrons (left) and pions (right) in the downscoped configuration.

In this case, it is easy to recover energy lost by Bremsstrahlung because the number of clusters per event is small, therefore making it easy to link a cluster to its particle. If an event presents two or more clusters, it is very likely that these additional clusters are formed by a Bremsstrahlung photon (less frequently by an electron) that is created before the LHCb magnet and whose life can be easily reconstructed. Therefore, to recover the most energy out of each event, the energy of each cluster is summed per event. This quantity is then compared to the momentum of the primary particle at the origin vertex, in this treatise called 'true_p' to indicate the fact that it is linked to the Monte Carlo truth. In fig. 4.9 it is possible to see the difference in the reconstructed energy between electrons and pions.

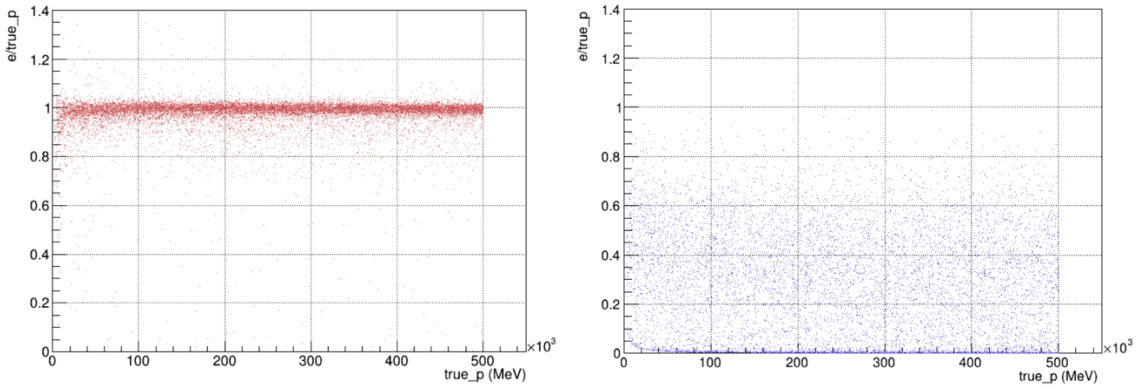

Figure 4.9: Scatter plot of the reconstructed energy with respect to the momentum of the particle at the primary vertex for electrons (left) and pions (right) in the downscoped configuration with pile-up = 1.

The reconstructed electron energy distribution is very close to the momentum of the primary electron, while for pions this value is more spread out and closer to 0. The reason for that is the nature of the electromagnetic and hadronic showers. While hadron showers do have an electromagnetic component,

their interaction with matter has longer distances resulting in a more spread-out energy deposition in the electromagnetic calorimeter.

The energy over primary momentum distribution is shown in fig. 4.10. The two distributions have minimum overlap, and to separate them, a simple cut on e/true_p = 0.8 value can be exploited.

The choice of using the primary particle momentum, i.e. the generated Monte Carlo particle momentum, has multiple reasons. First, this quantity is used to assess the reconstruction performance of the electromagnetic calorimeter. Secondly, the track information should have been obtained by using other detectors than the calorimeter. Since this study was only performed with ECAL-reconstructed quantities, track momentum was not available. Lastly, the Monte Carlo truth was needed in order to train a Machine Learning model for the development of a discrimination algorithm, which will be discussed in the following sections.

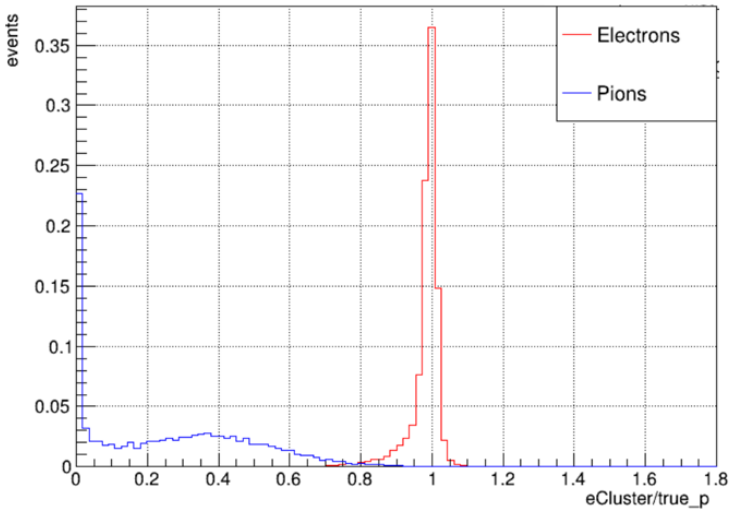

Figure 4.10: Distribution of total cluster energy with respect to the true primary energy of the particle in the downscoped configuration.

4.3.2 Energy distributions in high pile-up conditions

Once the procedure of identifying the original particle was consolidated and the relevant quantities connected to the different particle shower behaviours were found, a new sample was generated with Minimum Bias at $1.5 \cdot 10^{34}$ cm^{-2}s^{-1} luminosity. This corresponds to the expected luminosity for Run 5, which results in several interactions per bunch crossing $\mu > 40$.

The procedure of the previous section was repeated, investigating the same quantity distributions. The difference with the previous case is that the number of clusters per event is such high that it is not possible to 'follow' the original particle in its interaction with matter. The cluster distribution is shown in Fig. 4.11. In this case, the primary particle energy is recovered only with the energy of the corresponding cluster, resulting in the loss of the energy due to upstream Bremsstrahlung emission. This effect is visible in the plots in Fig. 4.12, especially for electrons, where the points are more spread with respect to fig. 4.9 where the energy fraction was around 1. Bremsstrahlung recovery algorithms in these conditions are being currently developed and were not available at the time of this study.

Figure 4.13 shows the reconstructed energy over the primary particle momentum (e/true_p) distributions for electrons and pions. The difference between these distributions in the case with $\mu = 1$ in Fig. 4.10 is quite evident. In that case, pion and electron distribution were well distinguished. Also, the same consideration could be done by comparing it to the same plot for the Run 3 configuration, shown in Fig. 2.9 of Chapter 2, where the electron left tail was less relevant with respect to pion energy.

In order to discriminate effectively electrons and pions, this distribution is not sufficient. Thus, a BDT algorithm is developed and will be presented in the next section.

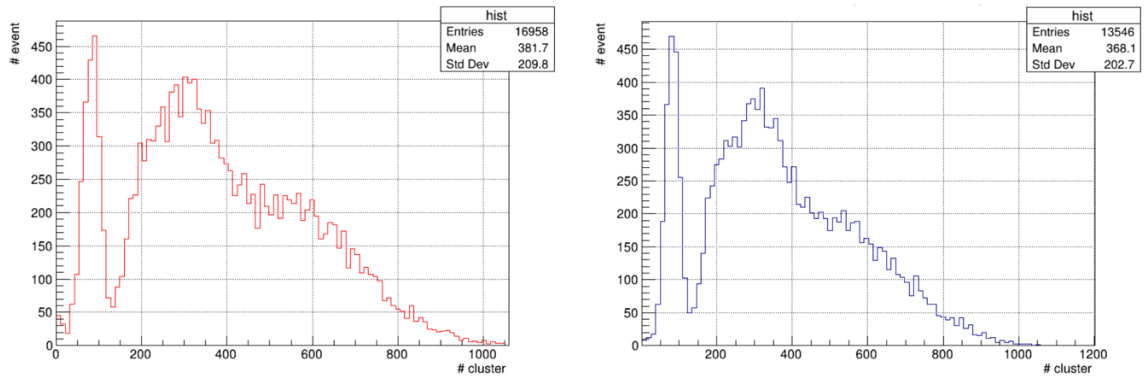

Figure 4.11: Number of cluster distribution for electrons (left) and pions (right) in the baseline configuration with pile-up.

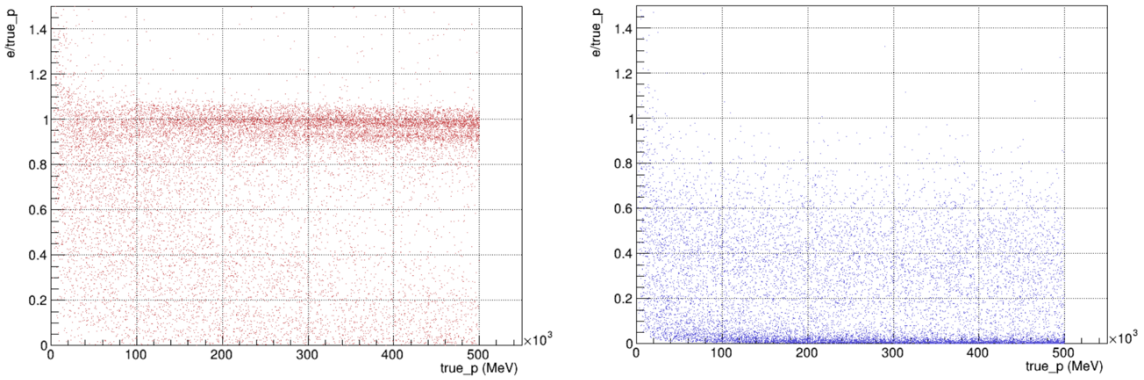

Figure 4.12: Scatter plot of the reconstructed energy with respect to the momentum of the particle at the primary vertex for electrons (left) and pions (right) in the baseline configuration with pile-up.

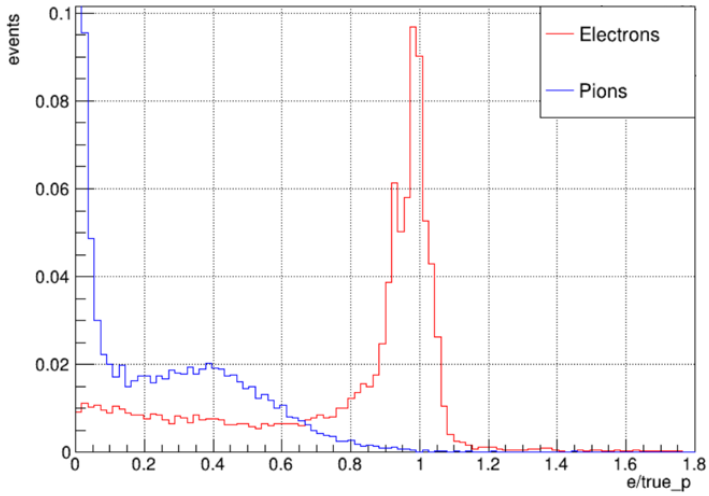

Figure 4.13: Distribution of total cluster energy with respect to the true primary energy of the particle in the baseline configuration with pile-up (left).

4.4 Boosted Decision Tree for discrimination algorithm

The current methodology to perform charged particle identification with ECAL relies strongly on the E/p distributions as explained in Chapter 3. In Upgrade II conditions, the presence of a high pile-up makes it impossible to rely solely on this method. The goal of this analysis is to find effective ways to perform electron-hadron discrimination while reconstructing particles with ECAL and the tool chosen is a BDT classifier.

To analyze complex datasets, advanced computational techniques are often used, among which is Machine Learning, which focuses on building algorithms that can learn and make predictions or decisions based on data patterns. In this section, the chosen Machine Learning algorithm, a Boosted Decision Tree (BDT) [31, 32], is described and its applications in electron-hadron discrimination are presented.

4.4.1 Boosted Decision Tree

A Boosted Decision Tree is a Machine Learning algorithm that combines the simplicity and interpretability of decision trees with the power of boosting. This falls back into the category of Multivariate Analysis, which refers to a set of statistical techniques used to analyze data that involve multiple variables or features simultaneously, searching for relationships and correlations. There are various types of Machine Learning, but in the context of particle physics, supervised learning is most commonly used, where algorithms are trained to classify particles or predict physical quantities based on a set of input features derived from experimental data.

To implement this analysis, I used the scikit-learn library [30], specifically, I employed the Gradient Boosting Classifier, an implementation of the BDT algorithm.

A Decision Tree is a simple algorithm that splits data into subsets based on the value of input features. It operates by making a series of binary decisions, forming a tree-like structure where each node represents a decision based on a specific feature, and each branch represents the outcome of that decision. The process continues until the model reaches a decision, or "leaf" node, which provides the final classification or prediction. There are various parameters that can be set such as learning rate and maximum depth. Lowering the learning rate increases the number of trees needed for a good model, while the depth of a tree is the longest path from the root node to a leaf node, a deeper tree can model more complex relationships.

To overcome the limitations of individual decision trees, Boosting is employed. Boosting is an ensemble technique that combines the outputs of multiple weak learners (in this case, decision trees) to create a strong predictive model. The key idea behind boosting is to build trees sequentially, where each new tree focuses on correcting the errors made by the previous ones. This is achieved by adjusting the weights of incorrectly classified data points, so the next tree pays more attention to them.

In a Gradient Boosting framework, each tree is trained to minimize a loss function, typically through a method similar to gradient descent as shown in Fig. 4.14.

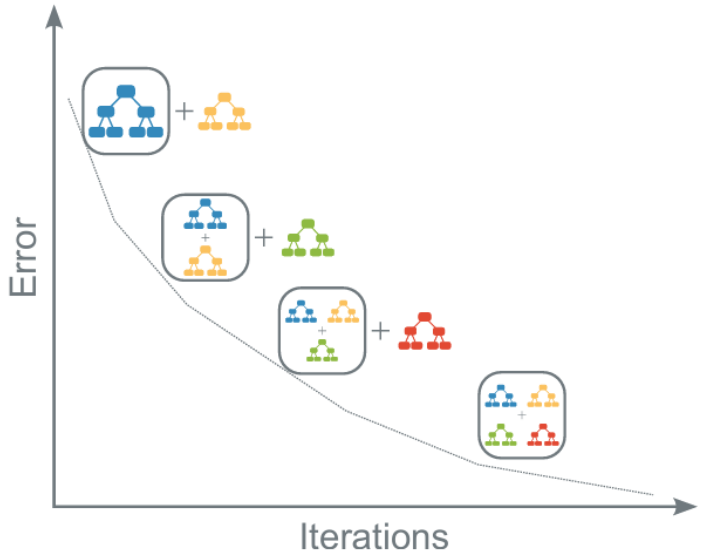

Figure 4.14: Operating steps of a Gradient Boosting Classifier [39].

4.4.2 Input variables

In order to investigate the quantities that better discriminate pions and electrons, several distributions are studied. Here, some of the most relevant ones are presented. The first distribution is chosen as the current criterion for electron-hadron discrimination: the reconstructed energy over the primary particle momentum (e_{tot}/true_p), shown in fig 4.10 for the downscoped configuration with $\mu = 1$ and in fig 4.13 for the high pile-up environment.

Another important quantity concerning the difference between the particle shower development is the energy of the clusters created in the front or back side of the calorimeter with respect to the total energy (e_F/e_{tot} or e_B/e_{tot}). In this case, it is necessary to distinguish the two configurations, baseline and downscoped. In the downscoped option, only the SpaCal modules are built with double-sided readout, and therefore able to measure the energy deposited in the front and back sections, while the Shashlik modules read only the energy deposited on one side. To keep the same notation, since all modules are aligned by their rear end, also for non-segmented Shashlik modules there will be references on its back side, simply meaning the readout-equipped face. This results in the energy deposited on the back side of Shashlik having the same distribution of the total energy, therefore being 1. Here only the quantities related to the SpaCal modules are presented, as shown in Fig. 4.15 for the case of $\mu = 1$ in the downscoped configuration and in fig. 4.16 for the baseline configuration in high pile-up conditions.

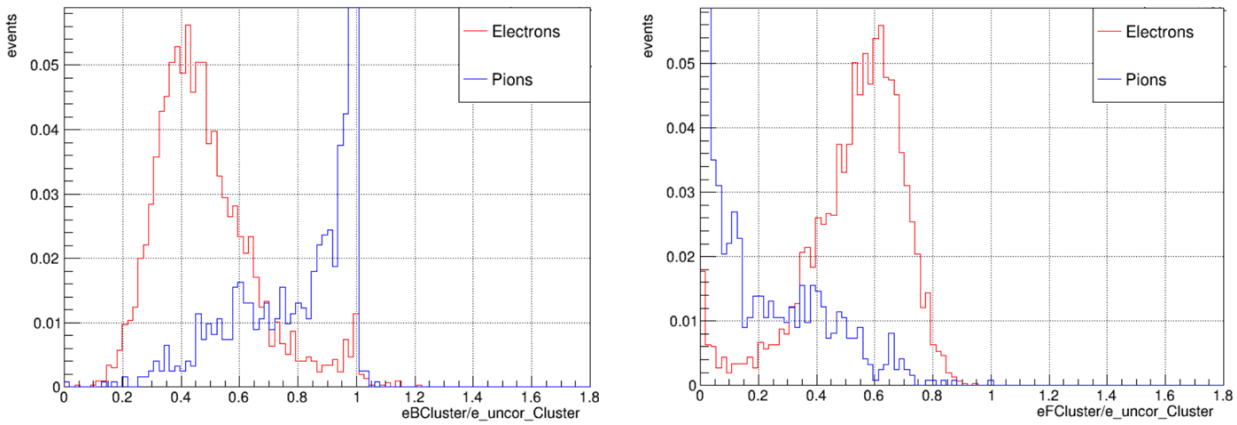

Figure 4.15: Distribution of total back (left) or front (right) cluster energy with respect to the total cluster energy (front and back) for the SpaCal modules in the downscoped configuration and pile-up of 1.

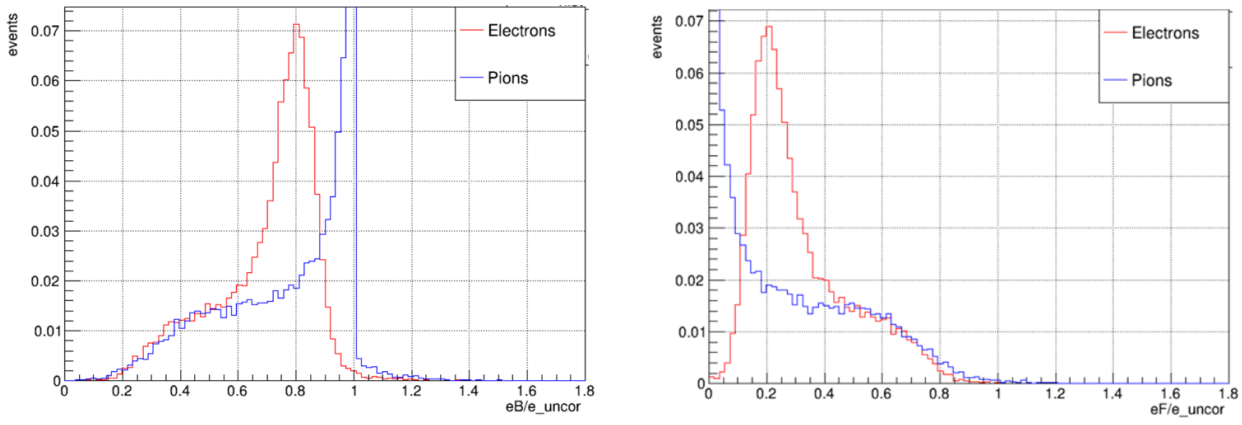

Figure 4.16: Distribution of back (left) or front (right) cluster energy with respect to the total cluster energy (front and back) in the case of high pile-up.

The difference in behavior of the particle shower is clearly visible in these plots. While the electron shower is more longitudinally developed and so the energy deposited is more or less equally distributed

on the front and back sides, the pion shower is shorter, and almost all the energy is deposited in the first material encountered, so the front side of the calorimeter.

A new quantity is investigated in the high pile-up case, that is the distribution of the deposited energy on each cluster cell with respect to the cluster energy (e_{Fcell}/e_F). The distributions were computed for each cell, and the results presented are the energy of the seed cell, namely cell 1, and the energy of cell 6 (middle-right), which can be seen in Fig. 4.17. The distribution of the seed cell energy to the cluster energy reflects the different behaviors of particle shower, providing a probably good discriminating variable. The deposited energy is more spread out in the whole cluster in the case of pion shower while it is more contained to the center for electrons. The distribution of the deposited energy in all cluster cells but the seed cell exhibits the same behavior as in the figure, the two-particle distributions do not seem so recognizable at first glance, but still, the pion energy is more spread, while the electron has a narrower peak around 0.

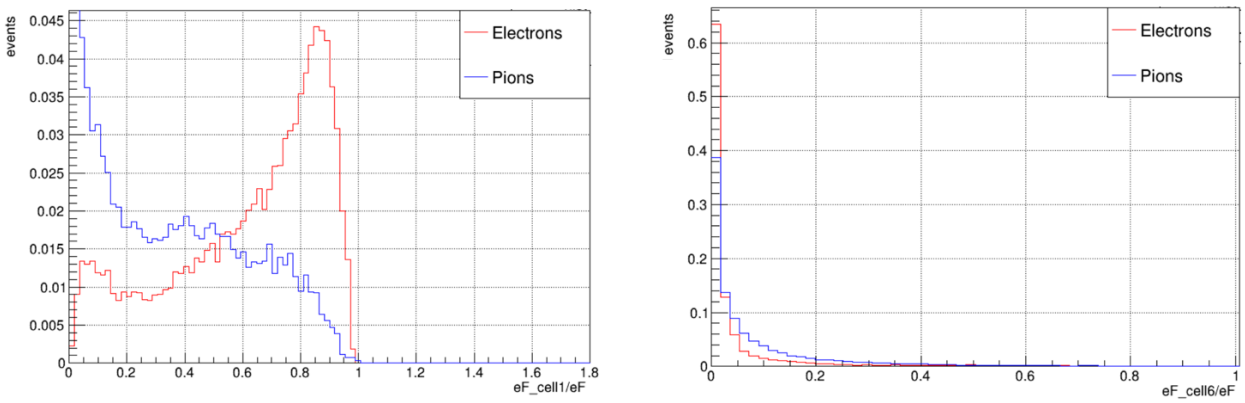

Figure 4.17: Distribution of the deposited energy in the seed cell of the front cluster with respect to the total cluster energy in the baseline configuration with pile-up (left). Distribution of the deposited energy in the outer cell number 6 of the front cluster with respect to the total cluster energy in the baseline configuration with pile-up (right).

4.4.3 BDT training

I developed a Machine Learning separation algorithm to distinguish between pions and electrons based on their shower development in the electromagnetic calorimeter. The pion shower is much more contained compared to the electron shower, resulting in a lower ratio of reconstructed energy to the original momentum of the particle. This difference in shower development is fundamental for particle identification, as can be observed from the energy over momentum distributions of the two datasets, abundantly shown in the previous section.

To exploit this difference, I implemented a BDT algorithm based on a Gradient Boosting Classifier. The classifier needs a data sample and an input variable. In this case, two samples of electron and pion reconstructed datasets in the case of high pile-up were used, and the training variables were chosen as explained above.

Training parameters In every training, which uses 70% over the input data, the same parameters were used:

1. the number of estimators (decision trees) is chosen to be 200.
2. the maximum depth of a decision tree is chosen to be 3. This parameter controls how deep each individual tree can grow.
3. the learning rate was set at 0.1. This parameter controls the contribution of each individual decision tree to the final ensemble model.

Scoring system A scoring system in the context of machine learning refers to the method or metric used to evaluate the performance of a model. This system determines how well a model's predictions align with the actual outcomes.

By applying a scoring metric, it is possible to compare different models or different configurations of the same model to determine which one performs best.

The chosen scoring metric is the ROC AUC (Receiver Operating Characteristic Area Under the Curve), a classification metric used to assign data points to predefined classes. After training the model, the classifier predicts the probability of each event in the test set belonging to the positive class (electrons). The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings. Finally, the AUC score is the area under the ROC curve. It provides a single value to summarize the performance of the classifier.

4.5 Performance evaluation results

In order to assess the impact of pile-up in electron-hadron discrimination, I trained separately the dataset with $\mu = 1$ and the dataset at high μ. I used the same input variables, which are the energy over the primary momentum (e_{tot}/true_p) in both baseline and downscoped datasets, and the front cluster energy over the total cluster energy (e_F/e_{tot}) only in the baseline configuration.

The performance of both configurations in both cases is shown in Fig. 4.18 where the ROC curves are shown. The orange curve represents the downscoped configuration performance while the blue one represents the baseline configuration performance. Table 4.1 summarizes the information about the scoring values for each model. From these plots it is possible to draw some conclusions: the $\mu = 1$ case obtains better discrimination performance, with a false positive rate (pion efficiency) reaching $\sim 2 \cdot 10^{-3}$ at 80% electron efficiency (true positive rate), while in the case of high pile-up, this value corresponds to $\sim 6 \cdot 10^{-2}$ at the same true positive rate. The value 80% is chosen to compare it to the value of Run 2 shown in Fig. 3.7. The second conclusion regards the performance of baseline and downscoped configuration: the baseline configuration performs better, with a significant difference in the case of high pile-up.

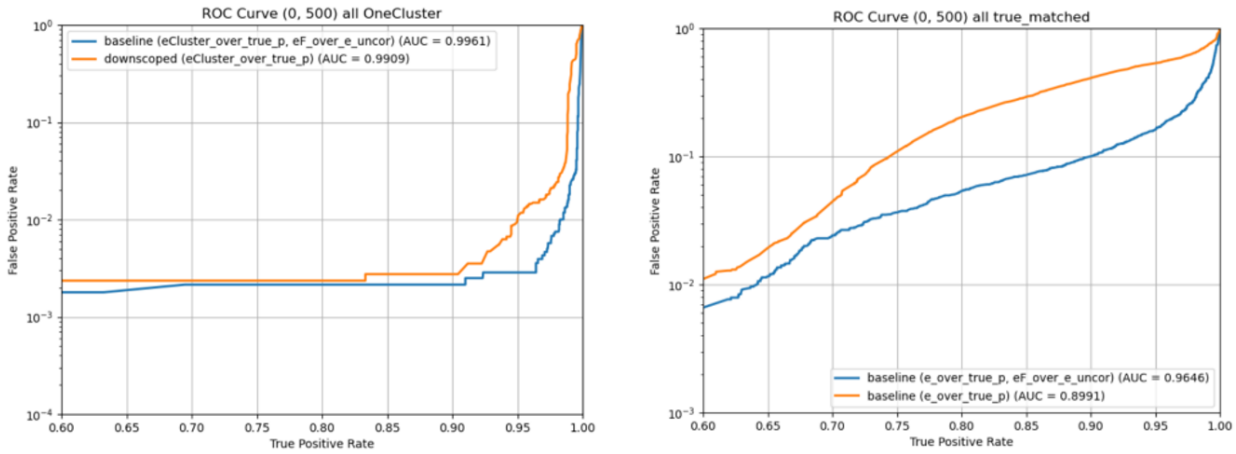

Figure 4.18: Receiver Operating Curve (ROC) for the BDT discrimination algorithm. Training set with $\mu = 1$ (left). Training set with high μ (right).

One note should be done in the case of high pile-up. In fact, in this case, the samples for baseline and downscoped configuration are the same. In order to assess the difference between the two configurations it was decided to use different input variables in the two cases: to simulate the performance of downscoped configuration, only the variable e_{tot}/true_p was used, while for baseline configuration it was decided to add e_F/e_{tot}, so a variable which depends on longitudinal segmentation, whose impact on the downscoped configuration is very little because it is given only by the innermost SpaCal modules.

Table 4.1: Scoring AUC results for BDT training in $\mu = 1$ and high μ conditions.

μ	Configuration	Training set	Input variable	AUC score
1	Baseline	12852	e_{tot}/true_p, e_F/e_{tot}	0.9961
1	Downscoped	11819	e_{tot}/true_p	0.9909
high	Baseline	21352	e_{tot}/true_p, e_F/e_{tot}	0.9646
high	Downscoped	21352	e_{tot}/true_p	0.8991

To improve the algorithm's performance, I incorporated additional training variables into the baseline configuration with a high pile-up. These included the energy of individual cluster cells in the front section of the calorimeter with respect to the total energy of the front cluster (e_{Fcell}/e_F). The inclusion of these variables led to a slight but non-negligible improvement in the algorithm's performance, enhancing its ability to differentiate between pions and electrons even in the presence of pile-up effects. Figure 4.19 shows the comparison between the Receiver Operating Curve (ROC) for BDT applied to the high μ dataset in the baseline configuration and trained with an additional training variable for each curve. The specifics are written in table 4.2.

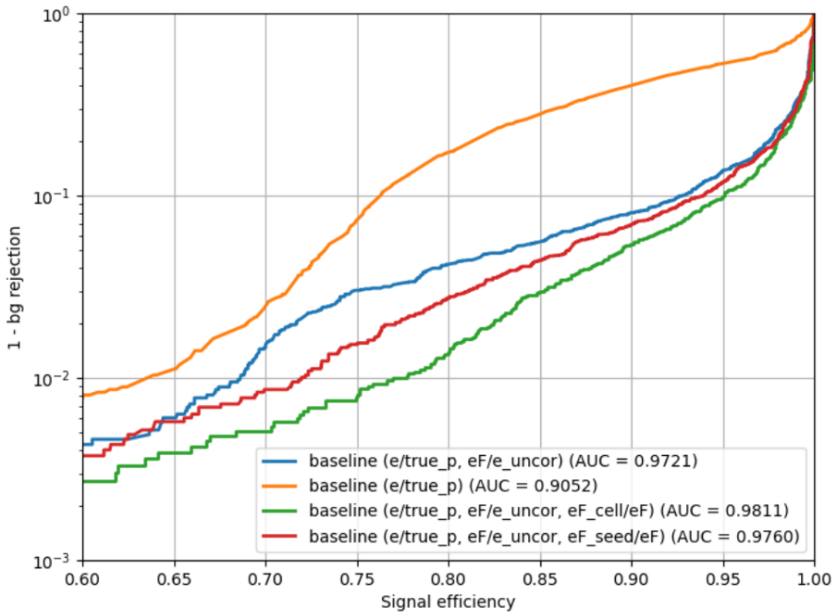

Figure 4.19: ROC curve comparison between BDT models with different training variables.

Table 4.2: AUC score results for baseline configuration in the case of high pile-up with different training variables.

Input variables	Training set	AUC score
e_{tot}/true_p	21352	0.9052
e_{tot}/true_p, e_F/e_{tot}	21352	0.9721
e_{tot}/true_p, e_F/e_{tot}, e_{Fseed}/e_F	21352	0.9760
e_{tot}/true_p, e_F/e_{tot}, e_{Fseed}/e_F, e_{Fcell}/e_F	21352	0.9811

From the green curve, obtained with 11 training variables ($e_{tot}/true\_p$, e_F/e_{tot} and 9 e_{Fcell}/e_F), it is possible to get the pion efficiency at 80% of electron efficiency, that is a little bit over $1 \cdot 10^{-2}$. This result is important because it proves not only that the downscoped configuration produces significantly worse discrimination power with respect to the baseline configuration, but also that the addition of a few input variables in the BDT training results in a discrimination performance that moves closer to the case with pile-up of 1.

The histogram presented in Fig. 4.20 illustrates the output of a BDT classifier applied to a dataset containing pions and electrons. The chosen BDT model is the one with all the input variables de-

scribed above, and which in Fig. 4.19 is represented by a green line. The dataset consists of pions and electrons generated events, mixed together, on which the classifier assigns a probability score, referred to as 'BDT_prediction', indicating how likely a given entry belongs to a particular class, with values ranging from 0 to 1. In order to understand the effectiveness of the chosen model, after the probabilities assignment, the dataset is divided into two parts representing electrons and pions, which are colored respectively in red and blue. The histogram was generated by plotting the frequency of these probability scores on a logarithmic scale. The distribution shows that the BDT model is highly effective at distinguishing between the two particle types, with most pions clustering near a prediction of 1 and most electrons clustering near 0. The minimal overlap between the two distributions in the intermediate range suggests that the classifier generally performs well. Overall, this histogram serves as a visual confirmation of the BDT classifier's ability to discriminate between pions and electrons.

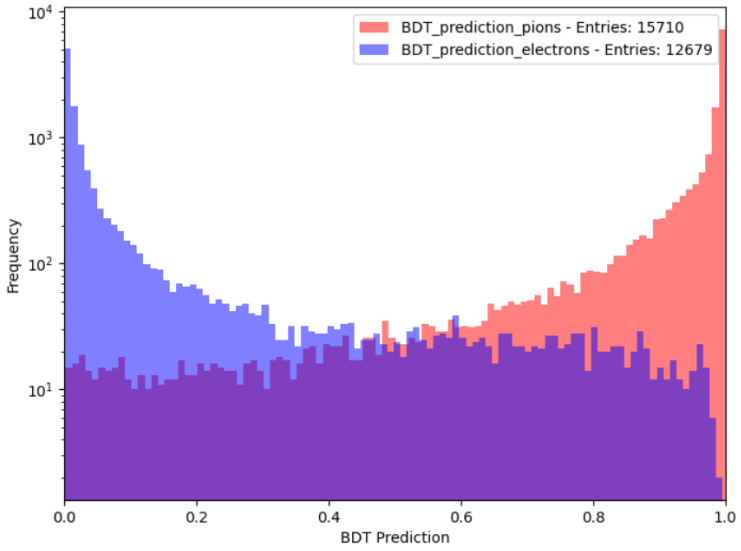

Figure 4.20: BDT prediction histogram with input variables e_{tot}/true_p, e_F/e_{tot}, e_{Fseed}/e_F, e_{Fcell}/e_F applied to pion and electron sample. Electron prediction values are drawn in red, while pion ones are in blue.

In conclusion, the energy over momentum quantity is a reliable variable to discriminate electrons and pions even in the case of high pile-up. Also, the longitudinal segmentation of all modules strongly helps the discrimination efficiency.

4.6 Application to physics process

The study presented in this Chapter does not take into account all the variables present in a physics process. Also, it is limited to the calorimeter information, and therefore it had to use the primary particle momentum, true_p, instead of the momentum measured by the tracking system. However, the model, although simple, can and has been adapted to the analysis of the LHCb benchmark physics process $B^0 \to K^{*0}e^+e^-$ for Upgrade II.

The same BDT model is used, limited to the low p_T range of $0.5 < p_T < 5$ GeV, with training variables whose distribution for this sample is shown in Fig. 4.21, respectively e/p, e_F/e and e_{Fseed}/e_F. Note that the energy of the front cluster lateral cells with respect to the total front cluster energy is not included.

Figure 4.22 shows the comparison between the downscoped and baseline configuration performance of $B^0 \to K^{*0}e^+e^-$ physics event generated with Minimum Bias of $1.0 \cdot 10^{34}$ luminosity.

Although the luminosity is lower with respect to the one simulated for the single particle study in this Chapter, it is clear that the same conclusions could be drawn also for the physics process. Particle identification strongly relies on longitudinal segmentation variables, since the downscoped option could not reach the same performance of the baseline. Moreover, at $\epsilon(ee)$ =80%, a pion identification

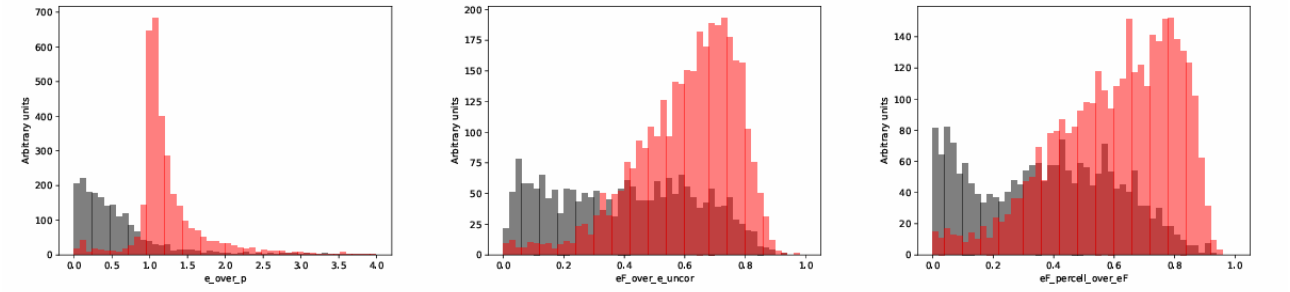

Figure 4.21: Distribution of e/p, e_F/e and e_{Fseed}/e_F for electrons (red) and pions (grey) from $B^0 \to K^{*0}e^+e^-$ event.

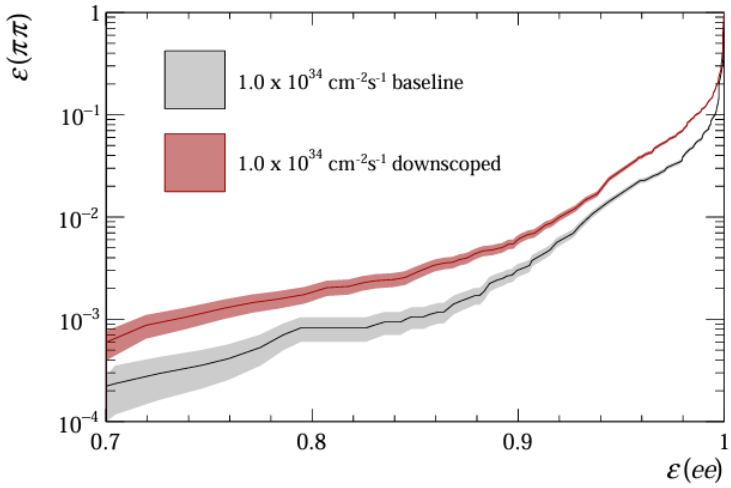

Figure 4.22: ROC curve comparison between BDT model applied to downscoped and baseline configurations for $B^0 \to K^{*0}e^+e^-$. From 2nd ECAL Upgrade II Workshop.

efficiency of $\sim 2 \cdot 10^{-3}$ for downscoped and $\sim 8 \cdot 10^{-4}$ for baseline is obtained. These results reproduce the same reconstruction efficiency obtained in Run 2 and shown in Fig. 3.7.

Chapter 5

Characterization of SpaCal module with simulation and test-beam data

The analysis presented here is based on data collected during the test beam conducted in June 2024 at CERN, focusing on a ECAL prototype: SpaCal with a lead absorber (Pb) and Polystyrene fibers. The primary objective of this study is to determine the energy and time resolution of this prototype.

The second objective is the testing of the performance of the separation algorithm described in Section 4 in the discrimination between electron and pions. Moreover, the energy resolution is computed also with data obtained with a standalone module simulation, in order to compare it to the test-beam results.

In this Chapter, this prototype will be referred to as SpaCal Pb/Poly.

5.1 The experimental setup

The experimental setup is illustrated in Figure 5.1. The reference system of the test-beam facility is defined as follows: the z-axis is aligned with the particle beam direction, with increasing values pointing towards the module, the y-axis is parallel and opposite to the direction of gravity, and the x-axis is chosen to form a right-handed coordinate system. Signal acquisition was triggered by detecting coincidences between two plates of plastic scintillation material. Two multichannel plates (MCPs) were used to provide two timestamps, calculated by a Constant Fraction Discriminator (CFD), whose average serves as the initial reference time t_0.

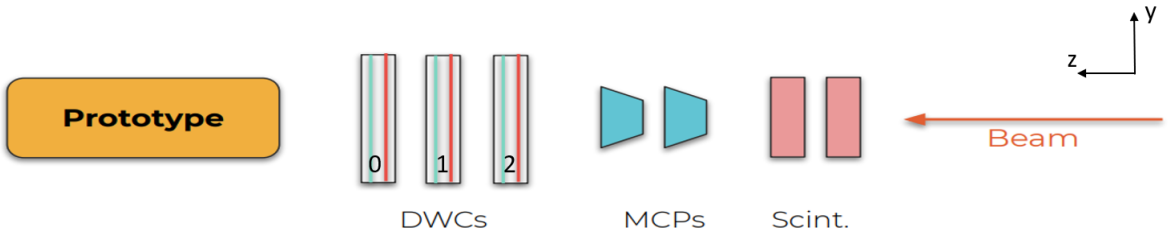

Figure 5.1: Layout of the experimental setup of SPS June 2024 test beam. The particle beam is directed toward the z-axis and hits in order two scintillator pads, two multichannel plates, three delay wire chambers, and lastly the prototype.

Three delay wire chambers (DWCs) were used for tracking information. These chambers, filled with a gas mixture of argon (Ar) and carbon dioxide (CO2), consisted of two grids of wires dedicated to detecting the x and y coordinates separately. In this chapter, the coordinates of the DWCs are denoted as (x_j, y_j, z_j), where the index $j = 0, 1, 2$ identifies the chamber, with 0 being closest to the SpaCal prototype and 2 being the farthest.

The SpaCal prototype, shown in Fig. 5.2, was enclosed in a dark box and could rotate in both azimuthal and polar directions.

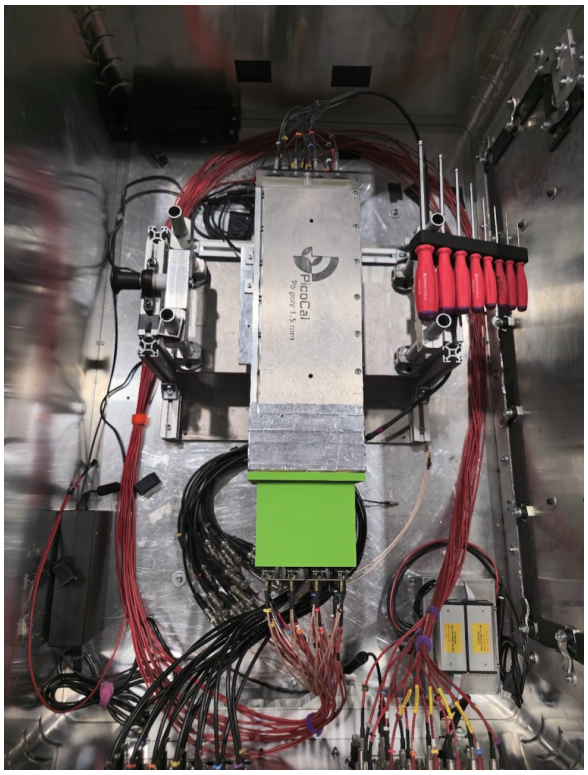

Figure 5.2: Top view of the aluminum box with the Spacal module already set up. The green piece is the 10 cm long light guide which was connected to the back side of the module.

The electron beam, with energies between 20 and 100 GeV, was directed as shown in Figure 5.1. All modules impacted by the beam lay in the xy-plane, orthogonal to the beam direction (z-axis). For the data taking, the prototype was tilted by 3°in both azimuthal and polar directions. This choice is due to the necessity of a compromise between the condition where the particle hits only a few fibers with a great energy deposit and the one where it hits a lot of fibers with a very small energy deposit.

The Data Acquisition (DAQ) system utilized several electronic modules, with trigger logic managed by NIM modules. A CAEN high-voltage power supply, which was controlled remotely via the CAEN GECO2020 interface, was used to tune the detector voltages.

Photodetector signals were processed by one of two VME modules:

- For energy measurements, the LeCroy 1182 ADC module was used, integrating signals over a 400 ns gate.
- For timing measurements, the CAEN V1742 Digitiser sampled waveforms at rates of up to 5 Gs/s with a 500 MHz bandwidth. A custom calibration reduced interchannel jitter to under 5 ps, and attenuators (10 to 70 dB) were applied to match signal levels.

Finally, the analysis of test-beam data was conducted using the software package ROOT [13].

5.1.1 Module geometry

The calorimeter module consists of a parallelepipedal block of 300 mm width and 121.24 mm side length. This block is filled with lead with the exception of small cylindrical holes which run through the longitudinal length of the module. In those holes, Polystyrene fibers are placed. A bundle then is used to group the fibers into 16 cylinders as shown in Fig. 5.3 to test different light collection techniques. A 10 cm long light guide is placed in front of the module as shown in Fig. 5.2. Finally, scintillation photons are collected by a R9880U Hamamatsu Photomultiplier tube [24]. Figure 5.4 shows the fiber bundle and the light guide.

The absorber is divided in the (x,y) plane into 16 squared cells of 1.5 mm side, which corresponds to a collection of holes in the absorber, where the crystals are placed, with a gap of air of 0.05 mm between the crystals and the absorber. Cell pitch, i.e. the distance between centers of two cells, is 2.160 mm. Figure 5.3 shows the front side of the module, both in the picture and in schematization. It is possible to see how the crystals are placed with respect to the absorber.

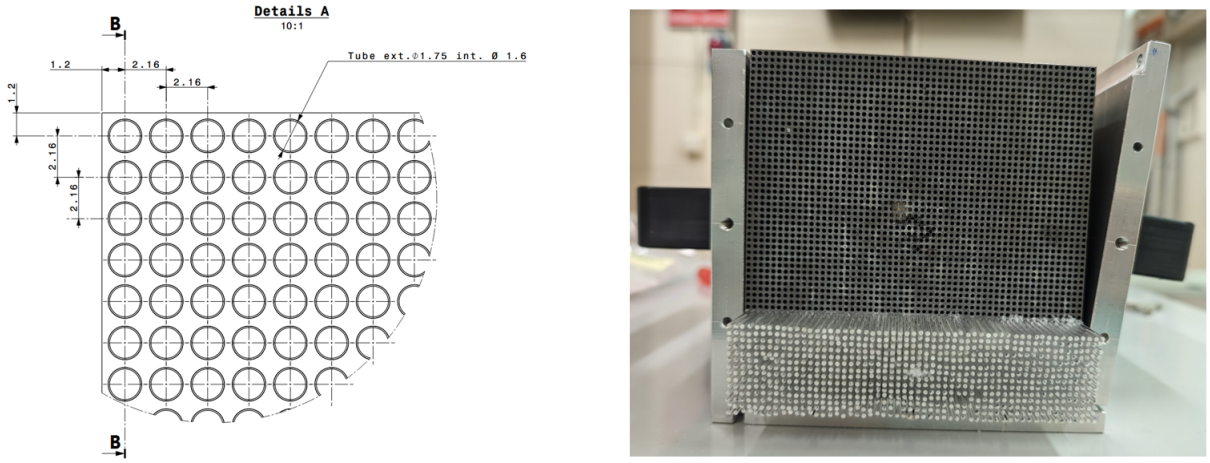

Figure 5.3: Schematic drawing of the front face of the absorber, with measurements (left). Picture of the front face of the prototype absorber while being instrumented with crystal fibers (right).

Figure 5.4: Picture Pb/Poly prototypes while being assembled. The back side of a prototype with visible fibers is separated into 9 cells and grouped with a bundle (left). The back side of the actual test-beam prototype with 16 cells already grouped in bundles (middle). Front view of a light guide (right).

5.2 SpaCal module simulation

In this section, the standalone prototype simulation is described in its main steps and tools. It differs from the LHCb simulation because it does not take into account the interaction of the particles with the other detectors.

5.2.1 The hybrid simulation

The simulation process for the standalone modules is similar to the one described in Chapter 4. However, for comparison with the data, it is more effective to use a particle gun configuration. This approach involves generating particles individually, rather than from a proton-proton interaction, from a plane in front of the module face. This setup removes interactions with other LHCb detectors, thus eliminating the need for Bremsstrahlung recovery.

The steps taken to simulate the module used in the test beam include defining the geometry and calibrating the module, as described below.

- Geometry definition: The simulation of the modules in all their geometrical characteristics is performed using GEANT4. In particular, the module's specifics are described in the modules dedicated section 5.1.1.

- Optical Calibration: The hybrid simulation involves first running a GEANT4 simulation without allowing the transport of scintillation photons with a full ray-tracing library, as previously explained in the Chapter section 4.2. Instead, a calibration technique is used to parametrically model the transport of these photons, from their production point to the readout volume.

  The energy deposits from primary particles to the module are converted into scintillation photons based on the light yield and emission spectrum of the active material. Then, the calibration is used as a probability distribution creating a list of photons with their timestamps. These photons are manipulated by applying characteristics of the chosen photodetector, such as its quantum efficiency, transit time spread, and single photoelectron response. Raw pulses are produced and determine the total signal collected by the module (in photoelectrons) and the timestamp for each detector channel.

  One million primary particles are directed at the module. Calibration points are chosen at 1 mm intervals along the z-axis, with the energy range corresponding to the crystal emission sampling energy: for Polystyrene this is (2.3, 3.5) eV.

  Full ray-tracing simulations indicate that the transport of scintillation photons in the SpaCal Pb/Poly module is essentially complete after 100 ns. This time interval is divided into 2000 parts, resulting in a time binning of 50 ps per bin (for transport only, not for the entire simulation). The emission angle of the photons in the longitudinal direction is divided into 10 parts. Thus, for the Polystyrene module, the input parameters of the calibration, namely the total transport time, the time division, and the emission angle division are set to 100, 2000, and 10.

- Energy Calibration: Energy calibration is conducted by producing a primary particle (electron) beam directed towards the center of the module and comparing the total photoelectrons produced by the simulation with the total energy deposited in the module, that, in the case of single-sided readout modules, corresponds to the beam energy. This allows for parametrizing the relationship between energy and number of photoelectrons. This procedure is essential for establishing an energy scale in the simulation study.

  A critical point regarding the beam source setup: the primary beam's main axis (square, 5x5 mm^2, with a 3+3 degrees stereo inclination relative to the module's main axis, i.e., the z-axis) crosses the xy-plane at (0,0) when $z = -42.72$ mm. This z-value corresponds to the separation between the front and back sections. The (0,0) values for x and y are chosen to maximize lateral shower containment by shooting at the module's center. Although lateral leakage in this module is not expected to be an issue (the module is 8x8 Molière radii wide), energy calibration is generally performed by shooting at the center of a sufficiently wide module.

After completing these steps, the simulation is ready to study the energy and time resolution of the modules. The source for the particle beam is set to a 6x6 cm^2 plane square with a uniform distribution, positioned 250 mm from the module. The beam is centered on one cell at a time and tilted at an angle of 3+3 degrees. Both pion and electron beams are simulated at test beam energies of 20, 40, 60, 80, and 100 GeV.

5.3 Energy calibration

In this section, the procedure to obtain an energy calibration both with test-beam data and simulated data is described. In particular, only one of the four central cells, specifically cell 11, was studied.

5.3.1 Energy calibration of test-beam data

To perform the energy calibration on the prototype, an electron beam at the fixed energy of 40 GeV was shot toward the center of each one of the 16 cells, and 10^4 signal events were taken.

To select the particles that enter a specific SpaCal cell, it is assumed that the (x,y) hit coordinates correspond to the coordinates of the closest DWC to the prototype.

Cuts on the DWC x and y coordinates are applied to select a 10 x 10 mm^2 center portion of each cell. This is done to obtain the best signal uniformity. Figure 5.5 shows the signal amplitude profile in Volts of cell 11 without and with the cut.

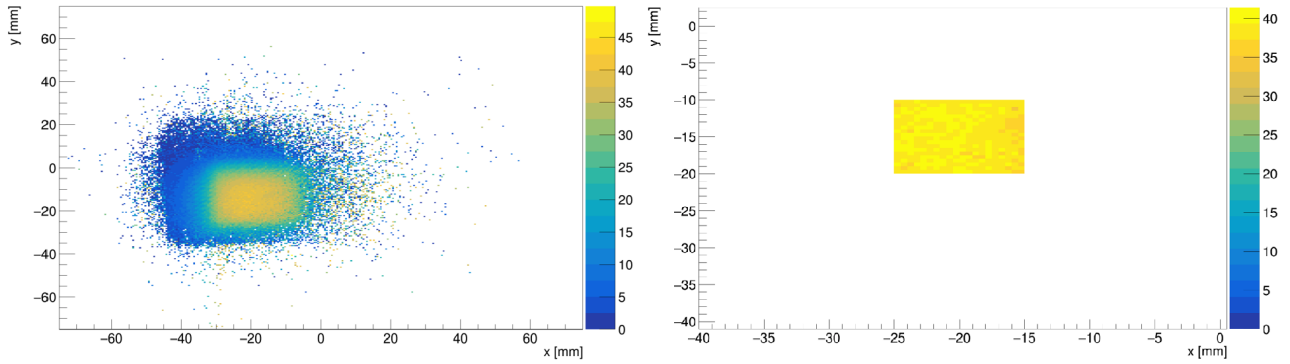

Figure 5.5: Amplitude profile in Volts as a function of the position for cell 11 corresponding to an incident electron energy of 40 GeV. To the right, the same profile is shown with a cut on the position $-25 < x < -15$ mm and $-20 < y < -10$ mm.

To convert the signal amplitude from Volts to GeV, calibration factors for each cell are needed. The signal amplitude distribution of each cell is computed with a test calibration factor of 1, and then fit with a Gaussian function to get its mean value. Calibration factors are then computed cell by cell as the energy of the beam (40 GeV) over the signal mean value, to obtain a quantity expressed in GeV/V.

By repeating the procedure with data taken at different energies, with the beam centered only on one chosen cell (in this case cell 11), it is possible to compute the linearity of the module response after calibration, i.e. the correspondence between nominal beam energy and reconstructed energy after applying calibration factors.

Figure 5.6 shows the energy distribution after the first calibration and the calorimeter response. It is clear that these calibration factors overcorrect the energy, and it is necessary to scale them. The linearity coefficient is given by the reconstructed energy with respect to the beam energy, and it is found to be 1.249, as shown in the figure.

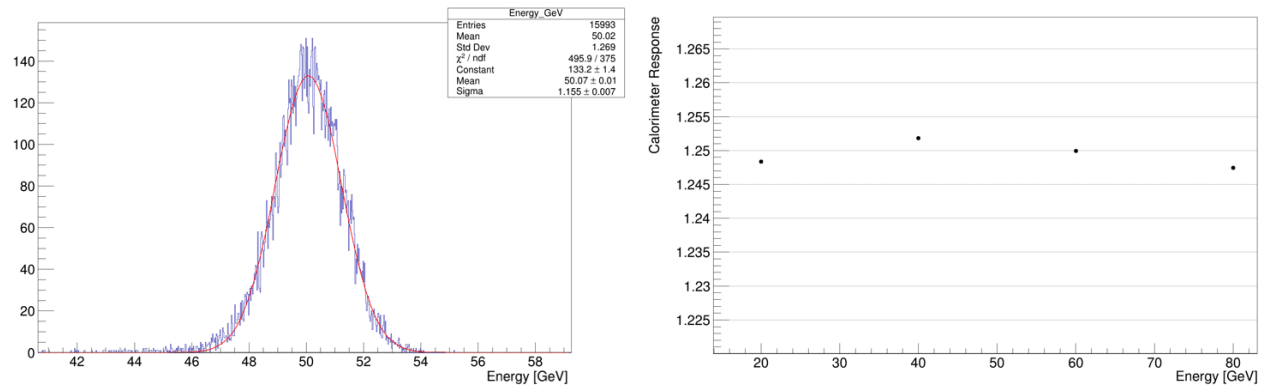

Figure 5.6: Energy distribution curve for a beam energy of 40 GeV after first calibration (left). Calibrated energy with respect to the beam energy, or calorimeter response (right).

The first calibration factors are scaled by 1.249, and the procedure is repeated again. Once it is finished, the energy curve correctly matches the primary energy, and the linearity coefficient is closer to 1 as shown in figure 5.7.

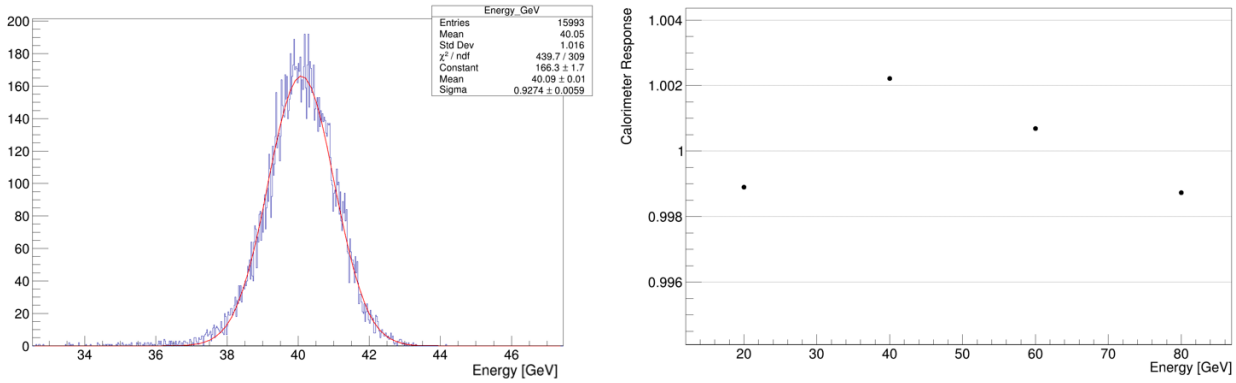

Figure 5.7: Energy distribution curve for a beam energy of 40 GeV after second calibration (left). Calibrated energy with respect to the beam energy, or calorimeter response (right).

Table 5.1 summarizes the first and second calibration factors for each cell, as well as the signal amplitude.

Table 5.1: Calibration factors for each cell of SpaCal Pb/Poly module hit by a 40 GeV electron beam.

cell	amplitude (V)	factors (GeV/V)	corrected factors (GeV/V)
1	1449	0.027605245	0.022101877
2	1398	0.028612303	0.022908169
3	1432	0.027932961	0.02236426
4	1176	0.034013605	0.02723267
5	1531	0.026126715	0.020918106
6	1504	0.026595745	0.021293631
7	1364	0.029325513	0.023479194
8	1314	0.0304414	0.024372618
9	1399	0.028591851	0.022891794
10	1372	0.029154519	0.023342289
11	1382	0.02894356	0.023173387
12	1322	0.030257186	0.024225129
13	1453	0.02752925	0.022041033
14	1384	0.028901734	0.023139899
15	1374	0.029112082	0.023308312
16	1252	0.031948882	0.025579569

5.3.2 Energy calibration of simulated data

The simulation was performed following the steps thoroughly described in section 5.2.1.

The module configuration for the simulation consists of 5 volumes, one absorber, two readouts, and two interfaces placed along the direction of the z-axis as respectively readout, interface, absorber, interface, and readout. Everything is placed in a stainless steel container filled with air. The interface consists of a gap of air with z-length of 5 millimeters each between the photodetectors and the absorber. The readout represents the volumes of the photodetectors, and its z length is 145 mm.

The optical calibration is performed by shooting 10^6 particles towards the module. This step allows us to successfully convert the signal collected from the detector channels to photoelectrons.

Then, the energy calibration allows us to convert the number of photoelectrons produced in the

simulation to the corresponding deposited energy in the module. It is achieved by shooting a single particle beam to the center of the module at different energies (in this case, an electron beam with 1000 particles of 1, 2, 5, 10, 20, 35, 50, and 100 GeV is used). The beam source is a square plane of 5 x 5 mm^2, from where particles are shot uniformly, with a 3+3 degrees stereo inclination with respect to the main z-axis of the module. The beam is centered to (0,0) in the (x,y) plane. This choice is necessary to maximize lateral containment of the shower in the module. Photoelectrons are then converted to energy by fit to a logarithmic function, defined as $y = p_0 \ln(2p_1 + 1)$. This choice is supported by the fact that at low energies the relation between photoelectrons and deposited energy is not linear, while at high energy the logarithmic function behaves as a linear function. The logarithm parameters are assumed to be the same for every cell following the assumption that every crystal has the same light output. The results of the calibration are shown in fig 5.8, which shows the deposited energy on the module with respect to the total number of photoelectrons produced.

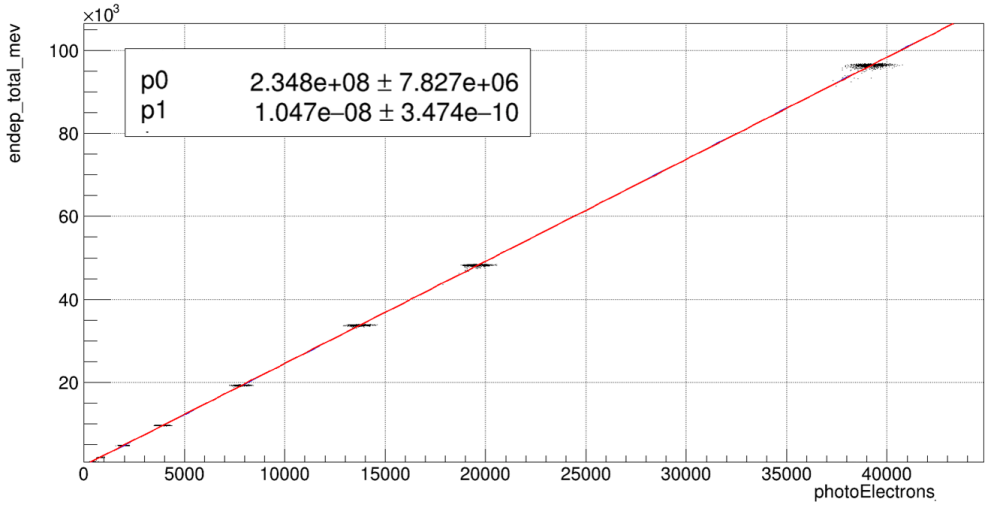

Figure 5.8: Calibration plot for Spacal Pb/Poly. The energy deposited on the module (MeV) was interpolated with the corresponding number of photoelectrons for each beam energy.

Once the energy calibration is completed, a proper energy resolution study can be performed by shooting a single electron beam of energies 1, 2, 5, 10, 20, 40, 60, 80, and 100 GeV. The beam source is again a plane square and tilted by a 3+3 degrees angle with respect to the z-direction, while the center of the beam now is chosen to hit the center of one single cell, in this case, cell 11. This is an arbitrary choice as long as the cell is one of the four central ones, see section 5.4.3.

The energy calibration is applied on each cell signal, and the energy distribution in GeV is fit with a Gaussian function, as shown in Fig. 5.9 for a 40 GeV electron beam.

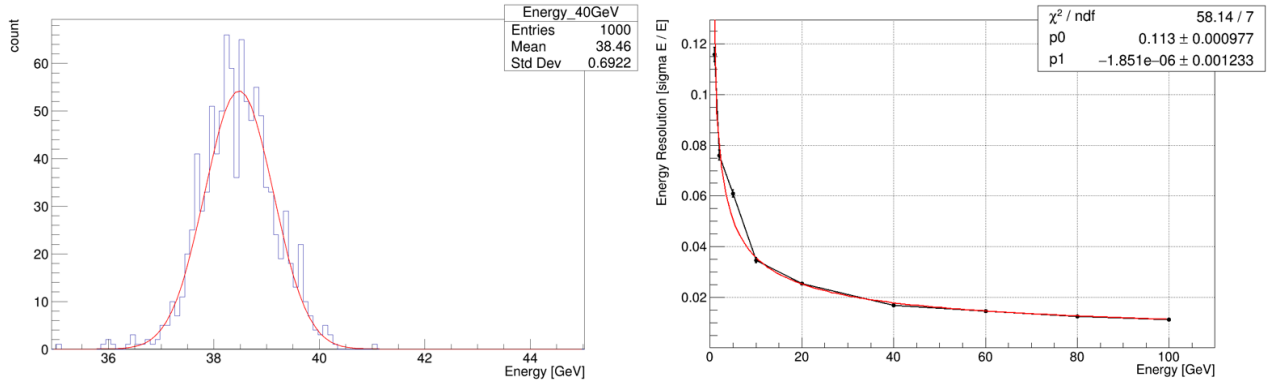

Figure 5.9: Total reconstructed energy distribution from simulated events corresponding to an electron incident energy of 40 GeV (left). Energy resolution plot as a function of energy for simulated electron events (right).

This fit is performed for each energy and its mean values are saved to produce the points for the

energy resolution plot (fig. 5.12). The sampling term results in 11.3% ± 0.1% while the constant term is $1.85 \cdot 10^{-6}$, compatible with zero as expected. In fact, the constant term is generated by detector geometry imperfections, something that is not taken into account in this simulation.

5.3.3 Correction of simulation procedure for detector inhomogeneities

In order to be able to fairly compare the energy resolution obtained with simulated data and test-beam data, it is necessary to "correct" the simulation. In this case, inhomogeneities came from the light guide, which had worse efficiency on the edges.

Looking at each cell of the test-beam prototype, the idea is to identify the response of each fiber, i.e. the number of photoelectrons counted over the number of photoelectrons generated. In the simulation, in order to reproduce real conditions, not all generated photons will be counted in the final photoelectron count, but they will be scaled by the response of the cell.

In the test-beam analysis, it is possible to look at the signal amplitude coming from the cell as shown in Fig. 5.11 with the darker part being the absorber, and the lighter being the fiber of a cell. After removing the background by applying a cut on the amplitudes, the fibers are identified. Each fiber is made by a collection of pixels with a certain amplitude, and an automated process is tasked to keep the seed pixels and discard border pixels.

The pixels group which is kept defines the response of the fiber as the mean value of the amplitude. The figure to the right represents the response map of each fiber of a cell, with its percentage superimposed on the corresponding fiber position. This map corresponds to the response of one cell, and this map is iterated over each cell in order to cover the whole module, under the assumption that light guide inefficiencies are at least homogeneous for each bundle.

Once this feature is implemented in the simulation configuration, the last part of the analysis is repeated, this time with a particle beam as large as the cell dimension. The results, shown in Fig. 5.10, show a constant term much closer to the one in test-beam data analysis. The comparison between the energy resolution curves and the one of the analysis is shown in fig. 5.12 and the procedure to obtain the last resolution curve is described below.

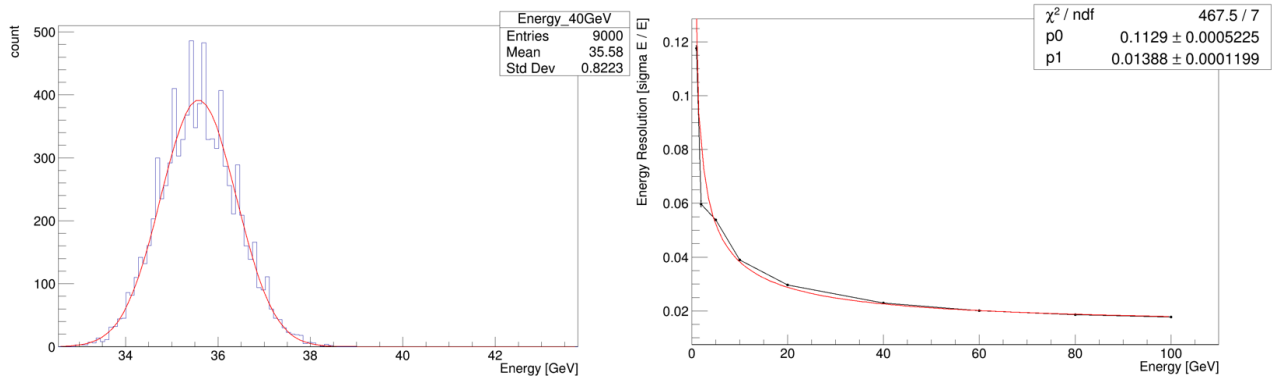

Figure 5.10: Total reconstructed energy distribution corresponding to an electron incident energy of 40 GeV obtained from simulated events with the response map correction (left). Energy resolution plot as a function of energy for simulated electron events with the response map correction (right).

5.4 Results

Here, the prototype performance in terms of energy and time resolution is presented, compared with simulation. Also, particle discrimination performance is studied.

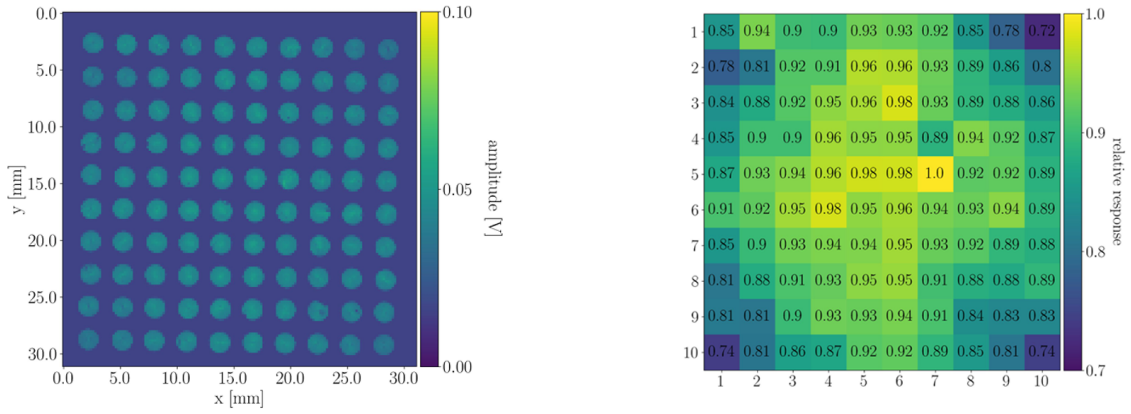

Figure 5.11: Signal amplitude of a cell in the back side of the prototype, the crystal fibers are well recognizable (left). Response map of a prototype cell defined for every fiber (right). From 2nd ECAL Upgrade II Workshop.

5.4.1 Energy resolution

For the energy resolution analysis with test-beam data, a scan of cell 11 was done with different electron beam energies. While doing the analysis, corrected calibration factors were applied, and energy resolution values were fit with the function 2.1 where the noise term was fixed to 0.16 GeV. The obtained result:

$$\frac{\sigma_E}{E} = \frac{(12.5 \pm 0.1)\%}{\sqrt{E}} \oplus \frac{(0.16 \pm 0.01)}{E} \oplus (1.12 \pm 0.03)\% \tag{5.1}$$

is compatible with the results of the simulation with the effect of the fibers inhomogeneities as shown in figure 5.12.

In the 5.12 plot, it is possible to see various energy resolution curves as a function of the energy. The red line, which fits the red marker points, is the result of the test-beam data analysis. Green lines with a 2σ confidence band (i.e. 95% confidence level) are the simulation results for energy resolution. The dark green line corresponds to simulation results obtained without taking into account light guide inhomogeneities, as described in section 5.3.2. This has a constant term which goes to 0 towards high energies. This result is expected because the constant term is connected to detector imperfections, which are not present in the simulation. The light green band, instead, is the simulation result for energy resolution obtained with the correction on light guide inhomogeneities described in section 5.3.3. This correction allows the inclusion of realism in the simulation, and the result is compatible with the one obtained from test-beam data.

Finally, the dashed grey line represents the actual ECAL energy resolution, as stated in 2.2, and corresponds to the threshold requirement for Upgrade II.

Overall, SpaCal Pb/Poly performance is compatible with simulation results, while they are not yet as close to the required energy resolution.

5.4.2 Time resolution

Time resolution is a new important feature of the LHCb ECAL upgrade, as described in Chapter 3. In a high pile-up environment, where it will not be easy to distinguish particles coming from different bunches, the addition of time information can be a crucial aspect to their association with tracks.

For this analysis, an electron beam was shot directly into the center of cell 1 with energies from 20 to 100 GeV.

As for the energy resolution, it is necessary to apply a spatial cut to select only events that hit the cell, which measures 30 x 30 mm^2. The cut is applied to the coordinate of the closest DWC to the prototype.

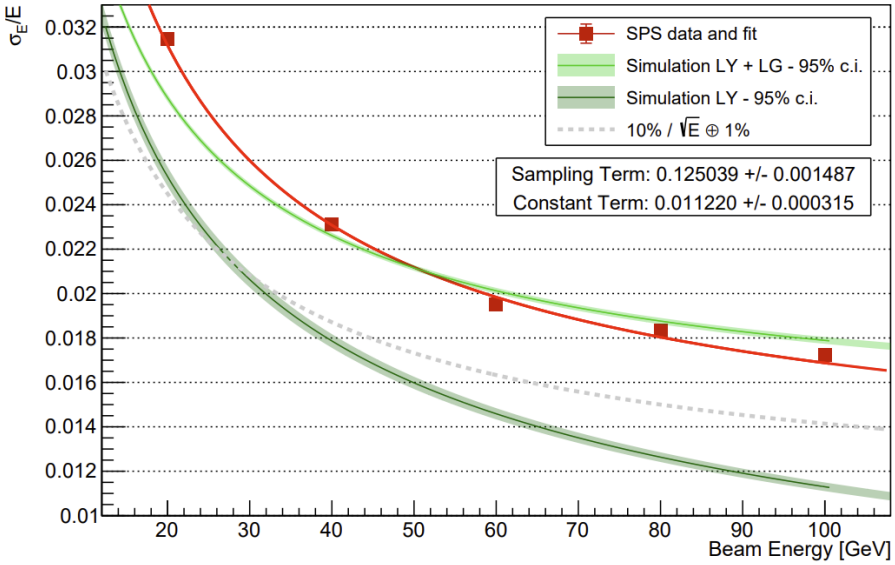

Figure 5.12: Energy resolution curves obtained with test-beam and simulated data. In red the points obtained by test-beam data analysis and fit by an exponential curve with sampling and constant terms written in the plot. In dark green, there is the fit to simulated data obtained without any inhomogeneities correction surrounded by a band which corresponds to 2σ. In light green, there is fit to simulated data obtained with inhomogeneities correction surrounded by a band which corresponds to 2σ. The dashed grey line is the current energy resolution of the ECAL.

The first quantity that is measured is the time reference, given by the two MCPs. The distribution of the difference between the timestamps of the second and first MCPs in coincidence fit with a Gaussian curve. Its standard deviation σ_{fit} is then the resolution of the initial time.

Assuming that both MCPs have equal time resolution ($\sigma_{t_1} = \sigma_{t_2}$), the time resolution for a single MCP can be expressed as $\sigma_{t_1} = \sigma_{\text{fit}}/\sqrt{2}$. The initial time t_0 is calculated as the average of the two MCP timestamps $\left(t_0 = \frac{t_1+t_2}{2}\right)$. Consequently, the resolution on t_0 is $\sigma_{\text{MCP}} = \frac{\sigma_{\text{fit}}}{2}$.

A time resolution of $\sigma_{\text{MCP}} = 19$ ps is obtained for electrons within the spatial range $x_0 \in [44, 50]$ mm and $y_0 \in [37, 43]$ mm.

Finally, the time resolution of the SpaCal prototype is calculated using the following procedure. For each event, the initial time t_0 is subtracted from the time t recorded by the SpaCal prototype. A Gaussian fit is applied to the resulting distribution, allowing us to determine the time resolution of the prototype as $\sigma_{\text{SpaCal}} = \sqrt{\sigma_{\text{fit}}^2 - \sigma_{\text{MCP}}^2}$, where σ_{SpaCal} represents the intrinsic time resolution of the prototype, σ_{fit} is the standard deviation from the Gaussian fit, and σ_{MCP} is the average time resolution of the MCPs as calculated earlier.

The SpaCal time resolution as a function of beam energy is shown in Fig. 5.13.

The time resolution reaches $\sim$ 36 ps at 100 GeV. This result is not auspicable since the requirement for Upgrade II is to have a time resolution under 15 ps.

However, there is still room for improvement. Light guides have been demonstrated to be a great source of time resolution worsening, coming both from their inefficiencies and from the air gap created between them and the PMT photocathode and crystal fibers. The possibility of gluing the respective interfaces is taken into consideration.

5.4.3 Electron-pion discrimination

Other than time and energy resolution, another important metric to evaluate the performance of the LHCb calorimeter prototype is particle identification. We are interested to see the electron and hadron distributions, and how well they are separated.

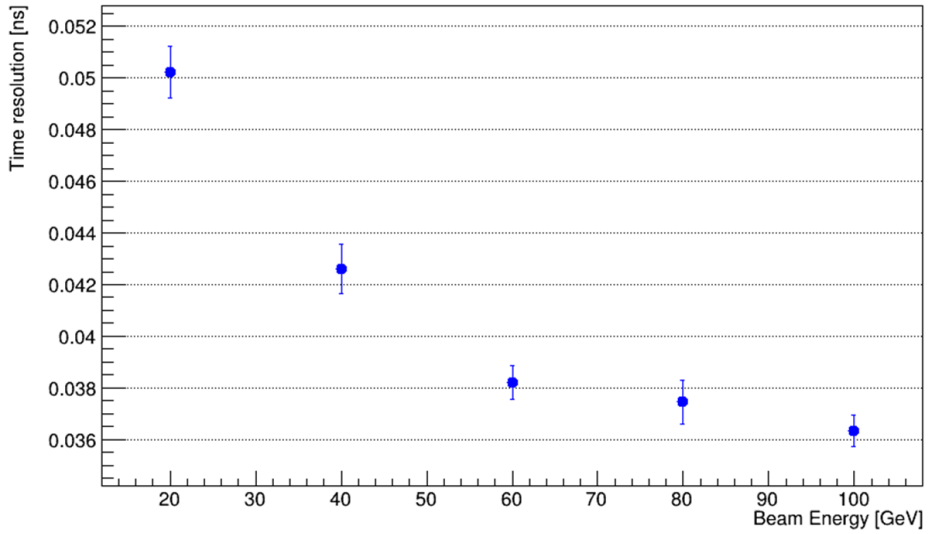

Figure 5.13: Time resolution plot for the SpaCal Pb/Poly module obtained by test-beam data.

To do so, pion and electron beams with different energies were shot toward the center of cell 11.

Since the module consists of 16 cells, a clustering algorithm should be used. After the signal was calibrated (5.3.2) to obtain the energy in GeV, a cut was applied to the track x and y position on the last DWC such that only the cell 11 was selected. It is possible to show that a particle hitting the module in a cell could deposit the majority of its energy also in other adjacent cells. In the full ECAL reconstruction, clusters are squares on the ECAL plane formed by 3x3 cells with the seed cell, namely the cell with the most energy deposit, being in the center. Therefore, in the case of the prototype, the energy deposited on all the cells was saved event per event, if the cell with the most energy deposit was in the four central cells (6,7,10,11), a cluster would be formed around it. An example is shown in figure 5.14, where central cells, which are possible seed cells, are underlined in red. The seed cell corresponds to cell 11, and its 3x3 cluster is highlighted in green. The total energy of the cluster is then saved.

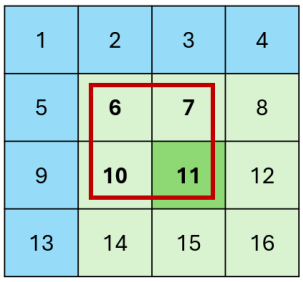

Figure 5.14: Schematization of the cells in the back side of the prototype. Underlined with the red square there are the allowed seed cells. In green, there is a cluster example centered in cell 11.

The energy over momentum distribution is computed for both pions and electrons, using the beam's primary energy as the primary momentum, resulting in an optimal separation between the two particles, since a simple cut on this variable is able to separate at 100% of the particles. A side note should be said about the overestimation of the electron energy with respect to the beam energy. This is due to an imprecision in the computation of the calibration factors. They were computed on the assumption that all the energy of the beam was deposited in the cell where the beam was centered, but this is not true, as fig. 5.15 shows.

Comparing e/e_{beam} obtained by test-beam data with the simulation data, whose e/e_{beam} distribution for electrons is shown in Fig. 5.17, it is possible to conclude that the energy is well reconstructed, as the peak is well defined.

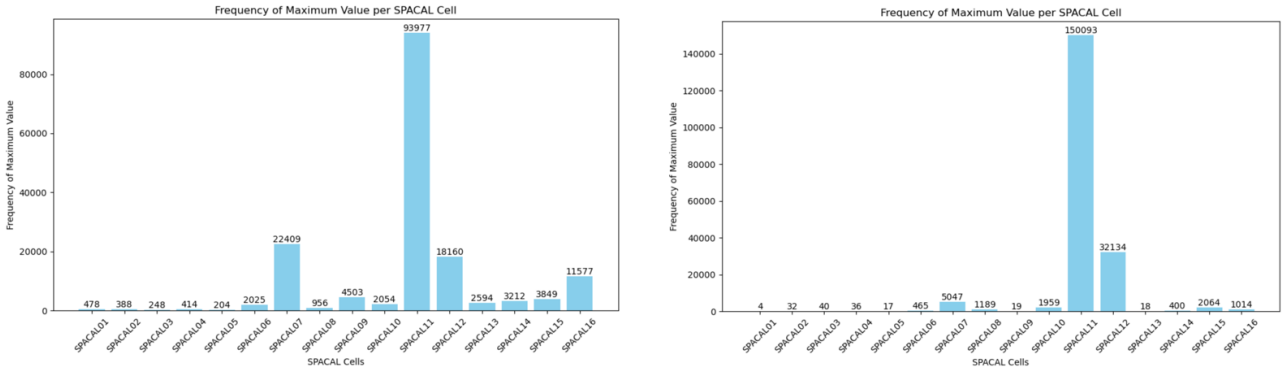

Figure 5.15: Number of events that deposited the maximum energy in each cell of the SpaCal prototype. To the left, there is the distribution associated with the pion beam at 100 GeV, and to the right the distribution associated with the electron beam at 100 GeV.

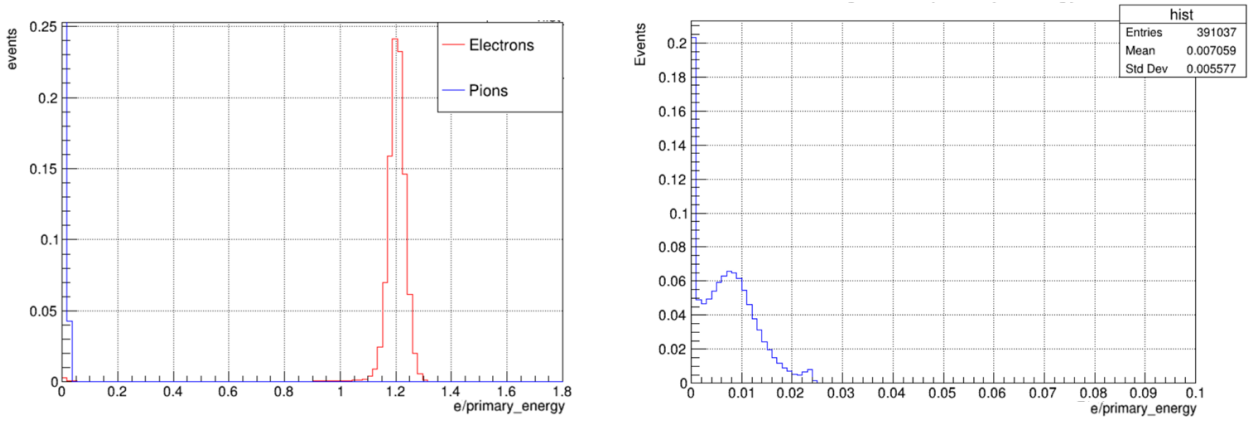

Figure 5.16: Reconstructed energy distribution on the 3x3 cluster over the primary energy of the beam for electrons (red) and pions (blue), and zoom of pion distribution.

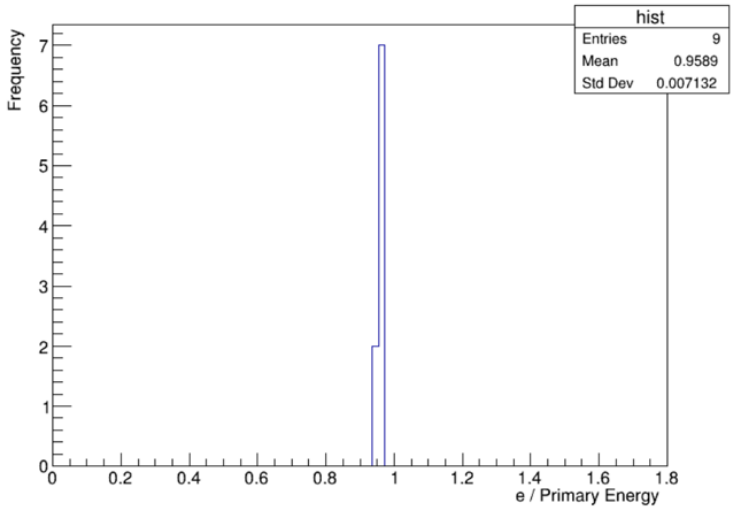

Figure 5.17: Reconstructed electron energy over beam energy with prototype simulation.

Other distributions, such as e_{cell1}/e and e_{cell6}/e, shown in fig. 5.18, were also plotted to check the results obtained in Chapter 4. The distribution behavior is the same as the one showed in Chapter 4, with two very distinct distributions for e_{cell1}/e and two distributions peaking near 0 for all the other cells, in this case, it is showed only e_{cell6}/e.

To conclude, test-beam data on the SpaCal Pb/Poly prototype obtained with pion and electron beams

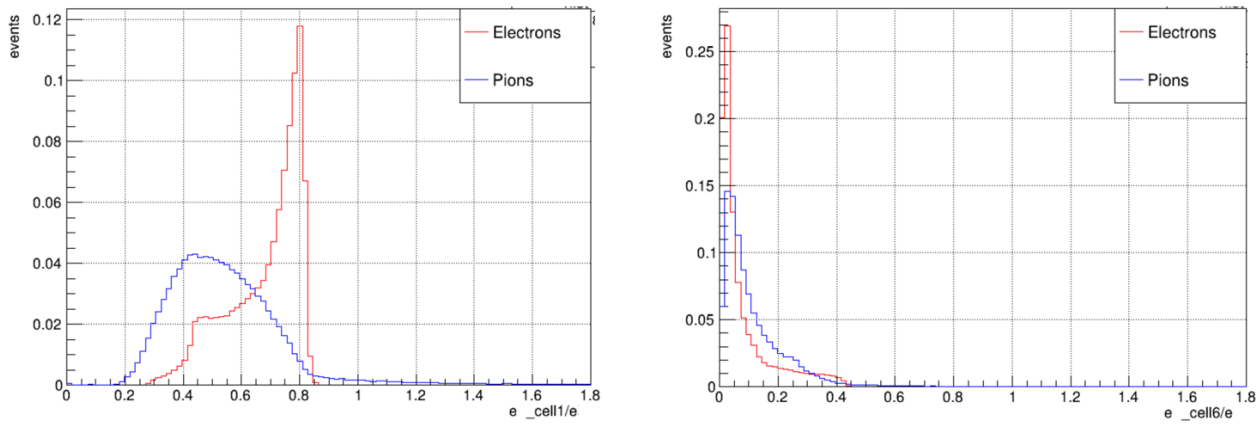

Figure 5.18: Distribution of cluster quantities for electrons and pions. (Left) Reconstructed energy distribution of the cluster seed cell over the total energy of the cluster. (Right) Reconstructed energy distribution of the middle-right cell of the cluster with respect to the total cluster energy.

can reproduce the separation performance distribution. However, this result is still limited by the poor statistics and energy range of pion beam measurements. The next step should point towards the same study but with a longitudinal segmented module, in order to check also the distribution of variables related to front and back cluster energy. This is expected to lead to some additional complications, especially in the energy calibration procedure, but with double-sided readout modules, we should be able to effectively test the PID performance of a veritable module of the upgraded ECAL.

Chapter 6

Reconstruction of $Z \to e^+e^-$ with calorimeter upgrade

This Chapter presents the reconstruction of the invariant mass of the Z boson and other relevant quantities from its decay channel in an electron-positron pair, $Z \longrightarrow e^+e^-$, obtained from simulation of Run 5 conditions.

The importance of reconstructing electrons and the performance of Z boson production at LHCb with the current configuration are thoroughly described in Chapter 1. The study here presented aims to assess the capabilities of the future upgraded calorimeter and eventually give an idea of the impact of high luminosity conditions and how this issue could be overcome.

This Chapter is structured in the following way:

- Section 1 is dedicated to the generation of Z events with the simulation framework
- Section 2 refers to Z event reconstruction and the association of ECAL clusters to its corresponding Monte Carlo event
- In section 3 the analysis of the reconstructed events is presented, first performed with requirements on p_T and $M(ee)$ and then with cuts based on BDT discrimination algorithm predictions.

6.1 Generated events

The samples necessary for this study were generated using the simulation tools already described in Chapter 4. Specifically, 9400 $Z \longrightarrow e^+e^-$ signal events were generated with PYTHIA8, with the requirement of both electron and positron to be in the detector acceptance and to have transverse momentum $p_T > 10$ GeV.

After applying the requirement on $p_T > 20$ GeV, the number of Z candidates goes down to 8634. This is due to the generator cut at the lowest p_T. The distributions of some relevant quantities were investigated, such as the invariant mass or the transverse momentum.

Figure 6.1 shows the transverse momentum distribution (in MeV) and the distribution of pseudorapidity, which covers all the angular acceptance of the LHCb detector ($2 < \eta < 5$).

Finally, fig. 6.2 shows the distribution of the invariant mass obtained with the generated sample. The distribution has a well-defined peak with a mean value of 90.91 GeV and 4.85 GeV resolution.

6.2 Event reconstruction

The ECAL response to the generated sample consists of 10^4 events and it is merged with the response to a Minimum Bias sample at $1.5 \cdot 10^{34}$ cm^{-2}s^{-1} luminosity, obtaining 6488195 particles which deposits

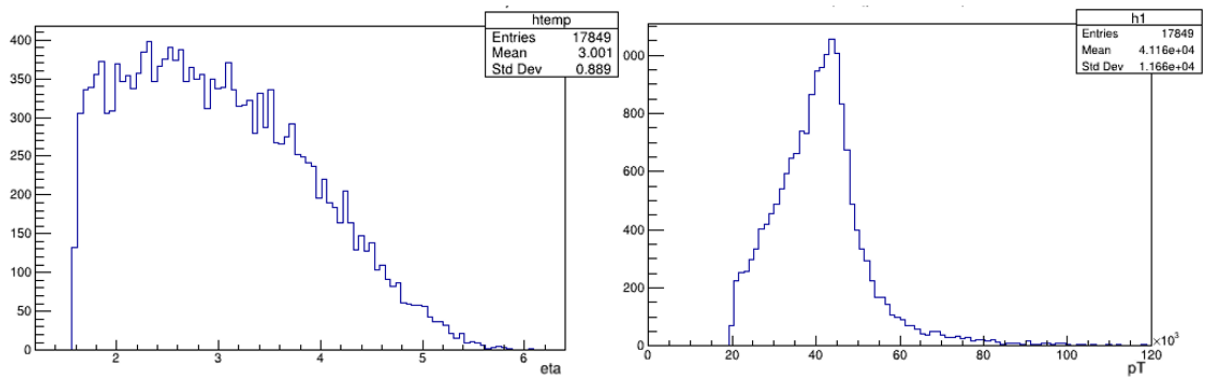

Figure 6.1: Pseudorapidity η and transverse momentum p_T (in MeV) distributions for generated $Z \to e^+e^-$ events.

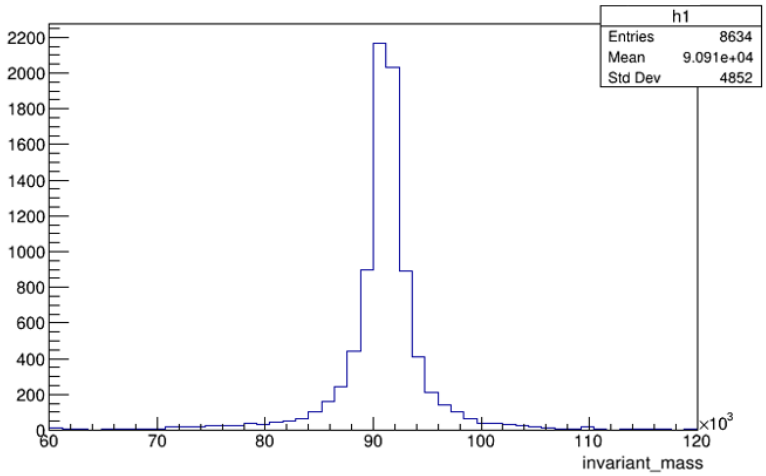

Figure 6.2: Invariant mass distribution in MeV for for generated $Z \to e^+e^-$ events.

energy on the calorimeter.

The energy is reconstructed by the Electromagnetic Calorimeter in its Run 5 configuration with the framework defined in Chapter 4 and without Bremsstrahlung recovery. In the case of Z boson decay in an electron-positron pair, the Z invariant mass is defined as such:

$$M = \sqrt{(E_1 + E_2)^2 - (p_{x1} + p_{x2})^2 - (p_{y1} + p_{y2})^2 - (p_{z1} + p_{z2})^2} \tag{6.1}$$

where $E_{1,2}$ and $\vec{p}_{1,2}$ defines respectively the energy of the two products, electron and positron, and their trimomentum. Both the energy and the momentum information come from the reconstructed energy deposited on the calorimeter, while, in standard procedure, the momentum comes from the LHCb tracking system.

Once the events are generated, in order to be able to correctly evaluate the reconstruction performance of ECAL, it is necessary to define a procedure to match the reconstructed cluster on ECAL to its corresponding generated particle. In Chapter 4, this procedure relied on the minimum distance between the reconstructed cluster and the particle position registered just before the calorimeter. This time, a new method was chosen, which is described in the section below.

6.2.1 Cluster matching with Monte Carlo particle

In order to extrapolate the candidates that correspond to the Z event, it was decided to follow a procedure to match the generated electrons with their corresponding cluster information (energy,

position). This match is necessary to remove pile-up clusters and study Z properties without a combinatorial background.

Event per event, the electron momentum direction is individuated by looking at the (η,ϕ) plane.

Taking into account the position vector (x, y, z) of the calorimeter cluster and the momentum vector (p_{x1}, p_{y1}, p_{z1}) at the production vertex, the difference in pseudorapidity η and the difference of the angle ϕ can be calculated, which is defined as the angle between the two particles produced in the transverse plane and its value is defined with the formula $\cos\phi = \frac{p_1 \cdot p_2}{p_{T1} \cdot p_{T2}}$.

The direction of the particle, when it hits the calorimeter and the direction of the generated electron, are defined respectively as $p_{Cluster} = \frac{z}{\sqrt{x^2+y^2+z^2}}$ and $p_{gen} = \frac{p_z}{\sqrt{p_x^2+p_y^2+p_z^2}}$, and the values $\Delta\eta$ and $\Delta\phi$ will be defined as such:

$$\Delta\eta = \eta_{Cluster} - \eta_{gen} = -\ln(\tan(\theta_{Cluster}/2)) + \ln(\tan(\theta_{gen}/2))$$

$$\Delta\phi = \phi_{Cluster} - \phi_{gen} = \arccos(\frac{p_{Cluster1} \cdot p_{Cluster2}}{p_{TCluster1} \cdot p_{TCluster2}}) - \arccos(\frac{p_{gen1} \cdot p_{gen2}}{p_{Tgen1} \cdot p_{Tgen2}})$$

Knowing the momentum of the two electrons close to the origin vertex from the generator information, it is possible to construct a cone with an apex on the production vertex. The aperture of the cone is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ and, assuming that the generated particle passes through this cone, the ΔR factor can put a constraint on the angular acceptance, selecting the particle that hits the calorimeter by increasing or decreasing its value.

Looking at the distribution of ΔR in Fig. 6.3 it appears that a suitable requirement could be $\Delta R < 0.1$ since there are two visible distributions. When more than one pair satisfies this requirement, the cluster with closer energy to the primary particle energy is chosen.

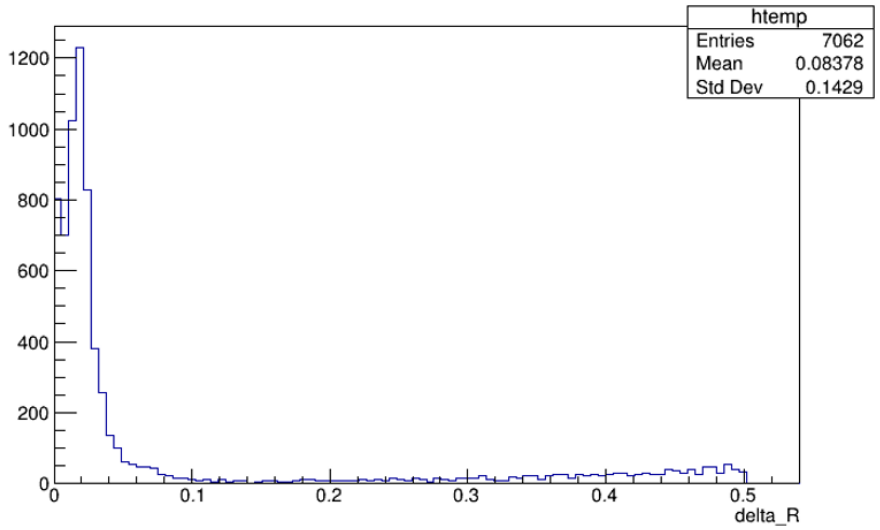

Figure 6.3: ΔR distribution obtained by matching the generated and reconstructed electron candidates for Z event.

Before narrowing this interval by applying the $\Delta R < 0.1$ cut, it was chosen to look at the transverse momentum distribution, and requiring a cut on the reconstructed $p_T > 20$ GeV. Once doing so, the event number drops at 6944, and after checking, this requirement corresponds to the $\Delta R < 0.1$ cut.

Figures 6.4 and 6.5 shows some distributions related to the reconstructed sample after the requirement on p_T such as the transverse momentum (in MeV), the pseudorapidity, $\cos\phi$, and the number of clusters per event.

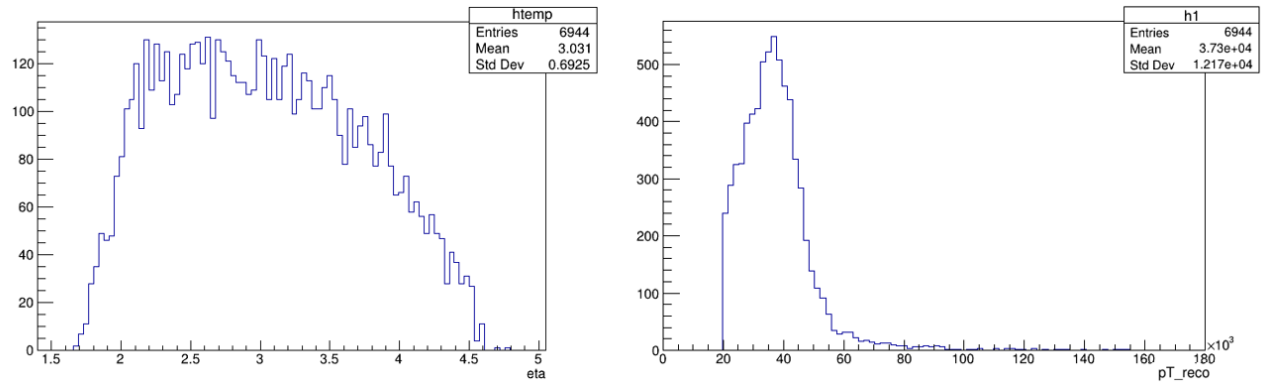

Figure 6.4: Pseudorapidity η and transverse momentum p_T (in MeV) distributions for reconstructed $Z \to e^+e^-$ events.

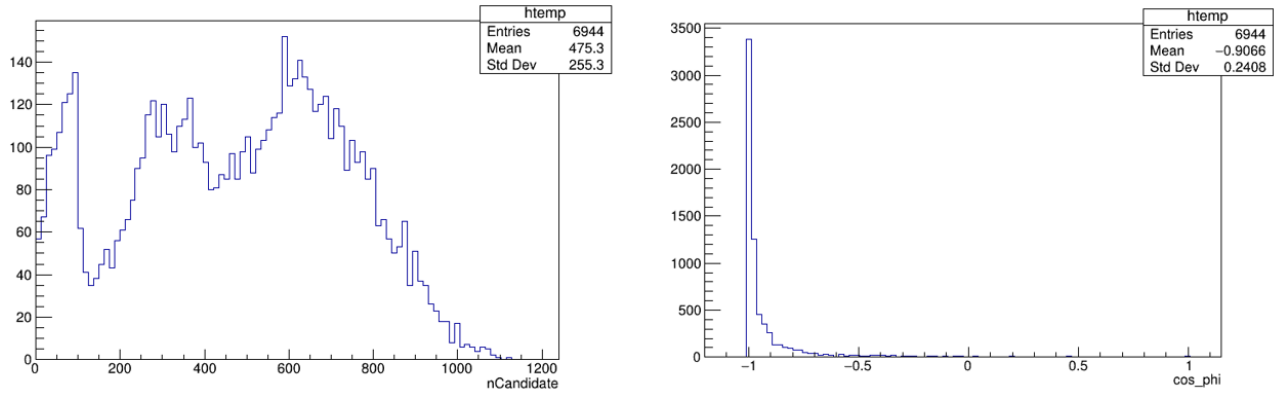

Figure 6.5: Cluster number distribution (left) and $\cos\phi$ distribution (right) for reconstructed $Z \to e^+e^-$ events.

Figure 6.6 shows the invariant mass distribution for the reconstructed sample after the requirement on p_T.

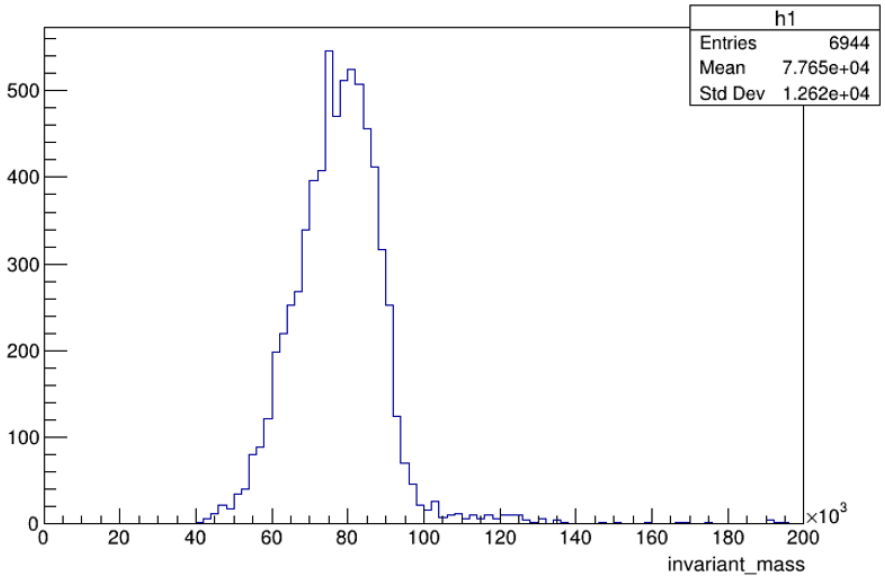

Figure 6.6: Invariant mass distribution in MeV for for reconstructed $Z \to e^+e^-$ events.

The peak efficiency is defined as the number of reconstructed and matched Z events over the total number of Z events generated. Those values are respectively 3472 and 9400, resulting in a 37% efficiency. This is most probably due to the chosen generator cut on $p_T > 10$ GeV.

The invariant mass peak corresponds to 77.650 GeV, with a resolution of 12.7 GeV. Although this value is not compatible $M_Z = 91.1880 \pm 0.0020$ GeV of 2024 PDG average [38], this result was anticipated because of the lack of Bremsstrahlung recovery.

It could be useful to look at the distribution of the generated p_T and the reconstructed particle transverse momentum p_{Treco}. Figure 6.7 shows this distribution from which it is clear that the generated momentum is underestimated after the reconstruction.

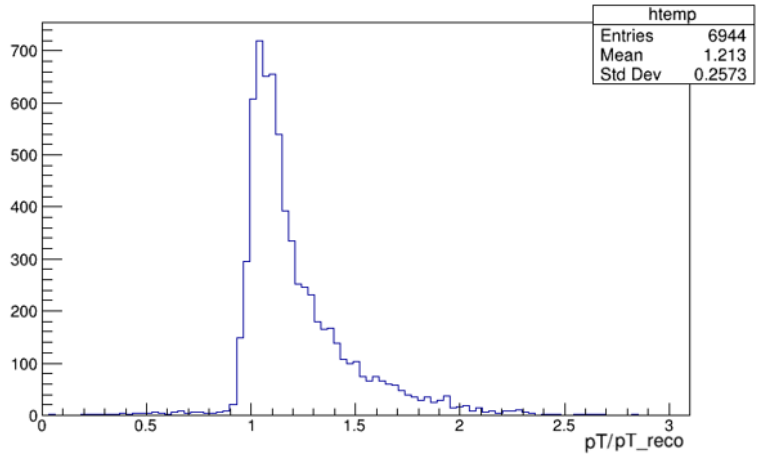

Figure 6.7: Generated events p_T distribution with respect to the reconstructed p_T (p_{Treco}) for $Z \to e^+e^-$ events.

Comparing FWHM/2, i.e. Full Width at Half Maximum, of the invariant mass plot and of the one obtained with LHCb data of Run 1 (fig. 1.7 in Chapter 1), it is possible to conclude that in Run 5 conditions we have a ∼17 % improvement on the peak resolution.

Energy distributions Here some energy distributions are presented for the reconstructed events. The figure of merit e/true_p, shown in Fig. 6.8 is used to assess the calorimeter energy reconstruction performance. This distribution peaked around 1 but with a left tail towards lower values. This means that the calorimeter is able to recover all the energy of the primary electrons only for some events, but for others, it is underestimating it.

Another relevant figure is the distribution of the front cluster energy with respect to the total energy reconstructed by the calorimeter e_F/e_tot. The electron distribution follows the expected behavior, already seen in Fig. 4.16, with the two distinguished distributions representing the difference in SpaCal and Shashlik modules reconstruction.

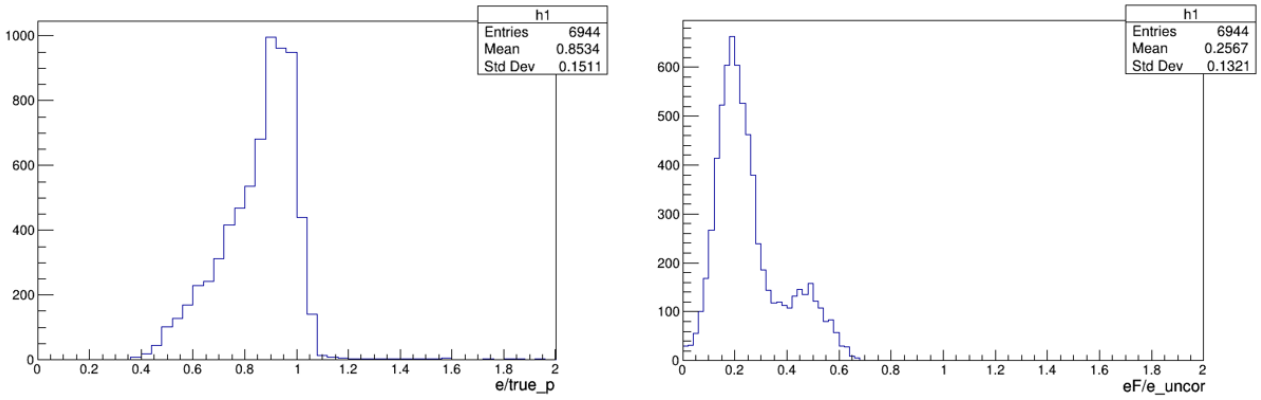

Figure 6.8: Cluster energy distribution with respect to primary particle momentum true_p (left), front cluster energy distribution with respect to the total deposited energy (front and back) (right) for reconstructed electrons from $Z \to e^+e^-$ event.

The distributions of the seed cell of the front cluster e_{Fseed}/e_F and the other cluster cells e_{Fcell}/e_F

with respect to the total energy of the cluster, shown in Fig. 6.9, have the same behaviour obtained with the generated single particle sample in Chapter 4.

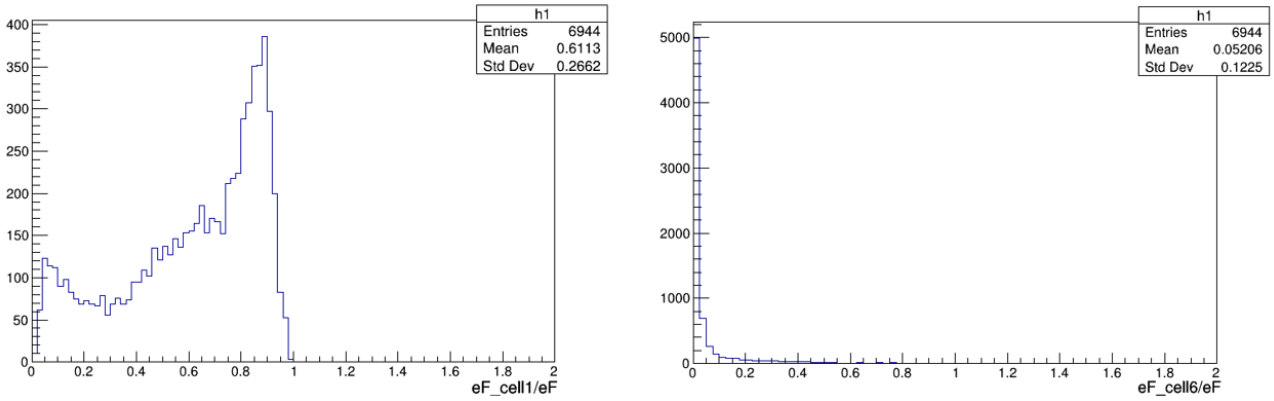

Figure 6.9: Distribution of front cluster seed cell energy with respect to front cluster energy (left), distribution of front cluster cell 6 (middle, right) energy with respect to front cluster energy (right) for reconstructed candidates from $Z \to e^+e^-$ event.

These distributions validate the results of the cluster reconstruction algorithm presented in Chapter 4.

6.3 $Z \to e^+e^-$ candidate selection

In this section, the analysis procedure for the identification of Z events is described. In the first instance, preliminary requirements on transverse momentum and dielectron invariant mass are used to filter the reconstructed data. Then, a BDT discrimination algorithm is applied to test its performance on electron candidates for the Z event.

6.3.1 Preliminary selection

In order to properly identify Z events from reconstructed data with pile-up, it is necessary to apply some preliminary cuts.

As explained in the previous section, the whole dataset consists of 10^4 events with 6488195 particles which deposit energy on the calorimeter. The number of cluster distributions is shown in Fig. 6.10. The high cluster number is given by the presence of pile-up events in the sample.

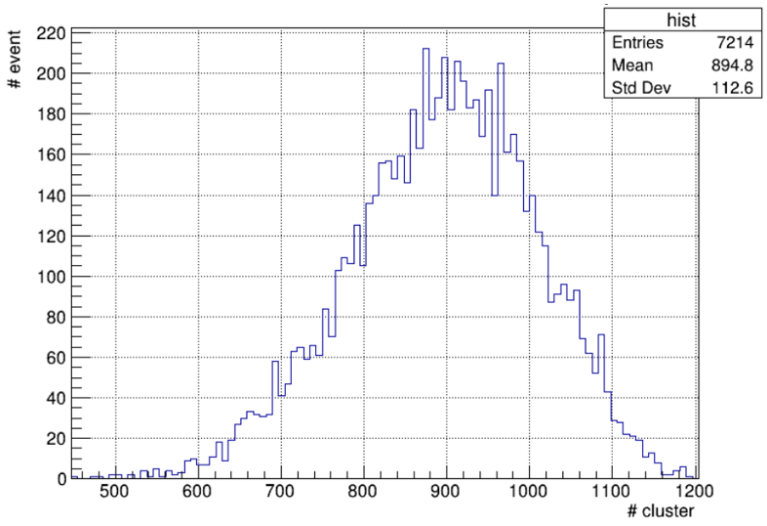

Figure 6.10: Number of clusters per event distribution for reconstructed data from Z production with pile-up.

The first requirement is on the transverse momentum, and it is keeping only particles that satisfy $p_T > 20$ GeV. This is a strong requirement that can lower the cluster number to 9949, around 1.5 ‰ of the original sample.

To apply the second requirement, the remaining dataset has to be manipulated in order to create pairs. Event per event, particles are paired together, and some quantities of interest are investigated, together with the computation of invariant mass for each combination. Now the cut can be applied, selecting only events with $M(ee) > 40$ GeV. This requirement justification is in the previous section 6.2.

The number of pairs after the two cuts is 3696. The invariant mass mean value obtained with this dataset is 77.65 GeV and the resolution is 13.84 GeV. The efficiency is 39%, which is higher than the 37% obtained with matching to the MonteCarlo events. This is due to the fact that in this measurement there will also be events from the combinatorial background.

The distribution is plotted together with the reconstructed events matched with generated events in Fig. 6.11, in order to understand the difference.

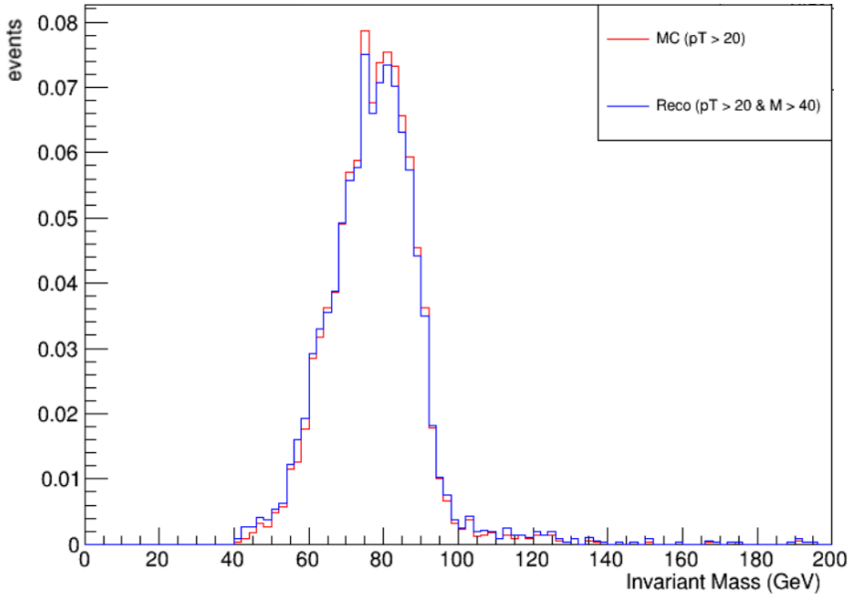

Figure 6.11: M_Z distribution comparison between reconstructed events obtained with matching with Monte Carlo events and reconstructed events with a cut on $p_T > 20GeV$ and $M(ee) > 40$ GeV.

The result after the two preliminary requirements is in very good agreement with the Monte Carlo truth.

6.3.2 Electron identification requirement

As the last step, a very similar BDT model to the one previously trained in the Chapter section 4.4.1 is applied, in order to study its performance in the identification of high-p_T electrons coming from Z boson production.

In this case, the reconstructed information includes only quantities measured by the calorimeter, and therefore the momentum of the particle track is not available. It was decided to train again the BDT without the input variable which depends on particle momentum which is e/true_p.

The model is trained with the same pion and electron samples with Minimum Bias generated in Chapter 4. The chosen input variables are the front cluster energy with respect to the total energy reconstructed (e_F/e_{tot}) and the energy of each front cluster cell with respect to the total energy of the front cluster e_{Fcell}/e_F. The other BDT parameters are left unvaried.

Figure 6.12 shows the ROC curve for this training sample. The AUC score is 0.9467, which implies worse discrimination power than the corresponding model trained with the e/true_p variable, see table 4.2. The pion identification efficiency at 80% electron efficiency is $\sim 8 \cdot 10^{-2}$.

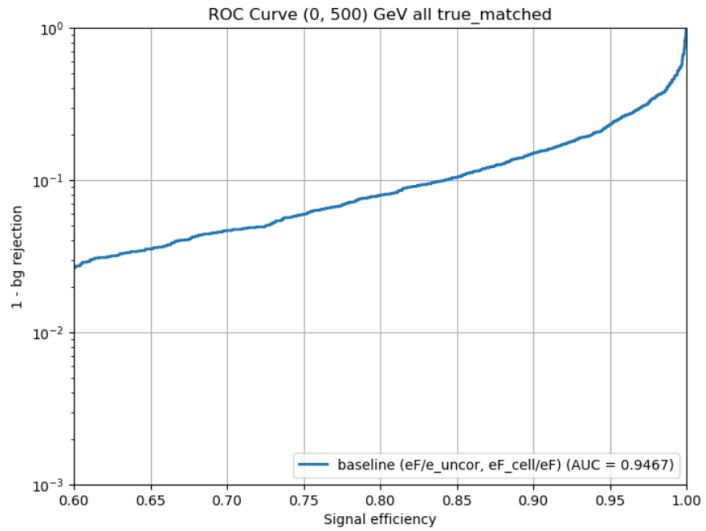

Figure 6.12: Receiver Operating Curve (ROC) for a BDT model with input variables e_F/e_{tot} and e_{Fcell}/e_F.

To test the BDT model performance on Z event reconstruction, it would be necessary to have also the simulation of the hadronic background of the process. Since this is not available, it was decided to compare the $Z \to e^+e^-$ event with the same pion sample used in Chapter 4.

The model is then applied to both the pion and the reconstructed electron sample from $Z \to e^+e^-$ with p_T and $M(ee)$ requirements. The model predictions are computed in both cases and their distribution is shown in Fig. 6.13. It has a long tail on the left.

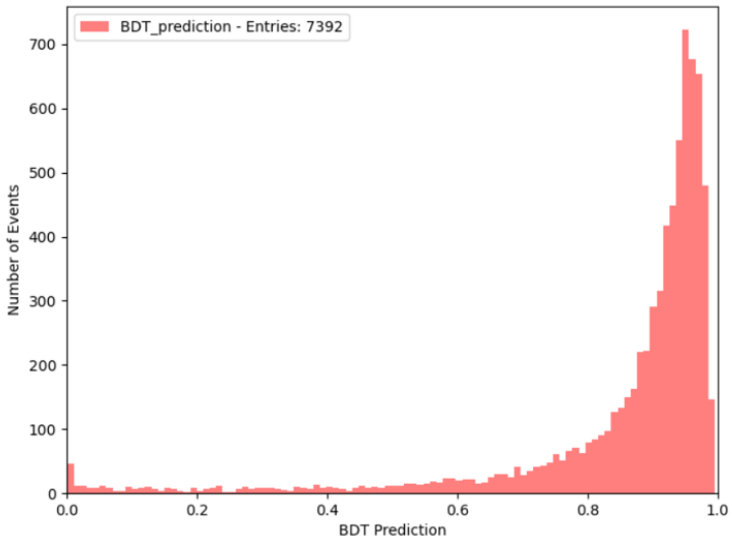

Figure 6.13: Distribution of BDT model predictions for reconstructed events with $p_T > 20$ GeV and $M(ee) > 40$ GeV.

It is possible to determine some working points, corresponding to different values of pions misidentification. Taking the pion misidentification rate at 10%, the cut on the BDT prediction is identified as 0.66. Figure 6.14 shows the two prediction distributions for pions and electrons on a logarithmic scale, where the BDT cut is identified with a red vertical line.

By adding a cut on BDT predictions > 0.66 on each electron, the dataset reduced to 3017 pairs which results in a reconstruction efficiency of 32%. The invariant mass resolution is 13.27 GeV and a mass

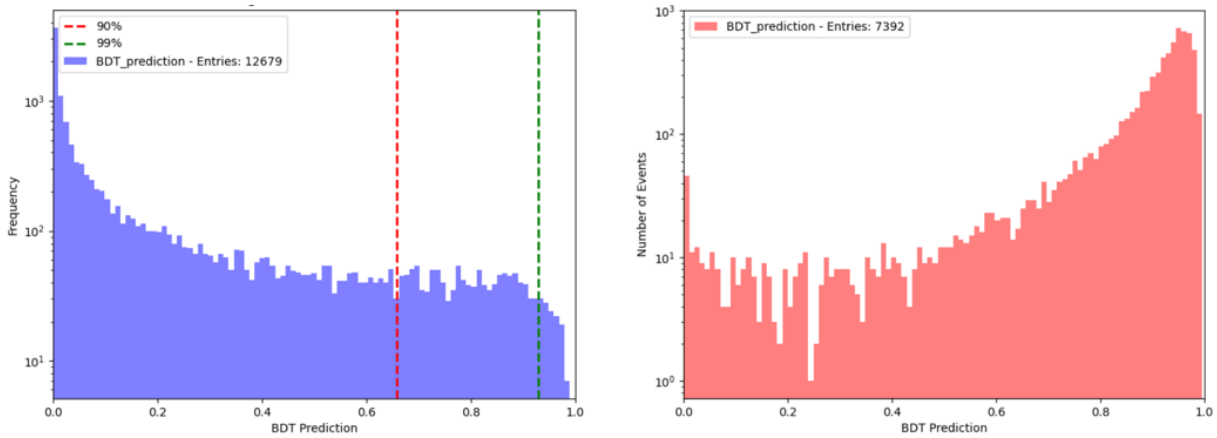

Figure 6.14: Comparison of BDT prediction distribution obtained after applying the BDT model to pions (left) and reconstructed electrons (right). The cut on the 90% and 99% pion efficiency are highlighted by the red and green dashed lines.

peak is around 77.7 GeV

A more stringent cut is applied at BDT predictions > 0.93, corresponding to accepting pion misidentification at 1%. With this, the dataset is reduced to 833 pairs, i.e. 9% efficiency, with an invariant mass resolution of 12.11 GeV around 78.56 GeV.

While this requirement does not have a great mass reconstruction efficiency, some important conclusions should be drawn from the comparison with the reconstructed mass obtained with matching with Monte Carlo truth. The resolution of the invariant mass peak is compatible to the one obtained with previous requirements.

This result is expected to improve once the tracking information becomes available, unlocking the possibility of using the E/p input variable.

An intermediate cut was chosen to 4% of pion misidentification, this led to having 2420 pairs, thus 26% efficiency.

Figure 6.15 shows the invariant mass distribution obtained with this requirement.

Table 6.1 summarises the reconstruction efficiency and invariant mass resolution for each cut.

Table 6.1: Signal efficiency and invariant mass resolution after each cut.

Cut	Candidates	π misID	Efficiency	M_{inv} (GeV)	Resolution (GeV)
$p_T > 20$ GeV $M(ee) > 40$ GeV	3696	-	39%	77.65	13.84
bdt > 0.806	833	4%	26%	77.97	13.23

In conclusion, the trained model proved effective in identifying electron candidates.

Further improvements of this study on event selection would be the addition of a simulated hadronic background and the addition of momentum information from the tracking system. This would allow the re-training of the BDT algorithm, exploiting also the e/p variable, and optimizing the requirement on BDT prediction tuning the pion misidentification rate.

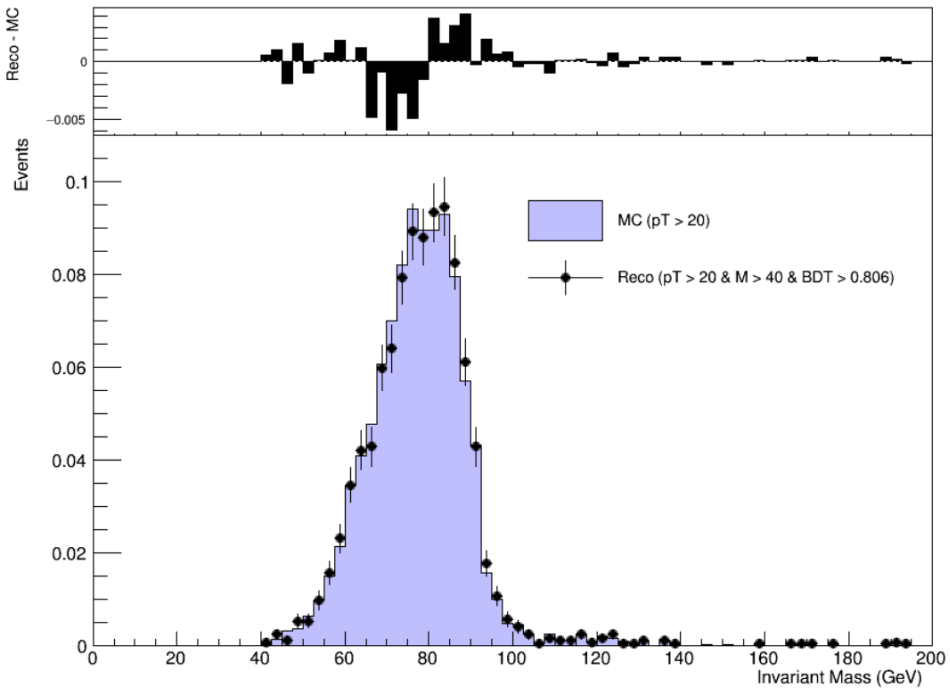

Figure 6.15: M_Z distribution comparison between reconstructed events obtained with matching with Monte Carlo events and reconstructed events with cut on $p_T > 20$ GeV, $M(ee) > 40$ GeV and BDT predictions > 0.806.

Chapter 7

Conclusions

In this thesis, the particle identification performance of the Electromagnetic Calorimeter of the LHCb experiment in the future Upgrade II configuration was evaluated both with simulation and test-beam data.

Pion and electron samples were generated within the new high pile-up environment and the energy deposited on the calorimeter was reconstructed. The energy over primary particle momentum, called (e/true_p), distribution for the two particles shows a significantly different behavior with respect to the one of Run 2, with pile-up $\sim$1, concluding that this discrimination variable is not sufficient to do a proper separation.

I developed a BDT algorithm to discriminate between electrons and pions, with input variables chosen to exploit the presence of longitudinal segmentation in the calorimeter modules. I demonstrated that adding just one input variable linked to the segmentation, such as the energy of the front cluster with respect to the total cluster energy, improves the pion identification efficiency of almost an order of magnitude from the efficiency obtained using only the E/p variable. This result was used also to affirm that the baseline configuration, i.e., the option where all ECAL modules have longitudinal segmentation, can offer significant improvements from the downscoped one, i.e. where only innermost modules are equipped with a double-side readout.

Moreover, the results of this model on the LHCb benchmark physics channel $B^0 \to K^{*0} e^+ e^-$ proved to reach the same particle identification efficiency as the one of Run 2.

I studied energy reconstruction and particle discrimination performance on a SpaCal prototype for Upgrade II with lead absorber and polystyrene fibers. The prototype was tested during the test-beam campaign that took place in June 2024 at CERN. The energy resolution is compatible with results obtained by simulating particle beam interaction with the detector exploiting the Hybrid simulation framework and provides a sampling term of 12.5 $\pm$ 0.1 % and a constant term of 1.12 $\pm$ 0.03 %, while the noise term is estimated to be 0.16 $\pm$ 0.01. This result is not optimal since the required performance is the current ECAL energy resolution, with a sampling term of 10%. However, there is still more space for improvements, for example, managing light guide inefficiencies. Also, test-beam data at lower energies are needed to obtain a more precise sampling term from the combined fit.

Time resolution was $\sim$36 ps at 100 GeV, and thus did not meet the performance requirement needed for the upgrade, which is 15 ps. The same consideration on improvements could be done as for the energy resolution. Finally, I evaluated the electron-pion discrimination performance of this prototype, concluding that the deposited energy distributions with respect to the beam energy of the two particles are very well separated and therefore do not necessitate further discrimination besides a cut on the E/p variable.

Furthermore, I studied the reconstruction of simulated $Z \to e^+ e^-$ events produced with high luminosity conditions. With selection requirements on transverse momentum and di-electron mass, I was able to reconstruct the Z invariant mass with an efficiency of 39% and a resolution of 13.84 GeV.

Finally, I applied the algorithm developed in the first part of this thesis to further select Z candidates. This study aimed to demonstrate the practical utility of the algorithm in a realistic physics analysis scenario.

Since a hadronic background sample was not available for this study, I estimated its impact by using the single pion simulated sample with all the necessary considerations made in Chapter 6. The BDT model was applied to electron candidates for the Z event. Several cuts on pion misidentification were tested, choosing 4%, which resulted in a 26% Z mass reconstruction efficiency and 13.23 GeV resolution.

Under these selections, I obtained an improvement on the Z mass resolution within $\sim$ 15-17% from the current measurement of LHCb.

Bibliography

[1] Global PID performance for charged particles. 2020.

[2] Framework TDR for the LHCb Upgrade II: Opportunities in flavour physics, and beyond, in the HL-LHC era. Technical report, CERN, Geneva, 2021.

[3] Charged PID performance in early 2024 data. 2024.

[4] R. et al. Aaij. Measurement of the forward-backward asymmetry in $z/\gamma^* \to \mu^+\mu^-$ decays and determination of the effective weak mixing angle. *Journal of High Energy Physics*, 2015(11), November 2015.

[5] R. et al. Aaij. Measurement of the forward z boson production cross-section in pp collisions at $\sqrt{s} = 13$ tev. *Journal of High Energy Physics*, 2016(9), September 2016.

[6] R. et al. Aaij. Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era. Technical report, CERN, Geneva, 2016. ISBN 978-92-9083-494-6.

[7] Roel et al. Aaij. Measurement of forward $Z \to e^+e^-$ production at $\sqrt{s} = 8$ TeV. *JHEP*, 05:109, 2015.

[8] Carlos et al. Abellán Beteta. Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC. Technical report, 2020.

[9] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[10] S Barsuk. The Shashlik Electro-Magnetic Calorimeter for the LHCb Experiment. Technical report, CERN, Geneva, 2004. Published in *Perugia 2004, Calorimetry in particle physics* 61-67.

[11] Tom Boettcher, Philip Ilten, and Mike Williams. Direct probe of the intrinsic charm content of the proton. *Physical Review D*, 93(7), April 2016.

[12] Maximilien Brice. The LHCb electromagnetic calorimeter. Le mur du calorimètre électromagnétique de LHCb a été entièrement monté en un petit mois seulement ! 2005.

[13] Rene Brun and Fons Rademakers, editors. *ROOT - An Object Oriented Data Analysis Framework*, volume Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) of *81-86*, 1996. See also ROOT [software], Release v6.YY/ZZ, dd/mm/yyyy.

[14] CDF Collaboration and T. Aaltonen et al. High-precision measurement of the w boson mass with the cdf ii detector. *Science*, 376(6589):170–176, 2022.

[15] LHCb collaboration and R. Aaij et al. The lhcb upgrade i, 2023.

[16] John C. Collins and Davison E. Soper. Angular Distribution of Dileptons in High-Energy Hadron Collisions. *Phys. Rev. D*, 16:2219, 1977.

[17] O et al. Deschamps. Photon and neutral pion reconstruction. Technical report, CERN, Geneva, 2003.

[18] J. Allison et al. Recent developments in geant4. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835:186–225, 2016.

[19] Johannes Haller et al. Status of the global electroweak fit with gfitter in the light of new precision measurements, 2022.

[20] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[21] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008.

[22] Stephen Farry. Electroweak Physics at LHCb. *Nucl. Part. Phys. Proc.*, 273-275:2181–2186, 2016.

[23] Vagelis Gkougkousis. An LHCb Vertex Locator (VELO) for 2030s. 2022.

[24] Hamamatsu Photonics. *Photomultiplier tube R9880U series*, June 2010.

[25] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[26] Ross Hunter. Measurement of the W boson mass at LHCb. In *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2021)*, volume 398, page 456, 2022.

[27] D Jacquet and F Follin. Implementation and experience with luminosity levelling with offset beam, 2014.

[28] Gordon L. Kane. *MODERN ELEMENTARY PARTICLE PHYSICS*. Cambridge University Press, 2 2017.

[29] Rolf Lindner. LHCb Particle Identification Enhancement Technical Design Report. Technical report, CERN, Geneva, 2023.

[30] Fabian Pedregosa et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[31] Alexander et al. Radovic. Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560(7716):41–48, 2018.

[32] Byron P. et al. Roe. Boosted decision trees, an alternative to artificial neural networks. *Nucl. Instrum. Meth. A*, 543(2-3):577–584, 2005.

[33] L Rossi and O Brüning. Introduction to the HL-LHC Project. *Adv. Ser. Dir. High Energy Phys.*, 24:1–17, 2015.

[34] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.

[35] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.

[36] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852–867, June 2008.

[37] Huayang Song, Xia Wan, and Jiang-Hao Yu. Custodial symmetry violation in scalar extensions of the standard model, 2022.

[38] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[39] Konrad Zuchniak. Multi-teacher knowledge distillation as an effective method for compressing ensembles of neural networks, 02 2023.